

# Investigating the assembly and function of the bacterial injectisome by hybrid structural methods

by

Dorothy Diana Majewski

B.Sc. McGill University, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2020

© Dorothy Diana Majewski, 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Investigating the assembly and function of the bacterial injectisome by hybrid structural methods

submitted by Dorothy Diana Majewski in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Biochemistry and Molecular Biology

**Examining Committee:**

Dr. Natalie C. J. Strynadka, Biochemistry and Molecular Biology

Supervisor

Dr. Brett B. Finlay, Microbiology and Immunology

Supervisory Committee Member

Dr. Rachel C. Fernandez, Microbiology and Immunology

University Examiner

Dr. Leonard J. Foster, Biochemistry and Molecular Biology

University Examiner

**Additional Supervisory Committee Members:**

Dr. Calvin K. Yip, Biochemistry and Molecular Biology

Supervisory Committee Member

## Abstract

The bacterial injectisome is an essential virulence factor for many Gram-negative pathogens. Resembling a 50-100 nm long syringe, the injectisome creates a continuous channel between the bacterial and host cytosols through which the bacterium secretes effector proteins to modulate host signalling. The aim of the following work is to contribute to the structural characterization of the injectisome, focusing on proteins involved in its assembly and substrate selection.

The outer membrane pore of the injectisome, termed the secretin, relies on a pilotin protein for its localization. The X-ray crystallographic structure of the *Salmonella enterica* SPI-1 pilotin InvH reveals an  $\alpha$ -helical dimer, confirmed to exist in solution through biophysical experiments. The pilotin-secretin interface, characterized by X-ray crystallography and NMR spectroscopy, is mutually exclusive with the InvH dimer and results in a 1:1 complex.

The inner membrane pore protein has a vital role in injectisomal secretion hierarchy. The cryo-EM structure of the IM pore EscV from enteropathogenic *Escherichia coli* (EPEC) demonstrates that the protein forms a nonameric ring in solution. Of its four subdomains, two contribute to ring formation while the remaining two have some rotational freedom.

The cytosolic ATPase complex is essential to separating effector proteins from their cognate chaperones prior to secretion. The cryo-EM analysis of the EPEC ATPase-stalk complex, EscN-EscO, yielded an asymmetric homohexameric structure resembling F<sub>1</sub>- and V-ATPases. Kinetic studies show that oligomerization and the presence of the stalk are required for efficient ATP hydrolysis. The similarity to rotary ATPases supports the hypothesis that the injectisomal ATPase complex is a rotary motor, with the stalk acting as the rotor. How this rotation contributes to secretion is yet to be discovered.

## Lay Summary

Antibiotic resistance is a growing health crisis. As bacteria such as *Salmonella* and pathogenic *E. coli* develop resistance to multiple drugs, the infection mortality rate increases. One method to lower the chances of bacteria developing resistance is to use therapeutics that target their host interaction, rather than killing the pathogen.

The bacterial injectisome is a promising target for such therapeutics. It functions as a molecular syringe, through which the pathogen injects harmful proteins into the host cell. To create a foundation for targeted drug design, we sought to characterize the injectisome structure at an atomic level. Through the structures of several injectisome components explored in this thesis, we now have a better understanding of how the system is assembled at a molecular level and how it functions. With this information, we are better armed to create drugs against the injectisome, and the harmful pathogens that encode it, in the future.

## Preface

Figure 1.3 from Chapter 1 was published in a review article (Majewski DD, Worrall LJ, Strynadka NCJ. 2018. Secretins revealed: structural insights into the giant gated outer membrane portals of bacteria. *Curr Opin Struct Biol.* **51**:61–72). I prepared the text and figures from Chapter 1, with revisions by my supervisor Dr. Natalie Strynadka.

Chapter 2 is a published article (Majewski DD, Okon M, Heinkel F, Robb CS, Vuckovic M, McIntosh LP, Strynadka NCJ. 2020. Characterization of the pilotin-secretin complex from the *Salmonella enterica* type III secretion system using hybrid structural methods. *Structure*. doi: 10.1016/j.str.2020.08.006). I performed the protein expression and purification, crystallization, NMR sample preparation, and biophysical assays. The NMR spectra were collected by Mark Okon and myself, and I carried out data analyses and structure determination with help from Florian Heinkel and Lawrence McIntosh. I collected and solved X-ray crystallographic data with help from Craig Robb. Marija Vuckovic cloned the constructs and I performed subsequent mutagenesis. I wrote the manuscript and made the figures, with revisions by Lawrence McIntosh and Natalie Strynadka.

Chapter 3 is a submitted manuscript. I performed the protein expression and purification, cryo-EM reconstruction, and model building. Cryo-EM data collection was performed by Claire Atkinson. Bronwyn Lyons ran the molecular dynamics simulations, and I performed the data analysis. I wrote the manuscript and made the figures, with revisions by Natalie Strynadka.

Chapter 4 is a published article (Majewski DD, Worrall LJ, Hong C, Atkinson CE, Vuckovic M, Watanabe N, Yu Z, Strynadka NCJ. 2019. Cryo-EM structure of the homo-hexameric T3SS ATPase-central stalk complex reveals rotary ATPase-like asymmetry. *Nat Commun.* **10**:626). I performed protein expression, purification, and kinetic assays. I cloned the constructs with help

from Marija Vuckovic. Nobuhiko Watanabe and Liam Worrall performed early purification optimization and tests of ATP transition state analogues. Cryo-EM data collection and initial processing was performed by Chuan Hong and Zhiheng Yu, and reprocessing was done by Claire Atkinson and Liam Worrall. I built and refined the model. I wrote the manuscript and made the figures, with revisions by Liam Worrall and Natalie Strynadka.

Chapter 5 was written by myself and revised by Natalie Strynadka.

# Table of Contents

<b>Abstract</b> .....	<b>iii</b>
<b>Lay Summary</b> .....	<b>iv</b>
<b>Preface</b> .....	<b>v</b>
<b>Table of Contents</b> .....	<b>vii</b>
<b>List of Tables</b> .....	<b>x</b>
<b>List of Figures</b> .....	<b>xi</b>
<b>Acknowledgements</b> .....	<b>xv</b>
<b>Dedication</b> .....	<b>xvi</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Secretion in Gram-negative bacteria .....	1
1.2 Type III Secretion System Pathogenesis .....	2
1.2.1 The role of the injectisome .....	2
1.2.2 The function of two different injectisomes in <i>Salmonella enterica</i> .....	4
1.2.3 Injectisome-mediated pathogenesis in EPEC and EHEC .....	6
1.2.4 The injectisome as a drug target .....	8
1.3 Structural characterization of the T3SS.....	9
1.3.1 The basal body .....	11
1.3.2 The cytosolic sorting platform .....	17
1.3.3 The needle, tip, and translocon .....	19
1.4 Role of pilotins in injectisome assembly .....	21
1.5 Secretion hierarchy .....	23
1.5.1 Regulation of needle length .....	23
1.5.2 Switch to late substrates.....	25
1.6 Energetics of secretion .....	26
1.7 Thesis objectives .....	28
<b>2 Characterization of the pilotin-secretin complex from the <i>Salmonella enterica</i> type III secretion system using hybrid structural methods</b> .....	<b>31</b>
2.1 Introduction .....	31
2.2 Methods .....	34
2.2.1 Expression and Purification of InvH and InvG.....	34
2.2.2 Crystallization of InvH <sup>70-147</sup> .....	36

2.2.3 Co-Crystallization of InvH <sup>84-147</sup> and InvG <sup>543-558</sup> .....	37
2.2.4 Nuclear Magnetic Resonance Spectroscopy .....	38
2.2.5 Size-exclusion coupled Multi-angle Light Scattering.....	39
2.2.6 Small Angle X-Ray Scattering .....	40
2.2.7 Isothermal Titration Calorimetry .....	41
<b>2.3 Results .....</b>	<b>41</b>
2.3.1 Crystallized InvH forms a homodimer .....	41
2.3.2 InvH forms the crystallographically-observed homodimer in solution .....	45
2.3.3 NMR spectroscopic studies on the InvH-InvG complex reveal ordered and dynamic regions .....	49
2.3.4 Structural studies of the InvH-InvG complex by NMR spectroscopy and X-ray crystallography....	52
2.3.5 Aromatic and small hydrophobic residues contribute to InvH-InvG binding.....	57
<b>2.4 Discussion .....</b>	<b>58</b>
<b>3 Cryo-EM structure of the EscV cytosolic domain nonamer .....</b>	<b>67</b>
3.1 Introduction .....	67
3.2 Methods .....	69
3.2.1 Purification of EscV <sub>C</sub> .....	69
3.2.2 Preparation of negative-stain TEM grids.....	70
3.2.3 Cryo-EM data collection.....	70
3.2.4 Cryo-EM Map reconstruction and model-building.....	71
3.3 Results .....	72
3.3.1 EscV <sub>C</sub> forms a nonameric ring in solution.....	72
3.3.2 EscV <sub>C</sub> shares similarities with injectisomal and flagellar inner membrane pores .....	77
3.4 Discussion .....	80
<b>4 Cryo-EM structure of the homohexameric T3SS ATPase-central stalk complex reveals rotary ATPase-like asymmetry .....</b>	<b>83</b>
4.1 Introduction .....	83
4.2 Methods .....	86
4.2.1 Expression and purification of EscN <sup>29-446</sup> and EscO <sup>1-125</sup> .....	86
4.2.2 Glycerol gradient centrifugation of EscN-EscO complex .....	87
4.2.3 Cryo-EM reconstruction of the EscN-EscO complex .....	88
4.2.4 Model-building into cryo-EM maps .....	89
4.2.5 EnzChek ATPase activity assays .....	90
4.3 Results .....	91
4.3.1 Cryo-EM analysis of the EscN-EscO complex.....	91

4.3.2 EscN forms an asymmetric oligomer conserved with F <sub>1</sub> /V <sub>1</sub> -ATPases .....	94
4.3.3 The EscO coiled coil interacts at the EscN C-terminal domain.....	97
4.3.4 Active site architecture and implications for catalysis.....	101
4.4 Discussion .....	106
<b>5 Conclusion .....</b>	<b>114</b>
5.1 Summary .....	114
5.2 The pilotin InvH .....	115
5.3 The inner membrane pore EscV .....	117
5.4 The ATPase complex .....	118
5.5 Closing Remarks .....	120
<b>References .....</b>	<b>122</b>
<b>Appendices.....</b>	<b>145</b>
Appendix A: Chapter 2 Supplementary Information .....	145
Appendix B: Chapter 3 Supplementary Information .....	155
Appendix C: Chapter 4 Supplementary Information .....	158

## List of Tables

Table 2.1: InvH X-ray crystallography data collection, refinement and validation statistics.....	44
Table 2.2: InvH-InvG NMR spectroscopy data collection and refinement statistics .....	53
Table 3.1: EscV <sub>C</sub> cryo-EM data collection, refinement and validation statistics .....	74
Table 4.1: EscN-EscO cryo-EM data collection, refinement and validation statistics.....	93
Table A.1: SAXS radius of gyration and molecular weight estimates for InvH .....	154
Table A.2: ITC-calculated binding affinities, enthalpies and entropies for InvH-InvG binding mutants .....	154

## List of Figures

Figure 1.1: Examples of Gram-negative secretion systems.....	2
Figure 1.2: Injectisome-mediated pathogenesis in <i>S. enterica</i> and <i>E. coli</i> . ....	5
Figure 1.3: Structural characterization of the injectisome. ....	10
Figure 1.4: Structures of injectisomal (T3SS) and homologous T2SS secretins.....	13
Figure 1.5: Schematic of the export apparatus and sorting platform proteins:.....	15
Figure 2.1: Crystal structure of the InvH <sup>70-147</sup> homodimer. ....	43
Figure 2.2: Evidence for InvH homodimer formation in solution. ....	48
Figure 2.3: NMR spectroscopic data of InvH <sup>70-147</sup> and InvG <sup>520-562</sup> . ....	50
Figure 2.4: Structural characterization of the InvH-InvG binding interface. ....	55
Figure 2.5: Structural comparison of InvH with other pilotins and secretin interfaces.....	61
Figure 2.6: Potential secretin assembly mechanisms.....	63
Figure 3.1: Map and model of EscV <sub>C</sub> . ....	73
Figure 3.2: Atomic details of the EscV <sub>C</sub> model.....	76
Figure 3.3: Comparison of EscV <sub>C</sub> with other characterized homologues. ....	78
Figure 4.1: Cryo-EM density and resolution. ....	94
Figure 4.2: EscN structure. ....	95
Figure 4.3: Overview of EscN-EscO complex. ....	99
Figure 4.4: Characterized EscN/EscO mutants.....	101
Figure 4.5: EscN catalysis.....	106
Figure 4.6: Function of EscN-EscO in the injectisome. ....	112
Figure A.1: Details of the InvH <sup>70-147</sup> crystal structure. ....	146
Figure A.2: Overview of SAXS data. ....	147
Figure A.3: SEC-MALS data on InvH mutants and the InvH-InvG complex.....	149
Figure A.4: NMR spectra of the InvH-InvG complex.....	150
Figure A.5: Details of the InvH-InvG binding interface.....	152
Figure A.6: Comparison of the InvH-InvG heterodimer with other structurally characterized pilotins and secretins. ....	153
Figure B.1: Cryo-EM data collection and refinement of EscV <sub>C</sub> .....	155
Figure B.2: EscV <sub>C</sub> surface and conformation. ....	156

Figure B.3: Multiple sequence alignment. ....	157
Figure C.1: EscN-EscO complex purification. ....	158
Figure C.2: Kinetics and glycerol gradient analysis. ....	159
Figure C.3: Cryo-EM imaging and reconstruction of the EscN/EscO complex. ....	160
Figure C.4: EscO structural analysis. ....	161
Figure C.5: Asymmetry of the EscN homohexamer. ....	162
Figure C.6: Differences in conformation between EscN chains. ....	163
Figure C.7: EscN homohexamer formation and comparison with F- and V-ATPases. ....	164
Figure C.8: EscO interaction interface with EscN. ....	165
Figure C.9: Sequence alignment of EscN homologues. ....	166
Figure C.10: Differences in magnesium ion coordination between the four ligand-bound active sites. ....	167

## List of Abbreviations

ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BAM	$\beta$ -barrel assembly machinery
BMRB	Biological magnetic resonance bank
CCCP	Carbonyl cyanide m-chlorophenyl hydrazone
CCD	Charge-coupled device
CTF	Contrast transfer function
CV	Column volume
DDM	n-Dodecyl $\beta$ -D-maltoside
DM	n-Decyl $\beta$ -D-maltoside
EDTA	Ethylenediaminetetraacetic acid
EHEC	Enterohemorrhagic <i>Escherichia Coli</i>
EM	Electron microscopy
EMDB	Electron microscopy data bank
EPEC	Enteropathogenic <i>Escherichia Coli</i>
ET	Electron tomography
ETEC	Enterotoxigenic <i>Escherichia Coli</i>
FSC	Fourier shell correlation
GFP	Green fluorescent protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	Heteronuclear single quantum coherence
ICP-MS	Inductively coupled plasma mass spectrometry
IM	Inner membrane
ITC	Isothermal titration calorimetry
LEE	Locus of enterocyte effacement
LoL	Localization of lipoproteins
MD	Molecular dynamics
MOPS	3-(N-morpholino)propanesulfonic acid
MWCO	Molecular weight cutoff
NMR	nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NTA	Nitrilotriacetic acid
OM	Outer membrane
PDB	Protein data bank
PEG	Polyethylene glycol
PMF	Proton motive force
RBM	Ring-building motif
RCI	Random coil index

RMSD	Root-mean-square deviation
SAD	Single-wavelength anomalous dispersion
SASBDB	Small angle scattering biological data bank
SAXS	Small-angle X-ray scattering
SCV	<i>Salmonella</i> containing vacuole
SD	Subdomain
SDS	Sodium dodecyl sulfate
SEC	Secretory
SEC-MALS	Size-exclusion coupled multi-angle light scattering
SPI	<i>Salmonella</i> pathogenicity island
T3SS	Type III secretion system
TAT	Twin-arginine translocase
TCEP	Tris(2-carboxyethyl)phosphine
TEM	Transmission electron microscopy
TLS	Translation, libration, and screw-rotation
TM	Transmembrane
TPR	Tetratricopeptide repeat
UV	Ultraviolet

## Acknowledgements

I would like to thank my supervisor, Dr. Natalie Strynadka, for giving me the opportunity to pursue my PhD in her lab. Under her guidance I have learned a large repertoire of structural biology techniques, attended courses to enrich my knowledge, and had the opportunity to write and publish several manuscripts.

I also thank my committee members, Dr. Brett Finlay and Dr. Calvin Yip. Their advice helped me to focus my experiments towards completing my degree, and their support throughout was a great encouragement.

I would like to thank our collaborator Dr. Lawrence McIntosh. With his help I learned a great deal about nuclear magnetic resonance, and his detailed feedback on our manuscript improved it to a state I was proud to submit. I also thank Dr. McIntosh's lab members, Mark Okon and Florian Heinkel, who took a lot of time to teach me about NMR data collection and processing.

I am thankful to several of our lab members for helping me get through this degree. Marija Vuckovic has done an incredible job of running the lab, and still managed to find the time to teach me cloning. Fred Rosell has always been a bedrock of biophysical knowledge, teaching me techniques and helping with instruments that would have otherwise been a nightmare. Natalie Zeytuni has always been very supportive and provided helpful advice with protein purification and crystallization. Craig Robb taught me most of what I know about solving crystal structures, and spent many afternoons helping out with tricky datasets. I'd also like to thank everyone in the lab who I haven't mentioned here; I have received guidance from all when experiments failed.

I would like to thank NSERC for funding throughout my degree. I would also like to thank the Zbarsky family, who contributed the S. H. Zbarsky scholarship in my third year.

## Dedication

To my parents:

You built up a love of science that got me this far in the first place.

# 1 Introduction

## 1.1 Secretion in Gram-negative bacteria

The bacterial cell envelope provides a substantial survival benefit in harsh extracellular environments. Gram-negative bacteria are equipped with two membranes, the inner and outer membrane (IM and OM), with a peptidoglycan cell wall in the periplasmic space between them<sup>1</sup>. The cell wall, made up of a mesh of cross-linked saccharide and peptide chains, gives the bacterium its shape and prevents turgor pressure-induced rupture of the bacterium in non-isotonic solutions. The OM provides an additional permeability barrier, protecting the cell wall from antibiotics and lysozyme<sup>2</sup>.

Despite the benefits of the bacterial envelope, it also presents a significant obstacle for transport of molecules. Bacteria rely on substrate secretion and import for several functions including adhesion, niche establishment, and pathogenicity<sup>3,4</sup>. Further complicating the issue, diverse types of molecules require cross-membrane transport including peptides, nucleic acids, lipids, and small molecules. To overcome the barrier posed by the envelope, bacteria have evolved a series of intricate secretion systems that allow macromolecular transport across membranes and peptidoglycan.

Several specialized secretion systems are unique to Gram-negative bacteria (Figure 1.1). Systems such as the type II secretion system (T2SS)<sup>5</sup> and type V secretion system (T5SS)<sup>6</sup> transport substrates across the OM and rely on an inner membrane translocase, such as the secretory (SEC) pathway<sup>7</sup> or twin-arginine translocase (TAT)<sup>8</sup>, to secrete cytoplasmic molecules in a two-step mechanism. Other systems cross the IM, peptidoglycan, OM, and in some cases a host membrane; these include the type III secretion system (T3SS)/injectisome<sup>9</sup>, the type IV

secretion system (T4SS)<sup>10</sup>, and the type VI secretion system (T6SS)<sup>11</sup>. As these systems are important for bacterial survival and virulence, they have been studied extensively through microbiology and structural biology experiments to provide a foundation for targeted therapeutic development.

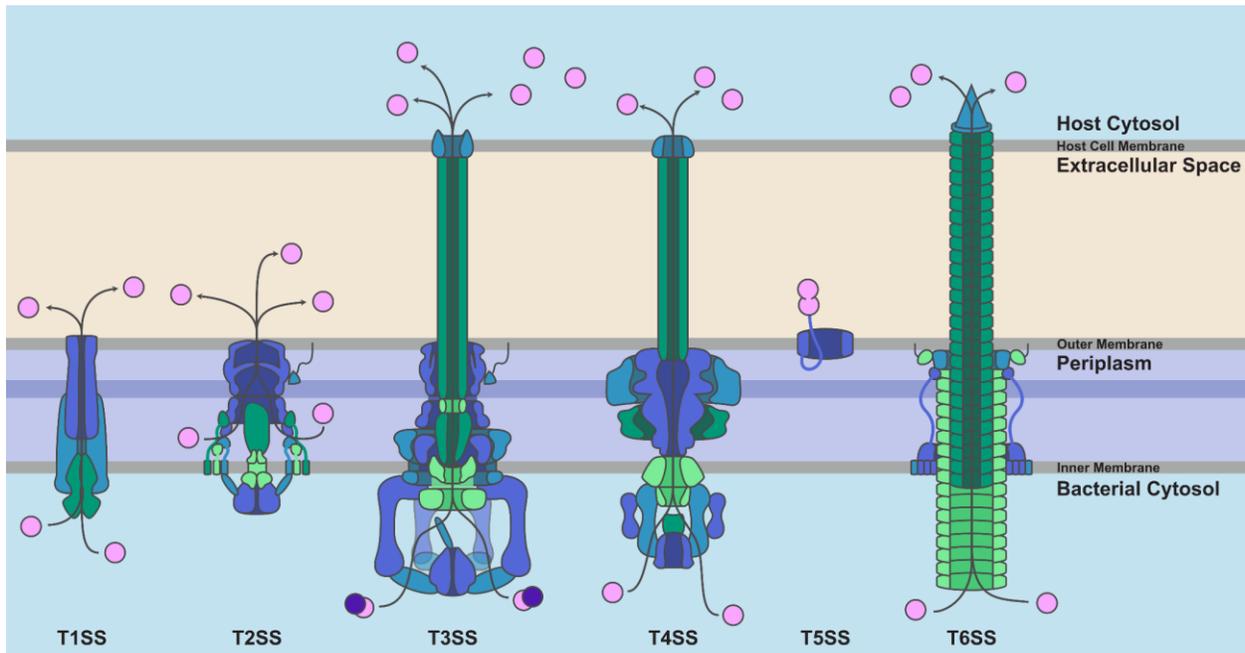


Figure 1.1: Examples of Gram-negative secretion systems. The T1SS, Injectisome, T4SS, and T6SS have a one step pathway where the substrate is secreted across multiple membranes. The T2SS and T5SS instead use a two-step pathway, relying on the SEC or TAT systems to transport the substrate across the inner membrane.

## 1.2 Type III Secretion System Pathogenesis

### 1.2.1 The role of the injectisome

Two Gram-negative bacterial nanomachines, the injectisome and the flagellum, represent the two types of type III secretion systems<sup>9,12</sup>. Both systems are made up of over 20 different proteins in varying oligomeric states, forming a complex that passes through the IM, OM, and peptidoglycan. Many of the cytoplasmic and inner membrane proteins of flagella have T3S homologues in the injectisome, which serve to secrete the components of the helical hook and

filament assemblies<sup>13</sup>. Once assembled, the flagellar T3S machinery ceases to secrete proteins; the flagellum takes on its role of providing motility through the rotation of the filament.

The injectisome, in contrast to the flagellum, relies on type III secretion for its function of increasing bacterial pathogenicity<sup>9</sup>. Resembling a syringe and needle, the injectisome forms a continuous channel from the bacterial to host cytoplasm through which proteins can pass in a partially unfolded state. Through this channel, bacteria secrete effector proteins that can alter host cell signalling to the pathogen's advantage. The injectisome is an important virulence machine in many clinically relevant pathogens, including enteropathogenic and enterohemorrhagic *Escherichia coli* (EPEC and EHEC respectively), *Salmonella enterica*, *Shigella flexneri*, *Pseudomonas aeruginosa*, *Chlamydia trachomatis*, and *Yersinia pestis*. As such, several injectisomal proteins have shown promise as targets for vaccine development and anti-virulence therapeutics<sup>14</sup>.

The function of the injectisome varies greatly depending on the life cycle, host, environment, and pathogenicity of the bacterium. It is encoded on pathogenicity islands that have been shared between species through horizontal gene transfer<sup>15</sup>. The genes encoded on these islands include structural proteins and some effectors; other effectors are encoded on separate loci. Effector proteins rely on cognate chaperones to protect them from degradation and to target them to the cytosolic machinery of the injectisome<sup>16,17</sup>. The chaperone is dissociated from the effector prior to secretion, and the effector is partially unfolded to permit passage through the narrow injectisome channel. Effectors can target a wide range of host cell machinery depending on the bacterial species.

### 1.2.2 The function of two different injectisomes in *Salmonella enterica*

Infection by *S. enterica* broadly causes two different pathologies, depending on whether the serovar is typhoidal or non-typhoidal. Typhoidal *S. enterica* serovars, serovar Typhi and Paratyphi A, B, and C, are the causative agents of typhoid fever and paratyphoid fever respectively. Both typhoid and paratyphoid fever are life-threatening systemic infections, causing an estimated 11.9 to 26.9 million cases and 128,000 to 216,500 deaths annually<sup>18,19</sup>. Other *S. enterica* serovars are non-typhoidal and instead cause Salmonellosis, an infection of the gut leading to diarrhea and fever. Salmonellosis is generally a self-limiting infection with low mortality, causing 95.1 million cases and 50,771 deaths in 2017<sup>20</sup>. However, non-typhoidal *Salmonella* can occasionally invade other tissues, resulting in bacteraemia, meningitis, or other infections; these invasive cases have a much higher mortality rate, and resulted in 535,000 infections and 77,500 deaths worldwide in 2017<sup>21</sup>. Antibiotic resistance is growing increasingly common in both types of *Salmonella*, leading to increased mortality rates<sup>18,22,23</sup>.

*S. enterica* encodes two injectisomes on *Salmonella* pathogenicity islands 1 and 2 (SPI-1 and SPI-2 respectively)<sup>24</sup>. Both typhoidal and non-typhoidal *S. enterica* are invasive pathogens; their SPI-1 injectisome is involved in initial engulfment by the host cell, while the SPI-2 system contributes to survival within the host cell (Figure 1.2). After ingestion by a host, *S. enterica* adheres to the epithelial cells in the gut and secretes effectors through the SPI-1 system. These secreted proteins promote actin cytoskeleton arrangement in the host, creating membrane ruffles that surround the bacterium and eventually lead to phagocytosis<sup>24</sup>. Multiple host proteins are targeted to create membrane ruffles. The effector SopB interacts with the small GTPase Rho and annexin A2 to promote actin rearrangement and accumulation<sup>25</sup>. The small GTPases Rac1 and Cdc42 are targeted by effectors SopE and SopE2, further stimulating actin nucleation factors to

promote membrane ruffling and engulfment<sup>26</sup>. Effectors SipA and SipC are secreted to promote actin bundling. Both non-typhoidal and typhoidal *S. enterica* use these SPI-1 effectors for host cell invasion with the exception of SopE2, which is a pseudogene in serovar Typhi<sup>27</sup>.

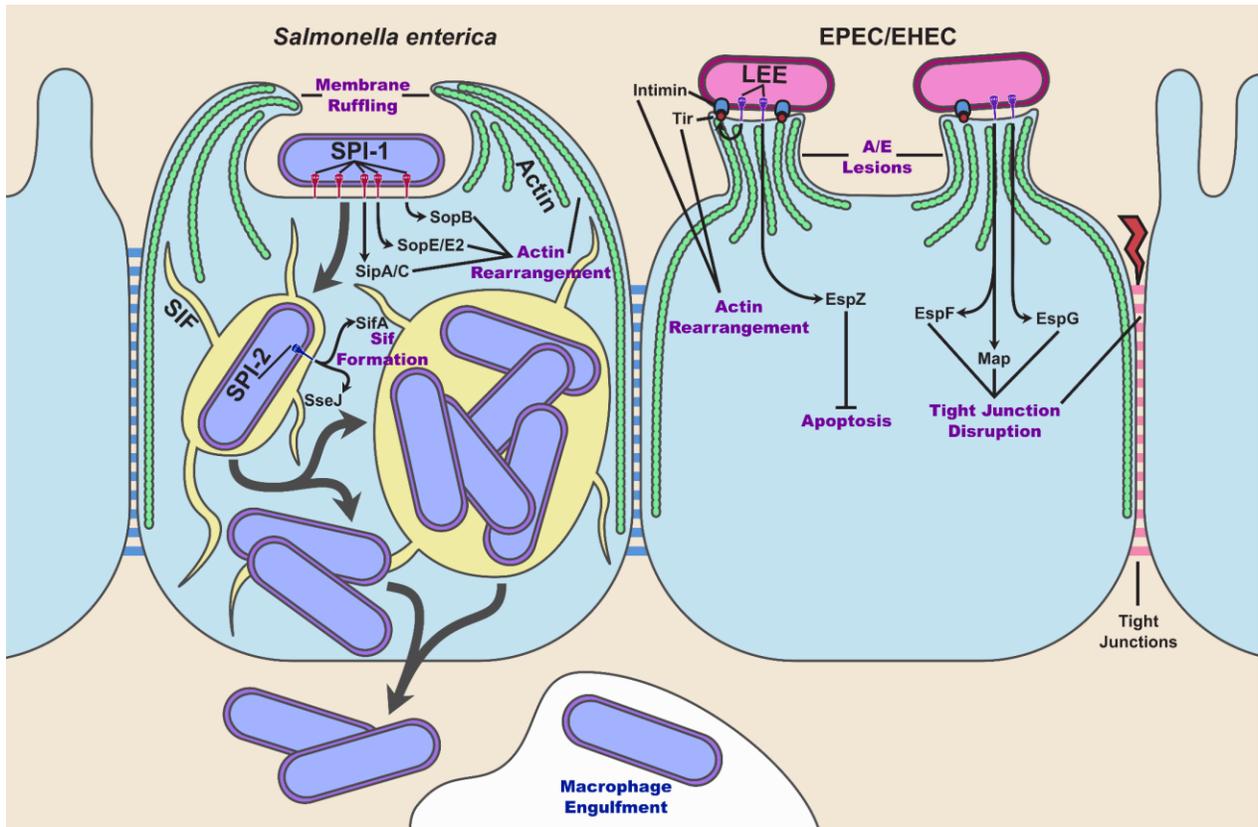


Figure 1.2: Injectisome-mediated pathogenesis in *S. enterica* and *E. coli*. *S. enterica* (left) uses the SPI-1 injectisome to promote actin cytoskeleton rearrangement, resulting in phagocytosis by the host cell. Once inside, it uses the SPI-2 injectisome to proliferate inside the SCV. EPEC and EHEC (right) use the LEE-encoded injectisome to form attaching and effacing lesions, adhering to raised pedestals on the host cell.

Once *S. enterica* has been phagocytosed by the epithelial cell, it begins to form a specialized compartment called the Salmonella-containing vacuole (SCV)<sup>24,28</sup>. The SCV has some characteristics of late endosomes, such as an acidic pH of ~5 and the presence of endosomal organelle markers. One key step in SCV maturation is the formation of *Salmonella*-induced filaments (Sifs), filaments containing lysosomal membrane glycoproteins (lgps) that are attached

to the SCV<sup>29</sup>. Sif formation is mediated by SPI-2 effector proteins SifA and SseJ, which work to induce formation of the filaments and to modify the lipid content respectively<sup>24</sup>. Interestingly, the role of Sifs in *S. enterica* survival is not certain, but they are hypothesized to be involved in nutrient acquisition<sup>28</sup>.

In Salmonellosis, the invasion is limited to the gut epithelial cells, with various SPI-2 effectors maintaining the SCV and prolonging host infection<sup>24</sup>. Non-typhoidal *S. enterica* that breach the epithelial cells do not cause systemic infection in the majority of cases<sup>27</sup>. However, in enteric fever, serovars Typhi and Paratyphi A, B, and C are adapted to survive within macrophages and spread to other tissues<sup>27</sup>. As these serovars are host-restricted to humans, the details of their infection are less understood and often based on the mouse model of enteric fever, caused by serovar Typhimurium<sup>30</sup>. It has been shown that the serovar Typhi SPI-1 injectisome is required for invasion of non-phagocytic host cells<sup>31</sup>; however, the bacterium's survival within macrophages was uncompromised by disruption of the SPI-2 injectisome<sup>32</sup>. It is not fully understood what effectors are essential in enteric fever, as many effectors are pseudogenes or absent from typhoidal serovars (see Johnson *et al* for a summary of effector functions in serovars Typhi and Paratyphi<sup>30</sup>).

### 1.2.3 Injectisome-mediated pathogenesis in EPEC and EHEC

Diarrhoeal disease causes an estimated 5.8 million deaths per year in children under the age of 5 worldwide<sup>33</sup>. EPEC and EHEC are food-borne pathogens that cause diarrhoea; both encode an injectisome on the locus of enterocyte effacement (LEE), and EHEC additionally encodes shiga toxin<sup>34,35</sup>. EPEC primarily colonizes the duodenum, while EHEC mainly targets Peyer's patches and the large intestine<sup>35</sup>. While most cases of diarrhoea are self-limiting, about 5 – 7 % of EHEC infections progress to life-threatening hemolytic-uremic syndrome<sup>36</sup>. In cases of moderate to

severe diarrhoea in children under the age of 24 months, EPEC infection increased the risk of death more than two-fold<sup>37</sup>.

As non-invasive enteric pathogens, EPEC and EHEC do not use their injectisome for cell invasion. Instead, they use it to create attaching and effacing (A/E) lesions in the gut (Figure 1.2). The bacteria adhere to the intestinal epithelial cells and, similarly to *S. enterica*, use effector proteins to modulate the host cell cytoskeleton. The actin cytoskeleton is rearranged in such a way that microvilli are flattened and form a raised pedestal to which the bacteria adhere. When expressed in non-pathogenic *E. coli* K-12, the LEE pathogenicity island from EPEC is sufficient for A/E lesion formation, but the LEE from EHEC is not<sup>38,39</sup>.

To form the A/E lesion, EPEC first adheres to the host cell using a type IV pilus (T4P)<sup>35</sup>. From there, it secretes proteins into the host cell via the LEE-encoded injectisome. A key effector EPEC translocates is Tir, which is inserted into the host cell membrane<sup>40</sup>. It acts as a receptor for intimin, a protein embedded in the bacterial outer membrane. Upon tir and intimin interaction, a signalling cascade triggers actin cytoskeleton rearrangement and pedestal formation. This interaction allows EPEC to remain intimately attached to the host cell, which is thought to give it a competitive advantage against other intestinal bacteria<sup>41</sup>.

The number of effectors encoded are variable depending on the strain, with approximately 22 in EPEC up to 39 in some strains of EHEC<sup>35</sup>. There are seven LEE-encoded effectors shared by EPEC and EHEC: Tir, Map, EspF, EspG, EspH, EspZ, and EspB<sup>35</sup>. The strains additionally encode unique effectors outside of the LEE pathogenicity island, termed non-LEE-encoded effectors, some of which have been shown to be essential for A/E lesion formation<sup>42</sup>. These effectors have variable targets and downstream effects in the host cell. For example, the disruption of tight junctions in the gut has been attributed to EspF, Map, and EspG, though each

of these effectors has different intracellular targets<sup>43</sup>. The effector EspZ inhibits host cell apoptosis and regulates effector translocation to limit cytotoxicity, prolonging the adhesion of the pathogen to the host<sup>44,45</sup>. Research is ongoing to fully understand the effects of LEE-translocated proteins on the host cell and on bacterial survival.

#### 1.2.4 The injectisome as a drug target

As the bacterial injectisome is important to the virulence of many clinically relevant pathogens, there have been many studies investigating its potential as a therapeutic or vaccine target. Drugs specifically targeting the injectisome, without unintended targets in the related flagellum or other systems, can reduce the virulence of a pathogen without decreasing its fitness<sup>14</sup>. This quality makes drugs targeting the injectisome have a lower risk of resistance development, as selective pressure is lower when viability is unaffected.

Several classes of inhibitors have been shown to target components of the injectisome, with one having been tested in a phase 1 clinical trial<sup>46</sup>. A review by Lyons *et al* provides detailed information on injectisome-targeting drugs<sup>14</sup>. Several compounds have been shown to target proteins regulating injectisome expression, while others target its assembly, the extracellular needle, and the cytosolic ATPase. For example, the antibody MEDI3902 was identified through phenotypic screening and has been shown to block assembly of the needle tip in *Pseudomonas aeruginosa*, protecting from infection in mouse models<sup>47,48</sup>. This antibody has been shown to be safe in phase 1 clinical trials, showing promise as a candidate for further development<sup>46</sup>.

Interestingly, the injectisome itself is being investigated and engineered as a potential delivery mechanism for drugs with intracellular targets<sup>49</sup>. Improved structural characterization of the injectisome in recent years is contributing to targeted therapeutic development, which will hopefully provide more treatment options for multi-drug resistant infections in the future.

Many proteins in the injectisome are promising candidates for vaccine development. Proteins including the needle tip<sup>50</sup>, translocators<sup>51</sup>, and the pilotin<sup>52</sup> have been shown to provide varying degrees of immunity to *S. enterica* and *P. aeruginosa* in mouse models. An attenuated EHEC vaccine secreting effectors including Tir resulted in reduced fecal shedding in calves, a promising result for reduced contamination in agriculture<sup>53</sup>. We may one day benefit from vaccines made up of injectisomal structural proteins and effectors.

### 1.3 Structural characterization of the T3SS

The injectisome has been the subject of structural studies for decades, an undertaking that is still not complete due to the sheer number of interacting proteins. Individual components have been characterized by X-ray crystallography and nuclear magnetic resonance (NMR), whereas holistic information on the system has generally come from negative-stain transmission electron microscopy (TEM) and cryo-electron tomography (cryo-ET). In the past decade, advances in cryo-electron microscopy (cryo-EM) technology, particularly in electron microscope detectors and data processing software, have pushed resolution limits of the technique into near-atomic resolution permitting *ab initio* modelling<sup>54,55</sup>. As single-particle cryo-EM is a technique well suited to studying large, symmetrical complexes, these advances have been instrumental in a goldrush of injectisome structures in the past five years. The combination of individual protein structures since the early 2000s and more recent structures of large, oligomeric T3SS protein complexes have contributed to a near-complete atomic model of the injectisome (Figure 1.3).

As early as 1998, negative stain TEM of purified SPI-1 injectisomes from *S. enterica* serovar Typhimurium (*S. Typhimurium*) revealed its syringe-like structure<sup>56</sup>. Since then, the *S. Typhimurium* SPI-1 injectisome has been the archetypical system, with most structural

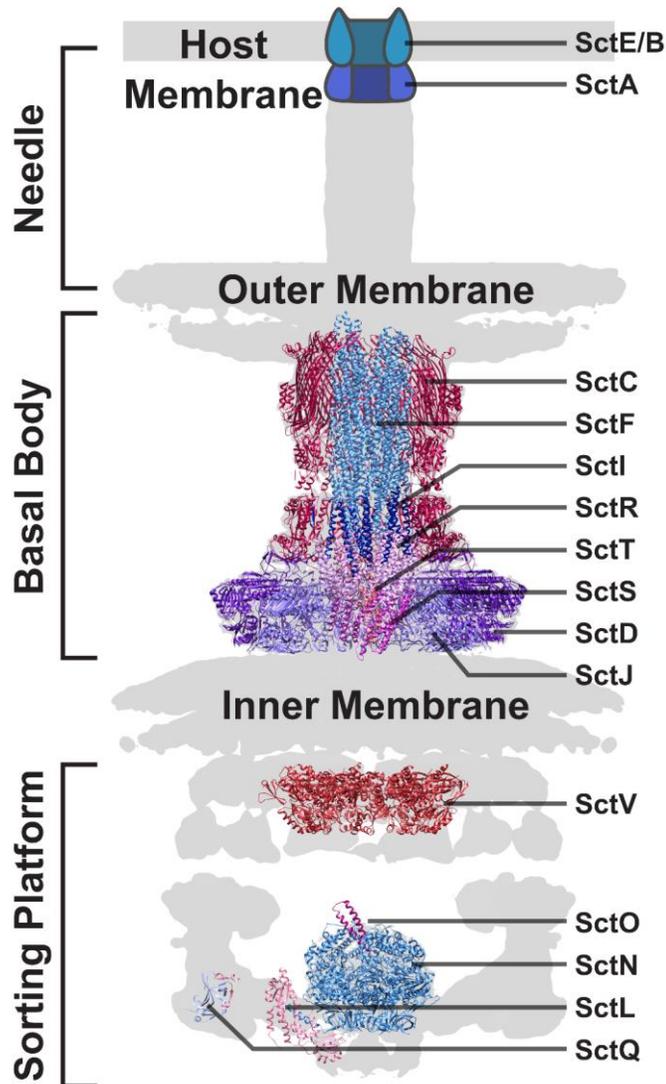


Figure 1.3: Structural characterization of the injectisome.

The basal body has the most complete structural characterization. The structures of solved injectisome components are overlaid on an *in situ* map of the *S. Typhimurium* injectisome.

characterization performed on this system. The injectisome can be divided into three sections, made up of several proteins each: the cytosolic sorting platform, the basal body, and the extracellular needle<sup>9</sup>. The sorting platform – the site of effector processing – consists of a cytosolic ATPase related to  $F_0/F_1$  ATPases, six supporting pods each comprising four different

proteins, and the cytosolic region of the inner membrane pore<sup>57</sup>. The basal body forms the cylindrical “syringe” portion of the complex, with several concentric protein rings passing through the Gram-negative envelope. These protein rings surround the periplasmic export apparatus and the filamentous rod and needle<sup>58</sup>. The extracellular portion of the injectisome consists of the hollow needle (~80 nm long in *S. typhimurium*<sup>56</sup>), tip protein, and translocon pore inserted into the host cell membrane. The following sections will provide detailed descriptions of the structural characterization of each injectisome subcomplex, using the universal (Sct) rather than species-specific nomenclature for individual proteins.

### 1.3.1 The basal body

To create a continuous channel from the bacterial to host cytosol, the injectisome must cross the Gram-negative IM, peptidoglycan wall, and OM. The primary function of the proteins comprising the basal body is to cross the bacterial envelope; thus, many of the proteins in this complex are membrane proteins, making them difficult to purify and characterize. The bulk of the basal body comes from oligomeric ring proteins: the 24-fold IM rings SctD and SctJ, and the 15-fold OM secretin pore SctC. SctD and SctJ each have a single IM TM and a periplasmic domain, with SctD also having a small cytosolic N-terminal region<sup>59</sup>. SctC has three N-terminal periplasmic domains termed N0, N1, and N3, and a large C-terminal  $\beta$ -barrel domain inserted into the OM. Early structural characterization by X-ray crystallography and NMR focused on soluble fragments of these proteins. A specific structural motif called the ring-building motif (RBM), characterized previously in unrelated oligomeric ring systems, was found to be present in the periplasmic domains of all three proteins<sup>60</sup>. Based on computational modelling, monomeric structures, and previous low-resolution reconstructions of the injectisome<sup>61,62</sup>, models were generated for the periplasmic regions of SctC, SctD and SctJ<sup>60,63,64</sup>. However, the

models lacked the accuracy of high resolution structures, especially in the case of SctC which had incorrect C12 stoichiometry applied.

In 2016, the cryo-EM structure of the SPI-1 *S. Typhimurium* basal body was solved to a resolution of 4.3 Å for 24-fold SctD-SctJ and 3.6 Å for 15-fold SctC59. Characterization of these three proteins involved purification of massive oligomeric rings spanning two separate membranes, a feat that would be extremely difficult to replicate in X-ray crystallography. The model gleaned from this map provided the first atomic insight into the OM secretin SctC  $\beta$ -barrel, which was the largest  $\beta$ -barrel characterized at the time. SctC formed a double-walled  $\beta$ -barrel made up of 60 strands, with the inner  $\beta$ -barrel kinked towards the centre of the structure to form a closed gate (Figure 1.4). The gate prevents free diffusion of solutes across the OM before assembly of the extracellular needle. The shear on the inner  $\beta$ -barrel was 0 – resulting in strands

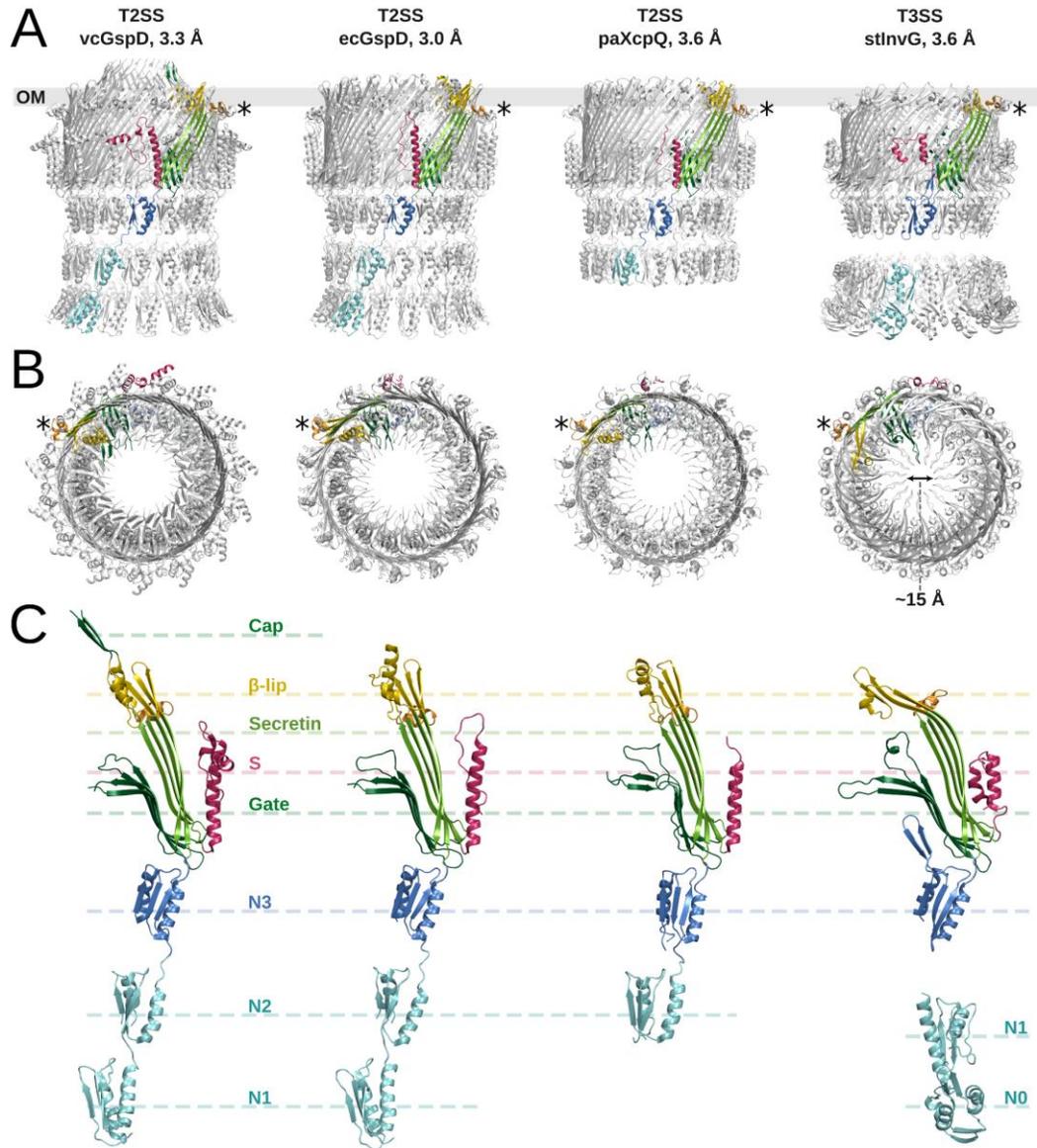


Figure 1.4: Structures of injectisomal (T3SS) and homologous T2SS secretins. Comparison of near-atomic resolution structures of T2SS *Vibrio cholerae* GspD, *E. Coli* GspD, and *P. aeruginosa* XcpQ, and injectisomal *S. Typhimurium* InvG. Domains are coloured as follows: membrane-inserting  $\beta$ -lip (gold), amphipathic helical loop (orange), secretin (green), periplasmic and cap gates (dark green), S domain (magenta), N3 domain (blue), and remaining N domains (turquoise). Asterisk denotes the position of the amphipathic helical loop. Comparison of secretin (a) side view, (b) top view, and (c) side view of a single monomer.

perfectly perpendicular to the outer membrane – a property that has only been observed in a

handful of other massive  $\beta$ -barrel structures<sup>65</sup>. A conserved amphipathic helix was resolved and

hypothesized to be instrumental in  $\beta$ -barrel assembly machinery (BAM)-independent OM insertion of the secretin based on mutational studies and its position near the OM inner leaflet<sup>59</sup>.

In 2019, the structure of the *S. Typhimurium* needle complex including the basal body and needle was solved to near-atomic resolution<sup>58</sup>. It was solved in several stages of assembly; in the fully assembled state, the secretin OM gate is open to accommodate the needle, resulting in a straightening of the bent inner  $\beta$ -strands to a position flush with the outer  $\beta$ -barrel. The needle complex structure has an unusual symmetry mismatch: though the secretin  $\beta$ -barrel and N3 domain have 15-fold symmetry, the N0 and N1 domains instead have 16-fold symmetry. The 16-fold symmetry results in an 8-fold symmetric interaction with the 24-fold inner membrane ring SctD, with two protomers of SctC interacting with three of SctD.

The core periplasmic export apparatus is an  $\alpha$ -helical complex nestled inside the inner membrane rings (Figure 1.5). It is made up of the proteins SctR, SctS, and SctT with a stoichiometry of R<sub>5</sub>S<sub>4</sub>T<sub>1</sub>; however, the ratio R<sub>5</sub>S<sub>5</sub>T<sub>1</sub> has also been observed in some cryo-EM 3D classes<sup>58,66</sup>. Several cryo-EM reconstructions of this complex have been solved since 2018, two in the context of the needle complex from *S. Typhimurium*<sup>58</sup> and *Shigella flexneri*<sup>67</sup> and three flagellar homologues purified from *S. Typhimurium*<sup>66,68</sup>, *Vibrio mimicus*, and *Pseudomonas savastanoi*<sup>69</sup>. These structures revealed that the export apparatus proteins, predicted to be integral membrane proteins due to their hydrophobicity, were instead localized in the periplasm within the SctD and SctJ rings. SctR and SctT form a pseudo-hexameric core with four subunits of SctS decorating the periphery. Each subunit protomer forms one to three  $\alpha$ -helical hairpins; SctR forms two, SctQ forms one, and SctT is a fusion of R and Q and forms three. The subunits are organized in a pseudo-helical arrangement, creating a narrow channel for substrate passage. A loop on the IM-facing side constricts the channel in the solved structures, preventing effector

accommodation even in a partially unfolded state. It is not yet known what causes the export apparatus to open; it is possible that interaction with inner membrane protein SctV could trigger a conformational change that would open the channel<sup>58</sup>. The exterior of the export apparatus is primarily hydrophilic, interacting with the inside of the SctJ ring; in addition, SctR projects a loop that interacts electrostatically with the secretin N0 domain<sup>58</sup>.

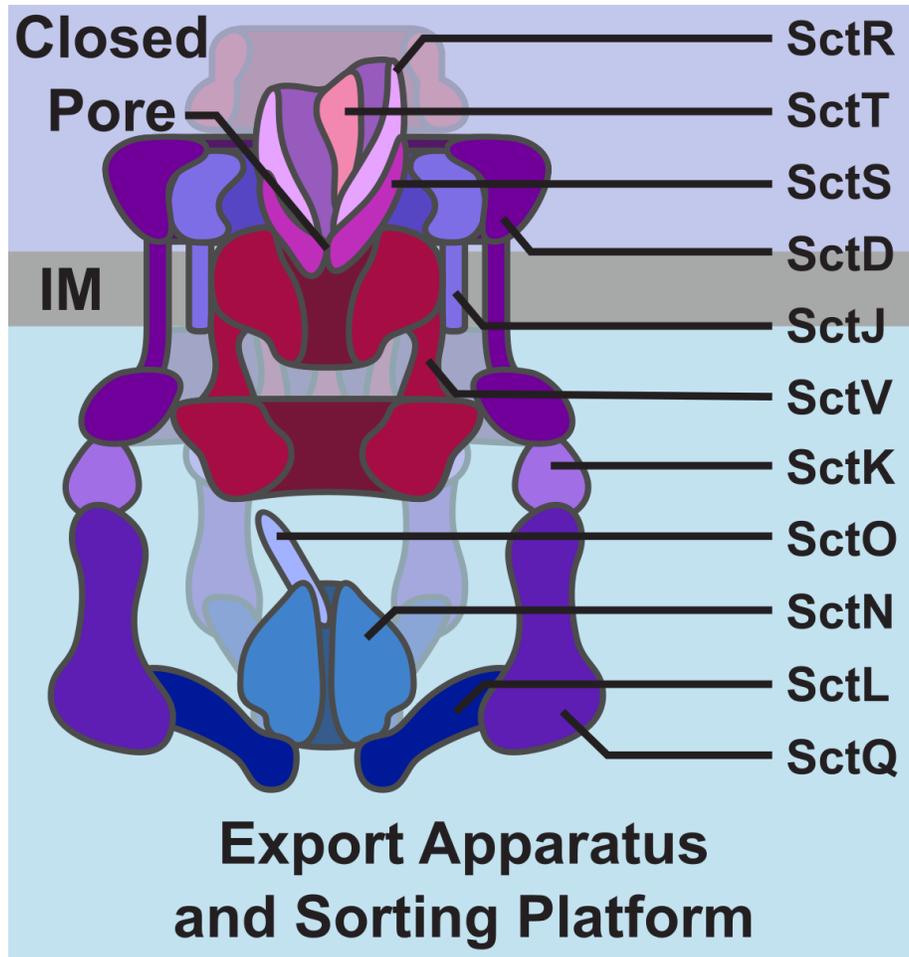


Figure 1.5: Schematic of the export apparatus and sorting platform proteins: The export apparatus is situated on the periplasmic side of the inner membrane, made up of SctRST in a 5:4:1 ratio. SctV creates a pore through the inner membrane. The sorting platform consists of the ATPase complex (SctLNO) and six supporting pods (SctQK). The pods are anchored to the inner membrane rings, SctD and SctJ.

Attached to the export apparatus is the inner rod protein SctI, which secures the needle to the basal body<sup>70</sup>. Six molecules of the inner rod SctI bind to the periplasmic face of the export apparatus, forming a homogenous platform for needle assembly<sup>58</sup>. SctI forms additional  $\alpha$ -helical hairpins arranged in a pseudo-helix, with the first subunit bound in a groove between SctR and SctT. At its N-terminus, it forms  $\beta$ -strand complementation interactions with the N1 domain of secretin SctC. In certain systems, the inner rod protein has been shown to interact with a lytic transglycosylase; in EPEC, the transglycosylase is called EtgA. EtgA is a small (~17 kDa), primarily  $\alpha$ -helical protein thought to clear the peptidoglycan near the injectisome IM complex, generating a hole large enough for interaction between the IM rings and the OM secretin<sup>71</sup>. The binding between EtgA and SctI has been characterized by ITC, but no structure of the complex has been solved. Thus, the inner rod serves an important role in assembly through recruitment of the lytic transglycosylase and later through binding of needle filaments.

In flagellar homologues, the interaction between the export apparatus and the autoprotease SctU has been structurally characterized. At the IM face, the export apparatus binds to the autoprotease SctU, a substrate switch protein whose function will be discussed in depth in section 1.5.1. In structures from flagellar homologues, one copy of SctU was bound to the export apparatus, forming four helices wrapped around the complex<sup>69</sup>. The SctU-bound complex is in the stoichiometry R<sub>5</sub>S<sub>4</sub>T<sub>1</sub>, with one helical hairpin occupying the site sometimes bound by a fifth subunit of SctS<sup>66,69</sup>.

The only basal body protein that has not been characterized fully is the inner membrane protein SctV, which likely interacts with the IM face of the export apparatus. The structure of its nonameric cytosolic domain has been solved by crystallography in a monomeric form<sup>72,73</sup> and nonameric ring form<sup>74</sup>. In addition, the cryo-EM structure of the EPEC homologue EscV has

been solved and is discussed in depth in chapter 4. In brief, SctV has a four-subdomain structure, of which two subdomains form the primary electrostatic oligomerization interface. It forms a nine-fold symmetric ring with a 5 nm pore diameter, through which substrates can pass. Despite the in depth characterization of the cytosolic domain, the only structural characterization of the SctV transmembrane domain comes from cryo-electron tomography (cryo-ET), revealing an inner membrane ring with a large pore<sup>57</sup>. Thus, SctV is the last missing piece of the basal body structural model and attempts to solve its full-length structure are ongoing.

### 1.3.2 The cytosolic sorting platform

The cytosolic components of the injectisome sorting platform are mainly involved in processing substrates prior to secretion (Figure 1.5). A key component is the ATPase SctN, homologous to F<sub>0</sub>/F<sub>1</sub> rotary ATPases. SctN interacts with two other homologues of ATPase complexes: the central stalk SctO and the peripheral stalk or stator SctL<sup>75-77</sup>. The remaining components, SctQ and SctK, form six pods visualized by cryo-ET that attach the stator SctL to the cytosolic region of the inner membrane ring SctD<sup>57,78,79</sup>.

The role of the ATPase is to separate substrates from their bound chaperones to be fed into the secretion system in a partially unfolded state<sup>80</sup>. It is homologous to the catalytic domain of F<sub>1</sub> ATPases and has been structurally characterized in monomeric form in EPEC, *S. Typhimurium*, and *S. flexneri*<sup>81-85</sup>. It has three subdomains: the N-terminal oligomerization domain, the catalytic domain, and the C-terminal stalk-interaction domain. The structure of the ATPase EscN from EPEC in complex with stalk EscO has been solved by cryo-EM, the results of which are reported in depth in chapter 4<sup>86</sup>. In brief, EscN forms a homohexameric complex with ATP binding sites situated at the interface between monomers. The inner stalk EscO is bound in the central “pore” formed by EscN. The asymmetry of the complex and its similarity to ATP synthase and V-

ATPases suggests that EscN acts in a similar rotary mechanism, with the stalk EscO acting as the rotor. The stalk itself has only about half of its residues resolved in the cryo-EM structure. However, SctO's structure has been solved prior by crystallography in *Chlamydia trachomatis*<sup>87</sup> and *Vibrio parahaemolyticus*<sup>88</sup>, and a homology model of the EPEC variant<sup>89</sup> is in agreement with the portion resolved in cryo-EM. SctO forms a coiled coil domain with its N- and C-termini bound by the ATPase. Further studies are required to determine how the hypothesized rotation of SctO results in the dissociation of the chaperone from its bound effector protein.

The role of the ATPase stator SctL is to immobilize the catalytic domains of SctN during rotary catalysis. SctL is fixed to the injectisome through SctQ<sup>90</sup>, which is in turn bound by SctK<sup>91</sup>, which presumably interacts with the cytosolic region of inner membrane ring SctD. While there is no structure solved of the injectisomal SctL, the flagellar homologue FliH has been solved in complex with the ATPase FliI<sup>92</sup>. This structure comprises the C-terminal half of SctL. Two anti-parallel molecules of SctL are bound to the ATPase N-terminal oligomerization domain. SctL is composed of a long  $\alpha$ -helix with a globular domain at one end consisting of a four-strand  $\beta$ -sheet and three  $\alpha$ -helices. The  $\beta$ -sheet interacts with one strand from SctN, and the N-terminal most helix of SctN interacts with both molecules of SctL to form a helical bundle. In an oligomerized ATPase, this would imply a stoichiometry of N<sub>6</sub>L<sub>12</sub>, but such a structure has yet to be solved.

The cytoplasmic ring SctQ is the largest member of the six pods supporting the ATPase. Its C-terminal half from *S. Typhimurium* has been solved in complex with a T4 lysozyme fusion with a small N-terminal peptide of SctL<sup>93</sup>. SctQ forms a curved  $\beta$ -sheet with two  $\beta$ -strand hairpins and two  $\alpha$ -helices in loop regions, made up of two intertwining “surface presentation of antigen” domains. It forms an extensive interface with the short EscL peptide, which consists of

a lariat containing three short  $\alpha$ -helices and a  $\beta$ -strand. The N-terminal half of SctQ has not been structurally characterized, nor has its interaction partner SctK that connects it to the basal body through SctD. As the cytosolic sorting platform has proven difficult to purify as a single complex, or to purify from native expression with the needle complex, it may take time to structurally characterize the entire substructure at high resolution.

### 1.3.3 The needle, tip, and translocon

The injectisome needle SctF forms a hollow helical filament through which secreted effectors pass in a partially unfolded state. The length of the needle varies between species (~80 nm long in *S. typhimurium*<sup>56</sup>), bridging the distance between the bacteria and host cell. The lumen has a diameter of approximately 1.5 nm, sufficient to closely fit an  $\alpha$ -helix<sup>94</sup>. The end of the needle is capped by the tip protein SctA. In some species, such as EPEC, the tip is elongated into a long filament similar to the flagellar filament<sup>95,96</sup>. The injectisome secretes the components of the translocon, made up of SctB and SctE, which insert into the host membrane and create a pore through which effectors can pass. Adherence of the tip protein to the assembled translocon is the final step in creating a continuous channel from the bacterial cell to the host.

There are numerous structures of the injectisome needle, most of which are monomeric or cognate chaperone-bound forms solved by X-ray crystallography and NMR<sup>97-102</sup>. The species characterized include *S. Typhimurium*, *Shigella flexneri*, *Yersinia pestis*, and *Bulkholderia pseudomallei*. Models of the filament structure were calculated based on low resolution cryo-EM envelopes<sup>97,103</sup>. The first oligomerized structure came from a needle assembled *in vitro* solved by solid state NMR<sup>104</sup>, and two similar structures followed<sup>105,106</sup>. However, the most recent structures from high-resolution cryo-electron microscopy have different helical parameters than those solved by solid state NMR, resulting in a poor fit between the structures<sup>94,107</sup>. As the cryo-

EM structures have similar helical parameters when the filament was assembled *in vitro* and *in vivo*, it is likely the differences stems from the experimental technique used<sup>107</sup>.

The SctF needle monomer forms an  $\alpha$ -helical hairpin with a bend in one of the helices. The monomers form a symmetrical helical assembly with a turn of  $63.34^\circ$  and a rise of  $4.33 \text{ \AA}$ <sup>94</sup>. This assembly forms a lumen in the shape of a right-handed helix, with an inner diameter ranging from 1.3 nm to 2.0 nm wide. The inner surface of the channel has a prominent positively charged groove, which arises from conserved C-terminal residues pointing into the lumen. The positively charged residues form a spiral pattern surrounding the lumen and are hypothesized to contribute in an active role for substrate secretion<sup>94</sup>.

The tip protein SctA binds to the end of the oligomerized needle filament, mediating the interaction with the translocon pore. Its structure has been solved by X-ray crystallography in *Bulkholderia pseudomallei*<sup>108–110</sup>, *Shigella flexneri*<sup>110–112</sup>, and *S. Typhimurium*<sup>113</sup>, and in one case, in complex with the needle protein<sup>102</sup>. The core of the tip protein is a coiled coil domain, neighbored by three shorter  $\alpha$ -helices and a small third domain of mixed  $\alpha$ -helical and  $\beta$ -sheet content. The tip protein has been found to bind bile salts; the presence of bile salts has variable effects on secretion depending on the species, increasing invasion in *S. flexneri* and repressing it in *S. Typhimurium*<sup>102</sup>. The bile salt binding interface is a hydrophobic patch on the central coiled coil, causing a subtle conformational change in helix orientation<sup>102,111,113</sup>; it is unclear how bile salt binding here affects secretion, but it may affect the affinity with which the tip binds to the needle<sup>111</sup>.

The structure of the tip has also been studied in the context of needle binding. Upon binding to the needle SctF, the conformation of SctA changed, with a shift in the coiled coil to form a four-helix bundle with the needle<sup>102</sup>. As there is no high-resolution cryo-EM structure of the tip

assembled onto the needle, estimates of the stoichiometry of binding come through indirect methods. Cheung *et al* used streptavidin labelling in *Shigella flexneri* and found that the end of the needle was bound by four copies of tip SctA along with one copy of a translocon component, SctE<sup>112</sup>. Based on cross-linking experiments, they were able to generate a model of the four SctA copies and fit them into a low-resolution cryo-EM map of the *S. flexneri* tip complex, with sufficient space for the fifth SctE to bind. More structural studies are required to model the tip complex bound to the needle through direct methods.

The translocon is a complex forming a pore through the host cell membrane. It is made up of the major translocon protein SctE and the minor translocon protein SctB. The complex has proved difficult to characterize structurally, with only small cognate chaperone-bound portions<sup>114–117</sup> and a soluble coiled coil domain from the major translocon<sup>118</sup> solved by X-ray crystallography. The minor and major translocon proteins are predicted to have one<sup>119,120</sup> and two<sup>121</sup> transmembrane helices respectively, and both have a predicted coiled coil region<sup>121–123</sup>. The minor translocator also has an amphipathic helix required for insertion into the membrane<sup>124</sup>, its structure solved by NMR<sup>125</sup>. Stoichiometry estimates vary, ranging from E<sub>1</sub>B<sub>4</sub>A<sub>6</sub> in *S. flexneri* (including tip SctA)<sup>126</sup>, an E<sub>8</sub>B<sub>8</sub> hexadodecamer in *P. aeruginosa*<sup>127</sup>, and a 500-700 kDa complex with 15-20 SctE/SctB subunits in *Yersinia enterocolitica*<sup>128</sup>. There is no structure to date of the membrane-associated form of the translocon pore, but its ~13.5 nm diameter envelope has been observed through AFM<sup>129</sup> and cryo-electron tomography<sup>130</sup>. The difficulties in purifying this membrane complex have prevented detailed structural knowledge of the translocon pore.

#### 1.4 Role of pilotins in injectisome assembly

Pilotins are a family of proteins that facilitate assembly of the OM secretin pore. The secretin protein SctC has a highly conserved, OM-inserted  $\beta$ -barrel domain and a variable number of

periplasmic N-terminal RBM domains<sup>131</sup>. At its C-terminus, it has a short motif termed the S domain that binds to its cognate pilotin protein. The secretin outer membrane pore is conserved not only between injectisome-expressing species, but also with the T2SS, the type IV pilus (T4P), and filamentous phage. In contrast with secretin conservation, the pilotins are made up of a large variety of non-homologous proteins, often with varying secondary and tertiary structures. This may be the reason why pilotin knockouts have different effects on secretin assembly, as they have varying mechanisms of action.

All pilotin proteins are small, OM-inserted periplasmic lipoproteins<sup>132</sup>, shown to use the localization of lipoproteins (LoL) pathway to reach the OM in *Klebsiella oxytoca*<sup>133</sup>. All function by binding the C-terminal S domain of the secretin, with the exception of secretins from the T4P that do not have an identified pilotin binding site<sup>134–136</sup>. Another exception are the self-targeting secretins, such as HxcQ from *P. aeruginosa*<sup>137</sup>; these have no S domain, and instead have a conserved lipidation signal at their N-terminus for OM localization. Pilotins generally function by protecting the secretin from degradation, and by either facilitating secretin localization, assembly, or both. Due to their lack of homology, they have species-specific names. Pilotins are thought to work by one of two general mechanisms: either they form a ternary complex with the secretin and LolA (a LoL pathway protein that binds the lipidated N-terminus) and co-migrate to the OM, or they are first inserted into the OM and use a long N-terminal linker to bind the secretin near the IM. The pilotin is thought to dissociate after secretin assembly in injectisomal systems, but there is evidence that it may remain associated in the T2SS<sup>138,139</sup>. A detailed discussion and diagram of these two mechanisms can be found in chapter 2.

The pilotin from *S. Typhimurium*, InvH, is essential for the OM localization of the secretin protein<sup>140,141</sup>. Knockout of InvH results in mis-targeting of SctC to the IM, and in lower levels of

SctC. In this system, SctC is able to auto-oligomerize in the absence of the pilotin; this results in SctC forming pores in the IM, permeabilizing the bacterium. This phenotype is shared by the well-studied T2SS pilotin PulS from *Klebsiella oxytoca*<sup>142</sup> and the injectisome pilotin YscW from *Y. enterocolitica*<sup>143</sup>, where YscW also increased the yield of SctC oligomers. In species where the pilotin is essential to secretin assembly, pilotin knockout can result in a completely monomeric secretin population. Most injectisome pilotins do not fall into this category; the pilotins PilF<sup>135</sup> and PilW<sup>144</sup> from T4P examples of pilotins essential for secretin assembly, with PilF also important for OM localization. The large diversity of pilotins suggests that there may be differences in secretin targeting unique to each system.

## 1.5 Secretion hierarchy

The selection of specific proteins for secretion through the injectisome is a tightly regulated process. Part of the assembly process, specifically assembly of the inner rod, needle, and translocon, are dependent on secretion; thus, these substrates must be secreted prior to effectors. Secreted proteins are divided into early (needle and inner rod SctF and SctI), middle (tip SctA, translocon proteins SctE and SctB), and late substrates (effectors). There are two substrate switches responsible for transitioning through the three groups. After the needle reaches a specific length, the ruler protein SctP and autoprotease SctU trigger the switch to secreting the tip and translocon proteins. The gatekeeper protein SctW is involved in the switch from middle to late effectors once the translocon pore has been assembled.

### 1.5.1 Regulation of needle length

The needle is oligomerized to a specific length before secretion specificity is switched to middle substrates. One of the proteins key to the switch is the inner membrane autoprotease SctU, which interacts with the export apparatus<sup>69</sup>. Autoproteolysis of the SctU cytoplasmic loop

is required for the switch to middle substrate secretion<sup>145-151</sup>; while it is unclear how the proteolysis is triggered, it is thought to be through interaction with needle length regulator SctP. Knockout of SctP, also called the ruler, results in the formation of abnormally long needles<sup>152-155</sup>. There are two main hypotheses for how the length of the needle is controlled: the molecular ruler model<sup>156</sup>, and the timer model<sup>70</sup>.

The molecular ruler model states that the physical length of SctP dictates the length of the needle. Early experiments demonstrate that the length of the region in between the SctP N-terminal and C-terminal domains is directly proportional to needle length<sup>156</sup>. It is thought that the N-terminus of SctP is secreted, and its C-terminal domain interacts with autoprotease SctU; once the needle is oligomerized to the length of SctP, an interaction with the SctP N-terminus signals the autoproteolysis of SctU. However, as one molecule of SctP is sufficient to control needle length, this model relies on concurrent secretion of the ruler and needle through the narrow injectisome channel<sup>157</sup>.

The timer model posits a key role for the inner rod SctI in controlling needle length. It is derived from the observation that overexpression of the needle SctF results in longer needles, while overexpression of inner rod SctI results in shorter needles; both mutants remain functional<sup>70</sup>. From this data, it was proposed that needle oligomerization continues until the inner rod is fully assembled, anchoring the needle and triggering the switch to middle substrates. In this model, the role of the ruler SctP is less clear; however, there is evidence that the inner rod does not assemble in the absence of SctP<sup>158</sup>. The timer model is supported by mathematical modelling, which generated a similar distribution of needle lengths as found in wild-type and SctI overexpression *Salmonella* strains<sup>159</sup>. However, a recent paper suggests that the ruler model

is more consistent with the data, proposing that overexpression of needle or inner rod affects needle length through secretion substrate competition<sup>160</sup>.

Further studies are needed to clarify whether needle length is controlled by a ruler, a timer, or a combination of the two models. In addition, there is little data on how the autoproteolysis of SctU changes the affinity from early to middle substrates; structural data of substrate-bound complexes could clarify this mechanism.

#### 1.5.2 Switch to late substrates

The gatekeeper protein SctW is responsible for the switch from secreting translocators to effectors (middle to late substrates). Knockout of SctW abolishes secretion of translocon proteins, instead over-secreting effectors<sup>138,161</sup>. SctW has been shown to bind IM pore SctV in EPEC, increasing the affinity of SctV for middle chaperone-substrate complex CesAB/EspA and decreasing its affinity for late complex CesT/Tir<sup>162</sup>. In the absence of SctW, SctV had high affinity for late chaperones, indicating it may facilitate a conformational change in SctV. Thus, the gatekeeper likely facilitates a conformational change in SctV to promote middle chaperone binding.

The signal for switching from middle to late effectors is not certain. The switch has been proposed to be triggered by calcium ion concentration, pH, or by physical contact of the needle tip and translocon with the host cell<sup>163-166</sup>. It is unclear how this signal is transmitted to the injectisome base; potentially, it is through conformational changes in the needle and inner rod. Structural studies on the needle complex at various stages of substrate export, or on the IM pore in complex with the gatekeeper, will improve our understanding of signalling in the injectisome.

## 1.6 Energetics of secretion

The mechanism by which the injectisome harnesses energy for secretion has proven elusive. Chaperone effector complexes are targeted to the cytosolic sorting platform. From there, the effector must be partially unfolded and fed through the inner membrane pore to be secreted. It is not clear how the effector is unfolded, or how it passes unidirectionally through the injectisome channel. However, there have been many microbiology and biochemical studies on the topic that provide insight into key proteins involved in the process.

Upon the discovery of the injectisome ATPase SctN in 1994, it was thought to be the source of energy for secretion<sup>167</sup>. Knockout of the ATPase in *Y. enterocolitica* abolished secretion of the effector Yop, and decreased secretion of YopD and YopE. However, this was not the complete story; in 2005, it was discovered that the injectisome relied on proton motive force, as treatment with the protonophore carbonyl cyanidem-chloro-phenylhydrazone (CCCP) inhibited secretion<sup>168</sup>. The same was true for the related flagellum<sup>168-171</sup>, and most of the studies concerning the role of proton motive force on T3S have been done on flagellar homologues. The ATPase was found to have a different primary role: to separate chaperone-effector complexes through ATP hydrolysis prior to secretion<sup>80,172,173</sup>.

It soon became clear that the role of the ATPase in injectisome energetics was not simple. Flagella knockouts of ATPase and stator SctN/SctL still allowed for a small amount of flagella to be assembled, leaving bacterial weakly motile<sup>171,174</sup>. It was then posited that one of the inner membrane export apparatus proteins, SctV, contains the proton translocation channel, as certain mutations in this protein prevented PMF-driven secretion<sup>175,176</sup>. Interestingly, the ATPase seems to be intimately associated with this process. While SctN/SctL knockouts require both a pH and potential difference gradient to function, wild-type flagella only require a potential difference

gradient<sup>175</sup>. In addition, SctV alone is able to translocate both sodium ions and protons, but the presence of the ATPase switches the flagella to only use proton motive force and makes the process less sensitive to environmental Na<sup>+</sup> and pH conditions<sup>177</sup>. The ATPase stalk protein SctO also has an important role to play. The presence of SctN/SctL are required for SctV/SctO interaction *in vivo*, which has been shown to be important for flagellar motility<sup>178</sup>. To complicate things further, a recent study using inverted inner membrane vesicles found that ATP hydrolysis by SctN in the absence of proton motive force was enough to assemble the flagellar hook<sup>179</sup>. From this data, it is clear that secretion is powered by proton motive force through cooperation of the ATPase and the inner membrane pore. The caveat to this conclusion is that all of these studies were performed on flagella; it is possible that the results would be different when testing the same mutations in the injectisome.

There is still much that is not understood about the mechanics of secretion. It is not known how the energy from proton motive force is coupled to protein unfolding and export, or how ATP hydrolysis can dissociate the effector from its cognate chaperone. One suggestion is that a conformational change occurs in the IM pore SctV in response to changes in local pH during proton translocation; this conformational change results in the cytoplasmic ring retracting towards the IM, which could push effectors towards the injectisome pore with proper timing of binding and release<sup>180</sup>. Another paper suggests that the rotation of the ATPase could facilitate a mechanism by which the effector is “screwed” into the helical lumen of the needle, rotating through its entire length<sup>181</sup>. Structural snapshots of the injectisome cytosolic proteins at various stages of secretion, combined with biophysical and microbiology assays, will be required to fully unravel how these proteins work to secrete effectors.

## 1.7 Thesis objectives

With the ongoing struggle against antibiotic resistant pathogens, the field is constantly in search of novel therapeutic targets. Drugs targeting bacterial virulence rather than survival are especially lucrative, as the selective pressure for the development of resistance is lower<sup>182</sup>. The injectisome's importance for virulence, multiple protein targets, and extracellular presence make it a good target for vaccine and drug development<sup>14</sup>. Understanding the assembly and function of the system is key to targeting it from a therapeutic perspective. As a massive nanomachine, the injectisome is also fascinating in terms of its energetics and several stable interactions with mismatched stoichiometry between its numerous protein rings. Structural studies therefore improve our understanding of how this massive protein assembly comes together, while also contributing to antimicrobial development. The main topics of this thesis cover proteins involved in injectisome assembly, effector sorting, and energetics.

Chapter 2 details the structural characterization of the binding interface between the pilotin InvH and the secretin InvG from *S. Typhimurium*. The crystal structure of apo InvH at 1.2 Å resolution forms an entwined dimer that also exists in solution. The hydrophobic binding interface between InvH and the InvG S domain is probed through both NMR- and X-ray crystallography-derived structures and confirmed through isothermal titration calorimetry (ITC) studies on mutants. The binding interface with InvG is mutually exclusive through steric clashes with the InvH dimerization interface. Size-exclusion coupled multi-angle light scattering and the ITC-calculated stoichiometry confirms that InvH forms a 1:1 complex with InvG. The nanomolar affinity of InvG for InvH supports the hypothesis that the pilotin InvH forms a dimer that is dissociated upon contact with the secretin. Though the structure of InvH is distinct from

previously characterized pilotins, the binding interface with InvG has some similarity to other systems such as a key pilotin-binding aromatic residue present on the S domain.

Chapter 3 compares the cryo-EM derived structure of the IM pore EscV cytosolic domain (EscV<sub>C</sub>) from EPEC to previous crystal structures. EscV<sub>C</sub> formed a double-stacked nonameric ring with D<sub>9</sub> symmetry, its structure solved through subtraction of one ring at C<sub>9</sub> symmetry to 4.7 Å resolution. Through homology modelling, the resolution was sufficient for model building. EscV<sub>C</sub> has four subdomains and a portion of the linker to the transmembrane region is resolved. Similar to previously characterized homologues, the interface between EscV<sub>C</sub> subunits is primarily electrostatic and mediated through subdomains 1 and 3. The resolved portion of the linker nestles into a groove on the adjacent protomer, contributing to the oligomerization interface. Mapping the chaperone-binding interface from flagellar homologue FlhA onto the structure reveals that a subtle conformational change between subdomains 2 and 4 to a more open state would be required to accommodate binding. EscV<sub>C</sub> represents the first structure of an oligomerized IM pore cytosolic domain in the solution state.

Chapter 4 describes the near-atomic resolution cryo-EM structure of the ATPase-stalk complex EscN-EscO from the EPEC injectisome. EscN forms a catalytically active homo-hexamer with structural similarity to the catalytic domains of F<sub>1</sub>- and V-ATPases. Whereas the F<sub>1</sub>- and V- systems have three catalytic subunits alternating with three non-catalytic subunits, EscN has six active subunits. The cryo-EM structure captures EscN in an asymmetrical ADP-AlF<sub>3</sub>-inhibited state, with four out of six active sites bound to nucleotide. The coiled coil stalk EscO is bound in a tilted orientation inside the EscN pore, with a similar interface as observed in the F<sub>1</sub>/F<sub>0</sub> γ-subunit and the V-ATPase D subunit. Both EscN oligomerization and the presence of EscO increase the rate of ATP hydrolysis. The similarities to previously characterized rotary

ATPases and its asymmetrical conformation suggest that EscN also undergoes rotary catalysis, with EscO acting as the rotor. A rotary catalytic mechanism may contribute to coupling the proton motive force with substrate secretion through interaction with the IM pore EscV.

## 2 Characterization of the pilotin-secretin complex from the *Salmonella enterica* type III secretion system using hybrid structural methods

### 2.1 Introduction

The pathogenicity of many clinically-relevant Gram-negative bacteria relies on a virulence nanomachine termed the type III secretion system (T3SS), or injectisome<sup>183</sup>. In the manner of a syringe, this multi-membrane spanning proteinaceous assembly allows passage of virulence effector proteins from the bacterial cytosol directly through a hollow needle filament and pore-forming tip into the cytosol of its infected host. The secreted bacterial effector proteins serve to target and manipulate host cell processes, resulting in various pathogenic effects. As such, the T3SS is essential to the virulence of important clinical and community pathogens including *Salmonella enterica*, *Pseudomonas aeruginosa*, and *Bordetella pertussis*. Whereas the effectors that are translocated vary greatly in nature and function based on each species' life cycle and niche, the T3SS injectisome apparatus itself is well conserved. In Salmonellosis, which is the infection of the gut by non-typhoidal *S. enterica*, invasion of host cells is carried out by an injectisome encoded on the *Salmonella* pathogenicity island 1 (SPI-1)<sup>184</sup>. Survival within *Salmonella*-containing vacuoles is then ensured by a second T3SS encoded on SPI-2<sup>184</sup>. As the discovery of novel antibiotic targets lags severely behind the rise in antibiotic resistant pathogens, including *Salmonella* and other pathogenic *Enterobacteriaceae*<sup>185</sup>, understanding the mechanics behind the injectisome is vital to target it for the development of new anti-virulence therapeutics.

The ~ 3 megadalton injectisome is composed of over 20 different proteins, many creating concentric rings of repeating units that pass through the inner and outer bacterial membranes<sup>183</sup>. It has three main subcomplexes: the cytosolic sorting platform with roles in substrate selection and in energizing secretion<sup>186–188</sup>; the basal body, which passes through the inner membrane, peptidoglycan wall, and outer membrane<sup>189</sup>; and the hollow needle and translocon pore, which create a channel directly into the host cell<sup>190</sup>. Assembly of such a complex system is no mean feat, and ever-evolving models have been proposed to explain observations collected from the prototypical T3SS, *Salmonella enterica*, as well as from a range of other Gram-negative bacteria (recently reviewed in Deng W *et al*<sup>183</sup>). Experiments are ongoing to test these models and more accurately describe the assembly modes for individual components of the T3SS.

The assembly of the outer membrane (OM) secretin pore of the injectisome basal body occurs by a highly regulated mechanism that, although generally conserved, differs between species in the details of its OM targeting steps<sup>191</sup>. The T3SS secretin is a single multi-domain polypeptide that oligomerizes at the OM into a massive (~ 1 MDa), gated, double-walled  $\beta$ -barrel with 15 subunits and 60  $\beta$ -strands<sup>189,192–194</sup>. It is a member of the broadly conserved OM secretin family with significant sequence and structural conservation amongst other distinct bacterial secretion systems, including the type II secretion system (T2SS), the type IV pilus system (T4PS), and the filamentous phage extrusion channel<sup>195</sup>. Secretin monomers are transported into the periplasm by the general SEC secretion system and upon oligomerization undergo BAM-independent insertion into the outer membrane<sup>196</sup>. In most cases, this process is mediated by a cognate T3SS-specific chaperone protein known as a pilotin. Pilotins are small lipoproteins anchored to the inner leaflet of the OM via the localization of lipoproteins (Lol) pathway; they serve to localize or assemble the outer membrane pore while protecting it from degradation<sup>197,198</sup>. The pilotins from the T2SS

and the injectisome bind a ~ 50 residue C-terminal region of their cognate secretins. These regions form a helix-loop-helix structure called the S domain<sup>199</sup>. In most cases, pilotin knockouts disrupt subsequent injectisome function through a range of effects including mis-targeting of the secretin to the inner membrane (*Salmonella*) or entirely preventing assembly<sup>200-202</sup>.

Despite their similar functions, pilotins are a highly diverse family of proteins. Most share little sequence identity, with differing predicted secondary structures. Of the T3SS pilotins, the two known crystallographic structures are both  $\beta$ -sheet containing proteins: *P. aeruginosa* ExsB forms a  $\beta$ -sandwich fold<sup>203</sup>, whereas *Shigella flexneri* MxiM is a cracked  $\beta$ -barrel protein with a deep hydrophobic cavity capable of binding either lipids or the secretin S domain<sup>204,205</sup>. The T2SS pilotins GspS, OutS, and PulS are on the other hand compact proteins made up of four  $\alpha$ -helices<sup>206-209</sup>, whereas AspS and EspS contain a mixture of  $\alpha$ -helices and  $\beta$ -sheets<sup>191,194,202</sup>. The T4PS pilotins are different still, forming much larger bundles of tetratricopeptide repeats and lacking an identified secretin binding site<sup>210-212</sup>. The complexes of three pilotins, T3SS MxiM, T2SS OutS, and T2SS AspS, with their respective secretin S domains (MxiD, OutX, and AspX) have been characterized at the atomic level<sup>194,205-207</sup>. Despite their differing architecture, all bind their corresponding secretin S domain via a hydrophobic interface lined with aromatic residues.

Historically, the prototypical *Salmonella* SPI-1 injectisome has been the most well-studied T3SS. However, the SPI-1 pilotin InvH had not been structurally characterized prior to this study. Predicted to be primarily  $\alpha$ -helical, it shares no sequence identity with other structurally characterized pilotins. InvH is required for OM targeting, but not oligomerization of the *Salmonella* secretin InvG<sup>200,201</sup>. Knockout of *invH* results in InvG pore formation in the IM, reducing invasion efficiency of *Salmonella* by 70-80 % in C57BL/6 mice<sup>213</sup>. As such, it is an essential player in *Salmonella* virulence and, given its periplasmic localization, a relatively

accessible and novel antivirulence target. Here, we present structures of the pilotin InvH, with and without the bound InvG S domain, determined by X-ray crystallography and NMR spectroscopy. Along with complementary biochemical experiments, these data show that InvH forms a homodimer in the absence of InvG, yet a 1:1 heterodimer when complexed with the S domain. InvH is the first pilotin to exhibit homodimer formation in solution, which suggests an alternative mechanism of secretin assembly to those proposed previously for the monomeric  $\beta$ -barrel fold pilotins such as MxiM<sup>205</sup>.

## 2.2 Methods

### 2.2.1 Expression and Purification of InvH and InvG

The genes encoding InvH and InvG from *Salmonella enterica* serovar Typhimurium were cloned into individual pET28a vectors. The constructs used were N-terminally His<sub>6</sub>-tagged InvH<sup>27-147</sup>, His<sub>6</sub>-tagged InvH<sup>70-147</sup>, His<sub>6</sub>-tagged InvH<sup>84-147</sup>, and His<sub>10</sub>-tagged InvG<sup>520-562</sup>, all cloned into the pET28a expression vector. All constructs were purified separately following the same basic protocol. *E. coli* BL21 ( $\lambda$ DE3) transformed with pET28a expressing the desired construct was cultivated in 1L of auto-induction media. Cultures were grown for 3.5 hours at 37°C, followed by 18 hours overnight at 22°C. Cells were harvested by centrifugation at 6,200  $\times$ g for 20 minutes. The pellet was resuspended in lysis buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 15 mM imidazole, one protease inhibitor mixture tablet by Roche per 100 mL) in a volume of 4 mL per gram of cell pellet. The suspension was lysed in two passes through an Avestin cell homogenizer. Insoluble material was pelleted by centrifugation at 40,000  $\times$ g for 30 minutes at 4°C. The supernatant was filtered using a 20  $\mu$ m syringe filter and loaded onto a 1 mL HisTrap HP Ni Sepharose column (GE Life Sciences) at 0.5 mL/minute. The column was washed with 40  $\times$  column volume (CV) of buffer A (20 mM HEPES pH 7.5, 500 mM NaCl) with 50 mM

imidazole, followed by 10 × CV of buffer A with 75 mM imidazole and then 10 × CV with 100 mM imidazole. InvH was eluted in buffer A with 250 mM imidazole, whereas InvG<sup>520-562</sup> required 1M imidazole. Fractions were pooled and dialyzed into buffer A with a 3.5 kDa MWCO membrane (Repligen), and the N-terminal His-tag was cleaved overnight at 4 °C with 1:1000 molar ratio of bovine  $\alpha$ -thrombin (HTI). The sample was passed over a 1 mL HisTrap HP column once more to remove the tag and any uncleaved protein. The protein was injected onto a Superdex 75 10/300 GL size exclusion column (GE Life Sciences) equilibrated in 20 mM HEPES pH 7.5, 500 mM NaCl, and pure fractions were pooled and concentrated using a 3 kDa MWCO Amicon Ultra centrifugal filter (Millipore Sigma).

For crystallization of InvH<sup>70-147</sup>, the protocol was modified as follows. The cell lysate was extracted in 1% DM for one hour at 4°C with gentle rotation prior to centrifugation to pellet insoluble material. All subsequent buffers were supplemented with 0.016% DDM. In the case of InvH<sup>84-147</sup> co-crystallization with the InvG<sup>543-558</sup> peptide (purchased at >95% purity from Biomatik), no detergent was added in this purification. Prior to size exclusion, InvG<sup>543-558</sup> was incubated with InvH<sup>84-147</sup> in a 1:1 molar ratio. The sample was injected on a Superdex 75 Increase 10/300 GL size exclusion column (GE Life Sciences), and the peak containing both InvH and InvG was pooled and concentrated using a 3 kDa MWCO Amicon Ultra centrifugal filter (Millipore Sigma). For SAXS experiments, the final size exclusion buffer was modified to 20 mM HEPES pH 7.5, 150 mM NaCl.

For NMR spectroscopy experiments, isotopically labelled InvH<sup>70-147</sup> and InvG<sup>520-562</sup> were each expressed in 1 L of M9 minimal media supplemented with 1 gm/L <sup>15</sup>NH<sub>4</sub>Cl and 3 gm/L <sup>13</sup>C<sub>6</sub>-glucose. The purification was performed following the general protocol, except that the final size exclusion step was done in 20 mM MOPS pH 6.5, 150 mM NaCl.

### 2.2.2 Crystallization of InvH<sup>70-147</sup>

Crystallization conditions were screened using the commercial screens Classics, JCSG+, and PACT (Qiagen) in sitting drop and microbatch plates. An initial microbatch hit from Classics G3 was optimized using grid screening and additives. For best crystal formation, purified InvH<sup>70-147</sup> (20 mM HEPES pH 7.5, 500 mM NaCl, 0.016% DDM) at 20 mg/mL was mixed 1:1 (1  $\mu$ L drop size) with precipitant solution (100 mM sodium acetate pH 4.6, 80 mM CdCl<sub>2</sub>, 20% PEG 300, 400 mM NaCl) in a microbatch plate coated with paraffin oil. Plates were set up at 30°C, and large crystals formed overnight. Crystals were soaked in a cryo-protectant solution made up of a 1:1 mixture of InvH<sup>70-147</sup> buffer and 100 mM sodium acetate pH 4.6, 80 mM CdCl<sub>2</sub>, 45% PEG 300 for 10 seconds prior to freezing in liquid nitrogen. For phasing, crystals were soaked in 10  $\mu$ L of mercury (II) acetate solution (4.2 mM mercury acetate, 50 mM sodium acetate pH 4.6, 20% PEG 300) for 10 minutes, then back soaked in the cryo-protectant.

InvH crystallized in the P1 space group, with unit cell dimensions  $a = 25.4 \text{ \AA}$ ,  $b = 53.2 \text{ \AA}$ ,  $c = 56.9 \text{ \AA}$ ,  $\alpha = 107.4^\circ$ ,  $\beta = 97.2^\circ$ , and  $\gamma = 96.1^\circ$ ; four copies of InvH (two dimers) were present in each unit cell. A 1.2  $\text{\AA}$  native dataset and a 1.4  $\text{\AA}$  phasing dataset (using a mercury-soaked crystal) were collected on the MX-ID and MX-BM beamlines respectively at the Canadian Light Source (Saskatoon, Canada). The phasing dataset was collected at the Hg<sup>2+</sup> L-III peak, at a wavelength of 1.006  $\text{\AA}$  ( $f' = -12.61$ ,  $f'' = 10.62$ ). 28 mercury atoms were placed, 7 per NCS group. X-ray datasets were processed with XDS<sup>214</sup> and merged with Aimless<sup>215</sup>.

SAD-phasing was performed on the mercury-soaked crystal dataset using Phenix Autosol<sup>216</sup>. Phenix Autobuild was used to build an initial structure, which was then refined into the higher-resolution native InvH<sup>70-147</sup> dataset using Phenix refine<sup>216</sup>. Refinement was completed iteratively using Phenix refine, Refmac, and Coot<sup>215-218</sup>, with anisotropic B-factors refined throughout the

entire model and alternative conformations modelled for several residues. The model encompasses residues 71-147 on two chains and residues 72-147 on the remaining two chains. Cadmium ions were modelled into any highly electron-dense regions with anomalous signal, and their occupancies were refined in Phenix refine. Final cadmium occupancies were highest in the inter-dimer cadmium chloride cluster (~ 80-99%), with lower occupancies in some peripherally bound cadmium ions (~ 30-70%). Several large densities peripheral to cadmium ions (but lacking anomalous signal) were refined as chloride ions, while one sodium ion was modelled into density coordinated by aspartate and glutamine; chloride and sodium occupancies were also refined in Phenix refine. All waters were refined at 100% occupancy. Final data statistics are shown in Table 2.1.

Structural figures were generated using PyMOL<sup>219</sup>, Chimera<sup>220</sup>, and ChimeraX<sup>221</sup>.

### 2.2.3 Co-Crystallization of InvH<sup>84-147</sup> and InvG<sup>543-558</sup>

The purified InvH<sup>84-147</sup>/InvG<sup>543-558</sup> complex (20 mM HEPES pH 7.5, 500 mM NaCl) at 4 mg/mL was mixed 1:1 (1  $\mu$ L drop size) with precipitant solution (100 mM sodium acetate pH 4.0, 0.5 M LiSO<sub>4</sub>, 8.3% PEG 6000) in a hanging drop well with 500  $\mu$ L of reservoir solution at 18°C. Large (> 100  $\mu$ m long) hexagonal bipyramidal crystals formed within one week. Crystals were soaked 10 seconds in a cryo-protectant solution made up of 80 % mother liquor and 20 % glycerol prior to freezing.

Data were collected at the Advanced Light Source (Berkeley Center for Structural Biology) on the 5.0.2 beamline. The best crystals diffracted to 2.0 Å resolution, in the P 6<sub>5</sub> 2 2 space group with unit cell dimensions  $a = 36.3$  Å,  $b = 36.3$  Å,  $c = 221.6$  Å,  $\alpha = 90^\circ$ ,  $\beta = 90^\circ$ , and  $\gamma = 120^\circ$ . There is one InvH-InvG complex present in the asymmetric unit. The data were processed using Dials<sup>222</sup> and merged with Aimless<sup>215</sup>.

The data was phased by molecular replacement in Phaser<sup>215,223</sup>, using a starting model from residues 90-146 of the InvH dimer chain B. Phenix Autobuild was used to build an initial structure<sup>216</sup>. Refinement was completed iteratively using Phenix refine, Refmac, and Coot<sup>215-218</sup>. TLS parameters were refined in Refmac, with one TLS group encompassing InvH<sup>84-147</sup> and a second encompassing InvG<sup>543-558</sup>. Occupancies of sulfate ions were refined using Phenix refine, whereas waters were modelled with 100% occupancy. Final data statistics are shown in Table 2.1.

#### 2.2.4 Nuclear Magnetic Resonance Spectroscopy

NMR spectra were recorded on Bruker Avance III 500 MHz or 850 MHz spectrometers. Experiments were done at 25°C with 0.5 mM labelled protein in 20 mM MOPS pH 6.5, 150 mM NaCl, 0.2 mM TCEP-HCl, and 10% D<sub>2</sub>O. In the first set of experiments, <sup>13</sup>C/<sup>15</sup>N InvH<sup>27-147</sup> was incubated with a 25% molar excess of unlabelled InvG<sup>520-562</sup>. The inverse case had <sup>13</sup>C/<sup>15</sup>N InvG<sup>520-562</sup> incubated with 25% molar excess of unlabelled InvH. The following spectra were collected using TopSpin (Bruker) for each sample: <sup>15</sup>N-HSQC, HNCO, HNCACB, CBCACO, HN(ca)CO, HCCH-TOCSY, HCC(co)NH-TOCSY, <sup>13</sup>C-HSQC, HBHA(cbcaco)NH, hbCBcgc dHD, hbCBcgc dceHE, HBCBcgcCH, <sup>13</sup>C/<sup>15</sup>N-NOESY-HSQC, and <sup>13</sup>C/<sup>15</sup>N-filtered-edited NOESY-HSQC. The NOESY mixing times were 100 msec, while the filtered-edited NOESY mixing times were 150 msec.

Data were visualized using TopSpin (Bruker), processed with NMRpipe<sup>224</sup> and analyzed using NMRFAM-SPARKY<sup>225</sup>. Assignments of signals from backbone <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N nuclei were automatically generated by PINE<sup>226</sup> and manually verified. Signals from sidechain nuclei were assigned with the help of simulated data from PACSY<sup>227</sup>. The remaining chemical shifts

were manually assigned using the collected spectra. Secondary structure and RCI-S<sup>2</sup> analyses based on chemical shifts was performed using Talos-N<sup>228</sup>.

Structure calculation was performed in CYANA<sup>229</sup>. Assigned chemical shifts, dihedral angle restraints generated by Talos-N, and unassigned peaks from <sup>13</sup>C/<sup>15</sup>N-NOESY HSQC spectra were used with Cyana's automated NOESY assignment and structure calculation algorithms. The two molecules, InvH<sup>70-147</sup> and InvG<sup>520-562</sup>, were attached by their C- and N- termini respectively during structure calculation by a 20-residue "ghost" linker. The ghost residues were required for structural calculations of the heterodimeric complex using CYANA, but did not provide any restraints to bias the positioning of the two constituent proteins. Histidine sidechains were modelled in their protonated imidazolium cation state. All X-Pro bonds were constrained to the trans conformation by chemical shift criteria<sup>230</sup>. Eight cycles of automated NOESY assignment and structure calculation were performed, generating 100 models at each cycle using simulated annealing with 10,000 torsion angle dynamics steps. The best 20 models from the final cycle were aligned and their coordinates saved in the final ensemble.

#### 2.2.5 Size-exclusion coupled Multi-angle Light Scattering

The oligomerization states of wild-type and mutant forms of InvH<sup>27-147</sup> and InvH<sup>70-147</sup>, and mutants were analyzed with a MiniDawn static light scattering unit (WYATT Technology) preceded by a Superdex 75 10/300 GL column (GE Life Sciences). The Superdex 75 column was equilibrated in 20 mM HEPES pH 7.5, 500 mM NaCl, and 100  $\mu$ L of protein was injected at 5 mg/mL with a flow rate of 0.2 mL/min. In the case of the InvH-InvG complex, the proteins were mixed in a 1:1 molar ratio and injected onto a Superdex 75 Increase 10/300 GL column (GE Life Sciences). The peak containing both InvH and InvG (assayed by SDS-PAGE) was injected onto the Superdex 75 column preceding the static light scattering unit at 5 mg/mL

(calculated using  $A_{280}$  and the extinction coefficient of a 1:1 InvH:InvG complex). The ASTRA software suite (WYATT Technology) was used to calculate the molecular masses of the protein particles based on their light scattering at each UV peak.

#### 2.2.6 Small Angle X-Ray Scattering

Experiments were performed using an in-house BioSAXS 2000 (Rigaku). InvH<sup>27-147</sup> and InvH<sup>70-147</sup> were injected onto a Superdex 75 10/300 GL column (GE Life Sciences) equilibrated in 20 mM HEPES pH 7.5, 150 mM NaCl, and the most concentrated fraction of each protein was diluted as follows: InvH<sup>27-147</sup> at 9.0, 4.5, 2.3, and 1.3 mg/mL and InvH<sup>70-147</sup> at 16.0, 8.0, 4.0, 2.0, and 1.0 mg/mL (concentration derived from  $A_{280}$  measurements from a Nanodrop 2000 (ThermoFisher Scientific) and predicted molar absorptivities). Scans were collected at 6°C at a wavelength of 1.54 Å with a  $q$  range from 0 to 0.639 Å<sup>-1</sup>. A blank scan using the size exclusion buffer was performed prior to each sample measurement as a control, and a standard of silver behenate was collected with an exposure time of 60 seconds for calibration. Six frames were collected per sample, with an exposure time of 10 minutes per frame.

Initial data processing was performed in SAXSLab 3.0 (Rigaku). Scattering intensities were averaged over the six collected frames, and the blank scan was subtracted from the averaged curve. Porod volume and molecular weight estimates ( $MW^{Vc}$ ,  $MW^{DAM}$ ) were calculated using the Automated Analysis Pipeline. Subsequent analysis was performed using the ATSAS software package. Primus was used for data analysis, merging, and Guinier analysis. Gnom was used to generate  $P(r)$  functions and to calculate radius of gyration. Dammif was used to compare envelopes generated at various protein concentrations. The most consistent envelopes were generated for InvH<sup>70-147</sup> from merging the 2 mg/mL and 16 mg/mL datasets, and for InvH<sup>27-147</sup> from merging the 4.5 mg/mL and 9 mg/mL datasets. Dammin was used to generate 10 final

envelopes for each dataset, and Damaver and Damfilt were used to create averaged and filtered envelopes respectively. Crysol was used to generate a predicted SAXS curve from the InvH homodimer crystal structure.

### 2.2.7 Isothermal Titration Calorimetry

ITC measurements were performed using a Microcal iTC200 (Malvern). Protein samples were dialyzed overnight into 2 L of 20 mM HEPES pH 7.5, 500 mM NaCl. In all experiments, InvH<sup>27-147</sup> was titrated into InvG<sup>520-562</sup>; mutant variants were always tested against wild-type binding partner. The InvH mutants probed were L123A, I133A and I133R, and InvG was mutated to W549A, Y550A, and Y553A. The concentrations of protein varied based on affinity. Lower affinity binding experiments (InvH I133R, InvG W549A and Y553A) were performed at 100  $\mu$ M InvH and 1 mM InvG, while higher affinity experiments (InvH I133A, InvG V550A, InvG R551A) were performed at 30  $\mu$ M InvH and 300  $\mu$ M InvG. In the case of InvH L123A titrated into wild-type InvG, InvH was at 20  $\mu$ M concentration while InvG was at 200  $\mu$ M. In all experiments, a volume of 200  $\mu$ L InvG<sup>520-562</sup> was titrated with twenty 2  $\mu$ L injections of InvH<sup>27-147</sup> at 25°C. The heat of dilution of the titrant was corrected for by subtracting a control (twenty 2  $\mu$ L injections of titrant into buffer). The heats of interaction were integrated, and the resulting data were fit to a one binding site model using the Origin-7 software (OriginLab Corporation) to obtain a binding stoichiometry **n**, equilibrium dissociation constant **K<sub>d</sub>**, binding enthalpy  **$\Delta$ H**, and binding entropy  **$\Delta$ S**.

## 2.3 Results

### 2.3.1 Crystallized InvH forms a homodimer

To gain a better understanding of the broad diversity in pilotin structures, we sought to crystallize heterologously expressed InvH. Secondary structure analysis by Phyre<sup>2</sup> for InvH

predicted an  $\alpha$ -helical C-terminal domain, joined by a long (~ 50 residue) disordered linker region to the signature (L-A/S-G/A-C) N-terminal lipobox<sup>231</sup>. Guided by the secondary structure predictions, we focused crystallization trials on two high-yielding soluble constructs: InvH<sup>27-147</sup>, lacking the lipobox sequence, and InvH<sup>70-147</sup>, truncated past the predicted linker region (Figure 2.1a). InvH<sup>27-147</sup> was soluble but failed to crystallize, likely due to the disordered linker region. However, using a microbatch approach under oil, InvH<sup>70-147</sup> formed large plate-like crystals. Single-wavelength anomalous diffraction (SAD) phasing with mercury acetate-soaked crystals was successful, and the final native dataset was refined at 1.15 Å resolution (Figure 2.1b, Table 2.1).

The asymmetric unit consists of four InvH<sup>70-147</sup> molecules (Figure A.1a), forming two homodimers (Figure 2.1c-d). No significant differences in backbone architecture were noted between the four protein molecules or between the two dimers, with pairwise C $\alpha$  RMSDs of ~ 0.2 Å. The homodimer architecture is compact, made up of three long  $\alpha$ -helices per subunit. Helix 1 is kinked, and its two halves termed  $\alpha$ 1a and  $\alpha$ 1b. Helix  $\alpha$ 1a (residues 73 to 88) interacts solely with the opposing protomer, burying a hydrophobic surface made up of Phe99, Phe119, Leu123, Ile133, and Leu137 (Figures 2.1a, e). Following the turn after  $\alpha$ 1a,  $\alpha$ 1b (90-101) and  $\alpha$ 2 (111-124) form a wide  $\alpha$ -helical hairpin. The 10-residue linker between  $\alpha$ 1b and  $\alpha$ 2 also contains a small four-residue  $\alpha$ -helix. A long loop separates  $\alpha$ 2 from  $\alpha$ 3 (133-146), and  $\alpha$ 3 interacts with  $\alpha$ 1b,  $\alpha$ 2, and the opposing protomer's  $\alpha$ 1a primarily through its N-terminal residues. Helix  $\alpha$ 1a resembles a domain swap between the protomers; however, due to the short linker length between  $\alpha$ 1a and  $\alpha$ 1b, it is unlikely that  $\alpha$ 1a could bind the same interface in monomeric InvH. In addition to the hydrophobic surface bound by  $\alpha$ 1a described above, the ~ 1200 Å<sup>2</sup> dimer

interface (calculated by PDBePISA<sup>232</sup>) involves polar interactions between the respective loops connecting  $\alpha 2$  to  $\alpha 3$  (Figure A.1b).

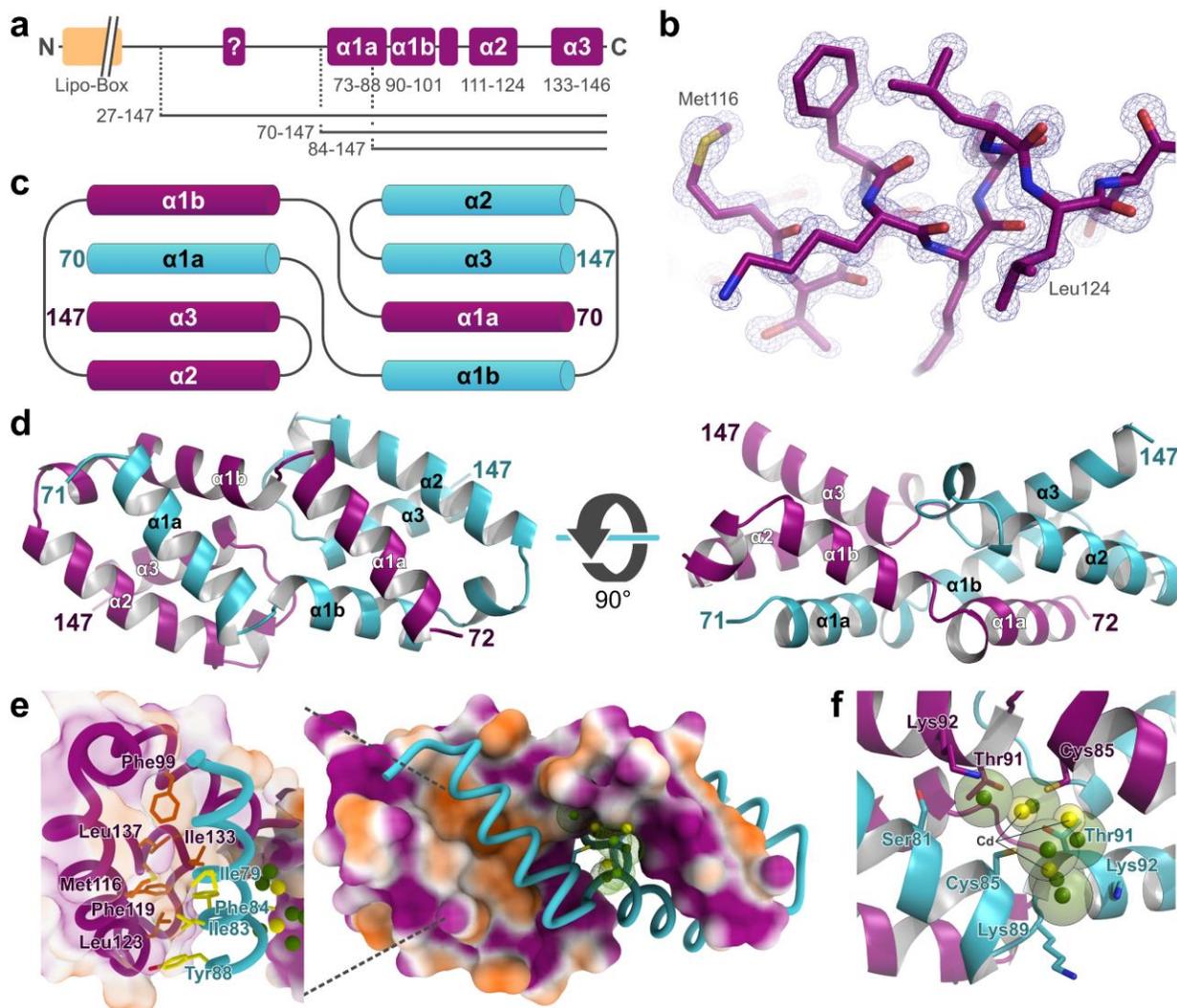


Figure 2.1: Crystal structure of the InvH<sup>70-147</sup> homodimer. (a) Illustration of the observed secondary structure of InvH, with the N-terminal linker (uncrystallized) predicted to be disordered by the Phyre2 server<sup>231</sup>. (b) A sample 1.2 Å resolution 2Fo-Fc electron density map contoured at 1.5  $\sigma$ , shown for residues 115-126. (c) A topology schematic illustrating the contacts helix  $\alpha 1a$  makes with the opposite protomer. (d) 90-degree views of the InvH<sup>70-147</sup> homodimer, with the protomers coloured in magenta and turquoise. (e) Surface depiction of one protomer from the InvH<sup>70-147</sup> dimer with high lipophilic potential (calculated by pyMLP<sup>233</sup>) coloured in orange. Interacting residues along the interface are detailed on the left. Spheres indicate positions of cadmium (yellow) and chloride (green) ions, detailed in next panel. (f) Cluster of electron-rich cadmiums (yellow) and chlorides (green) centred between Cys85 on each protomer. The van der Waals radii are illustrated as transparent spheres<sup>234</sup>.

Table 2.1: InvH X-ray crystallography data collection, refinement and validation statistics

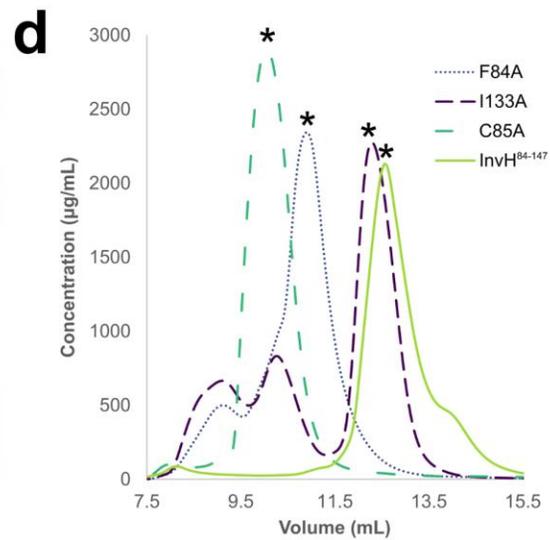
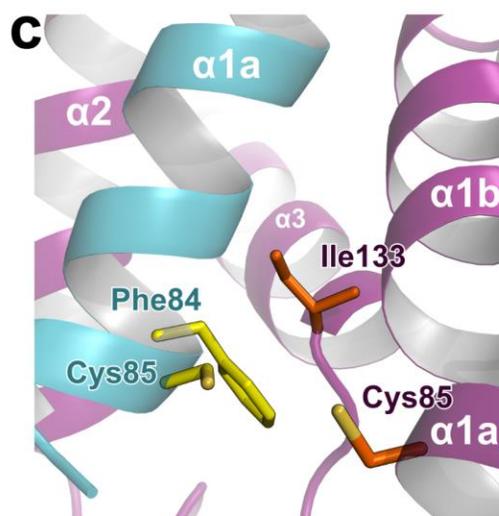
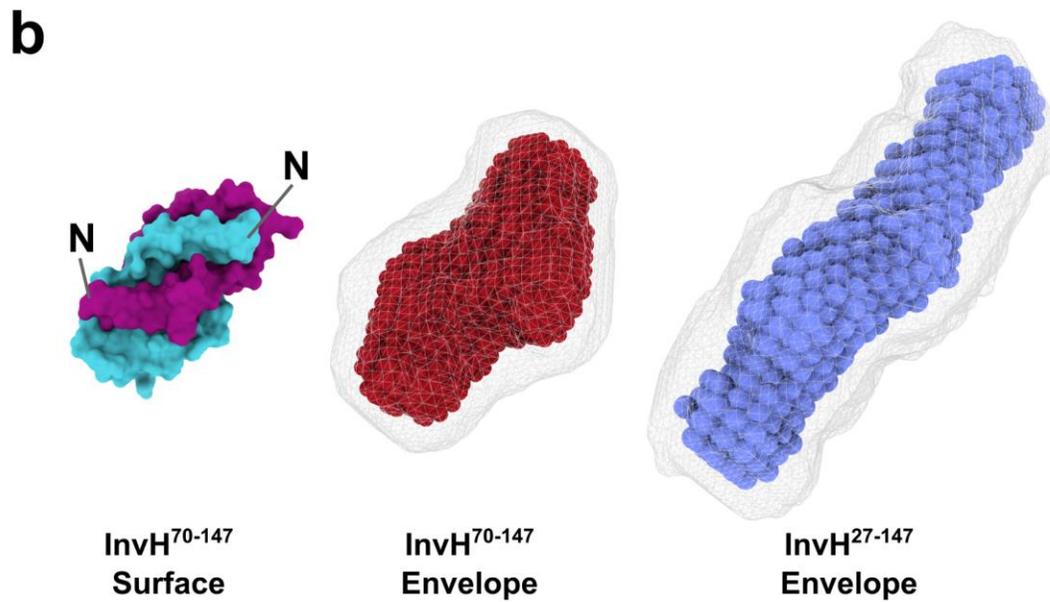
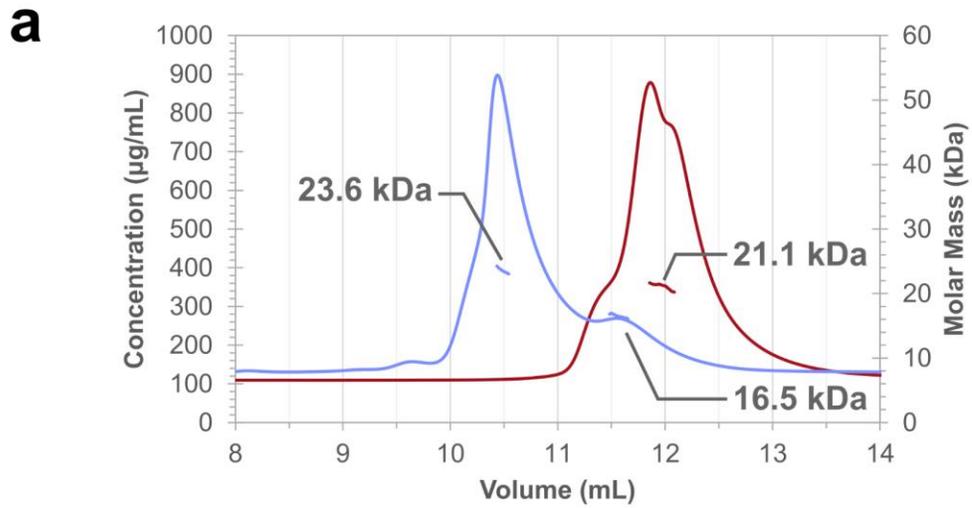
	InvH <sup>70-147</sup>	InvH <sup>70-147</sup> Hg <sup>2+</sup> Soaked	InvH <sup>84-147</sup> /InvG <sup>543-558</sup>
<b>Data Collection</b>			
Wavelength (Å)	0.98	1.01	1.00
Resolution range (Å)	44.33 - 1.20 (1.24 - 1.20)	44.28 - 1.37 (1.42 - 1.37)	31.45 - 1.85 (1.92 - 1.85)
Space group	P 1	P 1	P 65 2 2
Unit cell parameters (Å)	a=25.5 b=53.2 c=56.9 α=107.7° β=97.2° γ=96.0°	a=25.4 b=53.2 c=56.9 α=107.4° β=97.2° γ=96.1°	a = 36.3 = 36.3 c = 221.6 α = β = 90° γ = 120°
Total reflections	325986	418857	135602
Unique reflections	82540	36439	8226
Multiplicity	3.9 (3.9)	11.5 (9.4)	16.5 (17.1)
Completeness (%)	94.5 (91.2)	61.6 (4.7)	99.5 (99.8)
Mean I/σI	11.1 (2.0)	27.7 (1.8)	16.8 (2.5)
Wilson B-factor (Å <sup>2</sup> )	7.7	12.1	41.4
R <sub>merge</sub>	0.083 (0.765)	0.061 (0.948)	0.069 (0.885)
R <sub>meas</sub>	0.097 (0.886)	0.063 (1.002)	0.071 (0.912)
R <sub>pim</sub>	0.049 (0.446)	0.019 (0.315)	0.018 (0.218)
CC <sub>1/2</sub>	0.998 (0.681)	0.999 (0.713)	1.000 (0.908)
CC*	0.999 (0.900)	1.000 (0.912)	1.000 (0.976)
<b>Refinement</b>			
Reflections used in refinement	82540		8225
Reflections used for R <sub>free</sub>	4111		420
R <sub>work</sub> /R <sub>free</sub>	13.2 / 16.8		21.8 / 24.7
CC <sub>work</sub> /CC <sub>free</sub> (%)	96.8 / 95.4		90.6 / 91.4
Number of non-hydrogen atoms	3048		708
macromolecules	2605		629
ligands	38		10
solvent	405		69
Protein residues	306		74
RMS(bonds, Å <sup>2</sup> )	0.016		0.019
RMS(angles, °)	1.86		1.88
Ramachandran favored (%)	99.7		100.0
Ramachandran allowed (%)	0.3		0.0
Ramachandran outliers (%)	0.0		0.0
Rotamer outliers (%)	0.3		0.0
Clashscore	1.9		0.8
Average B-factor (Å <sup>2</sup> )	17.0		48.3
macromolecules	14.8		46.3
ligands	21.0		59.8
solvent	31.0		64.9
Number of TLS groups	None		2

Several highly electron-dense species are observed at protein-protein contacts in the InvH crystal. We initially predicted the densities were cadmium ions, as the crystallization solution included 80 mM cadmium chloride ( $\text{CdCl}_2$ ). Creating an anomalous difference map revealed several high-intensity peaks ( $>5 \sigma$ ), each bordered by two non-anomalous electron dense species (Figure A.1c). Based on this pattern, it is most likely that the anomalous scatterers are cadmiums, neighbored by two chlorides. Although several cadmium chloride groups are present throughout the asymmetric unit, the most notable is a cluster of three bridging the dimerization interface. These cadmiums are coordinated primarily by Cys85 from each protomer, but also contact Lys89, Thr91, and Lys92 (Figure 2.1f). One cadmium is localized at the kink between helices  $\alpha 1a$  and  $\alpha 1b$  by coordination with Gln90 and Glu93 (Figure A.1d). Crystal contacts between Asp111 and Glu113 are also bridged by cadmium. To determine whether purified InvH binds ions at the discussed coordination sites, we performed inductively coupled plasma mass spectrometry (ICP-MS) on purified InvH samples. No significant metal species were found out of the elements tested (cadmium, copper, manganese, nickel, lead, and zinc; data not shown). Therefore, the  $\text{CdCl}_2$  was found to aid InvH crystal formation, but is not likely to represent a physiologically relevant bound species.

### 2.3.2 InvH forms the crystallographically-observed homodimer in solution

No previously characterized pilotins have been shown to form a homodimer in solution. Thus, size-exclusion coupled multi-angle light scattering (SEC-MALS) experiments were performed to investigate the oligomerization states of InvH<sup>27-147</sup> and InvH<sup>70-147</sup> in solution. The molecular masses observed for both constructs approximately corresponded to those of the homodimer, with a small additional monomer population present for InvH<sup>27-147</sup> (Figure 2.2a).

To further characterize InvH in solution, in house small-angle X-ray scattering (SAXS) was performed on the same two constructs (Figure A.2a-c). Consistent with the SEC-MALS result, SAXS analysis of both constructs resulted in calculated molecular masses indicative of homodimers (Table A.1). The SAXS data could be fit to obtain consistent envelopes for both InvH<sup>70-147</sup> and InvH<sup>27-147</sup>, and averages for each construct were generated from ten envelopes (Figure 2.2b). In the case of InvH<sup>70-147</sup>, the size and shape of the envelope broadly matched the shape of the crystallized dimer, though the SAXS envelope is slightly wider and flatter (Figure 2.2b). Using the software Crysol to generate a simulated SAXS curve from the InvH dimer structure resulted in a relatively poor fit to the experimental SAXS data ( $\chi^2 = 17.5$ ; we attribute to dynamic motion as discussed further below; Figure A.2d). In the case of the InvH<sup>27-147</sup> construct, which includes the predicted disordered linker region, fitting of the SAXS data produced an envelope with additional scattering density extending from both poles of the InvH<sup>70-147</sup> structure (Figure 2.2b). These poles correspond to the locations of the N-termini of the InvH<sup>70-147</sup> crystal structure where the linker residues would be joined. Collectively, the SEC-MALS and SAXS data confirm the formation of an InvH homodimer in solution, with the SAXS data additionally supporting the position of the InvH<sup>27-147</sup> N-terminal linker region.



**Figure 2.2: Evidence for InvH homodimer formation in solution. (a)** SEC-MALS traces of InvH<sup>27-147</sup> (blue) and InvH<sup>70-147</sup> (red), with concentration (left Y axis) estimated based on A<sub>280</sub> and the molar absorptivities of each construct. Horizontal lines beneath each peak correspond to the calculated molecular weights, plotted on the secondary Y axis (right). In each case, the estimated molecular mass of the predominant population is consistent with that of a homodimer (molecular masses of monomeric InvH<sup>27-147</sup> and InvH<sup>70-147</sup> are 14 kDa and 9 kDa, respectively). **(b)** Comparison of the surface representation of the InvH<sup>70-147</sup> homodimer with envelopes generated from SAXS data. The crystal surface represents the solvent-excluded surface as calculated by ChimeraX, using a probe radius of 1.4 Å, approximating a water molecule<sup>221</sup>. The outer (grey) mesh of each SAXS model represents the average of ten envelopes, while the inner (coloured) envelope is a filtered average (cut-off at the mean envelope volume). **(c)** Stick representation of residues chosen to be mutated for InvH dimerization assays. **(d)** Overlaid S75 size-exclusion traces of InvH<sup>84-147</sup> and three InvH<sup>27-147</sup> mutants, with concentration calculated based on extinction coefficients of each. Predominant peaks (\*) had their molecular masses confirmed by SEC-MALS (Figure A.2d). The marked peaks of InvH<sup>84-147</sup> and InvH<sup>27-147</sup> mutant C85A are dimeric, whereas those of the F84A and I133A mutants are monomeric. Note that the molecular mass of dimeric InvH<sup>84-147</sup> is similar to that of monomeric InvH<sup>27-147</sup>.

---

Several InvH residues were mutated to probe whether the crystallographic interface reflects the dimer conformation in solution. The mutants were tested within the context of the InvH<sup>27-147</sup> construct. Phe84 and Ile133 were individually mutated to alanine based on their interactions with the opposite protomer. In addition, Cys85 was chosen because of its coordination with cadmium in the crystal (Figure 2.2c). Each of these InvH<sup>27-147</sup> mutants were purified on a Superdex 75 size exclusion column, and the predominant elution peak collected and assayed by SEC-MALS. The I133A mutant had the strongest effect, shifting most of the InvH population to the monomeric state (Figure 2.2d, Figure A.3a). The F84A mutation was deleterious to dimerization, with both monomeric and, to a lesser extent, dimeric species detected. Both mutants had minor high-molecular mass populations, attributed to aggregation. The C85A mutant also eluted as a homodimer peak, confirming that cadmium coordination was not important for self-association outside of the crystal condition (Figure 2.2d, Figure A.3a). Unexpectedly, an additional construct investigated in this series of experiments, InvH<sup>84-147</sup>, was still predominantly a homodimer

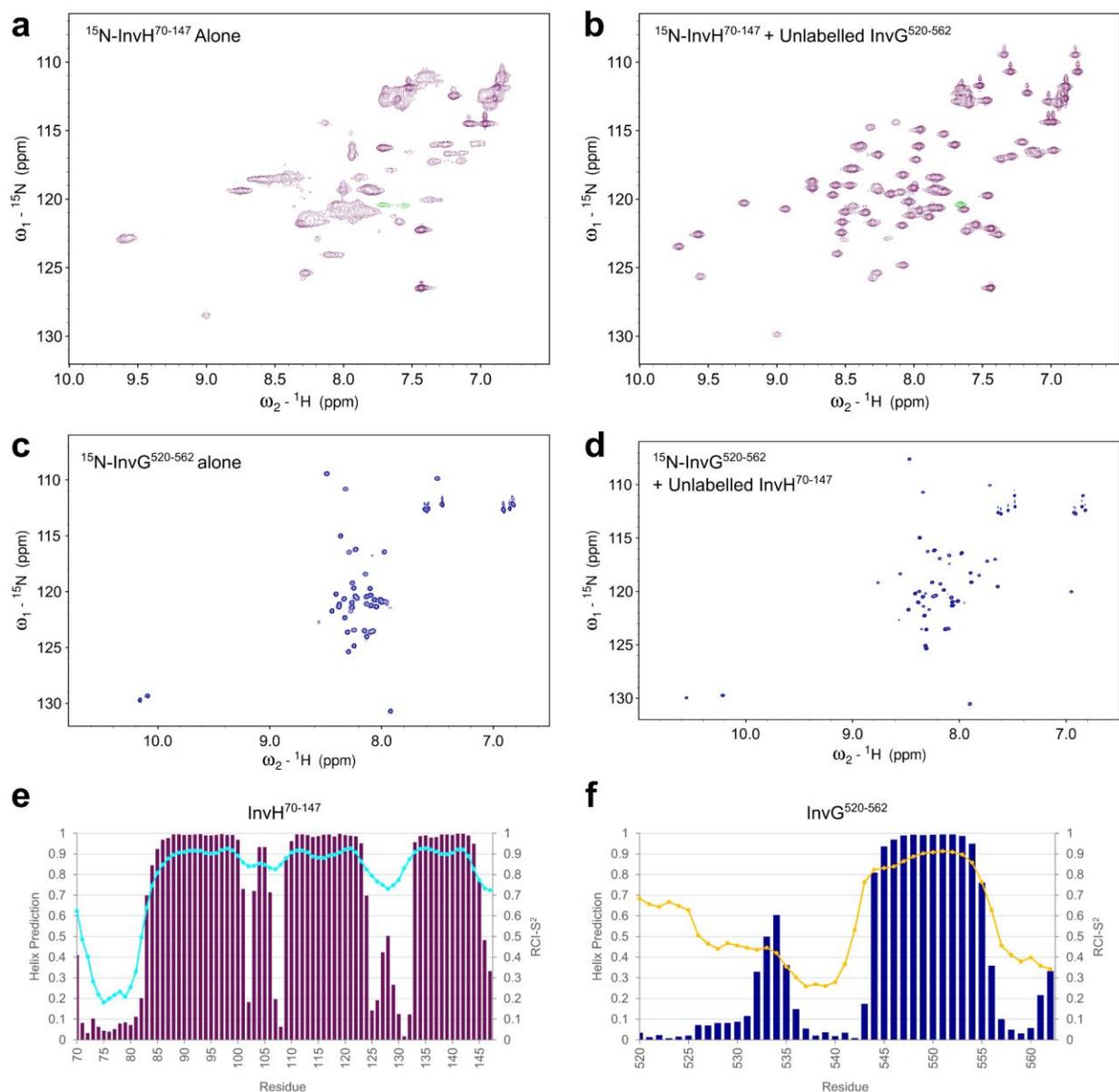
despite truncation of the protomer-bridging helix  $\alpha$ 1a. These data suggest that residues 70-83 are not essential for dimerization, whereas Phe84 and Ile133 have a role in dimer formation.

### 2.3.3 NMR spectroscopic studies on the InvH-InvG complex reveal ordered and dynamic regions

The importance of InvH in injectisome assembly is hypothesized to lie in its ability to bind the secretin InvG and guide it to the outer membrane in a monomeric state, preventing premature oligomerization at the inner membrane. To probe the binding mechanics of InvH and the InvG S domain, NMR spectroscopy was used to characterize the two proteins separated and in complex. An N-terminally His-tagged InvG<sup>520-562</sup> construct was chosen based on the cryo-EM structure of assembled InvG<sup>189</sup>. This fragment encompasses the entirety of the pilotin-binding S domain helix-loop-helix motif.

The InvG<sup>520-562</sup> construct was shown previously to bind to InvH<sup>27-147</sup> with high affinity<sup>189</sup>. As a step towards its structural characterization, the binding and oligomerization state of InvG<sup>520-562</sup> with InvH<sup>27-147</sup> and InvH<sup>70-147</sup> was analyzed using SEC-MALS. Both InvH constructs formed a complex with InvG with a molecular mass corresponding to that of a 1:1 heterodimer (Figure A.3b). This result indicated that the InvH homodimers, which would otherwise form under these conditions, were disrupted upon addition of the InvG S domain.

Strikingly, both <sup>15</sup>N-labelled InvH constructs yielded unassignable <sup>15</sup>N-HSQC spectra with only a small number of broad, overlapping <sup>1</sup>H<sup>N</sup>-<sup>15</sup>N signals (Figure 2.3a, Figure A.4a). Although uniformly broad signals were expected for dimeric proteins in this size range, such poor-quality spectra indicated that InvH<sup>27-147</sup> and InvH<sup>70-147</sup> were undergoing conformational exchange on a msec- $\mu$ sec timescale that lead to severe signal broadening<sup>235,236</sup>. The exchange could be between the monomer and homodimer states or within the predominant homodimeric state.



**Figure 2.3:** NMR spectroscopic data of InvH<sup>70-147</sup> and InvG<sup>520-562</sup>. **(a)** The <sup>15</sup>N-HSQC spectrum of <sup>15</sup>N-InvH<sup>70-147</sup> has broad overlapping signals indicative of conformational exchange broadening. **(b)** Addition of a 1.2 molar excess of unlabelled InvG<sup>520-562</sup> resulted in a high quality <sup>15</sup>N-HSQC spectrum with well-dispersed signals for InvH<sup>70-147</sup>. Negative peaks (green) in (a) and (b) are aliased signals from arginine sidechains. **(c)** The <sup>15</sup>N-HSQC spectrum of <sup>15</sup>N-InvG<sup>520-562</sup> has well-dispersed signals but a narrow <sup>1</sup>H<sup>N</sup> chemical shift range, indicative of disorder. **(d)** The <sup>15</sup>N-HSQC spectrum of <sup>15</sup>N-InvG<sup>520-562</sup> in complex with a 1.2 molar excess of unlabelled InvH<sup>70-147</sup> is also well-dispersed and indicative of an ordered conformation. Assignments for the latter two spectra are provided in Figure A.4c-d. The chemical shift-derived helical secondary structure predictions (bars, left axis) and RCI-S<sup>2</sup> values (lines, right axis) are plotted for **(d)** InvH<sup>70-147</sup> and **(e)** InvG<sup>520-562</sup> in the presence of their unlabelled partner.

In contrast to its NMR spectroscopic features when separated, isotopically labeled InvH<sup>70-147</sup> yielded high quality <sup>15</sup>N-HSQC spectra with sharp dispersed signals when combined with its unlabelled InvG<sup>520-562</sup> partner (Figure 2.3b). Thus InvH<sup>70-147</sup> formed a stable, well-ordered complex with InvG<sup>520-562</sup> that was amenable for further NMR spectroscopic characterization. The <sup>15</sup>N-HSQC spectrum from InvH<sup>27-147</sup> also showed increased dispersion upon binding InvG<sup>520-562</sup> (Figure A.4a-b). Furthermore, additional amides not present in the InvH<sup>70-147</sup> complex have <sup>1</sup>N<sup>H</sup> chemical shifts in the narrow range of 8.0 - 8.5 ppm. Consistent with sequence-based predictions<sup>231,237,238</sup>, these chemical shifts indicate that the InvH linker is conformationally disordered<sup>239</sup>. However, its high molecular mass, the presence of additional signals from the linker residues, and possible exchange broadening still precluded detailed analysis of InvH<sup>27-147</sup>.

The <sup>15</sup>N-HSQC spectrum of <sup>15</sup>N-labelled InvG<sup>520-562</sup> contained sharp signals with limited <sup>1</sup>H<sup>N</sup> dispersion (Figure 2.3c). This is diagnostic of a conformationally disordered polypeptide<sup>239</sup>. Upon addition of unlabelled InvH<sup>70-147</sup>, many well-dispersed <sup>1</sup>H<sup>N</sup>-<sup>15</sup>N signals were observed in its <sup>15</sup>N-HSQC spectrum (Figure 2.3d). Thus, InvG<sup>520-562</sup> also adopted an ordered conformation when bound to its cognate pilotin.

Using standard heteronuclear correlation experiments, the <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N signals from both protein components of the InvH<sup>70-147</sup>/InvG<sup>520-562</sup> complex were assigned (Figure A.4c-d). Using these chemical shifts, the secondary structure and mainchain dynamics of the two proteins were predicted with Talos-N<sup>228</sup>. The latter were reflected by the random coil index squared order parameter (RCI-S<sup>2</sup>), with decreasing values corresponding to increasing conformational flexibility<sup>240</sup>. Similar to the X-ray crystallographic homodimer, InvH<sup>70-147</sup> residues 109-124 ( $\alpha$ 2) and 133-145 ( $\alpha$ 3) are also helical when bound to InvG<sup>520-562</sup> (Figures 2.1a and 2.3e). However, whereas the crystallized homodimer has two helices  $\alpha$ 1a and  $\alpha$ 1b (residues 73-88 and 90-101

respectively), the heterodimer in solution has one extended helix spanning residues 83-101 and residues 70-82 are disordered. The NMR-based predicted secondary structure of InvH-bound InvG<sup>520-562</sup> also diverged from the observed helix-loop-helix motif in the near-atomic resolution cryo-EM structure of the assembled secretin pore<sup>189</sup> (Figure 2.3f). In contrast to the latter, only residues 543-555 have  $\alpha$ -helical character, corresponding to the C-terminal helix of the S domain helix-loop-helix motif. The remainder of the protein fragment is disordered, perhaps not surprising in the absence of the  $\beta$ -barrel domain against which the S domain is intimately packed in the assembled secretin<sup>189</sup>.

#### 2.3.4 Structural studies of the InvH-InvG complex by NMR spectroscopy and X-ray crystallography

The structure of InvH with the InvG S domain was characterized with both NMR spectroscopy and X-ray crystallography. Using an extensive set of NOE-derived distance and chemical shift-derived dihedral angle restraints, the structural ensemble of InvH<sup>70-147</sup>-InvG<sup>520-562</sup> was determined (Table 2.2). Of particular importance, filtered-edited NOESY spectra were used to identify interproton NOEs between labeled and unlabeled members of the heterodimer.

Table 2.2: InvH-InvG NMR spectroscopy data collection and refinement statistics

<b>NMR Distance and Dihedral Restraints</b>	
Distance restraints	
Total nuclear Overhauser effect	1,670
Intraresidue	466
Sequential ( $ i - j  = 1$ )	553
Medium range ( $ i - j  < 4$ )	378
Long range ( $ i - j  > 5$ )	359
Dihedral angle restraints	
$\Phi$	336
$\Psi$	336
Hydrogen bond restrains	62
<b>Structure Statistics</b>	
Violations (mean $\pm$ SD)	
Distance restraints ( $\text{\AA}$ )	0.013 $\pm$ 0.001
Dihedral angle restraints ( $^\circ$ )	0.97 $\pm$ 0.07
Max. distance restraint violation ( $\text{\AA}$ )	0.014
Max. dihedral angle violation ( $^\circ$ )	1.05
Ramachandran plot (%)	
Most favored	79.0
Additionally allowed	20.7
Generously allowed	0.4
Disallowed	0.0
Pairwise root-mean-square deviations (mean $\pm$ SD, $\text{\AA}$ )	
Backbone atoms	0.25 $\pm$ 0.07
Heavy atoms	0.8 $\pm$ 0.1

**Pairwise root-mean-square deviation calculated among the 20 refined structures for residues 85–145, 544–555.**

The structural ensemble revealed that residues 543-556 of the InvG S domain fold as a single  $\alpha$ -helix (Figure 2.4a-b), whereas the remaining residues are conformationally disordered. The InvG helix binds along the InvH hydrophobic cleft that also forms most of its homodimerization interface observed by X-ray crystallography. Most notably, part of InvH helix  $\alpha$ 1a is displaced from the hydrophobic binding pocket by InvG and becomes disordered in solution in the absence of any stabilizing interface. Although skewed slightly (C $\alpha$  RMSD of  $\sim 1.6$  Å over 56 residues), the overall architecture of the remaining InvH helices  $\alpha$ 1b through  $\alpha$ 3 is similar to that observed crystallographically for the InvH<sup>70-147</sup> homodimer (Figure 2.1d and 2.4a). However, helix  $\alpha$ 1b is elongated and thus renamed as helix  $\alpha$ 1 (residues 84-101 rather than 90-101). These results are consistent with helical propensities and mainchain RCI-S2 values from initial chemical shift analyses (Figure 2.3e-f).

Residues 525-537 of oligomerized InvG form a helix in the cryo-EM structure of the assembled secretin pore<sup>189</sup>, where they pack against the adjacent InvG protomer  $\beta$ -barrel. In contrast, these residues are disordered in the 1:1 InvH<sup>70-147</sup>-InvG<sup>520-562</sup> complex and do not interact with InvH. This result is in line with previous data showing that S domains are disordered in the absence of pilotin binding, and likely only fully ordered upon final secretin assembly<sup>189,241,242</sup>. Based on such insights from NMR spectroscopy, crystallization trials were undertaken with a synthetic peptide of InvG<sup>543-558</sup> bound to InvH<sup>84-147</sup>. These species lacked disordered residues, including 70-82 of InvH that would otherwise form part of helix  $\alpha$ 1a in the homodimer.

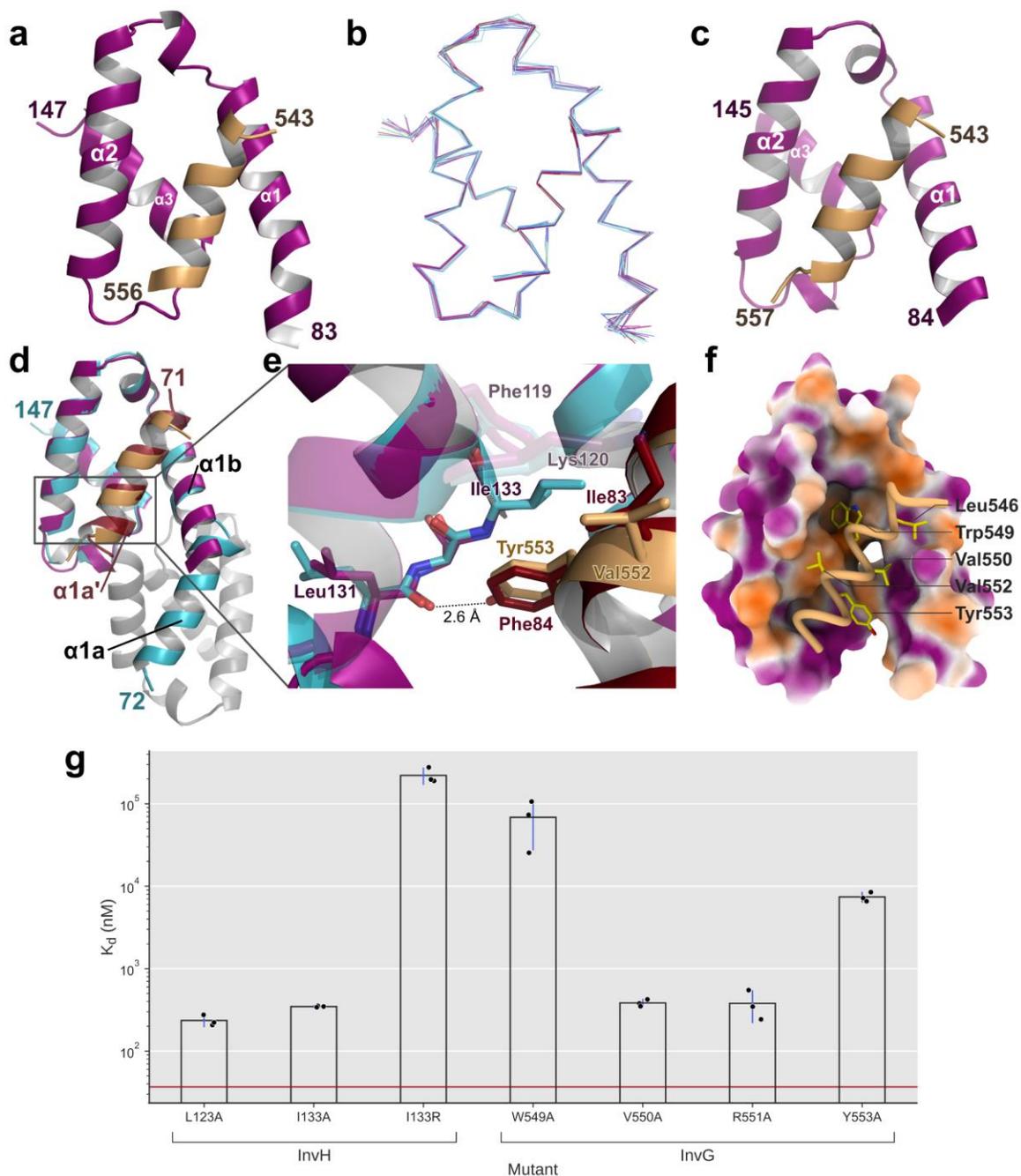


Figure 2.4: Structural characterization of the InvH-InvG binding interface. **(a,b)** Cartoon representation and superimposed 20-member structural ensemble of InvH<sup>70-147</sup>-InvG<sup>520-562</sup> determined by NMR spectroscopy (InvH magenta; InvG tan; disordered regions omitted). **(c)** Cartoon representation of InvH<sup>84-147</sup>-InvG<sup>543-558</sup> determined by X-ray crystallography. InvH is coloured magenta and InvG is coloured tan. **(d)** Overlay of the InvH<sup>84-147</sup>-InvG<sup>543-558</sup> heterodimer on the InvH<sup>70-147</sup> homodimer (turquoise and grey, with  $\alpha 1a'$  red). **(e)** Zoomed view of the overlay from part (d), demonstrating the close alignment between InvH helix  $\alpha 1a$  Phe84 (red) and InvG Tyr553 (tan). **(f)** Surface representation of the hydrophobic cleft on InvH<sup>70-147</sup> from the NMR-

derived structure (hydrophobic regions coloured bright orange), with bound InvG (residues 543-556, tan cartoon). Interacting residues of InvG are portrayed as sticks. **(g)** Summary of binding data from ITC experiments performed on InvH<sup>27-147</sup> and InvG<sup>520-562</sup> mutants. The  $K_d$  value of each mutant with its wild-type binding partner is plotted on a logarithmic y axis. Each point represents the  $K_d$  value calculated from one experiment, with bars and vertical blue lines marking the average and standard deviation respectively. The horizontal red line marks the previously characterized  $K_d$  (40 nM) for the two wild-type proteins<sup>189</sup>.

---

The InvH<sup>84-147</sup>-InvG<sup>543-558</sup> complex readily crystallized in hanging drop vapour diffusion, forming elongated hexagonal bipyramids. The crystals diffracted to 2.0 Å resolution, and the structure was solved by molecular replacement using a model from the InvH homodimer (Table 1). The resulting structure (Figure 2.4c) aligns well with the InvH<sup>70-147</sup>-InvG<sup>520-562</sup> ensemble (Supplemental Figure A.4c).

In the InvH<sup>84-147</sup>-InvG<sup>543-558</sup> heterodimer, the conformation of residues 90-147 is very similar to that observed in the InvH homodimer, with a  $C\alpha$  RMSD of 0.43 Å over 49 residues (chain A) (Figure 2.4c-d). As in the NMR-derived structural ensemble of InvH<sup>70-147</sup>-InvG<sup>520-562</sup>, helix  $\alpha 1b$  is elongated to residue 84, and renamed  $\alpha 1$  (Figure A.5a). Similarly, the InvG S domain helix (residues 544-555) takes the corresponding place of helix  $\alpha 1a'$  from the other protomer of the homodimer structure (Figure 2.4d). At a more detailed level, overlaying the crystal structures of the InvH<sup>70-147</sup> homodimer with the InvH<sup>84-147</sup>-InvG<sup>543-558</sup> heterodimer reveals differences in the interface of InvH with its partner protomer versus with InvG. The InvG helix is bound at a slightly different angle, causing its N-terminus to be about 3 Å further from the interface (Figure 2.4d). However, the C-terminus of the InvG helix aligns closely with InvH helix  $\alpha 1a'$ . Notably, InvG residues Val552 and Tyr553 are well aligned with InvH residues Ile83' and Phe84' respectively, at the start of  $\alpha 1a'$  ( $C\alpha$  distances of 0.8 and 0.7 Å respectively) (Figure 2.4e). Tyr553 is embedded in the same hydrophobic cleft as Phe84 in the homodimer, surrounded by

InvH residues Phe119, Leu123, Ser132, and Ile133. The common interface for InvH self-association and InvG-binding provides a simple structural explanation for the mutually exclusive formation of the InvH-InvH homodimer versus the InvH-InvG heterodimer.

### 2.3.5 Aromatic and small hydrophobic residues contribute to InvH-InvG binding

InvG and InvH bind through primarily hydrophobic interfaces on both proteins (Figure 2.4f). The InvG helix lies anti-parallel to InvH helix  $\alpha_1$ , and interacts mainly with helix  $\alpha_1$ ,  $\alpha_2$ , and the N-terminal half of  $\alpha_3$ . Two aromatic residues on InvG, Trp549 and Tyr553, are embedded in the binding interface and interact with InvH residues Ala95, Phe98, Glu112, Met116, Leu123, Ser132, Ile133, and Leu137 (Figure 2.4f, Figure A.5b-d). Tyr553 also forms a hydrogen bond between its phenolic hydroxyl and the backbone carbonyl of Leu131 (Figure 2.4e). Other InvG hydrophobic residues also line the interface, including Leu546, Val550, Val552, and Leu 554. The opposite solvent-exposed face of the amphipathic InvG helix is made up of polar and charged residues, some of which may be positioned close enough to also contribute to binding.

The importance of these residues for heterodimer formation was tested via mutagenesis. Specifically, Leu123 and Ile133 from the InvH hydrophobic groove were chosen for mutation to alanines and, in the case of Ile133, also to an arginine. The binding interface of the InvG helix is lined primarily with aromatic and hydrophobic sidechains, and thus Trp549, Val550, and Tyr553, were also chosen for mutation to alanine. One charged residue near the interface, Arg551, was mutated to an alanine. ITC experiments were conducted on each of these constructs to assay their binding affinity to their partner wild-type protein (Figure 2.4g, Figure A.5e-f).

Our previous ITC studies revealed that the binding interaction between wild-type InvH<sup>27-147</sup> and InvG<sup>520-562</sup> occurs in a 1:1 ratio, with a high affinity  $K_d$  value of 40 nM<sup>189</sup>. As InvG binding also involves dissociation of the InvH homodimer, the heats measured by ITC and the fit  $K_d$

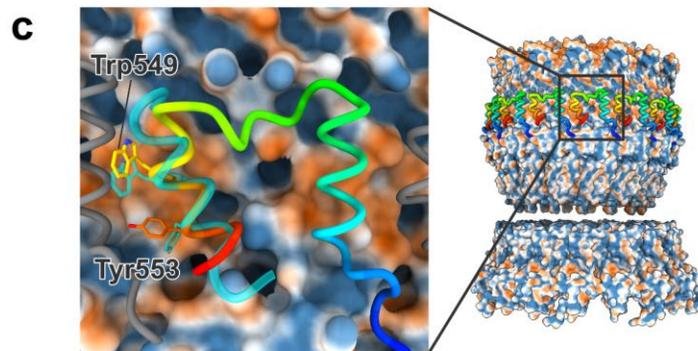
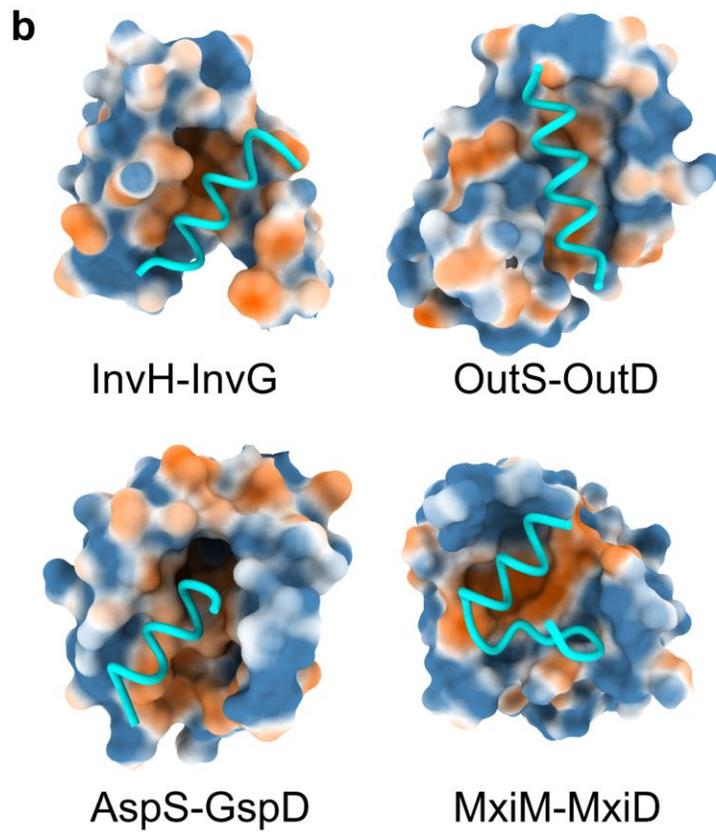
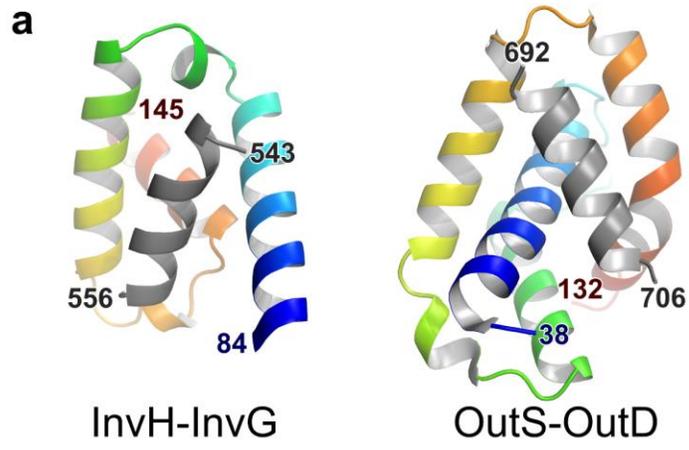
values reflect the contributions of both steps. Of the mutants assayed (Supplemental Table S2), the greatest decrease in binding affinity was from the InvH I133R mutant, with a 2000-fold increase in  $K_d$  value relative to the wild-type protein (Figure 2.4g, Figure A.5e-f). This decreased affinity is expected, as the mutation was intentionally deleterious and placed a positively charged residue along the hydrophobic interface. The I133A and L123A mutations had more modest effects, causing 10-fold and 5-fold increased  $K_d$  values respectively. As noted above, the I133A substitution also strongly destabilized InvH homodimer formation (Figure 2.2d, Figure A.3a). From the InvG<sup>520-562</sup> residues, the W549A mutation had the largest impact, increasing the  $K_d$  value by over 2000-fold. Following this, mutation of Tyr553 to Ala reduced affinity 200-fold, whereas substitutions of Val550 or of Arg551 had the smallest changes, causing modest 10-fold increases in  $K_d$ . Collectively, this mutational analysis is consistent with the description of the InvH-InvG interface derived from the structural ensemble of InvH<sup>70-147</sup>-InvG<sup>520-562</sup> and the crystal structure of InvH<sup>84-147</sup>-InvG<sup>543-558</sup>.

## 2.4 Discussion

Pilotins are a surprisingly diverse family of proteins, despite their common role in OM targeting of the highly conserved secretin pore. Secretin localization mechanisms have previously been proposed based on structures of individual pilotin-secretin complexes. However, it appears that no single mechanism can account for the range of distinct tertiary folds exhibited by the pilotins. The structural characterization of InvH provides insights into the first  $\alpha$ -helical T3SS pilotin. Notably, InvH has the distinguishing ability to form a stable homodimer in solution. From these data, we can build upon our existing pilotin knowledge to highlight essential functional elements shared between species and systems.

In terms of secondary structure, InvH most resembles the  $\alpha$ -helical T2SS pilotins OutS, PulS, and GspS<sup>206–209</sup>. Nevertheless, it does not share any sequence identity, nor does it align structurally with these proteins. InvH adopts a superficially similar helical hairpin fold to OutS and PulS, but the positions of its N- and C-termini are inverted (Figure 2.5a). Any similarities are likely a product of convergent evolution, considering these proteins all bind a short amphipathic  $\alpha$ -helix. In addition, InvH is ultimately a three-helix protein when bound to the InvG S domain (due to  $\alpha$ 1a unfolding), whereas OutS and PulS each have four  $\alpha$ -helices. The other characterized pilotins are markedly different from InvH, with primarily  $\beta$ -sheet secondary structure in the case of AspS, EspS, ExsB, and MxiM, and with type IV pilotins folding into tetratricopeptide repeats.

The binding interfaces between pilotins and their cognate secretin S domains are consistently hydrophobic among different species in type II and type III secretion systems (Figure 2.5b). The S domain forms a helix-loop-helix motif on the assembled secretin<sup>189,192,193</sup>, with the C-terminal of the two helices contributing the majority of the pilotin-binding surface in the cases of MxiD, OutD, and GspD<sup>194,205–207</sup>. Our NMR-derived structural ensemble of the InvH<sup>70–147</sup>/InvG<sup>520–562</sup> complex shows the same trend, with the N-terminal of the two S domain helices disordered and no intermolecular NOEs detectable to indicate any persistent interactions with InvH. Furthermore, all structurally characterized C-terminal S domain helices are amphipathic, and able to form an extensive hydrophobic interface with their pilotin binding partner<sup>194,205–207</sup> to mediate high affinity binding with  $K_d$  values of less than 100 nM<sup>189,194,205,242</sup>. All are lined to some extent with aromatic residues. In the case of AspS, interaction with GspD is abolished by the S domain mutation F639A, which parallels the near-loss in InvH-InvG binding affinity with the W549A substitution<sup>194</sup>.



**Figure 2.5: Structural comparison of InvH with other pilotins and secretin interfaces. (a)** Cartoon representation of InvH and OutS (coloured rainbow, blue to red from N- to C-terminus) with the bound S domain helix of InvG and OutD, respectively (grey). **(b)** Surface representation of hydrophobic cleft (hydrophobic surface orange; hydrophilic blue; calculated by pyMLP<sup>237</sup>) used for S domain (cyan) binding among four characterized pilotins. S domain helix shown in cyan. **(c)** Lipophilicity surface of the InvG  $\beta$ -barrel cryo-EM structure with its bound S domain represented as a rainbow tube, with stick views of Trp549 and Tyr553. The S domain from the InvH-InvG crystal structure is overlaid in cyan.

---

In cryo-EM structures of oligomerized secretin pores, the S domain acts as a molecular “clamp”. This helix-loop-helix motif packs against the  $\beta$ -barrel of the  $n+2$  adjacent protomer, an interdigitation believed critical to the hallmark stability of secretin pores<sup>189,192,193</sup>. The 15-fold packing of the S domain with the InvG  $\beta$ -barrel is mainly hydrophobic (Figure 2.5c), but the residues involved differ compared to the interaction with pilotin InvH. Where the aromatic residues Trp549 and Tyr553 of InvG are intimately bound to InvH, they form peripheral interactions against the InvG  $\beta$ -barrel in the assembled pore (Figure A.6b). The main interface in oligomerized InvG involves the smaller hydrophobic valines and leucines, with some contribution from polar and charged residues such as Arg551. These differences in binding interfaces may facilitate transfer of the C-terminal S domain helix from the chaperone InvH-bound state to its interface against the InvG  $\beta$ -barrel.

The homodimerization interface of InvH observed crystallographically and supported by biophysical assays distinguishes it from previously-characterized pilotins. Though *in vivo* experiments are required to confirm the importance of the InvH homodimer on *Salmonella enterica* injectisome assembly, its structure hints at a potential function. Dimerization may protect the extensive hydrophobic surface that InvH otherwise uses to bind the InvG S domain. The presence of larger aggregate peaks in size exclusion chromatography in some dimerization-deficient mutants supports a stabilizing role for the homodimer (Figure 2.2d). Similar binding-

site protecting mechanisms can be observed in other known pilotins. In the absence of the T2SS secretin GspD S domain, the  $\beta$ -sheet of pilotin AspS curves to occlude many hydrophobic side chains that would otherwise be exposed to an aqueous environment (Figure A.6a)<sup>194,202</sup>. In the case of T3SS pilotin MxiM, the protein crystallized with a lipid bound in its central solvent occluded hydrophobic channel. This interaction led to the hypothesis that it can bind its own lipidated N-terminus and protect the secretin binding site until it encounters its partner MxiD<sup>204,205</sup>. A *Parabulkoheria* pilotin was shown to be similar to MxiM, with the one key addition of an N-terminal helix that occludes the hydrophobic cavity for S domain binding from solvent<sup>243</sup>. Exceptions to this observation are found in the structures of GspS, OutS, and PulS, which have an exposed hydrophobic pocket, albeit a smaller one than what would be exposed on an InvH monomer<sup>206–209</sup>.

Our biophysical data highlight the dynamic nature of the InvH dimer in solution. While the generated SAXS envelope and molecular weight calculations of InvH<sup>70-147</sup> in solution correspond to the size of the crystallized dimer, the Crysol-generated SAXS curve of the InvH<sup>70-147</sup> dimer does not fit well to the data (Figure 2.2b, Figure A.2d). It is possible that the discrepancy arises from dynamicity in the InvH dimer that is not reflected in the crystallized state. Further support for the plasticity of the InvH dimer is seen in the severe exchange broadening of <sup>1</sup>H<sup>N</sup>-<sup>15</sup>N signals in the <sup>15</sup>N-HSQC spectra of InvH<sup>27-147</sup> and InvH<sup>70-147</sup> (Figure 2.3a, Figure A.4a). This broadening is diagnostic of conformational changes, with resulting differences in chemical shifts, occurring on the msec- $\mu$ sec time scale. Regardless of the exact nature, the flexibility of the InvH dimer may be important for its ability to dissociate and accommodate the InvG S domain.

Comparison of the structures of the InvH homodimer and InvH-InvG heterodimers provides clues towards how the InvG S domain could preclude InvH self-association. At the simplest

level, the homo- and heterodimer are mutually exclusive to a steric competition for the same binding interface of InvH. In more detail, residues 73-88 form helix  $\alpha$ 1a in the homodimer due to interactions with the opposing protomer. In contrast, residues 70-82 are dynamic in the heterodimer with a stable helix  $\alpha$ 1 beginning from residue 83. However, truncated InvH<sup>84-147</sup> can still form a primarily dimeric population, with a shift towards monomers upon mutation of Phe84 to alanine in InvH<sup>27-147</sup>. These data indicate Phe84 forms central interactions in the homodimer while residues 70-83 contribute to a minor extent (Figure 2.2d, Figure A.3a). We know from the ITC data that Trp549 is important for high affinity binding to InvH (Figure A.5e); it is therefore likely that Trp549 is able to displace the lower-affinity InvH residues 70-83 from the hydrophobic cleft. InvG Tyr553 can then take the place of InvH Phe84, while also forming an additional hydrogen bond with the Leu131 carbonyl (Figure 2.4e). Collectively, these interactions may lead to a higher propensity for InvH to bind InvG versus binding itself.

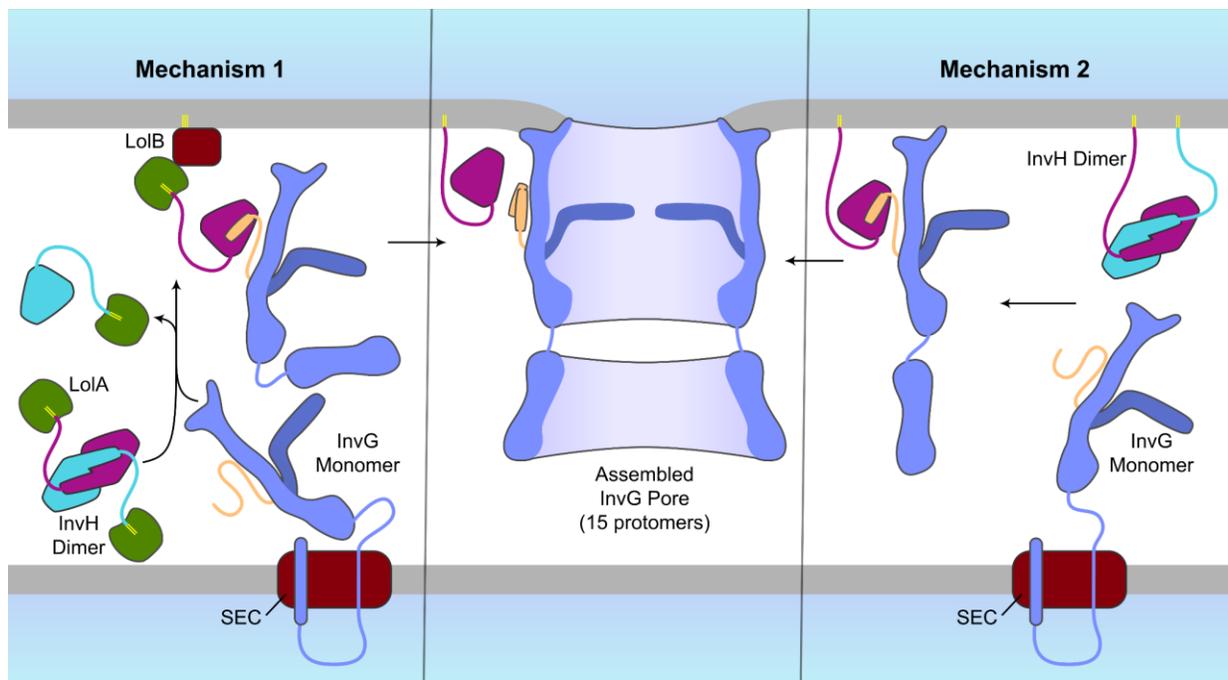


Figure 2.6: Potential secretin assembly mechanisms. In mechanism 1, the C-terminal S domain (tan) of InvG (blue) is free to bind to InvH (magenta, turquoise) once translocated to the

periplasm by the SEC transporter. This dissociates one member of the InvH homodimer, creating a pilotin-secretin-LolA complex that is transported to the outer membrane. There, LolA (green) interacts with LolB (red) to insert the InvH lipidated N-terminus (yellow) into the outer membrane, which brings InvG close enough to auto-oligomerize and perform BAM-independent insertion. In mechanism 2, InvH has already reached the OM via the Lol pathway before encountering InvG. InvH is on a long linker (~ 22 nm), which enables it to contact translocated InvG and form the InvH-InvG complex. InvG, now tethered to the outer membrane by InvH, is free to auto-oligomerize and insert once its local concentration is high enough.

---

Based on key similarities between pilotins noted above, we can begin to build a more generalized mechanism of pilotin-mediated secretin assembly (Figure 2.6). Collin *et al* were able to capture an R43L mutant of LolA R43L trapped in complex with the *Klebsiella* pilotin PulS, indicating that LolA is likely involved in shuttling the pilotin to the OM<sup>197</sup>. Given that all pilotins share a conserved N-terminal lipobox motif, it is expected that they also utilize the Lol pathway for their localization in a similar manner. Pilotins are required for either the OM localization and/or assembly of the secretin pore, depending on the species<sup>198</sup>. In the pilotins characterized to date, this relies on a hydrophobic interaction interface with the S domain, or on a tetratricopeptide repeat-mediated interaction. It has been suggested in the *Klebsiella* system that the secretin PulD traverses the peptidoglycan wall in a monomeric state, as PulD lacking the S domain cannot “hitchhike” with full-length PulD to reach the outer membrane<sup>197</sup>. It is unclear how the presence of the pilotin prevents insertion into the IM, as some species of secretins are capable of oligomerizing within minutes of synthesis *in vitro* and have high affinity for lipids<sup>244,245</sup>. One model suggests that the LolA-bound pilotin binds the C-terminal secretin S domain immediately after translocation by the SEC system, and the heterotrimeric complex is localized to the OM together (Figure 2.6). However, no LolA-pilotin-secretin complex has yet been isolated, indicating this may be a very short-lived state or does not form at all<sup>197</sup>. An alternative mechanism involves the pilotin being anchored to the OM prior to secretin binding

and utilizing its long N-terminal linker region to reach the secretin as it is translocated into the periplasm<sup>197</sup>. The InvH linker (from residues 16-72), would be approximately 22 nm long, assuming a disordered peptide length of  $\sim 4 \text{ \AA}$ <sup>246</sup>. With the distance between the inner and outer membranes spanning  $\sim 32 \text{ nm}$  in *Salmonella enterica* tomography<sup>247</sup>, this would allow InvH to extend further than half the distance from the outer to inner membrane. From there, InvH could contact the disordered C-terminal InvG S domain during SEC translocation, in turn preventing premature inner membrane insertion. Once the secretin has reached the OM, it undergoes assembly facilitated by multiple ring building motifs<sup>189</sup> and subsequent BAM-independent membrane insertion<sup>196</sup>, the latter perhaps facilitated by a localized bilayer perturbation the presence of multiple lipidated pilotins would presumably instill<sup>189</sup>.

It is unclear whether pilotins in general remain associated to the secretin after assembly. There is little evidence for InvH remaining associated with InvG after assembly; attempts to isolate the InvH-InvG complex have been unsuccessful<sup>200</sup>. In cryo-electron tomography of the *Shigella* T3SS, faint density surrounding the S domain may be attributed to the pilotin MxiM<sup>189,248</sup>. Such density, however, is not present in tomography of the *Salmonella* injectisome<sup>247</sup>. In contrast, there is direct evidence of type II pilotins binding assembled secretins. Type II secretion complexes including the *Klebsiella pneumoniae* pilotin and secretin, PulS and PulD, have been co-purified from recombinant expression<sup>249</sup>. The cryo-EM structure of *Escherichia coli* O78:H11 (ETEC) pilotin AspS in complex with oligomerized secretin GspD reveals that AspS can remain bound to the C-terminal helix of the S domain, preventing it from forming the clamp like n+2 interactions with the secretin  $\beta$ -barrel as observed for the apo secretin form in this Type II (Figure A.6c)<sup>192-194,250,251</sup> and other Type III secretins<sup>189</sup>. Although a structural feature apparently important for stability of the outer membrane pore, Yin *et al* suggest that the pilotin

itself may have a role in stabilizing the secretin due to interactions between AspS and the adjacent GspD protomer, at least in the ETEC T2SS GspD<sup>194</sup>. However, as the C-terminal helix of the InvG S domain is much more intimately associated with the  $\beta$ -barrel of InvG and with the neighboring S domain helix (interface surface area  $\sim 840 \text{ \AA}^2$  over S domain residues 543-558) than in GspD, it is unlikely that InvH would remain bound after assembly (interface surface area  $\sim 720 \text{ \AA}^2$  with InvG<sup>543-558</sup>). Further *in vivo* experiments and biochemical work can help to clarify the differences among the diverse pilotin mechanisms of action.

## 3 Cryo-EM structure of the EscV cytosolic domain nonamer

### 3.1 Introduction

The bacterial type III secretion system (T3SS) is divided into two types: the injectisome and the flagella, both multi-megadalton nanomachines<sup>9</sup>. Whereas the assembled flagellum grants motility to the bacteria, the injectisome instead functions as a “molecular syringe”. The latter complex forms a continuous channel through the Gram-negative envelope and host plasma membrane, allowing the bacterium to secrete effector proteins directly into the host cytosol. Effector proteins then modulate host cell signalling to benefit the pathogen. For example, effectors are used in enteropathogenic *Escherichia coli* (EPEC) to form attaching and effacing lesions in the gut<sup>35</sup>. Many clinically relevant pathogens including EPEC and enterohemorrhagic *E. coli* (EHEC)<sup>35</sup>, *Salmonella enterica*<sup>252</sup>, *Shigella flexneri*<sup>253</sup>, *Pseudomonas aeruginosa*<sup>254</sup>, and *Bordetella pertussis*<sup>255</sup> rely on injectisome expression for their virulence.

The injectisome is made up of several functional units resembling a syringe, which enable it to cross three membrane barriers simultaneously. The cytosolic sorting platform consists of an ATPase related to F<sub>o</sub>F<sub>1</sub> ATPases and six homomeric but asymmetrical pods. Effectors are stabilized by cognate chaperones and targeted to the sorting platform; there, they are separated from the chaperone and partially unfolded to fit through the 1.5 to 2.5 nm inner channel of the secretion system<sup>9,61,86,256,257</sup>. The sorting platform is docked to the basal body, a structural subdomain that forms a scaffold through the Gram-negative inner and outer membranes<sup>59</sup>. Three proteins form concentric oligomeric rings, and the periplasmic export apparatus creates a gated channel for effector proteins<sup>58,66,68</sup>. In addition, evidence suggests the inner membrane (IM) pore protein, part of the export apparatus, contains a proton channel involved in PMF-driven secretion<sup>168,175,177</sup>. Several of the sorting platform and basal body proteins are conserved in the

flagellum, notably the inner membrane rings, export apparatus, and the cytosolic ATPase<sup>9</sup>. The extracellular portion of the injectisome is the needle, made up of a hollow helical filament that anchors into the basal body<sup>257</sup>. The translocon pore is secreted through the needle and punctures the host plasma membrane, allowing the needle to adhere to the host<sup>130</sup>. The functional domains of the injectisome coalesce to create a continuous channel from the bacterium to the host.

A vital step in injectisome assembly and function is substrate selection and secretion, which is mediated primarily by the inner membrane pore protein<sup>162,258</sup>. This homonameric assembly, termed EscV in the EPEC injectisome and FlhA in its flagellar homologue, is made up of a soluble cytosolic domain and a seven or eight transmembrane (TM) helix integral membrane domain<sup>180</sup>. The cytosolic domain functions to select effectors for secretion; it has been shown to be vital to switching from middle (translocator) to late (effectors) substrates in the EPEC injectisome<sup>162</sup>, and its binding site to effector-chaperone complexes has been characterized in flagellar FlhA<sup>258</sup>. The cytosolic domain of both the injectisome and flagella counterparts has been characterized by crystallography<sup>72-74,258-262</sup>, with MxiA from the *Shigella flexneri* injectisome forming a nonameric ring architecture in the crystal<sup>74</sup> that agrees with dimensions from *in situ* studies<sup>78</sup>. The transmembrane region of the IM pore, characterized in *Salmonella* injectisome InvA by sub-tomogram averaging<sup>57,79</sup>, forms a pore through the inner membrane through which secreted effectors enter the T3SS. Interestingly, Minamino *et al* suggest that the transmembrane region of FlhA forms a proton channel, implicating the IM pore in the coupling of energy from proton motive force to secretion<sup>175,177</sup>. Knockout of EscV has a detrimental effect on secretion and prevents formation of host pedestals by EPEC<sup>263,264</sup>.

We present the single-particle cryo-electron microscopy structure of the cytosolic domain of EscV (EscV<sub>C</sub>) at 4.7 Å resolution, allowing model building based on homologous structures

from injectisomes and flagella. EscV<sub>C</sub> formed two stacked nonameric rings in solution, with the membrane face forming the interaction interface between the two rings. The overall structure of the cytosolic domain is similar to homologous crystal structures, with subdomains (SD) SD3 and SD4 the most conserved. The interfaces between the EscV<sub>C</sub> protomers are held together mainly through electrostatic interactions, and the assembled ring forms a large electropositive interface presented towards the inner membrane. The EscV<sub>C</sub> model provides structural insight into the oligomerized cytosolic ring in solution.

## 3.2 Methods

### 3.2.1 Purification of EscV<sub>C</sub>

DNA encoding EscV residues 338-675 was amplified from EPEC genomic DNA (*Escherichia coli* O127:H6 E2348/69) and cloned into pET28a with a thrombin-cleavable N-terminal 6His tag. The vector was transformed into *E. coli* BL21 ( $\lambda$ DE3, Sigma), and the transformed bacteria were used to inoculate a starter culture in 2xYT media. Starter culture (50 mL) was used to inoculate 1 L ZYP-5052 auto-induction media; the culture was grown with shaking at 200 rpm for 4 hours at 37 °C followed by 16 hours at 22 °C for expression. Cells were pelleted at 6,200  $\times$ g for 20 minutes and resuspended in 4 mL lysis buffer per gram of bacterial pellet (20 mM HEPES pH 7.5, 500 mM NaCl, 0.5 mM TCEP-HCl, 5 % (v/v) glycerol, 15 mM imidazole, EDTA-free protease inhibitor tablet (Roche)). Cells were lysed by two passes through an EmulsiFlex-C5 homogenizer (Avestin), and insoluble material was pelleted by centrifugation at 40,000  $\times$ g for 30 minutes. The supernatant was filtered and allowed to flow over 1 mL HisPur™ Ni-NTA Resin (Thermo Scientific) by gravity. The resin was washed with high salt buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 0.5 mM TCEP-HCl, 5 % (v/v) glycerol) with the following imidazole concentrations: 40  $\times$  CV 50 mM imidazole, 10  $\times$  CV 75 mM imidazole, 10

× CV 100 mM imidazole. Protein was eluted in the same buffer containing 300 mM imidazole. Protein-containing fractions were pooled and the thrombin tag was cleaved by addition of 1:1000 molar ratio bovine alpha-thrombin (Haematologic Technologies Incorporated) at 4 °C overnight. During thrombin cleavage, the protein was dialyzed into size exclusion buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP-HCl, 5 % (v/v) glycerol). The following day, the protein was concentrated using a stirred cell (Amicon) with a 100 kDa MWCO membrane. Protein was injected onto a superose 6 10/300 GL size-exclusion column (GE Life Sciences), and protein eluting in a peak at about 14 mL was pooled for subsequent grid preparation. Prior to applying sample to cryo-EM grids, the protein was desalted into a lower glycerol buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP-HCl, 1.5 % (v/v) glycerol) using a PD-10 gravity column (GE Healthcare) and concentrated to 1.5 mg/mL using an Amicon Ultra centrifugal filter 100 kDa MWCO (EMD Millipore).

### 3.2.2 Preparation of negative-stain TEM grids

Purified EscV<sub>C</sub> nonamers at 400 nM concentration was applied to glow-discharged 400 mesh copper TEM grids coated with a carbon film (CF400-CU, Electron Microscopy Sciences). The protein sample (3 µL) was applied to the grid for 30 seconds, followed by two 30 second washes in water and two 30 second stains in 2% uranyl acetate solution. Negative stain TEM images were taken using a 200 kV FEI Tecnai G2 TEM on an FEI Eagle 4k CCD camera.

### 3.2.3 Cryo-EM data collection

Purified EscV<sub>C</sub> (3 µL) at 1.5 mg/mL in cryo-EM buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP-HCl, 1.5 % (v/v) glycerol) was applied to a glow-discharged Quantifoil grid (Copper, 300 mesh, R 2/2). Using a Vitrobot Mark 4 (FEI) at 100% humidity, the sample

was immediately blotted with a blot force of -5 and a time of 1.5 seconds and plunge-frozen in liquid ethane.

Cryo-EM data collection was performed on a 300 kV Titan Krios (FEI) equipped with a Falcon 3 detector (FEI) at the HRMEM facility at the University of British Columbia. Using EPU collection software (ThermoFisher Scientific), 2037 movies were collected at  $96,000 \times$  magnification with a pixel size of  $0.852 \text{ \AA}/\text{pix}$ . The exposure time was 60 seconds with a total dose of 60 electrons per  $\text{\AA}^2$ . With 38 frames per movie, each frame had an exposure time of 1.52 seconds and a dose of 1.47 electrons per  $\text{\AA}^2$ . Movies were collected at a defocus range of -1 to -4  $\mu\text{m}$ .

#### 3.2.4 Cryo-EM Map reconstruction and model-building

The 2037 collected movies were motion corrected in Relion3 using MotionCor2, and the contrast transfer function was calculated for each micrograph using CTFFind4.1<sup>265-267</sup>.

Approximately 1000 particles were picked manually and 2D averaged in Relion3 to create templates; the templates were used to auto-pick particles from all micrographs, yielding 137,341 particles in total. 2D averaging was used to remove poor particle picks, and 127,583 particles were used to generate an *ab initio* model. The particles were refined by repeated rounds of 2D and 3D classification, finally yielding a  $5.1 \text{ \AA}$  D9 structure in Relion3's 3D auto-refine using 14,908 particles. After per-particle CTF refinement, particle polishing, and post-processing, the resolution of the D9 reconstruction was  $4.6 \text{ \AA}$ .

To obtain a C9 reconstruction, 127,583 particles were used to generate a C1 map that was subsequently subtracted to a single ring in Relion3. Further rounds of 2D and 3D classification, and subtraction using a tighter mask, improved the resolution of the C9 ring to  $5.3 \text{ \AA}$  with 83,566

particles. After correcting beam tilt, refining per-particle CTF, and post-processing, the final resolution was 4.7 Å.

Homology models of EscV<sub>C</sub> were generated using Phyre2<sup>231</sup> from crystal structures of MxiA (4A5P<sup>74</sup>), InvA (2X49<sup>73</sup> and 3LW9<sup>72</sup>) and FlhA (3A5I<sup>262</sup>, 3MIX<sup>259</sup>, 3MYD<sup>260</sup>). Each subdomain of the models was rigid body fit into the cryo-EM density in Chimera<sup>220</sup>. The aligned homology models were then combined to create the best agreement with the experimental density using RosettaCM with C9 symmetry<sup>268</sup>. Following this, Rosetta was used for C9 symmetric refinement of the best RosettaCM model into the density<sup>269</sup>. Refinement was performed iteratively using Coot, Phenix real space refine, and Rosetta<sup>218,269,270</sup>.

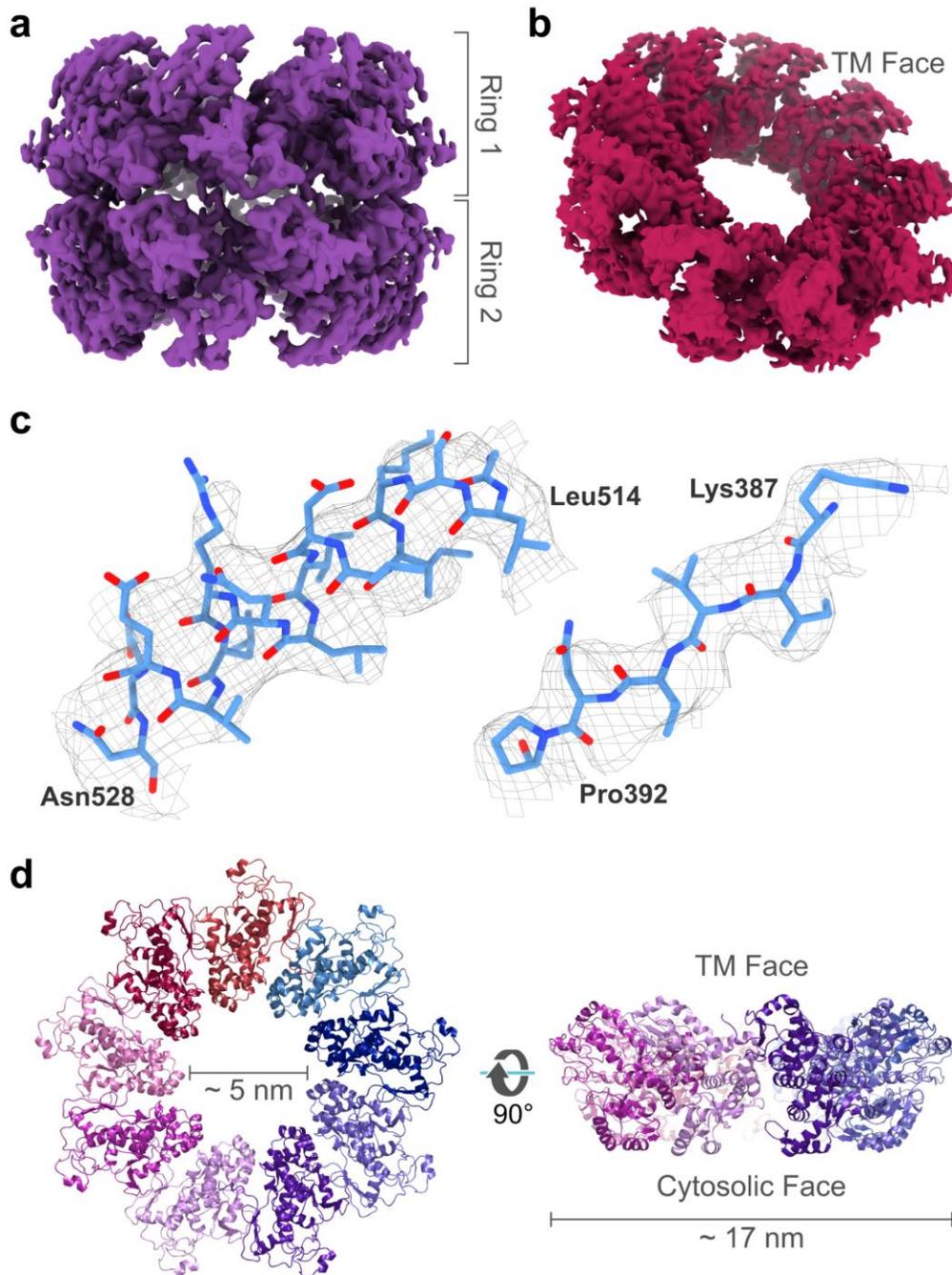
Figures were generated using PyMOL, Chimera, and ChimeraX<sup>219–221</sup>. Electrostatic potential surface was calculated using APBS<sup>271</sup>, and lipophilicity potential surface was calculated using PyMLP<sup>233</sup>.

### 3.3 Results

#### 3.3.1 EscV<sub>C</sub> forms a nonameric ring in solution

To determine the structure of cytosolic EscV in solution, a construct of the cytosolic domain EscV<sub>C</sub> (encompassing residues 338-675) was recombinantly expressed in *E. coli* BL21. In high salt buffers, EscV<sub>C</sub> eluted from a size exclusion column at a volume expected for a monomer (data not shown); however, at a salt concentration of 150 mM, the primary population of EscV<sub>C</sub> eluted at a volume indicating a large, monodisperse complex (Figure B.1a). When analyzed by negative stain transmission electron microscopy, clear ~15 nm rings were visible (Figure B.1b). This size was consistent with what was expected based on the crystal structure of the *S. flexneri* homologue MxiA, crystallized as a nonameric ring<sup>74</sup>. Quantifoil cryo-EM grids were prepared

with protein from the primary size exclusion peak, and 2037 micrographs were collected using a Titan Krios with a Falcon 3 detector (Figure B.1c).



**Figure 3.1: Map and model of EscV<sub>C</sub>.** (a) 4.6 Å reconstruction of the D9 symmetry stacked EscV<sub>C</sub> rings. (b) 4.7 Å reconstruction of a single C9 ring obtained through particle subtraction. (c) Sample electron density from (b). (d) View of model from transmembrane face and 90 degree rotation.

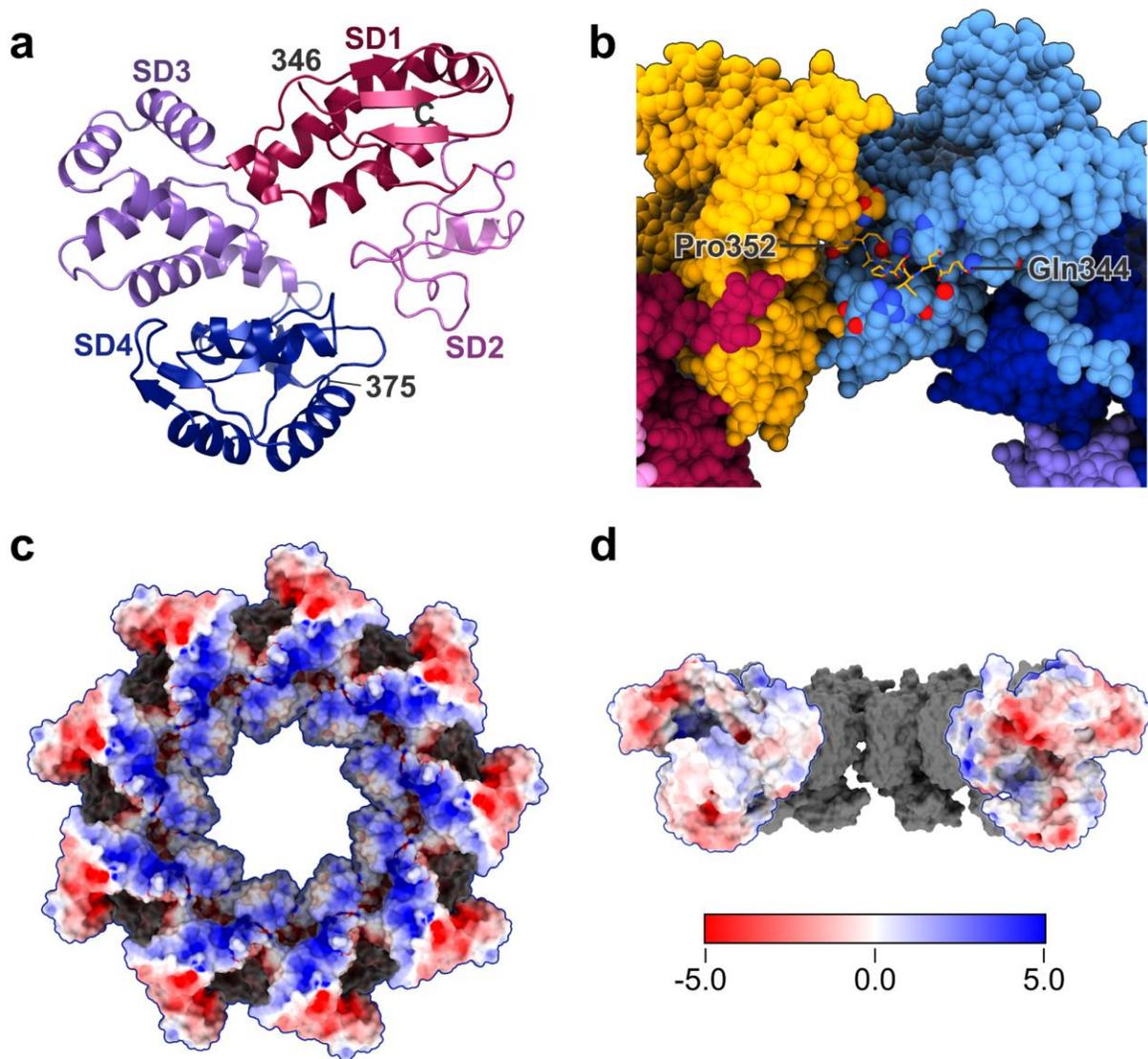
Table 3.1: EscV<sub>C</sub> cryo-EM data collection, refinement and validation statistics

	EscV <sub>C</sub> C9	EscV <sub>C</sub> D9
<b>Data collection and processing</b>		
Magnification	96000	96000
Voltage (kV)	300	300
Total dose (e <sup>-</sup> /Å <sup>2</sup> )	60	60
Dose per frame (e <sup>-</sup> /Å <sup>2</sup> )	1.47	1.47
Defocus range (µm)	-1 to -4	-1 to -4
Pixel size (Å)	0.852	0.852
Symmetry imposed	C9	D9
Initial number of particles	127583	127583
Final number of particles	83566	14908
Map resolution (Å)	4.7	4.6
FSC threshold	0.143	0.143
<b>Refinement</b>		
Initial model used (PDB code)	3A5I, 3MYD, 2X49, 3MIX, 4A5P, 3LW9	
Model resolution (Å)	4.7	
FSC threshold	0.143	
Model resolution range (Å)		
Map sharpening <i>B</i> factor (Å <sup>2</sup> )	-235.7	
Model composition		
Non-hydrogen atoms	23985	
Protein residues	2988	
Ligands	0	
<i>B</i> factors (Å <sup>2</sup> )		
Protein	101.4	
Ligand	N/A	
R.m.s. deviations		
Bond lengths (Å)	0.024	
Bond angles (°)	2.52	
Validation		
MolProbity score	1.75	
Clashscore	5.72	
Poor rotamers (%)	0.0	
Ramachandran plot		
Favored (%)	93.26	
Allowed (%)	5.59	
Disallowed (%)	1.15	

A preliminary reconstruction at 4.6 Å resolution was calculated from 14,908 particles, after several rounds of 2D and 3D classification in Relion3 (Figure B.1d-e). EscV<sub>C</sub> formed two stacked nonameric rings in solution with D<sub>9</sub> symmetry (Figure 3.1a). The rings stacked head-to-head along the interface abutting the inner membrane; this stacking is likely an artefact of purifying the cytosolic construct, as the presence of the transmembrane domain in full-length EscV would block this interface. Despite the D<sub>9</sub> symmetry of the reconstruction, 2D averaging made it clear that one ring was the most stable, while the second ring was sometimes missing or incomplete (Figure B.1d). To improve the resolution, particle subtraction was performed in Relion 3 to generate particles with a single ring. While the FSC-calculated resolution did not improve, local side-chain details became much more pronounced (Figure 3.1b). The final C<sub>9</sub> reconstruction was generated from 83,566 particles, subtracted to a single ring, which yielded a resolution of 4.7 Å (Figure B.1f). The EscV<sub>C</sub> ring has a diameter of 14 and 17 nm at its narrowest and widest points respectively, with an inner pore diameter of 5 nm. The height of a single ring is about 5.5 nm.

At the map resolution of 4.7 Å, *ab initio* modelling was not feasible (Figure 3.1c); to build a model, homology models were generated in Phyre2<sup>231</sup> from injectisome homologues MxiA<sup>74</sup> and InvA<sup>72,73</sup>, and from several species of flagellar FlhA<sup>259,260,262</sup>. Rosetta comparative modelling was used to combine features best fitting the density from each model, and for subsequent refinement<sup>268,269</sup>. The final model has a good fit to the density and good geometry (Figure 3.1d, Table 1). As in homologous structures, the EscV<sub>C</sub> monomer is made up of four subdomains (Figure 3.2a). The best resolved regions are located towards the centre of the ring, with subdomains SD3 and SD4 modelled into the highest resolution density and SD1 in the lowest resolution region (Figure B.2a-b). The model encompasses residues 344-675, with linker

residues 344-352 nestled into a groove along the adjacent protomer (Figure 3.2b).



**Figure 3.2: Atomic details of the EscV<sub>C</sub> model.** (a) View of an EscV<sub>C</sub> monomer, with subdomains 1 through 4 coloured (red, pink, purple, blue respectively) and labelled. (b) Stick representation of the EscV linker (orange), nestled into a groove on the adjacent protomer (light blue). Oxygens and nitrogens are coloured red and blue respectively, both on the linker and on nearby residues. (c) Electrostatic potential surface view, generated by APBS, of the transmembrane face of EscV<sub>C</sub>. A distinct electropositive ring (blue) is visible on this face. (d) Electrostatic surface of the EscV<sub>C</sub> protomer interface, with complementary electrostatics on each of the two faces (left and right).

To determine the nature of the interactions EscV<sub>C</sub> exhibits between protomers and with other interacting proteins, electrostatic potential analysis was performed using APBS. EscV<sub>C</sub> has many charged surface residues and is primarily electronegative around the outside of the ring. However, it has strong electropositive character along the inside of the pore, and in a discrete circle on the face presented towards the transmembrane region (Figure 3.2c). The intermolecular interfaces between protomers in the EscV<sub>C</sub> ring show complementary electrostatic regions (Figure 3.2d). As in *Shigella* homologue MxiA<sup>74</sup>, the majority of the ~1000 Å<sup>2</sup> interacting interface (calculated by PDBePISA) comes from SD1 and SD3.

### 3.3.2 EscV<sub>C</sub> shares similarities with injectisomal and flagellar inner membrane pores

The EscV<sub>C</sub> monomer shares the same four-subdomain architecture as injectisome homologues MxiA<sup>74</sup> and InvA<sup>72,73</sup>, and flagella homologue FlhA<sup>258–262</sup>. SD3 is the most structurally conserved, with C $\alpha$  RMSDs of ~ 1.1 Å<sup>2</sup> when compared to eight homologous structures (Figure 3.3a, Table 2.2). SD1 is similar but has differences in peripheral regions, with C $\alpha$  RMSDs of ~ 1.8 Å<sup>2</sup>. SD2 and SD4 have the poorest structural conservation but are better conserved between EscV and injectisome homologues than in flagella.

FlhA has been shown previously to take on either an open, semi-closed, or closed conformation, with the protein alternating between these states in molecular dynamics simulations<sup>261</sup>. EscV<sub>C</sub> takes on a closed conformation, similar to other injectisome crystal structures. Overlaying SD3 of EscV<sub>C</sub> with an open or semi-closed form of FlhA from *Salmonella enterica* shows a stark difference in the orientation of SD2, with a much larger opening in open FlhA<sup>261</sup>. However, the difference between EscV<sub>C</sub> and closed homologous structures (MxiA and

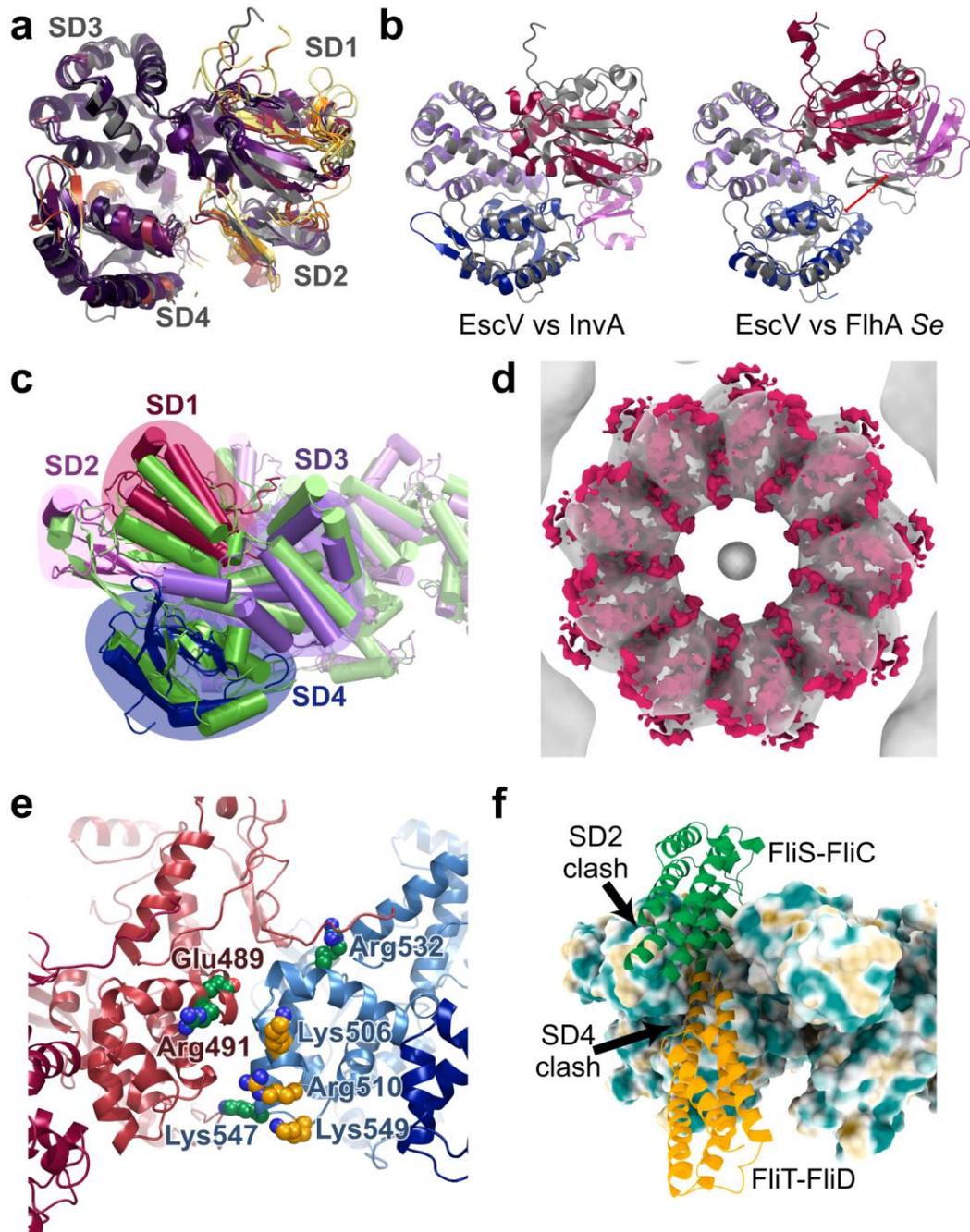


Figure 3.3: Comparison of EscV<sub>C</sub> with other characterized homologues. (a) Alignment of EscV<sub>C</sub> monomer (grey) with six homology models (MxiA 4A5P, InvA 2X49 3LW9, FlhA 3A5I 3MYD 3MIX), aligned individually on each subdomain and coloured based on the distance between homologous residues (0 Å black; > 8 Å yellow; PyMOL script ColorByRMSD). SD1 and SD2 have the most variable regions. (b) Overlay of EscV<sub>C</sub> monomer (grey) with InvA (left; 2X49; coloured by subdomains in red, pink, purple, blue sequentially) and FlhA (right; 3A5I; same colouration as InvA), demonstrating the slightly more closed conformation of InvA and the

more open conformation of FlhA. (c) EscV<sub>C</sub> (coloured by subdomain as InvA in (b)) aligned onto the crystallized nonameric MxiA ring (green; 4A5P). MxiA is rotated slightly towards the outside of the ring relative to EscV<sub>C</sub>. (d) Overlay of EscV<sub>C</sub> map (threshold 0.017) with *in situ* InvA density (threshold 1.4). (e) Mapped conserved residues from MxiA mutations. Green residues affect MxiA oligomerization, while orange residues line the inner pore and affect secretion. (f) Overlay of chaperone-bound FlhA complexes (FliT-FliD-FlhA orange, 6CH2; FliS-FliC-FlhA, green, 6CH3) aligned on SD1 and SD2 of EscV<sub>C</sub>. EscV<sub>C</sub> is represented as a hydrophobic surface (hydrophilic teal; hydrophobic tan; calculated by pyMLP). The chaperones bind into the hydrophobic cleft between SD2 and SD4, but clash with EscV<sub>C</sub>, indicating a conformational change would be required for binding of a similarly shaped EPEC chaperone.

---

InvA) is more subtle, with EscV<sub>C</sub> having a slightly larger distance between SD2 and SD4 (Figure 3.3a). EscV<sub>C</sub> takes on a closed conformation, albeit slightly less compact than in other closed homologues.

To determine whether EscV<sub>C</sub> can alternate between closed and open forms, a molecular dynamics simulation was run on the monomer. Unlike previous simulations on FlhA<sup>261</sup>, EscV<sub>C</sub> maintained a closed form, indicated by a distance ranging from 23 to 33 Å between the SD2 and SD4 centres of mass (Figure B.2d). The subdomains do show flexibility, deviating from the 29.5 Å distance observed between them in the cryo-EM structure, but not opening completely to a 40 to 45 Å distance as in FlhA<sup>261</sup>. The FlhA G368C mutant forces the protein into a closed form, possibly by limiting its flexibility<sup>261</sup>; interestingly, this glycine is not conserved with injectisome homologues (Figure B.3), suggesting that these proteins have evolved to naturally take on a closed conformation.

The only other EscV homologue to be characterized to high resolution in nonameric form is MxiA from the *Shigella flexneri* injectisome<sup>74</sup>. The EscV<sub>C</sub> ring is overall similar in size and architecture to MxiA, with a C $\alpha$  RMSDs of 4.2 over 2162 residues (Figure 3.3c). There are differences in the orientations of the subdomains relative to one another (as described above),

causing the overall RMSD to decrease. In addition, MxiA is tilted further towards the outside of the ring than EscV. Notably, the crystal packing of MxiA rings does not share the same head-to-head interface as in stacked EscV<sub>C</sub> but are instead packed head-to-tail.

Several cryo-electron tomography structures of injectisomes have been published with high detail on the IM pore region<sup>57,78,79,256</sup>. The dimensions of the ring in all species are similar, with a diameter of about 15 nm and a pore size of 5 nm. The highest resolution tomogram, from the *Salmonella enterica* injectisome<sup>57</sup>, shares the same C9 symmetry in its IM pore InvA as in EscV<sub>C</sub>. The model of EscV<sub>C</sub> agrees well with the InvA density, indicating that its architecture in solution is similar to what has been observed *in situ* (Figure 3.3d).

### 3.4 Discussion

EscV is a vital protein in the EPEC injectisome IM pore complex, acting as a substrate selection checkpoint, proton channel, and inner membrane pore. The cryo-EM map and model provide insight into EscV<sub>C</sub> ring formation in solution, permitting direct comparison with related injectisomal and flagellar homologue structures solved by X-ray crystallography.

The EscV<sub>C</sub> ring relies primarily on electrostatic interactions for its assembly. Electrostatic surface potential analysis reveals complementary charges between interacting protomers, similar to those observed previously in the *Salmonella* flagella FlhA<sub>C</sub> dimer<sup>262</sup> and *Shigella* injectisome MxiA ring<sup>74</sup>. In addition, the ring relies on low salt conditions for stability in solution, a trait noted previously in FlhA<sup>258</sup>. The cryo-EM map resolution was sufficient to model a portion of the linker, which nestles into a groove on the adjacent subunit fueled by mainly polar interactions. The intimate position of the linker with the preceding protomer supports its importance for oligomerization, which has previously been observed in the structure of FlhA<sub>C</sub><sup>262</sup>. Conserved residues from SD3 of MxiA that have been shown to be important for

oligomerization<sup>74</sup> map to the oligomerization interface and to the binding pocket of the linker region (Figure 3.3d, Figure B.3). In addition, three positively charged residues that affect secretion in MxiA<sup>74</sup> are conserved and map to the inner pore, as in *S. flexneri*. It is unclear why the presence of a positive charge inside the pore is important for secretion. EscV<sub>C</sub> shares many similarities with injectisomal and flagellar homologues in its ring formation properties.

The binding site between FlhA<sub>C</sub> and two chaperone-effector complexes, FliS-FliC and FliT-FliD, has been characterized by NMR spectroscopy and X-ray crystallography<sup>258</sup>. Aligning the ternary complexes to the assembled EscV<sub>C</sub> ring demonstrates that the conserved hydrophobic pocket on EscV<sub>C</sub> could also accommodate chaperone binding (Figure 3.3f). However, there are some steric clashes. The FliC (flagellin) complex clashes with SD2, while the FliD (filament cap) complex clashed with SD4. The open conformation of FlhA has been proposed to switch specificity of the IM pore from early (hook) to late (flagellin and tip) substrates<sup>261</sup>; in agreement with this model, EscV<sub>C</sub> would require a more open conformation to accommodate these late chaperone-substrate complexes. The injectisomal equivalent of flagellin and the cap are the middle substrates, the tip/filament and translocon. The specificity switch in EscV<sub>C</sub> is governed by the gatekeepers SepD and SepL, which have been shown to increase the affinity of EscV<sub>C</sub> for middle rather than late chaperone complexes<sup>162</sup>. It is possible that the gatekeepers mediate the open conformation of EscV<sub>C</sub>, allowing binding of the middle chaperone-substrate complexes.

Previous studies on FlhA have outlined its importance in PMF-dependent secretion and its potential role as a proton channel<sup>175,177</sup>. Erhardt *et al* found that a protonation mimicking mutant, D158N (D148 in EscV; Figure B.3), on a conserved cytosolic loop of FlhA<sub>TM</sub> caused a conformational change in the protein, which they attribute to a physiological secretion mechanism fueled by local pH<sup>180</sup>. Interestingly, the transmembrane face of EscV<sub>C</sub> has a notably

electropositive surface (with a large contribution from conserved R374), which could play a role in this mechanism. Deprotonated, negatively charged D148 could form a favourable interaction with the electropositive face of EscV<sub>C</sub>; protonation of this residue due to high local pH would be less favourable and could result in a conformational change of EscV. Further studies into the affect of pH on EscV conformation would improve our understanding of how energy from proton motive force is coupled to secretion.

## 4 Cryo-EM structure of the homohexameric T3SS ATPase-central stalk complex reveals rotary ATPase-like asymmetry

### 4.1 Introduction

Rotary ATPases are a biologically important and well-conserved protein family, fuelling vital life processes from archaea to humans. One of the earliest examples of molecular machines, their well-studied mechanism of ATP binding and hydrolysis fuels conformational changes to generate torque. The related F- and V- ATPases are composed of a soluble catalytic complex that can synthesize or hydrolyze ATP ( $F_1/V_1$ ), which is coupled to a transmembrane proton (or sodium) channel ( $F_o/V_o$ ). These motors have opposite roles depending on their cellular context: F-ATPases use membrane potential to rotate the  $F_o$  complex and synthesize ATP at the coupled  $F_1$  domain, while V-ATPases use energy derived from ATP hydrolysis to pump protons across the membrane and acidify intracellular compartments such as vacuoles.

The broad  $F_1/V_1$ -ATPase family – with characteristic Rossmann fold, Walker A and B motifs, and hexameric stoichiometry – also encompasses distant relatives including ATPases associated with the bacterial injectisome and flagellum<sup>272</sup>. These two proteinaceous assemblies, involved in virulence and motility, use an evolutionarily related type III secretion system (T3SS; referred to here as *f*T3SS (flagellar) and *v*T3SS (virulence injectisome)) to secrete self-assembly and host-cell manipulating substrates. The T3SS ATPase homologues are made distinct by their homohexameric construction, a presumed ancestral precursor to the heterohexameric rotary  $F_1/V_1$ -ATPases. Homology of the type III secretion (T3S) ATPases to the catalytic  $F_1/V_1$  subunits was first predicted based on sequence comparison<sup>273,274</sup>, and subsequently demonstrated with monomeric structures of the *f*T3SS ATPase FliI (PDB: 2DPY<sup>275</sup>) and the enteropathogenic

*E. coli* (EPEC)  $\nu$ T3SS ATPase EscN (2OBL, 2OBM)<sup>276</sup>. Further evolutionary similarities were found in soluble accessory components of the F<sub>1</sub>/V<sub>1</sub>- ATPases, with sequence and structural homology demonstrated for the peripheral<sup>277,278</sup> and central<sup>279–283</sup> stalks. Accessory subunits form the foundation of ATPase function: the central stalk acts as a rotor and couples the F<sub>1</sub>/V<sub>1</sub> and F<sub>o</sub>/V<sub>o</sub> complexes, while the peripheral stalks act as stators to prevent rotation of the catalytic subunits. The evolutionary relationship and conserved elements of this core complex raise the intriguing possibility that T3S may utilize a related torque-generating motor in bacterial secretion.

The injectisome is a highly complex nanomachine that secretes specific bacterial effector proteins directly into the cytoplasm of infected host cells, a process which allows subsequent subversion of cell signalling to the pathogen's advantage. The injectisome is essential to the virulence of a broad range of pathogens, including EPEC/EHEC, *Salmonella*, *Shigella*, *Pseudomonas*, *Yersinia*, *Chlamydia*, and *Bordetella*: causative agents of food poisoning, dysentery, nosocomial infections, plague, sexually transmitted infections, and whooping cough. Although the T3S-delivered effectors vary amongst the different pathogens, the secretion apparatus itself is well conserved and thus of interest as a multivalent target of anti-virulence therapeutics and vaccines.

The injectisome must overcome the significant hurdles of passaging effectors through the inner and outer membranes of the Gram-negative envelope, the intervening cell-wall peptidoglycan mesh, the extracellular space (~90 nm in EPEC), and the host cell membrane. The major structural scaffold of the injectisome is the basal body, comprising concentric, highly-oligomerized membrane-spanning protein rings that pass through the Gram-negative envelope. The attached extracellular needle terminates with a translocon pore-forming complex that

punctures the host cell membrane, creating a continuous channel from pathogen to host. This assembly results in the characteristic syringe-like appearance, with a wide cylindrical body adjoining a hollow needle of ~20 Å inner diameter through which effector proteins can pass in a semi-folded state (for a T3SS review, see Deng *et al.*<sup>284</sup>). Substrate selection occurs in the cytoplasm and inner membrane, mediated by the export gate sub-complex (EscRSTUV in EPEC, and FliPQR, FlhB, FlhA in flagella) and the cytoplasmic ATPase complex (ATPase EscN, peripheral stalk EscL, and central stalk EscO; flagellar FliI, FliH and FliJ respectively). Together, these components must presumably work to select, prepare, and secrete effector proteins through the injectisome.

The T3SS has co-opted an early ancestor of F<sub>1</sub>/V<sub>1</sub>-ATPases for its distinct, customized action. The energy for secretion was historically proposed to solely derive from the T3S ATPase<sup>285</sup>, but was subsequently shown in *ϕ*T3S to depend on proton motive force (PMF)<sup>286,287</sup>, mediated in part by the partnering export gate component FlhA (homologue of EPEC EscV)<sup>288,289</sup>. The ATPase has been proposed to specifically enhance PMF-induced secretion efficiency<sup>289</sup>, as well as to play a role in effector targeting, release from chaperones, and unfolding prior to secretion<sup>290-293</sup>. Most T3S effectors rely on protective chaperones to deliver them in a partially-unfolded state to the cytosolic complex<sup>294</sup> and it has been demonstrated that the *v*T3SS ATPase can dissociate T3S effector cargo from its cognate chaperone<sup>291-293</sup>. It is hypothesized that the partially-unfolded effector (including structural elements as large as single helices<sup>295</sup>) is subsequently passaged into the adjacent export gate protein pore (EscV in EPEC) and funnelled into the hollow T3SS needle for secretion.

Although predicted to function as a hexamer by analogy to the rotary ATPases, all structures to date of T3SS ATPase orthologues<sup>275,276,278,296-298</sup> have been monomeric, resulting in a poor

understanding of the precise atomic details of oligomerization and cooperative catalytic function in these homohexameric variants. Similarly, little is known about the ATPase-central stalk interaction and how that relates to its role in T3S. To address these questions, we have determined here the cryo-EM structure of a homohexameric T3S ATPase EscN in complex with its central stalk protein EscO. The 3.3 Å resolution is sufficient to trace conformational differences, model amino acid sidechains, and observe bound ligands and metal ions. Remarkably, EscN's asymmetric hexamer reveals similarity to the rotary ATPases that extends beyond the monomer, despite not sharing their heterohexameric construction. The different catalytic states observed in its active sites and relative disposition of the central stalk are in close agreement with previous structures of various assembled F<sub>1</sub> and V<sub>1</sub> heterohexameric ATPases, showing conservation of the necessary conformational elements required for a rotary ATPase mechanism. This resulting structural evidence for EscN/EscO torque generation brings us closer to unravelling the mechanism of ATPase action in T3S substrate selection and energetics.

## 4.2 Methods

### 4.2.1 Expression and purification of EscN<sup>29-446</sup> and EscO<sup>1-125</sup>

The genes for EscN and EscO were cloned from EPEC genomic DNA (*Escherichia coli* O127:H6 E2348/69)<sup>299</sup> into individual pET28a vectors, each with a thrombin-cleavable N-terminal 6His tag. The first 28 residues were omitted from EscN, as this improved expression and stability; EscO was expressed as a full-length protein. The vectors pET28<sup>EscN-6NHis</sup> and pET28<sup>EscO-6NHis</sup> were transformed separately into *E. coli* BL21 (λDE3, Sigma) and expressed in 2L and 0.5L ZYP-5052 auto-induction media respectively (grown at 37 °C for 3.5 hours followed by 22 °C for 16 hours). Cells were harvested by centrifugation at 6,200 ×g for 20 minutes, yielding approximately 25g pET28<sup>EscN-6NHis</sup>-transformed cells and 7g pET28<sup>EscO-6NHis</sup>-

transformed cells. Pellets were resuspended in lysis buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP-HCl, 15 mM imidazole, EDTA-free protease inhibitor tablet (Roche)) at a ratio of 4 mL buffer to 1 g cell pellet. The cells were lysed by two passes through an EmulsiFlex-C5 homogenizer (Avestin), and insoluble material from the lysate was removed by centrifugation at 40,000 ×g for 30 minutes. The lysate was filtered and passed over a 5mL HisTrap HP (GE Healthcare) in the case of EscN, and over a 1mL HisTrap HP (GE Healthcare) for EscO. The columns were washed with 20 CV of 50 mM imidazole wash buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP-HCl, 50 mM imidazole), followed by 5 CV of 75 mM imidazole wash buffer and 5 CV 100 mM imidazole wash buffer; protein was eluted in 300 mM imidazole buffer. To remove the His-tag, the proteins were desalted into EscN-EscO buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP-HCl) and incubated with bovine alpha-thrombin (Haematologic Technologies Incorporated) at 4 °C overnight at a molar ratio of 1:1000. Imidazole stock was added to the proteins to a final concentration of 15 mM, and the proteins were passed once more over a HisTrap HP (GE Healthcare) column to remove any uncleaved sample. The proteins were once again desalted into EscN-EscO buffer using a PD-10 desalting column (GE Healthcare), and concentrated with Amicon Ultra centrifugal filters (EMD Millipore) to a final concentration of ~25 mg/mL for EscN<sup>29-446</sup> and ~5 mg/mL for EscO<sup>1-125</sup>. Mutants of EscN<sup>29-446</sup> and EscO<sup>1-125</sup> were made using the QuikChange mutagenesis kit (Stratagene), and purified using the same protocol.

#### 4.2.2 Glycerol gradient centrifugation of EscN-EscO complex

EscN<sup>29-446</sup> and EscO<sup>1-125</sup> were incubated together at 4 °C for a minimum of 8 hours at 5.0 mg/mL and 0.8 mg/mL respectively (1:2 molar ratio EscO:EscN, 3-fold excess EscO from the expected 1:6 stoichiometry) in EscN-EscO buffer containing aluminium fluoride ATP hydrolysis

transition state analogue (EscN-EscO-ADP-AlF<sub>3</sub>: 20 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP-HCl, 1 mM ADP, 3 mM MgCl<sub>2</sub>, 6.25 mM KF, 1.25 mM AlCl<sub>3</sub>). A 10-25% glycerol gradient was made with EscN-EscO-ADP-AlF<sub>3</sub> buffer using a Gradient Station (BioComp Instruments), and allowed to cool for 30 minutes at 4 °C. The EscN-EscO mixture (200µL) was added to the top of the gradient and centrifuged at 367,600 ×g at 4 °C for 5 hours in a SW 55 Ti rotor (Beckman Coulter). The gradient was fractionated into 14 fractions using the Piston Gradient Fractionator (BioComp Instruments), and protein was visualized with the Triax UV detector (BioComp Instruments) and by SDS-PAGE. Fractions containing the assembled protein complex (approximately halfway down the gradient) were combined and desalted into cryo-EM buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 2% glycerol, 0.5 mM TCEP-HCl). The desalted protein complex was concentrated to ~2 mg/mL in Amicon Ultra centrifugal filters (EMD Millipore).

For analytical gradient centrifugation runs, the same protocol was used but with EscN at 2.0 mg/mL, and EscO at 0.7 mg/mL.

#### 4.2.3 Cryo-EM reconstruction of the EscN-EscO complex

Aliquots of 3 µl sample of the EscN-EscO complex were applied to glow-discharged (60s on carbon side) Quantifoil grids (Copper, 300 mesh, R1.2/1.3). The grids were blotted for 3 s at 100% humidity and plunge-frozen into liquid ethane using a Vitrobot Mark IV. Grids were imaged on a 300 kV Titan Krios cryo-electron microscope equipped with Gatan K2 Summit direct electron detector. Images were taken on the K2 camera in dose-fractionation mode at a calibrated magnification of 49020, corresponding to 1.02 Å per physical pixel (0.51 Å per super-resolution pixel). The dose rate on the specimen was set to be 9.61 electrons per Å<sup>2</sup> per s and total exposure time was 6 s, resulting in a total dose of 57.67 electrons per Å<sup>2</sup>. With dose

fractionation set at 0.2 s per frame, each movie series contained 30 frames and each frame received a dose of 1.92 electrons per  $\text{\AA}^2$ . Fully automated data collection was carried out using SerialEM with a nominal defocus range set from  $-1.2$  to  $-2.5 \mu\text{m}$ <sup>300</sup>.

A total of 2645 movies were collected at super resolution ( $0.51 \text{\AA}/\text{pixel}$ ). Initial processing was carried out in Relion3<sup>301</sup> using MotionCor2<sup>302</sup> to bin ( $1.02 \text{\AA}$  per pixel after binning) and align the 30 frames, and sum to a single micrograph with dose filtering; and CTFFIND4 to determine the contrast transfer functions<sup>303</sup>. Using Relion3,  $\sim 600,000$  particles were automatically picked, reference-free 2D classification was performed and  $\sim 198,000$  particles were kept in the good class averages. An *ab-initio* model was generated in cryoSPARC v2<sup>304</sup>. Relion3 was used for 3D classification without imposing any symmetry, which generated one good 3D map out of four with  $\sim 110,000$  particles. Within this good class, a second round of 3D classification was done in Relion3 that generated two distinct classes (class 1 with  $\sim 58,000$  particles and 2 with  $\sim 55,000$ ). 3D auto-refinements for classes 1 and 2 were subsequently done in Relion3 followed by CTF refinement and Bayesian Polishing<sup>305</sup>. Semi-automated post-processing of the refined maps, including automated soft masking, modulation transfer function and B-factor sharpening, were performed in Relion3 and yielded the final maps. Fourier Shell Correlation at a criteria of 0.143 reported a  $3.34 \text{\AA}$  resolution for class 1 map and a  $3.29 \text{\AA}$  resolution for class 2 map,, both using gold-standard refinement procedures and high-resolution noise substitution to correct soft-mask effects<sup>306</sup>.

#### 4.2.4 Model-building into cryo-EM maps

To build EscN<sup>29-446</sup> into the class 1 reconstruction ( $3.34 \text{\AA}$  resolution), an initial model based on the crystal structure of EscN<sup>103-446</sup> (2OBM)<sup>276</sup> was used in combination with a Phyre2<sup>307</sup> homology model of the N-terminal domain (residues 29-102) from flagellar homologue FliI

(5B0O)<sup>278</sup>. These models were first rigid-body fit into the map using Chimera<sup>308</sup>, then refined with phenix.real\_space\_refine<sup>309</sup>. Iterations of manual refinement using Coot<sup>310</sup> were followed by runs of phenix.real\_space\_refine and phenix.refine. Ligands were fit into the active sites and key interactions were linked using phenix.link\_edits.

The model of EscN<sup>29-446</sup> from the class 1 map was then refined into the class 2 map (3.29 Å resolution, with density of ~60 residues from EscO<sup>1-125</sup>) using phenix.real\_space\_refine. Two poly-alanine helices were then fit into the density of EscO using Coot, with their directionalities clearly visible based on the orientation of residue Cβs. The side chain density, especially for the regions extending into the EscN hexamer, was of sufficient quality to allow confident sequence assignment for residues 0 (remaining post cleavage affinity tag residue ordered) - 30 on helix 1 and 92 - 122 on helix 2.

Structural figures were made using PyMOL<sup>311</sup> and Chimera<sup>308</sup>, and the ABPS plugin<sup>312</sup> for electrostatic surface representation.

#### 4.2.5 EnzChek ATPase activity assays

Steady-state EscN ATPase activity was assayed using the EnzChek phosphate assay kit (Thermo Scientific). EscN<sup>29-446</sup> and EscO<sup>1-125</sup> (and mutants thereof) were incubated overnight in EscN-EscO buffer at a 1:1 molar ratio (25 μM each) at 4 °C to induce complex formation. The EscN-EscO incubation was performed in the presence of ADP-AIF<sub>3</sub> (1 mM ADP, 3 mM MgCl<sub>2</sub>, 6.25 mM KF, 1.25 mM AlCl<sub>3</sub>) to improve EscN<sup>29-446</sup> hexamerization unless otherwise specified; due to the inhibitory nature of ADP-AIF<sub>3</sub> combined with its requirement for complex stabilization, it was diluted out of the enzyme mixture immediately prior to the beginning of the assay (50× diluted in the final reaction mixture). All assays were performed at 37 °C at 3 mM ATP with 500 nM EscN (following EnzChek protocol), in 24 μL reactions in 384 well plates

(Corning 3540), using a Synergy H4 Microplate Reader (BioTek Instruments). Absorption reads at 360 nm were taken at 10 second intervals, and a linear portion over a one minute timespan was used to calculate relative steady-state reaction rate.

## 4.3 Results

### 4.3.1 Cryo-EM analysis of the EscN-EscO complex

Various constructs of EPEC EscN were screened for expression and purification, with the aim of isolating a stable oligomer for structural characterization. The most stable construct, EscN<sup>29-446</sup>, was predominantly monomeric but readily formed oligomers in the presence of the transition state analogue Mg<sup>2+</sup>ADP-AlF<sub>3</sub>, according to size exclusion chromatography, glycerol gradient centrifugation, and negative stain EM (Figure C.1). The catalytic activity of monomeric EscN was barely detectable at the 500 nM enzyme concentration used in our assay; however, pre-incubation with Mg<sup>2+</sup>ADP-AlF<sub>3</sub> (but omission from the reaction) resulted in a substantial increase in enzymatic activity, supporting the formation of functional oligomers (Figure C.2a–b). Full-length EPEC EscO was purified separately and found to further enhance EscN oligomerization and ATPase activity (Figure C.2). Studies have shown the V-ATPase central stalk protects A<sub>3</sub>B<sub>3</sub> heterohexamers against ATP-induced dissociation<sup>313</sup> and the *f*T3SS orthologue stabilizes the FliI ATPase complex<sup>279</sup>. Thus, the increased activity observed here may be due to similar stabilization of the EscN homohexamer. To isolate the complex, EscO was incubated in stoichiometric excess with EscN in the presence of Mg<sup>2+</sup>ADP-AlF<sub>3</sub>, with both co-sedimenting in a glycerol gradient. The complex formed visible rings (~10 nm diameter) in negative-stain TEM and cryo-EM (Figure C.1c, Figure C.3a).

The EscN-EscO complex was subjected to single-particle cryo-EM, with 3D classification resulting in two primary classes that refined to high resolution (Figure C.3). Class 1 (58,066

particles, 3.34 Å) reveals a homohexameric EscN pore lacking visible EscO density (Figure 4.1a–b, Figure C.3c–f), while class 2 (55,259 particles, 3.29 Å) looks similar but with additional  $\alpha$ -helical density for approximately half of a single EscO molecule resolved (Figure 4.1d–e, Figure C.3h–l, Figure C.4). Six copies of EscN were modelled into the class 1 reconstruction (EscN alone) guided by prior crystal structures of monomeric EscN and flagellar homologue FliI (PDBs 2OBM and 5B0O<sup>276,278</sup> Figure 4.2b). For clarity, we refer to the six chains of EscN as N<sub>A</sub> through N<sub>F</sub>. The final refined model spans residues 35–446 for all chains, excepting unresolved loop 323–329 on chain N<sub>F</sub> (Figure 4.2a, c). The EscN model remained similar when refined into the class 2 map (C $\alpha$  RMSD of ~0.5 Å over 2130 residues), with unfilled density from two associated anti-parallel helices attributed to EscO. The quality of the side chain density was sufficient for sequence assignment of EscO residues 0 (from affinity tag) to 30 and 91 to 122 out of a total of 125 residues (Figure 4.1e, Figure C.3i). The assigned EscO register fits well in the observed side chain densities (EMRinger<sup>314</sup> score for the EscO coordinates alone of 2.55), and agrees with prior experimental data<sup>280</sup>. In four of the six EscN inter-subunit active sites, the Mg<sup>2+</sup>ADP-AlF<sub>3</sub> transition state analogue was clearly visible (Figure 4.1c, 4.2c); the resolution was insufficient to differentiate between AlF<sub>3</sub> and AlF<sub>4</sub>, so AlF<sub>3</sub> was used as the more abundant species at pH 7.5<sup>315</sup>. The final overall models fit well to the EM data (with EMRinger scores of 3.5 and 3.4 for class 1 and 2) and good stereochemical indicators (Table 1).

Table 4.1: EscN-EscO cryo-EM data collection, refinement and validation statistics

	Class 1 (EMDB-9390) (PDB 6NJO)	Class 2 (EMDB-9391) (PDB 6NJP)
<b>Data collection and processing</b>		
Magnification	49020	49020
Voltage (kV)	300	300
Total dose (e <sup>-</sup> /Å <sup>2</sup> )	57.67	57.67
Dose per frame (e <sup>-</sup> /Å <sup>2</sup> )	1.92	1.92
Defocus range (µm)	-1.2 to -2.5	-1.2 to -2.5
Pixel size (super resolution) (Å)	1.02	1.02
Symmetry imposed	C1	C1
Initial number of particles	376000	376000
Final number of particles	58000	55000
Map resolution (Å)	3.34	3.29
FSC threshold	0.143	0.143
<b>Refinement</b>		
Initial model used (PDB code)	2OBM, 5B0O	
Model resolution (Å)	3.34	3.29
FSC threshold	0.143	0.143
Map sharpening <i>B</i> factor (Å <sup>2</sup> )	-98.5	-80.3
Model composition		
Non-hydrogen atoms	19090	19606
Protein residues	18942	19458
Ligands	128	128
<i>B</i> factors (Å <sup>2</sup> )		
Protein	55.8	55.3
Ligand	24.3	26.8
R.m.s. deviations		
Bond lengths (Å)	0.0100	0.0089
Bond angles (°)	1.45	1.40
Validation		
MolProbity score	1.78	1.77
Clashscore	5.91	6.23
Poor rotamers (%)	0.49	0.67
Ramachandran plot		
Favored (%)	92.90	93.55
Allowed (%)	7.10	6.45
Disallowed (%)	0.00	0.00

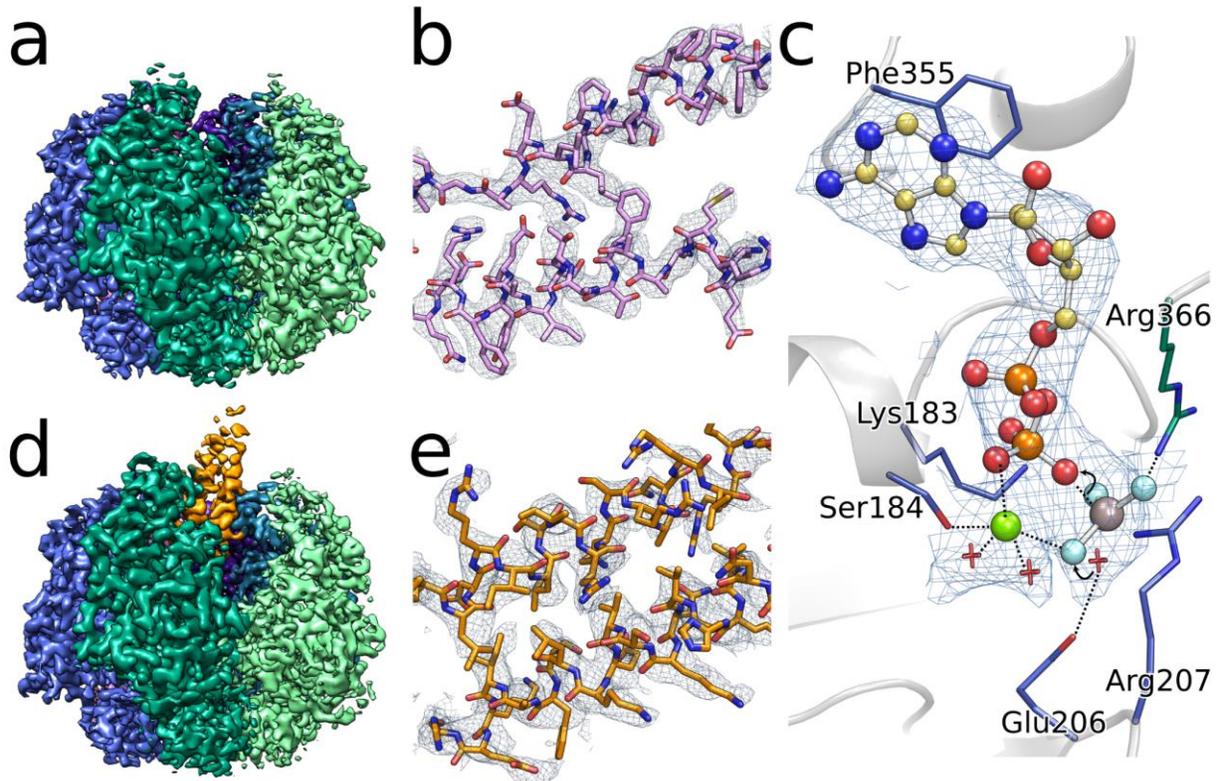


Figure 4.1: Cryo-EM density and resolution. Cryo-EM maps, coloured by subunit, of the (a) class one reconstruction comprising the EscN hexamer (without visible EscO density) at 3.34 Å resolution, and of the (d) class two reconstruction of the EscN-EscO complex at 3.29 Å resolution. Representative density is shown for (b) class one (EscN chain C residues 238–258 and 290–311) contoured at 6  $\sigma$  and (e) class two (EscO residues 1–19 and 103–122) contoured at 4  $\sigma$ . (c) Sample density of the class two catalytic site ligands ADP,  $\text{AlF}_3$  (grey and cyan), and  $\text{Mg}^{2+}$  (green), contoured at 4  $\sigma$ . Ligand coordination is represented with dotted lines, while the curved lines demarcate the reaction mechanism.

#### 4.3.2 EscN forms an asymmetric oligomer conserved with $F_1/V_1$ -ATPases

Each EscN monomer displays three sub-domains: the N-terminal oligomerization domain (residues 1-102), a cluster of curved  $\beta$ -sheets; the ATPase domain (103-372), made up of a central Rossmann fold with seven parallel  $\beta$ -strands flanked by four  $\alpha$ -helices on one side and three on the other; and the C-terminal domain (373-446), made up of four helices lining the central pore (Figure 4.2a). Comparison to the monomeric EscN<sup>103-446</sup> crystal structure lacking the

N-terminal domain<sup>276</sup> reveals the most prominent structural difference at the conserved C-terminal domain, where the helices are kinked away from the ATPase domain (Figure 4.2b). This difference likely stems from the helix-interrupting Val393Pro mutation, a screened functional mutant<sup>316</sup> required for crystallization in the earlier study.

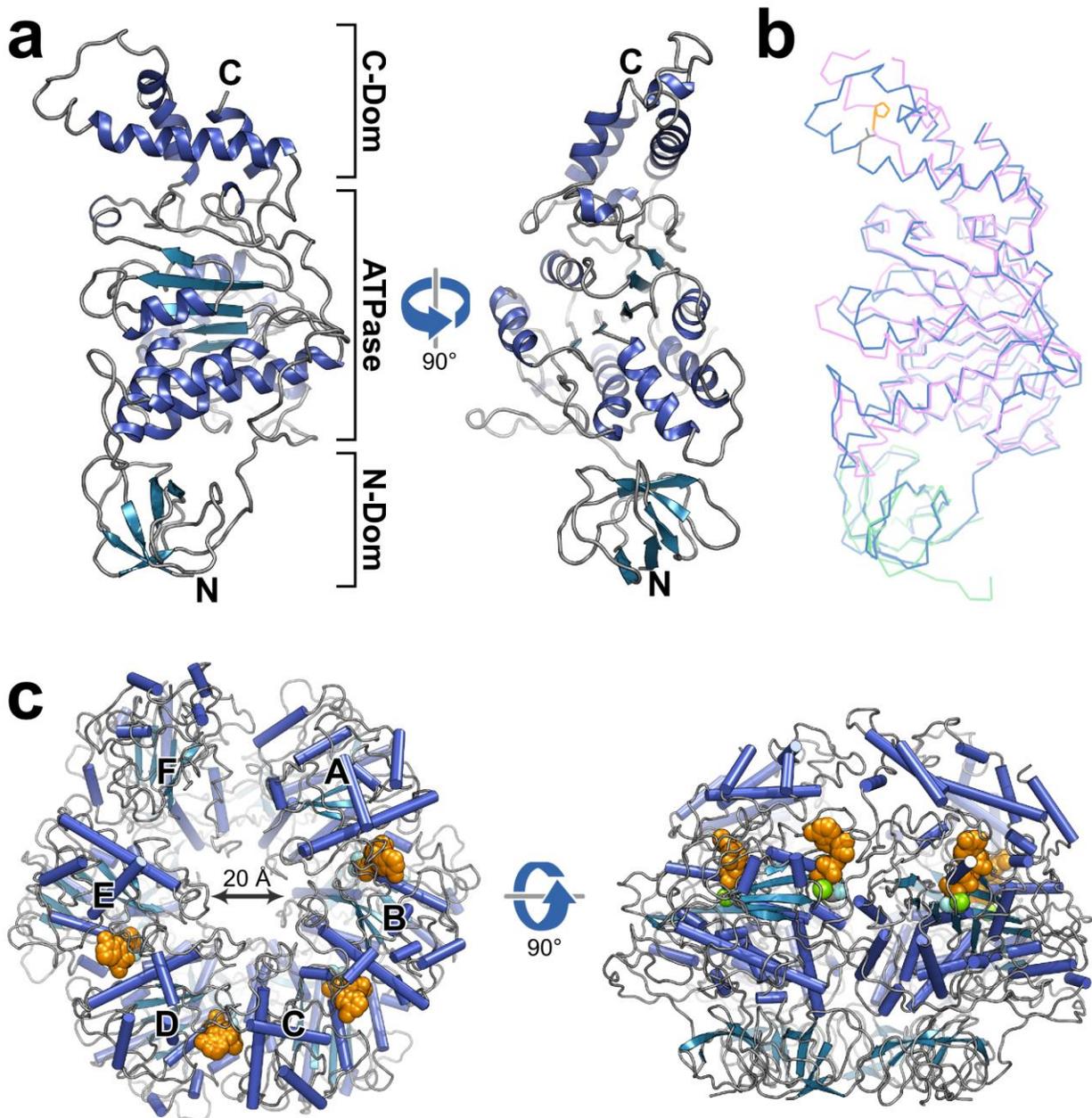


Figure 4.2: EscN structure. (a) Cartoon representation of the EscN chain D monomer from the class one cryo-EM model, shown from two views and coloured with  $\alpha$ -helices in blue,  $\beta$ -sheets

in teal, and loops in grey. **b)** Overlay of EscN chain D (blue) with EscN<sup>103-446</sup> crystal structure (pink, PDB 2OBM) and homology model from FliI (PDB 5B0O) N-terminal domain (green), highlighting a highly similar structure save for differences in the C-terminal domain due to the Val393Pro mutation in 2OBM (shown in orange). **(c)** Overview of hexameric EscN class one model from top and side views, with modelled ADP (orange), Mg<sup>2+</sup> (green), and aluminium fluoride (grey and cyan) shown as spheres.

---

The EscN hexameric structure shows marked asymmetry, with a prominent cleft between chains N<sub>A</sub> and N<sub>F</sub> (Figure 4.2c, 4.3a). This hexameric arrangement has not been captured in the previous crystal structures<sup>275,276,278,296-298</sup>, nor accurately predicted from subsequent modelling<sup>275,276,278,296,297</sup>. Comparison to published catalytic heterohexamers of the F<sub>1</sub>- and V<sub>1</sub>-ATPases reveals a remarkable degree of conservation in asymmetric subunit orientation (Figure C.5), despite their composition of alternating catalytic ( $\beta$ /A) and inert ( $\alpha$ /B) subunits.

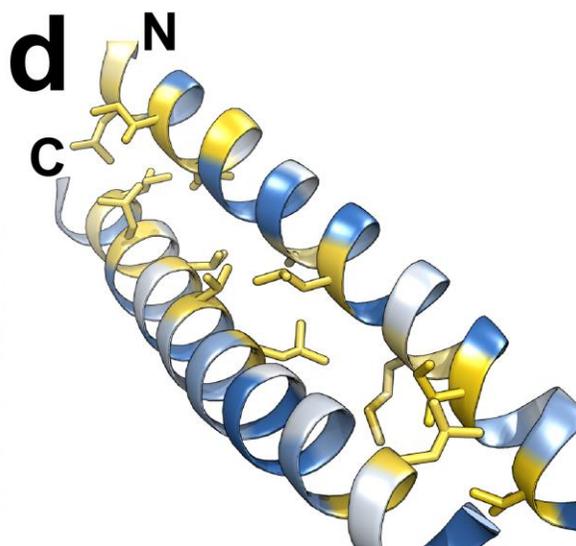
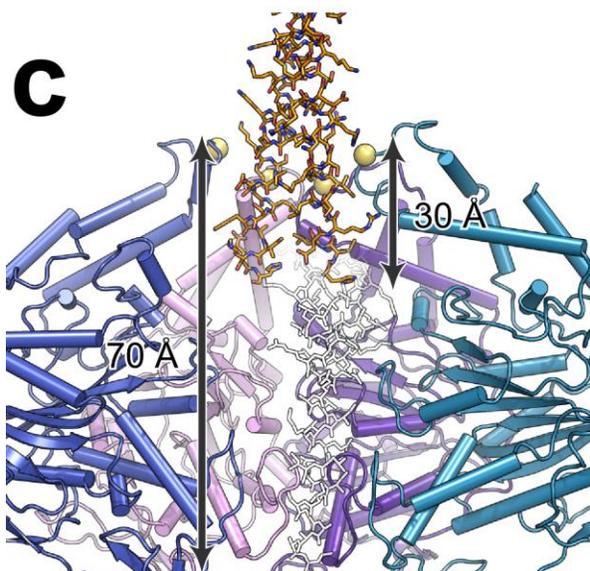
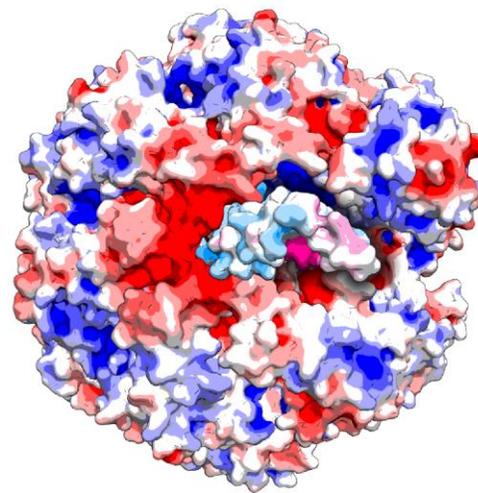
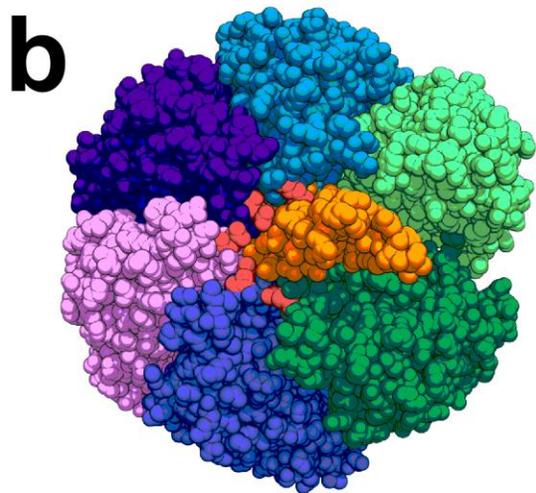
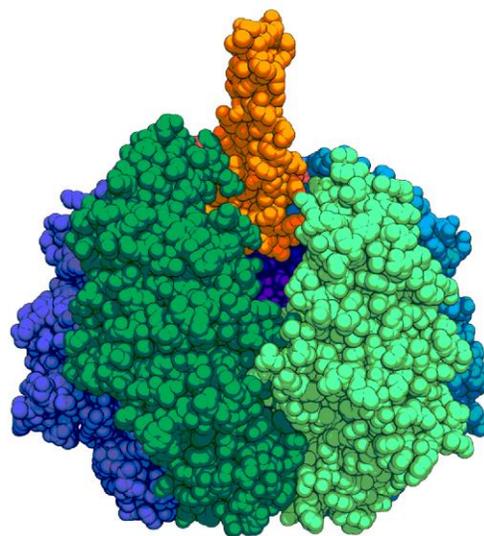
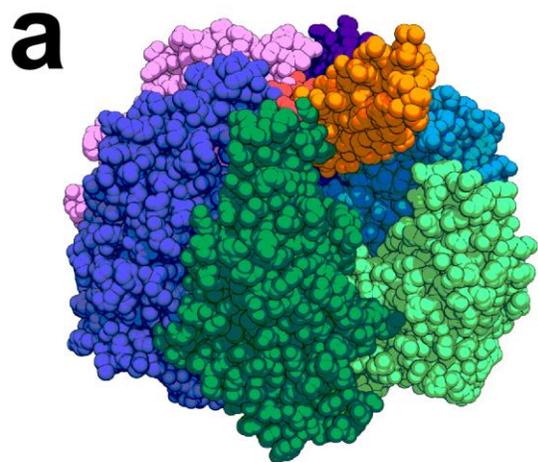
Superposition of the EscN hexamer with the bovine F<sub>1</sub>-ATPase bound to transition state analogue ADP-AlF<sub>3</sub> (PDB 1E1R<sup>315</sup>; ~28 % sequence identity with F<sub>1</sub> $\beta$ , N<sub>B</sub> C $\alpha$  RMSD ~1.5 Å over 314 residues) or the *E. hirae* V<sub>1</sub>-ATPase in complex with AMP-PNP (PDB 3VR6<sup>317</sup>; ~26 % sequence identity with V<sub>1</sub>A, N<sub>B</sub> C $\alpha$  RMSD ~1.3 Å over 311 residues) highlights the conserved variation in inter-subunit packing from open to tightly bound (Figure C.5). The functional significance of EscN's asymmetry is discussed below.

To probe where the key asymmetry-inducing conformational changes occur, careful comparison of each EscN subunit was undertaken. The N-terminal domain is largely static between subunits (Figure C.6, Figure C.7b), with approximate C<sub>6</sub> symmetry around the pore axis (Figure C.7a). Aligning each EscN chain by the N-terminal domain highlights a key pivot point between the N- and ATPase domains, with the ATPase and C-terminal domains essentially moving as a rigid body from subunit to subunit (Figure C.6, Figure C.7b). Subunit N<sub>F</sub>, the most dynamic monomer with the least resolved density, is tilted back nearly 30 degrees relative to the

most inward angled subunit N<sub>D</sub>, indicating impressive motion at this pivot point (Figure C.6a, Figure C.7b).

#### 4.3.3 The EscO coiled coil interacts at the EscN C-terminal domain

EscO is the proposed EPEC homologue of the *f*T3SS central stalk protein FliJ<sup>280</sup>, which in turn is homologous to the central stalk proteins of the rotary ATPases<sup>279,280,283</sup>. Consistently, the resolved region here (EscO residues 1-30 and 92-122) forms a coiled coil similar in nature to the coiled coil region of the F<sub>1</sub>-ATPase  $\gamma$ -subunit (1E1R)<sup>315</sup> and the V<sub>1</sub>-ATPase D subunit (3VR6)<sup>317</sup>, as well as to the isolated structures of T3SS central stalks FliJ (3AJW)<sup>279</sup>, CdsO (3K29)<sup>281</sup>, and YscO (4MH6)<sup>282</sup>; the typical coiled coil amphipathic side-chain packing is evident and supports our EscO structural model (Figure 4.3d, Figure C.4). Extending the structural and functional similarity to the rotary ATPases, both the N- and C-termini of the EscO coiled coil are observed to penetrate the EscN C-terminal pore opening (Figure 4.3c, Figure C.4). In contrast to the F<sub>1</sub>- and V<sub>1</sub>-ATPase complexes where the central stalk subunit extends ~70 Å into the catalytic core, EscO only penetrates ~30 Å, preventing interaction with all but the C-terminal domains (Figure 4.3c). Previous studies on F<sub>1</sub>-ATPase stalks have shown that truncation of the N-terminal 50 residues<sup>318</sup> or C-terminal 36 residues<sup>319</sup> in the  $\gamma$ -subunit from thermophilic *Bacillus* only resulted in an approximate 50 % decrease in torque generation. A 36-residue C-terminal  $\gamma$ -subunit truncation brings the end of the F<sub>1</sub> rotor a similar distance into the pore as EscO, suggesting that while the increased pore entry distance by the central stalk improves torque, it is not essential for function. Isolated structures of other T3SS EscO homologues have a C-terminal helix more similar in length to the F<sub>1</sub>- and V<sub>1</sub>-ATPase components, making EscO an outlier (Figure C.4). Counterintuitively, two studies



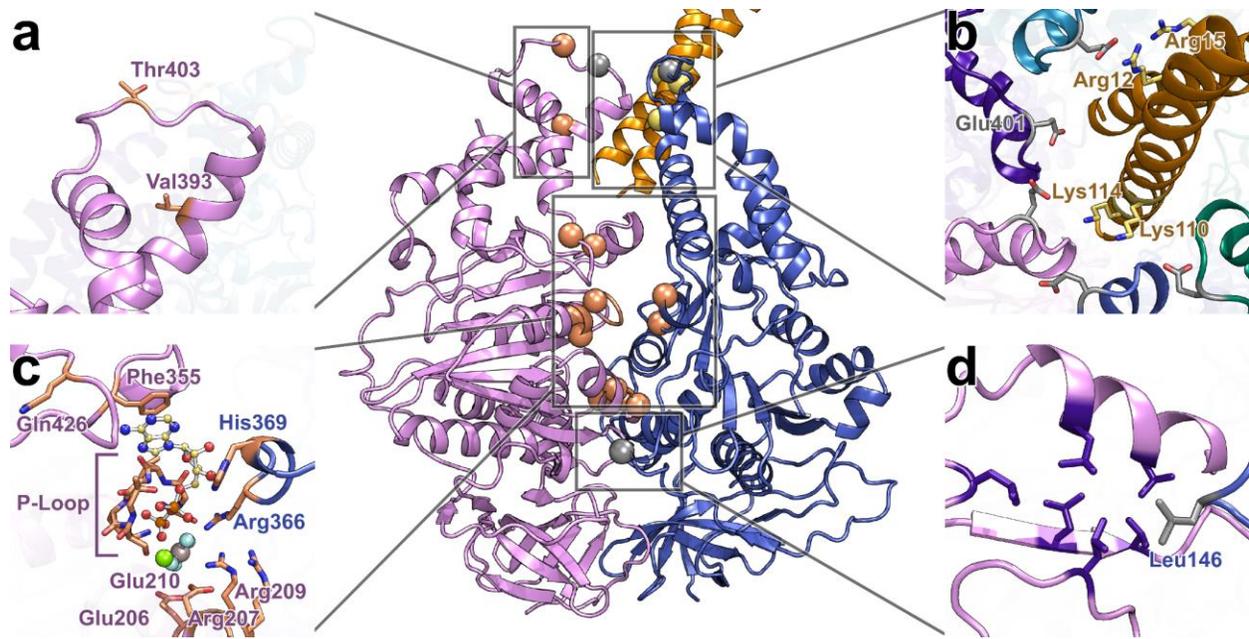
**Figure 4.3: Overview of EscN-EscO complex.** (a) Sphere-representation of the EscN-EscO complex (class 2), coloured by subunit and shown from an angled view and side view to highlight the cleft (located between the light and dark green subunits). The EscO stalk (orange) tilts towards the cleft. (b) Top-view of EscN-EscO complex, with negatively-charged glutamates shown in red lining the pore, compared with the APBS-calculated electrostatic surface demonstrating the complementary charges of EscN (red, white, blue) and EscO (pink, white, light blue). (c) Stick depiction of EscO insertion into the EscN pore, where it penetrates  $\sim 30$  Å; the F<sub>1</sub>  $\gamma$  subunit (PDB 1H8E) is overlaid in white, demonstrating its longer  $\sim 70$  Å extension into the F<sub>1</sub> ATPase pore. EscN Glu401 lining the pore is represented as yellow spheres. (d) EscO coloured by hydrophobicity, with hydrophobic residues coloured yellow and hydrophilic residues coloured teal; hydrophobic residues line the coiled coil interface, characteristic of this motif.

---

have demonstrated that the *Salmonella* EscO homologue InvI can be tagged with an N-terminal fluorescent protein (GFP) without impacting secretion<sup>320,321</sup>. Given the proximity of the EscO N-terminus to the EscN hexamer cleft opening, it is plausible that the fluorescent tag (even with a relatively short linker) could reside outside of the hexamer pore where it would not impede function. Alternatively, a C-terminally GFP-tagged construct of the V-ATPase central stalk homologue Vma8p was shown to be incorporated into functional V-ATPase complexes, suggesting that GFP can be accommodated within the central cavity without disrupting function<sup>322</sup>.

The EscN-EscO complex structure is sufficiently resolved to characterize the binding interface between stalk and ATPase. The two resolved helices of EscO are bound in a semicircle formed by the same EscN helix-loop region (residues Leu395–Glu401) from chains N<sub>A</sub> through N<sub>E</sub>, which congregate along one half of the pore due to EscN's tilt; this region in chain N<sub>F</sub>, the most loosely packed chain, is directly opposite the bound EscO and poorly ordered (Figure 4.3b, Figure C.8a–b). EscN creates a cradle of negative charge provided predominantly by clustered Glu401 from chains N<sub>A</sub> through N<sub>E</sub> (conserved in injectisome orthologues (Figure C.9) and structurally conserved with the F<sub>1</sub>  $\beta$ -subunit DELSEED motif), situated on the apex of the pore-

proximal loop (Figure 4.3b). In our model, EscO presents many electropositively-charged lysines and arginines near the EscN pore entrance; specifically, Glu401 in chains A-C are positioned to interact directly with EscO Lys110 and Lys114 (Figure 4.4b, Figure C.8c). In support, mutations of either EscN Glu401 to alanine or simultaneously mutated EscO Lys110 and Lys114 to glutamate abrogated the ability of EscO to form a stable complex with EscN as measured by co-sedimentation during glycerol gradient centrifugation (Figure C.2c–e). The EscO double lysine mutant resulted in a ~50 % decrease in catalytic rate, nearing that of EscN without EscO (Figure C.2a–b). Interestingly, Glu401Ala negatively affected the EscN ATPase rate both in the presence and absence of EscO, suggesting a cooperative role of the C-terminal domain (discussed below). In contrast, simultaneous mutation of conserved EscO Arg12 and Arg15 to glutamates (which interact with Glu401 on chain E, Figure 4.4b) had no significant effect on either complex formation or catalytic activity compared to wild-type EscN-EscO (Figure C.2a–b,d–e), indicating that their position is less important for the EscO-EscN interaction and function. In addition to these electrostatic interactions, conserved residues Leu396 and Ile399 (Figure C.9) contribute to a hydrophobic collar below the Glu401 ring. These residues form close interactions with hydrophobic residues on EscO (Figure C.8d), and may have a similar rotation-facilitating role as in F<sub>1</sub>-ATPases, where the homologous region was termed a molecular bearing<sup>323</sup>.



**Figure 4.4: Characterized EscN/EscO mutants.** Side-chain views of previously studied EscN mutants (shown in orange) mapped to the model, as well as EscN mutants (grey) and EscO mutants (yellow) characterized in this paper. The central view of the T and T' interface provides context for the location of each site, with detailed stick views denoting sidechain locations in the **(a)** chaperone binding site **(b)** EscN-EscO interaction interface, **(c)** active site, and **(d)** hydrophobic oligomerization interface.

#### 4.3.4 Active site architecture and implications for catalysis

The resolution of our EscN reconstructions has permitted detailed modelling of the active sites, including bound substrates and conserved active site waters (Figure 4.1c, 4.4c–d, Figure C.10) that reveal strong parallels with the F<sub>1</sub>- and V<sub>1</sub>-ATPase active site architectures. The EscN active site location at the subunit interface and key motifs are conserved with the latter, including the P-loop (residues 179-185), catalytic glutamate (Glu206), and arginine finger (Arg366) (Figure 4.4c, Figure C.9). Clear density is observed for the transition state analogue Mg<sup>2+</sup>ADP-ALF<sub>3</sub> in four of the six active sites, with the most detailed density observed for the chain N<sub>B</sub> active site at the interface with chain N<sub>A</sub>. EscN's ADP nucleotide is stabilized in the binding pocket by an aromatic stacking interaction with Phe355 (conserved in F<sub>1</sub>-ATPases (bovine

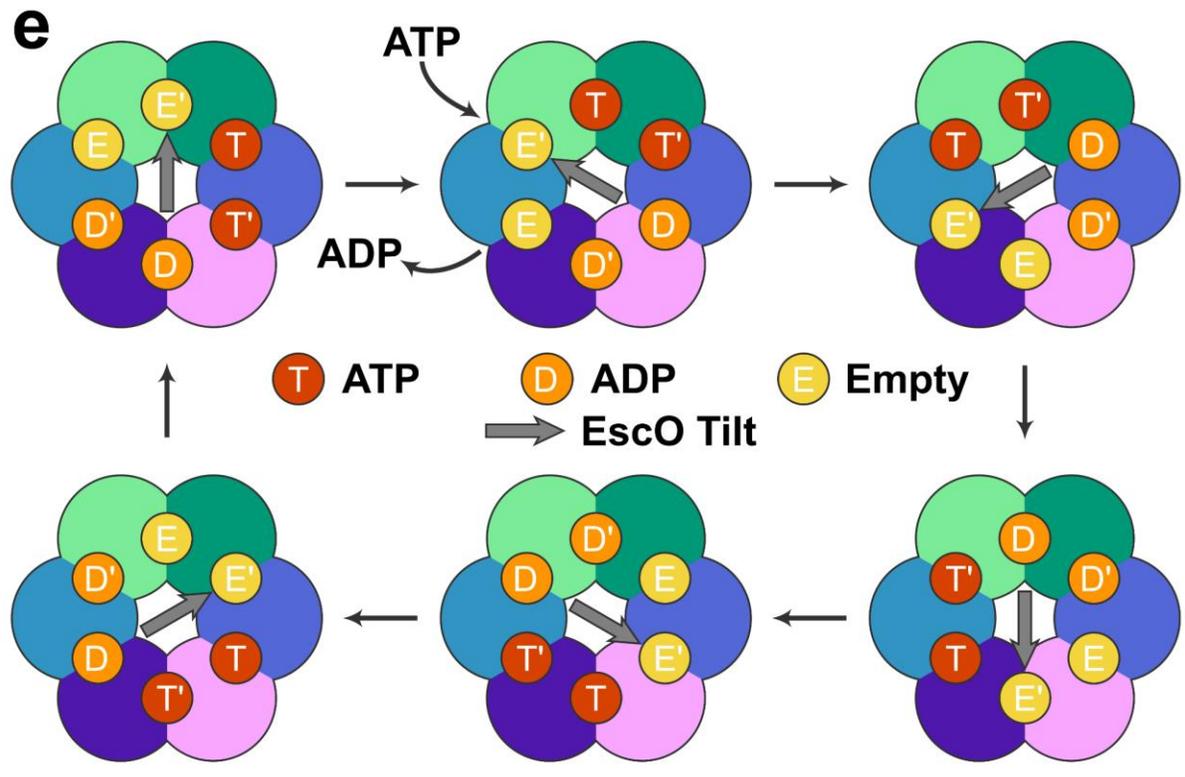
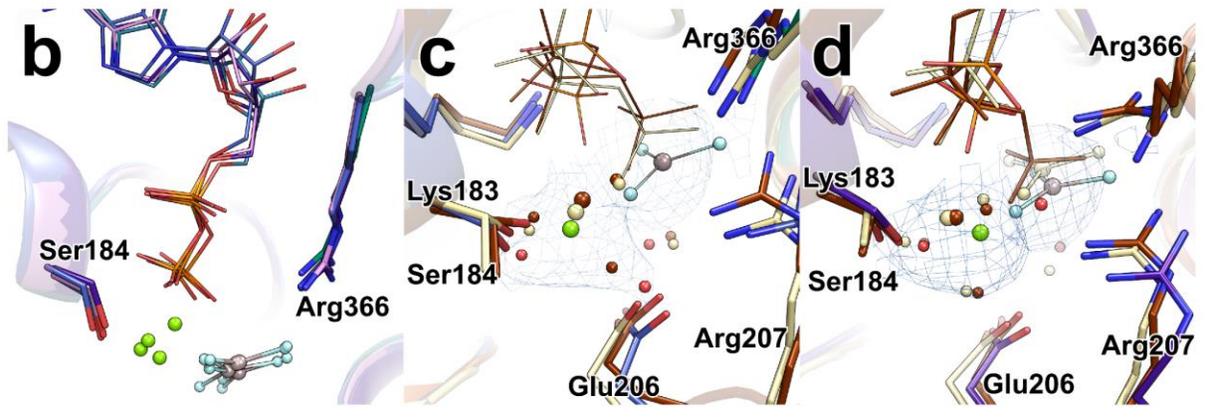
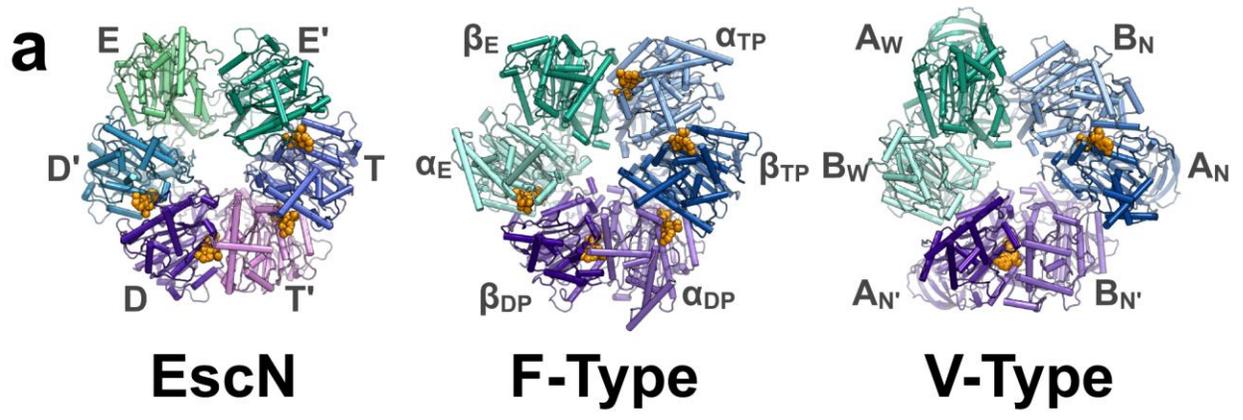
$\beta$ Tyr345)<sup>315</sup> and V<sub>1</sub>-ATPases (*E. hirae*  $\alpha$ Phe425))<sup>317</sup>, which accesses the site through a turn flanked by conserved His354 and Pro356. However, EscN lacks a large loop found in F<sub>1</sub>-ATPases (bovine  $\beta$ 419-427) and a similarly placed bulky phenylalanine in V<sub>1</sub>-ATPases (*E. hirae*  $\alpha$ Phe506) that stabilize the nucleotide ribose moiety. A looser nucleotide interface could contribute to EscN's observed nucleotide-dependent oligomerization, and may relate to the recycling of the T3S ATPase complex, which has been shown to be in dynamic exchange between a membrane associated hexameric assembly and a cytoplasmic hetero-trimeric complex with the peripheral stalk component<sup>324</sup>. The P-loop stabilizes the  $\alpha$ - and  $\beta$ -phosphates; multiple backbone amides point towards the  $\beta$ -phosphate, with conserved Gly182 facilitating the required tight loop conformation. Further, the planar AlF<sub>3</sub> aligns well with the position of the  $\gamma$ -phosphate (or analogues) in several bovine F<sub>1</sub>  $\beta$ <sub>TP</sub> sites (PDB 1E1R<sup>315</sup>, 1H8E<sup>325</sup>) and in the *E. hirae* V<sub>1</sub>-ATPase (3VR6) (Figure 4.5c–d). Coordination is by three conserved, electropositive residues with similar conformations amongst homologues: Lys183 from the P-loop, Arg207, and the arginine finger Arg366 from the preceding chain, essential for oligomerization (Figure 4.5c–d). The catalytic magnesium ion is coordinated by the side chain hydroxyl of Ser184 (homologous to the conserved P-loop threonine in F<sub>1</sub> and V<sub>1</sub>-ATPases), one fluoride from AlF<sub>3</sub>, and one  $\beta$ -phosphate oxygen. In ATP synthase, the remaining hexadentate coordination comes from three water molecules; variable B-factor sharpening here reveals additional density peaks around the magnesium ion, consistent with the coordinating waters that we have modelled (Figure 4.5c–d, Figure C.10), see below. Density for more active site waters is also observed at positions conserved with the F<sub>1</sub> active site, notably the key water bridging the catalytic Glu206 and the AlF<sub>3</sub> (Figure 4.5c–d, Figure C.10). Our data is in keeping with the classic catalytic ATPase mechanism, where Glu206 activates the catalytic water for nucleophilic attack on the ATP  $\gamma$ -

phosphate, while the  $Mg^{2+}$  and  $\gamma$ -phosphate-coordinating residues (Lys183, Arg207, and Arg366) draw electron density from the phosphorous atom to render it a more attractive electrophile (Figure 4.1c).

The assembled EscN-EscO structure determined here allows atomic level understanding of its similarities to and customized differences with the  $F_1/V_1$ -ATPases. Foremost, the different conformational states at each EscN active site mirror what has been observed in EscN's rotary relatives. Although EscN's homohexameric complex distinguishes it from the alternating inactive ( $\alpha/B$ ) and active ( $\beta/A$ ) subunits of  $F_1$  and  $V_1$ , clear conformational parallels are present. Historically, the three catalytic states in  $F_1$  are termed the empty state ( $\beta_E$ ), the ATP bound state ( $\beta_{TP}$ ) and the ADP bound state ( $\beta_{DP}$ ) based on the first structure with identifiable bound substrates<sup>323</sup>.  $V_1$  has analogous sites, described as wide-open ( $A_W$ ) and narrowly closed ( $A_N$  and  $A_{N'}$ )<sup>326</sup>. The empty site ( $\beta_E$  or  $A_W$ ) allows nucleotide exchange and transitions to the first nucleotide-bound site upon ATP binding ( $\beta_{TP}$  or  $A_N$ ), where the inter-subunit distance is tighter and the C-terminal domain (lever domain) moves from open to closed<sup>327</sup>. From there, the subunits show subtle active site repositioning as they shift to the proceeding  $\beta_{DP}$  or  $A_{N'}$  state, thought to reflect the ATP hydrolysis-competent conformation<sup>317,328</sup>. The cycling between these sites has been shown to generate torque in  $F_1$  and  $V_1$ -ATPases.

We observe a series of conformational changes between EscN's subunits and active sites consistent with the observed catalytic states of  $F_1/V_1$ -ATPases (Figure C.7c–d). Based on the EscN active site features and supported by similarities to the  $F_1/V_1$ -ATPase active sites discussed below, we will refer to the active site states as E (empty; equivalent to the  $F_1 \beta_E$  and  $V_1 A_W$  states), E' ( $F_1 \alpha_{TP}$ ,  $V_1 B_N$ ), T (analogous to ATP bound state;  $F_1 \beta_{TP}$ ,  $V_1 A_N$ ), T' ( $F_1 \alpha_{DP}$ ,  $V_1 B_{N'}$ ), D (analogous to ADP bound state;  $F_1 \beta_{DP}$ ,  $V_1 A_{N'}$ ) and D' ( $F_1 \alpha_E$ ,  $V_1 B_W$ ) (Figure 4.5a). The

EscN E state subunit is tilted furthest back from the pore and is the least well resolved, superposing with the F<sub>1</sub> β<sub>E</sub> and V<sub>1</sub> A<sub>W</sub> states (2.3 Å and 1.5 Å Cα RMSD respectively; Figure C.7c–d) and displaying no clear density for bound substrates. Moving clockwise, the EscN E' active site is positioned at the solvent-exposed cleft and is unoccupied by substrate, corresponding to the inactive F<sub>1</sub> α<sub>TP</sub> and V<sub>1</sub> B<sub>N</sub> sites (between the catalytically active empty and nucleotide-bound states). The E' state subunit conformation and active site accessibility is well placed for ATP binding, which would trigger the associated conformational changes. The subsequent four EscN sites show the tightest subunit interfaces, with prominent densities for ADP, Mg<sup>2+</sup>, and AlF<sub>3</sub> (Figure 4.5a); we propose these sites represent the progression from ATP binding, through hydrolysis, and eventually to the ADP-bound state. The interface area at each site gradually increases from the T site through the D' site, indicating a gradual tightening through the catalytic cycle. In both the F<sub>1</sub> and V<sub>1</sub> structures, the transitions from the β<sub>TP</sub>/A<sub>N</sub> to β<sub>DP</sub>/A<sub>N'</sub> states are characterized by a tighter subunit interface and accompanying shift of the catalytically critical arginine finger located on the neighbouring subunit; this arginine is thought to be involved in stabilizing the transiently formed penta-coordinated state of the γ-phosphate<sup>329</sup>. Comparison of the EscN active sites with bound substrates reveals a similar pattern, with the distance between Arg366 guanidinium carbon and the aluminium atom of AlF<sub>3</sub> decreasing by ~0.5 Å between the T and D active sites (Figure 4.5b). At this resolution, we observe the active site Mg<sup>2+</sup> to have pentadentate coordination with two water molecule ligands compared to the more typical hexadentate coordination that is observed in the F<sub>1</sub> and V<sub>1</sub> structures; this is consistent with the only published T3SS ATPase with a defined active site Mg<sup>2+</sup> coordination (*S. flexneri* Spa47<sup>84-430</sup> PDB 5YBI<sup>298</sup>). In EscN's T site, the Mg<sup>2+</sup> coordination shows clear trigonal bipyramidal geometry, with the water positions incrementally shifting to



**Figure 4.5: EscN catalysis. (a)** Top-view of EscN homohexamer without EscO compared to F<sub>1</sub> (1E1R) and V<sub>1</sub> (3VR6) heterohexamers, with subunit names labelled along the periphery; nucleotides are shown as orange spheres. **(b)** Comparison of the EscN inter-site distances for all four nucleotide bound sites, with C $\alpha$  distances between Ser184 and Arg366 progressively decreasing from 11.9, 11.2, 11.1, and 10.8 Å through sites T to D'. Overlay of EscN (subunits coloured as in (A)), F<sub>1</sub>-ATPase (coloured tan, 1E1R), and V<sub>1</sub>-ATPase (coloured brown, 3VR6) **(c)** T,  $\beta_{TP}$ , and A<sub>N</sub> sites and **(d)** D,  $\beta_{DP}$ , and A<sub>N'</sub> sites, demonstrating the similarities in side chain residues and water positions. Blue density is carved around Mg<sup>2+</sup> and waters from a class 2 map (B-factor sharpened by a factor of -150 and contoured at 10  $\sigma$ ), showing the broad density that encompasses the coordinating waters. **(e)** Possible rotational catalysis mechanism of EscN and EscO, with the central arrow representing the direction of EscO's tilt.

---

adopt a square pyramidal geometry in the D site (Figure C.10). Combining these details, we hypothesize that the T site (Figure 4.5c), having the loosest interface, houses freshly-bound ATP; the D site (Figure 4.5d), showing the most similarity to the  $\beta_{DP}/A_{N'}$  sites, represents the catalytic conformation<sup>328</sup>. Sites T' and D', unique to homohexameric rotary ATPases, represent accordingly the intermediate transitions between binding, hydrolysis, and ADP and/or phosphate release (Figure 4.5e).

#### 4.4 Discussion

The structure of the EscN homohexamer in complex with substrate analogue and inner stalk EscO represents a significant advance over previously available structures of monomeric orthologues<sup>275,276,278,296–298</sup>, which lacked the varied conformational states observed here and the interface with the inner stalk. The remarkable conservation of sequential active site states from F<sub>1</sub>/V<sub>1</sub>-ATPase heterohexamers provides direct atomic evidence of a related torque-generating catalytic cycle. A role for rotation of the *f*T3SS central stalk FliJ in flagellar assembly has been previously hypothesized but not directly shown, and remained a conundrum given the homohexameric nature. The *v*T3SS EscO stalk characterized here has been shown to

complement the *f*T3SS homologue FliJ *in vivo*, upregulating ATPase activity of the *f*T3SS FliI and partially rescuing motility of a *fliJ* deletion mutant<sup>280</sup>. FliJ, in turn, has been shown to bind and upregulate a V<sub>1</sub>-ATPase catalytic A<sub>3</sub>B<sub>3</sub> heterohexamer, and can be rotated in an imperfect manner (likely a result of its lower affinity binding)<sup>283</sup>. We propose that the EscN-EscO atomic structures presented here are therefore generally representative of the flagellar and virulence T3S ATPase variants, and directly support a role of torque generation and subsequent rotation of the central stalk in the process of T3SS.

In the F<sub>1</sub>- and V<sub>1</sub>-ATPases, the C-terminal helical lever domain is important for torque generation. ATP binding, hydrolysis, and release serve to induce conformational changes within this domain that provide the force to rotate the central stalk<sup>327</sup>, with hydrolysis of three ATP molecules resulting in one full rotation. The structure presented here shows that EscN can adopt similar C-terminal domain conformations; the lever is closed in nucleotide-bound sites T through D', and open in nucleotide-free sites E and E'. Large-scale conformational changes occur upon nucleotide binding (E' to T; closing toward the pore) and release (D' to E; opening away from the pore), fuelled by movement between EscN's N-terminal and ATPase domains (Figure C.6a, 8b). These alternating conformations create a rippling dynamic as the subunits undulate in and out during ATP hydrolysis. By analogy to the F<sub>1</sub>/V<sub>1</sub>-ATPases, this could be translated into torque on the EscO stalk delivered by the C-terminal helical domain. As EscN's subunits are pushed towards and away from the pore during a rotary cycle, the electronegative pockets we observe formed by successive Glu401s would circle around one half of the pore (Figure 4.3b); residues along the EscO stalk, particularly Lys110 and Lys114 (the former conserved with vT3SS central stalks<sup>280</sup>), likely follow these glutamates via complementary electrostatics, causing it to rotate (Figure C.8c). Simultaneously, hydrophobic interactions from the Leu396 and Ile399 molecular

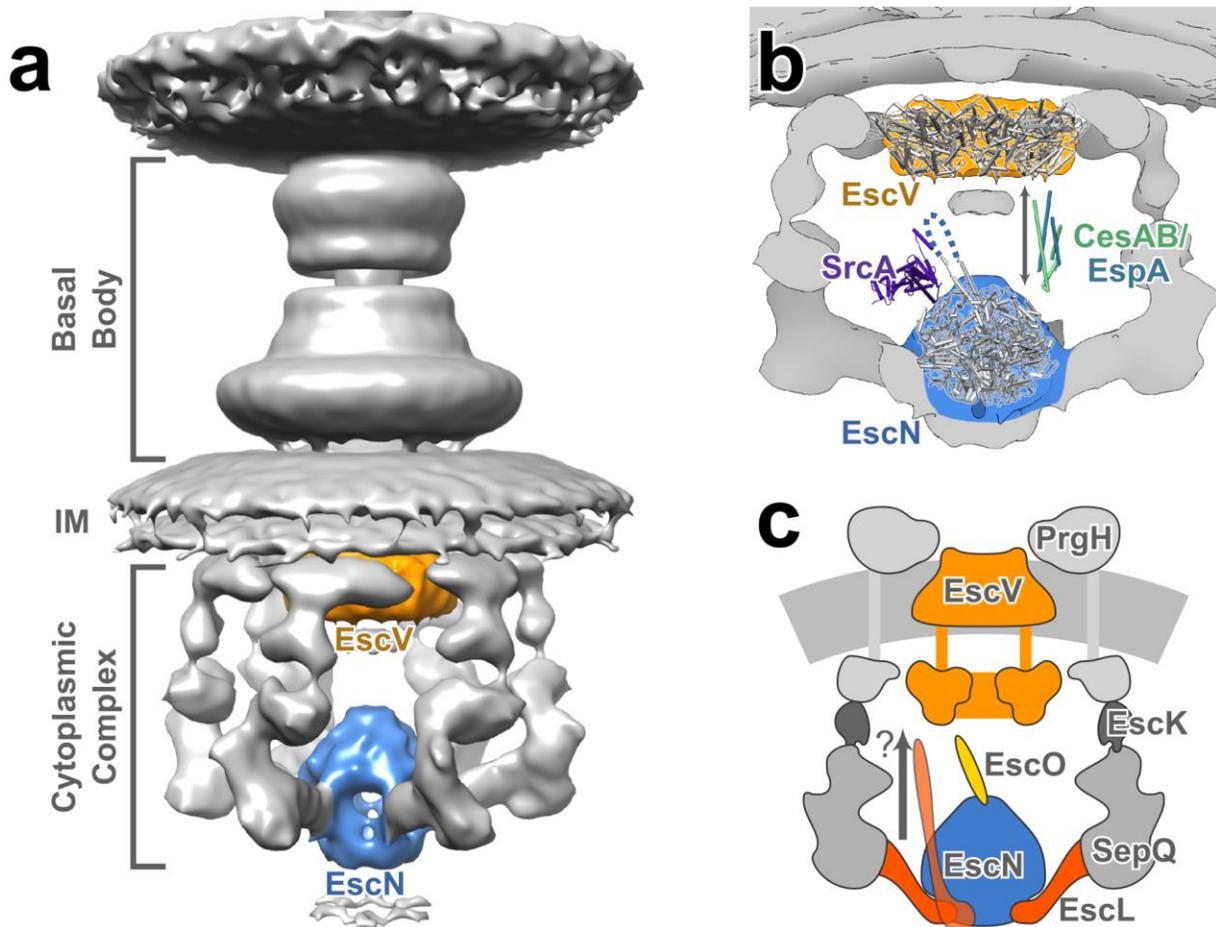
bearing<sup>323</sup> help to secure EscO within the pore while permitting smooth rotation (Figure C.8d). We hypothesize this interaction efficiently translates the torque from EscN's conformational changes into a rotational force on EscO. Our structural data provides the foundation to probe other customized differences that may be unique to the homohexameric T3SS catalytic cycle compared to the well-studied F- and V-type ATPases. The presence of six (presumably active) catalytic sites suggests EscN is less efficient than the latter, requiring hydrolysis of six molecules of ATP rather than three for a full 360° rotation. Future biophysical experiments will be needed to clarify how much torque is generated per ATP, and to identify if and where energy is being lost in this presumably more primitive homohexameric motor.

The precise role(s) played by the ATPase in assembly and function of the T3S injectisome and flagellum remain poorly understood. At the heart of both nanomachines, the conserved T3SS export apparatus consists of a membrane embedded export apparatus (EscRSTUV or FliPQR, FlhB, and FlhA respectively) and the cytoplasmic ATPase complex. T3S requires both ATP hydrolysis and the movement of protons across the cytoplasmic membrane by the export gate for efficient secretion<sup>290,330</sup>. PMF is the prominent energy source, with disruption by the proton ionophore CCCP inhibiting secretion<sup>287</sup>. The export gate EscV (flagellar FlhA) is composed of a nonameric cytoplasmic domain and a structurally uncharacterized transmembrane region proposed to be a potential proton channel and function as a proton/protein antiporter<sup>331</sup>. Export gate FlhA has been shown to be capable of translocating protons and sodium ions<sup>331</sup> and supports secretion even in the absence of the ATPase in certain mutant backgrounds<sup>332</sup>. The ATPase, however, improves the efficiency of this process, with interaction of the ATPase complex required to fully activate the export gate<sup>289</sup>. These studies point to a PMF-dependent cooperative mechanism for T3S energetics involving both the export gate and the ATPase complex.

In keeping, direct interactions between the ATPase complex and export gate have been documented. Data from both the  $\nu$ T3SS<sup>333</sup> and the  $f$ T3SS<sup>289</sup> suggests the central stalk protein directly interacts with the major export gate protein EscV/FlhA<sup>334</sup>, which could draw the ATPase-central stalk complex within interacting distance of the gate. Specifically, the central stalk FliJ in complex with ATPase FliI and peripheral stalk FliH interact with FlhA through conserved Phe72 and Leu76<sup>335</sup>, while mutants mapping to the same region of *P. aeruginosa* central stalk PscO upregulated secretion via a suggested interaction with the export gate PcrD<sup>333</sup>. For both, it was hypothesized that the interaction increased the efficiency of substrate secretion by modulating the efficiency with which the T3SS uses the PMF. The region implicated in the interaction of the central stalk and export gate is not resolved in our structure, although FliJ residues Phe72 and Leu76 are conserved in EscO (Tyr51 and Leu55, Figure C.4). Assuming a continuous coiled coil motif in keeping with the characteristic primary sequence, these residues would be positioned approximately at the apex of the EscO structure, ideal for interaction with the export gate. Such an interaction could allow EscO to affect the conformational state of EscV, or to facilitate passage of substrates from EscN to EscV. Remarkably, recent experiments have shown that the ATPase is capable of energizing secretion even in the absence of bulk PMF<sup>336</sup>; based on this observation, it was proposed that the ATPase complex may itself be involved in the generation of a local proton gradient that is used by the export gate. Our structure of the EscN-EscO complex and its extensive structural homology with the F<sub>1</sub>/V<sub>1</sub>-ATPases (which can fuel outward pumping of protons) supports such a hypothesis. These similarities raise the intriguing possibility that elements of the export gate might function in a manner similar to the F<sub>o</sub>/V<sub>o</sub> components, acting as a channel through which protons are pumped across the inner membrane.

A further role of the T3S ATPase in substrate targeting and unfolding has been suggested from prior work. Most T3S effector substrates are localized to the T3S apparatus in a partially unwound state, a result of complex formation with specialized chaperones<sup>294</sup>. Studies have demonstrated that some of these chaperone-effector complexes can both bind to the T3S ATPase and be dissociated and unfolded in an ATP dependent manner<sup>291–293</sup> (typically mediated by the chaperone whether alone<sup>296,337</sup> or in complex with effector<sup>337,338</sup>, although interaction of the effector signal-sequence with the ATPase has also been demonstrated<sup>339,340</sup>). Published mutations mapping to the EscN C-terminal domain in our structures here have suggested an important role in substrate recognition<sup>276,291,296,341</sup>. Specifically, Allison *et al.* identified mutants in a conserved region of the C-terminal helical domain of *S. Typhimurium* SPI-2 T3S ATPase SsaN that abrogated interaction with the multicargo chaperone SrcA<sup>296</sup> (Figure 4.4a, 4.6c). Our structures of the EscN hexamer and its varying catalytic states provides a potential structural basis by which ATP hydrolysis induced effector-chaperone dissociation can be achieved. As discussed above, we observe a significant degree of conformational change in the C-terminal domain throughout the catalytic cycle detailed by our structure and we propose that these, especially the transition to the empty (E, E') states, will disrupt chaperone binding to the effector and/or ATPase C-terminal domain. In the case of EPEC CesAB-EspA, Chen *et al.*<sup>338</sup> found that it bound only to oligomerized EscN (with deletion of EscN's N-terminal oligomerization domain abrogating the interaction), suggesting some chaperones may selectively bind subunit conformations only adopted in the assembled state. The function of the central stalk in this is not clear; interactions of specific chaperones and central stalk homologues have been demonstrated<sup>342,343</sup>, and Allison *et al.* identified the SrcA binding site based on its structural similarity to the globular portion of the F<sub>1</sub>-ATPase central stalk  $\gamma$ -subunit<sup>296</sup> (absent in the

coiled-coil structure of EscO and other T3S homologues; Figure C.4). For the studies with  $\nu$ T3SS ATPase homologues demonstrating chaperone-effector dissociation<sup>291–293</sup> the central stalk component was not included, suggesting the ATP hydrolysis induced conformational changes alone are sufficient. Stabilization of the hexamer by the central stalk could potentiate this or, given the proposed role in activation of the export gate discussed above, the central stalk could facilitate communication between the ATPase and export gate upon effector release (Figure 4.6b). Indeed, some chaperone-effector complexes have been shown to bind both the ATPase and the export gate including  $\nu$ T3SS CesAB (chaperone)-EspA (EPEC filament)<sup>338,344</sup> and  $f$ T3SS FliT (chaperone)-FliD (flagellar filament cap)<sup>341,345</sup> suggesting a cooperative role in chaperone recruitment, targeting and secretion regulation (Figure 4.6). FliT-FliD in particular has been shown to bind to monomeric  $f$ T3SS ATPase FliI in the presence of peripheral stalk FliH, a complex hypothesized to facilitate escort of the effector to the export gate<sup>341</sup>. In keeping with this, a recent study has demonstrated that  $f$ T3SS ATPase subunits are in a continual state of exchange between membrane associated and cytoplasmic complexes<sup>324</sup>. As discussed above, the looser inter-subunit active site packing and shallower central stalk binding of the EscN-EscO complex compared with F/V-ATPases could allow for readier subunit dissociation to facilitate this exchange. Finally, we note that the structures of both CesAB-EspA and FliT-FliD complexes have coiled-coil structures with similar characteristics to the central stalk components (PDBs 1XOU<sup>346</sup> and 6CH2<sup>345</sup>, Figure C.4) and it is plausible that some chaperones may interact with the ATPase in a manner related to the observed EscO interaction here, potentially providing an additional path for subsequent targeting to the export gate. The latter possibility is intriguing given the breadth of apparatus and effector substrates requiring passage in the various T3SS species.



**Figure 4.6: Function of EscN-EscO in the injectisome. (a)** Overview of EscN and EscV location in the context of the T3SS nanomachine (based on *in situ* tomography of the *Salmonella* injectisome, EMDB 8544). **(b)** Putative binding sites of type III chaperone-substrate complexes, based on previous studies on SrcA/SsaN and CesA/EspA/EscV. Each binding site brings the cargo into the lumen between EscN and EscV. The unresolved region of EscO is represented by a dotted line. **(c)** Schematic of proteins present in the cytoplasmic subcomplex. It has been hypothesized that the entire ATPase complex may rise towards the inner membrane to interact with the export gate.

In summary, we present the atomic details of an assembled T3SS ATPase in complex with its central stalk protein. The structure provides a molecular foundation for the differential rotational states of the homohexamer, for chaperone/substrate recognition, and for subsequent ATPase-driven secretion initiation. In turn, we can begin to probe deeper into the underlying biophysics of T3SS ATPase torque generation and rotational mechanism and how this ancient protein

family has been preserved through its cooperation with the T3SS in widespread bacterial pathogenicity.

## 5 Conclusion

### 5.1 Summary

The complete structural characterization of the injectisome is a real possibility in the near future. With the feasible resolution of cryo-EM improving in steady increments each year, the biggest limitation lies in our ability to purify a stable, homogeneous complex for single-particle analysis. The structures and biochemical experiments described in this thesis were designed to characterize binding interfaces between injectisomal proteins and within oligomeric assemblies, making use of structural techniques best suited for each case. In doing so, we have gained understanding of how several of these proteins interact, information that can contribute to the development of targeted therapeutics in the future.

The assemblies characterized in this work come from two subdomains of the injectisome: the basal body and the sorting platform. The pilotin InvH works to localize the OM secretin pore by interacting through its hydrophobic groove with an amphipathic helix on the InvG S domain. In its apo form, InvH forms a homodimer along the same interface, perhaps protecting the hydrophobic surface prior to InvG binding. At the sorting platform, the cytosolic domain of the IM pore EscV forms a nonameric ring held together through electrostatic interactions. Based on chaperone binding sites in the flagellar homologue FlhA, EscV would likely undergo a conformational change to accommodate a chaperone interaction. At the base of the system, EscN forms a homohexameric rotary ATPase complex. Hydrolysis of ATP likely results in the rotation of the coiled coil stalk EscO, but how this process dissociates chaperones from their effectors is unclear. The following sections will summarize the conclusions from each paper and focus on identifying future directions for each of these projects.

## 5.2 The pilotin InvH

InvH forms a helical bundle distinct from other characterized pilotin structures. In its apo state *in vitro*, it forms a primarily dimeric population where the helix of one protomer covers the hydrophobic cleft of its binding partner. This hydrophobic patch is also the binding interface with the InvG S domain, resulting in the dimer dissociating upon secretin binding. The InvH-InvG interface is lined with many hydrophobic residues and a key InvG tryptophan, mutation of which decreased binding affinity nearly 2000-fold. The C-terminal helix of the S domain interacts with InvH, whereas the remainder of the helix-loop-helix motif remains disordered until secretin assembly is complete.

No other pilotin has been shown to form a dimer in solution. As InvH binds InvG in a monomeric state, it is unclear what function the dimerization serves. One hypothesis is that it simply protects the extensive hydrophobic surface from solvent exposure. A similar mechanism occurs in the *S. flexneri* pilotin MxiM, which binds to its cognate secretin through a deep hydrophobic cleft. The crystallized MxiM had a molecule of lipid bound in this cleft, leading to the proposed mechanism wherein MxiM binds its own lipidated N-terminus until it interacts with the secretin MxiD<sup>347,348</sup>. This proposed mechanism posits that MxiM forms a heterotrimer with the secretin and the localization of lipoproteins pathway LolA. LolB would then insert MxiM into the OM, and MxiD would undergo BAM-independent OM insertion. However, as the pilotin-secretin-LolA complex has not been observed, there is another possible mechanism for secretin localization: the pilotin may be inserted into the OM prior to secretin binding, using its long N-terminal linker to interact with the S domain as the secretin passes through the SEC translocon<sup>133</sup>. The latter proposal provides another possible explanation for the dimerization of InvH. As the secretin forms a 15-mer pore (possibly with 16 subunits in its periplasmic rings<sup>58</sup>),

an increased local concentration of InvH through dimerization would facilitate bringing 15 InvG protomers together. With the diversity of pilotin structures and linker lengths, it is possible that both mechanisms may exist in different species.

Several microbiology experiments could clarify the InvH-InvG localization mechanism *in vivo*. To identify whether the dimerization interface is physiologically important, secretion studies can be performed in *S. Typhimurium* on InvH knockouts complemented with dimerization-deficient mutants. Similar studies can be performed on the InvG binding mutants, identified *in vitro* through ITC experiments, to confirm the importance of these residues on injectisome assembly. Experiments with the LoL pathway can be performed to determine whether InvH forms a complex with InvG and LolA, supporting the co-localization hypothesis. The LolA R43L mutant prevents insertion of the lipoprotein into the membrane, and was used to study how the T2SS pilotin PulS localizes secretin PulD<sup>133</sup>. Affinity purification of complemented His-LolA-R43L in *S. Typhimurium* followed by immunoblotting for InvH and InvG would clarify what complex is transported by LolA, and separation of the IM and OM through gradient centrifugation would probe localization. To test the hypothesis that InvH is localized to the OM before secretin binding, mutations on the linker region of InvH would be informative. If shortening the linker results in progressively less efficient InvG localization, it could indicate that the probability of InvH reaching InvG across the periplasm is decreasing. However, if InvH is co-localized with InvG to the OM, the linker length need only be sufficient to bridge the gap between the OM and S domain and gradually shortening it should have less of an effect on assembly. A strong microbiological analysis of InvH and InvG is still required to determine the exact mechanism by which the secretin pore is assembled in *Salmonella*.

### 5.3 The inner membrane pore EscV

The IM-pore protein is a multi-functional mystery, and the only basal body protein yet to be characterized completely. It has roles in chaperone binding, substrate hierarchy, and potentially as a proton pore through the IM<sup>162,175,177,349</sup>, and has been proposed to be responsible for coupling the energy from proton motive force to substrate export<sup>180</sup>. The cryo-EM structure of the EPEC EscV cytosolic domain (EscV<sub>C</sub>) demonstrates that EscV<sub>C</sub> forms a nonameric ring in solution. The oligomerization interface is primarily electrostatic, similar to what has been observed in crystallized *S. flexneri* homologue MxiA, and only relies on two of its four subdomains. The remaining two subdomains are in a closed state, with the only open state structure observed in flagellar homologue FlhA<sup>350</sup>.

EscV<sub>C</sub> has many proposed binding partners that have yet to be probed and structurally characterized. It has been shown to interact with the gatekeeper proteins SepL and SepD, which lowers its affinity for late chaperone complex CesT/Tir while increasing affinity for middle chaperone complex CesAB/EspA<sup>162</sup>. The only structure of the IM pore in complex with chaperones comes from flagellar FlhA in a monomeric state. Performing cryo-EM analysis on EscV<sub>C</sub> in complex with SepL-SepD and in complex with middle or late chaperones would permit analysis of conformational changes that alter its substrate specificity, providing insight into how secretion hierarchy is controlled. Residues important to EscV<sub>C</sub> oligomerization and protein interaction could be identified through biophysical techniques such as ITC and SEC-MALS, then probed *in vivo* through secretion assays with EPEC EscV mutants. These experiments would provide a much deeper understanding of how the IM pore is involved in substrate selection.

The transmembrane region of the IM pore has yet to be structurally characterized. The full-length construct of the IM pore is difficult to purify in any species, including EscV, which is not

surprising as it is predicted to have seven or eight transmembrane helices and form a nonamer<sup>180</sup>. The attempts I made to purify it had yield too low even for single-particle cryo-EM analysis, and the protein was unstable and prone to aggregation. Optimization of its purification, both recombinantly and from native expression, are ongoing in the lab. The structure of the EscV transmembrane region, especially in complex with the export apparatus, would be extremely informative. It could reveal whether there is indeed a proton channel in this protein<sup>175,177</sup>, clarifying its role in the coupling of proton motive force to secretion. The interaction with the export apparatus could provide insight into how the closed pore in characterized structures<sup>58,66-68</sup> could be opened to accommodate effectors. Analysis of any interaction between the transmembrane and cytosolic domain at various pHs, even at low resolution, would determine whether the cytosolic domain indeed moves closer to the transmembrane region as a result of local proton concentration<sup>180</sup>. As the protein is difficult to work with, these experiments will take some time to fulfil; once completed, they will provide a wealth of information on how injectisome secretion is energized.

#### 5.4 The ATPase complex

The EPEC ATPase EscN forms a homohexameric, asymmetrical ring complex with EscO bound in its pore. This architecture confirmed the homology of the sequence and monomeric structures to F<sub>1</sub>- and V-ATPases, taking on different conformations at each of its six interprotomer nucleotide binding sites. Based on this similarity, it is likely that EscN undergoes rotary catalysis, with coiled coil EscO acting as the rotor. The chaperone binding site in the *S. Typhimurium* SPI-2 system maps to an interface near EscO<sup>82</sup>, suggesting that the effector could interact with the stalk to mediate its separation from the chaperone. As previous experiments on

flagella have shown that the ATPase complex has a role in PMF-mediated secretion<sup>177–179</sup>, it is possible that the stalk couples ATP hydrolysis to proton transport as in rotary ATPases.

Previously studied rotary ATPases have been characterized extensively through biophysical experiments. To confirm whether EscN works to rotate EscO, some assays of this nature would be helpful. The simplest assay involves affixing EscN to a substrate, such as mica, and using labelled EscO to identify whether rotation occurs during ATP hydrolysis. This labelling can be done by linking fluorescent actin to the predicted apex of EscO through streptavidin, followed by fluorescence microscopy<sup>351</sup>. Such an experiment would confirm that rotation occurs, while also probing its speed and direction. More sophisticated experiments can be performed using high-speed imaging, such as atomic force microscopy<sup>352</sup>, total internal reflection dark-field illumination using a gold colloidal bead probe<sup>353</sup>, or single-photon counting of light scattering intensities using a gold nanorod probe<sup>354</sup>. These experiments would provide much more detailed information on the thermodynamic steps of ATP hydrolysis and the angle of rotation produced per hydrolysis event, but require specialized equipment. Biophysical analysis of the injectisome ATPase complex would provide interesting information on how this homohexameric system functions in contrast to heterohexameric F<sub>1</sub>- and V-ATPases.

Information on the proteins anchoring the injectisome ATPase in the cytosol is lacking. EscN is held in place by six pods made up of peripheral stalk SctL, cytoplasmic ring SctQ, and possibly ATPase cofactor SctK<sup>90,91</sup>. The best resolution of the cytosolic complex comes from sub-tomogram averaging<sup>57,79</sup>, as the entire subregion has failed to remain associated with purified needle complexes analyzed by cryo-EM<sup>58,61</sup>. The most direct way to characterize this complex would be by single-particle cryo-EM, either through a gentler purification of native needle complexes, heterologous co-expression of the desired components, or recombinant

expression and purification of each separate component to reconstitute the complex *in vitro*. As the sorting platform is associated with the IM ring SctD, which itself requires interaction with the export apparatus for stability, it may be difficult to purify the entire complex in the absence of the basal body. However, even the SctL-EscN-EscO complex would be informative to observe how the peripheral stalk stabilizes the EscN catalytic domains, which would be realistically attainable as the monomeric SctL-SctN complex has been characterized in flagella<sup>92</sup>.

Interactions between the ATPase and chaperone-effector complexes have not been structurally characterized. Using either a catalytic mutant or inhibited form of the EscN-EscO complex, it may be possible to create a stable assembly with the chaperone for single-particle analysis. A molecular snapshot of this interaction could reveal how the ATPase separates the effector prior to secretion. As the cytosolic domain of IM pore EscV also interacts with chaperones, it would be interesting to also add EscV<sub>C</sub> to the ATPase-chaperone complex. It is possible that both EscV<sub>C</sub> and EscN would bind the chaperone-effector complex simultaneously, or that one of them would have the higher affinity of the two. As the stalk has been shown to interact with the IM pore in flagella<sup>355</sup>, it could bridge the two proteins and interact with both EscN and EscV<sub>C</sub>. I have attempted to combine the EscN-EscO complex with purified EscV, which resulted in the previously stable assemblies precipitating despite identical buffer conditions; this result is promising, suggesting that the proteins do interact. Structural analysis of this system, including addition of various chaperone-effector complexes, would provide valuable insight into how the ATPase and IM pore cooperate to prepare effectors for secretion.

## 5.5 Closing Remarks

The goal of this body of work was to further our understanding of the injectisome using structural methods. Through characterization of InvH, we have uncovered a novel pilotin fold

with a homodimeric stoichiometry, adding to the possible mechanisms pilotins may use to assemble the secretin pore. Work on the IM pore EscV has confirmed that the protein's cytosolic domain forms a nonameric ring in solution, with a conformation suggesting that opening of two subdomains is required to accommodate chaperone-effector binding. The cryo-EM structure of the ATPase-stalk complex supports a rotational catalytic mechanism similar to F<sub>1</sub>- and V-ATPases, with interesting implications in secretin energization and substrate selection.

The results from this thesis provide a strong base for future projects in the lab. While the next steps for the InvH project are mainly microbiology experiments, there remains much to be understood about the structure of the cytosolic sorting platform. The purification protocols of EscV<sub>C</sub> and the EscN-EscO complex are a foundation for cryo-EM characterization of co-complexes with chaperones and effectors. Such structures would provide many clues to how substrates are selected, including any conformational changes involved in EscV<sub>C</sub>. A key direction to be explored is the co-structure of EscV<sub>C</sub> and the ATPase complex, which could clarify the mystery of how proton motive force and ATP hydrolysis are coupled to secretion. Uncovering the mechanisms involved in substrate selection and secretion energization will contribute not only to our understanding of type III secretion, but to the functions behind all bacterial nanomachines.

## References

1. Dik, D. A., Fisher, J. F. & Mobashery, S. Cell-Wall Recycling of the Gram-Negative Bacteria and the Nexus to Antibiotic Resistance. *Chem. Rev.* **118**, 5952–5984 (2018).
2. Hobot, J. A. Bacterial Ultrastructure. in *Molecular Medical Microbiology* 7–32 (Elsevier, 2015). doi:10.1016/B978-0-12-397169-2.00002-0.
3. Christie, P. J. The Rich Tapestry of Bacterial Protein Translocation Systems. *Protein J* **38**, 389–408 (2019).
4. Green, E. R. & Meccas, J. Bacterial Secretion Systems: An Overview. in *Virulence Mechanisms of Bacterial Pathogens, Fifth Edition* (eds. Kudva, I. T. et al.) 215–239 (American Society of Microbiology, 2016). doi:10.1128/microbiolspec.VMBF-0012-2015.
5. Korotkov, K. V. & Sandkvist, M. Architecture, Function, and Substrates of the Type II Secretion System. in *Protein Secretion in Bacteria* (eds. Sandkvist, Cascales & Christie) 227–244 (American Society of Microbiology, 2019). doi:10.1128/ecosalplus.ESP-0034-2018.
6. Meuskens, I., Saragliadis, A., Leo, J. C. & Linke, D. Type V Secretion Systems: An Overview of Passenger Domain Functions. *Front. Microbiol.* **10**, 1163 (2019).
7. Tsirigotaki, A., De Geyter, J., Šoštaric', N., Economou, A. & Karamanou, S. Protein export through the bacterial Sec pathway. *Nat Rev Microbiol* **15**, 21–36 (2017).
8. Frain, K. M., Robinson, C. & van Dijl, J. M. Transport of Folded Proteins by the Tat System. *Protein J* **38**, 377–388 (2019).
9. Deng, W. *et al.* Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology* **15**, 323–337 (2017).
10. Sgro, G. G. *et al.* Bacteria-Killing Type IV Secretion Systems. *Front. Microbiol.* **10**, 1078 (2019).
11. Coulthurst, S. The Type VI secretion system: a versatile bacterial weapon. *Microbiology* **165**, 503–515 (2019).
12. Nakamura, S. & Minamino, T. Flagella-Driven Motility of Bacteria. *Biomolecules* **9**, 279 (2019).
13. Minamino, T. Protein export through the bacterial flagellar type III export pathway. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 1642–1648 (2014).
14. Lyons, B. J. E. & Strynadka, N. C. J. On the road to structure-based development of anti-virulence therapeutics targeting the type III secretion system injectisome. *Med. Chem. Commun.* **10**, 1273–1289 (2019).
15. Brown, N. F. & Finlay, B. Potential origins and horizontal transfer of type III secretion systems and effectors. *Mobile Genetic Elements* **1**, 118–121 (2011).
16. Page, A.-L. & Parsot, C. Chaperones of the type III secretion pathway: jacks of all trades: Chaperones of the type III secretion pathway. *Molecular Microbiology* **46**, 1–11 (2002).
17. Fattori, J., Prando, A., Martini Martins, A., Henrique dos Santos Rodrigues, F. & Tasic, L. Bacterial Secretion Chaperones. *PPL* **18**, 158–166 (2011).

18. Gibani, M. M., Britto, C. & Pollard, A. J. Typhoid and paratyphoid fever: a call to action. *Current Opinion in Infectious Diseases* **31**, 440–448 (2018).
19. Fowler, C. C. *et al.* Emerging insights into the biology of typhoid toxin. *Current Opinion in Microbiology* **35**, 70–77 (2017).
20. Roth, G. A. *et al.* Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1736–1788 (2018).
21. Stanaway, J. D. *et al.* The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Infectious Diseases* **19**, 1312–1324 (2019).
22. Mała, Ł. & Popowska, M. Antimicrobial resistance of *Salmonella* spp. isolated from food. *Rocz Panstw Zakl Hig* **67**, 343–358 (2016).
23. Dahiya, S. *et al.* Current antibiotic use in the treatment of enteric fever in children. *Indian J Med Res* **149**, 263 (2019).
24. Pucciarelli, M. G. & García-Del Portillo, F. Salmonella Intracellular Lifestyles and Their Impact on Host-to-Host Transmission. *Microbiology Spectrum* **5**, (2017).
25. LaRock, D. L., Chaudhary, A. & Miller, S. I. Salmonellae interactions with host processes. *Nat Rev Microbiol* **13**, 191–205 (2015).
26. Schlumberger, M. C. & Hardt, W.-D. Triggered Phagocytosis by Salmonella: Bacterial Molecular Mimicry of RhoGTPase Activation/Deactivation. in *Bacterial Virulence Factors and Rho GTPases* (eds. Boquet, P. & Lemichez, E.) vol. 291 29–42 (Springer-Verlag, 2005).
27. Dougan, G. & Baker, S. *Salmonella enterica* Serovar Typhi and the Pathogenesis of Typhoid Fever. *Annu. Rev. Microbiol.* **68**, 317–336 (2014).
28. Liss, V. & Hensel, M. Take the tube: remodelling of the endosomal system by intracellular *Salmonella enterica*: Intracellular *Salmonella*. *Cell Microbiol* **17**, 639–647 (2015).
29. Garcia-del Portillo, F., Zwick, M. B., Leung, K. Y. & Finlay, B. B. Salmonella induces the formation of filamentous structures containing lysosomal membrane glycoproteins in epithelial cells. *Proceedings of the National Academy of Sciences* **90**, 10544–10548 (1993).
30. Johnson, R., Mylona, E. & Frankel, G. Typhoidal *Salmonella* : Distinctive virulence factors and pathogenesis. *Cellular Microbiology* **20**, e12939 (2018).
31. Bishop, A. *et al.* Interaction of *Salmonella enterica* serovar Typhi with cultured epithelial cells: roles of surface structures in adhesion and invasion. *Microbiology* **154**, 1914–1926 (2008).
32. Forest, C. G., Ferraro, E., Sabbagh, S. C. & Daigle, F. Intracellular survival of *Salmonella enterica* serovar Typhi in human macrophages is independent of *Salmonella* pathogenicity island (SPI)-2. *Microbiology* **156**, 3689–3698 (2010).
33. Troeger, C. *et al.* Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Infectious Diseases* **17**, 909–948 (2017).

34. Pacheco, A. R. *et al.* CRISPR Screen Reveals that EHEC's T3SS and Shiga Toxin Rely on Shared Host Factors for Infection. *mBio* **9**, e01003-18, /mbio/9/3/mBio.01003-18.atom (2018).
35. Gaytán, M. O., Martínez-Santos, V. I., Soto, E. & González-Pedrajo, B. Type Three Secretion System in Attaching and Effacing Pathogens. *Frontiers in Cellular and Infection Microbiology* **6**, (2016).
36. Kaper, J. B. & O'Brien, A. D. Overview and Historical Perspectives. *Microbiology Spectrum* **2**, (2014).
37. Levine, M. M. *et al.* Diarrhoeal disease and subsequent risk of death in infants and children residing in low-income and middle-income countries: analysis of the GEMS case-control study and 12-month GEMS-1A follow-on study. *The Lancet Global Health* **8**, e204–e214 (2020).
38. McDaniel, T. K. & Kaper, J. B. **A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12 .** *Molecular Microbiology* **23**, 399–407 (1997).
39. Elliott, S. J., Yu, J. & Kaper, J. B. The cloned locus of enterocyte effacement from enterohemorrhagic *Escherichia coli* O157:H7 is unable to confer the attaching and effacing phenotype upon *E. coli* K-12. *Infect. Immun.* **67**, 4260–4263 (1999).
40. Kenny, B. *et al.* Enteropathogenic *E. coli* (EPEC) Transfers Its Receptor for Intimate Adherence into Mammalian Cells. *Cell* **91**, 511–520 (1997).
41. Vallance, B. A. & Finlay, B. B. Exploitation of host cells by enteropathogenic *Escherichiacoli*. *Proceedings of the National Academy of Sciences* **97**, 8799–8806 (2000).
42. Cepeda-Molero, M. *et al.* Attaching and effacing (A/E) lesion formation by enteropathogenic *E. coli* on human intestinal mucosa is dependent on non-LEE effectors. *PLoS Pathog* **13**, e1006706 (2017).
43. Ugalde-Silva, P., Gonzalez-Lugo, O. & Navarro-Garcia, F. Tight Junction Disruption Induced by Type 3 Secretion System Effectors Injected by Enteropathogenic and Enterohemorrhagic *Escherichia coli*. *Front. Cell. Infect. Microbiol.* **6**, (2016).
44. Berger, C. N. *et al.* EspZ of Enteropathogenic and Enterohemorrhagic *Escherichia coli* Regulates Type III Secretion System Protein Translocation. *mBio* **3**, e00317-12 (2012).
45. Roxas, J. L. *et al.* The Enteropathogenic *Escherichia coli*-Secreted Protein EspZ Inhibits Host Cell Apoptosis. *Infect. Immun.* **80**, 3850–3857 (2012).
46. Ali, S. O. *et al.* Phase 1 study of MEDI3902, an investigational anti-*Pseudomonas aeruginosa* PcrV and Psl bispecific human monoclonal antibody, in healthy adults. *Clinical Microbiology and Infection* **25**, 629.e1-629.e6 (2019).
47. DiGiandomenico, A. *et al.* Identification of broadly protective human antibodies to *Pseudomonas aeruginosa* exopolysaccharide Psl by phenotypic screening. *The Journal of Experimental Medicine* **209**, 1273–1287 (2012).
48. DiGiandomenico, A. *et al.* A multifunctional bispecific antibody protects against *Pseudomonas aeruginosa*. *Science Translational Medicine* **6**, 262ra155-262ra155 (2014).

49. Chabloz, A. *et al.* Salmonella-based platform for efficient delivery of functional binding proteins to the cytosol. *Commun Biol* **3**, 342 (2020).
50. Aguilera-Herce, J., García-Quintanilla, M., Romero-Flores, R., McConnell, M. J. & Ramos-Morales, F. A Live *Salmonella* Vaccine Delivering PcrV through the Type III Secretion System Protects against *Pseudomonas aeruginosa*. *mSphere* **4**, e00116-19, /msphere/4/2/mSphere116-19.atom (2019).
51. Martinez-Becerra, F. J. *et al.* Characterization and Protective Efficacy of Type III Secretion Proteins as a Broadly Protective Subunit Vaccine against *Salmonella enterica* Serotypes. *Infect Immun* **86**, e00473-17, /iai/86/3/e00473-17.atom (2018).
52. Dehghani, B., Rasooli, I., Jalali-Nadoushan, M., Owlia, P. & Rasooli, Z. Immunoprotectivity of *Salmonella enterica* serovar Enteritidis virulence protein, InvH, against *Salmonella typhi*. *Iran J Basic Med Sci* **17**, 560–565 (2014).
53. Allen, K. J., Rogan, D., Finlay, B. B., Potter, A. A. & Asper, D. J. Vaccination with type III secreted proteins leads to decreased shedding in calves after experimental infection with *Escherichia coli* O157. *Can. J. Vet. Res.* **75**, 98–105 (2011).
54. Cheng, Y. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell* **161**, 450–457 (2015).
55. Bai, X., McMullan, G. & Scheres, S. H. W. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences* **40**, 49–57 (2015).
56. Kubori, T. Supramolecular Structure of the *Salmonella typhimurium* Type III Protein Secretion System. *Science* **280**, 602–605 (1998).
57. Butan, C., Lara-Tejero, M., Li, W., Liu, J. & Galán, J. E. High-resolution view of the type III secretion export apparatus in situ reveals membrane remodeling and a secretion pathway. *Proc Natl Acad Sci USA* **116**, 24786–24795 (2019).
58. Hu, J. *et al.* T3S injectisome needle complex structures in four distinct states reveal the basis of membrane coupling and assembly. *Nat Microbiol* **4**, 2010–2019 (2019).
59. Worrall, L. J. *et al.* Near-atomic-resolution cryo-EM analysis of the *Salmonella* T3S injectisome basal body. *Nature* **540**, 597–601 (2016).
60. Spreter, T. *et al.* A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nature Structural & Molecular Biology* **16**, 468–476 (2009).
61. Schraidt, O. & Marlovits, T. C. Three-Dimensional Model of *Salmonella*'s Needle Complex at Subnanometer Resolution. *Science* **331**, 1192–1195 (2011).
62. Marlovits, T. C. *et al.* Structural Insights into the Assembly of the Type III Secretion Needle Complex. *Science* **306**, 1040–1042 (2004).
63. Bergeron, J. R. C. *et al.* A Refined Model of the Prototypical *Salmonella* SPI-1 T3SS Basal Body Reveals the Molecular Basis for Its Assembly. *PLoS Pathogens* **9**, e1003307 (2013).
64. Bergeron, J. R. C. *et al.* The Modular Structure of the Inner-Membrane Ring Component PrgK Facilitates Assembly of the Type III Secretion System Basal Body. *Structure* **23**, 161–172 (2015).

65. Zeytuni, N. *et al.* Near-atomic resolution cryoelectron microscopy structure of the 30-fold homooligomeric SpoIIIAG channel essential to spore formation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* **114**, E7073–E7081 (2017).
66. Kuhlen, L. *et al.* Structure of the core of the type III secretion system export apparatus. *Nat Struct Mol Biol* **25**, 583–590 (2018).
67. Lunelli, M. *et al.* Cryo-EM structure of the Shigella type III needle complex. *PLoS Pathog* **16**, e1008263 (2020).
68. Johnson, S., Kuhlen, L., Deme, J. C., Abrusci, P. & Lea, S. M. The Structure of an Injectisome Export Gate Demonstrates Conservation of Architecture in the Core Export Gate between Flagellar and Virulence Type III Secretion Systems. *mBio* **10**, e00818-19, /mbio/10/3/mBio.00818-19.atom (2019).
69. Kuhlen, L. *et al.* The substrate specificity switch FlhB assembles onto the export gate to regulate type three secretion. *Nat Commun* **11**, 1296 (2020).
70. Marlovits, T. C. *et al.* Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* **441**, 637–640 (2006).
71. Burkinshaw, B. J. *et al.* Structural Analysis of a Specialized Type III Secretion System Peptidoglycan-cleaving Enzyme. *J. Biol. Chem.* **290**, 10406–10417 (2015).
72. Lilic, M., Quezada, C. M. & Stebbins, C. E. A conserved domain in type III secretion links the cytoplasmic domain of InvA to elements of the basal body. *Acta Crystallogr D Biol Crystallogr* **66**, 709–713 (2010).
73. Worrall, L. J., Vuckovic, M. & Strynadka, N. C. J. Crystal structure of the C-terminal domain of the Salmonella type III secretion system export apparatus protein InvA. *Protein Science* **19**, 1091–1096 (2010).
74. Abrusci, P. *et al.* Architecture of the major component of the type III secretion system export apparatus. *Nature Structural & Molecular Biology* **20**, 99–104 (2013).
75. Pallen, M. J. Evolutionary links between FliH/YscL-like proteins from bacterial type III secretion systems and second-stalk components of the FoF1 and vacuolar ATPases. *Protein Science* **15**, 935–941 (2006).
76. Ibuki, T. *et al.* Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. *Nature Structural & Molecular Biology* **18**, 277–282 (2011).
77. Kishikawa, J. *et al.* Common Evolutionary Origin for the Rotor Domain of Rotary ATPases and Flagellar Protein Export Apparatus. *PLoS ONE* **8**, e64695 (2013).
78. Hu, B. *et al.* Visualization of the type III secretion sorting platform of *Shigella flexneri*. *Proceedings of the National Academy of Sciences* **112**, 1047–1052 (2015).
79. Hu, B., Lara-Tejero, M., Kong, Q., Galán, J. E. & Liu, J. In Situ Molecular Architecture of the Salmonella Type III Secretion Machine. *Cell* **168**, 1065-1074.e10 (2017).
80. Akeda, Y. & Galán, J. E. Chaperone release and unfolding of substrates in type III secretion. *Nature* **437**, 911–915 (2005).

81. Zarivach, R., Vuckovic, M., Deng, W., Finlay, B. B. & Strynadka, N. C. J. Structural analysis of a prototypical ATPase from the type III secretion system. *Nature Structural & Molecular Biology* **14**, 131–137 (2007).
82. Allison, S. E. *et al.* Identification of the Docking Site between a Type III Secretion System ATPase and a Chaperone for Effector Cargo. *Journal of Biological Chemistry* **289**, 23734–23744 (2014).
83. Burgess, J. L. *et al.* Structural and Biochemical Characterization of Spa47 Provides Mechanistic Insight into Type III Secretion System ATPase Activation and *Shigella* Virulence Regulation. *Journal of Biological Chemistry* **291**, 25837–25852 (2016).
84. Gao, X. *et al.* Structural Insight Into Conformational Changes Induced by ATP Binding in a Type III Secretion-Associated ATPase From *Shigella flexneri*. *Frontiers in Microbiology* **9**, (2018).
85. Demler, H. J. *et al.* Interfacial amino acids support Spa47 oligomerization and *shigella* type three secretion system activation. *Proteins* **87**, 931–942 (2019).
86. Majewski, D. D. *et al.* Cryo-EM structure of the homohexameric T3SS ATPase-central stalk complex reveals rotary ATPase-like asymmetry. *Nat Commun* **10**, 626 (2019).
87. Lorenzini, E. *et al.* Structure and Protein-Protein Interaction Studies on *Chlamydia trachomatis* Protein CT670 (YscO Homolog). *Journal of Bacteriology* **192**, 2746–2756 (2010).
88. Minasov, G. *et al.* 2.8 Angstrom Crystal Structure of Type III Secretion Protein YscO from *Vibrio parahaemolyticus*. (2013) doi:10.2210/pdb4mh6/pdb.
89. Romo-Castillo, M. *et al.* EscO, a Functional and Structural Analog of the Flagellar FliJ Protein, Is a Positive Regulator of EscN ATPase Activity of the Enteropathogenic *Escherichia coli* Injectisome. *Journal of Bacteriology* **196**, 2227–2241 (2014).
90. Biemans-Oldehinkel, E., Sal-Man, N., Deng, W., Foster, L. J. & Finlay, B. B. Quantitative Proteomic Analysis Reveals Formation of an EscL-EscQ-EscN Type III Complex in Enteropathogenic *Escherichia coli*. *Journal of Bacteriology* **193**, 5514–5519 (2011).
91. Soto, E. *et al.* Functional Characterization of EscK (Orf4), a Sorting Platform Component of the Enteropathogenic *Escherichia coli* Injectisome. *J Bacteriol* **199**, e00538-16, e00538-16 (2017).
92. Imada, K., Minamino, T., Uchida, Y., Kinoshita, M. & Namba, K. Insight into the flagella type III export revealed by the complex structure of the type III ATPase and its regulator. *Proceedings of the National Academy of Sciences* **113**, 3633–3638 (2016).
93. Notti, R. Q., Bhattacharya, S., Lilic, M. & Stebbins, C. E. A common assembly module in injectisome and flagellar type III secretion sorting platforms. *Nat Commun* **6**, 7125 (2015).
94. Hu, J. *et al.* Cryo-EM analysis of the T3S injectisome reveals the structure of the needle and open secretin. *Nat Commun* **9**, 3840 (2018).
95. Crepin, V. F., Martinez, E., Shaw, R. K., Frankel, G. & Daniell, S. J. Structural and Functional Properties of Chimeric EspA-FliCi Filaments of EPEC. *Journal of Molecular Biology* **378**, 243–250 (2008).

96. Nadler, H. *et al.* Deciphering the Mechanical Properties of Type III Secretion System EspA Protein by Single Molecule Force Spectroscopy. *Langmuir* **34**, 6261–6270 (2018).
97. Deane, J. E. *et al.* Molecular model of a type III secretion system needle: Implications for host-cell sensing. *Proceedings of the National Academy of Sciences* **103**, 12529–12533 (2006).
98. Zhang, L., Wang, Y., Picking, W. L., Picking, W. D. & De Guzman, R. N. Solution Structure of Monomeric BsaL, the Type III Secretion Needle Protein of Burkholderia pseudomallei. *Journal of Molecular Biology* **359**, 322–330 (2006).
99. Wang, Y. *et al.* Differences in the Electrostatic Surfaces of the Type III Secretion Needle Proteins PrgI, BsaL, and MxiH. *Journal of Molecular Biology* **371**, 1304–1314 (2007).
100. Sun, P., Tropea, J. E., Austin, B. P., Cherry, S. & Waugh, D. S. Structural Characterization of the Yersinia pestis Type III Secretion System Needle Protein YscF in Complex with Its Heterodimeric Chaperone YscE/YscG. *Journal of Molecular Biology* **377**, 819–830 (2008).
101. Poyraz, Ö. *et al.* Protein refolding is required for assembly of the type three secretion needle. *Nat Struct Mol Biol* **17**, 788–792 (2010).
102. Lunelli, M., Hurwitz, R., Lambers, J. & Kolbe, M. Crystal Structure of PrgI-SipD: Insight into a Secretion Competent State of the Type Three Secretion System Needle Tip and its Interaction with Host Ligands. *PLoS Pathog* **7**, e1002163 (2011).
103. Fujii, T. *et al.* Structure of a type III secretion needle at 7-Å resolution provides insights into its assembly and signaling mechanisms. *Proceedings of the National Academy of Sciences* **109**, 4461–4466 (2012).
104. Loquet, A. *et al.* Atomic model of the type III secretion system needle. *Nature* **486**, 276–279 (2012).
105. Loquet, A. *et al.* Atomic Structure and Handedness of the Building Block of a Biological Assembly. *J. Am. Chem. Soc.* **135**, 19135–19138 (2013).
106. Demers, J.-P. *et al.* High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nat Commun* **5**, 4976 (2014).
107. Guo, E. Z. *et al.* A polymorphic helix of a Salmonella needle protein relays signals defining distinct steps in type III secretion. *PLoS Biol* **17**, e3000351 (2019).
108. Erskine, P. T. *et al.* High Resolution Structure of BipD: An Invasion Protein Associated with the Type III Secretion System of Burkholderia Pseudomallei. *Journal of Molecular Biology* **363**, 125–136 (2006).
109. Pal, M., Erskine, P. T., Gill, R. S., Wood, S. P. & Cooper, J. B. Near-atomic resolution analysis of BipD, a component of the type III secretion system of Burkholderia pseudomallei. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 990–993 (2010).
110. Johnson, S. *et al.* Self-chaperoning of the Type III Secretion System Needle Tip Proteins IpaD and BipD. *Journal of Biological Chemistry* **282**, 4035–4044 (2006).
111. Barta, M. L. *et al.* Identification of the bile salt binding site on IpaD from Shigella flexneri and the influence of ligand binding on IpaD structure. *Proteins* **80**, 935–945 (2012).

112. Cheung, M. *et al.* Three-dimensional electron microscopy reconstruction and cysteine-mediated crosslinking provide a model of the type III secretion system needle tip complex: Pseudoatomic model of the T3SS needle tip complex. *Molecular Microbiology* **95**, 31–50 (2015).
113. Chatterjee, S. *et al.* The crystal structures of the *Salmonella* type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate: Salmonella SipD-Bile Salts Crystal Structures. *Protein Science* **20**, 75–86 (2011).
114. Lokareddy, R. K., Lunelli, M., Eilers, B., Wolter, V. & Kolbe, M. Combination of Two Separate Binding Domains Defines Stoichiometry between Type III Secretion System Chaperone IpgC and Translocator Protein IpaB. *J. Biol. Chem.* **285**, 39965–39975 (2010).
115. Lunelli, M., Lokareddy, R. K., Zychlinsky, A. & Kolbe, M. IpaB-IpgC interaction defines binding motif for type III secretion translocator. *Proceedings of the National Academy of Sciences* **106**, 9661–9666 (2009).
116. Job, V., Mattei, P.-J., Lemaire, D., Attree, I. & Dessen, A. Structural Basis of Chaperone Recognition of Type III Secretion System Minor Translocator Proteins. *J. Biol. Chem.* **285**, 23224–23232 (2010).
117. Discola, K. F. *et al.* Membrane and Chaperone Recognition by the Major Translocator Protein PopB of the Type III Secretion System of *Pseudomonas aeruginosa*. *J. Biol. Chem.* **289**, 3591–3601 (2014).
118. Barta, M. L. *et al.* The Structures of Coiled-Coil Domains from Type III Secretion System Translocators Reveal Homology to Pore-Forming Toxins. *Journal of Molecular Biology* **417**, 395–405 (2012).
119. Francis, M. S., Aili, M., Wiklund, M.-L. & Wolf-Watz, H. A study of the YopD-LcrH interaction from *Yersinia pseudotuberculosis* reveals a role for hydrophobic residues within the amphipathic domain of YopD. *Mol Microbiol* **38**, 85–102 (2000).
120. Kuwae, A. *et al.* *Shigella* Invasion of Macrophage Requires the Insertion of IpaC into the Host Plasma Membrane: FUNCTIONAL ANALYSIS OF IpaC. *J. Biol. Chem.* **276**, 32230–32239 (2001).
121. Hume, P. J., McGhie, E. J., Hayward, R. D. & Koronakis, V. The purified *Shigella* IpaB and *Salmonella* SipB translocators share biochemical properties and membrane topology: *Shigella* IpaB. *Molecular Microbiology* **49**, 425–439 (2003).
122. Pallen, M. J., Dougan, G. & Frankel, G. Coiled-coil domains in proteins secreted by type III secretion systems. *Molecular Microbiology* **25**, 423–425 (1997).
123. Terry, C. M. *et al.* The C-terminus of IpaC is required for effector activities related to *Shigella* invasion of host cells. *Microbial Pathogenesis* **45**, 282–289 (2008).
124. Costa, T. R. D. *et al.* YopD Self-assembly and Binding to LcrV Facilitate Type III Secretion Activity by *Yersinia pseudotuberculosis*. *J. Biol. Chem.* **285**, 25269–25284 (2010).
125. Tengel, T., Sethson, I. & Francis, M. S. Conformational analysis by CD and NMR spectroscopy of a peptide encompassing the amphipathic domain of YopD from *Yersinia*:

- Tertiary structure of the YopD amphipathic domain. *European Journal of Biochemistry* **269**, 3659–3668 (2002).
126. Veenendaal, A. K. J. *et al.* The type III secretion system needle tip complex mediates host cell sensing and translocon insertion: Shigella T3SS needle tip complex. *Molecular Microbiology* **63**, 1719–1730 (2007).
  127. Romano, F. B. *et al.* Type 3 Secretion Translocators Spontaneously Assemble a Hexadecameric Transmembrane Complex. *J. Biol. Chem.* **291**, 6304–6315 (2016).
  128. Montagner, C., Arquint, C. & Cornelis, G. R. Translocators YopB and YopD from *Yersinia enterocolitica* Form a Multimeric Integral Membrane Complex in Eukaryotic Cell Membranes. *Journal of Bacteriology* **193**, 6923–6928 (2011).
  129. Ide, T. *et al.* Characterization of translocation pores inserted into plasma membranes by type III-secreted Esp proteins of enteropathogenic *Escherichia coli*. *Cell Microbiol* **3**, 669–679 (2001).
  130. Park, D. *et al.* Visualization of the type III secretion mediated Salmonella–host cell interface using cryo-electron tomography. *eLife* **7**, e39514 (2018).
  131. Silva, Y. R. de O., Contreras-Martel, C., Macheboeuf, P. & Dessen, A. Bacterial secretins: Mechanisms of assembly and membrane targeting. *Protein Science* **29**, 893–904 (2020).
  132. Koo, J., Burrows, L. L. & Lynne Howell, P. Decoding the roles of pilotins and accessory proteins in secretin escort services. *FEMS Microbiology Letters* **328**, 1–12 (2012).
  133. Collin, S., Guilvout, I., Nickerson, N. N. & Pugsley, A. P. Sorting of an integral outer membrane protein via the lipoprotein-specific Lol pathway and a dedicated lipoprotein pilotin: Lol pathway sorting of an integral outer membrane protein. *Molecular Microbiology* **80**, 655–665 (2011).
  134. Kim, K. *et al.* Crystal structure of PilF: Functional implication in the type 4 pilus biogenesis in *Pseudomonas aeruginosa*. *Biochemical and Biophysical Research Communications* **340**, 1028–1038 (2006).
  135. Koo, J. *et al.* PilF Is an Outer Membrane Lipoprotein Required for Multimerization and Localization of the *Pseudomonas aeruginosa* Type IV Pilus Secretin. *Journal of Bacteriology* **190**, 6961–6969 (2008).
  136. Trindade, M. B., Job, V., Contreras-Martel, C., Pelicic, V. & Dessen, A. Structure of a Widely Conserved Type IV Pilus Biogenesis Factor that Affects the Stability of Secretin Multimers. *Journal of Molecular Biology* **378**, 1031–1039 (2008).
  137. Viarre, V. *et al.* HxcQ Liposecretin Is Self-piloted to the Outer Membrane by Its N-terminal Lipid Anchor. *Journal of Biological Chemistry* **284**, 33815–33823 (2009).
  138. Deng, W. *et al.* Regulation of Type III Secretion Hierarchy of Translocators and Effectors in Attaching and Effacing Bacterial Pathogens. *IAI* **73**, 2135–2146 (2005).
  139. Nouwen, N. *et al.* Secretin PulD: Association with pilot PulS, structure, and ion-conducting channel formation. *Proceedings of the National Academy of Sciences* **96**, 8173–8177 (1999).

140. Crago, A. M. & Koronakis, V. Salmonella InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization. *Mol. Microbiol.* **30**, 47–56 (1998).
141. Daefler, S. & Russel, M. The Salmonella typhimurium InvH protein is an outer membrane lipoprotein required for the proper localization of InvG. *Mol. Microbiol.* **28**, 1367–1380 (1998).
142. Guilvout, I., Chami, M., Engel, A., Pugsley, A. P. & Bayan, N. Bacterial outer membrane secretin PulD assembles and inserts into the inner membrane in the absence of its pilotin. *The EMBO Journal* **25**, 5241–5249 (2006).
143. Burghout, P. *et al.* Role of the Pilot Protein YscW in the Biogenesis of the YscC Secretin in *Yersinia enterocolitica*. *Journal of Bacteriology* **186**, 5366–5375 (2004).
144. Carbonnelle, E., Hélaïne, S., Prouvensier, L., Nassif, X. & Pelicic, V. Type IV pilus biogenesis in *Neisseria meningitidis*: PilW is involved in a step occurring after pilus assembly, essential for fibre stability and function: Type IV pilus biogenesis in pathogenic *Neisseria*. *Molecular Microbiology* **55**, 54–64 (2004).
145. Hirano, T., Yamaguchi, S., Oosawa, K. & Aizawa, S. Roles of FliK and FlhB in determination of flagellar hook length in *Salmonella typhimurium*. *Journal of Bacteriology* **176**, 5439–5449 (1994).
146. Zarivach, R. *et al.* Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS. *Nature* **453**, 124–127 (2008).
147. Minamino, T. & Macnab, R. M. Domain Structure of Salmonella FlhB, a Flagellar Export Component Responsible for Substrate Specificity Switching. *J. Bacteriol.* **182**, 4906–4914 (2000).
148. Fraser, G. M. *et al.* Substrate specificity of type III flagellar protein export in *Salmonella* is controlled by subdomain interactions in FlhB: Role of the Salmonella FlhB protein in flagellar protein export. *Molecular Microbiology* **48**, 1043–1057 (2003).
149. Lavander, M. *et al.* Proteolytic Cleavage of the FlhB Homologue YscU of *Yersinia pseudotuberculosis* Is Essential for Bacterial Survival but Not for Type III Secretion. *JB* **184**, 4500–4509 (2002).
150. Ferris, H. U. *et al.* FlhB Regulates Ordered Export of Flagellar Components via Autocleavage Mechanism. *J. Biol. Chem.* **280**, 41236–41242 (2005).
151. Deane, J. E. *et al.* Crystal structure of Spa40, the specificity switch for the *Shigella flexneri* type III secretion system. *Molecular Microbiology* **69**, 267–276 (2008).
152. Magdalena, J. *et al.* Spa32 Regulates a Switch in Substrate Specificity of the Type III Secretion of *Shigella flexneri* from Needle Components to Ipa Proteins. *JB* **184**, 3433–3441 (2002).
153. Edqvist, P. J. *et al.* YscP and YscU Regulate Substrate Specificity of the *Yersinia* Type III Secretion System. *JB* **185**, 2259–2266 (2003).
154. Tamano, K., Katayama, E., Toyotome, T. & Sasakawa, C. *Shigella* Spa32 Is an Essential Secretory Protein for Functional Type III Secretion Machinery and Uniformity of Its Needle Length. *JB* **184**, 1244–1252 (2002).

155. Monjaras Feria, J. *et al.* Role of EscP (Orf16) in Injectisome Biogenesis and Regulation of Type III Protein Secretion in Enteropathogenic *Escherichia coli*. *Journal of Bacteriology* **194**, 6029–6045 (2012).
156. Journet, L. The Needle Length of Bacterial Injectisomes Is Determined by a Molecular Ruler. *Science* **302**, 1757–1760 (2003).
157. Wagner, S., Stenta, M., Metzger, L. C., Dal Peraro, M. & Cornelis, G. R. Length control of the injectisome needle requires only one molecule of Yop secretion protein P (YscP). *Proceedings of the National Academy of Sciences* **107**, 13860–13865 (2010).
158. Lefebvre, M. D. & Galan, J. E. The inner rod protein controls substrate switching and needle length in a *Salmonella* type III secretion system. *Proceedings of the National Academy of Sciences* **111**, 817–822 (2014).
159. Nariya, M. K., Israeli, J., Shi, J. J. & Deeds, E. J. Mathematical Model for Length Control by the Timing of Substrate Switching in the Type III Secretion System. *PLoS Comput Biol* **12**, e1004851 (2016).
160. Wee, D. H. & Hughes, K. T. Molecular ruler determines needle length for the *Salmonella* Spi-1 injectisome. *Proc Natl Acad Sci USA* **112**, 4098–4103 (2015).
161. Deng, W. *et al.* Dissecting virulence: Systematic and functional analyses of a pathogenicity island. *Proceedings of the National Academy of Sciences* **101**, 3597–3602 (2004).
162. Portaliou, A. G. *et al.* Hierarchical protein targeting and secretion is controlled by an affinity switch in the type III secretion system of enteropathogenic *Escherichia coli*. *The EMBO Journal* **36**, 3517–3531 (2017).
163. Cherradi, Y. *et al.* Interplay between predicted inner-rod and gatekeeper in controlling substrate specificity of the type III secretion system: MxiC-MxiI interaction controls effectors secretion by *Shigella*. *Molecular Microbiology* **87**, 1183–1199 (2013).
164. Martinez-Argudo, I. & Blocker, A. J. The *Shigella* T3SS needle transmits a signal for MxiC release, which controls secretion of effectors: Regulatory cascade leading to T3SS activation. *Molecular Microbiology* **78**, 1365–1378 (2010).
165. Yu, X. J., McGourty, K., Liu, M., Unsworth, K. E. & Holden, D. W. pH Sensing by Intracellular *Salmonella* Induces Effector Translocation. *Science* **328**, 1040–1043 (2010).
166. Armentrout, E. I. & Rietsch, A. The Type III Secretion Translocation Pore Senses Host Cell Contact. *PLoS Pathog* **12**, e1005530 (2016).
167. Woestyn, S., Allaoui, A., Wattiau, P. & Cornelis, G. R. YscN, the putative energizer of the *Yersinia* Yop secretion machinery. *J. Bacteriol.* **176**, 1561–1569 (1994).
168. Wilharm, G. *et al.* *Yersinia enterocolitica* Type III Secretion Depends on the Proton Motive Force but Not on the Flagellar Motor Components MotA and MotB. *Infection and Immunity* **72**, 4004–4009 (2004).
169. Macnab, R. M. Type III flagellar protein export and flagellar assembly. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1694**, 207–217 (2004).
170. Minamino, T. & Namba, K. Self-Assembly and Type III Protein Export of the Bacterial Flagellum. *Journal of Molecular Microbiology and Biotechnology* **7**, 5–17 (2004).

171. Paul, K., Erhardt, M., Hirano, T., Blair, D. F. & Hughes, K. T. Energy source of flagellar type III secretion. *Nature* **451**, 489–492 (2008).
172. Lorenz, C. & Buttner, D. Functional Characterization of the Type III Secretion ATPase HrcN from the Plant Pathogen *Xanthomonas campestris* pv. *vesicatoria*. *Journal of Bacteriology* **191**, 1414–1428 (2009).
173. Thomas, J., Stafford, G. P. & Hughes, C. Docking of cytosolic chaperone-substrate complexes at the membrane ATPase during flagellar type III protein export. *Proceedings of the National Academy of Sciences* **101**, 3945–3950 (2004).
174. Minamino, T. & Namba, K. Distinct roles of the FliI ATPase and proton motive force in bacterial flagellar protein export. *Nature* **451**, 485–488 (2008).
175. Minamino, T., Morimoto, Y. V., Hara, N. & Namba, K. An energy transduction mechanism used in bacterial flagellar type III protein export. *Nature Communications* **2**, (2011).
176. Erhardt, M., Mertens, M. E., Fabiani, F. D. & Hughes, K. T. ATPase-Independent Type-III Protein Secretion in *Salmonella enterica*. *PLoS Genetics* **10**, e1004800 (2014).
177. Minamino, T., Morimoto, Y. V., Hara, N., Aldridge, P. D. & Namba, K. The Bacterial Flagellar Type III Export Gate Complex Is a Dual Fuel Engine That Can Use Both H<sup>+</sup> and Na<sup>+</sup> for Flagellar Protein Export. *PLOS Pathogens* **12**, e1005495 (2016).
178. Minamino, T. *et al.* FliH and FliI ensure efficient energy coupling of flagellar type III protein export in *Salmonella*. *MicrobiologyOpen* **5**, 424–435 (2016).
179. Terashima, H. *et al.* *In Vitro* Reconstitution of Functional Type III Protein Export and Insights into Flagellar Assembly. *mBio* **9**, (2018).
180. Erhardt, M. *et al.* Mechanism of type-III protein secretion: Regulation of FlhA conformation by a functionally critical charged-residue cluster: Critical charged-residue site in FlhA. *Molecular Microbiology* **104**, 234–249 (2017).
181. Wilharm, G., Dittmann, S., Schmid, A. & Heesemann, J. On the role of specific chaperones, the specific ATPase, and the proton motive force in type III secretion. *International Journal of Medical Microbiology* **297**, 27–36 (2007).
182. Maura, D., Ballok, A. E. & Rahme, L. G. Considerations and caveats in anti-virulence drug development. *Current Opinion in Microbiology* **33**, 41–46 (2016).
183. Deng, W. *et al.* Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology* **15**, 323–337 (2017).
184. van der Heijden, J. & Finlay, B. B. Type III effector-mediated processes in *Salmonella* infection. *Future Microbiology* **7**, 685–703 (2012).
185. Wang, X. *et al.* Antibiotic Resistance in *Salmonella* Typhimurium Isolates Recovered From the Food Chain Through National Antimicrobial Resistance Monitoring System Between 1996 and 2016. *Front. Microbiol.* **10**, 985 (2019).
186. Lara-Tejero, M. The Type III Secretion System Sorting Platform. in (Springer Berlin Heidelberg, 2019). doi:10.1007/82\_2019\_167.

187. Minamino, T., Morimoto, Y. V., Hara, N., Aldridge, P. D. & Namba, K. The Bacterial Flagellar Type III Export Gate Complex Is a Dual Fuel Engine That Can Use Both H<sup>+</sup> and Na<sup>+</sup> for Flagellar Protein Export. *PLOS Pathogens* **12**, e1005495 (2016).
188. Minamino, T., Morimoto, Y. V., Hara, N. & Namba, K. An energy transduction mechanism used in bacterial flagellar type III protein export. *Nature Communications* **2**, (2011).
189. Worrall, L. J. *et al.* Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body. *Nature* **540**, 597–601 (2016).
190. Dey, S., Chakravarty, A., Guha Biswas, P. & De Guzman, R. N. The type III secretion system needle, tip, and translocon. *Protein Science* (2019) doi:10.1002/pro.3682.
191. Howard, S. P. *et al.* Pilotin from *Vibrio vulnificus* type 2 secretion system, EpsS. (:unav) (2019) doi:10.2210/pdb6i2v/pdb.
192. Yan, Z., Yin, M., Xu, D., Zhu, Y. & Li, X. Structural insights into the secretin translocation channel in the type II secretion system. *Nature Structural & Molecular Biology* **24**, 177–183 (2017).
193. Hay, I. D., Belousoff, M. J. & Lithgow, T. Structural Basis of Type 2 Secretion System Engagement between the Inner and Outer Bacterial Membranes. *mBio* **8**, e01344-17 (2017).
194. Yin, M., Yan, Z. & Li, X. Structural insight into the assembly of the type II secretion system pilotin–secretin complex from enterotoxigenic *Escherichia coli*. *Nature Microbiology* **3**, 581–587 (2018).
195. Majewski, D. D., Worrall, L. J. & Strynadka, N. C. Secretins revealed: structural insights into the giant gated outer membrane portals of bacteria. *Current Opinion in Structural Biology* **51**, 61–72 (2018).
196. Dunstan, R. A. *et al.* Assembly of the secretion pores GspD, Wza and CsgG into bacterial outer membranes does not require the Omp85 proteins BamA or TamA: Assembly of the secretion pores. *Molecular Microbiology* **97**, 616–629 (2015).
197. Collin, S., Guilvout, I., Nickerson, N. N. & Pugsley, A. P. Sorting of an integral outer membrane protein via the lipoprotein-specific Lol pathway and a dedicated lipoprotein pilotin: Lol pathway sorting of an integral outer membrane protein. *Molecular Microbiology* **80**, 655–665 (2011).
198. Koo, J., Burrows, L. L. & Lynne Howell, P. Decoding the roles of pilotins and accessory proteins in secretin escort services. *FEMS Microbiology Letters* **328**, 1–12 (2012).
199. Daefler, S., Guilvout, I., Hardie, K. R., Pugsley, A. P. & Russel, M. The C-terminal domain of the secretin PulD contains the binding site for its cognate chaperone, PulS, and confers PulS dependence on pIVf1 function. *Mol. Microbiol.* **24**, 465–475 (1997).
200. Crago, A. M. & Koronakis, V. Salmonella InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization. *Mol. Microbiol.* **30**, 47–56 (1998).
201. Daefler, S. & Russel, M. The Salmonella typhimurium InvH protein is an outer membrane lipoprotein required for the proper localization of InvG. *Mol. Microbiol.* **28**, 1367–1380 (1998).

202. Dunstan, R. A. *et al.* Assembly of the Type II Secretion System such as Found in *Vibrio cholerae* Depends on the Novel Pilotin AspS. *PLoS Pathogens* **9**, e1003117 (2013).
203. Izoré, T. *et al.* Structural Characterization and Membrane Localization of ExsB from the Type III Secretion System (T3SS) of *Pseudomonas aeruginosa*. *Journal of Molecular Biology* **413**, 236–246 (2011).
204. Lario, P. I. *et al.* Structure and biochemical analysis of a secretin pilot protein. *The EMBO Journal* **24**, 1111–1121 (2005).
205. Okon, M. *et al.* Structural Characterization of the Type-III Pilot-Secretin Complex from *Shigella flexneri*. *Structure* **16**, 1544–1554 (2008).
206. Gu, S., Rehman, S., Wang, X., Shevchik, V. E. & Pickersgill, R. W. Structural and Functional Insights into the Pilotin-Secretin Complex of the Type II Secretion System. *PLoS Pathogens* **8**, e1002531 (2012).
207. Rehman, S., Gu, S., Shevchik, V. E. & Pickersgill, R. W. Anatomy of secretin binding to the *Dickeya dadantii* type II secretion system pilotin. *Acta Crystallographica Section D Biological Crystallography* **69**, 1381–1386 (2013).
208. Tosi, T. *et al.* Pilotin-secretin recognition in the type II secretion system of *Klebsiella oxytoca*: Pilotin-secretin interactions in T2SS. *Molecular Microbiology* **82**, 1422–1432 (2011).
209. Korotkov, K. V. & Hol, W. G. J. Crystal structure of the pilotin from the enterohemorrhagic *Escherichia coli* type II secretion system. *Journal of Structural Biology* **182**, 186–191 (2013).
210. Kim, K. *et al.* Crystal structure of PilF: Functional implication in the type 4 pilus biogenesis in *Pseudomonas aeruginosa*. *Biochemical and Biophysical Research Communications* **340**, 1028–1038 (2006).
211. Koo, J. *et al.* PilF Is an Outer Membrane Lipoprotein Required for Multimerization and Localization of the *Pseudomonas aeruginosa* Type IV Pilus Secretin. *Journal of Bacteriology* **190**, 6961–6969 (2008).
212. Trindade, M. B., Job, V., Contreras-Martel, C., Pelicic, V. & Dessen, A. Structure of a Widely Conserved Type IV Pilus Biogenesis Factor that Affects the Stability of Secretin Multimers. *Journal of Molecular Biology* **378**, 1031–1039 (2008).
213. Pati, N. B. *et al.* Deletion of *invH* gene in *Salmonella enterica* serovar Typhimurium limits the secretion of Sip effector proteins. *Microbes and Infection* **15**, 66–73 (2013).
214. Kabsch, W. *XDS*. *Acta Crystallographica Section D Biological Crystallography* **66**, 125–132 (2010).
215. Winn, M. D. *et al.* Overview of the CCP 4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* **67**, 235–242 (2011).
216. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallographica Section D Structural Biology* **75**, 861–877 (2019).

217. Vagin, A. A. *et al.* REFMAC 5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D Biological Crystallography* **60**, 2184–2195 (2004).
218. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallographica Section D Biological Crystallography* **66**, 486–501 (2010).
219. *The PyMOL molecular graphics system, version 1.7.2.1.* (2014).
220. Pettersen, E. F. *et al.* UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605–1612 (2004).
221. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis: UCSF ChimeraX Visualization System. *Protein Science* **27**, 14–25 (2018).
222. Beilsten-Edmands, J. *et al.* Scaling diffraction data in the *DIALS* software package: algorithms and new approaches for multi-crystal scaling. *Acta Crystallographica Section D Structural Biology* **76**, 385–399 (2020).
223. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658–674 (2007).
224. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, (1995).
225. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
226. Lee, W., Westler, W. M., Bahrami, A., Eghbalnia, H. R. & Markley, J. L. PINE-SPARKY: graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy. *Bioinformatics* **25**, 2085–2087 (2009).
227. Lee, W. *et al.* PACSY, a relational database management system for protein structure and chemical shift analysis. *Journal of Biomolecular NMR* **54**, 169–179 (2012).
228. Shen, Y. & Bax, A. Protein Structural Information Derived from NMR Chemical Shift with the Neural Network Program TALOS-N. in *Artificial Neural Networks* (ed. Cartwright, H.) vol. 1260 17–32 (Springer New York, 2015).
229. Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *Journal of Biomolecular NMR* **62**, 453–471 (2015).
230. Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on <sup>13</sup>C chemical shift statistics. *J. Biomol. NMR* **24**, 149–154 (2002).
231. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).
232. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* **372**, 774–797 (2007).
233. Laguerre, M., Saux, M., Dubost, J. P. & Carpy, A. MLPP: A Program for the Calculation of Molecular Lipophilicity Potential in Proteins. *Pharmacy and Pharmacology Communications* **3**, 217–222 (1997).

234. *CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data.* (CRC Press, 2001).
235. Yee, A. A. *et al.* NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *J. Am. Chem. Soc.* **127**, 16512–16517 (2005).
236. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1814**, 942–968 (2011).
237. Cheng, J., Randall, A. Z., Sweredoski, M. J. & Baldi, P. SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Research* **33**, W72–W76 (2005).
238. Jones, D. T. & Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857–863 (2015).
239. Konrat, R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance* **241**, 74–85 (2014).
240. Berjanskii, M. V. & Wishart, D. S. A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts. *J. Am. Chem. Soc.* **127**, 14970–14971 (2005).
241. Nickerson, N. N. *et al.* Outer Membrane Targeting of Secretin PulD Protein Relies on Disordered Domain Recognition by a Dedicated Chaperone. *Journal of Biological Chemistry* **286**, 38833–38843 (2011).
242. Gu, S., Rehman, S., Wang, X., Shevchik, V. E. & Pickersgill, R. W. Structural and Functional Insights into the Pilotin-Secretin Complex of the Type II Secretion System. *PLoS Pathogens* **8**, e1002531 (2012).
243. Jaudzems, K., Serrano, P., Michael, G., Reto, H. & Wuthrich, K. NMR structure of the protein YP\_557733.1 from *Burkholderia xenovorans*. (*:unav*) (2011)  
doi:10.2210/pdb2la7/pdb.
244. Guilvout, I., Nickerson, N. N., Chami, M. & Pugsley, A. P. Multimerization-defective variants of dodecameric secretin PulD. *Research in Microbiology* **162**, 180–190 (2011).
245. Guilvout, I. *et al.* In Vitro Multimerization and Membrane Insertion of Bacterial Outer Membrane Secretin PulD. *Journal of Molecular Biology* **382**, 13–23 (2008).
246. Ainaravaru, S. R. K. *et al.* Contour Length and Refolding Rate of a Small Protein Controlled by Engineered Disulfide Bonds. *Biophysical Journal* **92**, 225–233 (2007).
247. Hu, B., Lara-Tejero, M., Kong, Q., Galán, J. E. & Liu, J. In Situ Molecular Architecture of the Salmonella Type III Secretion Machine. *Cell* **168**, 1065-1074.e10 (2017).
248. Hu, B. *et al.* Visualization of the type III secretion sorting platform of *Shigella flexneri*. *Proceedings of the National Academy of Sciences* **112**, 1047–1052 (2015).
249. Chernyatina, A. A. & Low, H. H. Core architecture of a bacterial type II secretion system. *Nature Communications* **10**, (2019).
250. Hay, I. D., Belousoff, M. J., Dunstan, R. A., Bamert, R. S. & Lithgow, T. Structure and Membrane Topography of the Vibrio-Type Secretin Complex from the Type 2 Secretion System of Enteropathogenic *Escherichia coli*. *J Bacteriol* **200**, e00521-17, /jb/200/5/e00521-17.atom (2017).

251. Howard, S. P. *et al.* Structure and assembly of pilotin-dependent and -independent secretins of the type II secretion system. *PLoS Pathog* **15**, e1007731 (2019).
252. Mills, D. M., Bajaj, V. & Lee, C. A. A 40 kb chromosomal fragment encoding Salmonella typhimurium invasion genes is absent from the corresponding region of the Escherichia coli K-12 chromosome. *Molecular Microbiology* **15**, 749–759 (1995).
253. Schuch, R., Sandlin, R. C. & Maurelli, A. T. A system for identifying post-invasion functions of invasion genes: requirements for the Mxi-Spa type III secretion pathway of Shigella flexneri in intercellular dissemination. *Mol Microbiol* **34**, 675–689 (1999).
254. Williams McMackin, E. A., Djapgne, L., Corley, J. M. & Yahr, T. L. Fitting Pieces into the Puzzle of *Pseudomonas aeruginosa* Type III Secretion System Gene Expression. *J Bacteriol* **201**, e00209-19, /jb/201/13/JB.00209-19.atom (2019).
255. Kuwae, A., Momose, F., Nagamatsu, K., Suyama, Y. & Abe, A. BteA Secreted from the Bordetella bronchiseptica Type III Secretion System Induces Necrosis through an Actin Cytoskeleton Signaling Pathway and Inhibits Phagocytosis by Macrophages. *PLoS ONE* **11**, e0148387 (2016).
256. Makino, F. *et al.* The Architecture of the Cytoplasmic Region of Type III Secretion Systems. *Sci Rep* **6**, 33341 (2016).
257. Hu, J. *et al.* Cryo-EM analysis of the T3S injectisome reveals the structure of the needle and open secretin. *Nat Commun* **9**, 3840 (2018).
258. Xing, Q. *et al.* Structures of chaperone-substrate complexes docked onto the export gate in a type III secretion system. *Nat Commun* **9**, 1773 (2018).
259. Bange, G. *et al.* FlhA provides the adaptor for coordinated delivery of late flagella building blocks to the type III secretion system. *Proceedings of the National Academy of Sciences* **107**, 11295–11300 (2010).
260. Moore, S. A. & Jia, Y. Structure of the Cytoplasmic Domain of the Flagellar Secretion Apparatus Component FlhA from *Helicobacter pylori*. *J. Biol. Chem.* **285**, 21060–21069 (2010).
261. Inoue, Y. *et al.* Structural Insights into the Substrate Specificity Switch Mechanism of the Type III Protein Export Apparatus. *Structure* **27**, 965-976.e6 (2019).
262. Saijo-Hamano, Y. *et al.* Structure of the cytoplasmic domain of FlhA and implication for flagellar type III protein export. *Molecular Microbiology* **76**, 260–268 (2010).
263. Gauthier, A., Puente, J. L. & Finlay, B. B. Secretin of the Enteropathogenic Escherichia coli Type III Secretion System Requires Components of the Type III Apparatus for Assembly and Localization. *IAI* **71**, 3310–3319 (2003).
264. Nadler, C., Shifrin, Y., Nov, S., Kobi, S. & Rosenshine, I. Characterization of Enteropathogenic Escherichia coli Mutants That Fail To Disrupt Host Cell Spreading and Attachment to Substratum. *IAI* **74**, 839–849 (2006).
265. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

266. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).
267. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**, 216–221 (2015).
268. Thompson, J. & Baker, D. Incorporation of evolutionary information into Rosetta comparative modeling. *Proteins* **79**, 2380–2388 (2011).
269. Wang, R. Y.-R. *et al.* Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).
270. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**, 213–221 (2010).
271. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite: Improvements to the APBS Software Suite. *Protein Science* **27**, 112–128 (2018).
272. Mulkidjanian, A. Y., Makarova, K. S., Galperin, M. Y. & Koonin, E. V. Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nature Reviews Microbiology* **5**, 892–899 (2007).
273. Vogler, A. P., Homma, M., Irikura, V. M. & Macnab, R. M. Salmonella typhimurium mutants defective in flagellar filament regrowth and sequence similarity of FliI to F0F1, vacuolar, and archaeobacterial ATPase subunits. *J. Bacteriol.* **173**, 3564–3572 (1991).
274. Eichelberg, K., Ginocchio, C. C. & Galán, J. E. Molecular and functional characterization of the Salmonella typhimurium invasion genes *invB* and *invC*: homology of *InvC* to the F0F1 ATPase family of proteins. *J. Bacteriol.* **176**, 4501–4510 (1994).
275. Imada, K., Minamino, T., Tahara, A. & Namba, K. Structural similarity between the flagellar type III ATPase FliI and F1-ATPase subunits. *Proceedings of the National Academy of Sciences* **104**, 485–490 (2007).
276. Zarivach, R., Vuckovic, M., Deng, W., Finlay, B. B. & Strynadka, N. C. J. Structural analysis of a prototypical ATPase from the type III secretion system. *Nature Structural & Molecular Biology* **14**, 131–137 (2007).
277. Pallen, M. J. Evolutionary links between FliH/YscL-like proteins from bacterial type III secretion systems and second-stalk components of the FoF1 and vacuolar ATPases. *Protein Science* **15**, 935–941 (2006).
278. Imada, K., Minamino, T., Uchida, Y., Kinoshita, M. & Namba, K. Insight into the flagella type III export revealed by the complex structure of the type III ATPase and its regulator. *Proceedings of the National Academy of Sciences* **113**, 3633–3638 (2016).
279. Ibuki, T. *et al.* Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. *Nature Structural & Molecular Biology* **18**, 277–282 (2011).
280. Romo-Castillo, M. *et al.* EscO, a Functional and Structural Analog of the Flagellar FliJ Protein, Is a Positive Regulator of EscN ATPase Activity of the Enteropathogenic Escherichia coli Injectisome. *Journal of Bacteriology* **196**, 2227–2241 (2014).

281. Lorenzini, E. *et al.* Structure and Protein-Protein Interaction Studies on Chlamydia trachomatis Protein CT670 (YscO Homolog). *Journal of Bacteriology* **192**, 2746–2756 (2010).
282. Minasov, G. *et al.* 2.8 Angstrom Crystal Structure of Type III Secretion Protein YscO from *Vibrio parahaemolyticus*. (2013) doi:10.2210/pdb4mh6/pdb.
283. Kishikawa, J. *et al.* Common Evolutionary Origin for the Rotor Domain of Rotary ATPases and Flagellar Protein Export Apparatus. *PLoS ONE* **8**, e64695 (2013).
284. Deng, W. *et al.* Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology* **15**, 323–337 (2017).
285. Woestyn, S., Allaoui, A., Wattiau, P. & Cornelis, G. R. YscN, the putative energizer of the *Yersinia* Yop secretion machinery. *J. Bacteriol.* **176**, 1561–1569 (1994).
286. Minamino, T. & Namba, K. Self-Assembly and Type III Protein Export of the Bacterial Flagellum. *Journal of Molecular Microbiology and Biotechnology* **7**, 5–17 (2004).
287. Wilharm, G. *et al.* *Yersinia enterocolitica* Type III Secretion Depends on the Proton Motive Force but Not on the Flagellar Motor Components MotA and MotB. *Infection and Immunity* **72**, 4004–4009 (2004).
288. Minamino, T. Protein export through the bacterial flagellar type III export pathway. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 1642–1648 (2014).
289. Minamino, T., Morimoto, Y. V., Hara, N. & Namba, K. An energy transduction mechanism used in bacterial flagellar type III protein export. *Nature Communications* **2**, (2011).
290. Erhardt, M., Mertens, M. E., Fabiani, F. D. & Hughes, K. T. ATPase-Independent Type-III Protein Secretion in *Salmonella enterica*. *PLoS Genetics* **10**, e1004800 (2014).
291. Akeda, Y. & Galán, J. E. Chaperone release and unfolding of substrates in type III secretion. *Nature* **437**, 911–915 (2005).
292. Lorenz, C. & Buttner, D. Functional Characterization of the Type III Secretion ATPase HrcN from the Plant Pathogen *Xanthomonas campestris* pv. *vesicatoria*. *Journal of Bacteriology* **191**, 1414–1428 (2009).
293. Yoshida, Y. *et al.* Functional Characterization of the Type III Secretion ATPase SsaN Encoded by *Salmonella* Pathogenicity Island 2. *PLoS ONE* **9**, e94347 (2014).
294. Burkinshaw, B. J. & Strynadka, N. C. J. Assembly and structure of the T3SS. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 1649–1663 (2014).
295. Stebbins, C. E. & Galán, J. E. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **414**, 77–81 (2001).
296. Allison, S. E. *et al.* Identification of the Docking Site between a Type III Secretion System ATPase and a Chaperone for Effector Cargo. *Journal of Biological Chemistry* **289**, 23734–23744 (2014).
297. Burgess, J. L. *et al.* Spa47 is an oligomerization-activated type three secretion system (T3SS) ATPase from *Shigella flexneri*: Characterization of the *Shigella* T3SS ATPase Spa47. *Protein Science* **25**, 1037–1048 (2016).

298. Gao, X. *et al.* Structural Insight Into Conformational Changes Induced by ATP Binding in a Type III Secretion-Associated ATPase From *Shigella flexneri*. *Frontiers in Microbiology* **9**, (2018).
299. Iguchi, A. *et al.* Complete Genome Sequence and Comparative Genome Analysis of Enteropathogenic *Escherichia coli* O127:H6 Strain E2348/69. *Journal of Bacteriology* **191**, 347–354 (2009).
300. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology* **152**, 36–51 (2005).
301. Zivanov, J. *et al.* RELION-3: new tools for automated high-resolution cryo-EM structure determination. (2018) doi:10.1101/421123.
302. Zheng, S., Palovcak, E., Armache, J.-P., Cheng, Y. & Agard, D. Anisotropic Correction of Beam-induced Motion for Improved Single-particle Electron Cryo-microscopy. (2016) doi:10.1101/061960.
303. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**, 216–221 (2015).
304. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, 290–296 (2017).
305. Zivanov, J., Nakane, T. & Scheres, S. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. (2018) doi:10.1101/384537.
306. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
307. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).
308. Pettersen, E. F. *et al.* UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605–1612 (2004).
309. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**, 213–221 (2010).
310. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallographica Section D Biological Crystallography* **66**, 486–501 (2010).
311. Shrodinger, LLC. *The PyMOL molecular graphics system, version 1.7.2.1*. (2014).
312. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite: Improvements to the APBS Software Suite. *Protein Science* **27**, 112–128 (2018).
313. Imamura, H., Funamoto, S., Yoshida, M. & Yokoyama, K. Reconstitution *in Vitro* of V<sub>1</sub> Complex of *Thermus thermophilus* V-ATPase Revealed That ATP Binding to the A Subunit Is Crucial for V<sub>1</sub> Formation. *Journal of Biological Chemistry* **281**, 38582–38591 (2006).

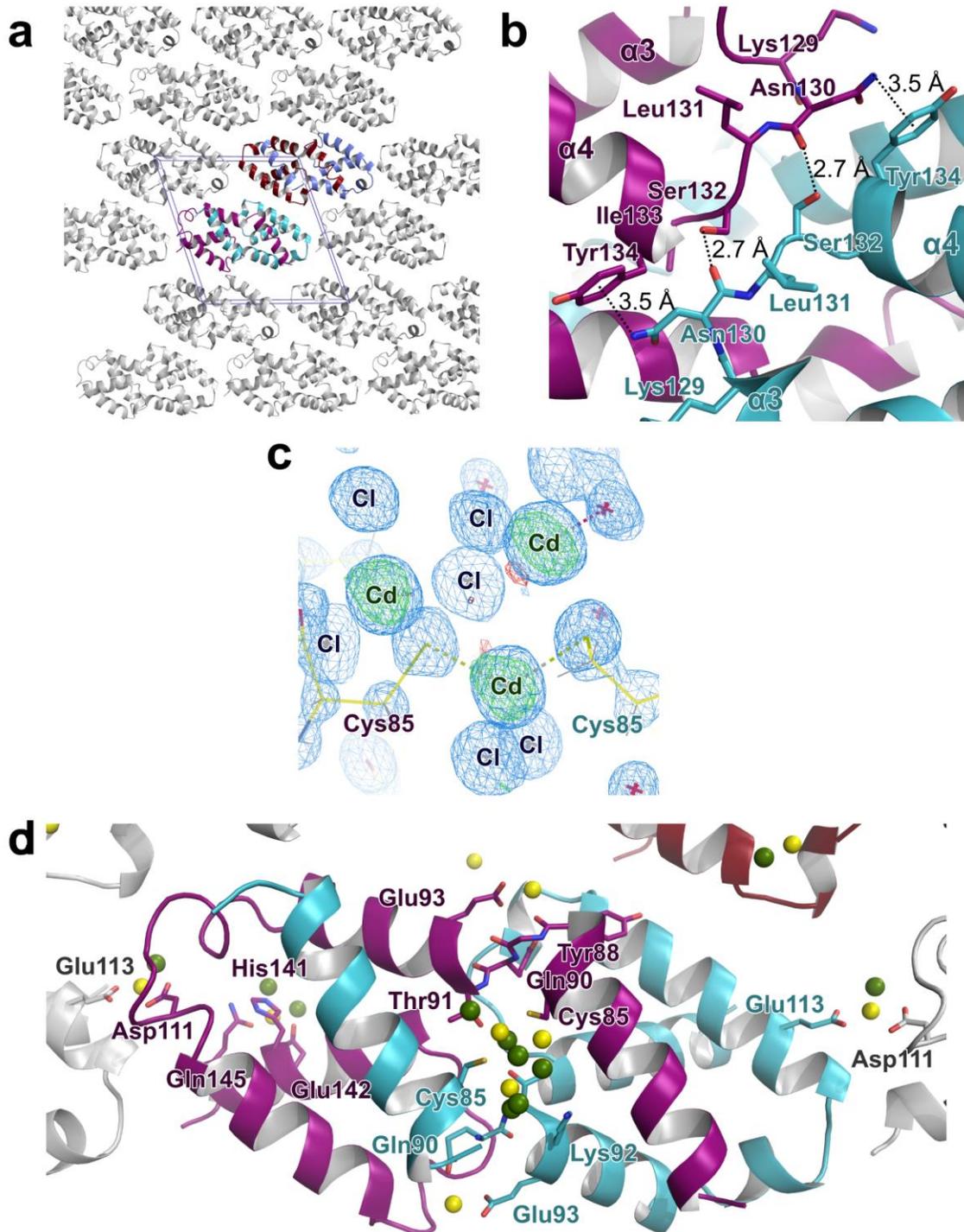
314. Barad, B. A. *et al.* EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nature Methods* **12**, 943–946 (2015).
315. Braig, K., Menz, R. I., Montgomery, M. G., Leslie, A. G. & Walker, J. E. Structure of bovine mitochondrial F1-ATPase inhibited by Mg<sup>2+</sup>+ADP and aluminium fluoride. *Structure* **8**, 567–573 (2000).
316. Akeda, Y. & Galan, J. E. Genetic Analysis of the Salmonella enterica Type III Secretion-Associated ATPase InvC Defines Discrete Functional Domains. *Journal of Bacteriology* **186**, 2402–2412 (2004).
317. Arai, S. *et al.* Rotation mechanism of Enterococcus hirae V1-ATPase based on asymmetric crystal structures. *Nature* **493**, 703–707 (2013).
318. Kohori, A. *et al.* Torque Generation in F1-ATPase Devoid of the Entire Amino-Terminal Helix of the Rotor That Fills Half of the Stator Orifice. *Biophysical Journal* **101**, 188–195 (2011).
319. Hossain, M. D. *et al.* Neither Helix in the Coiled Coil Region of the Axle of F1-ATPase Plays a Significant Role in Torque Production. *Biophysical Journal* **95**, 4837–4844 (2008).
320. Zhang, Y., Lara-Tejero, M., Bewersdorf, J. & Galán, J. E. Visualization and characterization of individual type III protein secretion machines in live bacteria. *Proceedings of the National Academy of Sciences* **114**, 6098–6103 (2017).
321. Hu, B., Lara-Tejero, M., Kong, Q., Galán, J. E. & Liu, J. In Situ Molecular Architecture of the Salmonella Type III Secretion Machine. *Cell* **168**, 1065-1074.e10 (2017).
322. Nowakowski, S., Mijaljica, D., Prescott, M. & Devenish, R. J. Vma8p-GFP Fusions Can Be Functionally Incorporated into V-ATPase, Suggesting Structural Flexibility at the Top of V1. *International Journal of Molecular Sciences* **12**, 4693–4704 (2011).
323. Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**, 621–628 (1994).
324. Bai, F. *et al.* Assembly dynamics and the roles of FliI ATPase of the bacterial flagellar export apparatus. *Scientific Reports* **4**, (2015).
325. Menz, R. I., Walker, J. E. & Leslie, A. G. W. Structure of Bovine Mitochondrial F1-ATPase with Nucleotide Bound to All Three Catalytic Sites. *Cell* **106**, 331–341 (2001).
326. Ueno, H., Suzuki, K. & Murata, T. Structure and dynamics of rotary V1 motor. *Cellular and Molecular Life Sciences* **75**, 1789–1802 (2018).
327. Martin, J. L., Ishmukhametov, R., Hornung, T., Ahmad, Z. & Frasch, W. D. Anatomy of F1-ATPase powered rotation. *Proceedings of the National Academy of Sciences* **111**, 3715–3720 (2014).
328. Kagawa, R., Montgomery, M. G., Braig, K., Leslie, A. G. W. & Walker, J. E. The structure of bovine F1-ATPase inhibited by ADP and beryllium fluoride. *The EMBO Journal* **23**, 2734–2744 (2004).
329. Komoriya, Y. *et al.* Principal Role of the Arginine Finger in Rotary Catalysis of F<sub>1</sub>-ATPase. *Journal of Biological Chemistry* **287**, 15134–15142 (2012).

330. Minamino, T. *et al.* FliH and FliI ensure efficient energy coupling of flagellar type III protein export in *Salmonella*. *MicrobiologyOpen* **5**, 424–435 (2016).
331. Minamino, T., Morimoto, Y. V., Hara, N., Aldridge, P. D. & Namba, K. The Bacterial Flagellar Type III Export Gate Complex Is a Dual Fuel Engine That Can Use Both H<sup>+</sup> and Na<sup>+</sup> for Flagellar Protein Export. *PLOS Pathogens* **12**, e1005495 (2016).
332. Paul, K., Erhardt, M., Hirano, T., Blair, D. F. & Hughes, K. T. Energy source of flagellar type III secretion. *Nature* **451**, 489–492 (2008).
333. Lee, P.-C., Zmina, S. E., Stopford, C. M., Toska, J. & Rietsch, A. Control of type III secretion activity and substrate specificity by the cytoplasmic regulator PcrG. *Proceedings of the National Academy of Sciences* **111**, E2027–E2036 (2014).
334. Hara, N., Morimoto, Y. V., Kawamoto, A., Namba, K. & Minamino, T. Interaction of the Extreme N-Terminal Region of FliH with FlhA Is Required for Efficient Bacterial Flagellar Protein Export. *Journal of Bacteriology* **194**, 5353–5360 (2012).
335. Ibuki, T. *et al.* Interaction between FliJ and FlhA, Components of the Bacterial Flagellar Type III Export Apparatus. *Journal of Bacteriology* **195**, 466–473 (2013).
336. Terashima, H. *et al.* *In Vitro* Reconstitution of Functional Type III Protein Export and Insights into Flagellar Assembly. *mBio* **9**, (2018).
337. Gauthier, A. & Finlay, B. B. Translocated Intimin Receptor and Its Chaperone Interact with ATPase of the Type III Secretion Apparatus of Enteropathogenic *Escherichia coli*. *Journal of Bacteriology* **185**, 6747–6755 (2003).
338. Chen, L. *et al.* Substrate-Activated Conformational Switch on Chaperones Encodes a Targeting Signal in Type III Secretion. *Cell Reports* **3**, 709–715 (2013).
339. Sorg, J. A., Blaylock, B. & Schneewind, O. Secretion signal recognition by YscN, the *Yersinia* type III secretion ATPase. *Proceedings of the National Academy of Sciences* **103**, 16490–16495 (2006).
340. Botteaux, A., Sory, M. P., Biskri, L., Parsot, C. & Allaoui, A. MxiC is secreted by and controls the substrate specificity of the *Shigella flexneri* type III secretion apparatus. *Molecular Microbiology* **71**, 449–460 (2009).
341. Minamino, T., Kinoshita, M., Imada, K. & Namba, K. Interaction between FliI ATPase and a flagellar chaperone FliT during bacterial flagellar protein export: Interaction between FliI and FliT. *Molecular Microbiology* **83**, 168–178 (2012).
342. Evans, L. D. B., Stafford, G. P., Ahmed, S., Fraser, G. M. & Hughes, C. An escort mechanism for cycling of export chaperones during flagellum assembly. *Proceedings of the National Academy of Sciences* **103**, 17474–17479 (2006).
343. Evans, L. D. B. & Hughes, C. Selective binding of virulence type III export chaperones by FliJ escort orthologues InvI and YscO. *FEMS Microbiology Letters* **293**, 292–297 (2009).
344. Portaliou, A. G. *et al.* Hierarchical protein targeting and secretion is controlled by an affinity switch in the type III secretion system of enteropathogenic *Escherichia coli*. *The EMBO Journal* **36**, 3517–3531 (2017).

345. Xing, Q. *et al.* Structures of chaperone-substrate complexes docked onto the export gate in a type III secretion system. *Nature Communications* **9**, (2018).
346. Yip, C. K., Finlay, B. B. & Strynadka, N. C. J. Structural characterization of a type III secretion system filament protein in complex with its chaperone. *Nature Structural & Molecular Biology* **12**, 75–81 (2005).
347. Okon, M. *et al.* Structural Characterization of the Type-III Pilot-Secretin Complex from *Shigella flexneri*. *Structure* **16**, 1544–1554 (2008).
348. Lario, P. I. *et al.* Structure and biochemical analysis of a secretin pilot protein. *The EMBO Journal* **24**, 1111–1121 (2005).
349. Xing, Q. *et al.* Structures of chaperone-substrate complexes docked onto the export gate in a type III secretion system. *Nature Communications* **9**, (2018).
350. Terahara, N. *et al.* Insight into structural remodeling of the FlhA ring responsible for bacterial flagellar type III protein export. *Sci. Adv.* **4**, eaao7054 (2018).
351. Noji, H., Yasuda, R., Yoshida, M. & Kinosita, K. Direct observation of the rotation of F<sub>1</sub>-ATPase. *Nature* **386**, 299–302 (1997).
352. Uchihashi, T., Iino, R., Ando, T. & Noji, H. High-Speed Atomic Force Microscopy Reveals Rotary Catalysis of Rotorless F<sub>1</sub>-ATPase. *Science* **333**, 755–758 (2011).
353. Watanabe, R., Minagawa, Y. & Noji, H. Thermodynamic analysis of F<sub>1</sub>-ATPase rotary catalysis using high-speed imaging: Thermodynamic Analysis of F<sub>1</sub>-ATPase Rotation. *Protein Science* **23**, 1773–1779 (2014).
354. Martin, J. L., Ishmukhametov, R., Hornung, T., Ahmad, Z. & Frasch, W. D. Anatomy of F<sub>1</sub>-ATPase powered rotation. *Proceedings of the National Academy of Sciences* **111**, 3715–3720 (2014).
355. Ibuki, T. *et al.* Interaction between FliJ and FlhA, Components of the Bacterial Flagellar Type III Export Apparatus. *Journal of Bacteriology* **195**, 466–473 (2013).

# Appendices

## Appendix A: Chapter 2 Supplementary Information



**Figure A.1: Details of the InvH<sup>70-147</sup> crystal structure.** **(a)** Unit cell of the InvH crystal, shown as a blue box. Two homodimers related by non-crystallographic symmetry fill the unit cell. **(b)** Interaction interface between the  $\alpha$ 2- $\alpha$ 3 loop on each protomer (oxygens red, nitrogens blue), sustained mainly by hydrogen bonds (dashed lines). **(c)** Anomalous difference map (green), contoured at 5  $\sigma$ , showing three clear densities for three cadmiums. Each is neighbored by two large densities attributed to chlorides (blue 2Fo-Fc map, 2  $\sigma$ ). **(d)** Location of cadmiums (yellow spheres) and chlorides (green spheres).

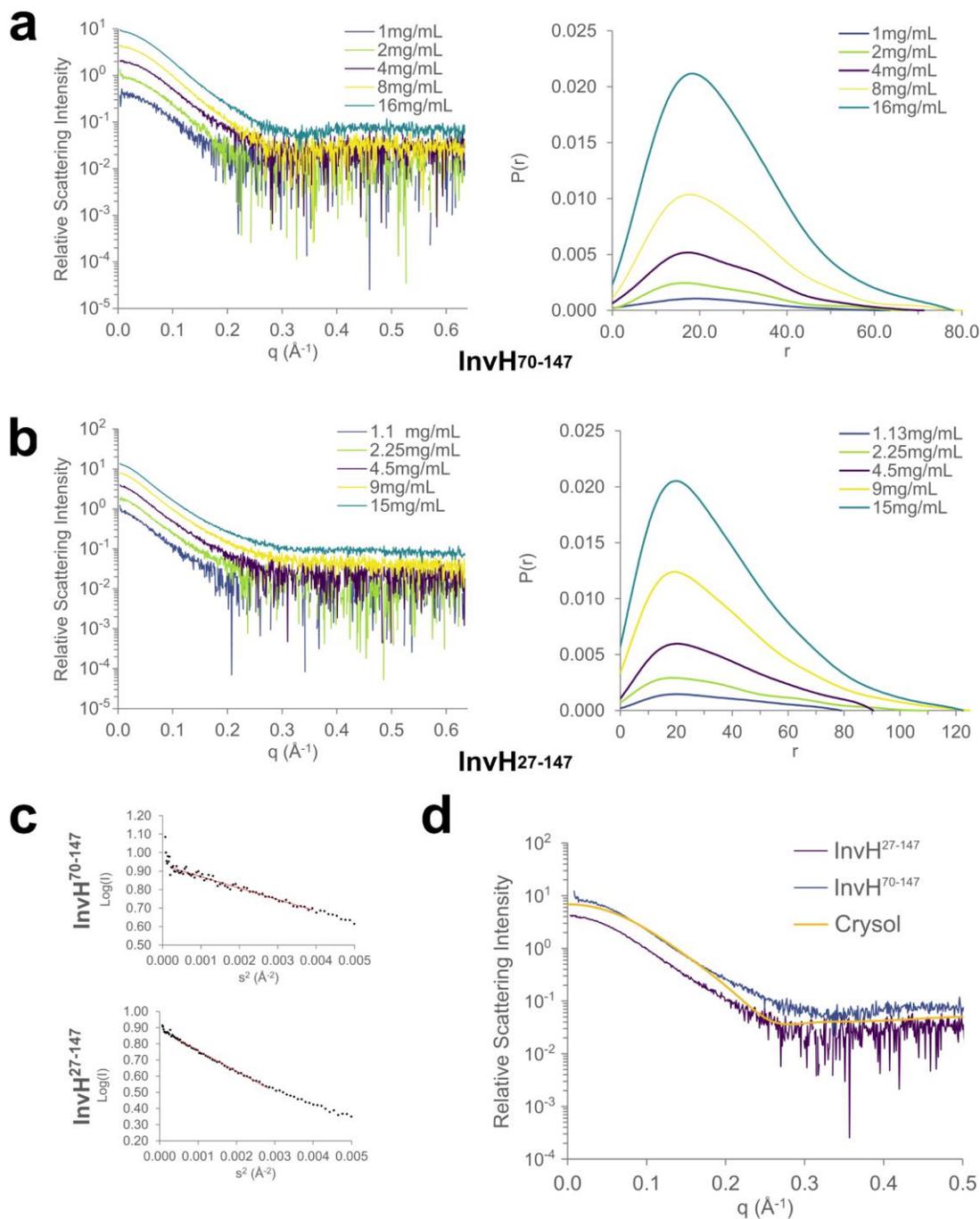
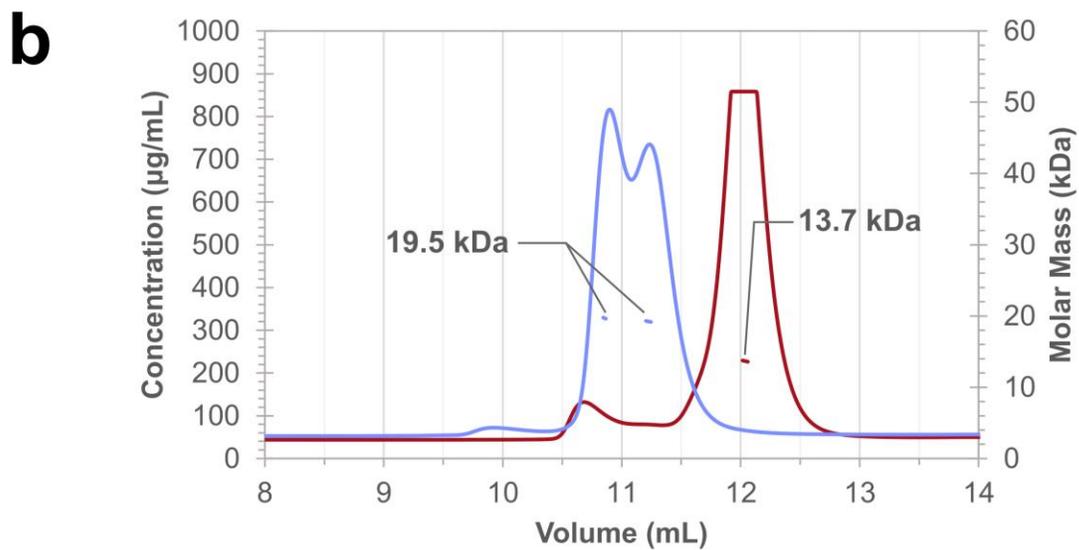
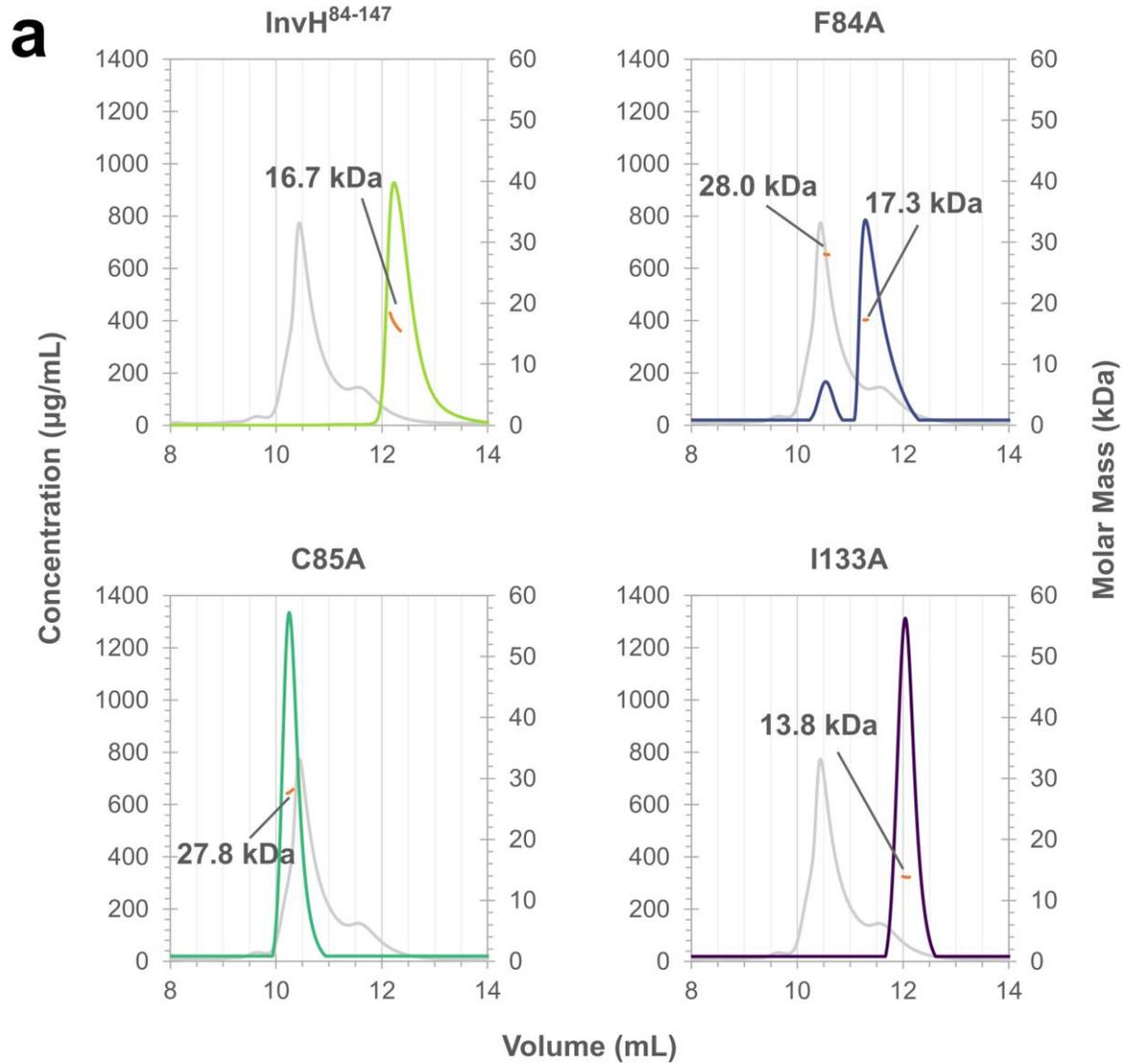


Figure A.2: Overview of SAXS data. (a, b) Relative scattering intensity (on log axis) vs  $q$  plot and  $P(r)$  vs  $r$  for several concentrations of (a) InvH<sup>70-147</sup> and (b) InvH<sup>27-147</sup>. (c) Guinier plots of InvH<sup>70-147</sup> (merged 2 and 16 mg/mL curves) and InvH<sup>27-147</sup> (merged 4.5 and 9 mg/mL curves). (d) Experimental SAXS data (relative scattering intensity on log axis vs  $q$ ) of InvH<sup>70-147</sup> (blue, merged concentrations 2 and 16 mg/mL) and InvH<sup>27-147</sup> (purple, merged concentrations 4.5 and 9 mg/mL). Overlaid is a simulated SAXS curve of the InvH<sup>70-147</sup> crystal dimer generated by Crysol (orange,  $\chi^2 = 17.5$ ).



**Figure A.3: SEC-MALS data on InvH mutants and the InvH-InvG complex. (a)** Chromatograms of SEC-MALS experiments with InvH<sup>84-147</sup> and with InvH<sup>27-147</sup> dimerization interface mutants. The UV absorbance was normalized based on the molar absorptivity of each mutant and plotted as concentration. The calculated molar masses are plotted as horizontal lines under each peak, using the secondary Y axis. The expected molecular mass for an InvH<sup>27-147</sup> monomer is 14 kDa whereas that of a dimer is 28 kDa. An InvH<sup>84-147</sup> dimer is expected to be 16 kDa. The SEC-MALS elution profile of wild-type InvH<sup>27-147</sup> is shown in grey. **(b)** SEC-MALS data of InvH<sup>70-147</sup> (red) and InvH<sup>27-147</sup> (blue) in complex with InvG<sup>520-562</sup>. The proteins were mixed in a 1:1 molar ratio and the complex purified prior to analysis by SEC-MALS. They each eluted with the molar mass expected for a 1:1 heterodimeric complex.

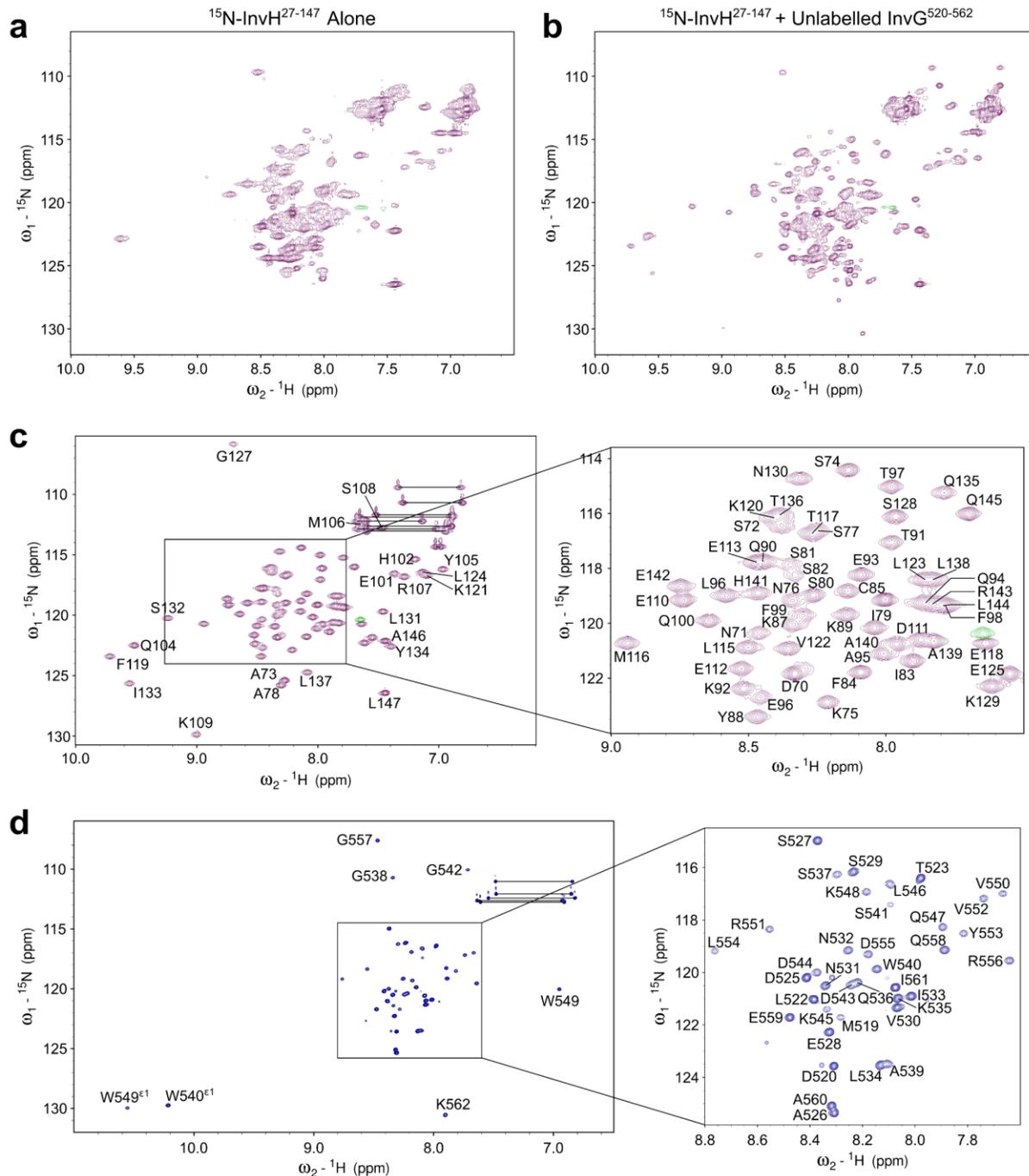
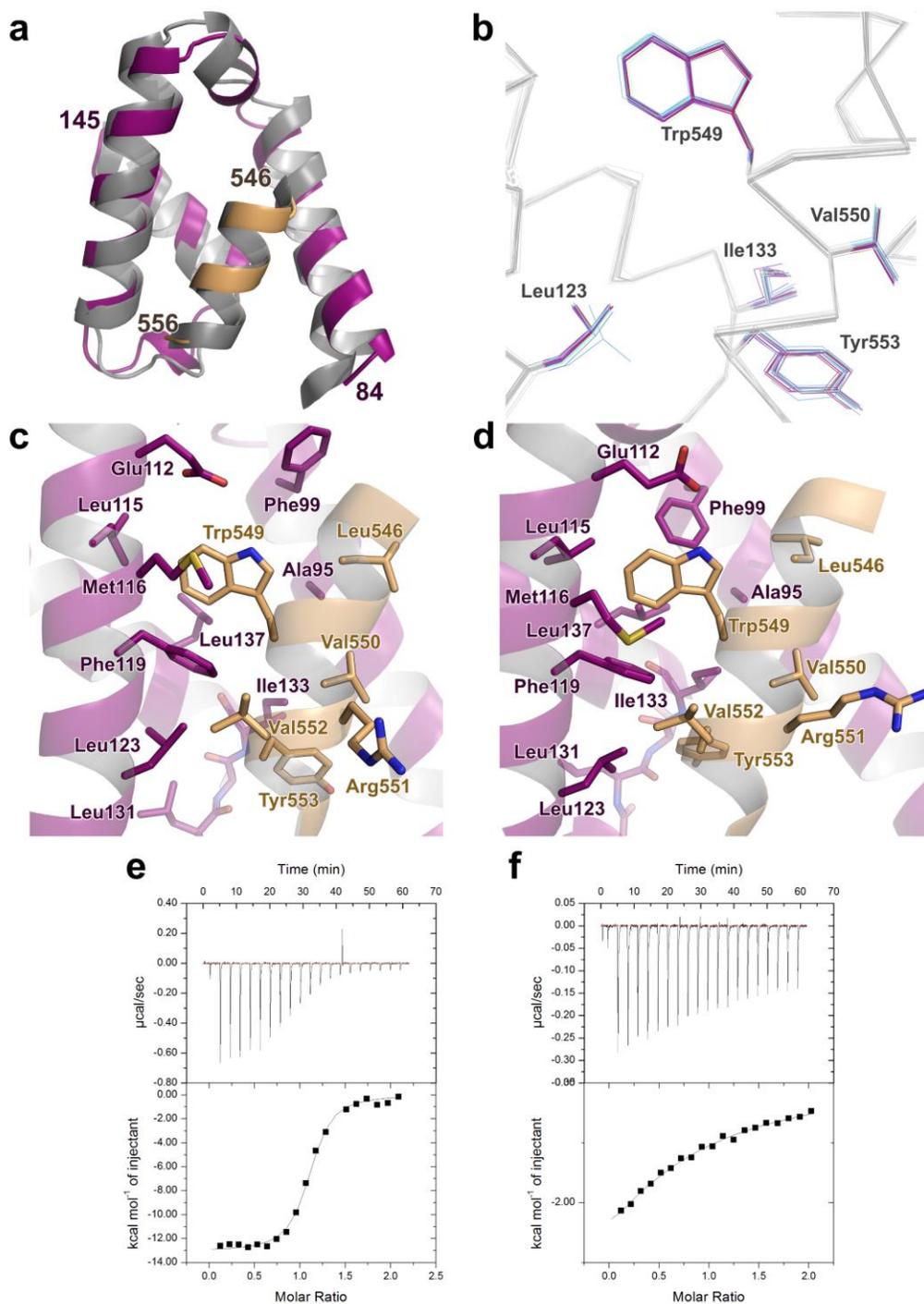


Figure A.4: NMR spectra of the InvH-InvG complex. (a) The  $^{15}\text{N}$ -HSQC spectrum of  $^{15}\text{N}$ -InvH<sup>27-147</sup>, showing broad peaks likely due to conformational exchange. (b) The  $^{15}\text{N}$ -HSQC spectrum of  $^{15}\text{N}$ -InvH<sup>27-147</sup> after addition of unlabelled InvG<sup>520-562</sup> in a 1:1 molar ratio, showing improved quality and dispersion. Signals, not present in the corresponding spectrum of the InvH<sup>70-147</sup> complex (Figure 3b), are attributed to the linker residues. Consistent with the predicted conformational disorder of the linker, these signals have  $^1\text{H}^{\text{N}}$  chemical shifts within a

relatively narrow window of approximately 8.0 - 8.5 ppm. (c) The assigned  $^{15}\text{N}$ -HSQC spectrum of  $^{13}\text{C}/^{15}\text{N}$ -InvH $^{70-147}$  in complex with 1.2 molar excess of unlabelled InvG $^{520-562}$ . (d) The assigned  $^{15}\text{N}$ -HSQC spectrum of  $^{13}\text{C}/^{15}\text{N}$ -InvG $^{520-562}$  in complex with 1.2 molar excess of unlabelled InvH $^{70-147}$ . In all spectra, negative peaks (green) are aliased signals from arginine sidechains.



**Figure A.5: Details of the InvH-InvG binding interface.** (a) Overlay of the InvH<sup>70-147</sup>-InvG<sup>520-562</sup> complex (NMR; grey; disordered residues omitted) with that of InvH<sup>84-147</sup>-InvG<sup>543-558</sup> (X-ray; InvH magenta; InvG tan). (b) Line representation of the 20 superimposed member structural ensemble of InvH<sup>70-147</sup>-InvG<sup>520-562</sup> determined by NMR spectroscopy. Highlighted are the important binding residues Leu123 and Ile133 from InvH, and Trp549, V550, and Tyr553 from InvG. (c) View of interacting residues in one member of the InvH<sup>70-147</sup>-InvG<sup>520-562</sup> ensemble (InvH magenta, InvG tan). (d) Similar view as (b) of the interacting residues from the InvH<sup>84-147</sup>-InvG<sup>543-558</sup> complex. ITC data of (e) wild-type InvG<sup>520-562</sup> titrated into InvH<sup>27-147</sup> L123A and of (f) InvG<sup>520-562</sup> W549A titrated into wild-type InvH<sup>27-147</sup>.

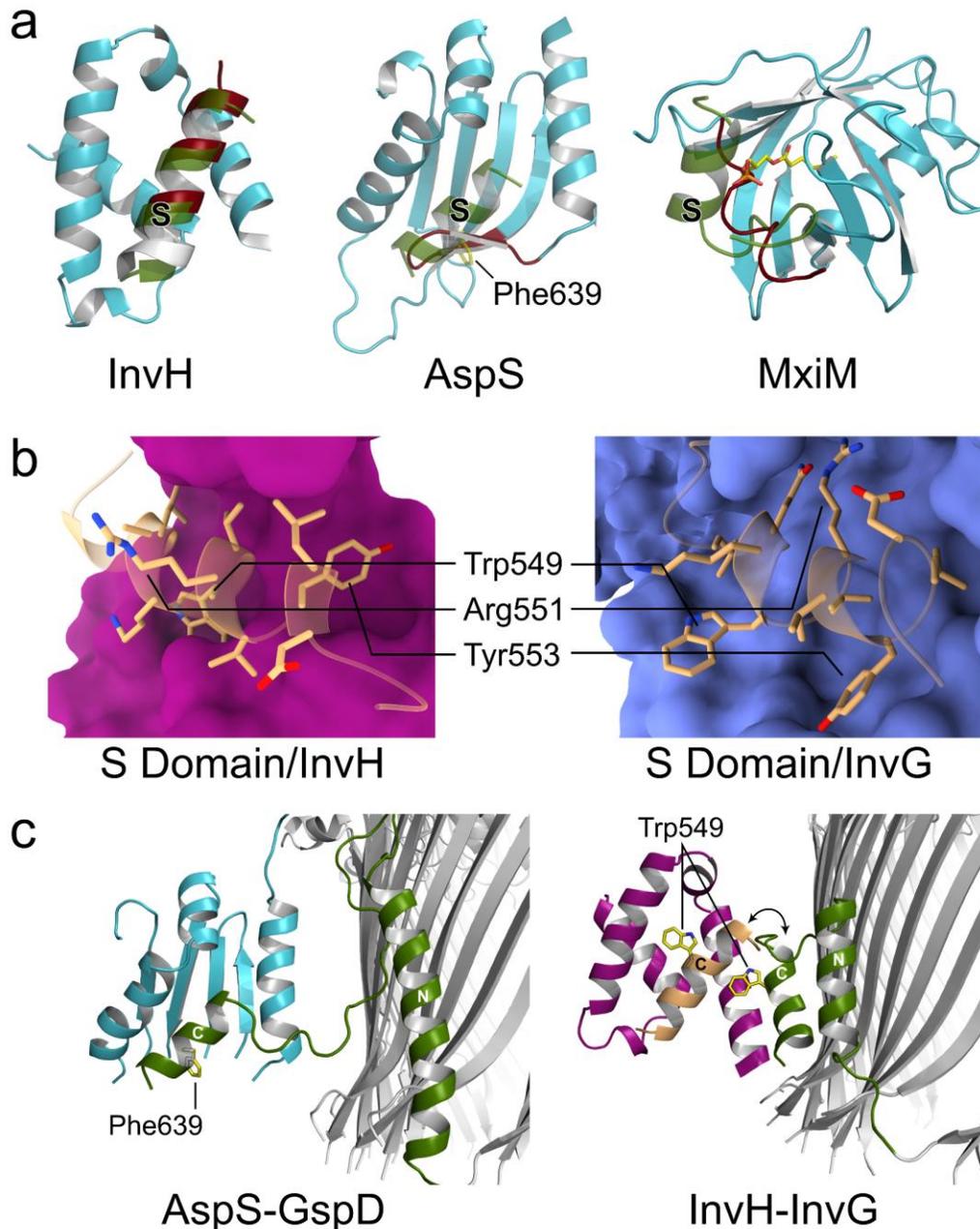


Figure A.6: Comparison of the InvH-InvG heterodimer with other structurally characterized pilotins and secretins. **(a)** Examples of pilotins (turquoise) protecting their hydrophobic S domain binding interface, with the S domain shown in green and the protecting residues coloured red. The lipid bound in the hydrophobic cavity of MxiM is illustrated in yellow. **(b)** Binding interface of the InvG<sup>544-558</sup> S domain helix (tan) with InvH (left, magenta) and with the assembled secretin  $\beta$ -barrel (right, blue). **(c)** The cryo-EM structure of AspS (turquoise; left) bound to assembled secretin GspD (grey), highlighting the pilotin-sequestered C-terminal S domain (green) helix. The right panel compares the cryo-EM structure of the assembled InvG secretin (grey, S domain green) with a hypothetical positioning of the InvH<sup>84-147</sup>-InvG<sup>543-558</sup> heterodimer (magenta and tan respectively) if it were bound to the assembled secretin.

Table A.1: SAXS radius of gyration and molecular weight estimates for InvH

InvH <sup>70-147</sup>						
Sample Concentration (mg/mL)	R <sub>g</sub> <sup>Guinier</sup> (Å)	R <sub>g</sub> <sup>P(r)</sup> (Å)	MW <sup>Porod</sup> (kDa)	MW <sup>Vc</sup> (kDa)	MW <sup>DAM</sup> (kDa)	Mean MW (kDa)
1.0	19.5	19.8	19.5	14.2	7.5*	16.9
2.0	20.5	20.9	17.7	16.6	18.8	17.2
4.0	20.4	20.9	15.8	16.3	16.9	16.1
8.0	21.4	22.2	17.8	20.0	19.1	18.9
16.0	22.1	22.6	18.4	20.3	19.7	19.4
Mean	20.8	21.3	17.8	17.5	18.3	17.7
InvH <sup>27-147</sup>						
Sample Concentration (mg/mL)	R <sub>g</sub> <sup>Guinier</sup> (Å)	R <sub>g</sub> <sup>P(r)</sup> (Å)	MW <sup>Porod</sup> (kDa)	MW <sup>Vc</sup> (kDa)	MW <sup>DAM</sup> (kDa)	Mean MW (kDa)
1.1	31.1	28.5	31.4	28.8	35.1	31.8
2.3	29.7	31.8	31.1	28.9	32.8	30.9
4.5	30.1	30.6	32.6	31.7	36.9	33.7
9.0	30.6	33.4	33.6	33.3	38.8	35.2
15.0	30.8	33.3	33.1	32.4	35.7	33.7
Mean	30.4	31.5	32.4	31.0	35.9	33.1

\*Outlier value not included in mean calculations

Table A.2: ITC-calculated binding affinities, enthalpies and entropies for InvH-InvG binding mutants

Mutant	N (sites ± SD)	Kd (μM ± SD)	ΔH (kcal/mol ± SD)	ΔS (cal/mol/° ± SD)
L123A	1.1 ± 0.1	0.23 ± 0.04	-13.1 ± 0.3	-14 ± 1
I133A	1.28 ± 0.03	0.35 ± 0.01	-13.7 ± 0.1	-16.4 ± 0.4
I133R	1.4 ± 0.1	220 ± 50	-8 ± 2	-8 ± 6
W549A	0.77 ± 0.06	70 ± 40	-4 ± 2	6 ± 7
V550A	1.05 ± 0.01	0.38 ± 0.04	-15.1 ± 0.2	-21 ± 1
R551A	1.15 ± 0.01	0.4 ± 0.1	-12.9 ± 1	-14 ± 4
Y553A	0.93 ± 0.05	7 ± 1	-7.8 ± 0.3	-3 ± 1

Appendix B: Chapter 3 Supplementary Information

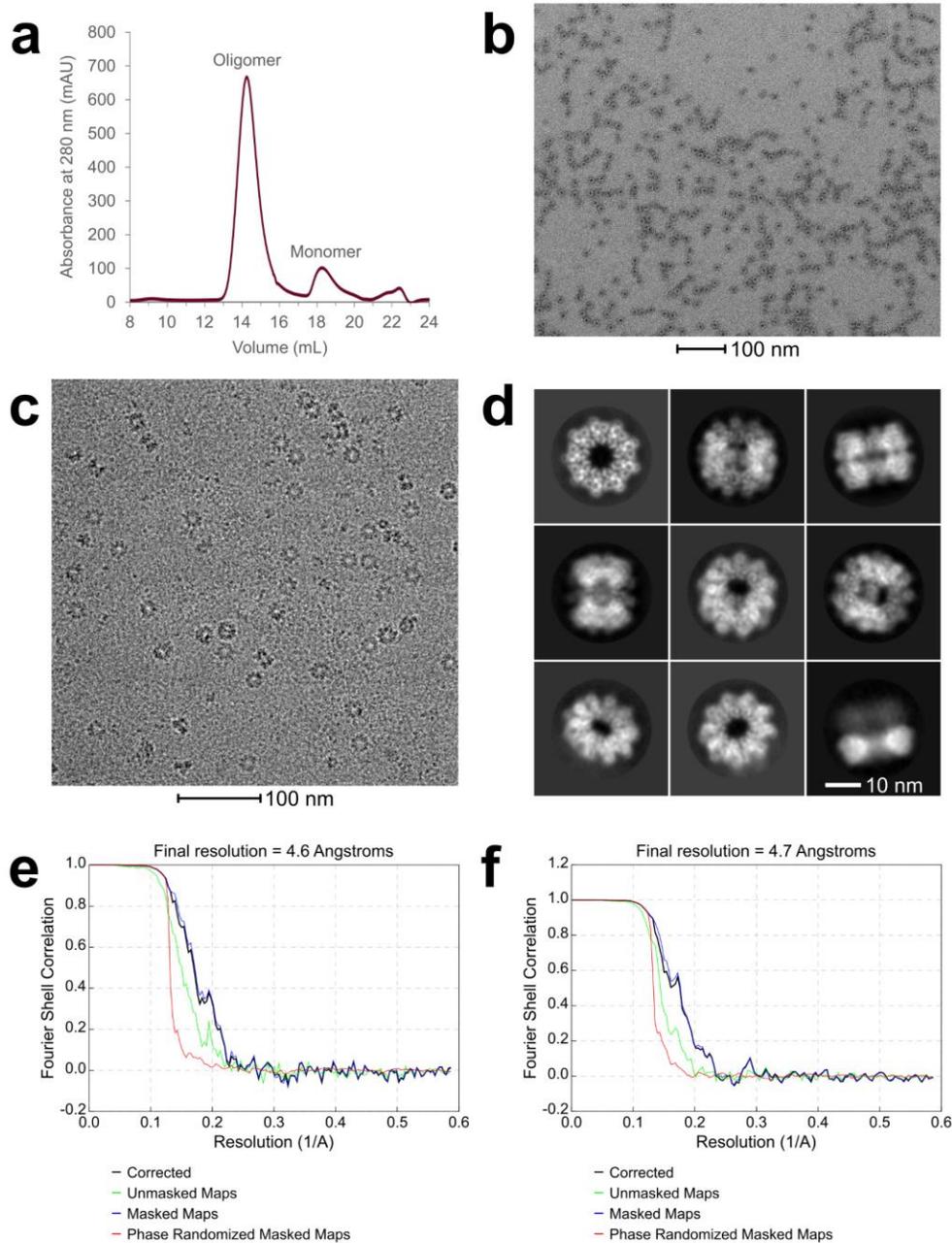


Figure B.1: Cryo-EM data collection and refinement of EscV<sub>C</sub>. (a) Size exclusion chromatogram of EscV<sub>C</sub>, with a primarily oligomeric population (estimated by elution volume off of a superpose 6 column). (b) Negative stain and (c) cryo-EM micrographs of EscV<sub>C</sub>, with clear ~15 nm rings. (d) 2D classification of EscV<sub>C</sub> from Relion3. Some particles lack the stacked ring, visible when comparing the top right and bottom right classes. FSC curves of (e) the D9 and (f) the C9 EscV<sub>C</sub> reconstructions.

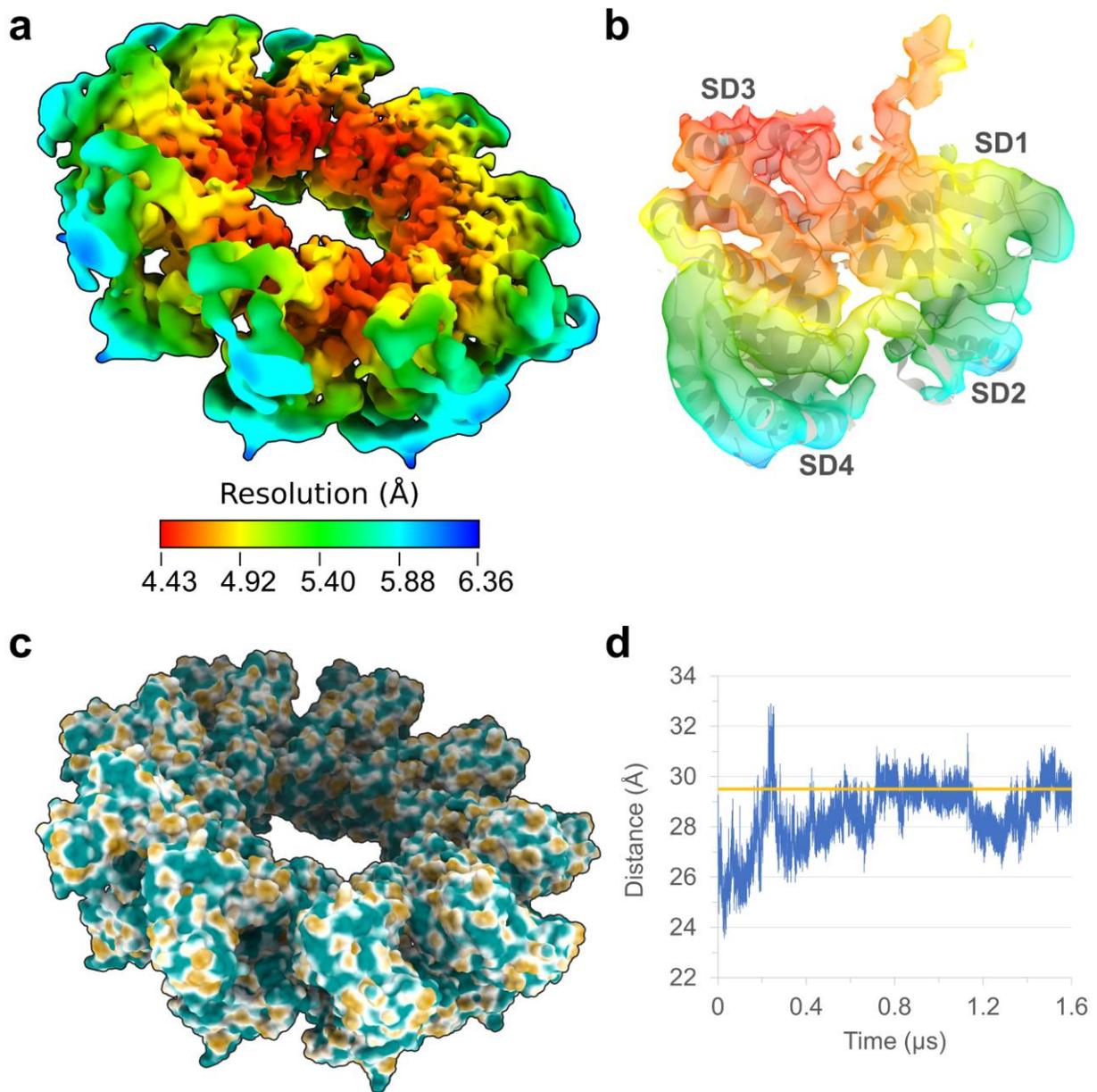


Figure B.2: EscV<sub>C</sub> surface and conformation. **(a)** Local resolution estimation by Relion3 of the EscV<sub>C</sub> C9 reconstruction. **(b)** Overlay of the local resolution map with an EscV<sub>C</sub> monomer, highlighting the lower resolution of SD2 and SD4. **(c)** hydrophobic surface of EscV<sub>C</sub>, with no large hydrophobic pockets but conservation of the hydrophobic chaperone binding interface identified in FlhA. **(d)** Distance between the centre of mass of SD2 and SD4 throughout a 1.6 μs molecular dynamics simulation. The same distance measured in the cryo-EM structure is denoted with an orange line.

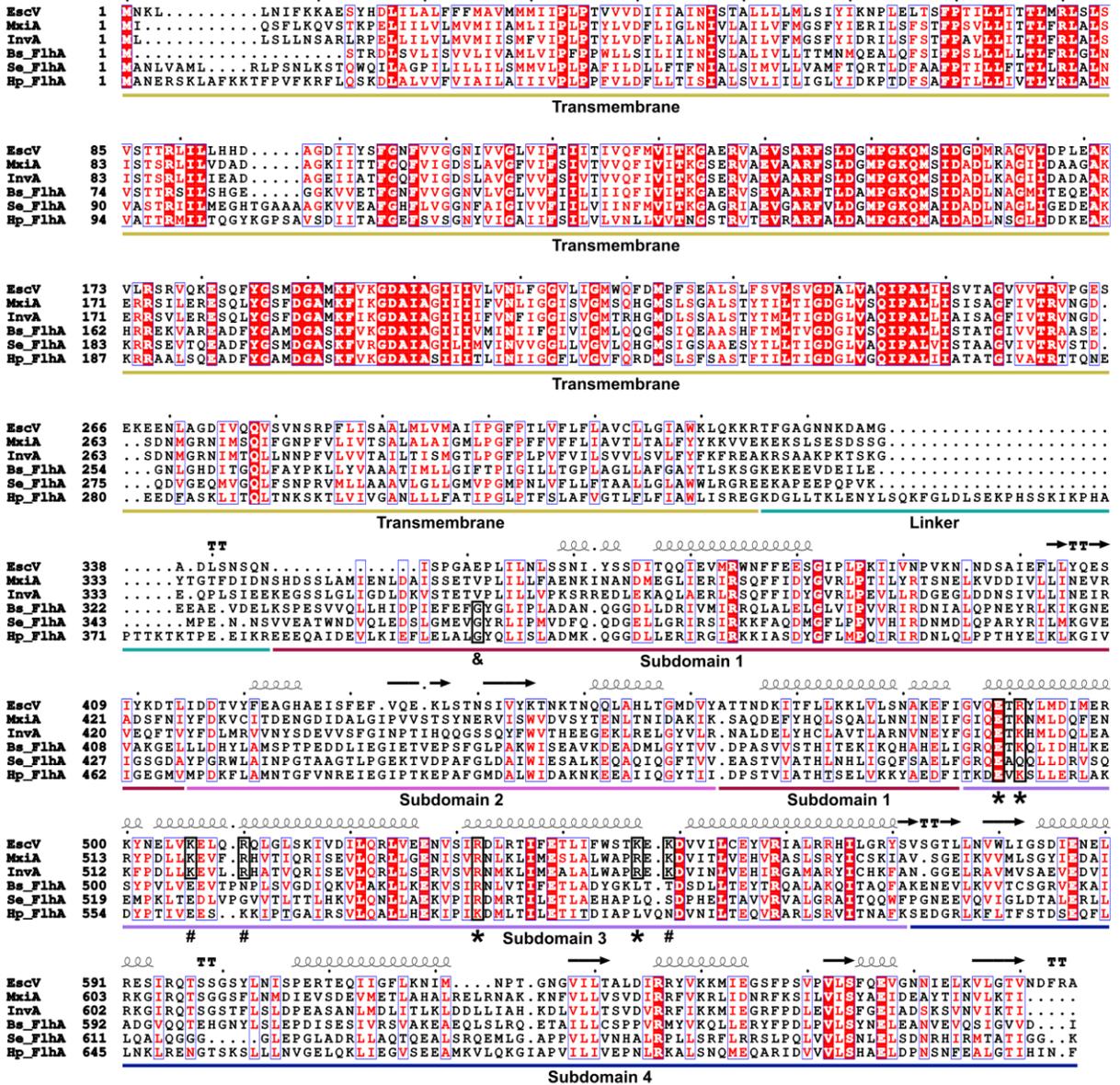
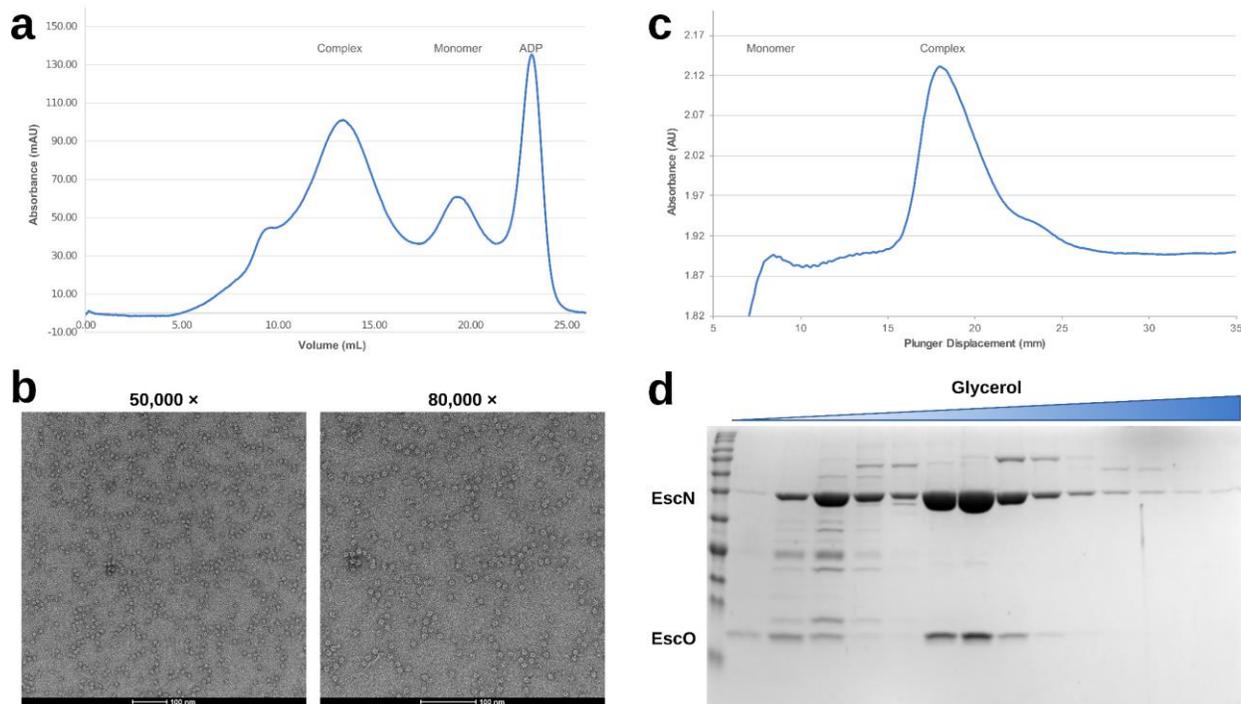


Figure B.3: Multiple sequence alignment. Alignment was created using PSI-Coffee and Esprpt3. Full-length EscV is aligned with two injectisome homologues (*Shigella flexneri* MxiA and *Salmonella enterica* InvA) and three flagellar homologues (FlhA from *Bacillus subtilis*, *Salmonella enterica*, and *Helicobacter pylori*). All homologues are structurally characterized and were used for model building. Conserved residues from SD3 marked with an asterisk have been shown to be important for MxiA oligomerization, while those marked with a pound sign line the pore and contribute to secretion. Conserved glycine from SD1 in FlhA is marked with an ampersand; note the lack of conservation with injectisome homologues.

## Appendix C: Chapter 4 Supplementary Information



**Figure C.1: EscN-EscO complex purification.** **(a)** Size-exclusion UV chromatogram of the EscN-EscO complex using a Superose 6 10/300 column (GE Healthcare), with EscN-EscO predominantly in an oligomeric population. 500  $\mu$ L of EscN (10 mg/mL) EscO (1 mg/mL) mixture were injected. **(b)** Negative-stain micrographs of the EscN-EscO complex with clear  $\sim$ 10 nm rings. **(c)** UV chromatogram and **(d)** SDS-PAGE of 10-25% glycerol gradient purification, showing EscO co-sedimenting with EscN in a large oligomeric population at high purity. A 200  $\mu$ L mixture of EscN (2.0 mg/mL) and EscO (0.7 mg/mL) were loaded onto the gradient.

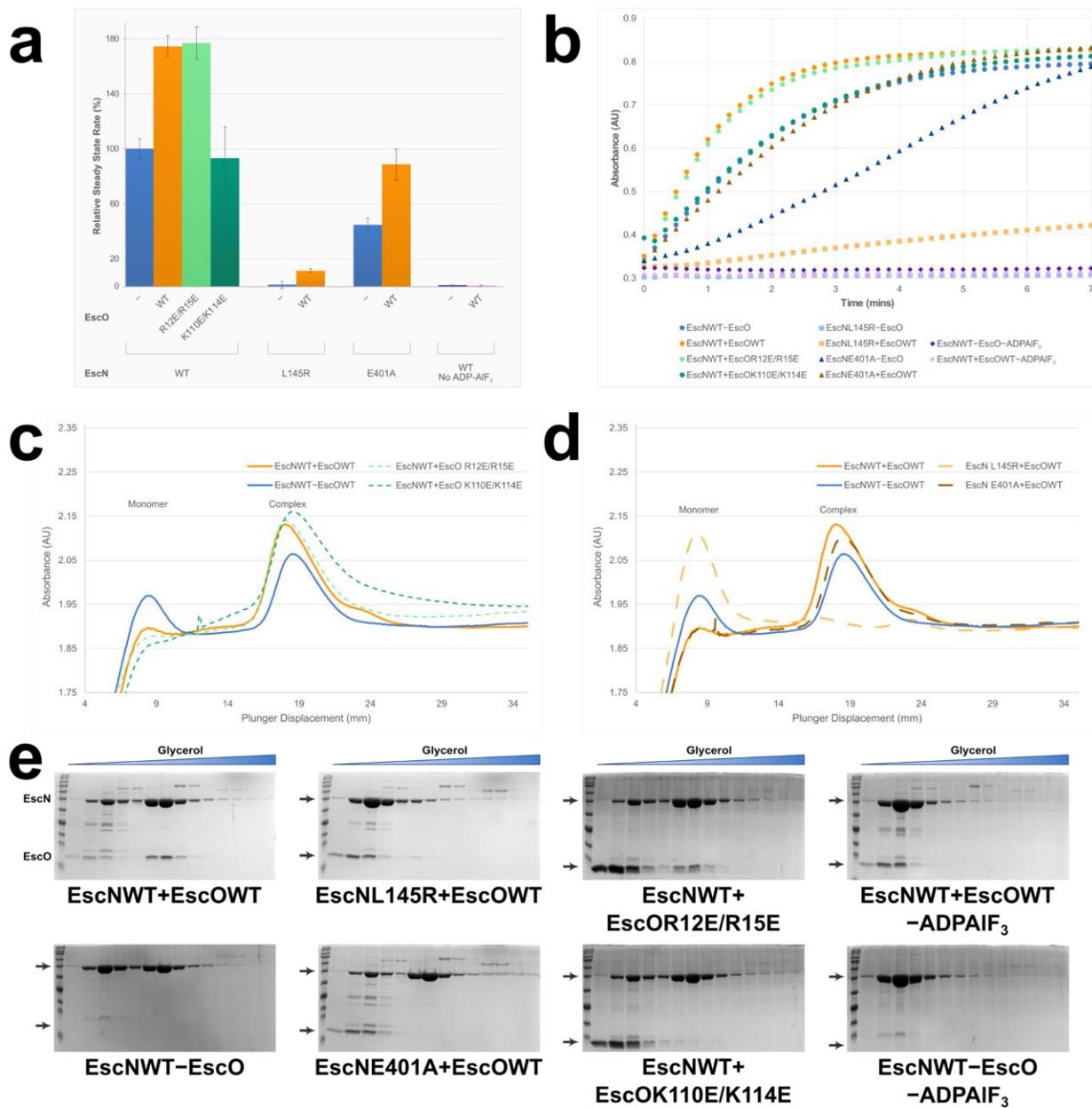


Figure C.2: Kinetics and glycerol gradient analysis. Steady-state kinetic analysis of EscN-EscO rate assayed by EnzChek phosphate detection assay, shown as a **(a)** bar graph of steady-state slope taken over a linear section of one minute (error bars show standard deviation from n=4 slopes), and **(b)** A<sub>360</sub> time-course showing EnzChek product formation over time. Both plots are averaged from four replicates. **(c,d)** Glycerol gradient UV chromatograms of EscN-EscO complex, showing monomeric vs oligomerized populations for various mutants. A 200  $\mu$ L mixture of EscN (2.0 mg/mL) and/or EscO (0.7 mg/mL) were loaded onto the gradient. **(e)** Gels of glycerol gradient purifications, highlighting the oligomerization incompetence of EscN L145R and EscN lacking ADPAIF<sub>3</sub>, and the destabilization of the EscN-EscO interface in EscN E401A and EscO K110E/K114E.

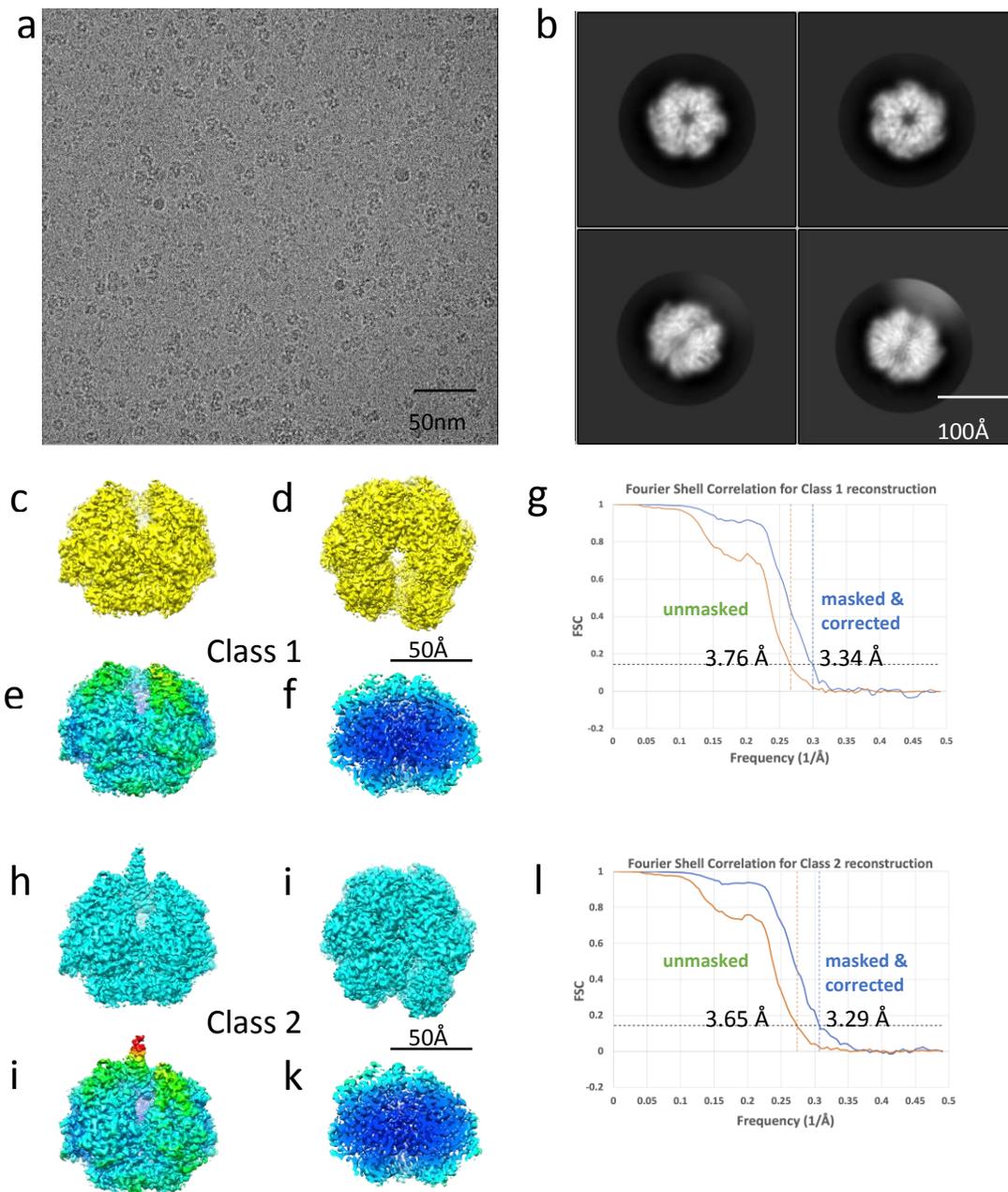
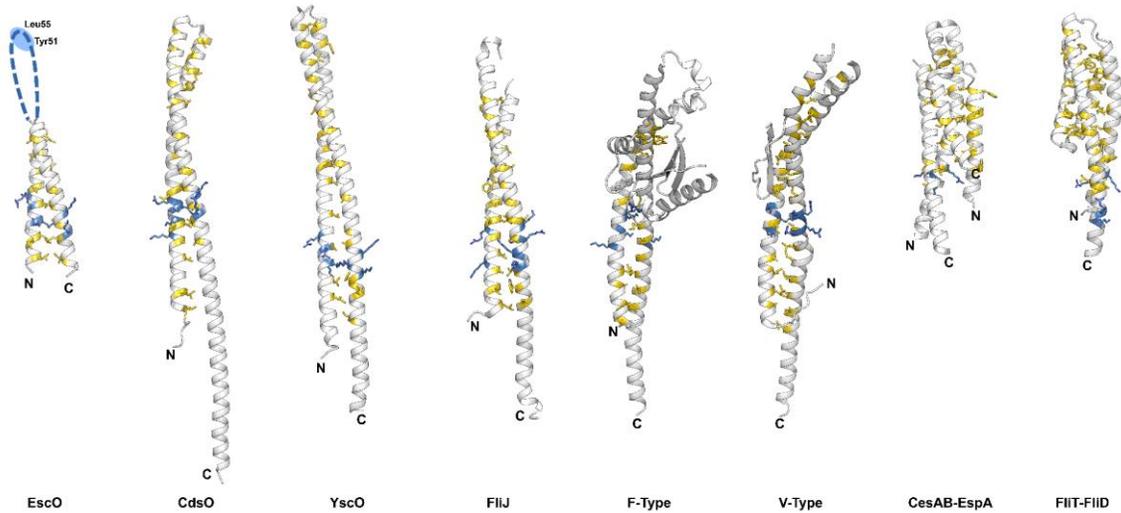


Figure C.3: Cryo-EM imaging and reconstruction of the EscN/EscO complex. **(a)** Representative micrograph. **(b)** Selected reference-free 2D class averages. **(c)** The top and **(d)** the side view of the class 1 reconstruction. **(e)** The side and **(f)** side-slab view of the class 1 reconstruction colored according to local resolution (blue – 3.2 Å, cyan – 3.8 Å, green – 4.4 Å, yellow – 5.0 Å, red – 5.6 Å). **(g)** FSC curve of the class 1 reconstruction using gold-standard refinement calculated from unmasked half maps and masked with the correction of soft masking effect. **(h)** The top and **(i)** the side view of the class 2 reconstruction. **(j)** The side and **(k)** side-slab view of the class 2 reconstruction colored according to local resolution as in **(e)** and **(f)**. **(l)** FSC curve of the class 2 reconstruction as in **(g)**.



**Figure C.4: EscO structural analysis.** Comparison of EscO coiled coil secondary structure with characterized vT3SS homologues CdsO (PDB 3K29) and YscO (4MH6), fT3SS FliJ (3AJW), F<sub>1</sub>  $\gamma$ -subunit (1H8E), V<sub>1</sub> D subunit (3VR6), and chaperone-effector complexes CesAB/EspA (1XOU) and FliT-FliD (6CH2). A ring of positively charged residues, mirroring those found in EscO to be positioned at the EscN insertion interface, are shown in blue; hydrophobic residues facilitating the coiled coil are coloured yellow. The unresolved region of EscO is demarcated with a dotted line, and the approximate location of export gate interacting residues Tyr51 and Leu55 is shown in a blue circle.

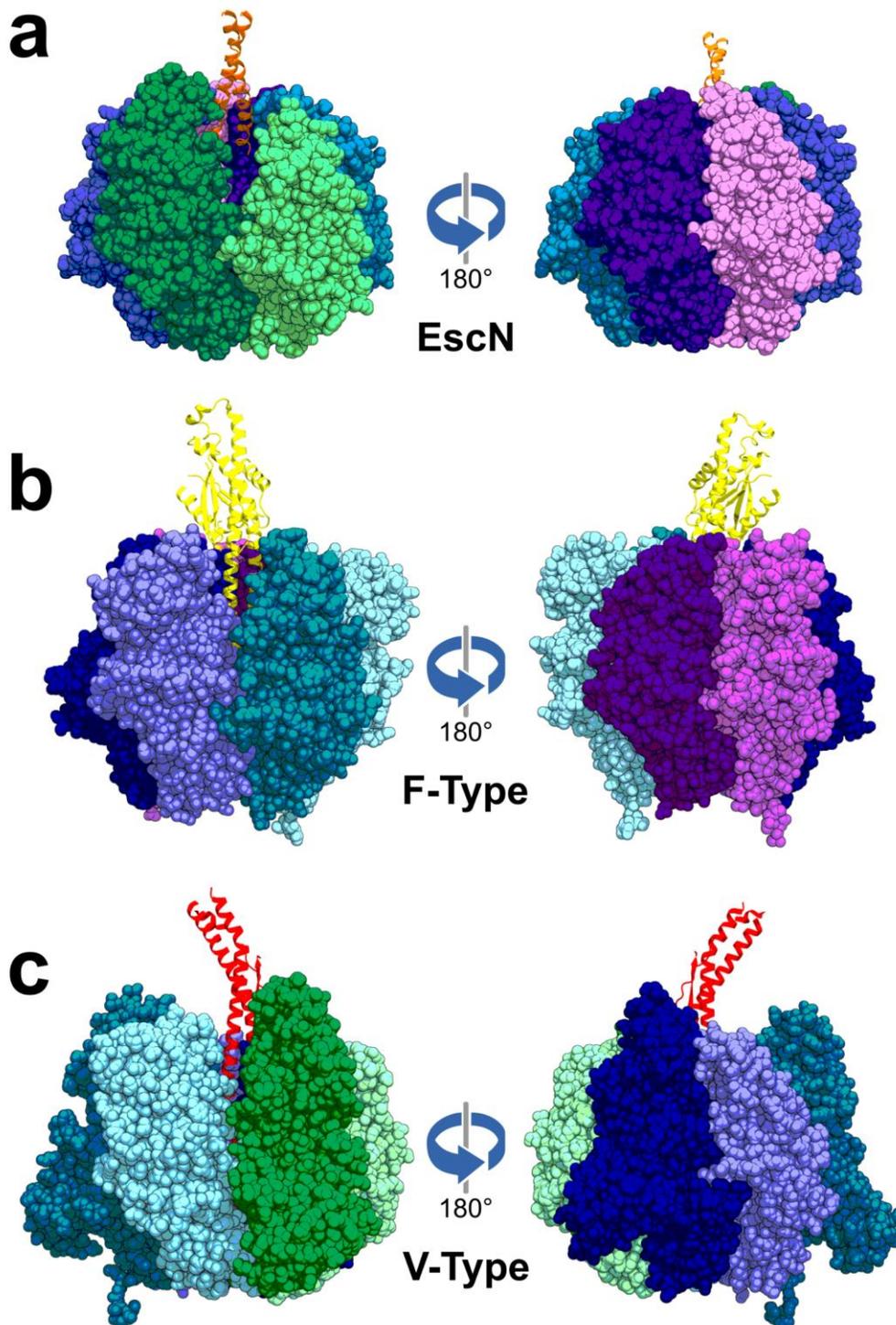


Figure C.5: Asymmetry of the EscN homohexamer. Comparison of quaternary structure of (a) EscN-EscO complex, (b) F-type ATPase  $\beta_3\alpha_3\gamma$  complex (1H8E), and (c) V-type ATPase  $A_3B_3D$  complex (3VR6), showing the front view (left) with a prominent cleft in each structure versus comparatively tight packing in the rear view (right).

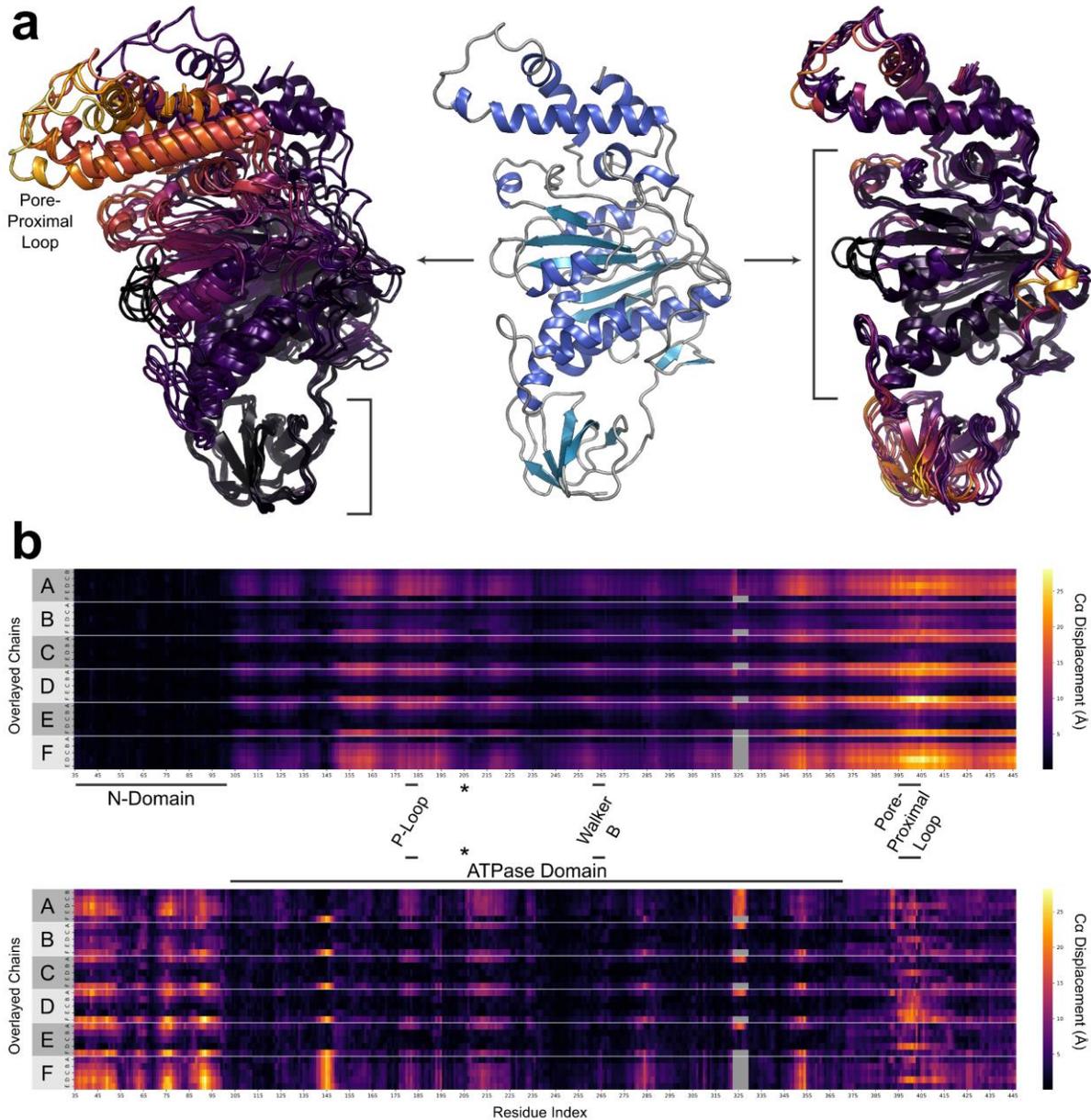


Figure C.6: Differences in conformation between EscN chains. **(a)** Alignment of EscN chains  $N_A$  through  $N_E$  on chain  $N_F$ , along the N-terminal domain (left) and the ATPase domain (right), coloured yellow in areas of high displacement between compared residues. **(b)** Heat maps of  $C\alpha$  displacement between residues when overlaying all combinations of EscN chains, aligned by the N-terminal domain (top) and ATPase domain (bottom). Locations of key motifs are labelled, with the catalytic glutamate highlighted with an asterisk.

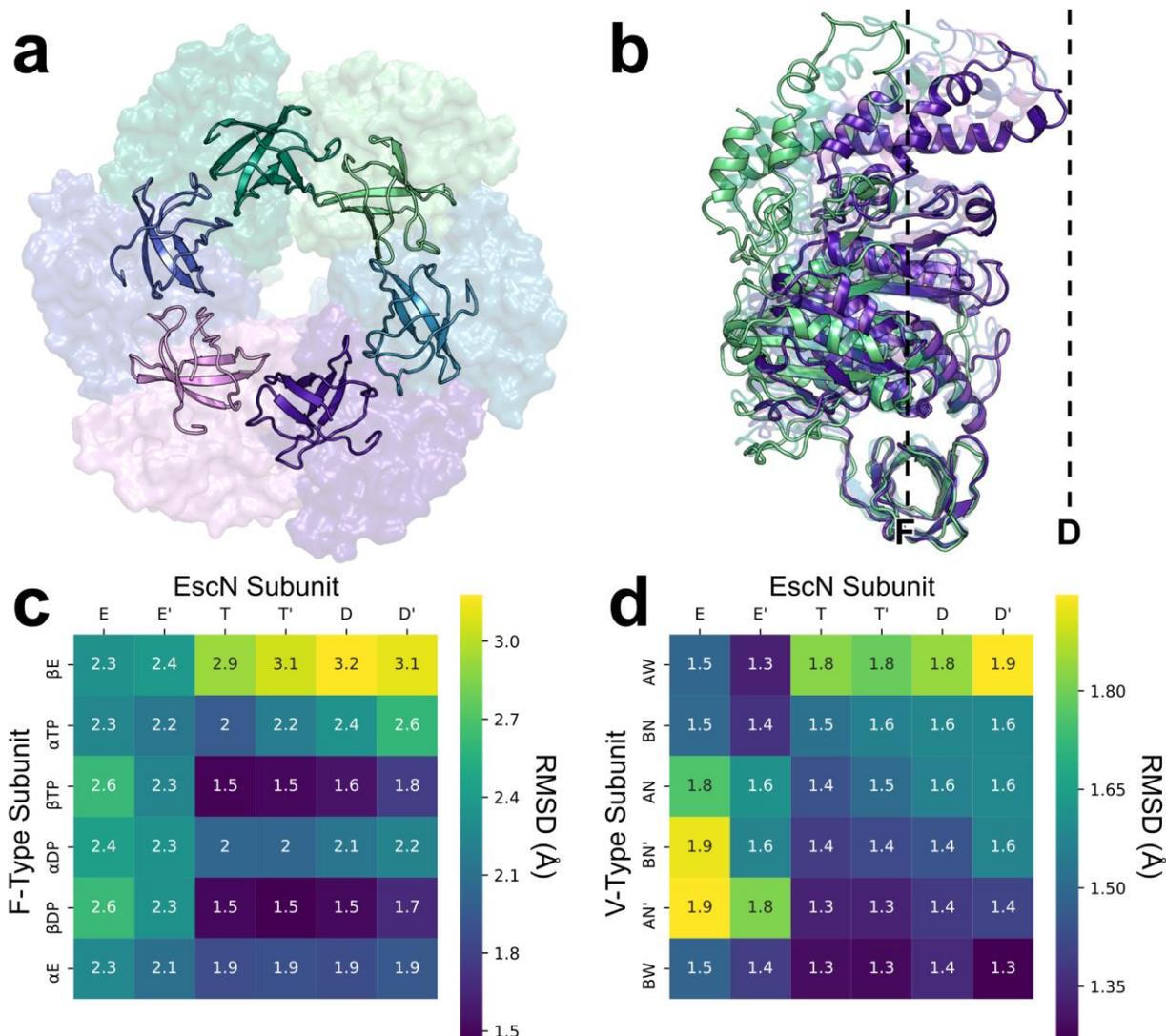


Figure C.7: EscN homohexamers formation and comparison with F- and V-ATPases. **(a)** Bottom view of the EscN class 1 N-terminal domains, highlighting their near-C6 symmetry. **(b)** Side view of EscN chains overlaid by their N-terminal domains, illustrating the strong tilt towards the pore by chain  $N_D$  (the D state) as compared to chain  $N_F$  (the E' state). **(c)** and **(d)** Heat maps of overall RMSDs comparing all ATPase catalytic conformations of EscN with those from  $F_1$ -ATPase (left, 1E1R) and  $V_1$ -ATPase (right, 3VR6).

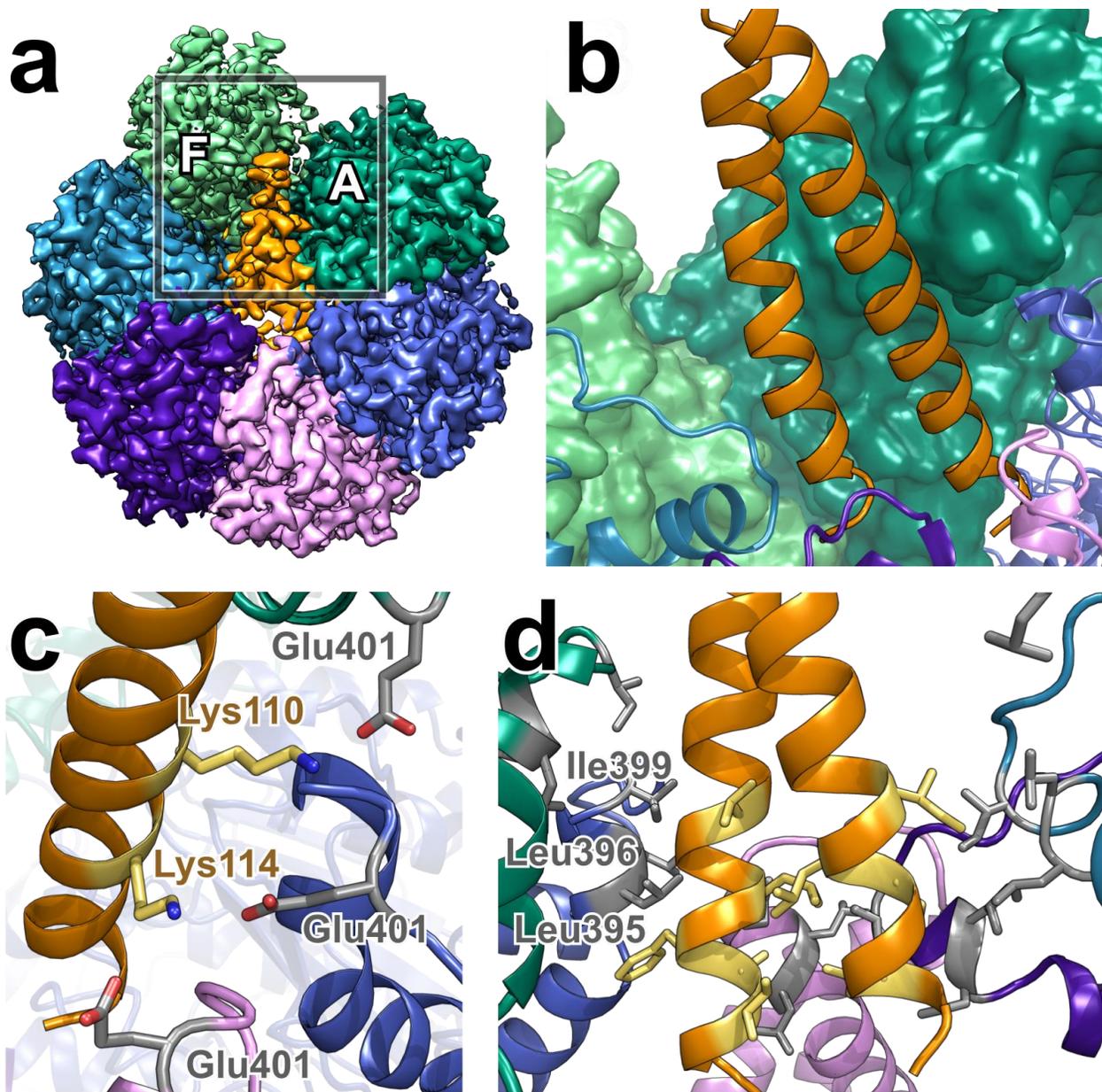


Figure C.8: EscO interaction interface with EscN. **(a)** Class 2 density of the EscN-EscO complex, demonstrating the relatively lower signal from chain N<sub>F</sub> (the E' state), indicating higher dynamicity. **(b)** Binding pocket of EscO in between EscN chains N<sub>F</sub> and N<sub>A</sub>. **(c)** Interaction between EscN Glu401 from chains N<sub>A</sub>, N<sub>B</sub>, and N<sub>C</sub> with EscO Lys110 and Lys114, an important interface for EscN-EscO complex formation and catalytic activity. **(d)** Hydrophobic ring of Leu395, Leu396, and Ile399 formed by EscN, which interacts with hydrophobic residues on EscO to stabilize it in the pore.



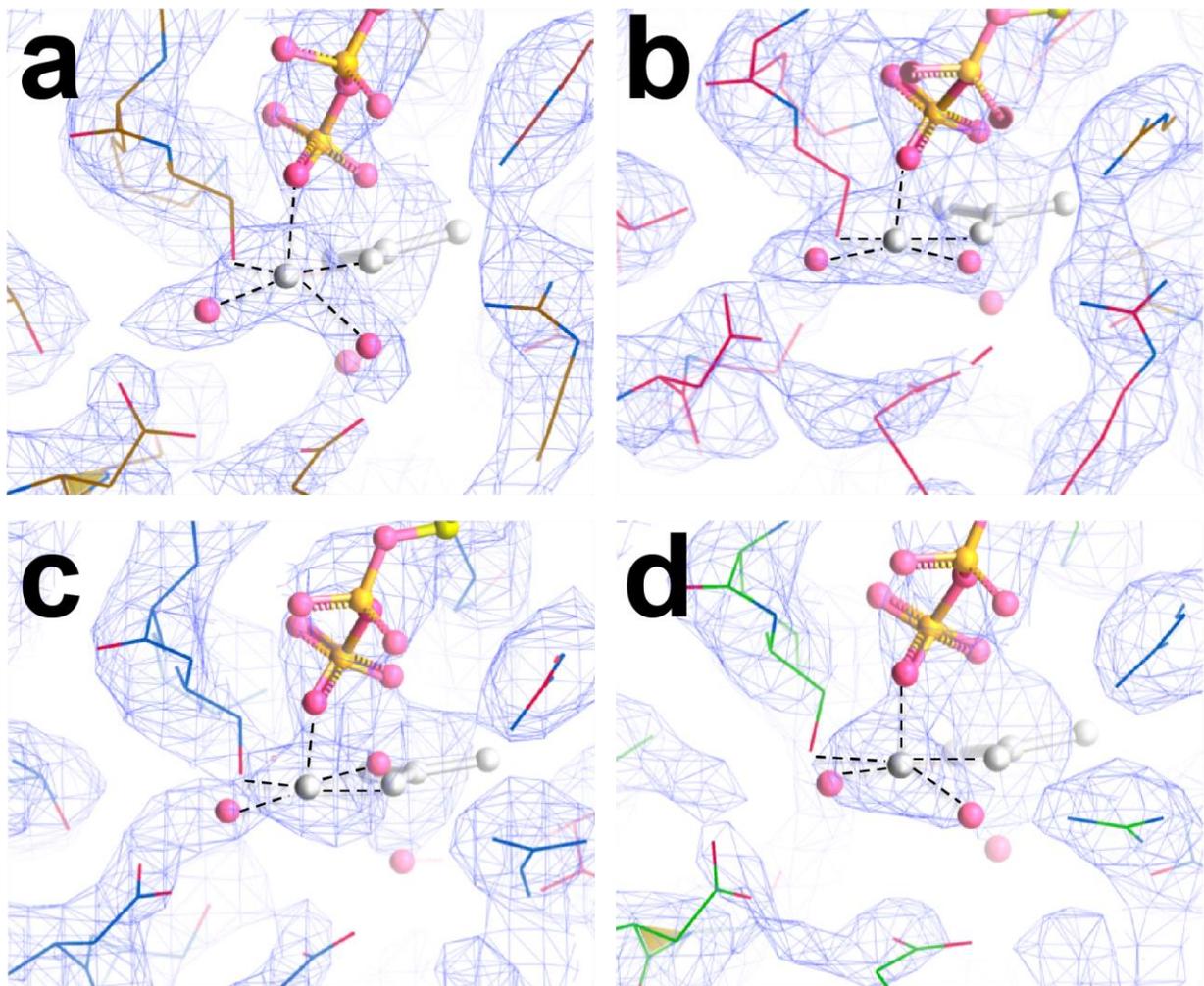


Figure C.10: Differences in magnesium ion coordination between the four ligand-bound active sites. The coordination moves from a trigonal bipyramidal geometry in **(a)** site T, with the waters migrating to reflect the expected octahedral geometry (with one water not resolved) in **(b)** site T', **(c)** site D, and **(d)** site D'.