
**Structural studies of the inner-
membrane platform of the bacterial type
II secretion system**

Hui Zhang

**QUEEN MARY, UNIVERSITY OF
LONDON**

Submitted in partial fulfilment of the requirement of
the degree of Doctor of Philosophy

February 2018

Declaration

I, Hui Zhang, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: 

Date: 25/01/2018

Acknowledgments

I would like to begin these acknowledgements by thanking my supervisor Professor Richard Pickersgill for giving me this fantastic opportunity to work in his group. I really appreciate for everything that he taught me, all the opportunities, encouragement and helps he has given to me.

This thesis would not be completed without expertise assistance and support from Alain Oregioni in Crick institute NMR centre, Tillmann Pape in Imperial Electron Microscopy centre, Dr Vidya Darbari in SBCS of Queen Mary. Yours unrelenting enthusiasm, patience and good inspired me a lot. It was lucky for me to be in a very lively and helpful lab. We shared lots of thoughts and laughs. Thank you all in the lab, Shuang, Liang, Ismail, Ben, Tania, Jimena, Jack, Yumi, Nadine and Dave.

Last but not least, I would like to thank my friends and family, especially my mum, dad and my husband. Their persistent encouragement, love and patience are with me all the time. I hope I've done you all proud.

The experience of undertaking this PhD will be with me forever.

Abstract

The type II secretion system (T2SS) is widespread in Gram-negative bacteria that cause disease in animals and plants. In human and animal pathogens toxins are secreted (e.g. cholera toxin) and in plant pathogens lytic enzymes that breakdown the plant cell wall are exported in to the extracellular milieu (e.g. pectate lyase). Structurally the T2SS comprises at least 11 core proteins that form three major subassemblies spanning the inner-membrane, periplasmic space and outer-membrane: (i) the inner-membrane platform and associated cytoplasmic ATPase (E); (ii) the pseudopilus, which consists of five pseudopilins, G to K; and (iii) a large, pore-forming outer-membrane complex secretin D.

The inner-membrane platform comprises three single transmembrane helix proteins, and one three transmembrane helix protein, OutF. The evidence from cryo-electron microscopy on the related type IVa pilus machine (T4PS) places the protein corresponding to OutF at the centre of this platform. This platform is responsible for assembling the pilus and for communicating between the periplasm and the cytoplasmic ATPase. To date, no high-resolution structure of a full-length OutF/PilC family protein is available. A low-resolution electron microscopy reconstruction of isolated PilG (PilC ortholog from *Neisseria meningitidis* T4PS) showed a tetrameric two lobed structure.

Here I report the results of studying the structure of the inner-membrane protein OutF from *Dickeya dadantii* and the complete inner-membrane platform comprising 9 proteins: OutEFGHIJKLM. This work involved cloning the corresponding operon, purifying the proteins, and using crystallography and electron microscopy. Key results reported here are the crystal structure of the first cytoplasmic domain of *Dickeya dadantii*, OutF⁶⁵⁻¹⁷² and a preliminary three-dimensional model of the

***Dickeya dadantii* inner-membrane platform. This model, and higher-resolution models to come, will provide valuable information about the oligomeric state, and arrangement of the inner-membrane proteins. These studies will help us to understand how the type II secretion system works.**

Table of content

Chapter 1	Introduction	20
1.1	Secretion in bacteria	20
1.1.1	Gram-negative bacteria secretion system.....	20
1.2	Type II secretion system	21
1.2.1	The discovery of T2SS.....	21
1.2.2	T2SS related disease	22
1.2.3	The function of T2SS	22
a	Assembly of surface organelles.....	22
b	Protein transport	23
1.2.4	The secretion model	23
1.3	The relationship between T2SS and other systems in bacteria and archaea	24
1.4	The structure of T2SS	26
1.4.1	Pseudopilus	29
a	Major pseudopilins.....	31
b	Minor pseudopilins	32
1.4.2	Inner-membrane platform	33
1.4.3	The secretin channel.....	34
1.4.4	Pilotins	36
1.5	The inner-membrane platform proteins	37
1.5.1	General secretion pathway protein, GspE	38
1.5.2	General secretion pathway protein, GspF	40
1.5.3	General secretion pathway protein, GspM.....	42
1.5.4	General secretion pathway protein, GspL.....	42
1.5.5	General secretion pathway protein, GspC.....	43
1.6	Interactions within the inner-membrane platform.....	46
Chapter 2	Material and methods	48
2.1	Materials.....	48
2.1.1	Water	48
2.1.2	Agarose gel electrophoresis	48
2.1.3	DNA clean-up	48
2.1.4	Polymerase Chain Reaction	49
2.1.5	Broth and media	49
a	Luria-Bertani (LB) and YT broth.....	49

b Terrific Broth (TB) media.....	49
c Amp/X-gal/IPTG plates.....	49
d M9 media.....	50
f 4 x additives.....	50
2.1.6 Competent cells.....	51
2.1.7 Buffers.....	51
2.1.8 Constructs and vectors.....	52
2.1.9 Vivaspin concentrators.....	54
2.1.10 Proteases and protease inhibitors.....	54
2.1.11 Inducers.....	54
2.1.12 Detergents.....	55
2.1.13 Chromatography.....	55
2.1.14 ThermoFluor assay.....	56
2.1.15 Circular Dichroism.....	56
2.1.16 Protein crystallization.....	56
2.1.17 NMR spectra.....	56
2.1.18 TEM.....	57
2.2 Methods for molecular biology.....	57
2.2.1 PCR.....	57
a Primer design.....	57
b PCR reaction and programs.....	61
c DNA analysis and purification.....	62
2.2.2 Restriction enzyme digestion.....	62
2.2.3 Dephosphorylation.....	63
2.2.4 Sticky-end ligation reactions.....	64
2.2.5 Blunt-end ligation.....	64
2.2.6 In-Fusion cloning.....	65
2.2.7 Construction of pET-14b-OutL ¹⁻²⁵⁷ -OutE ¹⁻⁵¹³ -OutF ¹⁻¹⁷² (or OutF ⁶⁵⁻¹⁷²).....	66
2.2.8 Transformation.....	68
2.2.9 Positive colony confirmation and sequencing.....	68
2.2.10 Making a recombinant construct.....	68
2.3 Protein purification techniques.....	69
2.3.1 Protein over-expression.....	69
2.3.2 Bacterial lysis.....	70
2.3.3 Membrane-fraction preparation.....	71
2.3.4 His tag purification.....	71
2.3.5 Strep-tag purification.....	72

2.3.6 GST tag purification.....	73
2.3.7 Gel filtration/size exclusion chromatography	73
2.3.8 Differential centrifugation.....	75
2.3.9 SDS-PAGE.....	77
2.3.10 Native gel	78
2.3.11 Western blot	78
2.3.12 Determination of protein concentration	79
2.4 Biochemical and biophysical techniques	80
2.4.1 Dynamic light scattering	80
2.4.2 Circular dichroism.....	81
2.4.3 Thermofluor assay.....	83
2.4.4 Full length of OutE, cytoplasmic domain of OutL and cytoplasmic domain of OutF pull down assay.....	84
2.4.5 NMR.....	85
a NMR principle.....	85
b NMR sample preparation.....	86
2.4.6 Peptide mass fingerprinting	87
2.4.7 Crystallization	87
a Crystallography theory.....	87
b X-ray crystallography	88
c Molecular Replacement	90
d Crystal vitrification	92
e Data collection and processing	92
f Synchrotron sources	93
g Structure validation.....	93
2.4.8 Single particle analysis using transmission electron microscopy	94
a Principle of single particle analysis	94
b Programs used for single particle analysis.....	97
c 2% Uranyl acetate negative staining preparation.....	98
d Sample preparation	98
e Nanogold labelling	99
Chapter 3 Structural studies of <i>D. dadantii</i> OutF	100
3.1 Overview	100
3.2 Results	101
3.2.1 Bioinformatics.....	101
a Signal peptide prediction	101

b Prediction of transmembrane regions of OutF.....	102
c Prediction of intrinsically disordered regions in OutF.....	103
3.2.2 Cloning <i>outF</i> , <i>outL</i> and <i>outE</i>	104
a Primer design	104
b PCR products of <i>outF</i> , <i>outL</i> and <i>outE</i>	104
c Construction of pET-24d-OutF	106
d Construction of pET-24d-OutF ⁵³⁻¹⁶⁸	106
e Construction of pET-14b-OutF ⁶⁵⁻¹⁷²	107
f Construction of pET-14b-OutF ²⁶⁷⁻³⁷⁴	108
g Construction of pOPINS3C-OutF ¹⁻⁴⁰⁸	108
h Construction of pET-14b-OutL ¹⁻²⁵⁷ -OutE ¹⁻⁵¹³ -OutF ¹⁻¹⁷² (or OutF ⁶⁵⁻¹⁷²)	108
3.2.3 Expression and purification.....	109
a Expression and purification of OutF ¹⁻⁴⁰⁸	109
b Expression and purification of SUMO-OutF ¹⁻⁴⁰⁸	110
c Expression and purification of OutF ⁵³⁻¹⁶⁸	112
d Expression and purification of OutF ⁶⁵⁻¹⁷²	114
3.2.4 Pull down assay of pET-14b-OutL ¹⁻²⁵⁷ -OutE ¹⁻⁵¹³ - OutF ⁶⁵⁻¹⁷²	115
3.2.5 Circular Dichroism (CD) Spectroscopy	116
3.2.6 Dynamic Light Scattering (DLS)	117
3.2.7 Crystallization trials	119
a Crystallization trial using SUMO-OutF ¹⁻⁴⁰⁸	119
b Crystallization of OutF ⁶⁵⁻¹⁷²	120
c Crystal structure determination.....	122
3.2.8 The crystal structure of OutF ⁶⁵⁻¹⁷²	125
a Sequence and structure comparisons between members of the GspF family	126
b The nature of the OutF ⁶⁵⁻¹⁷² surface	133
c Potential dimerization factors.....	135
d Predicted model of OutF cytoplasmic domain II.....	138
3.2.9 Interaction of OutE, OutF and OutL assessed using NMR.....	140
a Sample preparation for NMR spectroscopy.....	141
b Thermoflour assay to determine sample conditions for NMR studies.....	142
c NMR spectroscopy	144
3.2.10 Study of the interaction of OutF ⁶⁵⁻¹⁷² and OutE ¹⁻⁵¹³ -OutL ¹⁻²⁵⁷ by thermoflour assay	148
3.3 Summary and discussion.....	150

Chapter 4 Structural studies of the inner-membrane platform of the *D. dadantii* T2SS

152

4.1 Overview	152
4.2 Results	153
4.2.1 Cloning the genes corresponding to the inner-membrane complex.....	153
a Constructs	153
b PCR products.....	154
c Constructions of pASK3c: T2SS _{E-M} , pASK3c: T2SS _{C-M} , pASK3c: T2SS _{E, F, L and M} and pASK3c: T2SS _{F-M}	156
d Enzyme cleavage to analyze the new constructs.....	156
e Expression of pASK3c: T2SS _{E-M} and pASK3c: T2SS _{O-M}	157
4.2.2 Purification of the expressed inner-membrane complexes	159
a Membrane fraction preparation.....	159
b Purification of membrane protein complex.....	159
c Sucrose gradient centrifugation of inner-membrane complex OutE-M	163
d Dynamic light scattering (DLS)	164
e Mass Spectrometry	165
4.2.3 Initial three-dimensional model of inner-membrane platform.....	167
a Sample preparation.....	167
b Data collection	167
c Data processing programs.....	168
d Particle selection and normalisation.....	168
e CTF estimation	168
f Getting templates for auto-picking.....	169
g Auto-picking and particles sorting	169
h 2D classification to remove bad particles	169
i Generating template for 3D classification using SIMPLE	171
j 3D model refinement.....	174
4.2.4 Domain localisation in the T2SS inner-membrane platform	178
a Gold labelling.....	178
b Pilus-deleted and OutE-deleted mutations	180
c ATPase localization in the inner-membrane platform	180
4.3 Conclusion	182
Chapter 5 Concluding comments.....	184
5.1 A status report on the type II secretion system	184

5.2 Summary of achievements in this thesis	185
5.3 Future work	187

List of figures

Figure 1.1 Secretion systems in Gram-negative bacteria (from Erin R, 2016).....	21
Figure 1.2 Key components in T4P, T2SS, archaeal flagella and the competence systems in Gram-positive bacteria	25
Figure 1.3 Structural model of the <i>V. cholerae</i> Eps T2SS in its resting state.....	28
Figure 1.4 Pseudopilus biogenesis steps.	30
Figure 1.5 The structures of the pseudopilins.	32
Figure 1.6 Structure and function of GspD.....	36
Figure 1.7 The <i>Dickeya dadantii</i> secretion-pilotin complex.....	37
Figure 1.8 The inner-membrane platform.....	38
Figure 1.9 Crystal structures of EpsF ⁵⁶⁻¹⁷¹ , PilC ⁵³⁻¹⁶⁸ and TcpE ¹⁻¹⁰⁴	41
Figure 1.10 Structure of GspC and GspC/GspD complex.	45
Figure 2.1 Construction of pET-14b-OutL ¹⁻²⁵⁷ -OutE ¹⁻⁵¹³ using the link and lock method.	67
Figure 2.2 A summary of steps for making a recombinant construct	69
Figure 2.3 Summaries the step for over-expression of proteins	70
Figure 2.4 Columns involved in purifying proteins with different affinity tags.....	74
Figure 2.5 Cell fractionation by differential centrifugation	75
Figure 2.6 Far-UV CD reference spectra	82
Figure 2.7 Crystallization phase diagram.....	88
Figure 2.8 Satisfaction of Bragg's law to obtain diffraction.....	89
Figure 2.9 Molecular replacement method	91
Figure 2.10 Example of a 3D reconstruction from 2D projections.....	97
Figure 3.1 The predicted signal peptide of OutF.	101
Figure 3.2 Predicted transmembrane regions of <i>D. dadantii</i> OutF.....	102
Figure 3.3 The predicted] intrinsically disordered regions of OutF.....	103
Figure 3.4 Amplification of <i>outF</i> full-length and fragments.	105
Figure 3.5 Diagnostic cleavage of pET-24d-OutF ¹⁻⁴⁰⁸	106
Figure 3.6 Diagnostic cleavage of pET-24d-OutF ⁵³⁻¹⁶⁸	107
Figure 3.7 Colony PCR and enzyme digest of pET-14b-OutF ⁶⁵⁻¹⁷²	108
Figure 3.8 Colony PCR and enzyme digestion of pET-14b-OutL ¹⁻²⁵⁷ -OutE ¹⁻⁵¹³ - OutF ⁶⁵⁻¹⁷²	109

Figure 3.9 Expression test of OutF ¹⁻⁴⁰⁸	110
Figure 3.10 Purification of SUMO-OutF ¹⁻⁴⁰⁸	112
Figure 3.11 Expression and purification of OutF ⁵³⁻¹⁶⁸	113
Figure 3.12 Purification of OutF ⁶⁵⁻¹⁷²	114
Figure 3.13 Pull down assay of pET-14b-OutL ¹⁻²⁵⁷ -OutE ¹⁻⁵¹³ - OutF ⁶⁵⁻¹⁷²	115
Figure 3.14 CD spectra of the first cytoplasmic domain, OutF ⁶⁵⁻¹⁷² , and full-length OutF, OutF ¹⁻⁴⁰⁸	116
Figure 3.15 DLS result for OutF ⁶⁵⁻¹⁷² and SUMO-OutF ¹⁻⁴⁰⁸	118
Figure 3.16 Crystal hits of Sumo-OutF ¹⁻⁴⁰⁸	119
Figure 3.17 Crystal hits of OutF ⁶⁵⁻¹⁷² in 96-well screening plate after 48 hours and 96 hours.....	121
Figure 3.18 Crystal hits from optimized conditions	122
Figure 3.19 Ramachandran plot statistics for the OutF ⁶⁵⁻¹⁷² crystal structure (PDB: 5NBG).....	125
Figure 3.20 Comparison of predicted and solved structure of cytoplasmic domain I of OutF	126
Figure 3.21 The predicted secondary structure of full-length of OutF and sequence alignment of GspF family members.....	129
Figure 3.22 Conserved residues in the OutF ⁶⁵⁻¹⁷² structure.	130
Figure 3.23 Hydrophobicity mapped on the surface of the OutF ⁶⁵⁻¹⁷²	133
Figure 3.24 Electrostatic potential mapped onto the OutF ⁶⁵⁻¹⁷² structure.	134
Figure 3.25 Interactions at the dimer interface of OutF ⁶⁵⁻¹⁷²	136
Figure 3.26 The predicted model of cytoplasmic domain II of OutF	138
Figure 3.27 Two models of OutF ²⁴⁵⁻³⁷⁹ dimer.....	139
Figure 3.28 Size exclusion chromatography of ¹⁵ N-labelled OutF ¹⁻¹⁷² , OutF ⁶⁵⁻¹⁷²	142
Figure 3.29 Thermofluor plots	143
Figure 3.30 Assessment of interactions between cytoplasmic domain I of OutF and OutE-OutL ¹⁻²⁵⁷	145
Figure 3.31 Elucidation of OutF ⁶⁵⁻¹⁷² and OutE or OutL ¹⁻²⁵⁷ or OutE-OutL ¹⁻²⁵⁷ complex interactions	148
Figure 3.32 Thermofluor plots of interaction between OutF ⁶⁵⁻¹⁷² and OutE-OutL ¹⁻²⁵⁷ complex.....	149
Figure 4.1 Flow chart of single particle reconstruction process in Relion2(Fernandez-Leiro and Scheres 2017).....	153

Figure 4.2 Genetic organization of T2SSs from <i>E. coli</i> IHE 3034 and <i>D. dadantii</i>	154
Figure 4.3 PCR products of T2SS gene operon from <i>D. dadantii</i> and IHE strain.....	155
Figure 4.4 Restriction enzyme results	157
Figure 4.5 Expression test of the inner-membrane platform from <i>D. dadantii</i> and the whole complex of the T2SS from IHE3034, respectively.	159
Figure 4.6 Purification of T2SS complexes and gene mutated complexes from <i>D. dadantii</i> and IHE using strep tactin resin.....	161
Figure 4.7 Chromatographic separation on Superose 6 10/300 GL column of T2SS _{E-M} complex	162
Figure 4.8 Analysis of sucrose gradient fraction	164
Figure 4.9 T2SS _{OutE-M} DLS summary.	165
Figure 4.10 Mass spec result of T2SS _{E-M} complex	166
Figure 4.11 plots of distribution and resolution of 2D classes.....	170
Figure 4.12 Data process to generate the initial model for 3D classification	173
Figure 4.13 3D classification	173
Figure 4.14 Refined 3D model of OutE-OutM complex	175
Figure 4.15 EM analysis of OutE-OutM.....	178
Figure 4.16 T2SS _{E-M/M-His} gold labelling.....	179
Figure 4.17 Domain localisation in the type II secretion system inner-membrane platform.....	181

List of tables

Table 2.1 Components of LB broth	49
Table 2.2 Components required for making the TB media.....	49
Table 2.3 Components used for making Amp/X-gal/IPTG plates.....	50
Table 2.4 Components used for making 200ml 100 x trace element stock	50
Table 2.5 Components used for making 1 Litre 10 x M9 Stock.....	50
Table 2.6 Components used for making 200ml 4 x additives.....	51
Table 2.7 Competent cells used for expression and cloning	51
Table 2.8 Buffers for protein purification and biochemistry	52
Table 2.9 Constructs kindly provided by Dr. Vladimir Shevchik (University of Lyon)	53
Table 2.10 List of vectors used in Chapter 3 and 4.....	53
Table 2.11 An example of 15 bases of homology in primers for infusion cloning	57
Table 2.12 List of the primers used in Chapter3 to amplify <i>outF</i> , <i>outL</i> and <i>outE</i> genes and parts of these genes.....	58
Table 2.13 List of DNA primers used in Chapter 4	59
Table 2.14 PCR reaction	61
Table 2.15 PCR program using Q5 High-Fidelity DNA Polymerase.....	61
Table 2.16 Typical restriction enzyme digest for PCR products and vectors.....	63
Table 2.17 20µl reaction for dephosphorylation	64
Table 2.18 Typical sticky-end ligation reaction in 10µl volume.	64
Table 2.19 Blunt ligation reaction in 10 µl volume	65
Table 2.20 In-Fusion cloning reaction	66
Table 2.21 Recipe for SDS-PAGE.....	77
Table 2.22 The buffer conditions screened in thermoflour assay	84
Table 3.1 Optimization of the crystal hits condition.....	121
Table 3.2 Crystallographic data and refinement statistics for OutF ⁶⁵⁻¹⁷²	124
Table 3.3 Nomenclature of GspF for related filament systems in Gram-negative, Gram- positive bacteria and archaeal flagella.	127
Table 3.4 Structure comparison of GspF family from T2SS, T4P in Gram-negative bacteria, T4P, Tad, Com in Gram-positive bacteria and archaeal flagella.	131

Abbreviations

T2SS	Type II secretion system
Sec	general secretion route
Tat	twin-arginine translocation pathway
GSP	general secretory pathway
ETEC	enterotoxigenic <i>Escherichia coli</i>
EHEC	enterohaemorrhagic <i>Escherichia coli</i>
HUS	haemolytic uraemia syndrome
Com	competence systems
TAD	tight adherence
TMS	transmembrane helices
PG	peptidoglycan
SRP	signal recognition particle
Cryo-EM	Cryo-electron microscopy
HR	the homology region
$\mu\text{l}/\mu\text{M}$	micro-litre/micro-molar
Å	Angstrom
EDTA	Ethylene diamine tetraacetic acid
TAE	Tris-acetate-EDTA

PCR	polymerase chain reaction
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB	Luria-Bertani
PMSF	phenylmethylsulfonyl fluoride
CD	circular dichroism
NMR	Nuclear Magnetic resonance
AHT	Anhydrotetracyclin
OD	Optical density
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
CV	column volumes
DDM	<i>n</i> -Dodecyl β -D-maltoside
LDAO	<i>N, N</i> -dimethyl-1-dodecanamine- <i>N</i> -oxide
DTT	Dithiothreitol
PBS	phosphate buffered saline
V _e	Elution volume
V _o	Void volume
V _t	total bed volume
TEMED	Tetramethylethylenediamine
TBS	Tris-buffered saline
AP	alkaline phosphate

DLS	Dynamic light scattering
Pd	polydispersity
qPCR	Real-time polymerase chain reaction
TCEP	tris(2-carboxyethyl) phosphine
HSQC	heteronuclear single quantum correlation
CCD	charge-coupled device
MR	Molecular Replacement
PG	propylene glycol
EG	Ethylene glycol
ESRF	European Synchrotron Radiation Facility
SPA	Single particle analysis
TEM	Transmission electron microscopy
CTF	contrast transfer function
FSC	Fourier shell correlation
SUMO	Small Ubiquitin-like Modifier
SEC	size exclusion column
IMAC	Immobilized metal ion affinity chromatography
XDS	X-ray Detector Software
CCP4	Collaborative Computational Project Number 4
COOT	Crystallographic Object-Oriented Toolkit

RMS	root-mean-square
TD-PCR	Touch-down polymerase chain reaction
PVDF	polyvinylidene difluoride
IHE3034	Extra-intestinal pathogenic, O18:K1:H7
DM-NPG	Decyl Maltose Neopentyl Glycol
MALDI-TOF spectrometry	matrix assisted laser desorption ionization-time of flight mass
EPU	<i>E Pluribus Unum</i>
SIMPLE	Single-particle Image Processing Linux Engine
RELION	REGularised LIkelihood Optimisation

Chapter 1 Introduction

1.1 Secretion in bacteria

Secretion in bacteria is the transport or translocation of effector molecules including proteins, enzymes or toxins from the bacterial cytoplasm into the environment or host. Secretion is therefore a very important mechanism for bacterial growth, adaptation, and survival in their natural surroundings.

1.1.1 Gram-negative bacteria secretion system

Gram-negative bacteria have two membranes, inner- (also known as cytoplasmic) and outer-membrane which makes secretion more complex than in Gram-positive bacteria which have only a single, cytoplasmic, membrane. In Gram-negative bacteria the space between the inner- and outer- membrane is called the periplasmic space. Transport out of the cells must proceed either directly from the cytosol through a single step involving a tunnel through the periplasm (as in the T1SS, T3SS, T4SS and T6SS; Fig. 1.1), or using a two-step process with a periplasmic intermediate. In this latter process, secreted proteins are first translocated into the periplasm through Sec (secretory) or Tat (twin-arginine translocation) secretion systems and then transported across the outer-membrane by a second transport system (Green and Mecsas 2016). Protein effectors transported through type III secretion system (T3SS) are secreted mainly as unstructured proteins using a one-step process and then the effectors fold into their native conformation post transport in the host cell (Demers et al. 2014). On the other hand, substrates of the type II secretion system (T2SS) are transported first to the periplasm via Sec or TAT systems and are fully folded in the periplasm before being secreted across the outer-membrane (Merritt et al. 1994, Voulhoux et al. 2001, Fries et al. 2007, Richardson et al. 2012)

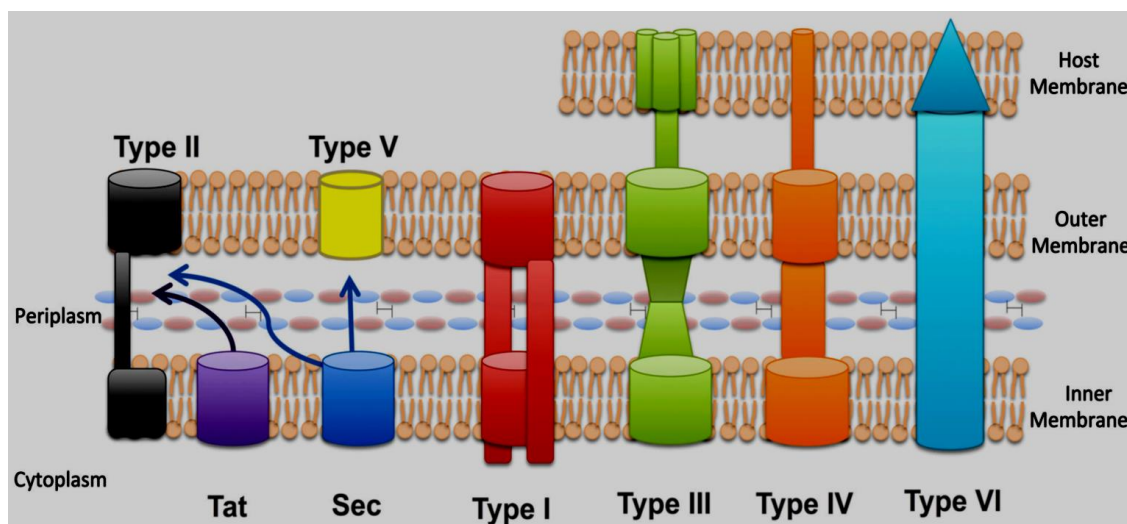


Figure 1.1 Secretion systems in Gram-negative bacteria (from Erin R, 2016).

Secretion systems types I, III, IV, VI secrete direct from cytosol across the two membranes of the bacterial cell. In fact, types III, IV and VI secrete also across the host membrane. The Type II and V systems secrete from the periplasmic space, between the inner- and outer- membranes into the environment.

1.2 Type II secretion system

The T2SS was initially considered as the main terminal branch of the general secretory pathway (Gsp), where unstructured substrates or native folded substrates first enter the periplasm via the SEC apparatus (Natale et al. 2008) or TAT pathway (Voulhoux, Ball et al. 2001) and are then secreted by the T2SS. Nonetheless, the term Gsp is often used to refer the ~15 protein components of the T2SS from GspA to GspO. The T2SS can secrete a variety of substrates out of the bacteria including biologically functional proteins like proteases, lipases and phosphatases (Korotkov et al. 2012).

1.2.1 The discovery of T2SS

The type II secretion system (T2SS) was discovered in the 1980s in the genus *Klebsiella* (d'Enfert et al. 1987) as the system responsible for the secretion of pullulanase A. Expression of the pullulanase secretion genes is required for the pullulanase to be

translocated across the outer-membrane. Cell surface localization of pullulanase is just a transient stage in the pullulanase secretion pathway, which eventually leads to the extracellular release of polypeptides (Pugsley et al. 1986).

1.2.2 T2SS related disease

Many bacterial pathogens use T2SSs to export virulence factors outside of the cells causing disease. Human pathogens include: *Vibrio cholerae* (secretes cholera toxin causing diarrhoea) (Sandkvist et al. 1997), enterotoxigenic *Escherichia coli* (ETEC) , secretes two types of enterotoxins heat-labile and heat-stable enterotoxin causing watery cholera-like diarrhoea in animals and humans) (Tauschek et al. 2002) and enterohaemorrhagic *Escherichia coli* (EHEC which causes diarrhoea or haemorrhagic colitis and even lethal haemolytic uraemia syndrome (HUS))(Lathem et al. 2002), *Pseudomonas aeruginosa* which secretes exotoxins causing death in models of pulmonary infection (Bally et al. 1992, Jyot et al. 2011), *Klebsiella pneumoniae* which secretes pullulanase causing pneumonia in the form of bronchopneumonia and bronchitis (D'Enfert and Pugsley 1989) and *Legionella pneumophila* secretes acid phosphatases, lipases, phospholipases C (Rossier et al. 2004). The plant pathogen *Dickeya dadantii*, which we will discuss in this Thesis, secretes plant cell-wall degrading enzymes including pectinases which cause crop spoilage and threaten food security (Toth and Birch 2005).

1.2.3 The function of T2SS

a Assembly of surface organelles

Like the type 4 pilus assembly system (T4PS), the T2SS can promote assembly of filaments via ATP hydrolysis. These filaments are localized in the periplasm and have been coined pseudo-pili which perform a range of functions in signalling, adhesion and

motility. Pseudopilus components and the assembly machinery are essential for exoprotein transport across the outer-membrane.

b Protein transport

A major function of the T2SS is nutrient acquisition. Most exoproteins identified so far are hydrolytic enzymes that degrade biopolymers (carbohydrates, lipids, proteins or nucleic acids). Additionally, the T2SS secretes toxins, slime proteins, adhesion proteins and cytochromes which affect respiration, motility or lead to biofilm formation facilitating other bacterial lifestyles and protecting the bacteria in the case of biofilm production.

The majority of secreted enzymes are released into the surroundings to generate small nutrient pools available for uptake. The *Klebsiella oxytoca* pullulanase, a lipoprotein which degrades branched maltotriose polymers is a rare example of an exoprotein which remains surface-associated after secretion (Pugsley and Kornacker 1991).

1.2.4 The secretion model

Compared to several other secretion systems which transport unfolded proteins, the T2SS recruits and transports fully folded proteins. Thus, it is suggested that the secretion depends on a conformational signal which could be a patchwork of structural signals on the substrate surface but conserved in various substrates (Palomaki et al. 2002, Korotkov, Sandkvist et al. 2012). Recent mutagenesis, cross-linking and functional studies in *D. dadantii* showed that a 9-residues loop of pectate lyase PelI acts as a secretion signal which could interact with GspC and GspD directly (Pineau et al. 2014). Recent structure and functional study on PulA substrate of *K. oxytoca* has also revealed that several structurally flexible regions of this large protein are important for its secretion (East et al.

2016). Therefore, in some systems, a few surface exposed and highly flexible regions of secreted substrates act as secretion signals via their transient conformation on T2SS components.

The recruitment of substrates into the periplasm is necessary to allow them to interact with T2SS components. Even though the periplasmic domain of GspD has been observed to interact with substrates in *D. dadantii*, *V. cholera* and *P. aeruginosa*, it is still not clear if these interactions represent periplasmic recruitment. It is generally acknowledged that recognition of secretion substrates by the periplasmic domain of GspC is important for their transport into the T2SS (Bouley et al. 2001, Gerard-Vincent et al. 2002). Once inside the secretion system, the secretion substrates are thought to lie on the top of the pseudopilus and as the pseudopilus assembles and grows it forces the substrate through the pore in the secretin and out of the cells. The energy for pilus assembly is provided by the ATP hydrolase GspE (Campos et al. 2013).

1.3 The relationship between T2SS and other systems in bacteria and archaea

The type II secretion system is closely related to several other systems in bacteria and archaea. The Type IV pili in Gram-negative bacteria (Korotkov et al. 2011), the competence systems (Com) and Tight adherence (Tad) in Gram-positive bacteria and archaeal flagella all have structural homologs of proteins found in the T2SS. The common feature is the ability to promote assembly of helical filaments composed of inner membrane-embedded pilin subunits. The common components in T2SS, T4P, archaeal flagella and the competence systems include (1) cytoplasmic ATPase, (2) the inner-membrane protein with multiple transmembrane helices, (3) the pilin or pseudopilin or

flagellin, and (4) the specific membrane protease (Fig. 1.2). However, there is no secretin in archaeal flagella as expected as archaea lack an outer-membrane. Interestingly, the archaeal flagella has more similarities with the T4P and T2SS than with the bacterial flagella (Hobbs and Mattick 1993, Peabody et al. 2003).

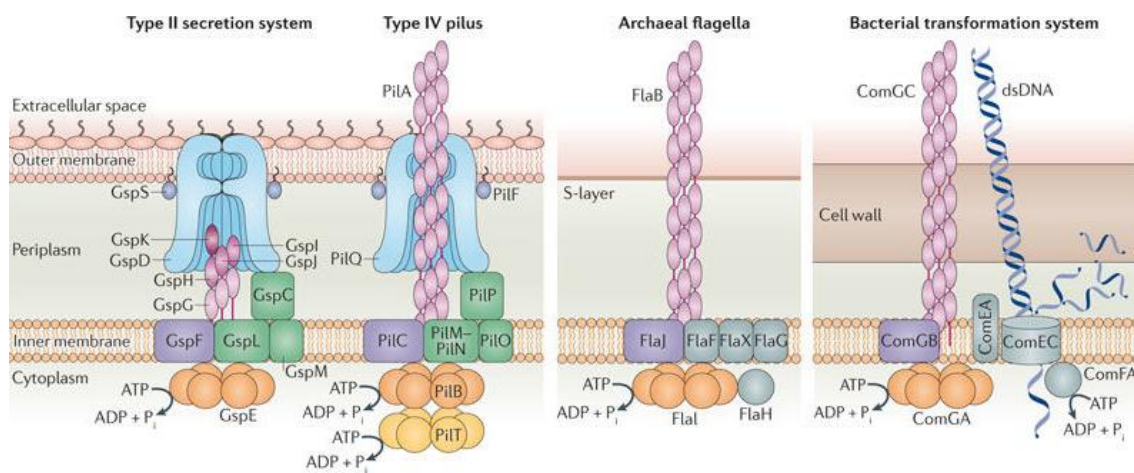


Figure 1.2 Key components in T4P, T2SS, archaeal flagella and the competence systems in Gram-positive bacteria

ATPase is in orange, the inner-membrane proteins with a single transmembrane helix are in green, the inner-membrane proteins with multiple transmembrane helices are in purple, the outer-membrane proteins are in blue, and the pilins, pseudopilins and flagellins are in pink (Korotkov, Sandkvist et al. 2012).

Although the protein components of the T2SS, T4P and archaeal flagella are widely distributed between the bacterial and archaeal domains, no horizontal transfer has occurred between prokaryotes and eukaryotes (Peabody, Chung et al. 2003). It has been revealed that structurally archaeal flagella are more similar to bacterial type IV pili rather than to bacterial flagella (Ghosh and Albers 2011). There is an evidence that the components of these systems evolved from a single primordial precursor system (Nguyen et al. 2000, Cao and Saier 2001, Peabody, Chung et al. 2003). Among these components, the ATPase and multi-spanning TM proteins are the largest and most slowly diverging

proteins although these components have undergone sequence divergence at different rates. The number of transmembrane α -helical hydrophobic segments (TMSs) arose by the intergenic duplication (Kuan and Saier 1993, Saier 2003). A good example for this is that *E. coli* CorA protein has three TMSs with an N-terminal soluble periplasmic domain and C-terminal cytoplasmic domain (Saier 2003). The two TMS homologues have both domains in the cytoplasm, the situation of which is quite like GspF we study here.

The first indication that Gram-positive bacteria possess Type IV pili (TFP-like) structures was provided by the Com system (Competent system) in *Bacillus subtilis*, in which the prepilin peptidase Pild homolog ComC cleaved the signal peptides producing a high molecular weight DNA-binding surface structure (Chung and Dubnau 1995, Chen and Dubnau 2004). The T4P in Gram-positive has been shown to be critical for biological processes *Clostridium perfringens pil* operon required for twitching motility and biofilm formation (Varga et al. 2006, Rodgers et al. 2011). *Bacillus anthracis* operon contains genes with some homology to Tad genes (tight adherence) which could be responsible for anthrax toxin secretion (Grynberg et al. 2007).

1.4 The structure of T2SS

There are 12 and 15 genes designated gsp (General Secretory Pathway) (Lindeberg and Collmer 1992) encoding T2S machines. This complex, two-membrane spanning machine comprises the following four sub-assemblies: (i) a secretion ATPase (GspE) providing the energy for secretion on the cytoplasmic side of the inner-membrane, which is closely associated with (ii) a sophisticated Inner-Membrane Platform (GspC, GspF, GspL, GspM), which senses the presence of a protein to be secreted and assembles (iii) the pseudopilus (GspG, GspH, GspI, GspJ and GspK) in the space between the membranes, whose tip pushes secreted proteins through and the prepilin peptidase, GspO, which is

responsible for processing of pseudo-pilin subunits. (iv) the large gated pore of the Outer-Membrane Complex (GspD) (Fig. 1.3). The core genes are in one large operon *gspCDEFGHIJKLMNO*. Several other genes encode factors involved in the localization and assembly of the out membrane channel-forming protein GspD (also called the secretin (Korotkov, Gonen et al. 2011)), including *gspS* (pilotins) and *gspAB* or *gspB*. In *V. cholera* and *D. dadantii*, T2S systems express pilotins, small lipoproteins each with a unique structure, which enhance the kinetics of secretin targeting and assembly in the outer-membrane (Daefler et al. 1997, Shevchik and Condemine 1998, Nickerson et al. 2011, Strozen et al. 2012). In *P. aeruginosa* Hxc and *X. campestris* Xps T2SSs, the secretin contains its own lipid anchor and they can pilot and anchor themselves (Viarre et al. 2009). A crystal structure is already available for *P. aeruginosa* protein PA3611, which shares similar structure with the *V. cholerae* pilotin, AspS, although its function as a pilotin has still to be confirmed experimentally (Seo et al. 2009).

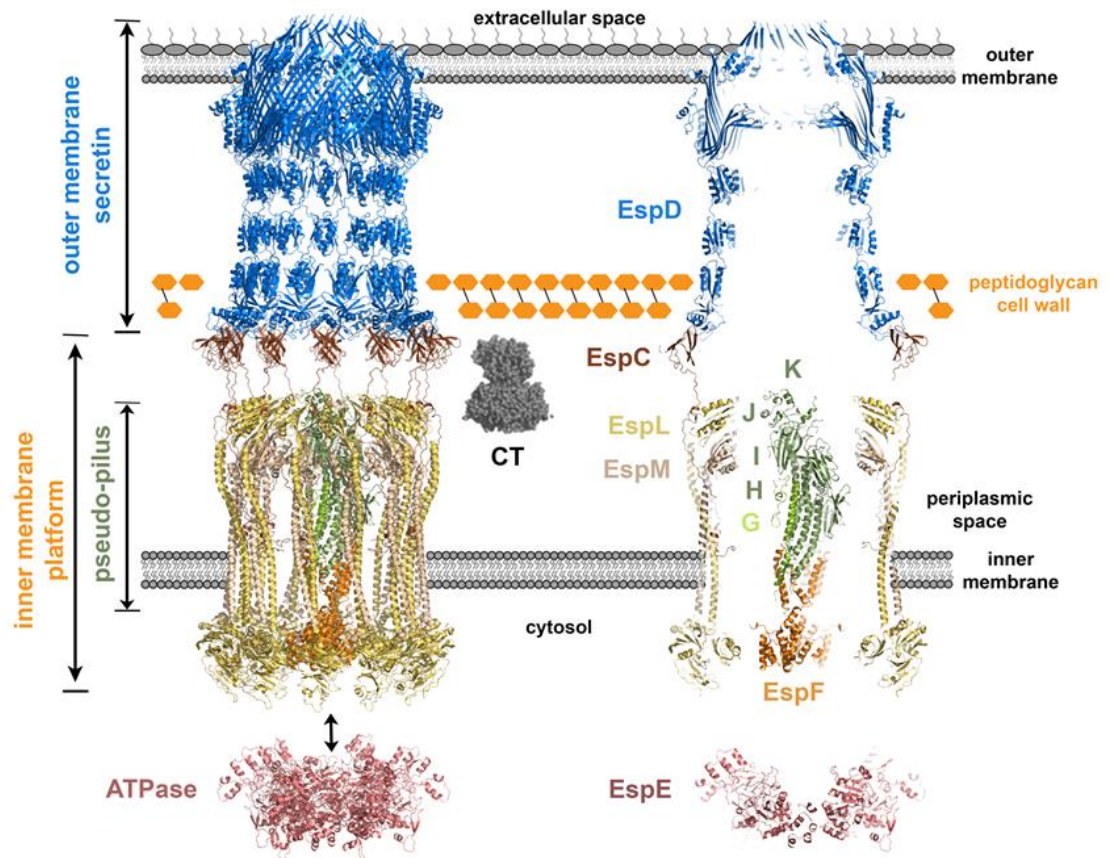


Figure 1.3 Structural model of the *V. cholerae* Eps T2SS in its resting state.

All T2SS structures are represented as cartoons and either from *V. cholera* (Peri-EpsM PDB Re. 1UV7, EpsF Re. 3C1Q, major pseudopilin Re. 3FU1, Cyto-EpsI Re. 1YF5, cytoplasmic domain of EpsE and EpsL 2BH1, EpsI and EpsJ 2RET, EpsH 1QV8, GspD 5WQ8) or modelled based on homologous structures (EpsF PDB Re. 3C1Q, ETEC N-terminal domain of GspD 3EZJ, *Thermus Thermophilus* N-terminal domain of PilC 2WHN, GspE hexamer 4KSS, GspE-cyto-GspL complex 4PHT) using the Phyre2. Structures were assembled using the EM model of the type Iva pilus system as a guide (Chang et al. 2016). The cholera toxin (CT) is shown as a grey surface (Gu et al. 2017) and to the same scale as the T2SS model (Zhang et al. 1995).

In *Aeromonas* and *Vibrio*, GspA and GspB span the inner-membrane once and form a multimeric complex together which is thought to modify or organise the peptidoglycan (PG) to allow assembly of secretin (Ast et al. 2002, Li and Howard 2010, Vanderlinde et al. 2014). The periplasmic domain of GspA interacts with peptidoglycan and forms a

complex with GspB. The GspAB complex is essential for type II secretion in *Aeromonas* but not in *Vibrio* (Strozen et al. 2011). However, in *D. dadantii*, only OutB is present. OutB of *D. dadantii* interacts with the cognate secretin OutD but its precise function remains unclear (Condemine and Shevchik 2000). Recent unpublished work from the Pickersgill group in collaboration with Shevchik, University of Lyon, has shown that in *D. dadantii* OutB serves to anchor the outer-membrane secretin to the inner-membrane a function ascribed to OutC in other systems. GspAB and possibly also GspB might attach the secretin complex to the peptidoglycan or help the secretin move across the peptidoglycan mesh (Li and Howard 2010, Strozen, Stanley et al. 2011, Vanderlinde, Zhong et al. 2014). Interestingly, although no additional genes are present in the *L. pneumophila* T2SS, its outer-membrane secretin LspD is predicted to contain a peptidoglycan binding domain at its N-terminus (Kelley et al. 2015).

1.4.1 Pseudopilus

The T2SS has been proposed to assemble a short filament termed “the pseudopilus” to push secreted protein substrates through the secretin channel by T4P-like extension and retraction dynamics acting which is a similar way to a piston (Douzi et al. 2011). T2SS pili are composed of one major subunit GspG and 4 minor pseudopilins GspH, GspI, GspJ and GspK (Johnson et al. 2006). The five pseudopilins have similar sequence: the N-terminal tail (several positively charged amino acid residues) followed by a Gly or Ala residue which is the cleavage site for the prepilin peptidase; a hydrophobic α -helix embedded into the inner-membrane and a C-terminal periplasmic globular domain (Korotkov, Sandkvist et al. 2012). GspG has been shown to use the SRP pathway to insert into the inner- membrane (Francetic et al. 2007). The minor pilins may also utilize this pathway for insertion due to the sequence similarity with GspG.

Such sequence conservation allows unambiguous identification of fibre assembly in bacteria and archaeal (Arts et al. 2007, Francetic, Buddelmeijer et al. 2007) and suggest that a strict selection pressure shaped these segments (Campos, Cisneros et al. 2013). Pseudopilus biogenesis steps include the following several steps: the transmembrane segment of pseudopilins interacts with cellular and type II secretion system partners via the signal recognition particle (SRP) complex to the Sec translocase; the positively charged pre-peptide of the pseudopilins are removed and then methylation of the new N-terminus by the prepilin peptidase after membrane insertion; the cleaved peptide interacts with at least one assembly factor in the membrane possibly the inner-membrane protein GspF/PilC; the TM segments contact with neighbouring pilins and become buried in the core of the fiber, while the globular pilin domains are exposed on the surface of pseudopilus (Campos, Cisneros et al. 2013), see Fig. 1.4.

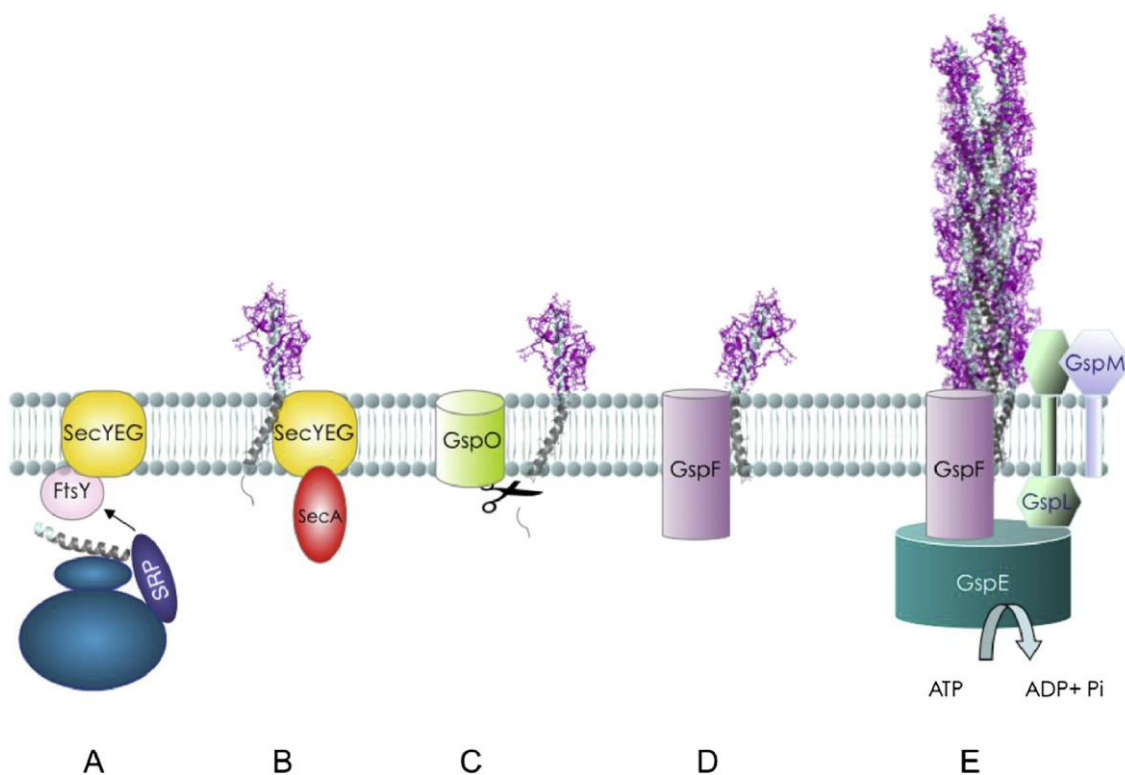


Figure 1.4 Pseudopilus biogenesis steps.

(A) The transmembrane segment of the pseudopilus (grey) target to the Sec translocase *via* the SRP complex; (B) Membrane insertion; (C) The prepilin peptidase cleaves the pre-peptide after the conserved Gly residue; (D) The cleaved peptide interacts with at least one component of the assembly inner-membrane platform; (E) The hydrophilic domains are exposed on the pseudopilus surface (Manuel, 2013), (Campos, Cisneros et al. 2013).

a Major pseudopilins

GspG which is restricted to the periplasm under physiological expression levels in liquid culture can build helical fibres (Kohler et al. 2004). Based on the predictions generated using structural modelling, the experimentally validated structure of T2SS pili revealed a right-handed helix protomer organization in the fibre (Campos et al. 2010). These hypothetical fibres could be visualized by overexpressing the *pul* genes in *Klebsiella oxytoca* (Sauvonnet et al. 2000), *Escherichia coli* (Vignon et al. 2003) and *P. aeruginosa* (Durand et al. 2003). Interestingly, the pseudopili were assembled only when the *xcpT* (a GspG homolog) was introduced in the whole *xcp* gene cluster deleted strain. It was concluded that assembly of the type II pseudopilus depended on the function of other Xcp components (Durand, Bernadac et al. 2003).

X-ray crystallography showed that the GspG has a long α -helical stem and a globular domain formed by three-strand β -sheet (Kohler, Schafer et al. 2004, Korotkov et al. 2009), see Fig. 1.5A. The C-terminal domain forms a stable loop folded around a calcium ion (Korotkov, Gray et al. 2009) and the residues binding the calcium atom are essential for secretion of the protein (Kohler, Schafer et al. 2004). In the Pul T2SS, the mutations of the calcium-binding sites could significantly reduce the level of the PulG (GspG) indicating the importance of calcium-ion in the major pseudopilin's structure and stability (Campos, Nilges et al. 2010).

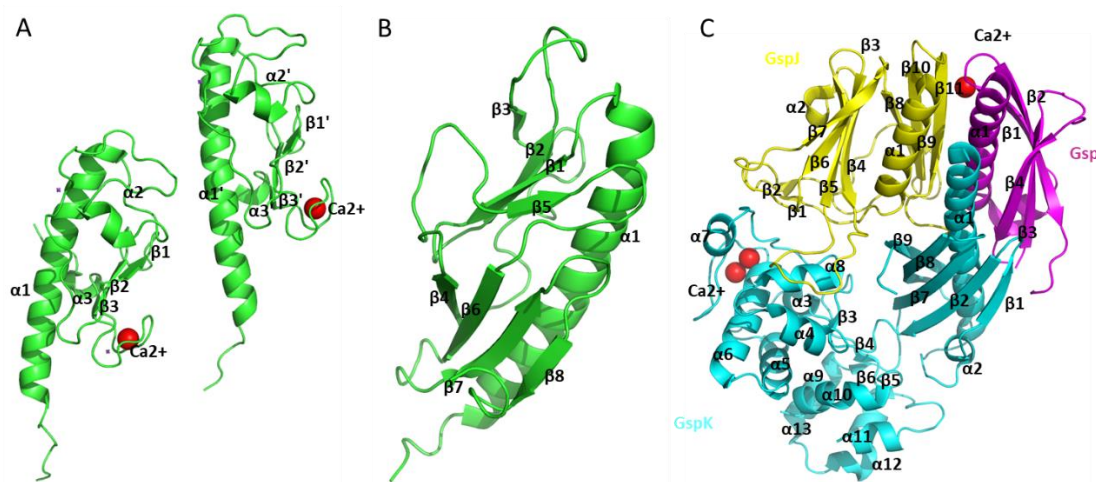


Figure 1.5 The structures of the pseudopilins.

A: Crystal structure of *Enterohaemorrhagic E. coli* GspG (PDB code: 3G20), calcium ions are shown as red balls; B: Solution structure of *E. coli* GspH (PDB code: 2KNQC); C: Crystal structure of *Enterotoxigenic E. coli* Gsp-K-J-I (PDB code: 3CI0). GspI is in magenta; GspJ yellow; GspK cyan and the calcium ions are shown as green balls.

b Minor pseudopilins

The studies of the *Xanthomonas campestris* T2SS show the presence of four minor pseudopilins under the physiological conditions (Kuo et al. 2005). The crystal structure of the GspJ–GspI–GspK periplasmic domain shows a complex with quasi-helical symmetry (Korotkov and Hol 2008), see Fig. 1.5C. GspI occupies a central place in the formation of the tip complex interacting with GspJ and GspK (Douzi et al. 2009). GspK caps this complex via its α -helical domain at the summit. Overexpression of GspK could reduce the pilus length while deletion mutant of *xcpX* (*GspK*) could assemble longer pili indicating that GspK could control the extension of the fibre (Vignon, Kohler et al. 2003, Durand et al. 2005). A recent study on minor pseudopilins by immunofluorescence microscopy revealed that *GspI* and *GspJ* mutants can result in fewer and longer pili (Cisneros et al. 2012). The fourth pseudopilin GspH binds to GspJ at the bottom of this trimer and forms a quaternary complex (Douzi, Durand et al. 2009). But GspH does not

play a role in the initiation or termination of the pseudopilus assembly since there are no changes of length of the pilus on deleting GspH (Cisneros, Bond et al. 2012). Furthermore, there is only a small secretion decrease when *GspH* mutants are overexpressed (Cisneros, Bond et al. 2012). Therefore GspH which binds this initiating tip complex *in vitro* via its globular domain (Douzi, Durand et al. 2009), may ensure a transition between initiation and ATPase-catalysed elongation which is required for efficient protein secretion (Cisneros, Bond et al. 2012). Once activated, GspG is then recruited and incorporated into the pilus which allows the pseudopili to extend. But, the quaternary complex of the minor pseudopilins does not bind to the periplasmic domain of GspG suggesting that the TM segment of the GspG is required for the elongation in T2SS (Douzi, Durand et al. 2009). The energy for assembly of the pilus has been established to be generated by conformational changes of ATPase GspE. A candidate to transduce the conformational change to promote pseudopilus assembly is the inner-membrane rotor, GspF (Nivaskumar et al. 2014). How the cytoplasmic ATPase controls the assembly of the pseudopilus remains to be elucidated in detail.

1.4.2 Inner-membrane platform

The inner-membrane assembly platform is a complex that contains: GspC, GspF, GspL and GspM. It is now established that the associated cytoplasmic ATPase, GspE, provides the energy for pseudopilus assembly and protein secretion (Patrick et al. 2011). In *Vibrio* and other T2SSs GspE has an extended N-terminal domain forming a complex with the cytoplasmic domain of GspL. The species-specific GspE–GspL interaction might be involved in energy transfer to inner-membrane assembly complex. The close genetic link between GspE and GspF suggests these proteins also may function as a complex. Yeast two-hybrid and pull-down studies suggest that the cytoplasmic domain I of GspF interacts with GspE and GspL (Py et al. 2001).

1.4.3 The secretin channel

The channel forming protein secretin GspD in T2SS is the central component of this system. It is typically associated with GspC (Tammam et al. 2013). In T2SS adjacent *gspCD* genes are always co-transcribed.

Analysis of *K. oxytoca* PulD and *V. cholera* GspD cryo-EM particles at low resolution reveals a dodecameric complex in the outer-membrane (Nouwen et al. 2000, Korotkov, Gonen et al. 2011), which is consistent with cryo-electron tomography and cryo-EM studies with *Myxococcus xanthus* and *Neisseria Meningitidis* in T4PS, respectively (Berry et al. 2012, Chang, Rettberg et al. 2016). Recent published structure for the *E. coli* K12 and *V. cholerae* GspD proteins at near atomic resolution display predominantly 15-fold symmetry (Yan et al. 2017) which is also observed in the *Salmonella* T3SS secretin (Worrall et al. 2016). While this difference may result from different recombinant expression strategies and sample preparations, the higher resolution of the latter structures reveals the pentadecameric structure of the secretin.

GspD has three regions: N-terminal periplasmic domains, the secretin domain and a short C-terminal S-domain (Fig. 1.6B). The four N-terminal domains termed N⁰ to N³ extend from the outer-membrane into the periplasm (Fig. 1.6B), they act to funnel substrates into the membrane pore. The N-domains penetrating the peptidoglycan communicate with the inner-membrane platform and also interact with substrates (Shevchik et al. 1997, Korotkov, Gonen et al. 2011, Pineau, Guschinskaya et al. 2014). The secretin domain is required to form the membrane embedded core of the channel while the S-domain stabilize the mature structure by decorating the adjacent subunit (Worrall, Hong et al. 2016, Yan, Yin et al. 2017).

The intrinsic disorder within GspD (Fig. 1.6A) has been highlighted by the cryo-EM structure of *E. coli* and *V. cholerae* GspD and secondary structure predictions (McGuffin et al. 2000, Ward et al. 2004). The first disordered region is localized to about 10-residue linker between N⁰ and N¹(Gu, Shevchik et al. 2017). The N⁰ domain is disordered and cannot be modelled in the absence of the inner-membrane platform to dock with in the structures of *E. coli* and *V. cholerae* GspD (Yan, Yin et al. 2017). Linkers connecting the N¹-N² and N²-N³ also show some flexibility between rings of N-domains (Sandkvist et al. 1999).

The second disordered region of GspD is located within the homologous loops of the N² and N³ domains, respectively. This loop in the N¹ domain is only several residues and well-ordered, while in N² there are up to 15 residues and disordered in all available structures (Korotkov et al. 2009, Van der Meeren et al. 2013, Yan, Yin et al. 2017). Moreover, this loop in the N³ contains up to 80 residues and revealed to be a flexible weak constriction site.

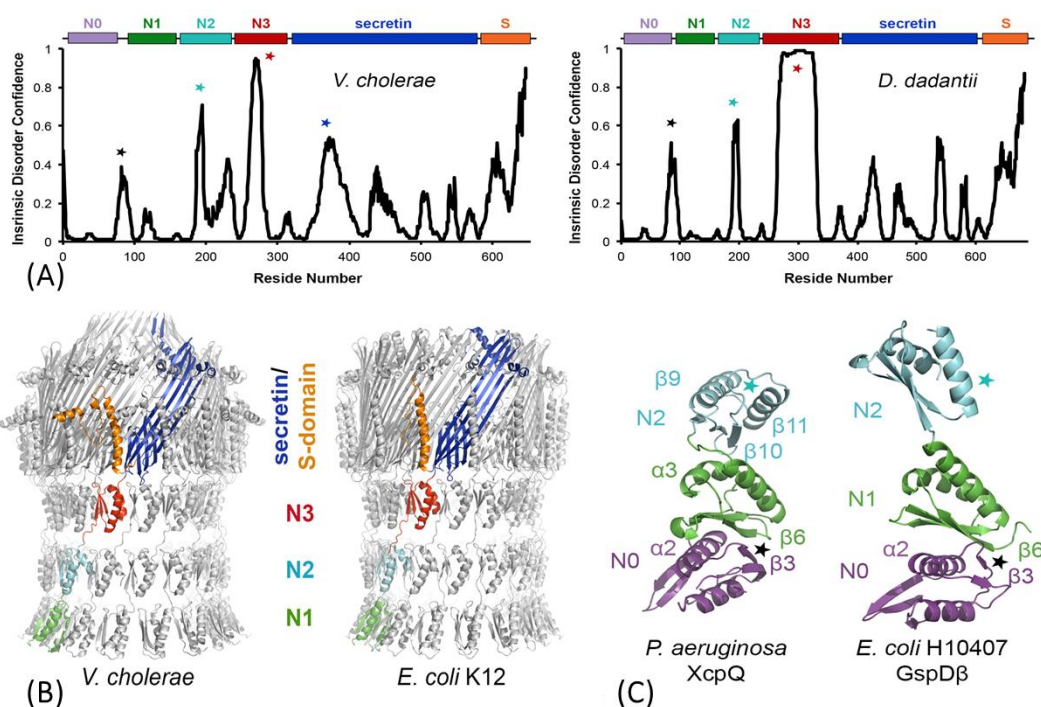


Figure 1.6 Structure and function of GspD.

A: Intrinsic disorder plots of EpsD from *V. cholerae* and OutD from *D. dadantii* (Ward, McGuffin et al. 2004). Stars represent significantly disordered regions. B: Cryo-EM structures of the *V. cholerae* and *E. coli* GspD proteins (Yan, Yin et al. 2017). C: N-domain of GspD in *P. aeruginosa* and *E. coli* (Van der Meeren, Wen et al. 2013, Yan, Yin et al. 2017) (from Gu et al, 2017).

1.4.4 Pilotins

Several factors have evolved to target the secretin subunits to the outer-membrane. Small OM lipoproteins of the GspS family referred to as ‘pilotins’ (Hardie et al. 1996) allow efficient targeting to the OM via the lipoprotein sorting pathway (Collin et al. 2011). Recent studies have revealed the interaction between PulS/GspS and the C-terminal S domain of the secretin PulD/GspD, which is intrinsically disordered (Nickerson, Tosi et al. 2011). *D. dadantii* OutS pilotin shows similar characters (Gu et al. 2012, Rehman et al. 2013) and these studies reveal the structure of the bound secretin / pilotin complex (Fig. 1.7). The absence of pilotin results in the mis-location and degradation of the secretin suggesting that the pilotin has a key role in targeting secretin to the outer-

membrane (Daefler, Guilvout et al. 1997, Shevchik, Robert-Baudouy et al. 1997, Shevchik and Condemine 1998).

During the T2SS biogenesis pilotins bind the S-domain of secretin which emerges from the Sec pathway in the inner-membrane and targets the outer-membrane via the LoL system (Collin, Guilvout et al. 2011).

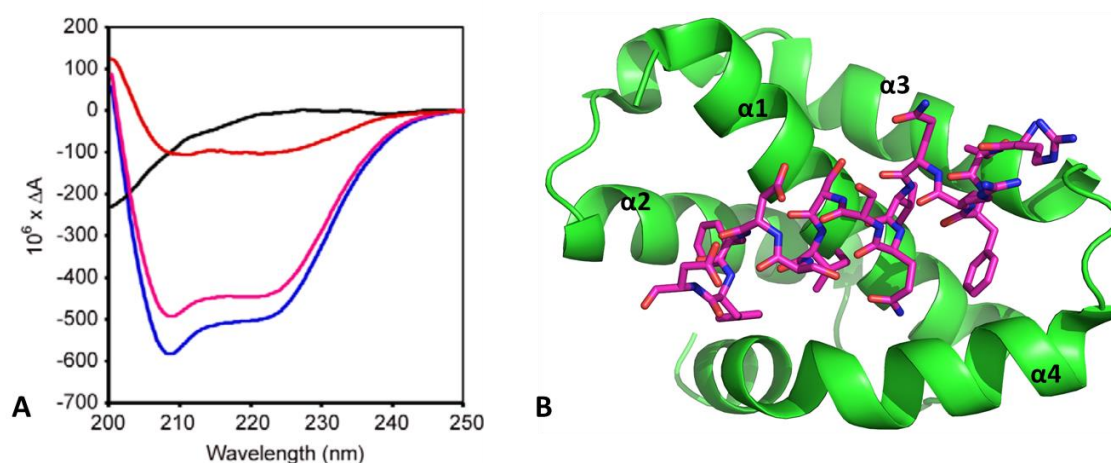


Figure 1.7 The *Dickeya dadantii* secretion-pilotin complex.

(A) The interaction between the secretin and pilotin detected by far uv circular dichroism spectra. 18-residue secretin peptide alone (black), pilotin alone (pink) and a stoichiometric ratio of both pilotin (OutS) and secretin peptide (OutD691–708) together (blue). The difference between the secretin/pilotin complex and the pilotin only is shown in red. (B) Structure of the secretin peptide bound (magenta) to the pilotin (green). (Gu, Rehman et al. 2012, Rehman, Gu et al. 2013).

1.5 The inner-membrane platform proteins

The inner-membrane platform, contains multiple copies of at least four different membrane proteins: the cytoplasmic “secretion ATPase” EpsE and the membrane proteins EpsL, EpsM, EpsC and EpsF (Fig. 1.8). Their role might be to coordinate substrate recruitment and pilus assembly (Nivaskumar and Francetic 2014).

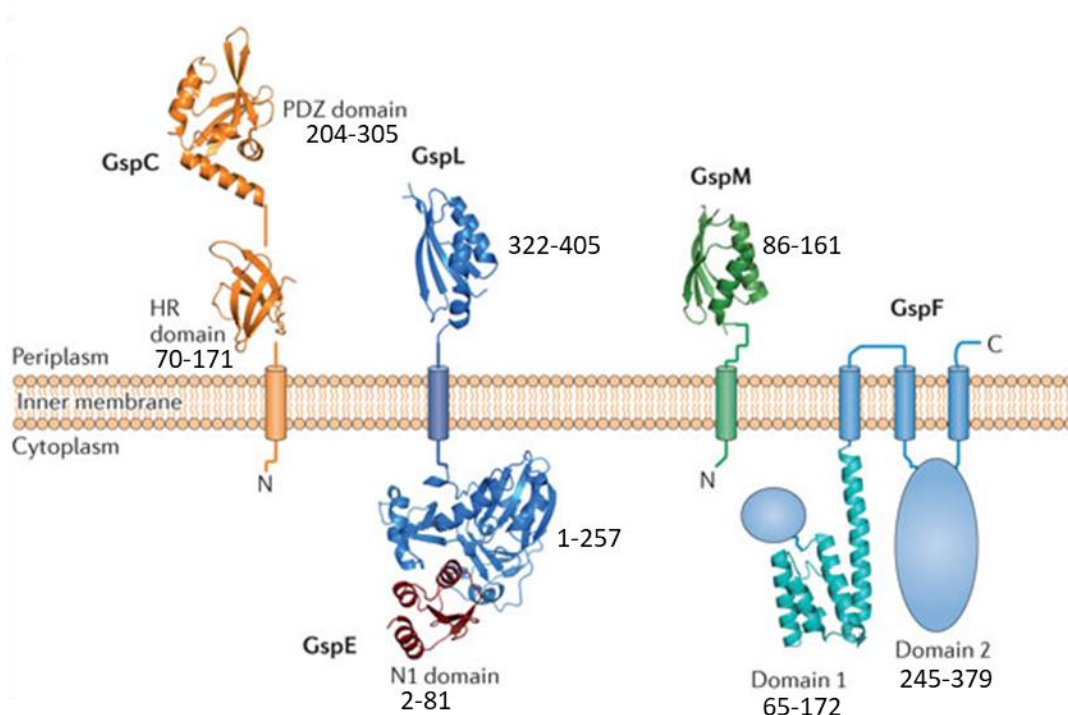


Figure 1.8 The inner-membrane platform.

Known structures are shown in ribbon representation; unknown domain structures in GspF are shown as circles. The structures shown are: the HR domain of ETEC GspC (PDB: 3OSS) and the PDZ domain of *V. cholerae* EpsC (PDB: 2I4S); the periplasmic domain of *V. parahaemolyticus* EpsL (PDB: 2W7V); the complex of the cytoplasmic domain of EpsL and the N1 domain of EpsE from *V. cholerae* (PDB: 2BH1); the periplasmic domain of *V. cholerae* EpsM (PDB: 1UV7); and the first cytoplasmic domain of *V. cholerae* EpsF (PDB: 3C1Q) (Korotkov et al 2013) (Residue numbers are indicated).

1.5.1 General secretion pathway protein, GspE

It has been noted that GspE is a reluctant hexamer preferring to crystallize in different oligomeric states. X-ray structures of EspE (GspE) from *V. cholera* and XpsE from *X.campestris* show a bilobed monomer structure with the ATP binding site between the N-terminal and the C-terminal domains (Robien et al. 2003, Chen et al. 2005). When fused to the hexameric protein, Hcp1, a quasi-C6 GspE hexamer was produced showing increased ATPase activity (Lu et al. 2013). GspE has an extended N-terminal domain compared to other ATPases and forms a stable complex with the cytoplasmic domain of

GspL which is the membrane anchor for the ATPase (Abendroth et al. 2005, Shiue et al. 2006). It has been shown that activation of the ATPase EpsE necessitates an interaction between the C-terminal end of the cytoplasmic domain of EpsL and the phospholipid bilayer (Camberg et al. 2007). This is consistent with EpsL displacement up onto the IM to activate the ATPase. Therefore, there is the possibility that coordinated displacements of the periplasmic and TM domains of GspC, GspL and GspM result in ATPase activation. Structural studies of the ATPase PilT shows that each PilT dimer has a different active site conformation corresponding to ATP binding and hydrolysis (Satyshur et al. 2007). This is consistent with two neighbouring subunits contributing to ATP-binding and hydrolysis resulting in a total of three ATPase active sites (Patrick, Korotkov et al. 2011). Interestingly, the arrangement of the pseudopilus initiation complex GspJ-GspI-GspK corresponds to the three state ATPase and may self-assemble to induce formation of the GspE motor (Cisneros, Bond et al. 2012). In a recent study, GspL was shown to interact with EpsG (GspG) by cross-linking in *V. cholera* (Gray et al. 2011) indicating that EpsL (GspL) transmits the ATPase driven conformational changes (Due to the presence of a flexible linker in the cytoplasmic ATPase (GspE), the N-terminal domain undergoes a large movement relative to the C-terminal domain on ATP hydrolysis and the release of inorganic phosphate(Gray et al. 2011)) to pseudopilins to force fibre assembly. Rotation could also be involved in the pilus assembly mechanism, by analogy with the bacterial flagella rotation (Mora et al. 2009), although in the T2SS the rotation element is probably part of the assembly machinery rather than the pilus. If GspL acts as an anchor for the ATPase in the inner-membrane, then GspF would be a candidate for the rotating component.

1.5.2 General secretion pathway protein, GspF

The central protein of the inner-membrane platform and the only polytopic inner-membrane protein GspF, which interacts with other members from the inner-membrane platform and is thought to be a key player in the T2SS and T4PB systems (Crowther et al. 2004). GspF is a highly hydrophobic protein containing approximately 400 amino acid residues. GspF and its orthologs interact in the cytoplasm with the conserved ATPase and function in complex with the other IM components; in T2SS there are GspL, GspM and GspC. Recent cryo-electron tomography of the T4PS in intact *Myxococcus xanthus* cells allowed to visualize PilC, the GspF ortholog, as a cytoplasmic dome located between the hexameric ATPase and the stem formed by pilin PilA (Chang, Rettberg et al. 2016). It is thought that ATP hydrolysis by GspE/PilB ATPase promotes rotation of GspF/PilC in complex with nascent pilus and hence facilitates incorporation of a new pilin subunit onto the base of the growing pilus (Nivaskumar, Bouvier et al. 2014, Chang, Rettberg et al. 2016).

To date, no high-resolution structure of a full-length GspF/PilC family protein is available. A low-resolution electron microscopy reconstruction of isolated PilG (PilC ortholog from *Neisseria meningitidis* T4PS) showed a tetrameric bilobed structure (Collins et al. 2007). However, this structure is significantly larger than that of PilC visualized by cryo-electron tomography of the intact T4PS and provoked steric clash. The authors have therefore proposed that PilC forms a dimer under physiological conditions (Chang, Rettberg et al. 2016). Several high-resolution structures of isolated N-terminal cytoplasmic domain of GspF/PilC have been reported, while the C-terminal domain has escaped from the crystallization attempts. More recently, Abendroth et al. have reported a crystal structure of truncated form of the first N-terminal of *Vibrio cholera* EpsF (Abendroth et al. 2009): This domain, termed cyto1-EpsF⁵⁶⁻¹⁷¹, shows a six helix bundle structure, with the final

helix terminating at the point where the first TM helix is predicted to begin (Fig. 1.9A). The structure forms a tight dimer in the crystal, with an interface formed from the $\alpha 1$, $\alpha 2$, and $\alpha 6$ helices, and the two monomers are related by twofold symmetry. However, PilC, the GspF homologue involved in T4P biogenesis in *Thermus thermophilus* showed quite different mode of dimerization (Fig. 1.9B). The PilC N-domain forms an asymmetric dimer through a “helical ladder” of hydrophobic residues (Karuppiyah et al. 2010). Each PilC^{53–168} monomer adopts a helical bundle structure, consisting of six helices arranged in a circular manner, with an up-down-up-down-up-down topology. Similar structures have been reported for the N-terminal cytoplasmic domain of TcpE from *V. cholerae* toxin-coregulated pilus (Kolappan and Craig 2013)(Fig. 1.9C). However, dimer interface of cyto-PilC is completely different from this of EpsF and cyto TcpE does not form a dimer in solution or in the crystal (Abendroth, Mitchell et al. 2009, Karuppiyah, Hassan et al. 2010, Kolappan and Craig 2013).

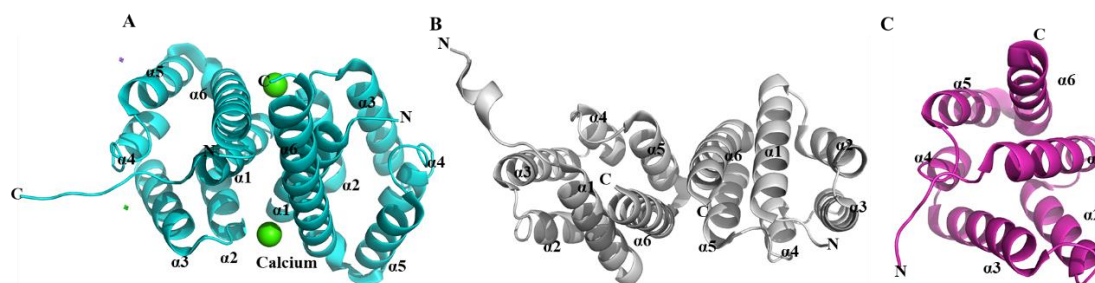


Figure 1.9 Crystal structures of EpsF⁵⁶⁻¹⁷¹, PilC⁵³⁻¹⁶⁸ and TcpE¹⁻¹⁰⁴

A: EpsF⁵⁶⁻¹⁷¹; B: PilC⁵³⁻¹⁶⁸; C: TcpE¹⁻¹⁰⁴. Calcium marked as green ball.

Bioinformatics analysis and *in vivo* topology studies using BlaM and alkaline phosphatase fusions indicate that GspF possesses three transmembrane segments (TMS) and two large cytoplasmic domains of high sequence similarity; only a small loop and a short C-terminal extension located in the periplasm, (Fig. 1.8 (Thomas et al. 1997, Arts,

de Groot et al. 2007, Abendroth, Mitchell et al. 2009)). A different topology, including four TMS and the second large domain in the periplasm has been suggested for *E. coli* EPEC T4P PilC homolog, BfpE (Lallemand et al. 2013) but it is not commonly accepted. Yet a more distantly related FlaJ component from archaeal flagella, is predicted to possess seven TMSs. Pull-down and Yeast two-hybrid studies suggest that GspF Cyto1 interacts with GspE and GspL, forming the so called T2SS assembly platform also including GspM (Py, Loiseau et al. 2001). GspF requires the presence of GspL and GspE for its full stability (Arts, de Groot et al. 2007).

1.5.3 General secretion pathway protein, GspM

GspM has a short cytoplasmic sequence, a transmembrane helix (TMH) and a periplasmic domain (Korotkov, Sandkvist et al. 2012). The closest homologues of GspM is EpsM is from *V. cholera*. Since full-length EpsM has not been crystallized so far. Jan Abendroth et al. crystallized the periplasmic soluble domain, which consists of two $\alpha\beta$ -subdomains forming a sandwich of two α -helices and a four-stranded antiparallel β -sheet repeats that form a cyclic permutation of the ferredoxin fold (Abendroth et al. 2004). Interestingly, the structure contains an extra electron density forming the cleft between the two $\alpha\beta$ -domains, suggesting that this site might be used for substrate binding. Co-immunoprecipitation experiments have shown that EpsM interacts with another cytoplasmic membrane protein, EpsL (Sandkvist, Hough et al. 1999). EpsM also appears to enhance the interaction between EpsE and EpsL (Sandkvist et al. 2000).

1.5.4 General secretion pathway protein, GspL

GspL consists of a cytosolic domain, a TMH and a periplasmic domain. The crystal structure of the cytoplasmic domain of *V. cholerae* EpsL consists of three β -sheet-rich domains. With domain I and III similar to the RNaseH-fold, cyto-EpsL shows structural

homology with the superfamily of actin-like ATPases (Abendroth, Murphy et al. 2005). However, no ATPase-binding site has been found in the structure. This is in contrast to a T4PS protein, PilM, which is homologous to the cytoplasmic domain of GspL and has a site that is capable of binding ATP (Karuppiyah and Derrick 2011). Interestingly, the chain of the periplasmic domain of *Vibrio parahaemolyticus* EpsL adopts the same circular permutation of the ferredoxin fold that was first observed in the periplasmic domain of EpsM (Abendroth et al. 2009). Both peri-EpsL and peri-EpsM form dimers; however, the site of subunit–subunit interaction appears to be entirely different. The cytoplasmic domain of GspL binds to the extended N-domain of the assembly ATPase GspE forming a hetero-tetramer (Abendroth, Murphy et al. 2005), while its periplasmic domain interacts with GspM (Sandkvist, Keith et al. 2000). The two GspL domains correspond to PilM and PilN in *P. aeruginosa* Type 4 Pilus System (Ayers et al. 2009).

1.5.5 General secretion pathway protein, GspC

The GspC protein in *D. dadantii* is called OutC and is a 272 residues inner-membrane protein. They consist of a short cytoplasmic segment, a TMH (transmembrane helix) followed by a flexible linker (TMHR) and two periplasmic domains (the homology region (HR) domain and a PDZ domain, Fig. 1.10 A and B). The sequence of the HR domains with a 6-stranded β sandwich fold (Korotkov et al. 2011, Gu, Rehman et al. 2012), is relatively well conserved, whereas the TMHR shows very little sequence homology (Gu, Shevchik et al. 2017). Intrinsic disorder and secondary structure predictions (McGuffin, Bryson et al. 2000, Ward, McGuffin et al. 2004) suggest that the TMHR linker is not fully unstructured and contains a single α -helix (Gu, Shevchik et al. 2017). This has been also confirmed by NMR for OutC in *D. dadantii* (Gu, Rehman et al. 2012). In some cases, the PDZ domain is replaced by a coiled-coil domain (Bleves et al. 1999). The domain can adopt open and closed conformations (Korotkov et al. 2006). Although the PDZ and

coiled-coil domains can be swapped without loss of function (Gerard-Vincent, Robert et al. 2002), in *D. dadantii* deletion of the PDZ domain abolishes secretion of all proteins but one, so the PDZ domain might be involved in the regulation of secretion specificity (Bouley, Condemine et al. 2001).

The GspC is thought to regulate the pore on the outer-membrane which can export the periplasmic proteins when it is open (Lee et al. 2004, Lee et al. 2005). GspC and GspD interact with each other via periplasmic domains. This interaction was confirmed by the following experiments: In *Enterotoxigenic Escherichia coli* (ETEC), the crystal structures of the GspC HR domain in complex with the GspD two or three N-terminal domains show that the HR domain adopt an all- β structure (Fig. 1.9C) (Korotkov, Johnson et al. 2011). And the HR domains is also reported to interact with the periplasmic domain of OutD biochemically by co-expressing method (Korotkov, Krumm et al. 2006). GspC was found partially in the Outer-membrane fraction in *Klebsiella oxytoca* (Bleves, Gerard-Vincent et al. 1999). This interaction between the GspC and GspD seems critical for the function and assembly of the T2SS (Lybarger et al. 2009).

GspC works as a bridge which can not only interact with the secretin in the outer-membrane but also communicate with the inner-membrane complex. The N-terminal region of the GspC can interact with GspL and GspM complex which could form a larger GspC-GspL-GspM complex (Robert et al. 2002, Tsai et al. 2002).

Additionally, the PDZ domain of the GspC is involved in substrate recognition (Korotkov, Krumm et al. 2006). In *Pseudomonas aeruginosa*, the periplasmic domain of the GspC has been shown to interaction with exoproteins directly (Douzi, Ball et al. 2011). Similarly, it is also reported that the exoproteins interact with the periplasmic domain of the secretin (Shevchik, Robert-Baudouy et al. 1997, Reichow et al. 2010, Douzi, Ball et

al. 2011). These results suggested that both GspC and secretin are involved in substrate recognition. However, recent study also demonstrates the interaction between the substrate and the pseudopilin by (Douzi, Ball et al. 2011). These studies may reflect different steps of the secretion. The GspC and GspD may involve in the early steps of the secretion which could be recognition and recruitment of the substrates(Pineau, Guschinskaya et al. 2014).

GspC alone or GspC and GspD together are considered as the gatekeeper for the entry into the T2SS since many substrates are too large to go through the vestibule which will destroy of the hetero-dimer formed by the GspC-HR domain and periplasmic N-domain of GspD. Several different orientation of GspC and GspD have been demonstrated via structural studies and further trapped with *in vivo* cross-linking (Korotkov, Krumm et al. 2006, Korotkov, Gonen et al. 2011, Wang et al. 2012, Van der Meeren, Wen et al. 2013).

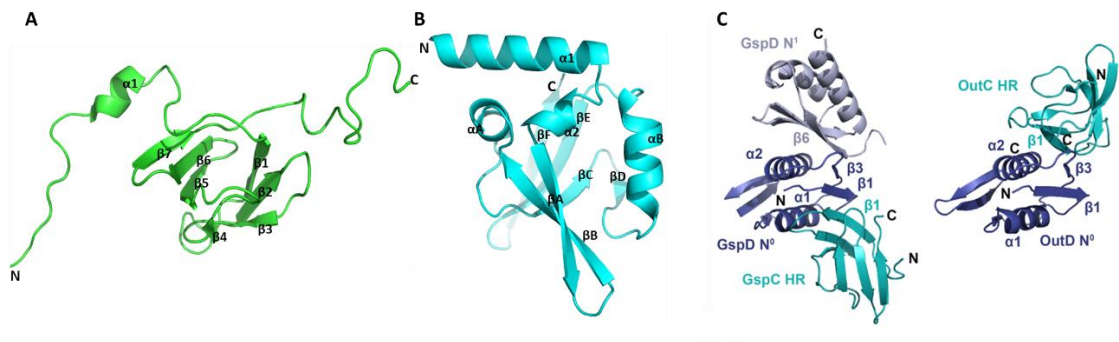


Figure 1.10 Structure of GspC and GspC/GspD complex.

A: NMR structure of the *D.dadantii* OutC HR domain; B: Crystal structure of the *V. cholerae* EpsC PDZ domain; C: Crystal structure of the *E. coli* GspC HR-GspD N01 complex and the model of the *D. dadantii* OutC HR-OutD N0 complex, GspC are in cyan, N0 are in blue and N1 is in grey (Gu et al 2017).

1.6 Interactions within the inner-membrane platform

Cytoplasmic domain I of OutF instead of cytoplasmic domain II could interact with both OutE and OutL using yeast two-hybrid system in *Erwinia chrysanthemi* (Py, Loiseau et al. 2001). OutE, OutL and OutF could form a stable complex *in vivo* and OutL is required for the OutE and OutF complex but OutF is not required for the OutE and OutL complex. The cytoplasmic loop of XcpS(GspF) was also shown to be stable in the presence of XcpR(GspE) and XcpY(GspL). And the additional XcpZ(GspM) has slightly increased the amount of the GspF compared with only XcpR and XcpY (Arts, de Groot et al. 2007). Crowther et al., showed that BfpE which is the ortholog of GspF in enteropathogenic *Escherichia coli* is involved in recruiting the ATPase to the cytoplasmic membrane (Crowther, Anantha et al. 2004). GspM is also involved in the formation of the complex and there is strong evidence for the interaction between GspM and GspL. This interaction was confirmed using two-hybrid between the periplasmic domains of GspL and GspM (Py, Loiseau et al. 2001). Sandkvist et al. also found that EpsL (GspL) and EpsM (GspM) could interact with each other and form a stable complex without the presence of other Eps proteins (Sandkvist, Hough et al. 1999). In *V.cholerae* GspL interacts with EpsG (GspG) by cross-linking *in vivo* (Gray, Bagdasarian et al. 2011) indicating that (EpsL) GspL transmits the ATPase driven conformational changes to pseudopilins to prompt fibre assembly. In *Neisseria meningitides*, the major pilin subunit PilE was shown to interact with PilG (GspF in T2SS) and with PilO (GspM in T2SS) (Georgiadou et al. 2012).

1.7 Aims and objectives

The overall aim of my work was to better understand the architecture of the type II secretion system and thereby illuminate the mechanism of secretion. With the emergence of antimicrobial resistance in pathogenic bacteria we need new ways of defeating these potent adversaries and inhibiting the secretion system used may be one method of achieving this.

The specific objectives of my Thesis are:

- 1) To produce, purify and study the structure of the inner-membrane protein OutF.

The first output from this work is the structure of the first cytoplasmic domain of *D. dadantii* OutF. The structure is described and compared with other similar structures.

- 2) To produce, purify, and study the structure of the intact inner-membrane complex, or sub-complexes of the platform complex. The major achievement from this work is the expression of protein complexes and determination of electron density envelopes of inner-membrane protein complexes from *D. dadantii*. This Thesis provides a contribution to higher-resolution studies of the inner-membrane complex and hence entire type II secretion system.

D. dadantii is a plant pathogen threatening our food security, but ultimately the lessons learnt will be important for understanding the generic properties of the type II secretion system across plant, human and animal pathogens.

Chapter 2 Material and methods

2.1 Materials

2.1.1 Water

The deionised water (dH₂O) used in the experiments was purified using a PureLab option (Elga) which provide water of 15M Ω .cm resistivity. When the buffer was to be used for crystallization, the water was further deionised to 18.2 M Ω .cm resistivity using a PureLab Classic (Elga). Double distilled water(ddH₂O) was achieved by autoclaving water from the PureLab purification system (15M Ω .cm resistivity) and used for molecular biology experiments.

2.1.2 Agarose gel electrophoresis

Agarose gels for analysing plasmids and inserts were made using:

1 x TAE (Tris-Acetate EDTA) was made using the 50 x TAE stock.

1% Agarose gel was made using agarose powder (VWR).

The DNA marker was HyperLadder purchased from Bioline.

2.1.3 DNA clean-up

Purification of PCR products, enzyme digests and DNA gel extraction were using QIAquick Gel Extraction Kit (QIAGEN).

2.1.4 Polymerase Chain Reaction

Primers for Polymerase Chain Reaction (PCR) reactions were ordered from Eurofins. Polymerase used for DNA amplification was Q5® High-Fidelity DNA Polymerase (NEB).

2.1.5 Broth and media

a Luria-Bertani (LB) and YT broth

Table 2.1 Components of LB broth

Component	LB (g/liter)	2YT (g/liter)
Tryptone (Sigma)	10	16
Yeast Extract (Sigma)	5	10
NaCl (Thermofisher)	5	5

b Terrific Broth (TB) media

Table 2.2 Components required for making the TB media

Components	Yeast extract	Tryptone	Buffer
g/liter	24	12	2.31 g of KH_2PO_4 and 12.54 g of K_2HPO_4

c Amp/X-gal/IPTG plates

Table 2.3 Components used for making Amp/X-gal/IPTG plates

Components	concentration	solvent
Ampicillin(Thermofisher)	100µg/ml	ddH2O
IPTG (Anatrace)	0.1mM	ddH2O
X-gal	80µg/ml	DMSO

d M9 media

Minimal M9 media was used for ^{15}N labelling of OutF⁶⁵⁻¹⁷² for 2D NMR spectra.

100 x trace element stock

Table 2.4 Components used for making 200ml 100 x trace element stock

EDTA	FeCl ₃ .6H ₂ O	CuCl ₂ .2H ₂ O	H ₃ BO ₃	CoCl ₂ .6H ₂ O	MuCl ₂ .6H ₂ O
1g	166mg	3mg	2mg	2mg	0.32mg

e 10 x M9 Stock

Table 2.5 Components used for making 1 Litre 10 x M9 Stock

Na ₂ HPO ₄	KH ₂ PO ₄	NaCl
30g	15g	2.5g

f 4 x additives

Table 2.6 Components used for making 200ml 4 x additives

Thiamine	Biotin	CaCl ₂	¹⁵ N-Ammonium chloride (Cambridge Isotope laboratories)	Glucose	MgSO ₄ · 7H ₂ O
4mg	4mg	1.2mg	4g	16g	1g

Sterile additives were obtained by filtration using a 0.02µm filter.

2.1.6 Competent cells

Table 2.7 Competent cells used for expression and cloning

Cells used for cloning	DH5α (BIOLINE), Stellar™ Competent Cells (Clontech)
Cells used for expression	BL21(DE3) (NEB), C41, C43 (Lucigen)

2.1.7 Buffers

Table 2.8 Buffers for protein purification and biochemistry

Buffer	composition
Nickel column binding buffer	20mM Tris pH 8.0, 500mM NaCl, 20mM Imidazole
Nickel column wash buffer	20mM Tris pH 8.0, 500mM NaCl, up to 70mM Imidazole
Nickel column elution buffer	20mM Tris pH 8.0, 500mM NaCl, 400mM Imidazole
Strep column binding and wash buffer	20mM Tris pH 7.4, 150-300mM NaCl, 1mM DTT, 0.04% <i>n</i> -Dodecyl β -D-maltoside (DDM, thermofisher) (with or without 1mM <i>N</i> , <i>N</i> -Dimethyldodecylamine <i>N</i> -oxide (LDAO)
Strep column elution buffer	20mM Tris pH 7.4, 150-300mM NaCl, 1mM DTT, 0.04%DDM (with or without 1mM LDAO), 2.5mM desthiobiotin (IBA)
Glutathione column binding and wash buffer	20mM Tris pH 7.4, 150-300mM NaCl
PreScission protease cleavage buffer	20mM Tris pH 7.0, 150-300mM NaCl, 1mM EDTA
Reduced glutathione buffer	20mM Tris pH 8.0, 10mM reduced glutathione
Size exclusion superdex 75/200	20mM Tris pH 7.0-8.0, 150-300mM NaCl, 1mM DTT
Biorad assay	Working Biorad solution-1.25ml of Biorad stock solution was mixed with 4.5ml of dH ₂ O Reaction: 47.5 μ l of working Biorad solution was mixed with 2.5 μ l of protein sample
Membrane protein solubilizing buffer	20mM Tris pH 7.4, 150-300mM NaCl, 1mM DTT, 1%DDM (with or without 10mM LDAO)

2.1.8 Constructs and vectors

Table 2.9 Constructs kindly provided by Dr. Vladimir Shevchik (University of Lyon)

Name	Vector	Protein
pGEX-6P-1-OutF	pGEX-6P-1	OutF ¹⁻⁴⁰⁸
pBAD-ELM	pBAD	OutE, OutL and OutM

Table 2.10 List of vectors used in Chapter 3 and 4

Plasmid	Description	Source
pET-14b	The pET-14b vector carries an N-terminal His•Tag® sequence followed by a thrombin site and three cloning sites. The vector carries the inducible T7 promoter/operator for the regulated expression of proteins and the ampicillin (Amp) Resistance cassette	Novagen
pET-3a	The pET-3a-d vectors carry an N-terminal T7•Tag® sequence and BamHI cloning site. These vectors are the precursors to many pET family vectors.	Novagen
pOPINS3C	The vector carries the inducible T7 promoter/operator for the regulated expression of proteins, the SUMO tag between the His-tag and the protein of interest can be cleaved from the protein of interest with 3C protease.	Oxford protein production facility
pASK-IBA3C	The vector carries the inducible tetracycline promoter/operator for the regulated expression of proteins, the <i>Strep</i> -tag for C-terminal fusion to the recombinant protein and the Chloramphenicol Resistance cassette	IBA solutions for life Sciences
pOPIN-3c-eGFP	The vector carries the inducible T7 promoter/operator for the regulated expression of proteins, the GFP for C-terminal fusion to the recombinant protein and the ampicillin (Amp) Resistance cassette	Oxford protein production facility

2.1.9 Vivaspin concentrators

Proteins were concentrated in Vivaspin columns using swing bucket rotors in a centrifuge at 4,000g. Vivaspin500 (capacity 500µl), Vivaspin6 (capacity 6ml) and Vivaspin20 (capacity 20ml) were purchased from Sartorius. The Molecular weight cut off of the polyethersulfone membranes were 10kDa, 30kDa or 30kDa.

2.1.10 Proteases and protease inhibitors

PreScission protease (GE healthcare) was used for GST tag cleavage. Thrombin (Novagen) was used for His tag cleavage.

For bacterial lysis, protease inhibitors are an important class of additives. Generally, cell disruption will release lots of proteolytic enzymes. To reduce protein losses and increase the homogeneity of the protein produced, it is necessary to add protease inhibitors to the cell lysate. Pierce™ Protease Inhibitor Mini Tablets were used to inhibit protease activity. One tablet of Mini is sufficient in 10ml extraction solution.

Alternatively, phenylmethylsulfonyl fluoride (PMSF) and EDTA protease inhibitors were used for serine and metalloproteases, respectively.

PMSF has a short half-life time in water. Therefore, a stock solution of 200 mM in ethanol or isopropanol was made and diluted into buffer just before use. Since EDTA is not compatible with needed divalent metal cations (Mg^{2+} , Ni^{2+}) in the buffer, no EDTA was added to the buffers used for metal affinity chromatography on Ni^{2+} -columns.

2.1.11 Inducers

Isopropyl β-D-1-thiogalactopyranoside (IPTG, Anatrace) was used to induce protein expression under the control of T7 promoter. For induction, a filtered 1 M stock solution

of IPTG was typically made and diluted 1000 -10000 times into a logarithmically growing bacterial culture, to give a final concentration of 0.1-1 mM.

L-arabinose (Thermofisher) was used to induce protein expression under the control of the pBAD promoter. And various concentration of arabinose (0.2%,0.02%,0.002%,0.0002%) were tested for the efficiency of protein induction.

AHT (Anhydrotetracycline, IBA) is an efficient inducer of the tetracycline promoter on pASK-IBA vectors(IBA)(Skerra 1994). Typically, 2mg/ml stock solution was prepared in DMF (N, N-Di-Methyl-Formamid) or ethanol and diluted into 200 µg/ml into the cell culture.

2.1.12 Detergents

Detergents were used to solubilize and enhance the stability of membrane proteins. Generally, 1% DDM was used to solubilize membrane fractions, alternatively 10mM LDAO was used. In the following purification step, 0.04% n-Dodecyl β-D-maltoside (DDM) or/and 1mM N, N-dimethyl-1-dodecanamine-N-oxide (LDAO) were added into the buffer to maintain the solubility of the proteins.10% DDM was prepared in water or buffer as a stock solution and kept in -20°C for long term storage or 4°C for short term storage.

2.1.13 Chromatography

Size exclusion chromatography made use of the ÄKTA FPLC machine (GE healthcare) connected to a Frac-950 fraction collector in the cold room (4°C). The system was monitored and operated by UNICORN control software.

2.1.14 Thermofluor assay

Protein folding/ unfolding was monitored by mixed with SYPRO-Orange dye (Invitrogen).

A CFX Connect™ Real-Time PCR Detection System (Bio-Rad) with 96-well PCR detection plate (AB gene AB-1100) was applied for the Thermofluor analysis.

2.1.15 Circular Dichroism

Far-UV circular dichroism (CD) measurements were made using a Jasco J-715 spectropolarimeter equipped with a PTC-348WI temperature controller. 1mm path length fused silica cuvettes were used for all CD measurements.

2.1.16 Protein crystallization

The commercial crystal screen kits (Wizard I and II, MemGold™ MD1-39 (molecular dimensions), MemGold™ Eco Screen MD1-39-ECO, MemStart™ & MemSys MD1-33 (molecular dimensions)) were used for initial screening.

96-well microplates (Thermofisher) were setup using a Mosquito pipettor at 18°C. For further optimization, 24-well plates (Hampton Research) and siliconized cover slides (Hampton Research) were used and sealed with Silicone vacuum grease bought from Dow Corning.

2.1.17 NMR spectra

NMR spectra were acquired using Bruker Avance 700 and 600 MHz spectrometers at Crick Institute. Protein samples were prepared in Wilmad 5mm tubes (Sigma) with 10% D₂O (Goss Scientific Instruments Ltd).

2.1.18 TEM

Images were collected using Jeol 1230 TEM in Nanovision Centre of QMUL or using the FEI Tecnai F20 electron microscope at Imperial College operating at 200 kV and equipped with an FEI Falcon II CMOS direct electron detection camera. Micrographs were collected automatically using the EPU software.

2.2 Methods for molecular biology

2.2.1 PCR

a Primer design

The primers were designed using the Primer 5 or online software Primer 3 and followed the general considerations:

The primers should be between 18-35 base pairs;

The GC% should be between 40-60%;

The primers within itself or between the forward and reverse should not contain any strong secondary structure such a hairpin.

A. Primers for infusion cloning also required:

12-18 bases homologous at the 5' end of the primer which is complementary to the termini of the linearized vector (Table 2.11).

Table 2.11 An example of 15 bases of homology in primers for infusion cloning

vector	Forward Primer Extension	Reverse Primer Extension
pOPINF	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA

B. Primer for sticky end ligation

The restriction site should be designed in the 5' end of the primer according to the multiple cloning sites in the vector. Importantly, 2-6 extra bases should be added on either side of the recognition site to increase cleavage efficiency.

C. Primer for blunt end ligation

The primers for blunt ligation without restriction site should be 5' phosphate which is optional for primer orders.

Table 2.12 List of the primers used in Chapter3 to amplify *outF*, *outL* and *outE* genes and parts of these genes.

<i>Primer</i>	<i>Sequence</i>	<i>Restriction site</i>
<i>pET-24d-OutF¹⁻⁴⁰⁸-F</i>	<i>CTAGCTAGCATGGCGCTGTTCCAGTATCA</i>	<i>NheI</i>
<i>pET-24d-OutF¹⁻⁴⁰⁸-R</i>	<i>CCGCTCGAGCATACTCATCAGGGTATT</i>	<i>XhoI</i>
<i>pET-24d-OutF⁵³⁻¹⁶⁸-F</i>	<i>CTAGCTAGCATGAGCGGCTTCTCCCTG</i>	<i>NheI</i>
<i>pET-24d-OutF⁵³⁻¹⁶⁸-R</i>	<i>CCGCTCGAGACGGCTGCGCATCTGCTG</i>	<i>XhoI</i>
<i>pET-14b-OutF⁵⁹⁻¹⁷²-F</i>	<i>CGCCATAIGTTGCGCCGTTT</i>	<i>NdeI</i>
<i>pET-14b-OutF⁵⁹⁻¹⁷²-R</i>	<i>CGCGGATCCCTACGCCTGTTGGATA</i>	<i>BamHI</i>
<i>pET-14b-OutF⁶⁵⁻¹⁷²-F</i>	<i>CGCCATAIGATTAGCGCCAGCGATC</i>	<i>NdeI</i>
<i>pET-14b-OutF²⁶⁷⁻³⁷⁴-F</i>	<i>CGCCATAIGGCGCGGGGGCTTAACA</i>	<i>NdeI</i>

<i>pET-14b-OutF²⁶⁷⁻³⁷⁴-R</i>	<i>CGCGGATCCCTACCCCAGCGCCAGC</i>	<i>BamHI</i>
<i>pET-14b-OutF²⁷⁰⁻³⁶³-F</i>	<i>CGCCATATGGGCCTTAACACCGCCCGCTA</i>	<i>NdeI</i>
<i>pET-14b-OutF²⁷⁰⁻³⁶³-R</i>	<i>CGCGGATCCCTAGCGATCCTGATTGTCC</i>	<i>BamHI</i>
<i>pOPIN3C-OutF¹⁻⁴⁰⁸-F</i>	<i>GGGGTACCATGGCGCTGTTCCAGTATCAGGC</i>	<i>KpnI</i>
<i>pOPIN3C-OutF¹⁻⁴⁰⁸-R</i>	<i>CCCAAGCTTCTACATACTCATCAGGGTATTCAAC</i>	<i>HindIII</i>
<i>pET-3a-OutF¹⁻¹⁷²-F</i>	<i>CGCCATATGGCGCTGTTCCAGTATCAGGC</i>	<i>NdeI</i>
<i>pET-3a-OutF¹⁻¹⁷²-R</i>	<i>CATGGATCCCTAGCTAGCTCACGCCTGTTGGATAC</i>	<i>BamHI,NheI</i>
<i>pET-3a-OutF⁶⁵⁻¹⁷²-F</i>	<i>CGCCATATGATTAGCGCCAGCGATC</i>	<i>NdeI</i>
<i>pET-3a-OutF⁶⁵⁻¹⁷²-R</i>	<i>CATGGATCCCTAGCTAGCTCACGCCTGTTGGATAC</i>	<i>BamHI,NheI</i>
<i>pET-3a-OutE¹⁻⁵¹³-F</i>	<i>CATCATATGAGCGATCAGCCCGTCAACACG</i>	<i>NdeI</i>
<i>pET-3a-OutE¹⁻⁵¹³-R</i>	<i>CATGGATCCCTAGCTAGCTCATTCTCTTTCGTT</i>	<i>BamHI,NdeI</i>
<i>pOPINF-OutE¹⁻⁵¹³-F</i>	<i>GGGGTACCATGAGCGATCAGCCCGTCAACACGCCT</i>	<i>KpnI</i>
<i>pOPINF-OutE¹⁻⁵¹³-R</i>	<i>CCCAAGCTTTCATTCTCTTTCGTTACCCGCACCA</i>	<i>HindIII</i>
<i>pET-14b-OutL¹⁻²⁵⁷-F</i>	<i>CATCATATGAACAGGGCCGAGAACGCCA</i>	<i>NdeI</i>
<i>pET-14b-OutL¹⁻²⁵⁷-R</i>	<i>CATGGATCCCTAGCTAGCTCACGCCTGTTGGATAC</i>	<i>BamHI,NheI</i>

Table 2.13 List of DNA primers used in Chapter

<i>Contrast name</i>	<i>Primer sequence (5'-3')</i>	<i>Enzy-</i>
----------------------	--------------------------------	--------------

		<i>me</i>
<i>pASK3c: T2SS_E-Mhis+strep</i>	F: TCCGAGCTCAGCGATCAGCCCGTCAACACGCCTGA	<i>SacI</i>
	R: CCGCTCGAGGTGATGATGATGATGATGCAGCACCCGTTCCAGCGACAACCGGGTCACCTC	<i>XhoI</i>
<i>pASK3c: T2SS_C-Mhis+strep</i>	F: TCCGAGCTCATGAATATCTCGAAATTGCCACCGCTATCTCCGTC	<i>SacI</i>
	R: CCGCTCGAGGTGATGATGATGATGATGCAGCACCCGTTCCAGCGACAACCGGGTCACCTC	<i>XhoI</i>
<i>pASK3c: T2SS_S-Ohis+strep</i>	F: TCCGAGCTCCATGTATCTTCGCTAAAAGTGGTCCTTTTTGGTGCTGTTGCCTG	<i>SacI</i>
	R: CCGCTCGAGGTGATGATGATGATGATGCAGCACCCGTTCCAGCGACAACCGGGTCACCTC	<i>XhoI</i>
<i>pASK3c: T2SS_O-Mhis+strep (IHE)</i>	F: TCCGAGCTCCTTTTTGATGTTTTTCAGCAATACCCGCGGCGATGCCATACT	<i>SacI</i>
	R: CCGCTCGAGGTGATGATGATGATGATGCCCCCGTCCAAACTCCAGCCGCTGCACATTCAC	<i>XhoI</i>
<i>Popin-3C-eGFP: T2SS_{E-M}</i>	F: AGGAGATATACCATGAGCGATCAGCCCGTCAACAC	N/A
	R: CAGAACTCCAGTTTCAGCACCCGTTCCAGCGACA	N/A
<i>pASK3c: T2SS_{Ehis-M strep}</i>	F: TCCGAGCTCCATCATCATCATCACAGCGATCAGCCCGTCAACACGCCTGAACTGCGACCG	<i>SacI</i>
	R: TCCGAGCTCGAATTCGGGACCGCGGTCTCCCATTT	<i>SacI</i>
<i>pASK3c: T2SS_{F-Mhis+strep}</i>	F: TCCGAGCTCGCGCTGTTCCAGTATCAGGCAT TGAACGCCAGGGAAAGAAAAG	<i>SacI</i>
	R: TCCGAGCTCGAATTCGGGACCGCGGTCTCCCAT TT	<i>SacI</i>
<i>pASK3c: T2SS_{E,L}-F, M his+strep</i>	F: CGGGGTACCATGAACAGGGCCGAGAACGCCAGCGGCAAACA ACA	<i>KpnI</i>
	R: CGGGGTACCCTACATACTCATCAGGGTATTCAACTGCAGAATCGGCTGGAGGATCGC	<i>KpnI</i>
<i>pASK3c: T2SS_O-Mhis+strep+ Dhis(IHE)</i>	F: CGGGGTACCCATCATCATCATCACGTGTTTTGGCGTGATATG ACGTTGTCTATCTGGCGTAAGAAGACA AACTGGCCTCAAACA	<i>KpnI</i>
	R: CGG GGT ACC TTAGCGCAGTGCGATGGAAATGTCGT	<i>KpnI</i>
<i>pASK3c: T2SS_O-M strep+Dhis(IHE)</i>	F: CCGCTCGAGGTGCACCTGCAGGGGGACCATGGTCTCAGCGC	<i>XhoI</i>
	R: CCGCTCGAGCCCCGTCCAAACTCCAGCCGCTGCA	<i>XhoI</i>
<i>T7 primer</i>	T7 primer: TAATACGACTCACTATAGGG	
	T7 Terminator: GCTAGTTATTGCTCAGCGG	

b PCR reaction and programs

PCR reactions were carried out using a PCR thermal cycler (BIO-RAD) and the Q5® High-Fidelity DNA Polymerase. All reactions were made in thin wall 200µl PCR tubes (Starlab) at the final volume of 50µl. A typical PCR reaction setup is shown in Table 2.14 and a typical PCR reaction program setup is shown in Table 2.15.

Table 2.14 PCR reaction

Where possible primers were designed to conform optimized parameters: GC content 40-60%, length 25-35 bases, G or C at the end of the primer.

Reagent	Final concentration	50µl reaction
5×Q5 Reaction Buffer	1×	10
10mM dNTPs	200µM	1
10µM Forward Primer	0.5µM	2.5
10µM Reverse Primer	0.5µM	2.5
Template DNA	~100ng	3
Q5 High-Fidelity DNA Polymerase	0.02U/µl	1
5*Q5 High GC Enhance	1×	10
Nuclease-Free Water		To 50µl

Table 2.15 PCR program using Q5 High-Fidelity DNA Polymerase

		Temperature(°C)	time
1	Initial denaturation	98	30s

3-step cycling		Number of cycles	25~35
2	Denaturation	98	10
3	Annealing	55	20
4	Extension	72	30s per kb
Final step			
5	Final extension	72	10min

c DNA analysis and purification

1% Agarose gels were prepared and run according to standard procedures to separate DNA according to the size. PCR product and DNA fragment were purified using DNA Gel Extraction Kit: QIAquick Gel Extraction Kit (QIAGEN). All plasmids were purified from overnight culture of *E. coli* alpha-select competent cells by QIAprep® Spin Miniprep Kit (Qiagen), according to manufacturer's instruction. The final DNA concentration was determined by UV spectrophotometer according to the following formula:

$$\text{Concentration } (\mu\text{g/ml}) = (A_{260} - A_{320}) \times 50 \mu\text{g/ml}$$

Where the A260 reading is where DNA absorbs light most strongly (260 nm) and A320 is measurement for turbidity (320 nm). DNA purity is estimated from the A260/A280 ratio. An A260/A280 ratio between 1.7 and 2.0 generally represents high-quality DNA with little protein contamination.

2.2.2 Restriction enzyme digestion

Whenever double restriction enzyme digests were performed, buffer conditions compatible with both enzymes were chosen (according to NEB restriction enzyme buffer

chart). All digestions were carried out at 37 °C for at least 1 hour. A typical reaction mixture is shown in Table 2.1.

Table 2.16 Typical restriction enzyme digest for PCR products and vectors.

Regent	Volume (µl)
Restriction enzyme1	1
Restriction enzyme2	1
Reaction Buffer (10x)	2
DNA	1~2 (~100ng/µl)
dH2O	add to reach 20µl total reaction volume

Double digested DNA products were purified by electrophoresis using 1% agarose gel running in TAE buffer (40mM Tris Acetate, 1mM EDTA pH 8.0). Gels were stained with SYBR™ Safe DNA Gel Stain (Thermo Fisher) and visualized on UV transilluminator. The desired vector or PCR product was excised from gel using sterile scalpel blade and purified using Gel Extraction Kit (Qiagen). Each reaction was eluted with 15 µl water.

2.2.3 Dephosphorylation

To avoid self-ligation of the linearized vector after digestion especially for one enzyme digestion, the vector was dephosphorylated just before ligation. Alkaline Phosphatase, Calf Intestinal Phosphatase (CIP) can non-specifically catalyze the dephosphorylation of 5' and 3' ends of DNA which prevents re-ligation of linearized plasmid DNA. The reaction is prepared as shown in Table 2.14 and incubated at 37°C for 30 minutes. After that, the DNA can be purified either by gel purification or spin-column.

Table 2.17 20µl reaction for dephosphorylation

DNA	1 pmol of DNA ends
CutSmart® Buffer (10X)	2 µl
Alkaline phosphatase (CIP)	1 unit
H ₂ O, purified	to 20 µl

2.2.4 Sticky-end ligation reactions

Ligation reactions were prepared as in Table 2.15 and incubated for at least 2 hours at room temperature using T4 ligation kit (NEB). For blunt-end ligation, the incubation time should be extended one more hour.

Table 2.18 Typical sticky-end ligation reaction in 10µl volume.

The ratio between insert DNA and vector is at least 3 to 1.

Reagent	Volume (µl)
DNA insert	7
Digested vector	1
T4 Ligase	1
T4 Ligase buffer (10X)	1

2.2.5 Blunt-end ligation

Blunt-end cloning is one of the easiest methods for cloning dsDNA into plasmid vectors.

It is easy because the blunt-ended insert or vector requires little preparation. In this

method, the DNA does not require enzymatic digestion and subsequent purification which is needed for cohesive-end cloning. Without this preparation, the yield of the DNA is increased and the ligation efficiency also improved.

In these studies, we used blunt-end ligation for site mutation which involved ligating dsDNA with no restriction sites at their termini. Since this reaction depended on the available 5' phosphate and 3' hydroxyl, a special option to add 5' phosphate in the primer was chosen when ordering primers. A typical blunt-end ligation reaction is shown in Table 2.19.

Table 2.19 Blunt ligation reaction in 10 μ l volume

Reagent	Volume (μ l)
DNA	5-8
T4 ligase	1
T4 ligase buffer	1
Water	Add to 10 μ l in total

2.2.6 In-Fusion cloning

In-Fusion cloning is designed for fast, directional cloning of one or more fragments of DNA into any vector. This technique could fuse DNA fragments (e.g. PCR-generated inserts and linearized vectors) efficiently and precisely by recognizing 15-bp overlaps at their ends. These 15-bp overlaps can be engineered by designing primers for amplification of the desired sequences. To achieve a successful In-Fusion reaction, I first generated a linearized vector (either by restriction enzymes or PCR). In this example, I linearized the pOPIN-3C-eGFP vector using double digestion with *NcoI* and *PmeI* at 37°C overnight. Meanwhile, the insert was amplified using TD-PCR with primers which shared 15

homologous bases at each end with the vector. Both the linearized vector and PCR product were purified from the agarose gel. Then the In-Fusion cloning reaction was set up: the purified vector and insert were mixed in a 1:1 ratio with 50~200ng respectively, incubated with In-Fusion HD Enzyme Premix (Clontech) 50°C for 15 minutes (Table 2.20). Finally, 2.5µl ligation reaction was transformed into stellar competent cells (Clontech).

Table 2.20 In-Fusion cloning reaction

Reagent	Volume (µl)
5X In-Fusion HD Enzyme Premix	2
Linearized vector	1-2
Insert	1-2
Cloning Enhancer	1
H ₂ O	Add to reach 10 µl in total

2.2.7 Construction of pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³-OutF¹⁻¹⁷²(or OutF⁶⁵⁻¹⁷²)

To generate pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³-OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²) plasmid (with His-tag on OutL¹⁻²⁵⁷), the link and lock method was used for the cloning (Fig. 2.1). OutE and OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²) genes were cloned into pET-3a without tags and the Cyto-OutL gene with hexa-histidine tag was cloned into pET-14b.

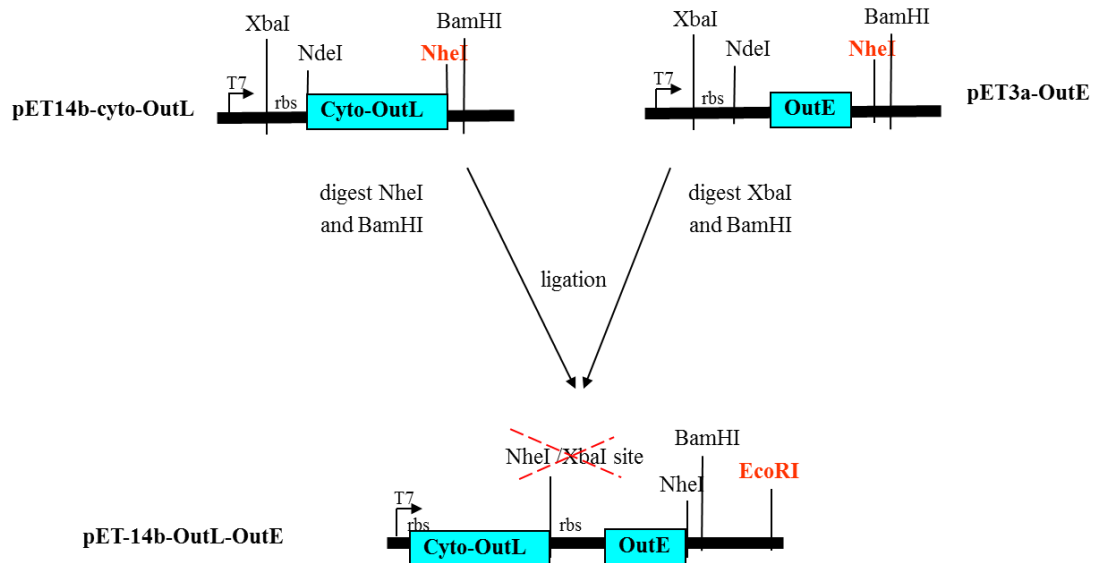


Figure 2.1 Construction of pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³ using the link and lock method.

The cartoon of construction of pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³ using link and lock strategy comprises: First, three plasmids were constructed harbouring *outL*¹⁻²⁵⁷, *outE*¹⁻⁵¹³, and *outF*⁶⁵⁻¹⁷² (or *OutF*¹⁻¹⁷²), respectively. The pET-14b vector was digested with *NdeI* and *BamHI* and then ligated with OutL¹⁻²⁵⁷ PCR fragment digested with the same enzymes to produce the pET14b-*NdeI*-OutL¹⁻²⁵⁷-*NheI*-*BamHI* plasmid.

For pET3a-OutE and pET3a-OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²), pET3a digested with *NdeI* and *BamHI* was ligated with OutE and OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²) PCR fragment digested with the same enzymes to give pET3a-*NdeI*-OutE-*NheI*-*BamHI* and pET3a-*NdeI*-OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²)-*NheI*-*BamHI* plasmids.

To link and lock, pET14b-OutL¹⁻²⁵⁷ digested with *NheI* and *BamHI* ligated with pET3a-OutE digested with *XbaI* and *BamHI* since *NheI* and *XbaI* have the same cohesive ends and could ligate into each other, but after ligation both sites are destroyed and locked. At the 3'-end of OutE, the *NheI* site introduced by PCR is still present and ready do add the

next gene OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²). Then pET14b-OutL¹⁻²⁵⁷-OutE digested with *NheI* and *BamHI* was ligated with pET3a- OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²) digested with *XbaI* and *BamHI* to yield the final pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³- OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²) plasmid.

2.2.8 Transformation

All transformations of bacterial competent cells were made according to the manufactures' instructions supplied with *E. coli* competent cells (NEB). A typical transformation was performed as follows:

1 µl of DNA (or 10ul DNA from ligation) was added and mixed into 50 µl of competent cells, incubated on ice for at least 30 minutes, heat-shocked at 42 °C for 30-45 seconds, incubated on ice for 2 minutes, then shaken for 60 minutes at 37 °C after the addition of 900µl of SOC and plated on appropriate media with antibiotic. All plates were incubated overnight at 37 °C.

2.2.9 Positive colony confirmation and sequencing

Colony PCR was used to confirm positive colonies using the T7 initiation primer and T7 terminator primer. Each positive colony was picked from the plate and grown in 10 ml LB culture overnight. The plasmids were extracted from cell pellets using QIAprep® Spin Miniprep Kit (Qiagen). The DNA sequence was confirmed by Sanger sequencing (Eurofins).

2.2.10 Making a recombinant construct

Figure 2.1 shows the whole procedure to make a new construct.

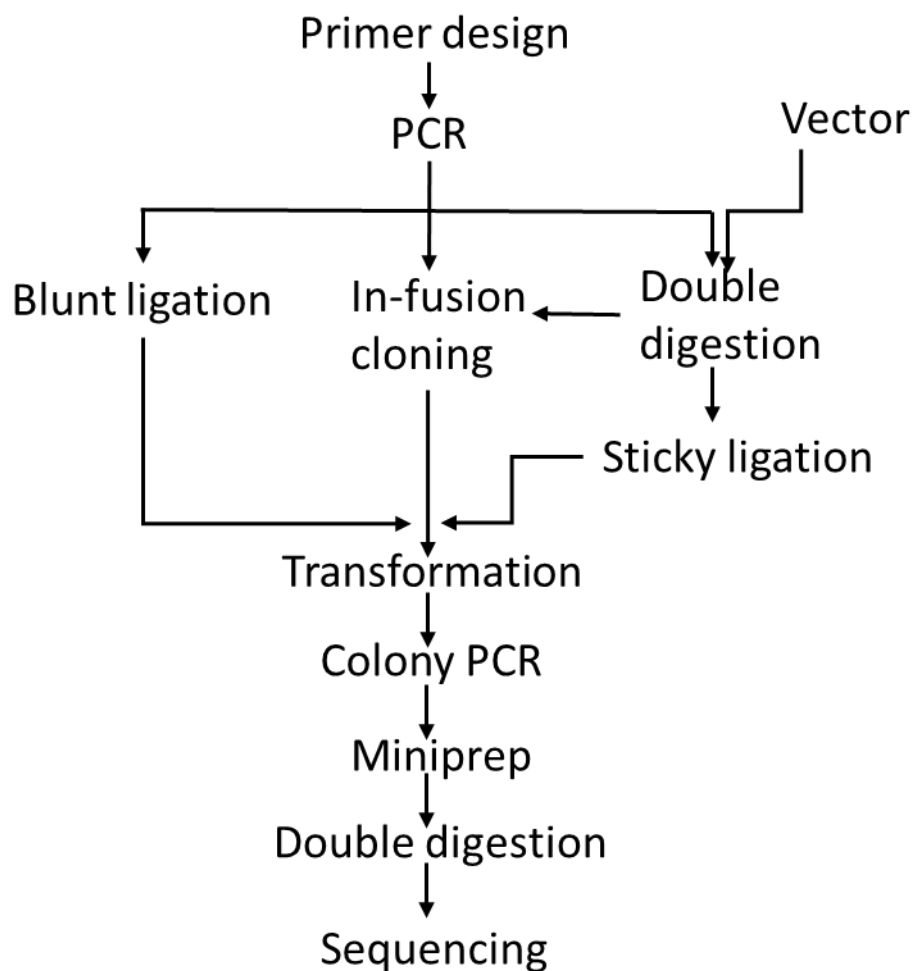


Figure 2.2 A summary of steps for making a recombinant construct

2.3 Protein purification techniques

2.3.1 Protein over-expression

Protein expression was tested using different conditions before large scale protein production. *E. coli* BL21(DE3) competent cells (NEB) and C41 (Lucigen) were selected, which were transformed according the manufacturer's protocol. One colony from an overnight plate was picked to inoculate 10ml of LB media containing the appropriate antibiotics. The culture grown overnight at 30 °C was used to initiate a large-scale culture.

1ml of starter culture was added to 1L of LB media or TB media with appropriate antibiotics. The culture was grown at 37 °C and induced with 1mM IPTG or 200µg/L AHT when the optical density (OD₆₀₀) reached 0.6-0.8. Cells were harvested by centrifugation at 7,000g for 15 minutes post induction at 18 °C for 16 hours. The pellets were either stored at -20°C or broken upon resuspension. Samples of pre- and post-induced cells were stored at -20°C for subsequent analysis by SDS-PAGE.

Figure 2.2 shows each step for over-expression and cell pellet preparation.

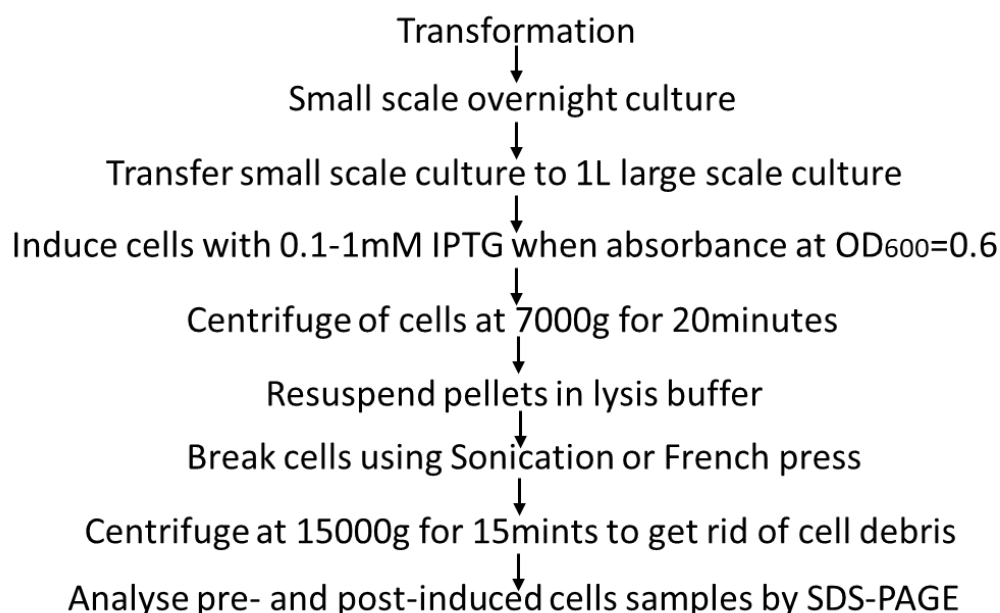


Figure 2.3 Summaries the step for over-expression of proteins

2.3.2 Bacterial lysis

All bacteria lysis was done at 4°C. Bacteria pellets were fully suspended in 10 ml/g (lysis buffer /pellet mass) lysis buffer according to individual purification protocol and supplemented with 0.2mg/ml lysozyme (Novagen) to aid lysis and reduce lysate viscosity. Cells were lysed either by sonication (VibraCell) on ice at 60% of power with 3 bursts of

10s and 30s cooling in between or by Emulsiflex Avestin C3 at 1200psi for 3 rounds. The soluble protein fraction was obtained by centrifugation at 15,000g for 15 minutes to remove insoluble protein. Samples of lysed cells, soluble protein fraction and insoluble pellet were again used for SDS-PAGE analysis.

2.3.3 Membrane-fraction preparation

For membrane protein purification, five essential steps are required: 1. Harvest cells and lyse, 2. Isolate of the membrane fraction, 3. Solubilisation of the membrane fraction, 4. Removal of non-soluble membrane fraction, 5 Chromatography. The first step is the same as used for soluble proteins. Cells were pelleted by centrifugation at 6000g for 20 minutes at 4 °C and then re-suspended in 20ml lysis buffer (20mM Tris-HCl pH8.0, 200mM NaCl, 1mM DTT, 1mM EDTA, 1mM lysozyme, 1mM PMSF and one tablet of EDTA-free protease inhibitor, Thermo Fisher Scientific). After 30 minutes' incubation on ice, cells were broken by sonication or French press. Cell debris was removed by low-speed centrifugation at 10,000g for 10 minutes at 4°C. Membranes were pelleted by ultracentrifugation (BECKMAN COULTER FUGE Optimatm L100XP Ultracentrifuge) at 100,000g for 1h at 4°C. The pellet was re-suspended and homogenized (20ml glass homogenizer) in lysis buffer (20mM Tris-HCl, pH 8.0, 200mM NaCl, 1mM DTT, 1mM PMSF) and ultracentrifuged again. Finally, the membrane pellet was re-suspended in 20mM Tris-HCl, pH 8.0, 200mM NaCl 10% glycerol, 1mM DTT, EDTA-free protease inhibitor. Aliquots were frozen in liquid nitrogen and stored at -80°C until use.

2.3.4 His tag purification

All proteins expressed as hexa-histidine tagged fusion proteins were purified as follows: the cell pellet was first lysed in Ni-column binding buffer (20mM Tris, 500mM NaCl, 20mM imidazole, pH8.0). The soluble protein fraction was then separated by

centrifugation at 15,000g for 15 minutes and loaded onto PD-10 column (GE Healthcare) packed with 1-2ml nickel beads. The column is pre-equilibrated in Ni-column binding buffer. Non-specifically binding proteins were eliminated by washing with 20 column volumes (CV) of Ni-column washing buffer (20mM Tris, 500mM NaCl, up to 70 mM imidazole, pH 8.0) until no more non-specifically bound protein was observed in the flow through (tested using Bradford solution).

For the membrane proteins, after solubilisation with 1% DDM, proteins were 2-fold diluted with Ni-column binding buffer including 0.03% DDM before being loaded onto the Ni-column.

2.3.5 Strep-tag purification

The frozen membranes were thawed on ice and solubilized for 1h at 4°C under gentle agitation in 1% DDM, 10mM LDAO, 20mM Tris-HCl, pH 7.4, 150mM NaCl 10% glycerol, 1mM DTT, 1mM PMSF, EDTA-free protease inhibitor. After ultracentrifugation (100,000 g, 1h at 4 °C) the supernatant was diluted 2-fold with 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.03%DDM (washing buffer). Then the supernatant bound for 2 h at 4 °C to Strep-Tactin® Sepharose® (IBA Lifesciences) or applied to the StrepTrap HP (GE Healthcare Life Sciences) using a flow rate 0.2ml/min. The column and the beads are pre-equilibrated in binding buffer. The beads were then loaded onto a spin column, washed with 5CV 20mM Tris-HCl, 200mM NaCl, 0.03% DDM, 1mM LDAO, 10% glycerol, 1mM TCEP, pH 7.4 at 4°C, and elute with the same buffer with 2.5mM desthiobiotin (IBA). Aliquot of each elute fraction were checked using SDS-PAGE.

2.3.6 GST tag purification

Cells from a 1L culture were harvested by centrifugation at 10,000g for 20min then resuspended in 20ml phosphate buffered saline (PBS) (containing 0.01M phosphate buffer, 0.0027M KCl and 0.137M NaCl, pH 7.4) and lysed on ice by sonication as described before. Cell debris was removed by centrifugation at 15,000g for 20min and the supernatant was incubated with 2.5ml of Glutathione Sepharose beads (G.E. Healthcare) for 2-6h. The glutathione sepharose was then collected in a PD-10 column and washed with 50ml of PBS.

On column cleavage of the GST tag required incubate Glutathione sepharose with PreScission protease buffer (containing 50mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.0) addition of 10 μ L of PreScission protease (G.E. Healthcare). Cleavage was done overnight at 4°C. Eluate containing proteins were then collected and analysed by SDS-PAGE.

2.3.7 Gel filtration/size exclusion chromatography

The eluate from affinity chromatography, described above, was concentrated to 1-2ml and loaded on to a Superdex prep grade S75/200 HR 10/30 column or Superose 6 10/30 GL (GE Healthcare) using ÄKTA purifier system (GE Healthcare). The column was pre-equilibrated with buffer and run at up to 0.5ml/min with 0.1-1 ml fractions collected. Fractions were checked using SDS-PAGE or the electron microscope and appropriate peaks were pooled, concentrated and stored at -80°C. Figure 2.3 shows different methods taken to purify different tag fused proteins.

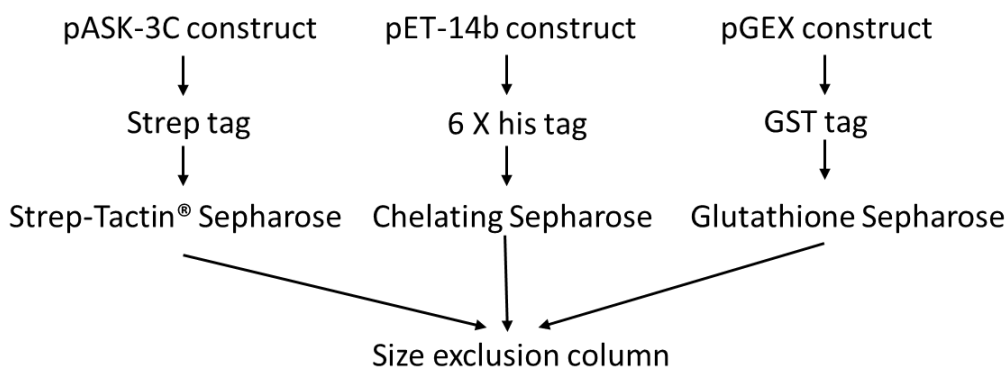


Figure 2.4 Columns involved in purifying proteins with different affinity tags.

Gel filtration was also used to determine the characteristics of the protein as it can usefully indicate the molecular mass of the protein and confirm the formation of a protein complex. The Superdex 75 and Superdex 200 were used to determine the molecular samples from calibration curves.

1. A fresh, filtered solution of Blue Dextran 2000 was used to determine the void volume (V_0).
2. Prepared the selected calibration references (BSA and lysozyme) in the running buffer (at concentrations recommended by the manufacturer) and applied the calibration solution to the column in a volume of 500 μ l.
3. Elution volumes (V_e) for the references were determined by measuring the volume of the eluent from the point of application to the centre of the elution peak.
4. K_{av} values were calculated for the references and used to prepare a calibration curve of K_{av} /logarithm of their molecular weights, as follows: $K_{av} = (V_e - V_0) / (V_t - V_0)$ where V_e = elution volume for the protein; V_0 = column void volume (elution volume for Blue Dextran 2000); V_t = total bed volume. The K_{av} value was then plotted against the corresponding logarithmic molecular weight for

- each standard protein and use Microsoft Excel to calculate the regression line.
5. Apply the sample in a volume of 500 μ l of the total column volume (V_t) and determine the elution volume (V_e) of the molecule of protein sample.
 6. Calculate the K_{av} for protein sample and determine its molecular weight from the calibration curve.

2.3.8 Differential centrifugation

Differential centrifugation is used to separate organelles from cells or particles based on their different mass or density. The larger and denser particles will sediment faster. At first, different parts of the cells are pelleted and removed by increasing the centrifuge force (Fig. 2.4). Then protein purification can be achieved through sucrose gradient centrifugation (equilibrium centrifuge) and collecting the desired layer. |

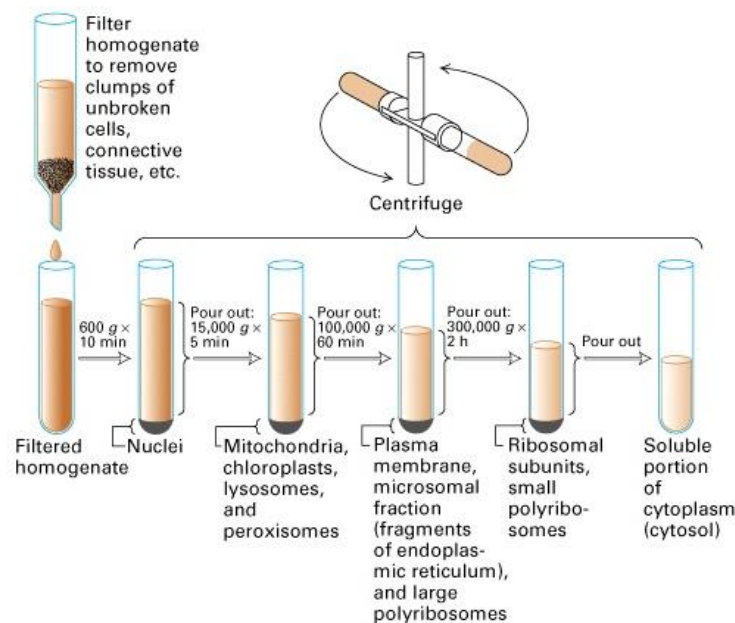


Figure 2.5 Cell fractionation by differential centrifugation

Organelles from cells were separated based on their different size or densities using increasing centrifugal force(2000).

Protocol for the preparation of a 10%-50% sucrose gradient:

1. Sucrose gradients were prepared at least a day before use. First step, wash out centrifuge tubes (Beckman No.344061).
2. Prepare 10% and 50% sucrose in 20mM Tris, 150mM NaCl, 0.03%DDM, pH 7.4.
3. Add 2.2ml of 50% sucrose. Place in -80°C to freeze.
4. Mix 1.65ml of 50% sucrose plus 0.55ml of 10% sucrose and layer over frozen sucrose. Place in -80°C to freeze.
5. Mix 1.1ml of 50% sucrose plus 1.1ml of 10% sucrose and layer over frozen sucrose. Place in -80°C to freeze
6. Mix 0.55ml of 50% sucrose plus 1.65ml of 10% sucrose and layer over frozen sucrose. Place in -80°C to freeze.
6. Add 2.2ml of 10% sucrose over frozen sucrose. Place in - 80°C to freeze
8. Seal with parafilm and store at -80°C until use. The night before use, thaw gradients in cold room for 1-2 hours.

After the sucrose gradient has thawed, layer the membrane fraction on to the sucrose gradient. Use a thin needle to trickle the solution down the side of the tube without disturbing the sucrose gradient.

1. Precool the centrifuge at 4°C and keep the rotor (SW32) in the cold room before centrifugation.
2. Load the sample into the centrifuge.
3. Set the centrifuge at 16,000g for 16 hours at 4°C with slow acceleration and deceleration (set acceleration and deceleration both to 7).

After centrifugation, label the tube with small scale using a ruler and collected each layer about 200-300µl volume using a plain tip needle. Check each fraction of the sample using SDS-PAGE, size exclusion chromatography, and/or transmission electron microscopy.

2.3.9 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), is a technique used to separate proteins according to their electrophoretic mobility, length, conformation and charge of molecule. It was used to track the expression and purification of proteins.

According to the size of the protein to be resolved on the gel, 12% and 15% polyacrylamide gels were used. Prepare separating gel and stacking gel in Table 2.18. The gels are made by adding TEMED just before pouring the gel mixture into the Bio-Rad Gel cassette (Bio-Rad). The running gel was poured into the cast first and followed by the stacking gel and a comb for 10 or 15 wells which was inserted before the gel was left to set. The cassette was filled to a level which allowed the comb to be inserted with 5mm between the wells and the top of the running gel.

Table 2.21 Recipe for SDS-PAGE.

Separating gel are used at 12% or 15% and stacking gel is 4%.

Solutions(ml)	4% stacking gel -(1.5ml)	12% running gel (5ml)	15% running gel
H ₂ O	1.02	1.65	4.4
Acrylamide/Bis-acrylamide (30%/0.8% w/v)	0.255	2	10
0.5 M Tris-HCl, pH 6.8	0.1875	-	-
0.5 M Tris-HCl, pH 8.8	-	1.25	5.2
10% SDS	0.015	0.05	0.2
10% (w/v) ammonium persulfate (APS)	0.015	0.05	0.2
TEMED	0.0015	0.002	0.02

Protein samples were prepared by mixing equal amount of 2 x SDS protein loading buffer (100mM Tris, 200mM DTT, 4% SDS, 0.2% bromophenol blue, 20% glycerol) before heated at 100 °C for 5 min. Gels were run in SDS running buffer (25mM Tris, 192mM

Glycine, 0.1% SDS) at 200V till the protein loading buffer dye ran out in a Bio-Rad Mini-Gel cell on a Bio-Rad powerPac 300 (Bio-Rad). Remove SDS-PAGE gel from glass and rinse once in ddH₂O before gel was stained in coomassie staining buffer (0.1% coomassie blue, 30% methanol, 10% acetic acid) for 20 minutes on shaker and destained in detaining buffer (10% acetic acid, 30% methanol) until the protein bands were clearly observed.

2.3.10 Native gel

The native gel was made using a similar method as SDS-PAGE except that SDS was replaced with water and the gel was running in the running buffer without SDS. The protein sample was mixed with 2x loading buffer (62.5 mM Tris-HCl, pH 6.8, 25% glycerol and 1% Bromophenol Blue). The gel was run at 4°C for 2 hours at 100V to avoid over-heating the gel. After that, the gel was stained using the same method as described for SDS-PAGE.

2.3.11 Western blot

To confirm the identity of the protein a Western blot was used. The protocol is as follows:

1. A semi-dry electrophoresis blot apparatus was used to transfer protein from gel to nitrocellulose filter membrane as follows:
 1. The nitrocellulose membrane and two extra filter papers were soaked with transfer buffer.
 2. The gel and membrane were sandwiched between the filter papers. The membrane was laid upon the gel. Each upper layer must be smaller than the one below to prevent a short circuit. The transfer was carried out at 15V for 15-30 minutes depending on the molecular mass of the protein.
 3. After transfer of the protein to the membrane, the membrane was blocked in 5%

semi-milk/ Tris-buffered saline (TBS) for 1h at room temperature or 4°C overnight. Blocking non-specific binding sites in this way can reduce the background and lead to a far clearer result.

4. Rinse the blot in TBS for 1-2min.
5. Incubate with primary antibody (mouse anti-His antibody) for 2 hours or overnight at 4°C.
6. Wash 3 times with TBS (five min each wash)
7. Incubate with secondary antibody (goat anti-mouse IgG, the second antibody was covalently coupled to alkaline phosphate) for 50 minutes at 4°C.
8. Wash three times with TBS (five min each wash)
9. Add alkaline phosphate. An AP detection kit from Novagen was used. Each tablet was dissolved in 10ml water and incubated with membrane until the band or bands appeared.

2.3.12 Determination of protein concentration

Protein concentration was determined by spectroscopic analysis using a UV-Vis spectrophotometer (HITACHI). Protein absorbance at 280 nm was measured.

Concentration was determined according to:

$$A_{280} = \epsilon \cdot c \cdot l$$

Where c is the molar concentration (M^{-1}), A_{280} is the protein absorbance value at 280nm, ϵ is the molar extinction coefficient (calculated using ExPASy ProtPrm online software) and l is the path-length of the cuvette.

2.4 Biochemical and biophysical techniques

2.4.1 Dynamic light scattering

Dynamic light scattering (DLS) is a fast, low resolution method for determination the protein particle size and molecular weight in solution. It is also used to screen the buffer condition for crystal growth.

In this method, the speed at which the particles are diffusing due to Brownian motion is measured. This is done by measuring the rate at which the intensity of the scattered light fluctuates when detected using a suitable optical arrangement. The rate of the scattering intensity fluctuation occurring depends on the size of particles. The small particles cause the intensity to fluctuate more rapidly than the large ones. The size of a particle is calculated from diffusion coefficient using the Stokes-Einstein equation:

$$D = k_B T / 6 \pi \eta r$$

where r is the radius of the protein, k_B is the Boltzmann constant, T is the temperature in Kelvin degrees and η is the viscosity of the solvent. In this study proteins' molecular size is calculated from the radius using Rayleigh sphere module.

Polydispersity (Pd)

Pd represents the particle size distribution width of the protein. As a rule of thumb, if the average % Pd is under 20%, it indicates the protein is monodispersing while above 20% then it indicates the protein is polydisperse. Thus, this value is an indicator for ease of crystallization (20% or less suggests the sample may be more easily crystallized as the protein exists in single homogeneous form).

Monomodal/ multi-modal distributions

The size distribution tends to be multi-modal when the protein is impurity or not stable in the buffer condition. When no impurity of proteins is detected, the buffer conditions (pH, salt and temperature) should be changed to make it more stable.

Dynamic light scattering was measured using a DynaPro Molecular sizing instrument and DYNAMICS V6 software. The quartz cuvettes (45 μ l) were washed with 1% Triton and subsequently with water and dried with compressed nitrogen air. The exterior surface of cuvette was wiped with ethanol and lens tissue to remove any dirt on the outside surface. The protein concentration used was 4mg/ml (or as stated). Each sample was passed through a 0.2 μ m filter and a minimal 20 measurements were recorded at 20°C.

2.4.2 Circular dichroism

Circular Dichroism (CD) is an excellent method to indicate the secondary structure of the protein (Greenfield 2006, Whitmore and Wallace 2008). Although it cannot give as detailed structure information as obtainable from crystallography and NMR, it can provide secondary structure information using small amounts of protein and very sensitive to changes in secondary structure and can also be used to monitor the changes in conformation. Unlike the polarised light which oscillates in one plane, the circularly polarised light has a rotating plane of oscillation which forms a helix as it travels. This helix can be clockwise or anticlockwise.

A Circular Dichroism spectrometer measures the absorption spectrum with left circularly polarised light minus that with right. Structural characteristics can cause slight difference between these two spectra. A polarised modulator was applied on a plane light provided by a xenon lamp making the plain polar light change to left or right circularly polarised.

The far UV CD of protein reflect the secondary structure, such as α -helix, β -sheet and unordered (irregular). (Fig. 2.5).

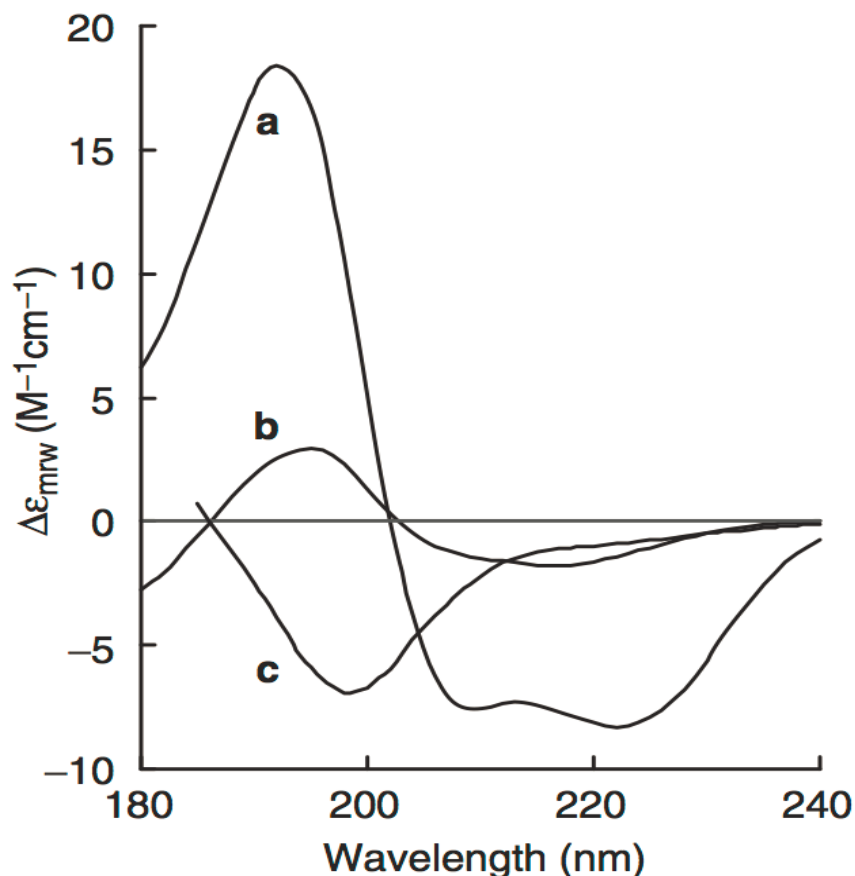


Figure 2.6 Far-UV CD reference spectra

Typical far-UV CD spectra of (A) myoglobin (all- α : 4mbn.pdb); (B) prealbumin (all- β : 2pab.pdb) (C) acid denatured staphylococcal nuclease (irregular)(Martin and Schilstra 2008).

All samples prepared for CD were at high purity level and the cuvettes were washed with 15 times of 5% Decon90 followed by 20 times of ddH₂O before totally being dried with nitrogen gas. The final protein concentration was determined using absorbance at 280 nm in 6M guanidine hydrochloride(Martin and Schilstra 2008). Protein samples used for CD were at a concentration of 50 μ M.

In this work, Far-UV CD measurements were made using a Jasco J-715 spectropolarimeter equipped with a PTC-348WI temperature controller. Spectra were recorded in 20mM Tris, 150mM NaCl (pH7.0) at 15 °C using 1mm path length fused silica cuvettes, spectra are presented as differential absorbance after baseline subtraction.

Direct CD value were converted to mean residue molar ellipticity:

$$[\theta]=100(\text{signal})/Cnl$$

$[\theta]$ =mean residue ellipticity in deg cm²dmol⁻¹; Signal=raw output in mdeg; C= protein or peptide concentration in mM; n=the number of amino acid residues; l=cell pathlength in cm.

Data was averaged and baseline correct using Chirscan software. Raw data was uploaded to DICHROWEB <http://dichroweb.cryst.bbk.ac.uk/html/home.shtml> for structural analysis after formatted. SELCON3, K2D and CONTIN were used to make a prediction of the secondary structure.

2.4.3 Thermofluor assay

Thermofluor is a high-throughput protein stability test. In this case, protein OutF⁶⁵⁻¹⁷², OutE¹⁻⁵¹³ and OutL¹⁻²³⁸ are used to test the stability of the proteins in different buffer conditions. The buffer that gives the higher melting point is providing the protein with greater stability.

Protein folding/ unfolding was monitored by the dye SYPRO-Orange (Invitrogen). The SYPRO-orange signal is highly quenched in an aqueous environment. As the protein unfolds, hydrophobic surfaces that are buried in the native protein become exposed to solvent and SYPRO-orange binds to these hydrophobic sites resulting in a signal.

Thermofluor analysis exploited a qPCR machine with 96-well PCR detection plate (AB gene AB-1100). The screened buffer conditions are shown in Table 2.22. For each screen mixture, a 25 μ l of screen mix was prepared, including 12.5 μ l screening buffer, 5 μ l 25xSypro orange (Invitrogen, prepared by mixing 5 μ l of 5000x stock with 955 μ l of water) and 7.5 μ l protein (starting with 5mg/ml and diluting if necessary). The screen was between 25-95 °C with one degree intervals. For each temperature, a 45 second incubation time was given between measurements. The emission signal was measured at 525nm.

Table 2.22 The buffer conditions screened in thermofluor assay

	100mM MES pH5.5	100mM MES pH6.0	20mM Tris-HCl pH7.0	20mM Tris-HCl pH7.5
0 mM NaCl	1	2	3	4
150 mM NaCl	5	6	7	8
300 mM NaCl	9	10	11	12

2.4.4 Full length of OutE, cytoplasmic domain of OutL and cytoplasmic domain of OutF pull down assay

Chelating Sepharose columns was self-prepared by adding the Chelating Sepharose resin to an PD-10 column. The column was equilibrated with 20mM imidazole, 150mM NaCl, 1mM TCEP, 20mM Tris-HCl, pH 8.0. The cell lysate containing the his-fusion OutL and no tag- fusion OutE and OutF was applied on the column and incubate for 30minutes at 4°C and the flow through were collected subsequently. The column then was washed with the same buffer and the his-fused OutL along with the protein bound to it as a result of pull down assay were eluted using 300mM imidazole, 150mM NaCl, 1mM tris(2-carboxyethyl) phosphine (TCEP), 20mM Tris-HCl, pH 8.0. Samples of the eluted protein, the beads before and after elution were analysed by SDS-PAGE.

2.4.5 NMR

Both Nuclear Magnetic resonance (NMR) and X-ray crystallography can be applied to the study of protein three-dimensional structures at atomic resolution. NMR spectroscopy is the only method that allows to determine three-dimensional structures of proteins in the solution. Additionally, NMR spectroscopy is a very useful method for the study of weak interactions between proteins (Vaynberg and Qin 2006) which cannot be detected by other methods like pull down assay.

NMR spectroscopy studies chemical properties by studying individual nuclei which is different from other methods.

NMR spectroscopy can be most easily applied to structure determination for proteins in the size range up to 25 kDa. Structure determination is comparatively easy for proteins in this size range. However, there are many failed examples of structure determinations of proteins due to protein aggregation and reduced solubility.

a NMR principle

The nuclei with even numbers (^{12}C and ^{14}N) may or may not rotate around a given axis. This problem can be solved by replacing with ^{13}C and ^{15}N . Both these isotopes can be incorporated into recombinant protein by using ^{13}C containing glucose and ^{15}N ammonium chloride. The nucleus with a spin is in other words a spinning particle, which is essentially an electric current in a closed circuit, well known to produce a magnetic field.

The difference between the magnetic field experienced by nuclei (which is caused by the environment of a nucleus and its type) and the applied field result in the chemical shift. Chemical shifts are valuable way of measuring interactions between proteins.

The simplest type of NMR is 1D ^1H . However, a proton spectrum of a large protein may be cloudy and hard to interpret. A 2D ^1H - ^{15}N HSQC spectrum (heteronuclear single quantum correlation) is a method used to record the coupling ^1H - ^{15}N in the peptide bond which is present in every backbone amino acid residue in a protein except the N-terminal and the proline residues. In this spectrum, there is a ^1H -chemical shift axis and a ^{15}N chemical shift axis.

b NMR sample preparation

^{15}N -labelled OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²) was expressed using the pET-14b vector in BL21(DE3), grown in M9 minimal media. 1L M9 minimal media contained Na₂HPO₄ (6g), KH₂PO₄ (3g), NaCl (0.5g), MgSO₄ (0.25g), CaCl₂ (0.015g), FeSO₄ (0.015g), thiamine (0.001g), Biotin (0.001g), $^{15}\text{NH}_4\text{Cl}$ (1g) (99% ^{15}N from Cambridge Isotope laboratories Inc.), and ampicillin (100 $\mu\text{g}/\text{ml}$). All minimal media was 0.2 μm filter sterilized and 1L induced expression cultures were grown at 16°C for 16 hours.

NMR spectra were acquired at 15°C using Bruker Avance 700 and 600 MHz spectrometers. NMR samples contained 0.05mM uniformly labelled OutF¹⁻¹⁷² (or 0.1mM OutF⁶⁵⁻¹⁷²) in 90% H₂O/ 10% D₂O containing 500mM NaCl and 20mM Tris-HCl, pH 7.0. OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²) and OutE-OutL¹⁻²⁵⁷ in low salt buffer showed a tendency to aggregate, which was attenuated in high-salt buffers. Therefore, a 500mM NaCl buffer was chosen resulting in higher signal to noise ratio from NMR spectroscopy. pH 7.0 was chosen as higher values of pH speed up the amide proton exchange rate reduce spectral resolution.

2.4.6 Peptide mass fingerprinting

The protein band was cut out of the Coomassie-stained gel and stored in a microtube containing 50µl double distilled water. The bands were sent to the Cambridge Centre for Proteomics for sequencing. The proteins were digested with trypsin followed by MALDI-TOF mass analysis. The proteins were analysed using Mascot (Matrix Science, London, UK; version 2.5.0) and Scaffold (version Scaffold_4.3.0, Proteome Software Inc., Portland, OR) was used to validate MS/MS based protein identifications.

2.4.7 Crystallization

a Crystallography theory

X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy were traditionally the major methods for solving the three-dimensional structures of proteins. Now transmission electron microscopy is ascendant as a structure-solution method as it can be used on larger complexes and allows heterogeneous samples to be used. The first and essential step of X-ray crystallography is to obtain well-diffracting protein crystals.

A chemically and conformationally homogeneous protein solution is the key point for protein crystallization. Protein crystallization involves three common stages of crystallization: nucleation, crystal growth and cessation of growth. To generate the nuclei or seeds of crystallization, a supersaturated protein is obtained using a precipitant. In a hanging-drop vapor equilibration experiment once nucleation occurs, the concentration of soluble protein decreases and the system enters the metastable zone where the crystal grows slowly (Fig. 2.6). If the protein is too concentrated, or its solubility too low due to its intrinsic characteristics or the addition of too much precipitant, then the protein will precipitate.

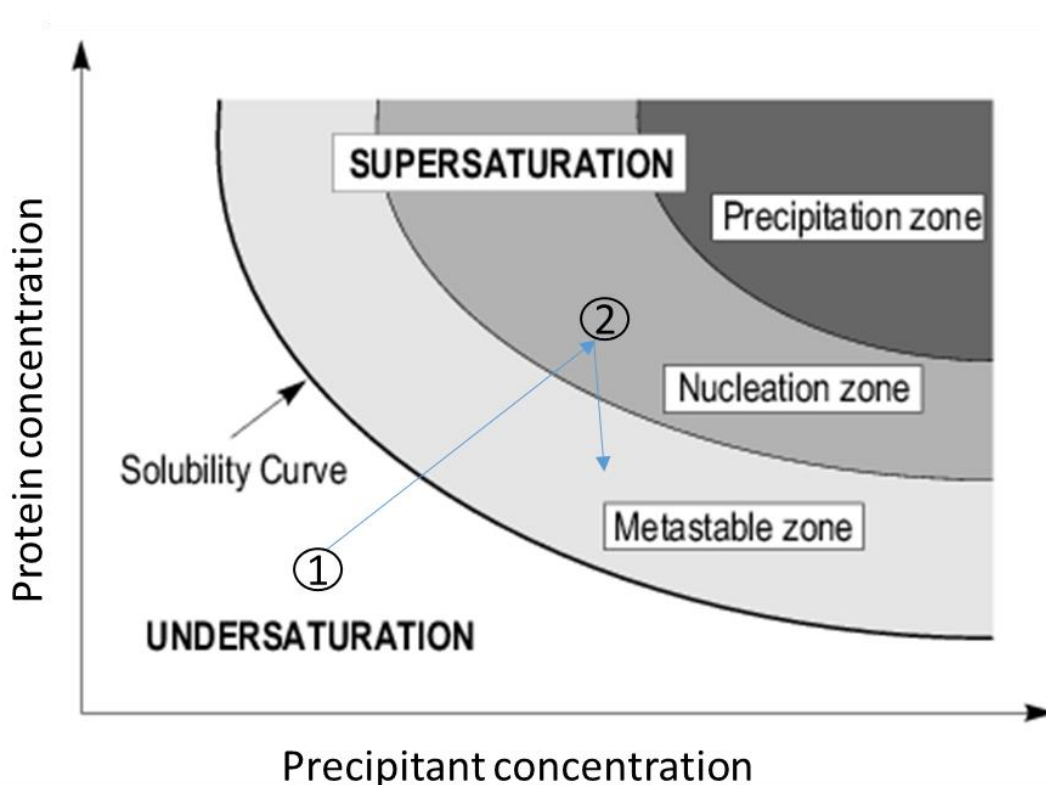


Figure 2.7 Crystallization phase diagram.

The solubility of protein is divided into undersaturation, saturation and supersaturation. The blue line indicates the crystal-growing path. (1) the starting point and (2) the formation of nuclei.

b X-ray crystallography

To be able to observe atomic level detail, electromagnetic radiation of similar wavelength to the distances to be resolved must be used. The wavelength of X-rays is around 1 \AA which is similar to the distances to be resolved in an organic structure.

Protein crystals have the ability to diffract X-rays. A protein crystal can be described as protein molecules arranged on a three-dimensional lattice. The crystal is made up of repeated unit cells related one to another by the crystal parameters: cell edges a , b , and c and α , β , γ the cell angles. The asymmetric unit is the basic building block of the unit cell, as the name implies it has no symmetry. When the symmetry operators are applied to the asymmetric unit, the unit cell is the result.

Structure factor

The X-ray beam beat the crystal and scattered X-ray waves provide information about the structure of the protein. To obtain diffraction, the scattered waves must combine in phase to give constructive interference. This satisfies Bragg's law which is $n\lambda=2d \sin\theta$, where λ is the wavelength, n is an integer and d is the distance between the two Miller planes of the crystal lattice (Fig. 2.7)(Blow 2002). The resultant spacing of the diffraction spots depends on the size of the unit cell (a , b , c) and the intensity of the spots depends on the structure of the protein.

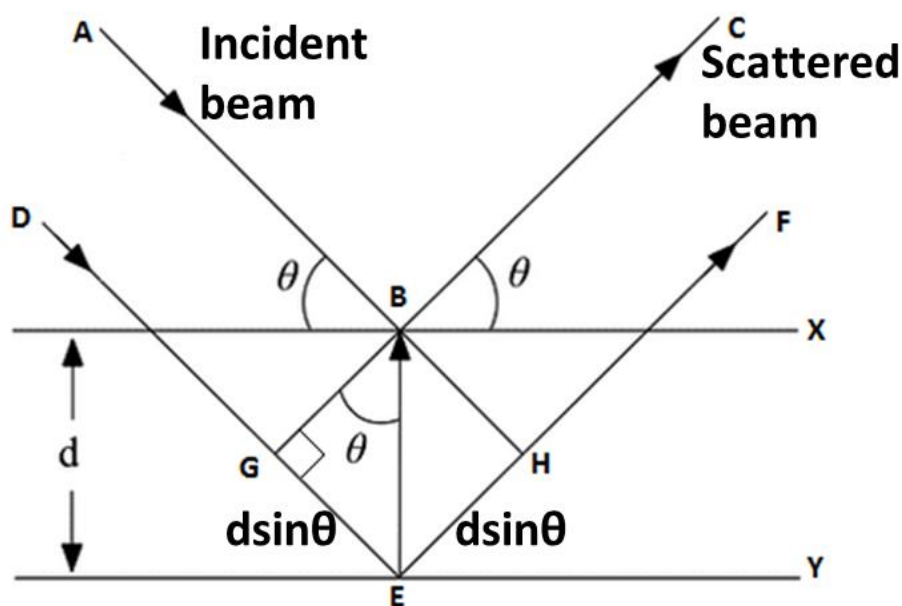


Figure 2.8 Satisfaction of Bragg's law to obtain diffraction.

The scattered waves from the crystal planes must combine in phase to give constructive interference.

To calculate the electron density at the certain point (x , y , z), it is necessary to know the amplitude and phase of the diffracted rays. The amplitude is the square root of the

intensity which can be detected by the CCD detectors. The phase cannot be measured directly and there are several ways to solve the phase problem including: use of anomalous dispersion, multiple isomorphous replacement, and molecular replacement (MR); MR is used in this work and will be discussed further.

Once the phase is known, an electron density can be generated by Fourier Transform Function:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

Where ρ is the electron density; $|F(hkl)|$ is the structure factor amplitude of reflection (h, k, l), x, y and z are the coordinates, and $\alpha(h, k, l)$ is the phase of reflections; V is the volume of the unit cell.

The resultant electron density map can then be used to build a model of the protein structure.

c Molecular Replacement

Molecular replacement was used in this project, as the structure of cytoplasmic domains of GspF from *Vibrio cholera* and *D. dadantii* are homologous, and the *Vibrio cholera* structure was deposited in the protein databank (PDB: 3C1Q). Molecular Replacement was used to determine first the orientation and then the position of the cytoplasmic domain I of OutF from *D. dadantii* in the new unit cell using the solved structure of GspF family protein as the search model. Molecular replacement works best when the search and target molecules have high sequence identity (25% or higher is helpful, but the success will also depend on the number of molecules to be found in the asymmetric unit).

In this case, the proteins have 52% sequence identity. Molecular replacement involves two steps: rotation and translation (Fig. 2.8). In the first step, the rotation function allows the correct orientation of the search molecule to be found using self-vectors from the Patterson (Vagin and Teplyakov 2000). The success is usually judged by the height of the peak and the height of this peak compared to the next peak on the hit list. The translation function uses cross-vectors in the Patterson function to discover the position of the rotated search model in the new unit cell. Therefore, the success of the translation function, and of molecular replacement overall was based on the highest correlation value.

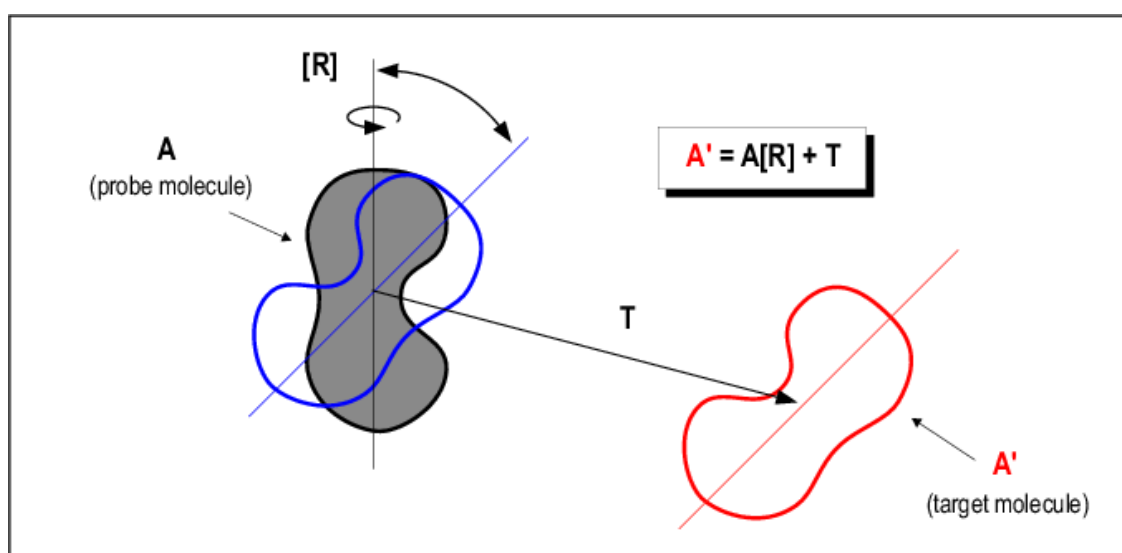


Figure 2.9 Molecular replacement method

Once the search molecule is correctly positioned in the unit cell, this model will then be refined and rebuilt until the final model has the sequence of the target molecule, good stereochemistry, and maximal agreement with the observed diffraction data as judged by the crystallographic residuals (R-factor and R-free).

d Crystal vitrification

In order to cryocool a crystal to liquid nitrogen temperature, it must be first passed through a cryoprotectant solution. Vitrification is achieved by plunging the loop with crystal into the liquid nitrogen quickly so the solution surrounding the crystal and within the crystal interstices forms a glass. Several common cryoprotectants were used (typically 10%) PEG-400, glycerol, PG (propylene glycol) and EG (Ethylene glycol) and 0.4 - 0.6 mm loops (Hampton Research) were used for transferring crystals and loading onto the goniometer head for data collection.

e Data collection and processing

Generally, two images were collected at 0 and 90 degrees and Mosflm (Battye et al. 2011) used to guide the data collection strategy. Auto-indexing was used to suggest the space group and crystal mis-setting parameters. Auto-indexing provides a list of space groups with a corresponding penalty, so the solution with highest symmetry and correspondingly low penalty was selected as hypothetical the space group. The mosaicity was also determined at this stage (ideally less than 0.5° as higher values indicate less order). Later the pattern of systematic absences and refinement statistics would confirm the correct choice of space group. Following this, the data collection strategy was calculated by running Testgen function providing the rotation angle and number of images to be collected.

Once the data were collected, they were processed using Aimless (Evans and Murshudov 2013) in CCP4 program suite (Winn et al. 2011) to scale and merge the data. 5% of the experimental data were flagged to be excluded from refinement and R-free was calculated using this set of reflections. The order of the crystal was determined by calculating Wilson B-factor (overall temperature factor, Å²). Matthews' coefficient (V_m) and solvent content

were calculated to find out the number of molecules present in the asymmetric unit. The reduced data were then used for phasing and further refinement.

Molecular replacement was used to determine the protein phases (Vagin and Teplyakov 2000). The initial model was built by the PHENIX autobuild procedure (Adams et al. 2010). Coot (Emsley and Cowtan 2004) was then used to rebuild and refine the model manually (Emsley and Cowtan 2004). The model was then used in cycles of refinement using REFMAC5 (Murshudov et al. 1997) or PHENIX refinement (Adams, Afonine et al. 2010). During refinement, the R-free and R-work values were monitored until good agreement was reached for a model with acceptable stereochemistry.

f Synchrotron sources

Data were collected from ESRF (European Synchrotron Radiation Facility, station ID 23-1) remotely.

g Structure validation

In crystallography, the R-factor (R_{work}) is a measure of the agreement between the crystallographic model and the experimental data. In other words, it is a measure of how well the refined structure predicts the observed data. It is defined by the following equation:

$$R = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

where F is the so-called structure factor and the sum extends over all the reflections measured and their calculated counterparts respectively. The structure factor is closely related to the intensity of the reflection. R_{free} computed according to the same formula as

R_{work} , but on a small of data which never used in the refinement. If the R_{free} is much more than R_{work} , it is suggesting that the crystal model is over fitted.

The stereochemistry of the structure was checked by plotting Ramachandran. Ramachandran plot is a way to visualize energetically allowed regions for backbone dihedral angles ψ against ϕ of amino acid residues in protein structure. Any distorted geometry was corrected manually using COOT.

2.4.8 Single particle analysis using transmission electron microscopy

a Principle of single particle analysis

Single particle analysis (SPA) is another method to determine the three-dimensional structure of proteins and protein complexes at near-atomic resolution apart from X-ray crystallography and NMR spectroscopy. It is a bridge between the X-ray crystallography and light microscopy. One of the greatest advantages of cryo-electron microscopy compared to conventional structural biology techniques is it can be used to analyze large proteins or complexes with different conformations, sometimes also with different compositions. It can also be used to study complexes that cannot be produced in the quantities needed for crystallography. These include many biologically important proteins, especially membrane proteins with high flexibility.

Transmission electron microscope is a microscope in which a beam of electrons is transmitted through an ultra-thin specimen interacting with the specimen as it passes through it. An image is formed from the interaction of the electrons transmitted through the specimen; the image is magnified and focused onto an imaging device or detected by a sensor such as a charge-coupled device.

TEMs are capable of imaging at a significantly higher resolution than light microscopes due to the small de Broglie wavelength of electrons. This instrument enables users to examine fine detail — even as small as a single atom. TEMs find application in cancer research, virology, materials science as well as pollution, nanotechnology and semiconductor research.

Single particle analysis involves a series of techniques for image processing to analyze images from transmission electron microscopy (TEM) (Frank 2006). Therefore, computational techniques for image processing and 3D structure reconstruction play a key role in single-particle TEM. In the single-particle studies, heterogeneity of the sample is one of the major challenges since particles can adopt different conformations. Most commonly used computational methods assume that the collected particles have homogeneous shape and quaternary-structure.

The procedure usually begins with extraction of single protein particles distributed in the raw micro-image. Contrast transfer function (CTF) parameters are used to help recover undistorted information that is buried in noise (Frank 2006). Single particles are then classified into different groups based on the similarity (2D classification), which represent various angular views of a 3D object that can then be used to estimate and iteratively reconstruct the 3D model (Fig. 2.9).

CTF is the function which modulates the amplitudes and phase of the electron diffraction pattern formed in the back focal plane of the objective lens. It can be represented as:

$$\text{CTF}(s) = \sin \gamma(s)$$

$$\gamma(s) = 2\pi [C_s \lambda^3 s^4 / 4 - \Delta z s^2 / 2],$$

where s is the spatial frequency, $\gamma(s)$ the wave aberration due to the spherical aberration C_s and the defocus Δz , and λ the wave length.

Given the equation above, it becomes apparent that the CTF would be 0 if imaging took place without spherical aberration and under in-focus conditions. This would result in the sample being invisible. Only spherical aberrations combined with defocus can lead CTF to a high contrast transfer.

Fourier shell correlation (FSC) also known as spatial frequency correlation function is employed to estimate resolution of 3D model. It measures the normalised cross-correlation coefficient between two 3D volumes over corresponding shells in Fourier space. Two 3D independent half that are always processed in parallel separately and then compared.

In this work, negative stain was used to provide contrast between the (unstained) protein particles and the stained background. Individual images of stained particles are very noisy and hard to interpret, but by combining several images of similar particles together gives an image with stronger signal to noise ratio. A three-dimensional model can then be built based on the 2D images. Cryo-electron microscopy is planned for future work as it is possible to generate reconstructions with sub-nanometer resolution and near-atomic resolution.

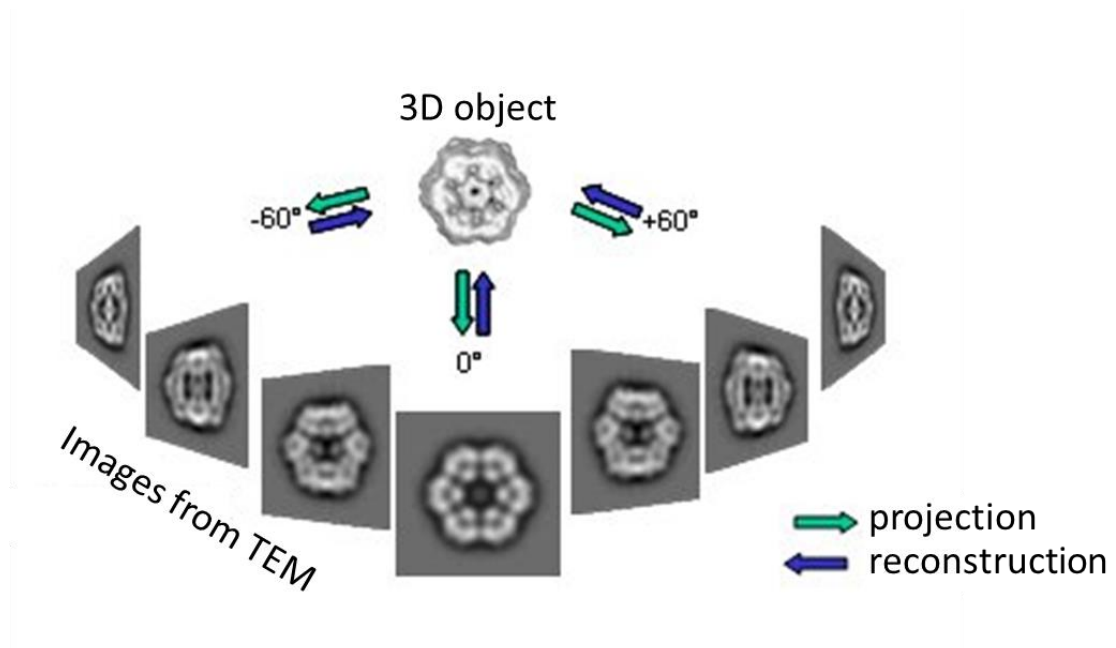


Figure 2.10 Example of a 3D reconstruction from 2D projections.

TEM images from micrographs corresponds to a 2D projection of the molecule and a 3D reconstruction is generated from these 2D projections.

b Programs used for single particle analysis

1. **Ctffind4** (Rohou and Grigorieff 2015) is a program for estimation of the contrast transfer function (CTF).
2. **SIMPLE** (Elmlund and Elmlund 2012) (Single-particle Image Processing Linux Engine). SIMPLE does *ab initio* 3D reconstruction from 2D projections only which applied in this work for generating initial 3D template. Its basis is global optimization with the use of Fourier common lines. The advance is the separation of the tasks of in-plane alignment and projection direction determination via bijective orientation search. Bijective orientation search divides the configuration space into two groups of paired parameters that are optimized separately. The first

group consists of the rotations and shifts in the plane of the projection; the second group consists of the projection directions and state assignments.

3. **RELION**(Fernandez-Leiro and Scheres 2017) (REgularised Likelihood OptimisatioN) is a software package used for particle selection, 2-D classification, 3-D reconstruction and automatic refinement of 3-D reconstructions.
4. **Chimera** is a 3-D data visualization, analysis and modelling program. In this Thesis, it was used for visualization and fitting of molecular models.

c 2% Uranyl acetate negative staining preparation

Uranyl Acetate (UA) is a mild gamma emitting radioisotope, this means that it is dangerous only in powder form if inhaled or ingested. The powder is stored in the poison cabinet. Weighing out was done in a fume hood. The stain was filtered through a 0.22 μm filter (VWR) that has been pre-rinsed with large volumes of double distilled water. The filtered stain was stored in the dark at 4°C.

d Sample preparation

1. TEM grids were made hydrophilic using a glow discharger;
2. TEM grids were handled using the special negative pressure TEM tweezers;
3. A grid was lifted from the slide and held in the air using the tweezers;
4. The sample was applied to the grid;
 - a. 5 μl of sample was added to the grid and left on grid for 1 minute;
 - b. 5 μl of ddH₂O was added to grid to wash off sample solution; left for 5 seconds (filter paper was used to soak up the solution by lightly touching the side of the grid between steps);
 - c. 5 μl of 2% Uranyl acetate was added; left between 1 and 30 seconds;
 - d. The grid was then left to dry.

e Nanogold labelling

Method 1: T2SS_{O-M/M-His} or T2SS_{E-M/M-His} complexes were loaded onto a glow-discharged carbon coated grid (SPI). After 1 min, excess liquid was blotted, and the grid was washed with a drop of cold purification buffer (20mM Tris pH 8, 100mM NaCl, 1mM DTT, 0.1% w/v digitonin, 0.06% w/v DM NPG) containing 50mM imidazole, quickly blotted and applied on a second drop of the same buffer in the presence of 5nM 5nm nanogold beads (Nanoprobes). After 2 min, the grid was rinsed sequentially for 20 s with one drop of purification buffer, one drop of the same buffer without detergent and three drops of 2% uranyl acetate. Incubation was for 1 min. Images were collected on a JEM-1230 microscope operating at a voltage of 80 kV. Particles were selected manually using Relion (Fernandez-Leiro and Scheres 2017).

Method 2: T2SS_{O-M/M-His} or T2SS_{E-M/M-His} complexes was incubated in the purified buffer supplemented with 5nM nanogold and 50mM imidazole at 4°C for 15 minutes. Then the TEM grids were prepared same as protocol described above.

Chapter 3 Structural studies of *D. dadantii* OutF

3.1 Overview

OutF is a key component of the inner-membrane platform of the type II secretion system for which only structural information on the first cytoplasmic domain is available. The objective of the work described in this Chapter was to explore the structure of *D. dadantii* OutF using X-ray crystallography. To study the structure and function of OutF at the molecular level, purified full-length and truncated variants of OutF were produced and crystallization trials made. Despite many attempts only the first cytoplasmic domain of OutF comprising 108 residues, OutF⁶⁵⁻¹⁷², could be crystallized and using the diffraction from these crystals the structure could be solved at 2.15Å resolution. The determined structure is compared with those obtained from other species.

The evidence from cryo-EM on the related type IVa pilus machine places OutF at the centre of the inner-membrane platform responsible for assembling the pilus and for communicating between the periplasm and the cytoplasmic ATPase. Thus, the interactions between the OutF⁶⁵⁻¹⁷² and cytoplasmic domain of OutL, OutF⁶⁵⁻¹⁷² and full length of OutE, OutF⁶⁵⁻¹⁷² and cytoplasmic domain of OutL-full length of OutE complex are also studied in this Chapter.

3.2 Results

3.2.1 Bioinformatics

a Signal peptide prediction

A signal peptide, sometimes referred to as signal sequence, leader or targeting sequence, is a short N-terminal peptide located at the N-terminal of proteins that are transported across a membrane, in Gram-negative bacterial cells this would be transport across the inner-membrane. The online SignalP server was used to establish if OutF possessed a signal sequence (<http://www.cbs.dtu.dk/services/SignalP/>).

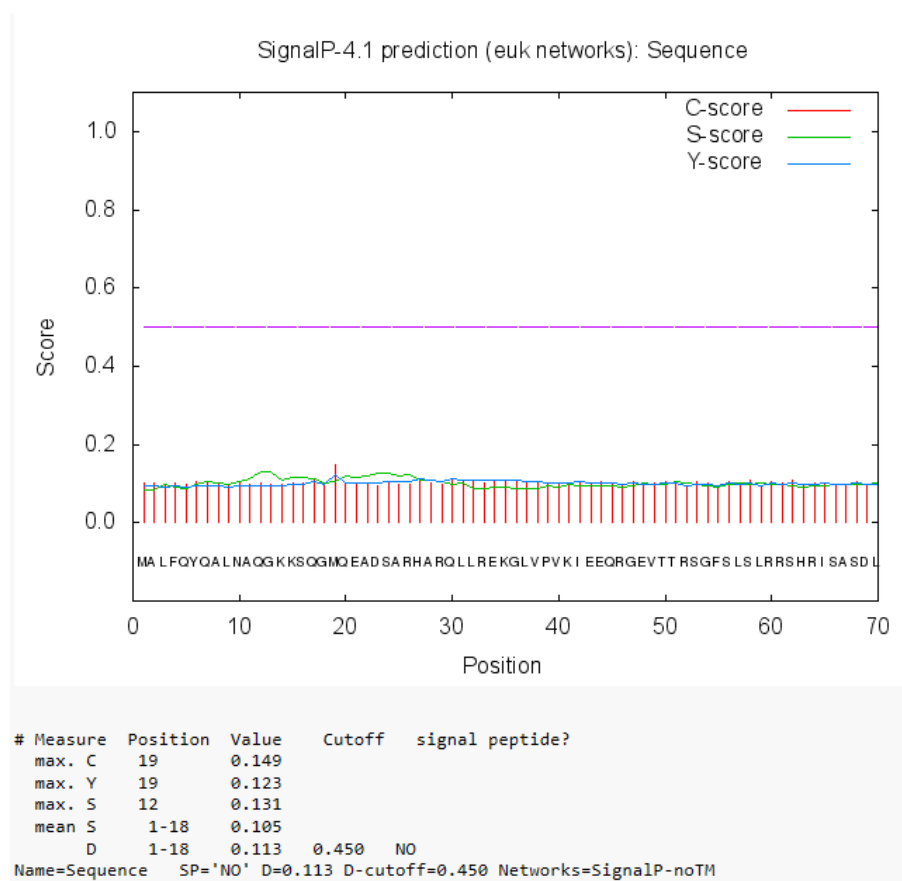


Figure 3.1 The predicted signal peptide of OutF.

Output from SignalP 4.1 suggests that *D. dadantii* OutF does not possess a signal sequence. The C-score, raw cleavage site score, is trained to be high at the position immediately after the

cleavage site, the first residue in the mature protein; S-score (signal peptide score); Y-score (combined cleavage site score)-A combination (geometric average) of the C-score and the slope of the S-score, The Y-score distinguishes between C-score peaks by choosing the one where the slope of the S-score is steepest.

I conclude that there is no discernable signal peptide in the OutF sequence (see output from SignalP in Fig. 3.1). Like most multi-spanning membrane-bound proteins which are targeted to the secretory pathway by the first transmembrane domain resembling a signal sequence but it is not cleaved due to no signal peptidase cleavage site (Goder and Spiess 2001).

b Prediction of transmembrane regions of OutF

The transmembrane regions of OutF were predicted using the online servers: <http://www.cbs.dtu.dk/services/TMHMM-2.0> and <http://www.sacs.ucsf.edu/cgi-bin/memsat.py>.

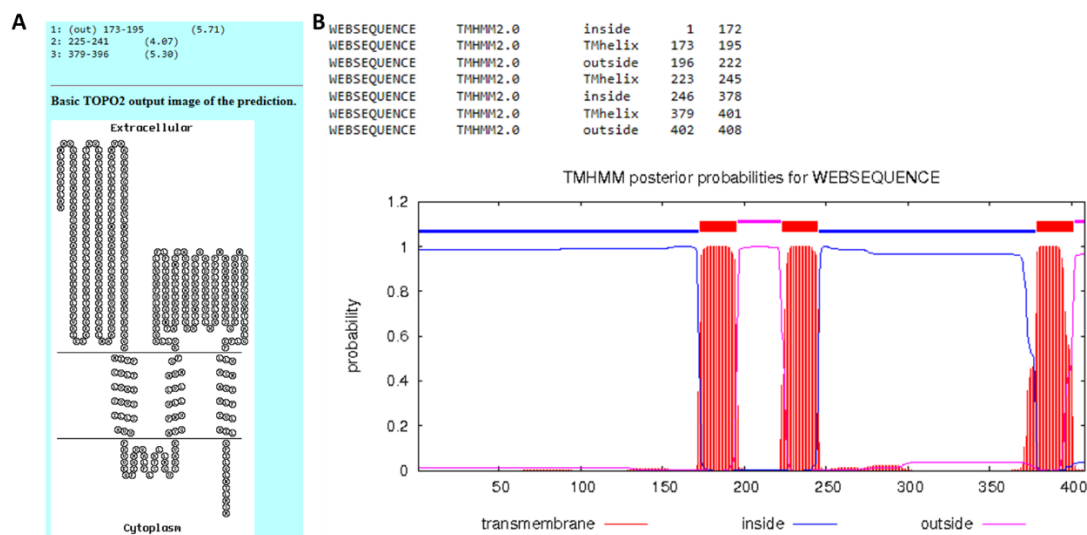


Figure 3.2 Predicted transmembrane regions of *D. dadantii* OutF.

A: predicted result from MEMSAT; B: predicted result from TMHMM.

Both predicted results from MEMSAT and TMHMM indicate there are three transmembrane domains in OutF. But for the prediction of two soluble domains with the first comprises residues 1 to 172 and the second one between the second and third transmembrane regions, two software got opposite results. Two domains are predicted as cytoplasmic domains from TMHMM while periplasmic domains from MEMSAT (Fig. 3.2).

c Prediction of intrinsically disordered regions in OutF

Intrinsic disorder within the OutF protein sequence were predicted using the online server <http://bioinf.cs.ucl.ac.uk/psipred/>.

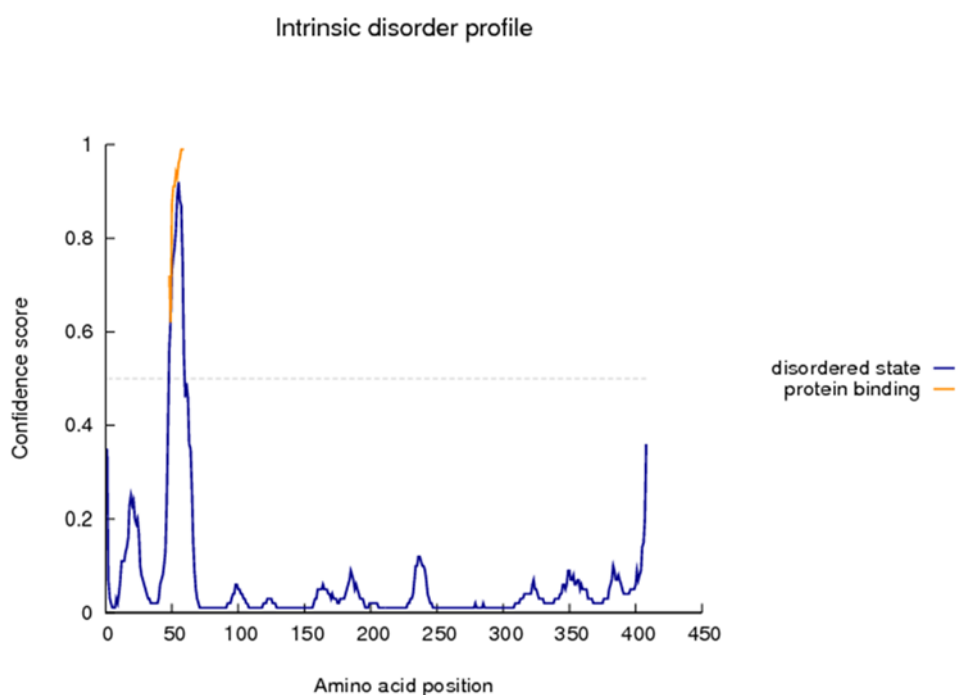


Figure 3.3 The predicted] intrinsically disordered regions of OutF.

The yellow line shows potential protein binding sites within the OutF sequence and the blue line shows the probability of the sequence being disordered.

The first 65 residues (cytoplasmic) and the extreme C-terminal residues (periplasmic) of OutF are predicted to be disordered (Fig. 3.3). Additionally, the N-terminal region, residues 1 to 65 might form a protein binding site.

3.2.2 Cloning *outF*, *outL* and *outE*

a Primer design

Based on the *outF*, *outL* and *outE* gene sequences and the vector restriction sites, primers were designed using program Primer3 as shown in Chapter 2.

b PCR products of outF, outL and outE

PCR products of *outF*, *outL* and *outE* were generated from standard three-step PCR method as described in Chapter 2.

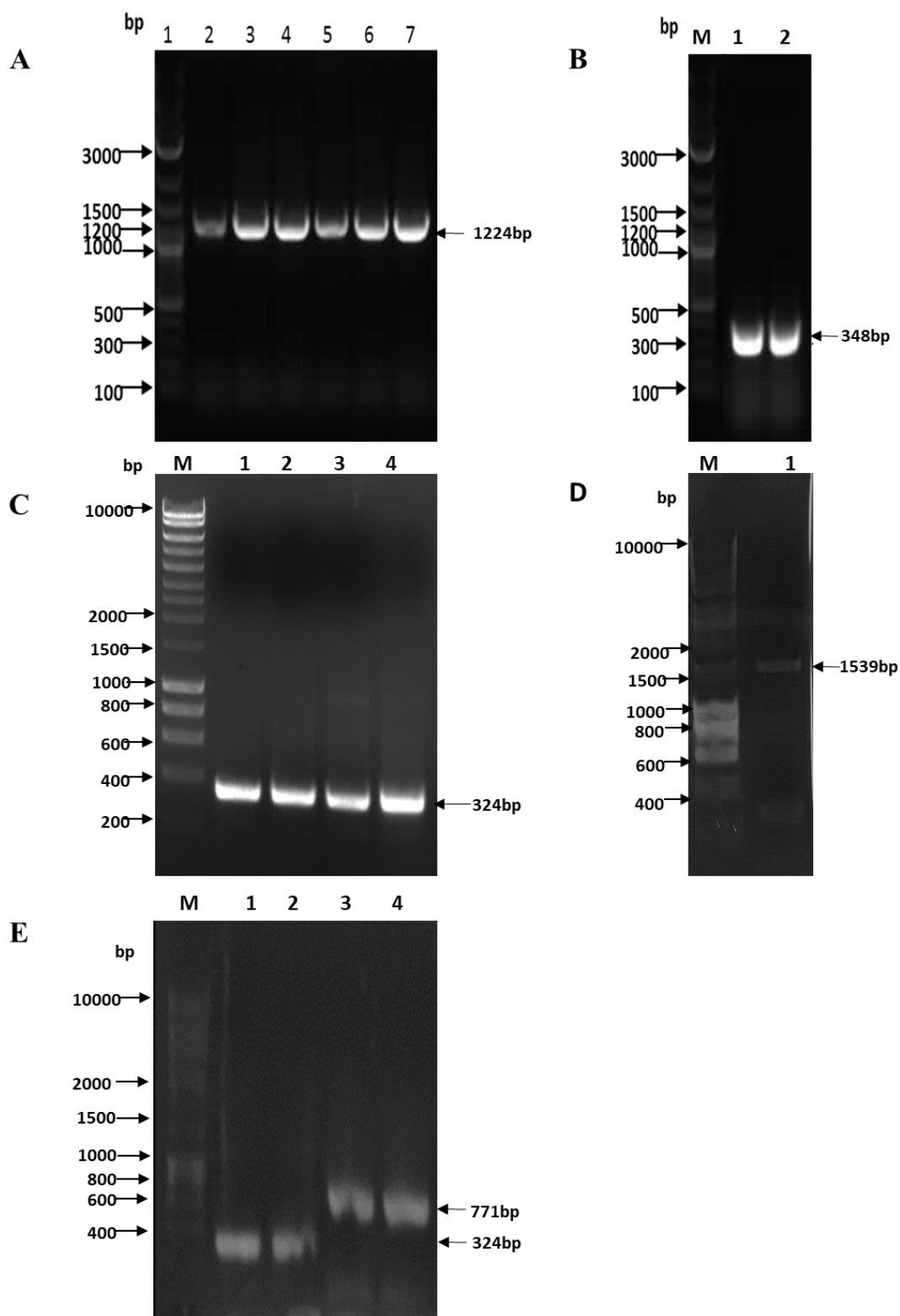


Figure 3.4 Amplification of *outF* full-length and fragments.

(A): the PCR products of full length OutF; (B) OutF⁵³⁻¹⁶⁸. (C) OutF²⁶⁷⁻³⁷⁴. (D) full length of OutE. (E) [1] and [2]: the PCR products OutF⁶⁵⁻¹⁷²; [3] and [4]: the PCR products OutL¹⁻²⁵⁷. The size of the PCR products are all consistent with that expected for the primers used.

c Construction of pET-24d-OutF

The DNA sequence coding for OutF was amplified from pGEX-6p-OutF (which had been constructed from the genome of the plant pathogen *D. dadantii* by Dr Vladimir Shevchek's group), using the primer in Table 2.1. The PCR product was digested using *NheI* and *XhoI* and ligated into the pET-24d expression vector (Novagen), to yield an expression construct with hexa-histidine tag at the C-terminus.

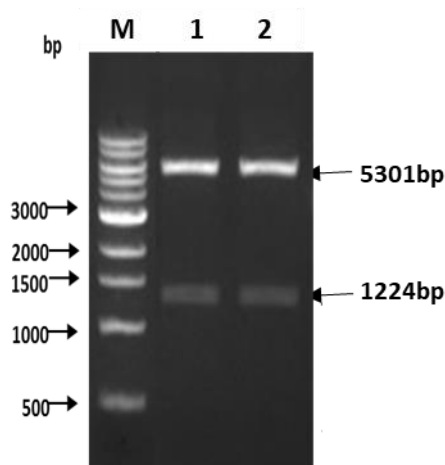


Figure 3.5 Diagnostic cleavage of pET-24d-OutF¹⁻⁴⁰⁸

Lane 1 and 2 show products following incubation of the pET-24d-outF plasmid with the restriction enzymes; two bands (1224 bp and 5301 bp) are visible at the expected positions.

Figure 3.5 shows the result of cleaving the pET-24d-OutF¹⁻⁴⁰⁸; the bands have the correct size so were sent for sequencing.

d Construction of pET-24d-OutF⁵³⁻¹⁶⁸

The DNA sequence coding for OutF⁵³⁻¹⁶⁸ was amplified from the pET-24d-OutF¹⁻⁴⁰⁸ plasmid. The ligation product was cleaved using the same restriction sites as pET-24d-OutF¹⁻⁴⁰⁸.

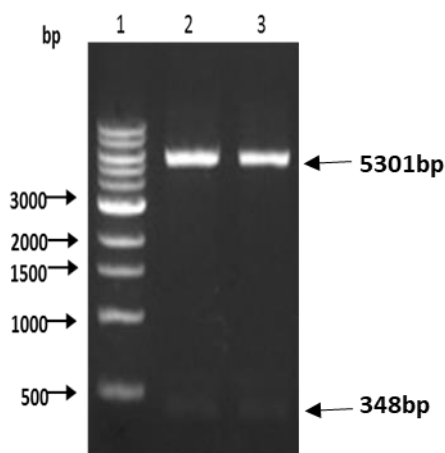


Figure 3.6 Diagnostic cleavage of pET-24d-OutF⁵³⁻¹⁶⁸.

Lanes 2 and 3 show products following incubation of the pET-24d-outF (53-168) plasmid with the restriction enzymes; the two bands of 348 bp and 5301 bp are visible at the expected masses, though the 348 bp product is faint on this reproduction of the agarose gel.

e Construction of pET-14b-OutF⁶⁵⁻¹⁷²

The DNA sequence coding for the first cytoplasmic domain of OutF (cytoplasmic domain I), OutF⁶⁵⁻¹⁷², was amplified from pGEX-6p-OutF. The PCR product was digested using *NdeI* and *BamHI* and ligated into the similarly cleaved pET-14b expression vector (Novagen), to yield an expression construct with hexa-histidine tag at the N-terminus separated from the protein sequence by a thrombin cleavage site.

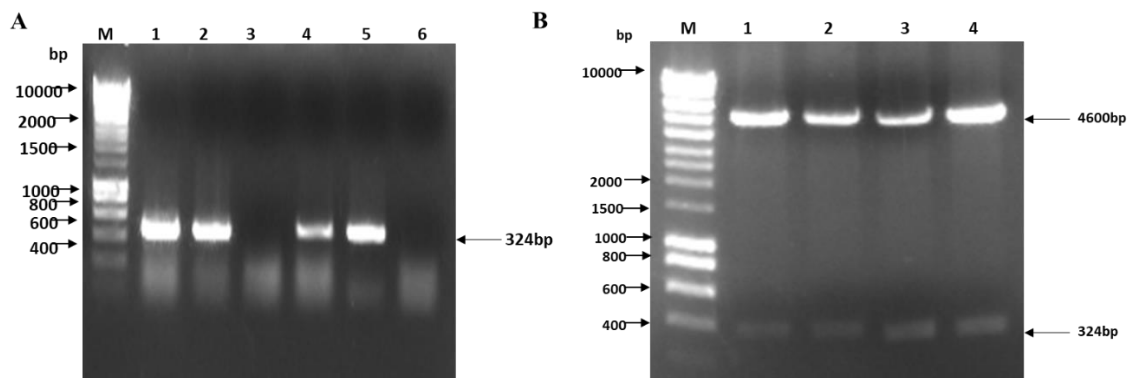


Figure 3.7 Colony PCR and enzyme digest of pET-14b-OutF⁶⁵⁻¹⁷²

(A) The colony PCR product of outF⁶⁵⁻¹⁷². PCR product from lane 1, 2, 4 and 5 are potential positive colonies. (B) The pET-14b-OutF⁶⁵⁻¹⁷² plasmids from four colonies are digested into two bands: the size of one is corresponding to that of the vector, the small band is at the expected size of OutF⁶⁵⁻¹⁷².

f Construction of pET-14b-OutF²⁶⁷⁻³⁷⁴

The DNA sequence coding for the second cytoplasmic domain (cytoplasmic domain II) OutF, OutF²⁶⁷⁻³⁷⁴, was amplified using the primers shown in Chapter 2 and digested and ligated using the same methods as OutF⁶⁵⁻¹⁷².

g Construction of pOPINS3C-OutF¹⁻⁴⁰⁸

The DNA sequence coding for full length OutF (OutF¹⁻⁴⁰⁸) was amplified using the primer in Table 2.1. The PCR product was digested using *KpnI* and *HindIII* and ligated into the pOPINS3C expression vector (Clontech), to yield an expression construct with hexahistidine tag and a fusion protein Small Ubiquitin-like Modifier (SUMO) at the N-terminus which can be cleaved with 3C protease (pOPINS3C-OutF¹⁻⁴⁰⁸).

h Construction of pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³-OutF¹⁻¹⁷²(or OutF⁶⁵⁻¹⁷²)

The pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³-OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²) plasmid (with His-tag on OutL¹⁻²⁵⁷) was created by the method of link and lock described in chapter 2.

The colony PCR and diagnostic cleavage are shown in Fig. 3.8.

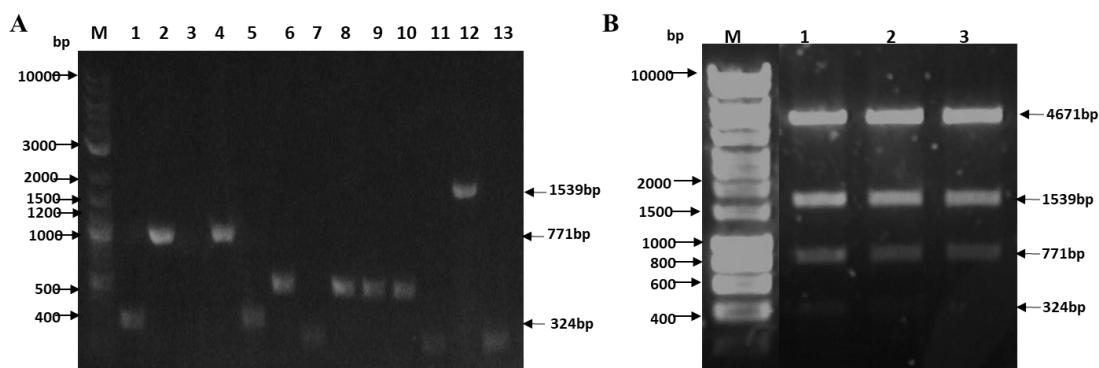


Figure 3.8 Colony PCR and enzyme digestion of pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³- OutF⁶⁵⁻¹⁷²

(A) Lanes [1] to [5] are the colony PCR product of OutL¹⁻²⁵⁷, lanes [2] and [4] are positive colonies; lanes [6] to [10] colony PCR product of OutF⁶⁵⁻¹⁷², lanes [6] and [8] to [10] are positive colonies; lanes [11] to [13] colony PCR product of OutE¹⁻⁵¹³, lanes [12] is positive colony. All the positive colony PCR product are consistent with the expected size. (B) shows products following incubation of the pET-14b-OutL¹⁻²⁵⁷-outE¹⁻⁵¹³-OutF⁶⁵⁻¹⁷² plasmid with the restriction enzymes NdeI and BamHI; four bands (4671bp, 1539bp, 771bp and 324bp) are visible at the expected positions. These four bands are corresponding to the pET-14b vector, OutE¹⁻⁵¹³, OutL¹⁻²⁵⁷ and OutF⁶⁵⁻¹⁷².

3.2.3 Expression and purification

a Expression and purification of OutF¹⁻⁴⁰⁸

pET-24d- OutF¹⁻⁴⁰⁸ was transformed into C41 *E. coli* competent cells (Lucigen) and grown in LB medium at 37°C until the absorbance of the media at 600nm reached 0.6-0.8. Expression was induced with 0.3mM IPTG for 16h at three different temperatures 16°C, 25°C and 37°C. Cells before and after induction from different temperature were collected and analysed with SDS-PAGE and western blot.

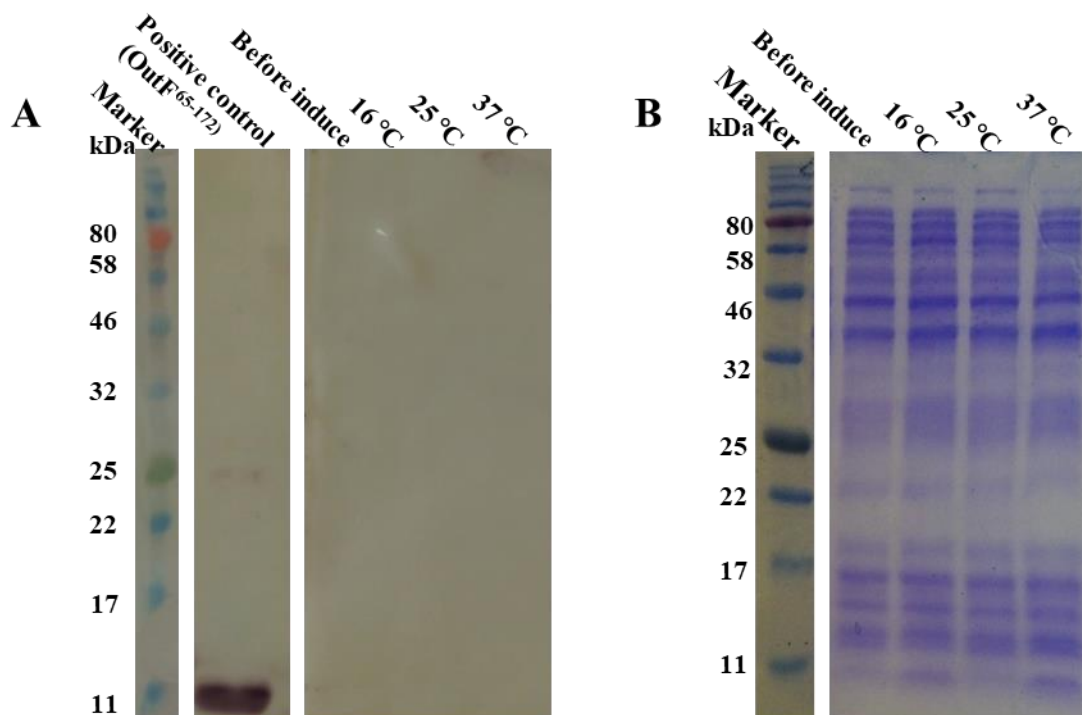


Figure 3.9 Expression test of OutF¹⁻⁴⁰⁸.

(A) Western blot and (B) SDS-PAGE. No OutF¹⁻⁴⁰⁸ was expressed after confirmed by SDS-PAGE and western blot. To help to enhance the expression and stability of OutF¹⁻⁴⁰⁸, a new construct with SUMO tag was generated.

b Expression and purification of SUMO-OutF¹⁻⁴⁰⁸

POPINS3C-OutF¹⁻⁴⁰⁸ was transformed into C41 *E. coli* competent cells and induced at different conditions as described above for pET-24d-OutF¹⁻⁴⁰⁸. After protein production was confirmed by SDS-PAGE. Large scale cells were harvested by centrifugation at 7000g for 30 min at 4 °C. The cells were disrupted by sonication and pelleted by low speed centrifugation at 10000g for 20 min at 4°C. Then the membrane fraction was purified by ultracentrifugation at 100000g for 1h at 4°C as described in Chapter 2.

The membrane fractions were suspended and homogenized in 20mM Tris-HCl, 200mM NaCl, 1% DDM, 10% glycerol, 1mM TCEP, pH 8.0 and solubilized at 4°C for 2 hours. Insoluble protein was removed by ultracentrifugation for 30min. The supernatant was then diluted 2-fold with wash buffer (20mM Tris-HCl, 200mM NaCl, 0.03% DDM, 20mM imidazole, 10% glycerol and 1mM TCEP, pH 8.0) and bound to nickel beads (0.5 ml bed volume) at 4 °C. After 2 hours, the beads were poured into a gravity flow column, washed with wash buffer (20 column volumes), and eluted with 3ml wash buffer augmented with 400mM imidazole. The eluted solution was subsequently concentrated using a Viva spin 50kDa cut-off concentrator to 2ml and loaded onto a Superdex200 10/300 GL column (GE HealthCare) pre-equilibrated with 20mM Tris-HCl, pH 8.0 200mM NaCl, 0.03% DDM, 10% glycerol and 1mM DTT.

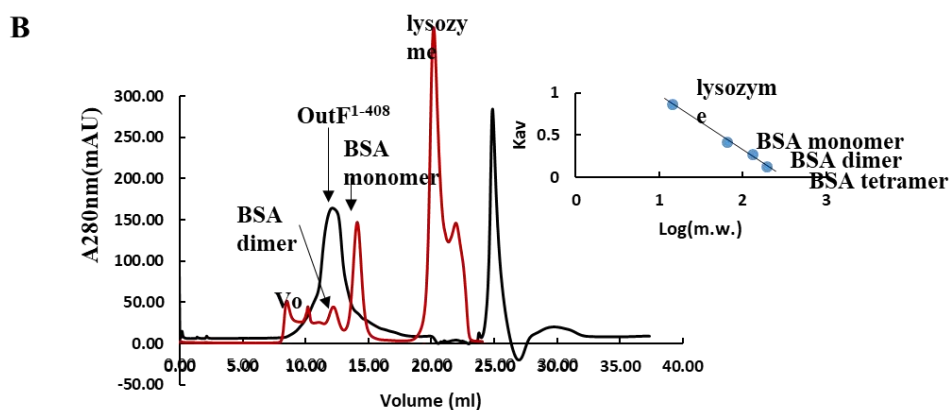
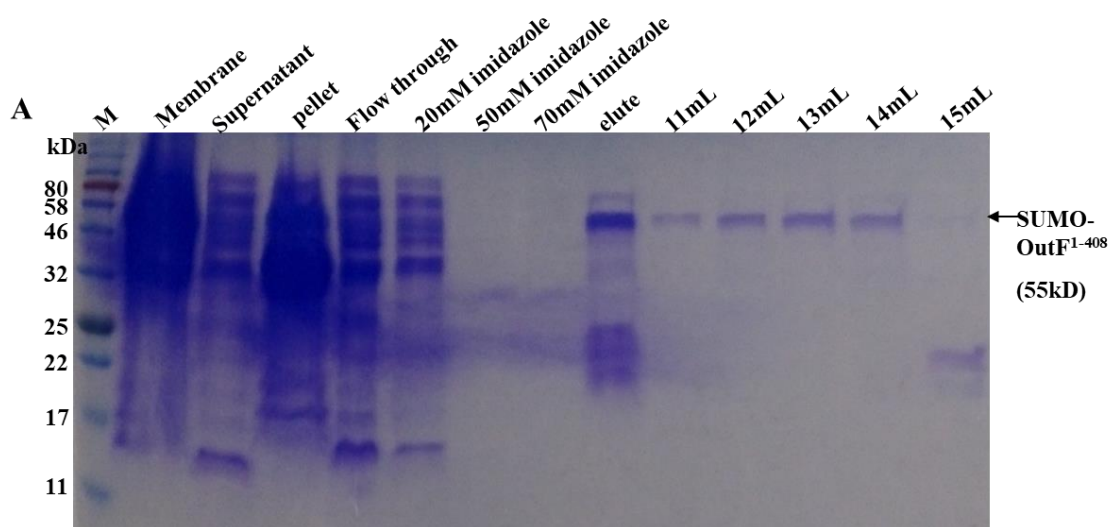


Figure 3.10 Purification of SUMO-OutF¹⁻⁴⁰⁸

(A) IMAC and SEC purification of SUMO-OutF¹⁻⁴⁰⁸. (B) SUMO-OutF¹⁻⁴⁰⁸ eluted at approximately the molecular mass of the dimer. A Superdex 200 10/300 GL size exclusion chromatography column was equilibrated with buffer (20mM Tris-HCl, 200mM NaCl, 0.03% DDM, pH 8.0) and calibrated with the following proteins used as standards: bovine serum albumin monomer (66 kDa), dimer, tetramer and lysozyme (14.4kDa). The molecular mass of OutF was calculated as 123kDa.

The expression level of SUMO fused OutF¹⁻⁴⁰⁸ increased substantially compared to OutF¹⁻⁴⁰⁸ alone indicating that the SUMO tag helped to increase protein solubility and stability. The mass of the SUMO- OutF¹⁻⁴⁰⁸ was calculated by calibrating the S200 column using several standard proteins. The results of size exclusion chromatography of OutF¹⁻⁴⁰⁸ were consistent with an OutF dimer a result confirmed later by the DLS results, these results are not consistent with measurements reported previously for PilG (Collins, Saleem et al. 2007), but the presence of the SUMO-tag could influence the oligomeric state.

c Expression and purification of OutF⁵³⁻¹⁶⁸

PET24d-outF⁵³⁻¹⁶⁸ was transformed into BL21 (DE3) *E. coli* cells (NEB) and grown in LB medium at 37°C until the absorbance of the media at 600 nm reached 0.6. Expression was then induced with 0.5mM IPTG for 16 hrs at 16°C and cells are harvested in the same way as described for OutF¹⁻⁴⁰⁸.

Protein was purified using the method described in Chapter 2. The eluted protein was concentrated using a Viva spin 10kDa cut-off concentrator to 2mL and loaded onto a Superdex75 10/300 GL column (GE HealthCare) pre-equilibrated with 20mM Tris-HCl (pH 8.0), 250mM NaCl and 1mM DTT.

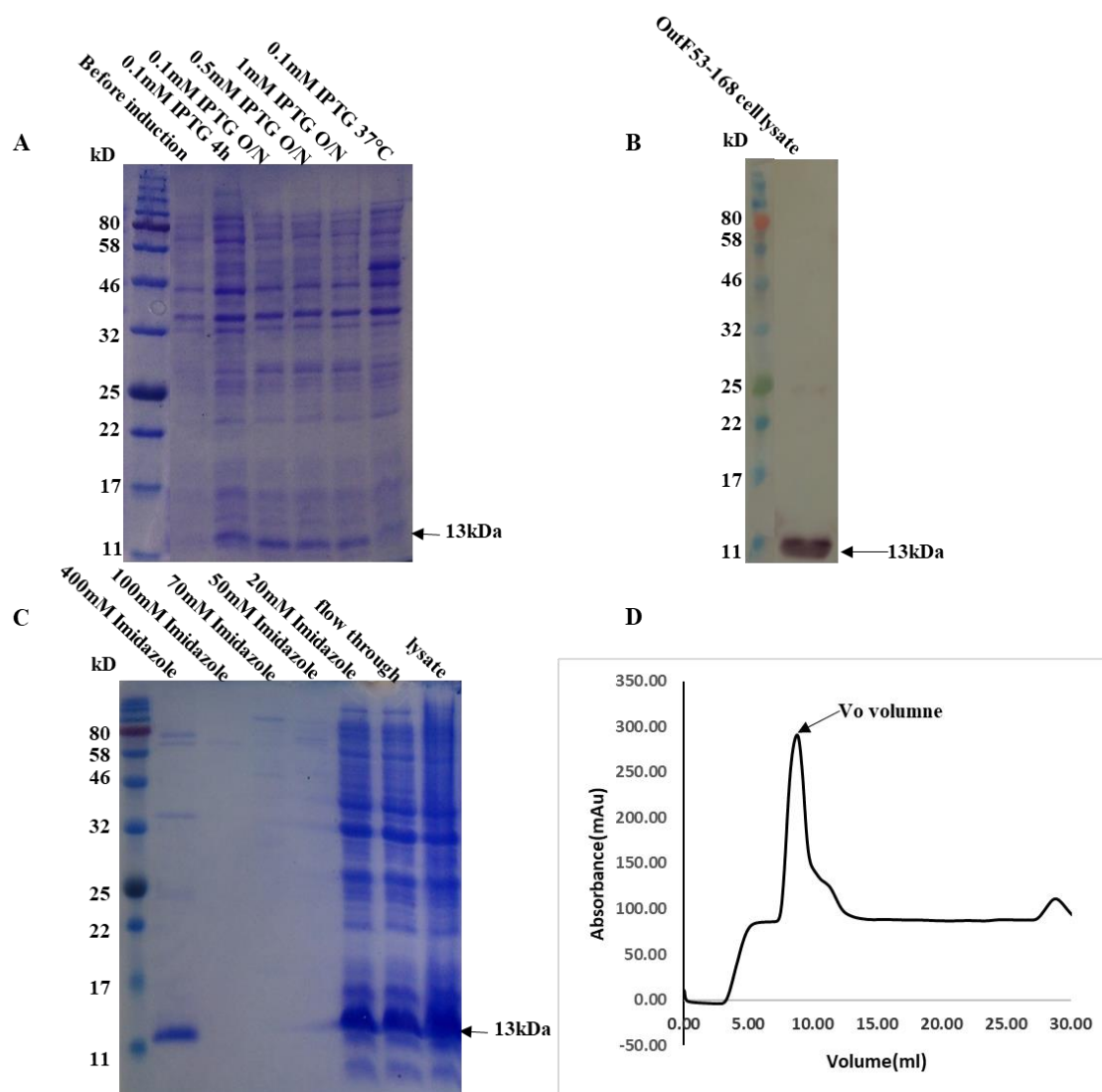


Figure 3.11 Expression and purification of OutF⁵³⁻¹⁶⁸.

(A) SDS-PAGE analysis of the expression level of OutF⁵³⁻¹⁶⁸ in different expression regimes. (B) Western blot using an anti-His tag antibody. (C) IMAC purification of OutF⁵³⁻¹⁶⁸ (the sample was washed with wash buffer plus 20, 50, 70 and 100mM imidazole sequentially and eluted with elution buffer containing 400mM imidazole). (D) Size exclusion chromatography of OutF⁵³⁻¹⁶⁸.

OutF⁵³⁻¹⁶⁸ can be expressed and purified from nickel column. But the result of the size exclusion indicated that the protein aggregated severely and almost all the protein went into the void volume. Buffer screening should be done to test the best condition for the

stability of this protein. An alternative route forward is to explore a different construct, as described in the next section.

d Expression and purification of OutF⁶⁵⁻¹⁷²

pET-14b-OutF⁶⁵⁻¹⁷² was transformed into BL21 and expressed and purified in the same method as OutF⁵³⁻¹⁶⁸.

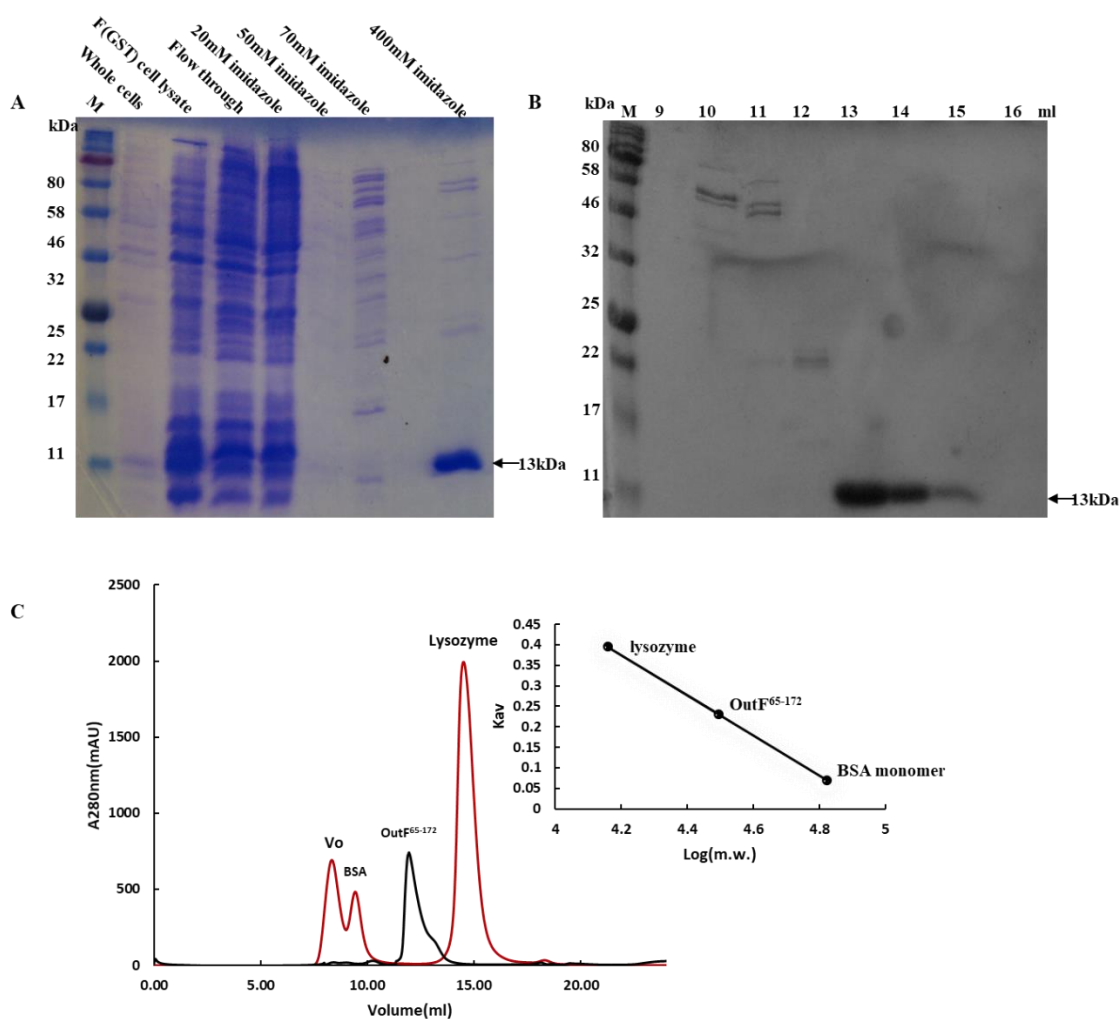


Figure 3.12 Purification of OutF⁶⁵⁻¹⁷²

A: IMAC purification of OutF⁶⁵⁻¹⁷²; B: SEC purification of OutF⁶⁵⁻¹⁷². C: Superdex 75 10/300 GL size exclusion chromatography of OutF⁶⁵⁻¹⁷² and protein elutes at the expected dimer molecular weight which is calculated using the method described in Chapter 2.

About 15mg of trials. The elution position from the Superdex 75 10/300 GL column (Fig. 3.13C) corresponds to a molecular mass of 31 kDa suggesting this domain is a dimer which is in agreement with the finding that the cytoplasmic domain of BfpE forms dimers in the yeast two-hybrid system (Crowther, Anantha et al. 2004). This measurement was confirmed by the light scattering results in a following section 3.2.6. However, in *Vibrio cholera*, gel permeation chromatography indicated that the corresponding domain, cyto1-EpsF⁵⁶⁻¹⁷¹, is monomeric in solution (Abendroth, Mitchell et al. 2009).

3.2.4 Pull down assay of pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³- OutF⁶⁵⁻¹⁷²

The pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³-OutF⁶⁵⁻¹⁷² was expressed using the same method as used with OutF⁶⁵⁻¹⁷². Non-specifically binding proteins were washed off using washing buffer with 20mM imidazole (Fig. 3.14A). Bound protein was eluted with elution buffer (20mM Tris, 150mM NaCl, 400mM imidazole, pH8.0). To further confirm the presence of OutF in the complex, four different proteins or complex were purified simultaneously as control (Fig. 3.13B).

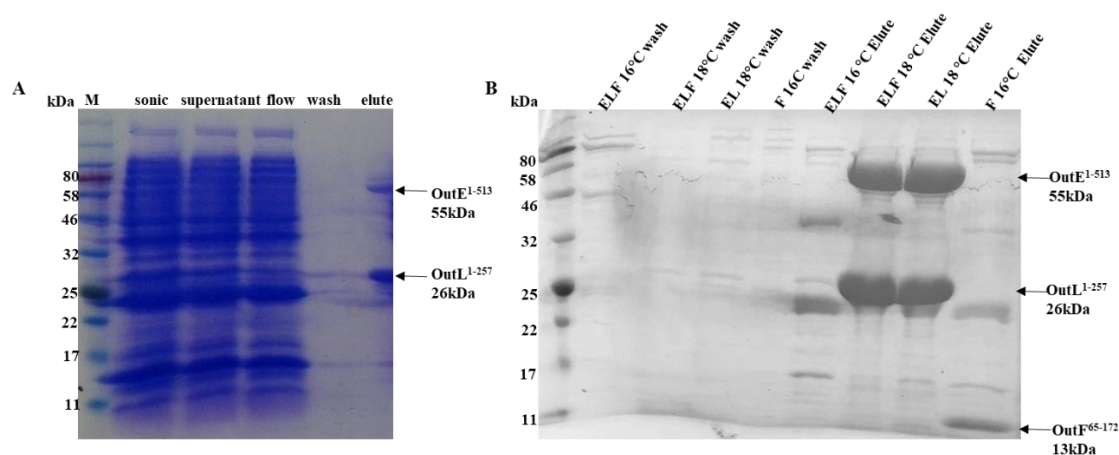


Figure 3.13 Pull down assay of pET-14b-OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³- OutF⁶⁵⁻¹⁷²

A: The OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³- OutF⁶⁵⁻¹⁷² complex was purified with his tagged OutL; B: The purification of OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³- OutF⁶⁵⁻¹⁷² complex compared to OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³, OutF⁶⁵⁻¹⁷² and different inducing temperature of OutL¹⁻²⁵⁷-OutE¹⁻⁵¹³- OutF⁶⁵⁻¹⁷² complex.

OutE¹⁻⁵¹³ can be pulled down with OutL¹⁻²⁵⁷, but not OutF⁶⁵⁻¹⁷². This result is not consistent with a previous study (Py, Loiseau et al. 2001). To further study these interactions NMR experiments were used.

3.2.5 Circular Dichroism (CD) Spectroscopy

Spectra of OutF⁶⁵⁻¹⁷² were recorded in 20mM Tris-HCl, 100mM NaCl (pH 8.0) and SUMO-OutF¹⁻⁴⁰⁸ were in 20mM Tris, 200mM NaCl, 10% glycerol and 0.03% DDM (pH 8.0) at 20 °C using 1 mm path length fused silica cuvettes. The spectra are presented as differential absorbance after baseline subtraction. Calculations employed K2D (Greenfield 2006), CONTIN and SELCON3 (Sreerama et al. 2000).

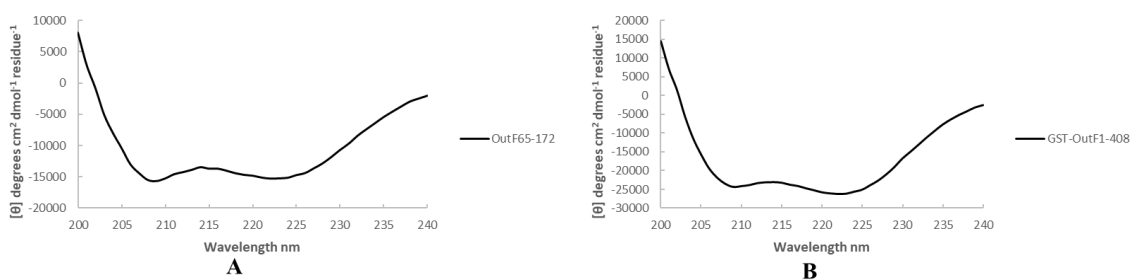


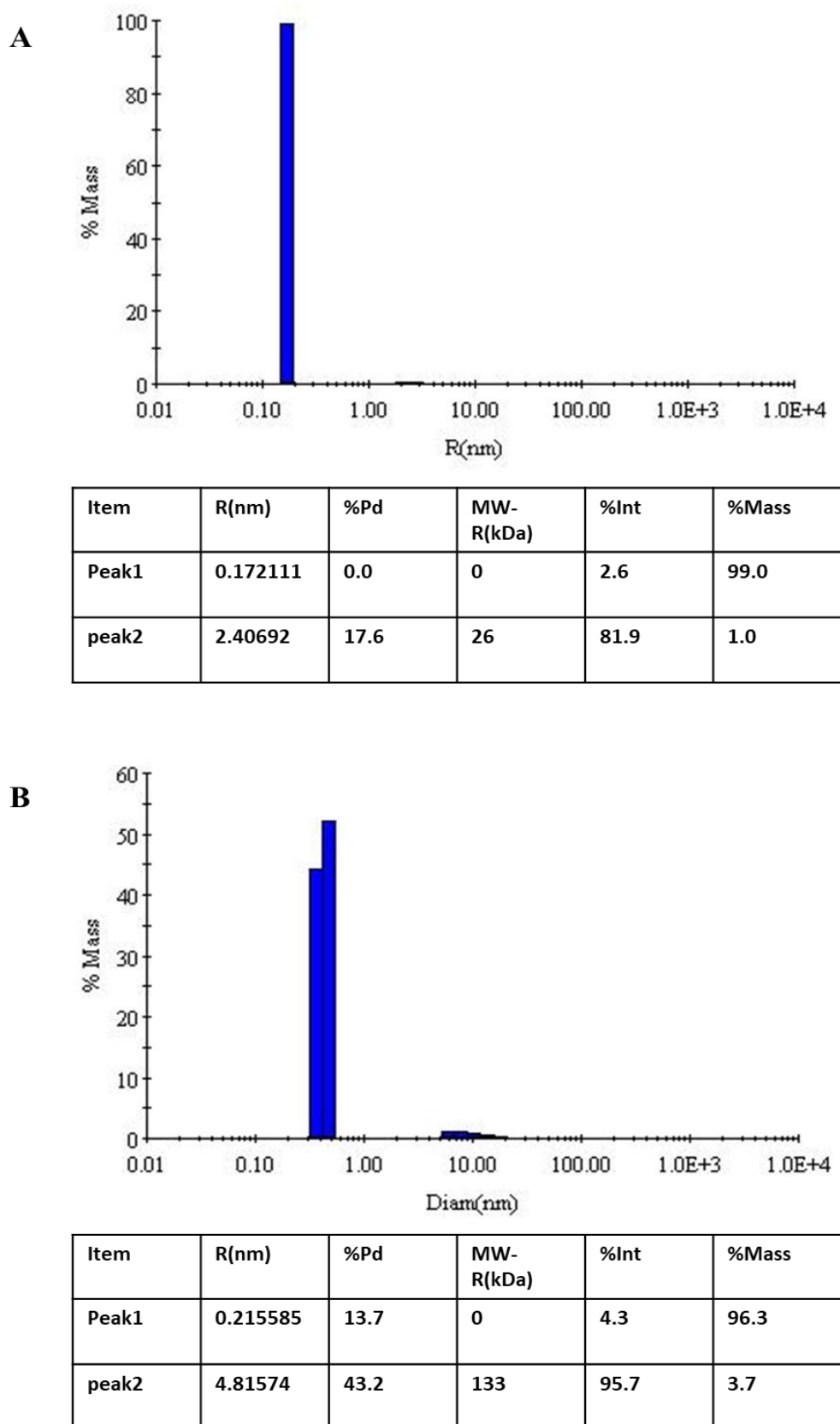
Figure 3.14 CD spectra of the first cytoplasmic domain, OutF⁶⁵⁻¹⁷², and full-length OutF, OutF¹⁻⁴⁰⁸.

The mean residue ellipticity of OutF⁶⁵⁻¹⁷², 0.21 mg ml⁻¹ (B) and OutF¹⁻⁴⁰⁸, 0.07mg/ml (A) in a 0.1 cm path length cell. Note: OutF⁶⁵⁻¹⁷² was in 20mM Tris-HCl, 100mM NaCl, pH 8.0. OutF¹⁻⁴⁰⁸ was in in 20mM Tris, 200mM NaCl, 10% glycerol and 0.03% DDM (pH 8.0). The protein concentration was determined using the predicted extinction coefficient of 0.1% of 0.217 and 0.810 respectively.

The circular dichroism spectra for both OutF⁶⁵⁻¹⁷² and Out¹⁻⁴⁰⁸ (Fig. 3.14A and B) have minima at 208 and 220 nm which is characteristic of proteins with a high helical content. Analysis of the spectra for Out⁶⁵⁻¹⁷² gives approximately 70% α -helical content which is consistently predicted using SELCON3, K2D and CONTIN.

3.2.6 Dynamic Light Scattering (DLS)

45 μ l of the eluent were transferred to a square cuvette for DLS measurements. OutF¹⁻⁴⁰⁸ and OutF⁶⁵⁻¹⁷² were measured and molecular weight determined by the instrument and the results shows in Fig 3.16.

Figure 3.15 DLS result for OutF⁶⁵⁻¹⁷² and SUMO-OutF¹⁻⁴⁰⁸(A) and (B) DLS results for OutF⁶⁵⁻¹⁷² and OutF¹⁻⁴⁰⁸, respectively.

The light scattering result of OutF⁶⁵⁻¹⁷² reveals a molecular mass of 26 kDa (17.6% polydispersity) a result consistent with that from size exclusion chromatography. 17.6% polydispersity indicates this protein is substantially monodispersing in solution and might therefore be crystallized.

The DLS results of SUMO-OutF¹⁻⁴⁰⁸ shows a mass of 133 kDa is consistent with the size exclusion chromatography but not consistent with measurements reported previously on PilG (Collins, Saleem et al. 2007).

3.2.7 Crystallization trials

a Crystallization trial using SUMO-OutF¹⁻⁴⁰⁸

The crystal screening of the SUMO-OutF employed the following commercial kits at 18°C using the hanging drop vapour diffusion method in 96-well microplates: MemGold™ MD1-39 (molecular dimensions), MemGold™ Eco Screen MD1-39-ECO, MemStart™ & MemSys MD1-33.



Figure 3.16 Crystal hits of Sumo-OutF¹⁻⁴⁰⁸

A: 0.1M Sodium chloride; 0.1M magnesium chloride hexahydrate; 0.1M CAPSO((3-(Cyclohexylamino)-2-hydroxy-1-propanesulfonic acid)) pH 9.5; 12% w/v PEG 4000; B: 1M sodium chloride; 0.1M magnesium chloride hexahydrate; 0.1M Sodium HEPES, pH7.5; 30% v/v PEG 400; C: 0.1M sodium chloride; 0.1M magnesium chloride hexahydrate; 0.1M sodium HEPES, pH 7.5 30% v/v PEG 400. Scale bar=200µm.

Several attempts were made to improve the size and quality of these crystals but no improvement in size or shape was observed.

b Crystallization of OutF⁶⁵⁻¹⁷²

Crystals of OutF⁶⁵⁻¹⁷² were grown at 18°C using the hanging drop vapor diffusion method in 96-well microplates. A Hampton Research sparse matrix screen was used to explore crystallization conditions. Crystallization trials were using 1µl OutF⁶⁵⁻¹⁷² and 1µl of 100mM sodium cacodylate pH 6.5, 200mM Li₂SO₄ and 30% v/v PEG-400 for optimization.

Data were collected from these crystals at ESRF ID23-1 and processed using XDS (Kabsch 2010). The structure was solved using molecular replacement and an initial model built using COOT (Emsley et al. 2010) and refined using CCP4 reffmac5. The final model comprises 234 amino acid residues (chain A 116, chain B 118) and 21 water molecules.

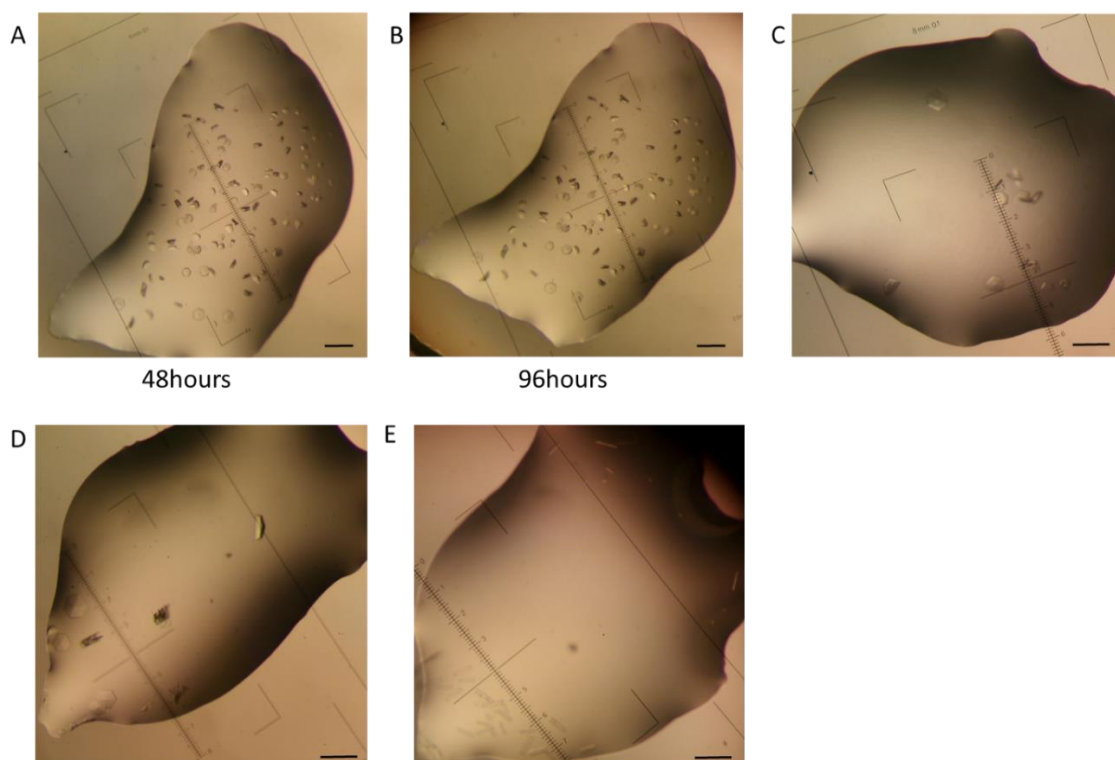


Figure 3.17 Crystal hits of OutF⁶⁵⁻¹⁷² in 96-well screening plate after 48 hours and 96 hours

(A) 100mM Na/K phosphate, pH 6.2, 200mM sodium chloride, 20% PEG-1000; (B) the same drop as (A) but after 96 hours; (C) 100mM cacodylate, pH6.5, 200mM magnesium chloride, 20%PEG-1000; (D) 100mM Na/K phosphate, pH 6.2, 10% w/v PEG-3000; (E) 100mM cacodylate, pH 6.5, 200mM lithium sulfate, 30% v/v PEG-400. Scale bar=200µm.

Two different shapes of crystals appeared in five different conditions and the crystal size did not change after 96 hours compared to after 48 hours. These conditions were further optimized in the following experiments to get larger crystals (Table 3.1).

Table 3.1 Optimization of the crystal hits condition

Sample:F65-172
 Concentration:15mg/ml
 Tray Number:18
 Title:20151020
 Date:Tuesday, October 20, 2015 2:28 AM

pH 6.2 pH6.5 pH6.7	26 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	28 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	30 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	30 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	32 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	34 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4
	26 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	28 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	30 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	30 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	32 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	34 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4
	26 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	28 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	30 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	30 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	32 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4	34 % PEG-400 0.1 M Cacodylate 0.2 M Li2SO4

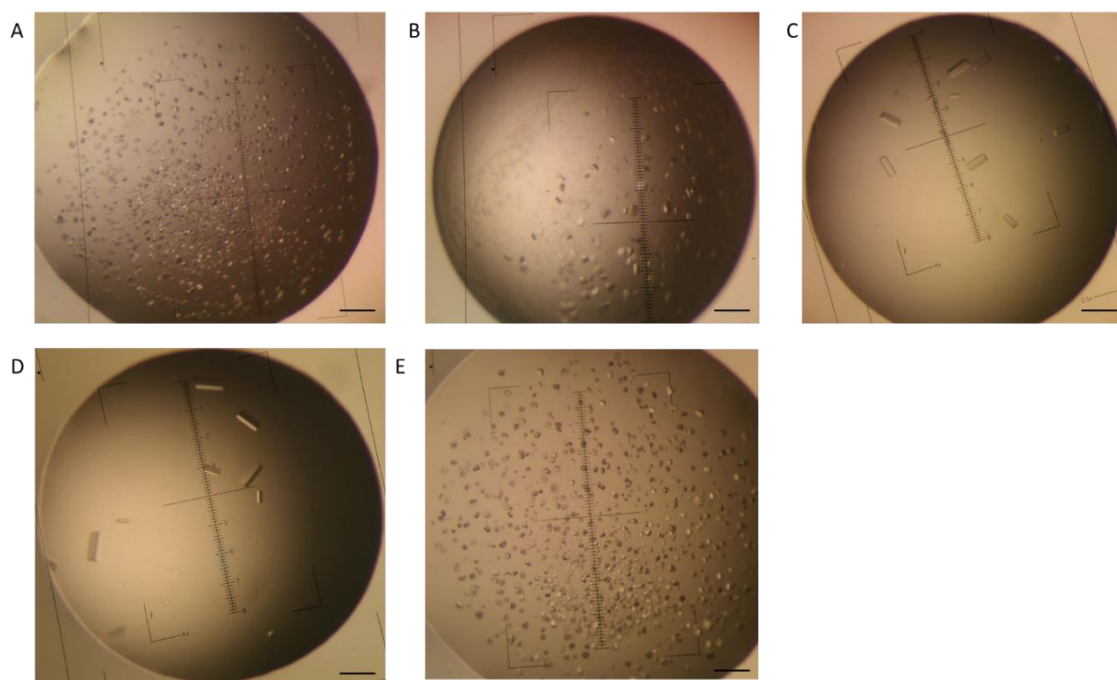


Figure 3.18 Crystal hits from optimized conditions

(A) 0.1M Na/K phosphate buffer, 0.2M NaCl, 20% PEG-1000, pH6.4; (B) 0.1M MES 0.1M NaCl 20% PEG-1000 0.1 M CaAc₂, pH 6.6; (C) 0.1M cacodylate, 32% PEG-400, 0.2M Li₂SO₄ pH 6.2; (D) 0.1M cacodylate 0.2M, 34% PEG-400, Li₂SO₄, pH 6.2; (E) 0.1M Na/K phosphate buffer, 0.2M NaCl, 20%PEG-1000, pH 6.4. Scale bar=200μm.

1 μl 15mg/ml of OutF⁶⁵⁻¹⁷² in 20mM Tris pH 8.0, 200mM NaCl and 1mM DTT buffer was mixed with an equal volume of precipitating solution. The tetragonal crystals in which grew using a reservoir of 0.1M cacodylate, 32% PEG-400, 0.2M Li₂SO₄ pH 6.2 gave the best diffraction and were used for structure determination.

c Crystal structure determination

Crystals of OutF⁶⁵⁻¹⁷² grew under several conditions with the best crystals obtained in the tetragonal space group P4₂12 with two molecules in the asymmetric unit (built using chains A and B, respectively). The *Vibrio cholera* GspF (PDB Re. 3C1Q) was used as the search model in molecular replacement using MOREP from CCP4 program suite(McCoy

et al. 2005). The model of OutF⁶⁵⁻¹⁷² was refined at a resolution of 2.15 Å to a R_{work} value of 20.7% (R_{free} value 24.6%) and consists of 216 residues of OutF (Table 3.2) with reasonable stereochemistry (Fig. 3.19 shows that 98.57% residues are located in the preferred region of the Ramachandran plot).

Table 3.2 Crystallographic data and refinement statistics for OutF⁶⁵⁻¹⁷²

Data collection	overall
Space group	P4 ₂ 12
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å) <i>α</i> , <i>β</i> , <i>γ</i> (°)	116.1, 116.1, 48.0 90, 90, 90
Number of images	840
Wavelength(Å)	0.972 Å
Resolution limits (Å)	41.45-2.15 (2.23-2.15)
Number of unique reflections	18328 (1764)
Completeness (%)	99.5(99.8)
Mean I/sigma(I)	16.5(2.3)
Wilson B-factor (Å ²)	42.76
Mn(I) half-set correlation CC _{1/2}	0.999(0.665)
<i>R</i> merge	0.058(0.804)
multiplicity	5.5(5.9)
Refinement	
Reflections	101006(10371)
R-factor/R-free (%)	20.7(24.6)
Rmsd bond(Å)/angle (°)	0.009(1.005)
Number of protein (solvent) atoms	1555 (66)
Ramachandran plot statistics (%)	
Residues in most favored regions	99.05%
Residues in additional allowed regions (%)	0.95%

Highest resolution shell is shown in parentheses.

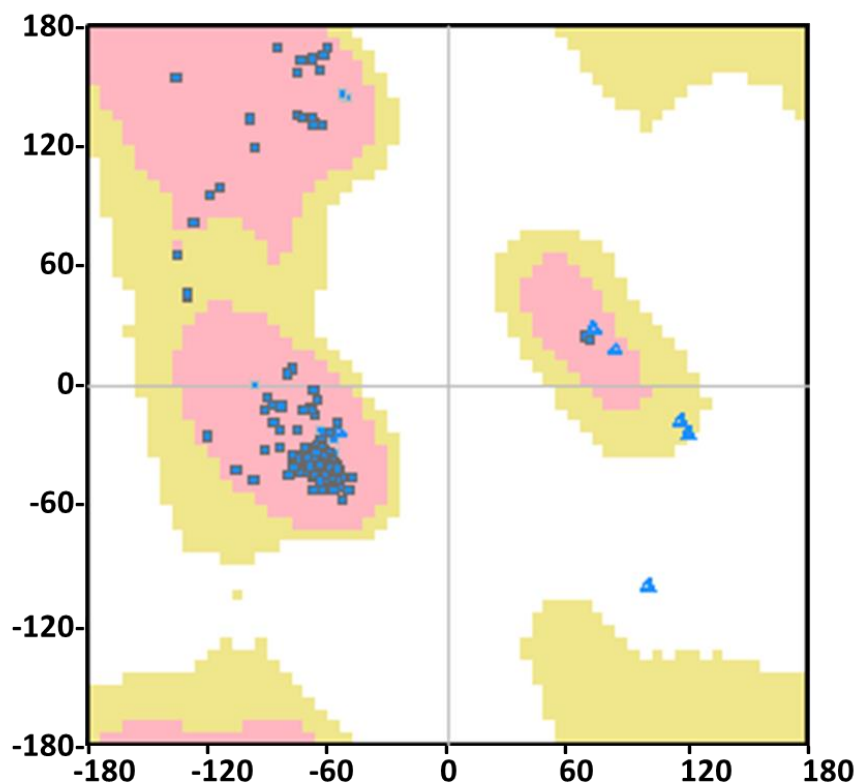


Figure 3.19 Ramachandran plot statistics for the OutF⁶⁵⁻¹⁷² crystal structure (PDB: 5NBG).

Pink: preferred regions; Yellow: allowed regions; White: Outliers. The percentage of residues in preferred region is 99.05% indicating a model with very good stereochemistry of OutF⁶⁵⁻¹⁷². Triangles are glycine residues, which are not subject to the same torsion angle restraints as the other residues.

3.2.8 The crystal structure of OutF⁶⁵⁻¹⁷²

OutF⁶⁵⁻¹⁷² has a structure predominantly α -helical structure: each domain has six antiparallel α helices and two domains come together to form a dimer. This structure is consistent with the predicted structure by online server Phyre2 (Fig. 3.20).

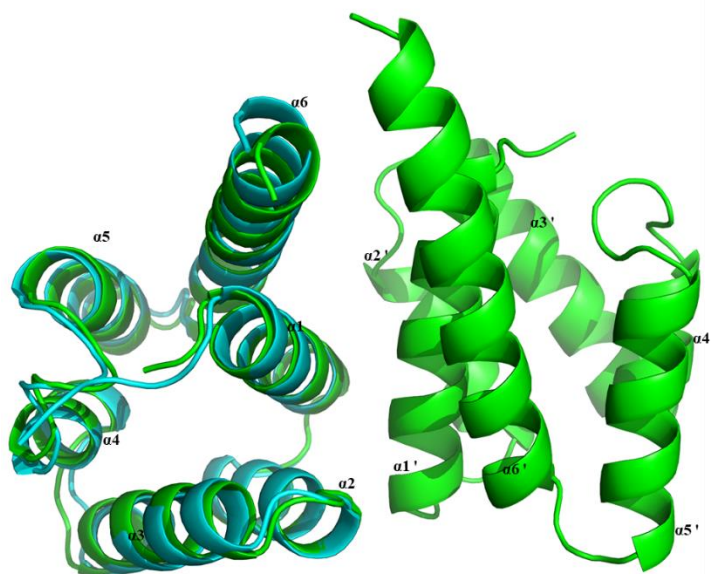


Figure 3.20 Comparison of predicted and solved structure of cytoplasmic domain I of OutF

The predicted structure of cytoI of OutF was superimposed onto the crystal structure and got a RMSD at 0.575 with 614 atoms. Cyan: predicted structure of cytoplasmic domain I of OutF by Phyre2; green: experimentally solved structure of OutF⁶⁵⁻¹⁷².

a Sequence and structure comparisons between members of the GspF family

The T2SS is ancestrally related to the Type 4 Pili System (T4PS) and both systems share common structurally equivalent components. Furthermore, some inner-membrane platform components are spread more broadly and present in T4PS, Tad and Com pili of Gram-positive bacteria as well as in archaeal flagella. We first compare the sequence of GspF family proteins across 9 species (Table 3.3 and Fig. 3.21).

Table 3.3 Nomenclature of GspF for related filament systems in Gram-negative, Gram-positive bacteria and archaeal flagella.

	Proteins in system					
	Gram-negative bacterial systems		Gram-positive bacterial systems			
Protein category	T2SS	T4P	TAD	COM	T4P	archaeal flagella
Inner-membrane core protein	GspF/OutF	PilC	TadB, TadC	ComGB	PilC	FlaJ

The Dali online server was then used to search for similar structures to OutF⁶⁵⁻¹⁷² and the three most similar structures were found: EpsF⁵³⁻¹⁷¹, TcpE¹⁻¹⁰² and PilC⁵³⁻¹⁶⁸. These solved GspF family proteins in addition with several unknown structure GspF family proteins were compared with OutF⁶⁵⁻¹⁷² based on sequence and structure (Fig. 3.21 and Table 3.4).

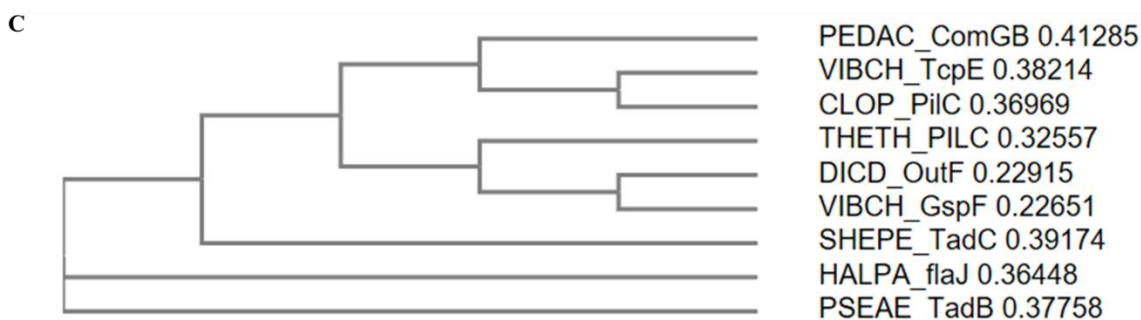
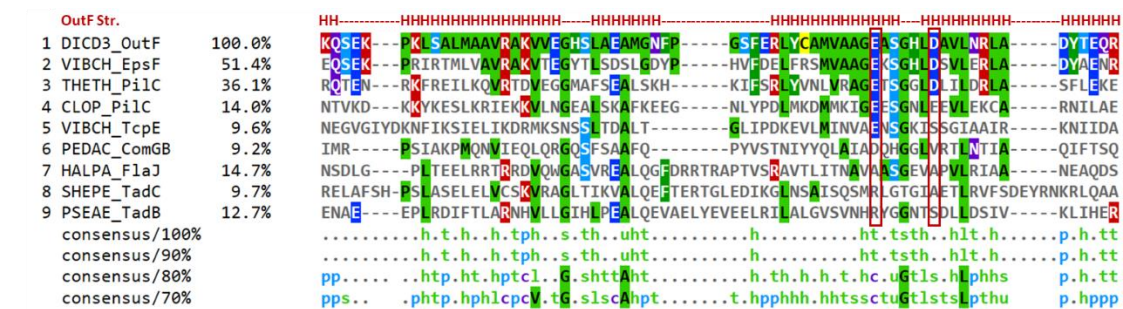
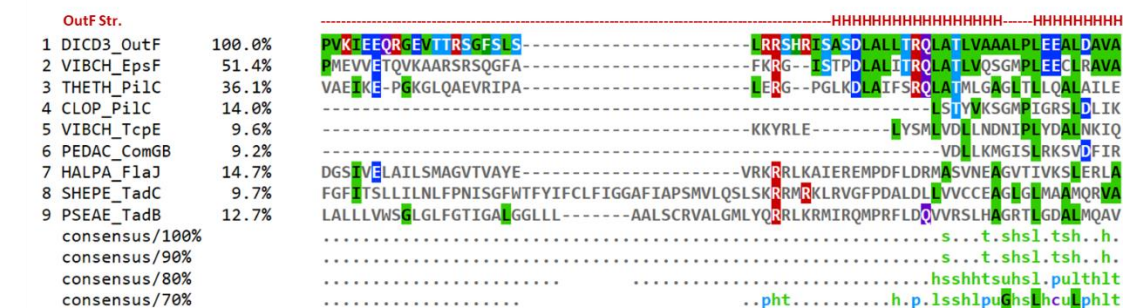
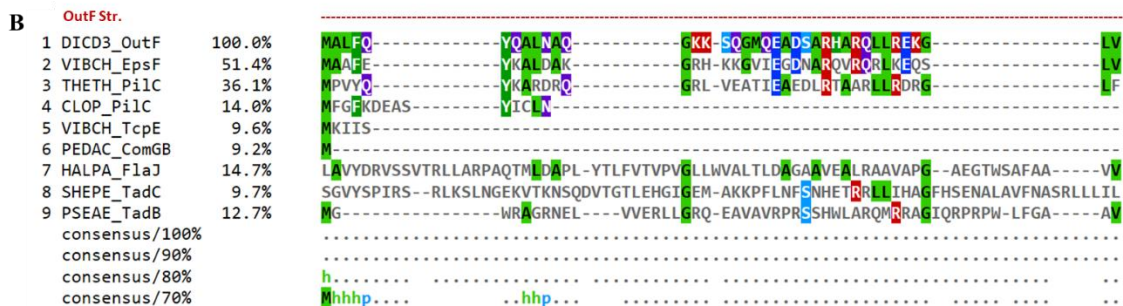


Figure 3.21 The predicted secondary structure of full-length of OutF and sequence alignment of GspF family members.

(A) The predicted secondary structure of full length of the OutF predicted using SACS, Tmpred Server and TMHMM Server. (B) The alignment of cytoplasmic domain I of GspF homologues in nine species: OutF [*Dickeya dadantii*] GenBank accession number WP_013318808.1; EpsF⁵³⁻¹⁷¹ [*Vibrio cholera*] accession NO. WP_000718700.1, PDB reference 3c1q; Toxin coregulated pilus biosynthesis protein E TcpE¹⁻¹⁰² [*Vibrio cholerae serotype O1*] accession NO. OJZ63583.1, PDB reference: 4hhx; PilC⁵³⁻¹⁶⁸ [*Thermus thermophilus*] accession NO. WP_011228203.1, PDB reference 2whn; pilus assembly protein TadC [*Shewanella pealeana*] accession NO. WP_012155257.1; Fimbrial assembly protein PilC [*Clostridium perfringens*] accession NO. WP_011590966.1; TadB [*Pseudomonas aeruginosa*] accession NO. CDO79668.1. competence protein; ComGB [*Pediococcus acidilactici*] accession NO. WP_008842339.1; flagella assembly protein j [*Halogeometricum pallidum*] accession NO. WP_008389602.1. The position of the α -helices is indicated by the letter H in red colour in the secondary structure row. Conserved residues are shaded in the consensus line. Red box: conserved and negatively charged residues in the bottom of the structure. (C) Phylogenetic tree of GspF family members. This is a neighbour-joining tree without distance corrections based on sequence alignment generated by Clustal Omega.

The sequence alignment and phylogenetic results support the close relationship between the T2SS and the T4PS. The sequence conservation was then mapped on the surface of the OutF⁶⁵⁻¹⁷² (Fig. 3.22).

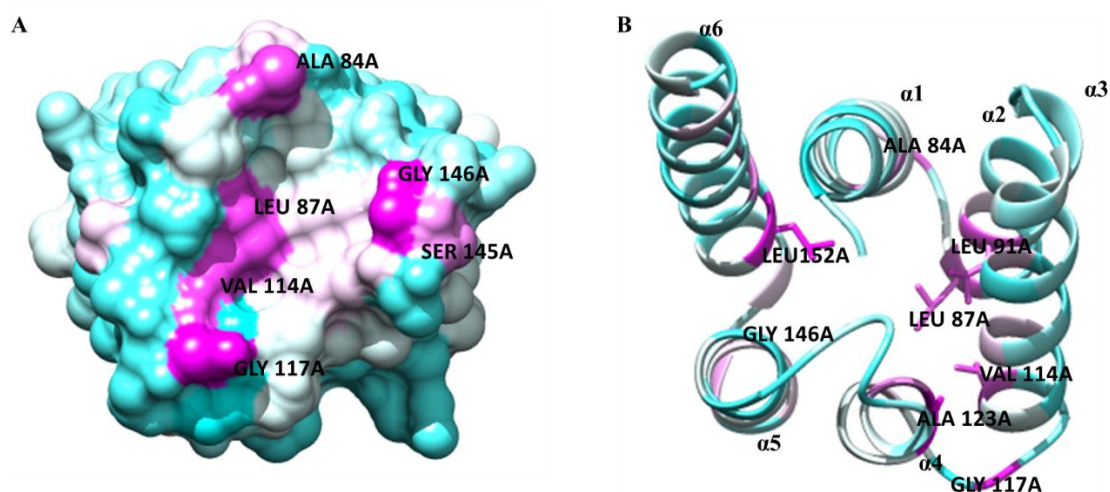


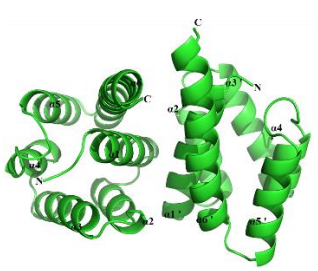
Figure 3.22 Conserved residues in the OutF⁶⁵⁻¹⁷² structure.

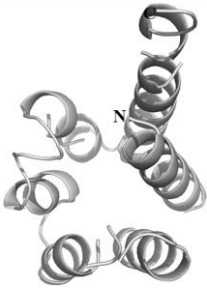
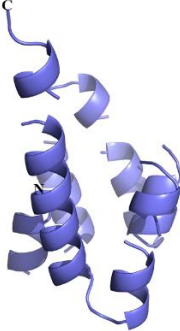
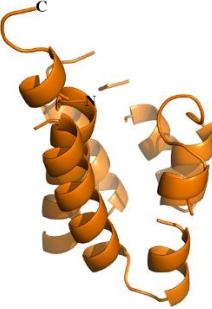
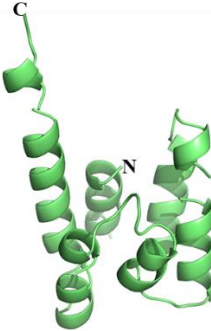
A: conserved residues in the “bottom view” of OutF⁶⁵⁻¹⁷². B: All the conserved residues mapped in the OutF⁶⁵⁻¹⁷² structure. Magenta: conserved residues.

There are eight conserved residues with at least 70% consensus in these nine species mapped on to the structure of OutF⁶⁵⁻¹⁷² (Fig. 3.22B): Ala 84, Leu 87, Leu 91, Val 114, Gly 117, Ala 123, Gly 146 and Leu 152. Five of the residues are located inside the cylinder formed by the six α helices and these residues play a key role in the helix bundle stabilization and the stability of the domain (Fig. 3.22). The conserved residues mapping to the surface (Fig. 3.22A) are presumably as a domain interface with another domain of OutF or other component of the T2SS (see below).

Table 3.4 Structure comparison of GspF family from T2SS, T4P in Gram-negative bacteria, T4P, Tad, Com in Gram-positive bacteria and archaeal flagella.

The comparison result was shown as RMS (root mean square deviation). The unknown structures are predicted by PHYRE2.

Structure 1	Structure 2	RMS
OutF ⁶⁵⁻¹⁷² [<i>Dickeya dadantii</i>]	EpsF ⁵⁶⁻¹⁷¹ [<i>Vibrio cholera</i>] PDB: 3C1Q	0.523(575 atoms); 1.232(1244 atoms)
	PilC ⁵³⁻¹⁶⁸ [<i>Thermus thermophilus</i>] PDB: 2whn	1.034(562atoms); 13.85(1279 atoms)
	TcpE ¹⁻¹⁰² [<i>Vibrio cholerae</i>] PDB: 4hhx	2.15(421 atoms)
	Fimbrial assembly protein PilC ¹²⁻¹¹⁰ [<i>Clostridium perfringens</i>] predicted	0.475(299 atoms)

	FlaJ ⁵⁰⁴⁻⁶¹⁴ [<i>Halogeometricum pallidum</i>] Predicted		9.401(475atoms)
	ComGB ¹⁻⁹¹ [<i>Pediococcus acidilactici</i>] Predicted		1.035(308atoms)
	TadB ⁹⁸⁻²⁰⁶ [<i>Pseudomonas aeruginosa</i>] Predicted		2.39(375atoms)
	TadC ¹⁷⁶⁻¹⁸² [<i>Shewanella pealeana</i>] Predicted		13.040(521 atoms)

Eight structures of GspF family members including the three solved structures and five predicted structures were compared to OutF⁶⁵⁻¹⁷² in Table 3.6. The most similar structure to OutF⁶⁵⁻¹⁷² is *V. Cholerae*-EspF⁵³⁻¹⁷¹ and the predicted structure of PilC¹²⁻¹¹⁰ in T4PS from Gram-positive bacteria with RMS 0.523 (575 atoms) and 0.475 (299 atoms) respectively. This result further supports the T4PS has close relationship with the T2SS.

b The nature of the OutF⁶⁵⁻¹⁷² surface

To study the surface characteristics of OutF⁶⁵⁻¹⁷², the hydrophobicity and electrostatic properties were mapped on to the surface of OutF⁶⁵⁻¹⁷² (Fig. 3.23 and 3.24).

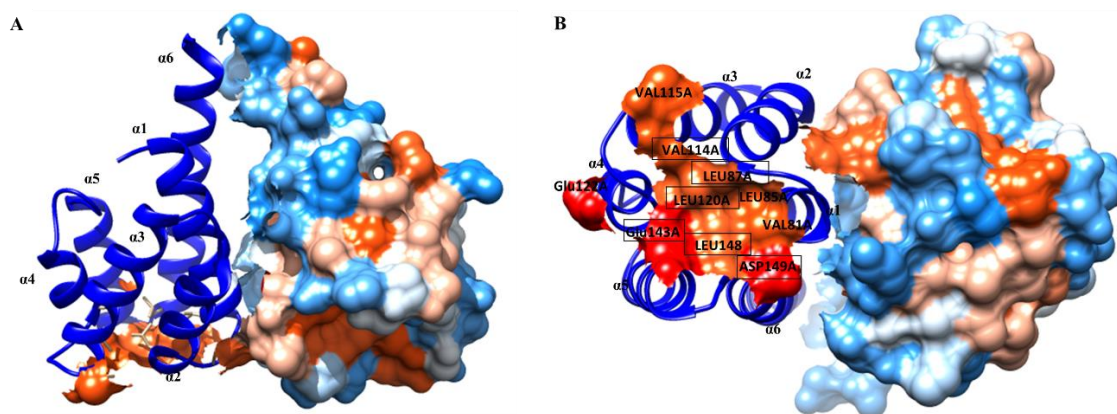


Figure 3.23 Hydrophobicity mapped on the surface of the OutF⁶⁵⁻¹⁷²

A: The front view of the OutF dimer. B: The bottom view of the dimer. Orange: hydrophobic residues; blue: hydrophilic residues; red: negatively charged residues; black rectangle: conserved residues.

Seven hydrophobic residues form a narrow and shallow groove in the bottom and Ala 114 and Leu 87, Leu 148 and Leu 120 are conserved in the GspF family proteins in T4PS and T2SS (Fig. 3.24). This hydrophobic area is located between two conserved negatively charged residues Glu 143 and Asp 149. This groove may provide the interaction area with other proteins in the secretion machine.

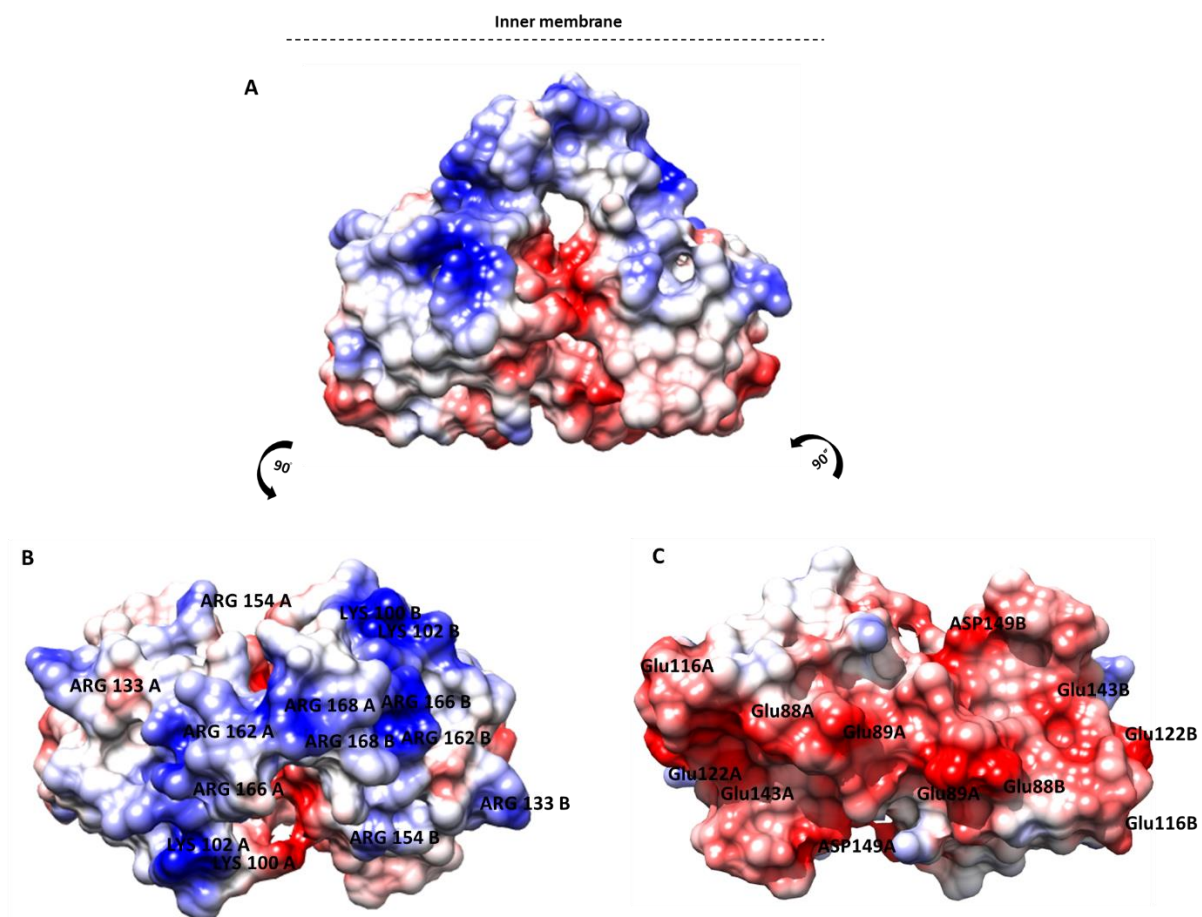


Figure 3.24 Electrostatic potential mapped onto the OutF⁶⁵⁻¹⁷² structure.

(A) The electrostatic potential mapped onto the OutF⁶⁵⁻¹⁷² surface. (B) Electrostatics of the top surface of the structure. Positive residues are labelled on the map: five arginine and two lysine from each chain of the dimer. (C) Electrostatic properties of the bottom surface of the structure. Negative residues are labelled: six glutamates and one aspartate from each chain of the dimer. Red and blue are negative and positive potential, respectively.

This positively charged top and negatively charged bottom suggests that the top of the molecule is close to the membrane, this orientation is consistent with an extension of the $\alpha 6$ helix extending into the inner-membrane as transmembrane helix 1.

c Potential dimerization factors

In the crystal, there is a dimer in the asymmetric unit which is consistent with the dynamic light scattering and size exclusion chromatography results in solution. The structure suggests how this dimerization may occur at the atomic level, in this analysis hydrogen bonding, hydrophobic burial, conserved residues and metal binding sites at the interface are taken into account.

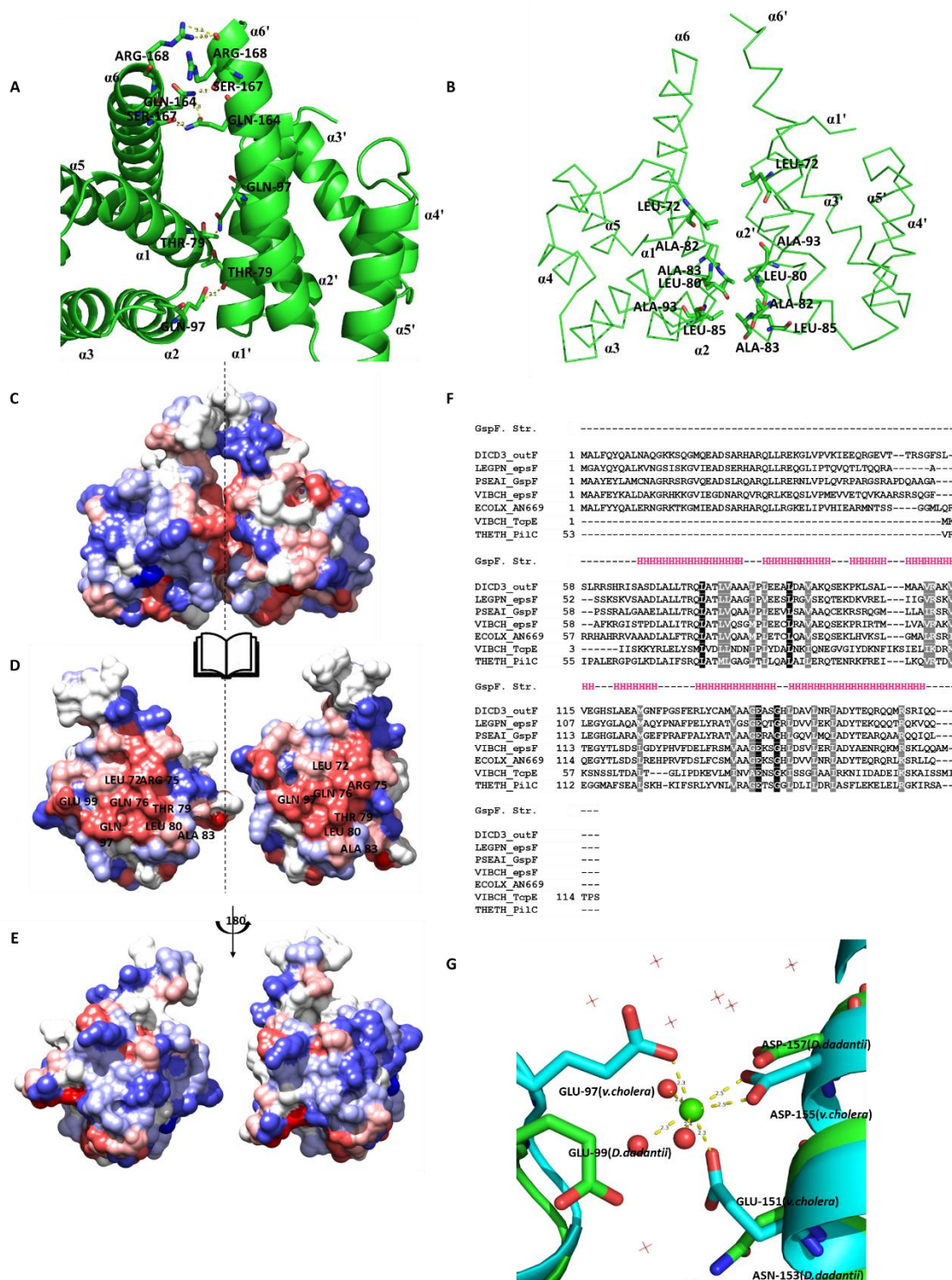


Figure 3.25 Interactions at the dimer interface of OutF⁶⁵⁻¹⁷².

(A) The hydrogen bonds at the OutF⁶⁵⁻¹⁷² homodimer interface. Seven hydrogen bonds at the top and bottom of dimer interface involve the following residues: Thr 79 provided by helix $\alpha 1$; Gln 97 from helix $\alpha 2$; Glu 99 from the loop between $\alpha 2$ and $\alpha 3$; Gln 164, Ser 167 and Arg 168 from $\alpha 6$; Thr 79 provide by helix $\alpha 1'$; Gln 97 from the loop between $\alpha 2'$ and $\alpha 3'$; Gln 164, Ser 167, Asn 153 and Arg 168 from $\alpha 6'$. (The prime indicates the symmetry related molecule.) Sticks:

residues forming hydrogen bond on the interface of the dimer; yellow dot: hydrogen bond. (B) Hydrophobic residues on the interface of the OutF⁶⁵⁻¹⁷² dimer. The hydrophobic residues are shown in sticks. There are total eight hydrophobic residues from chain A and chain B forming a hydrophobic core at the bottom of the OutF⁶⁵⁻¹⁷² dimer: Leu 72, Leu 80, Ala 82, Ala 83 from $\alpha 1$, Leu 85 from the loop between $\alpha 1$ and $\alpha 2$, Ala 93 from $\alpha 2$; Leu 72, Leu 80, Ala 82, Ala 83 from $\alpha 1'$, Leu 85 from the loop between $\alpha 1'$ and $\alpha 2'$; Ala 93 from $\alpha 2'$. (C) Conservation of the OutF⁶⁵⁻¹⁷² in the front side based on the sequence alignment of seven GspF family members from T2SS and T4P in Gram-negative bacteria: conservation of the interface the OutF⁶⁵⁻¹⁷² dimer (D) and backside of interface (E). (F) Sequence alignment. (G) Schematic of the calcium-binding site of EspF (cyan) and superimposed the OutF structure (green) that lacks the calcium-binding site. The calcium-binding site of *V. cholera* EspF⁵³⁻¹⁷¹ (cyan) is absent in *D. dadantii* OutF⁶⁵⁻¹⁷² (green). Notably glutamate 151 in *V. cholera* is substituted by asparagine in *D. dadantii* and calcium no longer binds, despite its inclusion in crystallization media. Calcium is shown as the small green sphere and waters coordinating the calcium by red spheres. Figure 3.26A, 3.26B and 3.26G were produced using PYMOL and Figure 3.26C, 3.26D and 3.26E were produced using Chimera.

Figure 3.25 shows the potential interaction on the interface which could stabilize the dimer. Unlike *V. cholera*-EspF⁵³⁻¹⁷¹ the most similar structure to OutF⁶⁵⁻¹⁷², there is no calcium or other metal in this structure. The reason appears to be because the calcium-binding site glutamate, Glu 151 in *V. cholera*, is substituted by asparagine in *D. dadantii*. Seven hydrogen bonds at the top and bottom of dimer interface and total eight hydrophobic residues from chain A and chain B forming a hydrophobic core at the bottom play an important role in the dimerization. Additionally, the conservation at the interface compared to the back of the molecule highlights the higher level of conservation at the dimerization interface (Fig. 3.25 C, D and E).

However, this dimerization interface is not the strongest association between molecules in the crystal as judged by PISA, http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver. This will be further discussed in the discussion section.

d Predicted model of OutF cytoplasmic domain II

Cytoplasmic domain II, OutF²⁴⁵⁻³⁷⁹, is 22% identical in sequence to cytoplasmic domain I, OutF⁶⁵⁻¹⁷² (Fig. 3.26A). When the sequence identity is 20% or more, spread evenly across the compared sequences, then the structures will be similar (Krissinel 2007).

OutF²⁴⁵⁻³⁷⁹ was also predicted to be a six helix bundle similar in structure to OutF⁶⁵⁻¹⁷² by

Phyre2

(3.27

B).

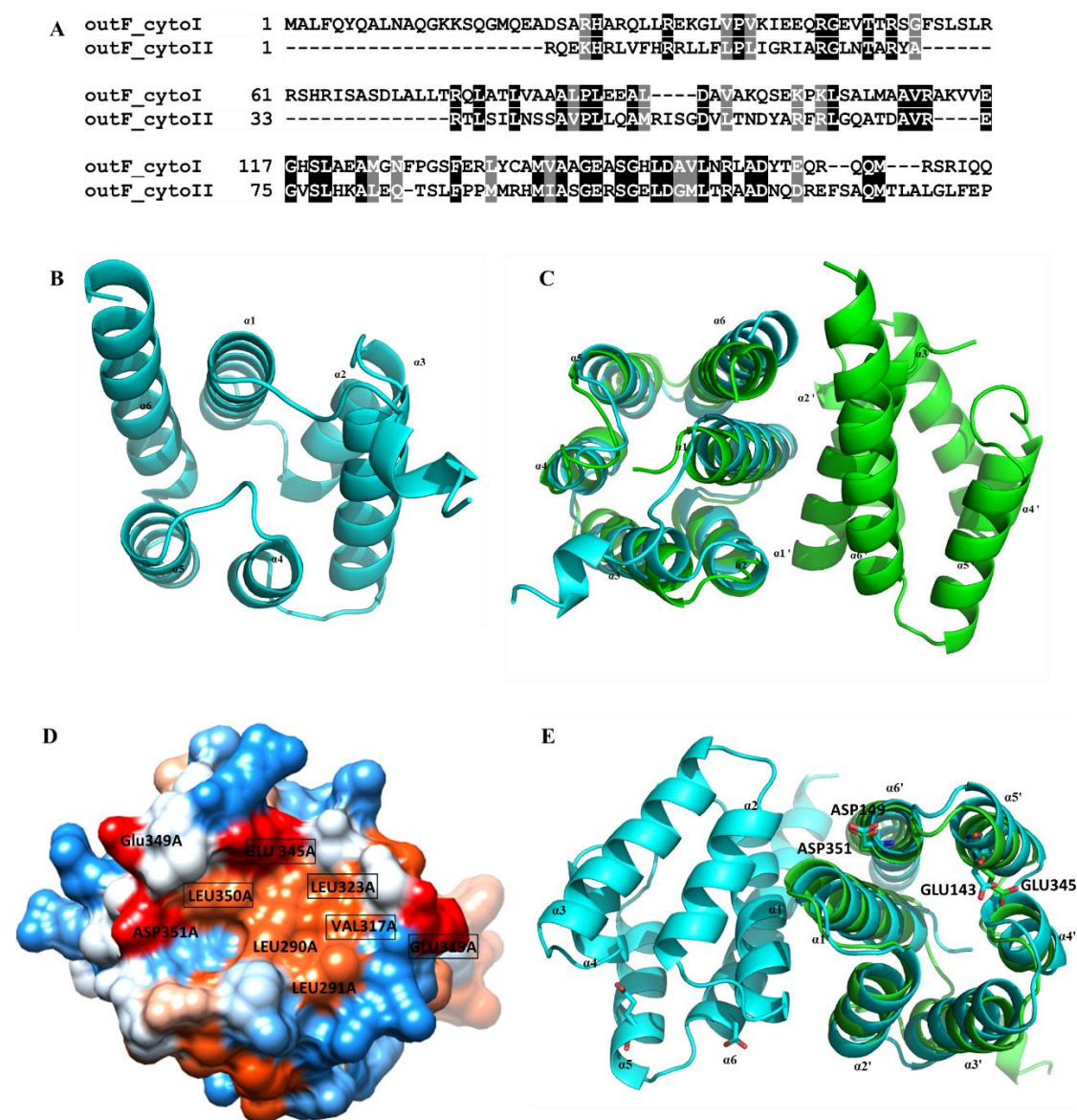


Figure 3.26 The predicted model of cytoplasmic domain II of OutF

(A) Sequence alignment of cytoplasmic domains I and II of OutF. (B) The predicted model of OutF

cytoplasmic domain II using Phyre2. (C) The structure of cytoplasmic domain II (cyan) superimposed on cytoplasmic domain I (OutF⁶⁵⁻¹⁷²) (green) with RMSD 2.448 Å (406 equivalent atoms). D Hydrophobicity of OutF²⁴⁵⁻³⁷⁹ mapped to the molecular surface. Orange: hydrophobic residues; red: negatively charged residues; black rectangle: conserved residues. The potential binding site in this area are compared with that in the OutF⁶⁵⁻¹⁷² in E. green: OutF²⁴⁵⁻³⁷⁹, cyan: OutF⁶⁵⁻¹⁷².

Two models of the OutF²⁴⁵⁻³⁷⁹ dimer were built based on the sequence similarity with OutF⁶⁵⁻¹⁷² and PilC⁵³⁻¹⁶⁸, respectively (Fig. 3.27).

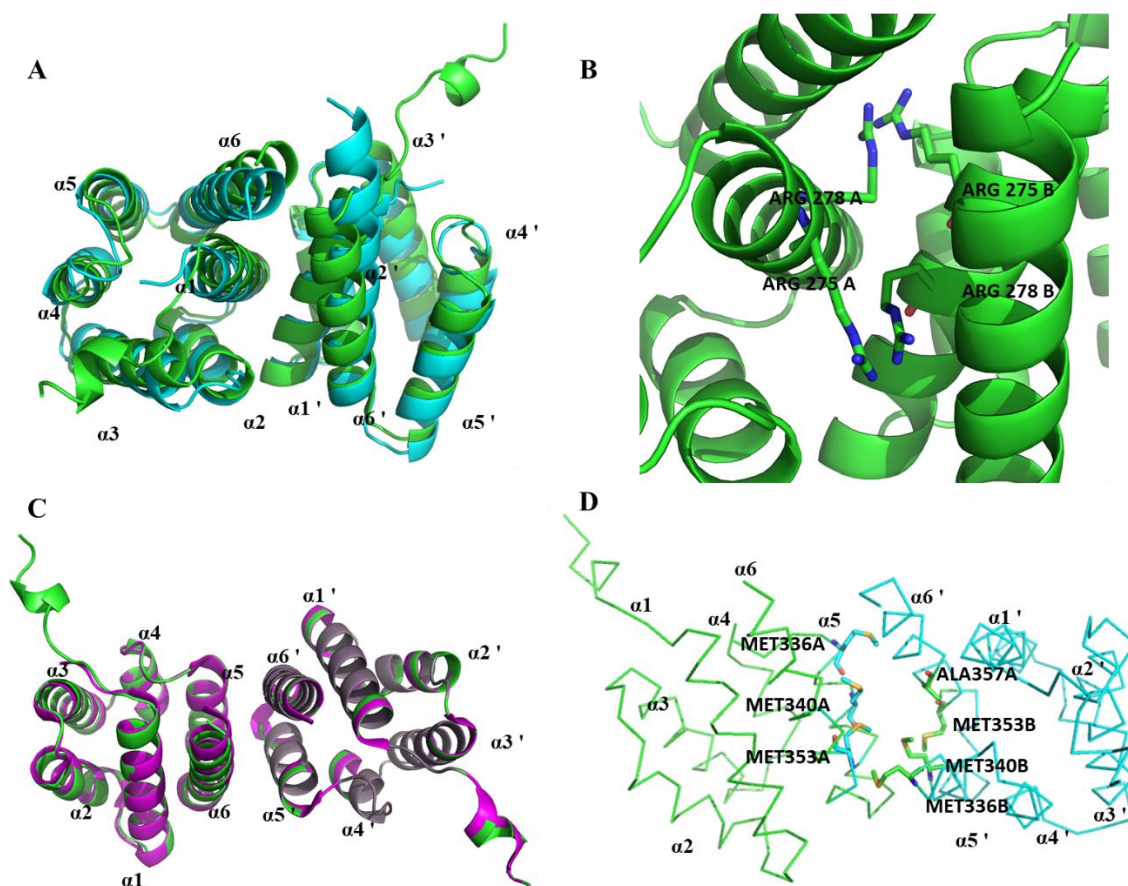


Figure 3.27 Two models of OutF²⁴⁵⁻³⁷⁹ dimer

A. The predicted monomer OutF²⁴⁵⁻³⁷⁹ was superimposed onto OutF⁶⁵⁻¹⁷² and a dimer built. Two arginines from each chain at the modelled dimer interface are shown in B. It seems unlikely that these arginine residues would be so close, unless they are in an unusual protonation state. A

monomeric structure or a different dimerization interface therefore seems more likely. Thus, the second possible dimer model was built by superimposing OutF²⁴⁵⁻³⁷⁹ onto PilC⁵³⁻¹⁶⁸ in *C. D.* Hydrophobic residues in the center of the dimer interface. These 7 hydrophobic residues are conserved based on sequence alignment of 5 species.

Figure 3.27 shows that when the OutF²⁴⁵⁻³⁷⁹ dimer is modelled on the OutF⁶⁵⁻¹⁷² dimer, where $\alpha 1$, $\alpha 2$ and $\alpha 6$ forms the interface, two arginine residues from each chain of the dimer are close, making dimer formation improbable, unless the ionization state of the arginine residues is unusual. A monomeric structure or a different dimerization interface seems more likely. Thus, a second dimer model was built using PilC⁵³⁻¹⁶⁸ as template where $\alpha 5$ and $\alpha 6$ helices from each chain form the interface. This dimer seems much more stable by forming a hydrophobic core in the center of the interface. Additionally, these hydrophobic residues are quite conserved.

3.2.9 Interaction of OutE, OutF and OutL assessed using NMR

The *outE*, *outF* and *outL* genes are organized on a single operon and an interaction between these proteins is known to be important for assembly of an inner-membrane platform in T2SS.

Beatrice Py *et al* (1999) have shown using a yeast two-hybrid experiment that the first N-terminal of 172 residues of GspF in *Erwinia chrysanthemi* could interact both with the cytoplasmic domain of OutL and OutE. The first cytoplasmic domain of OutF, the cytoplasmic domain of OutL and OutE were shown to form a stable complex *in vivo* using co-immunoprecipitation experiments in *Erwinia chrysanthemi* (Py, Loiseau *et al.* 2001). GspL is required for the formation of the GspE-GspF complex while OutL and OutE were able to form a complex without the requirement of OutF (Py *et al.* 2001). This OutF region

contains the cytoplasmic domain I helical domain preceded by a 64 residue, disordered region, with high intrinsic disorder predicted for the first 50~60 residues (Fig. 3.3). The pull-down assay results in section 3.2.4 show a strong interaction between the cytoplasmic domain of OutL and full length of OutE but no interaction between these proteins and cytoplasmic domain of OutF. To further analyze whether this intrinsically disordered OutF region could participate in the interaction with OutE and OutL and whether OutF could interact with OutE and OutL independently, NMR spectroscopy and thermofluor assay were used.

a Sample preparation for NMR spectroscopy

^{15}N -labelled OutF¹⁻¹⁷² (or OutF⁶⁵⁻¹⁷²) was expressed using the pET-14b vector in BL21(DE3) cells, grown in M9 minimal media. 1L M9 minimal media was prepared as the method in Chapter 2 supplemented with ampicillin (100 $\mu\text{g}/\text{ml}$). All minimal media was 0.2 μm filter sterilized and 1L induced expression cultures were grown at 16°C.

All protein samples for NMR experiments: OutE¹⁻⁵¹³, OutL¹⁻²⁵⁷, ^{15}N -OutF¹⁻¹⁷², ^{15}N -OutF⁶⁵⁻¹⁷² and OutE¹⁻⁵¹³-OutL¹⁻²⁵⁷ complex were purified using size exclusion chromatography after elution from the Nickel affinity column. The buffer condition for ^{15}N -labelled or non-labelled proteins are determined by Thermofluor assay.

Figure 3.28 shows the purified ^{15}N -labelled proteins prepared from gel filtration chromatography for NMR experiments.

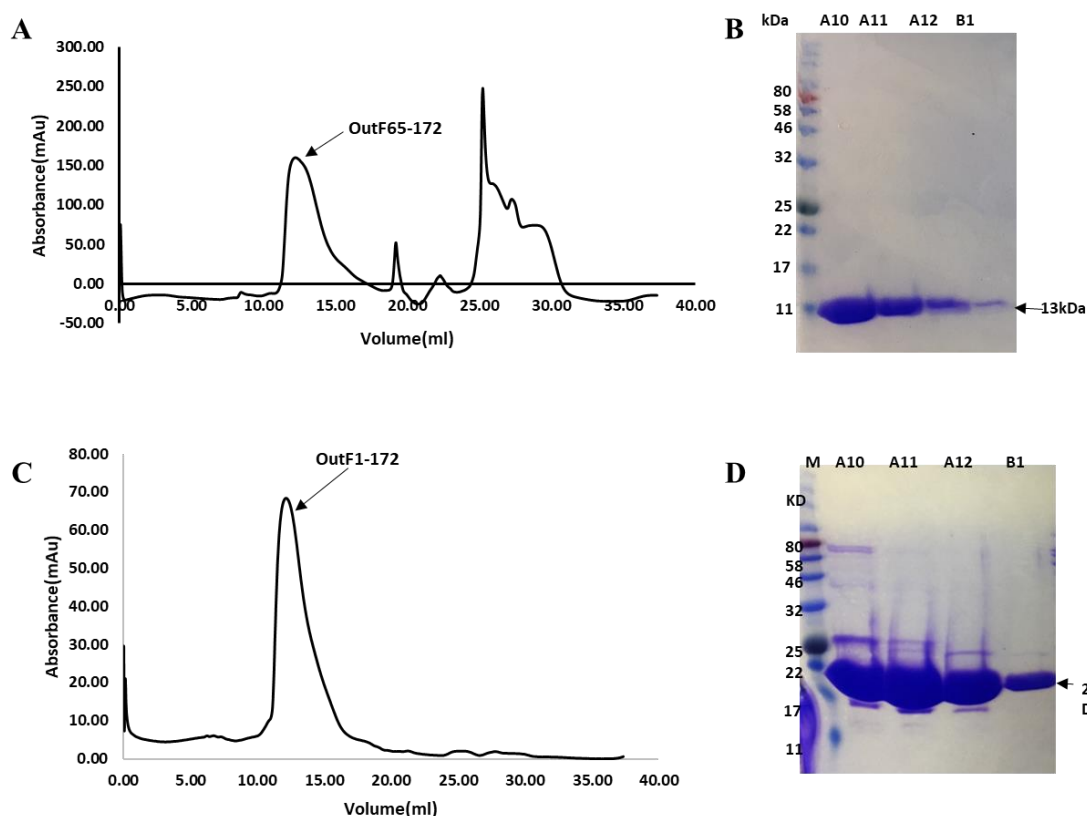


Figure 3.28 Size exclusion chromatography of ^{15}N -labelled OutF¹⁻¹⁷², OutF⁶⁵⁻¹⁷²

Superdex 75 10/300 GL exclusion chromatography column was equilibrated with buffer (20mM Tris-HCl, 150mM NaCl, 1mM DTT, pH 7.0)

A: ^{15}N -labelled OutF⁶⁵⁻¹⁷² S75 column size exclusion trace measured at 280 nm. B: Purified ^{15}N -labelled OutF⁶⁵⁻¹⁷² protein. Corresponding fractions were collected from after the size exclusion. C: ^{15}N -labelled OutF¹⁻¹⁷² S75 column size exclusion trace measured at 280 nm. D: Purified ^{15}N -labelled OutF¹⁻¹⁷² protein.

After purification using size exclusion chromatography, the proteins were concentrated to 0.1-1 mM for subsequent NMR experiments.

b Thermoflour assay to determine sample conditions for NMR studies

Sypro orange dye diluted from stock solution was added to the protein (at least 5 μM protein). The screening mixture was then made by mixing 5 μl of protein sample and 20 μl of screening buffer. The temperature screen was between 25-95 $^{\circ}\text{C}$ at one-degree

intervals with measurements every minute. The fluorescence emission signal was measured at 525 nm.

Buffer conditions used for the thermofluor assays were described in Chapter 2.

Thermofluor experiments used buffers made for NMR experiment purposes. The highest screening temperature is measured at pH 7.5. At pH values above 7.5, the amide proton exchange rate is high and it is difficult to study interaction by NMR. Four different pH values under 7.5 were chosen for thermofluor assay.

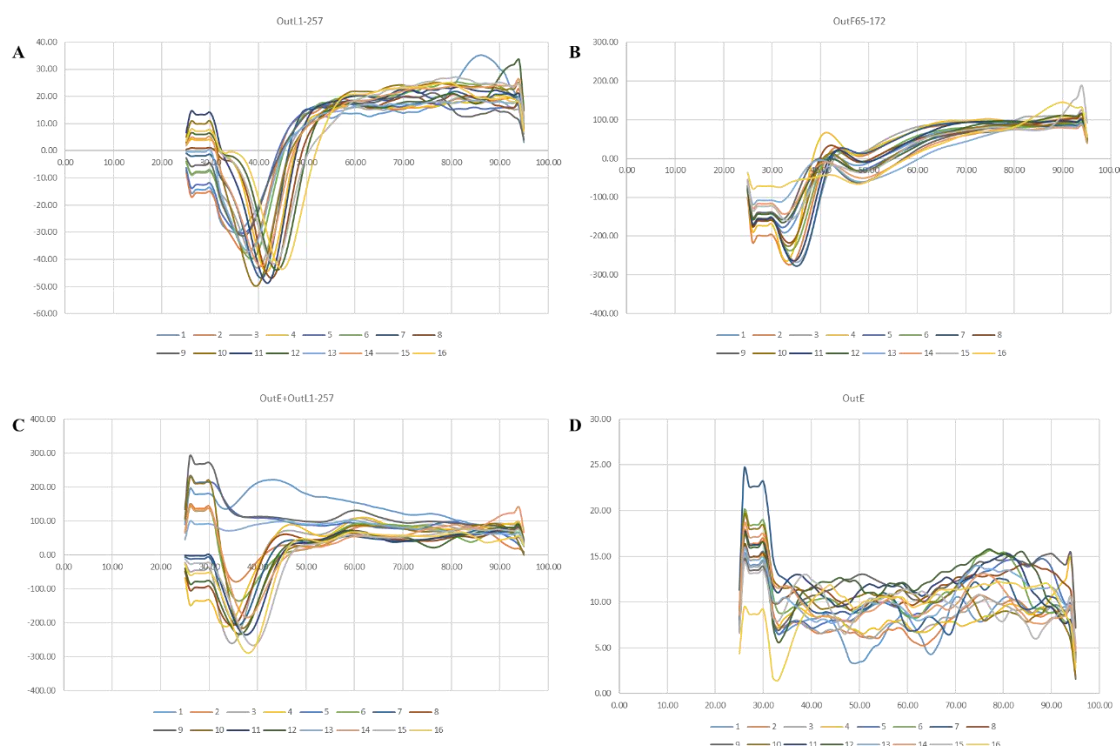


Figure 3.29 Thermofluor plots

(A) Thermofluor plot of OutL¹⁻²⁵⁷. (B) Thermofluor plot of OutF⁶⁵⁻¹⁷². (C) Thermofluor plot of OutE¹⁻⁵¹³ and OutL¹⁻²⁵⁷ complex. (D) Thermofluor plot of OutE¹⁻⁵¹³.

Finally, the condition 11 (20mM Tris-HCl 150mM NaCl pH 7.0) was chosen as the buffer to be used for NMR experiments since only this buffer ensures the stability of: OutL¹⁻²⁵⁷, OutF⁶⁵⁻¹⁷², OutE¹⁻⁵¹³ and OutE¹⁻⁵¹³ / OutL¹⁻²⁵⁷ complex.

c NMR spectroscopy

NMR spectra were acquired at 15°C using Bruker Avance 700 and 600 MHz spectrometers. At first, since no buffer screening was performed before the NMR experiments, a 500mM NaCl buffer was chosen to avoid protein aggregation resulting in higher signal to noise ratio in the recorded spectra. NMR samples contained 0.05-1 mM uniformly labelled OutF¹⁻¹⁷² (or 0.1mM OutF⁶⁵⁻¹⁷²) in 90% H₂O/ 10% D₂O. pH 7.0 was chosen for the data collection as higher pH values speed up the amide proton exchange rate reduce the spectral resolution.

¹H NMR spectra were first measured to evaluate the quality of the protein sample. Then ¹H-¹⁵N-HSQC spectra of OutF⁶⁵⁻¹⁷² (or OutF¹⁻¹⁷²) were measured in the absence and presence of OutE-OutL¹⁻²⁵⁷ binary complex.

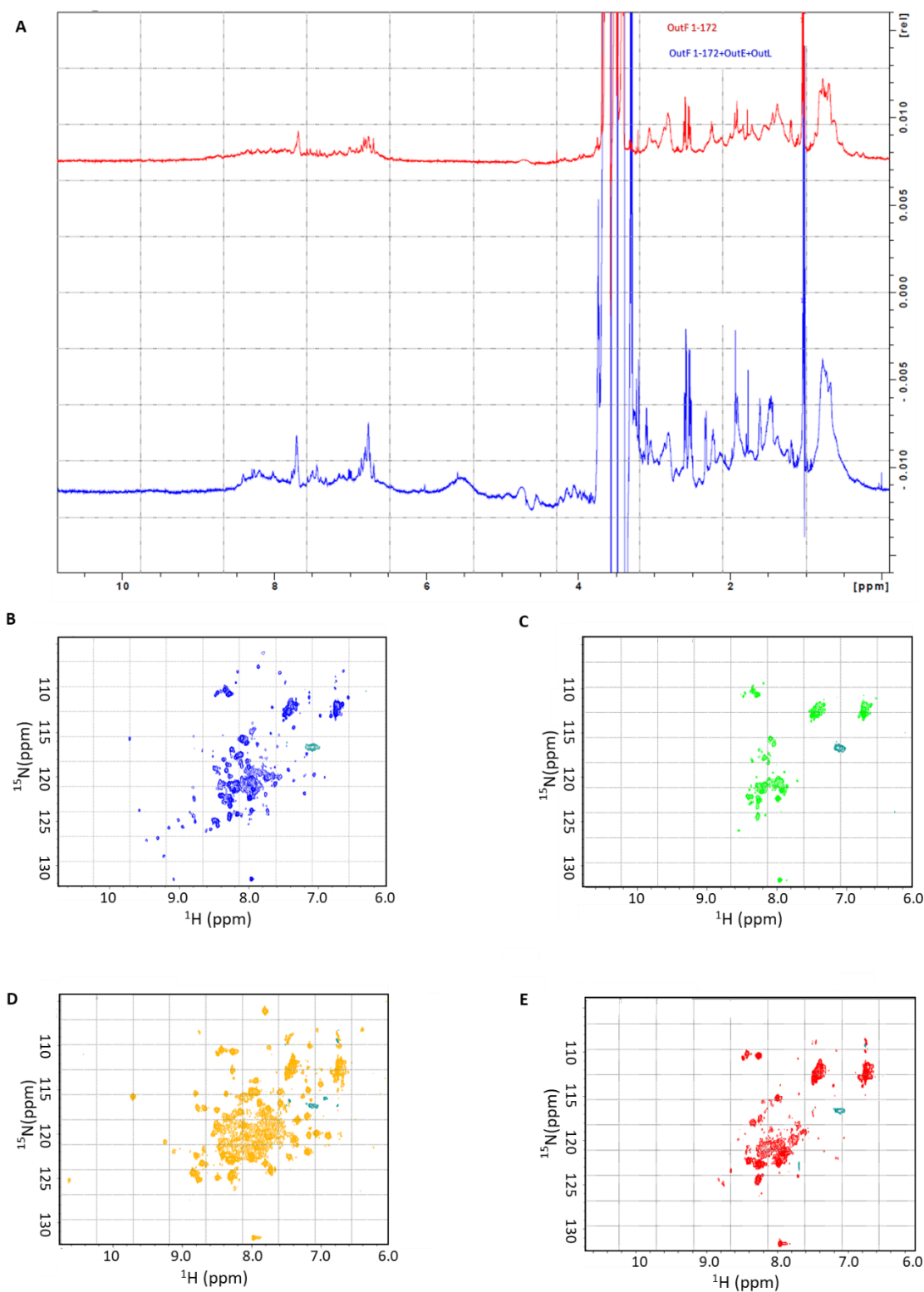


Figure 3.30 Assessment of interactions between cytoplasmic domain I of OutF and OutE-OutL¹⁻²⁵⁷

(A) ^1H spectra of OutF¹⁻¹⁷² in the absence (red) and presence of OutE-OutL¹⁻²⁵⁷ (blue). The ratio between OutF¹⁻¹⁷² and OutE-OutL¹⁻²⁵⁷ is 1:1.2. There is no peak shift in 1D spectra indicating that

the pH of Tris buffer doesn't change at all. Purified OutE-OutL¹⁻²⁵⁷ complex with approximate 1:1 stoichiometric ratio. Titration of OutE-OutL¹⁻²⁵⁷ into labelled OutF¹⁻¹⁷²(or OutF⁶⁵⁻¹⁷²). OutF¹⁻¹⁷² concentration is 0.05mM and OutF⁶⁵⁻¹⁷² concentration is 0.1mM. (B) ¹H-¹⁵N-HSQC spectra of OutF¹⁻¹⁷² in the absence of OutE-OutL¹⁻²⁵⁷ (blue spectra). (C) Spectra of OutF¹⁻¹⁷² in the presence of OutE-OutL¹⁻²⁵⁷(green spectra). Binding results in loss of signal due to spectral broadening. (D) OutF⁶⁵⁻¹⁷² in the absence of OutE-OutL¹⁻²⁵⁷ (yellow spectra). (E) Spectra of OutF⁶⁵⁻¹⁷² in the presence OutE-OutL¹⁻²⁵⁷(red spectra).

Both ¹⁵N-OutF⁶⁵⁻¹⁷² and ¹⁵N-OutF¹⁻¹⁷² signals decrease upon addition of OutE-OutL¹⁻²⁵⁷ due to the formation of a high molecular weight complex (Fig. 3.30) (the initial size of OutF⁶⁵⁻¹⁷², OutF¹⁻¹⁷² and OutE-OutL¹⁻²⁵⁷ are 13kDa, 20kDa and 81kDa, respectively).

These results indicate that the first cytoplasmic domain of OutF interacts with the OutE-OutL¹⁻²⁵⁷ complex and that the first 65 residues seem not to be essential for this interaction. The N-terminal intrinsic disorder region which was predicted to be a potential protein binding site therefore plays no role in this interaction.

To further investigate the interaction between the OutF⁶⁵⁻¹⁷² and OutE-OutL¹⁻²⁵⁷ complex, ¹H-¹⁵N-HMQC spectra of OutF⁶⁵⁻¹⁷² were measured in the absence or in the presence of OutE, OutL¹⁻²⁵⁷ and OutE-OutL¹⁻²⁵⁷ complex respectively. 20mM Tris-HCl 150mM NaCl pH 7.0 buffer was used as selected from thermofluor assay shown in Fig. 3.29.

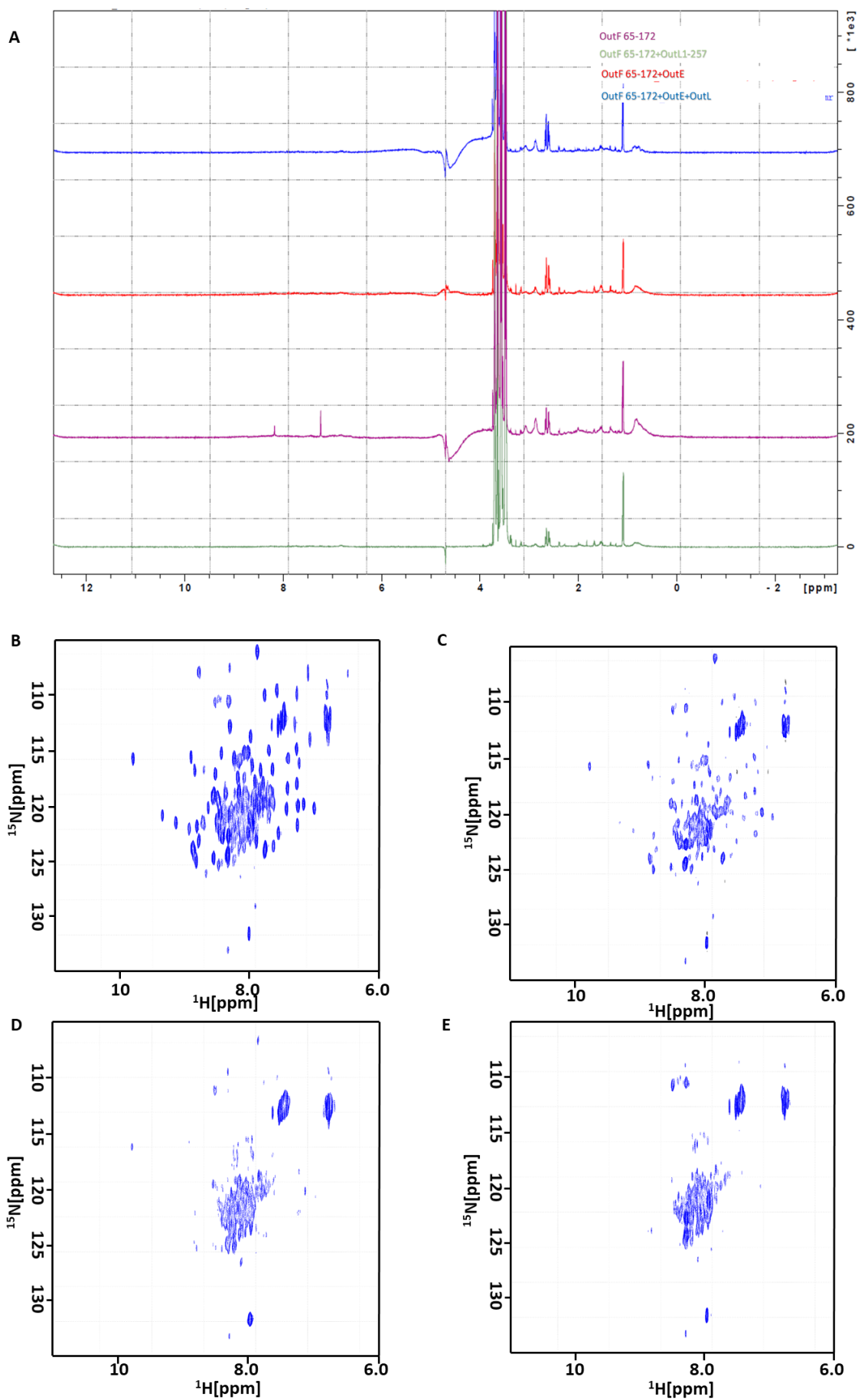


Figure 3.31 Elucidation of OutF⁶⁵⁻¹⁷² and OutE or OutL¹⁻²⁵⁷ or OutE-OutL¹⁻²⁵⁷ complex interactions

A: 1D spectra of OutF⁶⁵⁻¹⁷² (purple) alone, in presence of OutE (red), OutL¹⁻²⁵⁷ (green) and OutF⁶⁵⁻¹⁷² and OutE-OutL¹⁻²⁵⁷ complex (blue). No peak shifts indicate that pH of Tris buffers is the same. B: ¹H-¹⁵N-HMQC spectra of the OutF⁶⁵⁻¹⁷² alone. C: OutF⁶⁵⁻¹⁷² in the presence of OutL¹⁻²⁵⁷. OutF⁶⁵⁻¹⁷² concentration is 0.05mM. The ratio between OutL¹⁻²⁵⁷ and OutF⁶⁵⁻¹⁷² is 1.2:1. D: Spectra of OutF⁶⁵⁻¹⁷² in the presence OutE¹⁻⁵¹³. (E) Spectra of OutF⁶⁵⁻¹⁷² in the presence of OutE¹⁻⁵¹³-OutL¹⁻²⁵⁷ complex.

Binding of OutF⁶⁵⁻¹⁷² with OutL¹⁻²⁵⁷ or OutE¹⁻⁵¹³ or OutE¹⁻⁵¹³/ OutL¹⁻²⁵⁷ results in loss of signal. These results show that both OutL¹⁻²⁵⁷ and OutE¹⁻⁵¹³ can individually interact with OutF⁶⁵⁻¹⁷² which are consistent with previous yeast two-hybrid assay in which the interactions depend on measurement of β -galactosidase activity (Py, Loiseau et al. 2001).

3.2.10 Study of the interaction of OutF⁶⁵⁻¹⁷² and OutE¹⁻⁵¹³-OutL¹⁻²⁵⁷ by thermofluor assay

To study the interaction between OutF⁶⁵⁻¹⁷² and the OutE¹⁻⁵¹³-OutL¹⁻²⁵⁷ complex, the thermofluor assay was used. The sample was prepared as described above. Two different ratios between OutF⁶⁵⁻¹⁷² and the OutE¹⁻⁵¹³-OutL¹⁻²⁵⁷ complex were prepared based on a previous study (Lu, Turley et al. 2013). Since the GspE forms a hexamer (Lu, Turley et al. 2013) and GspF may form a dimer in this study a 3:1 ratio of OutF⁶⁵⁻¹⁷² : OutE¹⁻⁵¹³-OutL¹⁻²⁵⁷ complex was prepared (Fig. 3.32). To make sure all the individual OutF⁶⁵⁻¹⁷² was titrated, a 1:1 ratio of that was also prepared (Fig. 3.32).

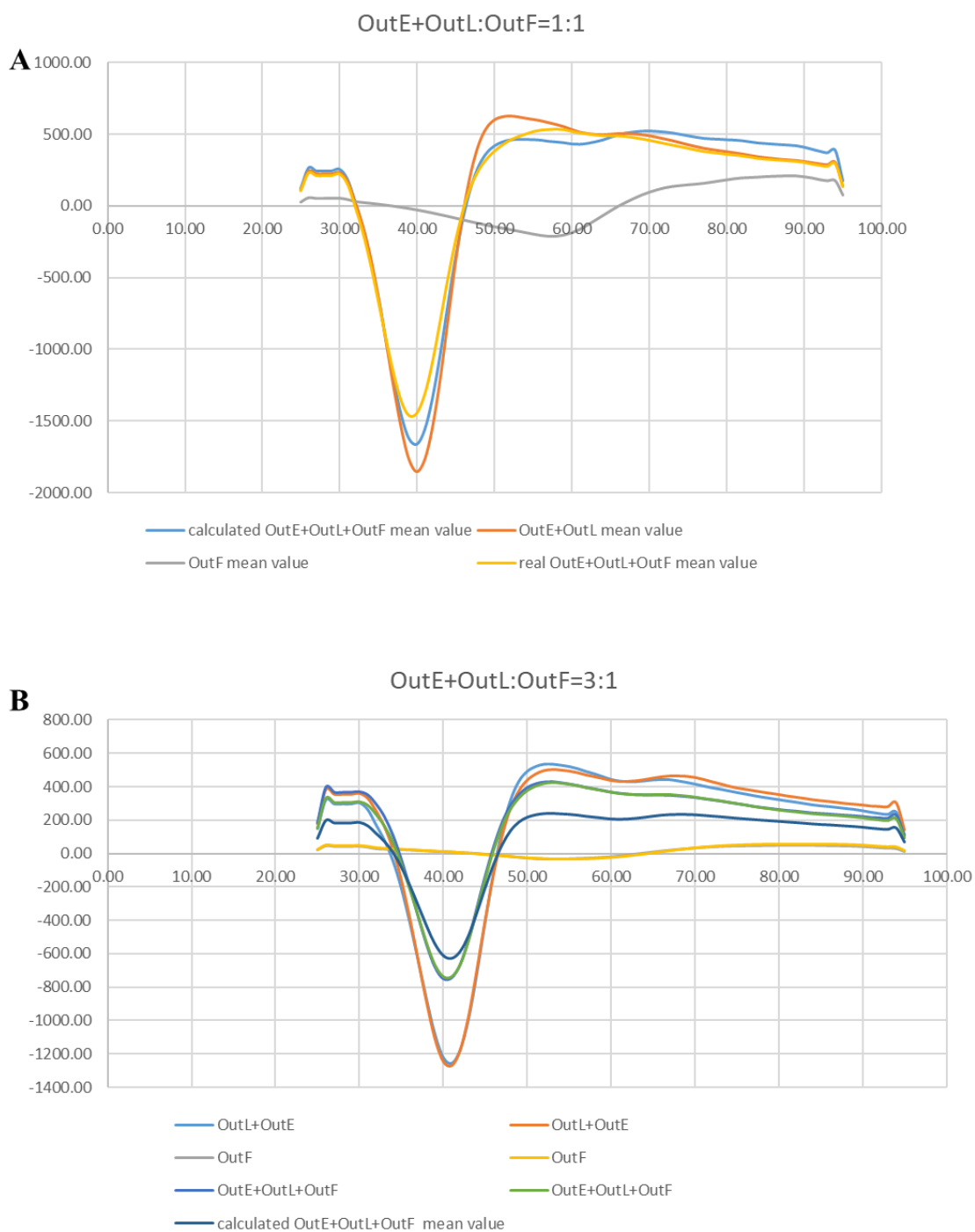


Figure 3.32 Thermofluor plots of interaction between OutF⁶⁵⁻¹⁷² and OutE-OutL¹⁻²⁵⁷ complex

A: 1:1 ratio between OutF⁶⁵⁻¹⁷² and OutE-OutL¹⁻²⁵⁷ complex; B: 3:1 ratio between OutF⁶⁵⁻¹⁷² and OutE-OutL¹⁻²⁵⁷ complex. The buffer was 20mM Tris, pH 8.0, 500mM NaCl.

This experiment was repeated several times but no obvious change was observed between the measured and calculated melting temperature of OutE-OutL¹⁻²⁵⁷ complex after adding OutF⁶⁵⁻¹⁷².

3.3 Summary and discussion

In this Chapter I report cloning, production, crystallization, and the crystal structure of the first cytoplasmic domain of OutF (residues 65 to 172; cyto I OutF) solved at 2.15 Å resolution. This domain consists of 108 residues which form a bundle of six anti-parallel helices. Two OutF⁶⁵⁻¹⁷² domains form a reasonably tight dimer organised so the following trans-membrane helices (TMH1, Fig. 3.21) can insert into the inner-membrane. In contrast to the structure of *Vibrio cholera* EpsF, OutF⁶⁵⁻¹⁷² does not have the calcium-binding sites that stabilize the dimer interface and does not need calcium for stabilization. I also compare the sequences and structures of nine species of GspF family proteins from T2SS and T4PS in Gram-negative bacteria, T4PS, Com and Tad in Gram-positive bacteria and archaeal flagella. Across these structures eight residues are conserved at the inside face of the cylinder formed by the cytoI α helix bundle. The sequence and structure comparison results show the close relationship between T2SS and T4PS. Two different dimer models of cytoplasmic domain II were built by similarity with known dimer structures, the one that was formed based on the PilC structure is more stable than that of cytoplasmic domain I of OutF indicating the two cytoplasmic domains of OutF may form a dimer in different ways, both allowing the insertion of the following transmembrane helices into the inner-membrane. In this Chapter, I also explore the interaction between OutF and the other inner-membrane proteins using NMR spectroscopy, pull down assay and thermofluor. OutF⁶⁵⁻¹⁷² was shown to interact with OutL, with OutE, and with the the OutL/OutE complex using NMR spectroscopy. While pull down assays detect the strong

interaction between OutL and OutE, the weaker interaction between OutF cyto I and the OutL/OutE complex can only be detected using NMR spectroscopy which can detect much weaker interactions.

Constructs producing full-length OutF were also successfully made, protein produced, solubilized using 1%DDM, and small crystals grown. Further work on the membrane protein OutF could be beneficial, but I decided to concentrate future studies on the inner-membrane complex of proteins as described in the next results Chapter (Chapter 4).

Chapter 4 Structural studies of the inner-membrane platform of the *D. dadantii* T2SS

4.1 Overview

The inner-membrane platform of T2SS comprises three single transmembrane helix proteins (OutL, OutM and OutC) and one polytopic membrane protein, OutF. In the *D. dadantii* secretion system there is an additional single transmembrane protein, OutB, which is responsible for locking the inner-membrane platform to the outer-membrane secretin. GspC has this function in other species including *Vibrio cholera*. Several crystal structures of the soluble domains of the inner-membrane proteins have been solved, however the overall architecture of this platform remains elusive. In this Chapter, I first cloned the gene operon *EFGHIJKLM* and purified the inner-membrane complex (including three inner-membrane proteins OutF, OutL, OutM, five pseudopilus subunits (OutG, OutH, OutI, OutJ and OutK) and the ATPase (OutE) with the molecular size of approximately 1million Dalton as estimated by gel filtration and dynamic light scattering (DLS). The purified protein complex was then subsequently analysed by mass spectrometry and negative staining transmission electron microscopy (TEM). Finally, I built a model at 32.7Å resolution solved by single particle analysis, using the workflow described in Figure 4.1, and fitted a ring in the structure with crystal structure of the hexameric ATPase, GspE. However, due to the conformational variability of the particles, this model was built from only 11952 particles of the 1300000 imaged in total. Future work will focus on increasing the homogeneity of the particles.

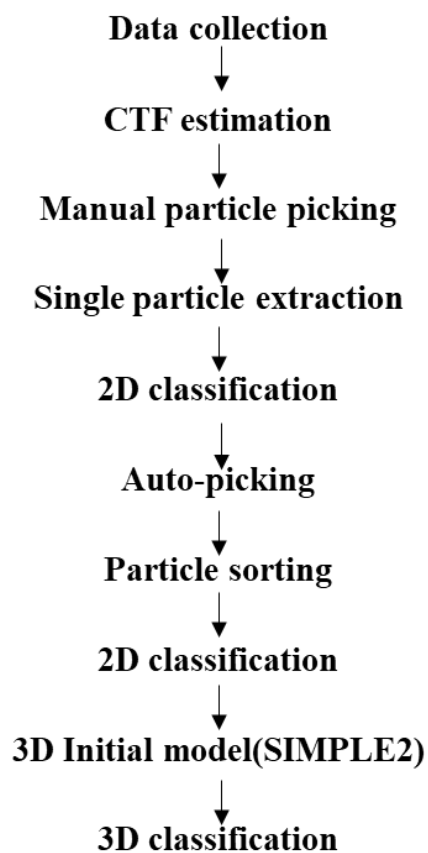


Figure 4.1 Flow chart of single particle reconstruction process in Relion2(Fernandez-Leiro and Scheres 2017).

4.2 Results

4.2.1 Cloning the genes corresponding to the inner-membrane complex

a Constructs

Constructs generated in this chapter are listed in Figure 4.2.

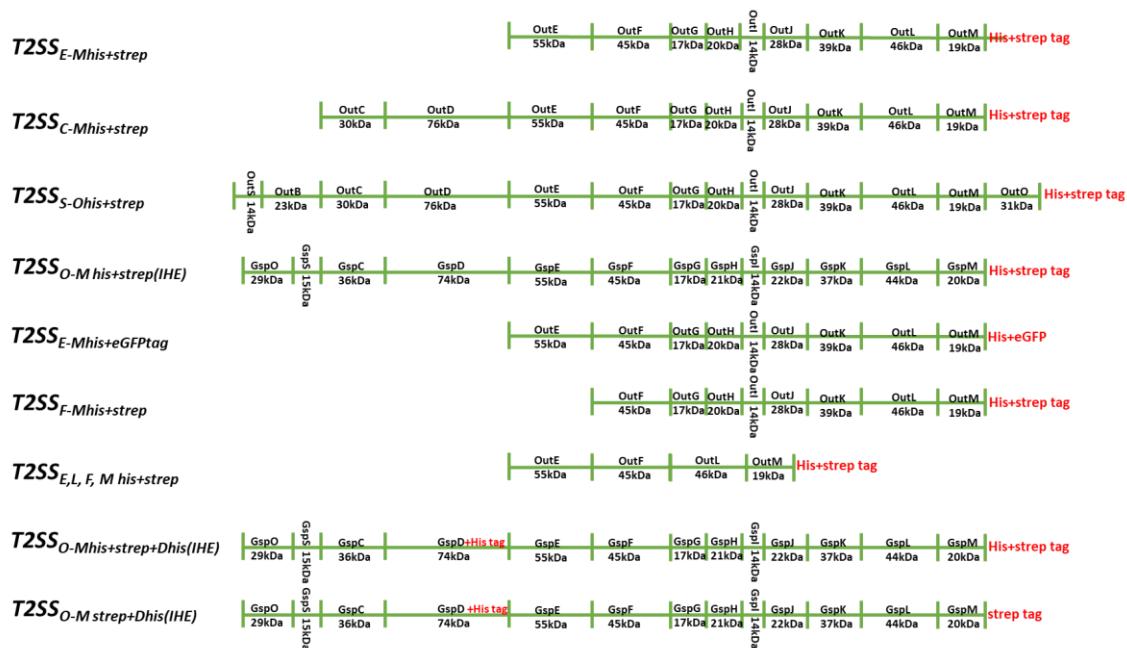


Figure 4.2 Genetic organization of T2SSs from *E. coli* IHE 3034 and *D. dadantii*.

Operons are separated by green lines.

b PCR products

The gene operon *EFGHIJKLM* from *outE* to *outM* was amplified from the genome of *D. dadantii*, using the primers presented in Table 4.3 and the gene operon *OSCDEFGHIJKLM* from *gspO* to *gspM* was amplified from the genome of *E. coli* IHE strain using the touch-down polymerase chain reaction as described in Chapter 2. Two further constructs *outE*-deleted and pseudopilus-deleted DNA were amplified from the pASK3c: T2SS_{E-M} plasmid.

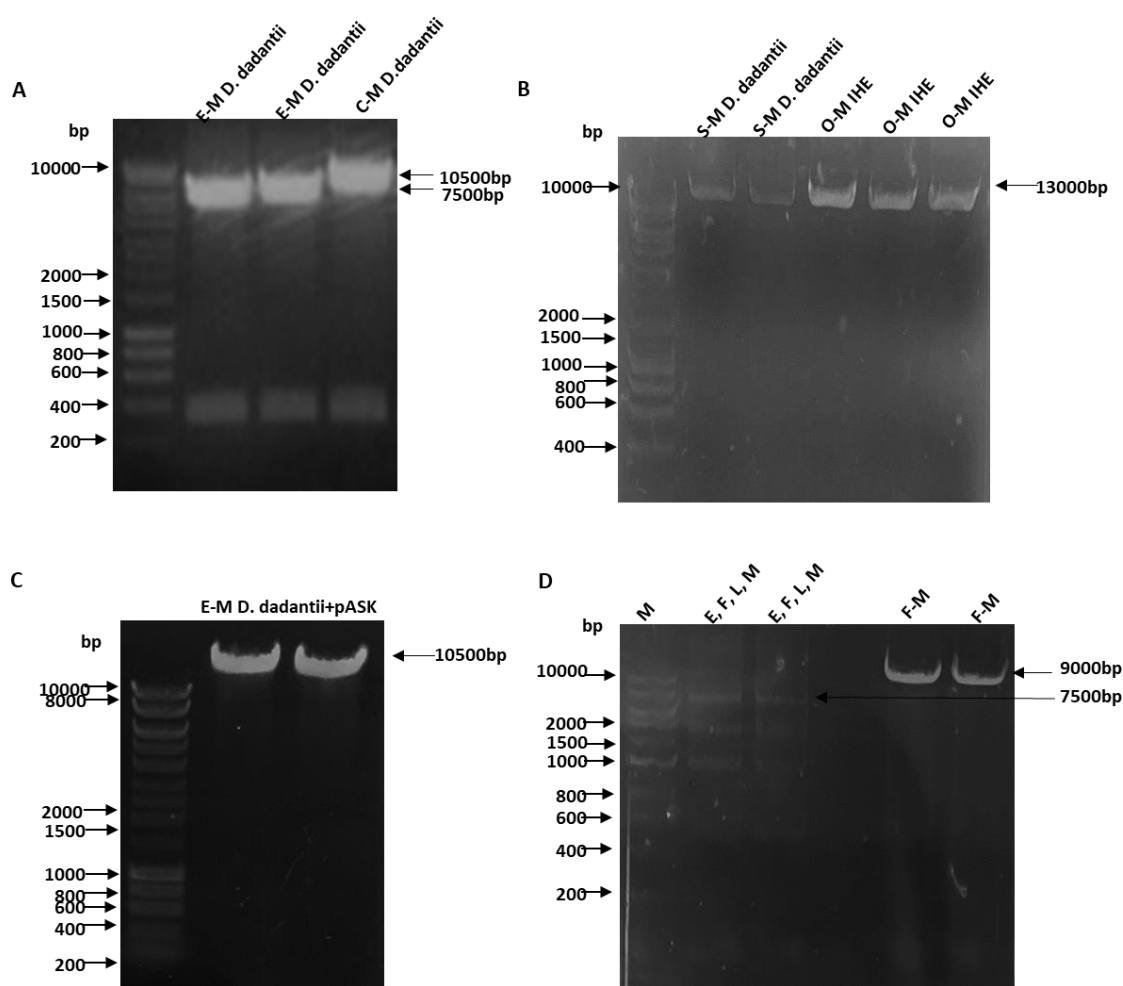


Figure 4.3 PCR products of T2SS gene operon from *D. dadantii* and IHE strain

A: PCR product of inner-membrane platform from *outE* to *outM* (lane1 and lane2) and T2SS proteins from *outC* to *outM* (lane3) with strep tagged OutM from *D. dadantii*; B: Inner- and outer-membrane protein gens *OutS-OutM* from *D. dadantii* (lane1 and lane 2) and *GspO-GspM* from IHE strain (lane 3 to lane 4); C: PCR product of *OutE* to *OutM* plus pask-3c vector with his tagged OutE; D: Inner-membrane proteins without pilus proteins (lane1 and lane 2) and OutE-deleted inner-membrane platform (lane 4 and lane 5) from *D. dadantii*.

The PCR products were then purified from the agarose gel and cut using restriction enzymes.

c Constructions of pASK3c: T2SS_{E-M}, pASK3c: T2SS_{C-M}, pASK3c: T2SS_{E, F, L and M} and pASK3c: T2SS_{F-M}

The PCR product *EFGHIJKLM* was digested using *SacI* and *XhoI* and ligated into the pASK-IBA3C (IBA) vector to produce the vector pASK3c: T2SS_{E-M} which was linearized by cutting using *SacI* and *XhoI*. Ligation was performed as described in chapter 2. In this construct, the His tag and Strep tag were incorporated at the C-terminal of OutM used for purification of the T2SS_{E-M} complex.

The PCR product *OSCDEFGHIJKLM* from *E. coli* IHE was cloned into the pASK-IBA3C (IBA) vector (pASK3c: T2SS_{C-M}) using the same method as the T2SS_{E-M} complex.

Two further constructs OutE-deleted and pseudopilus-deleted PCR product included the vector sequence and subsequent to self-ligation with sticky or blunt end as described in Chapter 2.

d Enzyme cleavage to analyze the new constructs

After miniprep from the positive colonies, the new constructs are confirmed by double digestion with *XhoI* and *SacI* restriction enzymes which are the same enzyme used before ligation.

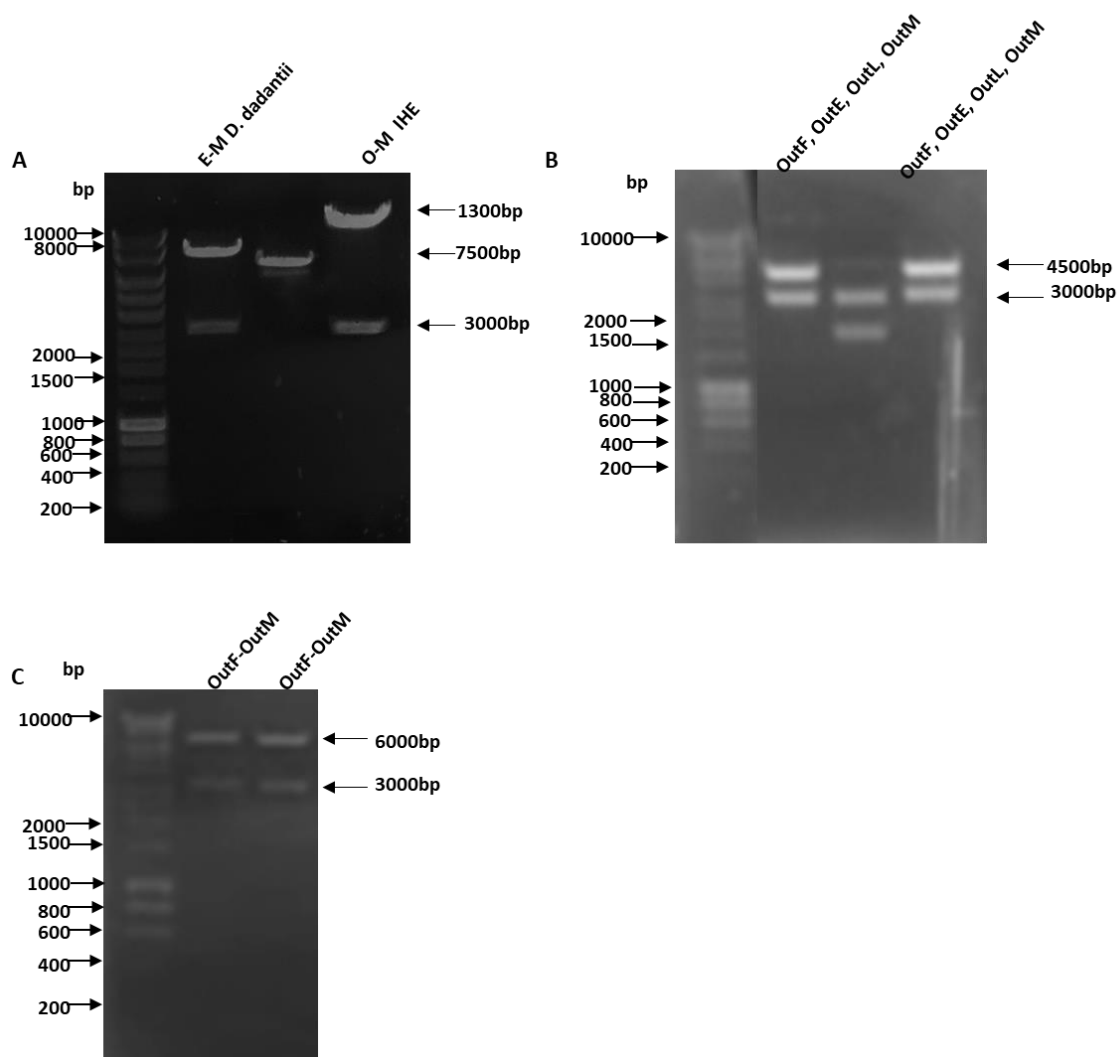


Figure 4.4 Restriction enzyme results

A: Lane 1 and lane 3 are the positive constructs of inner-membrane platform from *D. dadantii* and T2SS proteins from *E. coli* IHE, respectively; B: Lane1 and lane 3 are the positive constructs of pilus-deleted inner-membrane platform from *D. dadantii*; C: The double digestion results of OutE-deleted inner-membrane platform.

The positive constructs were then sent for sequencing before transformation into the expression cells.

e Expression of $pASK3c: T2SS_{E-M}$ and $pASK3c: T2SS_{O-M}$

Small volume expression tests were used to explore the usefulness of the two vectors:

pASK3c: T2SS_{E-M} and pASK3c: T2SS_{O-M}. The expression method was as described in Chapter 2. Samples (200µl) were collected before and after induction using AHT. 20µl of each sample was used for SDS-PAGE analysis. The proteins were then transferred from the gel to the PVDF membrane and a Western blot was employed to detect the presence of the target proteins.

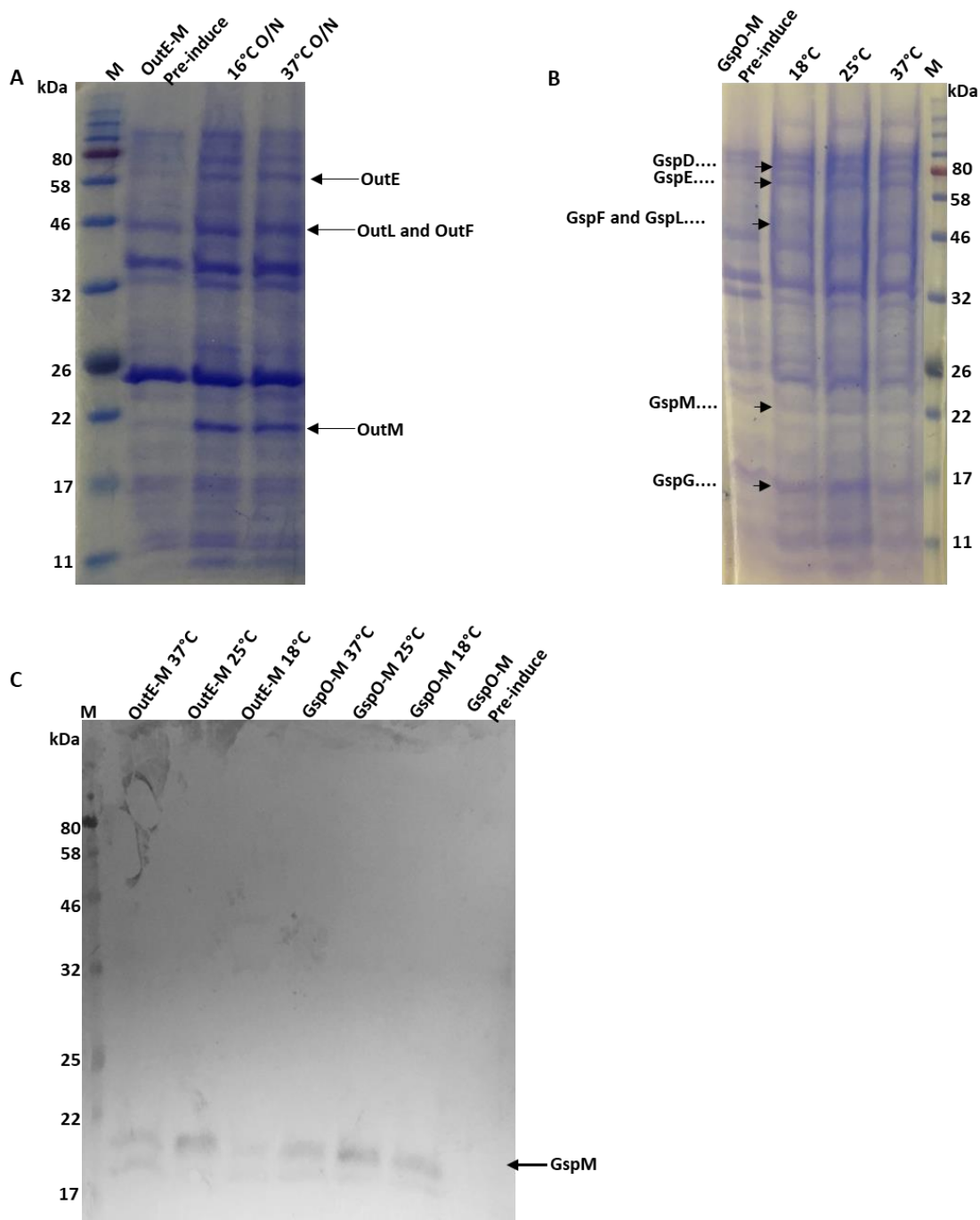


Figure 4.5 Expression test of the inner-membrane platform from *D. dadantii* and the whole complex of the T2SS from IHE3034, respectively.

A: Production of the inner-membrane platform proteins OutE, OutF, OutL and OutM from the pASK3c: T2SS_{E-M} construct. The level of protein expression at 16°C (lane 2) is similar to that at 37°C (lane 3) and greater to the level before induction (lane 1). B: Expression of the inner-membrane and outer-membrane proteins including the prepilin peptidase GspO from GspO to OutM: it is hard to see the proteins expression either at 16°C or at 37°C from SDS-PAGE analysis. However, the subsequent purification step indicated that at least some of the proteins were produced successfully. C: Western blot result of expression of inner-membrane platform and T2SS. The anti-His antibody was used for detection.

It is clear from the SDS-PAGE and Western blot that proteins are being successfully produced from the *D. dadantii* construct. OutM, and presumably a breakdown product of OutM, with His-tag are being detected on the western, and there is reasonable evidence of a greater quantity of OutM, OutL, and OutE being produced after induction.

4.2.2 Purification of the expressed inner-membrane complexes

a Membrane fraction preparation

All the membrane fractions were prepared using the method described in Chapter 2 except the cells were broken using French press rather than sonication. The cell pellets for outer-membrane expression were used immediately after harvesting without freezing. Both membrane fractions of inner- and outer-membrane were used immediately without freezing.

b Purification of membrane protein complex

T2SS_{E-M} complex, OutE deleted complex and OutE, OutF, OutL and OutM complex were purified using the methods described in Chapter 2, but with modifications as described below: 0.5% DDM (Fisher Scientific), 0.75% w/v Decyl Maltose Neopentyl Glycol (DM-

NPG, Anatrace), 0.1% w/v digitonin (Anatrace) in 20mM Tris, 150mM NaCl, pH 7.4 buffer was used to solubilize the membrane fraction at room temperature for 40 min. The suspension was then clarified by centrifugation at 100,000g for 20 min. The supernatant was diluted two-fold and loaded onto a PD-10 column packed with Strep-Tactin® Sepharose® and then washed with 20mM Tris-HCl pH 7.4, 150mM NaCl, 0.06% w/v DM-NPG, 0.1% w/v digitonin, 1mM DTT and 1mM EDTA at 4°C. The purified T4SS_{O_M} complex was eluted in the same wash buffer supplemented with 2.5mM desthiobiotin (IBA). The sample was used immediately for electron microscopy.

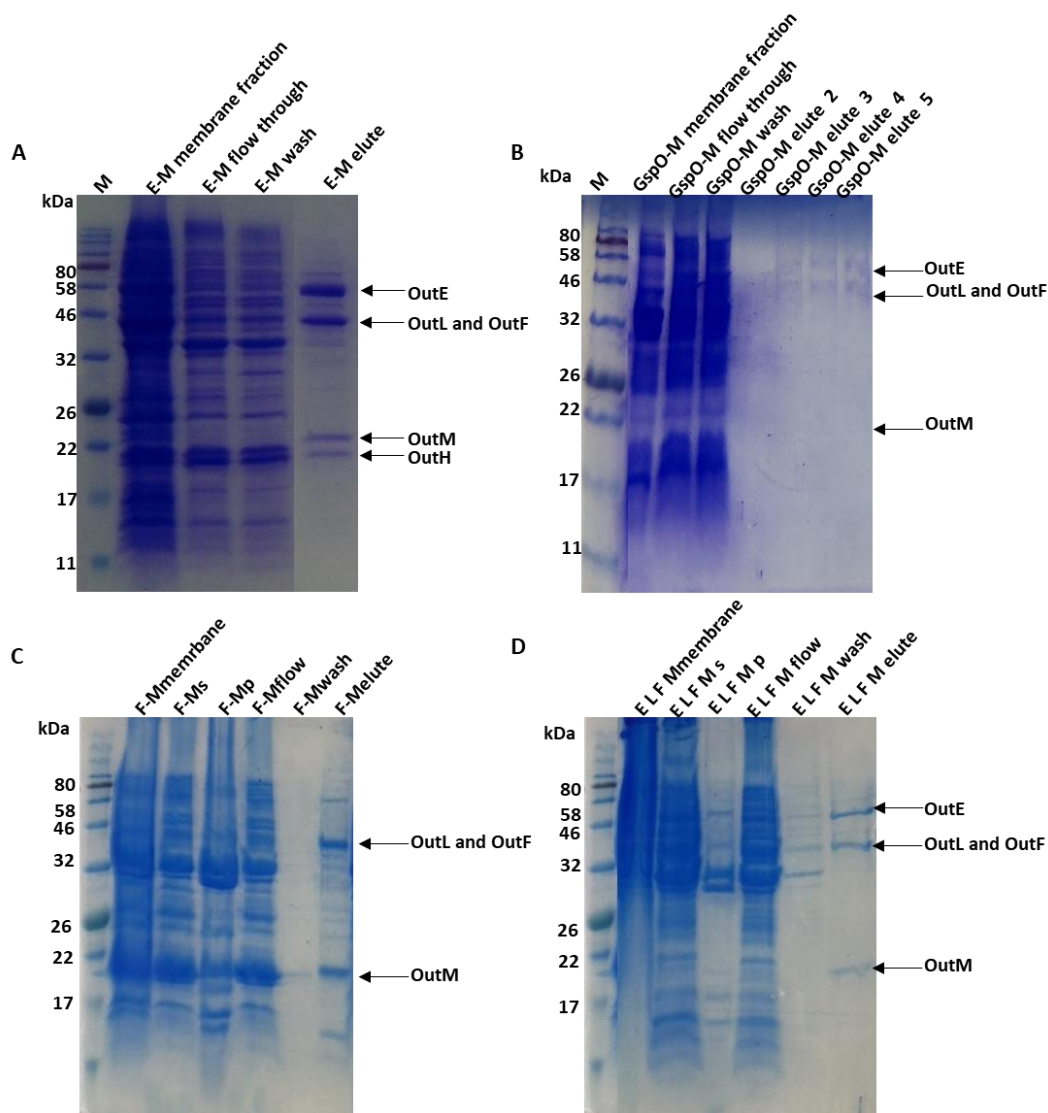


Figure 4.6 Purification of T2SS complexes and gene mutated complexes from *D. dadantii* and IHE using strep tactin resin.

A: Purification of the inner-membrane platform from the pASK3c: $T2SS_{E-Mhis+strep}$ construct coding from OutE to OutM. B: Purification of the inner- and Outer-membrane complex from the pASK3c: $T2SS_{O-Mhis+strep}$ construct coding from GspO to GspM. C: Purification of the OutE deleted mutation of inner-membrane platform from the pASK3c: $T2SS_{F-Mhis+strep}$ construct coding from OutF to OutM (F-Ms: supernatant, F-Mp: pellet). D: Purification of the inner-membrane platform from the pASK3c: $T2SS_{E, L, F, Mhis+strep}$ construct coding OutE, OutL, OutF and OutM.

After purification from the strep tactin, the proteins are loaded onto the size exclusion column Superose 6 and further purified according to proteins size after concentrating the elution fraction using a 100kDa cutoff concentrator.

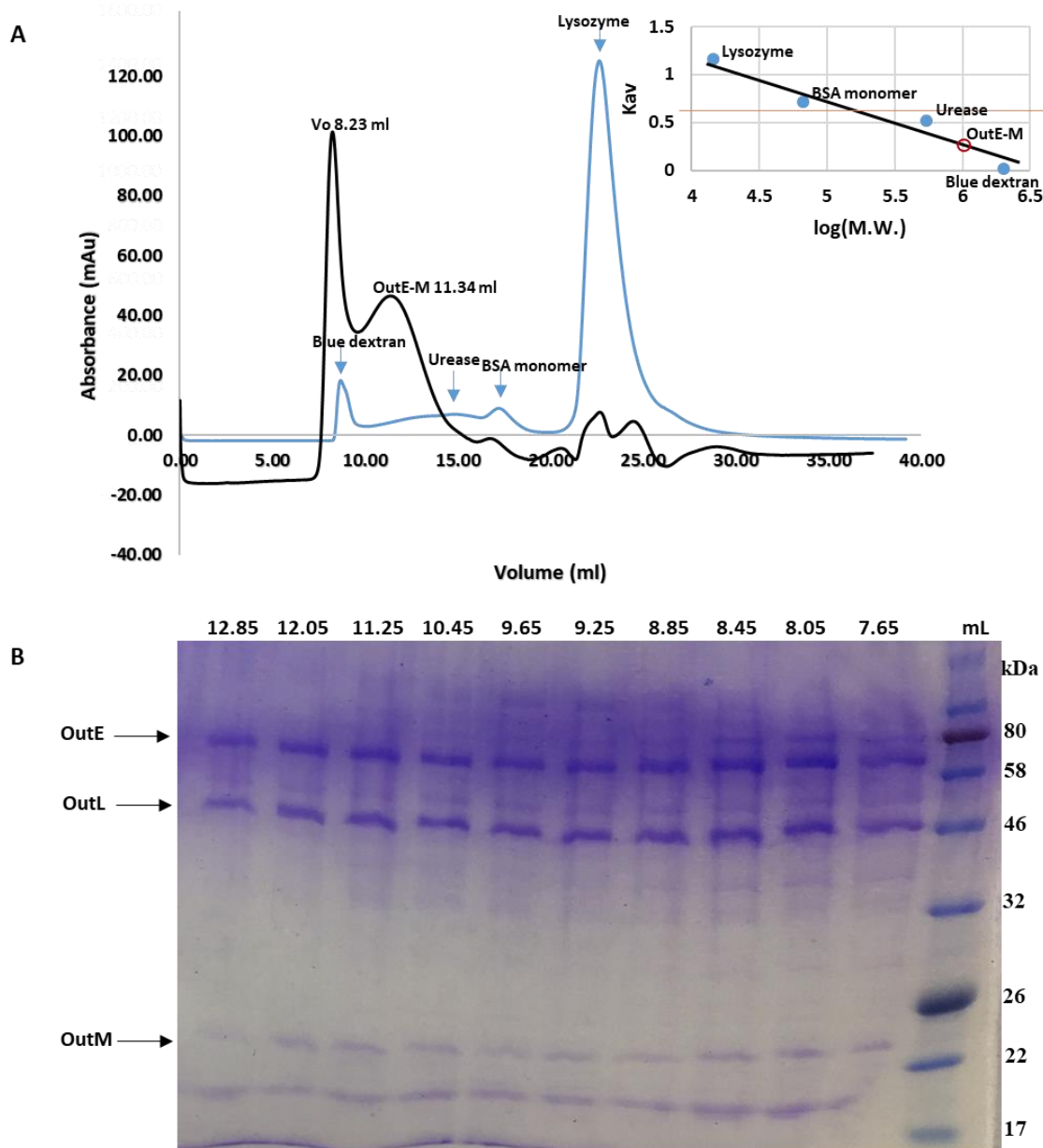


Figure 4.7 Chromatographic separation on Superose 6 10/300 GL column of T2SS_{E-M} complex

A: Superose 6 10/300 GL size exclusion chromatography column was equilibrated with buffer (20mM Tris-HCl, 150mM NaCl, 0.05%DDM, 1mM DTT, pH 8.0). The inner-membrane complex elutes at molecular weight of around 1000kDa based on calibration result with standard proteins (blue dextran 2000 (2000kDa), Urease (545kDa), BSA (66kDa) and lysozyme (14.4kDa)). Approximately 24 ml volume were collected with each fraction corresponding to 400µl retention volumes and 30 µl of each were analyzed by 15% SDS-PAGE (B).

c Sucrose gradient centrifugation of inner-membrane complex OutE-M

After the membrane fraction was solubilized and cleared by ultracentrifugation the supernatant was loaded onto a 10-50% linear sucrose gradient prepared as described in Chapter 2 but in addition with 0.03% DDM. The gradients were centrifuged at 150,000 g for 16 h at 4°C in a swinging-bucket rotor in a Beckman ultracentrifuge. Each fraction was collected and analysed by TEM and SDS-PAGE. Figure 4.8 D-H show the TEM images of the sucrose gradient fractions prepared in Figure 4.8A and B.

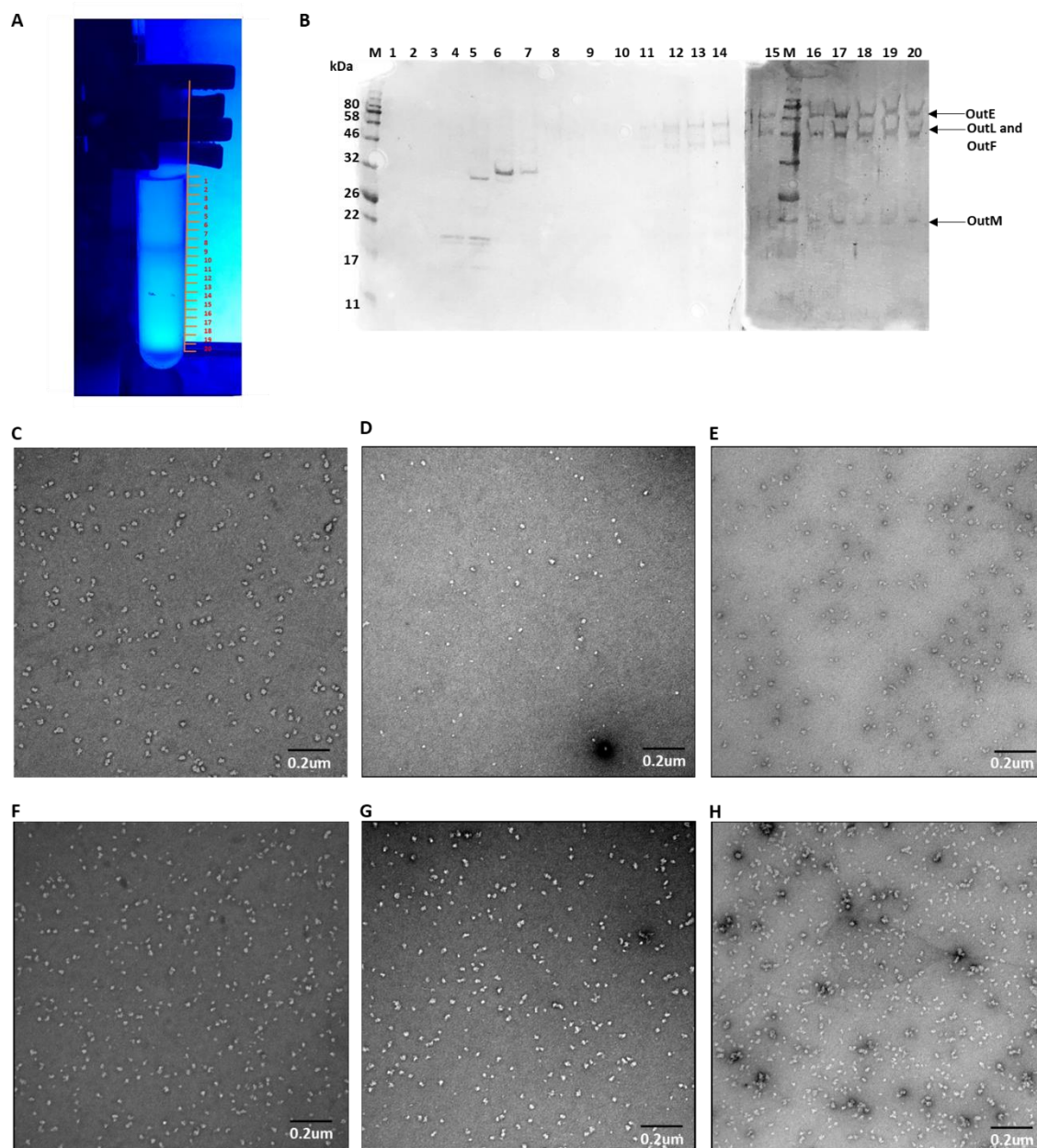


Figure 4.8 Analysis of sucrose gradient fraction

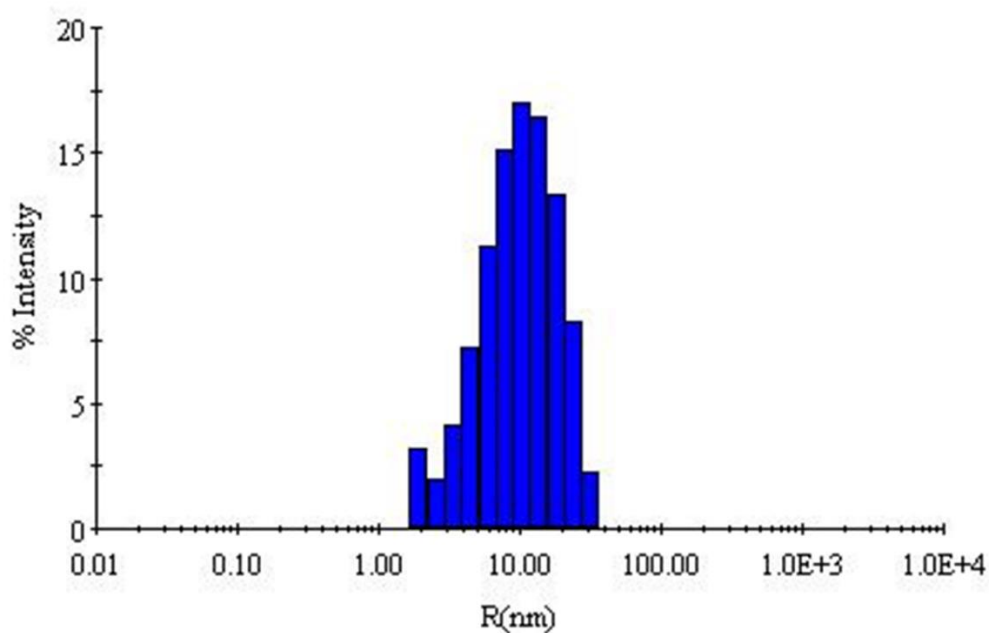
A: Gradient fractionate; B: SDS-PAGE of each fraction of sucrose gradient. C: Image of peak fraction from size exclusion column. D-H: TEM images of sucrose gradient fractions number 11, 13, 14, 15, 17 (B) respectively.

Since the molecular weight of the OutE-M particle is around 1000kDa according to the SEC and DLS results (Fig. 4.7A and Fig. 4.9), the protein is anticipated to be located in approximately 30% sucrose solution which is a little below the middle of the gradient (Fig. 4.8 A – numbers 13, 14, 15). However, quite a lot of protein was found close to the bottom of the gradient (40%~50% sucrose) indicating larger molecular masses formed which is consistent with the SEC result with a large peak in the void volume.

As dispersed protein complexes are anticipated in fraction 13, 14 and 15, the images of particles from these fractions were then compared with that from gel filtration. The results confirmed that the size of particles from number 11 fraction is a little smaller than that from SEC peak fraction. The proteins tend to aggregate in fraction 17 and form very large particles. The size of particles from fractions 13, 14, 15 is correct but the particles are still heterogeneous. This result is consistent with predicted location of protein according to the concentration of the sucrose gradient.

d Dynamic light scattering (DLS)

45µl of the eluent from the peak of the gel filtration column were transferred to a square cuvette for DLS measurements. The light scattering intensity caused by the Brownian motion of the protein were measured and molecular weight determined by the instrument.



Item	R(nm)	%Pd	MW-R(kDa)	%Int	%Mass
Peak1	2.19	14.0	21	5.1	65.1
peak2	12.14	54.1	1159	94.9	34.9

Figure 4.9 T2SS_{OutE-M} DLS summary.

OutE-OutM was measured in 20mM Tris, 150mM NaCl, 0.04%DDM, pH7.4 at 15°C.

The OutE-OutM sample shows a polydispersity of 54%, which indicates OutE-OutM complex is heterogenous in solution consistent with single particle analysis result. The average molecular mass calculated from the radius of the protein is approximately 1159kD, the size of which is close to that from size exclusion column (Fig. 4.7A).

e Mass Spectrometry

Mass spectrometry was used to identify the proteins expressed and purified from the T2SS_{E-M} construct. Two different methods were used to prepare the protein samples for

protein fingerprint. The elution from the Superpose 6 column was exchanged into low salt and low detergent buffer using PD10 column. Proteins were separated by SDS-PAGE analysis and stained using coomassie brilliant blue G-250 (1% coomassie G-250, 40% methanol, and 10% acetic acid). The bands of interest (55 kDa, 46kDa, 22kDa and 20kDa) were cut out cleanly and put in Eppendorf tubes (with 50 μ L HPLC water) separately. The eluted solution and gel bands were sent to Cambridge Centre for Proteomics for sequencing. The proteins were analysed using Mascot (Matrix Science, London, UK; version 2.5.0) and Scaffold (version Scaffold_4.3.0, Proteome Software Inc., Portland, OR) was used to validate MS/MS based protein identifications.

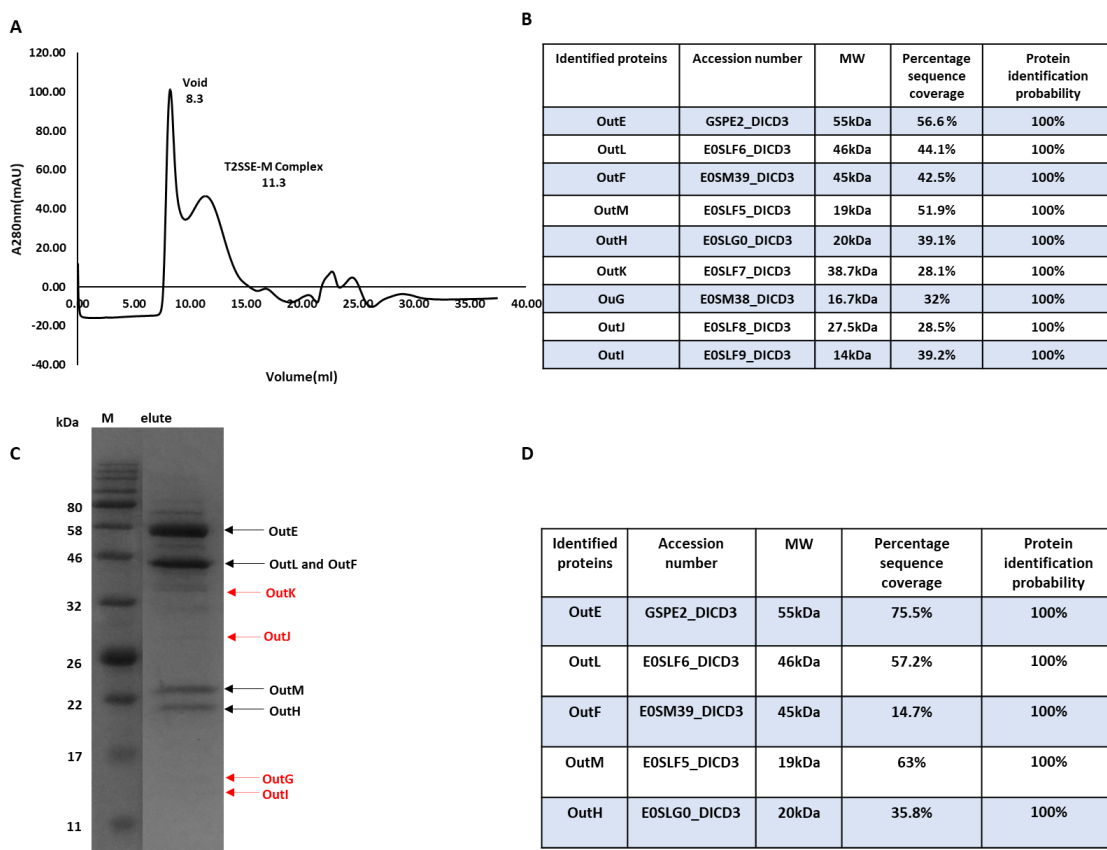


Figure 4.10 Mass spec result of T2SS_{E-M} complex

(A) T2SS_{E-M} co-elute as single peak during gel filtration. T2SS_{E-M} complex elute at 11.3 ml with molecular weight of ~1million Dalton. (C) SDS-PAGE of T2SS_{E-M}. Gel bands of OutE, OutL and OutF, OutM and OutH were confirmed by following mass spectrometry. Black labelled proteins:

identified protein after MALDI mass spec of gel slices. Red labelled proteins: the possible existing proteins according to the protein size and N-terminal sequencing result from elution. (B) Summary of the mass spec (MALDI-TOF) result of identified proteins from the elute solution. (D) Summary of mass spec result of proteins from gel slices (fragment tolerance: 0.100Da (Monoisotopic), parent tolerance: 10.0 PPM (Monoisotopic), fixed modification: +57 on C (Carbamidomethyl), digestion enzyme: trypsin, mass missed cleavages: 2, peptide threshold: 95% minimum, protein threshold: 95% minimum, peptide FDR: 0.5%, protein FDR: 0.0%).

Figure 4.10 shows the purified T2SS_{E-M} complex from strep tactin resin and SEC column included all the proteins expressed from gene operon *EFGHIJKLM*. The mass spec result also confirmed the 4 clear bands on SDS-PAGE to be OutE, OutL (and OutF), OutM and OutH respectively.

4.2.3 Initial three-dimensional model of inner-membrane platform

a Sample preparation

TEM grids were glow discharged in air and used immediately. 5 μ l of 5 μ g/mL T2SS_{E-M} complex in 20mM Tris-HCl, 150mM NaCl, 1mM DTT, 0.04%DDM, pH7.4 buffer was deposited on glow discharged carbon coated grids (SPI) (300 meshes) for 1 min followed by blotting with or without washing with ddH₂O and staining with 2% (v/v) uranyl acetate for 0.5-1.5 minutes once or twice.

b Data collection

The JEOL JEM-1230 was used to check sample quality before large scale data collection, gold labelling image collection, and mutated sample image collection in QMUL. Large data sets were collected at Imperial College using the FEI Tecnai F20.

Images taken on FEI Tecnai 20 transmission electron microscope operating at 200 kV were recorded with FEI Falcon II CMOS direct electron detection camera. 834 micrographs were collected automatically without intermediate frames using the EPU (Tan et al. 2016) software at a nominal magnification of 29,000 corresponding to a pixel size of 3.48 Å/pixel using an electron dose of 7 e/Å²/s, a defocus range of -0.5 to -3.5 μm and an exposure time of 1s.

c Data processing programs

Programs used for data processing and model evaluation are described in Chapter 2.

d Particle selection and normalisation

After image collection, single T2SS_{OutE-M} particles were selected using manual particle picking in RELION 2. Since the maximum dimension measured for a T2SS_{OutE-M} particle was around 250 Å shown in the TEM micrographs (Fig. 4.8C), the box size was set to ~556.8 Å (160 pixels) to fully accommodate the particle in the box and also cover enough background for processing. Picked particles were digitally cut out of the micrograph and stored in an array of individually addressable pictures in a star file. The particles were normalised by matching the grey value in RELION.

e CTF estimation

The CTF can be estimated since the exact value of defocus and astigmatism values of an images can be obtained by comparing the power spectra of micrographs with simulated CTF. The defocus values are required to correct CTF on images. In this work, CTF was estimated by running CTFIND4 in RELION.

f Getting templates for auto-picking

2213 particles were manually picked and the Reference-free 2D classification (RELION2) was then executed for these particles. 10 2D classes were obtained with 350Å mask size. In this method, the entire data set was used for alignment. All particle images are initially averaged to centre the particles before doing maximum likelihood calculations for each particle to belong a separate class.

g Auto-picking and particles sorting

The particle auto-picking process was used with 10 2D classes as template and in total, 127283 single OutE-OutM particles were extracted from 834 micrographs.

h 2D classification to remove bad particles

300 class averages were generated and each contained a few hundred particles that shared similar size, shape and orientation so that the signal was intensified and physical detail of the protein could be seen. The class distribution and resolution plot were generated by gnuplot and used to get rid of bad classes (Fig. 4.11).

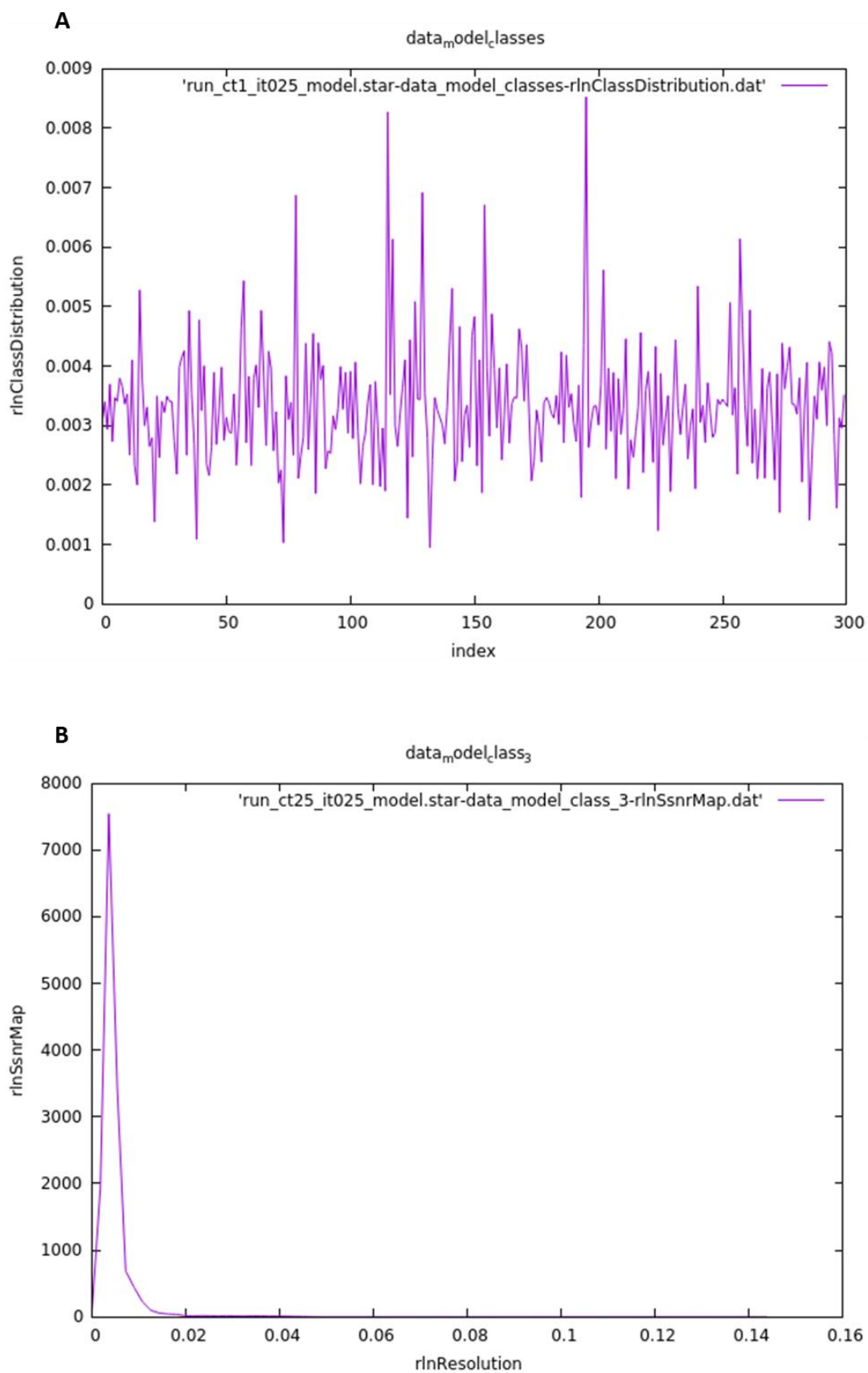


Figure 4.11 plots of distribution and resolution of 2D classes

A: Plot of distribution of 2D classes; B: Plot of resolution of 2D class (e. g. No. 3 class). Classes with few particles are low resolution while particles with many classes are high-resolution. Lower signal-to-noise ratios will lead to lower resolution.

The plots of distribution and resolution of 2D classification were used to monitor the quality of particles in each of 2D classes and get rid of bad particles. Bad particles do not average well as there are often few particles in each bad class, and the resolution of the corresponding class average is thus very low. These classes look like ghost. These bad classes are removed using Subset Selection in RELION. 220 2D classes were selected from 300 classes with 90580 particles selected.

i Generating template for 3D classification using SIMPLE

SIMPLE software was used to generate an initial 3D model. This model was generated from 220 2D classes with 350Å circular mask same as 2D classification.

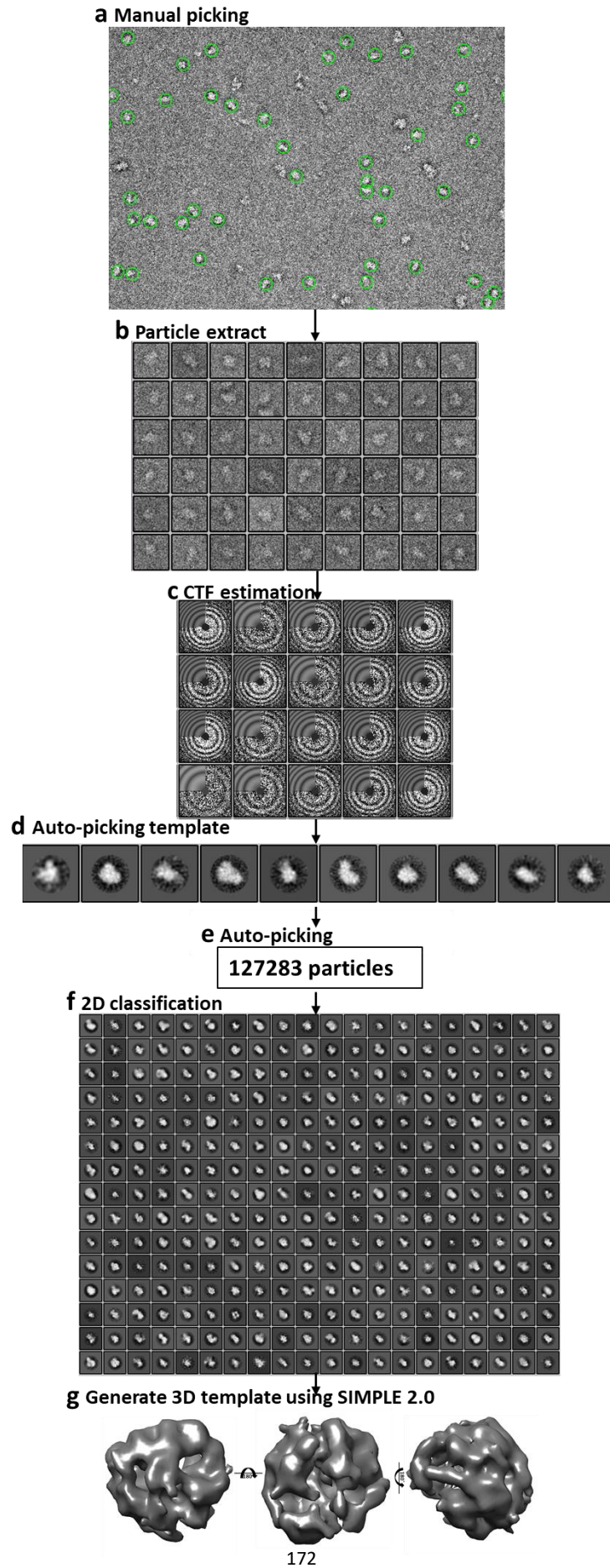


Figure 4.12 Data process to generate the initial model for 3D classification

The process was described as following:

a and b: 2213 manually selected particles using RELION were extracted and stored in a star file.
 c: CTF were estimated by using CTFFIND4 and images with good CTF were selected.
 d: 10 2D classes were generated from manually picked particles and 8 classes were selected as templates for auto-picking.
 e: 127283 particles were auto-picked using the 2D reference and 120139 particles were selected after particle sorting and selection.
 f: 300 2D classes were generated from those sub-selected particles and the quality of the classes were monitored by the plots of distribution and resolution. 220 classes were selected and used for initial 3D model building.
 g: Initial 3D model was obtained from SIMPLE and were used for 3D classification in Relion.

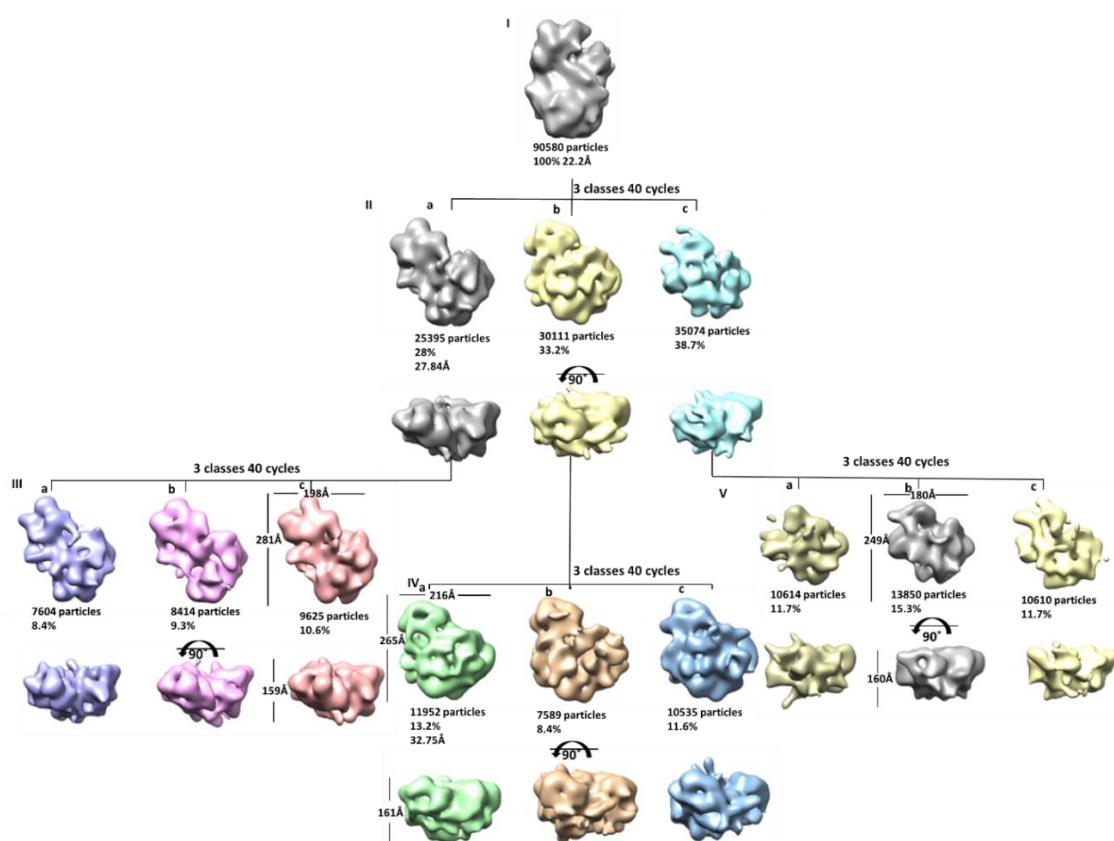


Figure 4.13 3D classification

Initial classification of 90580 particles into three 3D classes (II), after performing rounds of 2D classification step in RELION. Selected particles from each of the three classes was used for another 40 iterations of 3D classification (III, IV and V). The particles contributing to the class exhibiting the highest level of structural detail was used for the 3D refinement (IV).

To avoid introducing reference bias into 3D classification, the initial low-pass filter value of 60 Å was used to starting the auto-refinement process before getting a reference for 3D classification. Since it is recommended to use a consensus model from the data themselves, I first combined all selected particles in a single 3D auto-refine procedure. This model was used as the starting model in the subsequent three-dimensional classification with 3 classes, an angular sampling of 1.8Å and 60 Å initial low-pass filter again. Particles from each of 3 classes occupied 25.6%, 29.5% and 35% of total 90850 particles. By checking the particle quality from each of the classes, particles from class IIa looks most heterogeneous with highly different shape and size. Even though particles from both class IIb and class IIc have good qualities, they share different shape and size from each other. And the second round of 3D classification from each class shows that the 3 sub-classes from class IIb have similar structure and lower noise while the sub-classes from class IIc looks quite noisy. I chose one of sub-class IVa with most particles for the following 3D refinement to get better resolution.

j 3D model refinement

To further clean bad particles, the 3D class with 11952 particles were subsequent to two rounds of 2D classification and finally 11621 particles were selected and used for 3D refinement.

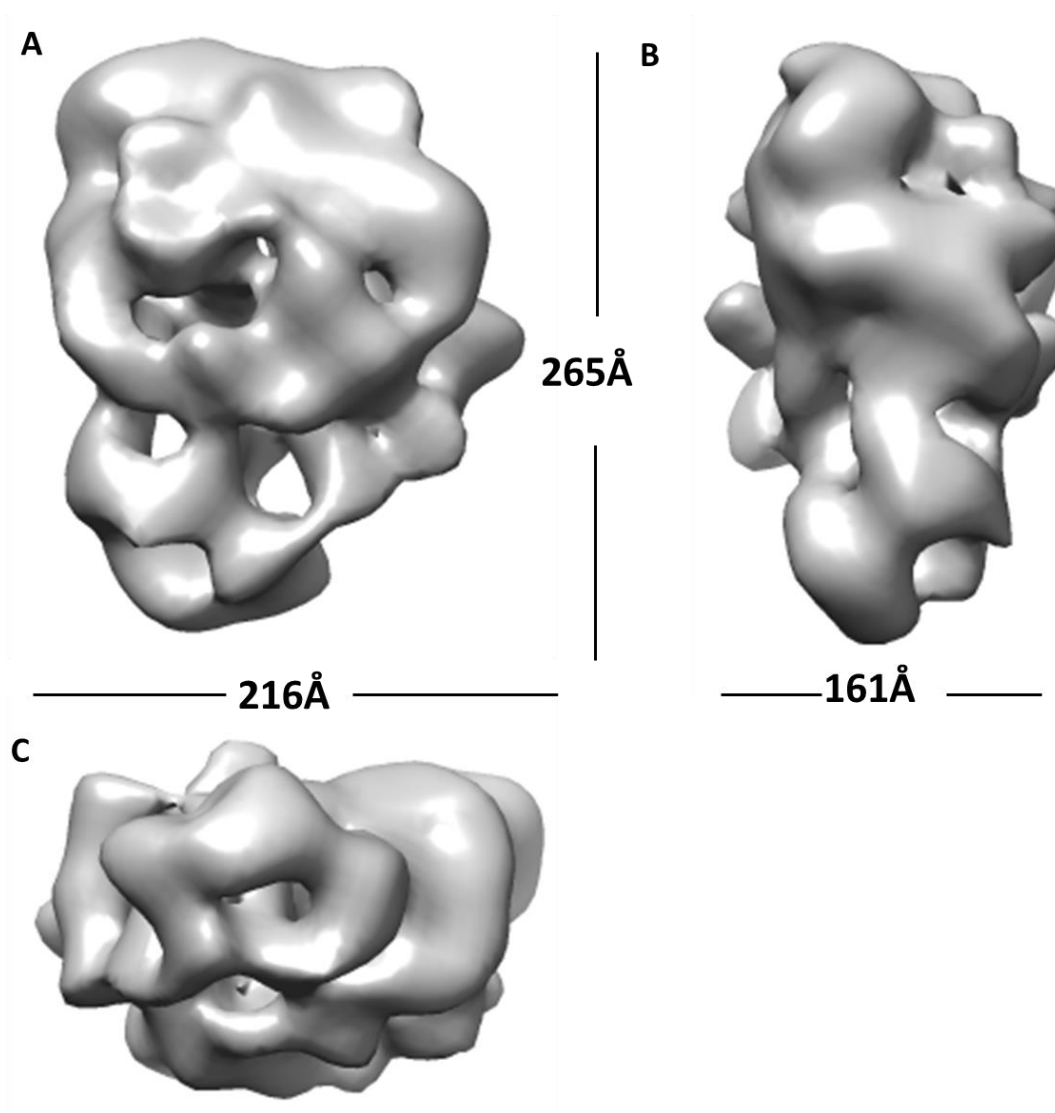


Figure 4.14 Refined 3D model of OutE-OutM complex

Three orthogonal views of the complex are shown in A, B and C.

The structure of OutE-OutM is ~161 Å in height, ~265 Å length, and ~216 Å width.

Assuming all proteins have approximately the same density about 1.37g/cm³(Erickson 2009). The protein mass of OutE-OutM can be calculated by using the average partial specific volume 0.73 cm³/g which is the reciprocal of the density.

$$V(\text{nm}^3) = \left(\frac{0.73 \text{cm}^3/\text{g} \times (10^{21} \text{nm}^3/\text{cm}^3)}{6.023 \times 10^{23}/\text{g}} \right) \times M(\text{Da})$$
$$= 1.212 \times 10^{-3} (\text{nm}^3/\text{Da}) \times M(\text{Da})$$

$$MW(\text{Da}) = 825 \text{Volume} (\text{nm}^3) = 825 \times (2.7248 \times 10^6) / 1000 = 906 \text{kDa}.$$

The volume was measured using Measure and Color blobs in Chimera with threshold 0.0243.

This calculated molecular weight from volume of the model is close to that from DLS result which is 1159kDa.

The “gold standard” refinement without symmetry applied yielded a final reconstruction (11621 particles) at 32.7 Å resolution, according to the FSC = 0.5 threshold. The back-projections of the EM models were obtained using the “Relion_project” tool in RELION.

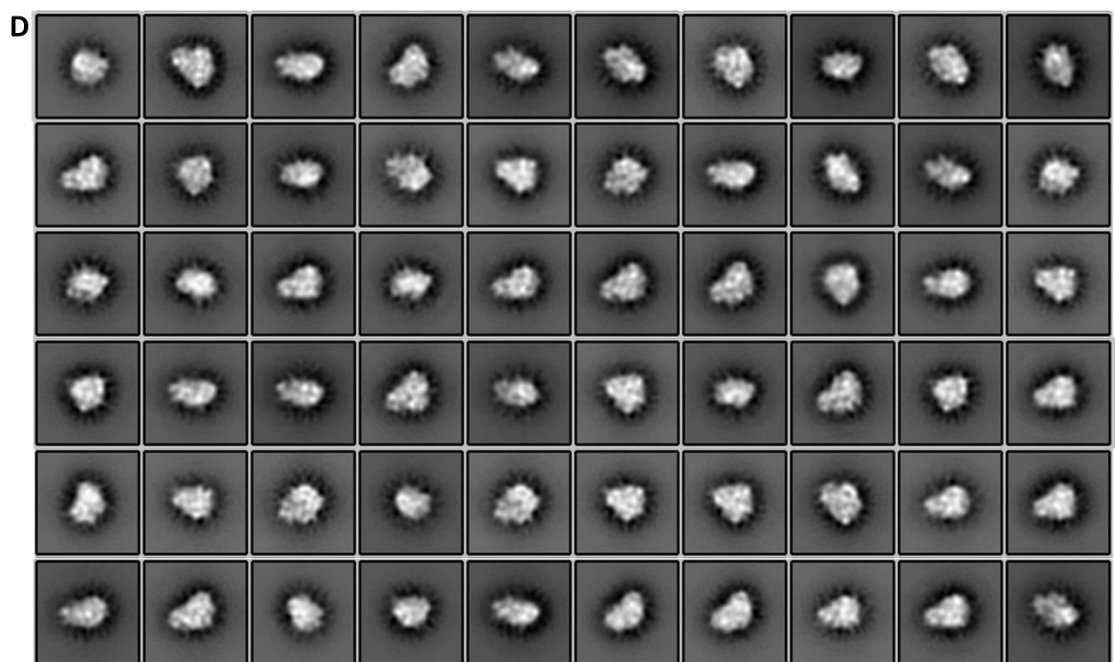
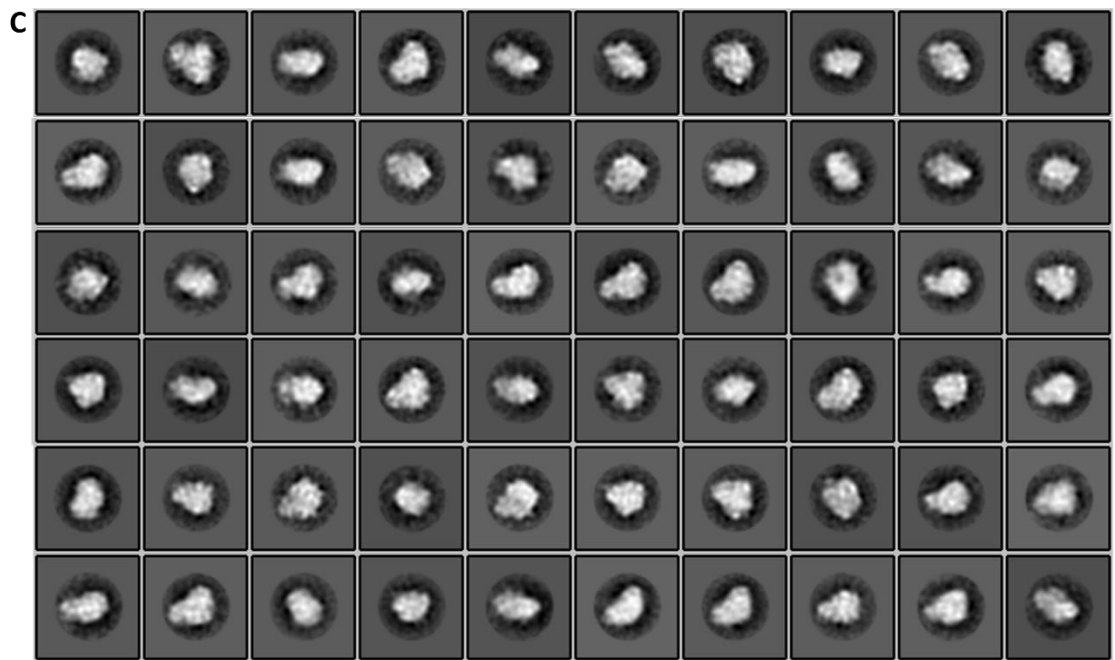
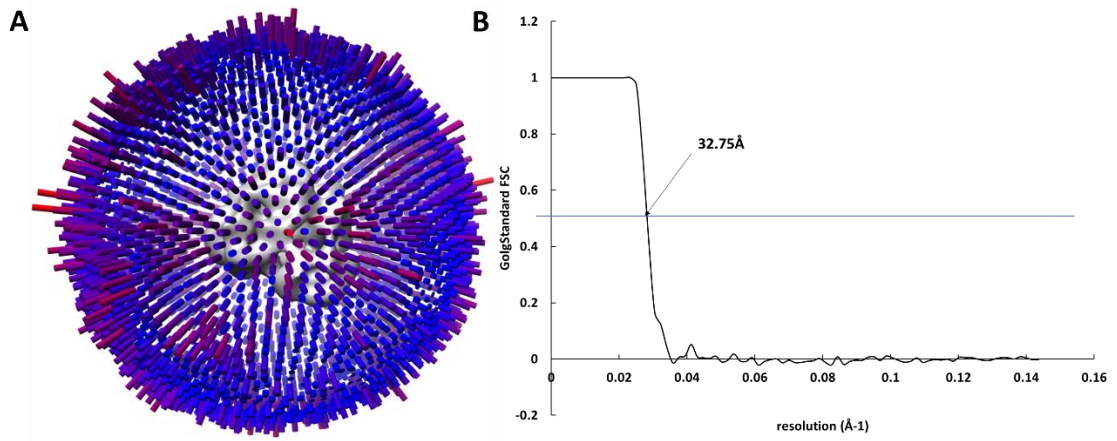


Figure 4.15 EM analysis of OutE-OutM

A: Euler plots for the angular distribution of OutE-OutM model; B: Plot of Fourier Shell Correlation curve for the OutE-OutM reconstruction. The blue line shows FSC = 0.5.; C and D: Comparison between the reference-free class averages (C), the back-projections of the non-symmetrized OutE-OutM model (D). Box size: 160 pixels; mask size: 350 Å.

Figure 4.15 shows even though the angular assignment of particle orientations is well distributed, few preferential orientations are observed. There is a clear correlation between the class averages and the model back - projections. The model shows an estimated resolution of 32.7 Å based on the Fourier Shell Correlation.

4.2.4 Domain localisation in the T2SS inner-membrane platform

To localise the individual components within the inner-membrane platform complex, the deleted mutations deletions of the cytoplasmic ATPase (E) and that of pilin subunits (G, H, I, J, K) have been cloned and expressed. In addition, nanogold labelling of OutM is also used to confirm the organisation of proteins within the complex. Initial fitting of crystal structures of ATPase(E) into the electron density envelopes was done using Chimera.

a Gold labelling

Two gold labelling methods were applied for T2SS_{E-M/M-His} as described in chapter 2. One is incubated with NI-NTA nanogold beads before loading the sample on the grid and the other is incubated with nanogold beads after loading the sample on the grid. These two methods were compared in the following figure.

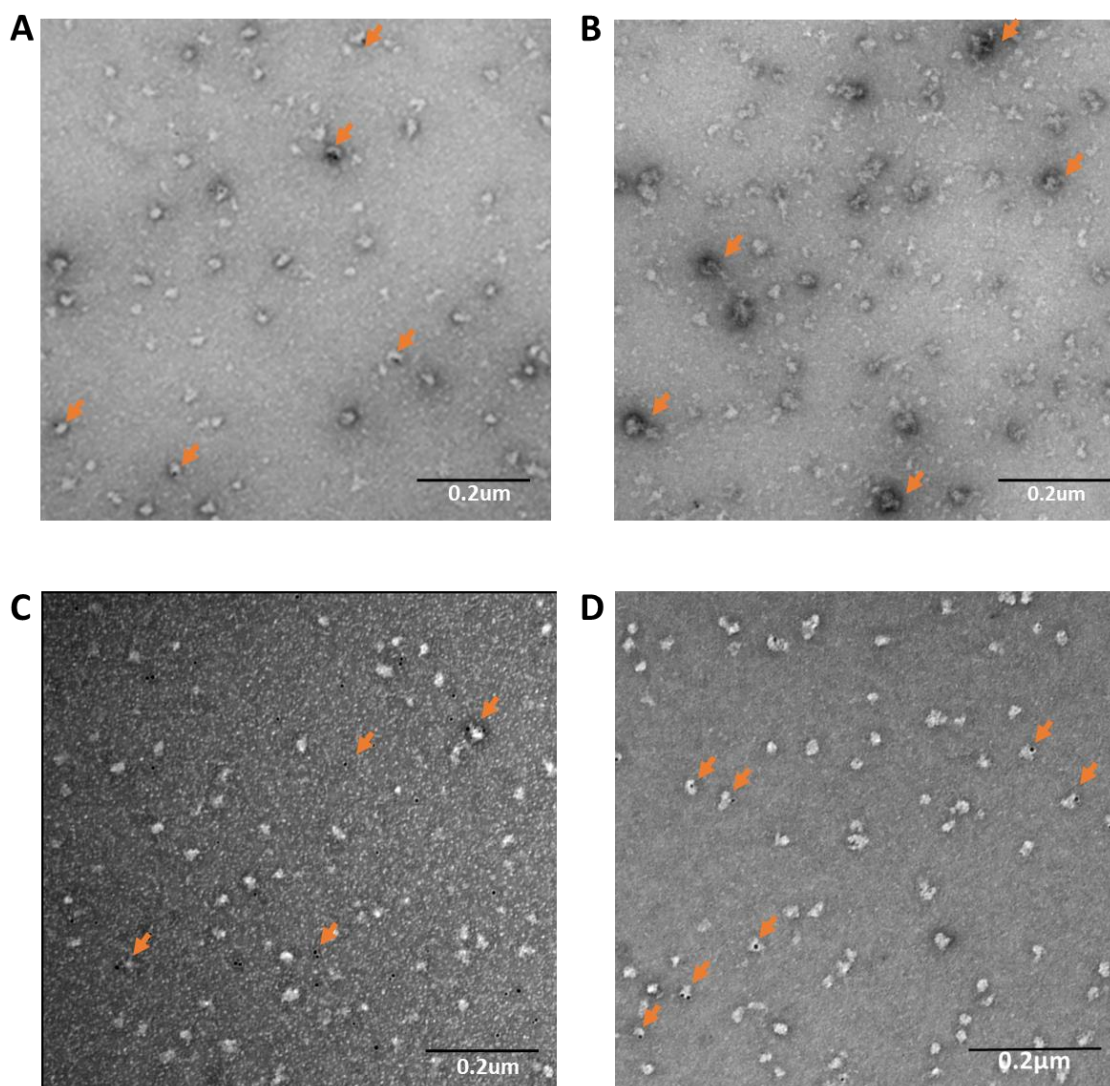


Figure 4.16 T2SS_{E-M/M-His} gold labelling

NI-NTA nanogold beads incubated with T2SS_{E-M/M-His} after sample was loaded on the grid for (A) 2 minutes, (B) 5 minutes and (C) 30 minutes respectively. D: NI-NTA nanogold beads were incubated with T2SS_{E-M/M-His} in solution at 4°C before loaded on grid. Ni-NTA Nanogold labelling of C-terminal His tagged OutM. Particles of Nanogold bound to the OutE-M complex are indicated in the micrograph using orange arrows.

More T2SS_{E-M/M-His} were labelled using the second method than the first without any severe non-specific labelling. However, the particle details are not clearly observed and it is

hard to compare with the unlabelled images. More and better-quality images need to be collected with gold labelled particles.

b Pilus-deleted and OutE-deleted mutations

To improve the homogeneity of the particles and locate the ATPase in the complex, two mutations were made: pilus-deleted complex and OutE-deleted complex (Fig. 4.17).

c ATPase localization in the inner-membrane platform

To fit crystal structure of hexamer GspE into the map of OutE-OutM, the generated EM map of OutE-OutM was segmented into 20 regions based on the immersive watershed algorithm and scale-spaced grouping(Pintilie et al. 2010).

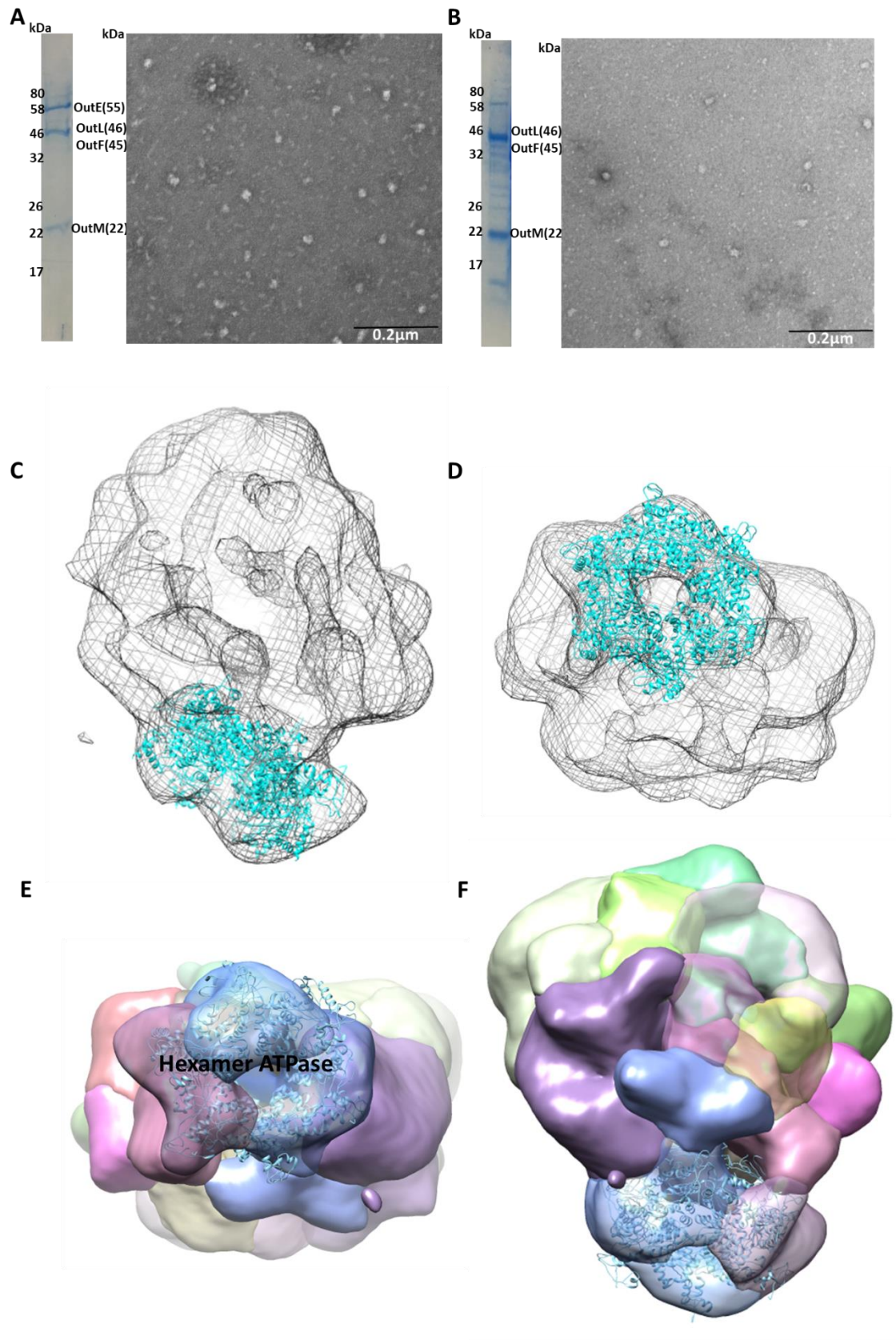


Figure 4.17 Domain localisation in the type II secretion system inner-membrane platform.

A: Pseudopilus subunits G-K deletion (i.e. E, F, L, M complex). B: F, L, M complex with ATPase OutE deleted. The left of each panel is the SDS-PAGE gel and to the right the negatively stained micrograph. The deletion complexes in A and B are less stable than the intact E-M complex showing that the presence of the ATPase and H pilin subunit stabilises the inner-membrane platform. C and D Preliminary fitting of the GspE homo-hexamer (PDB code: 4PHT) into the preliminary inner-membrane platform density of the T2SS using Chimera. E and F: The ATPase hexamer was fitted into three regions in the segmented OutE-OutM map (correlation: 0.85 density occupation: 0.71).

Figure 4.17 A and B shows the TEM images of pilus-deleted complex and OutE-deleted complex from 6 litres of cell cultures after Strep tag column and Superose 6 column without any dilution respectively. These two mutations are less stable than the intact E-M complex showing that the presence of the ATPase and H pilin subunit stabilises the inner-membrane platform.

To locate the ATPase, the crystal structure of hexamer GspE (PDB: 4PHT) from *Vibrio vulnificus* was fitted in the map of OutE-OutM using Chimera software and got 0.8547 correlation and 0.03013 average map value. In addition, the single ring in this structure can only be segmented into one or three regions indicating the ATPase hexamer may be formed by 3 dimers rather than 6 monomers.

4.3 Conclusion

In this Chapter, I reported cloning and expression of the inner-membrane platform OutE to OutM which includes three trans-membrane proteins F, L and M, the ATPase OutE and five pseudopilus subunits and further analysed the product by SDS-PAGE and mass spectrometry. However, only 5 single and clear bands were observed on the SDS-PAGE and the result from mass spec confirmed these bands are OutE, OutL and OutF, OutM and OutH respectively. The heterogeneity/instability of the inner-membrane platform

could be due to improper localisation of pseudopilus in the complex because of the absence of the prepilin peptidase O and the other core inner-membrane proteins OutC and OutB. The membrane domains of the pseudopilus cannot be cleaved without prepilin peptidase O, which may lead to faulty assembly of the pilins. OutB plays an important role in connection between the inner- and outer-membrane of T2SS, which may result in an unstable inner-membrane platform. Future plan will include adding these two genes into the inner-membrane platform I am producing.

Finally, I built an initial model of the inner-membrane complex. Even though the resolution is limited to 32Å because of the heterogeneous nature of the particles a clear ring located in the bottom of this structure can be fitted with the crystal structure of ATPase hexamer with a correlation of 0.855. By using the segmented function in chimera, this area can only be segmented into 1 or 3 regions indicating the hexamer ATPase may be formed by three dimers rather than six monomers.

To decrease the heterogeneous of the particles, I plan to stall larger complexes using non-hydrolysable ATP analogues, Walker motif mutations, and directed cross-linking according to *in vivo* engineering studies.

Chapter 5 Concluding comments

5.1 A status report on the type II secretion system

In the past decade, the structure of many isolated soluble protein domains of the T2SS have been solved. Several high-resolution structures of the T2SS secretin channel have been solved using Cryo-TEM (Hay et al. 2017, Yan, Yin et al. 2017); the cytoplasmic regions of GspF and L and the periplasmic regions of GspC, L and M were solved by X-ray crystallography (Abendroth, Rice et al. 2004, Abendroth, Murphy et al. 2005, Abendroth, Kreger et al. 2009, Abendroth, Mitchell et al. 2009); a GspE hexamer and several GspL-GspE complexes in 1:1 ratio have also been solved by X-ray crystallography (Abendroth, Murphy et al. 2005, Lu, Turley et al. 2013, Lu et al. 2014). A structural model of the *V. cholera* T2SS was built based on the solved structures and models by phyre2 (Gu, Shevchik et al. 2017). However, there is still limited knowledge about how the components assemble to form a fully functional secretion system.

In this thesis, I focused on the inner-membrane complex, the heart of the T2SS which recognizes secretion substrates and assembles the short pilus that pushes the substrate through the outer-membrane secretin. I successfully produced and solved the crystal structure of a cytoplasmic domain of OutF from the inner-membrane platform assembly which comprises three major integral membrane proteins: OutF, OutL, OutM (augmented by OutB/C see below and Chapter 1) and OutE. The pilus subunits and auxiliary components are also involved.

5.2 Summary of achievements in this thesis

In Chapter 3, I describe how I produced and solved the structure of cytoplasmic domain I of *D. dadantii* OutF, OutF⁶⁵⁻¹⁷², using X-ray crystallography. This domain is a bundle of six anti-parallel α -helices and two of these domains form a homodimer in the asymmetric unit of the unit cell. This structure shares high similarity with the first cytoplasmic domain of EpsF from *Vibrio cholera* (Abendroth, Mitchell et al. 2009), but OutF⁶⁵⁻¹⁷² doesn't possess the calcium-binding sites which help to stabilize the homodimer of EpsF. I compared the structure and sequence of OutF⁶⁵⁻¹⁷² with nine species of GspF family proteins from T2SS and T4PS in Gram-negative bacteria, T4PS, Com and Tad in Gram-positive bacteria and archaeal flagella. The eight totally conserved residues face inside of the cylinder formed by 6 α -helices and help to stabilize the helical bundle structure. The sequence and structure comparisons further confirmed the close relationship between components of the T2SS and T4PS. By sequence comparison and structure superimposition, two different dimer models of cytoplasmic domain II were built, the one formed in the same way as PilC is more stable than that of cytoplasmic domain I of OutF indicating the two cytoplasmic domains of OutF may associate to form dimers in different ways. In Chapter 3, the interactions between OutF and the inner-membrane proteins OutL and OutE were also studied using several methods including NMR, pull-down and thermofluor assays. NMR results suggest that OutF⁶⁵⁻¹⁷² interacts with OutL and OutE, but the interaction is not sufficiently strong to be confirmed by pull-down or thermofluor assay. Pull-down assay demonstrated a strong interaction between OutL and OutE and a 1:1 ratio was also observed on the SDS-PAGE. This interaction has been previously characterized in the *Vibrio cholerae* system (Py, Loiseau et al. 2001).

In Chapter 4, I described cloning and expression of the inner-membrane platform from OutE to OutM in *D.dadantii* and T2SS from prepilin peptidase to GspM in *E.coli* (Fig. 5.1). The inner-membrane complex from OutE to OutM was further purified by size exclusion column Superose 6 and eluted from a single peak eluting after the void volume. The molecular size of this complex was initially checked by dynamic light scattering and shows average size of ~1000 kDa with high heterogeneity. The purified complexes, comprising mainly OutF, OutL, OutM and OutE, were subsequently used for negative stained transmission electron microscopy.

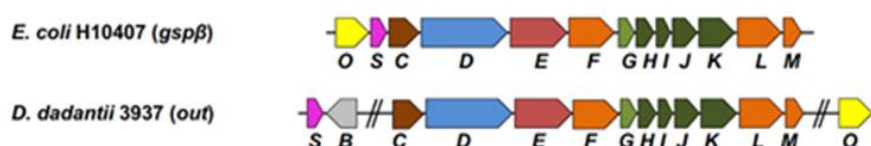


Figure 5.1 Genetic organization of T2SSs from *E. coli* and *D. dadantii*

Genes are labelled as single letters based on the Gsp nomenclature. Genes encoding GspC proteins are coloured brown, the secretins are coloured blue, the inner-membrane platform proteins are orange, the pseudo-pilins are coloured green, pre-pilin peptidases are yellow and accessory components are grey. Pilotins are coloured magenta. Operons are separated by double lines (Gu, Shevchik et al. 2017).

Also in Chapter 4, I describe the collection of 834 micrographs which were collected automatically using the FEI EPU (Tan, Cheng et al. 2016) (*E Pluribus Unum*) software at a magnification of 29,000x corresponding to a pixel size of 3.48 Å/pixel on FEI Tecnai 20 transmission electron microscope operating at 200 kV. The contrast transfer function was estimated using CTFFIND4 (Rohou and Grigorieff 2015) and 127283 particles were auto-picked using the 2D reference generated from 2213 manually selected particles. 220 of 300 2D classes were selected and used for initial 3D template building by SIMPLE

2.0(Elmlund and Elmlund 2012). This template was used for 3D classification in Relion(Fernandez-Leiro and Scheres 2017).

After two rounds of 3D classification, the best 3D class was chosen and used for 3D refinement. Even though the final resolution was only 32Å due to the heterogeneity of the particles, a ring can be observed in the bottom of model. To further validate the location of ATPase and the assembly of the complex, ATPase deleted and pilus deleted mutations were cloned and expressed. However, there were quite limited protein expressed resulting in poor-quality purifications and low quality TEM images. The fitting is therefore speculative at this point and will be confirmed by future work. Global searching and segment fitting were used to fit each of the known crystal structures into the density map of OutE-OutM complex. A high correlation 0.8547 was achieved after fitting the simulated model of crystal structure of ATPase hexamer into the map (model resolution was 20Å).

5.3 Future work

There is much work still to do to reveal the architecture of the inner-membrane complex of the type II secretion system of *D. dadantii*. Only four bands are observed on the gel after purification of the inner-membrane sub-complex from the construct OutE - OutM from *D. dadantii* and even for the entire gene operon GspO - GspM from *E. coli*. Even though these bands have been confirmed to be OutE, OutL and OutF, OutM and OutH respectively by mass spectrometry, the reason why pseudopilus subunits were excluded from the complex is still unknown, but is presumably because of the absence of the *D. dadantii* prepilin peptidase O.

Future work includes investigating the following:

- a) Supplementation of the sub-complex with OutB/C may facilitate production of more complete and stable inner-membrane complexes.
- b) To localise the individual components within the inner-membrane platform complex, the deletion of ATPase and pilin subunits have been cloned and expressed. However, only few proteins were expressed even from 6 litres of culture which resulted in bad quality of TEM images. To improve the expression and purification of these two constructs, different expression vectors, induction reagents, expression conditions and detergents need to be screened in the future work. More deletion constructs and nanogold labelling of other members in this complex will be made to help to confirm the organisation of proteins within the complex validate and aid the fitting of crystal structures of protein domains and homology models into the electron density envelopes using Chimera.
- c) Even though ~130000 particles were used to building the 3D model, only ~10000 particles were included in the final model with the highest resolution at 32.7Å due to the serious heterogeneity of the particles. To decrease the heterogeneity of the particles, I therefore plan to add a non-hydrolysable ATP analogue to inhibit the ATPase, as well as optimize the purification protocols including use of additional tags, detergent screening and gradient centrifugation techniques to increase compositional homogeneity before initiating cryo-TEM trials.
- d) The lack of the prepilin peptidase might prevent the full assembly of the system. ATP hydrolysis mutants of OutE/GspE might lead to assembly of a more stable complex mimicking the preassembled inactive state.
- e) Introducing purified secretion substrates might provide interesting insights into the secretion intermediates especially when combined with the inactive ATPase.

- f) Establish cryo-EM conditions for the complexes including the use of thin-film carbon-coated grids. Depending on the resolution of the cryo-EM analysis, which might not reach residue level, it may be important to apply more advanced strategies for structural modelling of the complexes.

Clearly there is much work to do, but the goal of understanding how the type II secretion system works is closer as a result of the work described in this Thesis.

References

(2000). Nature reviews. Molecular cell biology. London, UK, Nature Pub. Group: volumes.

Abendroth, J., A. C. Kreger and W. G. Hol (2009). "The dimer formed by the periplasmic domain of EpsL from the Type 2 Secretion System of *Vibrio parahaemolyticus*." *J Struct Biol* **168**(2): 313-322.

Abendroth, J., D. D. Mitchell, K. V. Korotkov, T. L. Johnson, A. Kreger, M. Sandkvist and W. G. Hol (2009). "The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of *Vibrio cholerae*." *J Struct Biol* **166**(3): 303-315.

Abendroth, J., P. Murphy, M. Sandkvist, M. Bagdasarian and W. G. Hol (2005). "The X-ray structure of the type II secretion system complex formed by the N-terminal domain of EpsE and the cytoplasmic domain of EpsL of *Vibrio cholerae*." *J Mol Biol* **348**(4): 845-855.

Abendroth, J., A. E. Rice, K. McLuskey, M. Bagdasarian and W. G. Hol (2004). "The crystal structure of the periplasmic domain of the type II secretion system protein EpsM from *Vibrio cholerae*: the simplest version of the ferredoxin fold." *J Mol Biol* **338**(3): 585-596.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 213-221.

Arts, J., A. de Groot, G. Ball, E. Durand, M. El Khattabi, A. Filloux, J. Tommassen and M. Koster (2007). "Interaction domains in the *Pseudomonas aeruginosa* type II secretory apparatus component XcpS (GspF)." *Microbiology* **153**(Pt 5): 1582-1592.

Ast, V. M., I. C. Schoenhofen, G. R. Langen, C. W. Stratilo, M. D. Chamberlain and S. P. Howard (2002). "Expression of the ExeAB complex of *Aeromonas hydrophila* is required for the localization and assembly of the ExeD secretion port multimer." *Molecular Microbiology* **44**(1): 217-231.

Ayers, M., L. M. Sampaleanu, S. Tammam, J. Koo, H. Harvey, P. L. Howell and L. L. Burrows (2009). "PilM/N/O/P proteins form an inner membrane complex that affects the stability of the *Pseudomonas aeruginosa* type IV pilus secretin." *J Mol Biol* **394**(1): 128-142.

Bally, M., A. Filloux, M. Akrim, G. Ball, A. Lazdunski and J. Tommassen (1992). "Protein secretion in *Pseudomonas aeruginosa*: characterization of seven xcp genes and processing of secretory apparatus components by prepilin peptidase." *Mol Microbiol* **6**(9): 1121-1131.

Battye, T. G. G., L. Kontogiannis, O. Johnson, H. R. Powell and A. G. W. Leslie (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM." *Acta Crystallographica Section D-Biological Crystallography* **67**: 271-281.

Berry, J. L., M. M. Phelan, R. F. Collins, T. Adomavicius, T. Tonjum, S. A. Frye, L. Bird, R. Owens, R. C. Ford, L. Y. Lian and J. P. Derrick (2012). "Structure and assembly of a trans-periplasmic channel for type IV pili in *Neisseria meningitidis*." *PLoS Pathog* **8**(9): e1002923.

Bleves, S., M. Gerard-Vincent, A. Lazdunski and A. Filloux (1999). "Structure-function analysis of XcpP, a component involved in general secretory pathway-dependent protein secretion in *Pseudomonas aeruginosa*." *J Bacteriol* **181**(13): 4012-4019.

- Blow, D. M. (2002). Outline of crystallography for biologists. Oxford ; New York, Oxford University Press.
- Bouley, J., G. Condemine and V. E. Shevchik (2001). "The PDZ domain of OutC and the N-terminal region of OutD determine the secretion specificity of the type II out pathway of *Erwinia chrysanthemi*." J Mol Biol **308**(2): 205-219.
- Camberg, J. L., T. L. Johnson, M. Patrick, J. Abendroth, W. G. Hol and M. Sandkvist (2007). "Synergistic stimulation of EpsE ATP hydrolysis by EpsL and acidic phospholipids." EMBO J **26**(1): 19-27.
- Campos, M., D. A. Cisneros, M. Nivaskumar and O. Francetic (2013). "The type II secretion system - a dynamic fiber assembly nanomachine." Res Microbiol **164**(6): 545-555.
- Campos, M., M. Nilges, D. A. Cisneros and O. Francetic (2010). "Detailed structural and assembly model of the type II secretion pilus from sparse data." Proc Natl Acad Sci U S A **107**(29): 13081-13086.
- Cao, T. B. and M. H. Saier, Jr. (2001). "Conjugal type IV macromolecular transfer systems of Gram-negative bacteria: organismal distribution, structural constraints and evolutionary conclusions." Microbiology **147**(Pt 12): 3201-3214.
- Chang, Y. W., L. A. Rettberg, A. Treuner-Lange, J. Iwasa, L. Sogaard-Andersen and G. J. Jensen (2016). "Architecture of the type IVa pilus machine." Science **351**(6278): aad2001.
- Chen, I. and D. Dubnau (2004). "DNA uptake during bacterial transformation." Nat Rev Microbiol **2**(3): 241-249.
- Chen, Y., S. J. Shiue, C. W. Huang, J. L. Chang, Y. L. Chien, N. T. Hu and N. L. Chan (2005). "Structure and function of the XpsE N-terminal domain, an essential component of the *Xanthomonas campestris* type II secretion system." Journal of Biological Chemistry **280**(51): 42356-42363.
- Chung, Y. S. and D. Dubnau (1995). "ComC is required for the processing and translocation of comGC, a pilin-like competence protein of *Bacillus subtilis*." Mol Microbiol **15**(3): 543-551.
- Cisneros, D. A., P. J. Bond, A. P. Pugsley, M. Campos and O. Francetic (2012). "Minor pseudopilin self-assembly primes type II secretion pseudopilus elongation." EMBO J **31**(4): 1041-1053.
- Collin, S., I. Guilvout, N. N. Nickerson and A. P. Pugsley (2011). "Sorting of an integral outer membrane protein via the lipoprotein-specific Lol pathway and a dedicated lipoprotein pilotin." Mol Microbiol **80**(3): 655-665.
- Collins, R. F., M. Saleem and J. P. Derrick (2007). "Purification and three-dimensional electron microscopy structure of the *Neisseria meningitidis* type IV pilus biogenesis protein PilG." J Bacteriol **189**(17): 6389-6396.
- Condemine, G. and V. E. Shevchik (2000). "Overproduction of the secretin OutD suppresses the secretion defect of an *Erwinia chrysanthemi* outB mutant." Microbiology-Uk **146**: 639-647.
- Crowther, L. J., R. P. Anantha and M. S. Donnenberg (2004). "The inner membrane subassembly of the enteropathogenic *Escherichia coli* bundle-forming pilus machine." Mol Microbiol **52**(1): 67-79.
- D'Enfert, C. and A. P. Pugsley (1989). "Klebsiella pneumoniae pulS gene encodes an outer membrane lipoprotein required for pullulanase secretion." J Bacteriol **171**(7): 3673-3679.
- d'Enfert, C., A. Ryter and A. P. Pugsley (1987). "Cloning and expression in *Escherichia coli* of the *Klebsiella pneumoniae* genes for production, surface localization and secretion of the lipoprotein pullulanase." EMBO J **6**(11): 3531-3538.

- Daefler, S., I. Guilvout, K. R. Hardie, A. P. Pugsley and M. Russel (1997). "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**(3): 465-475.
- Demers, J. P., B. Habenstein, A. Loquet, S. Kumar Vasa, K. Giller, S. Becker, D. Baker, A. Lange and N. G. Sgourakis (2014). "High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy." *Nat Commun* **5**: 4976.
- Douzi, B., G. Ball, C. Cambillau, M. Tegoni and R. Voulhoux (2011). "Deciphering the Xcp Pseudomonas aeruginosa type II secretion machinery through multiple interactions with substrates." *J Biol Chem* **286**(47): 40792-40801.
- Douzi, B., E. Durand, C. Bernard, S. Alphonse, C. Cambillau, A. Filloux, M. Tegoni and R. Voulhoux (2009). "The XcpV/GspI pseudopilin has a central role in the assembly of a quaternary complex within the T2SS pseudopilus." *J Biol Chem* **284**(50): 34580-34589.
- Durand, E., A. Bernadac, G. Ball, A. Lazdunski, J. N. Sturgis and A. Filloux (2003). "Type II protein secretion in Pseudomonas aeruginosa: the pseudopilus is a multifibrillar and adhesive structure." *J Bacteriol* **185**(9): 2749-2758.
- Durand, E., G. Michel, R. Voulhoux, J. Kurner, A. Bernadac and A. Filloux (2005). "XcpX controls biogenesis of the Pseudomonas aeruginosa XcpT-containing pseudopilus." *J Biol Chem* **280**(36): 31378-31389.
- East, A., A. E. Mechaly, G. H. Huysmans, C. Bernarde, D. Tello-Manigne, N. Nadeau, A. P. Pugsley, A. Buschiazzi, P. M. Alzari, P. J. Bond and O. Francetic (2016). "Structural Basis of Pullulanase Membrane Binding and Secretion Revealed by X-Ray Crystallography, Molecular Dynamics and Biochemical Analysis." *Structure* **24**(1): 92-104.
- Elmlund, D. and H. Elmlund (2012). "SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles." *J Struct Biol* **180**(3): 420-427.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallographica Section D-Biological Crystallography* **60**: 2126-2132.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132.
- Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4): 486-501.
- Erickson, H. P. (2009). "Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy." *Biological Procedures Online* **11**(1): 32-51.
- Evans, P. R. and G. N. Murshudov (2013). "How good are my data and what is the resolution?" *Acta Crystallographica Section D-Biological Crystallography* **69**: 1204-1214.
- Fernandez-Leiro, R. and S. H. W. Scheres (2017). "A pipeline approach to single-particle processing in RELION." *Acta Crystallogr D Struct Biol* **73**(Pt 6): 496-502.
- Fernandez-Leiro, R. and S. H. W. Scheres (2017). "A pipeline approach to single-particle processing in RELION." *Acta Crystallographica Section D-Structural Biology* **73**: 496-502.
- Francetic, O., N. Buddelmeijer, S. Lewenza, C. A. Kumamoto and A. P. Pugsley (2007). "Signal recognition particle-dependent inner membrane targeting of the PulG Pseudopilin component of a type II secretion system." *J Bacteriol* **189**(5): 1783-1793.
- Frank, J. (2006). *Electron tomography : methods for three-dimensional visualization of structures in the cell*. New York ; London, Springer.

- Frank, J. (2006). Three-dimensional electron microscopy of macromolecular assemblies : visualization of biological molecules in their native state. Oxford ; New York, Oxford University Press.
- Fries, M., J. Ihrig, K. Brocklehurst, V. E. Shevchik and R. W. Pickersgill (2007). "Molecular basis of the activity of the phytopathogen pectin methylesterase." EMBO J **26**(17): 3879-3887.
- Georgiadou, M., M. Castagnini, G. Karimova, D. Ladant and V. Pelicic (2012). "Large-scale study of the interactions between proteins involved in type IV pilus biology in *Neisseria meningitidis*: characterization of a subcomplex involved in pilus assembly." Mol Microbiol **84**(5): 857-873.
- Gerard-Vincent, M., V. Robert, G. Ball, S. Bleves, G. P. Michel, A. Lazdunski and A. Filloux (2002). "Identification of XcpP domains that confer functionality and specificity to the *Pseudomonas aeruginosa* type II secretion apparatus." Mol Microbiol **44**(6): 1651-1665.
- Ghosh, A. and S. V. Albers (2011). "Assembly and function of the archaeal flagellum." Biochem Soc Trans **39**(1): 64-69.
- Goder, V. and M. Spiess (2001). "Topogenesis of membrane proteins: determinants and dynamics." FEBS Lett **504**(3): 87-93.
- Gray, M. D., M. Bagdasarian, W. G. Hol and M. Sandkvist (2011). "In vivo cross-linking of EpsG to EpsL suggests a role for EpsL as an ATPase-pseudopilin coupling protein in the Type II secretion system of *Vibrio cholerae*." Mol Microbiol **79**(3): 786-798.
- Gray, M. D., M. Bagdasarian, W. G. J. Hol and M. Sandkvist (2011). "In vivo cross-linking of EpsG to EpsL suggests a role for EpsL as an ATPase-pseudopilin coupling protein in the Type II secretion system of *Vibrio cholerae*." Molecular Microbiology **79**(3): 786-798.
- Green, E. R. and J. Meccas (2016). "Bacterial Secretion Systems: An Overview." Microbiology Spectrum **4**(1).
- Greenfield, N. J. (2006). "Using circular dichroism spectra to estimate protein secondary structure." Nat Protoc **1**(6): 2876-2890.
- Greenfield, N. J. (2006). "Using circular dichroism spectra to estimate protein secondary structure." Nature protocols **1**(6): 2876-2890.
- Grynberg, M., Z. W. Li, E. Szcurek and A. Godzik (2007). "Putative type IV secretion genes in *Bacillus anthracis*." Trends in Microbiology **15**(5): 191-195.
- Gu, S., S. Rehman, X. Wang, V. E. Shevchik and R. W. Pickersgill (2012). "Structural and functional insights into the pilotin-secretin complex of the type II secretion system." PLoS Pathog **8**(2): e1002531.
- Gu, S., V. E. Shevchik, R. Shaw, R. W. Pickersgill and J. A. Garnett (2017). "The role of intrinsic disorder and dynamics in the assembly and function of the type II secretion system." Biochim Biophys Acta **1865**(10): 1255-1266.
- Hardie, K. R., A. Seydel, I. Guilvout and A. P. Pugsley (1996). "The secretin-specific, chaperone-like protein of the general secretory pathway: separation of proteolytic protection and piloting functions." Mol Microbiol **22**(5): 967-976.
- Hay, I. D., M. J. Belousoff and T. Lithgow (2017). "Structural Basis of Type 2 Secretion System Engagement between the Inner and Outer Bacterial Membranes." Mbio **8**(5).
- Hobbs, M. and J. S. Mattick (1993). "Common components in the assembly of type 4 fimbriae, DNA transfer systems, filamentous phage and protein-secretion apparatus: a general system for the formation of surface-associated protein complexes." Mol Microbiol **10**(2): 233-243.
- Johnson, T. L., J. Abendroth, W. G. Hol and M. Sandkvist (2006). "Type II secretion: from structure to function." FEMS Microbiol Lett **255**(2): 175-186.

- Jyot, J., V. Balloy, G. Jouvion, A. Verma, L. Touqui, M. Huerre, M. Chignard and R. Ramphal (2011). "Type II Secretion System of *Pseudomonas aeruginosa*: In Vivo Evidence of a Significant Role in Death Due to Lung Infection." *Journal of Infectious Diseases* **203**(10): 1369-1377.
- Kabsch, W. (2010). "Xds." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 125-132.
- Karuppiah, V. and J. P. Derrick (2011). "Structure of the PilM-PilN inner membrane type IV pilus biogenesis complex from *Thermus thermophilus*." *J Biol Chem* **286**(27): 24434-24442.
- Karuppiah, V., D. Hassan, M. Saleem and J. P. Derrick (2010). "Structure and oligomerization of the PilC type IV pilus biogenesis protein from *Thermus thermophilus*." *Proteins* **78**(9): 2049-2057.
- Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." *Nat Protoc* **10**(6): 845-858.
- Kohler, R., K. Schafer, S. Muller, G. Vignon, K. Diederichs, A. Philippsen, P. Ringler, A. P. Pugsley, A. Engel and W. Welte (2004). "Structure and assembly of the pseudopilin PulG." *Mol Microbiol* **54**(3): 647-664.
- Kolappan, S. and L. Craig (2013). "Structure of the cytoplasmic domain of TcpE, the inner membrane core protein required for assembly of the *Vibrio cholerae* toxin-coregulated pilus." *Acta Crystallogr D Biol Crystallogr* **69**(Pt 4): 513-519.
- Korotkov, K. V., T. Gonen and W. G. Hol (2011). "Secretins: dynamic channels for protein transport across membranes." *Trends Biochem Sci* **36**(8): 433-443.
- Korotkov, K. V., M. D. Gray, A. Kreger, S. Turley, M. Sandkvist and W. G. Hol (2009). "Calcium is essential for the major pseudopilin in the type 2 secretion system." *J Biol Chem* **284**(38): 25466-25470.
- Korotkov, K. V. and W. G. Hol (2008). "Structure of the GspK-GspI-GspJ complex from the enterotoxigenic *Escherichia coli* type 2 secretion system." *Nat Struct Mol Biol* **15**(5): 462-468.
- Korotkov, K. V., T. L. Johnson, M. G. Jobling, J. Pruneda, E. Pardon, A. Heroux, S. Turley, J. Steyaert, R. K. Holmes, M. Sandkvist and W. G. Hol (2011). "Structural and functional studies on the interaction of GspC and GspD in the type II secretion system." *PLoS Pathog* **7**(9): e1002228.
- Korotkov, K. V., B. Krumm, M. Bagdasarian and W. G. Hol (2006). "Structural and functional studies of EpsC, a crucial component of the type 2 secretion system from *Vibrio cholerae*." *J Mol Biol* **363**(2): 311-321.
- Korotkov, K. V., E. Pardon, J. Steyaert and W. G. Hol (2009). "Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody." *Structure* **17**(2): 255-265.
- Korotkov, K. V., M. Sandkvist and W. G. J. Hol (2012). "The type II secretion system: biogenesis, molecular architecture and mechanism." *Nature Reviews Microbiology* **10**(5): 336-351.
- Krissinel, E. (2007). "On the relationship between sequence and structure similarities in proteomics." *Bioinformatics* **23**(6): 717-723.
- Kuan, J. and M. H. Saier, Jr. (1993). "The mitochondrial carrier family of transport proteins: structural, functional, and evolutionary relationships." *Crit Rev Biochem Mol Biol* **28**(3): 209-233.
- Kuo, W. W., H. W. Kuo, C. C. Cheng, H. L. Lai and L. Y. Chen (2005). "Roles of the minor pseudopilins, XpsH, XpsI and XpsJ, in the formation of XpsG-containing pseudopilus in *Xanthomonas campestris* pv. *campestris*." *J Biomed Sci* **12**(4): 587-599.

- Lallemand, M., F. H. Login, N. Guschinskaya, C. Pineau, G. Effantin, X. Robert and V. E. Shevchik (2013). "Dynamic interplay between the periplasmic and transmembrane domains of GspL and GspM in the type II secretion system." *PLoS One* **8**(11): e79562.
- Lathem, W. W., T. E. Grys, S. E. Witowski, A. G. Torres, J. B. Kaper, P. I. Tarr and R. A. Welch (2002). "StcE, a metalloprotease secreted by *Escherichia coli* O157 : H7, specifically cleaves C1 esterase inhibitor." *Molecular Microbiology* **45**(2): 277-288.
- Lee, H. M., J. R. Chen, H. L. Lee, W. M. Leu, L. Y. Chen and N. T. Hu (2004). "Functional dissection of the XpsN (GspC) protein of the *Xanthomonas campestris* pv. *campestris* type II secretion machinery." *J Bacteriol* **186**(10): 2946-2955.
- Lee, M. S., L. Y. Chen, W. M. Leu, R. J. Shiau and N. T. Hu (2005). "Associations of the major pseudopilin XpsG with XpsN (GspC) and secretin XpsD of *Xanthomonas campestris* pv. *campestris* type II secretion apparatus revealed by cross-linking analysis." *J Biol Chem* **280**(6): 4585-4591.
- Li, G. and S. P. Howard (2010). "ExeA binds to peptidoglycan and forms a multimer for assembly of the type II secretion apparatus in *Aeromonas hydrophila*." *Molecular Microbiology* **76**(3): 772-781.
- Lindeberg, M. and A. Collmer (1992). "Analysis of eight out genes in a cluster required for pectic enzyme secretion by *Erwinia chrysanthemi*: sequence comparison with secretion genes from other gram-negative bacteria." *J Bacteriol* **174**(22): 7385-7397.
- Lu, C., K. V. Korotkov and W. G. J. Hol (2014). "Crystal structure of the full-length ATPase GspE from the *Vibrio vulnificus* type II secretion system in complex with the cytoplasmic domain of GspL." *J Struct Biol* **187**(3): 223-235.
- Lu, C., S. Turley, S. T. Marianni, Y. J. Park, K. K. Lee, M. Patrick, R. Shah, M. Sandkvist, M. F. Bush and W. G. J. Hol (2013). "Hexamers of the Type II Secretion ATPase GspE from *Vibrio cholerae* with Increased ATPase Activity." *Structure* **21**(9): 1707-1717.
- Lybarger, S. R., T. L. Johnson, M. D. Gray, A. E. Sikora and M. Sandkvist (2009). "Docking and assembly of the type II secretion complex of *Vibrio cholerae*." *J Bacteriol* **191**(9): 3149-3161.
- Martin, S. R. and M. J. Schilstra (2008). "Circular dichroism and its application to the study of biomolecules." *Methods Cell Biol* **84**: 263-293.
- McCoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni and R. J. Read (2005). "Likelihood-enhanced fast translation functions." *Acta Crystallogr D Biol Crystallogr* **61**(Pt 4): 458-464.
- McGuffin, L. J., K. Bryson and D. T. Jones (2000). "The PSIPRED protein structure prediction server." *Bioinformatics* **16**(4): 404-405.
- Merritt, E. A., T. K. Sixma, K. H. Kalk, B. A. van Zanten and W. G. Hol (1994). "Galactose-binding site in *Escherichia coli* heat-labile enterotoxin (LT) and cholera toxin (CT)." *Mol Microbiol* **13**(4): 745-753.
- Mora, T., H. Yu, Y. Sowa and N. S. Wingreen (2009). "Steps in the Bacterial Flagellar Motor." *Plos Computational Biology* **5**(10).
- Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." *Acta Crystallogr D Biol Crystallogr* **53**(Pt 3): 240-255.
- Natale, P., T. Bruser and A. J. M. Driessen (2008). "Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane - Distinct translocases and mechanisms." *Biochimica Et Biophysica Acta-Biomembranes* **1778**(9): 1735-1756.
- Nguyen, L., I. T. Paulsen, J. Tchieu, C. J. Hueck and M. H. Saier, Jr. (2000). "Phylogenetic analyses of the constituents of Type III protein secretion systems." *J Mol Microbiol Biotechnol* **2**(2): 125-144.

- Nickerson, N. N., T. Tosi, A. Dessen, B. Baron, B. Raynal, P. England and A. P. Pugsley (2011). "Outer membrane targeting of secretin PulD protein relies on disordered domain recognition by a dedicated chaperone." *J Biol Chem* **286**(45): 38833-38843.
- Nivaskumar, M., G. Bouvier, M. Campos, N. Nadeau, X. Yu, E. H. Egelman, M. Nilges and O. Francetic (2014). "Distinct docking and stabilization steps of the Pseudopilus conformational transition path suggest rotational assembly of type IV pilus-like fibers." *Structure* **22**(5): 685-696.
- Nivaskumar, M. and O. Francetic (2014). "Type II secretion system: a magic beanstalk or a protein escalator." *Biochim Biophys Acta* **1843**(8): 1568-1577.
- Nouwen, N., H. Stahlberg, A. P. Pugsley and A. Engel (2000). "Domain structure of secretin PulD revealed by limited proteolysis and electron microscopy." *EMBO J* **19**(10): 2229-2236.
- Palomaki, T., R. Pickersgill, R. Riekkki, M. Romantschuk and H. T. Saarilahti (2002). "A putative three-dimensional targeting motif of polygalacturonase (PehA), a protein secreted through the type II (GSP) pathway in *Erwinia carotovora*." *Molecular Microbiology* **43**(3): 585-596.
- Patrick, M., K. V. Korotkov, W. G. Hol and M. Sandkvist (2011). "Oligomerization of EpsE coordinates residues from multiple subunits to facilitate ATPase activity." *J Biol Chem* **286**(12): 10378-10386.
- Peabody, C. R., Y. J. Chung, M. R. Yen, D. Vidal-Ingigliardi, A. P. Pugsley and M. H. Saier, Jr. (2003). "Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella." *Microbiology* **149**(Pt 11): 3051-3072.
- Pineau, C., N. Guschinskaya, X. Robert, P. Gouet, L. Ballut and V. E. Shevchik (2014). "Substrate recognition by the bacterial type II secretion system: more than a simple interaction." *Mol Microbiol* **94**(1): 126-140.
- Pintilie, G. D., J. Zhang, T. D. Goddard, W. Chiu and D. C. Gossard (2010). "Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions." *J Struct Biol* **170**(3): 427-438.
- Pugsley, A. P., C. Chapon and M. Schwartz (1986). "Extracellular pullulanase of *Klebsiella pneumoniae* is a lipoprotein." *J Bacteriol* **166**(3): 1083-1088.
- Pugsley, A. P. and M. G. Kornacker (1991). "Secretion of the cell surface lipoprotein pullulanase in *Escherichia coli*. Cooperation or competition between the specific secretion pathway and the lipoprotein sorting pathway." *J Biol Chem* **266**(21): 13640-13645.
- Py, B., L. Loiseau and F. Barras (2001). "An inner membrane platform in the type II secretion machinery of Gram-negative bacteria." *EMBO Rep* **2**(3): 244-248.
- Py, B., L. Loiseau and F. Barras (2001). "An inner membrane platform in the type II secretion machinery of Gram-negative bacteria." *Embo Reports* **2**(3): 244-248.
- Rehman, S., S. Gu, V. E. Shevchik and R. W. Pickersgill (2013). "Anatomy of secretin binding to the *Dickeya dadantii* type II secretion system pilotin." *Acta Crystallogr D Biol Crystallogr* **69**(Pt 8): 1381-1386.
- Reichow, S. L., K. V. Korotkov, W. G. Hol and T. Gonen (2010). "Structure of the cholera toxin secretion channel in its closed state." *Nat Struct Mol Biol* **17**(10): 1226-1232.
- Richardson, D. J., J. N. Butt, J. K. Fredrickson, J. M. Zachara, L. Shi, M. J. Edwards, G. White, N. Baiden, A. J. Gates, S. J. Marritt and T. A. Clarke (2012). "The 'porin-cytochrome' model for microbe-to-mineral electron transfer." *Mol Microbiol* **85**(2): 201-212.
- Robert, V., F. Hayes, A. Lazdunski and G. P. Michel (2002). "Identification of XcpZ domains required for assembly of the secretin of *Pseudomonas aeruginosa*." *J Bacteriol* **184**(6): 1779-1782.

- Robien, M. A., B. E. Krumm, M. Sandkvst and W. G. J. Hol (2003). "Crystal structure of the extracellular protein secretion NTPase EpsE of *Vibrio cholerae*." Journal of Molecular Biology **333**(3): 657-674.
- Rodgers, K., C. G. Arvidson and S. Melville (2011). "Expression of a *Clostridium perfringens* type IV pilin by *Neisseria gonorrhoeae* mediates adherence to muscle cells." Infect Immun **79**(8): 3096-3105.
- Rohou, A. and N. Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs." J Struct Biol **192**(2): 216-221.
- Rossier, O., S. R. Starkenburg and N. P. Cianciotto (2004). "Legionella pneumophila type II protein secretion promotes virulence in the A/J mouse model of Legionnaires' disease pneumonia." Infect Immun **72**(1): 310-321.
- Saier, M. H., Jr. (2003). "Tracing pathways of transport protein evolution." Mol Microbiol **48**(5): 1145-1156.
- Sandkvist, M., L. P. Hough, M. M. Bagdasarian and M. Bagdasarian (1999). "Direct interaction of the EpsL and EpsM proteins of the general secretion apparatus in *Vibrio cholerae*." J Bacteriol **181**(10): 3129-3135.
- Sandkvist, M., J. M. Keith, M. Bagdasarian and S. P. Howard (2000). "Two regions of EpsL involved in species-specific protein-protein interactions with EpsE and EpsM of the general secretion pathway in *Vibrio cholerae*." J Bacteriol **182**(3): 742-748.
- Sandkvist, M., L. O. Michel, L. P. Hough, V. M. Morales, M. Bagdasarian, M. Koomey, V. J. DiRita and M. Bagdasarian (1997). "General secretion pathway (eps) genes required for toxin secretion and outer membrane biogenesis in *Vibrio cholerae*." J Bacteriol **179**(22): 6994-7003.
- Satyshur, K. A., G. A. Worzalla, L. S. Meyer, E. K. Heiniger, K. G. Aukema, A. M. Misic and K. T. Forest (2007). "Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility." Structure **15**(3): 363-376.
- Sauvonnet, N., G. Vignon, A. P. Pugsley and P. Gounon (2000). "Pilus formation and protein secretion by the same machinery in *Escherichia coli*." EMBO J **19**(10): 2221-2228.
- Seo, J., A. Brencic and A. J. Darwin (2009). "Analysis of Secretin-Induced Stress in *Pseudomonas aeruginosa* Suggests Prevention Rather than Response and Identifies a Novel Protein Involved in Secretin Function." Journal of Bacteriology **191**(3): 898-908.
- Shevchik, V. E. and G. Condemine (1998). "Functional characterization of the *Erwinia chrysanthemi* OutS protein, an element of a type II secretion system." Microbiology **144** (Pt 11): 3219-3228.
- Shevchik, V. E., J. Robert-Baudouy and G. Condemine (1997). "Specific interaction between OutD, an *Erwinia chrysanthemi* outer membrane protein of the general secretory pathway, and secreted proteins." EMBO J **16**(11): 3007-3016.
- Shiue, S. J., K. M. Kao, W. M. Leu, L. Y. Chen, N. L. Chan and N. T. Hu (2006). "XpsE oligomerization triggered by ATP binding, not hydrolysis, leads to its association with XpsL." EMBO J **25**(7): 1426-1435.
- Skerra, A. (1994). "Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*." Gene **151**(1-2): 131-135.
- Sreerama, N., S. Y. Venyaminov and R. W. Woody (2000). "Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis." Anal Biochem **287**(2): 243-251.
- Strozen, T. G., G. Li and S. P. Howard (2012). "YghG (GspSbeta) is a novel pilot protein required for localization of the GspSbeta type II secretion system secretin of enterotoxigenic *Escherichia coli*." Infect Immun **80**(8): 2608-2622.

- Strozen, T. G., H. Stanley, Y. Gu, J. Boyd, M. Bagdasarian, M. Sandkvist and S. P. Howard (2011). "Involvement of the GspAB complex in assembly of the type II secretion system secretin of *Aeromonas* and *Vibrio* species." *J Bacteriol* **193**(9): 2322-2331.
- Tammam, S., L. M. Sampaleanu, J. Koo, K. Manoharan, M. Daubaras, L. L. Burrows and P. L. Howell (2013). "PilMNOPQ from the *Pseudomonas aeruginosa* type IV pilus system form a transenvelope protein interaction network that interacts with PilA." *J Bacteriol* **195**(10): 2126-2135.
- Tan, Y. Z., A. Cheng, C. S. Potter and B. Carragher (2016). "Automated data collection in single particle electron microscopy." *Microscopy (Oxf)* **65**(1): 43-56.
- Tauschek, M., R. J. Gorrell, R. A. Strugnell and R. M. Robins-Browne (2002). "Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of *Escherichia coli*." *Proceedings of the National Academy of Sciences of the United States of America* **99**(10): 7066-7071.
- Thomas, J. D., P. J. Reeves and G. P. Salmond (1997). "The general secretion pathway of *Erwinia carotovora* subsp. *carotovora*: analysis of the membrane topology of OutC and OutF." *Microbiology* **143** (Pt 3): 713-720.
- Toth, I. K. and P. R. Birch (2005). "Rotting softly and stealthily." *Curr Opin Plant Biol* **8**(4): 424-429.
- Tsai, R. T., W. M. Leu, L. Y. Chen and N. T. Hu (2002). "A reversibly dissociable ternary complex formed by XpsL, XpsM and XpsN of the *Xanthomonas campestris* pv. *campestris* type II secretion apparatus." *Biochem J* **367**(Pt 3): 865-871.
- Vagin, A. and A. Teplyakov (2000). "An approach to multi-copy search in molecular replacement." *Acta Crystallographica Section D-Biological Crystallography* **56**: 1622-1624.
- Van der Meeren, R., Y. Wen, P. Van Gelder, J. Tommassen, B. Devreese and S. N. Savvides (2013). "New insights into the assembly of bacterial secretins: structural studies of the periplasmic domain of XcpQ from *Pseudomonas aeruginosa*." *J Biol Chem* **288**(2): 1214-1225.
- Vanderlinde, E. M., S. Zhong, G. Li, D. Martynowski, P. Grochulski and S. P. Howard (2014). "Assembly of the Type Two Secretion System in *Aeromonas hydrophila* Involves Direct Interaction between the Periplasmic Domains of the Assembly Factor ExeB and the Secretin ExeD." *Plos One* **9**(7).
- Varga, J. J., V. Nguyen, D. K. O'Brien, K. Rodgers, R. A. Walker and S. B. Melville (2006). "Type IV pili-dependent gliding motility in the Gram-positive pathogen *Clostridium perfringens* and other *Clostridia*." *Mol Microbiol* **62**(3): 680-694.
- Vaynberg, J. and J. Qin (2006). "Weak protein-protein interactions as probed by NMR spectroscopy." *Trends Biotechnol* **24**(1): 22-27.
- Viarre, V., E. Cascales, G. Ball, G. P. F. Michel, A. Filloux and R. Voulhoux (2009). "HxcQ Liposecretin Is Self-piloted to the Outer Membrane by Its N-terminal Lipid Anchor." *Journal of Biological Chemistry* **284**(49): 33815-33823.
- Vignon, G., R. Kohler, E. Larquet, S. Giroux, M. C. Prevost, P. Roux and A. P. Pugsley (2003). "Type IV-like pili formed by the type II secretin: specificity, composition, bundling, polar localization, and surface presentation of peptides." *J Bacteriol* **185**(11): 3416-3428.
- Voulhoux, R., G. Ball, B. Ize, M. L. Vasil, A. Lazdunski, L. F. Wu and A. Filloux (2001). "Involvement of the twin-arginine translocation system in protein secretion via the type II pathway." *Embo Journal* **20**(23): 6735-6741.
- Wang, X., C. Pineau, S. Gu, N. Guschinskaya, R. W. Pickersgill and V. E. Shevchik (2012). "Cysteine scanning mutagenesis and disulfide mapping analysis of arrangement

- of GspC and GspD protomers within the type 2 secretion system." *J Biol Chem* **287**(23): 19082-19093.
- Ward, J. J., L. J. McGuffin, K. Bryson, B. F. Buxton and D. T. Jones (2004). "The DISOPRED server for the prediction of protein disorder." *Bioinformatics* **20**(13): 2138-2139.
- Whitmore, L. and B. A. Wallace (2008). "Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases." *Biopolymers* **89**(5): 392-400.
- Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." *Acta Crystallographica Section D-Biological Crystallography* **67**: 235-242.
- Worrall, L. J., C. Hong, M. Vuckovic, W. Deng, J. R. Bergeron, D. D. Majewski, R. K. Huang, T. Spreter, B. B. Finlay, Z. Yu and N. C. Strynadka (2016). "Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body." *Nature*.
- Yan, Z. F., M. Yin, D. D. Xu, Y. Q. Zhu and X. M. Li (2017). "Structural insights into the secretin translocation channel in the type II secretion system." *Nature Structural & Molecular Biology* **24**(2): 177-+.
- Zhang, R. G., D. L. Scott, M. L. Westbrook, S. Nance, B. D. Spangler, G. G. Shipley and E. M. Westbrook (1995). "The three-dimensional crystal structure of cholera toxin." *J Mol Biol* **251**(4): 563-573.

Appendix Paper to be published in Journal of Frontiers in Microbiology

**Structure of the first cytoplasmic domain of OutF and
assembly of the inner-membrane platform proteins of the *D.*
dadantii type II secretion system**

Hui Zhang¹, Shuang Gu¹, Vladimir E. Shevchik², Richard W. Pickersgill^{1‡}

¹Queen Mary University of London, School of Biological and Chemical Sciences, Mile End Road, London E1 4NS, England. ²Université de Lyon, F-69003, Université Lyon 1, Lyon, F-69622, INSA-Lyon, Villeurbanne, F-69621, CNRS, UMR5240, Microbiologie Adaptation et Pathogénie, Lyon, F-69622, France.