

Structural studies of the Type IX Secretion System inner membrane complex

Rory Hennell James

Sir William Dunn School of Pathology,

Department of Biochemistry,

and

Worcester College

University of Oxford



A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at
the University of Oxford.

Trinity Term 2021

48,000 words

Structural studies of the Type IX Secretion System inner membrane complex

Rory Hennell James

Sir William Dunn School of Pathology, Department of Biochemistry, and Worcester College, University of Oxford.

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford.

Abstract

The Type IX Secretion System (T9SS) is a protein transport system found exclusively in the *Bacteroidetes* phylum of Gram-negative bacteria. The T9SS transports folded substrates across the outer membrane, including virulence factors of human pathogens, polysaccharide utilisation enzymes of environmental bacteria, and cell-surface adhesin molecules involved in *Bacteroidetes* gliding motility

The inner membrane proteins GldL and GldM use the proton-motive force to energise both the T9SS and the gliding motility apparatus. I present structures from five different species of the transmembrane core of the GldLM complex, solved using electron cryo-microscopy. The GldLM complex has two copies of GldM with one transmembrane helix each, surrounded by five copies of GldL with two transmembrane helices each. Most of the cytoplasmic domain of GldL was unresolved. I also present a low-resolution structure of the full length orthologous PorLM complex from *Porphyromonas gingivalis*.

Several polar residues are present in the transmembrane helices of GldL and GldM, suggesting that they may have a role in conducting protons across the inner membrane. I generated mutant strains where these polar residues were substituted with non-polar alternatives and found that some of these strains were deficient in T9SS and gliding motility. A universally conserved glutamate residue in

GldL was essential to T9SS and gliding motility activity. In the structures of GldLM, one copy of this glutamate residue appears to form a salt bridge with a conserved arginine residue in one copy of GldM. I propose that a proton flow-dependent alternation of which copy of GldM is involved in salt bridge formation drives rotation of the GldM dimer, which spans the periplasm to connect the inner membrane and outer membrane components of the T9SS.

These results reveal the GldLM complex to be a third class of ion-driven rotary motor after the ATP synthase and the MotAB flagellar stator family.

Acknowledgements

Firstly, I would like to thank my supervisors Susan Lea and Ben Berks for giving me the opportunity to work on this project and for all their help and advice over the last four years, particularly the last few months. I'd also like to thank all the members of the Lea and Berks labs, past and present, for making me feel welcome at the beginning, for all the cultures started and plates taken out of the incubator, and for all the fun times, scientific and otherwise. I'm especially grateful to Fred and Fliss from the Berks lab for helping me get started when I first arrived, and to Justin from the Lea lab for teaching me half of what I know about proteins and practically everything I know about electron microscopy.

I'm very grateful to the Dunn School and the Biochemistry Department for providing such a great environment to do science in. That particularly includes Adam and the rest of the COSMIC staff for EM training and everyone else who gave me training or helped me out over the years. Thanks are also due to the CSB DPhil Programme and the Wellcome Trust for funding me and my science so generously.

I'd like to thank my housemates for keeping me going over the last four years: James for being the best officemate; Gerard for all the cake, ice cream and homemade alcohol I could wish for; Ciaran for not being the worst *possible* housemate. I'm also grateful to Andy, David, James and Molly for all the fun adventures and providing probably half of my social life over the last 18 months. I'd like to thank all the members of Worcester College MCR for creating such a fun and friendly community, except for those who were too lazy to help on the Committee and stuck me with the job of Treasurer for three and a half years.

Lastly, I'd like to thank my family for supporting me and my love of learning over the last 26 years. I promise you'll see more of me now, at least for a few months.

Publications resulting from this study

Hennell James, R. *, Deme, J. C. *, Kjær, A., Alcock, F., Silale, A., Lauber, F., Johnson, S., Berks, B. C., & Lea, S. M. (2021). Structure and mechanism of the proton-driven motor that powers type 9 secretion and gliding motility. *Nature Microbiology*, 1–13. <https://doi.org/10.1038/s41564-020-00823-6>

* These authors contributed equally to this study. R.H.J. carried out all genetic and biochemical work except as credited otherwise. J.C.D. and S.M.L. collected electron microscopy data and determined the structures. A.K. developed and carried out the SprB tracking experiments including strain construction. F.A. carried out the pulse chase analysis of protein export. A.S. assayed cellular ATP levels. F.L. constructed the $\Delta g/dL$ strain and produced the GldL and GldM antibodies. B.C.B. and S.M.L. conceived the project. All authors interpreted data and wrote the manuscript. Some material from this paper is reproduced in modified form in this thesis. The work of other authors is acknowledged where appropriate.

Deme, J. C., Johnson, S., Vickery, O., Aron, A., Monkhouse, H., Griffiths, T., James, R. H., Berks, B. C., Coulton, J. W., Stansfeld, P. J., & Lea, S. M. (2020). Structures of the stator complex that drives rotation of the bacterial flagellum. *Nature Microbiology*, 5(12), 1553–1564. <https://doi.org/10.1038/s41564-020-0788-8>

J.C.D. carried out all biochemical work except as credited otherwise, prepared the cryo-EM grids, collected and processed the EM data and determined the structures. J.C.D., S.J. and S.M.L. designed the project, interpreted the data, built the models and wrote the first draft of the paper. S.J. performed the MALS experiments. O.V. and P.J.S. performed the molecular dynamics simulations. A.A., H.M. and T.G. carried out the biochemical work on *Pseudomonas* TonB-ExbB-ExbD. R.H.J. and B.C.B. contributed to the GldLM structure. J.W.C. initiated and provided materials for the ExbBD project. All authors commented on drafts of the manuscript.

Covid Statement

The closure of laboratories due to the Covid-19 pandemic caused a delay to the purification and characterisation of proteins described in Chapter 6. However, I was able to access computing facilities remotely to process one cryoEM dataset collected at the start of the laboratory closure and to prepare a manuscript for submission. The laboratory closure also delayed my experiments to identify cytoplasmic binding partners of *F. johnsoniae* GldL. Once laboratories reopened, I was able to complete all the experiments I had planned.

List of abbreviations

A ₂₈₀	Absorbance at 280 nm
<i>Ame</i>	<i>Apibacter mensalis</i>
ATP	Adenosine triphosphate
<i>Cca</i>	<i>Capnocytophaga canimorsus</i>
cryoEM	cryo-electron microscopy
cryoET	cryo-electron tomography
CTF	Contrast transfer function
CV	Column volume
DDM	Dodecyl maltopyranoside
DNA	Deoxyribonucleic acid
<i>Dor</i>	<i>Draconibacterium orientale</i>
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen (German Collection of Microorganisms and Cell Cultures)
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMDB	Electron Microscopy Data Bank
<i>Fjo</i>	<i>Flavobacterium johnsoniae</i>

FOM	Fluorinated octyl maltoside
<i>Fps</i>	<i>Flavobacterium psychrophilum</i>
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IM	Inner membrane
LMNG	Lauryl maltose neopentyl glycol
MALS	Multi-angle light scattering
MWCO	Molecular weight cut-off
NA	Numerical aperture
NAT	<i>Flavobacteriales</i> bacterium isolate NAT55 387147
OD ₆₀₀	Optical density at 600 nm
OM	Outer membrane
PCR	Polymerase chain reaction
PDB	Protein data bank
<i>Pgi</i>	<i>Porphyromonas gingivalis</i>
PMF	Proton-motive force
<i>Psa</i>	<i>Pseudopedobacter saltans</i>
<i>Ran</i>	<i>Riemerella anatipestifer</i>
RBS	Ribosome binding site

rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
<i>Sth</i>	<i>Schleiferia thermophila</i>
<i>Swe</i>	<i>Sphingobacterium wenxiniae</i>
T1SS	Type I secretion system
T2SS	Type II secretion system
T3SS	Type III secretion system
T4SS	Type IV secretion system
T4P	Type IV pilus
T5SS	Type V secretion system
T6SS	Type VI secretion system
T7SS	Type VII secretion system
T8SS	Type VIII secretion system
T9SS	Type IX secretion system
T10SS	Type X secretion system
T11SS	Type XI secretion system

TMH	Transmembrane helix
Tris	Trisaminomethane
<i>Tro</i>	<i>Thermonema rossianum</i>
UEG	Uncultured <i>Sphingobacteriia</i> bacterium sampled from the North Atlantic Ocean
UV	Ultraviolet
<i>Zga</i>	<i>Zobellia galactanivorans</i>

Contents

Table of Contents

Abstract	i
Acknowledgements	iii
Publications resulting from this study.....	iv
Covid Statement.....	v
List of abbreviations	vi
Contents	x
Table of Contents	x
List of Figures	xv
List of Tables.....	xviii
1 Introduction	1
1.1 Bacterial protein secretion systems.....	1
1.2 Bacterial motility systems	7
1.3 Asymmetry in protein function	11
1.4 <i>Bacteroidetes</i> bacteria use the T9SS and gliding motility for diverse purposes	14
1.5 The Type IX Secretion System and gliding motility	17
1.5.1 Early studies of gliding motility	17
1.5.2 Discovery of the Type IX Secretion System	18
1.5.3 The current model of the Type IX Secretion System.....	24

1.5.4	The current model of <i>F. johnsoniae</i> gliding motility	29
1.5.5	Current knowledge about GldL and GldM	37
1.6	Electron microscopy	42
1.7	Project aims	43
2	Methods	44
2.1	Culturing of bacteria	44
2.2	Bioinformatics analysis	49
2.3	Molecular Biology	50
2.3.1	Polymerase chain reactions	67
2.3.2	DNA sources	67
2.3.3	Restriction cloning	68
2.3.4	HiFi Assembly	68
2.3.5	Site directed mutagenesis	69
2.3.6	Agarose gels	69
2.4	Genetic modification of <i>F. johnsoniae</i>	70
2.5	Protein sample analysis	74
2.5.1	SDS-PAGE	74
2.5.2	Immunoblotting	74
2.5.3	Protein Concentration determination	75
2.5.4	Size-Exclusion Chromatography with Multi-Angle Light Scattering	75

2.5.5	Mass spectrometry.....	76
2.6	Protein Expression and Purification	76
2.6.1	Expression screens for proteins expressed from pRHJ004 plasmids	76
2.6.2	Expression screens for proteins expressed from pT12 plasmids	78
2.6.3	Purification of GldLM homologues from pRHJ004-derived vectors	79
2.6.4	Lysis and centrifugation conditions trial	83
2.6.5	Purification of GldLM homologues from pT12 vectors	84
2.7	Electron microscopy.....	86
2.7.1	Grid preparation.....	86
2.7.2	Imaging and data collection	87
2.7.3	Model building and refinement	88
2.8	<i>F. johnsoniae</i> gliding motility assays	89
2.9	<i>F. johnsoniae</i> type IX secretion assays and assessment of GldLM stability	90
3	Purification and characterisation of the full-length <i>Porphyromonas gingivalis</i> PorLM complex ..	92
3.1	Introduction	92
3.2	Construct design and expression screening.....	92
3.3	Purification of <i>Pgi</i> PorLM	98
3.4	Initial characterisation of <i>Pgi</i> PorLM.....	101
3.5	Purification of <i>Pgi</i> PorLM in LMNG	104
3.6	Purification of <i>Fjo</i> GldLM and <i>Tro</i> GldLM.....	109

3.7	Further <i>Pgi</i> PorLM characterisation	110
3.8	Discussion	114
4	Cryo-electron microscopy structure of the <i>Flavobacterium johnsoniae</i> GldLM' complex	117
4.1	Introduction	117
4.2	Design of <i>Fjo</i> GldLM' _A , <i>Fjo</i> GldLM' _B and <i>Pgi</i> PorLM' _A constructs.....	119
4.3	Purification of <i>Fjo</i> GldLM' _A , <i>Fjo</i> GldLM' _B and <i>Pgi</i> PorLM' _A	121
4.4	Solving the structure of <i>Fjo</i> GldLM'	125
4.5	A 3.9 Å structure of <i>Fjo</i> GldLM'	130
4.6	Discussion	138
5	Mutational analysis of <i>Fjo</i> GldLM.....	141
5.1	Introduction	141
5.2	Design and generation of mutants.....	142
5.3	Protein stability	142
5.4	Assessment of gliding motility on agar and glass.....	144
5.5	Assessment of T9SS protein secretion	147
5.6	Analysis of SprB propulsion.....	151
5.7	Interpretation of <i>gldLM</i> mutant phenotypes in the context of the <i>Fjo</i> GldLM' structure....	153
5.8	Discussion.....	157
6	Cryo-electron microscopy structures of GldLM' from four different species	167
6.1	Introduction	167

6.2	Purification of <i>PgiPorLM'</i> _C , <i>SthGldLM'</i> , and <i>TroGldLM'</i>	168
6.3	Expression screening and purification of new GldLM constructs from 14 species.....	171
6.3.1	Sequence selection and construct design	171
6.3.2	High throughput expression screening of different length constructs	176
6.3.3	Purification of different length GldLM constructs	180
6.4	Data processing of six GldLM cryoEM datasets	185
6.4.1	<i>TroGldLM'</i>	185
6.4.2	<i>PgiPorLM'</i> _C	187
6.4.3	<i>SthGldLM'</i>	190
6.4.4	<i>CcaGldLM''</i>	194
6.4.5	<i>CcaGldLM</i>	202
6.4.6	<i>SweGldLM'</i>	205
6.5	Comparison of GldLM' structures from five different species	210
6.6	Purification and structural characterisation of GldLM complexes with substitutions of functionally important residues.....	219
6.6.1	Purification of variant <i>SweGldLM'</i> constructs	219
6.6.2	Solving the structures of variant <i>SweGldLM'</i> constructs	222
6.6.3	Structures of <i>SweGldL_{K29A}M'</i> , <i>SweGldL_{E47D}M'</i> , <i>SweGldL_{E47Q}M'</i> , and <i>SweGldLM'_{Y17F}</i>	229
6.7	Discussion.....	233
7	Conclusions and future directions.....	236
7.1	Project Summary	236

7.2	Comparison to other ion-driven motors	241
7.3	Future directions	244
8	Bibliography	250
9	Appendix	271
9.1	Unsuccessful experiments to identify binding partners of <i>F. johnsoniae</i> GldL and GldM ...	271
9.2	Oligonucleotides used in this study	275

List of Figures

Figure 1-1	One-step secretion systems in Gram-negative bacteria.....	4
Figure 1-2	Two-step secretion systems in Gram-negative bacteria.....	5
Figure 1-3	Bacteria move with four main motility systems.	9
Figure 1-4	Symmetry mismatch in protein complex rotation.	13
Figure 1-5	Proteins involved in the T9SS and gliding motility in <i>P. gingivalis</i> (a) and <i>F. johnsoniae</i> (b).23	
Figure 1-6	SprA is the T9SS outer membrane translocon.	26
Figure 1-7	A rack and pinion model for adhesin propulsion proposed by Shrivastava and Berg.	34
Figure 1-8	A model for adhesin propulsion involving a linear motor.	36
Figure 1-9	Predicted topology and secondary structure of <i>F. johnsoniae</i> GldL.....	39
Figure 1-10	Topology and structure of <i>F. johnsoniae</i> GldM.....	41
Figure 2-1	The pYT354 genetic manipulation system.	71
Figure 3-1	Map of the empty expression vector pRHJ004, based on pETDuet-1.	94
Figure 3-2	Fluorescence measurements from the pRHJ004 expression screen.	96
Figure 3-3	Initial purification of PgiPorLM in DDM from C41 (a) and Mt56 (b) cells.	99
Figure 3-4	SEC from initial purification of <i>PgiPorLM</i> in DDM.	100

Figure 3-5 Negative stain and cryo-EM images from the initial <i>PgiPorLM</i> purification (Figure 3-4d).	102
Figure 3-6 Exemplar purification of <i>PgiPorLM</i> in DDM from 16 l of Mt56 cells.	103
Figure 3-7 Successful purification of <i>PgiPorLM</i> in LMNG.....	106
Figure 3-8 Lysis and centrifugation conditions screen.....	107
Figure 3-9 Purification of <i>PgiPorLM</i> from the cellular debris.	108
Figure 3-10 Purification of <i>FjoGldLM</i> and <i>TroGldLM</i>	109
Figure 3-11 SEC-MALS analysis of <i>PgiPorLM</i> analysed using a Superose 6 Increase 10/300 column.	110
Figure 3-12 Data processing workflow for the <i>PgiPorLM</i> map.	113
Figure 3-13 Low resolution cryo-EM map of <i>PgiPorLM</i>	114
Figure 4-1 Map of the empty pT12 vector.	119
Figure 4-2 Construct design for GldLM' constructs.....	120
Figure 4-3 Purification of <i>PgiPorLM'</i> _A , <i>FjoGldLM'</i> _A , and <i>FjoGldLM'</i> _B by StrepTactin XT affinity and size-exclusion chromatography (SEC).....	122
Figure 4-4 Negative-stain electron microscopy analysis of SEC fractions.....	123
Figure 4-5 SEC-MALS analysis of <i>PgiPorLM'</i> _A gave a calculated mass for <i>PgiPorLM'</i> _A of 236 kDa.	124
Figure 4-6 Representative 2D class averages for the <i>PgiPorLM'</i> _A (a), <i>FjoGldLM'</i> _A (b), and <i>FjoGldLM'</i> _B (c) constructs.	125
Figure 4-7 <i>FjoGldLM'</i> data processing workflow.....	128
Figure 4-8 Architecture of the <i>FjoGldLM'</i> complex.	131
Figure 4-9 The positions of the D1 domains of <i>FjoGldLM'</i> and <i>FjoGldM</i> _{peri} are incompatible.....	134
Figure 4-10 Conservation analysis of the <i>FjoGldLM'</i> complex.	135
Figure 4-11 Conserved polar residues are found in the transmembrane region of <i>FjoGldLM'</i>	137
Figure 5-1 Most proteins with substituted polar residues in the transmembrane region are stably expressed.	143

Figure 5-2 Assessment of the colony spreading phenotypes of <i>gldLM</i> mutant strains.....	145
Figure 5-3 Secretome analysis of strains with mutations in <i>gldL</i> and <i>gldM</i>	150
Figure 5-4 Measurement of SprB propulsion in selected strains.....	152
Figure 5-5 Positions of the residues substituted in this chapter on the <i>FjoGldLM'</i> structure.	153
Figure 5-6 The potential salt bridge between GldM Chain B Arg9 and GldL Chain C Glu49 causes rearrangements of GldM Chain B and GldL Chain C.	157
Figure 5-7 A model of <i>FjoGldLM</i> rotation.	161
Figure 5-8 Four models for GldLM anchoring.	164
Figure 6-1 Purification of new <i>P. gingivalis</i> , <i>S. thermophila</i> , and <i>T. rossianum</i> truncated GldLM constructs.....	171
Figure 6-2 Phylogenetic trees of GldL (a) and GldM (b) homologues.....	174
Figure 6-3 SDS-PAGE analysis of an expression screen for GldLM' constructs.	176
Figure 6-4 SDS-PAGE analysis of an expression screen for GldLM'' constructs.	178
Figure 6-5 SDS-PAGE analysis of an expression screen for GldLM''' constructs.....	179
Figure 6-6 SDS-PAGE analysis of an expression screen for GldLM constructs.....	180
Figure 6-7 Large scale purification of GldLM' constructs.....	182
Figure 6-8 Large scale purification of GldLM'' and GldLM complexes.	184
Figure 6-9 Data processing workflow for the <i>TroGldLM'</i> dataset.....	186
Figure 6-10 Data processing workflow for the <i>PgiPorLM'</i> map.	189
Figure 6-11 Data processing workflow for the <i>SthGldLM'</i> map.	194
Figure 6-12 Data processing workflow for the <i>CcaGldLM''</i> map.....	198
Figure 6-13 Data processing workflow for the <i>CcaGldLM'</i> map.	200
Figure 6-14 Comparison of the <i>CcaGldLM'</i> and the <i>CcaGldM''_{peri}</i> structures.	201
Figure 6-15 Data processing of the <i>CcaGldLM</i> dataset.....	204

Figure 6-16 Data processing workflow for SweGldLM' map.	207
Figure 6-17 All five GldLM' structures have the same architecture.	211
Figure 6-18 The CcaGldLM', SthGldLM', and SweGldLM' structures are at higher resolution than the FjoGldLM' structure.	212
Figure 6-19 <i>P. gingivalis</i> PorL has an N-terminal helix that is not present in GldL from the other structurally characterised GldLM/PorLM motor complexes.	213
Figure 6-20 The cytoplasmic region of GldL proximal to the membrane forms a cage.	216
Figure 6-21 The CcaGldLM', SthGldLM', and SweGldLM' maps have peripheral EM density that may correspond to detergent molecules.	218
Figure 6-22 Purification of variant SweGldLM' complexes.	221
Figure 6-23 Data processing workflow for SweGldL _{K29A} M' map.	223
Figure 6-24 Data processing workflow for SweGldL _{E47Q} M' map.	224
Figure 6-25 Data processing workflow for SweGldL _{E47D} M' map.	225
Figure 6-26 Data processing workflow for SweGldLM' _{Y17F} map.	226
Figure 6-27 Structures of four variant SweGldLM' structures.	230
Figure 6-28 Purification of SweGldLM' complexes with residues 75-215 of SweGldL removed (SweGldL _{Δcyto} M').	232
Figure 7-1 The GldLM motor has the same core structure as the MotAB flagellar stator unit.	243

List of Tables

Table 1-1 Proteins involved in gliding motility and the T9SS in <i>F. johnsoniae</i> and <i>P. gingivalis</i>	20
Table 2-1 Composition of the growth media used in this study.	44
Table 2-2 Antibiotics used in this study.	45
Table 2-3 <i>E. coli</i> strains used in this study.	46
Table 2-4 <i>F. johnsoniae</i> strains used in this study.	48

Table 2-5 Plasmids used for recombinant expression of proteins in this study.	51
Table 2-6 Plasmids used for generating mutant strains of <i>F. johnsoniae</i> in this study.	62
Table 2-7 Antibodies used in this study.	75
Table 3-1 Expression vectors based on pRHJ004 for co-expression of <i>gldL</i> and <i>gldM</i> homologues.	95
Table 3-2 Conditions used to screen for protein expression from pRHJ004-based vectors.	95
Table 3-3 Data collection and processing statistics for <i>PgiPorLM</i>	112
Table 4-1 Cryo-EM data collection, refinement, and validation statistics for the <i>FjoGldLM'</i> structure.	129
Table 4-2 Contacts between the periplasmic loops of GldL and GldM.	132
Table 5-1 Summary of gliding motility and T9SS phenotypes of <i>gldLM</i> mutant strains.	146
Table 6-1 Bacteroidetes strains used in this study.	173
Table 6-2 Expected sizes of proteins from the constructs described in this chapter.	175
Table 6-3 Cryo-EM data collection, refinement, and validation statistics for the <i>PgiPorLM'</i> _C , <i>SthGldLM'</i> , <i>CcaGldM''</i> , <i>CcaGldLM'</i> , and <i>SweGldLM'</i> structures.	208
Table 6-4 Cryo-EM data collection, refinement, and validation statistics for the <i>SweGldL</i> _{K29A} <i>M'</i> , <i>SweGldL</i> _{E47Q} <i>M'</i> , <i>SweGldL</i> _{E47D} <i>M'</i> , and <i>SweGldLM'</i> _{Y17F} structures.	227
Table 9-1 Strains used to identify binding partners of <i>F. johnsoniae</i> GldL and GldM.	271
Table 9-2 Plasmids used to generate <i>F. johnsoniae</i> strains used to identify binding partners of GldL and GldM.	273
Table 9-3 Oligonucleotides used in this study.	275

1 Introduction

1.1 Bacterial protein secretion systems

All living cells are enclosed by a cell envelope, which separates their cytoplasm from the environment. In Gram-negative bacteria this envelope consists of a phospholipid bilayer inner membrane (IM), a peptidoglycan cell wall a few nanometres thick, and a lipid bilayer outer membrane (OM) containing phospholipids in the inner leaflet and glycolipids in the outer leaflet (Kamio & Nikaido, 1976; Costerton *et al*, 1974). Many Gram-negative bacteria, as well as some Gram-positive bacteria and archaea, have a proteinaceous S-layer surrounding the OM and lipopolysaccharide layer (Sleytr *et al*, 2014). To interact with their environment cells must transport substances into and out of the cell. One major category of substances transported out of the cell are proteins. These transported proteins can have diverse roles, including adhering the bacterial cell to substrate surfaces, contributing to bacterial motility, breaking down extracellular polymers into smaller molecules which can be taken up by the bacterial cell, forming a biofilm matrix, interfering with the host immune response during pathogenesis, and killing other bacterial cells during inter-strain competition for nutrients (Taglialegna *et al*, 2016; Russell *et al*, 2014; Dhillon *et al*, 2021; Berg & Anderson, 1973; Stanier, 1947; Beachey, 1981). To secrete proteins out of the cell, bacteria have evolved numerous secretion systems (Costa *et al*, 2015; Christie, 2019). Some secretion systems are particular to a small group of proteins or species, but several are widespread enough to have been designated general secretion systems. Ten such systems (Type I to Type IX and the chaperone-usher system) have been well studied so far and two more (Type X and Type XI) have recently been proposed (Christie, 2019; Palmer *et al*, 2021; Grossman *et al*, 2021).

In Gram-negative bacteria the Type I Secretion System (T1SS), Type III Secretion System (T3SS), Type IV Secretion System (T4SS), and Type VI Secretion System (T6SS) are all one-step secretion systems, where the substrate protein is transported directly from the cytoplasm to the extracellular

environment, usually in a partially unfolded state (**Figure 1-1**; (Miletic *et al*, 2021; Costa *et al*, 2021; Wang *et al*, 2017; Morgan *et al*, 2017)). In Gram-negative bacteria the Type II Secretion System (T2SS), Type V Secretion System (T5SS), Type VIII Secretion System (T8SS), Type IX Secretion System (T9SS), and chaperone-usher pathway (CU system) are two-step secretion systems, where a substrate is transported across the IM by the Sec or Tat system into the periplasm and then across the OM by the secretion system machinery (**Figure 1-2**; (Naskar *et al*, 2021; Leo *et al*, 2012; Goyal *et al*, 2014; Lauber *et al*, 2018; Busch & Waksman, 2012)). No Type VII Secretion Systems (T7SS) have been identified in Gram-negative bacteria. The Type X Secretion System (T10SS) and Type XI Secretion System (T11SS) have only been recently described and their mechanisms are at present unclear (Palmer *et al*, 2021; Grossman *et al*, 2021).

Substrates of the T1SS are released directly into the environment, but substrates of the T3SS, T4SS and T6SS can be injected into other cells (**Figure 1-1**). The T1SS consists of an IM component that hydrolyses ATP to energise the system, a membrane fusion protein (MFP) that spans the periplasm and separates the substrate protein from the periplasmic environment, and TolC, an OM protein channel that substrates pass through to leave the cell (Morgan *et al*, 2017).

There are two types of T3SS, the virulence-associated T3SS and the flagellar T3SS (Erhardt *et al*, 2010). The virulence-associated T3SS is often used by pathogenic bacteria to inject virulence factors into host cells, and recent evidence suggests that some bacteria use the T3SS injectisome to extract nutrients from the host cell cytoplasm (Pal *et al*, 2019). The flagellar T3SS is not used to inject proteins but instead assembles the flagellar filament for cellular propulsion, which can be 10-15 µm long and composed of thousands of protein subunits (DePamphilis & Adler, 1971; Erhardt *et al*, 2010). The cytoplasmic sorting platform of the T3SS uses ATP hydrolysis to power the unfolding of substrate proteins (Akedo & Galán, 2005), which are then fed through the export gate apparatus and

into the lumen of the flagellum or injectisome needle (Kuhlen *et al*, 2018). Substrates pass along the lumen of the flagellum or injectisome needle to the tip (Hu *et al*, 2018; Miletic *et al*, 2021). At the tip of the flagellum substrates are assembled to extend the flagellum (Al-Otaibi *et al*, 2020). At the tip of the injectisome needle substrates are transferred across the target cell membrane into the target cell (Epler *et al*, 2012).

The T4SS is capable of injecting both proteins and DNA into target cells (Cascales & Christie, 2004). The T4SS is used in this way both to transfer plasmids during bacterial conjugation and to manipulate host gene expression during pathogenesis (Christie *et al*, 2014; Wallden *et al*, 2010). The T4SS in Gram-negative bacteria consists of an IM complex, an OM complex, and a central cylinder between the two membranes (Hu *et al*, 2019; Low *et al*, 2014). ATPases associated with the IM complex unfold substrates and feed them into the central cylinder (Bhatty *et al*, 2013). Substrates pass through the central cylinder to the OM complex, which is usually associated with a conjugative pilus that attaches the T4SS to the target cell (Arutyunov & Frost, 2013). In *Bordetella pertussis*, pertussis toxin is exported to the periplasm as two subunits via the Sec pathway and then the holotoxin is secreted from the across the OM via the T4SS (Locht *et al*, 2011). This makes the *B. pertussis* T4SS a hybrid one-step/two-step system.

The T6SS is used as a “bacterial speargun”, structurally similar and genetically homologous to some bacteriophage contractile tails (Leiman *et al*, 2009; Zoued *et al*, 2014), to propel effector proteins into a target bacterial cell. The spear is composed of Hcp proteins and is surrounded by an extended sheath complex in the cytoplasm (Cascales, 2008). A cell envelope-spanning complex orients the T6SS toward a target cell (Cherrak *et al*, 2018). When a target cell is detected the sheath contracts, driving the spear into the target cell (Wang *et al*, 2017; Brackmann *et al*, 2018). Effector proteins are attached to the VgrG spike at the tip of the spear and are driven through the envelope of the target

cell by sheath contraction, releasing the effector proteins into the periplasm or cytoplasm of the target cell (Böck *et al*, 2017). These effector proteins usually contribute to killing of non-kin cells during inter-strain competition over nutrients (Jurėnas & Journet, 2021). The ClpV ATPase disassembles the contracted sheath so that subunits can be reassembled into a new extended sheath (Kapitein *et al*, 2013).

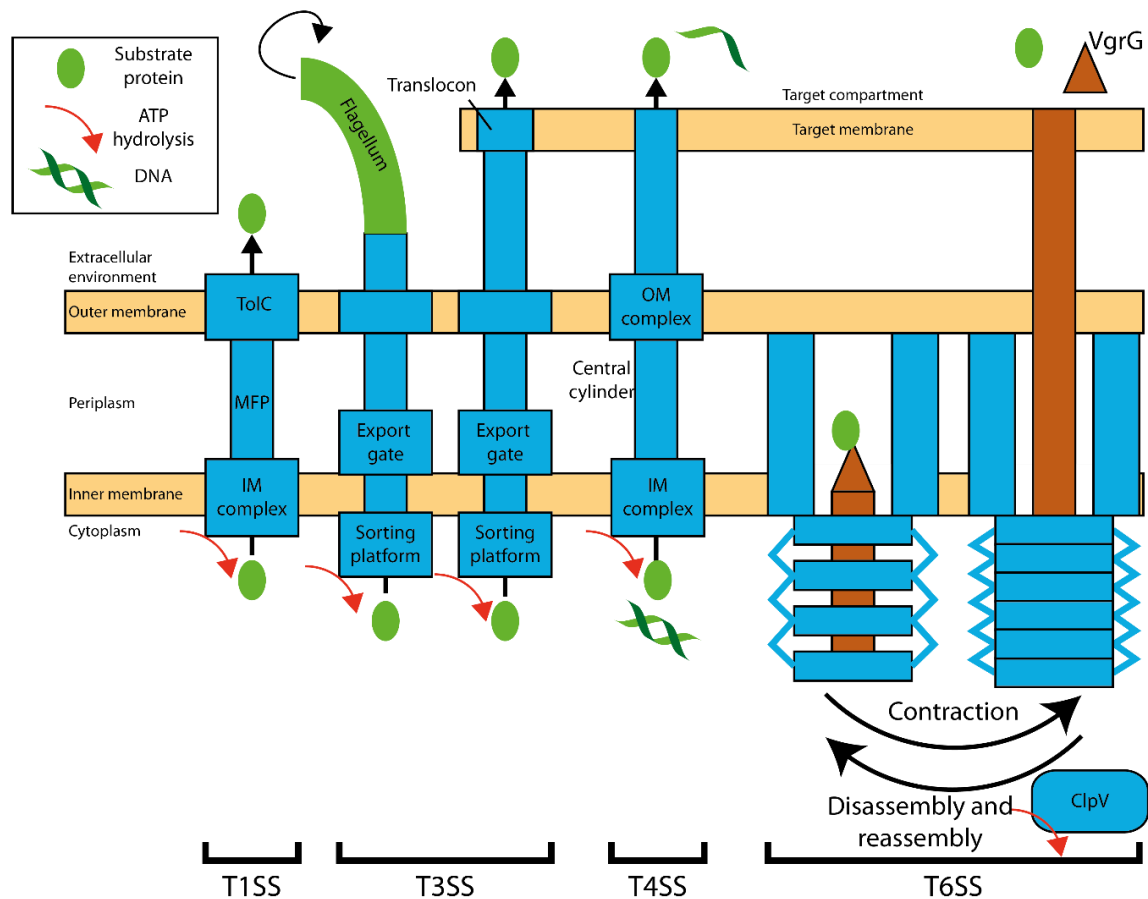


Figure 1-1 One-step secretion systems in Gram-negative bacteria. Components are not drawn to scale. Green ovals represent substrate proteins, black arrows represent the routes across the cell envelope taken by substrate proteins. Red arrows indicate sites of ATP hydrolysis. The T1SS secretes substrates from the cytoplasm to the extracellular environment. Secretion is powered by ATP hydrolysis. The flagellar T3SS secretes subunits of the flagellum and assembles them. The virulence-associated T3SS injects effector proteins into target cells. ATP hydrolysis powers unfolding of substrates and loading into the export gate. The T4SS secretes proteins and single stranded DNA (green helices) from the cytoplasm into target cells. ATP hydrolysis by cytoplasmic ATPases associated with the IM complex powers secretion. The T6SS injects effector proteins into target cells. Contact with a target cell causes contraction of the extended sheath, driving the VgrG spike and associated effector proteins across the target cell membrane. ATP hydrolysis by ClpV powers disassembly of the contracted sheath so the components can be reassembled into a new extended sheath.

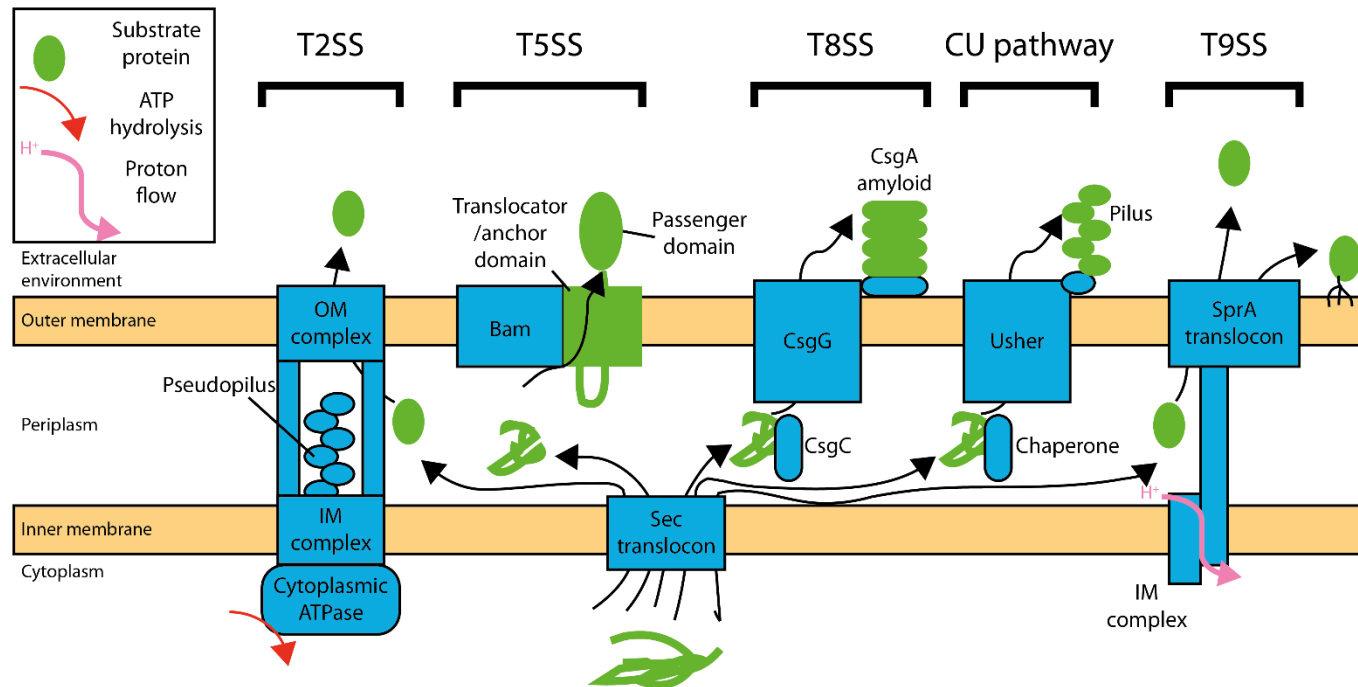


Figure 1-2 Two-step secretion systems in Gram-negative bacteria. Components are not drawn to scale. Green ovals represent substrate proteins, black arrows represent the routes across the cell envelope taken by substrate proteins. Substrates of two-step secretion systems are transported across the IM in an unfolded state by the Sec machinery. The T2SS uses a T4P-like pseudopilus to push substrates out of the periplasm through an OM complex. Pseudopilus extension is powered by cytoplasmic ATP hydrolysis (red arrow). T5SS substrates have C-terminal translocator/anchor domains that form a hybrid β -barrel with BamA from the Bam machinery. The N-terminal passenger domain passes through the hybrid barrel in an unfolded state and folds outside the cell. Most passenger domains remain attached to their translocator/anchor domains, but some are released into the environment. T8SS and CU pathway substrates are bound by chaperone proteins in the periplasm. They then pass through OM β -barrel proteins (CsgG or Usher) and assemble into amyloid fibres (T8SS) or pili (CU pathway). T9SS substrates cross the OM in a folded state through the SprA β -barrel. The T9SS is energised by proton flow (pink arrow) from the periplasm to the cytoplasm through the IM complex.

Of the two-step systems that transport substrates across just the OM, the T2SS and T9SS have components in both the IM and OM, whilst the T5SS and T8SS do not have IM components (**Figure 1-2**). The T2SS consists of an IM platform that anchors the complex, a cytoplasmic ATPase that binds the IM platform and hydrolyses ATP to energise the system, an OM complex through which substrate proteins are secreted and a periplasmic pseudopilus that pushes substrates through the OM complex when extended (Naskar *et al*, 2021). The components of the T2SS are related to the Type IV Pilus (T4P) machinery, but unlike the T4P the T2SS pseudopilus does not leave the periplasm (Denise *et al*, 2019).

T5SS substrates are termed autotransporters, with a C-terminal translocator/anchor domain that forms a β -barrel in the OM through which the N-terminal passenger domain is transported (Leo *et al*, 2012). Recent evidence suggests that the translocator domain does not transport the passenger domain by itself, but instead forms a hybrid barrel with the β -barrel assembly machinery (BAM) protein Bama whilst the translocator domain is being inserted into the OM (Doyle & Bernstein, 2021). The passenger domain crosses the OM through this hybrid channel and then the translocator domain finishes insertion and is released from the BAM complex. Most T5SS substrates remain attached to the OM via their β -barrel translocator/anchor domain, but some are released into the environment by proteolysis.

T8SS substrates are bound in the periplasm by chaperone proteins. In the *E. coli* curli system the chaperone CsgC binds the substrate CsgA (Evans *et al*, 2015). CsgC transfers CsgA to a 36-stranded β -barrel made of 9 CsgG subunits through which unfolded CsgA subunits pass (Goyal *et al*, 2014), before folding into amyloid fibrils that attach to the exterior of the cell and contribute to biofilm formation (Wang *et al*, 2007; Hammer *et al*, 2012). The CU pathway assembles pili on the cell surface and functions similarly to the T8SS (Busch & Waksman, 2012). Substrate proteins are bound in the periplasm by a chaperone protein and then transferred to the usher, an OM β -barrel protein. The substrate protein crosses the OM through the usher barrel and then assembles into a multimeric pilus that remains

anchored on the cell surface by the usher. As they lack IM components, the T5SS, T8SS, and CU pathway do not have access to the PMF across the IM or cytoplasmic ATP as an energy source. Instead, they are thought to rely on the energetic favourability of the substrate folding process after secretion and the Brownian ratchet created by the impossibility of retro-translocation once the substrate has folded (Goyal *et al*, 2014; Peterson *et al*, 2010).

The T9SS consists of a large β -barrel in the OM through which substrates are transported, using energy from the proton-motive force (PMF) across the IM (Lauber *et al*, 2018; McBride, 2019) as discussed further below.

1.2 Bacterial motility systems

It has been known since the very first microscopic observations of bacteria by Antonie van Leeuwenhoek that bacteria can move through their environments (van Leewenhoek, 1677). Since then, improvements in microscopic, genetic, and biochemical techniques have led to the discovery of several different motility systems used by various species of bacteria in various environments (Henrichsen, 1972). Motility allows bacteria to move away from crowded or toxic unfavourable environments and toward unoccupied or nutrient-rich favourable environments. Usually bacteria move along chemical gradients (chemotaxis) (Adler, 1966), but bacteria that respond to light, as well as electric and magnetic fields have also been observed (Shi *et al*, 1996; Blakemore, 1975; Chau *et al*, 2017). Motility behaviours observed in bacteria have included swimming, swarming, twitching, and gliding (**Figure 1-3**; (Henrichsen, 1972)).

The most-well studied form of bacterial motility is flagellar swimming. The flagellum is a long filament secreted via a dedicated T3SS that makes up part of the basal body of the flagellum (Erhardt *et al*, 2010). Rotation of the flagellum propels cells through liquid environments (**Figure 1-3a** (Berg & Anderson, 1973)). Flagellar swimming usually consists of long straight runs interrupted by random reorientations.

By reorienting more often in unfavourable conditions and less often in favourable conditions, bacteria undergo a biased random walk and on average move from unfavourable conditions to favourable conditions (Macnab & Koshland, 1972; Berg & Brown, 1972). The mechanism of reorientation varies between species depending on the exact nature of the flagellar machinery. *Escherichia coli* is peritrichous, having many flagella all over the cell envelope. When the flagella rotate counterclockwise, they form a bundle and propel the cell forward. When the flagella rotate clockwise, the bundle breaks up and the cell tumbles randomly. When the flagella rotate counterclockwise again the cell moves in a straight line in a new direction (Larsen *et al*, 1974). Bacteria with single flagellum (monotrichous bacteria) cannot tumble but use other mechanisms to reorient. *Vibrio alginolyticus* has a single polar flagellum that can pull the cell when rotating clockwise or push the cell when rotating counterclockwise. In *V. alginolyticus*, the hook at the base of the flagellum where it joins the cell is unstable and buckles when the direction of flagellar rotation changes (Son *et al*, 2013). This causes the flagellum to flick and orient the cell in a new direction, equivalent to tumbling in *E. coli*. The single flagellum of *Rhodobacter sphaeroides* only rotates clockwise, propelling the cell in a single direction. When the flagellum of *R. sphaeroides* stops rotating, it coils against the cell body and the cell reorients by Brownian motion before moving in a new direction when the flagellum rotates again (Armitage & Macnab, 1987).

Whilst swimming cells use their flagella to move in three dimensions in a liquid, swarming cells use their flagella to move in two dimensions on a solid or semi-solid surface (**Figure 1-3b** (Henrichsen, 1972)). Swarming cells often differentiate into a swarming morphology with elongated cell bodies and hyperflagellation compared to swimming cells. Swarming appears to require multiple cells to form a raft together. It is hypothesised that groups of cells can trap more water to push against with their flagella and lubricate their interaction with the surface than isolated cells can (Mezanges *et al*, 2012). Some species also secrete surfactants to reduce the friction between the cell and the surface (Harshey, 1994).

It has been shown that swarming behaviour increases resistance to some antibiotics in *Salmonella typhimurium* (Kim *et al*, 2003), so it may provide “safety in numbers” by an as yet unknown mechanism.

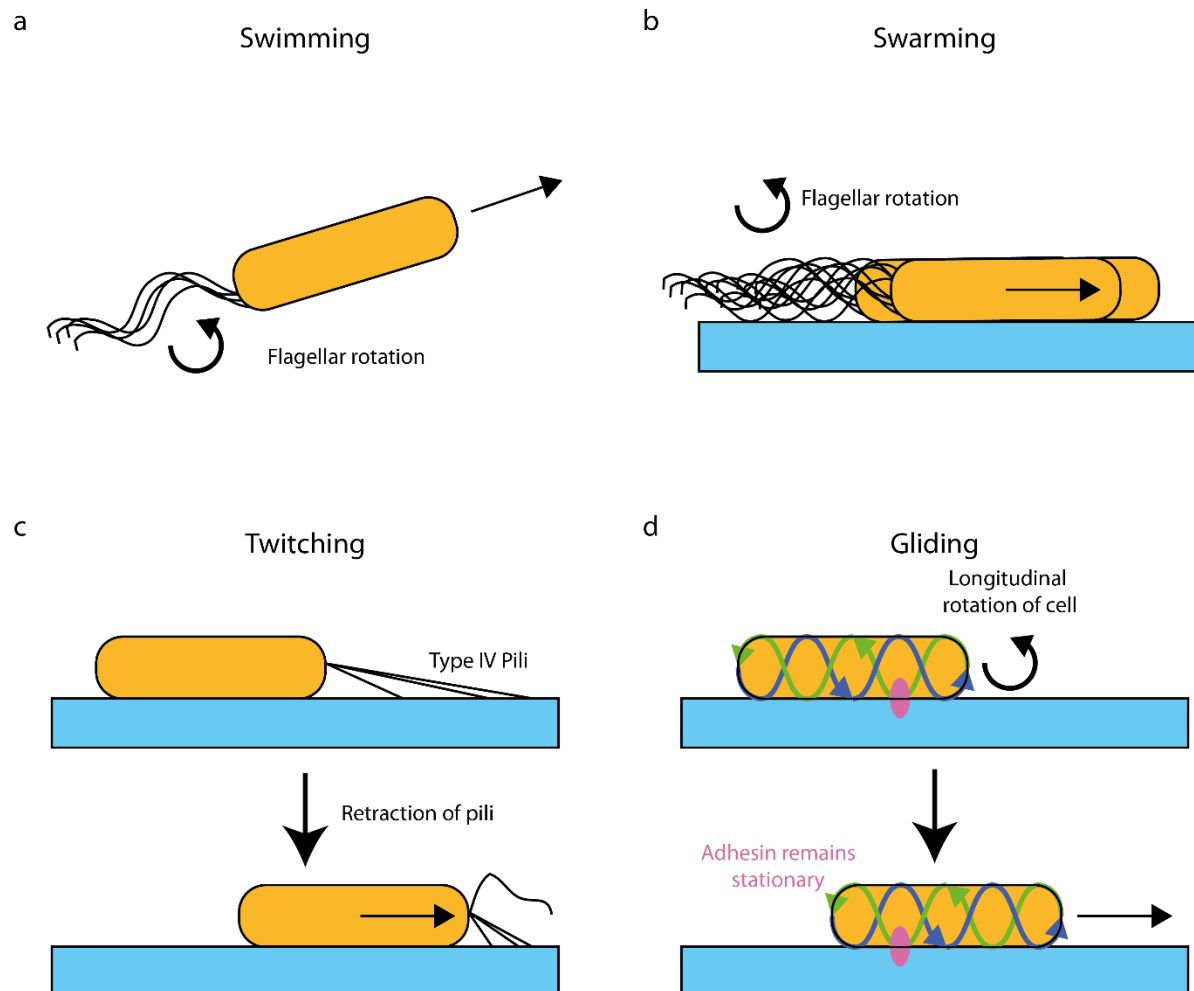


Figure 1-3 Bacteria move with four main motility systems. **a** Flagellated strains can swim in liquid media by rotating their flagella. **b** Some flagellated strains can swarm in groups on solid surfaces by rotating their flagella. **c** Some piliated strains demonstrate twitching, pulling themselves along surfaces by extending and retracting pili. **d** Some strains of *Myxococcales*, *Mycoplasma*, and *Bacteroidetes* bacteria can glide on surfaces. In *Myxococcales* and *Bacteroidetes* bacteria it is thought that external adhesin molecules are propelled along an internal track. When the adhesin is immobilised by adhering to a surface the cell is instead propelled forward in a corkscrew motion (Wartel & Mignot, 2014). Despite their similar mechanisms the proteins involved in *Myxococcales* and *Bacteroidetes* bacteria are unrelated. *Mycoplasma* bacteria are proposed to use proteins fixed on the cell surface which “walk” along the substrate surface (Miyata & Hamaguchi, 2016).

Twitching is a form of surface motility in which cells move in a jerky fashion by extension and retraction of T4Ps (**Figure 1-3c**; (Merz *et al*, 2000)). Twitching uses a T4P, which is extended by ATP-dependent

polymerisation and has been likened to a grappling hook (Burrows, 2012; Chang *et al*, 2016). Once the tip of the pilus engages with a surface, ATP-dependent depolymerisation of the pilus occurs, retracting the pilus back into the cell and generating tension to pull the cell forward. In rod-shaped bacteria such as *Pseudomonas aeruginosa*, the pili are localised at one pole and the direction of crawling can be switched by changing which pole the pilus machinery is localised to, allowing cells to achieve chemotaxis in a nutrient gradient (Oliveira *et al*, 2016). Spherical cells such as *Neisseria gonorrhoeae* extend pili in all directions, with all pili being under tension. When one pilus breaks or detaches from the substrate surface, the cell jerks in the opposite direction, generating random walk behaviour (Marathe *et al*, 2014).

Gliding motility is the movement of bacteria on a surface independent of propulsive structures such as flagella or pili (**Figure 1-3d**). This activity has been observed in many different species but has been best studied in *Myxococcus xanthus*, *Flavobacterium johnsoniae*, and *Mycoplasma mobile* (Wartel & Mignot, 2014; Miyata, 2010), although it should be noted that *M. xanthus* also exhibits twitching motility. *M. xanthus* and *F. johnsoniae* are found in separate phyla and glide through systems that are mechanistically similar but have evolved independently and do not use related proteins. In *M. xanthus* and *F. johnsoniae*, adhesin proteins are attached to the cell OM and adhere to substrate surfaces. Internal motor proteins propel the adhesin molecules along an internal helical track and, as the adhesins are fixed to the substrate surface, the cell is propelled like a corkscrew in the opposite direction (Wartel & Mignot, 2014). *M. xanthus* gliding motility and *F. johnsoniae* gliding motility operate at different speeds, *M. xanthus* typically moves at $5\ \mu\text{m min}^{-1}$ whilst *F. johnsoniae* moves at $2\text{-}5\ \mu\text{m s}^{-1}$ (Sun *et al*, 2011; Lapidus & Berg, 1982). *M. xanthus* gliding appears to rely on the cytoskeleton and is inhibited by the drug A22, which disrupts actin-like MreB filaments (Mauriello *et al*, 2010). *F. johnsoniae* gliding motility is not inhibited by A22 and is presumably not dependent on the cytoskeleton (Shrivastava & Berg, 2020). *F. johnsoniae* gliding motility and its relationship to the T9SS is discussed further below.

M. mobile is a parasitic bacterium that infects the gills of fish (Razin *et al*, 1998). *M. mobile* gliding motility does not resemble *M. xanthus* or *F. johnsoniae* gliding motility. *M. mobile* cells have a membrane protrusion that contains an internal “jellyfish structure” with a “bell” in the pole of the protrusion and “tentacles” within the cell that extend from the bell back toward the main body of the cell (Miyata & Hamaguchi, 2016). External filaments bind to sialylated oligosaccharides on the surface of host cells and are anchored by the tentacle fibres. It is proposed that ATP-dependent conformational changes of the external filaments drive a cycle of substrate binding, contraction, substrate release, and extension that lets them “walk” across the substrate surface (Miyata & Hamaguchi, 2016). The combined activity of hundreds or thousands of external filaments pulls the bacterium across the host cell surface in the direction of the membrane protrusion pole.

1.3 Asymmetry in protein function

Most proteins are predicted to form oligomeric assemblies. One survey of *E. coli* proteins found that 81% of proteins which have been studied form oligomers rather than monomers and that 79% form homooligomers rather than heterooligomers (Goodsell & Olson, 2000). Many homooligomeric protein complexes display symmetry at the global level, though this is usually broken by differing sidechain orientations and small loop movements at the local level. In cyclic oligomers, it is often the case that each subunit is in a different conformation that leads to a different state of competence for catalysis or substrate binding. For instance, in the F_1 domain of the ATP synthase, three pairs of $\alpha_1\beta_1$ dimers form a heterohexamer surrounding the central γ subunit stalk (Abrahams *et al*, 1994). The γ stalk presents a different interaction surface to each $\alpha\beta$ pair, so each pair adopts a different conformation. As the γ stalk rotates due to rotation of the c subunit ring in the F_0 domain, each $\alpha\beta$ pair is driven to cycle through three states: L, T, and O. The L state loosely binds ADP and P_i , the T state tightly binds ADP and P_i to catalyse the formation of ATP, and the O state releases ATP and binds fresh ADP and P_i molecules (Walker, 1998).

The cycling of $\alpha\beta$ subunit pairs in the F_1 ATPase through locally asymmetric pairs is driven by a mismatch between the threefold symmetry of the $(\alpha\beta)_3$ assembly and the onefold symmetry of the γ stalk. Symmetry mismatches have been seen in several other rotating systems. When studying how DNA is packaged into the heads of bacteriophages, Roger Hendrix proposed that the mismatch between the 6-fold symmetry of the bacteriophage tail and the 5-fold symmetry of the attachment point on the icosahedral head facilitated rotation of the connector between the head and tail to wind DNA into the head (Hendrix, 1978). He noted that if two cyclic homohexamers A_6 and B_6 rotate against each other each A:B interface is the same, so when one A subunit is in its most energetically favourable position all six A subunits are at their most energetically favourable position (**Figure 1-4**). The additive effect of all six A subunits being at their most energetically favourable position at once presents a very high energy barrier that prevents rotation away from this position and makes an A_6B_6 dodecamer very stable. However, if an A_5 homopentamer rotates against a B_6 homohexamer only one A subunit can be in the most energetically favourable position at once (**Figure 1-4**). The other subunits will be in less energetically favourable positions and so the energy barrier resisting rotation is much lower in the A_5B_6 complex than in the A_6B_6 complex. Hendrix proposed that the connector between the bacteriophage tail and the head had 6-fold symmetry to facilitate rotation against the head during DNA packaging (Hendrix, 1978). Later structural evidence showed that the connector in fact has 12-fold symmetry and can rotate against the bacteriophage head during phage maturation, but it is no longer thought that the connector rotates during DNA packaging (Simpson *et al*, 2000; Xu *et al*, 2019).

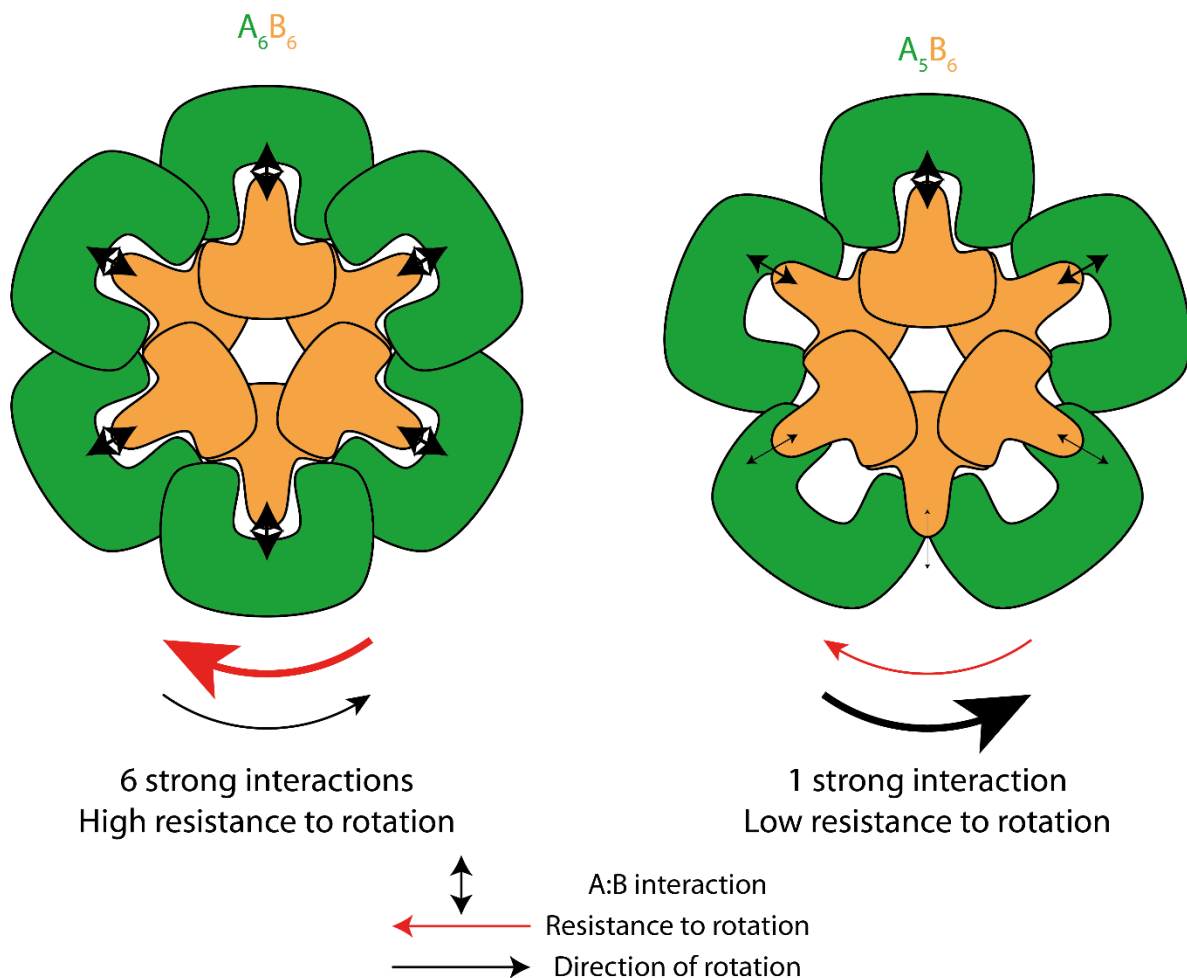


Figure 1-4 Symmetry mismatch in protein complex rotation. In an A_6B_6 complex (left) all six A:B interfaces are in the same conformation at once so the complex adopts a stable conformation that is resistant to rotation. In an A_5B_6 complex (right) all the A:B interfaces are in different conformations so the overall A_5B_6 interaction is weaker and free rotation about the central axis is possible. Black arrows represent direction of rotation, red arrows represent the resistance to rotation, double headed arrows represent A:B interactions. The size of each arrow corresponds to the magnitude of each force/movement.

Another role for symmetry mismatch facilitating rotation has been proposed in the ClpAP and ClpXP bacterial proteases. The protease ClpP forms a tetradecamer arranged as two rings of seven subunits, with the two rings joined tail to tail (Wang *et al*, 1997). The unfoldases ClpA and ClpX form hexamers that are members of the same AAA+ protein family as the MCM helicase (Grimaud *et al*, 1998). One or both ClpP heptameric rings is capped by ClpA or ClpX to form a functional complex (Ortega *et al*, 2004; Lopez *et al*, 2020; Ripstein *et al*, 2020). On the basis of this mismatch it has been proposed that ClpA or

ClpX rotates against ClpP to processively unfold substrate proteins and feed the unfolded polypeptides into the ClpP protease (Ripstein *et al*, 2020; Lopez *et al*, 2020). However, crosslinked ClpAP complexes where rotation is not possible are capable of proteolysis at a slower rate than non-crosslinked complexes (Kim *et al*, 2020). This suggests that rotation is not essential to ClpAP function but that local movements about the ClpA:ClpP interface still plays a role in proteolysis.

Recently determined structures of the MotAB flagellar stator unit show that it has a MotA₅B₂ stoichiometry, rather than the previously assumed MotA₄B₂ (Deme *et al*, 2020; Santiveri *et al*, 2020). The two TMHs of the MotB dimer are surrounded by a ring of five pairs of TMHs from the MotA pentamer. This has led to a model whereby ion flow drives alternating conformations of the MotB TMHs that lead to rotation of the MotA pentamer around the MotB dimer, discussed further in 7.2. It has been proposed that *Bacteroidetes* gliding motility involves a rotary motor (see 1.5.4 and (Shrivastava *et al*, 2015)), which may provide a new example of symmetry mismatch in protein function.

1.4 *Bacteroidetes* bacteria use the T9SS and gliding motility for diverse purposes

The *Bacteroidetes*¹ are a diverse phylum of Gram-negative bacteria (Thomas *et al*, 2011). Members of the *Bacteroidia* class are one of the dominant groups in mammalian gut microbiota, whilst other *Bacteroidetes* bacteria have been found in freshwater, marine, soil, and aerial environments (Lauber *et al*, 2009; Kirchman, 2002; Glöckner *et al*, 1999; Buczolits *et al*, 2002). Environmental *Bacteroidetes* bacteria are often specialised in digesting high molecular weight molecules such as polysaccharides and proteins (Church, 2008). The *Bacteroidetes* phylum includes human pathogens such as *Bacteroides fragilis*, *Elizabethkingia meningoseptica*, and *Porphyromonas gingivalis* (Kim *et al*, 2005; Brook, 2010;

¹ Note that it has been proposed to rename the *Bacteroidetes* to *Bacteroidota* as part of standardising names of phyla for inclusion in the International Code of Nomenclature of Prokaryotes (Whitman *et al*, 2018), for consistency with current usage I refer to *Bacteroidetes* throughout this thesis.

Shah & Collins, 1988) as well as agriculturally relevant fish, poultry, and plant pathogens (Segers *et al*, 1993; Li *et al*, 2017; Liao & Wells, 1986).

The T9SS is found exclusively in the *Bacteroidetes* phylum of Gram-negative bacteria (McBride & Zhu, 2013). The gliding motility system used by *F. johnsoniae* is also found exclusively in bacteria of the *Bacteroidetes* phylum and is linked to the T9SS, as discussed in 1.5. Genes encoding the T9SS and gliding motility machineries have been found in *Bacteroidetes* species inhabiting a diverse range of environments (McBride & Zhu, 2013). Most species with a T9SS also have genes encoding the gliding motility machinery. However, it is not known if all strains with these gliding genes are motile as it is difficult to experimentally verify whether apparently non-motile strains are truly non-motile or just non-motile in the conditions tested (Bernardet *et al*, 2002; McBride & Zhu, 2013). Gliding motility and T9SS substrates play a variety of roles in the lifestyles of *Bacteroidetes* bacteria, some of which are described below.

P. gingivalis is a non-motile T9SS⁺ bacterium. It is the keystone pathogen of chronic periodontitis and has been linked to other chronic conditions including rheumatoid arthritis and an increased risk of developing certain cancers (Qi *et al*, 2020; Nakayama, 2015; Potempa *et al*, 2017; Herzberg & Meyer, 1996). *P. gingivalis* uses the T9SS to secrete a host of virulence factors, including gingipain proteases which are attached to the cell surface and degrade host immune factors (Pathirana *et al*, 2007). *P. gingivalis* secretes these gingipain proteases in such large numbers that a T9SS-dependent electron dense surface layer is visible surrounding cells (Glew *et al*, 2012). *P. gingivalis* also secretes peptidylarginine deaminase, an enzyme that citrullinates proteins to disrupt the host immune response. Protein citrullination causes an immune cascade thought to result in the development of rheumatoid arthritis (Potempa *et al*, 2017). Other human oral pathogens such as *Tannerella forsythia* also use the T9SS to secrete virulence factors (Narita *et al*, 2014). *T. forsythia* strains with T9SS genes deleted were

more prone to forming biofilms, but another oral pathogen *Capnocytophaga ochracea* was less prone to forming biofilms when the T9SS gene *gldK* was deleted (Kita *et al*, 2016; Narita *et al*, 2014). Deletion of the gene encoding the gliding motility adhesin *sprB* also reduced biofilm formation in *C. ochracea*, suggesting that biofilm formation in this species relies on gliding motility rather than the T9SS directly. Animal pathogens also rely on the T9SS for virulence. *Flavobacterium psychrophilum* and *Flavobacterium columnare* both cause disease in farmed fish and have reduced virulence if T9SS genes are deleted (Barbier *et al*, 2020; Li *et al*, 2017). The T9SS is so essential to virulence in the poultry pathogen *Riemerella anatipestifer* that deletion of one T9SS gene rendered cells avirulent and suitable for use as an attenuated live vaccine in ducklings, possibly because they no longer secreted proteases that interfere with the host complement system (Guo *et al*, 2017).

Many non-pathogenic environmental bacteria such as *F. johnsoniae* and *Cytophaga hutchinsonii* use the T9SS to secrete carbohydrate digesting enzymes (McBride *et al*, 2003; Zhu & McBride, 2014). This allows them to digest large polysaccharides in their environment into smaller fragments which they can import and metabolise. Metagenome studies have identified candidate species that appear to use T9SSs to secrete cellulose-digesting enzymes in the guts of capybara and domestic cattle (Naas *et al*, 2018; Cabral *et al*, 2021). The T9SS is therefore important in the complex microbiome that allows these mammals to access the energy stored in plant cellulose even though they lack the necessary enzymes themselves. Interestingly, even though *Bacteroidetes* are the dominant phylum in the adult human gut microbiome (Rajilić-Stojanović & de Vos, 2014), no species resident in the human gut have yet been described that have T9SS-encoding genes. *Flavobacterium* sp. strain F52 has been shown to rely on gliding motility to colonise plant roots and seeds (Kolton *et al*, 2014). Colonisation with *Flavobacterium* sp. strain had a T9SS-dependent protective effect for tomato seedlings challenged with the pathogen *Clavibacter michiganensis* (Kolton *et al*, 2014), so there may be a role for the T9SS in beneficial plant-bacteria relationships.

The involvement of the T9SS and gliding motility in human and animal diseases means that better understanding of the T9SS and gliding motility will lead to better understanding of these diseases, as well as the agriculturally- and ecologically- relevant processes that non-pathogenic T9SS⁺ species are involved in.

1.5 The Type IX Secretion System and gliding motility

1.5.1 Early studies of gliding motility

Members of the *Bacteroidetes* phylum have been observed for over a century to exhibit gliding motility and cells of the soil bacterium *Flavobacterium johnsoniae* were first described as exhibiting gliding motility over 70 years ago (Stanier, 1947; Hutchinson & Clayton, 1919). Based on this observation, *F. johnsoniae* (formerly *Cytophaga johnsonae*) was originally classified as a non-fruiting myxobacterium related to *M. xanthus* (Stanier, 1947), but later genetic evidence showed that it belonged in the *Bacteroidetes* phylum (Bernardet *et al*, 1996; Krieg *et al*, 2010). *F. johnsoniae* cells are non-flagellated thin rods of 0.5 µm by 5-10 µm which form spreading colonies on agar. Early workers observed that in droplets of media on glass *F. johnsoniae* cells move in the direction of their long axis and change direction by either reversing their direction of motion or by spinning about one pole (Henrichsen, 1972; Lapidus & Berg, 1982). How *F. johnsoniae* cells moved without flagella was unclear, but some workers isolated structures from *F. johnsoniae* that resembled the basal bodies of flagella and proposed that they might form rotary surface assemblies that drove motility (Pate & Chang, 1979). Recently published structures of the membrane remodelling protein PspA closely resemble the structures isolated by Pate and Chang, suggesting that these structures may be PspA and are not in fact involved in gliding motility (Junglas *et al*, 2021; Liu *et al*, 2021).

Once genetic techniques to manipulate *F. johnsoniae* and related species were established (Kempf & McBride, 2000; McBride & Kempf, 1996; Zhu *et al*, 2017), the McBride laboratory were able to identify the *gld* genes necessary for gliding motility (Braun *et al*, 2005; Agarwal *et al*, 1997; McBride & Braun,

2004; McBride *et al*, 2003; Hunnicutt *et al*, 2002; Hunnicutt & McBride, 2001, 2000; Braun & McBride, 2005). *gld* cells do not form spreading colonies on agar or glide as individual cells on glass. These workers also identified *spr* genes, which are necessary for colony spreading on agar but not for individual cell motility (Nelson *et al*, 2008; Rhodes *et al*, 2011a; Nelson *et al*, 2007). The *F. johnsoniae* protein SprB was found to form filaments on the cell surface that moved continuously in helical paths along the length of the cell when immunofluorescently labelled (Nelson *et al*, 2008; Liu *et al*, 2007). On this basis it was proposed that SprB is an adhesin that binds to the substrate surface and is moved relative to the cell by a motor complex to propel the cell (**Figure 1-3d and 0**). *rem* genes have also been identified, mutations of which caused additional motility defects in *sprB*⁻ cells (Shrivastava *et al*, 2012). The roles of the other *gld*, *spr*, and *rem* gene products remained unclear, but it was noted that as well as being non-motile, *gld* mutants were unable to utilise chitin or adhere to surfaces and were resistant to infection by some bacteriophages (Braun *et al*, 2005; Chang *et al*, 1984).

1.5.2 Discovery of the Type IX Secretion System

At the same time as workers in the McBride laboratory were identifying genes involved in *F. johnsoniae* gliding motility, workers in other laboratories were investigating secretion of virulence factors by the human oral pathogen *Porphyromonas gingivalis*. The three gingipain proteases in *P. gingivalis* have conserved C-terminal domains (CTDs) which are important for secretion and surface attachment (Seers *et al*, 2006). Similar CTD sequences were found in other secreted proteins in *P. gingivalis* and other *Bacteroidetes* species including *F. johnsoniae*. It was therefore suggested that these CTDs might be involved in a novel secretion system (Sato *et al*, 2005; Nguyen *et al*, 2007). In 2007 the homologous OM proteins SprA in *F. johnsoniae* and Sov in *P. gingivalis* were found to be involved in gliding motility and gingipain secretion, respectively (Nelson *et al*, 2007; Saiki & Konishi, 2007). Further work identified several more genes necessary for gingipain secretion in *P. gingivalis* that had homologues in *F. johnsoniae* necessary for gliding motility (Sato *et al*, 2010). These genes were deemed to constitute a

novel secretion system intimately involved in gliding motility, termed the Por Secretion System and later the Type IX Secretion System (Sato *et al*, 2010; McBride & Zhu, 2013).

The dual role of *gld* genes in gliding motility and secretion explains the observation that gliding-deficient *F. johnsoniae* cells cannot utilise chitin and are resistant to some bacteriophages. The chitinase ChiA and the cell surface adhesin are T9SS substrates, which cannot be secreted by *gld*⁻ cells. Without ChiA secretion cells cannot utilise chitin. Several bacteriophages use SprB or other adhesins secreted via the T9SS as receptors for adhering to target *F. johnsoniae* cells. *gld*⁻ cells lack SprB on their surfaces so bacteriophages are unable to bind to and infect the *gld*⁻ cells.

Table 1-1 lists proteins that have been identified as involved in gliding motility or the T9SS in *F. johnsoniae* and *P. gingivalis*. **Figure 1-5** shows the cellular localisations of proteins involved in gliding motility or the T9SS in *F. johnsoniae*. In this thesis I use the *F. johnsoniae* Gld-/Spr-based nomenclature unless specifically referring to *P. gingivalis* proteins. The reader should note that some papers in the literature refer to *F. johnsoniae* proteins with the names for the equivalent *P. gingivalis* proteins and vice versa.

Table 1-1 Proteins involved in gliding motility and the T9SS in *F. johnsoniae* and *P. gingivalis*. Protein names are given for the strain *F. johnsoniae* UW101.

<i>F. johnsoniae</i> protein	<i>P. gingivalis</i> protein	Localisation	Role
Fjoh_0599, Fjoh_1000, Fjoh_1688, Fjoh_1689	Omp17	Periplasm	Skp chaperone-like protein, required for PorU stability in <i>P. gingivalis</i> (Taguchi <i>et al</i> , 2015)
Fjoh_1647, Fjoh_2275, Fjoh_3476, Fjoh_3950, Fjo_3973, Fjoh_4540	PorE	Outer membrane inner leaflet lipoprotein	Anchors outer membrane components to the peptidoglycan cell wall (Trinh <i>et al</i> , 2020). PorE homologues in <i>F. johnsoniae</i> lack a conserved cysteine residue for lipidation and have not been characterised. Each PorE homologue in <i>F. johnsoniae</i> is encoded in an operon with a <i>porP/sprF</i> homologue.
GldA	-	Cytoplasm	ATP-binding cassette (ABC) transporter subunit required for gliding, substrate unknown (Hunnicut <i>et al</i> , 2002)
GldB	-	Outer membrane inner leaflet lipoprotein	Required for gliding, role unknown (Hunnicut & McBride, 2000)
GldD	-	Outer membrane inner leaflet lipoprotein	Required for gliding, role unknown (Hunnicut & McBride, 2001)
GldF	-	Inner membrane	ABC transporter subunit required for gliding, substrate unknown (Hunnicut <i>et al</i> , 2002)
GldG	-	Inner membrane	ABC transporter subunit required for gliding, substrate unknown (Hunnicut <i>et al</i> , 2002)
GldH	-	Outer membrane inner leaflet lipoprotein	Required for gliding, role unknown (McBride <i>et al</i> , 2003)
GldI	-	Outer membrane inner leaflet lipoprotein	Required for gliding, peptidyl-prolyl <i>cis/trans</i> isomerase (McBride & Braun, 2004)
GldJ	-	Outer membrane inner leaflet lipoprotein	Required for gliding. Proposed gliding motility track protein (Braun & McBride, 2005; Nakane <i>et al</i> , 2013)
GldK	PorK	Outer membrane inner leaflet lipoprotein	Required for gliding and T9SS activity. Interface between GldLM and other components (Gorasia <i>et al</i> , 2016)
GldL	PorL	Inner membrane	Required for gliding and T9SS activity. Energises secretion and motility using the PMF (Hennell James <i>et al</i> , 2021)
GldM	PorM	Inner membrane	Required for gliding and T9SS activity. Energises secretion and motility using the PMF (Hennell James <i>et al</i> , 2021)

GldN	PorN	Periplasm	Required for gliding and T9SS activity. Interface between GldLM and other components (Gorasia <i>et al</i> , 2016)
GldO	-	Periplasm	Required for gliding and T9SS activity. Partially redundant homologue of GldN (Rhodes <i>et al</i> , 2010)
Plug	Plug	Periplasm	Prevents leakage through the SprA translocon (Lauber <i>et al</i> , 2018)
PPI	-	Outer membrane outer leaflet lipoprotein	Associated with the SprA translocon, role unknown (Lauber <i>et al</i> , 2018)
SprA	Sov	Outer membrane	Required for gliding and T9SS activity. Protein secretion translocon (Lauber <i>et al</i> , 2018)
SprB	-	Externally attached to outer membrane	Required for gliding on agar. Gliding motility adhesin (Nelson <i>et al</i> , 2008)
SprC	-	Periplasm	Required for gliding on agar, role unknown (Rhodes <i>et al</i> , 2011a)
SprD	-	Outer membrane β -barrel protein with large periplasmic domain	Required for gliding on agar, role unknown (Rhodes <i>et al</i> , 2011a)
SprE	PorW	Outer membrane inner leaflet lipoprotein	Required for gliding on agar and T9SS secretion, role unknown (Rhodes <i>et al</i> , 2011b; Sato <i>et al</i> , 2010)
SprF	PorP	Outer membrane	Removal of SprB from the SprA translocon (Rhodes <i>et al</i> , 2011a; Kulkarni <i>et al</i> , 2019) Fjoh_1646, Fjoh_1986, Fjoh_3477, Fjoh_3951, Fjoh_3972, Fjoh_4359, and Fjoh_4749 are SprF homologues in <i>F. johnsoniae</i> encoded in operons with <i>sprB</i> homologues. Fjoh_1677 and Fjoh_2274 are SprF homologues in <i>F. johnsoniae</i> not encoded near <i>sprB</i> homologues on the chromosome.
SprT	PorT	Outer membrane	Required for gliding and T9SS secretion, role unknown (Sato <i>et al</i> , 2005)
-	PorG	Outer membrane β -barrel protein	Required for T9SS secretion in <i>P. gingivalis</i> , associates with PorK and PorN. No known homologues in <i>F. johnsoniae</i> . (Naito <i>et al</i> , 2019)
PorQ	PorQ	Outer membrane	Attachment complex (Glew <i>et al</i> , 2017)
PorU	PorU	Cell surface	Attachment complex, removes substrate CTD and links to A-LPS (Gorasia <i>et al</i> , 2015)
PorV	PorV	Outer membrane	Removal of substrates from the SprA translocon (Lauber <i>et al</i> , 2018)
PorX	PorX	Cytoplasm	Regulation of gene expression (Vincent <i>et al</i> , 2016)

PorY	PorY	Inner membrane	Regulation of gene expression(Vincent <i>et al</i> , 2016)
PorZ	PorZ	Cell surface	Attachment complex, transfers A-LPS to PorU (Madej <i>et al</i> , 2021)

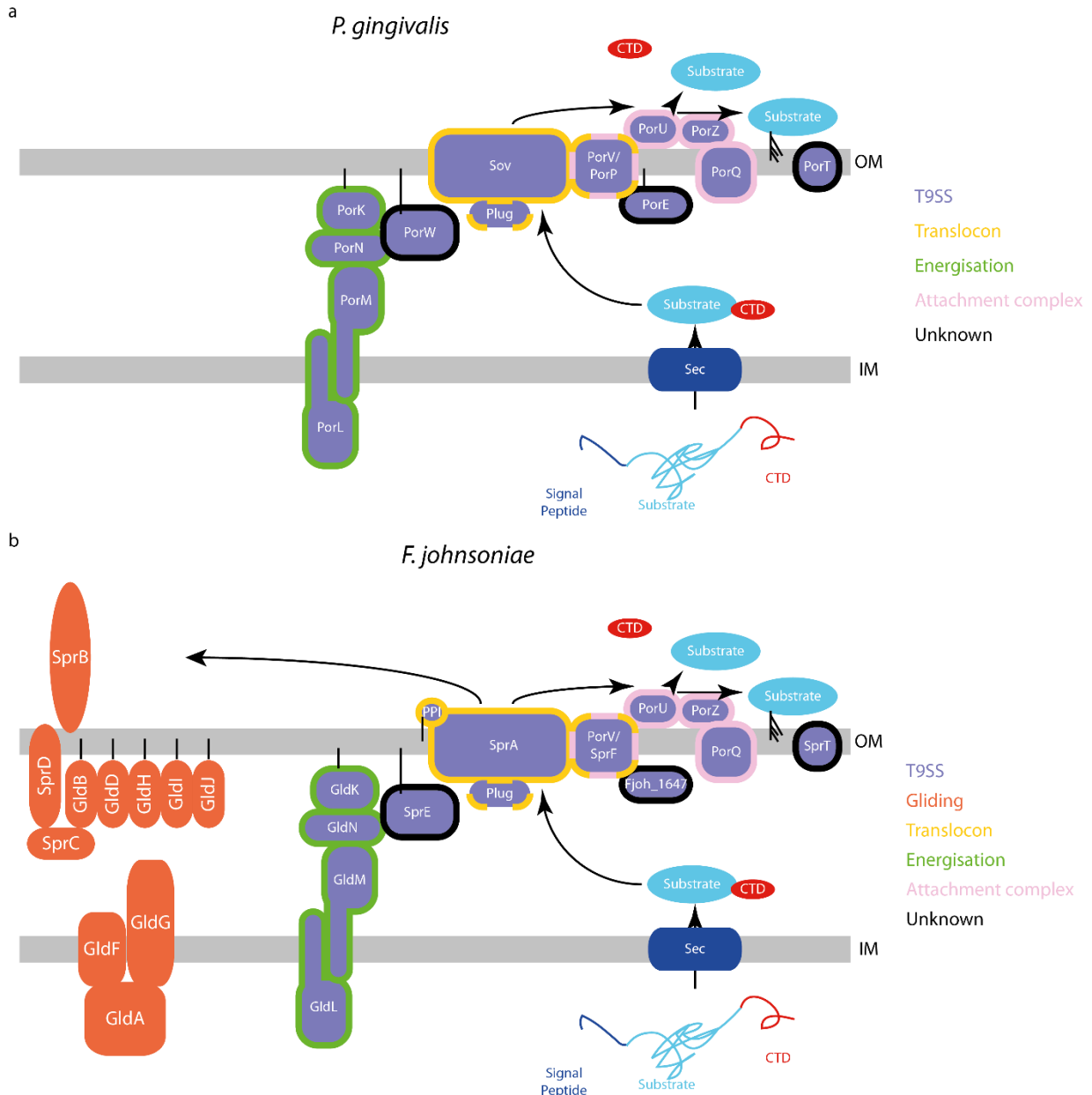


Figure 1-5 Proteins involved in the T9SS and gliding motility in *P. gingivalis* (a) and *F. johnsoniae* (b). T9SS proteins are shown in purple with a border to show their hypothesised role in the T9SS (yellow – translocon, green – energisation, pink – attachment complex, black – unknown). PorV/PorP and PorV/SprF are shown with a pink and yellow dashed border to indicate that PorV is thought to shuttle substrate proteins between the translocon and the attachment complex. It is unknown if PorP or the PorP orthologue SprF interact with the attachment complex. Plug is shown with a dashed border to indicate that it only associates with the translocon complex when the complex is in a non-translocating state. Gliding-associated proteins are shown in orange. Orthologous proteins are shown in the same position in each panel. In addition to Fjoh_1647 shown in **b**, Fjoh_275, Fjoh_3950, Fjoh_3973, and Fjoh_4540 are also homologous to PorE. Note that no T9SS substrate has yet been identified which is lipidated by the attachment complex in *F. johnsoniae*.

1.5.3 The current model of the Type IX Secretion System

In the 11 years since the T9SS was first described (Sato *et al*, 2010), a model has been developed to explain how substrate proteins are secreted and then either released into the environment or attached to the cell envelope (McBride, 2019). Nevertheless, no part of the process is yet fully understood.

The T9SS is a two-step secretion system (**Figure 1-2**), meaning that substrate proteins are transferred from the cytoplasm to the periplasm by a T9SS-independent mechanism and then transferred from the periplasm to the external environment via the T9SS. Substrates have signal sequences directing them to the Sec translocon for export across the IM. Substrates are targeted to the T9SS via their folded CTD, so it is presumed that they are secreted in a folded state, but it is possible that only the CTD is folded and other regions are in an unfolded state during secretion. Sequence alignments of proteins identified as T9SS substrates showed that in *F. johnsoniae* there are two families of CTDs, termed Type A and Type B CTDs (Kulkarni *et al*, 2017). In *P. gingivalis* there is only one protein with a Type B CTD, PG1035 (strain W83 locus, (Kulkarni *et al*, 2019)). The Type A CTDs have a YPNP motif approximately 70 amino acids from the C-terminus, a G(I/L/V)Y motif approximately 20 amino acids from the C-terminus and a (K/R)(I/L/V)(I/L/V)K motif immediately before the C-terminal residue (Kulkarni *et al*, 2017). Crystal structures of the Type A CTD show that it has an immunoglobulin-like β -sandwich fold and is often preceded by another immunoglobulin-superfamily (IgSF) domain (de Diego *et al*, 2016). Deleting the IgSF domain or disrupting the interface between the IgSF domain and the Type A CTD reduced the stability of at least one T9SS substrate in *P. gingivalis* (Sato *et al*, 2018). However, attachment of a Type A CTD alone is sufficient for transport of heterologous substrates in both *F. johnsoniae* and *P. gingivalis* (Kulkarni *et al*, 2017; Shoji *et al*, 2011). Type B CTDs are less well studied. They have an F(T/S)PNGDGXND motif approximately 80 amino acids from the C-terminus, an IFXR(W/Y)G motif approximately 55 amino acids from the C-terminus and a GX(L/F)X(L/I)X(R/K) motif

at the C-terminus (Kulkarni *et al*, 2017). Most genes encoding proteins with Type B CTDs are genetically linked to *porP/sprF*-type genes, suggesting that each Type B CTD interacts with a specific PorP family protein. In *F. johnsoniae* a fusion protein containing sfGFP and a Type B CTD was not secreted when over-expressed alone, but was secreted when the PorP family protein associated with the Type B CTD was over-expressed at the same time (Kulkarni *et al*, 2017, 2019).

Substrates are transported across the OM through the SprA translocon. CryoEM structures of SprA show that it forms a 36-strand β -barrel, the largest known single-chain β -barrel (**Figure 1-6a-d**, (Lauber *et al*, 2018)). Large insertions between strands 7 & 8 and strands 11 & 12 form cap domains that occlude the extracellular end of the β -barrel. However, there is an opening in one side of the barrel, termed the lateral gate, that allows access to the extracellular space. The structure of SprA was solved in two states, termed the PorV complex and the Plug complex. In the PorV complex PorV, the fourteen stranded OM β -barrel protein PorV, associates with the side of the SprA barrel and occludes the lateral gate (**Figure 1-6a, c**). The inter-strand extracellular loop between strands 3 and 4 of PorV is inserted into the lumen of the SprA β -barrel through the lateral gate (**Figure 1-6e**). In the Plug complex the soluble periplasmic protein Plug associates with the base of the SprA barrel and occludes the periplasmic end of the barrel (**Figure 1-6b, d**). In both states a lipoprotein peptidyl-prolyl *cis-trans* isomerase (PPI) was found on the extracellular side of the complex, on the opposite face of the barrel from the lateral gate, bound to two insertion loops in the SprA β -barrel. The role of the PPI is unclear. *F. johnsoniae* cells with the gene encoding the PPI deleted had no apparent phenotype but there are several homologous genes in *F. johnsoniae* which are probably functionally redundant and might compensate for the loss of PPI.

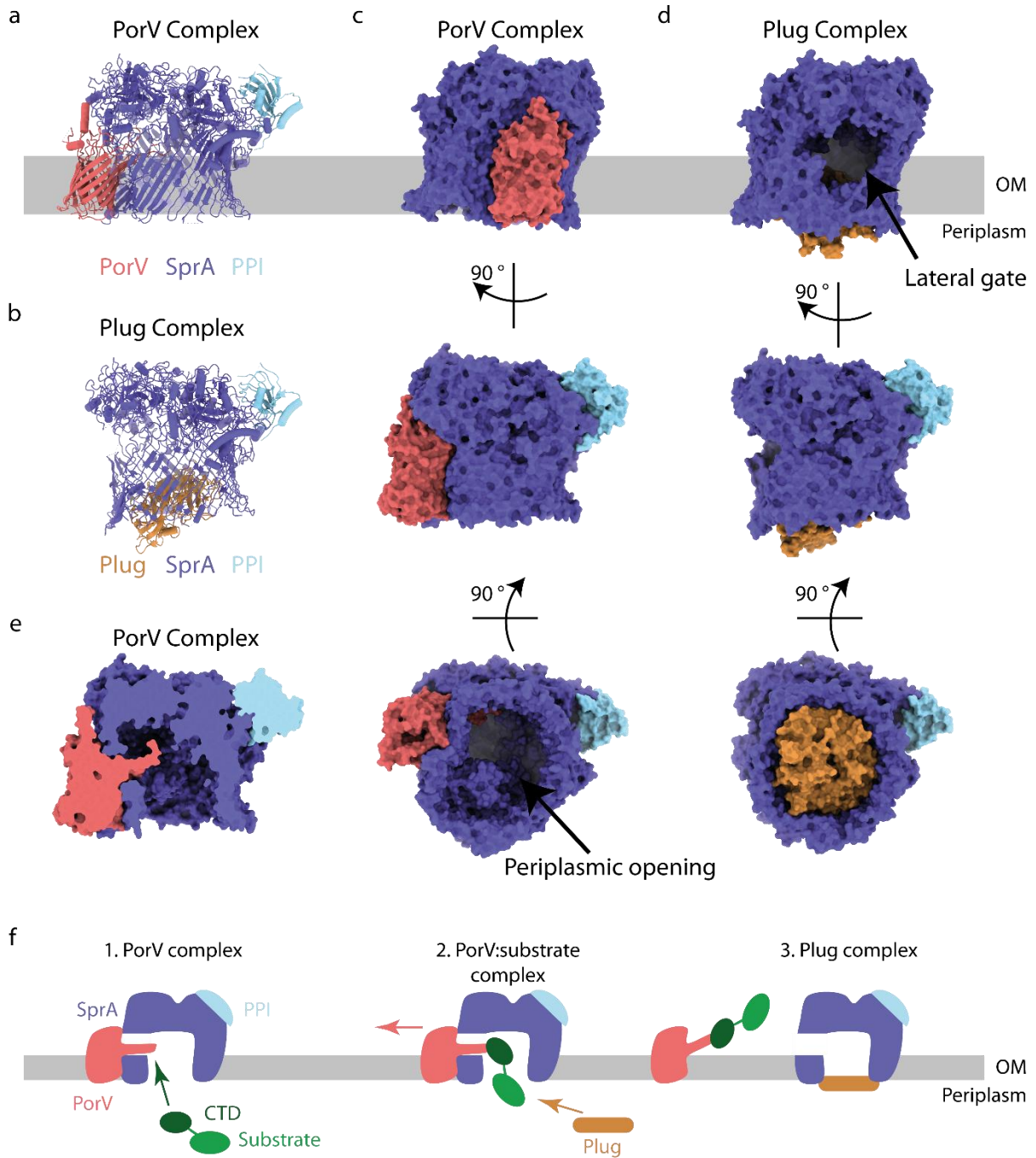


Figure 1-6 SprA is the T9SS outer membrane translocon. In all panels the proteins are coloured as follows: SprA – purple, PPI – light blue, PorV – pale red, Plug – gold, substrate protein – light green, substrate – dark green. **a** Cartoon representation of the PorV complex. **b** Cartoon representation of the Plug complex. **c** Surface representation of the PorV complex. **d** Surface representation of the Plug complex. **e** Surface representation of the PorV complex to show the PorV loop inserted through the SprA lateral gate. **f** Model for the T9SS translocon mechanism after (Lauber *et al*, 2018).

There is no energy source available at the OM to power the T9SS. The T5SS and T8SS are thought to be driven by the energetic favourability of substrate proteins folding once they have crossed the OM. This mechanism is not available to the T9SS as substrate proteins are believed to cross the OM in a folded state. Pulse-chase experiments using a radiolabelled mCherry-CTD fusion protein have shown that T9SS secretion is dependent on the proton-motive force across the IM (Hennell James *et al*, 2021). Many protein complexes across all domains of life use the energy released by proton transport from the periplasm to the cytoplasm to power processes such as ATP synthesis, flagellar rotation, and solute transport. T9SS secretion in *F. johnsoniae* is inhibited by the protonophore carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP), which transports protons across the IM independently of any protein complexes, but not by arsenate, which inhibits ATP synthesis (Hennell James *et al*, 2021).

Five proteins associated with gliding motility and T9SS in *F. johnsoniae* are predicted to be embedded in the IM. GldF and GldG have no homologues in *P. gingivalis*, so cannot be responsible for energising the T9SS and must have a role specific to gliding motility. PorY is found in both *F. johnsoniae* and *P. gingivalis* but is thought to be involved in regulation of T9SS activity or gene expression with PorX (Vincent *et al*, 2016; Yang *et al*, 2021; Jiang *et al*, 2021). GldL and GldM are found in both *F. johnsoniae* and *P. gingivalis* (named PorL and PorM in *P. gingivalis*). BTH experiments have shown that *P. gingivalis* PorL and PorM interact via their transmembrane helices (TMHs) and that the periplasmic domain of PorM interacts with the OM-associated T9SS components PorK and PorN (GldK and GldN in *F. johnsoniae*) (Vincent *et al*, 2017). PorK, PorL, PorM, and PorN are encoded by an operon and have been isolated together in a >1.2 MDa complex from *P. gingivalis* (Sato *et al*, 2010). GldL and GldM are therefore the most likely candidates for the proteins that use the PMF to energise T9SS. As they are the focus of this thesis, GldL and GldM, together with GldK and GldN, are discussed in more detail below.

After they have been transported across the OM, substrates may undergo further processing. In *P. gingivalis* PorQ, PorU, PorV, and PorZ form the attachment complex, which proteolytically removes the CTD from substrates and then either covalently links them to anionic lipopolysaccharide (A-LPS) to attach them to the OM or releases them into the environment. PorU and PorZ are both CTD-containing proteins whose CTD is not removed after secretion, whilst PorV and PorQ are outer-membrane embedded β -barrel proteins (Glew *et al*, 2012, 2017; Lasica *et al*, 2016). PorV is thought to function as a shuttle to transfer substrates from the Sov translocon (SprA in *F. johnsoniae*) to the attachment complex (Glew *et al*, 2017). PorU is the peptidase that removes the CTD from substrate proteins and can attach the new C-terminal residue of the substrate to A-LPS by a sortase-like mechanism (Gorasia *et al*, 2015). The structure of PorZ shows it has two β -propellor domains (Lasica *et al*, 2016). β -propellor domains are often involved in binding to sugar residues and it appears that the β -propellor domains of PorZ bind to an A-LPS molecule and then provide this A-LPS molecule to PorU for covalent attachment to a substrate protein (Madej *et al*, 2021; Lasica *et al*, 2016). PorQ is an OM β -barrel. In a $\Delta porU$ background a subcomplex of PorZ and PorQ was isolated, suggesting that the role of PorQ may be to anchor PorZ. *F. johnsoniae* has homologues of PorQ, PorV, PorU, and PorZ. In $\Delta porV$ *F. johnsoniae* strains proteins with Type A CTDs cannot be secreted but proteins with Type B CTDs, which have individual cognate PorP/SprF type-proteins with a similar function to PorV (see above) can be secreted (Kharade & McBride, 2015; Kulkarni *et al*, 2019). $\Delta porU$ strains are still capable of secreting the chitinase ChiA, the adhesin RemA (which has a Type A CTD) and the adhesin SprB (which has a Type B CTD). This suggests that these proteins are not attached to the OM via A-LPS and that PorU is not necessary for the removal of the CTD from ChiA. The CTDs of adhesins appear to be retained after secretion so do not require removal by PorU (Kulkarni *et al*, 2019). No protein has yet been identified in *F. johnsoniae* that is attached to the OM via A-LPS. The presence of PorQ, PorU, and PorZ in *F. johnsoniae* suggests that either such a protein exists but has yet to be

discovered, or that PorQ, PorU, and PorZ have a different role in *F. johnsoniae* from their role in *P. gingivalis*.

1.5.4 The current model of *F. johnsoniae* gliding motility

In the current model for gliding motility in *F. johnsoniae*, adhesin molecules are propelled over the cell surface on a left-handed helical path (**Figure 1-3d**; (McBride, 2019)). The adhesin proteins bind to the substrate surface, with different adhesins binding different substrate surfaces (Shrivastava *et al*, 2012). When the adhesin is bound to the substrate surface, helical motion of the adhesin relative to the cell surface leads to the cell moving in a corkscrew fashion across the substrate surface (**Figure 1-3d**, see (Nakane *et al*, 2013) for a video of this model). Roles in this model have been suggested for some of the *gld* and *spr* genes, but how they perform these roles and the roles of several more genes are still unknown.

SprB and RemA, together with homologous proteins, are the adhesin molecules in *F. johnsoniae*. SprB is a 669 kDa protein that forms 190 nm long filaments on the cell surface (Nelson *et al*, 2008; Liu *et al*, 2007) and is necessary for colony spreading by gliding motility on agar. *sprB*⁻ cells show reduced gliding motility on glass but are not completely immobile (Nelson *et al*, 2008). RemA is a 150 kDa protein that is not required for colony spreading on agar but does contribute to gliding on glass. *sprB*⁻ *remA*⁻ cells have a larger motility defect on glass than *sprB*⁻ cells. RemA contains lectin domains, so it is suggested that *F. johnsoniae* secretes polysaccharides that adhere to surfaces such as glass and that RemA binds these polysaccharides, adhering the *F. johnsoniae* cell to the surface. SprB and RemA are transported to the cell surface via the T9SS. However, they are not attached to the cell-surface by lipidation (Shrivastava *et al*, 2013). Secretion of SprB requires the OM β -barrel protein SprF (Kulkarni *et al*, 2019) and proteomics data indicate that SprB remains associated with SprF after secretion and in this manner is anchored to the OM (Gorasia *et al*, 2020). Observations of fluorescently labelled SprB show that it moves on a closed helical loop (Nakane *et al*, 2013), therefore

it must be associated with a helical structure of some sort. Whilst adhesin molecules move on a helical loop, on the surface of the cell in contact with the substrate surface there will adhesin molecules moving in both directions. Some mechanism must coordinate the binding activity of adhesins so that only adhesins moving in one direction bind to the substrate surface at any given time. Switching which set of adhesins are bound to the surface would allow the reversals of direction commonly observed for gliding cells (Lapidus & Berg, 1982). SprC and SprD are encoded by genes in an operon with *sprB* and *sprF*, so they may function in linking SprB to the helical track or controlling SprB adhesion to substrate surfaces.

GldJ is a lipoprotein anchored in the inner leaflet of the OM that is proposed to form part of the helical track in *F. johnsoniae* (Johnston *et al*, 2017). Immunogold labelling of GldJ showed an apparent helical distribution and in *F. johnsoniae* cells where the OM was removed by freeze fracture helical structures were visible in the periplasm (Braun & McBride, 2005; Katayama *et al*, 2019). These helical structures were not confirmed to be GldJ but attempts by members of the Berks laboratory to isolate GldJ found that it enters the pellet at low centrifugation speeds, suggesting that it forms part of a very large assembly (Kjær *et al*. unpublished data). Mutations in any of *gldA*, *gldB*, *gldD*, *gldF*, *gldG*, *gldH*, or *gldI* lead to a complete absence of GldJ protein in cells, but the nature of the interactions between the products of these genes and GldJ is unknown (Braun & McBride, 2005). GldJ is not thought to play a direct role in the T9SS, but the presence of GldJ is necessary for the stability of GldK, which is in turn necessary for T9SS activity (Johnston *et al*, 2017). Deletion of the 13 most C-terminal amino acids of GldJ abrogated gliding motility but did not destabilise GldK, suggesting that these 13 residues play a key role in gliding motility.

GldA, GldB, GldD, GldF, GldG, GldH, and GldI are essential for gliding motility in *F. johnsoniae* (Agarwal *et al*, 1997; Hunnicutt *et al*, 2002; Hunnicutt & McBride, 2000, 2001; McBride *et al*, 2003;

McBride & Braun, 2004). GldA, GldF, and GldG are predicted to form an ABC transporter (Hunnicuttt *et al*, 2002), the substrates of which are unknown. Substitutions that disrupted the predicted active site of GldA destabilised GldJ, suggesting that the substrates transported by GldAFG contribute to GldJ stability (Johnston *et al*, 2017). Sulphonolipids found in the OM of *F. johnsoniae* are necessary for gliding motility (Godchaux & Leadbetter, 1988; Abbanat *et al*, 1986). It has been suggested that these lipids may be the substrates for the GldAFG transporter (Hunnicuttt *et al*, 2002) but this has not to my knowledge been investigated. Some *Bacteroidetes* species capable of gliding motility lack homologues of GldA, GldF, and GldG so their role in gliding motility is specific to *F. johnsoniae* (McBride & Zhu, 2013; Zhu & McBride, 2016). No roles are known for GldB, GldD, or GldH, but they are all essential for gliding motility in *F. johnsoniae* (Hunnicuttt & McBride, 2000, 2001; McBride *et al*, 2003). Genome sequencing suggests that GldB and GldH are found in almost all *Bacteroidetes* species regardless of the presence of other gliding motility or T9SS genes (McBride & Zhu, 2013), so they may play an indirect role in gliding motility. GldI is a predicted PPI that is suggested to assist in folding of motility-associated proteins in the periplasm (McBride & Braun, 2004). *Cellulophaga algicola gldI* cells had a severe but not total defect in gliding motility (Zhu & McBride, 2016). Other species have been identified which are capable of gliding motility but lack GldI homologues (McBride & Zhu, 2013). These two results suggest that GldI is less critical than other proteins in gliding motility and may have an additional role specific to *F. johnsoniae*.

As well as energising the T9SS and so allowing adhesin proteins to be secreted, GldL and GldM are thought to energise gliding motility (Shrivastava *et al*, 2013). Like the T9SS, gliding motility in *F. johnsoniae* is dependent on the PMF across the IM (Dzink-Fox *et al*, 1997; Ridgway, 1977). Within a few seconds of the addition of the protonophore CCCP cells stop gliding and immunofluorescently-labelled SprB molecules stop moving on the cell surface. Within 10 s of CCCP being removed both SprB molecules and cells start moving again. GldL and GldM are embedded in the IM and their

orthologues in *P. gingivalis*, PorL and PorM, have been shown to interact with the OM T9SS components PorK and PorN (orthologous to GldK and GldN). The other gliding motility associated IM proteins, GldA, GldF, and GldG are not found in all gliding species of the *Bacteroidetes*. GldL and GldM are therefore the most likely candidates for energising gliding motility, as well as the T9SS (see 1.5.3). How GldL and GldM convert electrochemical potential energy from the PMF into mechanical movement of adhesins and cells was not known at the outset of my project and was the focus of my thesis research. What is known about the proteins GldL and GldM is discussed below in 1.5.5. Models for how GldL and GldM power gliding motility are discussed in the next paragraph.

When *F. johnsoniae* cells glide on glass, they sometimes spin about one pole, with the other pole of the cell detaching from the surface (Lapidus & Berg, 1982). Shrivastava and colleagues have investigated this behaviour by using anti-SprB antibodies to attach *F. johnsoniae* cells to glass slides via SprB. To reduce the number of SprB molecules on the cell surface they passed cells through a needle (Shrivastava *et al*, 2015). Passing *E. coli* cells through a needle has been used to shear their flagella (Block *et al*, 1982), which are made of thousands of non-covalently linked polypeptides. It is unclear if the same treatment would shear SprB filaments, which are thought to be a single polypeptide chain. SprB may be non-covalently linked to the cell surface by binding SprF (Gorasia *et al*, 2020), so it is possible that this interaction could be disrupted by shearing and most SprB filaments removed from the surface of the cell. When sheared *F. johnsoniae* cells were immobilised via SprB, Shrivastava *et al*. observed that cells spin continuously about one stationary pole at a constant frequency of 3 Hz (Shrivastava *et al*, 2015). They suggested that this indicated that gliding motility is driven by a rotary motor propelling baseplates below the OM to which SprB adhesins are attached. They proposed that passing cells through a needle before preparing slides sheared SprB filaments and disrupted the baseplates so that a single baseplate with one SprB filament was attached to one rotary motor at the pole of the cell. Rotation of the rotary motor drove rotation of the immobilised SprB

relative to the cell and hence rotation of the cell relative to the microscope slide (Shrivastava *et al*, 2015).

GldL and GldM are proposed to energise gliding motility, so Shrivastava and Berg proposed in a later paper that GldL and GldM might form part of this rotary motor. Shrivastava and Berg labelled GldL with a YFP fusion protein and observed that GldL-YFP foci were close to the centre of rotation of tethered cells and that foci were on average 90 nm from the path taken by SprB molecules on the surfaces of untethered cells. They therefore proposed that these GldL-YFP foci represent rotary motor complexes. However, they found that cells expressing only GldL-YFP were non-motile, but cells expressing WT GldL from the chromosome and GldL-YFP from a plasmid were motile (Shrivastava & Berg, 2020). It is therefore unclear if the GldL-YFP foci were part of functional motor complexes. They further proposed that these rotary motor complexes drive movement of adhesins on the cell surface by a rack and pinion mechanism (**Figure 1-7**, (Shrivastava & Berg, 2020)). In this model a rotary motor which contains GldL and GldM drives rotation of a toothed pinion (**Figure 1-7a-b**). This circular pinion engages with the teeth of a mobile rack. The mobile rack is mounted on a stationary track, so that rotation of the pinion pushes the mobile rack along the stationary track. The stationary track is composed in part by GldJ and is arranged in a helical pattern around the cell (**Figure 1-7c**). The mobile rack follows the helical stationary track so is pushed around the cell in a helical pattern. Adhesin molecules such as SprB are attached to the mobile rack such that they also move in a helical pattern relative to the cell body. Binding of the adhesins to a substrate surface means that motion of the adhesin relative to the cell is converted into motion of the cell relative to the substrate surface.

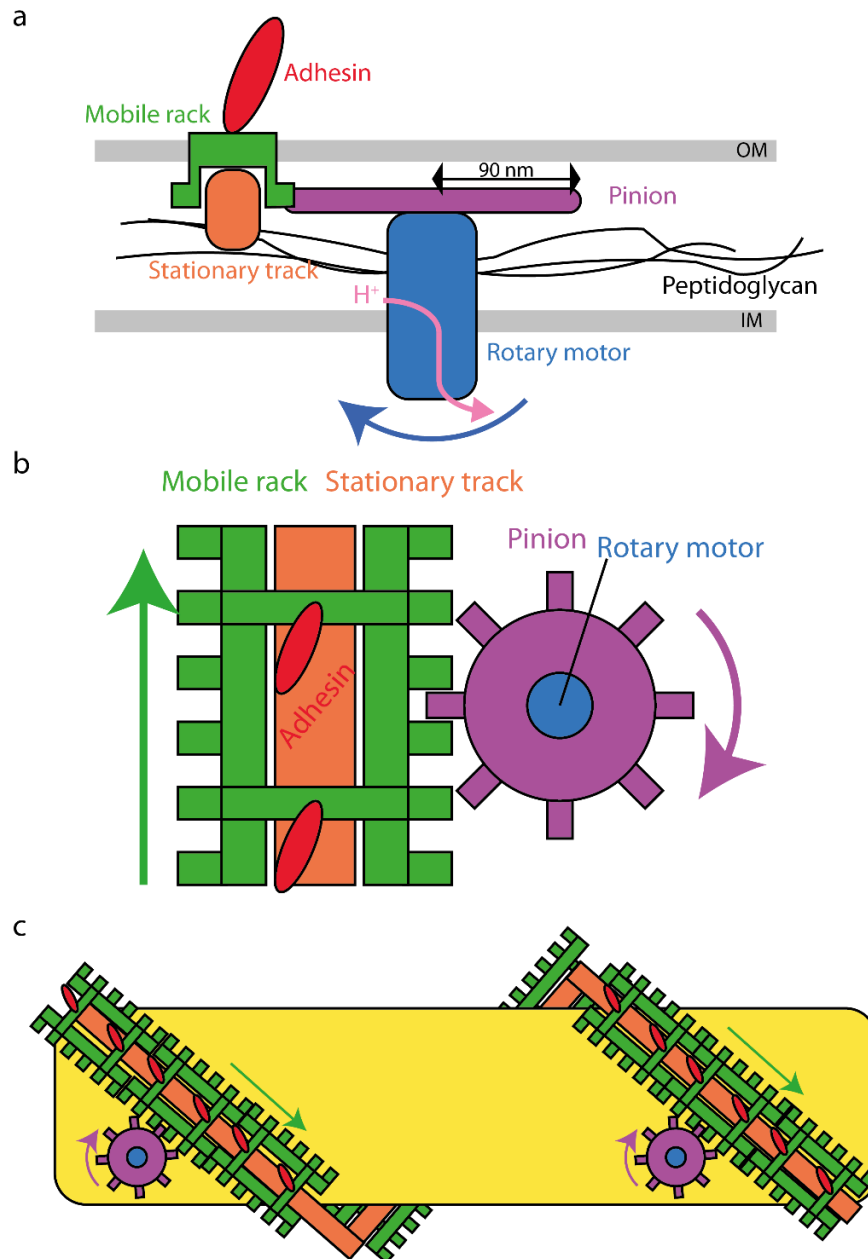


Figure 1-7 A rack and pinion model for adhesin propulsion proposed by Shrivastava and Berg. **a-b** View within the periplasm (**a**) and outside the cell (**b**) of the rack and pinion. The IM, OM, and peptidoglycan cell wall are not shown in **b**. A rotary motor (blue) is embedded within the IM and fixed in place relative to the cell body. The rotary motor rotates using energy from the PMF. Rotation of the rotary motor drives rotation of a toothed pinion (purple) of radius 80 nm. The teeth of the pinion engage with the teeth of the mobile rack (green) which can slide along the stationary track (orange). The track is kept stationary by binding to the peptidoglycan cell wall. The mobile rack can only move along the stationary track, so rotary motion of the pinion is converted into linear motion of the mobile rack. Adhesins (red) are anchored on the mobile rack so move along with it. **c** The stationary track is arranged in a helical pattern around the cell. The mobile rack and adhesins follow the helical arrangement of the stationary track, so the adhesins move in a helical pattern around the length of the cell. After (Shrivastava & Berg, 2020).

The rack and pinion model is an attractive explanation for how rotation of a rotary motor could be converted into helical motion of adhesins around the cell surface. However, several questions about how this model explains gliding motility remain. The GldL-YFP foci were on average 90 nm from the path followed by SprB, suggesting that if these foci were indeed rotary motors the pinion must have a radius of 90 nm to be mounted on the rotary motor and engaged with the mobile rack at the same time. No structures of this size have been reported from *F. johnsoniae*. Gorasia and colleagues have reported that in *P. gingivalis* PorK and PorN (orthologous to GldK and GldN in *F. johnsoniae*) form rings 25 nm in diameter (Gorasia *et al*, 2016, 2020). However, these structures have not been reported in *F. johnsoniae*. If these rings do exist in *F. johnsoniae*, additional components would be needed to give the ~90 nm radius required by the rack and pinion model. It is unknown what components contribute to the stationary track and the mobile rack, but several proteins of unknown function are known to be involved with gliding motility and could fulfil these roles. Cells expressing only GldL-YFP were non-motile, suggesting that GldL-YFP proteins are non-functional (Shrivastava & Berg, 2020). Shrivastava and Berg instead used cells expressing both WT GldL and GldL-YFP. It is possible therefore that the GldL-YFP foci they observed do not represent functional rotary motors and are not functionally relevant. It may be the case that the actual rotary motors, if they exist, were located closer than 90 nm to the SprB paths and correspondingly have a smaller pinion attached. It is also possible that there are no rotary motors involved in *F. johnsoniae* gliding motility. The mobile rack could instead be propelled along the stationary track by a back-and-forth motor that binds to the mobile rack, propels the mobile rack forward by a short distance, releases the mobile rack, resets to its original position, and then binds to a new section of the mobile rack (**Figure 1-8**). In this model it would not be necessary to have a continuous mobile rack. So long as they remained confined on the stationary track, units of the mobile rack could be physically separate from one another. This model could also explain the Shrivastava and colleagues' observation that cells tethered by SprB spin about

their poles (Shrivastava *et al*, 2015). If passing cells through a needle led to disruption of the internal stationary track small circular loops of track could form. Propulsion of adhesins around this circular track would cause the cell to spin about that point. The explanation is supported by the fact that fluorescently labelled SprB molecules sometimes move in circles around the poles of cells (Kjær & Miller, unpublished), these circles could represent circular loops of the stationary track. Considerably more research is needed to understand the roles of all the proteins involved in gliding motility before a model of adhesin propulsion can be confidently proposed.

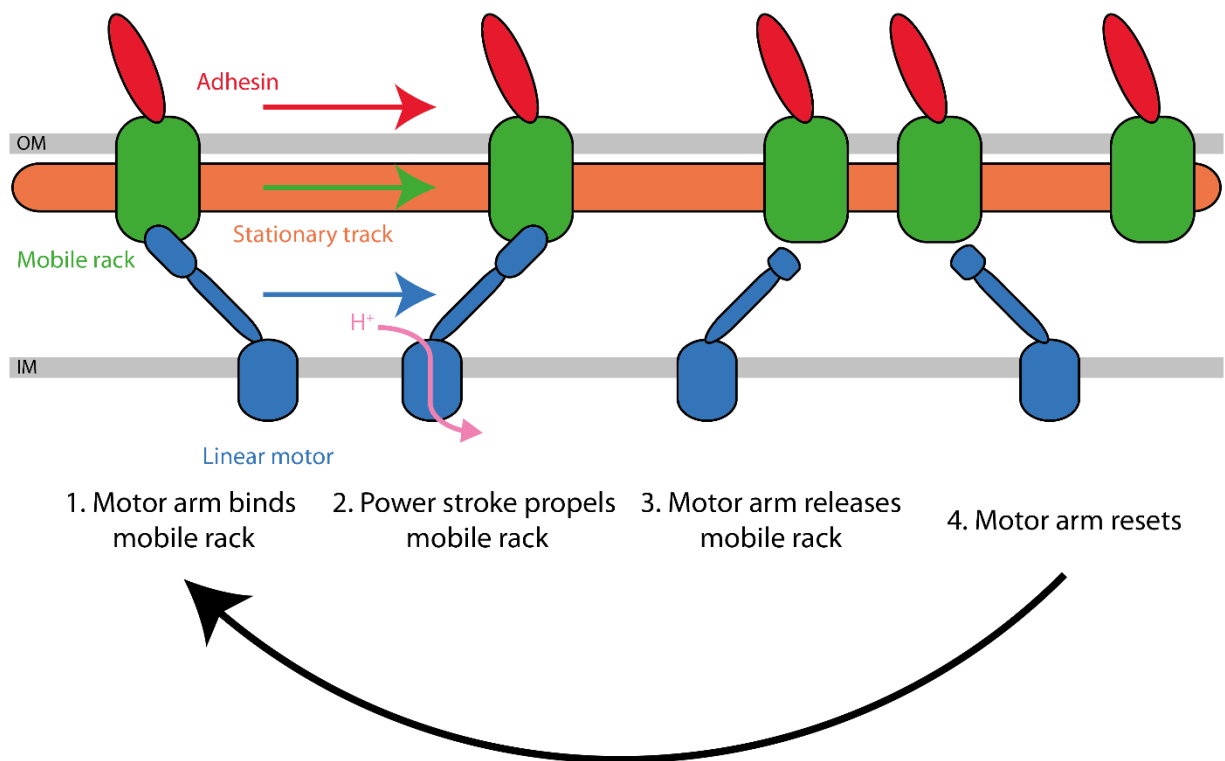


Figure 1-8 A model for adhesin propulsion involving a linear motor. In stage 1 the arm of the linear motor binds to the mobile rack, to which the extracellular adhesin is attached. In stage 2 the linear motor arm swings forward, using energy from the PMF. This propels the mobile rack and attached adhesin along the stationary track. In stage 3 the motor arm releases from the mobile rack. In stage 4 the motor arm resets to its original position and is ready to bind a new mobile rack unit. In this model the mobile rack may be one continuous structure or could be composed of many separate structures, as shown here.

1.5.5 Current knowledge about GldL and GldM

GldL and GldM are integral inner membrane proteins which are proposed to form the complex that energises the T9SS and gliding motility in *Bacteroidetes* (Shrivastava *et al*, 2013; Vincent *et al*, 2017). In *F. johnsoniae* GldL and GldM are encoded in an operon with GldK and GldN. Northern blot and RT-PCR evidence suggest that both *gldKLMN* and *gldLMN* transcripts are produced from this locus (Braun *et al*, 2005; Shrivastava *et al*, 2013). This operon structure is present in other species, usually with a PorP/SprF-type protein encoded upstream of *gldL*, although it has only been confirmed that the *porP/sprF*-like gene is co-transcribed with the other genes in *P. gingivalis* (Vincent *et al*, 2017). The operonic structure of the genes suggests that their protein products function together, a hypothesis supported by the isolation of a >1.2 MDa complex containing PorK, PorL, PorM, and PorN from *P. gingivalis* (orthologous to GldK, GldL, GldM, and GldN in *F. johnsoniae*, see **Table 1-1**) (Sato *et al*, 2010). Interactions between PorL and PorM have also been detected using BTH experiments (Vincent *et al*, 2017). Interactions between PorK, PorM, PorN, and PorP have also been detected using BTH and immunoprecipitation (Vincent *et al*, 2017). A separate study isolated > 4.5 MDa ring-shaped complexes from *P. gingivalis* 50 nm in diameter with 32-36-fold symmetry that contain PorK and PorN but not PorL and PorM (Gorasia *et al*, 2016). These rings have also been observed *in situ* using cryo-electron tomography (Gorasia *et al*, 2020), so it is unclear if the >1.2 MDa complex containing PorK, PorL, PorM, and PorN was an artefact of the isolation procedure or if it is a separate functional complex.

Secondary structure prediction programs predict that *F. johnsoniae* GldL has two N-terminal TMHs separated by a periplasmic loop and a c-terminal cytoplasmic domain (**Figure 1-9**). This topology has been confirmed for *P. gingivalis* PorL using cysteine accessibility experiments with recombinantly produced proteins in *E. coli* (Vincent *et al*, 2017). The cytoplasmic domain is predicted to have both disordered regions and a long coiled coil region (**Figure 1-9a**), which may facilitate interactions

between subunits within a GldLM complex or between subunits in different GldLM complexes. Vincent and colleagues analysed a recombinant construct corresponding to the cytoplasmic domain of *P. gingivalis* PorL. On the basis of size-exclusion chromatography and multi-angle light scattering (SEC-MALS) this construct was reported to form a mixture of monomers and trimers in solution (Vincent *et al*, 2017). However, the calculated masses for the monomer and trimer were 113 kDa and 44 kDa, respectively, differing substantially from the predicted masses of 87 kDa and 29 kDa, respectively. This suggests that the assignment of *P. gingivalis* PorL being a trimer is incorrect. Purification of the cytoplasmic domain of *F. johnsoniae* GldL has been reported but no structural information is yet available (Trinh *et al*, 2021). My own SEC-MALS analysis of the *P. gingivalis* PorLM complex and structures of *F. johnsoniae* GldLM and *P. gingivalis* PorLM are discussed in Chapters 3, 4, and 6. PorX, a CheY-like response regulator that co-activates transcription of T9SS-related genes with the transcription factor SigP (Jiang *et al*, 2021), has been shown in BTH experiments to bind the C-terminus of *P. gingivalis* PorL, but the role of this interaction is unclear (Vincent *et al*, 2016). In *F. johnsoniae*, deletion of PorX has no apparent effect on either T9SS or gliding motility activity (Wheeler, Lauber *et al.*, unpublished data). The second TMH of GldL has a universally conserved glutamate residue (Glu49 in *F. johnsoniae*, **Figure 1-9a**). The T9SS and gliding motility are both dependent on the PMF (Hennell James *et al*, 2021; Dzink-Fox *et al*, 1997), so it is proposed that this glutamate residue is involved in conducting protons across the IM to convert the electrochemical potential energy of the PMF into mechanical work (Shrivastava *et al*, 2013).

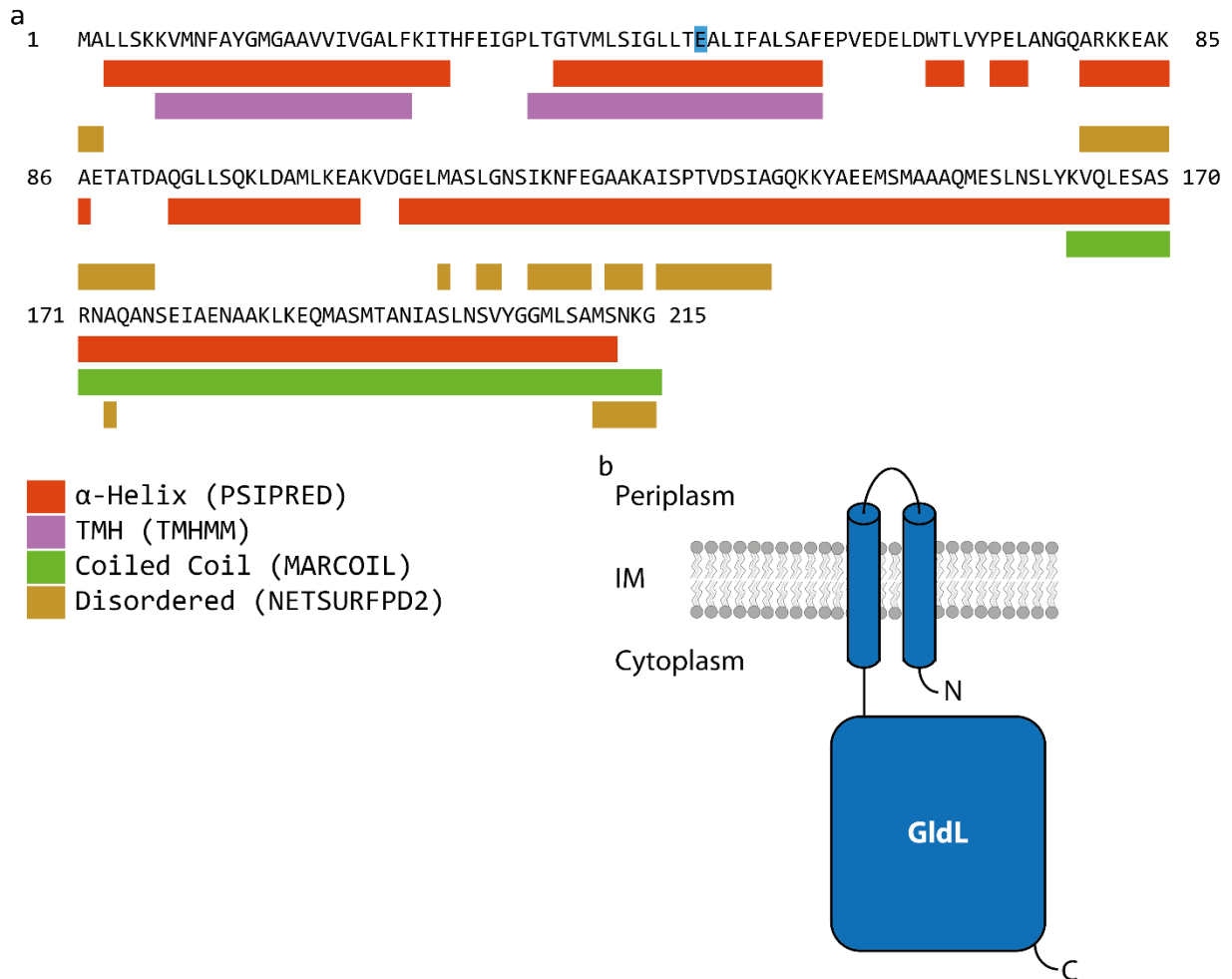


Figure 1-9 Predicted topology and secondary structure of *F. johnsoniae* GldL. **a** Predicted secondary structure elements of *F. johnsoniae* GldL. The universally conserved Glu49 is highlighted in blue. α -helices (red) were predicted with PSIPRED (Jones, 1999). TMHs (pink) were predicted with TMHMM (Krogh *et al*, 2001). Coiled-coil regions (green) were predicted with MARCOIL (Delorenzi & Speed, 2002). Disordered regions (brown) were predicted with NETSURFPD2 (Klausen *et al*, 2019). **b** Predicted topology of *F. johnsoniae* GldL.

Secondary structure prediction programs predict that GldM has one amino-terminal TMH and a large carboxy-terminal periplasmic region containing α -helices and β -strands (**Figure 1-10a-c**). This topology has been confirmed for *P. gingivalis* PorM using cysteine accessibility experiments with recombinantly produced proteins in *E. coli* (Vincent *et al*, 2017). The same study also used BTH to confirm that PorL and PorM interact via their TMHs when recombinantly expressed in *E. coli*. A crystal structure of the periplasmic region of *F. johnsoniae* GldM (*FjoGldM_{peri}*) shows that it forms an

elongated dimer with four domains (**Figure 1-10c**, (Leone *et al*, 2018)), consistent with SEC-MALS data reported for the periplasmic domain of *P. gingivalis* PorM (*PgiPorM_{peri}*) (Vincent *et al*, 2017). The most amino-terminal domain (D1) is entirely helical and the two copies in the dimer are arranged parallel to each other. The middle D2 and D3 domains are composed of β -strands and exchange strands between the two protomers. The most carboxy terminal domain (D4) has a mixture of α -helices and β -strands. The total length of the dimeric periplasmic region is 180 Å, sufficient to span the periplasm and interact with both IM and OM components of the T9SS and gliding motility machineries. BTH experiments suggest that the D4 domain of *P. gingivalis* PorM interacts with PorN, which is localised to the OM by association with the lipoprotein PorK (Leone *et al*, 2018). Sato *et al.* have solved a crystal structure of the periplasmic region of *P. gingivalis* PorM (*PgiPorM_{peri}*) (Sato *et al*, 2020). They observed kinks between the D1 and D2 domains and between the D2 and D3 domains (**Figure 1-10d**). Sato *et al.* suggested that Arg224 in the D2 domain of one copy of *PgiPorM_{peri}* forms hydrogen bonding interactions with the main chain carbonyls of D219 in D1 of both copies of *PgiPorM_{peri}*, appearing to stabilise the bend between the domains (**Figure 1-10d-e**). An R224D substitution reduced T9SS secretion activity as measured by colony pigmentation on blood agar, a measure of gingipain secretion (Sato *et al*, 2020). This suggests that the interaction between Arg224 and the two copies of Asp219 and therefore the bend between D1 and D2 are important to T9SS function, although it is possible that Arg224 plays a different role in T9SS function not apparent from the crystal structure. More research is needed to determine if the differences between the straight arrangement of D1-D4 domains in the *FjoGldM_{peri}* crystal structure and the kinked arrangement of D1-D4 domains in the *PgiPorM_{peri}* structure represent real differences between motile *F. johnsoniae* and non-motile *P. gingivalis* or if they represent different conformational states of GldM and PorM captured by different crystallisation conditions.

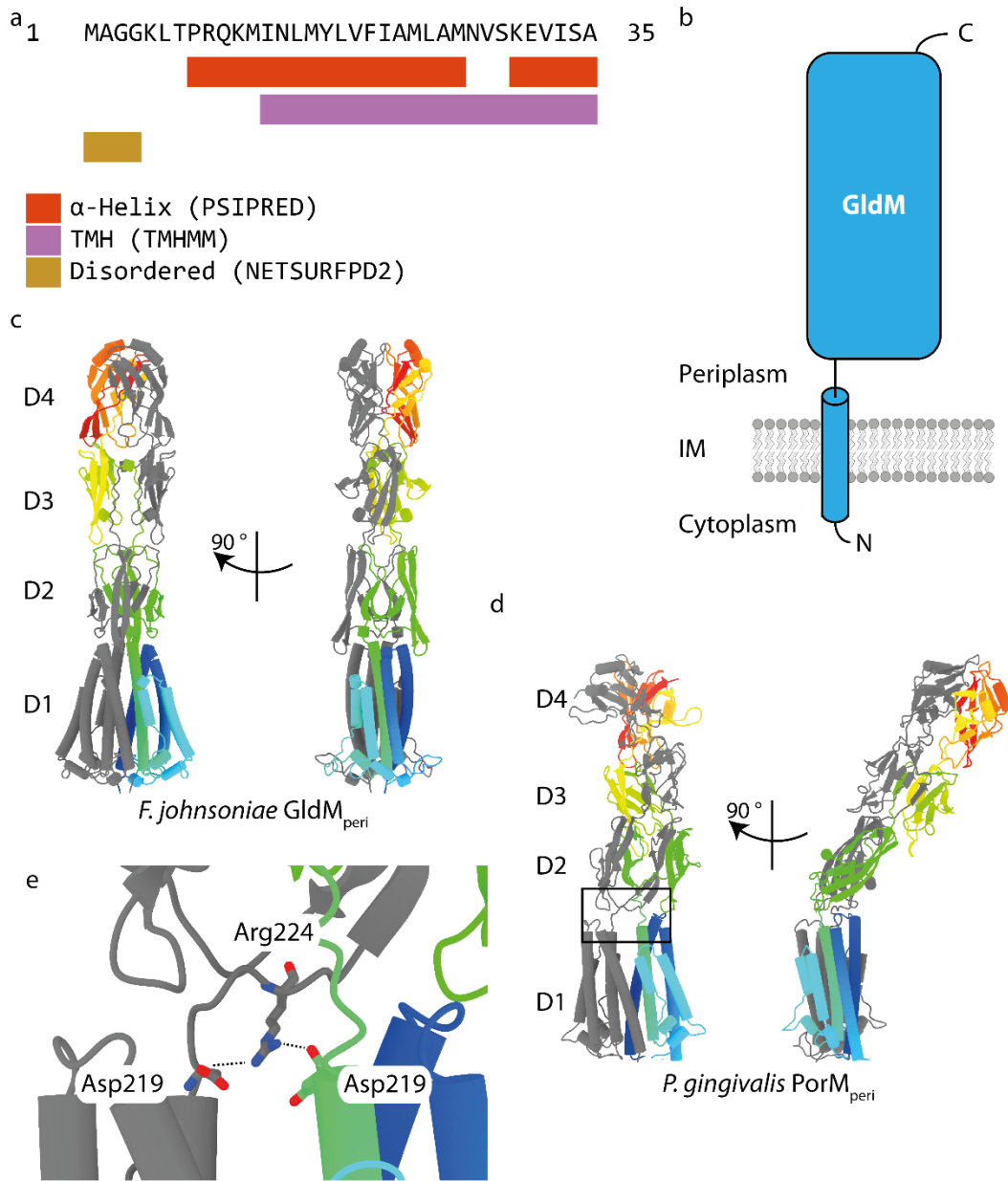


Figure 1-10 Topology and structure of *F. johnsoniae* GldM. **a** Predicted secondary structure elements of *F. johnsoniae* GldM. Only the residues not included in the crystal structure shown in **c** are shown. α -helices (red) were predicted with PSIPRED (Jones, 1999). TMHs (pink) were predicted with TMHMM (Krogh *et al*, 2001). Disordered regions (brown) were predicted with NETSURFPD2 nets (Klausen *et al*, 2019). **b** Predicted topology of *F. johnsoniae* GldM. **c** Crystal structure of *F. johnsoniae* GldM periplasmic region (PDB 6EY4, (Leone *et al*, 2018)). One chain is coloured grey and the other is coloured as a rainbow from the N-terminus (blue) to the C-terminus (red). **d** Crystal structure of *P. gingivalis* PorM periplasmic region (PDB 7CMG, (Sato *et al*, 2020)). One chain is coloured grey and the other is coloured as a rainbow from the N-terminus (blue) to the C-terminus (red). **e** Expansion of the boxed region in **d** showing Arg224 from one chain and Asp219 from both chains. Potential hydrogen bonds between Arg224 and the mainchain carbonyl groups of Asp219 are shown as dotted lines. Note that the potential hydrogen bonds were inferred by Sato *et al*. based on the proximity of the residues rather than the geometry of their arrangement due to the limited resolution of the structure (3.7 Å).

1.6 Electron microscopy

The highest resolution that can be achieved in an optical system is limited by the wavelength of the radiation used. Although modern visible light microscopy techniques can circumvent the theoretical Abbe diffraction limit, observation of biological macromolecules at atomic or near-atomic resolution requires shorter wavelength radiation than visible light, either X-rays or electrons. X-rays are usually used in diffraction experiments with crystals formed of many thousands of copies of a biological macromolecule, whilst electrons can be used to image either crystals by diffraction or individual biological macromolecules. In this study, electron microscopy was used to image individual protein complexes.

Proteins are composed of low atomic number elements which do not scatter electrons well, so give low signal in electron microscopy experiments. In negative stain electron microscopy, the signal is improved by embedding proteins in heavy atom stains such as uranyl acetate. The name negative stain refers to the fact that the protein is not imaged directly but instead the heavy atom stain is imaged and proteins are visible as absences of stain. Because the stain is imaged rather than the protein, only the shape of the surface of the protein can be determined and resolution is limited to ~ 20 Å. In cryo-electron microscopy experiments (cryoEM) proteins or biological structures suspended in vitreous ice are imaged directly. Using vitreous ice avoids diffraction patterns caused by crystalline ice. The electron dose must be kept low to avoid sample damage. This results in very low signal to noise ratios, so many protein particles must be imaged and averaged together. Computational methods are used to align 2D projection images from similar views and average them to remove noise. Further computation determines the relative 3D orientations of these views and reconstructs 3D maps of cryoEM density. The development of direct electron detectors over the last decade and accompanying developments in software have made it possible to determine the 3D structures of optimised protein samples relatively routinely at resolutions below 4 Å and in some cases at atomic

resolutions (Yip *et al*, 2020; Nakane *et al*, 2020; Callaway, 2015). These recent improvements in cryoEM techniques have led to improved understanding of many areas of biology in recent years, including various bacterial protein secretion systems (Lauber *et al*, 2018; Kuhlen *et al*, 2018; Bunduc *et al*, 2021).

CryoEM has lower sample requirements than other structural biology techniques. This is beneficial when studying membrane proteins which can be difficult to purify in large amounts. Membrane proteins are usually purified after solubilisation in detergents which can make them difficult to crystallise for crystallography experiments as the detergent micelle prevents protein-protein contacts forming. I therefore selected cryoEM as my main technique to structurally characterise the GldLM complex in this project.

1.7 Project aims

If the GldLM complex does indeed energise the T9SS and gliding motility by utilising the PMF, the interaction between GldL and GldM via their TMHs must be important to the function of the complex. If the GldLM complex does conduct protons across the inner membrane, protons must flow through the transmembrane region. Very little research has been carried out on this region of the GldLM complex, despite it being key to the proposed function of the complex. Therefore, in this thesis I set out to characterise the GldLM complex as a whole and the transmembrane region in particular. I did this through the following structural and biochemical approaches:

- Recombinant expression and purification of full-length GldLM complexes and their characterisation using cryoEM (Chapter 3)
- Recombinant expression and purification of periplasmically-truncated *F. johnsoniae* GldLM complexes and their characterisation using cryoEM (Chapter 4)
- Generation of *F. johnsoniae* strains with point mutations in *gldL* and *gldM* and their characterisation using motility and secretion assays (Chapter 5)
- Recombinant expression and purification of GldLM constructs from diverse species and their characterisation using cryoEM (Chapter 6)

2 Methods

Except where otherwise noted chemicals were purchased from Merck (previously Sigma-Aldrich).

2.1 Culturing of bacteria

Media used for growing *E. coli* and *F. johnsoniae* are detailed in Table 2-1. Agar plates were made by adding 15 g l⁻¹ agar to the relevant medium. Where appropriate, media were supplemented with antibiotics at the concentrations detailed in Table 2-2. Strains of *E. coli* and *F. johnsoniae* used in this study are listed in Table 2-3 and Table 2-4, respectively. *E. coli* cells were routinely grown in LB at 37 °C with shaking. *F. johnsoniae* cells were routinely grown in CYE at 30 °C with shaking.

Table 2-1 Composition of the growth media used in this study.

Medium	Components	Concentration	Reference
Lysogeny Broth (LB)	Tryptone Yeast Extract Sodium Chloride	10 g l ⁻¹ 5 g l ⁻¹ 10 g l ⁻¹	(Bertani, 1951)
Terrific Broth (TB)	Tryptone Yeast Extract Glycerol Potassium phosphate, monobasic Potassium phosphate, dibasic	12 g l ⁻¹ 24 g l ⁻¹ 0.4% 23.13 g l ⁻¹ 123.41 g l ⁻¹	(Tartof & Hobbs, 1987)
2xYT	Tryptone Yeast Extract Sodium Chloride	16 g l ⁻¹ 10 g l ⁻¹ 5 g l ⁻¹	(Miller, 1972)
CYE	Casitone Yeast Extract MgSO ₄ Tris-HCl pH 7.5	10 g l ⁻¹ 5 g l ⁻¹ 8 mM 10 mM	(McBride & Kempf, 1996)
PY2	Peptone Yeast Extract pH 7.3	2 g l ⁻¹ 0.5 g l ⁻¹	(Agarwal <i>et al</i> , 1997)
Motility Medium (MM)	Casitone Yeast Extract Tris-HCl pH 7.5	3.3 g l ⁻¹ 1.7 g l ⁻¹ 3.3 mM	(Liu <i>et al</i> , 2007)
MYA	MgSO ₄ FeSO ₄ EDTA Potassium phosphate pH 7.25 Yeast Extract	0.5 mM 0.05 mM 0.04 mM 20 mM 0.1 g l ⁻¹	(McBride & Braun, 2004)

Table 2-2 Antibiotics used in this study.

Antibiotic	Concentration used
Ampicillin	100 $\mu\text{g ml}^{-1}$
Chloramphenicol	25 $\mu\text{g ml}^{-1}$
Erythromycin	100 $\mu\text{g ml}^{-1}$
Kanamycin	30 $\mu\text{g ml}^{-1}$ (cultures in LB and 2xYT media) 50 $\mu\text{g ml}^{-1}$ (cultures in TB medium)
Spectinomycin	50 $\mu\text{g ml}^{-1}$

Table 2-3 *E. coli* strains used in this study.

Strain	Parental strain	Genotype	Description	Resistance	Source
TOP10	MC1061	<i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74 nupG</i> <i>recA1 araD139</i> Δ (<i>ara-leu</i>)7697 <i>galE15 galK16 rpsL(StrR)</i> <i>endA1</i> λ^-	Used for general cloning and storing plasmids as glycerol stocks.	Streptomycin	Invitrogen
DH5α Turbo	DH5 α	F' <i>proA⁺B⁺</i> <i>lacI^q</i> Δ <i>lacZ</i> M15/ <i>fhuA2</i> Δ (<i>lac-proAB</i>) <i>glnV galK16 galE15 R(zgb-210::Tn10)</i> Tet ^S <i>endA1</i> <i>thi-1</i> Δ (<i>hsdS-mcrB</i>)5	Used for general cloning and storing plasmids as glycerol stocks.	-	New England Biolabs
BL21 (DE3)	B	<i>ompT lon hsdSB (rB⁻ mB⁻) gal dcm</i> (DE3)	Deficient in OmpT and Lon proteases. T7 RNA polymerase (T7RNAP) is encoded by the DE3 lysogen. Used for protein expression.	-	Invitrogen
BL21 Star (DE3)	BL21 (DE3)	<i>ompT lon hsdSB (rB⁻ mB⁻) gal dcm rne131</i> (DE3)	RNase E is non-functional, leading to extended mRNA transcript half-life. Used for protein expression	-	Invitrogen
Rosetta 2 (DE3)	BL21 (DE3)	<i>ompT lon hsdSB (rB⁻ mB⁻) gal dcm rne131</i> (DE3) pRARE2	pRARE2 confers expression of 7 tRNAs for rare codons	Cam	Novagen
C41 (DE3)	BL21 (DE3)	<i>ompT lon hsdSB (rB⁻ mB⁻) gal dcm rne131</i> (DE3) <i>lacUV5[*]</i>	Mutations in the <i>lacUV5</i> promoter controlling T7RNAP expression lead to reduced T7RNAP expression levels. Can improve yields of overexpressed toxic and membrane proteins.	-	(Miroux & Walker, 1996)
Mt56 (DE3)	BL21 (DE3)	<i>ompT lon hsdSB (rB⁻ mB⁻) gal</i>	Variant T7RNAP has reduced affinity	-	(Baumgarten <i>et al</i> , 2017)

		<i>dcm rne131</i> (DE3) <i>tzrna_{mt56}</i>	for the T7 promoter. Can improve yields of overexpressed toxic and membrane proteins.		
S17-1	294	<i>pro res⁻ hsdR17 (rK⁻ mK⁺) recA⁻ RP4-2-Tc::Mu-Km::Tn7 Tp^r</i>	Used to introduce suicide and expression plasmids into <i>F. johnsoniae</i> by biparental mating.	Trimethoprim, Streptomycin	(Simon <i>et al</i> , 1983)

Table 2-4 *F. johnsoniae* strains used in this study.

Strain	Genotype	Source
UW101		(McBride and Braun, 2004)
FI_082	UW101 $\Delta gldL$	(Hennell James <i>et al</i> , 2021)
Rhj_006	UW101 $\Delta gldM$	This study
Rhj_011	UW101 $gldL_{H30A}$	This study
Rhj_012	UW101 $gldM_{Q10A}$	This study
Rhj_013	UW101 $gldL_{S43A}$	This study
Rhj_014	UW101 $gldM_{Y17A}$	This study
Rhj_015	UW101 $gldM_{Y17F}$	This study
Rhj_016	UW101 $gldM_{N14A}$	This study
Rhj_017	UW101 $gldL_{N10A}$	This study
Rhj_018	UW101 $gldL_{Y13A}$	This study
Rhj_019	UW101 $gldL_{K27A}$	This study
Rhj_020	UW101 $gldL_{T48A}$	This study
Rhj_021	UW101 $gldL_{E49A}$	This study
Rhj_022	UW101 $gldL_{E49Q}$	This study
Rhj_023	UW101 $gldL_{S56A}$	This study
Rhj_024	UW101 $gldL_{Y13F}$	This study
Rhj_025	UW101 $gldL_{E49D}$	This study
Rhj_028	UW101 $gldM_{N27A}$	This study
Rhj_029	UW101 $gldM_{R9E}$	This study
Rhj_030	UW101 $gldM_{K11A}$	This study
Rhj_032	UW101 $gldM_{R9K}$	This study
Rhj_035	UW101 $gldM_{R9A}$	This study
Rhj_057	UW101 $gldL_{\Delta 64-215}$	This study
Rhj_058	UW101 $gldL_{\Delta 78-215}$	This study
Rhj_059	UW101 $gldL_{\Delta 131-215}$	This study
Rhj_060	UW101 $gldL_{\Delta 145-215}$	This study
Rhj_061	UW101 $gldL_{\Delta 162-215}$	This study
Rhj_062	UW101 $gldL_{\Delta 206-215}$	This study
Ak_73	UW101 $\Delta porV$ <i>halotag::sprB</i>	(Hennell James <i>et al</i> , 2021)
Ak_203	AK_73 $\Delta gldM$	(Hennell James <i>et al</i> , 2021)
Ak_205	AK_73 $\Delta gldL$	(Hennell James <i>et al</i> , 2021)
Ak_196	AK_73 $gldL_{Y13A}$	(Hennell James <i>et al</i> , 2021)
Ak_197	AK_73 $gldL_{H30A}$	(Hennell James <i>et al</i> , 2021)
Ak_198	AK_73 $gldL_{K27A}$	(Hennell James <i>et al</i> , 2021)
Ak_199	AK_73 $gldM_{Y17F}$	(Hennell James <i>et al</i> , 2021)
Ak_289	Ak_73 $gldM_{R9K}$	(Hennell James <i>et al</i> , 2021)

2.2 Bioinformatics analysis

Secondary structure elements and disordered regions of proteins were predicted using the Quick2D suite (Gabler *et al*, 2020; Zimmermann *et al*, 2018).

To ensure that the GldLM homologues screened for expression sampled a broad range of the available sequence diversity, phylogenetic trees were constructed. For the tree of GldL sequences, sequences were obtained by BLAST searches against the UniRef90 database using the *F. johnsoniae* GldL, *P. gingivalis* PorL and *S. wenxiniae* GldL sequences as queries (Altschul *et al*, 1990; Suzek *et al*, 2015). Duplicate sequences and sequences from organisms outside the *Bacteroidetes* phylum were removed. So that the phylogenetic tree could be coloured by order for ease of understanding, clusters of sequences where the lowest common taxon was higher than order were removed. The GldL sequence homologues selected for construct design from *Apibacter mensalis* DSM 100903, *Capnocytophaga canimorsus* strain Cc5, *Draconibacterium orientale* DSM 25947, *Flavobacterium johnsoniae* UW101, *Flavobacterium psychrophilum* DSM 21280, *Flavobacteriales* bacterium isolate NAT55 387147, *Porphyromonas gingivalis* ATCC 33277, *Pseudopedobacter saltans* DSM 12145, *Riemerella anatipestifer* DSM 15868, *Schleiferia thermophila* str Yellowstone, *Sphingobacterium wenxiniae* DSM 22789, *Thermonema rossianum* DSM 10300, an uncultured *Sphingobacteriia* bacterium sampled from the North Atlantic Ocean and *Zobellia galactanivorans* DSM 12808 were added to the set of sequences. The sequences were then aligned using Clustal Omega (Madeira *et al*, 2019) and phylogenetic trees were inferred from the alignments using the Maximum Likelihood function of program MEGA X with a JTT matrix-based model with default settings (Kumar *et al*, 2018; Jones *et al*, 1992). A phylogenetic tree of GldM sequences was generated using the same approach.

Sequence logos were generated to assess the conservation of different residues of GldL and GldM. These represent the different residues at each position in a sequence alignment as single letter codes, with the size of each letter representing how well-conserved it is in the alignment. The same

sets of sequences obtained to produce phylogenetic trees above were used, except that the sequences selected for construct design were not included. These sequences were aligned using Clustal Omega (Madeira *et al*, 2019) then WebLogo 3.7.4 (Schneider & Stephens, 1990; Crooks *et al*, 2004) was used to generate sequence logos for the first and second helices of GldL. The *FjoGldLM'* structure (see Chapter 4) was used to decide helix boundaries. A sequence logo for the first GldM helix was generated using the same approach.

2.3 Molecular Biology

Plasmids used for recombinant expression of proteins or for the generation of mutant strains of *F. johnsoniae* are listed with descriptions of the cloning strategy in Table 2-5 and Table 2-6, respectively.

Table 2-5 Plasmids used for recombinant expression of proteins in this study.

Plasmid	Description	Resistance	Strategy/Source
pET28a (+)	Expression of one <i>orf</i> under the control of a T7lac promoter with sequence encoding an N-terminal His ₆ tag and thrombin cleavage site; <i>ori</i> ColE1	Kanamycin	Novagen
pWALDO_sfGFPd	pET28(a+)-derived expression vector fusing a TEV cleavage site, superfolder GFP, and a His ₈ tag to the C-terminus of the expressed protein; template for sfGFP	Kanamycin	(Drew <i>et al</i> , 2006)
pCDFDuet-1	Co-expression of two <i>orfs</i> under the control of T7lac promoters; <i>ori</i> CloDF13	Spectinomycin	Novagen
pSufIss-GFPHis	pCDFDuet-1 coding for a SufI signal sequence fused to GFP-His in the first multiple cloning site,	Spectinomycin	(Huang <i>et al</i> , 2017)
pETDuet-1	Co-expression of two <i>orfs</i> under the control of T7lac promoters; <i>ori</i> ColE1	Ampicillin	Novagen
pT12_SpaPQR^{3XFLAG}	Encodes <i>Salmonella enterica</i> serovar Typhimurium <i>SpaPQR</i> operon with a C-terminal ^{3XFLAG} tag on SpaR under the control of the E. coli <i>rhaB</i> promoter; <i>ori</i> cloDF13,	Kanamycin	(Dietsche <i>et al</i> , 2016)
pMA-RQ	Cloning vector for synthesised DNA; <i>ori</i> ColE1	Ampicillin	Invitrogen GeneArt
pMA-RQ_SthPKLMN	Synthesised <i>S. thermophila</i> <i>porPgldKLMN</i> operon in cloning vector, <i>ori</i> ColE1	Ampicillin	Synthesised by Invitrogen GeneArt
pTAH	Cloning vector for synthesised DNA; <i>ori</i> ColE1	Ampicillin	Twist Bioscience
pTAH_NAT_GldLM	Synthesised <i>Flavobacteriales</i> isolate NAT55 387147 <i>gldLM</i> operon in cloning vector, <i>ori</i> ColE1	Ampicillin	Synthesised by Twist Bioscience
pTAH_UEG_GldKLMN	Synthesised GenBank accession FQ032818 (UEG) <i>gldKLMN</i> operon in cloning vector, <i>ori</i> ColE1	Ampicillin	Synthesised by Twist Bioscience
pRHJ001	pSufIss-GFPHis with the GFPHis sequence replaced by a sequence encoding TEV cleavage site, superfolder GFP, and His ₈ tag	Spectinomycin	Restriction cloning of a BamHI and HindIII fragment of pSufIss-GFPHis into pWALDO-sfGFPd
pRHJ002	pRHJ001 with a sequence encoding a C-terminal TwinStrep tag inserted into the using the second multiple cloning site	Spectinomycin	KLD mutagenesis of pRHJ001 using primers RHJ025 and RHJ026
pRHJ003	pRHJ003 with the NcoI site in the sfGFP sequence removed	Spectinomycin	QuikChange mutagenesis of pRHJ002 using primers RHJ046 and RHJ047

pRHJ004	pETDuet-1 with a sequence encoding a C-terminal TEV cleavage site, superfolder GFP and His8 tag inserted into the first multiple cloning site and a sequence encoding a C-terminal TwinStrep tag fused to the second multiple cloning site	Ampicillin	Restriction cloning of an NcoI and XhoI fragment of pRHJ003 into pETDuet-1
pRHJ005	pRHJ004 with <i>P. gingivalis</i> <i>porL</i> sequence in MCS 1	Ampicillin	Restriction cloning of a BamHI and NcoI fragment of a PCR product amplified using primers RHJ051 and RHJ052 from <i>P. gingivalis</i> gDNA into pRHJ004
pRHJ006	pRHJ004 with <i>P. gingivalis</i> <i>porL</i> sequence in MCS 1 and <i>porM</i> sequence in MCS 2	Ampicillin	Restriction cloning of an NdeI and KpnI fragment of a PCR product amplified using primers RHJ057 and RHJ058 from <i>P. gingivalis</i> gDNA into pRHJ005
pRHJ007	pT12 <i>F. johnsoniae</i> <i>gldL gldM</i> (1-225)-twinstrep (<i>FjogldLM'</i> _A)	Kanamycin	Gibson assembly using pT12_SpaPQR ^{3XFLAG} linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ164 and RHJ165
pRHJ008	pT12 <i>F. johnsoniae</i> <i>gldL gldM</i> (1-232)-twinstrep (<i>FjogldLM'</i> _B)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ166 and RHJ167
pRHJ009	pT12 <i>P. gingivalis</i> <i>porL porM</i> (1-216)-twinstrep (<i>PgiporLM'</i> _A)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>P. gingivalis</i> gDNA using primers RHJ168 and RHJ169
pRHJ101	pRHJ004 with <i>T. rossianum</i> <i>gldM</i> sequence in MCS 1	Ampicillin	Restriction cloning of a BamHI and NcoI fragment of a PCR product amplified using primers RHJ011 and RHJ012 from <i>T. rossianum</i> gDNA into pRHJ004
pRHJ102	pRHJ004 with <i>P. gingivalis</i> <i>porM</i> sequence in MCS 1	Ampicillin	Restriction cloning of a BamHI and NcoI fragment of a PCR product amplified using primers RHJ055 and RHJ056 from <i>P. gingivalis</i> gDNA into pRHJ004

pRHJ103	pRHJ004 with <i>T. rossianum gldM</i> sequence in MCS 1 and <i>gldL</i> sequence in MCS 2	Ampicillin	Restriction cloning of an NdeI and AsiSI fragment of a PCR product amplified using primers RHJ021 and RHJ044 from <i>T. rossianum</i> gDNA into pRHJ101
pRHJ104	pRHJ004 with <i>F. johnsoniae gldL</i> sequence in MCS 1	Ampicillin	Restriction cloning of a BamHI and NcoI fragment of a PCR product amplified using primers RHJ001 and RHJ002 from <i>F. johnsoniae</i> gDNA into pRHJ004
pRHJ105	pRHJ004 with <i>F. johnsoniae gldM</i> sequence in MCS 1	Ampicillin	Restriction cloning of a BamHI and NcoI fragment of a PCR product amplified using primers RHJ003 and RHJ004 from <i>F. johnsoniae</i> gDNA into pRHJ004
pRHJ106	pRHJ004 with <i>F. johnsoniae gldM</i> sequence in MCS 1 and <i>gldL</i> sequence in MCS 2	Ampicillin	Restriction cloning of an NdeI and KpnI fragment of a PCR product amplified using primers RHJ013 and RHJ014 from <i>F. johnsoniae</i> gDNA into pRHJ105
pRHJ107	pRHJ004 with <i>T. rossianum gldL</i> sequence in MCS 1	Ampicillin	Restriction cloning of a BamHI and NcoI fragment of a PCR product amplified using primers RHJ009 and RHJ010 from <i>T. rossianum</i> gDNA into pRHJ004
pRHJ108	pRHJ004 with <i>T. rossianum gldL</i> sequence in MCS 1 and <i>gldM</i> sequence in MCS 2	Ampicillin	Restriction cloning of an NdeI and AsiSI fragment of a PCR product amplified using primers RHJ023 and RHJ045 from <i>T. rossianum</i> gDNA into pRHJ107
pRHJ109	pRHJ004 with <i>F. johnsoniae gldL</i> sequence in MCS 1 and <i>gldM</i> sequence in MCS 2	Ampicillin	Restriction cloning of an NdeI and KpnI fragment of a PCR product amplified using primers RHJ015 and RHJ016 from <i>F. johnsoniae</i> gDNA into pRHJ105
pRHJ110	pRHJ004 with <i>P. gingivalis porM</i> sequence in MCS 1 and <i>porL</i> sequence in MCS 2	Ampicillin	Restriction cloning of an NdeI and KpnI fragment of a PCR product amplified using primers RHJ053 and RHJ054 from <i>P. gingivalis</i> gDNA into pRHJ102
pRHJ113	pMA-RQ_SthPKLMN with restriction sites removed by	Ampicillin	Q5 site-directed mutagenesis of pMA-

	mutagenesis		RQ_SthPKLMN using primers RHJ073, RHJ074, RHJ075, RHJ076, RHJ077, and RHJ078
pRHJ117	pT12 <i>S. thermophila</i> <i>gldL gldM</i> (1-229)- <i>twinstrep</i> (<i>SthgldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and fragments amplified from pRHJ113 using primers RHJ177, 178, 179 and RHJ203
pRHJ118	pT12 <i>P. gingivalis</i> <i>porL porM</i> (1-227)- <i>twinstrep</i> (<i>PgiporLM'</i> _c)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>P. gingivalis</i> gDNA using primers RHJ168 and RHJ202
pRHJ119	pT12 <i>T. rossianum</i> <i>gldL gldM</i> (1-242)- <i>twinstrep</i> (<i>TrogldLM''</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>T. rossianum</i> gDNA using primers RHJ170 and RHJ204
pRHJ122	pT12 <i>F. johnsoniae</i> <i>gldL gldM</i> (1-321)- <i>twinstrep</i> (<i>FjogldLM''</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ166 and RHJ205
pRHJ123	pT12 <i>P. gingivalis</i> <i>porL porM</i> (1-316)- <i>twinstrep</i> (<i>PgiporLM''</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>P. gingivalis</i> gDNA using primers RHJ168 and RHJ206
pRHJ124	pT12 <i>T. rossianum</i> <i>gldL gldM</i> (1-336)- <i>twinstrep</i> (<i>TrogldLM''</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>T. rossianum</i> gDNA using primers RHJ170 and RHJ208
pRHJ125	pT12 <i>S. thermophila</i> <i>gldL gldM</i> (1-321)- <i>twinstrep</i> (<i>SthgldLM''</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and fragments amplified from pRHJ113 using primers RHJ177, 178, 179 and RHJ207
pRHJ155	pT12 <i>F. johnsoniae</i> <i>gldL</i> (E49Q) <i>gldM</i> (1-232)- <i>twinstrep</i> (<i>FjogldL_{E49Q}M'</i> _B)	Kanamycin	Gibson mutagenesis of pRHJ008 using primers RHJ322 and RHJ323
pRHJ156	pT12 <i>P. gingivalis</i> <i>porL</i> (E60Q) <i>porM</i> (1-227)- <i>twinstrep</i>	Kanamycin	Gibson mutagenesis of pRHJ118 using primers

	(<i>Pgipor</i> _{LE60Q} <i>M'</i> c)		RHJ430 and RHJ431
pRHJ157	pT12 <i>S. thermophila</i> <i>gldL</i> (E51Q) <i>gldM</i> (1-229)- <i>twinstrep</i> (<i>Sthgld</i> _{LE51Q} <i>M'</i>)	Kanamycin	Gibson mutagenesis of pRHJ117 using primers RHJ432 and RHJ433
pRHJ158	pT12 <i>T. rossianum</i> <i>gldL</i> (E54Q) <i>gldM</i> (1-242)- <i>twinstrep</i> (<i>Trogld</i> _{LE54Q} <i>M'</i>)	Kanamycin	Gibson mutagenesis of pRHJ119 using primers RHJ428 and RHJ429
pRHJ162	pT12 <i>A. mensalis</i> <i>gldL</i> <i>gldM</i> (1-229)- <i>twinstrep</i> (<i>AmegldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>A. mensalis</i> gDNA using primers RHJ449 and RHJ451
pRHJ163	pT12 <i>C. canimorsus</i> <i>gldL</i> <i>gldM</i> (1-232)- <i>twinstrep</i> (<i>CcagldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>C. canimorsus</i> gDNA using primers RHJ452 and RHJ454
pRHJ164	pT12 <i>D. orientale</i> <i>gldL1</i> <i>gldM</i> (1-252)- <i>twinstrep</i> (<i>DorgldL1M'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>D. orientale</i> gDNA using primers RHJ455 and RHJ458
pRHJ165	pT12 <i>D. orientale</i> <i>gldL2</i> <i>gldM</i> (1-252)- <i>twinstrep</i> (<i>DorgldL2M'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>D. orientale</i> gDNA using primers RHJ456 and RHJ458
pRHJ166	pT12 <i>F. psychrophilum</i> <i>gldL</i> <i>gldM</i> (1-235)- <i>twinstrep</i> (<i>FpsgldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>F. psychrophilum</i> gDNA using primers RHJ459 and RHJ461
pRHJ167	pT12 NAT55 387147 <i>gldL</i> <i>gldM</i> (1-276)- <i>twinstrep</i> (<i>NATgldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from pTAH_NAT_GLdLM using primers RHJ462 and RHJ464
pRHJ168	pT12 <i>P. saltans</i> <i>gldL</i> <i>gldM</i> (1-230)- <i>twinstrep</i> (<i>PsagldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163

			and a fragment amplified from <i>P. saltans</i> gDNA using primers RHJ465 and RHJ467
pRHJ169	pT12 <i>R. anatipestifer</i> <i>gldL gldM</i> (1-237)- <i>twinstrep</i> (<i>RangldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>R. anatipestifer</i> gDNA using primers RHJ468 and RHJ470
pRHJ170	pT12 <i>S. wenxiniae</i> <i>gldL gldM</i> (1-224)- <i>twinstrep</i> (<i>SwegldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>S. wenxiniae</i> gDNA using primers RHJ471 and RHJ473
pRHJ171	pT12 UEG <i>gldL gldM</i> (1-254)- <i>twinstrep</i> (<i>UEGgldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from pTAH_UEG_GldKLMN using primers RHJ474 and RHJ476
pRHJ172	pT12 <i>Z. galactanivorans</i> <i>gldL gldM</i> (1-230)- <i>twinstrep</i> (<i>ZgagldLM'</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>Z. galactanivorans</i> gDNA using primers RHJ477 and RHJ479
pRHJ173	pT12 <i>A. mensalis</i> <i>gldL gldM</i> (1-323)- <i>twinstrep</i> (<i>AmegldLM''</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>A. mensalis</i> gDNA using primers RHJ449 and RHJ480
pRHJ174	pT12 <i>C. canimorsus</i> <i>gldL gldM</i> (1-330)- <i>twinstrep</i> (<i>CcagldLM''</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>C. canimorsus</i> gDNA using primers RHJ452 and RHJ481
pRHJ175	pT12 <i>D. orientale</i> <i>gldL1 gldM</i> (1-324)- <i>twinstrep</i> (<i>DorgldL1M''</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>D. orientale</i> gDNA using primers RHJ455 and RHJ482
pRHJ176	pT12 <i>D. orientale</i> <i>gldL2 gldM</i> (1-324)- <i>twinstrep</i> (<i>DorgldL2M''</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG

			linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>D. orientale</i> gDNA using primers RHJ456 and RHJ482
pRHJ177	pT12 <i>F. psychrophilum</i> <i>gldL gldM(1-321)-twinstrep</i> (<i>FpsgldLM</i> "")	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>F. psychrophilum</i> gDNA using primers RHJ459 and RHJ483
pRHJ178	pT12 NAT55 387147 <i>gldL gldM(1-380)-twinstrep</i> (<i>NATgldLM</i> "")	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from pTAH_NAT_GldLM using primers RHJ462 and RHJ484
pRHJ179	pT12 <i>P. saltans</i> <i>gldL gldM(1-321)-twinstrep</i> (<i>PsagldLM</i> "")	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>P. saltans</i> gDNA using primers RHJ465 and RHJ485
pRHJ180	pT12 <i>R. anatipestifer</i> <i>gldL gldM(1-331)-twinstrep</i> (<i>RangldLM</i> "")	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>R. anatipestifer</i> gDNA using primers RHJ468 and RHJ486
pRHJ181	pT12 <i>S. wenxiniae</i> <i>gldL gldM(1-315)-twinstrep</i> (<i>SwegldLM</i> "")	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>S. wenxiniae</i> gDNA using primers RHJ471 and RHJ487
pRHJ182	pT12 UEG <i>gldL gldM(1-378)-twinstrep</i> (<i>UEGgldLM</i> "")	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from pTAH_UEG_GldKLMN using primers RHJ474 and RHJ488
pRHJ183	pT12 <i>Z. galactanivorans</i> <i>gldL gldM(1-329)-twinstrep</i> (<i>ZgagldLM</i> "")	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>Z.</i>

			<i>galactanivorans</i> gDNA using primers RHJ477 and RHJ489
pRHJ184	pT12 <i>F. johnsoniae</i> <i>gldL gldM-twinstrep</i> (<i>FjogldLM</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ166 and RHJ296
pRHJ185	pT12 <i>P. gingivalis</i> <i>porL porM-twinstrep</i> (<i>PgiporLM</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>P. gingivalis</i> gDNA using primers RHJ168 and RHJ297
pRHJ186	pT12 <i>S. thermophila</i> <i>gldL gldM-twinstrep</i> (<i>SthgldLM</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and fragments amplified from pRHJ113 using primers RHJ177, 178, 179 and RHJ298
pRHJ187	pT12 <i>T. rossianum</i> <i>gldL gldM-twinstrep</i> (<i>TrogldLM</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>T. rossianum</i> gDNA using primers RHJ170 and RHJ299
pRHJ188	pT12 <i>A. mensalis</i> <i>gldL gldM-twinstrep</i> (<i>AmegldLM</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>A. mensalis</i> gDNA using primers RHJ449 and RHJ450
pRHJ189	pT12 <i>C. canimorsus</i> <i>gldL gldM-twinstrep</i> (<i>CcagldLM</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>C. canimorsus</i> gDNA using primers RHJ452 and RHJ453
pRHJ190	pT12 <i>D. orientale</i> <i>gldL1 gldM-twinstrep</i> (<i>DorgldL1M</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>D. orientale</i> gDNA using primers RHJ455 and RHJ457
pRHJ191	pT12 <i>D. orientale</i> <i>gldL2 gldM-twinstrep</i> (<i>DorgldL2M</i>)	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>D. orientale</i>

			gDNA using primers RHJ456 and RHJ457
pRHJ192	pT12 <i>F. psychrophilum gldL gldM-twinstrep (FpsgldLM)</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>F. psychrophilum</i> gDNA using primers RHJ459 and RHJ460
pRHJ193	pT12 NAT55 387147 <i>gldL gldM-twinstrep (NATgldLM)</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from pTAH_NAT_GldLM using primers RHJ462 and RHJ463
pRHJ194	pT12 <i>P. saltans gldL gldM-twinstrep (Psa gldLM)</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>P. saltans</i> gDNA using primers RHJ465 and RHJ466
pRHJ195	pT12 <i>R. anatipestifer gldL gldM-twinstrep (RangldLM)</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>R. anatipestifer</i> gDNA using primers RHJ468 and RHJ469
pRHJ196	pT12 <i>S. wenxiniae gldL gldM-twinstrep (SwegldLM)</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>S. wenxiniae</i> gDNA using primers RHJ471 and RHJ472
pRHJ197	pT12 UEG <i>gldL gldM-twinstrep (UEGgldLM)</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from pTAH_UEG_GldKLMN using primers RHJ474 and RHJ475
pRHJ198	pT12 <i>Z. galactanivorans gldL gldM-twinstrep (ZgagldLM)</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>Z. galactanivorans</i> gDNA using primers RHJ477 and RHJ478

pRHJ204	pT12 <i>A. mensalis</i> <i>gldL gldM(1-410)-twinstrep (AmegldLM^{'''})</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>A. mensalis</i> gDNA using primers RHJ449 and RHJ579
pRHJ205	pT12 <i>C. canimorsus</i> <i>gldL gldM(1-415)-twinstrep (CcagldLM^{'''})</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>C. canimorsus</i> gDNA using primers RHJ452 and RHJ580
pRHJ206	pT12 <i>D. orientale</i> <i>gldL1 gldM(1-432)-twinstrep (DorgldL1M^{'''})</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>D. orientale</i> gDNA using primers RHJ455 and RHJ581
pRHJ207	pT12 <i>D. orientale</i> <i>gldL2 gldM(1-432)-twinstrep (DorgldL2M^{'''})</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>D. orientale</i> gDNA using primers RHJ456 and RHJ581
pRHJ208	pT12 <i>F. johnsoniae</i> <i>gldL gldM(1-410)-twinstrep (FjogldLM^{'''})</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ166 and RHJ582
pRHJ209	pT12 <i>F. psychrophilum</i> <i>gldL gldM(1-408)-twinstrep (FpsgldLM^{'''})</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>F. psychrophilum</i> gDNA using primers RHJ459 and RHJ583
pRHJ210	pT12 NAT55 387147 <i>gldL gldM(1-483)-twinstrep (NATgldLM^{'''})</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from pTAH_NAT_GLdLM using primers RHJ462 and RHJ584
pRHJ211	pT12 <i>P. gingivalis</i> <i>porL porM(1-403)-twinstrep (PgiporLM^{'''})</i>	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>P. gingivalis</i>

			gDNA using primers RHJ168 and RHJ585
pRHJ212	pT12 <i>P. saltans</i> <i>gldL gldM(1-408)-twinstrep</i> (<i>PsagldLM</i> '')	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>P. saltans</i> gDNA using primers RHJ465 and RHJ586
pRHJ213	pT12 <i>R. anatipestifer</i> <i>gldL gldM(1-418)-twinstrep</i> (<i>RangldLM</i> '')	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>R. anatipestifer</i> gDNA using primers RHJ468 and RHJ587
pRHJ214	pT12 <i>S. thermophila</i> <i>gldL gldM(1-409)-twinstrep</i> (<i>SthgldLM</i> '')	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and fragments amplified from pRHJ113 using primers RHJ177, 178, 179 and RHJ588
pRHJ215	pT12 <i>S. wenxiniae</i> <i>gldL gldM(1-401)-twinstrep</i> (<i>SwegldLM</i> '')	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>S. wenxiniae</i> gDNA using primers RHJ471 and RHJ589
pRHJ217	pT12 UEG <i>gldL gldM(1-478)-twinstrep</i> (<i>UEGgldLM</i> '')	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from pTAH_UEG_GldKLMN using primers RHJ474 and RHJ591
pRHJ218	pT12 <i>Z. galactanivorans</i> <i>gldL gldM(1-415)-twinstrep</i> (<i>ZgagldLM</i> '')	Kanamycin	Gibson assembly using pT12_SpaPQR3XFLAG linearised using primers RHJ162 and RHJ163 and a fragment amplified from <i>Z. galactanivorans</i> gDNA using primers RHJ477 and RHJ592
pRHJ225	pT12 <i>S. wenxiniae</i> <i>gldL(K29A) gldM(1-224)-twinstrep</i> (<i>SwegldL_{K29A}M</i> '')	Kanamycin	Gibson mutagenesis of pRHJ170 using primers RHJ547 and RHJ548
pRHJ226	pT12 <i>S. wenxiniae</i> <i>gldL gldM(Y17F)(1-224)-twinstrep</i> (<i>SwegldLM'_{Y17F}</i> '')	Kanamycin	Gibson mutagenesis of pRHJ170 using primers RHJ549 and RHJ550
pRHJ227	pT12 <i>S. wenxiniae</i> <i>gldL(H32A) gldM(1-224)-twinstrep</i> (<i>SwegldL_{H32A}M</i> '')	Kanamycin	Gibson mutagenesis of pRHJ170 using primers RHJ595 and RHJ596

pRHJ228	pT12 <i>S. wenxiniae</i> <i>gldL</i> (E47A) <i>gldM</i> (1-224)- <i>twinstrep</i> (<i>SwegldL_{E47A}M'</i>)	Kanamycin	Gibson mutagenesis of pRHJ170 using primers RHJ597 and RHJ598
pRHJ229	pT12 <i>S. wenxiniae</i> <i>gldL</i> (E47D) <i>gldM</i> (1-224)- <i>twinstrep</i> (<i>SwegldL_{E47D}M'</i>)	Kanamycin	Gibson mutagenesis of pRHJ170 using primers RHJ599 and RHJ600
pRHJ230	pT12 <i>S. wenxiniae</i> <i>gldL</i> (E47Q) <i>gldM</i> (1-224)- <i>twinstrep</i> (<i>SwegldL_{E47Q}M'</i>)	Kanamycin	Gibson mutagenesis of pRHJ170 using primers RHJ601 and RHJ602
pRHJ231	pT12 <i>S. wenxiniae</i> <i>gldL</i> <i>gldM</i> (R9K)(1-224)- <i>twinstrep</i> (<i>SwegldLM'_{R9K}</i>)	Kanamycin	Gibson mutagenesis of pRHJ170 using primers RHJ603 and RHJ604
pRHJ233	pT12 <i>S. wenxiniae</i> <i>gldL</i> (1-74) <i>gldM</i> (1-224)- <i>twinstrep</i> (<i>SwegldL_{Δcyto}M'_{R9K}</i>)	Kanamycin	Gibson mutagenesis of pRHJ170 using primers RHJ607 and RHJ608

Table 2-6 Plasmids used for generating mutant strains of *F. johnsoniae* in this study.

Plasmid	Description	Resistance in <i>E. coli</i> (Resistance in <i>F. johnsoniae</i>)	Strategy/Source
pGEM-T	General cloning vector; <i>ori</i> ColE1	Ampicillin	Promega
pYT354	<i>sacB</i> -containing mobilizable suicide vector; <i>ori</i> ColE1	Ampicillin (Erythromycin)	(Zhu <i>et al</i> , 2017)
pRHJ010	intermediate vector for deletion of <i>F. johnsoniae gldM</i> ; pGEM-T with 2.7-kbp upstream and 2.5 kbp downstream of <i>gldM</i>	Ampicillin	Gibson assembly using pGEM-T linearised using primers RHJ144 and RHJ145 and fragments amplified from <i>F. johnsoniae</i> gDNA using primers RHJ158, RHJ159, RHJ160 and RHJ161
pRHJ011	mobilizable suicide vector for deletion of <i>F. johnsoniae gldM</i> ; pYT354 with 2.7-kbp upstream and 2.5-kbp downstream of <i>gldM</i>	Ampicillin (Erythromycin)	Restriction cloning of a BamHI and KpnI fragment of a PCR product amplified from pRHJ010 using primers pRHJ110 and pRHJ113
pRHJ012	Intermediate vector for generating mutations in <i>F. johnsoniae gldL</i> ; pGEM-T with <i>gldL</i> with 2.5-kbp upstream and 2.5-kbp downstream sequence	Ampicillin	Gibson assembly using pGEM-T linearised using primers RHJ144 and RHJ145 and a fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ146 and RHJ149
pRHJ013	Intermediate vector for generating mutations in <i>F. johnsoniae gldM</i> ; pGEM-T with <i>gldM</i> with 2.5-kbp upstream and 2.5-kbp downstream sequence	Ampicillin	Gibson assembly using pGEM-T linearised using primers RHJ144 and RHJ145 and a fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ147 and RHJ148

pRHJ014	pGEM <i>gldL</i> (N10A)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ310 and RHJ311
pRHJ015	pGEM <i>gldL</i> (Y13A)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ312 and RHJ313
pRHJ016	pGEM <i>gldL</i> (Y13F)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ348 and RHJ349
pRHJ017	pGEM <i>gldL</i> (K27A)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ273 and RHJ274
pRHJ018	pGEM <i>gldL</i> (H30A)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ275 and RHJ276
pRHJ019	pGEM <i>gldL</i> (S43A)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ314 and RHJ315
pRHJ020	pGEM <i>gldL</i> (T48A)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ279 and RHJ280
pRHJ021	pGEM <i>gldL</i> (E49A)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ277 and RHJ278
pRHJ022	pGEM <i>gldL</i> (E49Q)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ322 and RHJ323
pRHJ023	pGEM <i>gldL</i> (E47D)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ350 and RHJ351
pRHJ024	pGEM <i>gldL</i> (S56A)	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ316 and RHJ317
pRHJ025	pGEM <i>gldM</i> (R9A)	Ampicillin	Gibson mutagenesis of pRHJ013 using primers RHJ352 and RHJ353
pRHJ026	pGEM <i>gldM</i> (R9K)	Ampicillin	Gibson mutagenesis of pRHJ013 using primers RHJ354 and RHJ355
pRHJ027	pGEM <i>gldM</i> (R9E)	Ampicillin	Gibson mutagenesis of pRHJ013 using primers RHJ356 and RHJ357
pRHJ028	pGEM <i>gldM</i> (Q10A)	Ampicillin	Gibson mutagenesis of pRHJ013 using primers RHJ320 and RHJ321
pRHJ029	pGEM <i>gldM</i> (K11A)	Ampicillin	Gibson mutagenesis of pRHJ013 using primers RHJ358 and RHJ359

pRHJ030	pGEM <i>gldM</i> (N14A)	Ampicillin	Gibson mutagenesis of pRHJ013 using primers RHJ308 and RHJ309
pRHJ031	pGEM <i>gldM</i> (Y17A)	Ampicillin	Gibson mutagenesis of pRHJ013 using primers RHJ269 and RHJ270
pRHJ032	pGEM <i>gldM</i> (Y17F)	Ampicillin	Gibson mutagenesis of pRHJ013 using primers RHJ318 and RHJ319
pRHJ033	pGEM <i>gldM</i> (N27A)	Ampicillin	Gibson mutagenesis of pRHJ013 using primers RHJ271 and RHJ272
pRHJ036	suicide plasmid used to introduce the N10A codon change into <i>gldL</i> ; <i>gldL</i> (N10A) in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ014 using primers RHJ341 and RHJ342
pRHJ037	suicide plasmid used to introduce the Y13A codon change into <i>gldL</i> ; <i>gldL</i> (Y13A) in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ015 using primers RHJ341 and RHJ342
pRHJ038	suicide plasmid used to introduce the Y13F codon change into <i>gldL</i> ; <i>gldL</i> (Y13F) in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ016 using primers RHJ341 and RHJ342
pRHJ039	suicide plasmid used to introduce the K27A codon change into <i>gldL</i> ; <i>gldL</i> (K27A) in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ017 using primers RHJ341 and RHJ342
pRHJ040	suicide plasmid used to introduce the H30A codon change into <i>gldL</i> ; <i>gldL</i> (H30A) in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ018 using primers RHJ341 and RHJ342
pRHJ041	suicide plasmid used to introduce the S43A codon change into <i>gldL</i> ; <i>gldL</i> (S43A) in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ019 using primers RHJ341 and RHJ342
pRHJ042	suicide plasmid used to introduce the T48A codon change into <i>gldL</i> ; <i>gldL</i> (T48A) in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ020 using primers RHJ341 and RHJ342
pRHJ043	suicide plasmid used to introduce the E49A codon change into <i>gldL</i> ; <i>gldL</i> (E49A) in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ021 using primers RHJ341 and RHJ342
pRHJ044	suicide plasmid used to introduce the E49Q codon	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall

	change into <i>gldL</i> ; <i>gldL(E49Q)</i> in pYT354		and a fragment amplified from pRHJ022 using primers RHJ341 and RHJ342
pRHJ045	suicide plasmid used to introduce the E49D codon change into <i>gldL</i> ; <i>gldL(E49D)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ023 using primers RHJ341 and RHJ342
pRHJ046	suicide plasmid used to introduce the S56A codon change into <i>gldL</i> ; <i>gldL(S56A)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ024 using primers RHJ341 and RHJ342
pRHJ047	suicide plasmid used to introduce the R9A codon change into <i>gldM</i> ; <i>gldM(R9A)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ025 using primers RHJ343 and RHJ344
pRHJ048	suicide plasmid used to introduce the R9K codon change into <i>gldM</i> ; <i>gldM(R9K)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ026 using primers RHJ343 and RHJ344
pRHJ049	suicide plasmid used to introduce the R9E codon change into <i>gldM</i> ; <i>gldM(R9E)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ027 using primers RHJ343 and RHJ344
pRHJ050	suicide plasmid used to introduce the Q10A codon change into <i>gldM</i> ; <i>gldM(Q10A)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ028 using primers RHJ343 and RHJ344
pRHJ051	suicide plasmid used to introduce the K11A codon change into <i>gldM</i> ; <i>gldM(K11A)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ029 using primers RHJ343 and RHJ344
pRHJ052	suicide plasmid used to introduce the N14A codon change into <i>gldM</i> ; <i>gldM(N14A)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ030 using primers RHJ343 and RHJ344
pRHJ053	suicide plasmid used to introduce the Y17A codon change into <i>gldM</i> ; <i>gldM(Y17A)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ031 using primers RHJ343 and RHJ344
pRHJ054	suicide plasmid used to introduce the Y17F codon change into <i>gldM</i> ; <i>gldM(Y17F)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ032 using primers RHJ343 and RHJ344
pRHJ055	suicide plasmid used to introduce the N27A codon	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall

	change into <i>gldM</i> ; <i>gldM(N27A)</i> in pYT354		and a fragment amplified from pRHJ033 using primers RHJ343 and RHJ344
pRHJ237	pGEM <i>gldL(64-74GSS)</i>	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ617 and RHJ618
pRHJ238	pGEM <i>gldL(Δ64-74)</i>	Ampicillin	Gibson mutagenesis of pRHJ012 using primers RHJ621 and RHJ622
pRHJ240	suicide plasmid to replace codons 64-74 of <i>gldL</i> with sequence encoding GSSGSSGSSGS; <i>gldL(64-74GSS)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ237 using primers RHJ341 and RHJ342
pRHJ241	suicide plasmid to delete codons 64-74 of <i>gldL</i> ; <i>gldL(Δ64-74)</i> in pYT354	Ampicillin (Erythromycin)	Gibson assembly of pYT354 linearised with BamHI and Sall and a fragment amplified from pRHJ238 using primers RHJ341 and RHJ342

2.3.1 Polymerase chain reactions

Polymerase chain reactions (PCRs) to produce plasmids were performed using the Q5 High-Fidelity DNA polymerase 2X Master Mix (NEB) according to the manufacturer's instructions. PCRs to determine the presence of desired DNA sequences in bacterial colonies (colony PCR) were performed using OneTaq Quick-Load 2X Master Mix with Standard Buffer according to the manufacturers' instructions. Primer annealing temperatures were determined using the manufacturer's recommendations. Primers used in this study are listed in the Appendix in Table 9-3.

2.3.2 DNA sources

F. johnsoniae UW101 and *P. gingivalis* ATCC 33277 genomic DNA (gDNA) was prepared from overnight cultures by Dr Frédéric Lauber. gDNA from the following strains was obtained from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ): *Apibacter mensalis* DSM100903 (*Ame*), *Capnocytophaga canimorsus* strain Cc5 (*Cca*), *Draconibacterium orientale* DSM 25947 (*Dor*), *Flavobacterium psychrophilum* DSM 21280 (*Fps*), *Pseudopedobacter saltans* DSM 12145 (*Psa*), *Riemerella anatipestifer* DSM 15868 (*Ran*), *Sphingobacterium wenxiniae* DSM 22789 (*Swe*), *Thermonema rossianum* DSM 10300 (*Tro*), and *Zobellia galactanivorans* DSM 12808 (*Zga*). gDNA stocks were amplified using a GenomiPhi V3 Ready-To-Go DNA Amplification Kit (Cytiva), which uses non-specific primers and the Φ 29 polymerase to amplify long stretches of the genome.

gDNA was not available from the DSMZ for *Schleiferia thermophila* str Yellowstone (*Sth*), *Flavobacteriales* bacterium isolate NAT55 387147 (GenBank accession NZWA01000084; henceforth referred to as NAT), or an uncultured *Sphingobacteriia* bacterium sampled from the North Atlantic Ocean (GenBank accession FQ032818; henceforth referred to as UEG), so genes from these species were commercially synthesized. Gene synthesis (Invitrogen GeneArt) was used to create plasmid pMA-RQ_SthPKLMN containing genes RS08900 (*porP* homologue), RS08905 (*gldK*), RS08910 (*gldL*), RS08915 (*gldM*), and RS08920 (*gldN*) from *S. thermophila* with a Twin-Strep tag coding sequence

added to the end of *gldL*. Q5 site-directed mutagenesis (NEB) was used as described above using primers RHJ073-078 to remove BamHI, NcoI and NdeI restriction sites by silent mutation, yielding plasmid pRHJ113. Gene synthesis (Twist Bioscience) was used to create pTAH_NAT_GldLM, containing genes CMD25_08895 and CMD25_08890 (*gldL* and *gldM*) and their intergenic region from NAT, and pTAH_UEG_GldKLMN, containing loci S18_1049_0017-0020 (*gldKLMN*) and their intergenic regions from UEG.

2.3.3 Restriction cloning

Sequences were amplified by PCR from appropriate plasmid or gDNA templates using primers that incorporated the appropriate restriction sites. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen). Vectors and purified PCR products were digested with appropriate restriction enzymes (NEB) at 37 °C for 1-16 h. Digested vectors were treated with recombinant shrimp alkaline phosphatase (NEB) at 37 °C for 1 h. Enzymes were inactivated by incubation at 65 °C for 15 min. Digested vectors and inserts were excised from 1% agarose gels after staining with SYBR Safe stain (Invitrogen) and gel-extracted using the QIAquick Gel Extraction Kit (Qiagen). Inserts were ligated into vectors at an insert:vector ratio of 3:1 with T4 DNA ligase (NEB) at room temperature for 1 h or at 4 °C overnight. Chemically competent *E. coli* TOP10 or DH5 α Turbo cells were transformed with ligation products and plated on LB agar plates with the appropriate antibiotics. Plasmids were screened for the correct inserts by colony PCR then selected clones were grown overnight in liquid culture. Plasmids were extracted using the QIAprep Spin Miniprep Kit (QIAGEN) and sequences confirmed by Sanger sequencing (Source Bioscience).

2.3.4 HiFi Assembly

Some plasmids were assembled using Gibson assembly, where the 5' ends of strands in DNA duplexes are digested by an exonuclease to expose ssDNA 3' overhangs. ssDNA overhang sequences that are complementary between vector and insert anneal and the missing bases are filled in by a DNA

polymerase. The resulting nick is then ligated by a ligase enzyme. Assembly reactions were performed using the NEBuilder HiFi DNA Assembly Kit (NEB) according to the manufacturer's instructions, using primers designed according to the manufacturer's recommendations. Clones were screened and plasmids extracted as above.

2.3.5 Site directed mutagenesis

Where specified in Table 2-5, point mutations, insertions, and deletions were generated using the Kinase-Ligase-DpnI enzyme mix from the Q5 Site-Directed Mutagenesis Kit (NEB) or the QuikChange Site-Directed Mutagenesis Kit (Agilent) according to the manufacturers' instructions. Primers for these reactions were designed using the NEBaseChanger tool (<http://nebasechanger.neb.com/>) or the QuikChange Primer Design tool (<https://www.agilent.com/store/primerDesignProgram.jsp>) as appropriate.

Other point mutations, insertions, and deletions were generated by Gibson assembly, as indicated in Table 2-5 and Table 2-6. Pairs of primers including the desired mutation in the overlapping regions were used to linearise plasmids by PCR amplification. Plasmids were re-circularised either by HiFi Assembly as described above or by directly transforming chemically competent bacteria with PCR products (*in vivo* recombination) (Watson & García-Nafría, 2019). It was found that *in vivo* recombination efficiently re-circularised single DNA fragments but, in my hands, did not efficiently combine multiple DNA fragments.

2.3.6 Agarose gels

5 parts DNA sample were mixed with 1 part 6X purple loading dye (NEB) and loaded on a 1% w/v agarose gel in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.5) containing 1X SYBR Safe (Invitrogen). Gels were run for 20-40 min at 100 V in a Sub-Cell GT agarose gel electrophoresis system (Bio-Rad) using TAE running buffer. Bands were visualised using either a Safe Imager 2.0 blue light transilluminator (Fisher Scientific) or a Gel Doc (Bio-Rad).

2.4 Genetic modification of *F. johnsoniae*

To generate scarless mutations in the chromosomal *gldLM* genes of *F. johnsoniae* a homologous recombination-based system developed in the McBride laboratory (Zhu *et al*, 2017) was used (**Figure 2-1**). In this system the desired mutated sequence together with 2.5 kb regions of flanking chromosomal sequence, is inserted into the mobilizable suicide vector pYT354. This vector can be transferred from *E. coli* to *F. johnsoniae* by conjugation, but *F. johnsoniae* is not able to initiate replication from the *ColE1* origin of replication. Consequently, the plasmid can only be retained by integration into the chromosome (**Figure 2-1b**). pYT354 encodes ErmF, which confers resistance to erythromycin, so erythromycin can be used to select for integration of the plasmid by a first recombination event. pYT354 also encodes SacB, which confers sucrose sensitivity. After a second recombination event, this allows for selection for clones which have excised the plasmid backbone from their chromosome. 50% of such strains should have reverted to the parental sequence, and 50% should have acquired the desired mutation with no remaining plasmid sequence. The presence of desired mutations can be screened for by colony PCR using primers that anneal either side of the mutated region if the parental and mutant chromosomes produce different sized PCR products. This is usually the case for gene deletions and fusion tags but is not the case with point mutations. To overcome this problem when making point mutants in *gldL* and *gldM*, deletion strains FI_086 ($\Delta gldL$) and Rhj_006 ($\Delta gldM$) were used as parental strains to ensure there was a clear difference between the parental and modified PCR products. This system can be used to generate single-base pair mutations as well as whole gene deletions and fusion proteins.

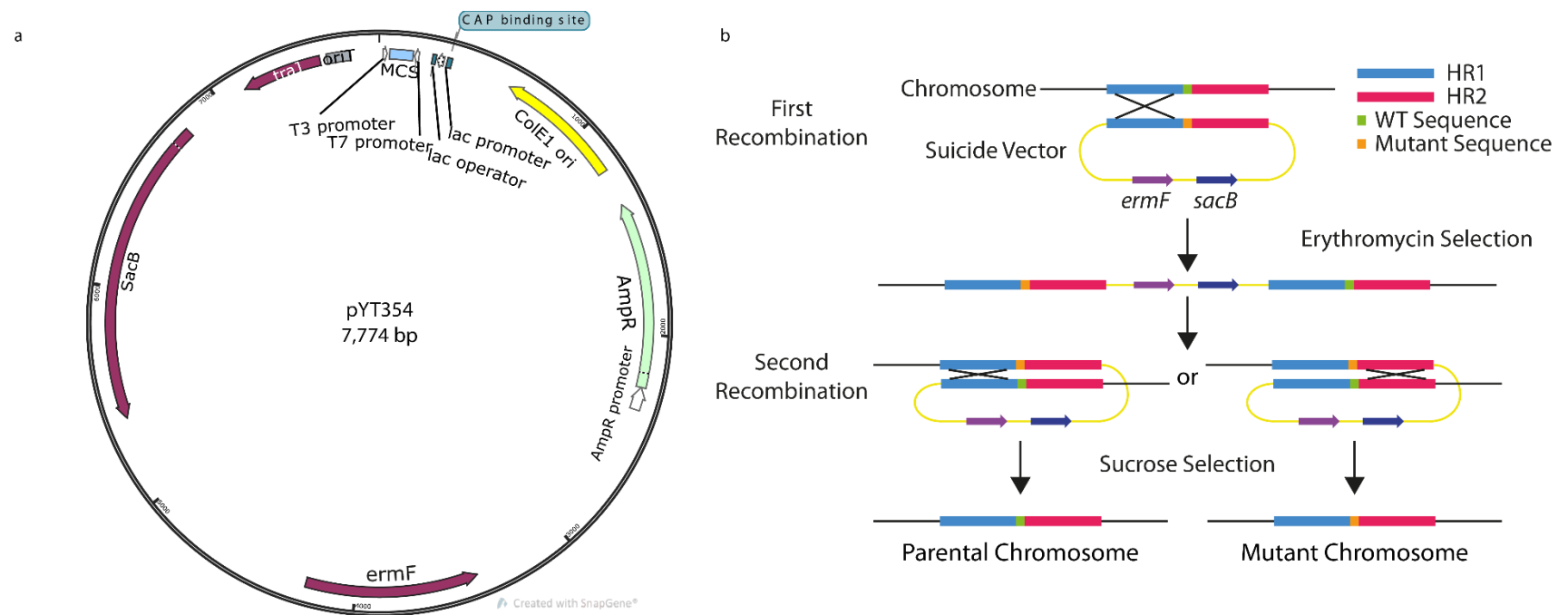


Figure 2-1 The pYT354 genetic manipulation system. **a** Map of the pYT354 vector. *ampR* encodes a beta-lactamase enzyme which confers resistance to ampicillin expressed from a promoter that is active in *E. coli* but not *F. johnsoniae*. *ermF* encodes an erythromycin methylase enzyme that confers resistance to erythromycin, under the control of a promoter that is active in *F. johnsoniae* but not *E. coli*. *sacB* encodes a sucrose levan synthase that confers sensitivity to sucrose, under the control of a promoter that is active in *F. johnsoniae* but only weakly active in *E. coli*. *traJ* encodes a component of the IncP-type plasmid conjugation system, the rest of the system is chromosomally encoded in the *E. coli* S17-1 donor strain. The mutated sequence and flanking regions are inserted into the multiple cloning site (MCS). The *oriT* site serves as the origin of transfer for conjugation. The ColE1 site can be used as an origin of replication by *E. coli* but not *F. johnsoniae*. **b** Schematic of the recombination and selection steps that lead to generation of scarless mutants. HR = homologous region.

A suicide vector to create an unmarked deletion of *gldM* was produced as follows. A 2.7 kb region containing the first 31 bp of *gldM* and the region upstream was amplified using primers RHJ148 and RHJ160. A 2.5 kb region containing the 38 bp of *gldM* and the region downstream was amplified using primers RHJ161 and RHJ147. The vector pGEM-T was linearised by PCR amplification using primers RHJ144 and RHJ145. These three fragments were combined by Gibson assembly to make pRHJ010 and then the fragment containing *F. johnsoniae* chromosomal sequences was amplified using RHJ110 and RHJ113. pYT354(Zhu *et al*, 2017) was linearised by digestion with BamHI and Sall then combined with the 5.2 kb fragment from pRHJ010 to give pRHJ011.

Intermediate cloning vectors to introduce mutations into *gldL* was made as follows. A 5.3 kb fragment containing *gldL* and the surrounding chromosomal regions was amplified using primers RHJ146 and RHJ149. pGEM-T was linearised by amplification using primers RHJ144 and RHJ145. These two fragments were combined by Gibson assembly to give pRHJ012. pRHJ013 containing *gldM* and surrounding regions was made similarly using primers RHJ147 and RHJ148.

Desired point mutations, short deletions, and fusion tags were introduced into pRHJ012 and pRHJ013 by Gibson assembly as described above, using the primer pairs described in

Table 2-6 and Table 9-3. Suicide vectors to introduce these mutations into *F. johnsoniae* were then constructed as follows. The mutated gene and flanking chromosomal sequences were PCR amplified using primers RHJ341 and 342 (*gldL*/pRHJ012) or primers RHJ343 and RHJ344 (*gldM*/pRHJ013). pYT354 was linearized by digestion with BamHI and Sall. Linearised pYT354 and the fragment containing the mutated gene were then assembled by Gibson cloning as described above.

Plasmids were introduced into *F. johnsoniae* by conjugation as follows. *E. coli* S17-1 cells carrying the appropriate plasmids were grown overnight in LB with ampicillin. *F. johnsoniae* parental cells were grown overnight in CYE. CYE agar plates were dried overnight then spread with 100 µl of 1 M CaCl₂.

1.5 ml samples of each culture were harvested by centrifugation at 6,000g for 3 min, then the *E. coli* S17-1 cells were washed to remove ampicillin by resuspending in fresh CYE and centrifuging at 6,000g for a further 3 min. Both cell pellets were then mixed by resuspending them in the same 750 µl of fresh CYE, then harvested by centrifugation at 6,000g for 3 min. The pellet was resuspended in 100 µl CYE and spotted onto the prepared CYE + CaCl₂ agar plate. After 20-24 h incubation at 30 °C, the resulting bacterial colonies were scraped up with an inoculation loop and resuspended in 750 µl TC Buffer (10 mM Tris-HCl pH 7.6, 8 mM CaCl₂). 75 µl of the resuspended cells were mixed with 675 µl fresh TC Buffer to give a 10⁻¹ dilution and both samples were centrifuged at 6,000g for 3 min to collect cells. The cell pellets were resuspended in 100 µl TC Buffer and spread on CYE agar plates supplemented with erythromycin. These plates were incubated at 30 °C for 48 h. Single *F. johnsoniae* colonies were re-streaked on fresh CYE agar with erythromycin plates and incubated at 30 °C for 24 h to ensure they had acquired erythromycin resistance. Strains carrying replicable plasmids were then grown in CYE medium with erythromycin overnight and stored as glycerol stocks at -80 °C. Strains with integrated pYT354-derived plasmids were grown overnight in two separate liquid CYE cultures, one with erythromycin to maintain the integrated plasmid and one without antibiotic to allow loss of the plasmid backbone by a second recombination event. Cells grown with erythromycin were used to make glycerol stocks of the co-integrand strain and stored at -80 °C. Cells grown without erythromycin were diluted in TC Buffer from 10⁰ to 10⁻⁷. 200 µl of each dilution was spread on a separate CYE agar plate supplemented with 5% sucrose. These plates were then incubated at 30 °C for 36 h. Individual *F. johnsoniae* colonies were re-streaked on fresh CYE agar plates with 5% sucrose and incubated at 30 °C overnight to ensure they had lost the plasmid backbone and were sucrose-insensitive. Colony PCR was used to screen for the presence of the desired chromosomal modifications. These were then confirmed by Sanger sequencing. Clones were grown overnight in CYE and stored as glycerol stocks at -80 °C.

2.5 Protein sample analysis

2.5.1 SDS-PAGE

Purified proteins were analysed by SDS-PAGE using 15% (for GldLM' constructs) or 12% (for other constructs) polyacrylamide gels made as a 15% resolving gel underneath a 5% stacking gel. The resolving gel was composed of 12% or 15% v/v acrylamide/bis-acrylamide (ratio 19:1), 375 mM Tris-HCl pH 8.8, 0.1% w/v SDS, 0.1% w/v ammonium persulphate and 0.1% v/v tetramethylethylenediamine. The stacking gel was composed of 5% v/v acrylamide/bis-acrylamide (ratio 19:1), 125 mM Tris-HCl pH 6.8, 0.1% w/v SDS, 0.1% w/v ammonium persulphate and 0.1% tetramethylethylenediamine. Cell pellets and secretome samples were analysed by SDS-PAGE using 4-20% gradient Mini-PROTEAN precast gels (Bio-Rad). The streptavidin elution fractions from the expression screen for proteins recombinantly expressed from pRHJ004-derived plasmids were analysed by SDS-PAGE using Mini-PROTEAN TGX Stain-Free precast gels, which contain a trihalo dye that crosslink with tryptophan residues when exposed to UV radiation to produce fluorescence. 2 parts sample were mixed with 1 part 3X loading buffer (150 mM Tris-HCl pH 6.8, 6% w/v SDS, 30% v/v glycerol, 0.3% w/v Bromophenol Blue, 300 mM DTT) before loading. Samples were not heated before running the gel as this can cause aggregation of membrane proteins. Gels were using a Mini-PROTEAN tetra cell (Bio-Rad) in SDS-PAGE running buffer (25 mM Tris, 192 mM glycine, 0.1% w/v SDS) at 180 V until the dye front reached the bottom of the resolving gel. Gels were stained in InstantBlue (Expedeon) or QuickBlue (Strattech) and imaged using a Gel Doc (Bio-Rad). Mini-PROTEAN TGX Stain-Free precast gels were imaged using the Stain-Free settings on a Gel Doc. sfGFP fusion proteins were imaged using a blue light filter on the Gel Doc before staining.

2.5.2 Immunoblotting

For detection of specific proteins by immunoblotting, samples were separated by SDS-PAGE as described above. After electrophoresis, proteins were transferred from the polyacrylamide gel to nitrocellulose membranes using a Transblot SD Semi-Dry Transfer Cell (Bio-Rad) in 48 mM Tris with 39

mM glycine, 0.0375% SDS, 20% v/v methanol, operated at 22 V for 75 minutes. After transfer, membranes were blocked with 2% w/v milk powder in PBS(T) (PBS with 0.02% Tween-20) for 1 h at room temperature or overnight at 4 °C. Membranes were probed with primary antibodies in 2% w/v milk powder in PBS(T) at the dilutions indicated in Table 2-7 for 1 h at room temperature or overnight at 4 °C. Membranes were washed three times for 5 min in PBS(T) then incubated with a secondary antibody conjugated to horseradish peroxidase in 2% w/v milk powder PBS(T) at 1:5,000 dilution for 1 h at room temperature then washed three times for 5 min in PBS(T). 6 ml Amersham ECL solution was added to membranes for 1-1.5 min, then luminescence was detected using a Gel Doc fitted with a ChemiDoc XRS+ camera (Bio-Rad).

Table 2-7 Antibodies used in this study.

Antibody	Dilution Used	Source	Reference
α GldL _{cyto}	1:5,000	Davids Biotechnologie	(Hennell James <i>et al</i> , 2021)
α GldM _{peri}	1:5,000	Davids Biotechnologie	(Hennell James <i>et al</i> , 2021)
α ChiA	1:1,000	Davids Biotechnologie	(Lauber <i>et al</i> , 2018)
anti-Rabbit IgG HRP conjugate	1:5,000	Promega (Batch Q4018)	-

2.5.3 Protein Concentration determination

Protein concentrations were determined spectrophotometrically using a NanoDrop 2000c spectrophotometer. An absorbance at 280 nm (A_{280}) value of 1 was taken to be equivalent to 1 mg ml⁻¹ of pure protein. For SEC-MALS experiments extinction coefficients were calculated using the ProtParam tool on the ExPASy server (Gasteiger *et al*, 2005).

2.5.4 Size-Exclusion Chromatography with Multi-Angle Light Scattering

The size of protein complexes can be estimated by comparing their elution volume on a size-exclusion column with standard proteins of known size. However, this relies on the proteins having a

spherical shape and not interacting with the column resin, which is not always the case. Size-Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS) can directly determine the absolute mass of particles using the fact that the amount of light scattered by particles in solution is proportional to their mass (Wyatt, 1993; Wen *et al*, 1996). Different refractive indices can be used to deconvolute the contributions of different components of the particle such as protein and detergent. SEC-MALS was carried out by Dr Steven Johnson (SJ) as follows. Proteins were concentrated to 1.2 mg ml⁻¹ in the same buffer they were purified in and 100 µl was injected onto a Superose 6 10/300 Increase column (Cytiva). Light scattering and refractive index were measured using a Dawn Heleos-II light scattering detector and an Optilab-TrEX refractive index monitor. Analysis was carried out by SJ using ASTRA 6.1.1.17 software assuming an $A_{280} = 0.84$ is equivalent to 1 mg ml⁻¹ for both *PgiPorLM'*_A and *PgiPorLM* and dn/dc values of 0.186 ml g⁻¹ for protein and 0.143 ml g⁻¹ for detergent.

2.5.5 Mass spectrometry

Mass spectrometry sample processing and peptide identification was carried out by the Advanced Proteomics Facility (University of Oxford), the Target Discovery Institute (University of Oxford) or the BSRC Mass Spectrometry and Proteomics Facility (University of St Andrews). Bands of interest were excised from gels and subjected to overnight in-gel trypsin digestion. The tryptic peptides were then subjected to liquid chromatography-mass spectrometry analysis.

2.6 Protein Expression and Purification

All cells grown for protein expression were freshly transformed with the appropriate plasmids.

2.6.1 Expression screens for proteins expressed from pRHJ004 plasmids

To decide which combinations of plasmid, expression strain, and induction conditions were best for purifying GldLM homologues from the pRHJ004-based plasmids, small scale cultures were screened as follows. *E. coli* C41(DE3), BL21(DE3) (Invitrogen), BL21 Star (DE3) (Invitrogen), and Mt56(DE3) chemically competent cells were transformed with each of the plasmids pRHJ006 (expressing *P.*

gingivalis PorL-sfGFP-His₈ and PorM-TS), pRHJ103 (expressing *T. rossianum* GldM-sfGFP-His₈ and GldL-TS), pRHJ106 (expressing *F. johnsoniae* GldM-sfGFP-His₈ and GldL-TS), pRHJ108 (expressing *T. rossianum* GldL-sfGFP-His₈ and GldM-TS), pRHJ109 (expressing *F. johnsoniae* GldL-sfGFP-His₈ and GldM-TS) and pRHJ110 (expressing *P. gingivalis* PorM-sfGFP-His₈ and PorL-TS). Single colonies were used to inoculate 2 ml cultures in Terrific Broth (TB; Fisher Chemical) supplemented with ampicillin in a 24-well plate (Promega) overnight at 37 °C with shaking. 3 x 100 µl of each overnight culture was used to inoculate 4 ml fresh TB with ampicillin in 3 separate 24-well plates A-C.

Plate A was grown at 37 °C with shaking for 3 h. Expression was then induced by adding 1 M isopropyl β-D-1-thiogalactopyranoside (IPTG; VWR) to each well to a final concentration of 1 mM. Cells were grown at 37 °C with shaking for a further 4 h. Cells were then harvested by centrifugation at 3,220g for 10 min at 4 °C. The supernatant was discarded, and the pellets stored at -20 °C overnight. Plates B and C were grown at 37 °C with shaking for 8 h, reaching OD₆₀₀ 3-4. Expression was then induced by adding 1 M IPTG to each well to a final concentration of 0.1 mM. Cells were grown for a further 16 h at 24 °C (Plate B) or 18 °C (Plate C) with shaking. Cells were then harvested by centrifugation at 3,220g for 10 min at 4 °C and the supernatants discarded. Plate A cell pellets were thawed at room temperature.

Except where otherwise noted, subsequent steps were carried out at room temperature. Each cell pellet from all 3 plates was resuspended in 1 ml Dulbecco A Phosphate Buffered Saline (PBS) then transferred to a fresh 96-well 2 ml block (Starlab). 10 µl of each suspension was diluted in 90 µl PBS in a separate 96-well fluorescence plate (Costar). These diluted samples were used to measure OD₆₀₀ and sfGFP fluorescence (λ_{ex} = 485 nm and λ_{em} = 538 nm) using a SpectraMax M3 plate reader (Molecular Devices). The remaining resuspensions were centrifuged at 3,220g for 10 min at 4 °C to re-pellet cells. Pellets were resuspended in 100 µl each 100 mM NaH₂PO₄ with 600 mM NaCl, 1 mg

ml⁻¹ lysozyme, 0.08 U ml⁻¹ benzonase, and 1 tablet SIGMAFAST protease inhibitor cocktail per 100 ml buffer, then the plate was agitated at 800 rpm for 30 minutes. To solubilise membranes 100 µl 2.5 % dodecyl maltoside (DDM; Anatrace) was added to each well. The plate was agitated at 800 rpm for a further 1 h. Unsolubilised material was removed by centrifugation at 3,220g for 30 minutes at 4 °C. To purify solubilised proteins, 15 µl streptavidin magnetic beads (Pierce) was dispensed into each well of a clean 96-well plate and 20 µl Ni-NTA magnetic beads (Ni-NTA) was dispensed into each well of a separate clean 96-well plate. 90 µl of the soluble supernatant was added to each well and the plates were shaken for 60 minutes. The plates were placed on a magnet for one minute and the supernatant removed. Beads were washed twice by adding 200 µl Wash Buffer (50 mM NaH₂PO₄, 300 mM NaCl, 0.03% w/v DDM) or 200 µl Wash Buffer with 30 mM imidazole to each well as appropriate and shaking for 5 minutes. Plates were placed on the magnet for 1 minute before the supernatant was removed. Bound proteins were eluted by adding 50 µl Wash Buffer with 10 mM D-biotin (IBA) or Wash Buffer with 250 mM imidazole to each well as appropriate and shaking for 1 minute before placing on the magnet for 1 minute and collecting the supernatant in a 96-well fluorescence plate. sfGFP fluorescence of the elution fractions was measured as above. Elution fractions were analysed by SDS-PAGE.

2.6.2 Expression screens for proteins expressed from pT12 plasmids

Some expression screens were carried out with assistance from Alicia Hunter (AH). For the screens of GldLM^{'''} and GldLM constructs only BL21 (DE3) cells were used.

To determine which GldLM constructs expressed well and could be purified as complexes of both proteins, small cultures were grown and proteins purified as follows. 24-well format plates with gas-permeable seals were used to grow BL21 (DE3) and Mt56 (DE3) cells carrying the appropriate plasmids in 3 ml of 2xYT medium with kanamycin at 37 °C with shaking for 8 h to OD₆₀₀ = 3-4. 25 µl of each culture was used to inoculate 3 x 4 ml TB medium with kanamycin and 1 mg ml⁻¹ L-rhamnose in

fresh 24-well plates. These cultures were grown at 37 °C with shaking for 12-13 h. Cells were harvested by centrifugation at 3,220g for 10 minutes at 4 °C. Except where otherwise noted, all subsequent steps were carried out at room temperature.

All three cell pellets for each strain/plasmid combination were resuspended in the same 300 µl of Tris-buffered saline with EDTA (TBSE; 100 mM Tris-HCl pH 8.0, 150 mM NaCl, and 1 mM EDTA) supplemented with 1 tablet SIGMAFAST protease inhibitors per 100 ml buffer, 30 µg ml⁻¹ DNase I, and 100 µg ml⁻¹ lysozyme using a 1 ml pipette then shaken vigorously for 1 h. To lyse the cells, 200 µl of each sample was transferred to a fresh 96 deep well plate and 50 µl TBSE with 5% w/v DDM was added to each well. DDM was found to lyse cells and solubilize proteins in this format more efficiently than LMNG so was used for expression screens, whilst LMNG was used in large purifications for reasons discussed below. Samples were then incubated with gentle shaking for 1 h. Samples were centrifuged at 3,220g for 30 min at 4 °C to pellet unsolubilised debris. To purify proteins, 200 µl of each sample was added to 5 µl StrepTactin XT MagStrep “type3” beads (IBA; 100 µl starting volume of 5% suspension) in a 96-well plate pre-equilibrated in TBSE, then incubated with gentle shaking for 1 h. The plate was placed on a magnet for 1 min to collect the beads at the bottom of the wells and the supernatant removed. The beads were washed twice by adding 200 µl TBSE with 0.025% DDM, incubating with gentle shaking for 4 min, placing on the magnet for 1 min and removing the supernatant. Proteins were eluted by adding 50 µl Buffer 5.E TBSE with 0.025% DDM and 50 mM D-biotin (IBA), incubating with gentle shaking for 5 min, placing on the magnet for 1 min then collecting the supernatant. Elution fractions were analysed by SDS-PAGE.

2.6.3 Purification of GldLM homologues from pRHJ004-derived vectors

PgiPorLM proteins were purified using plasmid pRHJ006 as follows. Modifications to the protocol are described at the end of this section and when they were used is described in the relevant chapters.

Cells of the appropriate strain carrying the pRHJ006 plasmid were grown in 50 ml TB with ampicillin at 37 °C with shaking overnight. Cultures were diluted 1:100 in fresh TB with ampicillin, using 1 l of medium in 2.5 l half-baffled Tunair flasks. Cultures were grown at 37 °C with 180 rpm shaking for 8 h, reaching OD₆₀₀ 3-4. Expression was then induced by adding 100 µl 1 M IPTG to a final concentration of 0.1 mM. Cultures were grown for a further 16 h at 24 °C with shaking at 180 rpm. Cells were harvested by centrifugation at 4,500g for 15 minutes at 4 °C. Cell pellets were washed once by resuspending in PBS then centrifuging at 3,220g for 20 minutes at 4 °C and stored at -20 °C until use. Except where otherwise noted, subsequent steps were carried out at 4 °C.

Purifications in DDM were carried out as follows. Cell pellets were thawed at room temperature then resuspended in 4 ml per g cell mass PBS supplemented with 1 mg ml⁻¹ lysozyme, 0.3 mg ml⁻¹ DNase I, and 1 tablet SIGMAFAST protease inhibitor per 100 ml buffer using a Dounce tissue grinder. Cells were then lysed by passing through an Emulsiflex C5 homogeniser (Avestin) once with no backpressure, then 5 times with a back pressure of 100 MPa. Cellular debris was removed by centrifugation at 27,200g for 30 minutes. Membranes were harvested by centrifugation at 210,000g for 60 minutes. Membrane pellets were stored at 4 °C overnight, then resuspended in 8 ml of PBS per g pellet. Membranes were disrupted using a spatula then homogenized using a Dounce tissue grinder using the loose pestle until no lumps were visible and then using the tight pestle for 10 strokes. To solubilise membranes, 1 ml 10% w/v DDM per g membrane and 1 M imidazole were added to approximate final concentrations of 1 % DDM and 20 mM imidazole. Resuspended membranes were gently stirred for 60 minutes. Insolubilised membranes were removed by centrifugation at 70,000g for 30 minutes at 10 °C.

Supernatants were loaded onto 5 ml Ni-NTA Superflow cartridges (Qiagen) equilibrated in PBS with 0.03% DDM and 20 mM imidazole using a peristaltic pump at 1 ml min⁻¹. The cartridges were then

washed with 10 column volumes (CV) PBS with 0.03% DDM and 20 mM imidazole followed by 10 CV Buffer PBS with 0.03% DDM and 40 mM imidazole. Bound proteins were eluted in 8 ml PBS with 0.03% DDM and 250 mM imidazole. Protein concentrations were estimated using a NanoDrop spectrophotometer (Thermo Scientific), assuming $A_{280} = 1$ to be equal to 1 mg ml^{-1} . TEV-His₆ protease was added at a 1:1 w:w ratio to remove the sfGFP-His₈ tag and samples dialysed overnight against PBS containing 0.02% DDM. Dialysed samples were centrifuged at $3,220g$ for 10 minutes to remove any precipitated material, then 1 M imidazole was added to a final concentration of 20 mM. To remove TEV protease and sfGFP-His₈ Samples were loaded on 5 ml Ni-NTA cartridges equilibrated in PBS with 0.02% DDM and 20 mM imidazole The flowthrough fractions were collected and loaded onto 1 ml StrepTactin XT Superflow cartridges (IBA) equilibrated in PBS with 0.03% DDM by recirculation at 1 ml min^{-1} for 2 hours. Columns were washed in 10 CV PBS + 0.02% DDM, then bound proteins were eluted in 12 1 ml fractions of PBS with 0.02% DDM and 50 mM D-biotin. Fractions containing protein were selected according to A_{280} values then pooled and analysed by SDS-PAGE. Selected fractions were pooled and concentrated to 500 μl using a 100-kDa molecular weight cut-off (MWCO) Amicon Ultra-15 centrifugal filter. The samples were then injected using a washed 0.5 ml loop onto either a Superdex 200 10/300 Increase column or a Superose 6 10/300 Increase column (Cytiva) mounted on an Äkta Pure FPLC system equilibrated in 20 mM HEPES pH 7.5 with 150 mM NaCl and 0.02% DDM. The same buffer was passed through the column at a flow rate of 0.5 ml min^{-1} and 0.5 ml fractions were collected. Fractions of interest were analysed by SDS-PAGE then stored at 4°C until further use.

Purifications in LMNG were carried out as follows. Cells were lysed and membranes were harvested and resuspended by the same method as for purifications in DDM, with the exception that 1 mM EDTA was added to all buffers. Membranes were solubilised by the same method as for purifications in DDM, but 1 ml of 10% w/v LMNG solution and no imidazole was added instead of DDM solution

and 1 M imidazole. After centrifugation to remove unsolubilised membranes, supernatants were passed through 5 ml StrepTactin XT Superflow cartridges equilibrated in PBS with 1 mM EDTA and 0.02% LMNG at 1 ml min⁻¹. The cartridges were washed with 10 CV PBS with 1 mM EDTA and 0.02% LMNG then 2 CV PBS with 0.02% LMNG. Bound proteins were then eluted in 8 ml PBS with 0.02% LMNG and 50 mM D-biotin. 1 M imidazole was added to a final concentration of 20 mM then the elution fractions were passed over 1 ml Ni-NTA Superflow cartridges equilibrated in PBS with 20 mM imidazole and 0.02% LMNG at 1 ml min⁻¹. The cartridges were washed with 10 CV PBS with 20 mM imidazole and 0.02% LMNG then 10 CV PBS with 40 mM imidazole and 0.02% LMNG. Bound proteins were eluted in 6 ml PBS with 0.02% LMNG and 250 mM imidazole. Protein concentrations were estimated using a NanoDrop spectrophotometer (Thermo Scientific), assuming $A_{280} = 1$ to be equal to 1 mg ml⁻¹. TEV-His₆ protease was added at a 1:1 w:w ratio to remove the sfGFP-His₈ tag and samples dialysed against PBS with 0.02% LMNG overnight. Dialysed samples were centrifuged at 3,220g for 10 minutes to remove any precipitated material, then 1 M imidazole added to a final concentration of 20 mM. Samples were passed through 1 ml Ni-NTA cartridges equilibrated in PBS with 0.02% LMNG and 20 mM imidazole to remove TEV protease and sfGFP-His₈. The flowthrough fractions were collected and analysed by SDS-PAGE. Selected fractions were then concentrated and further purified by size-exclusion chromatography by the same method used for purifications in DDM, with the exception that 0.02% DDM in buffers was replaced with 0.02% LMNG.

Judging by the green colour, most of the recombinant proteins were associated with the cellular debris fraction recovered by centrifugation at 27,200g following cell breakage rather than the supernatant fraction. Therefore, for some proteins purifications were purified from the cellular debris as follows. Cells were resuspended and lysed as normal, then centrifuged at 27,200g to pellet the cellular debris. The cellular debris pellets were weighed and resuspended in 8 ml PBS with 1 mM EDTA per g pellet by the same method used for membrane pellets. 1 ml 10% w/v LMNG solution was

added per g pellet to a concentration of approximately 1% LMNG. Unsolubilised material was removed by centrifugation at 70,000g at 10 °C for 30 min and purification was continued following the method for purifications in LMNG.

FjoGldLM and *TroGldLM* proteins were purified as follows. Mt56 cells carrying pRHJ109 (expressing *FjoGldLM*) were grown in TB with ampicillin at 37 °C with shaking for 8 h. protein expression was induced by addition of 1M IPTG to a final concentration of 0.1 mM and cells were grown for a further 16 h at 24 °C with shaking. Cells were then harvested as described for *PgiPorLM* proteins. Mt56 cells carrying pRHJ103 (expressing *TroGldLM*) were grown in TB with ampicillin at 37 °C with shaking for 3 h, reaching $OD_{600} = 0.6$. Protein expression was induced by addition of 1 M IPTG to a final concentration of 1 mM. Cells were grown for a further 4 h at 37 °C with shaking. Cells were then harvested as described for *PgiPorLM* proteins. Proteins were then purified according to the method for purifications in LMNG above, with the exception that DDM was used in place of LMNG as it generally solubilises a greater fraction of the protein present in membranes. To solubilise membranes 10% w/v DDM solution was used in place of 10% w/v LMNG solution. In other buffers 0.02% DDM was used in place of 0.02% LMNG.

2.6.4 Lysis and centrifugation conditions trial

Most sfGFP was seen in the cellular debris pellet after cell lysis when purifying *PgiPorLM* from cells carrying pRHJ006. Different lysis conditions and centrifugation speeds were therefore screened. Cells were grown, harvested, and lysed as in 2.6.3. After cell lysis, the lysate was split into two equal fractions. Fraction A had additional lysozyme added to a final concentration of 2.5 mg ml⁻¹ and was stirred gently at 4 °C for 50 minutes. Fraction B was immediately centrifuged at 4,000g for 10 minutes. The supernatant was transferred to a fresh centrifuge tube and centrifuged at 8,000g for 10 minutes. The supernatant was transferred to a fresh centrifuge tube and centrifuged at 12,000g for

10 minutes. After stirring, Fraction A was subjected to the same centrifugation procedure. Samples of the supernatants and pellets were resuspended in equal volumes and analysed by SDS-PAGE.

2.6.5 Purification of GldLM homologues from pT12 vectors

Some purifications were carried out with assistance from AH.

Proteins expressed from pT12-based vectors were generally purified as follows. Variations to the protocol are described at the end of this section and in Chapters 4 & 6. BL21(DE3) cells carrying the appropriate plasmids were grown in 50 ml Terrific Broth (TB) supplemented with kanamycin at 37 °C with shaking for 8 h. Cells were then diluted to an OD₆₀₀ of 0.02 in 1 l fresh TB medium in 2.5 litre half-baffled flasks supplemented with kanamycin and 0.2% L-rhamnose and grown at 37 °C with shaking at 200 rpm for 12-14 h. Experience in the Lea laboratory, and in the Samuel Wagner laboratory that developed the pT12 vector, shows that cells continue growing during expression from the pT12 vector induced by L-rhamnose. Growth conditions were therefore selected to give high final cell densities. Cells were harvested by centrifugation at 4,500g for 15 minutes, washed once with PBS, and stored at -20 °C until further use.

Cell pellets were thawed in a room temperature water bath then resuspended in 3.3 ml per g of cells of PBS with 1 mM EDTA, 30 µg ml⁻¹ DNase I, 400 µg ml⁻¹ lysozyme, and 1 SIGMAFAST protease inhibitor cocktail tablet per 100 ml of buffer. Cell pellets were resuspended using a Dounce tissue grinder. The following steps were carried out at 4 °C unless otherwise noted. Cells were disrupted by passing them through an Avestin Emulsiflex homogeniser operated at 15,000-20,000 PSI 4-5 times. The lysate was centrifuged at 27,200g for 30 min to pellet cell debris and then the supernatant was centrifuged at 200,000g for 1 h to collect membranes. Membrane pellets were stored overnight then disrupted using a spatula and homogenised with a Dounce tissue grinder in 8 ml of PBS with 1 mM EDTA per g membrane. The loose pestle was used until no more pieces of membrane were visible,

then the tight pestle was used for 10 strokes to break up membrane fragments as much as possible. To solubilise membranes, 1 ml of 10% LMNG solution was added per g membrane and the suspension was gently stirred in a beaker for 2 h. Non-solubilised material was removed by centrifugation at 75,000g at 10 °C for 30 min.

The resulting supernatants were passed over a StrepTactin XT 5 ml Superflow cartridge (IBA) equilibrated in PBS with 1 mM EDTA and 0.02% LMNG at a flow rate of 1 ml min⁻¹ to allow binding of the Twin-Strep tag. The cartridge was washed with 10 column volumes (CV) PBS with 1 mM EDTA and 0.02% LMNG, then proteins were eluted in 2 CV PBS with 1 mM EDTA, 0.02% LMNG and 50 mM D-biotin over several 1 ml fractions. Both washing and elution steps used a flow rate of 1 ml min⁻¹. Protein concentration was determined spectrophotometrically using a NanoDrop and fractions with A₂₈₀ values greater than 0.1 were analysed by SDS-PAGE with Coomassie Blue staining. Elution fractions containing the proteins of interest were concentrated to 500 µl using a 100-kDa MWCO Amicon Ultra-15 centrifugal filter and injected using a 0.5 ml washed loop onto a Superose 6 10/300 Increase GL SEC column equilibrated in 20 mM HEPES pH 7.5 with 150 mM NaCl and 0.02% LMNG and mounted on an ÄKTA pure system (Cytiva). The same buffer was passed through the column at a flow rate of 0.5 ml min⁻¹ and 0.5 ml fractions were collected. Protein concentrations were determined spectrophotometrically assuming that A₂₈₀ = 1 is equivalent to 1 mg ml⁻¹ and protein purity was assessed by SDS-PAGE. Fractions containing proteins of interest were concentrated using a GE Healthcare 100-kDa MWCO Vivaspin 500 concentrator and stored at 4 °C until further use.

*FjoGldLM'*_A, *FjoGldLM'*_B, and *PgiPorLM'*_A proteins were purified as described above. Other proteins expressed from pT12 vectors were purified with the following modifications or with modifications for specific constructs described in Chapter 6. To avoid the need to change buffer systems and bring the protocol in line with that recommended by the manufacturer of the StrepTactin XT Superflow

cartridges, all buffers were replaced with Tris-based systems. PBS-containing buffers were modified by replacing the PBS with Tris-buffered saline (TBS; 100 mM Tris-HCl pH 8.0, 150 mM NaCl). The SEC column was run using TBS with 0.01% LMNG instead of the HEPES-based buffer. LMNG micelles are large enough to be concentrated in a 100-kDa MWCO concentrator and when concentrated can form string-like objects that reduce contrast in cryoEM and interfere with particle picking. To avoid this issue, the concentration of LMNG in the StrepTactin XT elution buffer and SEC buffer was reduced to 0.01% for some purifications, as indicated in Chapter 6. To make it easier to complete the protocol in two days with time to prepare cryoEM grids at the end of the second day, membranes were resuspended and solubilised immediately after harvesting, rather than storing them at 4 °C overnight. Once the StrepTactin XT Superflow cartridges had been washed, they were stored at 4 °C overnight. Bound proteins were then eluted the next morning as described above.

2.7 Electron microscopy

2.7.1 Grid preparation

Some negative stain EM grids were prepared by or with the assistance of Dr Justin Deme (JCD). Some cryoEM grids were prepared by JCD or Dr Emily Furlong (EF).

For negative stain EM, copper grids coated with carbon were glow discharged at 15 mA for 20 s using a Leica EM ACE200 glow discharge unit or Quorum Technologies Q150R ES Dual Carbon/Sputter Coater. 10 µl of protein sample was applied to the grid, left for 2 min then removed by blotting. The grid was washed 3 times by placing the grid on a 30 µl droplet of deionized water for 5 s then blotting. The grid was stained by placing on a 30 µl droplet of 2% w/v uranyl acetate solution for 5 s, blotting, then placing on a second 30 µl droplet of 2% w/v uranyl acetate solution for 2 min before blotting. Grids were imaged at least 12 h after preparation to allow them to dry completely.

For cryoEM, holey carbon-coated grids (Quantifoil 300 mesh, Au 1.2/1.3) were glow discharged as above and mounted in a Vitrobot Mark IV (FEI) set to 100% relative humidity at 4 °C. 4 µl aliquots of

purified protein (usually $A_{280} = 0.2-2.0$ as indicated in later chapters) were applied to the grids and left to adsorb for 10 s, then blotted for 2 s. The ideal blot force can vary depending on both the sample and changes to the equipment set-up caused by user handling, so typically at least 2 grids were prepared per sample at relative blot forces from 0 to -15. After blotting, grids were frozen by plunging into liquid ethane, cooling the sample fast enough to prevent formation of crystalline ice. After freezing, grids were stored in liquid nitrogen until imaging.

2.7.2 Imaging and data collection

Some negative stain EM data samples were imaged by JCD. Negatively stained grids were imaged on a Tecnai T12 transmission electron microscope (TEM) operating at 120 kV at 68,000x magnification.

CryoEM grids were screened to assess protein aggregation state, particle density and ice quality using a Talos Arctica TEM operating at 200 kV with a Falcon 3 Direct Electron Detector. Some screening was carried out by JCD or EF.

All cryoEM data collections were carried out by JCD or Professor Susan Lea (SML). Data were collected using a Titan Krios (FEI) microscope operated at 300 kV and equipped with either a GIF energy filter and K2 Summit detector or a BioQuantum imaging filter and K3 direct detection camera (all Gatan) as indicated for each dataset in later chapters. The K2 Summit detector was used in counting mode at 165,000x magnification with a pixel size of 0.822 Å and a total dose of $48 \text{ e}^- \text{ Å}^{-2}$ over 20 or 32 fractions. With the K3 direct detection camera, the *PgiPorLM'*_C dataset was collected in counted super-resolution mode at 81,000x magnification with a pixel size of 1.068 Å and a total dose of $55.6 \text{ e}^- \text{ Å}^{-2}$ over 40 fractions. Other datasets were collected using the K3 direct detection camera in counted super-resolution mode at 105,000x magnification with a pixel size of 0.832 Å and total doses ranging from 59.1 to $62.4 \text{ e}^- \text{ Å}^{-2}$ over 40 fractions.

*FjoGldLM'*_A, *FjoGldLM'*_B, *PgiPorLM*, and *PgiPorLM'*_A datasets were primarily processed by JCD and SML. Data processing strategies changed over the course of this project as different software became available. For the *FjoGldLM'*_A, *FjoGldLM'*_B, *PgiPorLM*, and *PgiPorLM'*_A datasets, motion correction and dose weighting were performed using MotionCor implemented in RELION 3.0. Contrast transfer functions were then calculated using CTFFIND4 (Zivanov *et al*, 2019; Rohou & Grigorieff, 2015). For other datasets, motion correction, dose weighting, and contrast transfer functions were calculated in real time using SIMPLE 3.0. Particle picking and initial particle extraction were carried out in SIMPLE 2.5 or 3.0 (Reboul *et al*, 2018; Caesar *et al*, 2020). Subsequent data processing depended on the particular dataset and is described with each data collection in later chapters.

2.7.3 Model building and refinement

The *FjoGldLM'* model was built by JCD and SML. JCD advised on and assisted with the building of other models. The particular building strategy depended on each dataset as described in later chapters. In general, initial models from crystal structures or homology models were rigid body fitted into cryoEM density maps using *Coot*, then additional residues built *de novo* as necessary using *Coot* (Emsley *et al*, 2010). Multiple rounds of rebuilding in both globally sharpened and local-resolution-filtered maps and real-space refinement in *PHENIX* using secondary structure, rotamer, and Ramachandran restraints lead to the final atomic models (Afonine *et al*, 2018; Liebschner *et al*, 2019). Models were validated using MolProbity (Williams *et al*, 2018) and further analysis was carried out using the Consurf (Landau *et al*, 2005; Ashkenazy *et al*, 2016) and PDBe-PISA servers (<http://www.ebi.ac.uk/pdbe/pisart.html>, (Krissinel & Henrick, 2007)). Structures were aligned using the CCP4i program LSQKAB function (Winn *et al*, 2011). Figures were prepared using the UCSF ChimeraX (Pettersen *et al*, 2021) and PyMOL (PyMOL Molecular Graphics System v2.3.3, Schrödinger) programs.

2.8 *F. johnsoniae* gliding motility assays

To assess gliding motility by colony spreading on agar, cells were grown overnight in PY2 medium at 30 °C with shaking. Cells were washed once in fresh PY2 medium, then resuspended in PY2 medium to $OD_{600} = 0.1$. 2 μ l of cells were spotted onto PY2 + 1% agar plates. Plates were incubated at 25 °C for 48 h then imaged with a Zeiss AXIO Zoom MRm CCD camera and Zeiss software (ZenPro 2012, v 1.1.1.0).

To observe individual cells gliding on glass, cells were grown overnight at 30 °C with shaking in MM then diluted 40-fold into fresh medium and grown at 25 °C with 50 rpm shaking. Aliquots were added to tunnel slides and incubated for 5 min before washing twice with 100 μ l fresh medium. Slides were then imaged in phase contrast using a Zeiss Axiovert 200 microscope fitted with a Photometrics Coolsnap HQ camera using a $\times 63/1.4$ numerical aperture Plan-Apochromat lens. Tif images were captured using Metamorph software (Molecular Devices) and then modified in ImageJ (Schneider *et al*, 2012). Videos were collected at 20 frames per second with a 15-ms exposure time using 2×2 binning.

Experiments to observe the propulsion of individual HaloTag-SprB fusion proteins were carried out by Andreas Kjær. The following text is adapted from (Hennell James *et al*, 2021). Strains expressing the HaloTag-SprB fusion were grown overnight in CYE medium, diluted 40-fold into fresh medium and grown at 30 °C with 180 rpm shaking to early stationary phase ($OD_{600} = 0.8-1.0$). The HaloTag-SprB fusion protein was labelled by mixing 1 ml of this culture with 1 μ l of a 10 μ M stock solution of PA Janelia Fluor 646 HaloTag ligand (Grimm *et al*, 2016, 2017) in dimethyl sulfoxide and then culturing for a further 30 min. Cells were collected by centrifugation at 9,000g for 3 min and washed three times with 650 μ l of PY2 medium, and then 2 μ l of cells were spotted onto a 1% agarose pad containing 50% PY2 medium. Fluorescence images were acquired at 25 °C using a Nanoimager (Oxford Nanoimaging) equipped with 405-nm and 640-nm 1 W diode-pumped solid-state lasers.

Optical magnification was provided by a $\times 100$ oil-immersion objective (Olympus, 1.4 numerical aperture) and images were acquired using an ORCA-Flash4.0 V3 CMOS camera (Hamamatsu). Cells were imaged using a 20-ms exposure time, with the 405-nm photoactivation laser at 10% power and the 640-nm measurement laser at 20% power. Different strains were imaged at different strobing frequencies to accommodate the large differences in adhesin velocities between the different *gldLM* mutants. Specifically, strain Ak_73 (wild-type *gldLM*) was imaged without strobing, strains Ak_196 (*gldL-Y13A*), Ak_197 (*gldL-H30A*) and Ak_199 (*gldM-Y17F*) were imaged with a 60-ms dark interval and Ak_198 (*gldL-K27A*) was imaged with a 480-ms dark interval. Fluorescent foci were tracked using the Nanoimager software. Helical tracks were exported as csv files and the average frame to frame displacement was calculated. The resulting data were analysed using GraphPad Prism v.7.03 for Windows, using a Kruskal–Wallis test followed by Dunn’s multiple comparisons test. The data were plotted using Prism.

2.9 *F. johnsoniae* type IX secretion assays and assessment of GldLM stability

Assessments of *Fjo*GldLM stability and analysis of secretomes by SDS-PAGE and immunoblotting were all carried out using the same cell samples. Cells were grown in CYE medium for 22 h, reaching $OD_{600} = 5.5$ - 6.5 .

To assess *Fjo*GldLM stability, 1 ml of each cell culture was centrifuged at $9,000g$ for 10 min at $4^{\circ}C$ to collect cells. Cells were washed once in phosphate-buffered saline (PBS), then resuspended in PBS to $OD_{600} = 5.0$. Samples were diluted tenfold in PBS and analysed by SDS-PAGE followed by immunoblotting with α GldL_{cyto} or α GldM_{peri} primary antibodies diluted 1:5,000 (Davids Biotechnologie (Hennell James *et al*, 2021)) and anti-rabbit IgG HRP conjugate (Promega) secondary antibodies diluted 1:5,000.

To analyse secretomes, 10 ml culture samples were centrifuged at 9,000g for 25 min at 4 °C to remove cells. Supernatants were filtered with a 0.22 µm syringe filter (Millipore) and then centrifuged at 150,000g for 75 min at 4 °C to remove extracellular vesicles. Supernatants were concentrated to 400 µl using a 10-kDa MWCO Amicon Ultra-15 centrifugal filter unit. For secretome analysis, 14 µl of the concentrated supernatant was subjected to SDS-PAGE with proteins visualised by Coomassie blue staining. The presence of ChiA was assessed by immunoblotting using ChiA primary antibodies diluted 1:1,000 (Lauber *et al*, 2018) and anti-rabbit IgG HRP conjugate secondary antibodies diluted 1:5,000.

To quantitatively measure chitinase secretion, cells were grown in Motility Medium for 15 h, reaching OD₆₀₀ = 0.75-1.25. 5 ml culture samples were centrifuged at 3,720g for 10 min to remove cells. Chitinase activity in the resulting supernatants was determined using a fluorometric chitinase assay kit (Sigma) with the synthetic substrate 4-methylumbelliferyl-N,N'-diacetyl-β-d-chitobioside. Statistical analysis of the results was carried out using GraphPad Prism v.7.03 for Windows (GraphPad Software, www.graphpad.com). Outliers were identified and removed using the ROUT test with Q = 1%. Datasets were compared with wild type using a Kruskal–Wallis test followed by Dunnett's multiple comparisons test.

3 Purification and characterisation of the full-length *Porphyromonas gingivalis* PorLM complex

3.1 Introduction

To understand how the GldLM complex functions in the T9SS and gliding motility I set out to recombinantly purify and then characterise GldL and GldM homologues. I initially attempted to purify and characterise full length GldL and GldM proteins, as these would provide the most physiological context for structures. To do this I targeted the proteins from the model organisms *Flavobacterium johnsoniae* and *Porphyromonas gingivalis*, and from the thermophilic bacterium *Thermonema rossianum*. Most work to characterise the T9SS and gliding motility has been carried out in *F. johnsoniae* and *P. gingivalis* (McBride, 2019; Nakayama, 2015), so any findings from structural or biophysical characterisation of their proteins could be interpreted in the context of existing knowledge. Proteins from thermophilic species have been useful in structural biology as they have evolved to be thermostable and usually have fewer highly flexible regions. This often makes them easier to crystallise for x-ray diffraction and can reduce conformational heterogeneity in cryoEM, leading to higher resolution structures. I chose to recombinantly express the targeted proteins in *E. coli* as this usually leads to higher purification yields than purification from the native organism, although some recent work has taken advantage of the low sample requirements of cryoEM to solve structure of proteins purified from *F. johnsoniae* and other species at their native expression level (Lauber *et al*, 2018; Sun *et al*, 2018).

3.2 Construct design and expression screening

To purify recombinant GldL and GldM homologue proteins, I designed the pRHJ004 co-expression vector (**Figure 3-1**). This vector was based on pETDuet-1, where two separate transcripts can be produced at the same time from separate IPTG-inducible T7 polymerase promoters. Proteins encoded under the first promoter were fused to a C-terminal tobacco etch virus (TEV) protease cleavage site followed by super-folder GFP (sfGFP) and a His₈ tag. Proteins encoded under the second

promoter were fused to a C-terminal Twin-Strep II tag. Fluorescent proteins are often fused to recombinant membrane proteins to make it easier to assess how much recombinant protein is present and which fractions it is associated with during purification. sfGFP can fold in the periplasm whilst GFP cannot (Pédélecq *et al*, 2006). I therefore used sfGFP as the C-terminus of GldM is predicted to be in the periplasm (Vincent *et al*, 2017). Having two affinity tags makes it possible to enrich for complexes which contain both proteins of interest by using tandem affinity purification. I constructed six expression vectors expressing the full-length GldL and GldM homologues from *F. johnsoniae*, *T. rossianum* and *P. gingivalis* in both promoter site combinations. These vectors are summarised in Table 3-1, for a full description of the plasmids and their cloning strategies see Table 2-5.

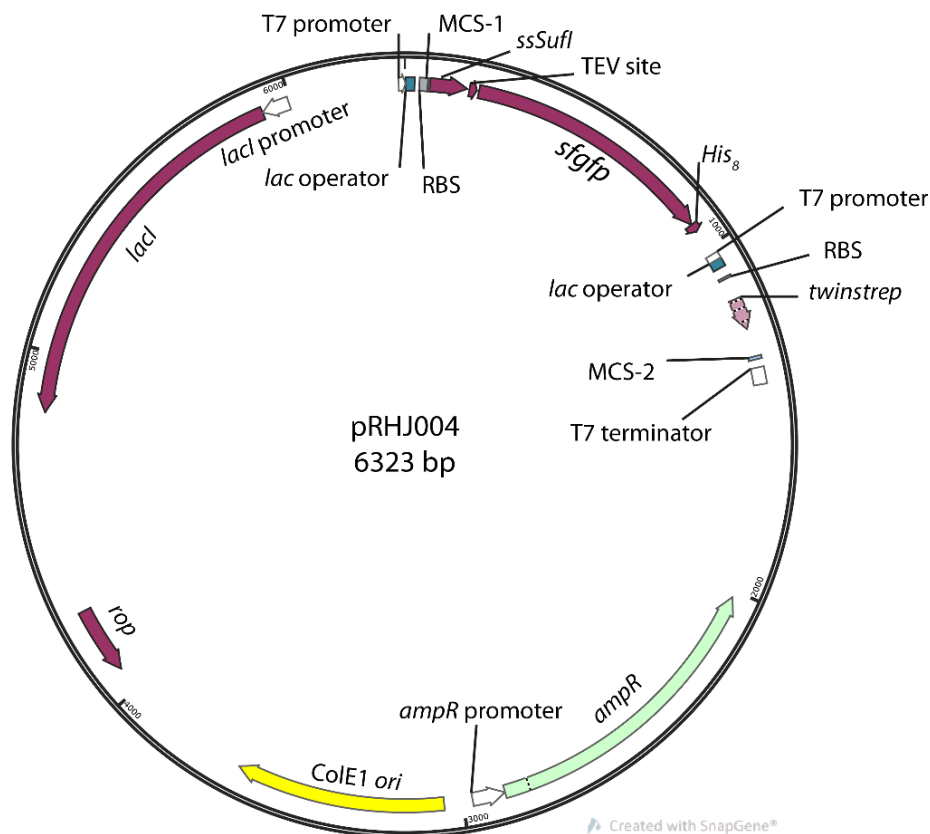


Figure 3-1 Map of the empty expression vector pRHJ004, based on pETDuet-1. *ampR* encodes a beta-lactamase enzyme that confers resistance to ampicillin. MCS – multiple cloning site. RBS – ribosome binding site. *lacI* encodes the *lac* repressor, which binds to the *lac* operators and inhibits transcription from the nearby T7 promoter sites in the absence of lactose or isopropyl- β -D-thiogalactopyranoside. *rop* encodes the Rop protein which limits the copy number of ColE1 plasmids.

To identify which constructs were best-expressed, and under what conditions the most protein was produced, I screened for expression in small-scale cultures under a range of conditions. As well as my six expression vectors, I screened for expression in four different *E. coli* strains. BL21 (DE3) is a commonly used protein expression strain, whilst BL21 Star (DE3), C41 (DE3) and Mt56 (DE3) are all derived from BL21 (DE3) with mutations that can improve recombinant protein yields. BL21 Star (DE3) lacks a functional RNase E enzyme so mRNA transcripts have a longer half-life and can be translated more times into more protein. C41 (DE3) has a mutation in the *lacUV5* promoter that controls expression of T7 RNA polymerase, whilst Mt56 (DE3) has a variant T7 RNA polymerase that has reduced affinity for the T7 promoter. Both these changes lead to a lower rate of mRNA

production and so slower recombinant protein production. Slower production of membrane proteins can in fact lead to higher yields, as it reduces the risk of overwhelming the membrane protein insertion machinery of the cell and thus alleviates toxicity. I also screened for expression in three different growth conditions, either induction with 1 mM IPTG for 4 h at 37 °C or induction with 0.1 mM IPTG for 16 h at 24 °C or 18 °C. Cells were grown in 24-well plates, the expression conditions of each plate are listed in Table 3-2.

Table 3-1 Expression vectors based on pRHJ004 for co-expression of *gldL* and *gldM* homologues.

Plasmid Number	Description
pRHJ006	pRHJ004 with <i>P. gingivalis</i> <i>porL</i> in MCS1 and <i>porM</i> in MCS2.
pRHJ103	pRHJ004 with <i>T. rossianum</i> <i>gldM</i> in MCS1 and <i>gldL</i> in MCS2.
pRHJ106	pRHJ004 with <i>F. johnsoniae</i> <i>gldM</i> in MCS1 and <i>gldL</i> in MCS2.
pRHJ108	pRHJ004 with <i>T. rossianum</i> <i>gldL</i> in MCS1 and <i>gldM</i> in MCS2.
pRHJ109	pRHJ004 with <i>F. johnsoniae</i> <i>gldL</i> in MCS1 and <i>gldM</i> in MCS2.
pRHJ110	pRHJ004 with <i>P. gingivalis</i> <i>porM</i> in MCS1 and <i>porL</i> in MCS2.

Table 3-2 Conditions used to screen for protein expression from pRHJ004-based vectors. Cells were grown at 37 °C with shaking for the length of the primary growth time. The indicated concentration of IPTG was added and cells were then grown at the expression temperature for the indicated expression time.

Condition	Primary growth time (h)	Expression time (h)	Expression temperature (°C)	IPTG added (mM)
Plate A	3	4	37	1
Plate B	8	16	24	0.1
Plate C	8	16	18	0.1

To assess which plasmid/strain/condition combination produced the most protein, I measured the fluorescence of the sfGFP fused to the protein cloned in MCS1 as a measure of recombinant protein production. This value was normalised to the cell density of each culture measured as OD₆₀₀ to get a measure of fluorescence per cell (**Figure 3-2a**). pRHJ006 (*P. gingivalis* PorL-TEV-sfGFP-His₈ and PorM-TS) generally produced the highest ratio of fluorescence to OD₆₀₀, indicating high expression of at least PorL-sfGFP-His₈. The combination of strain Mt56 (DE3) and overnight induction at 24 °C led to

almost twice as much normalised fluorescence as the other combinations for the plasmids pRHJ006, pRHJ109 (*F. johnsoniae* GldL-TEV-sfGFP-His₈ and GldM-TS), and pRHJ110 (*P. gingivalis* PorM-TEV-sfGFP-His₈ and PorL-TS) but not for the other plasmids. Beyond this effect, there was no apparent trend in the results for different strain and expression condition combinations.

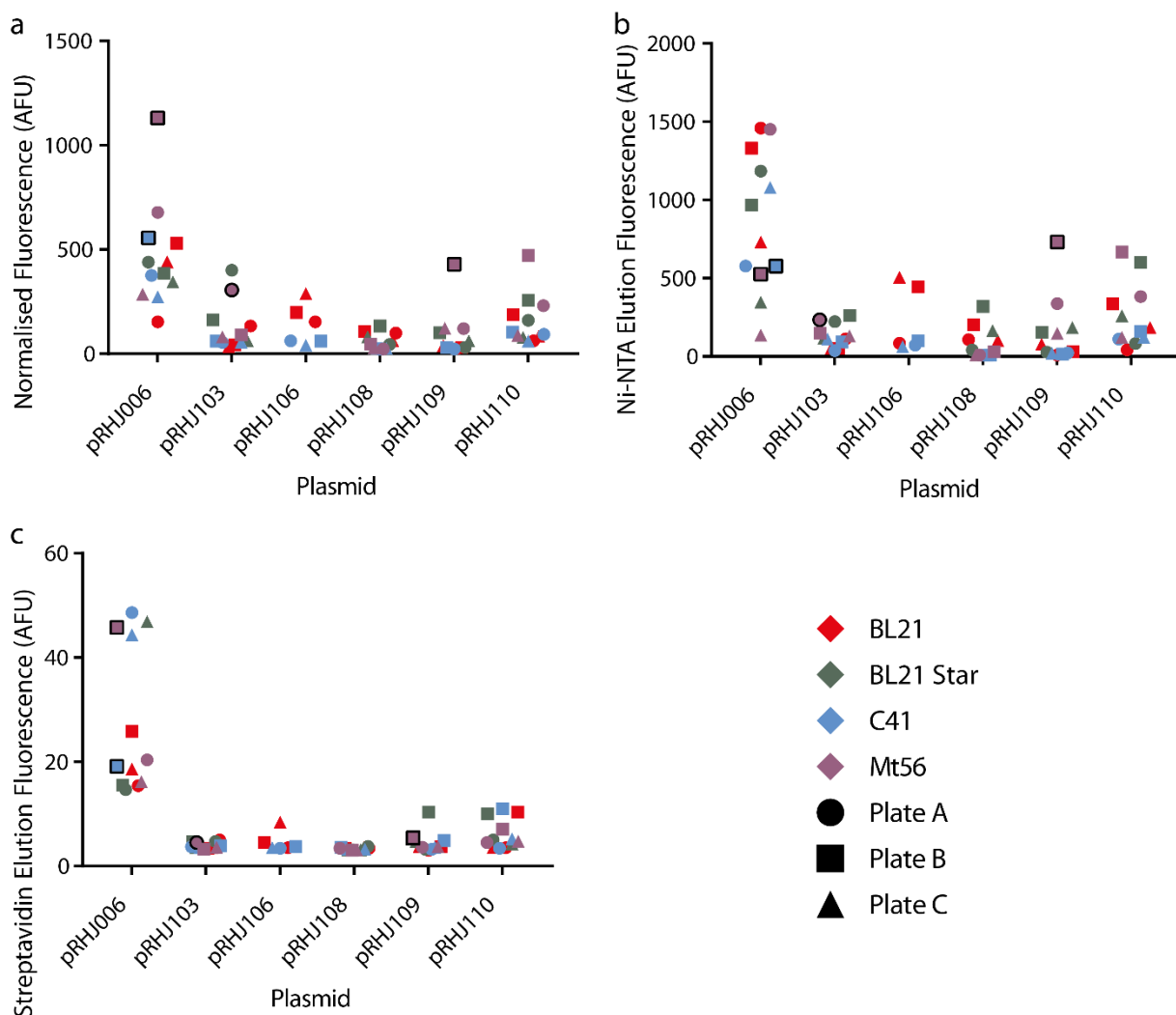


Figure 3-2 Fluorescence measurements from the pRHJ004 expression screen. The expression strain is indicated by symbol colour: Red – BL21 (DE3), Green = BL21 Star (DE3), Blue – C41 (DE3), Purple – Mt56 (DE3). Expression conditions are indicated by symbol shape (see Table 3-1): Circles – Plate A, Squares – Plate B, Triangles – Plate C. Symbols for the combinations used in later purifications are outlined in black. **a** Fluorescence measurements for expression screen cultures normalized against OD₆₀₀ (Raw Fluorescence/OD₆₀₀). **b** Fluorescence measurements for Ni-NTA elution fractions. **c** Fluorescence measurements for Streptavidin elution fractions.

To confirm that this fluorescence I measured for the cells was due to GldLM/PorLM complexes and not just free proteolyzed sfGFP I lysed the cells and used magnetic Ni-NTA or Streptavidin-coated beads to purify His₈- and TwinStrep-tagged proteins, respectively. I then measured the fluorescence in these elution fractions. Ni-NTA elution fractions showed a very similar pattern to the normalised cellular fluorescence values (**Figure 3-2b**), which makes sense as both sfGFP fusion proteins and proteolyzed free sfGFP should have His₈-tags and be purified by Ni-NTA beads. The fluorescence values for Streptavidin elution fractions were much lower than those for Ni-NTA fractions (**Figure 3-2c**). This follows from the fact that sfGFP would only be purified as a fusion protein in complex with the TwinStrep-tagged partner protein, but the much lower fluorescence values for Streptavidin elutions compared to Ni-NTA elutions may indicate that the Ni-NTA elutions contained large amounts of proteolyzed free sfGFP or that the recombinant proteins do not form a complex. Despite the much lower fluorescence values, I saw a similar pattern to the cellular fluorescence, whereby pRHJ006-carrying cells produced the most fluorescence. Streptavidin elutions from C41 cells carrying pRHJ006 induced for 4 h at 37 °C produced more fluorescence than those from Mt56 cells carrying pRHJ006 induced overnight at 24 °C, but only by a small amount. I then analysed the Streptavidin elution fractions by SDS-PAGE with Stain-Free gels and UV fluorescence imaging. The protein bands visualised by UV fluorescence were very faint, so it was not possible to confidently interpret the gels. The combination of pRHJ006 (*P. gingivalis* PorL-TEV-sfGFP-His₈ and PorM-TS) with expression in Mt56 (DE3) cells and induction at 24 °C overnight produced by far the most cellular fluorescence so I selected this combination for large scale purification. I also attempted larger scale purifications with the combinations pRHJ006 + C41 (DE3) + induction overnight at 24 °C, pRHJ103 (*T. rossianum* GldM-TEV-sfGFP-His₈ and GldL-TS) + Mt56 (DE3) + induction for 4 h at 37 °C, and pRHJ109 (*F. johnsoniae* GldL-TEV-sfGFP-His₈ and GldM-TS) + Mt56 (DE3) + induction overnight at 24 °C.

3.3 Purification of *PgiPorLM*

Expression screening suggested that pRHJ006, expressing *PgiPorL*-sfGFP-His₈ and *PgiPorM*-TS produced the most protein. I initially attempted to purify *PgiPorL*-sfGFP-His₈ and *PgiPorM*-TS proteins expressed from pRHJ006 in either C41 or Mt56 cells, induced with 0.1 mM IPTG for 16 h at 24 °C. However, after cell breakage, the pellet after centrifugation at 27,200g to remove cellular debris was bright green. This suggested that a large amount of sfGFP, and probably therefore the fusion protein, was either in inclusion bodies or otherwise associated with the debris. Nevertheless, the supernatant was also green, so I continued with protein purification. I initially solubilised membrane proteins in the detergent dodecyl maltoside (DDM) as experience in the Lea laboratory shows that it efficiently solubilises most proteins. After tandem Ni-NTA and StrepTactin XT affinity purification, full length PorL and PorM were detectable in the elution fraction from the second column by SDS-PAGE, but PorL was also detectable in the flowthrough fraction (Figure 3-3). This could occur if either *PgiPorL* had been produced in excess over the amount needed to form complexes with *PgiPorM* or if not all *PgiPorL* and *PgiPorM* monomers had successfully formed heteromeric complexes.

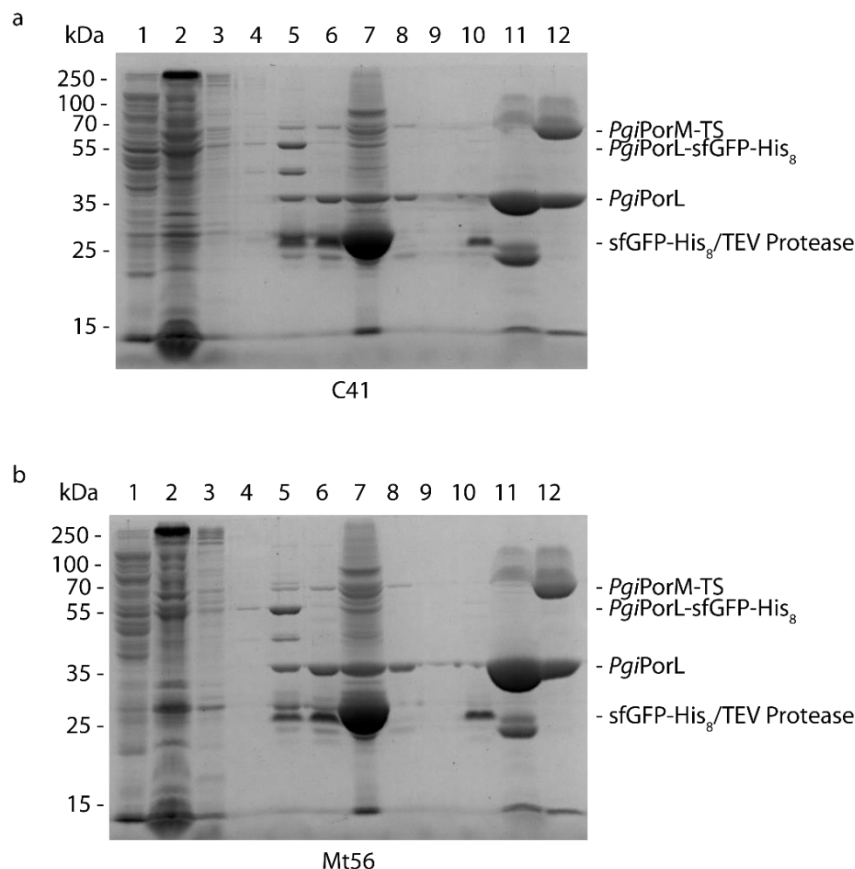


Figure 3-3 Initial purification of PgiPorLM in DDM from C41 (a) and Mt56 (b) cells. 1 – Ni-NTA input; 2 – Ni-NTA flowthrough; 3 – Ni-NTA 20 mM imidazole wash; 4 – Ni-NTA 40 mM imidazole wash; 5 – Ni-NTA elution; 6 – soluble material post-dialysis and TEV digestion; 7 – precipitated material post-dialysis and TEV digestion; 8 – reverse Ni-NTA flowthrough; 9 – reverse Ni-NTA wash; 10 – reverse Ni-NTA elution; 11 – StrepTactin XT flowthrough; 12 – StrepTactin XT elution.

I pooled the StrepTactin XT elution fractions purified from both C41 and Mt56 cells before injecting them on an S200 10/300 Increase column (Figure 3-4d-f). Most of the protein ran in the void volume of the column, suggesting that either it was aggregated or that the *PgiPorLM* complex is larger than can be resolved by an S200 10/300 Increase column. I concentrated fractions containing *PgiPorL* and *PgiPorM* and reinjected them on a Superose 6 10/300 Increase column which can resolve complexes up to 5 MDa in size. This separated the proteins into two main peaks. Both peaks contained both *PgiPorL* and *PgiPorM*. Given that detergent micelles can be of variable sizes it is not possible to accurately estimate the mass of a membrane protein complex from a SEC trace so I still could not be

sure of the stoichiometry of the protein complex. Given it was a monodisperse peak I presumed that Peak 2 contained *PgiPorLM* complexes rather than separate *PgiPorL* and *PgiPorM* proteins. Peak 1, which partly overlapped with the void volume may have contained higher order associations of *PgiPorLM* complexes or aggregates. Based on the results of cryo-EM imaging (Figure 3-5d-f) and the fact that in future purifications when I completed the purification more quickly this peak was smaller it seems more likely that these were aggregates.

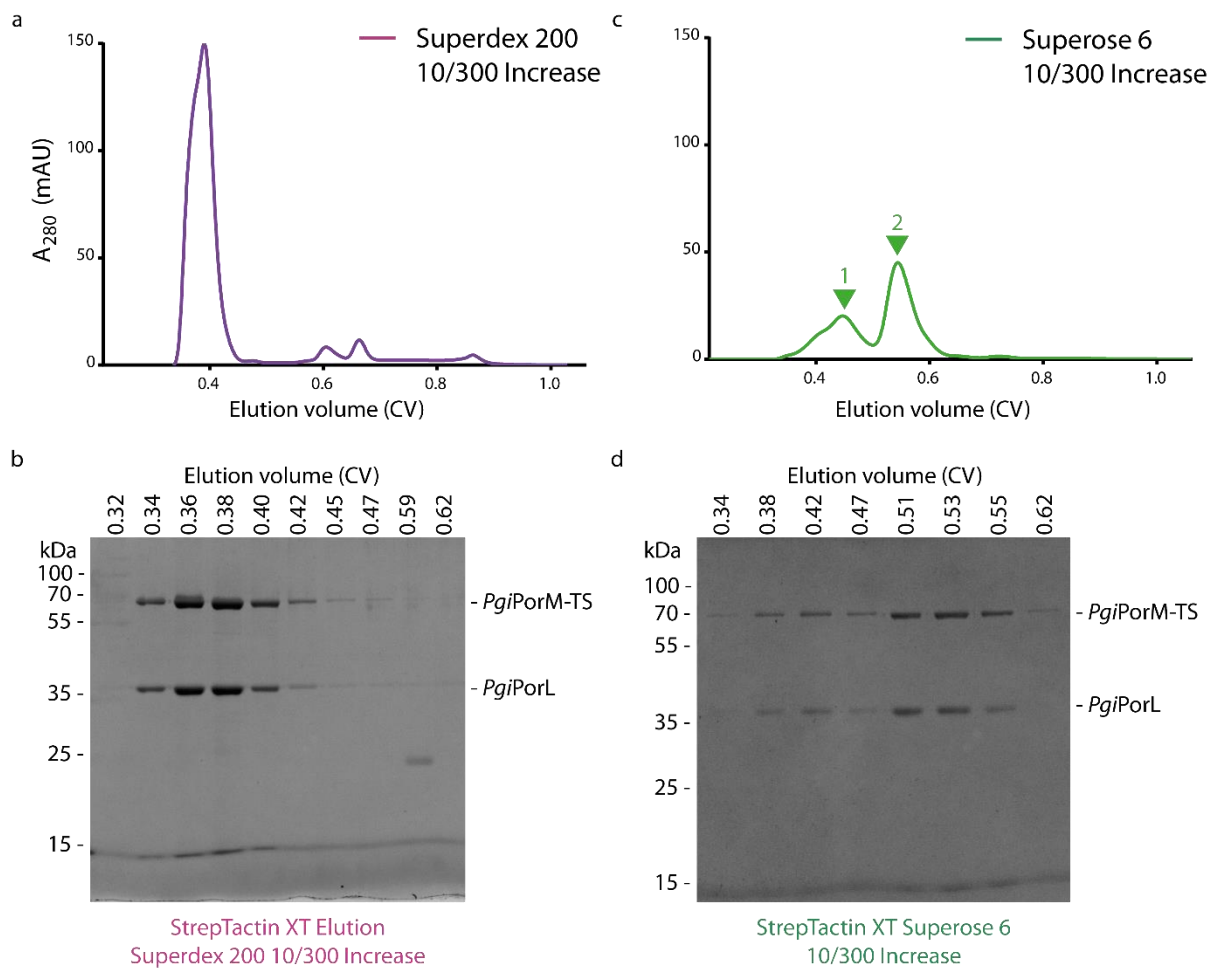


Figure 3-4 SEC from initial purification of *PgiPorLM* in DDM. **a** Superdex 200 10/300 Increase SEC traces for the pooled StrepTactin XT elution fractions from C41 and Mt56 cells. **b** SDS-PAGE gel of selected fractions from the Superdex 200 10/300 Increase trace shown in **a**. **c** Superose 6 10/300 Increase SEC trace for the main peak fractions from the Superdex 200 10/300 Increase SEC trace shown in **a**. Arrowheads indicate peaks analysed by negative stain electron microscopy. **d** SDS-PAGE gel of selected fractions from the Superose 6 10/300 Increase trace shown in **c**. The 0.51-0.55 CV fractions were analysed by cryoEM (Figure 3-5c).

3.4 Initial characterisation of *PgiPorLM*

Having purified *PgiPorLM*, I next set out to characterise my sample using electron microscopy. Dr Justin Deme (JCD) imaged samples from both the peaks from the Superose 6 Increase fractionation of the StrepTactin XT elutions by negative stain electron microscopy (Figure 3-5a-b). Peak 1 contained a lot of aggregates, with some individual particles visible. These particles resembled a piston with the head presumably corresponding to the micelle and transmembrane bundle, and the shaft corresponding to the extended PorM periplasmic region seen in the crystal structure published by Sato *et al.* (Sato *et al.*, 2020). Peak 2 contained far fewer aggregates and many more individual particles with the same piston shape. I concentrated the 0.51-0.55 CV fractions to 2.3 mg ml⁻¹ and JCD prepared cryoEM grids from this sample (Figure 3-5d). Imaging of these grids showed similar particles to those seen by negative stain as well as some particles without the “shaft”. These particles could be empty micelles, PorL alone or top views where the shaft is not visible. Like the PorL-only fractions, the particle density was too low for data collection and aggregated protein could be seen at the edges of the grid holes and adhered to the carbon support.

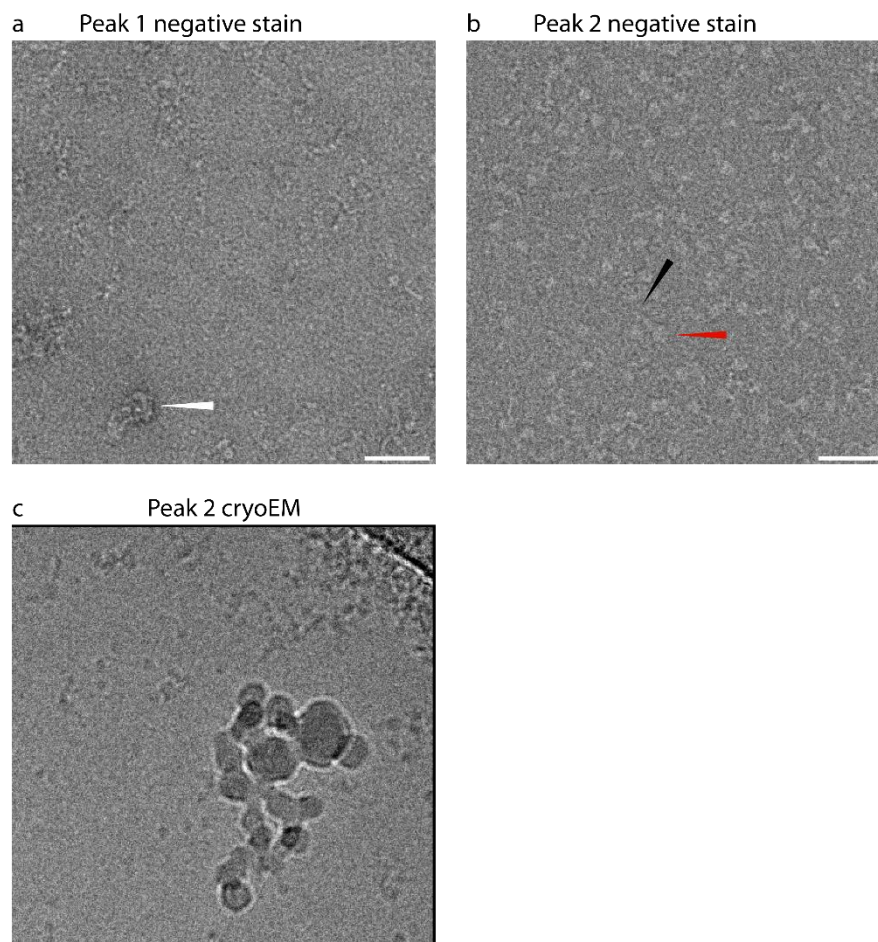


Figure 3-5 Negative stain and cryo-EM images from the initial *PgiPorLM* purification (Figure 3-4d). Scale bars 500 Å. **a** Negative stain image of Peak 1 from the Superose 6 10/300 Increase SEC of pooled StrepTactin XT elutions. Arrowhead indicates an aggregate. **b** Negative stain image of Peak 2 from the Superose 6 10/300 Increase pooled StrepTactin XT elutions. Arrowheads indicate the “head” (red) and “shaft” (black) of one particle. **c** Cryo-electron micrograph of Peak 2 from the Superose 6 10/300 Increase pooled StrepTactin XT elutions.

Solving the structure of the PorLM construct would require a higher particle density on cryoEM grids than I had achieved with my first purification. Experience in the cryoEM field shows that proteins solubilised in DDM tend to adhere to the carbon support of cryoEM grids rather than entering the holes and being frozen in vitreous ice (Cheung *et al*, 2013). I first attempted to overcome this issue by purifying protein from a larger volume of cells (16 l of Mt56 cells processed as described above), in the hopes of reaching a higher protein concentration to saturate the carbon surface and force proteins into the ice (Figure 3-6a-b). I was able to concentrate a sample of PorLM protein to 6.1 mg

ml⁻¹ and JCD made cryoEM grids from this sample. Although this concentration did saturate the carbon surface, with better particle density in the ice, the protein also aggregated at this concentration (Figure 3-6c). Most of the aggregates seemed to consist of groups of PorLM particles interacting via the tips of the “shaft”. Data collection was not possible with this aggregated sample.

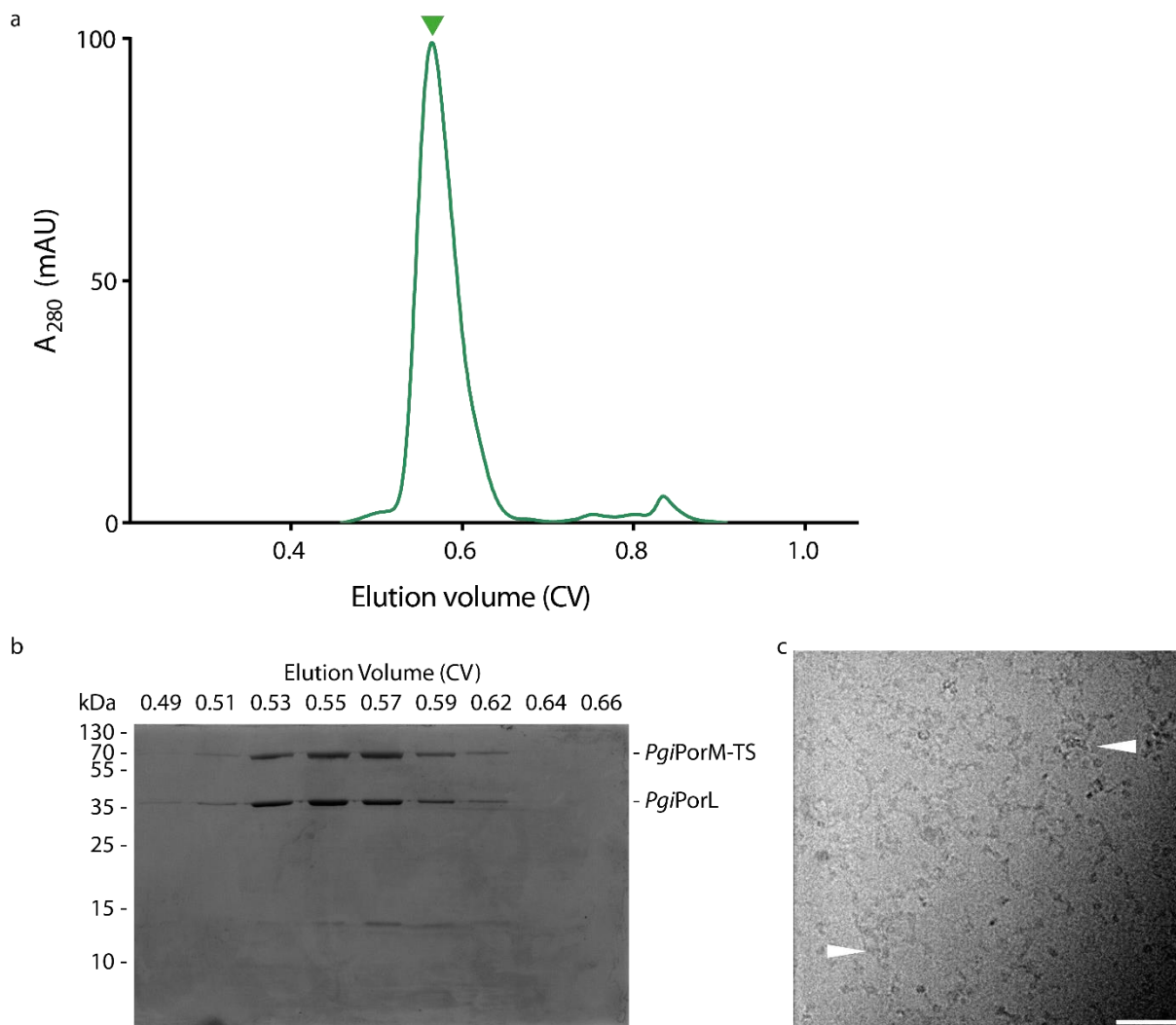


Figure 3-6 Exemplar purification of *PgiPorLM* in DDM from 16 l of Mt56 cells. a SEC trace from a Superose 6 10/300 Increase column. Arrowhead indicates the peak analysed by cryoEM. **b** SDS-PAGE analysis of fractions from **a**. **c** Cryo-electron micrograph of 0.53-0.57 CV fractions concentrated to 6.1 mg ml⁻¹. Arrowheads indicate aggregates. Scale bar 500 Å.

3.5 Purification of *PgiPorLM* in LMNG

As purifying *PgiPorLM* in DDM did not lead to producing good quality cryoEM grids, I instead attempted to purify *PgiPorLM* complexes solubilised in the detergent lauryl maltose neopentyl glycol (LMNG). Experience in the Lea laboratory shows that samples solubilised in LMNG are less prone than samples solubilised in DDM to adhering to the carbon support of cryoEM grids. As well as changing the detergent used, I reversed the order of the StrepTactin XT and Ni-NTA cartridges. Strep tag purification is generally more specific than Ni-NTA purification so this should reduce the presence of non-specific contaminants early in the purification process. It should also have prevented purification of free sfGFP-His₈ or *PgiPorL*-sfGFP-His₈ by itself but may have allowed purification of *PgiPorM*-TS by itself. My initial purification in LMNG produced a good amount of protein (Figure 3-7), but when I repeated the purification to produce more protein to optimise grid-making I had difficulties producing large amounts of protein when solubilising in LMNG. LMNG usually solubilises less protein material than DDM, and my purification yields were already quite low, so it seems that the less-efficient solubilisation pushed the yields over a tipping point below a useful amount. Although I could use lower protein concentrations to make grids with protein solubilised in LMNG than in DDM, I still needed to concentrate the protein at least 20-fold. As LMNG micelles have a molecular mass of 80-100 kDa, empty detergent micelles in the buffer can become concentrated in the 100 kDa MWCO concentrators I was using to concentrate my protein. At high concentrations LMNG micelles merge into string-like objects that reduce image contrast and interfere with proper particle picking on cryo-EM grids.

Even when protein purifications produced very low amounts of protein, the cells expressing protein were bright green, so clearly expressed a large amount of sfGFP and presumably a lot of *PgiPorLM*. My Ni-NTA purification steps did purify some free sfGFP, but in all purifications the cellular debris pellet from a low-speed centrifugation was bright green. This was unlikely to be free sfGFP alone,

which should be soluble. *PgiPorLM-sfGFP-His₈* could have entered the cellular debris pellet for three main reasons; because it was in inclusion bodies, because the long PorM periplasmic “stalk” was entangled with the peptidoglycan layer in the *E. coli* periplasm, or because the complexes had formed large higher-order assemblies or aggregates that were easily pelleted by centrifugation. In case the *PgiPorM* “stalk” was enmeshed with the peptidoglycan cell wall, I tried adding additional lysozyme to digest the peptidoglycan after cell lysis. I also tried centrifuging at lower speeds to avoid pelleting large protein complexes. However, even at low speeds there was a large amount of fluorescent material in the cellular debris pellet (**Figure 3-8**).

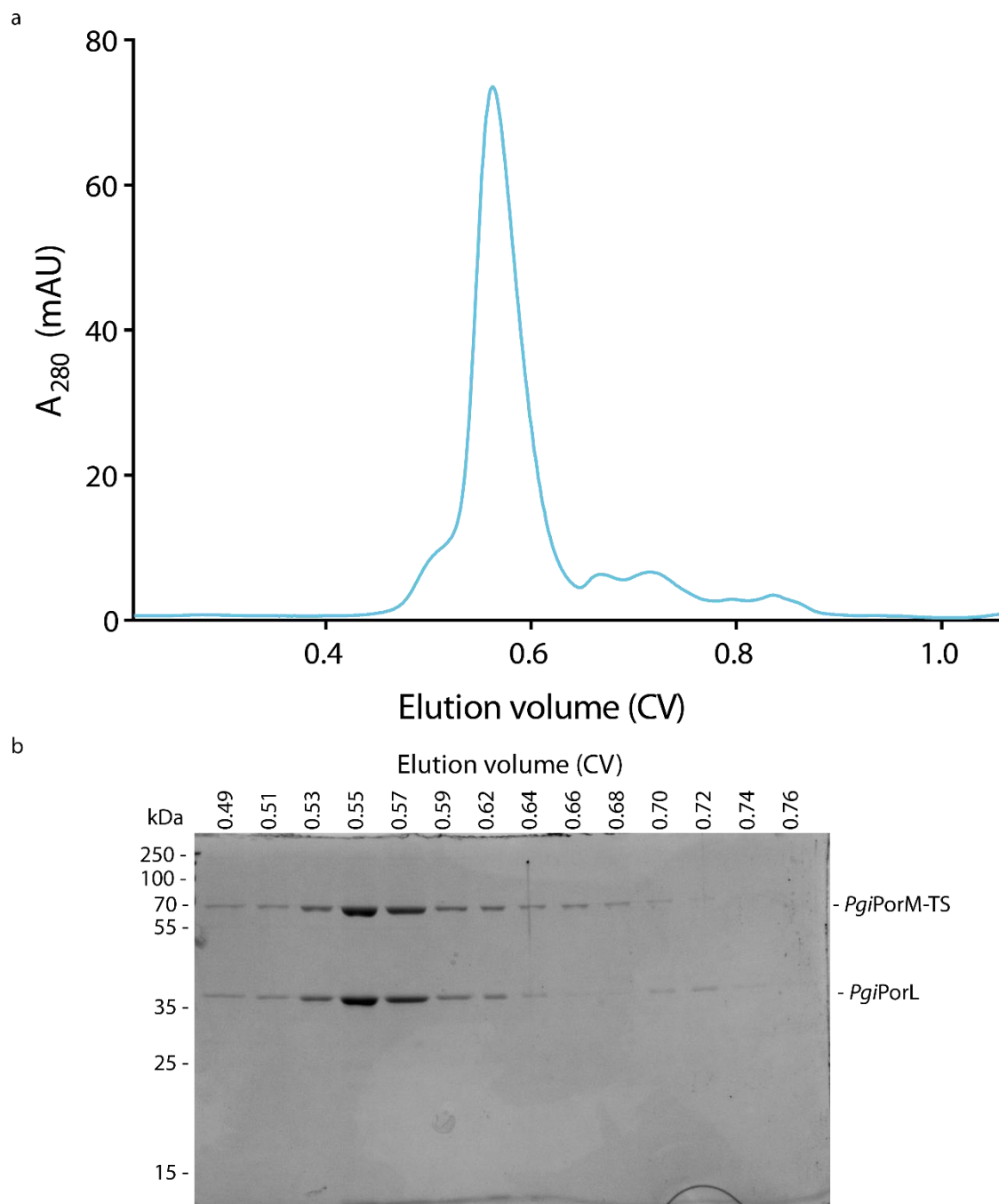


Figure 3-7 Successful purification of *PgiPorLM* in LMNG. **a SEC trace for *PgiPorLM* in LMNG on a Superose 6 10/300 Increase column. **b** SDS-PAGE analysis of fractions from the size exclusion trace.**

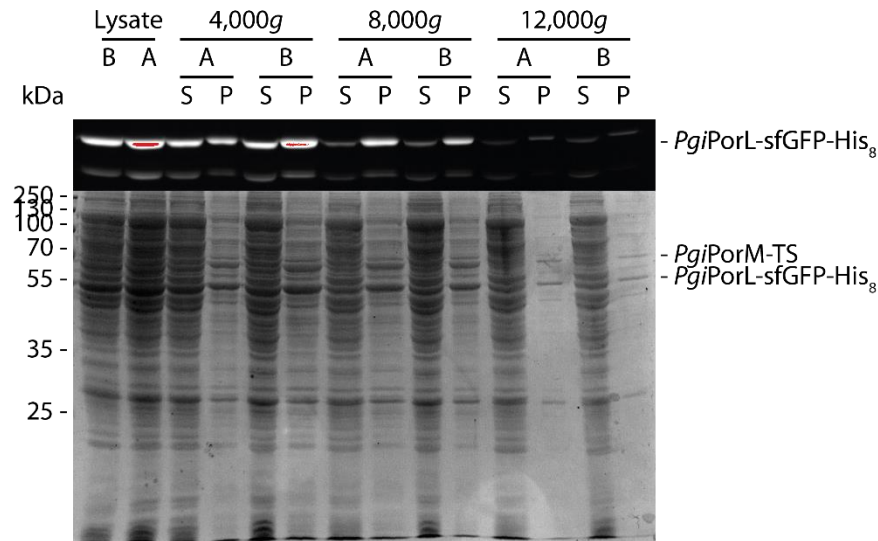


Figure 3-8 Lysis and centrifugation conditions screen. A – additional lysozyme treatment, B – no additional lysozyme treatment, S – supernatant, P – pellet. After cell breakage, samples were successively centrifuged at the indicated *g*-forces. The supernatant from the 4,000*g* centrifugation step was centrifuged at 8,000*g*, and then the supernatant from the 8,000*g* step was centrifuged at 12,000*g*. Loading was balanced for each supernatant/pellet pair, but loading is only approximate across the whole gel. Top panel shows sfGFP fluorescence, bottom panel shows sfGFP staining. Red indicates saturated pixels

Since varying the cell lysis conditions and centrifugation speeds in the purification did not improve the amount of *PgiPorLM* protein in the supernatant fraction after centrifugation to pellet the cellular debris I instead attempted to solubilise and purify protein from the cellular debris pellet. I reasoned that if the protein was folded but entangled with the peptidoglycan it may still be possible to solubilise it and purify protein. I added LMNG to cellular debris at a ratio of 1 g LMNG per 10 g cellular debris, equivalent to my purifications from harvested membranes in which I added 1 g of LMNG per 10 g membrane and then proceeded with the purification as usual. This approach succeeded in purifying more protein than my previous attempts solubilising membranes with LMNG (Figure 3-9). However, the unsolubilised fraction was still bright green and contained large amounts of sfGFP-His₈ and *PgiPorL*-sfGFP-His₈ according to in-gel fluorescence (Figure 3-9a lanes 1-3). This indicates that most protein remained in the cellular debris, either in inclusion bodies or irretrievably enmeshed with

the peptidoglycan cell wall. Nevertheless, I proceeded with characterising the protein I had purified. For future purifications with new constructs, I did not revisit this approach.

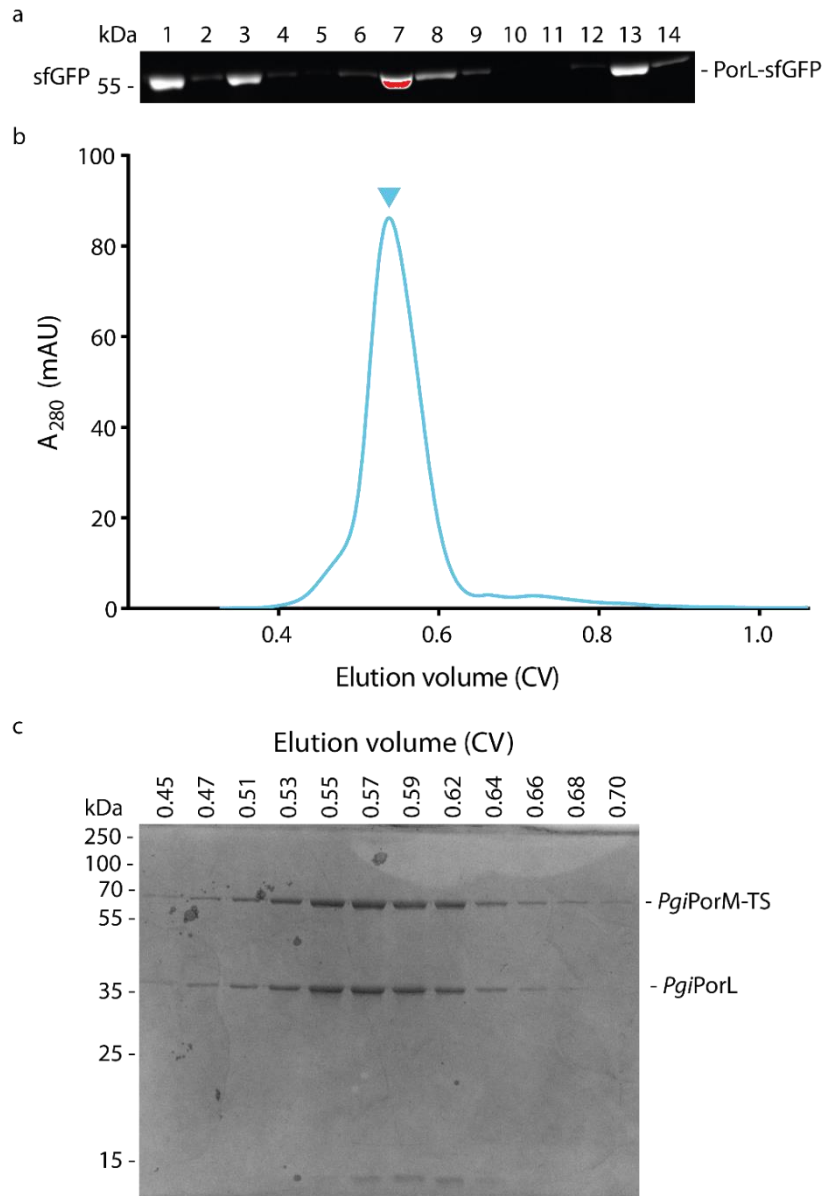


Figure 3-9 Purification of *PgiPorLM* from the cellular debris. **a** sfGFP fluorescence image of SDS-PAGE analysis of purification fractions. 1 – cellular debris, 2 – material solubilised with LMNG from the cellular debris, 3 – unsolubilised material, 4 – StrepTactin XT flowthrough, 5 – StrepTactin XT wash fraction, 6-8 StrepTactin XT elution fractions, 9 – Ni-NTA flowthrough, 10 – Ni-NTA wash with 20 mM imidazole, 11 – Ni-NTA wash with 40 mM imidazole, 12-14 – Ni-NTA elution fractions. Red indicates over-exposed pixels. *PgiPorL*-sfGFP-His₈ was cleaved with TEV protease prior to injecting on the SEC column. **b** SEC trace using a Superose 6 10/300 Increase column for *PgiPorLM* solubilised from cellular debris. Arrowhead indicates the peak analysed by SEC-MALS and cryoEM. **c** SDS-PAGE analysis of fractions from SEC of *PgiPorLM* purified from cellular debris.

3.6 Purification of *FjoGldLM* and *TroGldLM*

I attempted one purification of the most promising plasmid/strain/condition combinations from my expression screen for *FjoGldLM* and *TroGldLM* in DDM. Almost no *TroGldLM* protein was recovered and the yield of *FjoGldLM* was very low. As the yields for these constructs were so poor, I did not pursue them any further and instead made new constructs as described in Chapters 4 and 6.

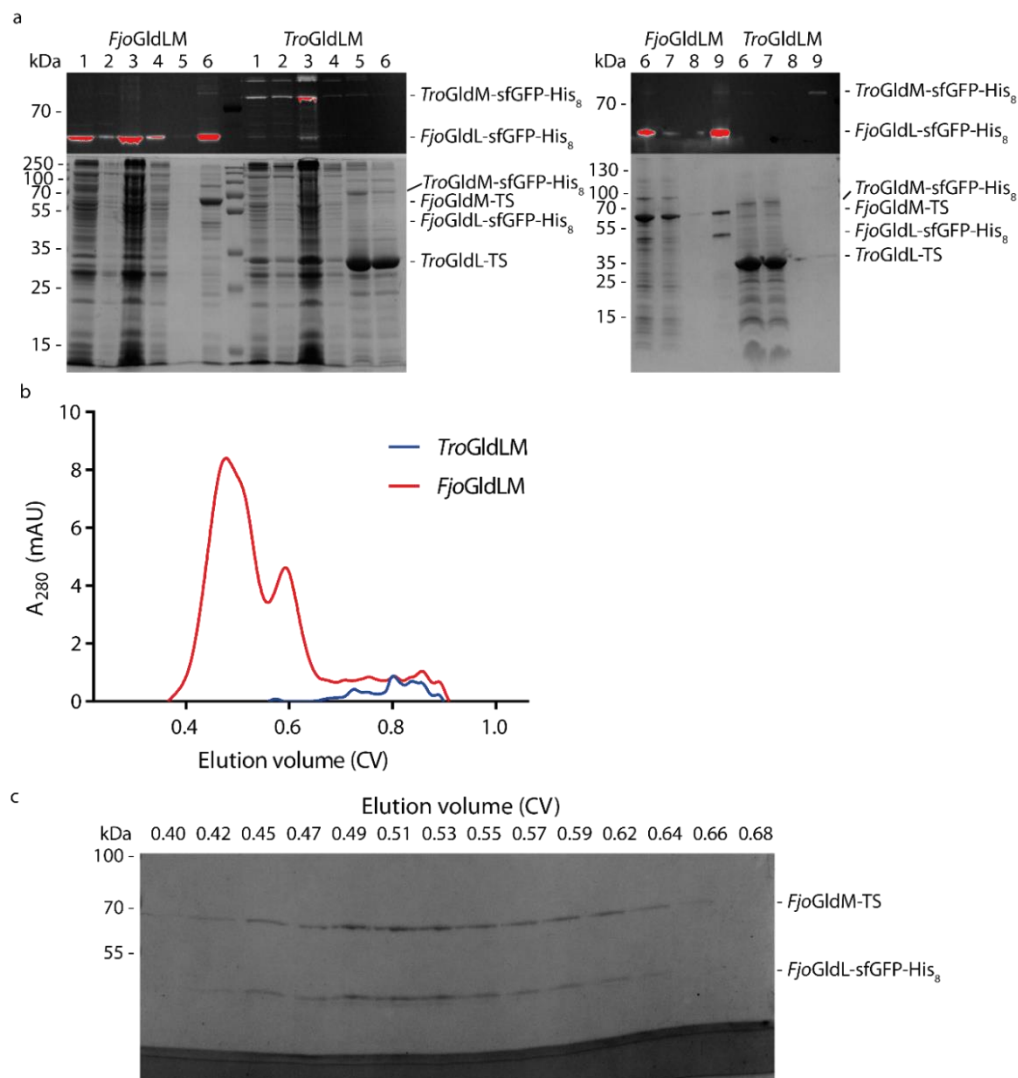


Figure 3-10 Purification of *FjoGldLM* and *TroGldLM*. **a** SDS-PAGE analysis of purification fractions, imaged by sfGFP fluorescence (top panel) or Coomassie blue staining. Red indicates overexposed pixels. 1 – DDM-solubilised material, 2 – unsolubilised material, 3 – StrepTactin XT flowthrough, 4 – StrepTactin XT wash, 5&6 – StrepTactin XT elution fractions, 7 – Ni-NTA flowthrough, 8 – Ni-NTA wash, 9 – Ni-NTA elution. **b** SEC traces for *FjoGldLM* (red) and *TroGldLM* (blue) run on a Superose 6 10/300 Increase column. **c** SDS-PAGE analysis of *FjoGldLM* SEC fractions.

3.7 Further *PgiPorLM* characterisation

Purification of *PgiPorLM* from the cellular debris produced enough material to allow for biophysical and structural characterisation. Dr Steven Johnson (SJ) analysed the *PgiPorLM* sample by size-exclusion chromatography multi-angle light scattering (SEC-MALS). The SEC-MALS data showed a main peak with a protein mass of 320 kDa and a small peak with a protein mass of 540 kDa. Based on the predicted molecular weights of 35.8 kDa for *PgiPorL* (after cleavage of sfGFP-His₈) and 59.7 kDa for *PgiPorM-TS*, 320 kDa is between the predicted molecular weights for *PgiPorL*₅M₂ (298.4 kDa) and *PgiPorL*₆M₂ (334.2 kDa). The structural data described later show that the *PgiPorL*₅M₂ stoichiometry is correct. A protein mass of 540 kDa is between the predicted masses of *PgiPorL*₈M₄ (525.2 kDa) and *PgiPorL*₉M₄ (561 kDa). However, the reported mass was not consistent across the 540 kDa peak, suggesting it contained aggregated protein rather than individual *PgiPorLM* complexes and should not be used to draw conclusions about physiological stoichiometries.

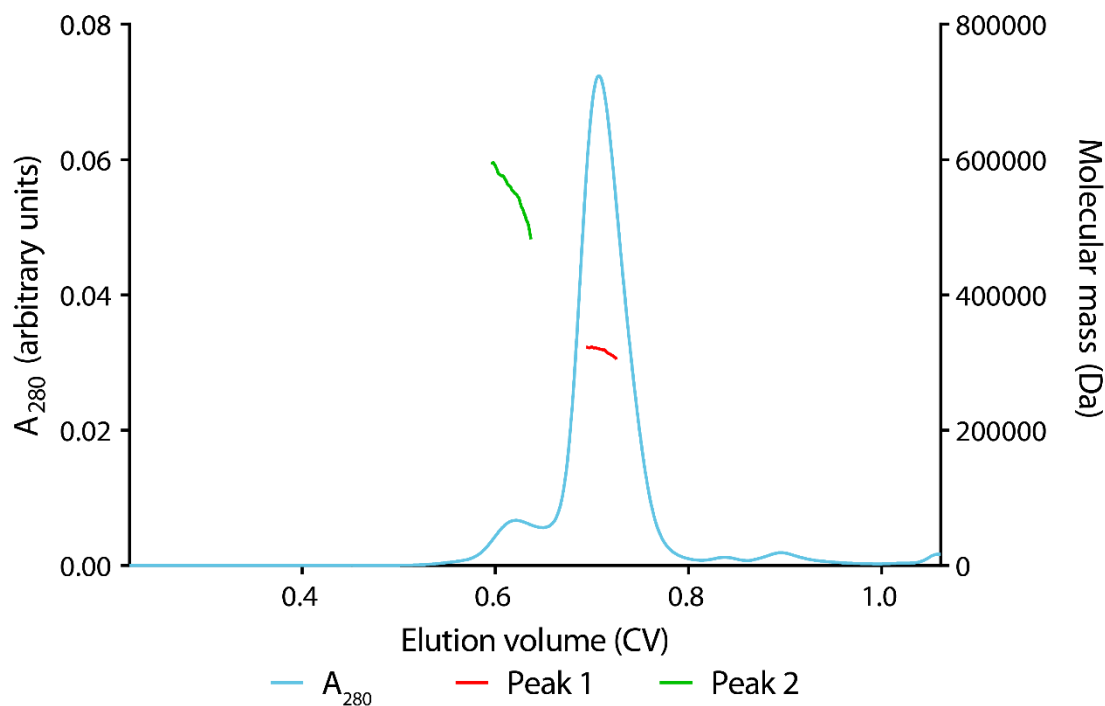


Figure 3-11 SEC-MALS analysis of *PgiPorLM* analysed using a Superose 6 Increase 10/300 column. A₂₈₀ trace is shown in light blue. The calculated protein mass for each peak is shown in green (Peak 1) or red (Peak 2).

CryoEM grids of *PgiPorLM* solubilised in LMNG had better particle density and less protein aggregation than *PgiPorLM* solubilised in DDM (**Figure 3-12a**). JCD collected a cryoEM dataset using a Titan Krios (FEI) with K2 Summit direct electron detector (Gatan). The K2 Summit detector was used in counting mode at 165,000x magnification with a pixel size of 0.822 Å and a total dose of 48 e⁻ Å⁻² over 20 fractions. After 2D classification of particles, transmembrane helices were visible in the detergent micelle in several classes (**Figure 3-12b**). Some secondary structure detail was visible for the first (D1) periplasmic domain of *PgiPorM*, and the other periplasmic domains were visible with less detail (**Figure 3-12b**). The blurring of domains D2-4 suggests there may be some flexibility or multiple conformations of the hinge region between domains that is amplified due to the distance from the micelle used as the centre point for alignment. The cytoplasmic domain of *PgiPorL* was not visible. After several rounds of 3D classification, no classes had resolved transmembrane helices. Micelle-focused refinements and classifications did not improve the transmembrane helices, so I stopped processing this dataset.

Once the structure of *FjoGldLM'* was available (Chapter 4; **Figure 4-7**, **Figure 4-8**), JCD used it as a reference to re-examine the *PgiPorLM* dataset and was able to solve a low-resolution structure as follows. JCD used SIMPLE 2.5 to automatically pick and extract 1,133,336 particles from 14,135 movies in 324 × 324-pixel boxes. JCD subjected these particles to two rounds of reference-free 2D classification in RELION 3.0 (Zivanov *et al*, 2018), then selected 495,572 particles in classes with secondary structure visible. JCD re-centred and re-extracted these particles in 256 x 256-pixel boxes for the initial 3D classification. This box size excluded most of the periplasmic density so that the initial classification was focused on the D1 domains, for which structural information was available from the *FjoGldLM'* map. JCD subjected these particles to 3D classification (4 classes, 15 iterations at 7.5° sampling then 10 iterations at 3.75° sampling) against a 40-Å low-pass-filtered *FjoGldLM'* map. JCD selected one class of 199,929 particles where the separation of the D1 domains was visible and

refined them against the corresponding map (low pass filtered to 40 Å) first using a soft spherical mask of 140 Å and then with a 180 Å mask. JCD then re-extracted the particles in 324 x 324-pixel to include the D2 and D3 domains and refined them again in a 400 Å spherical mask. JCD then re-extracted the particles again in 480 x 480-pixel boxes to include all domains and refined them with a mask based on the previous refinement and extended to cover the entire periplasmic region of PorM. This yielded a map of 8.6 Å resolution in which the two PorM transmembrane helices were separate, but the transmembrane helices of PorL were not resolved (**Figure 3-12c** and **Figure 3-13a**). The data processing workflow for *Pgi*PorLM is summarised in **Figure 3-12** and data collection and refinement statistics are given in **Table 3-3**.

Table 3-3 Data collection and processing statistics for *Pgi*PorLM. Adapted from (Hennell James *et al*, 2021).

***P. gingivalis* PorLM (EMD-10894)**

Data collection and processing	
Magnification	165,000
Voltage (kV)	300
Electron exposure (e⁻/Å²)	48
Defocus range (µm)	1.0-3.0
Pixel size (Å)	0.822
Symmetry imposed	C1
Initial particle images (no.)	1,133,336
Final particle images (no.)	199,929
Map resolution (Å)	8.6
FSC threshold	0.143

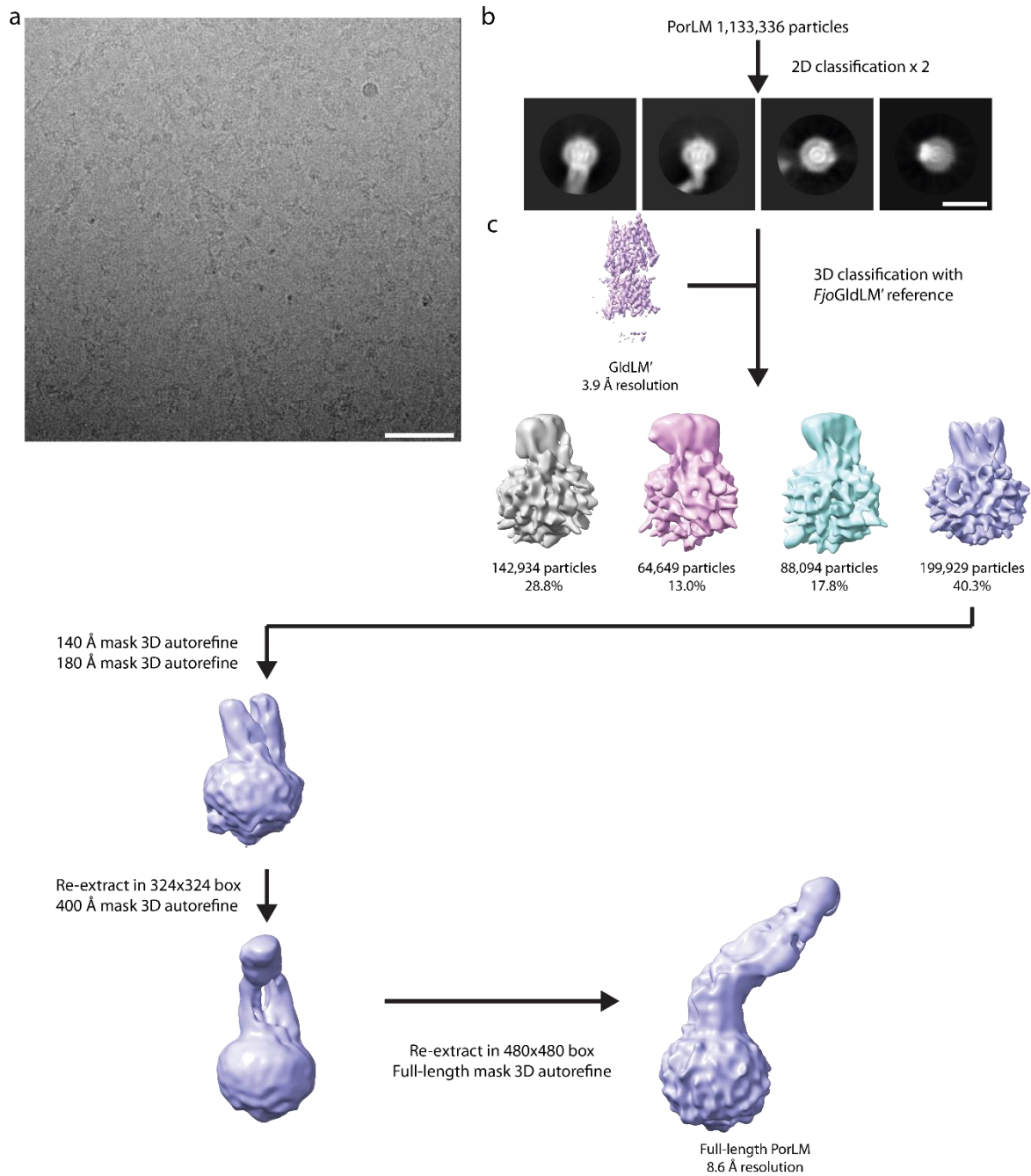


Figure 3-12 Data processing workflow for the *PgiPorLM* map. **a** Representative micrograph of *PgiPorLM* sample at approximately 1.8 μm defocus. Scale bar 500 Å. **b** representative 2D class averages of *PgiPorLM*. Scale bar 100 Å. **c** Data processing workflow.

The full-length structure of *PgiPorLM* had two separate transmembrane helices of *PgiPorM* were visible and separated D1 domains of *PgiPorM*, but the transmembrane helices of *PgiPorLM* could not

be resolved and the cytoplasmic domain of *PgiPorL* was not visible. A bend between the D1 and D2 was apparent in the structure, in contrast to previously published crystal structures of *PgiPorM*_{peri} and *FjoGldM*_{peri} (Leone *et al*, 2018), but consistent with the structure of *PgiPorM*_{peri} published later by a different group (Sato *et al*, 2020). This bend places the tip of the *PgiPorM* D4 domain roughly 110 Å away from the central axis of the micelle. Similarities and differences between the published crystal structure of *FjoGldM* and *PgiPorM* and the cryoEM structures presented here are discussed further in Chapter 4.

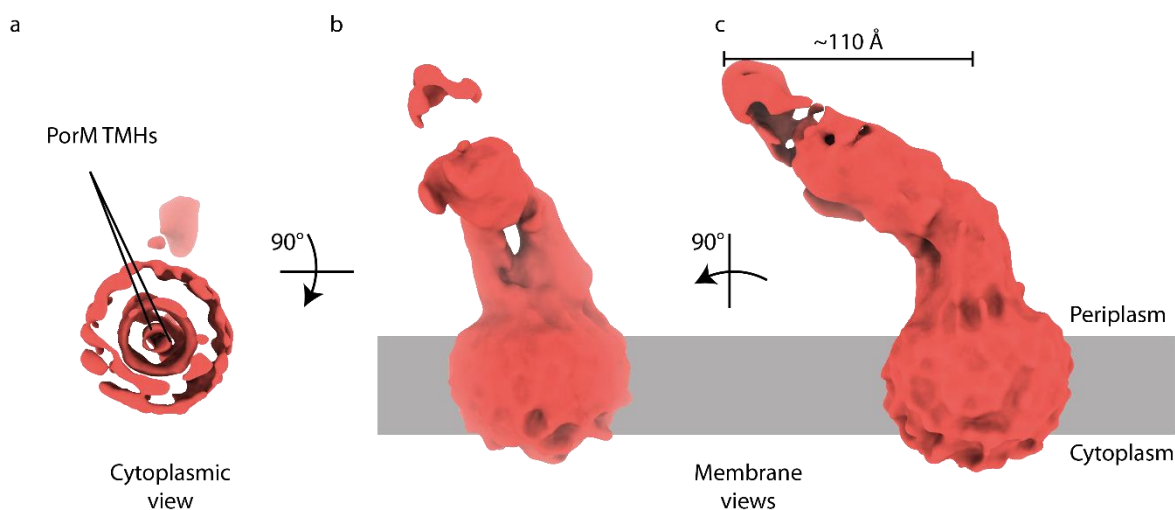


Figure 3-13 Low resolution cryo-EM map of *PgiPorLM*. **a** Cytoplasmic view. The same contour level and clipping pane is adjusted to show the two *PgiPorM* transmembrane helices. **b** View parallel to the membrane. The map contour level is adjusted to highlight the separation of the periplasmic D1 domains of *PgiPorM*. **c** Alternative view parallel to the membrane. The contour level is adjusted to show the full length of *PgiPorLM*.

3.8 Discussion

In this chapter I presented the purification and characterisation of full length *PgiPorLM*, expressed from a pETDuet-1-based vector. Samples purified in DDM adhered to the carbon support of cryoEM grids and so were unsuitable for data collection. I instead purified *PgiPorLM* proteins in LMNG. During the purifications I found that most of the *PgiPorLM* proteins, as judged by sfGFP fluorescence, entered the cellular debris pellet after cell lysis, even at low centrifugation speeds. This suggests that it was associated with a larger object, either inclusion bodies, the peptidoglycan cell wall, or

aggregates. I primarily based my choice of expression conditions on the fluorescence of whole cells, which may have included fluorescence from proteolyzed free sfGFP, fusion proteins in inclusion bodies, fusion proteins in aggregates and fusion proteins entangled with the peptidoglycan. In my expression screen I did not harvest membranes before purification with Streptavidin beads, but instead lysed cells directly detergent then pelleted cellular debris with a low-speed centrifugation step. This means that my Streptavidin bead elution fractions may have included proteins in aggregates that were too small to enter the pellet after 3,220g centrifugation. This would have inflated the fluorescence values obtained for Streptavidin elution fractions but not actually represented assembled and isolated *PgiPorLM* complexes. I did succeed in solubilising and purifying proteins from the cellular debris fraction, suggesting that not all *PgiPorLM* proteins were totally unfolded in inclusion bodies. The *PgiPorLM* protein complexes I solubilised from the cellular debris using LMNG appeared to be properly folded and assembled, even if they did represent a small fraction of the protein.

Initial characterisation of *PgiPorLM* by negative stain microscopy and cryo-EM suggested it had a piston-like shape with the detergent micelle forming the “head” and the periplasmic region of *PgiPorM* forming the “shaft”. The low-resolution 3D reconstruction of *PgiPorLM* had the same shape. No density corresponding to the cytoplasmic region of *PgiPorL* was visible in the 2D class averages or 3D reconstruction. This suggests that it was mobile and so blurred out when averaged. The region the cytoplasmic domain of PorL that is proximal to the membrane is predicted to be disordered so it makes sense that this domain would be mobile. The periplasmic D1 domains of *PgiPorM* were splayed apart in a manner similar to those of *FjoGldM'* in the *FjoGldLM'* structure presented in Chapter 4 (**Figure 4-8**) but different from those of *FjoGldM_{peri}* and *PgiPorM_{peri}* in published crystal structures (Leone *et al*, 2018; Sato *et al*, 2020). The arrangement of the periplasmic domains of GldM and PorM is discussed further in Chapters 4 and 6 (**Figure 4-9**, **Figure 6-14**). This kink may have allowed

for flexibility or multiple conformations of the periplasmic region which reduced the resolution of the reconstruction in the periplasmic region and prevented the determination of a high-resolution structure.

The full length *Pgi*PorLM construct did not lead to residue-level structural data or unambiguous determination of the stoichiometry of the *Pgi*PorLM complex by SEC-MALS. The full length *Fjo*GldLM and *Tro*GldLM constructs did not produce enough protein for any characterisation. I had much more success with the truncated constructs and the L-rhamnose inducible pT12 vector as described in Chapters 4 and 6, so I did not revisit the pETDuet-1-based system.

4 Cryo-electron microscopy structure of the *Flavobacterium johnsoniae* GldLM' complex

Much of the work described in this chapter was published in (Hennell James *et al*, 2021). The work of other authors is acknowledged in the text. Where indicated, some figures were adapted from initial drafts produced by RHJ for that publication.

4.1 Introduction

I initially tried to purify and solve the structure of full-length *P. gingivalis* PorLM but could not solve a high-resolution structure, as described in Chapter 3. The long periplasmic domain of PorM seemed to induce preferential orientations of *Pgi*PorLM on cryoEM grids, so lack of “top-down” views prevented solving a high-resolution structure. There also seemed to be some conformational heterogeneity at the boundaries between the different periplasmic domains of PorM, which lead to poor particle alignment and weak density for these regions. I therefore reasoned that truncating the periplasmic domain of PorM to only include the D1 domain would improve the chances of solving the PorLM structure as there would be no conformational heterogeneity of the periplasmic domain interfaces and the shorter complex might be less prone to preferential orientations. I was primarily interested in studying the interaction between PorL and PorM, which was believed to be via their transmembrane helices (Vincent *et al*, 2017), so loss of the D2-D4 domains should not compromise this objective. For future studies of the interactions between the PorLM complex and the outer membrane T9SS and gliding motility machinery, it may be necessary to purify only the periplasmic domains of PorM. I therefore set out to purify truncated forms of *Fjo*GldLM and *Pgi*PorLM.

Other members of the Lea Group had had success expressing membrane protein complexes from the L-rhamnose-inducible pT12 plasmid (Dietsche *et al*, 2016) rather than IPTG-inducible plasmids (Kuhlen *et al*, 2018; Deme *et al*, 2020). The pT12 plasmid contains the *rhaR* and *rhaS* genes and the *P_{rhaB}* promoter from the *E. coli* *rha* gene cluster (Figure 4-1). RhaB is constitutively expressed at a low

rate and in the presence of L-rhamnose upregulates expression of both RhaS and RhaR (Egan & Schleif, 1993). RhaS in turn upregulates transcription from the P_{rhaB} promoter. This system has less leaky expression in the absence of inducer than an IPTG-based system and can be further repressed by addition of glucose (Haldimann *et al*, 1998). Preventing leaky expression of proton-conducting proteins is important as they have the potential to cause toxicity when expressed exogenously by depleting the *E. coli* PMF. The combination of the pT12 vector with L-rhamnose induction during overnight exponential growth leads to “low-and-slow” expression which seems to avoid misfolding issues when highly overexpressed membrane proteins overwhelm the *E. coli* membrane protein-insertion machinery, resulting in more useable protein. As it had been successful for other, similar, projects, I decided to try using the pT12 vector to produce my truncated *PgiPorLM* and *FjoGldLM* constructs.

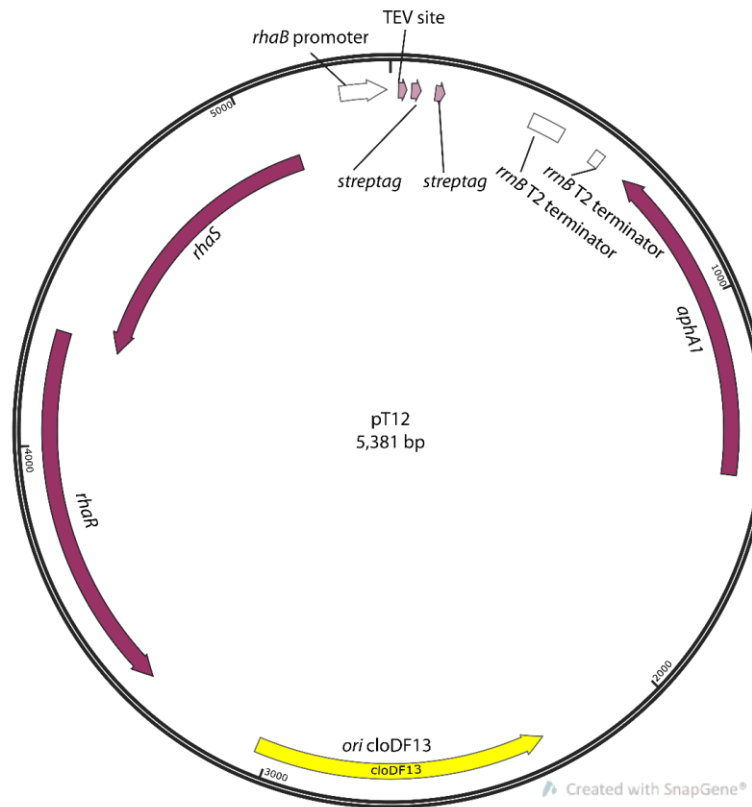


Figure 4-1 Map of the empty pT12 vector. *aphA1* encodes an aminoglycoside 3'-phosphotransferase that confers resistance to kanamycin. *rhaS* and *rhaR* encode components of the *E. coli* L-rhamnose-sensing system that induce transcription from the *rhaB* promoter in the presence of L-rhamnose.

4.2 Design of *FjoGldLM'*_A, *FjoGldLM'*_B and *PgiPorLM'*_A constructs

I designed my truncated *FjoGldLM* and *PgiPorLM* constructs to encode the entirety of the *FjoGldL* or *PgiPorL* protein and the sequence of the *FjoGldM* or *PgiPorM* protein up to the end of the periplasmic D1 domain. Genes were PCR-amplified from gDNA and assembled into the pT12 vector by Gibson cloning, as described in 2.3. The genes were transcribed as a single transcript with the chromosomal intergenic region retained in the hopes that this would lead to proper stoichiometry of translation and proper assembly. A sequence encoding a TEV cleavage site and a TwinStrep II tag was fused to the 3'-end of the truncated GldM/PorM to allow for affinity purification.

To decide where to truncate GldM/PorM I compared secondary structure predictions from PSI-PRED (Jones, 1999) and the crystal structures of *FjoGldM*_{peri} and *PgiPorM*_{peri} previously published (Leone *et*

al, 2018) (**Figure 4-2**). For *PgiPorM* the end of the final α -helix in D1 predicted by PSI-PRED was Lys216 and the final residue in the crystal structure was Leu214. I therefore designed the constructs *PgiPorLM'*_A, which extended to residue Lys216, and *FjoGldLM'*_A, which extended to the equivalent residue Lys225. However, the structure of *PgiPorM*_{peri} presented by Leone *et al.* was made of three fragments of *PgiPorM* crystallised separately and their construct for the D1 domain only included residues up to Ser217 so if the helix extended past this point it would not be resolved in the structure. I later addressed this by producing a *PgiPorLM'*_C construct, which extended to residue Ser227 (see Chapter 6). PSI-PRED predicted that Ser232 would be the last residue of the final α -helix in *FjoGldM* D1, which agreed with the crystal structure for which Ala231 was the last residue of that α -helix (**Figure 4-2a**). I therefore designed the construct *FjoGldLM'*_B, which extended to residue Ser232 of GldM.

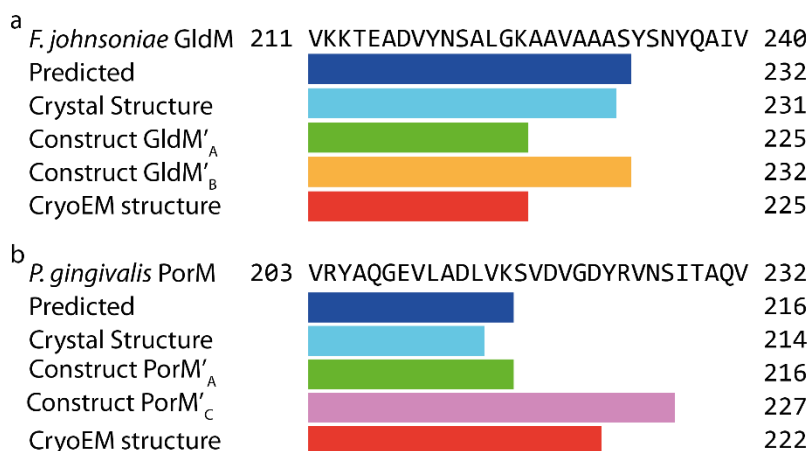


Figure 4-2 Construct design for GldLM' constructs. Aligned sequences from *F. johnsoniae* GldM (**a**) and *P. gingivalis* PorM (**b**). The lengths of the final helix in the D1 domain according to PSI-PRED (dark blue), the previously-solved crystal structures (light blue, PDB codes 6EY4 (**a**) and 6EY0 (**b**)) and the cryoEM structures presented here (red). The extent of the constructs created are shown in green, orange and pink. Construct PorM'_C and the related cryoEM structure are described in Chapter 6. N.B. the constructs expressed had an additional 39 C-terminal residues consisting of the TEV cleavage site and TwinStrep II tag.

4.3 Purification of *FjoGldLM'*_A, *FjoGldLM'*_B and *PgiPorLM'*_A

Once I had cloned these vectors, I attempted to purify proteins expressed from them. I used a purification protocol (2.6.5) based on the successful protocol used by Dr Lucas Kuhlen for purification of small membrane protein complexes involved in the bacterial Type III Secretion System (Kuhlen *et al*, 2018), but replaced the Tris-based buffers with PBS- or HEPES-based buffers as I had found that full-length *PgiPorLM* was stable in these conditions. I later found that GldLM homologues were also stable in Tris-based buffers (Chapter 6). I had found that *PgiPorLM* purified in dodecyl maltoside (DDM) adhered to the carbon support of cryoEM grids rather than entering the grid holes (Chapter 3). It seemed possible this would be the case for the truncated constructs as well. I therefore used the detergent lauryl maltose neopentyl glycol (LMNG) for my initial purification as *PgiPorLM* purified in LMNG did not seem to adhere to the carbon to such an extent (Chapter 3). LMNG has been used to solubilise bacterial membrane proteins for high-resolution structure determination in several other projects in the Lea laboratory (Kuhlen *et al*, 2018; Lauber *et al*, 2018; Deme *et al*, 2020). All three complexes expressed well, with bands corresponding to both proteins visible by SDS-PAGE after elution from the StrepTactin XT column (Figure 4-3a lanes 9).

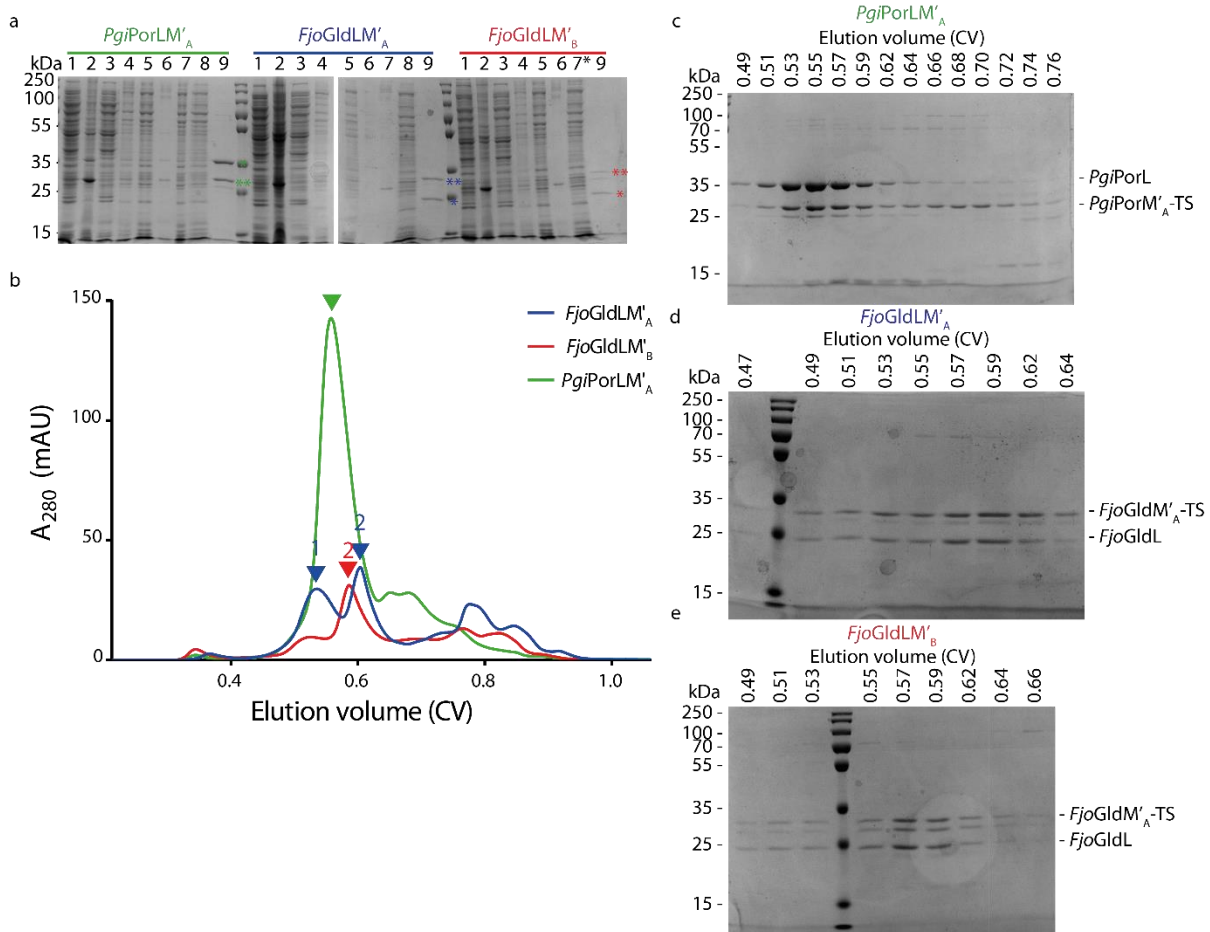


Figure 4-3 Purification of *PgiPorLM'*_A, *FjoGldLM'*_A, and *FjoGldLM'*_B by StrepTactin XT affinity and size-exclusion chromatography (SEC). **a** SDS-PAGE analysis of fractions from the purification process. 1 – 27,200g centrifugation supernatant, 2 – 27,200g centrifugation pellet, 3 – 200,000g centrifugation supernatant, 4 – 200,000g centrifugation pellet, 5 – 75,000g centrifugation supernatant, 6 – 75,000g centrifugation pellet, 7 – StrepTactin XT flowthrough, 8 – StrepTactin XT wash, 9 – StrepTactin XT elution, 7* - combined StrepTactin XT flowthrough and wash fractions. Lanes 1-6 are balanced for each construct so that approximately the same proportion of the fraction is loaded. * and ** indicate *FjoGldL/PgiPorL* and *FjoGldM'*_A-TS/*FjoGldM'*_B-TS/*PgiPorM'*_A-TS, respectively. **b** SEC traces using a Superose 6 Increase 10/300 GL column for *FjoGldLM'*_A (blue), *FjoGldLM'*_B (red), and *PgiPorLM'*_A (green). Arrowheads indicate the peaks analysed by electron microscopy. **c-e** SDS-PAGE analysis of SEC fractions for *PgiPorLM'*_A (c), *FjoGldLM'*_B (d), and *FjoGldLM'*_A (e). Lane labels indicate retention volumes for the SEC fractions analysed.

When analysed by SEC, *PgiPorLM'*_A formed a monodisperse peak at 0.55 CV that contained *PgiPorL* and *PgiPorM'*_A-TS (Figure 4-3b-c). Assessment by negative stain electron microscopy showed few aggregates in this peak (Figure 4-4c). There were some smaller subsequent peaks that appeared to contain dissociated *PgiPorL* and *PgiPorM'*_A-TS as well as some trace contaminants (Figure 4-3b, c). When analysed by SEC (Figure 4-3b, d-e), *FjoGldLM'*_A and *FjoGldLM'*_B both had peaks at 0.53 CV (peak

1) and 0.59 CV (peak 2) that both contained *FjoGldL* and *FjoGldM'*-TS. Negative-stain electron microscopy showed that peak 1 contained many aggregates, whilst peak 2 was mostly separated detergent micelles (**Figure 4-4b-c**). Similar to *PgiPorLM'*_A, the subsequent small peaks contained *FjoGldL* and *FjoGldM'*-TS and probably represent proteins from dissociated *FjoGldLM'* complexes (**Figure 4-3b, d-e**). In total, I purified approximately 130 µg of *PgiPorLM'*_A and approximately 40 µg each of *FjoGldLM'*_A and *FjoGldLM'*_B from 2 l each of cell culture.

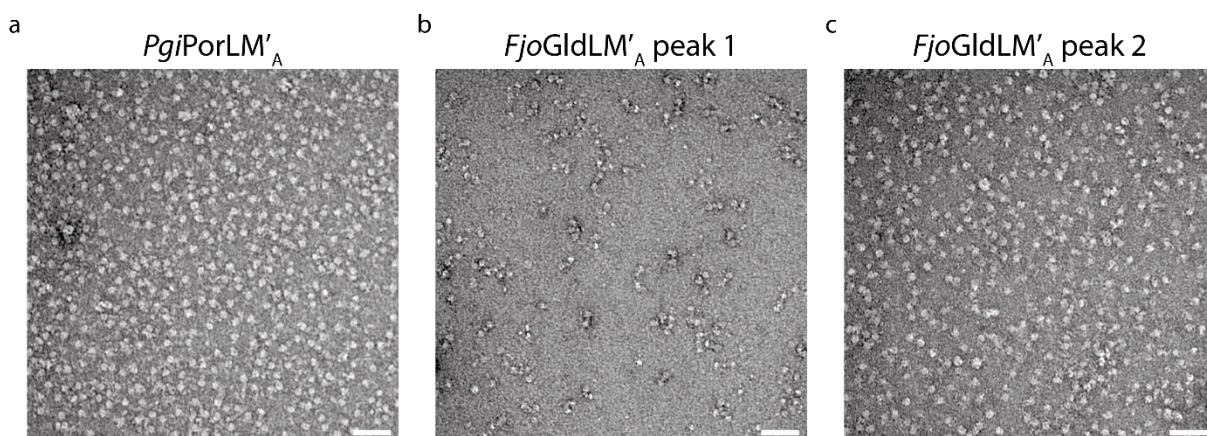


Figure 4-4 Negative-stain electron microscopy analysis of SEC fractions. **a** *PgiPorLM'*_A main peak. **b** *FjoGldLM'*_A peak 1. **c** *FjoGldLM'*_A peak 2. Scale bars 500 Å.

Knowing the mass of the *PgiPorLM'*_A or *FjoGldLM'* complexes would allow estimation of the protein stoichiometries, so Dr Steven Johnson (SJ) analysed the pooled 0.53-0.57 CV fractions of *PgiPorLM'*_A by SEC-MALS (**Figure 4-5**). SEC-MALS requires a minimum of 100 µg per experiment, so it was not possible to analyse *FjoGldLM'*_A or *FjoGldLM'*_B by SEC-MALS. SEC-MALS gave a calculated mass for *PgiPorLM'*_A of 236 kDa. The trace of the calculated mass over the main peak of the SEC trace was sloped, suggesting that not all particles in the peak were the same mass. This is probably due to the overlap of the main peak with the earlier-eluting peak which contained larger particles. The predicted masses of *PgiPorL* and *PgiPorM'*_A-TS are 34.9 kDa and 28.1 kDa, respectively. 236 kDa is therefore between the predicted masses for *PgiPorL*:*PgiPorM'*_A stoichiometries of 5:2 (231 kDa) and 6:2 (266

kDa). SJ previously calculate the mass of full length *PgiPorLM* as 320 kDa which is also between the predicted masses of *PgiPorLM* complexes with 5:2 and 6:2 stoichiometries (298 kDa and 334 kDa respectively, **Figure 3-11**). At the time, the only published data on *PgiPorL* suggested the cytoplasmic domain in isolation formed a trimer, so I presumed that the *PgiPorL*₆*M*₂ stoichiometry was correct. The *FjoGldL*₅*M*₂ structure presented later in 4.5 indicated that the *PgiPorL*₅*M*₂ stoichiometry was in fact correct, which was confirmed by the *PgiPorL*₅*M*₂ structure presented in Chapter 6.

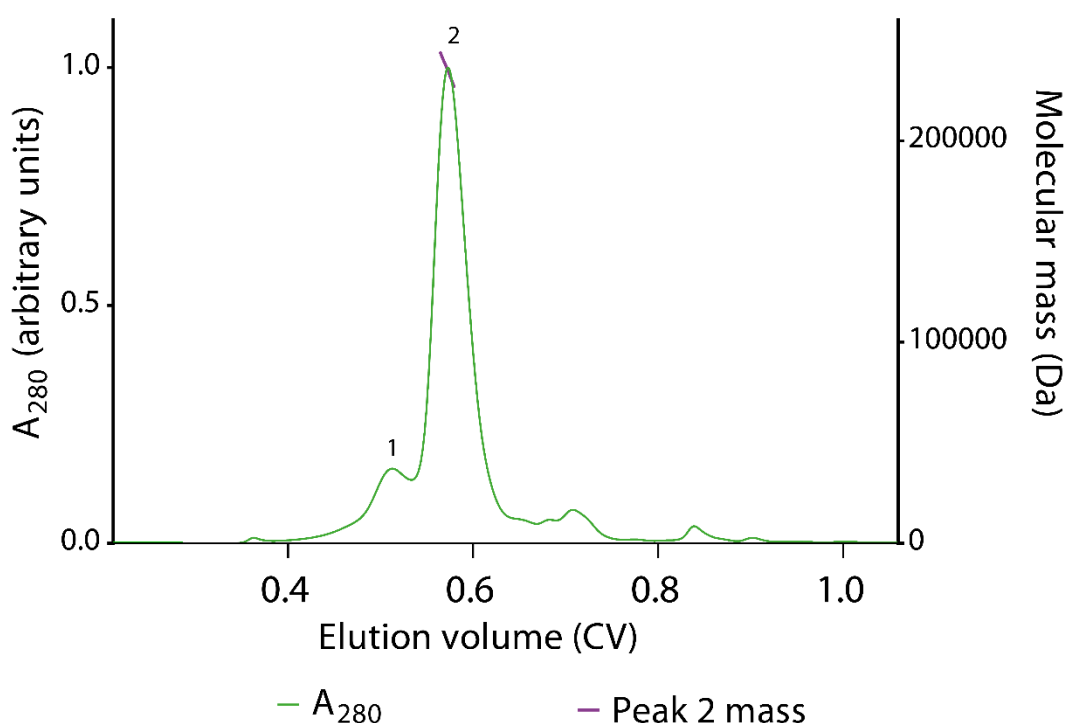


Figure 4-5 SEC-MALS analysis of *PgiPorLM*'A gave a calculated mass for *PgiPorLM*'A of 236 kDa. A₂₈₀ trace is shown in green, calculated protein mass across the second peak is shown in purple.

4.4 Solving the structure of *FjoGldLM'*

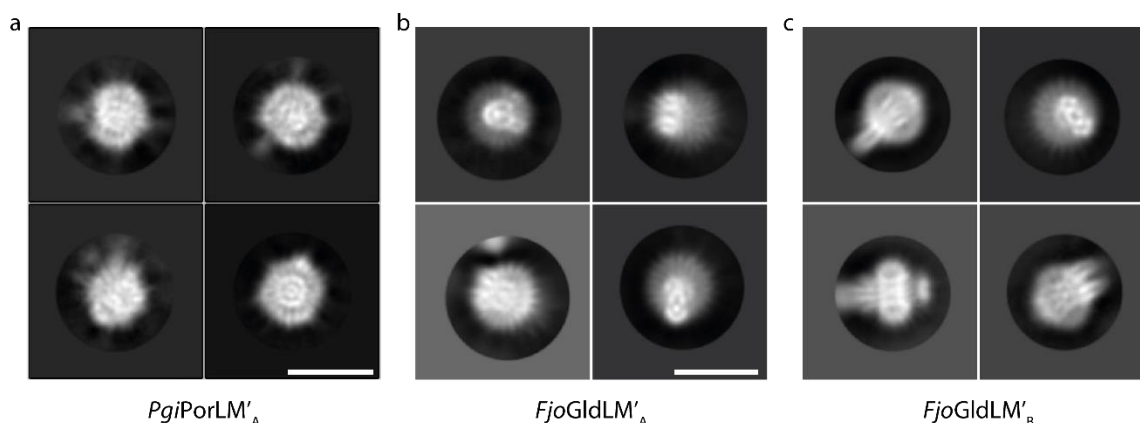


Figure 4-6 Representative 2D class averages for the *PgiPorLM'*_A (a), *FjoGldLM'*_A (b), and *FjoGldLM'*_B (c) constructs. b & c are repeated in Figure 4-7, which is in turn adapted from (Hennell James *et al*, 2021). Scale bars 100 Å.

Having purified *PgiPorLM'*_A, *FjoGldLM'*_A, and *FjoGldLM'*_B proteins, I next analysed them by cryoEM to attempt to solve the structures of *PgiPorLM'* and *FjoGldLM'*. Dr Justin Deme (JCD) prepared and imaged cryoEM grids from peak 2 of *FjoGldLM'*_A and *FjoGldLM'*_B, and the main peak of *PgiPorLM'*_A. After screening the grids for ice quality and particle density, JCD collected data for all three constructs using a Titan Krios microscope (FEI) with K2 Summit direct electron detector (Gatan). The K2 Summit detector was used in counting mode at 165,000x magnification with a pixel size of 0.822 Å and a total dose of 48 e⁻ Å⁻² over 20 fractions. A representative micrograph is shown in Figure 4-7a. After pre-processing, I used SIMPLE 2.5 (Reboul *et al*, 2018) to automatically pick and extract protein particles in 256x256 pixel boxes. I then subjected extracted particles to reference-free 2D classification in RELION 3.0 (Zivanov *et al*, 2018) and selected classes with clear secondary structure elements. There were no classes with clear detail for the D1 domains of *PgiPorLM'*_A, suggesting that they were not well folded and that I may have selected the wrong domain boundaries for my construct. This is supported by the structural data I later obtained for the *PgiPorLM'*_C construct (see Chapter 6, Figure 6-17). Secondary structure elements were visible in both the transmembrane and D1 domains in

several classes for *FjoGldLM'*_A and *FjoGldLM'*_B (Figure 4-6b-c), suggesting that these domains were well folded in both constructs.

Further processing was carried out by JCD with input from RHJ and SML. The data processing workflow is summarised in Figure 4-7c. JCD generated an *ab initio* initial model with C2 symmetry from 500 *FjoGldLM'*_A particles using SIMPLE 2.5. C2 symmetry was selected based on the presumed *FjoGldL₅M'*_A stoichiometry and the C2 symmetry of the published *FjoGldM_{peri}* structure (Leone *et al*, 2018). All subsequent processing was carried out in RELION 3.0. JCD low-pass filtered the initial model to 60 Å and used it for 3D classification (4 classes, 7.5 ° sampling, C2 symmetry) against the 335,887 *FjoGldLM'*_A particles. One class of 76,096 particles was selected and used for 3D autorefinement with C2 symmetry. This produced a 6.8 Å map in which density corresponding to the D1 domains of *GldM'*_A were visible, but splayed apart, rather than parallel as in the *FjoGldM_{peri}* structure (Figure 4-7c orange maps). No density corresponding to the cytoplasmic domain of *GldL* was visible and individual helices could not be resolved. Similar initial processing of the *FjoGldLM'*_B data produced similar results, so JCD combined the two datasets into a 531,515-particle dataset. JCD then subjected this dataset to 3D classification (3 classes, 15 iterations with 7.5 ° sampling then 10 iterations with 3.75 ° sampling) using the 6.8 Å *FjoGldLM'*_A model with a 40 Å low-pass filter as a reference and mask. This classification used C1 symmetry to avoid biasing the map to an inaccurate symmetry. The maps resulting from this classification all had visible α-helices in both the D1 domains and transmembrane region. 10 *GldL* helices were visible surrounding 2 *GldM* helices, suggesting a *FjoGldL₅M'*₂ stoichiometry. 5:2 stoichiometry is incompatible with C2 symmetry, explaining why the C2-refined map did not have defined α-helices. JCD selected the two classes with well-defined helices and little noise (382,460 particles) and subjected them to masked C1 autorefinement, yielding a 4.0 Å map. Bayesian particle polishing, which uses Bayesian statistics to correct for motion of particles induced by the electron beam (Zivanov *et al*, 2019), followed by another round of 3D classification

and autorefinement, improved this map to 3.9 Å resolution with 119,230 particles. Gold-standard Fourier shell correlations (FSC) using the 0.143 criterion and local resolution estimations were then calculated within RELION 3.0 (Figure 4-7d-e).

A 3.9 Å map allows the docking of domains and modelling of helices, but only bulky side chains are visible in the EM density (Figure 4-7f). JCD docked the D1 domains (residues 36–224) of the *FjoGldM_{peri}* crystal structure (PDB 6EY4, (Leone *et al*, 2018)) into the *FjoGldLM'* EM density using *Coot* (Emsley *et al*, 2010). The individual domains fitted neatly into the EM density but were splayed apart compared to the parallel dimer in the *FjoGldM_{peri}* crystal structure (discussed below). JCD built the transmembrane helices *de novo* using *Coot*, assigning the register based on the presence of bulky sidechains and TMH predictions from TMHMM (Krogh *et al*, 2001; Sonnhammer *et al*, 1998). Multiple rounds of rebuilding in both the globally sharpened and local-resolution-filtered maps and real-space refinement in *PHENIX* (Afonine *et al*, 2018; Liebschner *et al*, 2019) using secondary structure, rotamer and Ramachandran restraints yielded the final model. Validation was performed using MolProbity (Williams *et al*, 2018). Statistics for the data collection, cryoEM map and atomic model are given in

Table 4-1.

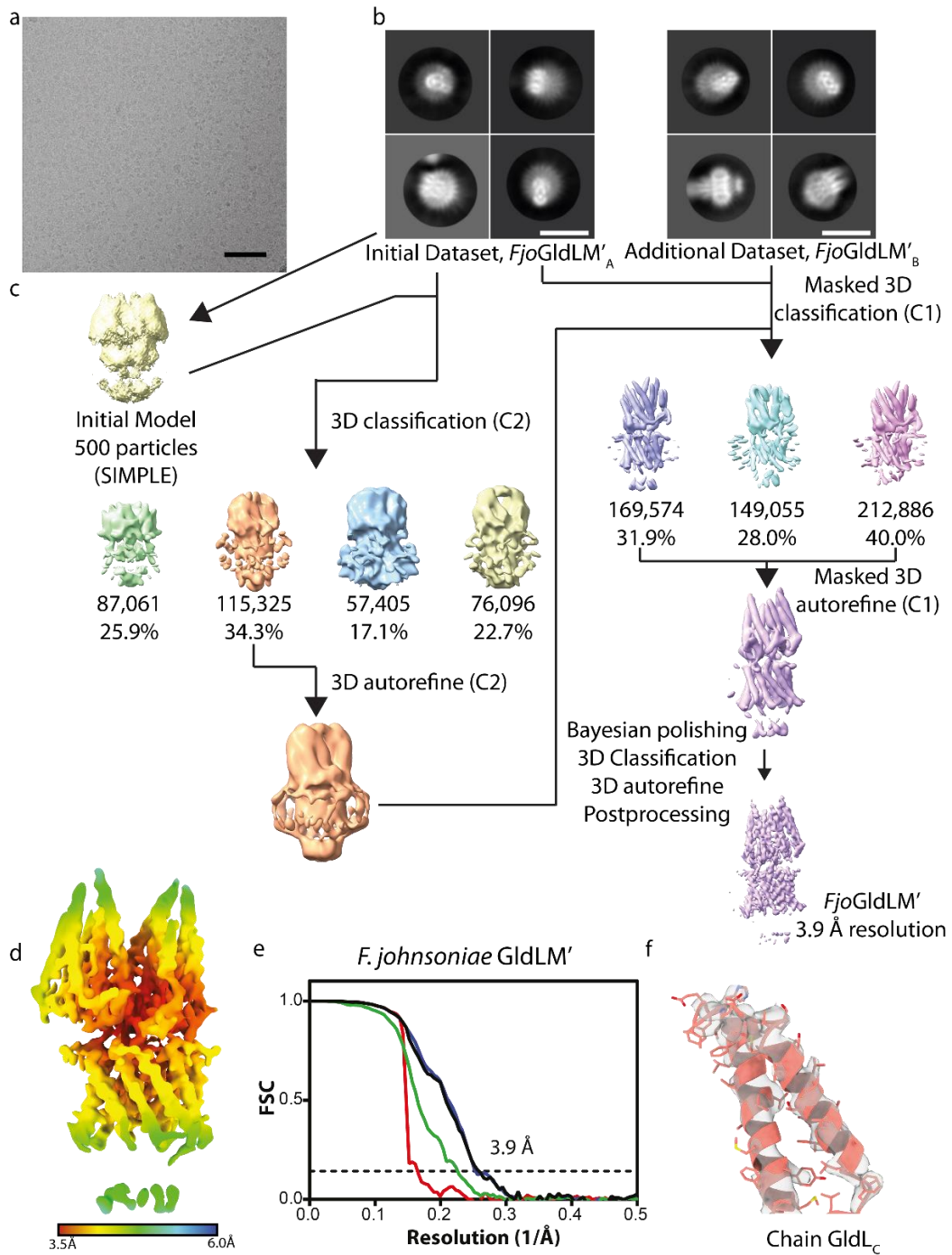


Figure 4-7 *FjoGldLM'* data processing workflow. Adapted from (Hennell James *et al*, 2021). **a** Representative micrograph at a defocus of approximately 2.5 μm . Scale bar 500 $\text{\AA}$. **b** Representative 2D class averages from the *FjoGldLM'*_A and *FjoGldLM'*_B datasets, repeated from **Figure 4-6**. Scale bar 100 $\text{\AA}$. **c** Structure-solving workflow for *FjoGldLM'*. **d** Fourier Shell Correlation (FSC) plot for the GldLM' structure. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomized; Green, unmasked; blue, masked; black, MTF-corrected. **e** Local resolution estimates (in $\text{\AA}$) for the sharpened GldLM' map. **f** Representative density with the atomic model built in.

Table 4-1 Cryo-EM data collection, refinement, and validation statistics for the *FjoGldLM'* structure. Adapted from (Hennell James *et al*, 2021).

<i>F. johnsoniae</i> GldLM' (EMD-10893, PDB 6YS8)	
Data collection and processing	
Magnification	165,000
Voltage (kV)	300
Electron exposure (e ⁻ Å ⁻²)	48
Defocus range (µm)	1.0-3.0
Pixel size (Å)	0.822
Symmetry imposed	C1
Initial particle images (no.)	3,284,887
Final particle images (no.)	119,230
Map resolution (Å)	3.9
FSC threshold	0.143
Map resolution range (Å)	3.5-5.8
Refinement	
Initial model used (PDB code)	None
Model resolution (Å)	3.9
FSC threshold	0.143
Model resolution range (Å)	3.5-5.8
Map sharpening B factor (Å ²)	-131
Model composition	
Non-hydrogen atoms	5624
Protein residues	737
Ligands	0
B factors (Å²)	
Protein	117
Ligand	NA
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.186
Validation	
MolProbity score	2.22
Clashscore	10.52
Poor rotamers (%)	1.70
Ramachandran plot	
Favoured (%)	91.42
Allowed (%)	6.36
Disallowed (%)	2.21

4.5 A 3.9 Å structure of *FjoGldLM'*

Only one conformation of *FjoGldLM'* was resolved, in which five copies of *FjoGldL* surround two copies of *FjoGldLM'* (**Figure 4-8a-c**). This is consistent with the SEC-MALS data for *PgiPorLM'*_A and suggests the previous reports that the cytoplasmic domain of *PgiPorL* forms a trimer in isolation are incorrect, or at least not applicable to the *PgiPorLM* complex ((Vincent *et al*, 2017), also discussed below). Residues 7-225 of *FjoGldM* could be modelled (**Figure 4-8d**), almost the entire length of *FjoGldM'*_A, but there was no density for the residues only present in *FjoGldM'*_B. The smaller dataset containing only *FjoGldLM'*_B particles could not be solved to sufficient resolution to build in an atomic model. As predicted, *FjoGldM'* has a single transmembrane helix and a helical periplasmic D1 domain connected by an ordered loop. The D1 domain adopts the same fold as seen in the published crystal structure (RMSD 0.7 Å for both copies), but the two copies are splayed apart rather than parallel, with each at a different angle from the TMH. This difference is discussed further below.

For *FjoGldL* in contrast, only residues 3-62 could be modelled (**Figure 4-8d**), leaving 153 residues at the C-terminus unbuilt. This supports the hypothesis that the cytoplasmic domain of *GldL* is at least partially disordered. The two helices of *FjoGldL* sit at an angle of ~25 ° from the membrane normal, with an ordered loop between them in the periplasm. The 10 *FjoGldL* helices surround the two *FjoGldM'* helices so that once the complex is assembled *FjoGldM'* does not contact the lipid bilayer. The periplasmic loops of *FjoGldL* contact the D1 domains of *FjoGldM'* and due to the inherently asymmetrical 5:2 stoichiometry of the complex, each copy of *FjoGldL* makes a different set of contacts with the D1 domains (**Table 4-2**). *FjoGldM'* Chain A makes more extensive contacts with all five chains of *FjoGldL* than Chain B, and Chain B does not make any contacts at all with *FjoGldL* Chains C and D.

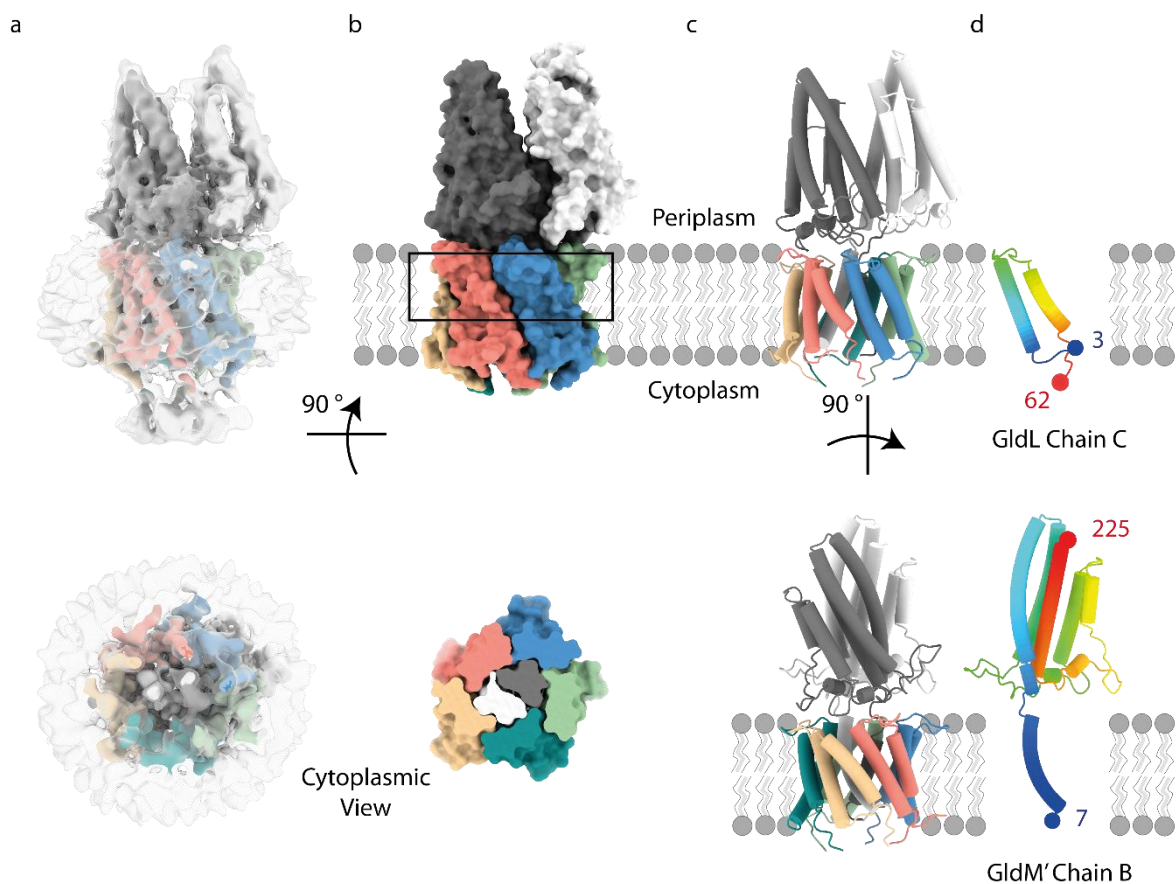


Figure 4-8 Architecture of the *FjoGldLM'* complex. Chains are coloured as follows: GldM' chain A – dark grey; GldM' chain B – white; GldL chain C – salmon; GldL chain D – blue; GldL chain E – green; GldL chain F – teal; GldL chain G – tan. Cartoon representations of the membrane are placed based on the detergent micelle. **a** EM density for *FjoGldLM'* viewed from within the membrane (upper) and the cytoplasm (lower). One copy of the map is shown at high contour level and coloured by chain to show protein structure, a second copy of the map is shown at low contour level and transparent to show the detergent micelle. **b** Surface representation of the atomic model viewed from within the membrane (upper) and the cytoplasm (lower). The black box in the upper panel indicates the clipping planes used in the lower panel. **c** Cartoon representation of the atomic model from two orthogonal views within the membrane. **d** Cartoon representation of GldL chain C (top) and GldM' chain B (bottom), coloured sequentially from the N-terminus (blue) to C-terminus (red). Adapted from (Hennell James *et al*, 2021).

Table 4-2 Contacts between the periplasmic loops of GldL and GldM. The PDBe-PISA server was used to calculate the sizes of the interfaces between the periplasmic loops of each copy of GldL and the periplasmic domains of GldM (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html, ref. (Krissinel & Henrick, 2007)). Adapted from (Hennell James *et al*, 2021).

GldL Copy	GldL Buried Surface area (Å ²)	Number of GldL Residues	GldL Residues	GldM Residues
GldM' Chain A				
C	265.4	8	Thr29, His30, Phe31, Glu 32, Thr37, Gly38, Thr39, Val40	Arg120, Asp122, Asp125, Asp126, Phe128, Thr129, Gly130, Lys193
D	107.4	3	Thr29, His30, Thr39	Met26, Lys30, Glu116, Ala117, Asp119, Arg120
E	139.5	4	Thr29, His30, Phe31, Glu32	Met26, Asn27, Ser29, Lys30, Glu31
F	125.8	3	His30, Thr37, Thr39	Lys180, Glu181, Ile183
G	183.8	4	Thr29, His30, Thr37, Thr39	Thr129, Gly130, Asp131, Asn179, Ile183, Lys184, Glu185
GldM' Chain B				
C	-	-	-	-
D	-	-	-	-
E	64.4	5	Phe31, Glu32, Thr37, Thr39, Val40	Arg120, Asp122, Lys180, Gly194
F	97.2	2	Thr29, His30	Val28, Ser29, Ile33, Glu116, Ala117, Asp119
G	18.1	1	His30	Met26

In the cryoEM structure of *FjoGldLM'*, the D1 domains are splayed apart by about 40 ° (Figure 4-9a). This contrasts with the crystal structure of *FjoGldM_{peri}*, where they are parallel as part of a dimer with C1 symmetry (Figure 4-9a, (Leone *et al*, 2018)). The fold of the D1 domain is the same between the two structures and one copy can be overlaid very well (Figure 4-9b; RMSD 0.7). However, the splaying in the cryoEM structure means that when the crystal structure of *FjoGldM_{peri}* and the cryoEM structure of *FjoGldLM'* are overlaid by aligning the D1 domain of GldM chain A, the D1 domains of chain B in each structure do not overlay at all (Figure 4-9c). The central D2 and D3 domains in the crystal structure feature β -strand exchanges between the two protomers and *PgiPorM_{peri}* was determined by SEC-MALS to be a dimer in solution (Vincent *et al*, 2017), so it seems very unlikely that

the dimerisation of *FjoGldM_{peri}* is an artifact of crystallisation conditions. On the other hand, in the presence of the transmembrane helices of *FjoGldM* and *FjoGldL* and the absence of crystal packing it is expected that the D1 domains would adopt a more physiological conformation in the cryoEM structure. To investigate this, I compared the *FjoGldLM'* structure to the *PgiPorLM* low-resolution volume described in Chapter 3 (Figure 4-9d). At the appropriate contour level, a gap between the *PgiPorM* D1 domains is visible, and they neatly accommodate the splayed D1 domains of *FjoGldLM'*. In contrast, when one copy of the *FjoGldM_{peri}* crystal dimer is docked into the *PgiPorLM* density the other copy does not fit neatly and occupies the gap between the EM density for the D1 domains (Figure 4-9e). No gap is visible between the D2 domains in the *PgiPorLM* volume, suggesting that the dimerisation of the D2-D4 domains seen in the crystal structure is valid (Figure 4-9d). There is, however, a kink between the D1 and D2 domains that is not seen in the *FjoGldM_{peri}* structure but is seen in a more recent *PgiPorM_{peri}* crystal structure (Figure 4-9f&h, (Sato *et al*, 2020)). This further suggests that the arrangement of the D1 domains in the *FjoGldM_{peri}* crystal structure is not physiologically accurate. Despite the presence of the kink, the D1 domains of *PgiPorM_{peri}* are parallel, so they align quite well with those of *FjoGldM_{peri}* (Figure 4-9g) and do not fit neatly into the *PgiPorLM* density when aligned to *FjoGldLM'* (Figure 4-9h) or manually docked (Figure 4-9i). The TMH of *PgiPorM* was present in the construct used by Sato *et al.*, but *PgiPorL* was absent and no detergent was included in the crystallisation buffer, so no TMH density was visible. It seems that the presence of GldL/PorL and a properly folded GldM/PorM TMH is necessary for the splaying of the D1 domains.

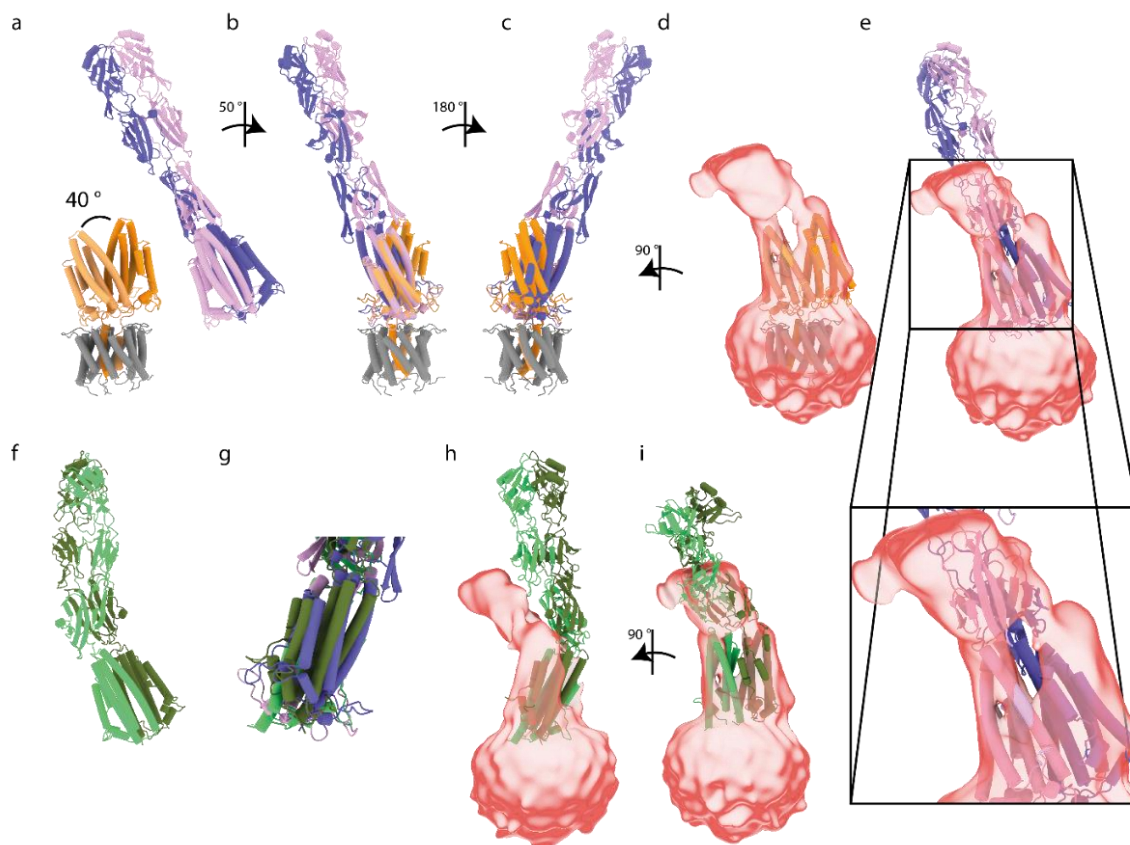


Figure 4-9 The positions of the D1 domains of *FjoGldLM'* and *FjoGldM_{peri}* are incompatible. **a** The D1 domains of *FjoGldLM'* are splayed about 40 ° from each other, whilst the D1 domains of *FjoGldM_{peri}* (PDB 6EY4(Leone *et al*, 2018)) are parallel to each other. **b-c** *FjoGldM_{peri}* aligned to *FjoGldLM'* according to the D1 domains of Chain A. The well-aligned Chain A (**b**, light purple) and misaligned Chain B (**c**, dark purple) are shown. **d** *FjoGldLM'* fitted into the full-length *PgiPorLM* map according to the D1 domains of GldM' (see Chapter 3). The contour level of the map is adjusted to highlight the gap between the D1 domains. **e** *FjoGldM_{peri}* fitted into the full length *PgiPorLM* map by alignment of Chain A to Chain A of *FjoGldLM'* in **c**. Right panel is an expanded view of the boxed region in the left panel showing how Chain B of *FjoGldM_{peri}* (dark purple) occupies the gap between the D1 domains of the full length *PgiPorLM* map. **f** *PgiPorM_{peri}* (PDB 7CMG(Sato *et al*, 2020)) from the same view as **a**. **g** Overlay of Chain B of *FjoGldM_{peri}* and Chain B of *PgiPorM_{peri}*, both aligned by Chain A to *FjoGldLM'* as in **c**. **h** *PgiPorM_{peri}* fitted in *PgiPorLM* by alignment to *FjoGldLM'* as in **c**, showing how the D1 domain of *PgiPorM_{peri}* does not fit in the EM density. **i** view of *PgiPorM_{peri}* manually fitted into the *PgiPorLM* map, viewed as in **d**, showing how the D1 domains of *PgiPorM_{peri}* occupy the gap between the D1 domains of *PgiPorLM*.

Analysis of amino acid conservation in the *FjoGldLM'* complex further supports the parallel arrangement of the D1 domains being inaccurate. ConSurf (Landau *et al*, 2005; Ashkenazy *et al*, 2016) analysis of *FjoGldL* and *FjoGldM'*, which colours residues according to sequence conservation at that site, shows that the faces of the transmembrane helices that interact in both proteins are well-conserved (Figure 4-10a-c). The outer faces of the *FjoGldL* helices, which interact with the lipid

bilayer, and the outer face of the D1 domains, which faces the periplasmic solvent, are not well-conserved. The inner face of *FjoGldM'* D1, which interacts with the other copy of D1, is weakly conserved (Figure 4-10b), suggesting that it does not form a strong interaction with its partner. This is further supported by co-evolution analysis of *FjoGldM'* using GREMLIN (Kamisetty *et al*, 2013), which found several co-evolving pairs of residues, but none for which an inter-molecular interaction was more plausible than an intra-molecular interaction (Figure 4-10d).

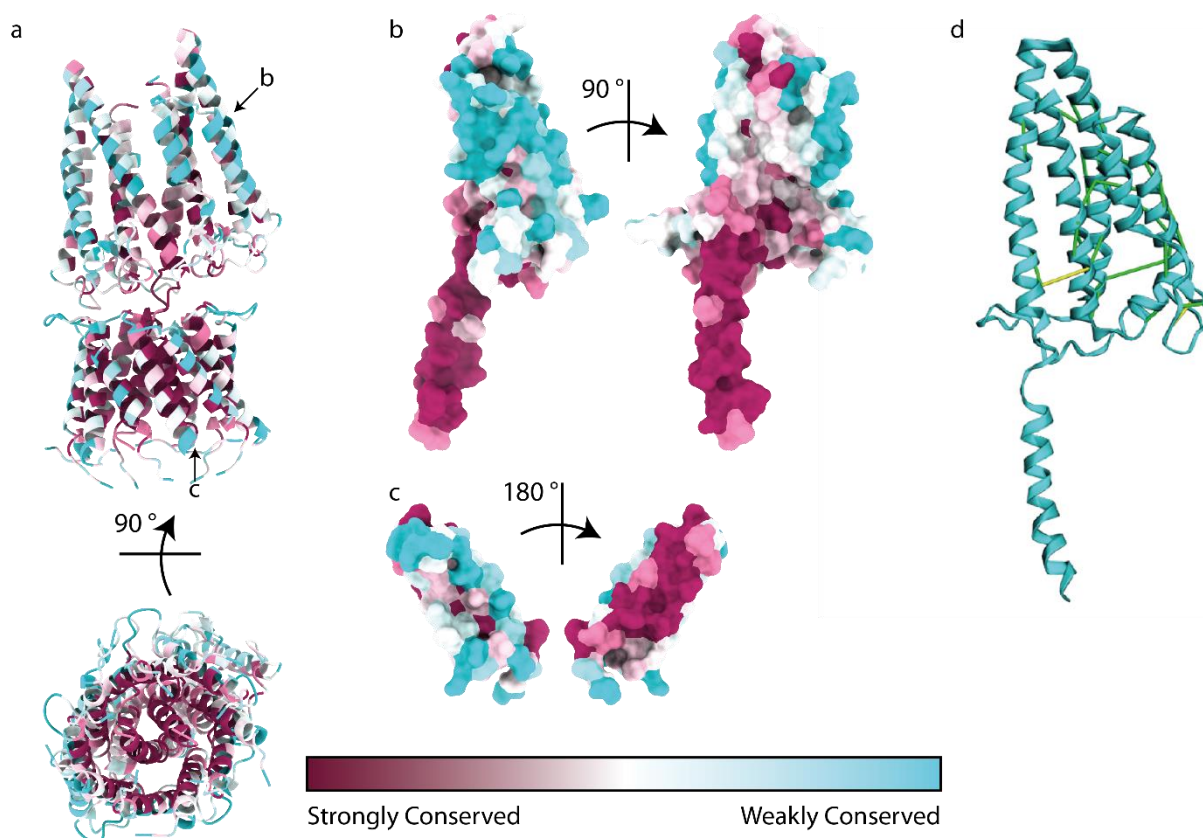


Figure 4-10 Conservation analysis of the *FjoGldLM'* complex. **a-c**, Sequence conservation assessed using ConSurf (Ashkenazy *et al*, 2016). **a** The whole complex in cartoon representation. The chains shown individually in **b** and **c** are indicated with black arrows. **b**, **c** GldM chain B and GldL chain F in surface representation. The left-hand panels show the chains in the same orientation as the upper panel of **a**. **d** Co-varying residue pairs in the first periplasmic domain of GldM identified using the Gremlin server (Kamisetty *et al*, 2013). The highest-scoring pairs (score cut-off 1.63) are shown on the structure of the GldLM' complex. Contacts with a minimum atom-atom distance of <5 Å are shown in green, and 5–10 Å in yellow. No high-scoring GldM-GldM intermolecular pairs were observed for D1. Dimerisation of the other periplasmic domains is supported by co-variance with the following strongly co-varying pairs identified in the other domains D2:Val253-Ile284, Phe248-Glu-293, Tyr236-Arg261; D3: Thr327-Ala349, Asp331-Lys371; D4: Ala456-Asp483. Adapted from (Hennell James *et al*, 2021).

The contacts between the transmembrane helices of *FjoGldL* and *FjoGldM'* are tight, leaving no channel for water or protons to flow through (**Figure 4-8b**). However, there are several polar or protonatable residues in the transmembrane helices of both *FjoGldL* and *FjoGldM'* (**Figure 4-11**). *FjoGldL* Ser56 is the only one of these residues which points into the detergent micelle rather than into the helical bundle. Sequence conservation analysis showed that most of these residues are strongly conserved (**Figure 4-11b**). The strong conservation of these polar and protonatable residues suggests that they are important to function of the *FjoGldLM* complex. Their unusual placement in the transmembrane region suggests that they might be involved in facilitating proton flow and the conversion of the PMF into mechanical work. To further investigate the roles of these residues, I carried out *in vivo* experiments using *F. johnsoniae* described in Chapter 5.

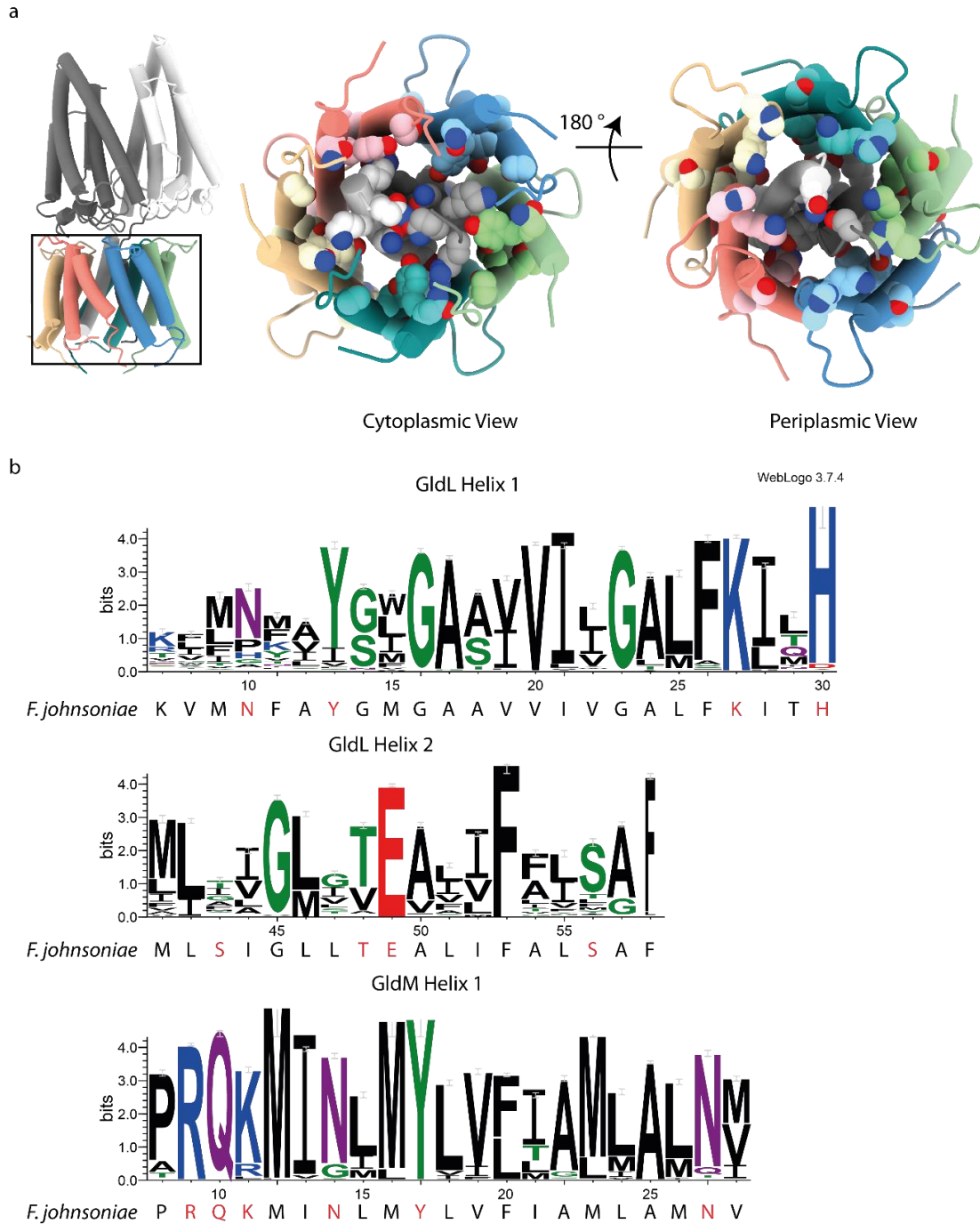


Figure 4-11 Conserved polar residues are found in the transmembrane region of *FjoGldLM*'. **a** Cytoplasmic and periplasmic views of the transmembrane helices of *FjoGldLM* with polar residues shown in sphere representation. For clarity, the D1 domains of *FjoGldM*' are hidden. **b** Sequence logos for the transmembrane helices of GldL and GldM, generated with WebLogo 3.7.4 (Crooks *et al*, 2004). The height of each letter corresponds to how common that residue is at that position. Residue numbers are from *F. johnsoniae* and the *F. johnsoniae* sequence is shown below with the residues shown as spheres in **a** coloured red.

4.6 Discussion

The combination of using truncated GldLM constructs and using the pT12 vector for protein expression instead of the pETDuet-1 derived pRHJ004 vector led to purification of *PgiPorLM'*_A, *FjoGldLM'*_A, and *FjoGldLM'*_B proteins and the determination of a structure of *FjoGldLM'* at 3.9 Å resolution. Although it expressed well, the *PgiPorLM'*_A construct did not have visible D1 domains in 2D class averages and a cryoEM structure could not be solved. This suggests that the D1 domains were not stably folded in this construct. Leone *et al.* were able to solve the crystal structure of a construct that ended at the same position (Lys216, (Leone *et al.*, 2018)), but this required a camelid nanobody to stabilise, suggesting it may also not have been rigidly folded. When I later made a construct that extended past the predicted end of this helix, I was able to solve its structure (Chapter 6). The *FjoGldLM'*_A construct, which was truncated to the equivalent position within the helix, seemed to be more stable and could be used to solve a structure, although analysis by SEC suggested that it was more prone to aggregate than the *FjoGldLM'*_B construct which terminated at the end of this helix (Figure 4-3b, Figure 4-4a&b). This may indicate that *FjoGldM* is more stably packed in this region than *PgiPorM*.

Consistent with published crystallographic and SEC-MALS analysis, there were two copies of *FjoGldM'* in the *FjoGldLM'* structure (Figure 4-8, (Vincent *et al.*, 2017; Leone *et al.*, 2018)). However, there were five copies of *FjoGldL*. This is inconsistent with the report (Vincent *et al.*, 2017) that the cytoplasmic domain of *PgiPorL* forms a trimer in isolation. My later structure of *PgiPorLM'*_C rules out the possibility that *FjoGldLM* and *PgiPorLM* form complexes with different stoichiometries (see Chapter 6). It is possible that the transmembrane region of GldL forms pentamers whilst the cytoplasmic domain forms trimers, but closer examination the published work suggests the SEC-MALS data in that paper may have been misinterpreted. The authors used a *PgiPorL*_{cyto} construct with predicted mass of 29 kDa and observed peaks with calculated masses of 113 kDa and 44 kDa, which they interpreted as

a trimer (predicted mass 87 kDa) and monomer, respectively. The authors suggested that the discrepancy between the predicted and calculated masses was due to the domain having an extended shape, but SEC-MALS calculates absolute masses which are unaffected by shape. It seems likely therefore that these calculated masses correspond to larger complexes, possibly a pentamer and a dimer, that had undergone degradation and were not actually of the predicted size. As in the full length *PgiPorLM* structure, density corresponding to the cytoplasmic domain of *FjoGldL* was not visible, suggesting that it is either disordered or mobile and preventing a structural assessment of its multimerization state.

The D1 domains of *FjoGldM'* were splayed apart in the cryoEM *FjoGldLM'* structure, whilst they were parallel in the crystal structures of *FjoGldM_{peri}* and *PgiPorM_{peri}* (**Figure 4-9** (Leone *et al*, 2018; Sato *et al*, 2020)). None of these constructs are entirely physiological as the *FjoGldLM'* structure lacks the D2, D3, and D4 domains of *FjoGldM*, whilst the *FjoGldM_{peri}* and *PgiPorM_{peri}* structures lack the TMHs of *FjoGldL* and *FjoGldM* or *PgiPorL* and *PgiPorM*, respectively. The full length *PgiPorLM* structure, *PgiPorM_{peri}* structure and conservation analysis of the D1 domain suggest that the splayed conformation in the *FjoGldLM* structure is the more valid conformation. It is possible that the parallel arrangement of D1 domains is an intermediate conformation during the motor complex's function, but the lack of co-varying residues within the crystallographic interface makes it unlikely it is the physiological resting state.

The TMHs of the *FjoGldLM'* complex are all tightly packed together, suggesting that no one subunit can move in isolation and any rearrangement will be coordinated across the entire complex. This would allow sidechain rearrangements due to proton flow to drive movements of all subunits, which can in turn be amplified by the long periplasmic domains of *FjoGldM* to do mechanical work in secretion and motility. Within these tightly packed interfaces there are several well-conserved polar

or protonatable residues. These residues are almost all oriented toward the interior of the complex, where they can interact and facilitate proton flow to generate movement. The roles of these residues in T9SS and gliding motility is the subject of the next chapter.

5 Mutational analysis of *FjoGldLM*

Much of the work described in this chapter was published in (Hennell James *et al*, 2021). The work of other authors is acknowledged in the text. Where indicated, some figures were adapted from initial drafts produced by RHJ.

5.1 Introduction

In Chapter 4 I presented the structure of *FjoGldLM*' at 3.9 Å resolution. This structure showed that several polar residues were present in the transmembrane helices of GldL and GldM. Conservation analysis indicated that some of these were well-conserved (**Figure 4-11**). It is unusual to find polar residues in transmembrane helices, which are usually surrounded by the hydrophobic lipid membrane. Both T9SS and gliding motility are known to be dependent on the proton motive force (PMF), and as GldL and GldM are the only inner-membrane proteins known to be conserved across T9SS⁺ bacteria, it is believed that the GldLM complex conducts protons across the inner membrane to convert the electrochemical potential energy of the PMF into mechanical work (Shrivastava *et al*, 2013; Vincent *et al*, 2017). I therefore reasoned that these polar residues might be involved in conducting protons across the inner membrane and set out to probe their role in T9SS and gliding motility.

To investigate the role of the polar residues I had identified in the transmembrane helices of the *FjoGldLM*' structure, I generated strains of *F. johnsoniae* with point mutations in *gldL* and *gldM*. *F. johnsoniae*, together with *P. gingivalis*, has also been used as a model organism for T9SS since the system was first described in 2010 (Sato *et al*, 2010), and gliding motility has been studied in *F. johnsoniae* for over 40 years (Pate & Chang, 1979). *F. johnsoniae* can therefore be used to study the phenotypic effects of mutations on both T9SS and gliding motility. In general, *F. johnsoniae* grows faster than *P. gingivalis*, is aerobic rather than an obligate anaerobe, has more genetic tools available, and is non-pathogenic in humans so only bio-safety level 1 procedures are necessary. *F. johnsoniae* is

hence a more convenient model organism than *P. gingivalis*. I therefore chose to conduct all my *in vivo* experiments with *F. johnsoniae*, rather than mutating those residues conserved in *P. gingivalis*.

5.2 Design and generation of mutants

I made strains with chromosomally encoded individual substitutions of each of the residues identified in Chapter 4. All residues were substituted with alanine, which has a small, non-polar side chain. GldL Tyr13 and GldM Tyr17 were also substituted with phenylalanine to examine functional effect of the hydroxyl groups of these residues. Analysis of the *FjoGldLM'* structure suggested there might be a salt bridge between one copy of GldL Glu49 and one copy of GldM Arg9 (discussed further below). To probe the roles of these residues in more detail, I substituted the negatively charged GldL Glu 49 with glutamine, which has the same size and shape but no charge, and aspartate, which has the same charge as glutamate but has a shorter side chain. I also substituted GldM Arg9 with lysine, which is also positively charged but has a different shape, and with glutamate, which is negatively charged. I generated these mutant strains using the mobilizable suicide vector pYT354, as described in 2.4 (Zhu *et al*, 2017).

5.3 Protein stability

Once I had generated mutant strains, I assessed whether GldL and GldM were properly expressed by immunoblotting of cell pellets (Figure 5-1). All the GldL variants were expressed, but GldL_{E49A} migrated faster than the WT GldL protein (GldL_{WT}), suggesting it may have undergone proteolysis (Figure 5-1a). GldL_{E49Q} also migrated slightly faster than GldL_{WT}, this small change may be due to minor proteolysis, or it may be due to the loss of charge from Glu49 affecting the variant protein's electrophoretic mobility. Later experiments showed that neither GldL_{E49A} nor GldL_{E49Q} were functional, so I did not investigate the aberrant migration of these variant proteins further.

Most of the GldM variants were stably expressed, but no band was visible corresponding to GldM_{R9A} and the band corresponding to GldM_{R9E} was very weak. This is likely because proper insertion of a

transmembrane helix requires a positive charge at the cytoplasmic end (von Heijne & Gavel, 1988), without this charge these variant proteins could not be properly inserted and so were degraded. I conducted most of the below experiments concurrently with these immunoblots, so the *gldM-R9A* and *gldM-R9E* strains are included in several of the figures, but as would be predicted, had the same phenotypes as the Δ *gldM* strain.

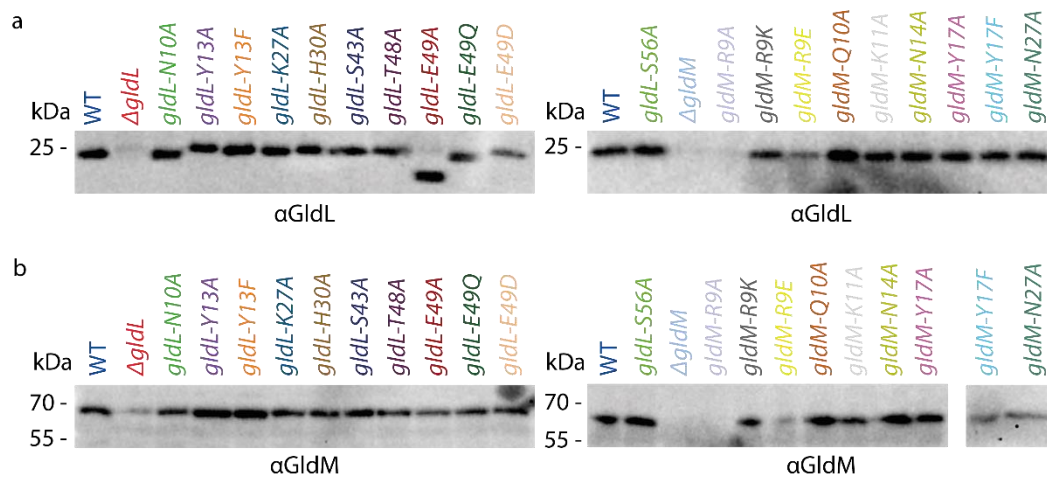


Figure 5-1 Most proteins with substituted polar residues in the transmembrane region are stably expressed. Cell pellets of the indicated strains were analysed by immunoblotting using α GldL (a) or α GldM (b) primary antibodies. The labelling colours used for each strain are used throughout this chapter. The immunoblots shown are representative of three biological repeats.

GldL was not stable in the absence of GldM (Figure 5-1a, right panel) and GldM was less stable in the absence of GldL (Figure 5-1b, left panel). This suggests that the GldLM complex must be properly assembled for GldL and GldM to be stable. In the absence of their partner proteins GldL and GldM probably either misfold and aggregate or are recognised as incorrectly inserted due to their exposed polar residues. In either case they will be removed from the membrane and degraded by quality-control pathways. In the strains where the variant protein was well-expressed, the partner protein was also well-expressed, indicating that the variant proteins could form GldLM complexes and stabilising their partners.

5.4 Assessment of gliding motility on agar and glass

Having established which variant proteins were stably expressed, I assessed the phenotypic effects of the *gldLM* point mutations on the strains' ability to move by gliding motility. I first assessed the ability of these mutant strains to glide on agar using a colony spreading assay. This serves as a convenient assay to assess gliding ability in lots of strains at once. WT colonies spread to about 15 mm in diameter after 48 h and had a ruffled edge morphology. Non-gliding $\Delta gldL$ and $\Delta gldM$ cells had much more compact colonies with a "coffee-ring" morphology caused by the drying of the spotted culture (Mampallil & Eral, 2018). However, the *gldL-Y13A*, *gldL-Y13F*, *gldL-K27A*, *gldL-E49A*, *gldL-E49Q*, *gldL-E49D*, and *gldM-Y17A* strains showed no colony spreading. *gldL-H30A*, *gldM-R9K*, and *gldM-Y17F* had intermediate spreading phenotypes. *gldL-H30A* and *gldM-Y17F* had much smaller colonies than the WT strain, whilst *gldM-R9K* had slightly smaller colonies than the WT strain.

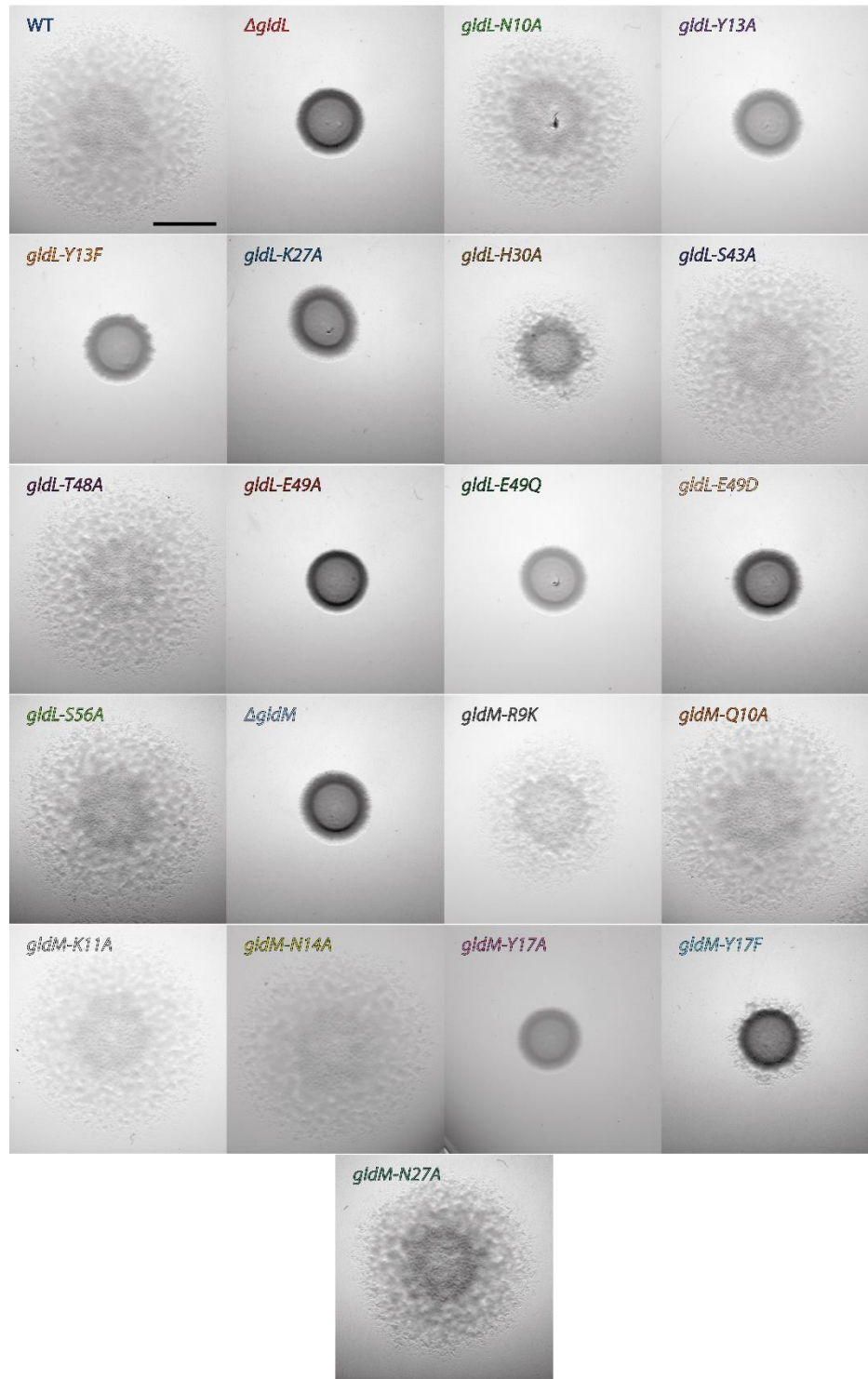


Figure 5-2 Assessment of the colony spreading phenotypes of *gidLM* mutant strains. 2 μ l each of overnight cultures of the indicated strains were spotted onto PY2 agar and imaged after 48 h incubation at 25 °C. Scale bar is 5 mm. Images are representative of at least three biological repeats.

Table 5-1 Summary of gliding motility and T9SS phenotypes of *gldLM* mutant strains. Adapted from (Hennell James *et al*, 2021). SprB propulsion data were collected by Andreas Kjær (AK).

Genotype	GldL stability	GldM stability	Gliding on PY2 agar	Gliding on glass	Chitinase assay (arbitrary units $\pm$ s.d.)	SprB propulsion ($\text{nm s}^{-1} \pm$ s.d.)
WT	Present	Present	+++	Glides normally	1.00 ± 0.2	1358 ± 900
ΔgldL	None	Reduced	-	Does not adhere	0.00 ± 0.0	n.d.
<i>gldL</i> _{N10A}	Present	Present	+++	Glides normally	1.09 ± 0.3	n.d.
<i>gldL</i> _{Y13A}	Present	Present	-	Adheres, does not glide	0.41 ± 0.1	275 ± 100
<i>gldL</i> _{Y13F}	Present	Present	-	Adheres, does not glide	1.37 ± 0.5	n.d.
<i>gldL</i> _{K27A}	Present	Present	-	Adheres, does not glide	0.40 ± 0.1	31 ± 20
<i>gldL</i> _{H30A}	Present	Present	++	Glides slowly	1.16 ± 0.2	524 ± 200
<i>gldL</i> _{S43A}	Present	Present	+++	Glides normally	0.96 ± 0.3	n.d.
<i>gldL</i> _{T48A}	Present	Present	+++	Glides normally	1.15 ± 0.6	n.d.
<i>gldL</i> _{E49A}	Clipped	Present	-	Does not adhere	0.00 ± 0.0	n.d.
<i>gldL</i> _{E49Q}	Present	Present	-	Does not adhere	0.00 ± 0.0	n.d.
<i>gldL</i> _{E49D}	Present	Present	-	Very few cells adhere	0.02 ± 0.0	n.d.
<i>gldL</i> _{S56A}	Present	Present	+++	Glides normally	0.88 ± 0.2	n.d.
ΔgldM	None	None	-	Does not adhere	0.00 ± 0.0	n.d.
<i>gldM</i> _{R9A}	None	None	-	Does not adhere	0.00 ± 0.0	n.d.
<i>gldM</i> _{R9K}	Present	Present	++	Glides slowly	0.95 ± 0.2	774 ± 400
<i>gldM</i> _{R9E}	Reduced	Reduced	-	Does not adhere	0.01 ± 0.0	n.d.
<i>gldM</i> _{Q10A}	Present	Present	+++	Glides normally	0.98 ± 0.3	n.d.
<i>gldM</i> _{K11A}	Present	Present	+++	Glides normally	0.91 ± 0.1	n.d.
<i>gldM</i> _{N14A}	Present	Present	+++	Glides normally	1.10 ± 0.3	n.d.
<i>gldM</i> _{Y17A}	Present	Present	-	Adheres, does not glide	0.11 ± 0.0	n.d.
<i>gldM</i> _{Y17F}	Present	Present	+	Glides slowly	1.13 ± 0.4	486 ± 200
<i>gldM</i> _{N27A}	Present	Present	+++	Glides normally	1.30 ± 0.5	n.d.

I then used phase-contrast microscopy to observe individual cells gliding on glass and qualitatively assessed their gliding capability (summarised in **Table 5-1**, videos of selected strains are included with (Hennell James *et al*, 2021)). Microscopic observation of cells on glass shows whether strains are incapable of adhering to a surface or can adhere but cannot move. *F. johnsoniae* cells use different

adhesins to glide on different surfaces (Shrivastava *et al*, 2012), so observing gliding on different surfaces can determine if a mutation causes a defect in secretion and propulsion of one particular adhesin or of adhesins in general. Strains with similar colony sizes to the WT strain also moved at similar speeds (*gldL-N10A*, *gldL-S43A*, *gldL-T48A*, *gldL-S56A*, *gldM-Q10A*, *gldM-K11A*, *gldM-N14A*, and *gldM-N27A*). The strains with partial spreading defects all moved on glass but were noticeably slower than WT cells (*gldL-H30A*, *gldM-R9K*, and *gldM-Y17F*). Of the strains that did not show colony spreading, most adhered to glass but did not glide (*gldL-Y13A*, *gldL-Y13F*, *gldL-K27A*, and *gldM-Y17A*), suggesting that these strains were capable of secreting adhesins but not of propelling them along the cell surface. The *gldL-E49A* and *gldL-E49Q* strain, in contrast, did not adhere to the glass, suggesting that these strains were not capable of secreting adhesins. Very few *gldL-E49D* cells adhered to the glass, and most of them seemed to be only attached at the poles, suggesting that very few adhesins had been properly secreted.

5.5 Assessment of T9SS protein secretion

I next investigated how these mutations in *gldL* and *gldM* polar transmembrane residues affected secretion of T9SS substrates. To analyse the whole secretome, I filtered and concentrated cell culture supernatants then subjected them to SDS-PAGE (Figure 5-3a). Strains which had similar gliding phenotypes to the WT strain also had similar secretome patterns (*gldL-N10A*, *gldL-S43A*, *gldL-T48A*, *gldL-S56A*, *gldM-Q10A*, *gldM-K11A*, *gldM-N14A*, and *gldM-N27A*), suggesting that they were fully competent for T9SS and gliding motility. The strains with null gliding motility phenotypes (*gldL-E49A* and *gldL-E49Q*) had similar secretomes to the Δ *gldL* and Δ *gldM* strains, with several protein bands present in the WT secretome missing. The T9SS-specific bands in the WT secretome were faintly visible in the *gldL-E49D* secretome, suggesting that this strain was capable of a small amount of secretion. Two new bands were visible at high molecular weight in the secretomes of Δ *gldL*, Δ *gldM*, *gldL-E49A*, *gldL-E49Q*, and *gldL-E49D* cells (Figure 5-3a, red dashes). I identified these proteins by

mass spectrometry peptide fingerprinting as Fjoh_2059 and Fjoh_2058. Neither of these proteins are characterised but they are part of a predicted operon with *Fjoh_2052-Fjoh_2057* that contains several genes with sequence similarity to phage genes from other bacterial species, so they may be part of a prophage that is released under stress caused by the non-functional T9SS. Strains which could adhere to glass but did not glide or glided more slowly than the WT strain (*gldL-Y13A*, *gldL-Y13F*, *gldL-K27A*, *gldL-H30A*, *gldM-Y17A*, and *gldM-Y17F*) all had similar secretome patterns to the WT strain, but with the T9SS-specific protein bands being fainter. This suggests that they were capable of secretion but were less efficient due to these substitutions.

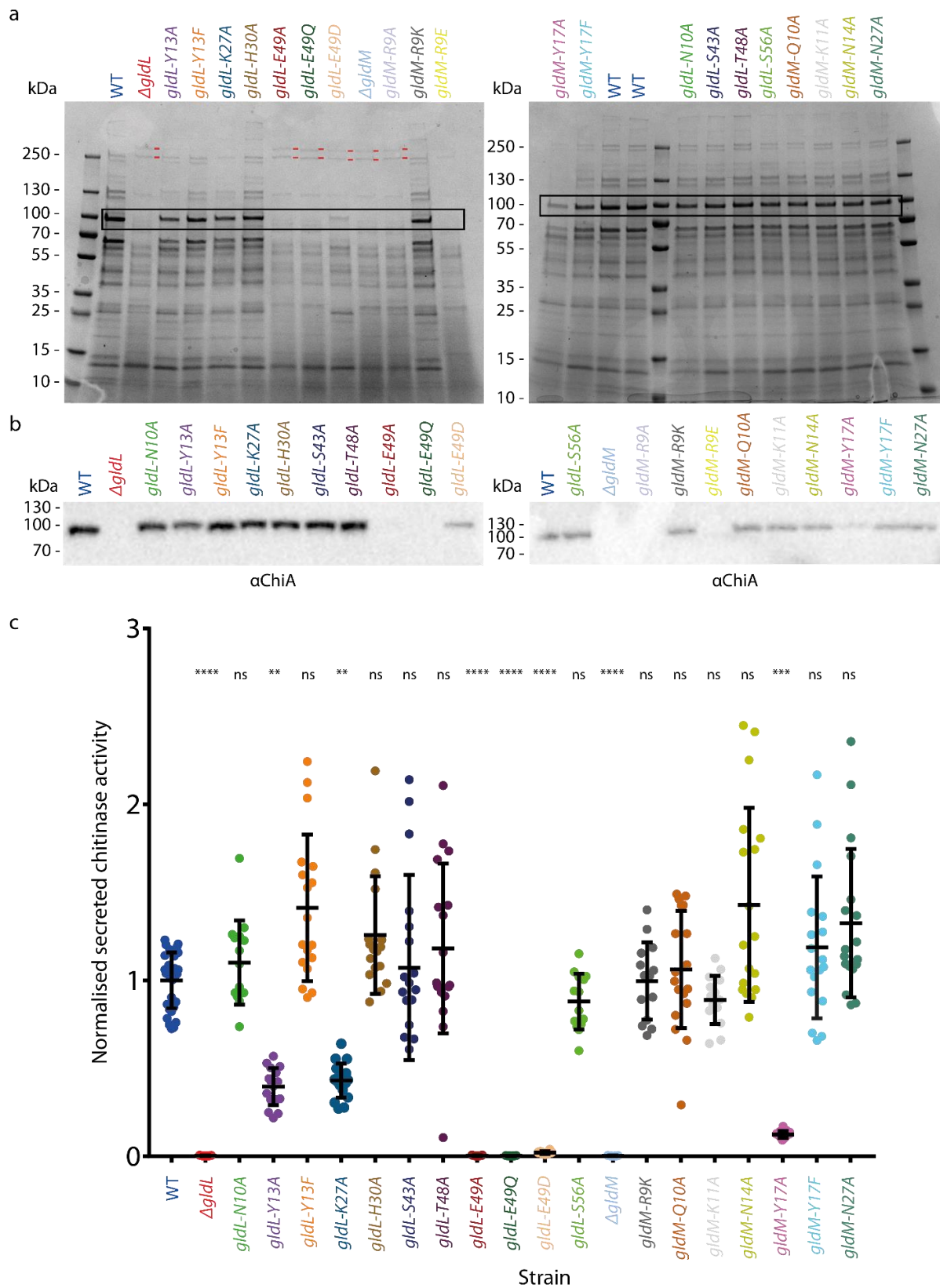


Figure 5-3 Secretome analysis of strains with mutations in *gldL* and *gldM*. (Previous page). **a** SDS-PAGE analysis of cell culture supernatants, stained with Coomassie blue. Black boxes indicate the bands corresponding to the ChiA chitinase, which is secreted by the T9SS. Red dashes indicate proteins secreted by T9SS-deficient strains, discussed in the text. **b** Immunoblot analysis of cell culture supernatants, probed with α ChiA primary antibody. Both **a** and **b** are representative of three biological repeats. **c** Secreted chitinase activity measured using the fluorescent substrate 4-methylumbelliferyl-N,N'-diacetyl- β -D-chitobioside. Horizontal bars indicate the mean for each strain and error bars represent one standard error. Significance of differences relative to the WT strain were assessed using a Kruskal-Wallis test followed by Dunnet's multiple comparison test. Ns - not significant, * - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$, **** - $p < 0.0001$. At least three technical replicates of at least three biological repeats were performed for each strain.

I then analysed secretion of the T9SS substrate ChiA, a chitinase that is one of the most abundant proteins in the secretome, visible as a distinct band by SDS-PAGE with Coomassie staining (**Figure 5-3a**). Immunoblotting showed a similar pattern to the SDS-PAGE analysis (**Figure 5-3b**), in that the totally adhesin-deficient strains ($\Delta gldL$, $\Delta gldM$, *gldL-E49A*, *gldL-E49Q*, *gldM-R9A*, and *gldM-R9E*) secreted no ChiA, whilst those capable of adhering to glass secreted at least some ChiA (**Figure 5-3b**). Immunoblotting seemed to have less dynamic range than SDS-PAGE analysis. Most of the bands were of similar intensity, except for those for the *gldL-E49D* and *gldM-Y17A* strains, despite differences in the intensity of Coomassie-stained bands.

To more quantitatively levels of secreted chitinases in spent culture media, I used a fluorescence-based assay for chitinase activity. This assay uses the substrate analogue 4-methylumbelliferyl-N,N'-diacetyl- β -D-chitobioside, which can be cleaved by endochitinases to release a fluorescent moiety. Previously published work with *F. johnsoniae* suggests that essentially all secreted chitinase activity is T9SS-dependent, so chitinase activity should be a good readout for T9SS activity (Sato *et al*, 2010). The results of this assay showed a high level of variability, probably due to pipetting errors when using a multichannel pipette with small volumes. In future experiments this might be ameliorated by using a pipette designed for smaller volumes or redesigning the experiment to avoid the need to transfer such small volumes. Nevertheless, the overall results were consistent with the other assays of ChiA secretion. As expected $\Delta gldL$, $\Delta gldM$, *gldL-E49A*, *gldL-E49Q*, *gldM-R9A*, and *gldM-R9E* all displayed no activity. Strains that behaved the same as the WT strain in previous

assays showed no significant difference from the WT strain in this assay. Strains that had intermediate secretome SDS-PAGE phenotypes also had intermediate chitinase activities. *gldL-Y13A* and *gldL-K27A* both had roughly 40% of the activity of the WT strain, whilst *gldM-Y17A* and *gldL-E49D* had activities of 11% and 2%, respectively. *gldL-Y13F*, *gldL-H30A*, and *gldM-Y17F*, which all had motility defects, had the same chitinase activity as the WT strain. This may indicate that the differences between these strains and the WT strain are below the sensitivity of this assay to detect, or it may be the case that since ChiA (~100 kDa) is much smaller than gliding motility adhesins (SprB = ~650 kDa) it can be more easily secreted by a partly functional T9SS. It may also be the case that GldLM function is more of a limiting factor for gliding motility than it is for protein secretion. This apparent discrepancy is discussed further below.

5.6 Analysis of SprB propulsion

Interpretation of the effects of *gldLM* mutations on gliding motility is confounded by the fact that gliding motility requires GldLM function for both secretion and propulsion of adhesins. To determine the effects of the mutations on adhesin movement specifically, Andreas Kjær (AK) developed a single molecule assay to directly track individual adhesin molecules. He constructed a fusion between the SprB adhesin and a HaloTag protein which can be covalently labelled with a fluorescent dye (Los *et al*, 2008). Using a membrane-impermeable dye ensures that only molecules which have been secreted are labelled. Adjusting the dye concentration to label only one adhesin molecule per cell allows direct measurement of the movement of individual molecules.

AK collected adhesin motility data for *gldL-Y13A*, *gldL-K27A*, *gldL-H30A*, *gldM-R9K*, and *gldM-Y17F*, all of which had a defect in motility on glass and agar but still secreted ChiA. Strains *gldL-H30A*, *gldM-R9K*, and *gldM-Y17F*, which all showed some colony spreading, propelled SprB more slowly than WT cells (Figure 5-4). Strains *gldL-Y13A* and *gldL-K27A*, neither of which showed colony spreading or gliding on glass, could both propel SprB along their surface, but much more slowly than WT cells

(Figure 5-4). Thus, these latter mutants are still capable of secreting and propelling adhesins, even though they cannot drive whole cell motility. This difference in phenotypic severity may be because the cells in the adhesin tracking assay were immobilised and so the moving adhesins were not under any load. The variant complexes may not be capable of propelling adhesins under load when they are engaged with a surface or of propelling enough adhesins to move the whole cell.

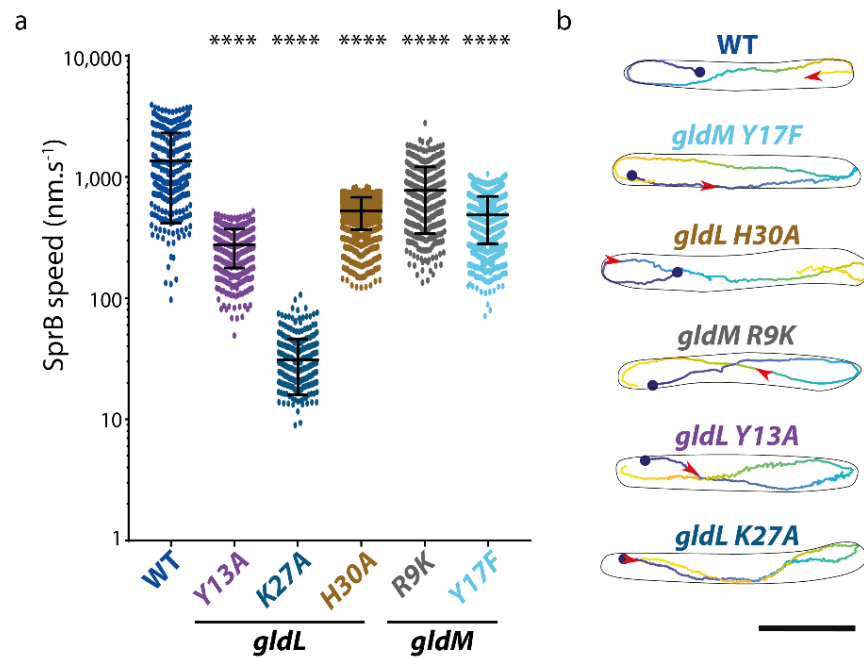


Figure 5-4 Measurement of SprB propulsion in selected strains. A HaloTag-SprB fusion protein was fluorescently labelled and observed by fluorescence microscopy. Adapted from ref. (Hennell James *et al*, 2021). Data were collected and tracking analysis was carried out by AK. Panel **a** was prepared by RHJ. Panel **b** was prepared by AK. **a** Average speed of movement for individually tracked SprB molecules. Error bars represent 1 standard error. Means were compared to the WT strain using a one-way Kruskal-Wallis test followed by Dunnet's multiple comparison test. **** - $p < 0.0001$. At least three biological repeats were carried out for each strain. **b** Representative tracks of single HaloTag-SprB molecules. Black lines represent the cell outline. Tracks represent the entire path recorded for one molecule, coloured from blue (beginning) to yellow (end). Red arrowheads represent the position recorded for each molecule 3.5 s after the beginning of the track. Scale bar 2 μm .

5.7 Interpretation of *gldLM* mutant phenotypes in the context of the *FjoGldLM'* structure

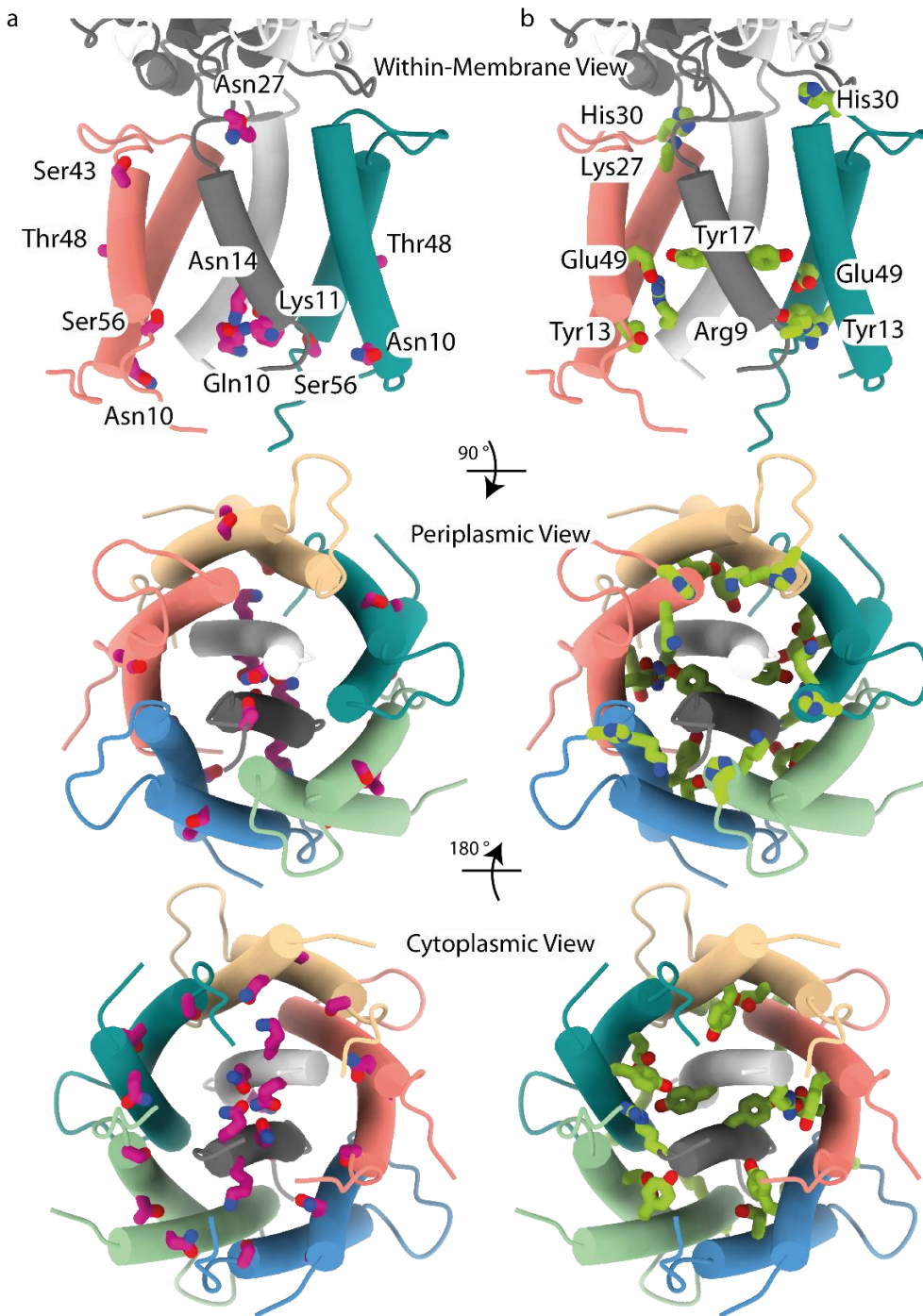


Figure 5-5 Positions of the residues substituted in this chapter on the *FjoGldLM'* structure. **a** Residues that were substituted without effect on gliding motility or type IX secretion shown as sticks with carbon atoms in purple. **b** Residues that affected gliding motility and/or type IX secretion shown as sticks with carbon atoms in green. In the top panels GldL Chains D, E, and G are hidden for clarity.

Some of the polar residues found in the transmembrane region of *FjoGldLM'* had effects on type IX secretion and gliding motility when substituted but others did not. I next compared the positions within the transmembrane region of *FjoGldLM'* of these two groups of residues.

The residues which did not influence secretion and motility when substituted were clustered at the ends of the transmembrane helices, except for GldL Thr48, which is approximately in the middle of GldL TMH2 (**Figure 5-5a**). Of these residues, GldL residues Ser43, Thr48, and Ser56 are oriented toward adjacent copies of GldL in the ring. GldM residues Gln10 and Asn14 form a hydrogen bonding network between the two copies at the cytoplasmic end of TMH1 (**Figure 5-5a**, bottom panel) whilst GldM Asn27 interacts with the other copy of itself at the periplasmic end of TMH1 (**Figure 5-5a**, middle panel). Of the residues that were unimportant for secretion and motility, only GldL Asn10 and GldM Lys11 appear close enough to interact between GldL and GldM (**Figure 5-5a**, bottom panel). These residues which had no phenotype when individually substituted may be important in aggregate for stabilising different conformations of GldLM and facilitating their rearrangement but individually dispensable. It would be interesting to test this in future work by combining substitutions or by targeting substitutions to break up the GldM Gln10-Asn14 and GldM Asn27-Asn27 interaction hydrogen bonding networks.

The residues of GldL and GldM that I experimentally identified as being functionally important form two clusters located in the cytoplasmic half of the TMH bundle and at the interface between the TMH domains and the GldM periplasmic D1 domains (**Figure 5-5b**). GldL Lys27 and His 30 are near the periplasmic interface and may be involved in coordinating motions of the periplasmic and transmembrane domains. Alternatively, they could be involved in shepherding proton flow into the transmembrane helices. The structure of *FjoGldLM'* does not show an obvious route through the

complex for protons (**Figure 4-8**), but in a different conformation there might be room for Lys27 to carry a proton deeper into the TMH bundle or for a hydroxonium ion to move between residues.

In the cluster of important residues toward the cytoplasmic face of the transmembrane helix bundle there is a potential salt bridge between Glu49 of GldL Chain C and Arg9 of GldM Chain B (**Figure 5-5b**). The sidechain density for Glu49_C is poor (**Figure 5-6a**), probably due to the fact glutamate sidechains are usually damaged by the electron beam (Hattne *et al*, 2018), but the Arg9_B sidechain can be confidently modelled pointing toward the Glu49_C position (**Figure 5-6c**). Indeed, the N-terminal end of the GldM Chain B transmembrane helix is bent upward toward Glu49_C compared to Chain A (**Figure 5-6d**), indicating that the salt bridge has caused rearrangements of the protein main chains. The side chain of aspartate is shorter than that of glutamate so will not be able to properly form this salt bridge despite conserving the side chain negative charge, explaining why the *gldL-E49D* mutant strain had essentially no T9SS or gliding motility function.

GldM Chain A Tyr17 and GldL Chain C Tyr13 can both be confidently modelled due to their bulky nature (**Figure 5-6 a & b**) and appear to bracket the salt bridge (**Figure 5-5b**, top panel). They may help stabilise the salt bridge or mediate other interactions in different conformations. For both these tyrosine residues, substitution with alanine had a more severe effect on type IX secretion than substitution with phenylalanine (**Figure 5-3**). In the case of GldM Tyr17, the *gldM-Y17A* strain showed no gliding motility whilst the *gldM-Y17F* strain did show gliding motility, albeit significantly less than the WT strain. This suggests that both the bulky phenyl rings and the hydroxyl groups of these two tyrosine residues are important for their function, so the role of these residues probably involves both hydrogen bonds and hydrophobic packing interactions. Superimposition of all five copies of GldL by alignment of TMH1 shows that TMH2 of GldL Chain C, which forms the salt bridge, is shifted toward the periplasm compared to the other copies of GldL (**Figure 5-6e**) and the periplasmic loop

between the helices is also rearranged (**Figure 5-6f**). This rearrangement does not appear to affect the orientations of Lys27 and His30, but Tyr17_c points in a different direction from the other copies, orienting its hydroxyl group toward the salt bridge (**Figure 5-6g**). The 5:2 stoichiometry of the *FjoGldLM'* complex is inherently asymmetrical, so only one such salt bridge can form at a time. Cycles of proton flow might disrupt and reform this salt bridge for alternating copies of GldM, which could lead to rotation in a model described below.

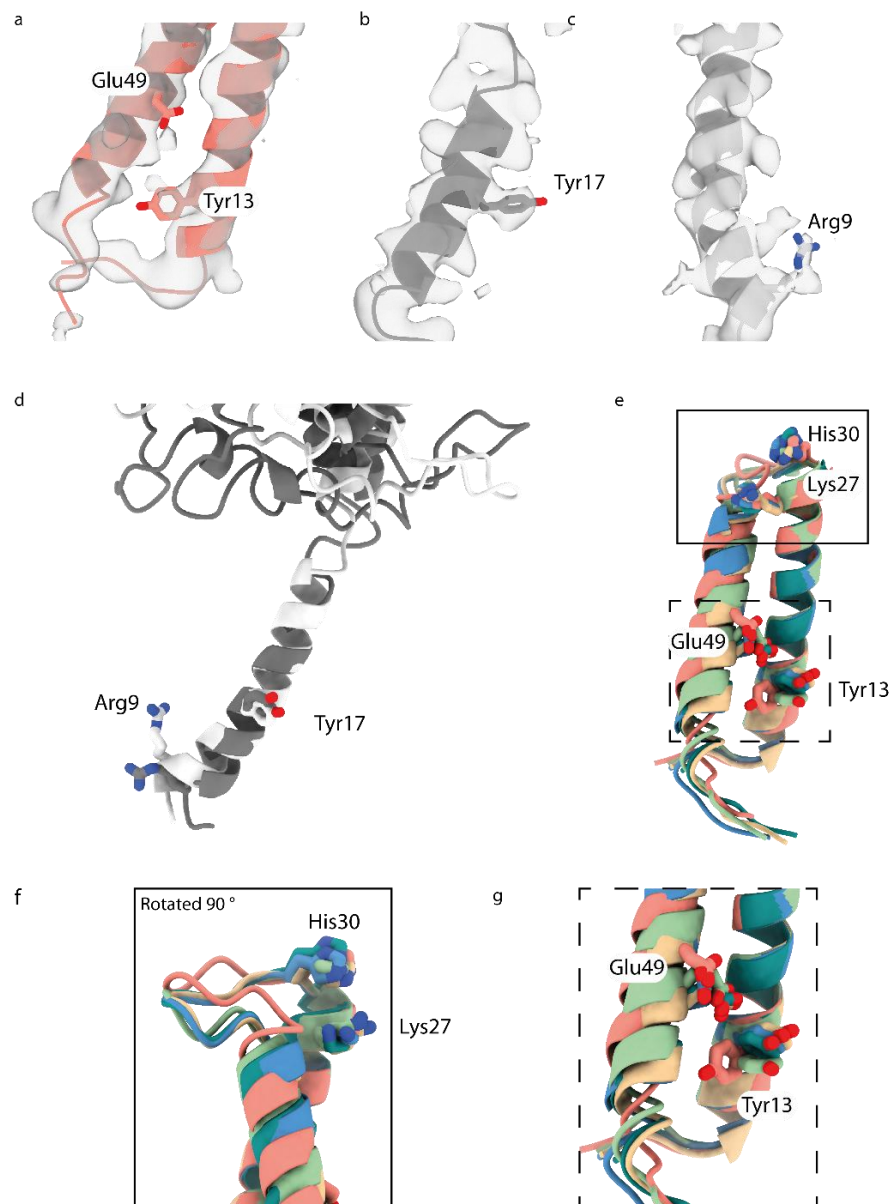


Figure 5-6 The potential salt bridge between GldM Chain B Arg9 and GldL Chain C Glu49 causes rearrangements of GldM Chain B and GldL Chain C. a-c The GldL Chain C (a) and GldM Chains A (b) and B (c) structural models overlaid with cryoEM density. **d** GldM Chains A (grey) and B (white) aligned according through 13-23. **e** All five copies of GldL aligned through residues 10-27 (TMH1). The boxed regions are expanded in **f** and **g**.

5.8 Discussion

In Chapter 4 I identified polar and protonatable residues in the transmembrane helices of GldL and GldM, several of which are well conserved (Figure 4-11). In this chapter I probed the role of these

residues in the T9SS and gliding motility by targeted mutagenesis in *F. johnsoniae*. Several of these residues appeared to be dispensable, at least individually, but several did have an impact on the T9SS and gliding when substituted with non-polar residues.

Several of the mutants that did have a phenotype (*gldL-Y13A*, *gldL-Y13F*, *gldL-K27A*, *gldM-R9K*, *gldM-Y17A*, and *gldM-Y17F*) had less severe defects in T9SS than in gliding motility (**Figure 5-2** & **Figure 5-3**). This may indicate that secretion and motility require the GldLM complex to act in two different ways that require different residues. However, this seems unlikely as these residues are conserved in both gliding and non-gliding species. Another possibility is that secretion and gliding require different levels of power output from the GldLM complex. The gliding motility adhesins such as SprB (~650 kDa) are much larger than ChiA (~100 kDa), so probably require more energy to secrete. Once secreted, ChiA is released into the environment, but adhesins must be propelled along the cell surface to generate motility. Experiments performed by AK with HaloTag-SprB showed that these strains with intermediate phenotypes were capable of secreting and propelling SprB but propelled it more slowly than WT cells. This suggests that the GldLM complex still functioned in this strain but either it did so less efficiently, or fewer functioning complexes assembled so less power was available for SprB propulsion.

All the methods I used to assess protein secretion were effectively end-point assays after overnight growth, when the cells would be in stationary phase. It would be interesting to compare protein secretion by the different mutant strains at different growth phases. A luminescence-based real-time assay for T9SS secretion has recently been developed in the Berks lab by Dr Xiaolong Liu (XL). It would be very interesting to examine the mutant strains I generated using this assay to assess differences in secretion rate rather than just total protein secreted between the strains.

GldL Glu49_C and GldM Arg9_B appear to form a salt bridge in the *FjoGldLM*' structure (**Figure 5-5, Figure 5-6**), although the resolution of the structure is not sufficient to be certain of the orientation of Glu49_C. Glu49 proved to be essential for secretion and motility – substitution with glutamine which lacks a negative charge totally removed activity whilst substitution with aspartic acid, which has a negative charge but is smaller, resulted in detectable but negligible activity (**Figure 5-2, Figure 5-3**). Arg9 could be substituted with lysine, which is also large and positively charged, but this strain did have reduced colony spreading and slower HaloTag-SprB propulsion (**Figure 5-2, Figure 5-3, and Figure 5-4**).

The GldL₅M₂ complex is inherently asymmetric, so only one copy of GldM can form the GldL Glu49 to GldM Arg9 salt bridge at a time and forming a new salt bridge with the other copy of GldM would require rearrangement of the complex. This salt bridge is dependent on Glu49_C being negatively charged and Arg9_B being positively charged. If Glu49_C were protonated and/or Arg9_B were deprotonated as protons were translocated through the complex, the salt bridge would be broken. This might “unlock” the conformation of the complex and allow the side chains on the other side of the complex to rearrange and form a new salt bridge between GldM Arg9_A and GldL Glu49_D or Glu49_G, accompanied by a rotation of 36 ° (**Figure 5-7a**). Another cycle of protonation and deprotonation would lead to a new salt bridge forming with GldM Arg9_B, driving another 36 ° rotation. 10 such cycles would lead to a full 360 ° rotation (**Figure 5-7**, see also (Hennell James *et al*, 2021) Supplementary Video 8), making GldLM a proton-driven rotary motor. It is not possible to say without direct observation in which direction GldLM rotates, or if it undergoes continuous rotation or switches back and forth between two positions. Rotation of the GldM dimer at the centre of the complex would then be transmitted to the outer membrane components of the T9SS and gliding motility machinery via the extended periplasmic domains.

Figure 5-7 A model of *FjoGldLM* rotation. (Previous page). **a** Schematic representation of the model. The GldM Arg9 to GldL Glu49 salt bridge (purple dashed line) on one side of the TMH bundle is broken for one copy of GldM and made for the other copy as protons flow across the membrane, driving a 36 ° rotation. Further proton flow breaks this second salt bridge and makes a third with the original copy of GldM, driving a further 36 ° rotation. Eight further rotations give a total rotation of 360 °. Anticlockwise rotation is shown but clockwise rotation is equally plausible based on the structure of *FjoGldLM*'. **b** Within-membrane, periplasmic, and cytoplasmic views of the same three positions shown in **a**. UCSF ChimeraX was used to generate a morph between copies of *FjoGldLM*. Selected frames from that morph are shown here, the full morph is available in Supplementary Video 8 of (Hennell James *et al*, 2021).

Exactly how rotary motion of the GldLM complex is converted into secretion and motility is as yet unclear. In *P. gingivalis* the GldM homologue PorM has been shown by bacterial two hybrid and co-immunoprecipitation to interact with PorK and PorN, which are encoded in the same operon as PorL and PorM (Vincent *et al*, 2017). This has not yet been confirmed in *F. johnsoniae*, but as the operon structure is conserved it is plausible that the interaction is also conserved. In *P. gingivalis* PorK and PorN have been shown to form multimeric rings 50 nm across attached to the inner face of the outer membrane by the lipid anchor of PorK (Gorasia *et al*, 2016, 2020). T9SS substrates are believed to be pulled through the *P. gingivalis* Sov translocon (equivalent to *F. johnsoniae* SprA) by PorV and its homologues (Lauber *et al*, 2018). Interaction between PorV and the PorKN rings would let PorLM movement drive movement of PorV to pull substrates across the outer membrane. The inner diameter of these PorKN rings is reported as 360 Å, but based on the *PgiPorLM* structure the D4 domain of PorM would describe a circle of only ~180 Å. Further investigation of the interaction between PorM and PorKN is needed to resolve this discrepancy.

In *F. johnsoniae*, it has been reported that cells attached to a surface using SprB antibodies rotate about the point of attachment (Shrivastava *et al*, 2015). The authors suggested that SprB was attached to a rotary motor so that when SprB is immobilised the cell rotates instead. However, direct attachment of the adhesins to a rotary motor is inconsistent with the observation that SprB moves in a helical pattern over the cell surface (Nakane *et al*, 2013). The authors have since proposed that GldLM drives rotation of a gear which propels a mobile tread to which SprB is attached along a helical

track around the cell in a rack-and-pinion mechanism (Shrivastava & Berg, 2020). GldJ, which is homologous to GldK but appears to have a distinct function (Johnston *et al*, 2017), is believed to form the helical track or the tread (Katayama *et al*, 2019; Nakane *et al*, 2013). How GldJ interacts with other components of the gliding machinery and which components move or remain stationary is still unclear. Experience in the Berks group shows that GldJ is recalcitrant to both recombinant expression in *E. coli* and fusion tagging in *F. johnsoniae*, so investigation has been difficult. The interactions of GldM with the outer-membrane components of the gliding machinery have not been well-studied in *F. johnsoniae*, but there are several proteins whose roles are unknown, and which could act to convert rotation or back-and-forth movement of GldM into movement of adhesins on the cell surface.

If the GldLM complex forms an ion-driven motor, it must have a stator and a rotor. In the structurally analogous MotAB flagellar stator complex (see (Deme *et al*, 2020) and 7.2), the central MotB component acts as the stator and binds to the peptidoglycan cell wall to immobilise the complex. However, in the GldLM motor, the central GldM component probably transmits movement to the outer membrane. It must therefore act as the rotor so cannot bind the peptidoglycan. Indeed, there probably needs to be an as-yet-uncharacterised mechanism to generate a hole in the peptidoglycan so that GldM can freely rotate. GldL must therefore act as the stator, but probably does not bind the peptidoglycan or any periplasmic proteins as its periplasmic loops are short and mostly covered by GldM's D1 domains. Presumably, therefore, GldL is held stationary via its cytoplasmic domain. In *P. gingivalis* an interaction between the cytoplasmic tail of PorL and the cytoplasmic PorX protein has been implicated in expression regulation of the T9SS and its substrates , but no interactions with partners that might anchor the complex have been reported.

Two possible objects which GldL could be anchored to are the cytoskeleton and the chromosome. Mutations which disrupted the cytoskeleton would likely be fatal and not be recovered in a mutagenesis screen. Gliding motility in *Myxococcus xanthus*, which uses machinery unrelated to the *F. johnsoniae* machinery, is inhibited by the drug A22, which depolymerises the MreB cytoskeleton (Mauriello *et al*, 2010). In *F. johnsoniae*, however, gliding motility is unaffected by A22 (Shrivastava & Berg, 2020). This suggests either that *F. johnsoniae* MreB is not depolymerised by A22 or that *F. johnsoniae* gliding motility is not dependent on MreB polymerisation, so GldL is not anchored by the cytoskeleton. The cytoplasmic domain of GldL may bind to the nucleoid but does not share homology to any known protein-protein or protein-DNA interaction motifs. I designed various constructs for pulldown experiments to identify binding partners of GldL, but they either did not express well or did not co-precipitate with any binding partners. The constructs I designed are summarised in the Appendix, see 9.1.

The C-terminal end of GldL's cytoplasmic domain is predicted to form a coiled-coil. If coiled coils formed between GldL subunits from different GldLM motor complexes, then when the motors rotated in the same direction the GldL rings would push against each other like interlocking gears and be unable to rotate around the GldM dimers. The GldM dimers, however, could rotate within the effectively immobilised GldL rings. If the five copies of GldL all form coiled coils with separate GldLM complexes, then large arrays of motors could potentially form without the need for additional components. I observed some aggregation of GldLM complexes during purification, but most proteins were in GldL₅M₂ complexes and there was no visible sign of GldL-mediated interactions between complexes on cryoEM grids. This may indicate that inter-motor interactions do not form. Alternatively, it may be that the interactions are not strong under my purification conditions or require proteins absent during recombinant expression to form.

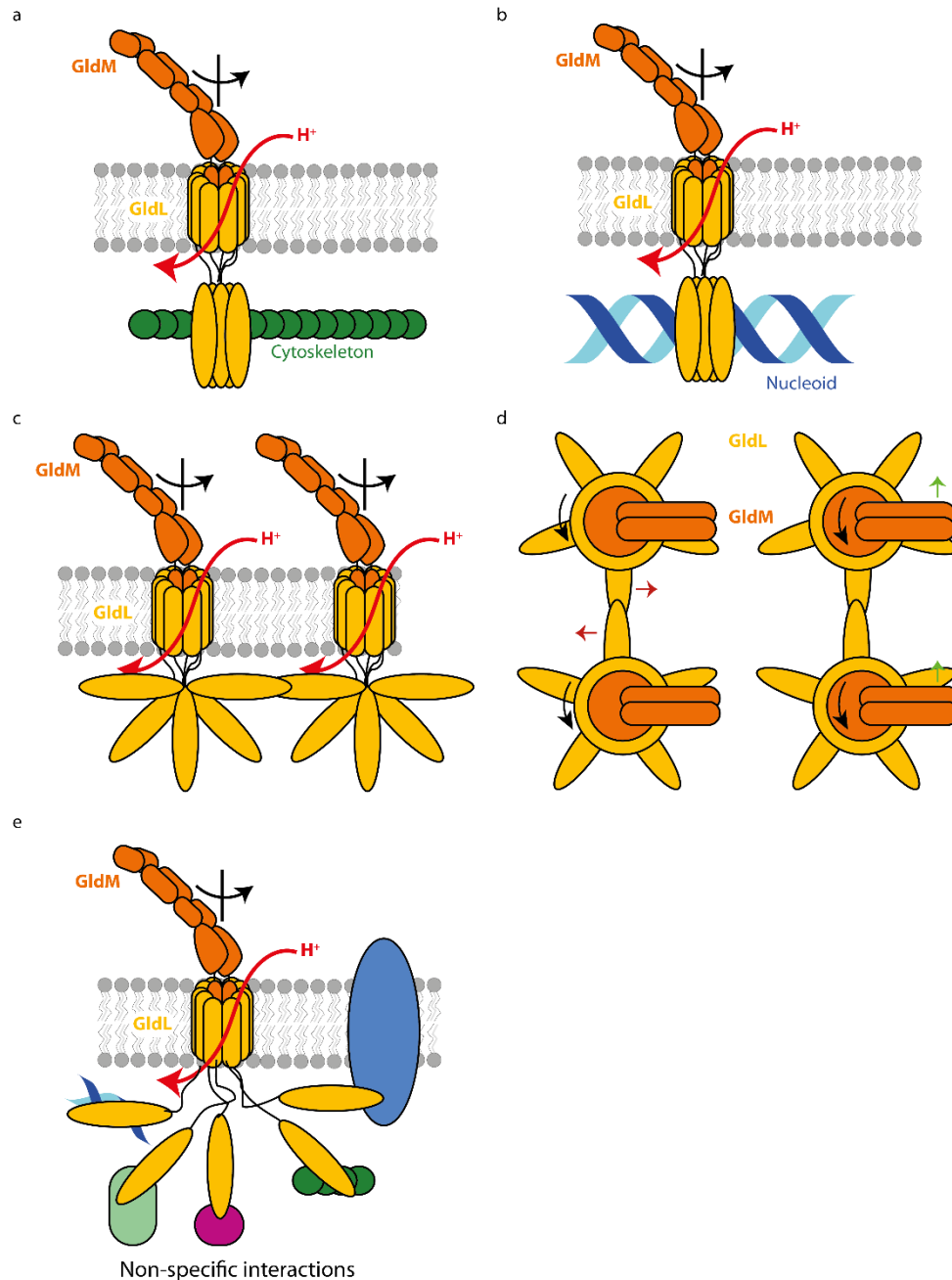


Figure 5-8 Four models for GldLM anchoring. **a** The GldL cytoplasmic domain binds the rigid cytoskeleton. **b** The GldL cytoplasmic domain binds the nucleoid via protein-protein or protein-DNA interactions (only a protein-DNA interaction is shown). **c** GldL subunits from different GldLM motors form coiled coils with each other. **d** Periplasmic view of **c**. If the GldL ring rotates and GldM remains stationary relative to the membrane (left panel), the GldL subunits which form the coiled coil must move in opposite directions and break their interaction. If GldL remains stationary and the GldM dimer rotates relative to the membrane (right panel), GldM can freely rotate. **e** The GldL cytoplasmic domain is extended and partly disordered so makes many non-specific interactions in the cytoplasm. Non-specific interactions with large objects such as the nucleoid, cytoskeleton, ribosomes and membrane-embedded supercomplexes could anchor the GldLM motor.

As well as coiled coils, the cytoplasmic domain of GldL is predicted to form extended helices or be disordered. If either of these predictions are correct, it might simply be the case that many non-specific interactions with large protein complexes, nucleic acids, and other molecules in the cytoplasm combine to make GldL effectively a very large molecular complex and anchor it in this way. These non-specific interactions would need to provide more resistance to the movement of GldL than the resistance that the T9SS and gliding motility machinery provide to the movement of GldM. GldM is assumed to drive motion of components in the outer membrane. The outer leaflet of the outer membrane of *E. coli* is extremely stiff, meaning that integral β -barrel proteins are essentially stationary (Rojas *et al*, 2018; Rassam *et al*, 2015). However, experiments in the gliding bacterium *M. xanthus*, which uses gliding genes unrelated to those of *F. johnsoniae*, have shown that it has a much more fluid outer membrane than *E. coli* (Cao & Wall, 2020). It would follow that as another gliding bacterium *F. johnsoniae* would also have a relatively fluid outer membrane. A fusion of the SprA T9SS translocon with mCherry was observed to be immobile in *F. johnsoniae* (Lauber *et al*, 2018), but it is not clear if this is due to the rigidity of the membrane or anchoring by protein-protein or protein-peptidoglycan interactions. This has not to my knowledge been directly investigated, but the outer membranes of both *F. johnsoniae* and *M. xanthus* have been shown to have high levels of lipids with branched-chain fatty acyl moieties, which are considered to increase membrane fluidity and are necessary for gliding motility (Dzink-Fox *et al*, 1997; Abbanat *et al*, 1986; Godchaux & Leadbetter, 1988; Pitta *et al*, 1989; Ware & Dworkin, 1973). Even so, anchoring GldL would require strong interactions with other objects, so it seems likely that these interactions are specific rather than non-specific.

All four of these possible anchoring mechanisms remain possibilities. How GldLM motor complexes are anchored, as well as the role of the GldL cytoplasmic domain more generally and how the motion of GldM is used to drive secretion and motility remain important avenues for further research.

At 3.9 Å resolution, the structure of *Fjo*GldLM' does not provide high confidence in the orientations of non-bulky side chains and no bound water molecules or ions are visible. Furthermore, a single conformation of the complex only gives limited information about how it might move. The next Chapter describes how I attempted to generate higher resolution structures of the GldLM complex in new conformations.

6 Cryo-electron microscopy structures of GldLM' from four different species

6.1 Introduction

In Chapter 4 I presented a structure of *FjoGldLM'*, the GldLM complex from *F. johnsoniae* truncated to exclude the D2-D4 domains of GldM. In Chapter 5 I showed that several polar residues in the transmembrane helices of GldL and GldM were important for T9SS and gliding motility function. The structure of *FjoGldLM'* was only determined to a resolution of 3.9 Å and showed a single conformation of the complex. Solving the structure at a higher resolution would improve confidence in the orientations of the side chains that are implicated in the motor mechanism and might make ordered water molecules visible. Structures of other complexes have shown that ordered water molecules often form part of “proton wires” that conduct protons from one part of the complex to another. A higher resolution structure might therefore give clues as to the path protons take through the GldLM motor complex. Rotation of the GldLM motor must require structural rearrangements. Solving the structure of the complex in a different conformation would give insight into these rearrangements, although it is possible that the conformation of the *FjoGldLM'* structure is the only stable conformation of GldLM proteins isolated in detergent.

With the aims of solving higher resolution and different conformation structures of the motor, I designed a series of new expression constructs using motor genes from a diverse range of *Bacteroidetes* species. The rationale for doing this was that protein sequences from different species can often behave quite differently in structural studies, even when key residues and the overall fold are conserved. Proteins from thermophilic species are often well-suited to x-ray crystallography projects as are more rigid at crystallisation temperatures than equivalent proteins from mesophilic species which aids packing in crystal growth. Crystal packing is not a concern in cryoEM experiments, but the greater rigidity of proteins from thermophilic species at grid preparation temperatures can

reduce conformational heterogeneity in samples and improve structure resolution. I screened motor sequences from 14 species for recombinant expression in *E. coli* and analysed the most promising samples by cryoEM. To increase the chances of solving the structure of GldLM in different conformations, I also purified and analysed constructs with different length truncations of GldM and with substitutions of some of the functionally important residues identified in Chapter 5.

6.2 Purification of *PgiPorLM'*_C, *SthGldLM'*, and *TroGldLM'*

I initially selected *P. gingivalis*, *S. thermophila*, and *T. rossianum* for protein expression and purification. on the basis that I already had the template DNA available from previous experiments (see Chapter 3). In Chapter 4 I presented a structure of truncated *F. johnsoniae* GldLM (*FjoGldLM'*). I solved this structure using samples of *FjoGldLM'*_A, which extended to GldM residue Lys225, and *FjoGldLM'*_B, which extended to GldM residue Ser232. I also presented the purification of *PgiPorLM'*_A, in which PorM was truncated after residue 216, the predicted final residue of the final α -helix of the PorM D1 domain and the equivalent position of Lys225 in *FjoGldLM'*_A. I could purify a useful amount of *PgiPorLM'*_A, but when I analysed the sample by cryoEM the D1 domains did not appear to be folded (**Figure 4-6**). To address this issue, I modified the position of the truncation in *PgiPorM*. Based on sequence alignment, *PgiPorM* Arg224 is equivalent to *FjoGldM* Ser232, which was the final residue included in the *FjoGldLM'*_B expression construct. To be sure that I had included enough residues so that the final α -helix of *PgiPorM* D1 could fold, I included the *PgiPorM* sequence up to Ser227 in my new construct *PgiPorLM'*_C, which was expressed from plasmid pRHJ118. pRHJ118 included the full length of *P. gingivalis* *porL*, *porM* up to codon 227, and the *porL-porM* intergenic region. A sequence encoding a TEV cleavage site and a TwinStrep tag at the C-terminus of PorM'_C was also included. I also designed the constructs *SthGldLM'* and *TroGldLM'*, with sequences from the thermophilic species *Schleiferia thermophila* and *Thermonema rossianum*, respectively. *SthGldLM'* and *TroGldLM'*

extended to the residues equivalent to *PgiPorM* Ser227. As I did not design constructs for these species equivalent to *FjoGldLM'*_A or *FjoGldLM'*_B, I omit the C subscript for simplicity.

I initially purified these proteins from 2 l each of cell culture using a similar approach to the one I used to purify *FjoGldLM'*_A and *FjoGldLM'*_B with the following modifications. The main changes were that I used Tris-based buffers as recommended by the StrepTactin XT manufacturer instead of PBS-based buffers and that I reduced the concentration of LMNG in the StrepTactin XT elution buffer and SEC buffer to 0.01% to reduce the formation of LMNG strings on concentration. The amount of *PgiPorLM'*_C purified was lower than the amount of *PgiPorLM'*_A I had previously purified (**Figure 6-1a-b**). The purification yields of *SthGldLM'* and *TroGldLM'* were somewhat smaller than that of *PgiPorLM'*_C, but I still purified enough material to prepare cryoEM grids of all three constructs (**Figure 6-1**). I snap froze remaining samples in liquid nitrogen and stored them at -80 °C.

Dr Justin Deme (JCD) collected cryoEM datasets for all three constructs and I processed them as described below in 6.4. After initial data processing, it became clear that the *SthGldLM'* dataset had the potential to produce a high-resolution structure but was limited by poor view distribution (see 6.4.3). Adding fluorinated detergents to buffers just before preparing cryoEM grids can change the view distribution of particles (Lauber *et al*, 2018). The exact mechanism for this effect is not known, but it is thought that the fluorinated detergents affect the interaction of proteins and detergent micelles with the air-water interface and carbon film to induce different preferred orientations. I decided to treat my *SthGldLM'* samples with fluorinated octyl maltoside (fOM) to generate new views of the protein complex. However, adding fOM to protein samples causes protein precipitation, so a higher than usual concentration of sample was needed to ensure good particle density on cryoEM grids. My initial purification yield for *SthGldLM'* was quite low, but I noticed that most of the protein in my *SthGldLM'* purification formed low-molecular-weight peaks, probably consisting of proteins

from *SthGldLM'* complexes which had dissociated (**Figure 6-1d-e**). I reasoned that as *S. thermophila* is a thermophile (Albuquerque *et al*, 2011), the *SthGldLM'* complex might be unstable when purified at 4 °C. I therefore attempted to purify *SthGldLM'* proteins at room temperature, with centrifugation steps carried out at 15 °C to avoid the centrifuges reaching excessively high temperatures. This did not significantly increase my purification yield for *SthGldLM'* (**Figure 6-1d and f**), even when starting from 6 l of cell culture, as the complex seemed more prone to precipitation during concentration steps at room temperature compared to 4 °C. Nevertheless, *SthGldLM'* formed a sharper monodisperse peak when purified at room temperature, with smaller low molecular weight peaks compared to purification at 4 °C (**Figure 6-1d**). I was able to concentrate this sample to $A_{280} = 3.2$ and prepare cryoEM grids with fOM, which led to a structure as described below.

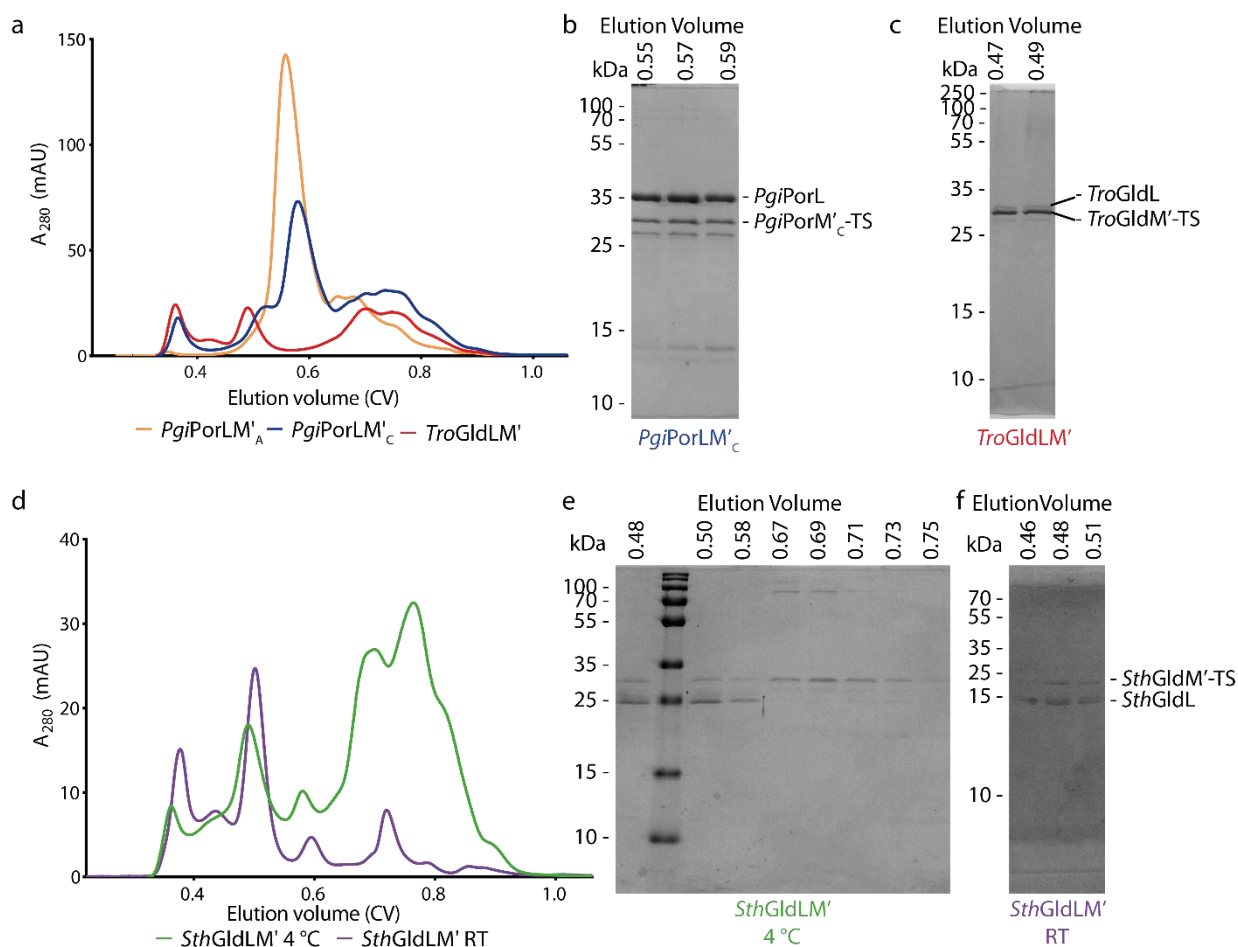


Figure 6-1 Purification of new *P. gingivalis*, *S. thermophila*, and *T. rossianum* truncated GldLM constructs. **a** SEC trace for *PgiPorLM'*_c (blue) and *TroGldLM'* (red). The SEC trace for *PgiPorLM'*_A (orange) is shown for comparison. **b** SDS-PAGE analysis of the *PgiPorLM'*_c fractions used to make cryoEM grids. **c** SDS-PAGE analysis of the *TroGldLM'* fractions used to make cryoEM grids. **d** SEC trace for *SthGldLM'* purifications at 4 °C (green) and room temperature (purple). **e** SDS-PAGE analysis of *SthGldLM'* purified at 4 °C. The 0.48 and 0.50 CV fractions were pooled to make cryoEM grids. **f** SDS-PAGE analysis of the *SthGldLM'* fractions purified at room temperature used to make cryoEM grids after mixing with fOM.

6.3 Expression screening and purification of new GldLM constructs from 14 species

6.3.1 Sequence selection and construct design

To increase the chances of solving a structure with higher resolution than or a different conformation from the *FjoGldLM'* structure I decided to screen sequences from a wider range of species. To do this I submitted the *F. johnsoniae* GldL sequence to the Fold and Function Assignment (FFAS) server (Xu *et al*, 2014; Jaroszewski *et al*, 2005, 2011). The FFAS server predicts various aspects of protein structure

and function, and as part of this conducts PSI-BLAST searches against the nr85s database (Altschul *et al*, 1990; Escapa *et al*, 2018; Kurokawa *et al*, 2007; NCBI Resource Coordinators, 2018). Sequences in this database are filtered so that no sequence is more than 85% identical to any other sequence, increasing the diversity of results. I selected sequences from the results of this PSI-BLAST search that had large insertions or deletions relative to *F. johnsoniae* GldL. I also included sequences from *Flavobacterium psychrophilum*, a fish pathogen closely related to *F. johnsoniae* and *Zobellia galactanivorans*, a marine bacterium which grows on red algae, as gDNA for these two species was already available in the Berks group. The 14 species I examined in this study are summarised in **Table 6-1**. Phylogenetic trees of GldL and GldM sequences show that these species are drawn from across the *Bacteroidetes* phylum (**Figure 6-2**).

To assess whether the species I selected were motile or non-motile, I consulted the publications in which they were described and examined their genomes for homologues to *F. johnsoniae* gliding motility genes. *C. canimorsus*, *F. psychrophilum*, *P. saltans*, *R. anatipestifer*, and *Z. galactanivorans* are all described as motile (Brenner *et al*, 1989; Bernardet *et al*, 1996; Cao *et al*, 2014; Barbeyron *et al*, 2001; McBride & Zhu, 2013). *A. mensalis*, *D. orientale*, *S. wenxiniae*, and *T. rossianum* are all described as non-motile, but all encode homologues of the *F. johnsoniae* gliding motility proteins GldJ and GldH in their genomes, so it is likely that they are in fact motile. It has been shown that species originally described as non-gliding may show gliding motility under the right conditions but examining all possible conditions for all species is extremely time consuming (McBride & Zhu, 2013; Bernardet & Bowman, 2011). *S. thermophila* does not encode a homologue of GldH in its genome but does encode a homologue of the gliding motility proteins GldJ, so it may be able to glide under appropriate conditions. *P. gingivalis* is described as non-motile and does not encode homologues of GldJ or GldH in its genome so is probably truly non-motile. Complete genome sequences are not available for NAT and UEG so I was not able to determine if they are motile.

Table 6-1 Bacteroidetes strains used in this study. The indicated prefixes are used for brevity in construct names and in Figure 6-2.

Strain	Prefix	Description
<i>Apibacter mensalis</i> DSM 100903	Ame	Honeybee gut commensal bacterium (Kwong <i>et al</i> , 2018)
<i>Capnocytophaga canimorsus</i> str. Cc5	Cca	Dog oral commensal bacterium, opportunistic pathogen in humans (Brenner <i>et al</i> , 1989)
<i>Draconibacterium orientale</i> DSM 25947	Dor	Marine bacterium isolated from sediment and a shark carcass (Du <i>et al</i> , 2014)
<i>Flavobacterium johnsoniae</i> UW101	Fjo	Soil bacterium (Stanier, 1947; Bernardet <i>et al</i> , 1996)
<i>Flavobacterium psychrophilum</i> DSM 21280	Fps	Fish pathogen (Bernardet <i>et al</i> , 1996)
<i>Flavobacteriales</i> bacterium isolate NAT55 387147	NAT	Marine bacterium (Tully <i>et al</i> , 2018)
<i>Porphyromonas gingivalis</i> ATCC 33277	Pgi	Human oral pathogen (Shah & Collins, 1988; Coykendall <i>et al</i> , 1980)
<i>Pseudopedobacter saltans</i> DSM 12145	Psa	Soil bacterium (Cao <i>et al</i> , 2014)
<i>Riemerella anatipestifer</i> DSM 15868	Ran	Poultry pathogen (Segers <i>et al</i> , 1993)
<i>Schleiferia thermophila</i> str Yellowstone	Sth	Thermophilic bacterium from a hot spring (Albuquerque <i>et al</i> , 2011)
<i>Sphingobacterium wenxiniae</i> DSM 22879	Swe	Isolated from chemical factory wastewater (Zhang <i>et al</i> , 2012)
<i>Thermonema rossianum</i> DSM 10300	Tro	Thermophilic bacterium from a hot spring (Tenreiro <i>et al</i> , 1997)
Uncultured <i>Sphingobacteriia</i> bacterium sampled from the North Atlantic Ocean	UEG	Identified from a marine metagenome sample (Gómez-Pereira <i>et al</i> , 2012)
<i>Zobellia galactanivorans</i> DSM 12808	Zga	Marine bacterium isolated from a red alga (Barbeyron <i>et al</i> , 2001)

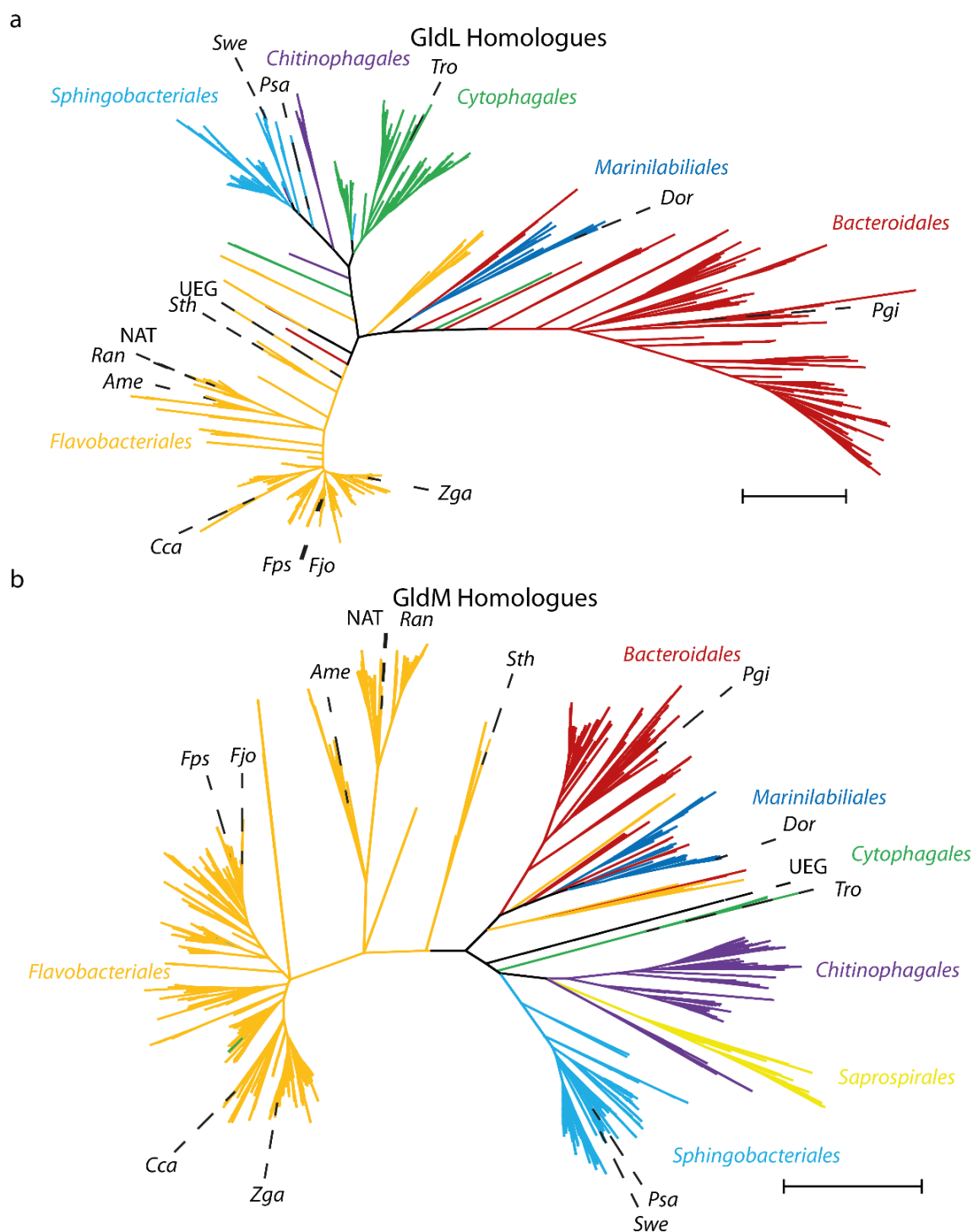


Figure 6-2 Phylogenetic trees of GldL (a) and GldM (b) homologues. Trees were inferred by Maximum-Likelihood from sequence alignments as described in 2.2. Branches are coloured by taxonomic order. Sequences used in this study are indicated with the prefixes defined in Table 6-1. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. Scale bars are of length 0.5 substitutions per site.

For each of the 10 new species I selected, I designed constructs equivalent to *PgiPorLM'*_c designated (*prefix*)GldLM', all expressed from the pT12 vector. In Chapter 3 I showed that the structure of the full length *PgiPorLM* structure could not be solved at high resolution due to preferential orientations of protein particles on cryoEM grids. In the hopes that I could find a homologue of *PgiPorLM* which did not adopt preferential orientations, I designed constructs encoding the GldL homologue and full length GldM homologue for each of the 14 species I studied. I designated these constructs (*prefix*)GldLM or *PgiPorLM* in the case of *P. gingivalis* proteins. I also designed intermediate constructs including GldL and GldM up to the end of the periplasmic D2 ((*prefix*)GldLM'') or D3 ((*prefix*)GldLM''') domains, using the crystal structure of the periplasmic region of *F. johnsoniae* GldM (*FjoGldM_{peri}*) (Leone *et al*, 2018) as a guide. Due to uncertainty in the start codon of *D. orientale* *gldL* I designed two versions of each *D. orientale* construct. A full list of expression plasmids is given in Table 2-5. The expected sizes of proteins from each construct are given in Table 6-2.

Table 6-2 Expected sizes of proteins from the constructs described in this chapter.

Species	Expected molecular mass (kDa)				
	GldL	GldM'-TS	GldM''-TS	GldM'''-TS	GldM-TS
<i>A. mensalis</i>	24.9	30.2	40.2	49.0	59.2
<i>C. canimorsus</i>	24.7	30.4	41.1	50.1	61.1
<i>D. orientale</i>	36.1 (GldL1) 33.5 (GldL2)	32.8	40.5	52.5	63.9
<i>F. johnsoniae</i>	22.8	29.8	39.1	48.1	59.8
<i>F. psychrophilum</i>	23.0	29.8	39.1	48.0	58.7
Isolate NAT55 387147	25.5	35.6	47.2	58.3	71.4
<i>P. gingivalis</i>	34.9	29.3	39.0	48.7	60.7
<i>P. saltans</i>	28.8	29.5	39.4	44.2	59.7
<i>R. anatipestifer</i>	25.7	31.1	41.0	49.6	55.9
<i>S. thermophila</i>	24.2	30.2	40.3	49.7	60.7
<i>S. wenxiniae</i>	23.1	28.5	38.2	47.0	58.1
<i>T. rossianum</i>	29.1	30.7	40.6	45.6	61.5
UEG	28.7	31.9	44.7	51.0	66.4
<i>Z. galactanivorans</i>	23.1	30.1	40.5	49.4	60.6

6.3.2 High throughput expression screening of different length constructs

Large scale purification of all 56 new constructs would be costly in terms of both time and reagents. Therefore, I used a protocol for high-throughput small-scale screening to determine which sequences were well-expressed in *E. coli* and which could be purified as intact GldLM complexes. I used 24-well plates to grow small cultures of cells carrying each plasmid in conditions similar to those I used for large-scale expression. I transferred harvested cells to 96-well plates and used lysozyme and the detergent DDM to lyse cells and solubilise membrane proteins in one step. I then used magnetic StrepTactin XT beads to purify TwinStrep-tagged GldM proteins. I then analysed the elutions from the StrepTactin XT beads by SDS-PAGE to determine the amount and purity of each protein complex. I screened my constructs in four sets (GldLM', **Figure 6-3**; GldLM'', **Figure 6-4**; GldLM''', **Figure 6-5**; GldLM, **Figure 6-6**), and included *PgiPorLM'*_c in each experiment as a reference.

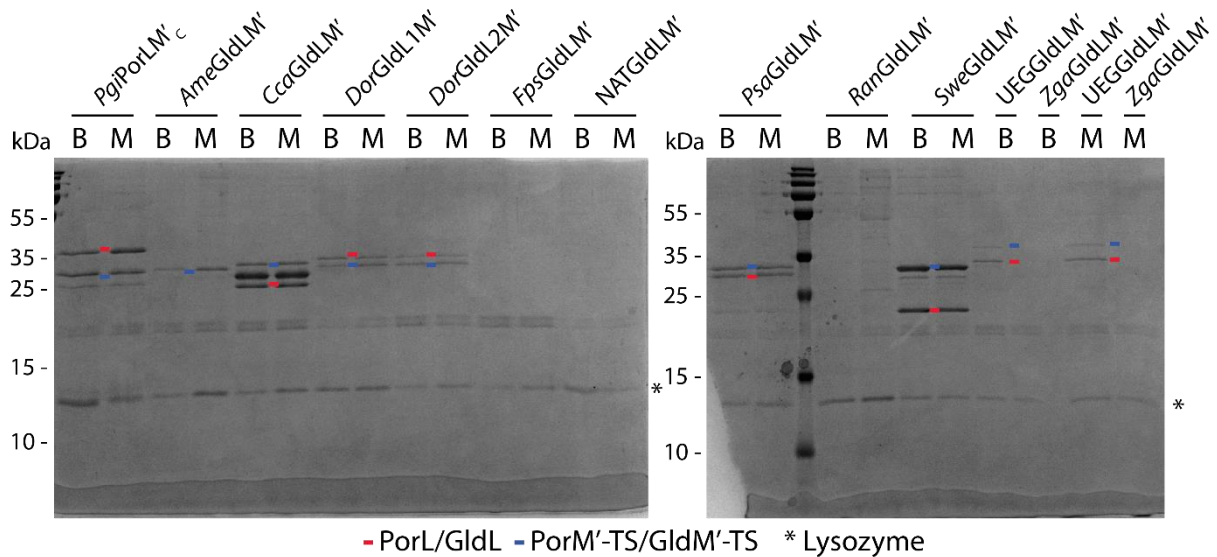


Figure 6-3 SDS-PAGE analysis of an expression screen for GldLM' constructs. The indicated constructs were expressed in either BL21 (DE3) cells (B lanes) or Mt56 (DE3) cells (M lanes). Proteins were solubilised in DDM and magnetic StrepTactin XT beads were used to purify TwinStrep-tagged GldM' proteins. Red and blue lines indicate bands of the size expected for GldL and GldM'-TS, respectively. The *RanGldLM'* vector was found to have a frameshift mutation, so a corrected vector was included in the GldLM'' screen in **Figure 6-4**. * Lysozyme from the lysis buffer.

In the screen of GldLM' constructs, *CcaGldLM'*, *DorGldL1M'*, *DorGldL2M'*, *PsaGldLM'*, *UEGGldLM'*, and *SweGldLM'* produced GldLM' complexes with *CcaGldLM'* and *SweGldLM'* both producing more protein than the *PgiPorLM'*_c reference (**Figure 6-3**). *CcaGldLM'* produced three bands in the range expected for *CcaGldL* and *CcaGldM'*-TS, suggesting that one of these bands represents a partial degradation product of one of the proteins. *AmeGldLM'* had no band corresponding to *AmeGldL*. *DorGldL1M'* and *DorGldL2M'* both produced *DorGldL* bands of the same size, suggesting that the shorter automatically annotated ORF is correct (GldL2). *PsaGldLM'* and *UEGGldLM'* both produced a little less protein than *PgiPorLM'*_c. *FpsGldLM'*, *NATGldLM'*, *RanGldLM'*, and *ZgaGldLM'* all produced no bands of the expected sizes. I later discovered that the *RanGldLM'* vector had a frame-shift mutation that prevented expression, so I generated a corrected construct to include in the GldLM' screen. This construct expressed at a similar level to *PgiPorLM'*_c (**Figure 6-4**). I screened for protein expression in BL21 (DE3) and Mt56 (DE3) cells as these two strains are commonly used for membrane proteins production in the Lea laboratory. The choice of strain did not make much difference to expression levels, which is unsurprising as Mt56 is a BL21 derivative that has a mutation in the gene encoding T7 polymerase, which is not relevant when expressing from the pT12 vector. I purified at larger scale all the constructs for which bands of the expected sizes were visible for both subunits, namely *CcaGldLM'*, *DorGldL1M'*, *PsaGldLM'*, *RanGldLM'*, *SweGldLM'*, and *UEGGldLM'*, as described below.

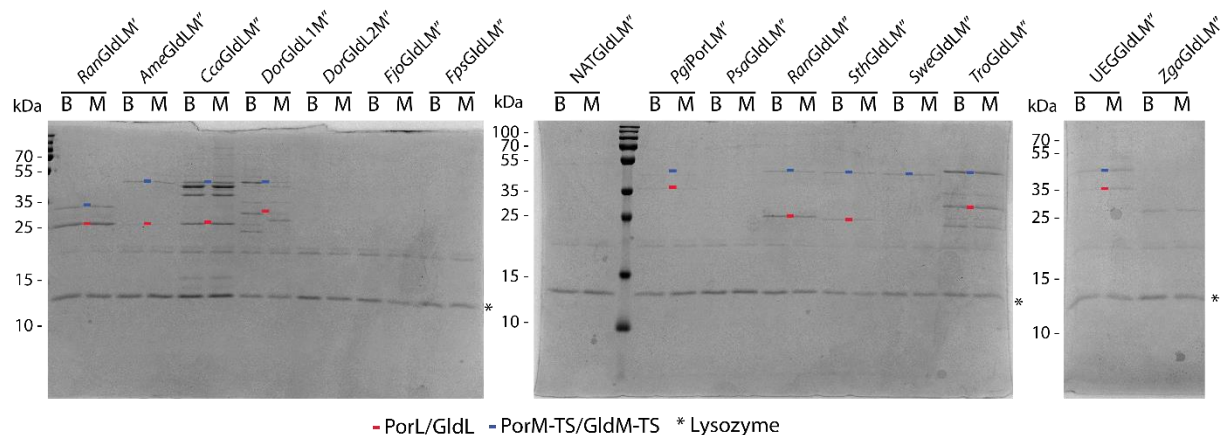


Figure 6-4 SDS-PAGE analysis of an expression screen for GldLM'' constructs. The indicated constructs were expressed in either BL21 (DE3) cells (B lanes) or Mt56 (DE3) cells (M lanes). Proteins were solubilised in DDM and magnetic StrepTactin XT beads were used to purify TwinStrep-tagged GldM'' proteins. Red and blue lines indicate bands of the size expected for GldL and GldM''-TS, respectively. *RanGldLM'* was included as the original vector used in the GldLM' screen was found to have a frameshift mutation. The *PgiPorLM'* reference sample was omitted by mistake. * Lysozyme from the lysis buffer.

In my screen of GldLM'' constructs, which include both the D1 and D2 domains of GldM, there were fewer constructs that expressed well than in the GldLM' screen (**Figure 6-4**). *CcaGldLM''* expressed strongly. Similarly to *CcaGldLM'* (**Figure 6-3**), there was a strong band between the predicted sizes for *CcaGldM''*-TS and *CcaGldL*. In this case the band was much closer to the expected size for *CcaGldM''*-TS than the expected size for *CcaGldL*, suggesting that it was a degradation product of *CcaGldM''*-TS, and that the band seen in **Figure 6-3** was a degradation product of *CcaGldM'*-TS. *DorGldL1M''* expressed in BL21 (DE3) cells produced bands of the expected size, but *DorGldL1M''* expressed in Mt56 (DE3) cells and *DorGldL2M''* expressed from either cell type did not. This was surprising as in the GldLM' screen both *DorGldL1M'* and *DorGldL2M'* expressed similarly in both strains. This may indicate an interaction between the construct sequence and the expression strains or that I made a mistake at some point in the cloning and expression process. Two bands of the expected sizes were visible in the lanes for *PgiPorLM''*, *RanGldLM''*, *SthGldLM''*, *TroGldLM''*, and *UEGGldLM''*, though they all appeared weaker than their GldLM' counterparts, suggesting they were less well expressed. None of the other constructs produced strong bands of the expected sizes for both proteins. Of my GldLM''

constructs, I selected *CcaGldLM''*, *DorGldL1M''*, *RanGldLM''*, *SthGldLM''*, *TroGldLM''*, and *UEGGldLM''* for larger-scale purification.

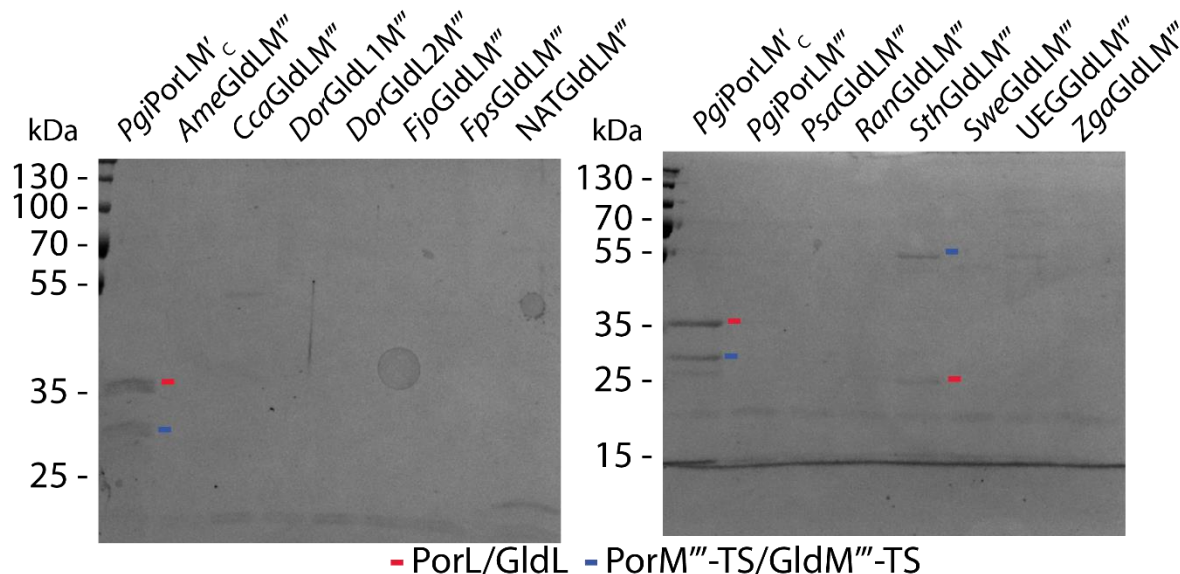


Figure 6-5 SDS-PAGE analysis of an expression screen for GldLM''' constructs. The indicated constructs were expressed in BL21 (DE3) cells. Proteins were solubilised in DDM and magnetic StrepTactin XT beads were used to purify TwinStrep-tagged GldM''' proteins. Red and blue lines indicate bands of the size expected for GldL and GldM'''-TS, respectively.

The choice of BL21 (DE3) or Mt56 (DE3) cells as the expression strain did not seem to make much difference to expression levels in the GldLM' or GldLM'' expression screens, so I only used BL21 (DE3) cells in my GldLM''' and GldLM expression screens. In the GldLM''' construct screen (**Figure 6-5**), two bands of the expected size were only visible for *SthGldLM'''*, and they were weaker than the *PgiPorLM'c* reference lane. This may indicate that I selected the wrong domain boundaries for the D3 domain for all species, or that the D3 domain cannot properly fold in *E. coli* in the absence of the D4 domain. In either case, the expression of all GldLM''' constructs was very poor, so I did not select any of them for larger scale purification.

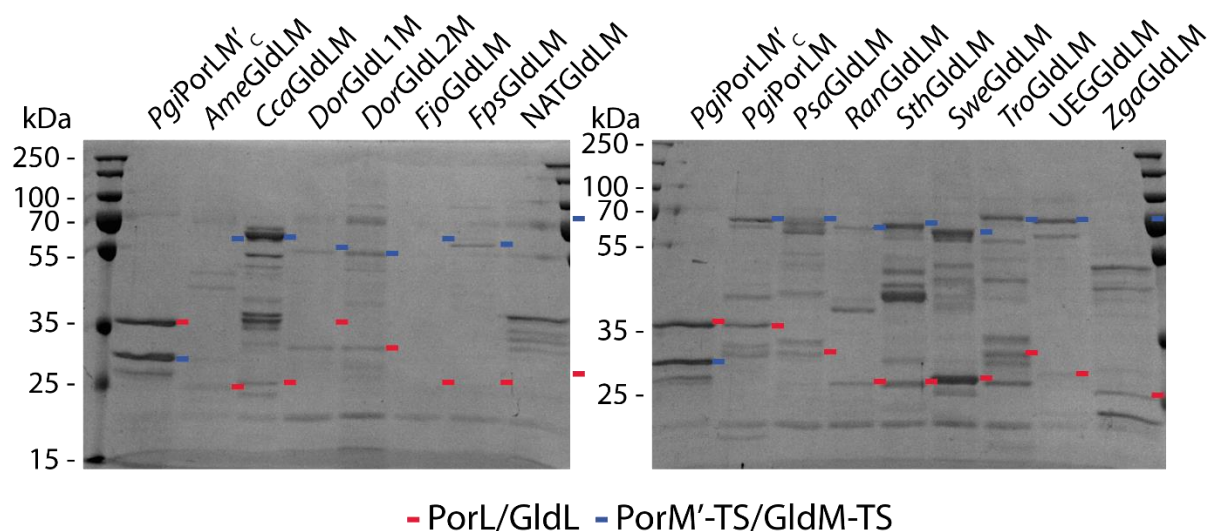


Figure 6-6 SDS-PAGE analysis of an expression screen for GldLM constructs. Red and blue lines indicate the size expected for GldL and GldM-TS, respectively.

In my screen of full length GldLM constructs (**Figure 6-6**), strong bands of the expected sizes were visible for *CcaGldLM*, *DorGldL2M*, *PgiPorLM*, *PsaGldLM*, *RanGldLM*, *SthGldLM*, *TroGldLM*, and *SweGldLM*, although for *DorGldL2M* and *TroGldLM* the bands of the expected sizes were not much stronger than the contaminating bands. Weak bands of the expected size were visible for *DorGldL1M*. Bands of the expected sizes for both proteins were not visible for *AmeGldLM*, *FjoGldLM*, *FpsGldLM*, and *UEGGldLM*. Of the full length GldLM constructs, I selected *CcaGldLM*, *PgiPorLM*, *PsaGldLM*, *RanGldLM*, *SthGldLM*, and *SweGldLM* for larger scale purification.

6.3.3 Purification of different length GldLM constructs

Once I had selected which new constructs to purify based on my expression screens I began purifying them at larger scale from 2 l of cell culture each, using the same protocol I used for the *PgiPorLM'*_c and *TroGldLM'* constructs. The behaviour of the constructs in these larger scale purifications was similar to that in the expression screen (**Figure 6-7a**). No strong bands of the expected sizes were visible on SDS-PAGE for *UEGGldLM'* after StrepTactin XT purification, so I did not analyse this sample by SEC. On SEC (**Figure 6-7b-c**) *CcaGldLM'*, *PsaGldLM'*, and *RanGldLM'* produced monodisperse peaks

at similar retention volumes to those seen for *FjoGldLM'*_B and *SthGldLM'*. *DorGldL1M'* appeared to have undergone some proteolysis, no band was visible at the expected size for *DorGldL1* (36.1 kDa), but a prominent band was visible at 25 kDa. The proteolyzed *DorGldL1M'* complex still formed a monodisperse peak which eluted ~0.1 CV later than *CcaGldLM'*, suggesting that the complex remained intact. As several other constructs were more promising in terms of yield, I did not investigate this proteolysis further or analyse *DorGldL1M'* by cryoEM. So much *SweGldLM'* protein eluted from the StrepTactin XT cartridge that I analysed it over two separate SEC runs, which produced identical peaks. *SweGldLM'* produced a peak with a trailing shoulder that may contain protein from dissociated complexes. I prepared cryoEM grids from the main peaks for *CcaGldLM'*, *PsaGldLM'*, *SweGldLM'*, and *RanGldLM'*, but due to constraints of time and expense only the *SweGldLM'* sample was analysed by cryoEM (see 6.4.6).

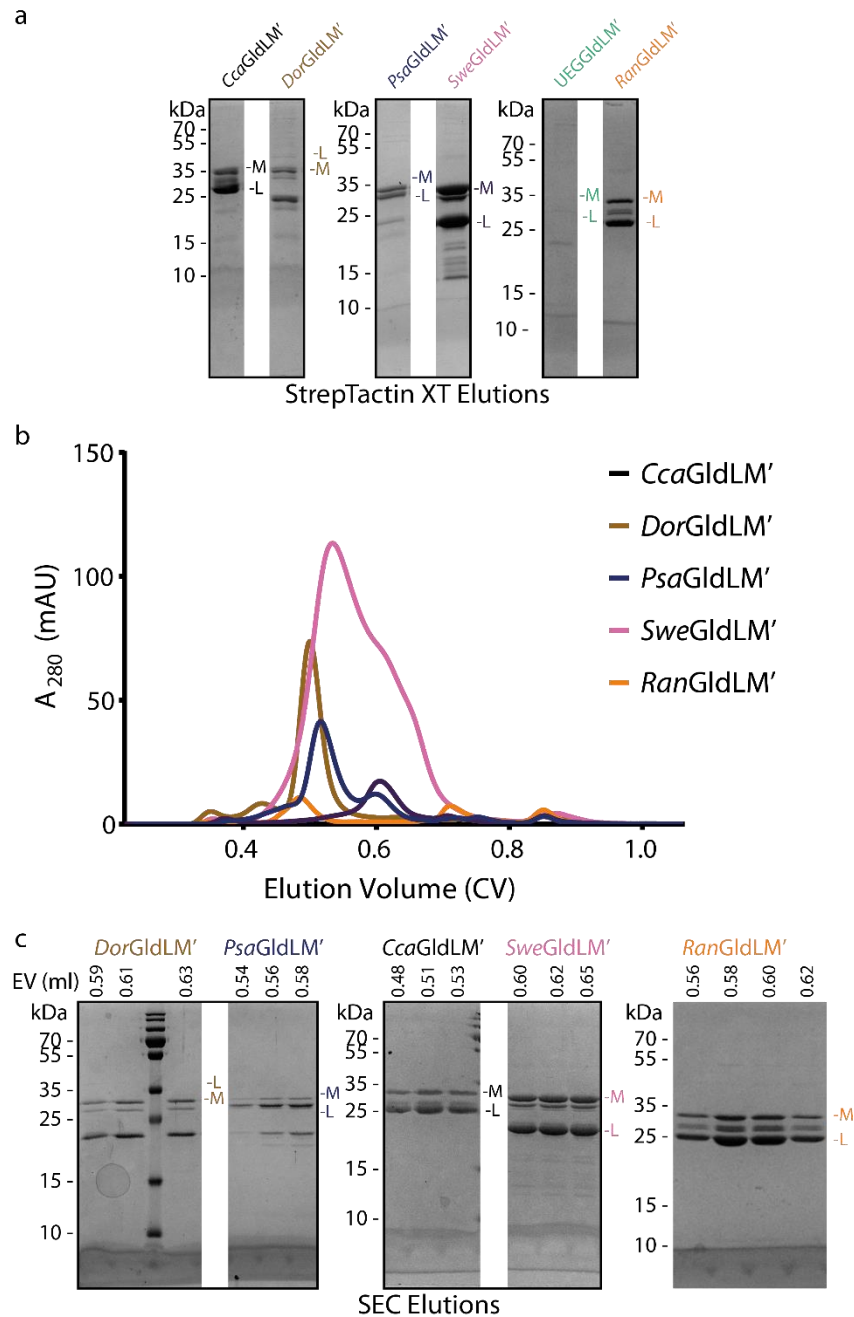


Figure 6-7 Large scale purification of GldLM' constructs. **a** SDS-PAGE analysis of StrepTactin XT elution fractions for the indicated constructs. **b** SEC traces for the indicated GldLM' constructs. All traces represent the total protein purified from 2 l of cell culture, except for *SweGldLM'* which uses material from 1 l of cell culture. **c** SDS-PAGE analysis of SEC fractions from **b** for the indicated constructs. The fractions shown on the gels were pooled and used to prepare cryoEM grids. EV – elution volume. Black boxes indicate lanes from the same gel. L and M indicate the expected sizes of GldL and GldM'-TS, respectively, for each construct. *DorGldLM'* underwent proteolysis, discussed in the text.

After purifying the new GldLM' constructs, I purified the longer GldLM'' and GldLM constructs in the same manner. After elution from StrepTactin XT cartridges, bands of the expected size were either not visible or very weak for *DorGldLM''*, *PsaGldLM''*, *SthGldLM''*, or *UEGGldLM''* (**Figure 6-8a**), so I did not analyse these samples by SEC. The StrepTactin XT elution fractions for *TroGldLM* were of such low protein concentration that I did not analyse them by SDS-PAGE or SEC. When analysed by SEC, the peaks for *SthGldLM* and *TroGldLM''* were very small so I did not analyse the elution fractions by SDS-PAGE (**Figure 6-8b**). On *CcaGldLM''* complexes eluted from the column slightly earlier than *CcaGldLM* complexes, whilst *RanGldLM''* complexes eluted from the column slightly later than *RanGldLM* complexes (**Figure 6-8b-c**). These small shifts in elution volume are probably due to using different columns and FPLC systems and so may not be meaningful. The main peak for *PgiPorLM* was only slightly larger than that of *SthGldLM* and *TroGldLM''* but did contain both PorL and PorM. *SweGldLM* once again produced the highest yield of these longer constructs, although it was somewhat less than the yield for *SweGldLM'* (**Figure 6-7b** and **Figure 6-8b**). Whilst the SEC peak for *SweGldLM'* had a trailing shoulder (**Figure 6-7b**), the peak for *SweGldLM* had a leading shoulder (**Figure 6-8b**). This suggests that the *SweGldLM* complex had mostly dissociated, confirmed by the fact that fractions in the shoulder contained bands for *SweGldL* and *SweGldM-TS* on SDS-PAGE with similar staining intensity, but fractions from the main peak contained much stronger bands for *SweGldM-TS* than *SweGldL*. This suggests that the *SweGldLM* complex was not properly assembled in *E. coli* or was unstable and dissociated, so that mostly free *SweGldM-TS* was purified. I prepared cryoEM grids of *CcaGldLM''*, *CcaGldLM*, *RanGldLM''*, *RanGldLM*, and *SweGldLM*, but due to constraints of time and expense only analysed the *CcaGldLM''* and *CcaGldLM* samples by cryoEM (6.4.4 and 6.4.5 below).

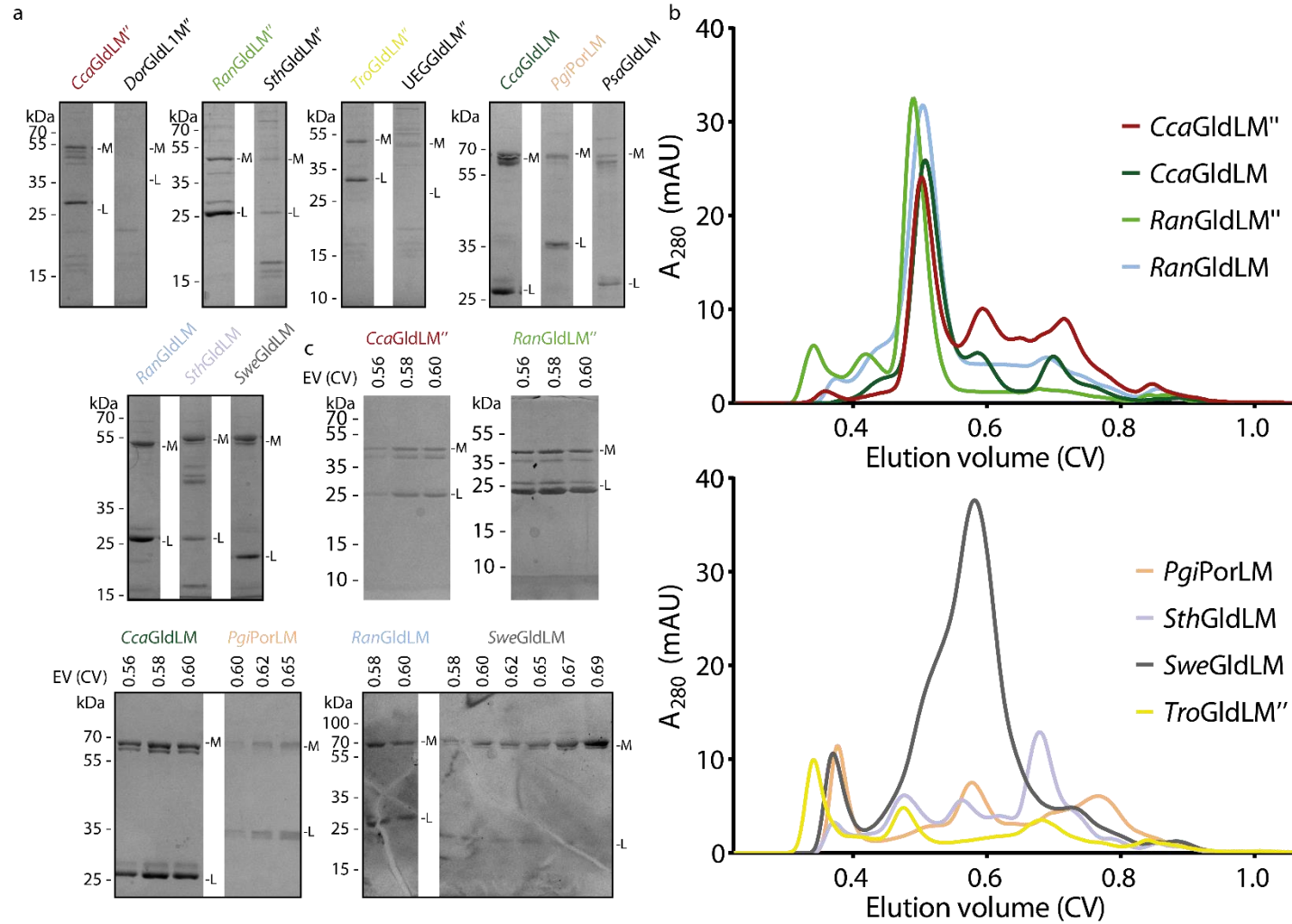


Figure 6-8 Large scale purification of GldLM'' and GldLM complexes. **a** SDS-PAGE analysis of StrepTactin XT elution fractions for the indicated constructs. **b** SEC traces for the indicated constructs. The traces are split over two graphs for clarity. **c** SDS-PAGE analysis of SEC elution fractions for the indicated constructs. EV – elution volume. Black boxes indicate lanes from the same gel. L and M indicate the expected sizes of GldL and GldM''-TS/GldM-TS, respectively.

6.4 Data processing of six GldLM cryoEM datasets

I analysed samples of *TroGldLM'*, *PgiPorLM'*, *SthGldLM'*, *CcaGldLM''*, *CcaGldLM*, and *SweGldLM'* proteins using cryoEM. All datasets were collected using Titan Krios electron microscope (FEI) operated at 300 kV with either a K2 Summit detector and GIF energy filter or a K3 direct detection camera and BioQuantum energy filter (all Gatan). The K2 Summit detector was used in counting mode at 165,000x magnification with a pixel size of 0.822 Å and a total dose of 48 e⁻ Å⁻² over 32 fractions. Except where otherwise noted, datasets collected using the K3 direct detection camera were collected in counted super-resolution mode at 105,000x magnification with a pixel size of 0.832 Å and total doses ranging from 59.1 to 62.4 e⁻ Å⁻² over 40 fractions. For the datasets presented in this chapter, motion correction, dose weighting, contrast transfer function calculations, particle picking, and initial particle extraction were all performed automatically using SIMPLE 3.0 (Caesar *et al*, 2020). Gold-standard Fourier shell correlations (FSC) using the 0.143 criterion and local resolution estimations were calculated within RELION 3.1 (Zivanov *et al*, 2018). The subsequent data processing strategies for each dataset are described below.

6.4.1 *TroGldLM'*

I concentrated the *TroGldLM'* sample to A₂₈₀ = 0.5 and JCD used this concentrated sample to prepare cryoEM grids. He then used these grids to collect a dataset of 3,352 movies using the K2 Summit detector (representative micrograph in **Figure 6-9a**). I used SIMPLE 3.0 to automatically pick and extract 671,591 particles. After 2 rounds of reference-free 2D classification I selected 487,686 particles in classes with good secondary structure detail, although very few of these classes represented side views of the complex. I then subjected these particles to 3D classification with 3 classes for 15 iterations with 7.5 ° sampling then 10 iterations with 3.75 ° sampling, using the *FjoGldLM'* map as a reference with a 60 Å low pass filter. None of the resulting classes were very well-defined, but I selected one class with 243,862 particles where some helices were defined and used this as reference and mask for another round of 3D classification with 15 iterations of 7.5 ° sampling

then 2 iterations of 3.75 °. After 2 iterations of finer sampling there was still no good secondary structure detail visible, so I did not continue processing this dataset. As the purification yields for *TroGldLM'* were poor compared to other samples with more promising cryoEM datasets I did not characterise *TroGldLM'* any further.

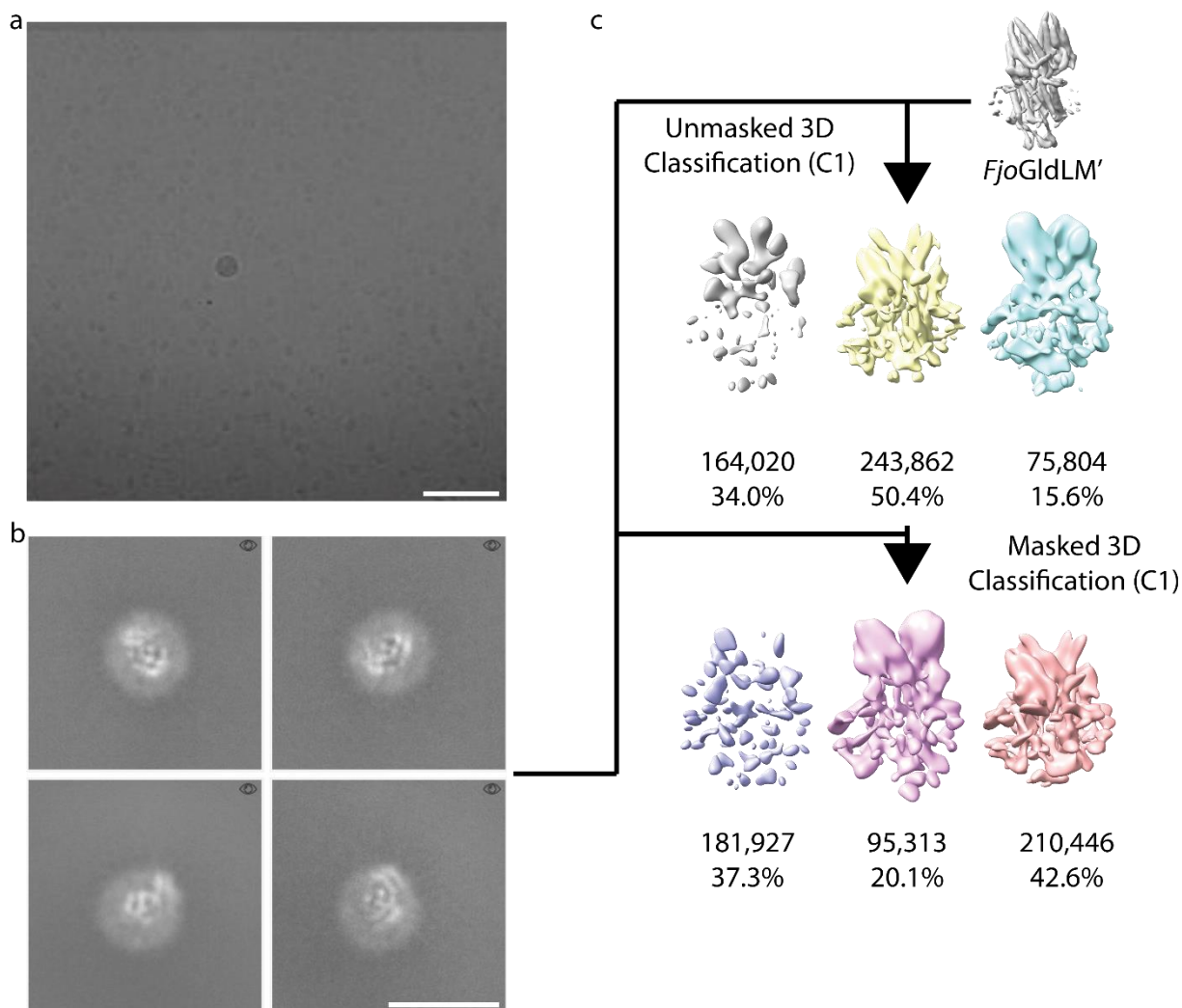


Figure 6-9 Data processing workflow for the *TroGldLM'* dataset. **a** Representative micrograph collected at approximately 2.7 μm defocus. Scale bar 500 Å. **b** Representative 2D class averages of *TroGldLM'* calculated in SIMPLE 3.0. Scale bar 100 Å. **c** Data processing workflow for the *TroGldLM'* dataset. After the masked 3D classification this dataset was not analysed further.

6.4.2 *PgiPorLM'*_C

I concentrated the *PgiPorLM'*_C sample to $A_{280} = 0.9$ and JCD used this concentrated sample to prepare cryoEM grids. He then used these grids to collect a dataset of 5,268 movies using the K2 Summit detector (representative micrograph in **Figure 6-10a**). I used SIMPLE 3.0 to automatically pick and extract 403,648 particles in 256 x 256-pixel boxes. I subjected these particles to two rounds of reference-free 2D classification in SIMPLE 3.0 and selected 328,271 particles from classes with visible secondary structure (**Figure 6-10b**). Folded D1 domains were visible in several of these classes, in contrast with the dataset previously collected for the *PgiPorLM'*_A construct (Chapter 4). I then low pass filtered the *FjoGldLM'* map to 60 Å and used it as a reference for 3D classification of the 328,271 *PgiPorLM'*_C particles in RELION 3.1 with 3 classes in C1 symmetry for 15 iterations with 7.5 ° sampling followed by 10 iterations with 3.75 ° sampling. I selected one class of 151,667 particles which resembled the *FjoGldLM'* map with little noise and used it for unmasked 3D autorefinement, yielding a 6.4 Å map in which distinct α -helices were visible. I used this 6.4 Å map as the reference (with 40 Å low pass filter) and mask for another round of 3D classification on the whole 328,271 particle dataset to improve the classification of particles. I selected one class of 174,135 particles which had defined helices and little noise. 3D autorefinement of this class produced a 5.0 Å map, but I could not improve this map further using further classifications or refinements of this dataset.

*PgiPorLM'*_C is quite a small target for cryoEM, which makes accurate alignments more difficult. I reasoned therefore that collecting a larger dataset might improve the resolution of the structure. I therefore thawed a sample of *PgiPorLM'*_C which had been concentrated to $A_{280} = 1.1$ then stored at -80 °C. Dr Emily Furlong (EF) used this thawed sample to prepare cryoEM grids which JCD imaged using the same Titan Krios microscope, but this time fitted with the K3 direct detection camera. With the K3 direct detection camera, the *PgiPorLM'*_C dataset was collected in counted super-resolution mode at 81,000x magnification with a pixel size of 1.068 Å and a total dose of 55.6 e⁻ Å⁻² over 40

fractions. From the 13,562 movies collected (representative micrograph **Figure 6-10c**) I used SIMPLE 3.0 to automatically pick and extract 8,205,503 particles in 256 x 256-pixel boxes. I subjected these particles to two rounds of reference-free 2D classification in SIMPLE 3.0, from which I selected 3,056,944 particles in classes with detailed secondary structure visible. I then subjected these particles to another round of 2D classification in RELION 3.1, from which I selected classes containing 2,513,045 particles (**Figure 6-10d**). I low pass filtered the 5.0 Å map from the initial dataset to 40 Å and used it as a mask and initial reference for 3D classification with 4 classes and 7.5 ° sampling. From this I selected one class of 633,283 particles with defined α -helices for 3D autorefinement, which yielded a 4.9 Å map. I used this map as reference and mask with a 15 Å low pass filter for another 3D autorefinement, yielding a 4.9 Å map. I used this map as the reference for Bayesian particle polishing, followed by 2D classification to remove any bad particles generated by polishing artefacts, from which I selected 649,359 particles. 3D autorefinement of these particles yielded a 4.0 Å map, which I then used as a reference for per-particle CTF refinement. 3D autorefinement of 649,359 per-particle CTF refined particles led to a 3.9 Å map, which I could not improve by further classification or refinement. The data processing workflow for the *PgiPorLM'*_C map is summarised with local resolution estimates and gold-standard Fourier shell correlations (FSC) in **Figure 6-10e-g**.

To build an atomic model of *PgiPorLM'* I used 1 to 1 threading with the Phyre² (Kelley *et al*, 2015) server to generate a model of *PgiPorLM'* based on the model of *FjoGldLM'*. I then docked this model into the *PgiPorLM'* EM density using rigid body fitting in *Coot* (Emsley *et al*, 2010). I then built additional residues at the N- and C-termini of PorL *de novo* using *Coot* and refined and validated the structure as described in 2.7.3. Data collection, model building, and refinement statistics are given in Table 6-3.

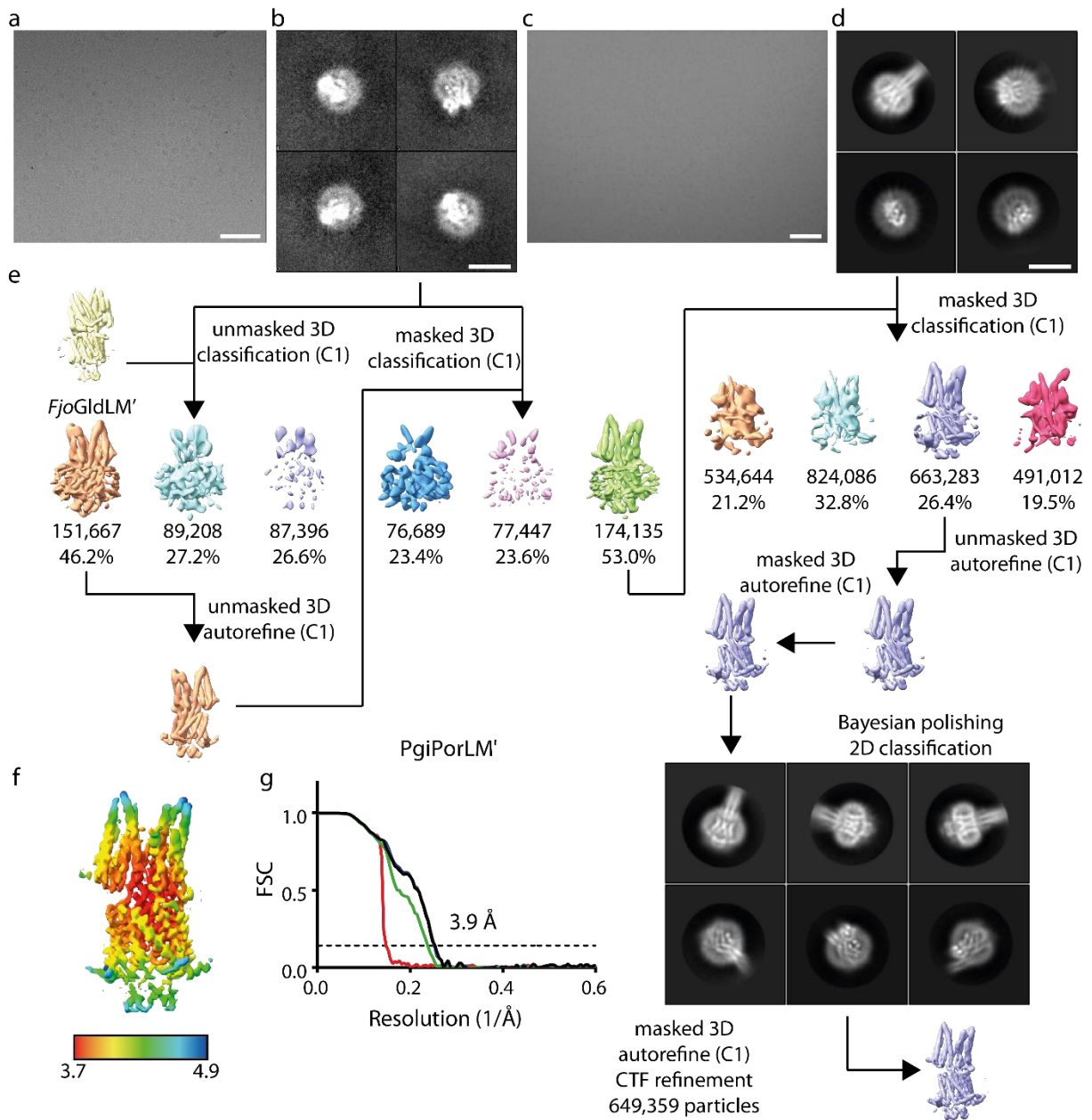


Figure 6-10 Data processing workflow for the *PgiPorLM'* map. **a** Example micrograph for the initial *PgiPorLM'* dataset collected with a K2 detector at approximately 1.8 μm defocus. Scale bar 500 $\text{\AA}$. **b** Representative 2D class averages from the K2 dataset used for initial processing. Scale bar 100 $\text{\AA}$. **c** Example micrograph from the second *PgiPorLM'* dataset collected with a K3 detector at approximately 1.8 μm defocus. Scale bar 500 $\text{\AA}$. **d** Representative 2D class averages from the K3 dataset used for final processing. Scale bar 100 $\text{\AA}$. **e** Data processing workflow for the *PgiPorLM'* map. Scale bar on 2D class averages 100 $\text{\AA}$. **f** Local resolution estimates (in $\text{\AA}$) for the sharpened *PgiPorLM'* map. **g** Fourier Shell Correlation (FSC) plot for the *PgiPorLM'* map. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomised; Green, unmasked; Blue, masked; black, MTF-corrected.

6.4.3 *SthGldLM'*

I concentrated the sample of *SthGldLM'* purified at 4 °C to a concentration of $A_{280} = 0.7$ and JCD used this concentrated sample to prepare cryoEM grids. He then used these grids to collect a dataset of 7,008 movies using the K2 Summit detector (**Figure 6-11a**). I used SIMPLE 3.0 to automatically pick and extract 1,466,521 particles, which I then subjected to two rounds of 2D classification in SIMPLE 3.0 (**Figure 6-11d**). I selected classes where detailed secondary structure was visible, containing 711,753 total particles. I low pass filtered the *FjoGldLM'* map to 60 Å and used this as a reference for 3d classification in RELION 3.1 with 4 classes for 15 iterations with 7.5 ° sampling then 10 iterations with 3.75 ° sampling (**Figure 6-11e**). From this classification I selected one class of 185,820 particles with defined α -helices for 3D autorefinement, which yielded a 6.0 Å map. I used this map as reference (with 40 Å low pass filter) and mask for another round of 3d classification using the 711,753-particle selection with four classes for 15 iterations with 7.5 ° sampling then 10 iterations of 3.75 ° sampling. I selected one class of 316,792 particles from this classification and used it for unmasked 3D autorefinement followed by masked 3D autorefinement, producing a 5.4 Å map. Despite the good detail visible in 2D class averages, I could not improve the resolution of this map by further classifications and refinements. A representation of the view distribution of particles for the 5.4 Å map (**Figure 6-11e** bottom left image) showed that almost all the particles were from “top-down” views (relative to the orientation shown in **Figure 6-11e**). I re-examined the 2D class averages and saw that there were very few classes showing side-views.

To improve the view distribution of particles, I purified a fresh sample of *SthGldLM'* at room temperature. I concentrated this sample to a concentration of $A_{280} = 1$ and removed an aliquot, then concentrated the remaining sample to $A_{280} = 3.2$. I added fOM to this $A_{280} = 3.2$ sample to a final fOM concentration of 0.7 mM. JCD then prepared cryoEM grids of the $A_{280} = 1$ sample without fOM and the $A_{280} = 3.2$ sample with fOM. He then imaged these grids using the K3 direct detection camera. In

total JCD collected 10,927 movies with fOM and 22,420 movies without fOM (**Figure 6-11b-c**). I used SIMPLE 3.0 to automatically pick and extract 4,539,707 particles in 256 x 256 pixel-boxes from the dataset with fOM, which I then subjected to two rounds of 2D classification in SIMPLE 3.0. I selected 817,030 particles in classes with good detail and subjected them to another round of reference-free 2D classification in RELION 3.1, from which I selected 386,285 particles. The 2D class averages for the dataset with fOM showed mostly side-views, but some showed string-like objects (**Figure 6-11f**). These objects were probably strings of LMNG formed during sample concentration, but they appeared to interact with the detergent micelles in some classes, so I retained them for further data processing. I used SIMPLE 3.0 to automatically pick and extract 9,203,748 particles in 256 x 256 pixel-boxes from the dataset without fOM, which I then subjected to two rounds of 2D classification in SIMPLE 3.0, from which I selected classes containing 3,135,006 particles. After a further round of reference-free 2d classification in RELION 3.1 I selected 1,422,930 particles, most of which were top-down views as in the previous dataset without fOM (**Figure 6-11d and f**). I combined the selected particles from the datasets with and without fOM to make a 1,809,215-particle dataset, which I used for 3d classification with 15 Å low pass filtered map from the previous dataset as an initial reference for 15 iterations with 7.5 ° sampling. The string-like object was not resolved in any of these classes, it appeared that particles containing strings were all classified into one class, which was stretched in one axis (**Figure 6-11g**, yellow class). I selected one well-defined class with 749,660 particles and used this for 3D autorefinement, generating a 3.7 Å map. I used this map as the reference for Bayesian particle polishing, which led to a 3.1 Å map. Per-particle CTF refinement did not improve the map, so I tried further 3D classifications without alignment in case there were slight conformational differences between particles that would reduce the resolution when averaged together. Classification without alignment using a mask around just the transmembrane helices followed by 3D autorefinement led to a final 3.0 Å map from 394,678 particles. The data processing workflow for the *SthGldLM'* map is

summarised with local resolution estimates and gold-standard Fourier shell correlations (FSC) in **Figure 6-11g-i**.

To build an atomic model of *SthGldLM'* I used 1 to 1 threading with the Phyre² (Kelley *et al*, 2015) server to generate a model of *SthGldLM'* based on the model of *FjoGldLM'*. I then docked this model into the *SthGldLM'* EM density using rigid body fitting in *Coot* (Emsley *et al*, 2010). I then built additional residues at the N and C-termini of GldL *de novo* using *Coot* and refined and validated the structure as described in 2.7.3. Data collection, model building, and refinement statistics are given in Table 6-3.

Figure 6-11 Data processing workflow for the *SthGldLM'* map. (Previous page) **a** Example micrograph for the initial *SthGldLM'* dataset collected with a K2 detector at a defocus of approximately 2.7 μm . **b** Example micrograph for the *SthGldLM'* + fOM dataset collected with a K3 detector at a defocus of approximately 1.7 μm . **c** Example micrograph for the *SthGldLM'* dataset collected with a K3 detector at defocus of approximately 1.7 μm . **d** Representative 2D class averages for the K2 dataset used in initial processing. **e** Data processing workflow for the initial dataset collected with a K2 detector. **f** Representative 2D class averages for the *SthGldLM'* + fOM (left) and *SthGldLM'* only (right) K3 datasets. **g** Data processing workflow for the combined datasets collected with a K3 detector. In the bottom-left panel each cylinder represents a view orientation, and the height of the cylinder corresponds to the number of views of that orientation. **h** Local resolution estimates (in $\text{\AA}$) for the sharpened *SthGldLM'* map. **i** FSC plot for the *SthGldLM'* map. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomised; Green, unmasked; Blue, masked; black, MTF-corrected. Scale bars for micrographs 500 $\text{\AA}$. Scale bars for 2D class averages 100 $\text{\AA}$.

6.4.4 *CcaGldLM''*

I concentrated the *CcaGldLM''* sample to $A_{280} = 1$ and JCD used this concentrated sample to prepare cryoEM grids. He then used these grids to collect 11,840 movies using the K3 direct detection camera (**Figure 6-12a**). I used SIMPLE 3.0 to automatically pick and extract 9,197,926 particles in 256 x 256-pixel boxes, which I subjected to one round of reference-free 2D classification with an 80 $\text{\AA}$ mask radius. I selected 4,607,771 particles from this classification with detailed secondary structure visible for the transmembrane and D1 regions. I re-extracted these particles in 412 x 412-pixel boxes and ran another round of reference-free 2D classification with a 180 $\text{\AA}$ mask radius to capture the D2 domain. Some of these classes had visible D2 domains, although the secondary structure detail for these regions was not as clear (**Figure 6-12b**) as for the TMH and D1 regions. Due to the large box size and mask radius, some classes showed two particles side-by-side. I selected 2,009,669 particles in classes where only one particle was visible and used these to generate an *ab initio* model in SIMPLE, imposing C2 symmetry to improve the density for the D2 domain, which I expected to be dimeric based on the *FjoGldM_{peri}* crystal structure (Leone *et al*, 2018). 3D classification should align just one particle from the double-particle classes, so I then made a larger selection of 4,282,288 particles including double-particle classes (**Figure 6-12c**). I re-scaled these 4,282,288 particles into 208 x 208-pixel boxes with a 1.648 $\text{\AA}$ pixel size to reduce the computational resources needed for downstream processing. I low pass filtered the *ab initio* model to 60 $\text{\AA}$ and used it as a reference for unmasked 3D classification of the 4,282,288 particles in RELION 3.1 with C1 symmetry with 5 classes and 7.5 $^{\circ}$

sampling. After 5 iterations all 5 classes had blobs of density at the edge of the spherical mask that were clearly not part of the protein (**Figure 6-12d** top image), so I selected one class and manually removed the extraneous density using ChimeraX to make a mask containing just the detergent micelle and D1 and D2 domains. I used this for another 3D classification of the whole dataset with 5 classes and 7.5 ° sampling for 15 iterations. I selected one class of 1,236,725 particles which showed good detail in the D1 domain but poor density in the D2 domain and used the map as a reference (with 15 Å low pass filter) for a round of unmasked 3d classification with this class for 15 iterations with 4 classes and 7.5 Å sampling, hoping that this would improve the D2 domain density. One of the resulting classes had good density for the D2 domain but poor density for the transmembrane helices, another had good density for the transmembrane helices but very poor density for the D2 domain. I then used a supervised 3D classification to sort all particles from the whole 4,282,288 particle dataset into each of these two classes. 1,793,574 particles were sorted into the class with good D2 domain density (D2 class) and 2,488,741 into the class with good transmembrane helix density (TMH class). At this stage, the orientation of the transmembrane helices did not match the other structures I had solved so I flipped the handedness of both maps. I now processed the D2 and TMH classes separately, with the aim of solving structures of both *CcaGldLM''* and *CcaGldLM'* from the same dataset.

To solve a structure of *CcaGldLM''* I used the D2 map as reference (with 15 Å low pass filter) for a 3D classification on the D2 dataset with four classes and 7.5 ° sampling for 15 iterations then 3.75 ° sampling for 10 iterations. I selected one class with 914,173 particles for 3D autorefinement, producing a 3.7 Å map. The transmembrane helices of this map were poorly defined and could not be improved by micelle-focused refinements. I instead ran a further 3D classification of these 914,173 particles with 4 classes and 7.5 ° sampling for 15 iterations, which yielded two classes totalling 595,559 particles with well-defined D2 domain density but poor TMH density and a class of 225,439

particles with improved TMH density but weakened D2 domain density. A periplasmic-domain focused refinement of the first class gave a 3.4 Å map of the D1 and D2 domains in a linear arrangement with the micelle only visible at low contour level. I attempted further processing of the 225,439-particle class in RELION 3.1 and cryoSPARC 2.15 (Punjani *et al*, 2017), but could not generate a map in which both the D2 and TMH regions were well defined and connected. I interpreted this as meaning that the linear arrangement of the D1 and D2 domains was incompatible with proper folding of the transmembrane helices, discussed further below. The data processing workflow for the *CcaGldLM* map is summarised with local resolution estimates and gold-standard Fourier shell correlations (FSC) in **Figure 6-12d-f**.

To build an atomic model of *CcaGldM*_{peri} I used 1 to 1 threading with the Phyre² (Kelley *et al*, 2015) server to generate a model of *CcaGldM*_{peri} based on the crystal structure of *FjoGldM*_{peri} (6EY4, (Leone *et al*, 2018)). I then docked this model into the *CcaGldLM* EM density using rigid body fitting in *Coot* (Emsley *et al*, 2010). I then remodelled sections using *Coot* and refined and validated the structure as described in 2.7.3. I did not extend the model into the poorly defined TMH region. Data collection, model building, and refinement statistics are given in Table 6-3.

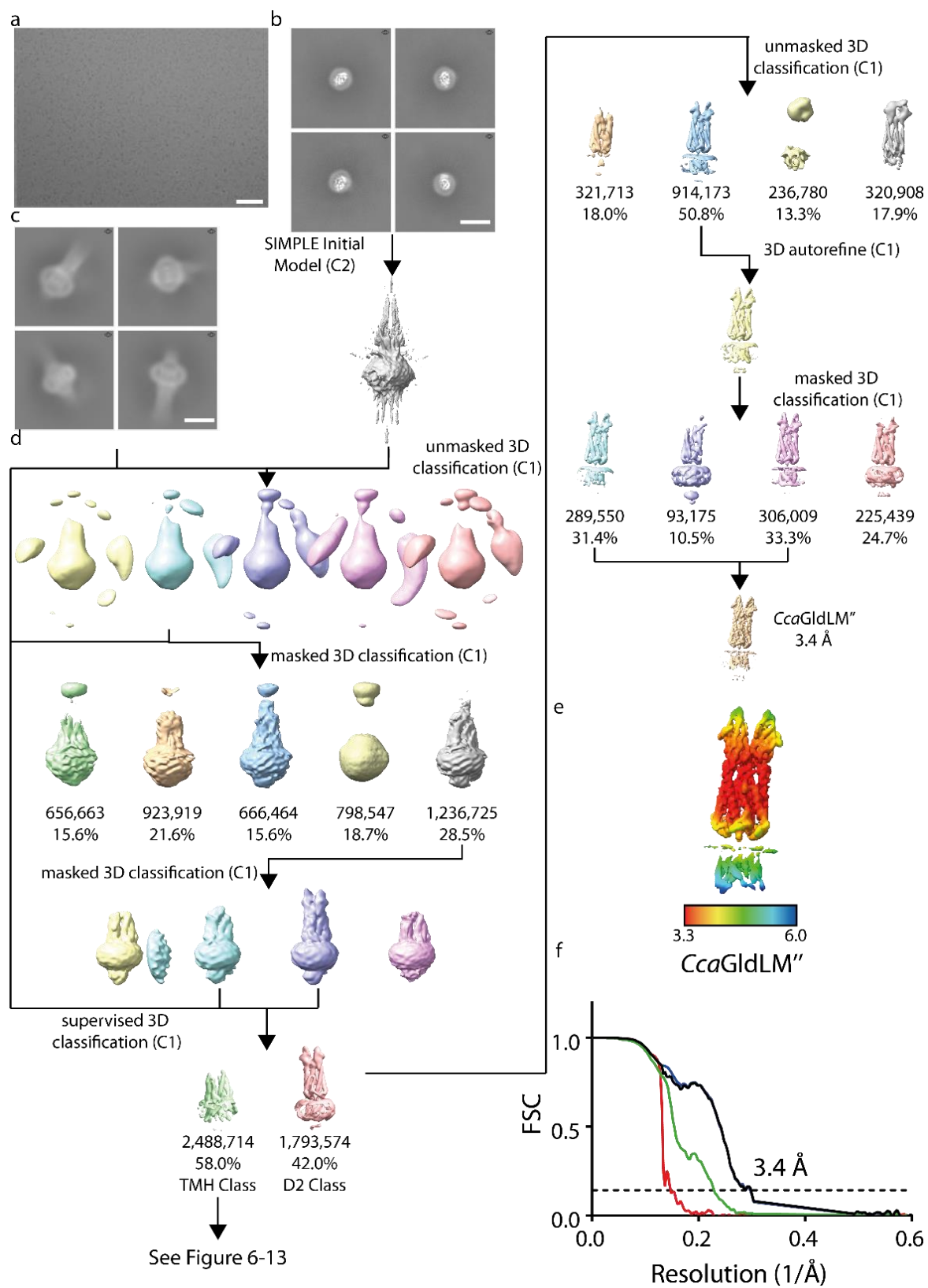


Figure 6-12 Data processing workflow for the *CcaGldLM*'' map. (Previous page) **a** Example micrograph for the *CcaGldLM*'' sample collected at approximately 2.6 μm defocus. Scale bar 500 Å. **b** Representative 2D class averages used to produce the initial model. Scale bar 100 Å. **c** Representative 2D class averages used to produce the *CcaGldLM*'' map. Scale bar 100 Å. **d** Data processing workflow for the *CcaGldLM*'' map. The handedness of the maps was flipped between the left and right columns once helices were visible. **e** Local resolution estimates (in Å) for the sharpened *CcaGldLM*'' map. **f** Fourier Shell Correlation (FSC) plot for the *CcaGldLM*'' map. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomised; Green, unmasked; Blue, masked; black, MTF-corrected.

To solve a structure of *CcaGldLM*' I processed the TMH class of 2,488,741 particles further in RELION 3.1, but all the resulting maps had distortions due to the high proportion of side views. Rather than collecting a new dataset with fOM to acquire new views of the complex as I had for *SthGldLM*', I instead took advantage of the large dataset JCD had collected for *CcaGldLM*'' and made a new selection from the 2D classification in 412 x 412-pixel boxes of 1,547,735 particles, selecting mostly classes with top-down views of the complex. I re-extracted these particles in 256 x 256-pixel boxes to exclude the D2 domain and ran a 3D classification using the *CcaGldLM*' class map with a 60 Å low pass filter as a reference with 4 classes for 15 iterations with 7.5 ° sampling. I selected one class with 531,709 particles for 3D autorefinement, which produced a 3.2 Å map. The helices of this map were a little distorted, appearing slightly flattened in one direction, so I used the *FjoGldLM*' model to generate a protein-only mask and ran a 3D classification without alignment. I selected one class of 131,883 particles and 3D autorefined it to produce a 3.1 Å map without distortions. I then applied Bayesian particle polishing and another 3D classification without alignment focused on the periplasmic domains, yielding a 3.0 Å map from 77,223 particles. The data processing workflow for the *CcaGldLM*' map is summarised with local resolution estimates and gold-standard Fourier shell correlations (FSC) in **Figure 6-13b-d**.

To build an atomic model of *CcaGldLM*' I used 1 to 1 threading with the Phyre² (Kelley *et al*, 2015) server to generate a model of *CcaGldM*' based on the model of *FjoGldLM*'. I then docked this model into the *CcaGldLM*' EM density using rigid body fitting in Coot (Emsley *et al*, 2010). I then remodelled

sections using *Coot* and refined and validated the structure as described in 2.7.3. Data collection, model building, and refinement statistics are given in Table 6-3.

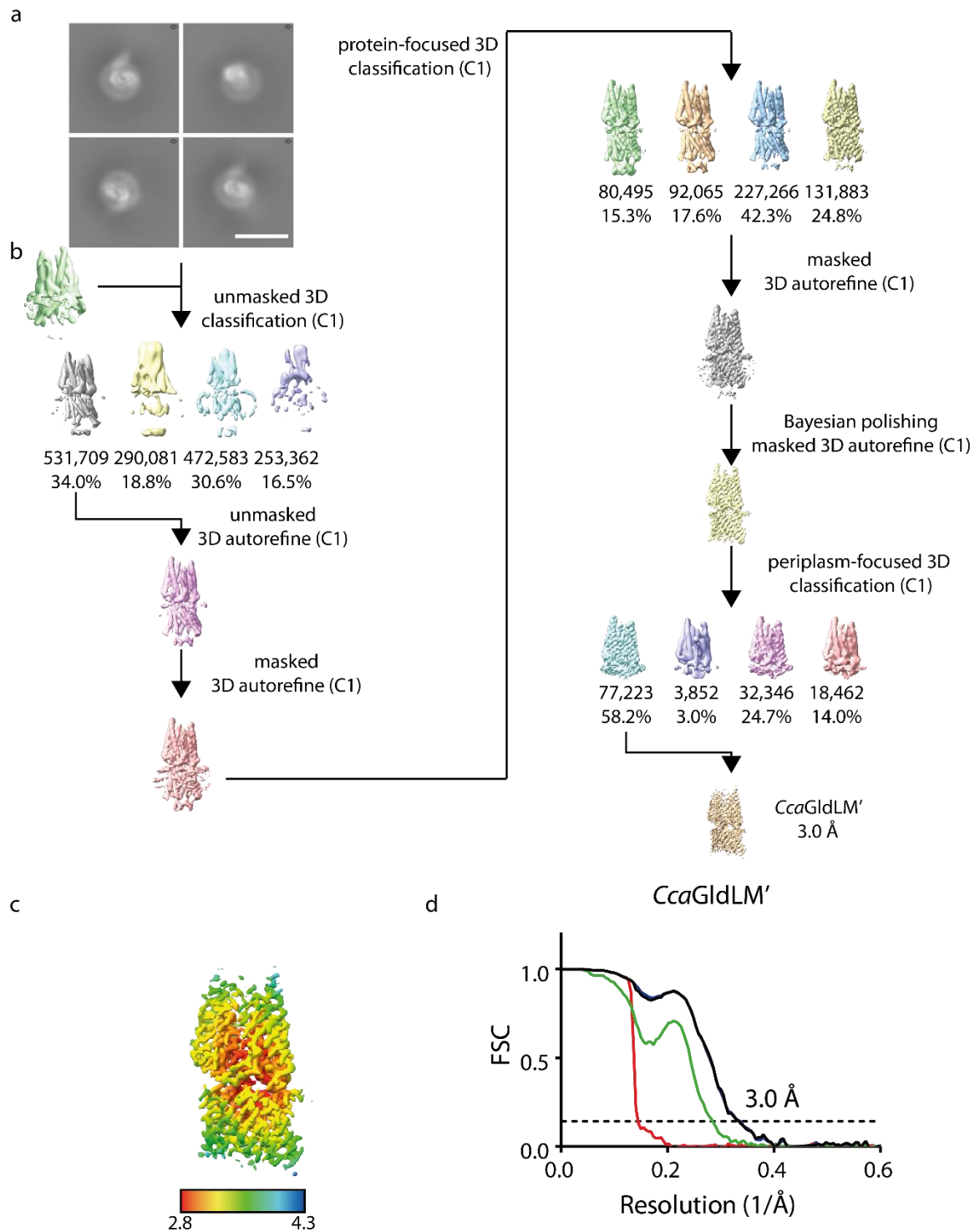


Figure 6-13 Data processing workflow for the *CcaGldLM'* map. **a** Representative 2D class averages used to produce the *CcaGldLM'* map. Scale bar 100 Å. **b** Data processing workflow for the *CcaGldLM'* map. **c** Local resolution estimates (in Å) for the sharpened *CcaGldLM'* map. **d** FSC plot for the *CcaGldLM'* map. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomised; Green, unmasked; Blue, masked; black, MTF-corrected.

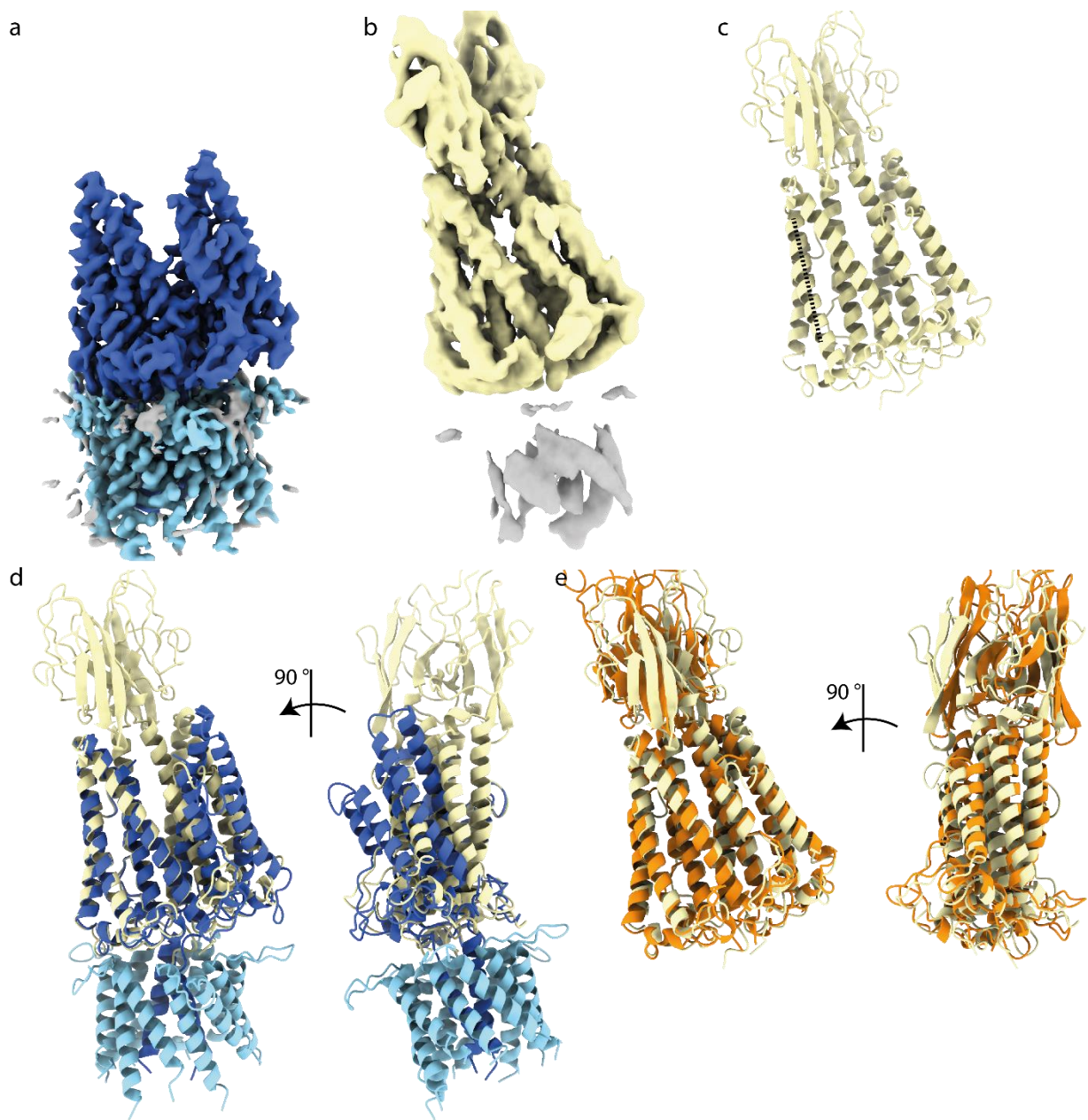


Figure 6-14 Comparison of the *CcaGldLM'* and the *CcaGldM''*_{peri} structures. **a** *CcaGldLM'* EM density map, coloured according to the protein model shown in **d**. GldL is coloured sky blue and GldM' is coloured royal blue. **b** *CcaGldLM''* EM density map, coloured according to the protein model shown in **c**. **c** Protein model of *CcaGldM''*_{peri}. It was not possible to build protein into the transmembrane density. All maps and models in this figure were aligned by to the second helix of *CcaGldM''*_{peri} Chain A, indicated with a dashed line. **d** Overlay of the *CcaGldLM'* and *CcaGldM''*_{peri} structures, showing the different relative orientations of the GldM D1 domains. **e** Overlay of *CcaGldM''*_{peri} and *FjoGldM*_{peri} (PDB 6ey4) structures, showing the agreement in GldM D1 domain orientations.

The cryoEM dataset collected for the *CcaGldLM''* construct seemed to contain particles that had either an ordered GldM'' D2 domain or an ordered TMH bundle but not both, according to 3D

classification (**Figure 6-12** and **Figure 6-14a-b**). Once I solved maps for these two classes separately and built in atomic models, I saw that the class where the D2 domain was ordered had parallel D1 domains (**Figure 6-14c**), similar to the *FjoGldM_{peri}* crystal structure (**Figure 6-14e**), whilst the class where the D2 domain was not visible had splayed D1 domains as in the GldLM' constructs (**Figure 6-14d**). This suggests that the parallel arrangement of D1 domains is incompatible with ordered transmembrane helices, further supporting the hypothesis that this arrangement is non-physiological as discussed in Chapter 4. I also noted in expression trials and large scale purifications (**Figure 6-3**, **Figure 6-4**, **Figure 6-7**, and **Figure 6-8**) that GldLM' constructs without the D2 domain were generally better expressed than GldLM'' constructs with the D2 domain. This may indicate that the parallel periplasmic arrangement disorders the transmembrane helices and prevents proper assembly of the GldLM'' complex in *E. coli*. It is worth bearing in mind that the GldLM''' and GldLM constructs were also generally less well-expressed than GldLM' constructs, so it may simply be the case that the periplasmic GldM dimer is not efficiently assembled in *E. coli*.

6.4.5 *CcaGldLM*

I was not able to solve a high-resolution structure of full length *PgiPorLM* due to the particles adopting a limited number of orientations in the ice and conformational heterogeneity between the domains of the PorM periplasmic region. To see if this was an issue specific to *PgiPorLM* or applied to full length GldLM constructs in general, I concentrated the *CcaGldLM* sample to $A_{280} = 1$ and used this concentrated sample to prepare cryoEM grids. JCD used these grids to collect a cryoEM dataset of 13,278 movies using the K3 direct detection camera (**Figure 6-16a**). I used SIMPLE 3.0 to automatically pick and extract 10,687,489 particles in 256 x 256-pixel boxes (equivalent to 213 x 213 Å), intending to generate a map with just the D1 domains first then re-extract in larger boxes and build out the other periplasmic domains later. I subjected these particles to two rounds of reference-free 2D classification in SIMPLE 3.0, after which I selected 4,608,042 particles in classes with secondary

structure detail (**Figure 6-15b**). I subjected these 4,605,042 particles to 3D classification in RELION 3.1 with 5 classes for 15 iterations at 7.5 ° sampling, using the *CcaGldLM*' map with a 60 Å low pass filter as a reference (**Figure 6-15c**). I selected two classes, containing 2,607,134 particles, where some detail was visible in the D1 domains and subjected them to a further round of masked 3D classification with 5 classes for 15 iterations at 7.5 Å sampling. I then selected one class, containing 529,492 particles, where some detail was visible in the D1 domains, and a blob of density was visible in the expected position for the D2 domain. I used this class for 3D autorefinement with no reference mask and a 140 Å spherical mask (**Figure 6-15c** bottom panel). This yielded a map of notional 3.8 Å resolution where there was some definition of the TMH and D1 regions but still no definition in the D2 domain. To provide more information for alignment of the D2-D4 domains I re-extracted the particles in 412 x 412-pixel boxes, then re-scaled them to 208 x 208-pixel boxes with 1.648 Å pixels to reduce the computational resources needed for processing. I used these re-extracted particles for 3D autorefinements with 240 Å and 300 Å spherical masks, but the D2 density did not improve. I next ran a 3D autorefinement using the *PgiPorLM* map as both a mask and a reference with a 60 Å low pass filter, but the D2 density was still not improved. To assess the view distribution of the re-extracted particles I subjected them to 2D classification in RELION 3.1 (**Figure 6-15d**). This showed that there were no side views of the complex in the particles I had selected after 3D classification. I re-examined the 2D class averages from classification in SIMPLE 3.0 and saw that although there were side views in that dataset, there were none with good detail in the periplasmic D1 domains. I therefore did not process the dataset any further.

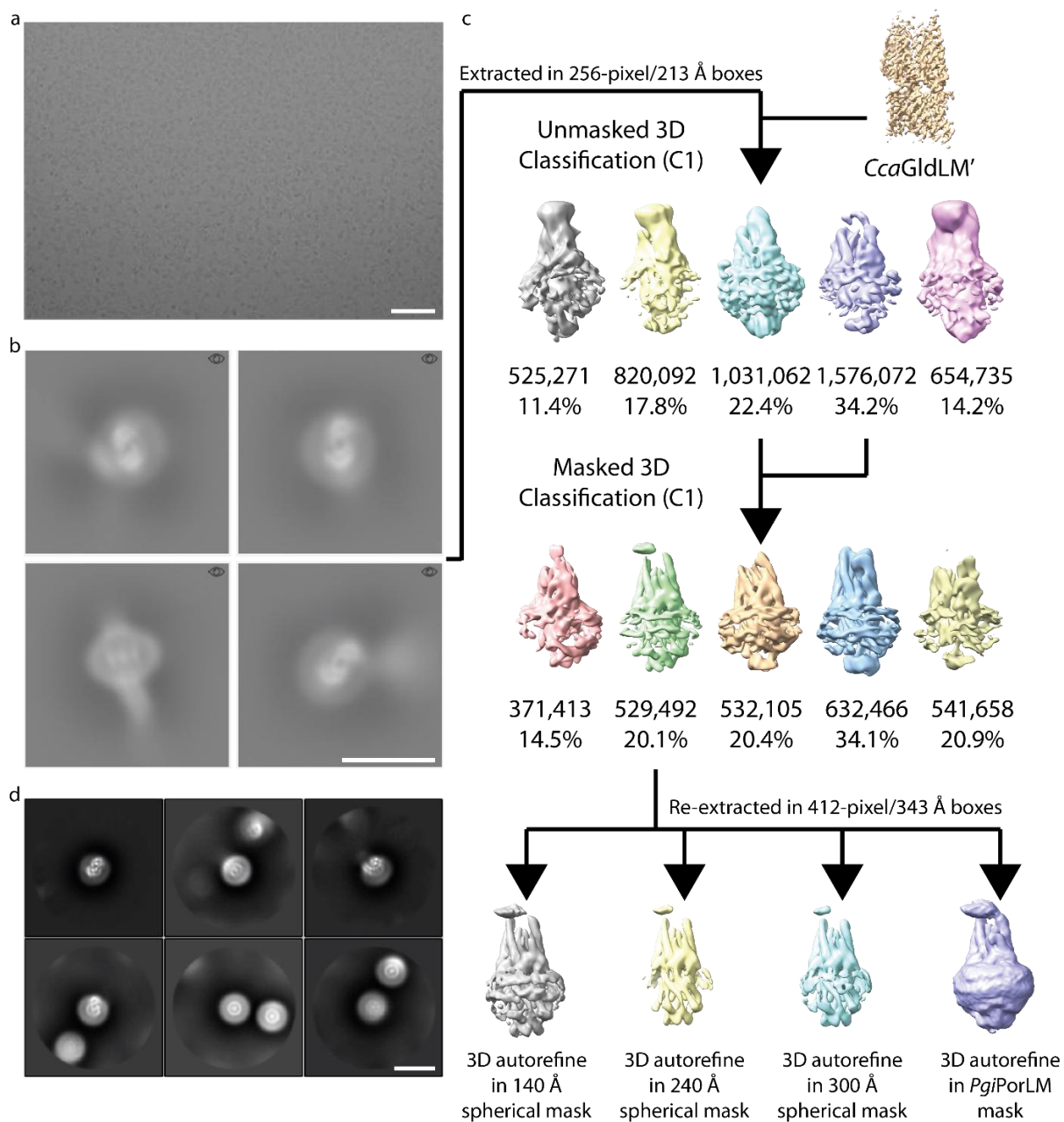


Figure 6-15 Data processing of the *CcaGldLM* dataset. **a** Representative micrograph collected at a defocus of approximately 2.3 μm . Scale bar 500 Å. **b** Representative 2D class averages of *CcaGldLM* calculated in SIMPLE 3.0 with a box size of 213 Å and a mask diameter of 183 Å. Scale bar 100 Å. **c** Data processing workflow for *CcaGldLM*. After the 3D autorefinements in the bottom panel no further processing was attempted. **d** 2D class averages of *CcaGldLM* calculated in RELION 3.1 with a box size of 343 Å and a mask diameter of 300 Å. Scale bar 100 Å.

6.4.6 *SweGldLM'*

I concentrated the *SweGldLM'* sample to a concentration of $A_{280} = 1$, then JCD used this concentrated sample to prepare cryoEM grids. He then used these grids to collect a cryoEM dataset of 12,495 movies using the K3 direct detection camera (**Figure 6-16a**). I used SIMPLE 3.0 to automatically pick and extract 7,167,266 particles in 256 x 256-pixel boxes. After two rounds of reference-free 2D classification in SIMPLE I selected classes with good secondary structure detail containing 3,873,460 particles (**Figure 6-16b**). I used the *SthGldLM'* map with a 60 Å low pass filter as a reference for 3D classification in RELION 3.1 with 5 classes for 15 iterations with 7.5 ° sampling (**Figure 6-16c**). I selected one class with 1,513,439 particles and used it for 3D autorefinement. The resulting 3.1 Å map had flattened helices caused by a lack of side views so I made a new selection of 1,360,637 particles from the 2D classification in SIMPLE, selecting classes with a high proportion of side-views (**Figure 6-16d**). I subjected these particles to 3D classification, using the distorted map with a 40 Å low pass filter as a reference and four classes with 7.5 ° sampling for 15 iterations (**Figure 6-16e**). I selected one class of 498,523 particles and used it for 3D autorefinement. This produced a 3.5 Å map where the periplasmic helices were still slightly flattened, so I ran a 3D classification without alignment focused on the periplasmic domains, from which I selected one class with 160,612 particles. 3D autorefinement of this class produced a 3.5 Å map without distortions in the periplasmic domains, which Bayesian particle polishing and a 3D classification without alignment focused on the transmembrane helices improved to a final 3.0 Å map from 111,727 particles. The data processing workflow for the *SweGldLM'* map is summarised with local resolution estimates and gold-standard Fourier shell correlations (FSC) in **Figure 6-16e-g**.

To build an atomic model of *SweGldLM'* I used 1 to 1 threading with the Phyre² (Kelley *et al*, 2015) server to generate a model of *SweGldLM'* based on the model of *FjoGldLM'*. I then docked this model into the *SweGldLM'* EM density using rigid body fitting in Coot (Emsley *et al*, 2010). I then remodelled

sections using *Coot* and refined and validated the structure as described in 2.7.3. Data collection, model building, and refinement statistics are given in Table 6-3.

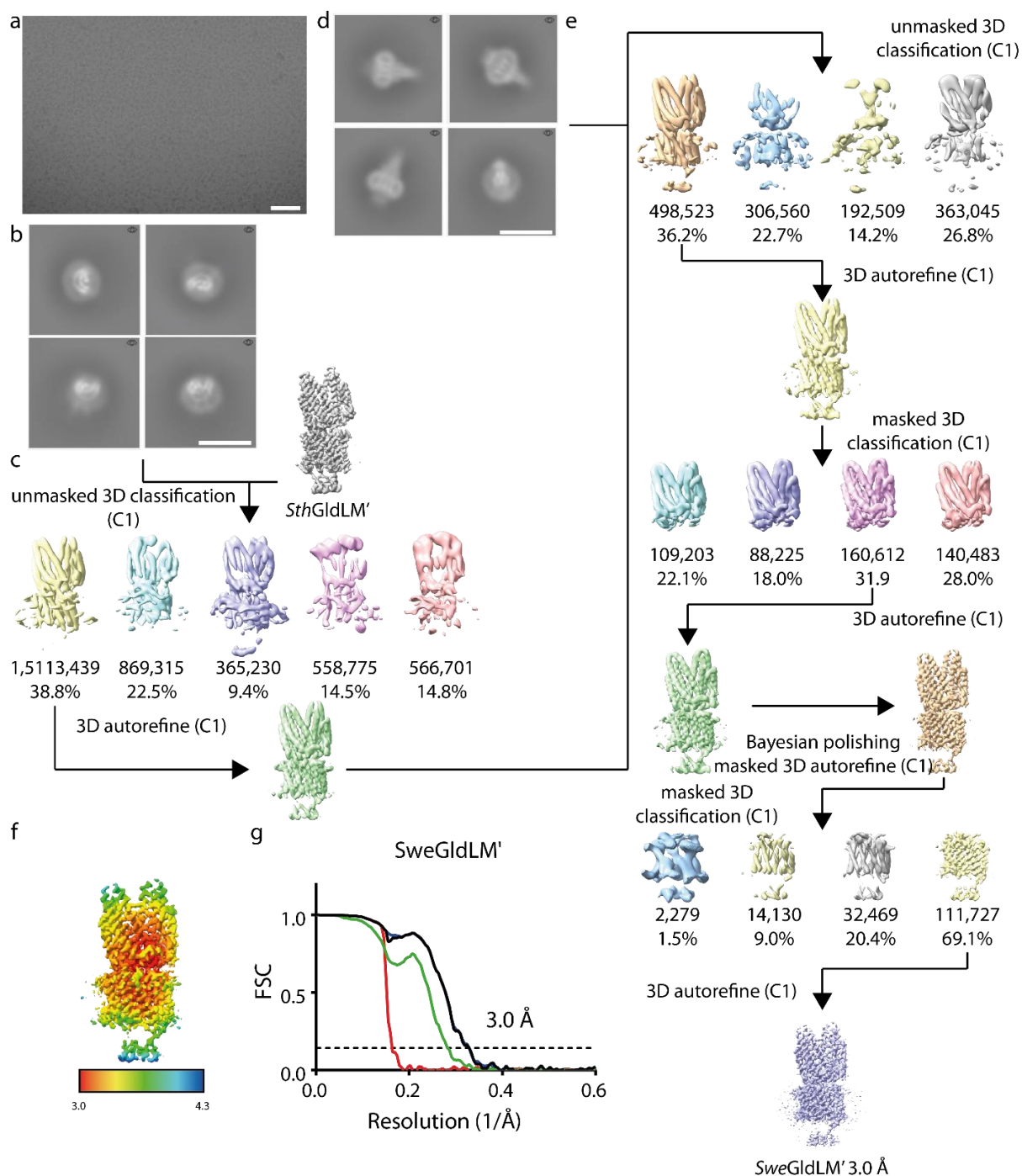


Figure 6-16 Data processing workflow for SweGldLM' map. **a** Example micrograph for SweGldLM' dataset collected with K3 detector at defocus of approximately 1.5 μm . Scale bar 500 $\text{\AA}$. **b** Representative 2D class averages for initial particle selection. Scale bar 100 $\text{\AA}$. **c** Data processing workflow for initial particle selection. **d** Representative 2D class averages for second, side-view focused particle selection. Scale bar 100 $\text{\AA}$. **e** Data processing workflow for second particle selection. **f** Local resolution estimates (in $\text{\AA}$) for the sharpened SweGldLM' map. **g** Fourier Shell Correlation (FSC) plot for the SweGldLM' map. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomised; Green, unmasked; Blue, masked; black, MTF-corrected.

Table 6-3 Cryo-EM data collection, refinement, and validation statistics for the *PgiPorLM'*, *SthGldLM'*, *CcaGldM''*, *CcaGldLM'*, and *SweGldLM'* structures.

	<i>P. gingivalis</i> PorLM'	<i>S. thermophila</i> GldLM'	<i>C. canimorsus</i> GldM''	<i>C. canimorsus</i> GldLM'	<i>S. wenxiniae</i> GldLM'
Data collection and processing					
Magnification	81,000	105,000	105,000	105,000	105,000
Voltage (kV)	300	300	300	300	300
Electron exposure (e ⁻ Å ⁻²)	55.6	62.4 (without fOM) 61.2 (with fOM)	59.1	59.1	56.9
Defocus range (μm)	1.0-3.0	1.0-3.0	1.0-3.0	1.0-3.0	1.0-3.0
Pixel size (Å)	0.832	0.832	0.832	0.832	0.832
Symmetry imposed	C1	C1	C1	C1	C1
Initial particle images (no.)	8,205,503	13,743,455	9,197,926	9,197,926	7,167,266
Final particle images (no.)	649,359	394,678	595,559	77,223	111,727
Map resolution (Å)	3.9	3.0	3.4	3.0	3.0
FSC threshold	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	3.7-4.9	2.8-3.7	3.3-6.0	2.8-4.3	3.0-4.3
Refinement					
Initial model used (PDB code)	None	None	None	None	None
Model resolution (Å)	3.9	3.0	3.4	3.0	3.0
FSC threshold	0.143	0.143	0.143	0.143	0.143
Model resolution range (Å)	3.7-4.9	2.8-3.7	3.3-6.0	2.8-4.3	3.0-4.3
Map sharpening B factor (Å ²)	-200	-93	-122	-58	-83
Model composition					
Non-hydrogen atoms	6581	6317	4728	5708	5301
Protein residues	822	791	600	722	689
Ligands	0	0	0	0	0
B factors (Å ²)					
Protein	72.16	51.61	93.02	55.39	63.20
Ligand	NA	-	-	-	-
R.m.s. deviations					
Bond lengths (Å)	0.004	0.005	0.005	0.005	0.009
Bond angles (°)	0.747	0.600	0.781	0.675	0.700

Validation					
MolProbity score	2.16	1.68	2.39	1.82	1.50
Clashscore	15.42	5.98	22.84	9.65	5.09
Poor rotamers (%)	0.28	0.15	0.00	0.00	0.00
Ramachandran plot					
Favoured (%)	92.57	94.85	90.34	95.48	96.44
Allowed (%)	7.43	5.15	9.66	4.38	3.56
Disallowed (%)	0.00	0.00	0.00	0.14	0.00

6.5 Comparison of GldLM' structures from five different species

Once I had solved the structures of GldLM' complexes from *C. canimorsus*, *P. gingivalis*, *S. thermophila*, and *S. wenxiniae* I compared them to the structure of GldLM' from *F. johnsoniae*. The overall architecture of the complexes is conserved, with five copies of GldL surrounding two copies of GldM (Figure 6-17a-c). Only the most membrane-proximal part of the GldL C-terminal domain was resolved in any of the structures with the major part of the domain too mobile to be resolved. The periplasmic D1 domain of GldM was visible in all structures. The precise angles between the two domains varied a little from 30 to 45 ° in, but all were splayed similarly to the previous *FjoGldLM'* structure and in contrast to the crystal structures of the whole periplasmic domain previously published for *FjoGldM_{peri}* and *PgiPorM_{peri}* (Sato *et al*, 2020; Leone *et al*, 2018).

The *CcaGldLM'*, *SthGldLM'*, and *SweGldLM'* structures were solved at 3.0 Å resolution. This was a higher resolution than the original *FjoGldLM'* structure, which was solved at 3.9 Å resolution, as was the *PgiPorLM'* structure. This improved resolution makes it possible to model the positions of side chains with more confidence than in the *FjoGldLM'* structure. The residues that I previously identified as important for function of the GldLM motor are well-conserved in the motors from other species and are found at similar positions in the membrane (Figure 6-17e). The exception is *FjoGldL* Tyr13 which is substituted with isoleucine (Ile15) in *S. wenxiniae*. The orientations of the functionally important polar side chains were mostly similar across species, although even with the improved resolution of the new structures it is not possible to be certain of the precise orientations of all side chains. The most variation was seen for GldL Glu49, which in the *FjoGldLM'* structure appeared to form a salt bridge with GldM Arg9. This variation is not surprising as glutamate sidechains are often damaged by the electron beam during cryo-EM experiments resulting in decarboxylation, meaning that they are not visible in EM density maps and cannot be accurately modelled (Hattne *et al*, 2018). In all structures, the arginine corresponding to *F. johnsoniae* GldM Arg9 adopts a similar position to

FjoGldM Arg9, suggesting that the salt bridge interaction with the GldL glutamate is conserved across the different motor complexes studied here.

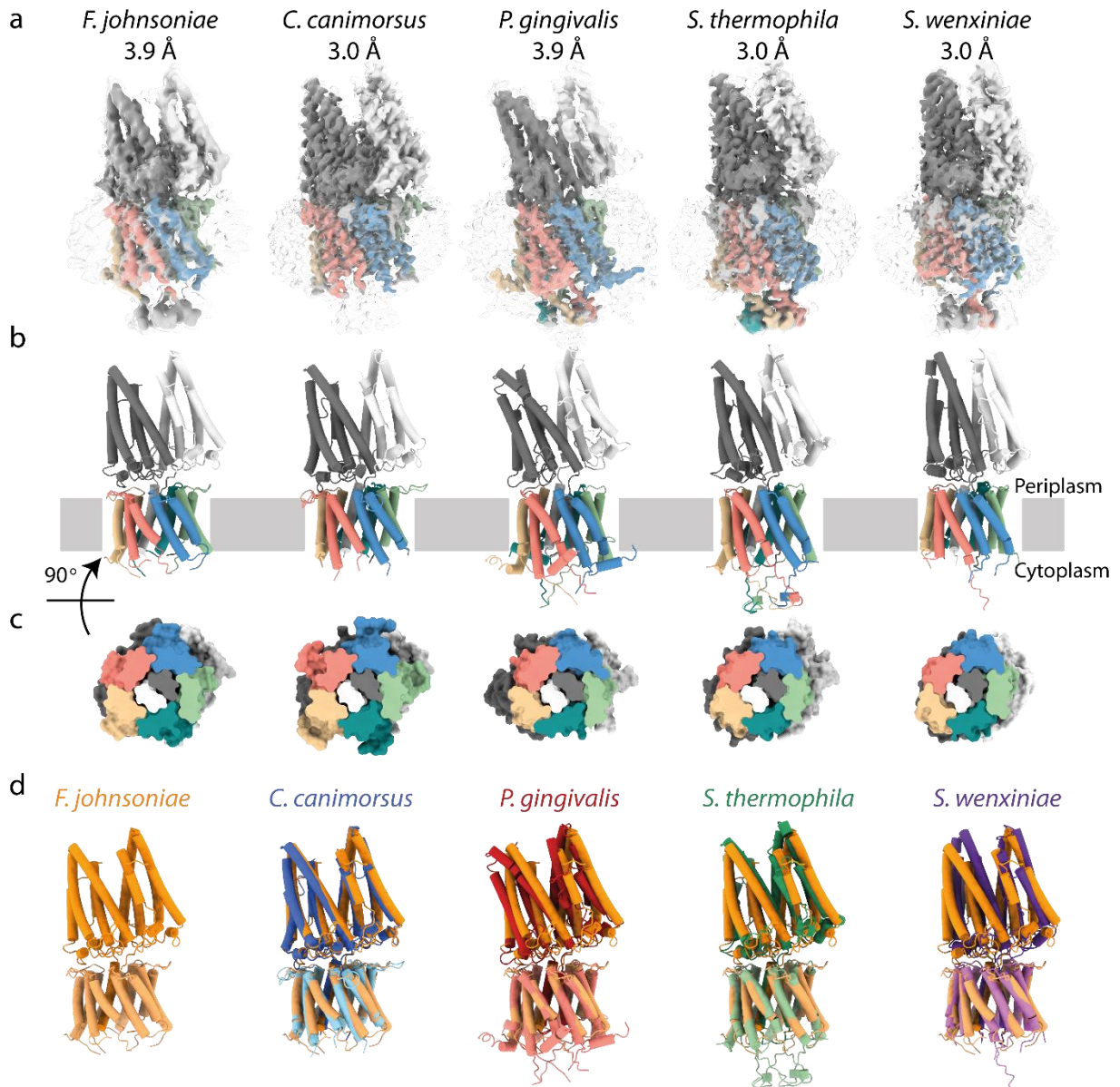


Figure 6-17 All five GldLM' structures have the same architecture. **a** EM density maps for each structure. A transparent second copy of the map is shown at low contour level for each structure to show the location of the detergent micelle. **b** Cartoon representations of each structure. **c** View from the cytoplasm of each structure in surface representation. Each structure is clipped approximately halfway through the TMH bundle. **d** Superimpositions of the five different structures in cartoon representation. For each structure GldM'/PorM' is coloured a darker shade than GldL/PorL. The indicated structures are overlaid on *FjoGldLM'* aligning on the TMH of GldM' Chain B.

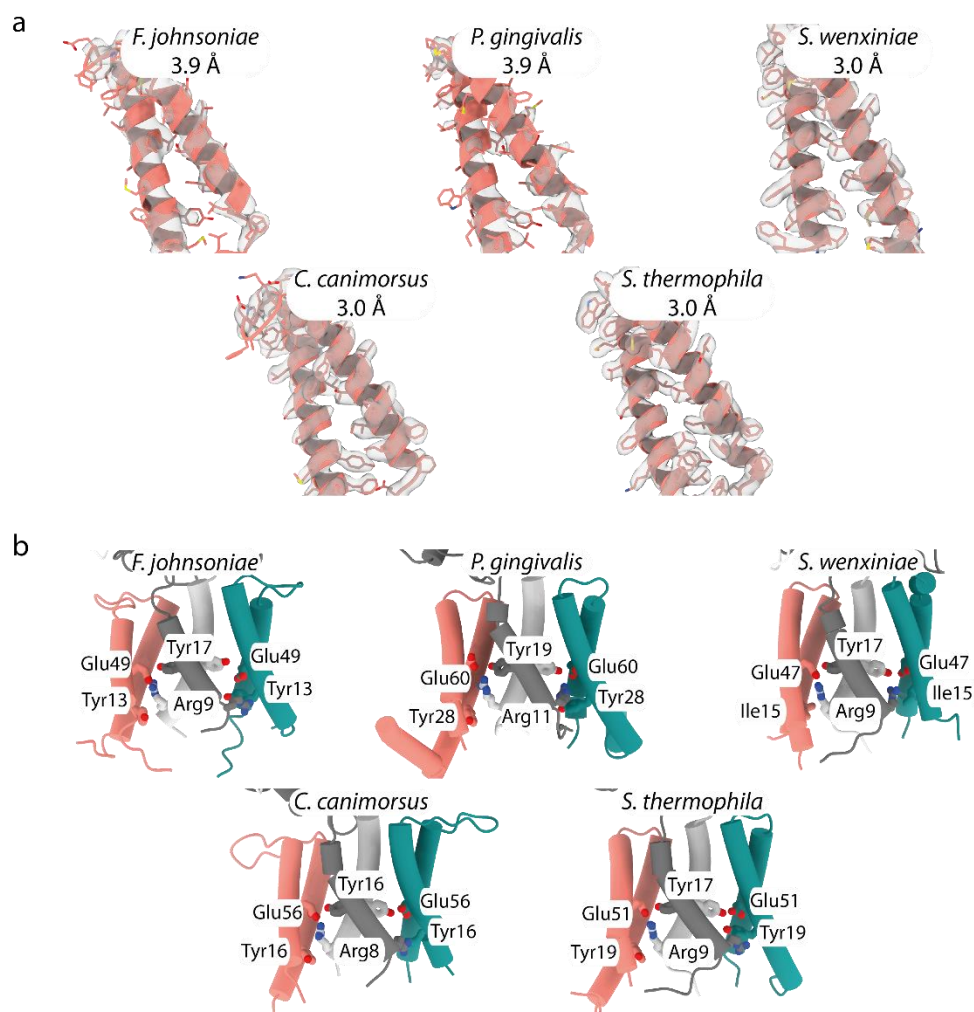


Figure 6-18 The *CcaGldLM'*, *SthGldLM'*, and *SweGldLM'* structures are at higher resolution than the *FjoGldLM'* structure. **a** EM density in transparent grey overlaid on GldL Chain C for each structure. **b** Views of each structure to show the residues equivalent to those identified as important for motor function in *FjoGldLM*. For clarity GldL Chains D, E, and G are hidden for each structure.

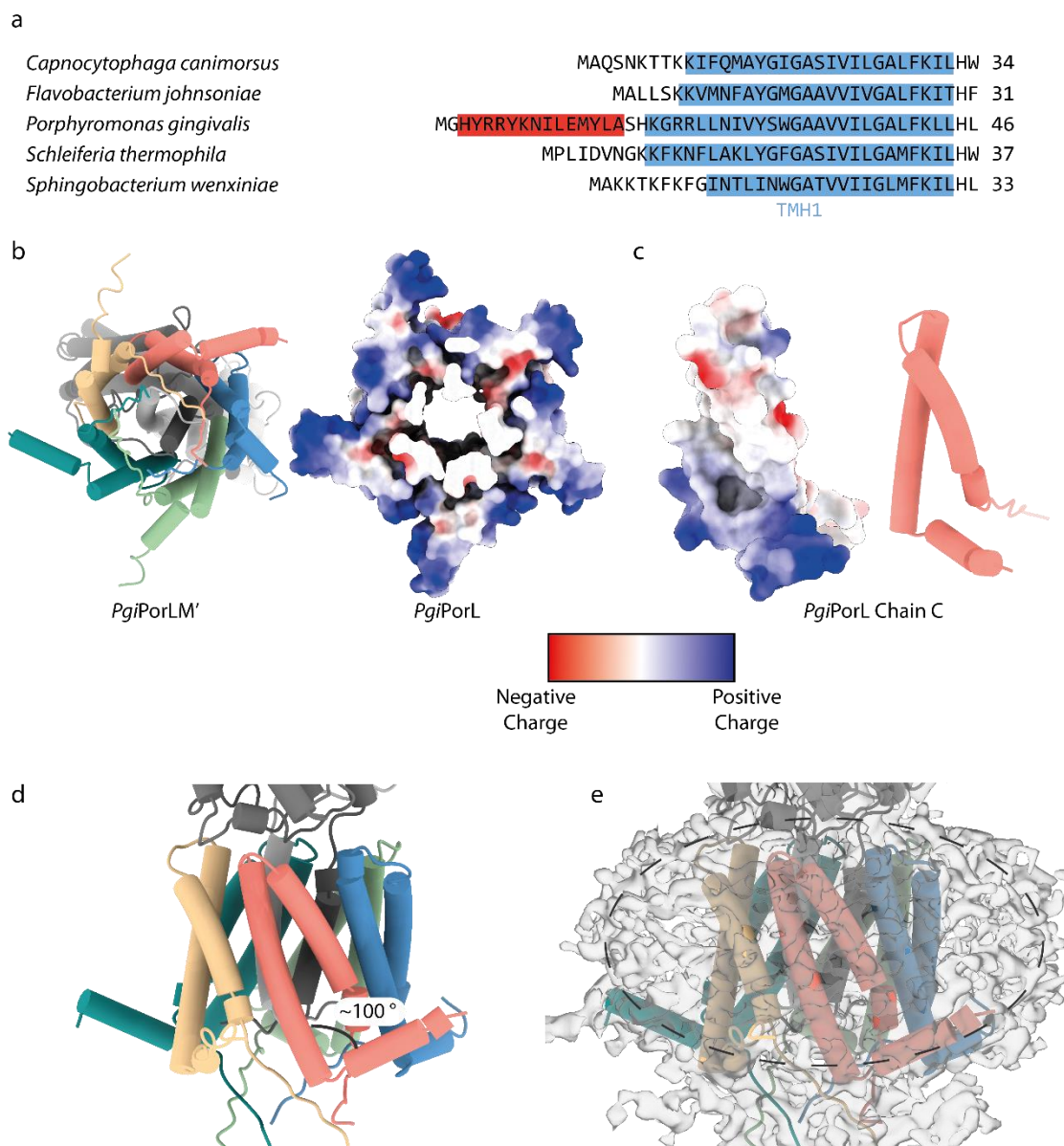


Figure 6-19 *P. gingivalis* PorL has an N-terminal helix that is not present in GldL from the other structurally characterised GldLM/PorLM motor complexes. **a** Sequence alignment of the N-terminal regions of GldL homologues. The first transmembrane helix of each sequence is highlighted in blue. The additional N-terminal helix of *PgiPorL* is highlighted in red. **b** View of *P. gingivalis* PorLM' from the cytoplasm in cartoon (left) and coulombic surface (right) representations. For clarity PorM' is hidden in the coulombic view and the cytoplasmic extensions of GldL sliced through. **c** Cartoon (left) and coulombic (right) surface representations of *P. gingivalis* PorL Chain C, from the membrane with the periplasm at the top. **d** Side view of *PgiPorLM'* in cartoon representation. The angle between the N-terminal helix and TMH1 was measured using the AngleBetweenHelices function in PyMol. **e** Side-view of the *PgiPorLM'* complex overlaid with the EM density map at low contour level. The approximate boundary of the detergent micelle is marked with a dashed line.

The *PgiPorL* sequence has a longer sequence before the first transmembrane helix than GldL from the other species I studied (**Figure 6-19a**). The *PgiPorLM'* structure revealed that these extra residues form a helix that is not present in the other structures (**Figure 6-19b**). The helix has a positively charged N-terminus and points tangentially away from the circumference of the transmembrane helix bundle (**Figure 6-19b-d**). The helix appeared to lie against the curved detergent micelle (**Figure 6-19e**). *In vivo*, it probably lies parallel to the membrane with the positive end interacting with the negatively-charged phospholipid head groups. This additional N-terminal helix may have a role related to *P. gingivalis* being non-motile, but as it is not known for certain whether *S. wenxiniae* and *S. thermophila* are motile or non-motile it is not possible to say if this helix has role in non-gliding T9SS systems.

My structures of GldLM' all lack EM density for most of the cytoplasmic domain of GldL. However, the region immediately adjacent to the inner face of the cytoplasmic membrane was resolved in the *PgiPorLM'*, *SthGldLM'*, and *SweGldLM'* structures. Extended coils from the C-terminal end of TMH2 formed a cage-like structure below the detergent micelle (**Figure 6-20a**). This region may be stabilised in *S. thermophila* and *P. gingivalis* because these species have higher optimum growth temperatures than *F. johnsoniae* (Albuquerque *et al*, 2011; Coykendall *et al*, 1980) so are expected to have evolved stronger inter-subunit interactions. In *SweGldLM'* the full coil was only fully resolved for GldL chain C, which is involved in the putative salt bridge with GldM chain B. *S. wenxiniae* does not have a higher growth temperature than *F. johnsoniae* (Zhang *et al*, 2012; Stanier, 1947), but in the *SweGldLM'* structure the N-terminus of GldL chain D appeared to brace the extended coil between the second transmembrane helix of GldL chain C and the cage-like structure. This interaction between the N-terminus and cytoplasmic domain of GldL is not seen in the other structures and may indicate a role for the N-terminus of GldL in species-specific functional tuning. Conservation analysis showed that the cage-like structure is quite well-conserved and has several negative charges (**Figure 6-20b-d**).

These negative charges might help keep the cage-like structure separate from the negatively charged head groups of the phospholipid bilayer. In *SthGldLM'*, for which this region was best resolved, the coils that make up the cage appear to be held together by a network of hydrogen bonds and hydrophobic packing between aspartate, tryptophan, and tyrosine residues in a five-fold pseudosymmetric arrangement (**Figure 6-20d**).

I investigated the role of the cage-like region in the genetically tractable *F. johnsoniae* system. Cells with the cage region of GldL deleted (residues 64-74) had a very slight gliding defect measured by colony spreading on PY2 agar, whilst cells with this region replaced with a GSS repeat linker showed no apparent gliding defect. This suggests that despite the sequence conservation in this region of the GldL protein it plays a minor role in motor function. Given that deleting the region caused a small defect in gliding motility, it may be the case that the main function of the region is to position the cytoplasmic domain of GldL at the appropriate distance from the membrane. Deleting the cage-like region would disrupt this positioning but not necessarily prevent any important protein-protein interactions that do not directly involve the deleted residues. It may also be the case that this region is important for function in some species but not in *F. johnsoniae*.

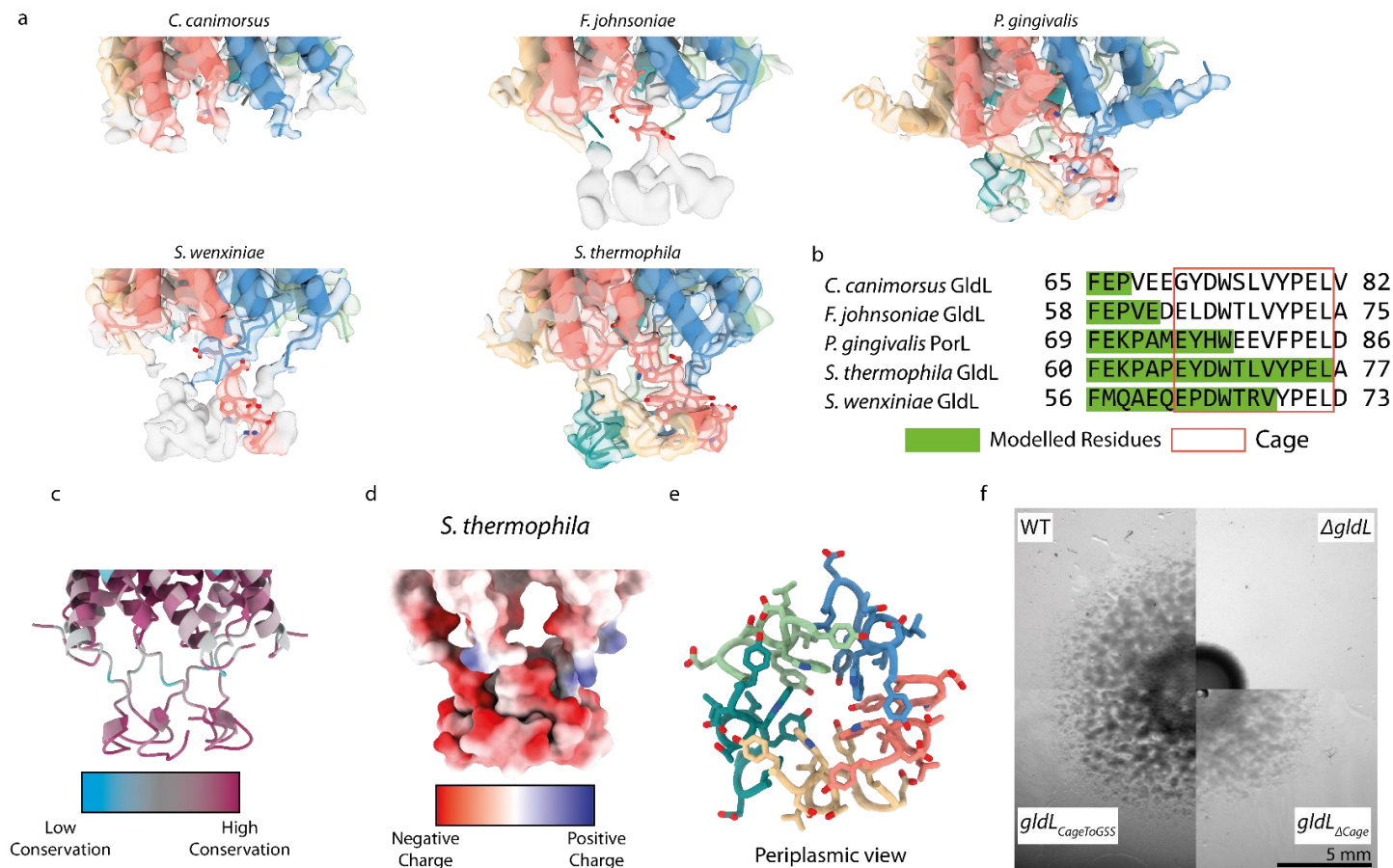


Figure 6-20 The cytoplasmic region of GldL proximal to the membrane forms a cage. **a** Overlay of EM density and the built model for the cytoplasmic region of each GldL structure. EM density is shown at the same contour level for all species. Side chains are shown for Chain C in the cage region. **b** Sequence alignment of the cage region for the five GldLM' structures. Residues which could be modelled for each structure are highlighted in green. Residues constituting the cage region are boxed in red. **c** Sequence conservation analysis of the cage region of *S. thermophila* GldL using the program Consurf (Landau *et al*, 2005; Ashkenazy *et al*, 2016). **d** Coulombic potential representation for the cage region of *S. thermophila* GldL. For clarity, the first TMH and N-terminal residues are hidden. **e** Interaction network at the base of the cage-like structure in *SthGldLM'*. Show from the periplasmic (top panel) and cytoplasmic (bottom panel) directions. **f** Effects of modifications in the GldL cytoplasmic cage on *F. johnsoniae* gliding motility (spreading) on plates. The results are representative of three independent experiments.

In the 3.0 Å resolution GldLM' structures (*Cca*GldLM', *Sth*GldLM', and *Swe*GldLM') there were some regions of density that resembled detergent rather than protein (**Figure 6-21a-d**). They were less well-ordered than the protein regions but several of them did resemble LMNG, the detergent I used to solubilise these proteins, with bulky hydrophilic sugar groups at the outer edge of the detergent micelle and long hydrophobic alkyl tails on the inside of the micelle (**Figure 6-21e**). All these regions of density were next to the periplasmic end of the TMH bundle, which may indicate that there are conserved interactions with lipids at these sites *in vivo*. However, interactions with LMNG do not necessarily reflect equivalent physiological interactions with lipids. It would be interesting therefore, to study the structure of GldLM' in a lipid environment. I did attempt to reconstitute *Swe*GldLM' into lipid nanodiscs, but without success. *Bacteroidetes* bacteria have different lipid compositions from the purified *E. coli* lipids I used so it may be necessary to reconstitute *Swe*GldLM' into lipids extracted from *F. johnsoniae* or *S. wexleyi*.

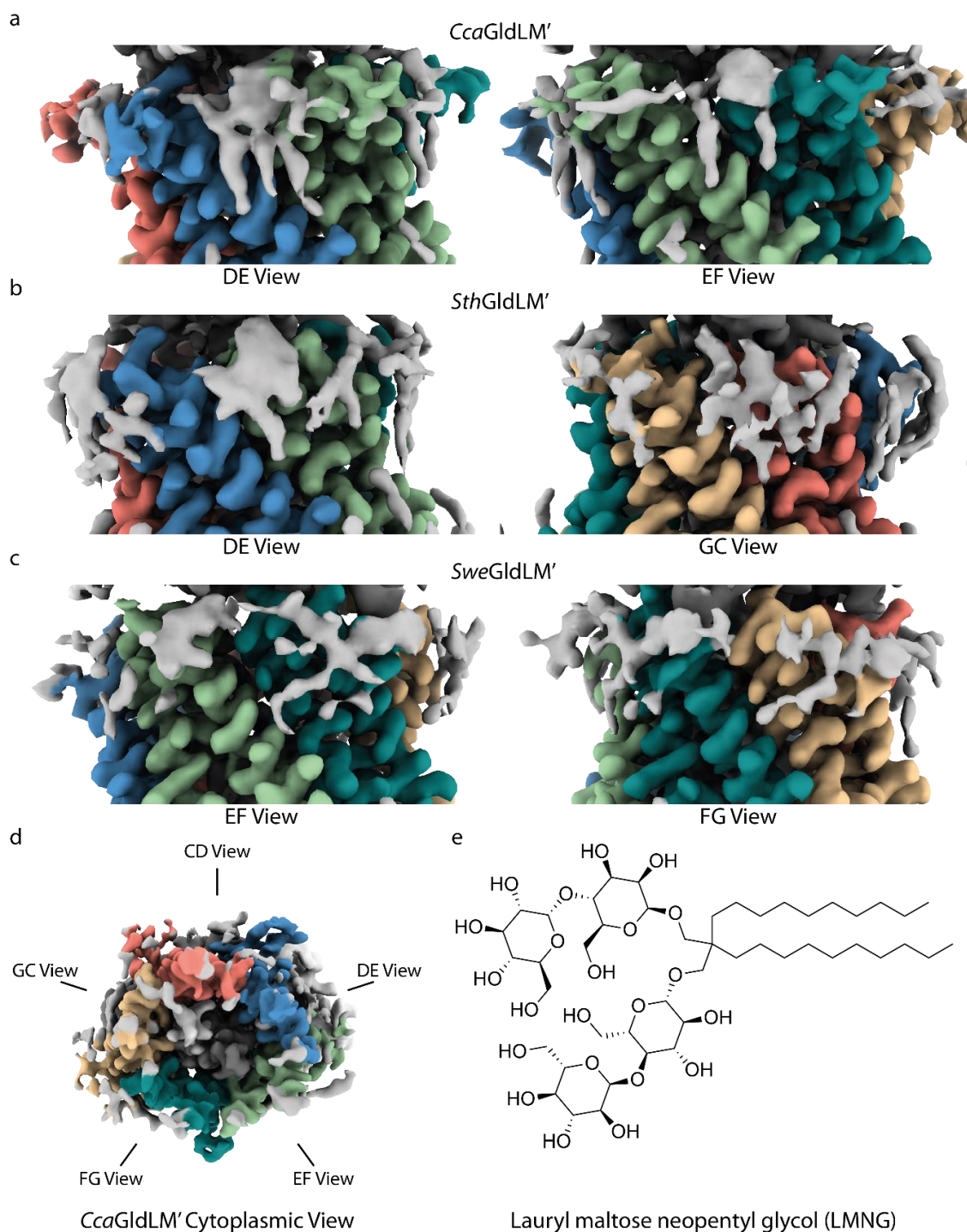


Figure 6-21 The *CcaGldLM'*, *SthGldLM'*, and *SweGldLM'* maps have peripheral EM density that may correspond to detergent molecules. **a-c** Within-membrane views of the top half of the TMH bundle for the indicated maps. Protein density is coloured by chain. The remaining density, some of which may correspond to detergent molecules, is coloured grey. For each structure, the pair of images show the map at the same contour level. Each view is labelled according to the pair of GldL chains in the centre of the view, as indicated in **d**. **d** The viewpoints used in **c-e** illustrated on a cytoplasmic view of the *CcaGldLM'* map. **e** Chemical structure of LMNG.

6.6 Purification and structural characterisation of GldLM complexes with substitutions of functionally important residues

The new structures of GldLM' homologues all showed the same conformation as the original *Fjo*GldLM' structure, so did not provide additional insight into the mechanism of proton transport and GldM rotation. I reasoned that solving structures of variant proteins with substitutions of functionally important amino acids might trap different conformations of the structure and provide hints as to the conformational changes occurring during motor activity. I therefore generated and attempted to purify variant constructs with substitutions of the residues I had previously identified as important for motor function.

6.6.1 Purification of variant *Swe*GldLM' constructs

After my expression screen and initial purifications of GldLM' constructs from new species, it was clear that *Swe*GldLM' was very well-expressed and I could purify large amounts of this complex. I reasoned that even if introducing substitutions reduced the purification yields for complexes, variant *Swe*GldLM' complexes should still give high enough purification yields to prepare cryoEM grids. I therefore generated new constructs with substitutions of the residues equivalent to *F. johnsoniae* GldL Lys27 (*S. wenxiniae* Lys29), GldL His30 (His32), GldL Glu49 (Glu47), GldM Arg9 (Arg9), and GldM Tyr17 (Tyr17), all of which caused a defect in gliding motility or T9SS when substituted in *F. johnsoniae*. The variants I generated were *Swe*GldL_{K29A}M', *Swe*GldL_{H32A}M', *Swe*GldL_{E47A}M', *Swe*GldL_{E47D}M', *Swe*GldL_{E47Q}M', *Swe*GldLM'_{R9K}, and *Swe*GldLM'_{Y17F}.

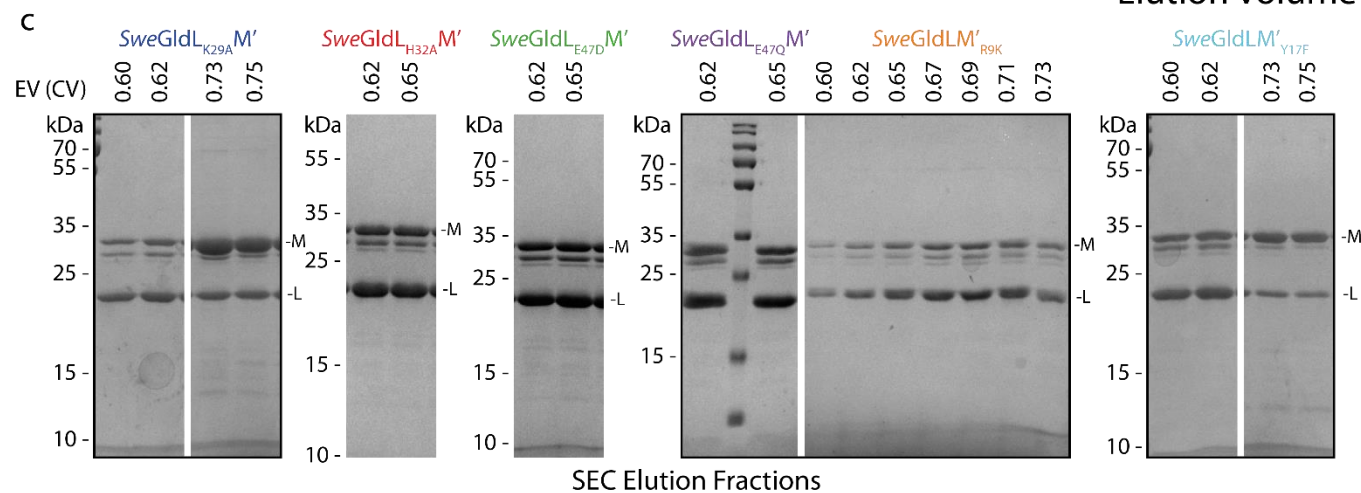
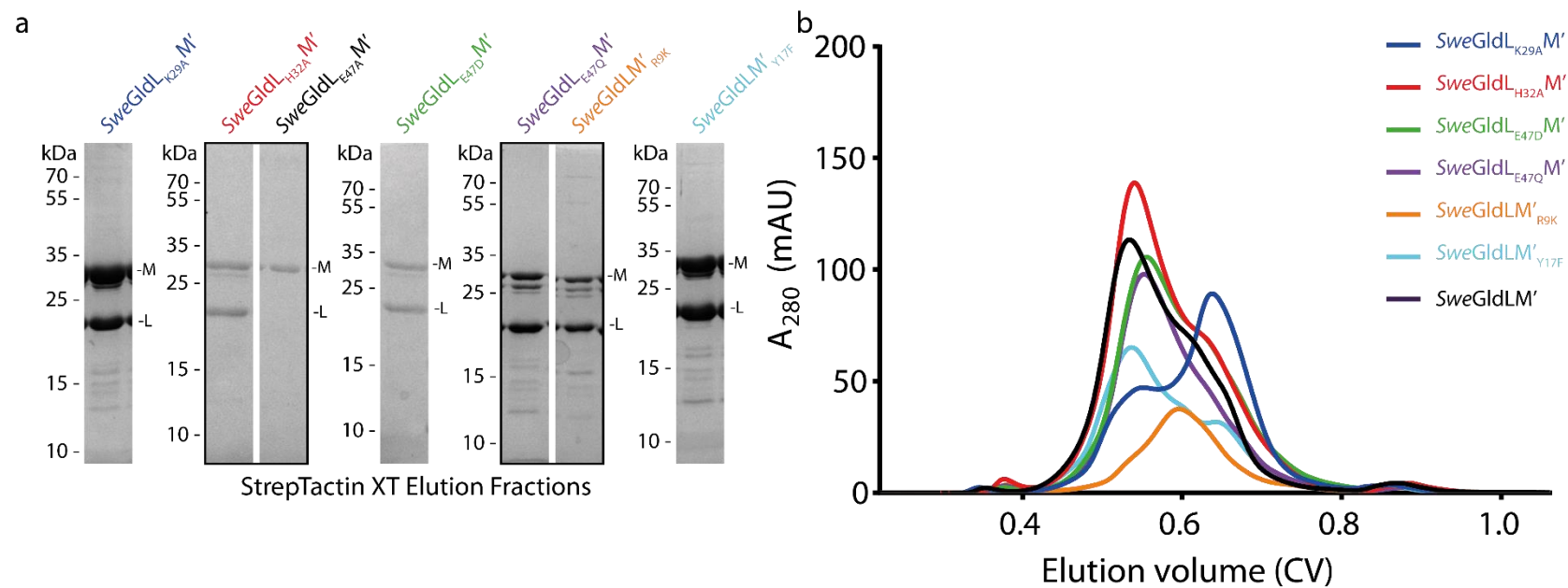


Figure 6-22 Purification of variant SweGldLM' complexes. (Previous page) **a** Elution fractions for the indicated complexes from StrepTactin XT cartridges. **b** SEC traces for the indicated complexes purified with a Superose 6 Increase 10/300 column. WT SweGldLM' is included as a reference. **c** SDS-PAGE analysis of SEC elution fractions for the indicated complexes. EV – elution volume. Black boxes indicate lanes from the same gel. L and M indicate the expected sizes for GldL and GldM'-TS for each construct.

I purified each of these complexes according to the same protocol used for WT SweGldLM', this time from 1 l of starting cell culture. No band corresponding to SweGldL_{E47A} was visible in the StrepTactin XT elution fraction for SweGldL_{E47A}M' (**Figure 6-22a**), suggesting that that variant complex had not properly assembled, so I did not analyse this complex by SEC. Bands at the expected sizes for SweGldL and SweGldM'-TS were visible in the StrepTactin XT elution fractions for all the other variant complexes (**Figure 6-22a**), so I further purified these complexes by SEC. SweGldL_{H32A}M', SweGldL_{E47D}M', SweGldL_{E47Q}M', and SweGldLM'_{Y17F} all behaved similarly to WT SweGldLM' by SEC, with a large peak at ~0.60 CV and a trailing shoulder or smaller peak, suggesting that they were properly assembled (**Figure 6-22b**). For SweGldL_{K29A}M', the second peak at 0.74 CV was larger than the first peak at 0.60 CV. Both peaks contained both SweGldL_{K29A} and SweGldM'-TS, but the band in the second peak corresponding to SweGldM'-TS was much stronger than the band corresponding to SweGldL_{K29A}. This suggests that most SweGldL_{K29A}M' complexes had dissociated and most of the protein purified by the StrepTactin XT cartridge was free SweGldM'-TS. For SweGldLM'_{R9K} there was one broad peak at ~0.67 CV which contained both SweGldL and SweGldM'_{R9K}-TS, suggesting that SweGldLM'_{R9K} complexes had also dissociated. These dissociated complexes suggest that GldL Lys29 and GldM Arg9 may be important for assembly of the GldLM complex, at least in *S. wexxiniae*, although I have not been able to test this hypothesis *in vivo* in *S. wexxiniae*. I pooled fractions from the 0.60 CV peaks and prepared cryoEM grids SweGldL_{K29A}M', SweGldL_{H32A}M', SweGldL_{E47D}M', SweGldL_{E47Q}M', and SweGldLM'_{Y17F} at a concentration of A₂₈₀ = 1. Due to constraints of time and expense the SweGldL_{H32A}M' sample was not analysed by cryoEM.

6.6.2 Solving the structures of variant *SweGldLM'* constructs

I solved the structures of all four of *SweGldL_{K29A}M'*, *SweGldL_{E47D}M'*, *SweGldL_{E47Q}M'*, and *SweGldLM'_{Y17F}* by a similar protocol to the WT *SweGldLM'* structure. JCD collected cryoEM movies using the K3 direct detection camera. I then used SIMPLE 3.0 to automatically pick and extract particles, which I subjected to reference-free 2D classification. I then selected classes with good secondary structure detail, including a high proportion of side view classes on the assumption that the variant proteins had the same preferential orientations as the WT *SweGldLM'*. I then subjected these particles to 3D classification in RELION 3.1 and selected the best classes for 3D autorefinement, followed by Bayesian particle polishing, per-particle CTF refinement and 3D classification without alignment using TMH-, periplasm- or whole protein-focused masks. This resulted in maps at the following resolutions: *SweGldL_{K29A}M'* - 3.5 Å from 108,876 particles, *SweGldL_{E47D}M'* - 3.4 Å from 99,732 particles, *SweGldL_{E47Q}M'* - 3.1 Å from 149,750 particles, and *SweGldLM'_{Y17F}* - 3.5 Å from 161,235 particles. The workflows for each dataset are summarised in **Figure 6-23**, **Figure 6-24**, **Figure 6-25**, and **Figure 6-26**.

To build atomic models of the variant *SweGldLM'* complexes, I docked the *SweGldLM'* model into the EM density using rigid body fitting in *Coot* (Emsley *et al*, 2010). I then used the mutate residue tool to change the substituted residues as appropriate and remodelled the structures as necessary using *Coot*. I refined and validated each structure as described in 2.7.3. Data collection, model building, and refinement statistics are given in Table 6-4.

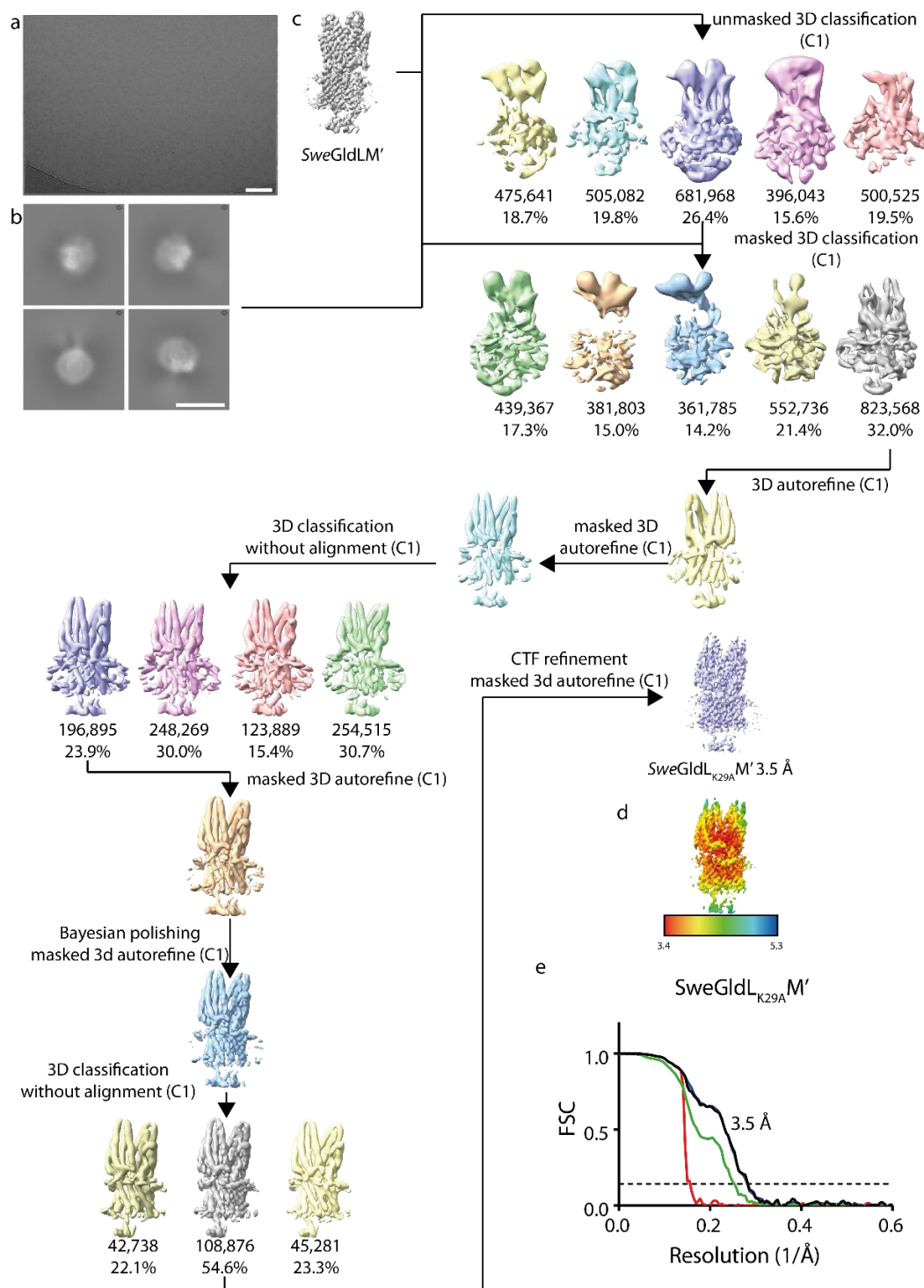


Figure 6-23 Data processing workflow for *SweGldL_{K29A}M'* map. **a** Example micrograph for *SweGldL_{K29A}M'* dataset collected with K3 detector at defocus of approximately 1.7 μm . Scale bar 500 $\text{\AA}$. **b** Representative 2D class averages. Scale bar 100 $\text{\AA}$. **c** Data processing workflow. **d** Local resolution estimates (in $\text{\AA}$) for the sharpened *SweGldL_{K29A}M'* map. **e** Fourier Shell Correlation (FSC) plot for the *SweGldL_{K29A}M'* map. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomised; Green, unmasked; Blue, masked; black, MTF-corrected.

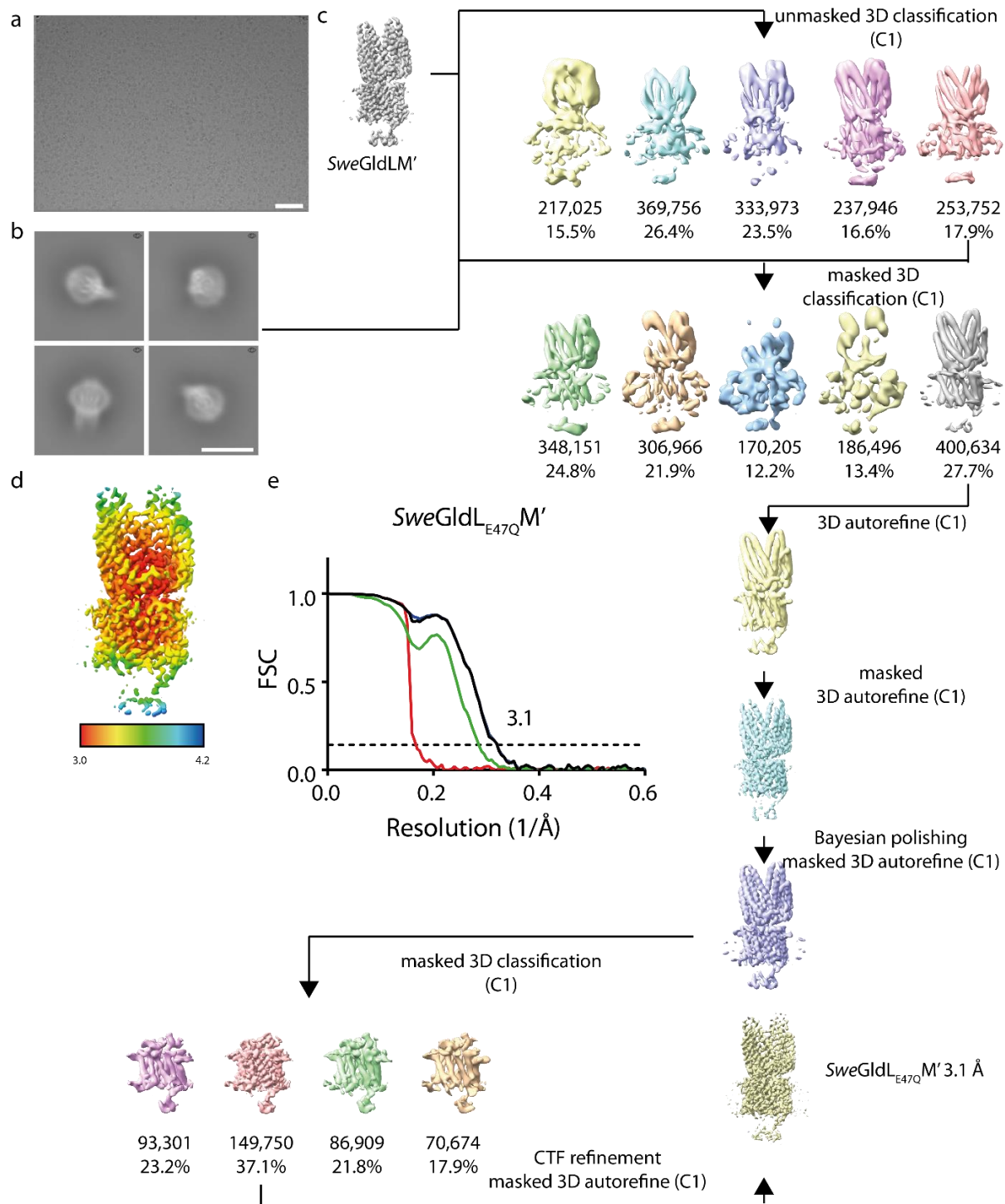


Figure 6-24 Data processing workflow for *SweGldL_{E47Q}M'* map. **a** Example micrograph for *SweGldL_{K29A}M'* dataset collected with K3 detector at defocus of approximately 2.5 μm . Scale bar 500 Å. **b** Representative 2D class averages. Scale bar 100 Å. **c** Data processing workflow. **d** Local resolution estimates (in Å) for the sharpened *SweGldL_{E47Q}M'* map. **e** Fourier Shell Correlation (FSC) plot for the *SweGldL_{E47Q}M'* map. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomised; Green, unmasked; Blue, masked; black, MTF-corrected.

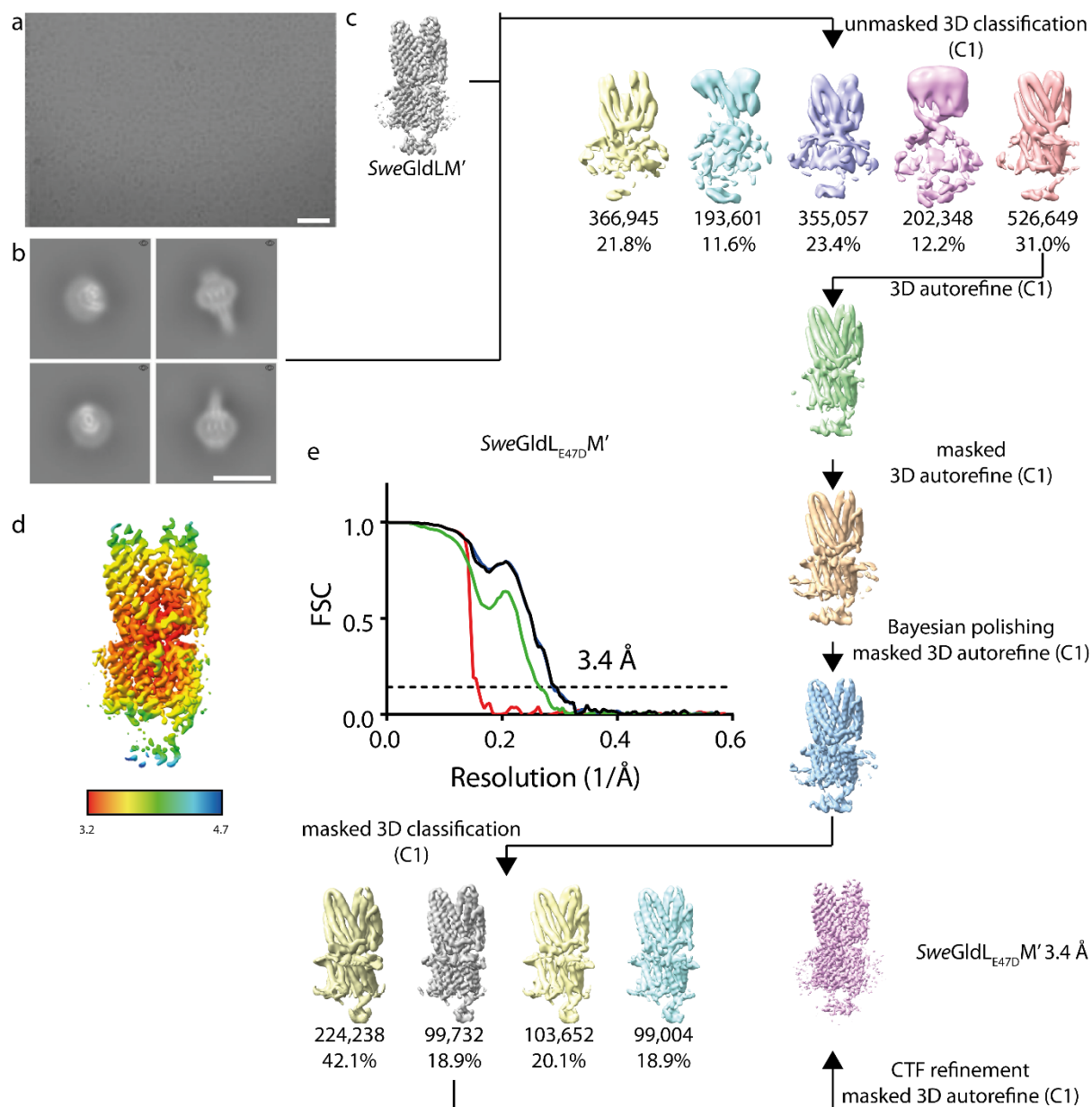


Figure 6-25 Data processing workflow for *SweGldL_{E47D}M'* map. **a** Example micrograph for *SweGldL_{E47D}M'* dataset collected with K3 detector at defocus of approximately 2.3 μm . Scale bar 500 $\text{\AA}$. **b** Representative 2D class averages. Scale bar 100 $\text{\AA}$. **c** Data processing workflow. **d** Local resolution estimates (in $\text{\AA}$) for the sharpened *SweGldL_{E47D}M'* map. **e** Fourier Shell Correlation (FSC) plot for the *SweGldL_{E47D}M'* map. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomised; Green, unmasked; Blue, masked; black, MTF-corrected.

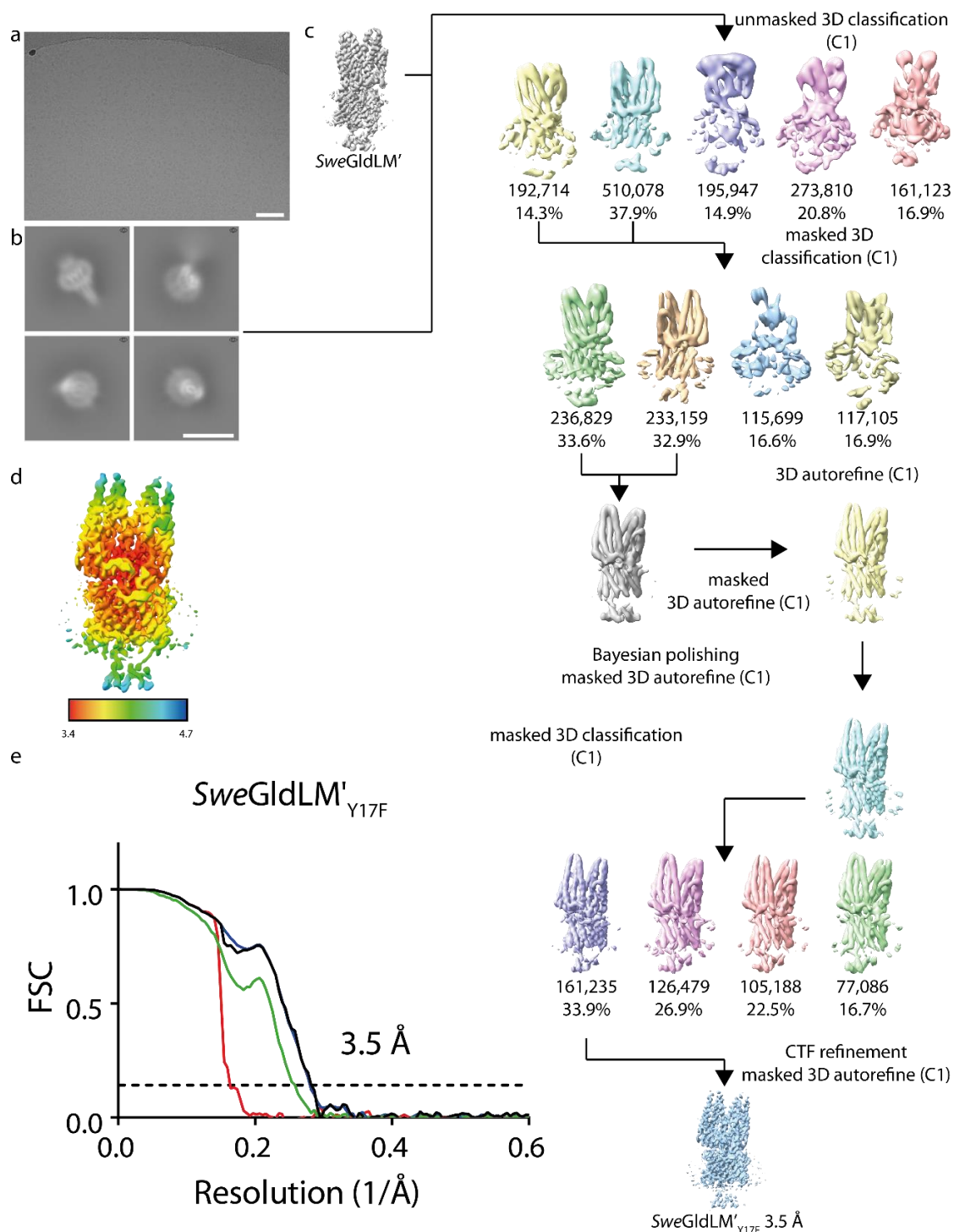


Figure 6-26 Data processing workflow for *SweGldLM'*_{Y17F} map. **a** Example micrograph for *SweGldLM'*_{Y17F} dataset collected with K3 detector at defocus of approximately 2.4 μm . Scale bar 500 $\text{\AA}$. **b** Representative 2D class averages. Scale bar 100 $\text{\AA}$. **c** Data processing workflow. **d** Local resolution estimates (in $\text{\AA}$) for the sharpened *SweGldLM'*_{Y17F} map. **e** Fourier Shell Correlation (FSC) plot for the *SweGldLM'*_{Y17F} map. The resolution at the gold-standard cut-off (FSC = 0.143) is indicated by the dashed line. Curves: Red, phase-randomised; Green, unmasked; Blue, masked; black, MTF-corrected.

Table 6-4 Cryo-EM data collection, refinement, and validation statistics for the *SweGld*_{K29A}M', *SweGld*_{E47Q}M', *SweGld*_{E47D}M', and *SweGld*LM'_{Y17F} structures.

	<i>S. wenxiniae</i> Gld _{K29A} M'	<i>S. wenxiniae</i> Gld _{E47Q} M'	<i>S. wenxiniae</i> Gld _{E47D} M'	<i>S. wenxiniae</i> GldLM' _{Y17F}
Data collection and processing				
Magnification	105,000	105,000	105,000	105,000
Voltage (kV)	300	300	300	300
Electron exposure (e ⁻ Å ⁻²)	59.9	59.2	59.2	59.2
Defocus range (μm)	1.0-3.0	1.0-3.0	1.0-3.0	1.0-3.0
Pixel size (Å)	0.832	0.832	0.832	0.832
Symmetry imposed	C1	C1	C1	C1
Initial particle images (no.)	4,751,486	5,157,115	5,107,067	4,833,448
Final particle images (no.)	108,876	149,750	99,732	161,235
Map resolution (Å)	3.5	3.1	3.4	3.5
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	3.4-5.3	3.0-4.2	3.2-4.7	3.4-4.7
Refinement				
Initial model used	<i>SweGld</i> LM'	<i>SweGld</i> LM'	<i>SweGld</i> LM'	<i>SweGld</i> LM'
Model resolution (Å)	3.5	3.1	3.4	3.5
FSC threshold	0.143	0.143	0.143	0.143
Model resolution range (Å)	3.4-5.3	3.0-4.2	3.2-4.7	3.4-4.7
Map sharpening B factor (Å ²)	-114	-89	-91	-116
Model composition				
Non-hydrogen atoms	5140	5248	5296	5268
Protein residues	672	682	689	685
Ligands	0	0	0	0
B factors (Å²)				
Protein	62.22	54.02	52.14	55.65
Ligand	-	-	-	-
R.m.s. deviations				
Bond lengths (Å)	0.005	0.008	0.005	0.004
Bond angles (°)	0.636	0.668	0.622	0.598

Validation				
MolProbity score	1.84	1.61	1.63	1.78
Clashscore	10.89	7.29	7.88	9.04
Poor rotamers (%)	0.00	0.18	0.00	0.00
Ramachandran plot				
Favoured (%)	95.90	96.71	96.74	95.68
Allowed (%)	4.10	3.29	3.26	4.17
Disallowed (%)	0.00	0.00	0.00	0.15

6.6.3 Structures of *SweGldL_{K29A}M'*, *SweGldL_{E47D}M'*, *SweGldL_{E47Q}M'*, and *SweGldLM'_{Y17F}*

Having solved the structures of four *SweGldLM'* variants and built atomic models, I compared them to the WT *SweGldLM'* structure, by aligning them to WT *SweGldLM'* GldM chain B TMH (**Figure 6-27a-e**). All the variant structures adopted very similar conformations to the WT structure (RMSD 0.2 to 0.3 Å). There were no overall conformational shifts, and the movements of individual residues were mostly very subtle. These small rearrangements of side chains may be due to uncertainties in the positions of residues at modest resolution (3.1 to 3.5 Å) rather than genuine differences between the structures. The only substantial movement was of GldL chain C Gln47 in the *SweGldL_{E47Q}M'* structure, which points up toward the periplasm where GldL Chain C Glu47 in the WT *SweGldLM'* structure points down toward the cytoplasm and GldM Chain B Arg9 (**Figure 6-27d**). Even in this case, it is not possible to be sure that the movement is genuine, as the Glu47 side chains in the WT *SweGldLM'* structure were damaged by the electron beam and decarboxylated. Damage to acidic side chains is commonly seen in cryoEM experiments (Hattne *et al*, 2018) and means that these side chains cannot be confidently modelled (**Figure 6-27f**). When I attempted to refine a model of WT *SweGldLM'* with GldL chain C Glu47 in a position equivalent to that of GldL chain C Gln47 in the *SweGldL_{E47Q}M'* structure the new orientation was unstable and after refinement GldL chain C Glu47 returned to its position in a salt bridge with GldM chain B Arg9. This suggests that the salt bridge between these two residues is genuine but does not offer conclusive proof. Despite these uncertainties, it does make sense that protonation of GldL chain C Glu47, mimicked here by substitution with glutamine, would result in disruption of the Glu47:Arg9 salt bridge. Whilst the protonated Glu47 moves up from the broken salt bridge, the also protonated Arg9 might move down to transfer its proton to another residue or to a water molecule in the cytoplasm. In the *SweGldL_{E47Q}M'* structure, the GldM Chain B Arg9 has only shifted downward very slightly compared to the WT structure (**Figure 6-27d**). It may be the case that further rearrangement requires a negative charge on another copy of GldL Glu47, which

is not possible in the SweGldL_{E47Q}M' variant as all five copies of GldL Glu47 are substituted with Gln47. These limitations make it difficult to draw strong conclusions from these structures.

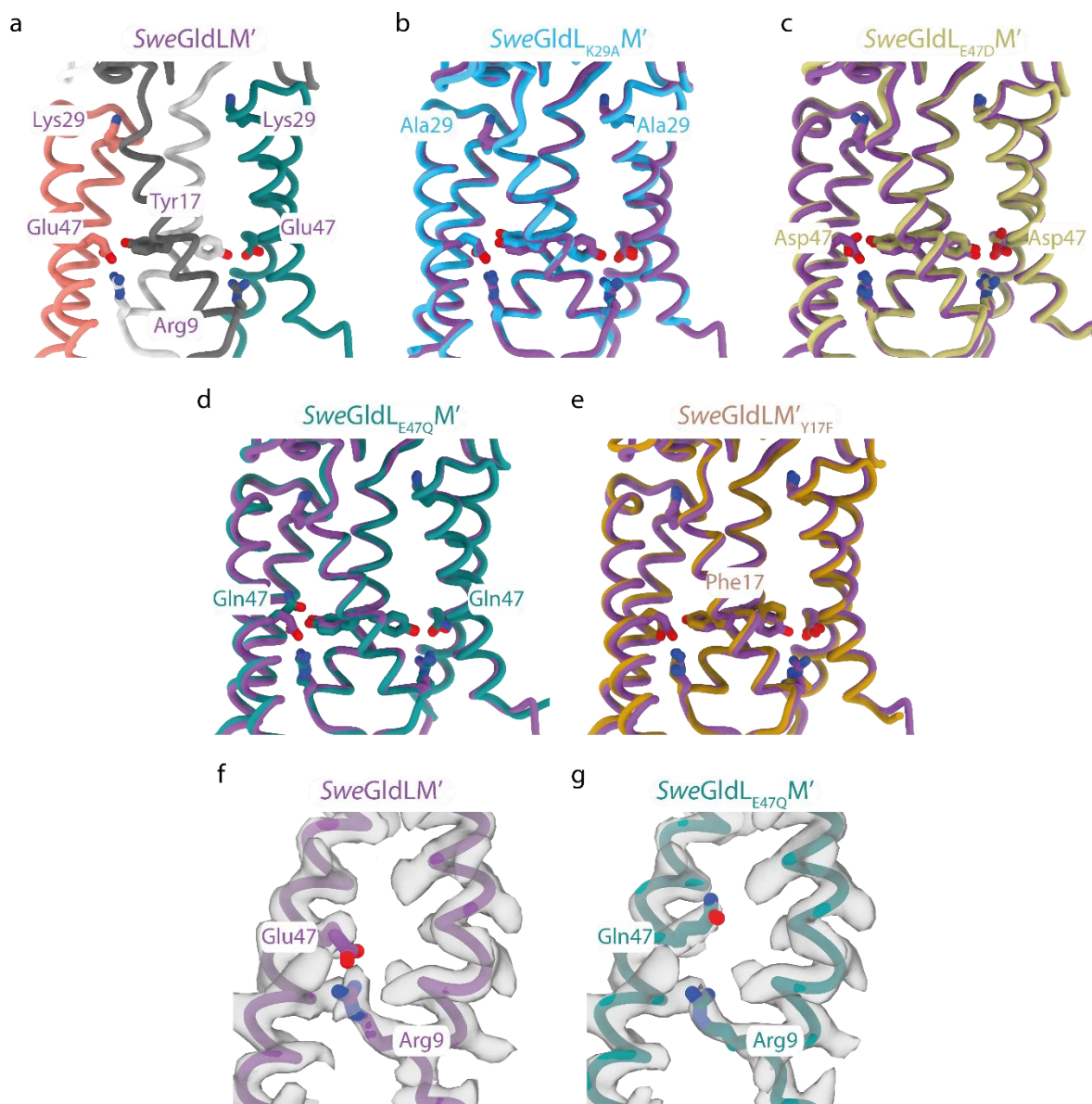


Figure 6-27 Structures of four variant SweGldLM' structures. **a** Transmembrane region of WT SweGldLM', coloured by chain as in **Figure 6-17**. For clarity, GldL chains D, E, and G are hidden. **b-e** Overlay of variant structures SweGldL_{K29A}M' (**b**), SweGldL_{E47D}M' (**c**), SweGldL_{E47Q}M' (**d**), and SweGldLM'_{Y17F} on WT SweGldLM', aligned by the TMH of GldM' Chain B. WT SweGldLM' is coloured purple in each panel. **f-g** Overlay of WT SweGldLM' (**f**) and SweGldL_{E47Q}M' (**g**) GldL chain C helix 2 and GldM chain B with their respective EM density maps.

It would be interesting to study a structure in which only one copy of GldL Glu47, or all but one copy, were substituted with Gln47. This might be possible by expressing a construct with two copies of *S. wexliniae gldL*, one with the WT sequence and one with a mutation encoding the E47Q substitution. However, this would produce complexes with a range of GldL_{WT}:GldL_{E47Q} stoichiometries and isolating complexes with the desired stoichiometry would not be technically feasible. I was able to purify SweGldLM' proteins with the cytoplasmic domain of GldL removed and found that it formed a monodisperse peak by SEC (**Figure 6-28**). This suggests that the TMH bundle of GldLM can assemble without the cytoplasmic domain of GldL. It might therefore be possible to generate a construct with five copies of the GldL TMHs joined by flexible linkers (SweGldL^{TMHx5}). Desired mutations could be introduced into each copy of GldL individually by an appropriate mutagenesis strategy and then the structures of the variant complexes could be solved. It would be necessary to solve the structure of SweGldL^{TMHx5} with the WT sequence to confirm any structural rearrangements in variant complexes were due to the residue substitution and not due to the fusion of copies of SweGldL^{TMH}. It would also be necessary to solve the structure of five variant complexes with the amino acid substitution in each copy of SweGldL^{TMH} to remove any effects of the position of the substitution in the polypeptide chain. This process would be time-consuming and expensive but should be technically feasible.

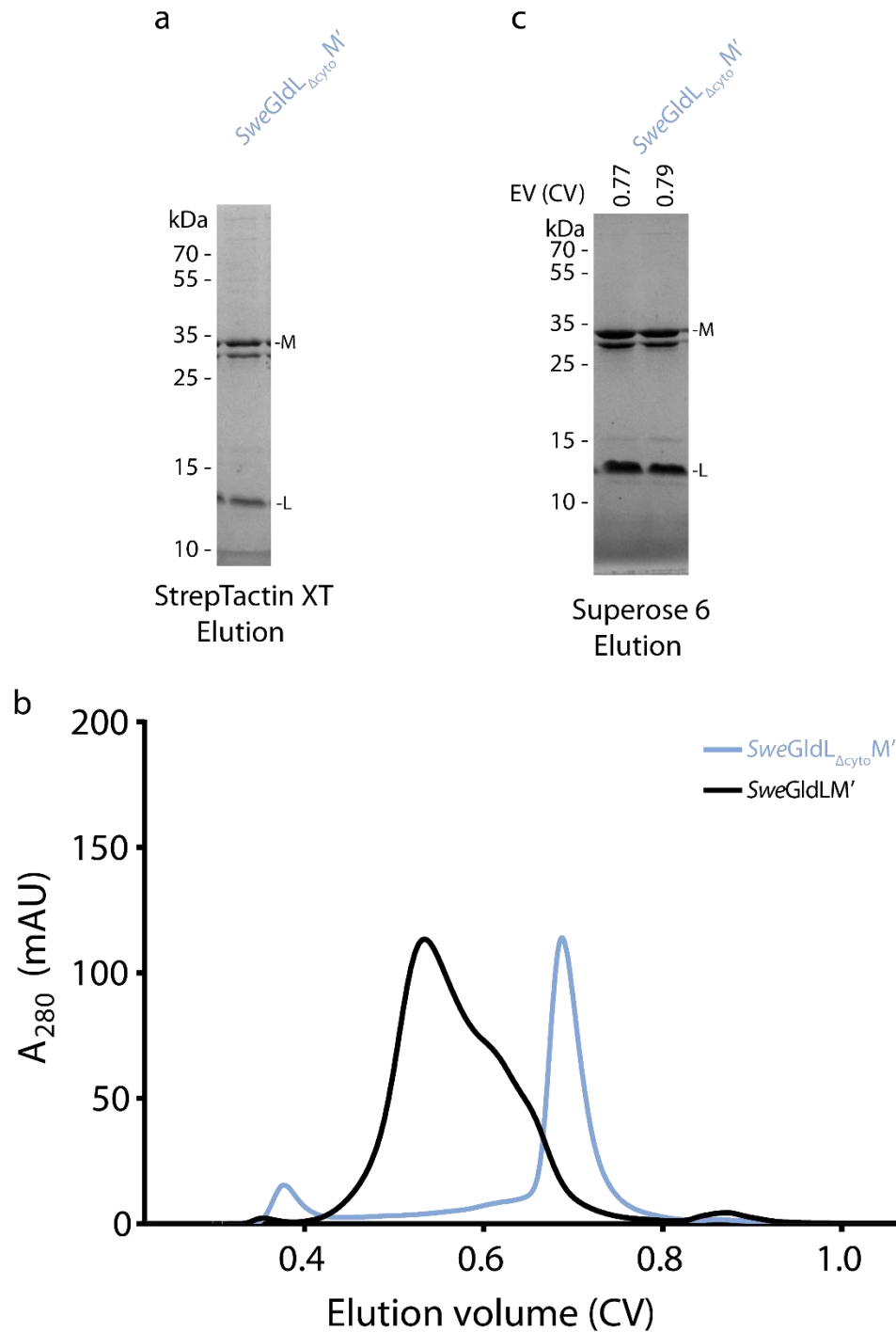


Figure 6-28 Purification of SweGldLM' complexes with residues 75-215 of SweGldL removed (SweGldL Δ cytoM'). **a** SDS-PAGE analysis of StrepTactin XT elution for SweGldL Δ cytoM'. **b** SEC trace for SweGldL Δ cytoM' using a Superose 6 10/300 Increase column. The trace for WT SweGldLM' from **Figure 6-7** is shown for comparison. Due to the removal of ~74 kDa of partly disordered protein the peak for SweGldL Δ cytoM' is sharper than that of SweGldLM' and elutes later. **c** SDS-PAGE analysis of SEC elution fractions for SweGldL Δ cytoM'. EV – elution volume. L and M indicate the expected sizes for GldL Δ cyto and GldM'-TS.

6.7 Discussion

In this chapter I presented structures of GldLM' complexes from *C. canimorsus*, *P. gingivalis*, *S. thermophila* and *S. wenxiniae*. All four structures of GldLM' have the same five GldL: two GldM stoichiometry as the *FjoGldLM'* structure presented in Chapter 4 with two transmembrane helices of GldM surrounded by five pairs of GldL helices. Conservation of this architecture across the whole *Bacteroidetes* phylum confirms that it is conserved across all T9SS systems.

The GldM D1 domains were splayed apart in all GldLM' structures, in contrast to previously solved crystal structures of the periplasmic domains of *F. johnsoniae* GldM (*FjoGldM_{peri}*) and *P. gingivalis* PorM (*PgiPorM_{peri}*) lacking the transmembrane helices (Sato *et al*, 2020; Leone *et al*, 2018), but matching the low-resolution structure of full length *PgiPorLM* (Chapter 3). Cryo-EM analysis of the *CcaGldLM'* construct showed a mix of classes with ordered D2 domains but disordered transmembrane helices and classes with ordered transmembrane helices but disordered D2 domains. Classes with ordered D2 domains had parallel D1 domains, unlike those with disordered D2 domains. This may indicate that the parallel arrangement of D1 domains seen in the crystal structures of *FjoGldM_{peri}* and *PgiPorM_{peri}* is an intermediate physiological state, in which the transmembrane helices have become mobile and poorly resolved. However, if the GldLM motor is driven by proton flow through the membrane, it follows that movements should be transmitted from the transmembrane helices to the periplasmic domains, not *vice versa*. This suggests that the parallel arrangement of the periplasmic D1 domains seen in the *FjoGldM_{peri}* and *PgiPorM_{peri}* structures is not a physiological state.

The conserved polar residues in the transmembrane domain of the GldLM complex that were previously identified as important for motor function (Chapter 5), all adopted similar positions in the motor complexes from different species. The improved resolution of *CcaGldLM'*, *SthGldLM'*, and *SweGldLM'* compared with *FjoGldLM'* allows more confidence in assigning the orientations of these

residues, but the similar orientations do not aid consideration of any protein rearrangements occurring during proton flow and motor rotation. Making substitutions of these residues did not affect the global conformation of the protein. *In vivo*, conformational changes in the GldLM motor occur in the presence of the PMF, a proton electrochemical gradient across the inner membrane. It is not possible to establish an electrochemical gradient across a micelle in solution, but it is possible to establish an electrochemical gradient across the lipid bilayer of a liposome with embedded proteins, which might induce conformational changes in the GldLM complex. Structures of a membrane protein embedded in a liposome have recently been reported (Yao *et al*, 2020), but they were extremely technically challenging to solve so the same technique cannot easily be applied to the small GldLM motor complexes. Conformational changes induced by a proton gradient might also lead to conformational heterogeneity that would make solving a high-resolution structure even more difficult.

PgiPorL has an N-terminal helix that is not found in the structures of the motor complex from other species, and which lies against the detergent micelle. *In vivo* this helix probably lies against the membrane and the positively charged N-terminus may interact with negatively charged phospholipid headgroups. The presence of the helix may be linked to the fact that *P. gingivalis* lacks gliding motility machinery and is non-motile. *F. johnsoniae* and *C. canimorsus* are both described as motile whilst *S. thermophila* and *S. wenxiniae* are described as non-motile. However, both *S. thermophila* and *S. wenxiniae* have homologues of most of the gliding motility machinery genes found in *F. johnsoniae*, so they may in fact be motile under conditions not yet examined. The extensive experimental work needed to determine whether a species truly non-motile mean that it is not possible to determine whether the extra N-terminal helix found in *PgiPorL* is specific to non-motile species by bioinformatic approaches alone. As and when more species are studied for gliding ability or more certain genomic determinants of gliding status are found this question may be better resolved.

Most of the cytoplasmic domain of GldL remains unresolved in my EM structures of the GldLM motor complex, but a small region near the transmembrane bundle could be resolved in *PgiPorLM*', *SthGldLM*', and *SweGldLM*'. This region forms a cage-like structure that may be involved in assembly of the complex or in coordinating motion between GldL subunits. However, deletion of this conserved region in *F. johnsoniae* only caused a small reduction in gliding activity, suggesting it is not critical for function in *F. johnsoniae*.

These structures show that the architecture and core mechanism of the GldLM motor complex is well-conserved across the *Bacteroidetes*. This conservation allows the possibility of conducting *in vitro* experiments that require large amounts of protein with well-expressing constructs from one species such as *S. wenxiniae* whilst conducting *in vivo* experiments in the genetically tractable *F. johnsoniae*. Findings from these experiments should in turn be applicable to important human pathogens such as *P. gingivalis* and *C. canimorsus*.

7 Conclusions and future directions

7.1 Project Summary

In this project I set out to structurally, biochemically, and genetically characterise the GldLM motor complex that powers both the Type IX Secretion System and gliding motility. At the beginning of my project there was no structural information about either protein, although early on in my project crystal structures were published that showed that the periplasmic region of GldM forms a 180 Å long homodimer containing four domains (D1-D4) (Leone *et al*, 2018). The available biochemical and genetic data on the GldLM motor were also quite limited. Both proteins were known to be essential for T9SS secretion and gliding motility. It was also known that GldL and GldM interact with each other and that GldM interacts with the GldK and GldN proteins encoded by the same operon (Vincent *et al*, 2017; Shrivastava *et al*, 2013).

I initially attempted to purify full-length GldL and GldM homologues for structural characterisation (Chapter 3). I optimised the purification of *P. gingivalis* PorLM but was unable to solve a high-resolution structure of the complex due to a combination of preferential orientations of the particles limiting the view distribution and conformational heterogeneity in the periplasmic region of PorM. The previously reported crystal structure of the *F. johnsoniae* GldM periplasmic region (*FjoGldM_{peri}*) showed the D1-D4 domains organised as a linear rod (Leone *et al*, 2018). My cryoEM data suggest that this linear domain organisation was a product of crystallisation and there is in fact flexibility at the boundaries between domains. I later tried to solve the structure of full length GldLM from *C. canimorsus* but was again unsuccessful for similar reasons (Chapter 6). Eventually, once the structure of truncated *F. johnsoniae* GldLM' was available (Chapter 4), Dr Justin Deme (JCD) was able to use this as a reference to solve the structure of *P. gingivalis* PorLM at 8.6 Å (**Figure 3-13**). This structure did not provide detailed structural insight but did resolve discrepancies between the cryoEM structure of *FjoGldLM'* and the crystal structure of *FjoGldM_{peri}*, as discussed below.

After my attempts to solve the structure of full length *PgiPorLM* were unsuccessful, I next purified and characterised truncated constructs of *F. johnsoniae* GldLM and *P. gingivalis* PorLM in which the D2-D4 domains of GldM/PorM had been removed to reduce the issues of conformational heterogeneity and preferential orientations seen with the full-length proteins (Chapter 4). The D1 domains of *P. gingivalis* PorM did not appear to be folded in my *PgiPorLM'*_A construct, suggesting I had not correctly determined the domain boundaries for *P. gingivalis* PorM. The D1 domains of GldM were folded in my *FjoGldLM'*_A and *FjoGldLM'*_B constructs, and by combining cryoEM data from these two constructs JCD was able to solve the structure of *FjoGldLM'* at 3.9 Å (**Figure 4-7** and **Figure 4-8**). The D1 domains of the *FjoGldM'* cryoEM structure had the same fold as in the *FjoGldM*_{peri} crystal structure but were splayed apart in the cryoEM structure compared to the parallel arrangement in the crystal structure. The splaying of the D1 domains was also seen in the 8.6 Å structure of full length *PgiPorLM*. Conservation analysis did not show any conserved interactions across the interface between the D1 domains. These two observations suggest that the splayed arrangement of D1 domains seen in the cryoEM structure of *FjoGldLM'* is more physiological and that the parallel arrangement of D1 domains seen in the crystal structure of *FjoGldM*_{peri} may be an artefactual or intermediate arrangement stabilised by crystal packing. I later attempted to solve the structure of *CcaGldLM''* which includes GldL and GldM up to the D2 domain from *C. canimorsus* (Chapter 6). I found a class of particles where the D1 and D2 domains were folded in a parallel arrangement like the *FjoGldM*_{peri} crystal structure, but the transmembrane helices of GldL and GldM were not well-ordered in *CcaGldLM''* (see 6.4.4). This further supports the hypothesis that the parallel arrangement of D1 domains does not represent a stable physiological state, although it does imply that movement of the periplasmic domains of GldM is coupled to movements of the TMHs of GldL and GldM during GldLM motor activity. In the same *CcaGldLM''* cryoEM dataset, I identified a class of particles where the

TMHs of GldL and GldM were ordered, but the D2 domains of GldM were not ordered. I used this class to solve a structure of *C. canimorsus* GldLM' at 3.0 Å.

After extensive expression screening of diverse protein sequences, I solved additional structures GldLM' complexes from *P. gingivalis*, *S. thermophila*, and *S. wenxiniae* at resolutions between 3.0 and 3.9 Å (Chapter 6). All five GldLM' structures had a very similar overall fold, with splayed GldM D1 domains and ten transmembrane helices from five copies of GldL surrounding two transmembrane helices from two copies of GldM in the detergent micelle. Outside the transmembrane bundle there were some differences between the structures. In the *Fjo*GldLM', *Pgi*PorLM', *Sth*GldLM', and *Swe*GldLM' structures a cage-like structure formed from the region of GldL just C-terminal to the second TMH could be resolved. No EM density corresponding to this structure was visible in the *Cca*GldLM' map. In the *Pgi*PorLM' and *Sth*GldLM' EM density maps the density was sufficiently strong to allow building of this structure (**Figure 6-20**). In the *Swe*GldLM' structure only the part of this cage from one copy of GldL could be built, this copy seemed to be stabilised by bracing from the N-terminus of the adjacent copy of GldL. The cage appeared to be stabilised by hydrogen bonding and packing interactions between conserved residues so may play a role in assembly of the complex or coordinating movement between subunits. However, deletion of this region in *F. johnsoniae* had a very minor impact on gliding motility, suggesting it does not play an important role in GldLM function. The rest of the cytoplasmic domain of GldL was unresolved in all the GldLM' structures, as well as in the full length *Pgi*PorLM low-resolution structure. The region of GldL immediately after the cage-like region is predicted to be intrinsically disordered (**Figure 1-9**). If this region is disordered it would not be resolved by cryoEM and the rest of the cytoplasmic domain would have no fixed position relative to the detergent micelle, making it impossible to resolve the cytoplasmic domain by aligning particle projections based on the micelle. None of the 2D class averages I saw in my data processing resembled the coiled-coil structure that the C-terminal end of the GldL cytoplasmic

domain is predicted to form. It seems likely therefore that the folded part of the GldL cytoplasmic domain is too small a structure to be solved by cryoEM. *P. gingivalis* PorL has a longer sequence N-terminal to the first TMH than the other GldL homologues. The structure of *PgiPorLM'* showed that this sequence forms an α -helix that sits against the detergent micelle. *In vivo* this helix probably sits against the cytoplasmic membrane, though its role is unclear.

Within the transmembrane bundle of the GldLM' structures there were several polar residues in both GldL and GldM. As polar residues are not often found in transmembrane helices and GldLM is believed to conduct the protons, I reasoned that these residues might be involved in proton movement and probed their role by targeted substitutions in *F. johnsoniae* (Chapter 5). Residues that appeared to mediate interactions between copies of the same protein had no impact on T9SS secretion and gliding motility when substituted with non-polar residues. Substitutions of residues that appeared to mediate interactions between GldL and GldM caused defects in T9SS secretion and gliding motility ranging from minor to severe. The strains *gldL-Y13F*, *gldL-H30A*, *gldM-R9A*, and *gldM-Y17F* had defects in gliding motility as measured by colony spreading on agar but had similar levels of T9SS secretion to the WT strain. I interpreted this to mean that secretion of the 100 kDa chitinase ChiA, which was my main readout for T9SS activity, requires less energy than secretion and propulsion of the 650 kDa adhesin SprB for gliding motility. This would mean that when a substitution reduces the power output of the GldLM motor it is still sufficient for ChiA secretion but not for gliding motility. This hypothesis was supported by single molecule-tracking experiments by Andreas Kjær, which showed that strains with large defects in gliding motility but small defects in ChiA secretion were capable of secreting SprB molecules and propelling them along the cell surface but propelled them at slower speeds than the WT strain. The residue GldL Glu49, which is universally conserved (Shrivastava *et al*, 2013) and forms a salt bridge between GldL Chain C Glu49 and GldM Chain B Arg9, proved to be essential to both T9SS secretion and gliding motility. *gldL-E49A* and *gldL-E49Q* strains

had no motility or T9SS activity, whilst the *gldL-E49D* strain was immotile and had only trace T9SS activity.

The combination of the GldLM' structures and *in vivo* data led me to propose a model for GldLM activity as an ion-driven rotary motor. The salt bridge between GldL Glu49 and GldM Arg9 is dependent on the protonation state of the residues, and the inherent asymmetry of the 5:2 stoichiometry means that only one such salt bridge can form at a time. I therefore suggest that flow of protons through the complex would lead to each copy of GldM Arg9 forming then unforming the salt bridge in an alternating pattern, driving rotation of 36 ° each time a new salt bridge forms (**Figure 5-7**). This would cause the GldM dimer to rotate relative to the GldL pentamer, and the rotation could be transmitted to the outer membrane components of the T9SS and gliding motility machinery by the long periplasmic region of GldM. It is unclear whether GldM would rotate continuously through 360 ° or alternate back and forth between two positions. Shrivastava and colleagues reported continuous rotation at constant speed of *F. johnsoniae* cells tethered to a glass slide via an anti-SprB antibody (Shrivastava *et al*, 2015). This might indicate that the GldLM ion-driven motor rotates continuously, but it seems unlikely that a single GldLM motor unit could generate sufficient force to spin an entire cell. Shrivastava *et al*. also reported that the gliding motor driving rotation of cells had similar torque output to the flagellar motor, which is known to depend on multiple GldLM-like MotAB stator units per flagellum ((Shrivastava *et al*, 2015; Reid *et al*, 2006); see also 7.2). It seems likely therefore that multiple GldLM complexes power the mechanism which propels adhesin molecules over the cell surface. Shrivastava and Berg have proposed a rack-and-pinion model in which multiple GldLM motors power a rotary motor (Shrivastava & Berg, 2020). In this model the rotary motor drives rotation of a pinion, which propels a mobile rack to which adhesin molecules are attached (**Figure 1-7**). If the GldLM ion-driven motor in fact moves back and forth between two positions, it might be the case that a linear motor propels the mobile rack around the cell (**Figure 1-8**). The complete gliding

motor probably also contains GldK and GldN, which are encoded by the same operon as GldL and GldM, and whose orthologues PorK and PorN in *P. gingivalis* have been shown to interact with the D4 domain of PorM (Leone *et al*, 2018). More structural and biochemical data are needed to understand the interactions involved in assembly of the gliding motor and how it then drives adhesin movement.

How rotation of the GldM dimer in the GldLM ion-driven motor might lead to protein secretion is not immediately clear. Bacterial two hybrid (BTH) experiments suggest that in *P. gingivalis* PorM interacts with PorP (orthologous to SprF and its homologues in *F. johnsoniae*) (Vincent *et al*, 2017). PorP and PorV are the 14-stranded β -barrel proteins that are respectively required for secretion of Type B and Type A CTD-containing proteins (Kulkarni *et al*, 2019). Structures of the SprA translocon from *F. johnsoniae* suggest that a loop from PorV extends into the lumen of the SprA barrel to contact the Type A CTD of a substrate and then pulls the substrate through the SprA lateral gate to the extracellular side of the outer membrane (**Figure 1-6**; (Lauber *et al*, 2018). SprF seem likely to fulfil the same role for Type B CTD-containing proteins. Rotation of the GldM dimer whilst it is interacting with SprF or PorV might drag it through the outer membrane to pull substrates through the SprA translocon. In *P. gingivalis*, PorM has been shown to interact with PorK and PorN as well as PorP (Vincent *et al*, 2017), so it may be the case that PorM rotation drives movement of PorK and PorN in the outer membrane. PorK and PorN have been reported to form rings 50 nm in diameter (Gorasia *et al*, 2016), but it is unclear how movement of these rings would contribute to protein secretion. As is the case for the involvement of GldLM in gliding motility, more structural and biochemical data are needed to understand how the T9SS assembles and how the GldLM ion-driven motor contributes to the function of the T9SS.

7.2 Comparison to other ion-driven motors

The GldLM complex is the third class of ion-driven motor discovered in bacteria, with the others being ATP synthase and the MotAB stator unit of the bacterial flagellar motor. At the sequence level,

there is no apparent homology between GldLM and either of these motors. The arrangement of subunits in ATP synthase is quite different from that of GldLM. The ATP synthase rotor is a ring of c subunits whilst the stator is an ab_2 rod that sits adjacent to, but does not surround, the rotor (Guo *et al*, 2019). Structures of MotAB, however, show remarkable structural similarity to GldLM (Santiveri *et al*, 2020; Deme *et al*, 2020). Like GldLM, MotA and MotB have a five-to-two subunit stoichiometry in which the single transmembrane helices of two copies of MotB are surrounded by a ring of ten transmembrane helices from MotA (**Figure 7-1a-c**). Note that each copy of MotA has an additional two transmembrane helices and an amphipathic helix that form an outer ring that has no counterpart in the GldLM complex. MotB also has a universally conserved aspartate residue (Asp24 in *Bacillus subtilis*) at a similar height in the membrane to GldL Glu49 (**Figure 7-1b**), although in this case the negatively charged residue is on the central component and is not involved in a salt bridge. A rotary motor model similar to the one I propose for GldLM has been proposed for MotAB (Santiveri *et al*, 2020; Deme *et al*, 2020). In the MotAB model, rather than a salt bridge as is the case in my GldLM model, Asp24 from one copy of MotB binds an H^+ or Na^+ counter-ion, whilst the other copy of Asp24 is unoccupied. Proton flow leads to alternation of which copy of MotB binds the counter-ion, with each alternation causing conformational change that leads to a 36° rotation. In this way the flow of 10 protons or Na^+ ions through the complex would drive one complete rotation, as in my GldLM rotary motor model. In the model of MotAB rotation the outer MotA ring rotates, whilst in my model the inner GldM dimer rotates. The relative motion of the subunits is the same in each case, but in the MotAB model MotB is anchored by binding of its periplasmic domain to the peptidoglycan cell wall, whilst in my GldLM model GldL is anchored via its cytoplasmic domain. Structures of the ExbBD complex, which energises TonB-dependent transport across the inner membrane, show a similar ExbB₅D₂ stoichiometry where the core helices align very well to the core helices of MotAB (Deme *et al*, 2020; Celia *et al*, 2019). Unlike GldLM, ExbBD is related to MotAB at the sequence level, so it

seems likely that other systems with sequence homology to MotAB such as the TolQRA system involved in outer membrane stability and cell division, the ZorAB proteins involved in phage defence, and the AglRQS system involved in *M. xanthus* gliding motility, all share a similar structure and mechanism (Doron *et al*, 2018; Sun *et al*, 2011; Szczepaniak *et al*, 2020).

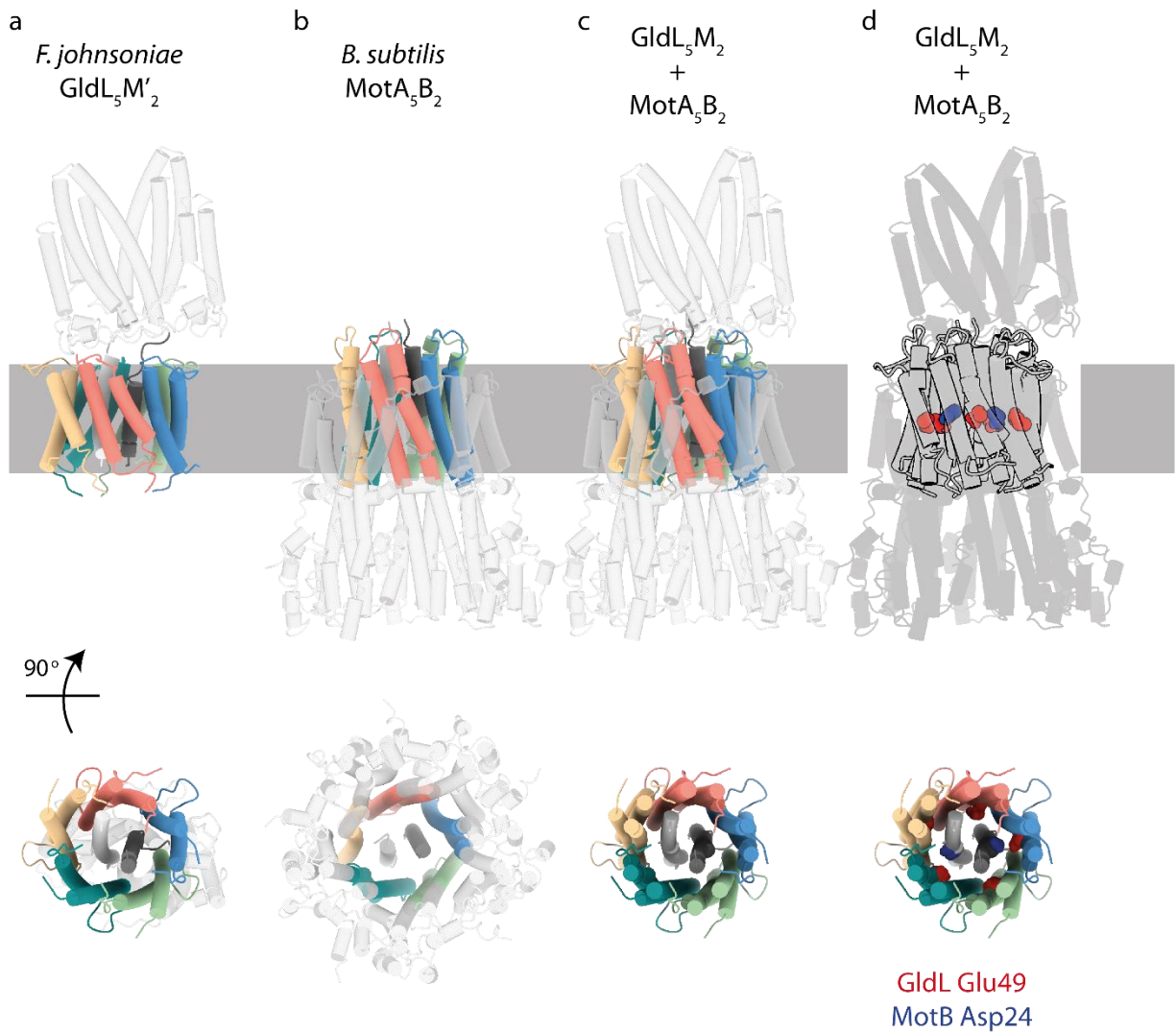


Figure 7-1 The GldLM motor has the same core structure as the MotAB flagellar stator unit. **a** Within-membrane and cytoplasmic views of *F. johnsoniae* GldLM'. The D1 domains of GldM' are shown as transparent. **b** Within-membrane and cytoplasmic views of *Bacillus subtilis* MotAB. Regions of MotA outside the transmembrane core of TMH1 and TMH2 are shown as transparent. Note that the periplasmic domains of MotB were present in the sample used to solve the structure but were not resolved in the structure (Deme *et al*, 2020). **c** Overlay of *Fjo*GldLM' and *B. subtilis* MotAB. **d** Overlay of *Fjo*GldLM' and *B. subtilis* MotAB with the universally conserved GldL Glu49 shown in red and the universally conserved MotB Asp24 shown in blue.

The MotAB-type motor has been adapted to various cellular functions, but so far, the GldLM-type motor has only been associated with gliding motility and the T9SS. Future work may uncover repurposing of GldLM-type motors to different cellular functions. For example, the genome sequences of several *Bacteroides fragilis* strains encode homologues of GldK, GldL, and GldM but not the T9SS translocon SprA or the gliding motility protein GldJ (McBride & Zhu, 2013). This suggests that *B. fragilis* may have repurposed the GldLM ion-driven motor to energise a process separate from the T9SS and gliding motility. Other components of the T9SS may also have been repurposed in different species. *Odoribacter splanchnicus* DSM 20712, *Rhodothermus marinus* DSM 4252, and *Salinibacter ruber* DSM 13855 all encode homologues of the T9SS proteins SprA and SprE in their genomes but lack homologues of the *gldKLMN* operon (McBride & Zhu, 2013). These species may have repurposed SprA to a non-secretory function or have evolved a mechanism to energise secretion that does not involve GldL and GldM. As well as fulfilling diverse cellular functions, some MotAB-type motors have been found to be powered by flow of Na⁺ ions instead of protons (Asai *et al*, 2003). It seems plausible that some GldLM motors, perhaps from marine species, might in future be found to utilise Na⁺ flow instead of proton flow as an energy source.

7.3 Future directions

My project revealed the structure of the core of the GldLM motor complex that drives the Type IX Secretion System and gliding motility, as well as identifying key residues involved in these processes. However, it leaves many unanswered questions and leads to new avenues that future research could explore. In my *in vivo* experiments (Chapter 5), I used targeted substitutions to probe the role of polar residues found in the transmembrane helices of *F. johnsoniae* GldL and GldM. I did not extensively examine the roles of residues that mediate interactions between the periplasmic loops of GldL and the D1 domains of GldM. I used PDBe-PISA to identify a number of residues involved in these interfaces (**Table 4-2**). Future work might therefore probe the role of residues at the periplasmic

interfaces between GldL and GldM by targeted substitutions. It would be interesting to confirm my results by targeted substitutions in another species, and to investigate the roles of structural features of the GldLM complex not seen in *FjoGldLM'*. Such features include the N-terminal cytoplasmic helix of *PgiPorL* (Figure 6-19) and the bracing interaction between the N-terminus of *SweGldL* chain D and the cytoplasmic cage region of *SweGldL* chain C (Figure 6-20). Tools exist for genetic manipulation of *P. gingivalis* and *C. canimorsus*, but no tools for genetic manipulation of *S. thermophila* and *S. wenxiniae* currently exist. It may be possible to adapt existing tools for genetic manipulation of other *Bacteroidetes* species for use in *S. thermophila* and *S. wenxiniae*.

My model of the GldLM complex as an ion-driven rotary motor explains how proton flow through the GldLM complex might lead to mechanical movement. However, there are still many questions remaining to be answered, both in terms of the molecular rearrangements within the GldLM complex during rotation and how these rearrangements are translated into protein secretion and gliding motility. Speculation about the precise route that protons take through the GldLM complex would require higher resolution structures than those I have presented here. Improvements in resolution might be possible through further exploration of sequence space or through further optimisation of cryoEM grids. Recent technological advances have led to sub-2.0 Å structures of membrane proteins solved by cryoEM (Yip *et al*, 2020; Nakane *et al*, 2020). However, the structures of other proteins with C1 symmetry have been solved to higher resolution than the structures presented here using the existing hardware available at the Central Oxford Structural Molecular Imaging Centre (Gérard *et al*, 2020; Harding *et al*, 2021), so it seems likely that sample factors, rather than technological factors, are what limits the resolution of GldLM structures.

I attempted to further investigate the effect of the substitutions which I found affected T9SS secretion and gliding motility function in *F. johnsoniae* GldLM by solving the structures of *SweGldLM'*

complexes with equivalent substitutions. These structures did not show any large conformational changes, so did not give much insight into how the GldLM complex functions. Solving the structure of GldLM at different pH values might lead to different protonation states of the key protonatable residues and drive the adoption of new conformational states, giving information about the conformational changes involved in GldLM function. Recently, structures of proteins embedded within liposomes have been solved using cryoEM (Yao *et al*, 2020). If this method can be recapitulated with GldLM it might make it possible to solve the structure of GldLM in the presence of a proton gradient similar to that which exists *in vivo*. This might induce new structural conformations which would be interesting to study, but this might in turn generate conformational heterogeneity that made solving a high-resolution structure difficult. If structures of GldLM in different conformational states cannot be solved, it might be possible to study the effects of proton flow through molecular dynamics simulations, as has recently been done for the yeast ATP synthase (Kubo *et al*, 2020). Such simulations might suggest a route for protons through the complex, which could then be investigated by further mutagenesis experiments.

Although both gliding motility and the T9SS are dependent on the PMF (Dzink-Fox *et al*, 1997; Hennell James *et al*, 2021), I did not formally prove that GldLM is responsible for conducting protons across the IM. This might be achieved by reconstituting GldLM into liposomes and using a pH sensitive dye such as pyranine to measure the acidification of the liposome lumen (Kano & Fendler, 1978; Clement & Gould, 1981; Parker *et al*, 2014). Techniques have also been developed to study ion flow through individual ion channels in bacterial membranes using patch clamping and giant spheroplasts (Martinac *et al*, 2013). Both liposome acidification and patch clamping could be used to study the effects of substitution of key amino acids on proton flow in GldLM. It is possible, however, that GldLM motors would not function to conduct ions in a liposome or spheroplast context as the periplasmic components of the T9SS and gliding motility machineries would not be present.

Researchers demonstrated the ATP synthase is a rotary motor by attaching a fluorescently labelled actin filament to isolated ATP synthase complexes and using microscopy to observe them spinning (Noji *et al*, 1997). It would be very interesting to attempt to recreate this experiment using the GldLM motor. Noji and colleagues were able to power rotation of the ATP synthase by providing ATP, but to drive rotation of GldLM it would be necessary to embed the GldLM complex in a lipid bilayer with a proton gradient. Doing so and then observing the GldLM complex by microscopy would be technically difficult and there is no guarantee that the GldLM motor would be functional in this context. Light microscopy experiments might also be used to further explore the function of the gliding motility motor and other components of the gliding motility apparatus *in vivo* beyond the work already done (Shrivastava & Berg, 2020; Nakane *et al*, 2013), but care must be taken to ensure that labelling of proteins does not interfere with their function.

In my project I did not explore the roles of either the cytoplasmic domain of GldL or the periplasmic domains of GldM in detail. The cytoplasmic domain of GldL has been proposed to play a role in regulation of the T9SS and gliding motility (Vincent *et al*, 2016), and presumably also plays a role in anchoring GldL so that GldM may rotate in the periplasm (**Figure 5-8**). The D3 and D4 domains of PorM in *P. gingivalis* have been shown to interact with PorK, PorN and PorP by bacterial two hybrid studies (Vincent *et al*, 2017). These interactions probably convert movement of PorM into protein secretion in *P. gingivalis*, and similar interactions probably convert movement of GldM into protein secretion and gliding motility in motile species. There may also be other periplasmic interactors of GldM that have not yet been discovered. To enable full understanding of these two proteins, further work should explore how GldL and GldM interact with other components of the T9SS and gliding motility machinery to convert rotation into the useful work of protein translocation and adhesin propulsion.

These experiments might include purification of GldLM complexes in their native context from *F. johnsoniae* or PorLM complexes from *P. gingivalis*. SprA from *F. johnsoniae* was co-purified with PorV, which was known to be involved in the T9SS, and Plug and PPI, which were not (Lauber *et al*, 2018). The structures of the PorV complex and Plug complex gave insights into how SprA functions as the T9SS translocon. Purification of GldLM complexes from their native context might reveal unknown interaction partners of GldL and GldM or reveal details of known interactions such as the interaction of GldM with GldK and GldN. My own attempts to purify GldLM complexes from *F. johnsoniae* using N-terminal or C-terminal affinity tags were unsuccessful (see

Appendix, 9.1), but introducing affinity tags into the disordered region of GldL between the TMH bundle and C-terminal coiled coil might lead to successful purifications. As well as interacting with GldK, GldN, and potentially the CTDs of substrates (Vincent *et al*, 2017, 2018), GldM must pass through the peptidoglycan cell wall to connect the IM and OM. No peptidoglycan remodelling enzymes have been found to be associated with the T9SS or gliding motility. It may be the case that the enzymes involved are essential to cell viability so cannot be recovered in a mutagenesis screen. These enzymes presumably associate with some component of the T9SS to remodel the peptidoglycan in the appropriate location and so might be identified by pulldown experiments. As well as biochemical and genetic techniques, it may be possible to identify structures involved in the T9SS and gliding motility by electron microscopy. The presence *in vivo* of the large rings formed by PorK and PorN has been confirmed in *P. gingivalis* by cryo-electron tomography (cryoET) and periplasmic “baseplates” attached to SprB fibres have been reported in *F. johnsoniae*, but no other structures have been identified (Liu *et al*, 2007; Gorasia *et al*, 2020). Shrivastava and Berg propose that gliding motility is driven a rotary motor similar in scale to the flagellar motor and attached to a pinion ~180 nm in diameter (Shrivastava & Berg, 2020). Electron tomography has been used to study flagellar motors from a variety of species *in situ* (Chen *et al*, 2011) so it ought to be possible to identify and study the gliding motility motor if it is of a similar size to the flagellar motor. GldJ was observed to have a helical distribution when imaged with immunogold labelling and helical structures in the periplasm of *F. johnsoniae* have been observed by quick-freeze-replica electron microscopy (Braun & McBride, 2005; Katayama *et al*, 2019). It remains to be determined, however, if these helical structures are composed of GldJ and what other components compose and anchor these structures. These questions might be probed by cryoET in combination with further genetic and biochemical analyses.

The structures I present here represent a step forward in our understanding of the GldL and GldM proteins. However, many further experiments are needed to fully understand the role of GldL and GldM in the overall T9SS and gliding motility systems.

8 Bibliography

- Abbanat DR, Leadbetter ER, Godchaux W & Escher A (1986) Sulphonolipids are molecular determinants of gliding motility. *Nature* 324: 367–369
- Abrahams JP, Leslie AGW, Lutter R & Walker JE (1994) Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* 370: 621–628
- Adler J (1966) Effect of amino acids and oxygen on chemotaxis in *Escherichia coli*. *Journal of Bacteriology* 92: 121–129
- Afonine PV, Poon BK, Read RJ, Sobolev OV, Terwilliger TC, Urzhumtsev A & Adams PD (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica D* 74: 531–544
- Agarwal S, Hunnicutt DW & McBride MJ (1997) Cloning and characterization of the *Flavobacterium johnsoniae* (*Cytophaga johnsonae*) gliding motility gene, *gldA*. *Proceedings of the National Academy of Sciences* 94: 12139–12144
- Akeda Y & Galán JE (2005) Chaperone release and unfolding of substrates in type III secretion. *Nature* 437: 911–915
- Albuquerque L, Rainey FA, Nobre MF & da Costa MS (2011) *Schleiferia thermophila* gen. nov., sp. nov., a slightly thermophilic bacterium of the phylum “Bacteroidetes” and the proposal of *Schleiferiaceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology* 61: 2450–2455
- Al-Otaibi NS, Taylor AJ, Farrell DP, Tzokov SB, DiMaio F, Kelly DJ & Bergeron JRC (2020) The cryo-EM structure of the bacterial flagellum cap complex suggests a molecular mechanism for filament elongation. *Nature Communications* 11: 3210
- Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410
- Armitage JP & Macnab RM (1987) Unidirectional, intermittent rotation of the flagellum of *Rhodobacter sphaeroides*. *Journal of Bacteriology* 169: 514–518
- Arutyunov D & Frost LS (2013) F conjugation: back to the beginning. *Plasmid* 70: 18–32
- Asai Y, Yakushi T, Kawagishi I & Homma M (2003) Ion-coupling Determinants of Na⁺-driven and H⁺-driven Flagellar Motors. *Journal of Molecular Biology* 327: 453–463
- Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T & Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* 44: W344–W350
- Barbeyron T, L’Haridon S, Corre E, Kloareg B & Potin P (2001) *Zobellia galactanovorans* gen. nov., sp. nov., a marine species of *Flavobacteriaceae* isolated from a red alga, and classification of [*Cytophaga*] *uliginosa* (ZoBell and Upham 1944) Reichenbach 1989 as *Zobellia uliginosa* gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 51: 985–997

- Barbier P, Rochat T, Mohammed HH, Wiens GD, Bernardet J-F, Halpern D, Duchaud E & McBride MJ (2020) The type IX secretion system is required for virulence of the fish pathogen *Flavobacterium psychrophilum*. *Applied and Environmental Microbiology* 86: e00799-20
- Baumgarten T, Schlegel S, Wagner S, Löw M, Eriksson J, Bonde I, Herrgård MJ, Heipieper HJ, Nørholm MHH, Slotboom DJ, *et al* (2017) Isolation and characterization of the *E. coli* membrane protein production strain Mutant56(DE3). *Scientific Reports* 7: 45089
- Beachey EH (1981) Bacterial Adherence: Adhesin-Receptor Interactions Mediating the Attachment of Bacteria to Mucosal Surfaces. *The Journal of Infectious Diseases* 143: 325–345
- Berg HC & Anderson RA (1973) Bacteria swim by rotating their flagellar filaments. *Nature* 245: 380–382
- Berg HC & Brown DA (1972) Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* 239: 500–504
- Bernardet J-F & Bowman J (2011) Genus I *Flavobacterium*. In *Bergey's Manual of Systematic Bacteriology*, Krieg N Staley J Brown D Hedlund B Paster B Ward N Ludwig W & Whitman W (eds) pp 112–154. New York: Springer
- Bernardet J-F, Nakagawa Y & Holmes B (2002) Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. *International Journal of Systematic and Evolutionary Microbiology* 52: 1049–1070
- Bernardet J-F, Segers P, Vancanneyt M, Berthe F, Kersters K & Vandamme P (1996) Cutting a gordian knot: Emended classification and description of the genus *Flavobacterium*, emended description of the family *Flavobacteriaceae*, and proposal of *Flavobacterium hydatis* nom. nov. (basonym, *Cytophaga aquatilis* Strohl and Tait 1978). *International Journal of Systematic Bacteriology* 46: 128–148
- Bertani G (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of Bacteriology* 62: 293–300
- Bhatty M, Laverde Gomez J & Christie PJ (2013) The expanding bacterial type IV secretion lexicon. *Research in microbiology* 164: 620–639
- Blakemore R (1975) Magnetotactic bacteria. *Science* 190: 377–379
- Block SM, Segall JE & Berg HC (1982) Impulse responses in bacterial chemotaxis. *Cell* 31: 215–226
- Böck D, Medeiros JM, Tsao H-F, Penz T, Weiss GL, Aistleitner K, Horn M & Pilhofer M (2017) In situ architecture, function, and evolution of a contractile injection system. *Science* 357: 713–717
- Brackmann M, Wang J & Basler M (2018) Type VI secretion system sheath inter-subunit interactions modulate its contraction. *EMBO reports* 19: 225–233
- Braun TF, Khubbar MK, Saffarini DA & McBride MJ (2005) *Flavobacterium johnsoniae* gliding motility genes identified by mariner mutagenesis. *Journal of Bacteriology* 187: 6943–52
- Braun TF & McBride MJ (2005) *Flavobacterium johnsoniae* GldJ is a lipoprotein that is required for gliding motility. *Journal of Bacteriology* 187: 2628–2637

- Brenner D, Hollis D, Fanning G & Weaver R (1989) *Capnocytophaga canimorsus* sp. nov. (formerly CDC group DF-2), a cause of septicemia following dog bite, and *C. cynodegmi* sp. nov., a cause of localized wound infection following dog bite. *Journal of clinical microbiology* 27: 231–235
- Brook I (2010) The role of anaerobic bacteria in bacteremia. *Anaerobe* 16: 183–189
- Buczolits S, Denner EBM, Vybiral D, Wieser M, Kämpfer P & Busse HJ (2002) Classification of three airborne bacteria and proposal of *Hymenobacter aerophilus* sp. nov. *International journal of systematic and evolutionary microbiology* 52: 445–456
- Bunduc CM, Fahrenkamp D, Wald J, Ummels R, Bitter W, Houben ENG & Marlovits TC (2021) Structure and dynamics of a mycobacterial type VII secretion system. *Nature* 593: 445–448
- Burrows LL (2012) *Pseudomonas aeruginosa* Twitching Motility: Type IV Pili in Action. *Annual Review of Microbiology* 66: 493–520
- Busch A & Waksman G (2012) Chaperone–usher pathways: diversity and pilus assembly mechanism. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 1112–1122
- Cabral L, Persinoti GF, Paixao DAA, Martins MP, Chinaglia M, Morais MAB, Pirolla RAS, Generoso WC, Maciel LF, Lombard V, *et al* (2021) Gut microbiome of capybara, the Amazon master of the grasses, harbors unprecedented enzymatic strategies for plant glycans breakdown. *PREPRINT (Version 1) available at Research Square*
- Caesar J, Reboul CF, Machello C, Kiesewetter S, Tang ML, Deme JC, Johnson S, Elmlund D, Lea SM & Elmlund H (2020) SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time. *Journal of Structural Biology: X* 4: 100040
- Callaway E (2015) The revolution will not be crystallized: A new method sweeps through structural biology. *Nature* 525: 172–174
- Cao J, Lai Q, Li G & Shao Z (2014) *Pseudopedobacter beijingensis* gen. nov., sp. nov., isolated from coking wastewater activated sludge, and reclassification of *Pedobacter saltans* as *Pseudopedobacter saltans* comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 64: 1853–1858
- Cao P & Wall D (2020) The Fluidity of the Bacterial Outer Membrane Is Species Specific. *BioEssays* 42: 1900246
- Cascales E (2008) The type VI secretion toolkit. *EMBO reports* 9: 735–741
- Cascales E & Christie PJ (2004) Definition of a Bacterial Type IV Secretion Pathway for a DNA Substrate. *Science* 304: 1170–1173
- Celia H, Botos I, Ni X, Fox T, de Val N, Lloubes R, Jiang J & Buchanan S (2019) Cryo-EM structure of the bacterial Ton motor subcomplex ExbB-ExbD provides information on structure and stoichiometry. *Communications Biology* 2: 358–358
- Chang LYE, Pate JL & Betzig RJ (1984) Isolation and characterization of nonspreading mutants of the gliding bacterium *Cytophaga johnsonae*. *Journal of Bacteriology* 159: 26–35

- Chang Y-W, Rettberg LA, Treuner-Lange A, Iwasa J, Sogaard-Andersen L & Jensen GJ (2016) Architecture of the type IVa pilus machine. *Science* 351: aad2001
- Chau R, Bhaya D & Huang K (2017) Emergent Phototactic Responses of Cyanobacteria under Complex Light Regimes. *mBio* 8: 330–16
- Chen S, Beeby M, Murphy GE, Leadbetter JR, Hendrixson DR, Briegel A, Li Z, Shi J, Tocheva EI, Müller A, *et al* (2011) Structural diversity of bacterial flagellar motors. *The EMBO Journal* 30: 2972–2981
- Cherrak Y, Rapisarda C, Pellarin R, Bouvier G, Bardiaux B, Allain F, Malosse C, Rey M, Chamot-Rooke J, Cascales E, *et al* (2018) Biogenesis and structure of a type VI secretion baseplate. *Nature Microbiology* 3: 1404–1416
- Cheung M, Kajimura N, Makino F, Ashihara M, Miyata T, Kato T, Namba K & Blocker AJ (2013) A method to achieve homogeneous dispersion of large transmembrane complexes within the holes of carbon films for electron cryomicroscopy. *Journal of Structural Biology* 182: 51–56
- Christie PJ (2019) The Rich Tapestry of Bacterial Protein Translocation Systems. *The Protein Journal* 38: 389–408
- Christie PJ, Whitaker N & González-Rivera C (2014) Mechanism and structure of the bacterial type IV secretion systems. *Biochimica et Biophysica Acta* 1843: 1578–1591
- Church MJ (2008) Resource control of bacterial dynamics in the sea. In *Microbial Ecology of the Oceans*, Kirchman DL (ed) pp 335–382. Hoboken, NJ: Wiley and Sons
- Clement NR & Gould JM (1981) Pyranine (8-Hydroxy-1,3,6-pyrenetrisulfonate) as a Probe of Internal Aqueous Hydrogen Ion Concentration in Phospholipid Vesicles. *Biochemistry* 20: 1534–1538
- Costa TRD, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M & Waksman G (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology* 13: 343–359
- Costa TRD, Harb L, Khara P, Zeng L, Hu B & Christie PJ (2021) Type IV secretion systems: Advances in structure, function, and activation. *Molecular Microbiology* 115: 436–452
- Costerton JW, Ingram JM & Cheng K-J (1974) Structure and Function of the Cell Envelope of Gram-Negative Bacteria. *Bacteriological Reviews*: 87–110
- Coykendall AL, Kaczmarek FS & Slots J (1980) Genetic Heterogeneity in *Bacteroides asaccharolyticus* (Holdeman and Moore 1970) Finegold and Barnes 1977 (Approved Lists, 1980) and Proposal of *Bacteroides gingivalis* sp. nov. and *Bacteroides macacae* (Slots and Genco) comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 30: 559–564
- Crooks GE, Hon G, Chandonia JM & Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Research* 14: 1188–1190
- Delorenzi M & Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics (Oxford, England)* 18: 617–625

- Deme JC, Johnson S, Vickery O, Aron A, Monkhouse H, Griffiths T, James RH, Berks BC, Coulton JW, Stansfeld PJ, *et al* (2020) Structures of the stator complex that drives rotation of the bacterial flagellum. *Nature Microbiology* 5: 1553–1564
- Denise R, Abby SS & Rocha EPC (2019) Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility. *PLOS Biology* 17: e3000390
- DePamphilis ML & Adler J (1971) Purification of Intact Flagella from *Escherichia coli* and *Bacillus subtilis*. *Journal of Bacteriology* 105: 376–383
- Dhillon A, Deme JC, Furlong E, Roem D, Jongerius I, Johnson S & Lea SM (2021) Molecular basis for *Bordetella pertussis* interference with complement, coagulation, fibrinolytic, and contact activation systems: The cryo-em structure of the vag8-c1 inhibitor complex. *mBio* 12: e02823-20
- de Diego I, Ksiazek M, Mizgalska D, Koneru L, Golik P, Szmigielski B, Nowak M, Nowakowska Z, Potempa B, Houston JA, *et al* (2016) The outer-membrane export signal of *Porphyromonas gingivalis* type IX secretion system (T9SS) is a conserved C-terminal β -sandwich domain. *Scientific Reports* 6: 23123
- Dietsche T, Tesfazgi Mebrhatu M, Brunner MJ, Abrusci P, Yan J, Franz-Wachtel M, Schärfe C, Zilkenat S, Grin I, Galán JE, *et al* (2016) Structural and Functional Characterization of the Bacterial Type III Secretion Export Apparatus. *PLOS Pathogens* 12: e1006071
- Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G & Sorek R (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359: ear4120
- Doyle MT & Bernstein HD (2021) BamA forms a translocation channel for polypeptide export across the bacterial outer membrane. *Molecular Cell* 81: 2000–2012
- Drew D, Lerch M, Kunji E, Slotboom DJ & de Gier JW (2006) Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods* 3: 303–313
- Du Z-J, Wang Y, Dunlap C, Rooney AP & Chen G-J (2014) *Draconibacterium orientale* gen. nov., sp. nov., isolated from two distinct marine environments, and proposal of *Draconibacteriaceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology* 64: 1690–1696
- Dzink-Fox JAL, Leadbetter ER & Godchaux W (1997) Acetate acts as a protonophore and differentially affects bead movement and cell migration of the gliding bacterium *Cytophaga johnsonae* (*Flavobacterium johnsoniae*). *Microbiology* 143: 3693–3701
- Egan SM & Schleif RF (1993) A regulatory cascade in the induction of *rhaBAD*. *Journal of Molecular Biology* 234: 87–98
- Emsley P, Lohkamp B, Scott WG & Cowtan KD (2010) Features and development of *Coot*. *Acta Crystallographica* D66: 486–501
- Epler CR, Dickenson NE, Bullitt E & Picking WL (2012) Ultrastructural analysis of IpaD at the tip of the nascent MxiH type III secretion apparatus of *Shigella flexneri*. *Journal of Molecular Biology* 420: 29–39

- Erhardt M, Namba K & Hughes KT (2010) Bacterial Nanomachines: The Flagellum and Type III Injectisome. *Cold Spring Harbor Perspectives in Biology* 2: a000299
- Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE & Lemon KP (2018) New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems* 3: e00187-18
- Evans ML, Chorell E, Taylor JD, Åden J, Göthesson A, Li F, Koch M, Sefer L, Matthews S, Wittung-Stafshede P, *et al* (2015) The bacterial curli system possesses a potent and selective inhibitor of amyloid formation. *Molecular cell* 57: 445–55
- Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN & Alva V (2020) Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics* 72: e108
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker JM (ed) pp 571–607. Humana Press
- Gérard SF, Hall BS, Zaki AM, Corfield KA, Mayerhofer PU, Costa C, Whelligan DK, Biggin PC, Simmonds RE & Higgins MK (2020) Structure of the Inhibited State of the Sec Translocon. *Molecular Cell* 79: 406–415
- Glew MD, Veith PD, Chen D, Gorasia DG, Peng B & Reynolds EC (2017) PorV is an Outer Membrane Shuttle Protein for the Type IX Secretion System. *Scientific Reports* 7: 8970
- Glew MD, Veith PD, Peng B, Chen Y-Y, Gorasia DG, Yang Q, Slakeski N, Chen D, Moore C, Crawford S, *et al* (2012) PG0026 Is the C-terminal Signal Peptidase of a Novel Secretion System of *Porphyromonas gingivalis*. *Journal of Biological Chemistry* 287: 24605–24617
- Glöckner FO, Fuchs BM & Amann R (1999) Bacterioplankton compositions of lakes and oceans: A first comparison based on fluorescence in situ hybridization. *Applied and Environmental Microbiology* 65: 3721–3726
- Godchaux W & Leadbetter ER (1988) Sulfonolipids are localized in the outer membrane of the gliding bacterium *Cytophaga johnsonae*. *Archives of Microbiology* 150: 42–47
- Gómez-Pereira PR, Schüler M, Fuchs BM, Bennke C, Teeling H, Waldmann J, Richter M, Barbe V, Bataille E, Glöckner FO, *et al* (2012) Genomic content of uncultured *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean. *Environmental Microbiology* 14: 52–66
- Goodsell DS & Olson AJ (2000) Structural Symmetry and Protein Function. *Annual Review of Biophysics and Biomolecular Structure* 29: 105–153
- Gorasia DG, Chreifi G, Seers CA, Butler CA, Heath JE, Glew MD, McBride MJ, Subramanian P, Kjær A, Jensen GJ, *et al* (2020) *In situ* structure and organisation of the type IX secretion system. *bioRxiv* doi: 10.1101/2020.05.13.094771-[PREPRINT]

- Gorasia DG, Veith PD, Chen D, Seers CA, Mitchell HA, Chen Y-Y, Glew MD, Dashper SG & Reynolds EC (2015) *Porphyromonas gingivalis* Type IX Secretion Substrates Are Cleaved and Modified by a Sortase-Like Mechanism. *PLOS Pathogens* 11: e1005152
- Gorasia DG, Veith PD, Hanssen EG, Glew MD, Sato K, Yukitake H, Nakayama K & Reynolds EC (2016) Structural Insights into the PorK and PorN Components of the *Porphyromonas gingivalis* Type IX Secretion System. *PLOS Pathogens* 12: e1005820
- Goyal P, Krasteva P v., Gerven N van, Gubellini F, Broeck I van den, Troupiotis-Tsaïlaki A, Jonckheere W, Péhau-Arnaudet G, Pinkner JS, Chapman MR, *et al* (2014) Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* 516: 250–253
- Grimaud R, Kessel M, Beuron F, Steven AC & Maurizi MR (1998) Enzymatic and structural similarities between the *Escherichia coli* ATP-dependent proteases, ClpXP and ClpAP. *The Journal of Biological Chemistry* 273: 12476–12481
- Grimm J, Brown T, English B, Lionnet T & Lavis L (2017) Synthesis of Janelia Fluor HaloTag and SNAP-Tag Ligands and Their Use in Cellular Imaging Experiments. In *Super-Resolution Microscopy: Methods and Protocols*, Erfle H (ed) pp 179–188. New York, NY: Springer New York
- Grimm JB, English BP, Choi H, Muthusamy AK, Mehl BP, Dong P, Brown TA, Lippincott-Schwartz J, Liu Z, Lionnet T, *et al* (2016) Bright photoactivatable fluorophores for single-molecule imaging. *Nature Methods* 2016 13:12 13: 985–988
- Grossman AS, Mauer TJ, Forest KT & Goodrich-Blair H (2021) A widespread bacterial secretion system with diverse substrates. *mBio* 12: e01956-21
- Guo H, Suzuki T & Rubinstein JL (2019) Structure of a bacterial ATP synthase. *eLife* 8: e43128
- Guo Y, Hu D, Guo J, Wang T, Xiao Y, Wang X, Li S, Liu M, Li Z, Bi D, *et al* (2017) *Riemerella anatipestifer* Type IX secretion system is required for virulence and gelatinase secretion. *Frontiers in Microbiology* 8: 2553
- Haldimann A, Daniels LL & Wanner BL (1998) Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the *Escherichia coli* phosphate regulon. *Journal of Bacteriology* 180: 1277–1286
- Hammer ND, McGuffie BA, Zhou Y, Badtke MP, Reinke AA, Brännström K, Gestwicki JE, Olofsson A, Almqvist F & Chapman MR (2012) The C-terminal repeating units of CsgB direct bacterial functional amyloid nucleation. *Journal of Molecular Biology* 422: 376–389
- Harding RJ, Deme JC, Hevler JF, Tamara S, Lemak A, Cantle JP, Szewczyk MM, Zuo X, Loppnau P, Seitova A, *et al* (2021) HAP40 orchestrates huntingtin structure for 1 differential interaction with polyglutamine 2 expanded exon 1. *bioRxiv* doi: 10.1101/2021.04.02.438217-[PREPRINT]
- Harshey RM (1994) Bees aren't the only ones: swarming in Gram-negative bacteria. *Molecular Microbiology* 13: 389–394
- Hattne J, Shi D, Glynn C, Zee C-T, Gallagher-Jones M, Martynowycz MW, Rodriguez JA & Gonen T (2018) Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. *Structure* 26: 759–766

- von Heijne G & Gavel Y (1988) Topogenic signals in integral membrane proteins. *European Journal of Biochemistry* 174: 671–678
- Hendrix RW (1978) Symmetry mismatch and DNA packaging in large bacteriophages. *Proceedings of the National Academy of Sciences* 75: 4779–4783
- Hennell James R, Deme JC, Kjær A, Alcock F, Silale A, Lauber F, Johnson S, Berks BC & Lea SM (2021) Structure and mechanism of the proton-driven motor that powers type 9 secretion and gliding motility. *Nature Microbiology* 6: 221–233
- Henrichsen J (1972) Bacterial surface translocation: a survey and a classification. *Bacteriological Reviews* 36: 478–503
- Herzberg MC & Meyer MW (1996) Effects of oral flora on platelets: possible consequences in cardiovascular disease. *Journal of Periodontology* 67: 1138–1142
- Hu B, Khara P, Song L, Lin AS, Frick-Cheng AE, Harvey ML, Cover TL & Christie PJ (2019) *In situ* molecular architecture of the *Helicobacter pylori* Cag type IV secretion system. *mBio* 10: e00846-19
- Hu J, Worrall LJ, Hong C, Vuckovic M, Atkinson CE, Caveney N, Yu Z & Strynadka NCJ (2018) Cryo-EM analysis of the T3S injectisome reveals the structure of the needle and open secretin. *Nature Communications* 9: 3840
- Huang Q, Alcock F, Kneuper H, Deme JC, Rollauer SE, Lea SM, Berks BC & Palmer T (2017) A signal sequence suppressor mutant that stabilizes an assembled state of the twin arginine translocase. *Proceedings of the National Academy of Sciences* 114: E1958–E1967
- Hunnicut DW, Kempf MJ & McBride MJ (2002) Mutations in *Flavobacterium johnsoniae* *gldF* and *gldG* disrupt gliding motility and interfere with membrane localization of GldA. *Journal of Bacteriology* 184: 2370–2378
- Hunnicut DW & McBride MJ (2000) Cloning and characterization of the *Flavobacterium johnsoniae* gliding- motility genes *gldB* and *gldC*. *Journal of Bacteriology* 182: 911–918
- Hunnicut DW & McBride MJ (2001) Cloning and characterization of the *Flavobacterium johnsoniae* gliding motility genes *gldD* and *gldE*. *Journal of Bacteriology* 183: 4167–4175
- Hutchinson HB & Clayton J (1919) On the decomposition of cellulose by an aerobic organism (*Spirochaeta cytophaga*, n. sp.). *The Journal of Agricultural Science* 9: 143–172
- Jaroszewski L, Li Z, Cai X, Weber C & Godzik A (2011) FFAS server: novel features and applications. *Nucleic acids research* 39: W38–W44
- Jaroszewski L, Rychlewski L, Li Z, Li W & Godzik A (2005) FFAS03: a server for profile--profile sequence alignments. *Nucleic acids research* 33: W284–W288
- Jiang C, Yang D, Hua T, Hua Z, Kong W & Shi Y (2021) A PorX/PorY and σ^P Feedforward Regulatory Loop Controls Gene Expression Essential for *Porphyromonas gingivalis* Virulence. *mSphere* 6: e00428-21

- Johnston JJ, Shrivastava A & McBride MJ (2017) Untangling *Flavobacterium johnsoniae* Gliding Motility and Protein Secretion. *Journal of Bacteriology* 200: e00362-17
- Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* 292: 195–202
- Jones DT, Taylor WR & Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8: 275–282
- Junglas B, Huber ST, Heidler T, Schlösser L, Mann D, Hennig R, Clarke M, Hellmann N, Schneider D & Sachse C (2021) PspA adopts an ESCRT-III-like fold and remodels bacterial membranes. *Cell* 184: 3674–3688
- Jurėnas D & Journet L (2021) Activity, delivery, and diversity of Type VI secretion effectors. *Molecular Microbiology* 115: 383–394
- Kamio Y & Nikaido H (1976) Outer membrane of *Salmonella typhimurium*: accessibility of phospholipid head groups to phospholipase c and cyanogen bromide activated dextran in the external medium. *Biochemistry* 15: 2561–2570
- Kamisetty H, Ovchinnikov S & Baker D (2013) Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences* 110: 15674–15679
- Kano K & Fendler JH (1978) Pyranine as a sensitive pH probe for liposome interiors and surfaces. pH gradients across phospholipid vesicles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 509: 289–299
- Kapitein N, Bönemann G, Pietrosiuk A, Seyffer F, Hausser I, Locker JK & Mogk A (2013) ClpV recycles VipA/VipB tubules and prevents non-productive tubule formation to ensure efficient type VI protein secretion. *Molecular microbiology* 87: 1013–1028
- Katayama E, Tahara YO, Bertin C & Shibata S (2019) Application of spherical substrate to observe bacterial motility machineries by Quick-Freeze- Replica Electron Microscopy. *Scientific Reports* 9: 14765
- Kelley LA, Mezulis S, Yates CM, Wass MN & Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10: 845–858
- Kempf MJ & McBride MJ (2000) Transposon insertions in the *Flavobacterium johnsoniae* *ftsX* gene disrupt gliding motility and cell division. *Journal of bacteriology* 182: 1671–1679
- Kharade SS & McBride MJ (2015) *Flavobacterium johnsoniae* PorV Is required for secretion of a subset of proteins targeted to the type IX secretion system. *Journal of Bacteriology* 197: 147–158
- Kim KK, Kim MK, Lim JH, Park HY & Lee ST (2005) Transfer of *Chryseobacterium meningosepticum* and *Chryseobacterium miricola* to *Elizabethkingia* gen. nov. as *Elizabethkingia meningoseptica* comb. nov. and *Elizabethkingia miricola* comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 55: 1287–1293

- Kim S, Zuromski KL, Bell TA, Sauer RT & Baker TA (2020) ClpAP proteolysis does not require rotation of the ClpAa unfoldase relative to ClpP. *eLife* 9: e6451
- Kim W, Killam T, Sood V & Surette MG (2003) Swarm-Cell Differentiation in *Salmonella enterica* Serovar Typhimurium Results in Elevated Resistance to Multiple Antibiotics. *Journal of Bacteriology* 185: 3111–3117
- Kirchman DL (2002) The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiology Ecology* 39: 91–100
- Kita D, Shibata S, Kikuchi Y, Kokubu E, Nakayama K, Saito A & Ishihara K (2016) Involvement of the Type IX Secretion System in *Capnocytophaga ochracea* Gliding Motility and Biofilm Formation. *Applied and Environmental Microbiology* 82: 1756–1766
- Klausen MS, Jespersen MC, Nielsen H, Jensen KK, Jurtz VI, S nderby CK, Sommer MOA, Winther O, Nielsen M, Petersen B, *et al* (2019) NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins: Structure, Function, and Bioinformatics* 87: 520–527
- Kolton M, Frenkel O, Elad Y & Cytryn E (2014) Potential Role of Flavobacterial Gliding-Motility and Type IX Secretion System Complex in Root Colonization and Plant Defense. *Molecular Plant-Microbe Interactions* 27: 1005–1013
- Krieg NR, Ludwig W, Whitman W, Hedlund BP, Paster BJ, Staley JT, Ward N, Brown D & Parte A eds. (2010) *Bergey’s Manual of Systematic Bacteriology* 2nd ed. New York: Springer-Verlag
- Krissinel E & Henrick K (2007) Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* 372: 774–797
- Krogh A, Larsson B, von Heijne G & Sonnhammer ELL (2001) Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *Journal of Molecular Biology* 305: 567–580
- Kubo S, Niina T & Takada S (2020) Molecular dynamics simulation of proton-transfer coupled rotations in ATP synthase FO motor. *Scientific Reports* 10: 8225
- Kuhlen L, Abrusci P, Johnson S, Gault J, Deme J, Caesar JJE, Dietsche T, Mebrhatu MT, Ganief T, Macek B, *et al* (2018) Structure of the core of the type III secretion system export apparatus. *Nature Structural & Molecular Biology* 25: 583–590
- Kulkarni SS, Johnston JJ, Zhu Y, Hying ZT & McBride MJ (2019) The carboxy-terminal region of *Flavobacterium johnsoniae* SprB facilitates its secretion by the type IX secretion system and propulsion by the gliding motility machinery. *Journal of Bacteriology* 201: e00218-19
- Kulkarni SS, Zhu Y, Brendel CJ & McBride MJ (2017) Diverse C-Terminal Sequences Involved in *Flavobacterium johnsoniae* Protein Secretion. *Journal of Bacteriology* 199: e00884-16
- Kumar S, Stecher G, Li M, Knyaz C & Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35: 1547–1549

- Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, *et al* (2007) Comparative Metagenomics Revealed Commonly Enriched Gene Sets in Human Gut Microbiomes. *DNA Research* 14: 169–181
- Kwong WK, Steele MI & Moran NA (2018) Genome sequences of *Apibacter* spp., gut symbionts of Asian honey bees. *Genome Biology and Evolution* 10: 1174–1179
- Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T & Ben-Tal N (2005) ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research* 33: W299–W302
- Lapidus IR & Berg HC (1982) Gliding motility of *Cytophaga* sp. strain U67. *Journal of Bacteriology* 151: 384–98
- Larsen SH, Reader RW, Kort EN, Tso W-W & Adler J (1974) Change in direction of flagellar rotation is the basis of the chemotactic response in *Escherichia coli*. *Nature* 249: 74–77
- Lasica AM, Goulas T, Mizgalska D, Zhou X, de Diego I, Ksiazek M, Madej M, Guo Y, Guevara T, Nowak M, *et al* (2016) Structural and functional probing of PorZ, an essential bacterial surface component of the type-IX secretion system of human oral-microbiomic *Porphyromonas gingivalis*. *Scientific Reports* 6: 37708
- Lauber CL, Hamady M, Knight R & Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and environmental microbiology* 75: 5111–5120
- Lauber F, Deme JC, Lea SM & Berks BC (2018) Type 9 secretion system structures reveal a new protein transport mechanism. *Nature* 564: 77–82
- van Leewenhoeck A (1677) Observations, communicated to the publisher by Mr. Antony van Leewenhoeck, in a Dutch letter of the 9th Octob. 1676. here English'd: concerning little animals by him observed in rain-well-sea- and snow water; as also in water wherein pepper had lain infus. *Philosophical Transactions of the Royal Society of London* 12: 821–831
- Leiman PG, Basler M, Ramagopal UA, Bonanno JB, Sauder JM, Pukatzki S, Burley SK, Almo SC & Mekalanos JJ (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proceedings of the National Academy of Sciences of the United States of America* 106: 4154–4159
- Leo JC, Grin I & Linke D (2012) Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 1088–1101
- Leone P, Roche J, Vincent MS, Tran QH, Desmyter A, Cascales E, Kellenberger C, Cambillau C & Roussel A (2018) Type IX secretion system PorM and gliding machinery GldM form arches spanning the periplasmic space. *Nature Communications* 9: 429
- Li N, Zhu Y, LaFrentz BR, Evenhuis JP, Hunnicutt DW, Conrad RA, Barbier P, Gullstrand CW, Roets JE, Powers JL, *et al* (2017) The Type IX Secretion System Is Required for Virulence of the Fish Pathogen *Flavobacterium columnare*. *Applied and Environmental Microbiology* 83: e01769-17

- Liao CH & Wells JM (1986) Properties of *Cytophaga johnsonae* strains causing spoilage of fresh produce at food markets. *Applied and environmental microbiology* 52: 1261–1265
- Liebschner D, Afonine P v., Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, *et al* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in *Phenix*. *Acta Crystallographica D* 75: 861–877
- Liu J, McBride MJ & Subramaniam S (2007) Cell surface filaments of the gliding bacterium *Flavobacterium johnsoniae* revealed by cryo-electron tomography. *Journal of Bacteriology* 189: 7503–7506
- Liu J, Tassinari M, Souza DP, Naskar S, Noel JK, Bohuszewicz O, Buck M, Williams TA, Baum B & Low HH (2021) Bacterial Vipp1 and PspA are members of the ancient ESCRT-III membrane-remodeling superfamily. *Cell* 184: 3660–3673
- Locht C, Coutte L & Mielcarek N (2011) The ins and outs of pertussis toxin. *The FEBS journal* 278: 4668–4682
- Lopez KE, Rizo AN, Tse E, Lin J, Scull NW, Thwin AC, Lucius AL, Shorter J & Southworth DR (2020) Conformational plasticity of the ClpAP AAA+ protease couples protein unfolding and proteolysis. *Nature Structural & Molecular Biology* 27: 406–416
- Los G v., Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, *et al* (2008) HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chemical Biology* 3: 373–382
- Low HH, Gubellini F, Rivera-Calzada A, Braun N, Connery S, Dujeancourt A, Lu F, Redzej A, Fronzes R, Orlova E, *et al* (2014) Structure of a type IV secretion system. *Nature* 508: 550–553
- Macnab RM & Koshland DE (1972) The Gradient-Sensing Mechanism in Bacterial Chemotaxis. *Proceedings of the National Academy of Sciences* 69: 2509–2512
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, *et al* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47: W636–W641
- Madej M, Nowakowska Z, Ksiazek M, Lasica AM, Mizgalska D, Nowak M, Jacula A, Bzowska M, Scavenius C, Enghild JJ, *et al* (2021) PorZ, an essential component of the type IX secretion system of *Porphyromonas gingivalis*, delivers anionic lipopolysaccharide to the PorU sortase for transpeptidase processing of T9SS cargo proteins. *mBio* 12: e02262-20
- Mampallil D & Eral HB (2018) A review on suppression and utilization of the coffee-ring effect. *Advances in Colloid and Interface Science* 252: 38–54
- Marathe R, Meel C, Schmidt NC, Dewenter L, Kurre R, Greune L, Schmidt MA, Müller MJ, Lipowsky R, Maier B, *et al* (2014) Bacterial twitching motility is coordinated by a two-dimensional tug-of-war with directional memory. *Nature Communications* 5: 3759
- Martinac B, Rhode PR, Cranfield CG & Nomura T (2013) Patch clamp electrophysiology for the study of bacterial ion channels in giant spheroplasts of *E. coli*. In *Bacterial Cell Surfaces: Methods and Protocols*, Delcour AH (ed) pp 367–380. New York, NY: Springer Science+Business Media

- Mauriello EMF, Mouhamar F, Nan B, Ducret A, Dai D, Zusman DR & Mignot T (2010) Bacterial motility complexes require the actin-like protein, MreB and the Ras homologue, MglA. *EMBO Journal* 29: 315–326
- McBride MJ (2019) *Bacteroidetes* Gliding Motility and the Type IX Secretion System. *Microbiology Spectrum* 7: PSIB-0002-2018
- McBride MJ & Braun TF (2004) GldI is a lipoprotein that is required for *Flavobacterium johnsoniae* gliding motility and chitin utilization. *Journal of bacteriology* 186: 2295–302
- McBride MJ, Braun TF & Brust JL (2003) *Flavobacterium johnsoniae* GldH Is a Lipoprotein that Is Required for Gliding Motility and Chitin Utilization. *Journal of Bacteriology* 185: 6648–6657
- McBride MJ & Kempf MJ (1996) Development of techniques for the genetic manipulation of the gliding bacterium *Cytophaga johnsonae*. *Journal of Bacteriology* 178: 583–590
- McBride MJ & Zhu Y (2013) Gliding motility and Por secretion system genes are widespread among members of the phylum *Bacteroidetes*. *Journal of Bacteriology* 195: 270–278
- Merz AJ, So M & Sheetz MP (2000) Pilus retraction powers bacterial twitching motility. *Nature* 407: 98–102
- Mezanges X, Regeard C, Gerin C, Deroulers C, Grammaticos B & Badoual M (2012) Modeling the role of water in *Bacillus subtilis* colonies. *Physical Review E* 85: 041913
- Miletic S, Fahrenkamp D, Goessweiner-Mohr N, Wald J, Pantel M, Vesper O, Kotov V & Marlovits TC (2021) Substrate-engaged type III secretion system structures reveal gating mechanism for unfolded protein translocation. *Nature Communications* 12: 1546
- Miller JH (1972) Experiments in molecular genetics Cold Spring Harbor, New York: Cold Spring Harbor Laboratory
- Miroux B & Walker JE (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology* 260: 289–298
- Miyata M (2010) Unique Centipede Mechanism of *Mycoplasma* Gliding. *Annual Review of Microbiology* 64: 519–537
- Miyata M & Hamaguchi T (2016) Prospects for the gliding mechanism of *Mycoplasma mobile*. *Current Opinion in Microbiology* 29: 15–21
- Morgan J, Acheson J & Zimmer J (2017) Structure of a Type-1 Secretion System ABC Transporter. *Structure* 25: 522–529
- Naas AE, Solden LM, Norbeck AD, Brewer H, Hagen LH, Heggenes IM, McHardy AC, Mackie RI, Paša-Tolić L, Arntzen MØ, et al (2018) “Candidatus *Paraporphyromonas polyenzymogenes*” encodes multi-modular cellulases linked to the type IX secretion system. *Microbiome* 6: 44
- Naito M, Tominaga T, Shoji M & Nakayama K (2019) PGN_0297 is an essential component of the type IX secretion system (T9SS) in *Porphyromonas gingivalis*: Tn-seq analysis for exhaustive identification of T9SS-related genes. *Microbiology and Immunology* 63: 11–20

- Nakane D, Sato K, Wada H, McBride MJ & Nakayama K (2013) Helical flow of surface protein required for bacterial gliding motility. *Proceedings of the National Academy of Sciences* 110: 11145–50
- Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PMGE, Grigoras IT, Malinauskaite L, Malinauskas T, Miehl J, *et al* (2020) Single-particle cryo-EM at atomic resolution. *Nature* 587: 152–156
- Nakayama K (2015) Porphyromonas gingivalis and related bacteria: from colonial pigmentation to the type IX secretion system and gliding motility. *Journal of Periodontal Research* 50: 1–8
- Narita Y, Sato K, Yukitake H, Shoji M, Daisuke N, Nagano K, Yoshimura F, Naito M & Nakayama K (2014) Lack of a surface layer in *Tannerella forsythia* mutants deficient in the type IX secretion system. *Microbiology* 160: 2295–2303
- Naskar S, Hohl M, Tassinari M & Low HH (2021) The structure and mechanism of the bacterial type II secretion system. *Molecular Microbiology* 115: 412–424
- NCBI Resource Coordinators (2018) Database resources of the National Center for Biotechnology Information. *Nucleic acids research* 46: D8–D13
- Nelson SS, Bollampalli S & McBride MJ (2008) SprB is a cell surface component of the *Flavobacterium johnsoniae* gliding motility machinery. *Journal of Bacteriology* 190: 2851–2857
- Nelson SS, Glocka PP, Agarwal S, Grimm DP & McBride MJ (2007) *Flavobacterium johnsoniae* SprA is a cell surface protein involved in gliding motility. *Journal of Bacteriology* 189: 7145–7150
- Nguyen K-A, Travis J & Potempa J (2007) Does the importance of the C-terminal residues in the maturation of RgpB from *Porphyromonas gingivalis* reveal a novel mechanism for protein export in a subgroup of Gram-negative bacteria? *Journal of Bacteriology* 189: 833–843
- Noji H, Yasuda R, Yoshida M & Kinosita K (1997) Direct observation of the rotation of F1-ATPase. *Nature* 386: 299–302
- Oliveira NM, Foster KR & Durham WM (2016) Single-cell twitching chemotaxis in developing biofilms. *Proceedings of the National Academy of Sciences* 113: 6532–6537
- Ortega J, Lee HS, Maurizi MR & Steven AC (2004) ClpA and ClpX ATPases bind simultaneously to opposite ends of ClpP peptidase to form active hybrid complexes. *Journal of Structural Biology* 146: 217–226
- Pal RR, Baidya AK, Mamou G, Bhattacharya S, Socol Y, Kobi S, Katsowich N, Ben-Yehuda S & Rosenshine I (2019) Pathogenic *E. coli* Extracts Nutrients from Infected Host Cells Utilizing Injectisome Components. *Cell* 177: 683–696
- Palmer T, Finney AJ, Saha CK, Atkinson GC & Sargent F (2021) A holin/peptidoglycan hydrolase-dependent protein secretion system. *Molecular Microbiology* 115: 345–355
- Parker JL, Mindell JA & Newstead S (2014) Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. *eLife* 3: e04273
- Pate JL & Chang LE (1979) Microbiology Evidence That Gliding Motility in Prokaryotic Cells Is Driven by Rotary Assemblies in the Cell Envelopes. *Current Microbiology* 2: 59–64

- Pathirana RD, O'Brien-Simpson NM, Brammar GC, Slakeski N & Reynolds EC (2007) Kgp and RgpB, but Not RgpA, Are Important for *Porphyromonas gingivalis* Virulence in the Murine Periodontitis Model. *Infection and Immunity* 75: 1436–1442
- Pédelacq JD, Cabantous S, Tran T, Terwilliger TC & Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* 24: 79–88
- Peterson JH, Tian P, Ieva R, Dautin N & Bernstein HD (2010) Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment. *Proceedings of the National Academy of Sciences of the United States of America* 107: 17739–17744
- Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH & Ferrin TE (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* 30: 70–82
- Pitta TP, Leadbetter ER & Godchaux W (1989) Increase of ornithine amino lipid content in a sulfonolipid-deficient mutant of *Cytophaga johnsonae*. *Journal of bacteriology* 171: 952–957
- Potempa J, Mydel P & Koziel J (2017) The case for periodontitis in the pathogenesis of rheumatoid arthritis. *Nature Reviews Rheumatology* 13: 606–620
- Punjani A, Rubinstein JL, Fleet DJ & Brubaker MA (2017) CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* 14: 290–296
- Qi Y-J, Jiao Y-L, Chen P, Kong J-Y, Gu B-L, Liu K, Feng D-D, Zhu Y-F, Ruan H-J, Lan Z-J, *et al* (2020) *Porphyromonas gingivalis* promotes progression of esophageal squamous cell cancer via TGFβ-dependent Smad/YAP/TAZ signaling. *PLOS Biology* 18: e3000825
- Rajilić-Stojanović M & de Vos WM (2014) The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiology Reviews* 38: 996–1047
- Rassam P, Copeland NA, Birkholz O, Tóth C, Chavent M, Duncan AL, Cross SJ, Housden NG, Kaminska R, Seger U, *et al* (2015) Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. *Nature* 523: 333–336
- Razin S, Yogev D & Naot Y (1998) Molecular Biology and Pathogenicity of Mycoplasmas. *Microbiology and Molecular Biology Reviews* 62: 1094–1156
- Reboul CF, Kiesewetter S, Eager M, Belousoff M, Cui T, de Sterck H, Elmlund D & Elmlund H (2018) Rapid near-atomic resolution single-particle 3D reconstruction with SIMPLE. *Journal of Structural Biology* 204: 172–181
- Reid S, Leake M, Chandler J, Lo C, Armitage J & Berry R (2006) The maximum number of torque-generating units in the flagellar motor of *Escherichia coli* is at least 11. *Proceedings of the National Academy of Sciences* 103: 8066–8071
- Rhodes RG, Nelson SS, Pochiraju S & McBride MJ (2011a) *Flavobacterium johnsoniae* *sprB* is part of an operon spanning the additional gliding motility genes *sprC*, *sprD*, and *sprF*. *Journal of Bacteriology* 193: 599–610

- Rhodes RG, Samarasam MN, van Groll EJ & McBride MJ (2011b) Mutations in *Flavobacterium johnsoniae sprE* result in defects in gliding motility and protein secretion. *Journal of Bacteriology* 193: 5322–5327
- Rhodes RG, Samarasam MN, Shrivastava A, van Baaren JM, Pochiraju S, Bollampalli S & McBride MJ (2010) *Flavobacterium johnsoniae gldN* and *gldO* are partially redundant genes required for gliding motility and surface localization of SprB. *Journal of Bacteriology* 192: 1201–11
- Ridgway HF (1977) Source of energy for gliding motility in *Flexibacter polymorphus*: effects of metabolic and respiratory inhibitors on gliding movement. *Journal of bacteriology* 131: 544–556
- Ripstein ZA, Vahidi S, Houry WA, Rubinstein JL & Kay LE (2020) A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery. *eLife* 9: e52158
- Rohou A & Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* 192: 216–221
- Rojas ER, Billings G, Odermatt PD, Auer GK, Zhu L, Miguel A, Chang F, Weibel DB, Theriot JA & Huang KC (2018) The outer membrane is an essential load-bearing element in Gram-negative bacteria. *Nature* 559: 617–621
- Russell AB, Wexler AG, Harding BN, Whitney JC, Bohn AJ, Goo YA, Tran BQ, Barry NA, Zheng H, Peterson SB, *et al* (2014) A Type VI Secretion-Related Pathway in *Bacteroidetes* Mediates Interbacterial Antagonism. *Cell Host and Microbe* 16: 227–236
- Saiki K & Konishi K (2007) Identification of a *Porphyromonas gingivalis* novel protein sov required for the secretion of gingipains. *Microbiology and Immunology* 51: 483–491
- Santiveri M, Roa-Eguirara A, Kühne C, Wadhwa N, Hu H, Berg HC, Erhardt M & Taylor NMI (2020) Structure and Function of Stator Units of the Bacterial Flagellar Motor. *Cell* 183: 244–257
- Sato K, Kakuda S, Yukitake H, Kondo Y, Shoji M, Takebe K, Narita Y, Naito M, Nakane D, Abiko Y, *et al* (2018) Immunoglobulin-like domains of the cargo proteins are essential for protein stability during secretion by the type IX secretion system. *Molecular Microbiology* 110: 64–81
- Sato K, Naito M, Yukitake H, Hirakawa H, Shoji M, McBride MJ, Rhodes RG & Nakayama K (2010) A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proceedings of the National Academy of Sciences* 107: 276–81
- Sato K, Okada K, Nakayama K & Imada K (2020) PorM, a core component of bacterial type IX secretion system, forms a dimer with a unique kinked-rod shape. *Biochemical and Biophysical Research Communications* 532: 114–119
- Sato K, Sakai E, Veith P, Shoji M, Kikuchi Y, Yukitake H, Ohara N, Naito M, Okamoto K, Reynolds E, *et al* (2005) Identification of a new membrane-associated protein that influences transport/maturation of gingipains and adhesins of *Porphyromonas gingivalis*. *The Journal of biological chemistry* 280: 8668–8677
- Schneider CA, Rasband WS & Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9: 671–675

- Schneider TD & Stephens RM (1990) Sequence logos: A new way to display consensus sequences. *Nucleic Acids Research* 18: 6097–6100
- Seers C, Slakeski N, Veith P, Nikolof T, Chen Y, Dashper S & Reynolds E (2006) The RgpB C-terminal domain has a role in attachment of RgpB to the outer membrane and belongs to a novel C-terminal-domain family found in *Porphyromonas gingivalis*. *Journal of Bacteriology* 188: 6376–6386
- Segers P, Mannheim W, Vancanneyt M, de Brandt K, Hinz K, Kersters K & Vandamme P (1993) *Riemerella anatipestifer* gen. nov., comb. nov., the causative agent of septicemia anserum exsudativa, and its phylogenetic affiliation within the *Flavobacterium-Cytophaga* rRNA homology group. *International Journal of Systematic Bacteriology* 43: 768–776
- Shah HN & Collins MD (1988) Proposal for Reclassification of *Bacteroides asaccharolyticus*, *Bacteroides gingivalis*, and *Bacteroides endodontalis* in a New Genus, *Porphyromonas*. *International Journal of Systematic and Evolutionary Microbiology* 38: 128–131
- Shi W, Stocker BAD & Adler J (1996) Effect of the surface composition of motile *Escherichia coli* and motile *Salmonella* species on the direction of galvanotaxis. *Journal of Bacteriology* 178: 1113–1119
- Shoji M, Sato K, Yukitake H, Kondo Y, Narita Y, Kadowaki T, Naito M & Nakayama K (2011) Por Secretion System-Dependent Secretion and Glycosylation of *Porphyromonas gingivalis* Hemin-Binding Protein 35. *PLoS ONE* 6: e21372
- Shrivastava A & Berg HC (2020) A molecular rack and pinion actuates a cell-surface adhesin and enables bacterial gliding motility. *Science Advances* 6: eaay6616
- Shrivastava A, Johnston JJ, van Baaren JM & McBride MJ (2013) *Flavobacterium johnsoniae* GldK, GldL, GldM, and SprA are required for secretion of the cell surface gliding motility adhesins SprB and RemA. *Journal of Bacteriology* 195: 3201–3212
- Shrivastava A, Lele PP & Berg HC (2015) A rotary motor drives *Flavobacterium* gliding. *Current Biology* 25: 338–341
- Shrivastava A, Rhodes RG, Pochiraju S, Nakane D & McBride MJ (2012) *Flavobacterium johnsoniae* RemA is a mobile cell surface lectin involved in gliding. *Journal of Bacteriology* 194: 3678–3688
- Simon R, Priefer U & Pühler A (1983) A broad host range mobilization system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria. *Bio/Technology* 1: 784–791
- Simpson AA, Tao Y, Leiman PG, Badasso MO, He Y, Jardine PJ, Olson NH, Morais MC, Grimes S, Anderson DL, et al (2000) Structure of the bacteriophage ϕ 29 DNA packaging motor. *Nature* 408: 745–750
- Sleytr UB, Schuster B, Egelseer E-M & Pum D (2014) S-layers: principles and applications. *FEMS Microbiology Reviews* 38: 823–864
- Son K, Guasto JS & Stocker R (2013) Bacteria can exploit a flagellar buckling instability to change direction. *Nature Physics* 2013 9:8 9: 494–498

- Sonnhammer ELL, von Heijne G & Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology* pp 175–182. AAAI Press
- Stanier RY (1947) Studies on Nonfruiting Myxobacteria: I. *Cytophaga johnsonae*, n.sp., a Chitin-decomposing Myxobacterium. *Journal of bacteriology* 53: 297–315
- Sun C, Benlekbir S, Venkatakrishnan P, Wang Y, Hong S, Hosler J, Tajkhorshid E, Rubinstein JL & Gennis RB (2018) Structure of the alternative complex III in a supercomplex with cytochrome oxidase. *Nature* 557: 123–126
- Sun M, Wartel M, Cascales E, Shaevitz JW & Mignot T (2011) Motor-driven intracellular transport powers bacterial gliding motility. *Proceedings of the National Academy of Sciences* 108: 7559–7564
- Suzek BE, Wang Y, Huang H, McGarvey PB & Wu CH (2015) UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31: 926–932
- Szczepaniak J, Press C & Kleanthous C (2020) The multifarious roles of Tol-Pal in Gram-negative bacteria. *FEMS Microbiology Reviews* 44: 490–506
- Taglialegna A, Lasa I & Valle J (2016) Amyloid Structures as Biofilm Matrix Scaffolds. *Journal of Bacteriology* 198: 2579–88
- Taguchi Y, Sato K, Yukitake H, Inoue T, Nakayama M, Naito M, Kondo Y, Kano K, Hoshino T, Nakayama K, *et al* (2015) Involvement of an Skp-like protein, PGN_0300, in the type IX secretion system of *Porphyromonas gingivalis*. *Infection and Immunity* 84: 230–240
- Tartof KDA & Hobbs CA (1987) Improved media for growing plasmid and cosmid clones. *Bethesda Research Labs Focus* 9: 12
- Tenreiro S, Nobre MF, Rainey FA, Miguel C & da Costa MS (1997) *Thermonema rossianum* sp. nov., a New Thermophilic and Slightly Halophilic Species from Saline Hot Springs in Naples, Italy. *International Journal of Systematic Bacteriology* 47: 122–126
- Thomas F, Hehemann J-H, Rebuffet E, Czejek M & Michel G (2011) Environmental and Gut *Bacteroidetes*: The Food Connection. *Frontiers in Microbiology* 2: 93
- Trinh NTT, Tran HQ, Dong Q van, Cambillau C, Roussel A & Leone P (2020) Crystal structure of Type IX secretion system PorE C-terminal domain from *Porphyromonas gingivalis* in complex with a peptidoglycan fragment. *Scientific Reports* 10: 7384
- Trinh TTN, Gaubert A, Melani P, Cambillau C, Roussel A & Leone P (2021) Crystal structures of two camelid nanobodies raised against GldL, a component of the type IX secretion system from *Flavobacterium johnsoniae*. *Acta Crystallographica* F77: 171–176
- Tully B, Graham E & Heidelberg J (2018) The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Scientific data* 5: 170103
- Vincent MS, Canestrari MJ, Leone P, Stathopoulos J, Ize B, Zoued A, Cambillau C, Kellenberger C, Roussel A & Cascales E (2017) Characterization of the *Porphyromonas gingivalis* Type IX

- Secretion Trans-envelope PorKLMNP Core Complex. *The Journal of biological chemistry* 292: 3252–3261
- Vincent MS, Chabali  r M & Cascales E (2018) A conserved motif of *Porphyromonas* Type IX secretion effectors C-terminal secretion signal specifies interactions with the PorKLMN core complex. *bioRxiv* doi: 10.1101/483123-[PREPRINT]
- Vincent MS, Durand E & Cascales E (2016) The PorX Response Regulator of the *Porphyromonas gingivalis* PorXY Two-Component System Does Not Directly Regulate the Type IX Secretion Genes but Binds the PorL Subunit. *Frontiers in Cellular and Infection Microbiology* 6: 96
- Walker JE (1998) ATP Synthesis by Rotary Catalysis (Nobel lecture). *Angewandte Chemie* 37: 2308–2309
- Wallden K, Rivera-Calzada A & Waksman G (2010) Type IV secretion systems: versatility and diversity in function. *Cellular Microbiology* 12: 1203–1212
- Wang J, Brackmann M, Casta  o-D  ez D, Kudryashev M, Goldie KN, Maier T, Stahlberg H & Basler M (2017) Cryo-EM structure of the extended type VI secretion system sheath–tube complex. *Nature Microbiology* 2: 1507–1512
- Wang J, Hartling JA & Flanagan JM (1997) The structure of ClpP at 2.3   resolution suggests a model for ATP-dependent proteolysis. *Cell* 91: 447–456
- Wang X, Smith D, Jones J & Chapman M (2007) In vitro polymerization of a functional *Escherichia coli* amyloid protein. *The Journal of biological chemistry* 282: 3713–3719
- Ware JC & Dworkin M (1973) Fatty acids of *Myxococcus xanthus*. *Journal of Bacteriology* 115: 253–261
- Wartel M & Mignot T (2014) The mechanism of bacterial gliding motility: Insights from molecular and cellular studies in the *Myxobacteria* and *Bacteroidetes*. In *Nanomicrobiology: Physiological and Environmental Characteristics*, Barton LL Bazylnski D & Xu H (eds) pp 127–143. Springer Science+Business Media New York
- Watson JF & Garc  a-Nafr  a J (2019) In vivo DNA assembly using common laboratory bacteria: A re-emerging tool to simplify molecular cloning. *Journal of Biological Chemistry* 294: 15271–15281
- Wen J, Arakawa T & Philo JS (1996) Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Analytical Biochemistry* 240: 155–166
- Whitman WB, Oren A, Chuvochina M, Costa MS da, Garrity GM, Rainey FA, Rossello-Mora R, Schink B, Sutcliffe I, Trujillo ME, *et al* (2018) Proposal of the suffix –ota to denote phyla. Addendum to ‘Proposal to include the rank of phylum in the International Code of Nomenclature of Prokaryotes.’ *International Journal of Systematic and Evolutionary Microbiology* 68: 967–969
- Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, *et al* (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* 27: 293–315

- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, *et al* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica D* 67: 235–242
- Wyatt PJ (1993) Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta* 272: 1–40
- Xu D, Jaroszewski L, Li Z & Godzik A (2014) FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics* 30: 660–667
- Xu J, Wang D, Gui M & Xiang Y (2019) Structural assembly of the tailed bacteriophage ϕ 29. *Nature Communications* 2019 10:1 10: 2366
- Yang D, Jiang C, Ning B, Kong W & Shi Y (2021) The PorX/PorY system is a virulence factor of *Porphyromonas gingivalis* and mediates the activation of the Type IX secretion system. *Journal of Biological Chemistry* 296: 100574
- Yao X, Fan X & Yan N (2020) Cryo-EM analysis of a membrane protein embedded in the liposome. *Proceedings of the National Academy of Sciences* 117: 18497–18503
- Yip KM, Fischer N, Paknia E, Chari A & Stark H (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature* 587: 157–161
- Zhang J, Zheng JW, Cho BC, Hwang CY, Fang C, He J & Li SP (2012) *Sphingobacterium wenxiniae* sp. nov., a cypermethrin-degrading species from activated sludge. *International Journal of Systematic and Evolutionary Microbiology* 62: 683–687
- Zhu Y & McBride MJ (2014) Deletion of the *Cytophaga hutchinsonii* type IX secretion system gene sprP results in defects in gliding motility and cellulose utilization. *Applied Microbiology and Biotechnology* 98: 763–775
- Zhu Y & McBride MJ (2016) Comparative analysis of *Cellulophaga algicola* and *Flavobacterium johnsoniae* gliding motility. *Journal of Bacteriology* 198: 1743–1754
- Zhu Y, Thomas F, Larocque R, Li N, Duffieux D, Cladière L, Souchaud F, Michel G & McBride MJ (2017) Genetic analyses unravel the crucial role of a horizontally acquired alginate lyase for brown algal biomass degradation by *Zobellia galactanivorans*. *Environmental Microbiology* 19: 2164–2181
- Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN & Alva V (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology* 430: 2237–2243
- Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJH, Lindahl E & Scheres SHW (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* 7: e42166
- Zivanov J, Nakane T & Scheres SHW (2019) A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCr* 6: 5–17
- Zoued A, Brunet YR, Durand E, Aschtgen M-S, Logger L, Douzi B, Journet L, Cambillau C & Cascales E (2014) Architecture and assembly of the Type VI secretion system. *Biochimica et biophysica acta* 1843: 1664–1673

9 Appendix

9.1 Unsuccessful experiments to identify binding partners of *F. johnsoniae*

GldL and GldM

I generated various strains expressing affinity-tagged variants of GldL and GldM to attempt to co-purify binding partners of GldL and GldM. None of these experiments were successful. The strains I used are listed, together with the results of pulldown experiments in TABLE. Plasmids that were used to construct these strains but were not used elsewhere in this thesis are listed in TABLE. Primers used to generate these plasmids are listed in Table 9-3.

Table 9-1 Strains used to identify binding partners of *F. johnsoniae* GldL and GldM.

Strain	Description	Result
Rhj_004	UW101 <i>gldL-twinstrep</i> ; encodes GldL with a C-terminal TwinStrep tag in the endogenous locus.	<i>gldL-TS</i> colonies spread to ~50% the size of WT colonies. GldL-TS copurified with GldM but no other proteins.
Rhj_063	UW101 pRHJ160; encodes a construct equivalent to <i>FjoGldLM'</i> _B with a His ₆ tag replacing the TwinStrep tag under the control of the endogenous promoter on a pCP11-derived plasmid.	Rhj_063 cells showed no differences from WT cells in colony spreading. No expression of GldM' _B -His ₆ was detected by immunoblotting with either αHis ₄ or αGldM antibodies.
Rhj_064	UW101 pRHJ161; encodes GldL and GldM up to residue 235 with a C-terminal His ₆ tag under the control of the endogenous promoter on a pCP11-derived plasmid.	Rhj_064 cells showed no differences from WT cells in colony spreading. No expression of GldM' _C -His ₆ was detected by immunoblotting with either αHis ₄ or αGldM antibodies.
Rhj_065	UW101 pRHJ203; encodes a construct equivalent to <i>FjoGldLM'</i> _B with a His ₆ tag replacing the TwinStrep tag under the control of the <i>remA</i> promoter on a pCP11-derived plasmid.	Rhj_065 cells showed no differences from WT cells in colony spreading. No expression of GldM' _B -His ₆ was detected by immunoblotting with either αHis ₄ or αGldM antibodies.
Rhj_066	UW101 <i>His₆-gldL</i> ; encodes GldL with an N-terminal His ₆ tag in the endogenous locus.	Rhj_066 cells showed no gliding motility as assessed by colony spreading on PY2 agar. No expression of His ₆ -GldL was detected by immunoblotting with either αGldL or αHis ₄ antibodies.
Rhj_067	UW101 <i>His₆-gldL₆₋₂₂₁</i> ; encodes GldL with residues 2-5 replaced with a His ₆ tag in the endogenous locus.	Rhj_067 cells showed no gliding motility as assessed by colony spreading on PY2 agar. No expression of His ₆ -GldL ₆₋₂₂₁ was detected by

		immunoblotting with either α GldL or α His ₄ antibodies.
Rhj_070	UW101 <i>twinstrep-gldL</i> ; encodes GldL with an N-terminal TwinStrep tag in the endogenous locus.	Rhj_070 cells showed no gliding motility as assessed by colony spreading on PY2 agar. No expression of TS-GldL was detected by immunoblotting with either α GldL antibodies or StrepTactin-HRP conjugate (IBA)
Rhj_071	UW101 pRHJ242; encodes the cytoplasmic region of GldL including the cage region with and N-terminal His ₆ tag under the control of the <i>ompA</i> promoter on a pCP11-derived plasmid.	Rhj_071 colonies spread to ~50% the size of WT colonies. A faint band corresponding to His ₆ -GldL ₆₂₋₂₂₁ was detected by immunoblotting with α His ₄ antibodies but no equivalent band was detected using α GldL antibodies. I interpreted this to mean that His ₆ -GldL ₆₂₋₂₂₁ was expressed below the detection level of α GldL antibodies but that this expression was still sufficient to interfere with chromosomally-encoded GldL function.
Rhj_072	UW101 pRHJ243; encodes the cytoplasmic region of GldL excluding the cage region with and N-terminal His ₆ tag under the control of the <i>ompA</i> promoter on a pCP11-derived plasmid.	Rhj_072 cells showed no differences from WT cells in colony spreading. No expression of His ₆ -GldL ₇₂₋₂₂₁ was detected by immunoblotting with either α His ₄ or α GldL antibodies.
Rhj_077	UW101 <i>gldL(31-36 to His₆)</i> ; encodes GldL with six residues of the periplasmic loop replaced with His ₆	Rhj_077 cells showed no gliding motility as assessed by colony spreading on PY2 agar. No expression of GldL _{31-36 to His₆} was detected by immunoblotting with either α GldL or α His ₄ antibodies.
Rhj_078	UW101 <i>gldL₁₋₃₄-His₆-gldL₃₅₋₂₂₁</i> ; encodes GldL with His ₆ inserted between residues 33 and 34 in the periplasmic loop.	Rhj_078 cells showed no gliding motility as assessed by colony spreading on PY2 agar. <i>gldL₁₋₃₄-His₆-gldL₃₅₋₂₂₁</i> was expressed less well than WT GldL in WT cells as assessed by immunoblotting. Membranes were solubilised with DDM and applied to Ni-NTA resin. No differences were observed between the elution fractions for WT and Rhj_078 cells.

Table 9-2 Plasmids used to generate *F. johnsoniae* strains used to identify binding partners of GldL and GldM.

Plasmid	Description	Strategy/source
pCP11	Expression vector for use in <i>F. johnsoniae</i> .	(McBride & Kempf, 1996)
pCP11 <i>P_{ompA}</i> mcherry	Vector to express mCherry in <i>F. johnsoniae</i> under the control of the <i>ompA</i> promoter	Paul Guy (Berks Laboratory)
pFL115	pCP11 <i>P_{remA}</i> <i>fjoh_0978</i>	Dr Frédéric Lauber (Berks Laboratory)
pYT313-GldL-YFP	Suicide vector to add a sequence encoding YFP to the 3' end of <i>gldL</i>	Dr Frédéric Lauber (Berks Laboratory)
pRHJ058	Suicide vector to add a sequence encoding a TwinStrep tag to the 3' end of <i>gldL</i>	Restriction cloning of a BamHI and NcoI fragment of a PCR product amplified from pRHJ007 using primers RHJ172 and RHJ173 into pYT313-GLdL-YFP
pRHJ160	pCP11 <i>gldL gldM(1-232)-His₆</i> , includes 100 bp upstream of codon 1 of <i>gldL</i> which is sufficient for expression (Shrivastava <i>et al</i> , 2013)	Gibson assembly using pCP11 linearised using primers RHJ434 and RHJ448 and a fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ442 and RHJ446
pRHJ161	pCP11 <i>gldL gldM(1-235)-His₆</i> , includes 100 bp upstream of codon 1 of <i>gldL</i> which is sufficient for expression (Shrivastava <i>et al</i> , 2013)	Gibson assembly using pCP11 linearised using primers RHJ434 and RHJ448 and a fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ442 and RHJ447
pRHJ199	pGEM <i>His₆-gldL</i>	Gibson mutagenesis of pRHJ012 using primers RHJ566 and RHJ567
pRHJ200	pGEM <i>His₆-gldL(6-221)</i>	Gibson mutagenesis of pRHJ012 using primers RHJ567 and RHJ568
pRHJ201	Suicide vector used to add a sequence encoding a His ₆ tag to the 5' end of <i>gldL</i>	Gibson assembly using pYT354 linearised using BamHI and Sall and a PCR fragment amplified from pRHJ199 using primers RHJ341 and RHJ342
pRHJ202	Suicide vector used to replace codons 2-5 of <i>gldL</i> with a sequence encoding a His ₆ tag	Gibson assembly using pYT354 linearised using BamHI and Sall and a PCR fragment amplified from pRHJ200 using primers RHJ341 and RHJ342
pRHJ203	pCP11 <i>P_{remA}</i> <i>gldL gldM(1-232)-His₆</i>	Gibson assembly using pFL115 linearised using primers RHJ448 and RHJ572 and a fragment PCR amplified from <i>F. johnsoniae</i> gDNA using primers RHJ569 and RHJ446
pRHJ219	pGEM <i>twinstrep-gldL</i>	KLD mutagenesis of pRHJ012 using primers RHJ573 and RHJ574
pRHJ220	Suicide vector used to add a sequence encoding a TwinStrep tag to the 5' end of <i>gldL</i>	Gibson assembly using pYT354 linearised using BamHI and Sall and a PCR fragment amplified from pRHJ219 using primers RHJ341 and RHJ342

pRHJ242	pCP11 <i>P_{ompA} His₆-gldL(62-221)</i>	Gibson assembly using pCP <i>P_{ompA} mcherry</i> linearised using primers RHJ623 and RHJ624 and a PCR fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ625 and RHJ626
pRHJ243	pCP11 <i>P_{ompA} His₆-gldL(75-221)</i>	Gibson assembly using pCP <i>P_{ompA} mCherry</i> linearised using primers RHJ623 and RHJ624 and a PCR fragment amplified from <i>F. johnsoniae</i> gDNA using primers RHJ625 and RHJ627
pRHJ252	pGEM <i>gldL(31-36 to His₆)</i>	Gibson mutagenesis of pRHJ012 using primers RHJ645 and RHJ646
pRHJ253	pGEM <i>gldL₁₋₃₄-His₆-gldL₃₅₋₂₂₁</i>	Gibson mutagenesis of pRHJ012 using primers RHJ647 and RHJ648
pRHJ255	Suicide vector to replace codons 31 to 36 of <i>gldL</i> with a sequence encoding a His ₆ tag	Gibson assembly using pYT354 linearised using BamHI and Sall and a PCR fragment amplified from pRHJ252 using primers RHJ341 and RHJ342
pRHJ256	Suicide vector to insert a sequence encoding a His ₆ tag between codons 33 and 34 of <i>gldL</i>	Gibson assembly using pYT354 linearised using BamHI and Sall and a PCR fragment amplified from pRHJ253 using primers RHJ341 and RHJ342

9.2 Oligonucleotides used in this study

Table 9-3 Oligonucleotides used in this study.

Oligonucleotide	Sequence	Used for:
RHJ001	actgc <u>ccatgg</u> gtatggcattattaagtaaaaaag	Construction of pRHJ001, restriction site is underlined
RHJ002	cgtggc <u>ggatc</u> ctctttgttactcattgcag	Construction of pRHJ001, restriction site is underlined
RHJ003	actgc <u>ccatgg</u> gaatggcaggaggaaaattaaccc	Construction of pRHJ105, restriction site is underlined
RHJ004	gcaccg <u>ggatc</u> cttgtatttcgtaaattaccggagcag	Construction of pRHJ105, restriction site is underlined
RHJ009	gacttg <u>ccatgg</u> gtatgagcgcaagaaaaagttaaac	Construction of pRHJ107, restriction site is underlined
RHJ010	gaatctgga <u>tc</u> ctgagcagcacctttc	Construction of pRHJ107
RHJ011	gaataa <u>ccatgg</u> gtatggccggaggagg	Construction of pRHJ101, restriction site is underlined
RHJ012	gcatcg <u>ggatc</u> ctcctgtatcacaaactc	Construction of pRHJ101, restriction site is underlined
RHJ013	gcgcg <u>ccat</u> atgatggcattattaagtaaaaaag	Construction of pRHJ106, restriction site is underlined
RHJ014	gactahgga <u>tc</u> ctcttgttactcattgcag	Construction of pRHJ106, restriction site is underlined
RHJ015	gcacgt <u>cat</u> atgatggcaggaggaaaattaacc	Construction of pRHJ109, restriction site is underlined
RHJ016	cgtgacgga <u>tc</u> cttgtatttcgtaaattaccggagc	Construction of pRHJ109, restriction site is underlined
RHJ021	cgggcg <u>cat</u> atgatgagcgcaagaaaaag	Construction of pRHJ103, restriction site is underlined
RHJ023	cattga <u>cat</u> atgatggccggaggaggaaaag	Construction of pRHJ108, restriction site is underlined
RHJ025	<u>tgatcgggaggttcagcgtggagccaccgcagttcgagaaataact</u> cgagtctggtaaagaaacc	Construction of pRHJ002, inserted sequence is underlined
RHJ026	<u>ccgccagaacctccaccttttcgaactcgggtggctccaagcgtggt</u> accgacgtcagcgat	Construction of pRHJ002, inserted sequence is underlined
RHJ044	aaaaaagcga <u>tc</u> gcctgagcagcacctttc	Construction of pRHJ103, restriction site is underlined
RHJ045	aaaaaagcga <u>tc</u> gcctctgtatcacaaactcaaatac	Construction of pRHJ108, restriction site is underlined
RHJ046	ccagagtcggccaaggaaccggcagtt	Construction of pRHJ003, altered bases are underlined
RHJ047	ccagagtcggccaaggaaccggcagtt	Construction of pRHJ003, altered bases are underlined
RHJ051	gcgcg <u>ccatgg</u> gtcattatagaagatacaag	Construction of pRHJ005, restriction site is underlined
RHJ052	aaaaaagga <u>tc</u> taagggtgagctgcc	Construction of pRHJ005, restriction site is underlined
RHJ053	gcgcg <u>ccat</u> atgggtcattatagaagatacaag	Construction of pRHJ110, restriction site is underlined

RHJ054	aaaaaaggtacctaagggtgagctgcc	Construction of pRHJ110, restriction site is underlined
RHJ055	tataagccatggcagtagg	Construction of pRHJ102, restriction site is underlined
RHJ056	aaccaaggatccgttcacaattacttcaatggc	Construction of pRHJ102, restriction site is underlined
RHJ057	gcacgtcatatggcagtaggttctaattgg	Construction of pRHJ006, restriction site is underlined
RHJ058	aaccaaggtaccgttcacaattacttcaatggc	Construction of pRHJ006, restriction site is underlined
RHJ073	tgcttcggcgatggcgctta	Construction of pRHJ113, altered bases are underlined
RHJ074	tacctcatatactttattcagggc	Construction of pRHJ113, altered bases are underlined
RHJ075	ctgcaaagcacatggagagtc	Construction of pRHJ113, altered bases are underlined
RHJ076	caagagtgatttgctcag	Construction of pRHJ113, altered bases are underlined
RHJ077	gccaacctcttgatccttcattg	Construction of pRHJ113, altered bases are underlined
RHJ078	cattgatgtagattttctggttttg	Construction of pRHJ113, altered bases are underlined
RHJ110	aaaaaaggatcctccgctggaagaagaggaac	Construction of pRHJ011, restriction site is underlined
RHJ113	aaaaaaggtaccacaagcctcctgcaattctcgaag	Construction of pRHJ011, restriction site is underlined
RHJ115	tagaaggtcgacctccagcgatagaacaatagc	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ116	aaaaaaggatccggctgagcctgtatttcgag	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ117	aaaaaagtcgaccataactttttacttaataatgccat	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ118	gatgagtattatgctgaagtaatgg	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ119	ggatataccgaaaaaatcgc	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ144	gtcgaccatatgggagagctcc	Linearisation of pGEM-T
RHJ145	gcatgcgacgtcgggcc	Linearisation of pGEM-T
RHJ146	cccgacgtcgcacgctgagcctgtatttcgagatg	Construction of pRHJ012, the region overlapping the vector is underlined
RHJ147	tctcccatatggtcgacacaagcctcctgcaattctcg	Construction of pRHJ013, the region overlapping the vector is underlined
RHJ148	cccgacgtcgcacgctcctggaagaagaggaacttg	Construction of pRHJ013, the region overlapping the vector is underlined
RHJ149	tctcccatatggtcgaccagtacttcacgtgcatttgggaag	Construction of pRHJ012, the region overlapping the vector is underlined
RHJ154	ctcaagtaacaggaggcgg	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ155	aagaaatgtctatggcagctgc	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ156	gcttcagcagcttttgatc	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ157	agtcgatctgagcaggag	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ160	agttcttggtctgtctagggttaattttcctcctgc	Construction of pRHJ010

RHJ161	cctagacagaccaagaactgctccggaatt	Construction of pRHJ010
RHJ162	gaaaacctgtacttccagg	Linearisation of pT12
RHJ163	ggtgaattcctcctgaatttc	Linearisation of pT12
RHJ164	<u>aaattcaggaggaattcaccatggcattattaagtaaaaaagtatgaa</u> ttt	pT12 vectors with <i>F. johnsoniae</i> genes, the region overlapping the vector is underlined
RHJ165	<u>cctggaagtacaggttttcttacctaatagcactgttgtaaac</u>	Construction of pRHJ007, the region overlapping the vector is underlined
RHJ166	<u>aaattcaggaggaattcaccatggcattattaagtaaaaaagtatgaa</u> ttt	Construction of pRHJ008, the region overlapping the vector is underlined
RHJ167	<u>cctggaagtacaggttttcagaagcagcagcaact</u>	Construction of pRHJ008, the region overlapping the vector is underlined
RHJ168	<u>aaattcaggaggaattcaccatgggcattatagaagatacaagaac</u>	pT12 vectors with <i>P. gingivalis</i> genes, the region overlapping the vector is underlined
RHJ169	<u>cctggaagtacaggttttcttaccagatcggaagt</u>	Construction of pRHJ009, the region overlapping the vector is underlined
RHJ170	<u>aaattcaggaggaattcaccatgagcgcaagaaaaagt</u>	pT12 vectors with <i>T. rossianum</i> genes, the region overlapping the vector is underlined
RHJ172	aaaaaaggatccgaaaacctgtacttccagg	Construction of pRHJ058, restriction site is underlined
RHJ173	aaaaaaccatgggtattttcgaactgcgg	Construction of pRHJ058, restriction site is underlined
RHJ174	cggctgtggaatgcagg	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ175	cagcacttccaatttcttcg	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ176	ggatataccgaaaaaatcgc	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ177	<u>aaattcaggaggaattcaccatgccactattgatgtaaagc</u>	pT12 vectors with <i>S. thermophila</i> genes, the region overlapping the vector is underlined
RHJ178	ttatatgtttacttgctaacgccat	pT12 vectors with <i>S. thermophila</i> genes
RHJ179	gcgttagcaagtaaacatataaacattaaatcagaataatttaaatg ct	pT12 vectors with <i>S. thermophila</i> genes
RHJ202	<u>cctggaagtacaggttttctactgttcacacgataatccc</u>	Construction of pRHJ118, the region overlapping the vector is underlined
RHJ203	<u>cctggaagtacaggttttcatcggtaaactttaatcggtttt</u>	Construction of pRHJ117, the region overlapping the vector is underlined
RHJ204	<u>cctggaagtacaggttttcgacgtcgaacttcaagg</u>	Construction of pRHJ119, the region overlapping the vector is underlined
RHJ205	<u>cctggaagtacaggttttcaggaactacaacatattttccagag</u>	Construction of pRHJ122, the region overlapping the vector is underlined
RHJ206	<u>cctggaagtacaggttttctggtccgtaacgaagtattcg</u>	Construction of pRHJ123, the region overlapping the vector is underlined
RHJ207	<u>cctggaagtacaggttttctggagctactgtatattttgtgtgg</u>	Construction of pRHJ125, the region overlapping the vector is underlined

RHJ208	<u>cctggaagtacagg</u> ttttcggggcgtaacc	Construction of pRHJ124, the region overlapping the vector is underlined
RHJ269	acctgatggctctggtttcatcgcaatg	Construction of pRHJ031, the altered bases are underlined
RHJ270	aaaccagag <u>cc</u> atcaggtaatcatcttctg	Construction of pRHJ031, the altered bases are underlined
RHJ271	tagcaatggccgatcaaaagaagtatttctgc	Construction of pRHJ033, the altered bases are underlined
RHJ272	ttgatacggccattgctaaccattgcatg	Construction of pRHJ033, the altered bases are underlined
RHJ273	ctttattc <u>ca</u> attactcactttgagattggg	Construction of pRHJ017, the altered bases are underlined
RHJ274	tgagtaatt <u>gc</u> gaataaagctccaacgattacc	Construction of pRHJ017, the altered bases are underlined
RHJ275	aaattactg <u>cc</u> tttgagattggccattaac	Construction of pRHJ018, the altered bases are underlined
RHJ276	atctcaaagg <u>c</u> agtaatttgaataaagctccaac	Construction of pRHJ018, the altered bases are underlined
RHJ277	ttttgactg <u>cg</u> gcattaatctttgcgttatctg	Construction of pRHJ021, the altered bases are underlined
RHJ278	attaatg <u>ccg</u> cagtcaaaagaccgatagaaag	Construction of pRHJ021, the altered bases are underlined
RHJ279	gtcttttgctgaggcattaatctttgcg	Construction of pRHJ020, the altered bases are underlined
RHJ280	tgctcagc <u>ca</u> aaagaccgatagaaagcatc	Construction of pRHJ020, the altered bases are underlined
RHJ281	gtctacacgaacccttgg	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ296	<u>gaagtacagg</u> ttttctgtatttcgtaaattaccggagc	Construction of pRHJ184, the region overlapping the vector is underlined
RHJ297	<u>gaagtacagg</u> ttttctgtcacaattacttcaatggc	Construction of pRHJ185, the region overlapping the vector is underlined
RHJ298	<u>gaagtacagg</u> ttttctgtaatatctatggagatcggact	Construction of pRHJ186, the region overlapping the vector is underlined
RHJ299	<u>gaagtacagg</u> ttttctcctgctatcacaacactcaa	Construction of pRHJ187, the region overlapping the vector is underlined
RHJ300	aggatcggacaacaccggtg	Binds near the 5' end of <i>sfGFP</i> , used for sequencing
RHJ308	agatgattg <u>cc</u> ctgatgtatctggttttcac	Construction of pRHJ030, the altered bases are underlined
RHJ309	acatcagggcaatcatcttctgtctagggg	Construction of pRHJ030, the altered bases are underlined
RHJ310	aagttatggcttctgcttatggtatgggag	Construction of pRHJ014, the altered bases are underlined
RHJ311	aagcgaaagccataactttttacttaataatgccataataatc	Construction of pRHJ014, the altered bases are underlined
RHJ312	atttcgctgctggtatgggagcggc	Construction of pRHJ015, the altered bases are underlined
RHJ313	ccataccagcagcgaaattcataactttttacttaataatg	Construction of pRHJ015, the altered bases are underlined
RHJ314	tgatgcttctatcggtctttgactgagg	Construction of pRHJ019, the altered bases are underlined
RHJ315	gaccgatag <u>ca</u> agcatcactgttcctg	Construction of pRHJ019, the altered bases are underlined
RHJ316	ttgcgttagctgcttttgaaccagttgag	Construction of pRHJ024, the altered bases are underlined
RHJ317	aaaagcag <u>ct</u> aacgcaaagattaatgcctc	Construction of pRHJ024, the altered bases are underlined

RHJ318	cctgatgttctggtttcatcgcaatg	Construction of pRHJ032, the altered bases are underlined
RHJ319	aaaccagaacatcagggttaacatctcttg	Construction of pRHJ032, the altered bases are underlined
RHJ320	cccctagagcgaagatgattaacctgatgtatctg	Construction of pRHJ028, the altered bases are underlined
RHJ321	tcattctgcctctaggggttaatttctcctg	Construction of pRHJ028, the altered bases are underlined
RHJ322	tttgactcaggcattaatctttgcgttatc	Construction of pRHJ022 and pRHJ155, the altered bases are underlined
RHJ323	taatgcctgagtcaaaagaccgatagaaagc	Construction of pRHJ022 and pRHJ155, the altered bases are underlined
RHJ341	cggccgctctagaactagtggatccgctgagcctgtatttcgag	pYT354 suicide vectors to introduce mutations in <i>F. johnsoniae gldL</i> , the region overlapping the vector is underlined
RHJ342	taccgggccccctcgaggctgaccagtacttcacg	pYT354 suicide vectors to introduce mutations in <i>F. johnsoniae gldL</i> , the region overlapping the vector is underlined
RHJ343	cggccgctctagaactagtggatcccgtggaagaagagg	pYT354 suicide vectors to introduce mutations in <i>F. johnsoniae gldM</i> , the region overlapping the vector is underlined
RHJ344	taccgggccccctcgaggctgaccaagcctcctgaattc	pYT354 suicide vectors to introduce mutations in <i>F. johnsoniae gldM</i> , the region overlapping the vector is underlined
RHJ345	cctgacttaccacagccg	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ348	tttcgcttttggtatgggagcggc	Construction of pRHJ016, the altered bases are underlined
RHJ349	ccataccaagcgaaattcataactttttacttaataatg	Construction of pRHJ016, the altered bases are underlined
RHJ350	ttgactgatgcattaatctttgcgttatctg	Construction of pRHJ023, the altered bases are underlined
RHJ351	attaatgcctcagtcacaaagaccgatagaaag	Construction of pRHJ023, the altered bases are underlined
RHJ352	taaccctgcacagaagatgattaacctgatgtatc	Construction of pRHJ025, the altered bases are underlined
RHJ353	tcttctgtcagggggttaatttctcctg	Construction of pRHJ025, the altered bases are underlined
RHJ354	aaccctaaacagaagatgattaacctgatg	Construction of pRHJ026, the altered bases are underlined
RHJ355	tcttctgttaggggttaatttctcctg	Construction of pRHJ026, the altered bases are underlined
RHJ356	taaccctgaacagaagatgattaacctgatg	Construction of pRHJ027, the altered bases are underlined
RHJ357	tcttctgttaggggttaatttctcctg	Construction of pRHJ027, the altered bases are underlined
RHJ358	ctagacagcgatgattaacctgatgtatctggtttc	Construction of pRHJ029, the altered bases are underlined
RHJ359	ttaatcatcgctgtctaggggttaatttcc	Construction of pRHJ029, the altered bases are underlined
RHJ423	gagaacgccttcaggtc	Sequencing of vectors to genetically modify <i>F. johnsoniae</i>
RHJ428	tttggtacgcaagctattatcttcttactactgc	Construction of pRHJ158, the altered bases are underlined
RHJ429	ataatagcttgcgtacccaaaccggc	Construction of pRHJ158, the altered bases are underlined

RHJ430	gatgatcacaca <u>attcttgggtgtttttatctccg</u>	Construction of pRHJ156, the altered bases are underlined
RHJ431	accaagaattgtgtgatcatccccacgaaaag	Construction of pRHJ156, the altered bases are underlined
RHJ432	tttatctacacaggcagtgatctttttctttc	Construction of pRHJ157, the altered bases are underlined
RHJ433	atcactgcctgtgtagataaaccaataattagcatcaag	Construction of pRHJ157, the altered bases are underlined
RHJ434	<u>cttgagttccccgaattttccgctg</u>	Construction of pRHJ160 and pRHJ161, the region overlapping the other fragment is underlined
RHJ442	<u>aaaattcgggggaactcaagtaacaggag</u>	Construction of pRHJ160 and pRHJ161, the region overlapping the other fragment is underlined
RHJ446	<u>ttagtggtgatggtgatgatgagaagcagcagcaac</u>	Construction of pRHJ160, pRHJ161, and pRHJ203 the region overlapping the other fragment is underlined
RHJ447	<u>ttagtggtgatggtgatggtgtgctgtaagaagcagc</u>	Construction of pRHJ161, the region overlapping the other fragment is underlined
RHJ448	<u>catcatcaccatcaccactaaaattcagcaatagaaacctcc</u>	Construction of pRHJ160, pRHJ161, and pRHJ203 the region overlapping the other fragment is underlined
RHJ449	<u>aaattcaggaggaattcaccatggcaatgaaaacaaattcaaa</u>	pT12 vectors with <i>A. mensalis</i> genes, the region overlapping the vector is underlined
RHJ450	<u>ccctggaagtacaggttttctgaacattataagaataggtgaaatact</u> tt	Construction of pRHJ188, the region overlapping the vector is underlined
RHJ451	<u>ccctggaagtacaggttttcggctttaattcaattcatcttcaag</u>	Construction of pRHJ162, the region overlapping the vector is underlined
RHJ452	<u>aaattcaggaggaattcaccatggcacaatcaacaaaac</u>	pT12 vectors with <i>C. canimorsus</i> genes, the region overlapping the vector is underlined
RHJ453	<u>ccctggaagtacaggttttcgtttgttatttcaatagatatcggag</u>	Construction of pRHJ189, the region overlapping the vector is underlined
RHJ454	<u>ccctggaagtacaggttttcatttgcacgaaatttgagac</u>	Construction of pRHJ163, the region overlapping the vector is underlined
RHJ455	<u>aaattcaggaggaattcaccatgcgttcgctctac</u>	pT12 vectors with <i>D. orientale</i> genes using the <i>gldL1</i> start codon, the region overlapping the vector is underlined
RHJ456	<u>aaattcaggaggaattcaccatgaatctgggagaactttttaa</u>	pT12 vectors with <i>D. orientale</i> genes using the <i>gldL2</i> start codon, the region overlapping the vector is underlined
RHJ457	<u>ccctggaagtacaggttttcttttatttgaatgatattgggtcgag</u>	Construction of pRHJ190 and pRHJ191, the region overlapping the vector is underlined
RHJ458	<u>ccctggaagtacaggttttcacggtaaatttaaatgagcc</u>	Construction of pRHJ164 and pRHJ165, the region overlapping the vector is underlined
RHJ459	<u>aaattcaggaggaattcaccatggcaattttaagcaaaaaaac</u>	pT12 vectors with <i>F. psychrophilum</i> genes, the region overlapping the vector is underlined

		underlined
RHJ460	<u>ccctggaagtacaggtttt</u> ctgtacttcgtatgtacaagttga	Construction of pRHJ192, the region overlapping the vector is underlined
RHJ461	<u>ccctggaagtacaggtttt</u> cattactataagaagcagccg	Construction of pRHJ166, the region overlapping the vector is underlined
RHJ462	<u>aaattcaggaggaattcaccat</u> gagttttgtacagaacattaac	pT12 vectors with NAT55 387147 genes, the region overlapping the vector is underlined
RHJ463	<u>ccctggaagtacaggtttt</u> ctgtaattcaagttctattactgtatttctat ctaadc	Construction of pRHJ193, the region overlapping the vector is underlined
RHJ464	<u>ccctggaagtacaggtttt</u> caccatcaaacttaataagtttctgag	Construction of pRHJ167, the region overlapping the vector is underlined
RHJ465	<u>aaattcaggaggaattcaccat</u> ggctggaagaaaaaatt	pT12 vectors with <i>P. saltans</i> genes, the region overlapping the vector is underlined
RHJ466	<u>ccctggaagtacaggtttt</u> ctcttctatttgatactcacagg	Construction of pRHJ194, the region overlapping the vector is underlined
RHJ467	<u>ccctggaagtacaggtttt</u> ctgcatccaaagtaactaatgcttc	Construction of pRHJ168, the region overlapping the vector is underlined
RHJ468	<u>aaattcaggaggaattcaccat</u> gagactaaatgataaaacattaaatt	pT12 vectors with <i>R. anatipestifer</i> genes, the region overlapping the vector is underlined
RHJ469	<u>ccctggaagtacaggtttt</u> ctgaacattatcactactggag	Construction of pRHJ195, the region overlapping the vector is underlined
RHJ470	<u>ccctggaagtacaggtttt</u> ccgcatcaaacttaatatccgc	Construction of pRHJ169, the region overlapping the vector is underlined
RHJ471	<u>aaattcaggaggaattcaccat</u> ggcaaagaaaacaaaattcaa	pT12 vectors with <i>S. wenxiniae</i> genes, the region overlapping the vector is underlined
RHJ472	<u>ccctggaagtacaggtttt</u> cattcgtcaccgggaac	Construction of pRHJ196, the region overlapping the vector is underlined
RHJ473	<u>ccctggaagtacaggtttt</u> ctgggtctaagttactactgcc	Construction of pRHJ170, the region overlapping the vector is underlined
RHJ474	<u>aaattcaggaggaattcaccat</u> ggcaaagaacatccat	pT12 vectors with UEG genes, the region overlapping the vector is underlined
RHJ475	<u>ccctggaagtacaggtttt</u> ctttactgaataagaaactgggtgc	Construction of pRHJ197, the region overlapping the vector is underlined
RHJ476	<u>ccctggaagtacaggtttt</u> cgaaaacctttgtgtgtacttg	Construction of pRHJ171, the region overlapping the vector is underlined
RHJ477	<u>aaattcaggaggaattcaccat</u> ggcacagtcaaatca	pT12 vectors with <i>Z. galactanivorans</i> genes, the region overlapping the vector is underlined
RHJ478	<u>ccctggaagtacaggtttt</u> cgttgttatctccactacaacc	Construction of pRHJ198, the region overlapping the vector is underlined
RHJ479	<u>ccctggaagtacaggtttt</u> cattcttaagtatacttggtg	Construction of pRHJ172, the region overlapping the vector is underlined
RHJ480	<u>ccctggaagtacaggtttt</u> ctggtaaatctttacaacatcttgaac	Construction of pRHJ173, the region overlapping the vector is underlined
RHJ481	<u>ccctggaagtacaggtttt</u> caggttttaggaattgttgaataacttg	Construction of pRHJ174, the region overlapping the vector is underlined
RHJ482	<u>ccctggaagtacaggtttt</u> ccggcgtttgtattttattaac	Construction of pRHJ175 and pRHJ176, the region overlapping the vector is underlined

RHJ483	<u>ccctggaagtacagggttttcaggcacaacaacataattacc</u>	Construction of pRHJ177, the region overlapping the vector is underlined
RHJ484	<u>ccctggaagtacagggttttcaggggccactctatagg</u>	Construction of pRHJ178, the region overlapping the vector is underlined
RHJ485	<u>ccctggaagtacagggttttcaggctctgtaccgtataac</u>	Construction of pRHJ179, the region overlapping the vector is underlined
RHJ486	<u>ccctggaagtacagggttttcctgcaatgctacttccttag</u>	Construction of pRHJ180, the region overlapping the vector is underlined
RHJ487	<u>ccctggaagtacagggttttcctggtttagccacttgataagtcac</u>	Construction of pRHJ181, the region overlapping the vector is underlined
RHJ488	<u>ccctggaagtacagggttttcaggaaaccagaaattccc</u>	Construction of pRHJ182, the region overlapping the vector is underlined
RHJ489	<u>ccctggaagtacagggttttcaggcttgctaatagtagc</u>	Construction of pRHJ183, the region overlapping the vector is underlined
RHJ547	tttgatgttc <u>gca</u> attttacacctcaaagggtgg	Construction of pRHJ225, the altered bases are underlined
RHJ548	gtgtaaaatt <u>gcga</u> acatcaaaccgataataacaac	Construction of pRHJ225, the altered bases are underlined
RHJ549	ggcatccttttttagtgctactaggcttg	Construction of pRHJ226, the altered bases are underlined
RHJ550	tagcactaaa <u>aaagg</u> atgcctatcatcc	Construction of pRHJ226, the altered bases are underlined
RHJ566	at <u>gcatcatcaccatcaccac</u> gcattattaagtaaaaaagttatgaatttcg	Construction of pRHJ199, the altered bases are underlined
RHJ567	<u>gtggtgatggtgatgatg</u> cataataattcttttttaggttttaattttaaatagattc	Construction of pRHJ199 and pRHJ200, the altered bases are underlined
RHJ568	at <u>gcatcatcaccatcaccaca</u> aaaaagttatgaatttcgcttatgg	Construction of pRHJ199, the altered bases are underlined
RHJ569	cacgcctactatggcattattaagtaaaaaagttatgaat	Construction of pRHJ203
RHJ572	ataatgccatagtaggcgtggtattaaa	Construction of pRHJ203
RHJ573	<u>ggatcgggaggtggatcgtggagccacccgcagttcg</u> aaaaagcattattaaagtaaaaaagttatgaatttc	Construction of pRHJ219, the inserted bases are underlined
RHJ574	<u>acctccggaacctccacctttctcgaactcg</u> gggtggctccacataataattcttttttaggttttaattttaaatag	Construction of pRHJ219, the inserted bases are underlined
RHJ579	<u>ccctggaagtacagggttttcgcgcttaggtacattcttaattc</u>	Construction of pRHJ204, the region overlapping the vector is underlined
RHJ580	<u>ccctggaagtacagggttttcagcacgaggaatatcctttatac</u>	Construction of pRHJ205, the region overlapping the vector is underlined
RHJ581	<u>ccctggaagtacagggttttcgggatcg</u> ggaacacg	Construction of pRHJ206 and pRHJ207, the region overlapping the vector is underlined
RHJ582	<u>gaagtacagggttttctggaccaggaatacctttaattc</u>	Construction of pRHJ208, the region overlapping the vector is underlined
RHJ583	<u>ccctggaagtacagggttttctggagcaggcaaacc</u>	Construction of pRHJ209, the region overlapping the vector is underlined
RHJ584	<u>ccctggaagtacagggttttctggatcaggtgttctttaac</u>	Construction of pRHJ210, the region overlapping the vector is underlined
RHJ585	<u>gaagtacagggttttcaggatcggggagagc</u>	Construction of pRHJ211, the region overlapping the vector is underlined
RHJ586	<u>ccctggaagtacagggttttcggaggaggaacctc</u>	Construction of pRHJ212, the region overlapping the vector is underlined

RHJ587	<u>ccctggaagtacaggttttccggagcaggaacac</u>	Construction of pRHJ213, the region overlapping the vector is underlined
RHJ588	<u>gaagtacaggttttctgcttgaggcaaacc</u>	Construction of pRHJ214, the region overlapping the vector is underlined
RHJ589	<u>ccctggaagtacaggttttcaggtttaggtatagctttcac</u>	Construction of pRHJ215, the region overlapping the vector is underlined
RHJ591	<u>ccctggaagtacaggttttcagggtggcgaaaatcac</u>	Construction of pRHJ218, the region overlapping the vector is underlined
RHJ592	<u>ccctggaagtacaggttttcaggcttggaatatccttaatcc</u>	Construction of pRHJ219, the region overlapping the vector is underlined
RHJ595	caaaattttag <u>ccctcaaagggtggtgaatgg</u>	Construction of pRHJ227, the altered bases are underlined
RHJ596	acctttgagg <u>gctaaaaatttgaacatcaaaccgataataac</u>	Construction of pRHJ227, the altered bases are underlined
RHJ597	ttggctgttgcagccctgcttttctttattatg	Construction of pRHJ228, the altered bases are underlined
RHJ598	aagcagggtcgaacagccaaacctacac	Construction of pRHJ228, the altered bases are underlined
RHJ599	tggctgttgatgccctgcttttctttattatg	Construction of pRHJ229, the altered bases are underlined
RHJ600	aaagcagggc <u>atcaacagccaaacctacac</u>	Construction of pRHJ229, the altered bases are underlined
RHJ601	tttgctgttcaagccctgcttttctttattatg	Construction of pRHJ230, the altered bases are underlined
RHJ602	agcagggttgaacagccaaacctacac	Construction of pRHJ230, the altered bases are underlined
RHJ603	gagacaccaa <u>aacagcgatgataggc</u>	Construction of pRHJ231, the altered bases are underlined
RHJ604	catccgctgttttgggtgtcttttttctctg	Construction of pRHJ231, the altered bases are underlined
RHJ607	gttgatgaataatgtattttaatctttaaggtagtatttgag	Construction of pRHJ233
RHJ608	aaatacattattcatccaactccgataaac	Construction of pRHJ233
RHJ617	<u>tctgggtcttcaagtggtcctcaggagctaacggacaagctag</u>	Construction of pRHJ237, the altered bases are underlined
RHJ618	<u>gagccacttgaagaccagatgatccatcctcaactggttcaaaag</u>	Construction of pRHJ237, the altered bases are underlined
RHJ621	agttgaggatgctaacggacaagctag	Construction of pRHJ238
RHJ622	gtccgttagcatcctcaactggttcaaaag	Construction of pRHJ238
RHJ623	taagagctcgtcgacc	Construction of pRHJ242 and pRHJ243
RHJ624	catacttaattttttaattacaatttagttaattacaag	Construction of pRHJ242 and pRHJ243
RHJ265	cccggtcgacgagctcttattatcctttgttactcattgcag	Construction of pRHJ242 and pRHJ243
RHJ626	aattaaaaaattaagtatg <u>catcatcaccatcaccacgaggatgaatt</u> agattggactc	Construction of pRHJ242, the inserted bases are underlined
RHJ627	aattaaaaaattaagtatg <u>catcatcaccatcaccacgctaacggaca</u> agctag	Construction of pRHJ243, the inserted bases are underlined
RHJ645	<u>ccatcaccatcaccatcacacagggacagtgatgc</u>	Construction of pRHJ252, the inserted bases are underlined
RHJ646	<u>tgtgatggtgatggtgatgtagtaattttgaataaagctcc</u>	Construction of pRHJ252, the inserted bases are underlined

RHJ647	<u>tc</u> atcaccatcaccatcacgggccattaacaggac	Construction of pRHJ253, the inserted bases are underlined
RHJ648	cgtgatggtgatggtgatgaatctcaaagtgagtaattttgaataaag	Construction of pRHJ253, the inserted bases are underlined