STRUCTURAL AND FUNCTIONAL STUDIES OF ALGK: A PROTEIN REQUIRED FOR THE SECRETION OF HIGH-MOLECULAR WEIGHT ALGINATE IN PSEUDOMONAS AERUGINOSA

By

Carrie-Lynn Keiski

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Biochemistry
University of Toronto

© Copyright by Carrie-Lynn Keiski (2010)
ABSTRACT

STRUCTURAL AND FUNCTIONAL STUDIES OF ALGK:
A PROTEIN REQUIRED FOR THE SECRETION OF HIGH-
MOLECULAR WEIGHT ALGINATE IN PSEUDOMONAS
AERUGINOSA

Carrie-Lynn Keiski
Ph.D. Thesis 2010
Department of Biochemistry
University of Toronto

Alginate is an exopolysaccharide secreted by Pseudomonas aeruginosa and is a
major component of biofilms that infect the lungs of cystic fibrosis patients. Ten
proteins have been implicated in alginate polymerization, modification and export,
and are believed to assemble into a multi-protein complex that spans the cell envelope
and coordinates the synthesis and secretion of alginate. AlgK is a protein encoded in
the alginate biosynthetic operon, which is required for the secretion of high-molecular
weight alginate. This study describes structural and functional studies of AlgK to
improve our understanding of AlgK’s role in alginate biosynthesis.

To shed light on the function of AlgK, C
14-palmitic acid labeling and sucrose
gradient fractionation studies confirmed that AlgK is an outer membrane lipoprotein.
Cellular fractionation experiments also found that AlgK is involved in the proper localization of AlgE, the alginate secretion pore in the outer membrane. The structure of AlgK was determined to 2.5 Å resolution by X-ray crystallography and revealed that the protein folds into 22 α-helices that pack into a flexible right-handed solenoid. Closer examination of the amino acid sequence revealed that AlgK carries 9.5 tetratricopeptide repeat (TPR)-like elements. Given the role that TPR motifs generally play in protein-protein interaction and the assembly of multi-protein complexes, the presence of these motifs in AlgK suggests that it can bind to one or more proteins.

Based on the results presented in this study, we propose that AlgK acts as a scaffold for the assembly of the alginate secretion complex. By mapping highly conserved residues onto the surface of our model, three putative sites of protein-protein interaction were identified. We hypothesize that the N-terminus of AlgK binds to AlgE in the outer membrane, and the C-terminus of AlgK binds to periplasmic and/or inner membrane Alg proteins, thereby acting as a linker between the inner and outer membrane components of the alginate biosynthetic complex. We further hypothesize that together AlgE and AlgK constitute a novel exopolysaccharide secretin. The alginate biosynthetic complex appears to be distinct from the canonical capsular polysaccharide systems currently described.
# TABLE OF CONTENTS

**CHAPTER 1** ...................................................................................................................... 1

## INTRODUCTION ............................................................................................................. 1

1.1 Bacterial Biofilms ........................................................................................................ 1
    1.1.1 What are biofilms? ............................................................................................. 1
    1.1.2 Biofilm structure ............................................................................................. 2
    1.1.3 Exopolysaccharides ......................................................................................... 3
    1.1.4 Biofilm life cycle ............................................................................................. 5
    1.1.5 Biofilm and disease ......................................................................................... 8

1.2 *P. aeruginosa* lung infections in cystic fibrosis patients ........................................ 9
    1.2.1 Cystic Fibrosis ............................................................................................... 10
    1.2.2 Bacterial pathogens that infect the CF lung ................................................... 11
    1.2.3 *P. aeruginosa* and the CF lung .................................................................... 12
        1.2.3.1 Biofilm mode of growth ........................................................................ 13
        1.2.3.2 Factors in CF lung that facilitate persistence of *P. aeruginosa* ............ 13
        1.2.3.3 Hypermutability and accumulation of genetic variants ....................... 13
        1.2.3.4 Mucoid phenotype ................................................................................ 14

1.3 Alginate ..................................................................................................................... 15
    1.3.1 Structure of alginate ..................................................................................... 15
    1.3.2 The alginate biosynthetic operon .................................................................... 16
    1.3.3 Regulation of alginate biosynthesis ............................................................... 18
    1.3.4 Alginate biosynthesis .................................................................................... 21
        1.3.4.1 Precursor production ............................................................................. 21
1.3.4.2 Alginate polymerization and transport across the inner membrane........ 23
1.3.4.3 Alginate modification ........................................................................... 25
1.3.4.4 Periplasmic conduit.............................................................................. 27
1.3.4.5 Translocation of alginate across the outer membrane....................... 28
1.4 Thesis Objective .......................................................................................... 29
1.5 Thesis Organization ...................................................................................... 30

CHAPTER 2 ......................................................................................................... 31

STRUCTURE DETERMINATION OF ALGK BY PROTEIN X-RAY CRYSTALLOGRAPHY ................................................................................................................. 31

2.1 Overview....................................................................................................... 31
2.2 Materials and Methods.................................................................................. 33
  2.2.1 Cloning....................................................................................................... 33
  2.2.2 Protein expression...................................................................................... 35
  2.2.3 Protein purification ................................................................................... 35
    2.2.3.1 28K1 purification.................................................................................. 36
    2.2.3.2 26K1 purification.................................................................................. 37
    2.2.3.3 Purification of 26K1ntag...................................................................... 38
  2.2.4 Limited proteolysis studies...................................................................... 39
  2.2.5 Western blots probing for His-tag on chymotrypsin-digested 26K1-PA and
    28K1-PA ......................................................................................................... 39
  2.2.6 Purification of chymotrypsin-digested proteins........................................ 40
    2.2.6.1 Purification of chymotrypsin-digested 26K1-PA ................................. 40
    2.2.6.2 Purification of chymotrypsin-digested 28K1-PA ................................. 41
2.2.7 Crystallization of 26K1-PF from *P. fluorescens*................................. 41
2.2.8 Native data collection ............................................................................ 42
2.2.9 Design and generation of selenomethionine (SeMet) triple mutant ......... 42
2.2.10 Expression, purification, crystallization and X-ray analysis of SeMet triple mutant ....................................................................................................................... 43
2.2.11 Structure solution and model refinement.............................................. 44
2.2.12 Superposition and comparison of the AlgK molecules present in the asymmetric unit......................................................................................................................... 45
2.2.13 Identification and characterization of TPRs in the AlgK structure.......... 45
2.2.14 Identification of structural homologues and superposition with AlgK model 45
2.3 Results............................................................................................................. 46
2.3.1 Crystallization of *P. aeruginosa* AlgK .................................................... 46
2.3.1.1 28K1-PA construct and recombinant protein ........................................ 46
2.3.1.2 26K1-PA construct and recombinant protein ........................................ 47
2.3.1.3 26K1ntag-PA construct and recombinant protein................................ 48
2.3.1.4 Production of a protease-resistant core of AlgK................................. 50
2.3.2 Crystallization of AlgK orthologues.......................................................... 52
2.3.2.1 *P. syringae* AlgK ............................................................................. 53
2.3.2.2 *P. fluorescens* AlgK ....................................................................... 54
2.3.3 Phasing......................................................................................................... 56
2.3.4 Structure solution and refinement.............................................................. 57
2.3.5 General overview of structure................................................................. 59
2.3.6 Conformational Flexibility of AlgK ........................................................ 60
2.3.7 AlgK is a member of the TPR superfamily ..................................................... 64

2.4 Discussion............................................................................................................... 69

2.4.1 Conformational flexibility of AlgK ................................................................. 69

2.4.2 AlgK’s TPR-like motifs................................................................................... 70

2.5 Conclusions............................................................................................................. 72

CHAPTER 3 .................................................................................................................... 73

ALGK ACTS AS A SCAFFOLD FOR THE ASSEMBLY OF THE ALGINATE

BIOSYNTHETIC COMPLEX ...................................................................................... 73

3.1 Overview................................................................................................................. 73

3.2 Materials and Methods............................................................................................ 74

3.2.1 Identification of putative AlgK binding sites................................................... 74

3.2.2 Bioinformatic characterization of proteins ...................................................... 75

3.3 Results..................................................................................................................... 75

3.3.1 Potential sites of protein-protein interaction.................................................... 75

3.3.2 Outer membrane proteins containing a TPR and porin domain ...................... 79

3.4 Discussion............................................................................................................... 80

3.4.1 Wzy-dependent biosynthesis of Group I/IV capsular polysaccharides ........... 82

3.4.2 ABC transporter-dependent biosynthesis of Group II/III capsular polysaccharides ........................................................................................................ 83

3.4.3 Exopolysaccharide production......................................................................... 85

3.4.4 Alginate is secreted by novel secretion complex............................................. 86

3.4.4.1 AlgK is a stable component of the periplasmic conduit ........................... 87

3.4.4.2 AlgK acts as a scaffold for the assembly of the complex ......................... 88
3.4.4.3 N-terminal binding sites................................................................. 89
3.4.4.4 C-terminal binding site ..................................................................... 90
3.4.5 Similarities between alginate, cellulose and poly β-1-6-GlcNAc biosynthetic systems...................................................................................................................... 91

CHAPTER 4.................................................................................................................... 95

TOWARDS THE IDENTIFICATION OF ALGK’S BINDING PARTNERS ........ 95

4.1 Overview.............................................................................................................. 95

4.2 Materials and Methods............................................................................................ 96

4.2.1 Expression, purification and refolding of *P. aeruginosa* AlgE .................... 96
4.2.2 Expression, purification and thrombin cleavage of *P. aeruginosa* 26K1ntag AlgK.......................................................................................................................... 97
4.2.3 Expression and purification of *P. aeruginosa* 26K1 AlgK .................................. 97
4.2.4 Nickel affinity co-elution studies..................................................................... 97
4.2.5 Western Blots for co-elution............................................................................ 98
4.2.6 Gel filtration studies......................................................................................... 98
4.2.7 Growth and harvesting of FRD bacterial cultures ........................................... 99
4.2.8 Western blots for stability assays................................................................. 100
4.2.9 PCR to verify insertion of Gm cassette in FRD∆alg44 and FRD∆algX......... 100

4.3 Results................................................................................................................... 101

4.3.1 NiNTA co-elution studies.................................................................................... 101
4.3.2 Gel filtration studies........................................................................................ 105
4.3.3 Western blots comparing Alg protein levels in FRD mutants and FRD1..... 108
4.3.4 PCR for FRD∆alg44 and FRD∆algX mutant ................................................... 114
4.4 Discussion ............................................................................................................. 116

4.4.1 K-E interaction studies ................................................................................... 116

4.4.2 Protein stability assays ................................................................................... 117

4.4.2.1 FRDΔalgK and FRDalgKC28S .................................................................. 118

4.4.2 Acetylation machinery mutants ..................................................................... 118

4.4.3 FRDΔalgX ...................................................................................................... 119

4.4.4 FRDΔalg44 .................................................................................................... 120

4.4.5 FRDΔalgG ..................................................................................................... 120

4.5 Conclusions ........................................................................................................... 121

CHAPTER 5 .................................................................................................................. 122

SUMMARY OF RESULTS AND FUTURE DIRECTIONS ........................................ 122

5.1 Summary of Results .............................................................................................. 122

5.1.1 AlgK is an outer membrane lipoprotein .......................................................... 122

5.1.2 AlgK promotes the proper localization of the porin AlgE ............................... 123

5.1.3 Structure of AlgK ........................................................................................... 124

5.1.4 AlgK acts as a scaffold for the assembly of the alginate biosynthetic complex ................................................................................................................................................................. 125

5.1.5 Identification of AlgK binding partners .......................................................... 126

5.2.1 AlgK’s role in biogenesis of AlgE .................................................................... 127

5.5.2 Identification and/or validation of AlgK binding partners ............................. 129

5.5.3 PgaA, BcsC and AlgK/AlgE as a novel secretins ............................................ 131

APPENDIX 1 ................................................................................................................. 132
ALGK IS AN OUTER MEMBRANE LIPOPROTEIN THAT PROMOTES THE
PROPER LOCALIZATION OF ALGE

A1.1 Overview

A1.2 Materials and Methods

A1.2.1 Palmitic acid labeling

A1.2.2 Membrane fractionation

A1.2.3 Succinate dehydrogenase activity

A1.2.4 Western blots for membrane fractionations

A1.3 Results

A1.3.1 AlgK is a lipoprotein

A1.3.2 AlgK is tethered to the outer membrane

A1.3.3 AlgK promotes the proper localization of AlgE in the outer membrane

A1.4 Discussion

A1.4.1 Palmitoylation experiment

A1.4.2 Outer membrane lipoprotein biogenesis

A1.4.3 Outer membrane β-barrel protein biogenesis

A1.4.4 AlgK’s role in AlgE biogenesis

A1.5 Conclusions

REFERENCES
LIST OF TABLES

CHAPTER 2

Table 2.1 AlgK expression vectors.................................................................34
Table 2.2 X-ray data collection and refinement statistics.................................58
Table 2.3 Summary of completeness of AlgK molecules in the asymmetric unit....59
Table 2.4 TPR motif lengths...........................................................................68
Table 2.5 Inter-repeat packing angles.............................................................69

CHAPTER 4

Table 4.1 Ratios and concentrations of AlgK and AlgE..................................99
Table 4.2 FRD deletion mutants......................................................................100
Table 4.3 Summary of PCR reactions and expected size of products...............101
Table 4.4 Summary of experiments and controls for AlgK-AlgE co-elution studies...102
Table 4.5 Summary of stability assay results...................................................108
LIST OF FIGURES

CHAPTER 1

Figure 1.1 Bacterial biofilms ................................................................. 4
Figure 1.2 Biofilm life cycle ................................................................. 7
Figure 1.3 P. aeruginosa ................................................................. 12
Figure 1.4 Alginate ................................................................. 17
Figure 1.5 Structure of alginate ................................................................. 18
Figure 1.6 Organization of the alginate biosynthetic operon in P. aeruginosa .... 18
Figure 1.7 Regulation of σ22 activity ................................................................. 19
Figure 1.8 Proposed alginate biosynthetic pathway .................................................. 23
Figure 1.9 Proposed alginate acetylation complex .................................................... 26
Figure 1.10 Proposed alginate biosynthetic complex .................................................... 28

CHAPTER 2

Figure 2.1 Purification and digestion of 28K1-PA protein ........................................... 47
Figure 2.2 Purified 26K1-PA protein ................................................................. 48
Figure 2.3 Purified untagged 26K1ntag-PA protein .................................................... 49
Figure 2.5 Crystals of AlgK from P. aeruginosa ................................................................. 50
Figure 2.6 Limited proteolysis studies of 28K1-PA and 26K1-PA AlgK .................. 51
Figure 2.7 Purification of digested and cloned protease resistant cores ........................................... 52
Figure 2.8 Purified 26K1-PS protein ................................................................. 53
Figure 2.9 Purification of 26K1-PF protein ................................................................. 54
Figure 2.10 Crystals of 26K1-PF AlgK .................................................................56
Figure 2.11 AlgK structure.............................................................62
Figure 2.12 Multiple sequence alignment of AlgK homologues..................63
Figure 2.13 Local displacements....................................................64
Figure 2.14 Structural homologues of AlgK.......................................66
Figure 2.15 TPR motif.................................................................67
Figure 2.16 Multiple sequence alignment of AlgK’s TPR-like repeats............67
Figure 2.17 TPR, SLR and AlgK consensus sequence...............................71

CHAPTER 3

Figure 3.1 Putative binding sites mapped onto multiple sequence alignment of AlgK homologues........................................................77
Figure 3.2 Putative sites of protein-protein interaction in the AlgK model...........78
Figure 3.3 Predicted TPR and porin domains.........................................81
Figure 3.4 Alginate and capsular polysaccharide secretion systems..................85
Figure 3.5 Bacterial polysaccharide biosynthetic complexes............................93

CHAPTER 4

Figure 4.1 Co-elution studies with N-terminally His-tagged AlgE and untagged AlgK.................................................................103
Figure 4.2 Co-elution studies with C-terminally His-tagged AlgK and untagged AlgE.................................................................104
Figure 4.3 Co-elution studies with N-terminally His-tagged AlgK and untagged AlgE

Figure 4.4 Chromatogram of AlgK and AlgE mixed in a 1:1 ratio

Figure 4.5 Chromatogram of AlgK and AlgE mixed in a 3:1 ratio

Figure 4.6 Chromatogram of AlgK and AlgE mixed in a 1:3 ratio

Figure 4.7 Western blots of FRD1, FRDΔalgK and FRDΔalgG

Figure 4.8 Western blots of FRD1, FRDΔalg8 and FRDΔalg44

Figure 4.9 Western blots of FRD1, FRDΔalgF, FRDΔalgI and FRDΔalgI

Figure 4.10 Western blots of FRD1 and FRDΔalgX

Figure 4.11 Western blots of FRD1, FRDalgKC28S and FRD2

Figure 4.12 Western blots of FRD1 and FRD mutant strains probed for Alg44

Figure 4.13 Orientation of the Gm cassette in FRDΔalg44 and FRDΔalgX mutants

APPENDIX 1

Figure A1.1 Lipidation and subcellular localization of AlgK

Figure A1.2 Biogenesis of AlgK and other outer membrane lipoproteins

Figure A1.3 Biogenesis of AlgE and other outer membrane β-barrel proteins
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>ATP-binding cassette</td>
</tr>
<tr>
<td>Aborku</td>
<td><em>Alcanivorax borkumensis</em></td>
</tr>
<tr>
<td>AHL</td>
<td>acyl homoserine lactone</td>
</tr>
<tr>
<td>ali</td>
<td>aliphatic side chain</td>
</tr>
<tr>
<td>aro</td>
<td>aromatic side chain</td>
</tr>
<tr>
<td>ASM</td>
<td>American Society for Microbiology</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine tri-phosphate</td>
</tr>
<tr>
<td>Avine</td>
<td><em>Azotobacter vinelandii</em></td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BAM</td>
<td>β-barrel assembly machinery</td>
</tr>
<tr>
<td>BCIP</td>
<td>5-Bromo-4-Chloro-3'-Indolyphosphate p-Toluidine Salt</td>
</tr>
<tr>
<td>BNL</td>
<td>Brookhaven National Laboratories</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>c-di-GMP</td>
<td>cyclic-di-guanylate</td>
</tr>
<tr>
<td>CF</td>
<td>cystic fibrosis</td>
</tr>
<tr>
<td>CFTR</td>
<td>cystic fibrosis transmembrane conductance regulator</td>
</tr>
<tr>
<td>DCPIP</td>
<td>2, 6-dichlorophenolindophenol</td>
</tr>
<tr>
<td>DDM</td>
<td>dodecyl-β-D-maltopyranoside</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>Gm</td>
<td>gentamicin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>GMD</td>
<td>GDP-mannose dehydrogenase</td>
</tr>
<tr>
<td>GMP</td>
<td>GDP-mannose pyrophosphorylase</td>
</tr>
<tr>
<td>hdp</td>
<td>hydrophobic side chain</td>
</tr>
<tr>
<td>His&lt;sub&gt;6&lt;/sub&gt;</td>
<td>hexahistidine</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>IM</td>
<td>inner membrane</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodaltons</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani medium</td>
</tr>
<tr>
<td>M</td>
<td>mixture of inner and outer membranes</td>
</tr>
<tr>
<td>Malg</td>
<td><em>Marinobacter algicola</em></td>
</tr>
<tr>
<td>Mbp</td>
<td>mega base pairs</td>
</tr>
<tr>
<td>MAP</td>
<td>modified alginate-producing</td>
</tr>
<tr>
<td>MES</td>
<td>2-(N-morpholino)ethanesulfonic acid</td>
</tr>
<tr>
<td>NaCN</td>
<td>sodium cyanide</td>
</tr>
<tr>
<td>NAD(H)</td>
<td>nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NBT</td>
<td>nitro-blue tetrazolium chloride</td>
</tr>
<tr>
<td>NCS</td>
<td>non-crystallographic symmetry</td>
</tr>
<tr>
<td>NHS</td>
<td>N-hydroxysuccinimide</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>nickel-nitrilotriacetic acid</td>
</tr>
<tr>
<td>NSLS</td>
<td>National Synchrotron Light Source</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>optical density at 600 nm</td>
</tr>
<tr>
<td>OGT</td>
<td>O-linked N-acetylglucosamine transferase</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>OM</td>
<td>outer membrane</td>
</tr>
<tr>
<td><em>palgD</em></td>
<td><em>algD</em> promoter</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td><em>Paeru</em></td>
<td><em>Pseudomonas aeruginosa</em></td>
</tr>
<tr>
<td>PCP</td>
<td>polysaccharide co-polymerase</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td><em>PfluorW</em></td>
<td><em>Pseudomonas fluorescens</em> WCS374r</td>
</tr>
<tr>
<td>PI</td>
<td>EDTA-free Roche Complete Protease Inhibitor</td>
</tr>
<tr>
<td>PIA</td>
<td><em>Pseudomonas</em> isolation agar</td>
</tr>
<tr>
<td>PMI</td>
<td>phosphomannose isomerase</td>
</tr>
<tr>
<td>PMS</td>
<td>phenazine methosulfate</td>
</tr>
<tr>
<td>poly-β-1,6-GlcNAc</td>
<td>poly-β1,6-N-acetyl-d-glucosamine</td>
</tr>
<tr>
<td>polyG</td>
<td>polyguluronate</td>
</tr>
<tr>
<td>polyM</td>
<td>polymannuronate</td>
</tr>
<tr>
<td>POTRA</td>
<td>Polypeptide-TRansport-Associated</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulphonyl fluoride</td>
</tr>
<tr>
<td>PPR</td>
<td>pentatricopeptide repeat</td>
</tr>
<tr>
<td>PVDF</td>
<td>polyvinylidene fluoride</td>
</tr>
<tr>
<td>R</td>
<td>repeat</td>
</tr>
<tr>
<td>r.m.s.</td>
<td>root mean square</td>
</tr>
<tr>
<td>SAD</td>
<td>single anomalous dispersion</td>
</tr>
<tr>
<td>SDH</td>
<td>succinate dehydrogenase</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>SeMet</td>
<td>selenomethionine</td>
</tr>
<tr>
<td>SLR</td>
<td>Sell-like repeat</td>
</tr>
<tr>
<td>sm</td>
<td>small side chain</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris-buffered saline</td>
</tr>
<tr>
<td>TBST</td>
<td>TBS containing 0.05% Tween-20</td>
</tr>
<tr>
<td>TM</td>
<td>transmembrane domain</td>
</tr>
<tr>
<td>TPR</td>
<td>tetratricopeptide repeat</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>V_m</td>
<td>Matthew’s coefficient</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Bacterial Biofilms

1.1.1 What are biofilms?

Bacteria display two distinct lifestyles: they can flourish as free-swimming planktonic cells or they can live within a biofilm. The canonical definition of a biofilm is a surface-attached community of cells encased in a hydrated polymeric matrix. Bacterial biofilms are ubiquitous; they can colonize all environmental niches (i.e. natural, industrial and clinical environments) and can grow on biotic or abiotic surfaces. For example, biofilms can grow in hot springs, deep-sea vents, oil pipelines and on the surface of teeth, medical implants or living tissues. Another defining characteristic of biofilms is their persistence. For reasons we are only beginning to understand, biofilms are incredibly difficult to eliminate. They are highly resistant to predators and tolerant to biocides. Biofilms are thought to be the prevalent mode of growth of bacteria in the natural environment as they appear to offer bacteria several advantages: 1) Surface-attachment can offer bacteria stability and can enable bacteria to remain in a more favourable environment; 2) the close proximity of cells within the biofilm can facilitate or promote intercellular interactions, e.g. horizontal exchange of genetic information, metabolite exchange between interdependent communities, communication between cells for coordinated adaptation to a changing environment and 3) biofilms provide a protected
mode of growth that enables bacteria to survive hostile environmental agents and conditions, *e.g.* UV light, metal toxicity, phagocytosis and other immune factors, salinity, desiccation, antibiotics, antimicrobial agents and shear forces\(^1,^2,^5\).

### 1.1.2 Biofilm structure

Generally, biofilms consist of bacterial microcolonies, *i.e.* aggregates of cells, interspersed within a hydrated polymeric matrix\(^1\). The varying density of the matrix creates water channels that enable the uptake of nutrients and disposal of waste into the bulk phase surrounding the biofilm\(^6\text{-}^8\) (*Figure 1.1*). The extracellular matrix is an essential component of the biofilm acting as a structural scaffold that keeps cells together and determines their spatial organization within the biofilm\(^9,^10\). The extracellular matrix is highly hydrated and contains a mixture of biopolymers such as exopolysaccharides, DNA and protein\(^9\). Membrane vesicles, rhamnolipids, other cell surface polysaccharides, environmental agents and cell detritus have also been shown to be components of the biofilm matrix\(^11\text{-}^13\).

Biofilms have been described as being “structurally complex and dynamic biological systems”\(^2\). At the gross level, biofilms can develop into a number of different morphologies, *e.g.* mushroom-like structures, flat structures, or elongated filamentous structures. The morphology is determined by the interplay between genetic and environmental influences, such as surface properties, nutrient availability, carbon source, organisms living within the biofilm, matrix composition and hydrodynamics\(^1,^2,^14,^15\). Single-species, mixed-species and even mixed-kingdom biofilms have been observed\(^16\). The heterogeneous structure of a biofilm creates different microniches that give rise to
phenotypically and genotypically distinct subpopulations of bacteria, thus favouring survival in a changing environment\textsuperscript{4,5}.

### 1.1.3 Exopolysaccharides

Bacterial exopolysaccharides can be a major extracellular matrix constituent\textsuperscript{10}. They are long-chain complex polysaccharides that are loosely associated with the bacterial cell surface. Other cell surface polysaccharides such as capsular polysaccharides and lipopolysaccharides are tightly anchored to the cell \textit{via} covalent attachment\textsuperscript{17}. Exopolysaccharides are involved in a number of different functions and are an essential component of bacterial biofilms. They have been implicated in different stages of biofilm development, the maintenance of its structural integrity and determination of its architecture\textsuperscript{18-21}. Exopolysaccharides have also been shown to play an important role in cell-cell and/or cell-surface interactions\textsuperscript{20,22-24}, protection against harmful environmental agents\textsuperscript{15,21,25,26} and pathogenesis\textsuperscript{26,27}. Examples of biofilm exopolysaccharides include alginate, pel- and psl-polysaccharides, cellulose, colanic acid and poly-\(\beta\)1,6-N-acetyl-D-glucosamine (poly-\(\beta\)-1,6-GlcNAc). Bacteria can produce multiple types of exopolysaccharides and the biofilm extracellular matrix can consist of a mixture of different exopolysaccharides\textsuperscript{9,10,21,28}.

There is a great variation in the composition and, as a consequence, the physicochemical properties of exopolysaccharides produced by bacteria. Negatively charged (\textit{e.g.} alginate), neutral (\textit{e.g.} homopolymers in oral biofilms) and positively charged (\textit{e.g.} poly-\(\beta\)-1,6-GlcNAc) exopolysaccharides have been observed.
Figure 1.1 Bacterial biofilms. Upper left-hand panel: Transmission electron microscope image of *Pseudomonas aeruginosa* cells surrounded by extracellular polymeric matrix. Scale bar, 1 µm. Figure reproduced from Singh *et al.* (2000)\(^29\) and reprinted with permission from Macmillan Publishers Ltd: Nature (copyright 2000). Upper right-hand panel: Epifluorescence microscopy image of mixed biofilm stained with 4,6-diamidino-2-phenylindole (DAPI). Scale bar, 20 µm. Figure reproduced from Donlan (2002)\(^30\). Lower panel: Scanning electron micrograph of a *Haemophilus influenzae* biofilm. Scale bar, 2 µm. Figure reproduced from Gallaher *et al.* (2006)\(^31\), an article originally published by BioMed Central.
In cellulose, β-1,4-glycosidic linkages confer rigidity to the polymer; meanwhile, the α-1,6-glycosidic linkages present in dextran enables the polysaccharide to remain more flexible\(^9\). Both homopolysaccharides and heteropolysaccharides have been observed in nature. Heteropolysaccharides are composed of repeating units containing 2-8 sugar moieties. The sugar moieties in the polysaccharide can be modified with substituents, \(e.g\). acetate, pyruvate, etc\(^{28}\).

### 1.1.4 Biofilm life cycle

The environmental cues that trigger cells to switch from planktonic to the biofilm lifestyle are not well understood and appear to vary from one genera of bacteria to another. For example, \textit{Pseudomonas aeruginosa} will form biofilms under almost all conditions that allow growth\(^{32}\). However, in other bacteria, nutrient content, temperature, carbon source, osmolarity, pH, iron and oxygen have all been shown to play a role in biofilm initiation\(^5,14,33\).

Recently, the intracellular level of the second messenger cyclic-di-guanylate (c-di-GMP) has been shown to play an important role in the transition from the planktonic to biofilm mode of growth. A positive correlation has been observed between c-di-GMP intracellular concentration and biofilm formation, exopolysaccharide production and sessility\(^{34,35}\). Meanwhile, a negative correlation has been observed between c-di-GMP and motility and virulence\(^{34,36}\). The steady-state levels of c-di-GMP within cells are regulated by the relative rates of enzymatic synthesis and degradation catalyzed by diguanylate cyclases and phosphodiesterases, respectively\(^{37}\). Environmental cues activate signal transduction pathways within cells that modify the activity and/or expression level
of these enzymes thereby altering intracellular levels of c-di-GMP\textsuperscript{38}. Surface-attachment has been shown to be an environmental cue that leads to an increase in intracellular c-di-GMP levels\textsuperscript{39}. Cyclic-di-GMP affects cellular function at numerous levels: 1) acts as an allosteric activator of cellulose synthase\textsuperscript{40}; 2) binds directly to other proteins implicated in exopolysaccharide production\textsuperscript{35,41}; 3) binds to riboswitches in mRNA thereby regulating expression of a wide variety of genes\textsuperscript{42}; and 4) modifies gene expression\textsuperscript{43}, in part by modifying the activity of transcription factors thereby promoting expression of exopolysaccharide biosynthetic genes\textsuperscript{44}.

Most biofilms appear to follow a similar developmental path however the specific mechanisms underlying each stage of development is different depending on the bacteria and environmental conditions\textsuperscript{14,45}. \textit{P. aeruginosa} biofilms have been extensively studied and will be used here as a paradigm to illustrate the different stages of the biofilm life cycle (Figure 1.2).

**Initial attachment** – The first step in biofilm formation involves the attachment of bacterial cells to a surface as a monolayer\textsuperscript{32,46}. Type IV pili, flagella, lipopolysaccharides, cup fimbriae and the psl-polysaccharide have all been shown to play some role in early attachment of \textit{P. aeruginosa} cells\textsuperscript{20,32,47,48}.

**Microcolony formation** – After initial attachment, cells will form multicellular clusters referred to as microcolonies\textsuperscript{32}. Both clonal expansion and type IV pili-mediated twitching motility have been implicated in microcolony formation depending on growth conditions\textsuperscript{32,45,49}.

**Growth and maturation** – The differentiation of microcolonies into a mature biofilm is characterized by an increased production of exopolysaccharides and tolerance
to antibiotics with a concomitant decrease in motility\textsuperscript{14}. Environmental factors have been shown to play a crucial role in determining the morphology of the mature biofilm. Depending on the carbon source, \textit{P. aeruginosa} biofilms have been shown to develop into two distinct morphologies: mushroom-like structures or flat uniform biofilms\textsuperscript{45,49}.

**Dispersal** – Dispersal is the final step of the biofilm life cycle where cells are shed from the biofilm to colonize new surfaces and reinitiate the biofilm growth cycle. Dispersal can occur passively by sloughing caused by shear forces acting on the biofilm, or actively via seeding dispersal\textsuperscript{50}. During seeding dispersal, a subpopulation of cells become motile and are released, leaving hollow cavities in the biofilm\textsuperscript{46,50,51}. Enzymatic degradation of the matrix, reactive oxygen species-induced cell lysis, cellular production of nitric oxide and quorum-sensing have all been implicated in seeding dispersal\textsuperscript{11,50,52,53}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{biofilm_cycle.png}
\caption{Biofilm life cycle. Stages of the biofilm life cycle: 1) attachment, 2) microcolony formation, 3) and 4) growth and maturation, 5) seeding dispersal. Graphic created by and reprinted with the permission of Peg Dirckx (copyright 2003), Center for Biofilm Engineering, Montana State University.}
\end{figure}
A form of cell-cell communication known as quorum sensing appears to play a role in biofilm development, maturation and dispersal. Quorum sensing involves the secretion or passive diffusion of bacterial-made molecules, called auto-inducers, into the environment that modify gene expression in other bacteria in a concentration-dependent manner\textsuperscript{54}. \textit{P. aeruginosa} has been shown to have two acyl homoserine lactone systems, \textit{las} and \textit{rhl}, and one non-AHL system, the Pseudomonas quinolone system\textsuperscript{55}. In 1998, Davies \textit{et al.} showed that \textit{P. aeruginosa las} and \textit{lasl/rhl} mutants formed flat undifferentiated biofilms unlike the wild-type strain that formed structured biofilms\textsuperscript{56}. Since then, other groups have found that the role of quorum sensing in \textit{P. aeruginosa} biofilm development is complex and is determined by the nutritional environment\textsuperscript{57-60}. Other functions regulated by quorum sensing include the production of secreted factors such as siderophores, proteases and toxins, and swarming motility\textsuperscript{55}.

\textbf{1.1.5 Biofilm and disease}

Given their ability to colonize and flourish in different environments, it is not surprising that biofilms cause chronic and persistent infections in humans. Greater than 60\% of infections in the developed world are thought to be caused by biofilms\textsuperscript{4}. Biofilms can cause serious infections by growing on the surface of indwelling medical devices such as central venous and urinary catheters, prosthetic heart valves, intrauterine devices and contact lenses\textsuperscript{61}. The most common culprits responsible for these infections are \textit{Staphylococci}, \textit{i.e. Staphylococcus aureus} and \textit{Staphylococcus epidermidis}, and \textit{P. aeruginosa}, two opportunistic pathogens that are part of the normal commensal flora\textsuperscript{2}. 
Biofilms can also grow on living tissues to cause infections such as middle ear infections (otitis media), native valve endocarditis, periodontitis and chronic bacterial prostatitis. Biofilm infections are notoriously difficult to eradicate. Even in healthy individuals, biofilms are rarely resolved by the immune system in part because they protect bacteria from phagocytes and antibodies. Instead, biofilms instigate a prolonged immune response that can damage surrounding host tissues. Furthermore, biofilm-embedded bacteria are approximately 1000 times more tolerant to antibiotics than their planktonic counterparts. During an infection, conventional antibiotics can effectively eliminate the planktonic bacterial cells, however the biofilm and bacteria embedded within it remain. Eventually, biofilms may need to be surgically removed from the patient or the medical device may need to be replaced. The increased tolerance of biofilms toward antibiotics is not well understood, however, several mechanisms have been proposed: 1) Exopolysaccharide may act as a physical barrier that binds, prevents entry or dilutes offensive agents, e.g. reactive oxygen species, charged metals, antibodies and large antimicrobial agents; 2) Subpopulations of cells within the biofilm are in a dormant/slow-growing phase rendering antibiotics that target cellular processes less effective; 3) The presence of genotypic variants within the biofilm gives rise to specialized subpopulations that favour the survival of the biofilm when confronted with a changing environment and antibiotic treatment and 4) novel biofilm-specific efflux pumps have also been implicated in the increased resistance of biofilms to antibiotics.

1.2 *P. aeruginosa* lung infections in cystic fibrosis patients
1.2.1 Cystic Fibrosis

One of the most well-studied biofilm-related diseases in humans is the colonization and persistence of *P. aeruginosa* biofilms in the lungs of cystic fibrosis patients. Cystic fibrosis (CF) is the most common lethal inherited disease in Caucasian populations\(^2\). CF is an autosomal recessive disease characterized by disturbances in the normal functioning of the airways, liver, pancreas, small intestine, skin and reproductive tract\(^2,3\). These physiological disturbances are caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. This gene encodes a chloride ion channel that is a member of the adenosine-triphosphate (ATP)-binding cassette transporter superfamily, which is localized to the apical membrane of some epithelial cells and is needed for their normal functioning. CFTR is responsible for chloride secretion and is involved the regulation of transport of other electrolytes (*e.g.* sodium and potassium) across the membrane\(^3,4\). To date, over 1600 CFTR mutations associated with the disease have been characterized [http://www.genet.sickkids.on.ca/cftr/StatisticsPage.html\(^{5}\)] with the most common being the ∆F508 mutation which lacks a phenylalanine at position 508 in the protein\(^4\). This mutation causes misfolding of the protein interfering with its trafficking and leading to its degradation\(^5-7\). How the ∆F508 and other mutations in CFTR lead to the plethora of clinical manifestations observed in CF patients is not entirely understood.

The major cause of morbidity and mortality in CF patients is lung failure caused by chronic bacterial lung infections, extensive tissue damage and airway plugging\(^2,3\). Several factors are thought to contribute to these clinical features. The inability of CFTR-deficient epithelial cells to secrete chloride leads to the production of an unusually thick
airway surface liquid that impairs mucociliary clearance, promotes biofilm formation, and impairs neutrophil migration and antimicrobial molecule diffusion. Other observed changes that might render the CF lung more susceptible to bacterial infection include defective alveolar macrophages and decreased secretion and/or inactivation of antimicrobial molecules. The CF lung exhibits an exaggerated inflammatory response to pathogens, which is believed to be caused by the increased secretion of pro-inflammatory cytokines and the chemokine interleukin-8. Recently, the accumulation of ceramide in CF airway epithelial cells was shown to indirectly facilitate bacterial adherence and promote inflammation. Oxidative bursts of polymorphonuclear leukocytes in response to P. aeruginosa biofilms in the lungs of CF patients are a contributing factor to the inflammation and tissue damage. The combination of the exaggerated inflammatory response exacerbated by chronic bacterial infections is believed to be responsible for the extensive tissue damage that leads to pulmonary failure.

1.2.2 Bacterial pathogens that infect the CF lung

From birth onwards, CF patients are plagued by bacterial lung infections that begin as recurrent temporary infections but eventually become chronic and punctuated by acute debilitating exacerbations. The most common pathogens detected in the CF lung are S. aureus, P. aeruginosa and Haemophilus influenzae. The CF lung is colonized by a mixture of bacteria, however one pathogen tends to be more prevalent depending on the age of the patient, i.e. during infancy S. aureus is the predominant pathogen which is succeeded by H. influenzae and by late adolescence, 80% of CF patients are infected primarily with P. aeruginosa. Once established, P. aeruginosa infections in the CF
lung cannot be eradicated with conventional antibiotics or by the host immune response, thus enabling the bacteria to persist in the lungs for years until the patient succumbs to lung failure\textsuperscript{91}.

1.2.3 \textit{P. aeruginosa} and the CF lung

\textit{P. aeruginosa} is a rod-shaped gram-negative bacterium (Figure 1.3). Their metabolic versatility enables Pseudomonads to be ubiquitous and flourish in many environmental niches. Generally, they are found in terrestrial and aquatic environments and as commensals associated with animals and plants. \textit{P. aeruginosa} is also an opportunistic human pathogen that can cause life-threatening infections in immunocompromised or injured individuals\textsuperscript{93}. It remains unclear why \textit{P. aeruginosa} becomes the major pulmonary pathogen in CF patients from adolescence onwards and what mechanisms enable it to colonize and persist for years in the lung environment. Several factors have been proposed to explain these phenomena.

\textbf{Figure 1.3} \textit{P. aeruginosa}. Scanning electron micrograph of \textit{P. aeruginosa} cells. Courtesy of Janice Haney Carr. Centers for Disease Control and Prevention's Public Health Image Library (PHIL), image 232. Scale bar, 2.5 \textmu m.
1.2.3.1 Biofilm mode of growth

Several observations strongly suggest that *P. aeruginosa* cells grow as a biofilm in the lungs of CF patients. Most convincingly, *P. aeruginosa* microcolonies encased in a polymeric matrix have been observed on lung tissues and in the sputum from CF patients\(^2^9,9^4\). The ratio of quorum-sensing molecules in the CF lung resembles more closely that of *in vitro* biofilms than planktonic cells\(^2^9\). As discussed earlier, the biofilm mode of growth is characterized by an increased tolerance to antibiotics and immune evasion, both of which probably play an important role in *P. aeruginosa* persistence. Interestingly, *P. aeruginosa* biofilms grown on mucin or in co-culture with airway epithelial cells display an increased tolerance to antibiotics when compared to *P. aeruginosa* biofilms grown on an abiotic surface\(^9^5,9^6\).

1.2.3.2 Factors in CF lung that facilitate persistence of *P. aeruginosa*

Conditions in the CF lung appear to favour the colonization and long term survival of *P. aeruginosa*. *P. aeruginosa* cells can grow in the hypoxic mucus layer that covers CF epithelia cells; conditions that also promote the production of the exopolysaccharide alginate\(^9^7,9^8\). In normal epithelial cells, CFTR has been suggested to contribute to the innate immune response by acting as a receptor for the internalization of *P. aeruginosa* cells; this claim has not been adequately proven experimentally and remains highly controversial \(^9^9\). Lastly, ΔF508 epithelial cells release more iron than normal cells which promotes *P. aeruginosa* biofilm formation\(^9^6\).

1.2.3.3 Hypermutability and accumulation of genetic variants
There appears to be a large phenotypic variation in *P. aeruginosa* cells isolated from lungs of CF patients\(^{100}\). The stressful and fluctuating conditions in the CF lung are thought to promote the emergence of hypermutable strains of *P. aeruginosa* that accumulate adaptive mutations that favour survival in the lung. In contrast, no hypermutable *P. aeruginosa* strains were found in acute infections of non-CF patients\(^ {101}\). Frequent mutagenesis targets included the *mexz* gene that encodes a negative regulator of the MexXY-OprM multidrug efflux pump and virulence-related genes\(^ {102}\). The increased rate of mutation observed in hypermutable strains appears to be caused by mutations that impair the mismatch repair system\(^ {103}\). Other phenotypic changes observed in *P. aeruginosa* cells isolated from the CF lung include a decreased expression of immunogenic molecules\(^ {102,104,105}\).

### 1.2.3.4 Mucoid phenotype

The conditions in the lung also select for genotypic variants of *P. aeruginosa* that secrete copious amounts of the exopolysaccharide alginate; these variants are referred to as being "mucoid"\(^ {98,106}\). Initially, the CF lung is intermittently colonized by non-mucoid *P. aeruginosa* that can be eradicated by aggressive antibiotic therapy. Eventually, cells convert to a mucoid phenotype and the bacteria can no longer be cleared from the lungs\(^ {107,108}\). The emergence of mucoid cells in the CF lung is associated with an increased tolerance to antibiotics and accelerated pulmonary decline and is thus an indicator of bad prognosis\(^ {91,108,109}\). Alginate is a major component of *P. aeruginosa* biofilms that infect the lungs of CF patients\(^ {110}\). Interestingly, alginate is not a major component of *P. aeruginosa* biofilms outside the CF lung and there is no known ecological niche for mucoid *P.
aeruginosa\textsuperscript{110,111}. Therefore, the mucoid phenotype appears to be specifically selected for by the conditions in CF lung presumably because it offers a survival advantage to \textit{P. aeruginosa}. Alginate plays a role in adherence\textsuperscript{112-114} and protection against dessication\textsuperscript{115}, free radicals\textsuperscript{116}, cationic antimicrobial peptides\textsuperscript{117,118} and aminoglycosides\textsuperscript{119-121}. Importantly, alginate contributes to the immune evasion and virulence of \textit{P. aeruginosa} in the CF lung\textsuperscript{122-125}.

1.3 Alginate

Alginate is an exopolysaccharide synthesized by brown algae as well as the two bacterial genera \textit{Pseudomonas} and \textit{Azotobacter} (Figure 1.4)\textsuperscript{126}. Alginate is used for a different purpose in each of these organisms. As discussed earlier, alginate is the major extracellular polymeric component of \textit{P. aeruginosa} biofilms infecting the lungs of CF patients. \textit{P. aeruginosa} FRD1 is a strain derived from a clinical isolate of the lungs of a CF patient that is often used to study alginate biosynthesis. Alginate is an essential component of the cyst formed by the gram-negative soil bacterium \textit{Azotobacter vinelandii} that enables dormant bacteria to survive long periods of dessication\textsuperscript{126}. In brown algae, alginate acts as a structural component\textsuperscript{127}. Recent genome sequencing efforts have suggested that other organisms may also produce alginate such as the oil-degrading marine bacteria \textit{Marinobacter algicola} and \textit{Alcanivorax borkumensis} \textsuperscript{128,129}.

1.3.1 Structure of alginate
Alginate is a negatively charged unbranched polymer of β-1,4-linked D-mannuronate and its C5 epimer, L-guluronate. *P. aeruginosa* produces polymannuronate (polyM) blocks (-M-M-M-M-) interspersed by mixed mannuronate-guluronate blocks (-M-G-M-M-G-). *A. vinelandii* and brown algae synthesize alginate that also contains polyguluronate (polyG) blocks (-G-G-G-G-). The hydroxyls of bacterial alginates are selectively acetylated at positions C2 and/or C3 of mannuronate residues (Figure 1.5)\textsuperscript{126}. The composition (*i.e.* the block structure, ratio of guluronate to mannuronate residues, degree of acetylation) and chain length determine the physicochemical properties of alginate\textsuperscript{111}. For example, polyG blocks increase the rigidity of the polymer and, in the presence of divalent cations, form rigid gels. Meanwhile, the polyM blocks contribute to the formation of more flexible gels\textsuperscript{130}. O-acetylation increases the hydration of the polymer, contributes to its viscosity and renders alginate less susceptible to enzymes that degrade alginate\textsuperscript{131,132}.

### 1.3.2 The alginate biosynthetic operon

In *P. aeruginosa*, thirteen genes have been implicated in the biosynthesis and secretion of alginate, and with the exception of *algC*, all are located in an operon found at 3.96 Mega base pairs (Mbp) on the 6.26 Mbp genome of *P. aeruginosa* PAO1 (Figure 1.6)\textsuperscript{133,134}. The expression of the alginate biosynthetic operon is under the control of the *algD* promoter (PalgD) which is silent in typical *P. aeruginosa* strains and highly active in mucoid strains\textsuperscript{135}. The *algC* gene is located at 5.99 Mbp on the chromosome, where it has its own promoter and its expression is upregulated in mucoid strains\textsuperscript{136,137}.
Figure 1.4 Alginate. Top panel: LB plate streaked with mucoid and non-mucoid *P. aeruginosa* cells. Figure adapted from Jain and Ohman (2005)\textsuperscript{138} with permission from the American Society for Microbiology (ASM, DOI: 10.1128/IAI.73.10.6429-6436.2005, copyright 2005). Bottom panel: Alginate from *P. aeruginosa* culture precipitated with ethanol. Figure reproduced from Govan and Deretic (1996)\textsuperscript{91} with permission from ASM (copyright 1996).
1.3.3 Regulation of alginate biosynthesis

The transcriptional regulation of the alginate biosynthetic operon is complex and involves a number of different proteins. $\sigma^{22}$ is an alternative sigma factor encoded by the $algU$/$algT$ gene that is essential for the expression of the alginate biosynthetic operon\textsuperscript{139,140}. $\sigma^{22}$ positively regulates the transcription of the $PalgD$ operon, other regulators of $PalgD$ (i.e. $algB$, $algR$, $amrZ/algZ$) and the $algUmucABCD$ operon, thereby regulating its own expression\textsuperscript{91,141-143}. The proteins encoded by $mucABCD$ are responsible for modulating the activity of $\sigma^{22}$\textsuperscript{144}. MucA is the primary regulator of $\sigma^{22}$ activity\textsuperscript{145}. MucA is an inner membrane anti-sigma factor with 1 transmembrane domain (TM) flanked by an N-terminal cytosolic domain that binds and sequesters $\sigma^{22}$, and a C-
terminal periplasmic domain (Figure 1.7a)\textsuperscript{144,146}. MucB is an inner membrane protein that negatively regulates $\sigma_{22}$ activity and is thought to bind to the periplasmic domain of MucA\textsuperscript{144,147,148}. The function of MucC is unknown\textsuperscript{132}. MucD is also a negative regulator of $\sigma_{22}$ and displays amino acid similarity to the \textit{E. coli} HtrA periplasmic serine protease\textsuperscript{149}.

Figure 1.7 Regulation of $\sigma_{22}$ activity. (a) Under non-stressful conditions, $\sigma_{22}$ is sequestered by the cytoplasmic domain of MucA, thereby preventing its activity. (b) \textit{P. aeruginosa} cells infecting the lungs of CF patients acquire a genetic mutation that leads to the expression of a non-functional form of the MucA protein, \textit{i.e.} mucA$_{22}$, leading to the deregulation of the $algD$ operon and alginate overproduction. The inner and outer membranes are abbreviated IM and OM, respectively.

Recently, Qiu and colleagues have obtained evidence suggesting that the regulation of $\sigma_{22}$ in \textit{P. aeruginosa} is analogous to that of $\sigma_{E}$ in \textit{E. coli} and have proposed the following model for the release and activation of $\sigma_{22}$\textsuperscript{150}. Briefly, environmental stress is thought to cause the accumulation and misfolding of the small
outer membrane or periplasmic protein MucE which binds and derepresses the AlgW protease located in the inner membrane\textsuperscript{145}. This results in the cleavage of the periplasmic domain of MucA by AlgW, followed by the cleavage of MucA’s TM by MucP\textsuperscript{150}. Subsequently, ClpXP proteases appear to be involved in degradation of the cytoplasmic domain of MucA promoting the release of $\sigma_22$\textsuperscript{151}.

Other factors shown to play a role in the regulation of the $P_{algD}$ operon include the response regulators AlgR\textsuperscript{152} and AlgB\textsuperscript{153}, the AlgZ transcription factor\textsuperscript{154}, the histone-like proteins IHF\textsuperscript{155} and AlgP, as well as AlgQ\textsuperscript{156,157}. Alginate production also appears to be regulated at the post-translational level by the second messenger c-di-GMP. The diguanylate cyclase activity of MucR and Alg44 binding to c-di-GMP have both been shown to be necessary for alginate production in $P. aeruginosa$\textsuperscript{35,158}.

The environmental signals that lead to the activation of $\sigma_22$ are bypassed in CF-isolated mucoid strains by mutations in the MucA anti-sigma factor. Approximately 84% of clinical isolates from the CF lung carry mutations in the $mucA$ gene\textsuperscript{124}. The most frequently observed mutation is the deletion of a guanine base within a homopolymeric stretch of five guanines located between positions 429 and 433 in the coding sequence. This deletion results in the premature termination of translation leading to the expression of a truncated non-functional form of the MucA protein (Figure 1.7b)\textsuperscript{124}. This allele is referred to as $mucA22$. The mucoid phenotype is unstable under laboratory conditions and cells spontaneously convert to a non-mucoid phenotype by acquiring second-site suppressor mutations that inactivate $\sigma_22$\textsuperscript{143,159}. Experiments by Mathee et al. suggest that oxidative stress in the CF lung promotes the emergence of mucoid variants. Non-mucoid $P. aeruginosa$ biofilms exposed to hydrogen peroxide or activated polymorphonuclear
leukocytes that produced reactive oxygen species lead to the emergence of mucoid bacteria that carried the *mucA*22 allele106.

**1.3.4 Alginate biosynthesis**

The proteins implicated in alginate production are believed to assemble into a large hetero-oligomeric complex that, like other bacterial secretion machinery, spans the entire cell envelope; this hypothesis is supported by the finding that both inner and outer membrane components are needed for alginate polymerization *in vitro*160.

**1.3.4.1 Precursor production**

GDP-mannuronate is the activated precursor of alginate and is synthesized *via* 4 sequential reactions catalyzed by 3 enzymes, AlgA, AlgC and AlgD (*Figure 1.8*)132.

**Step 1: AlgA – Phosphomannose isomerase** - The first step in the production GDP-mannuronate corresponds to the conversion of fructose 6-phosphate to mannose 6-phosphate, a reaction catalyzed by the phosphomannose isomerase (PMI) domain of AlgA161. AlgA is a bifunctional enzyme that catalyzes both PMI and GDP-mannose pyrophosphorylase activities (GMP)162. A combination of mutagenesis and limited proteolysis studies have demonstrated that AlgA is composed of two independent enzymatic domains, where the PMI activity is localized to the C-terminus of the protein and the GMP activity is localized to its N-terminus163.
**Step 2: AlgC – Phosphomannomutase** - Next, mannose 6-phosphate is converted to mannose 1-phosphate by AlgC\(^1\)\(^\text{137}\). AlgC is also involved in the biosynthesis of rhamnolipids and O-antigen of lipopolysaccharide where it acts as a phosphoglucomutase\(^1\)\(^\text{164-166}\). The crystal structure of AlgC revealed that the protein consists of four domains, each consisting of a mixed $\alpha/\beta$ structure, arranged in a “heart”. The active site of AlgC corresponds to a large cleft in the centre of the protein formed by residues from each of the four domains\(^1\)\(^\text{167}\). Several high-resolution structures of AlgC combined with mutagenesis studies have provided substantial mechanistic insight into substrate binding and catalysis of this enzyme\(^1\)\(^\text{167-170}\).

**Step 3: AlgA – GDP-mannose pyrophosphorylase** - The GDP-mannose pyrophosphorylase domain of AlgA catalyzes the third step of precursor production where mannose 1-phosphate and GTP are used to synthesize GDP-mannose\(^1\)\(^\text{162}\).

**Step 4: AlgD – GDP-mannose dehydrogenase** - AlgD catalyzes the committed step of alginate biosynthesis given that GDP-mannuronate’s sole purpose is for alginate production. AlgD is a GDP-mannose dehydrogenase (GMD) that catalyzes the conversion of GDP-mannose to GDP-mannuronic acid\(^1\)\(^\text{171,172}\). The structure of AlgD complexed with its cofactor NAD(H) and reaction product GDP-mannuronic acid showed that AlgD consists of two dinucleotide binding motifs connected by a long $\alpha$-helix where the N-terminal domain binds to NAD(H) and the C-terminal domain binds to the substrate/product\(^1\)\(^\text{173}\).
**Figure 1.8 Proposed alginate biosynthetic pathway.** Proteins responsible for alginate production are coloured in red. The inner and outer membranes are coloured in green and blue, respectively.

1.3.4.2 Alginate polymerization and transport across the inner membrane

Two proteins have been implicated in alginate polymerization and are believed to responsible for transporting the nascent polymer across the inner membrane: Alg8 and
Alg44. Alg8 is believed to catalyze the polymerization of GDP-mannuronate into polyM. Alg8 is an integral inner membrane protein that is essential for alginate polymerization\textsuperscript{160}. Based on alkaline phosphatase fusion experiments, two similar models have been proposed where Alg8 carries four TMs and a large cytoplasmic domain exhibiting sequence similarity to family 2 glycosyltransferases such as the SpsA protein from \textit{Bacillus subtilis}\textsuperscript{160,174,175}. Alanine substitution mutagenesis of putative active site residues confirmed that the cytoplasmic loop is involved in alginate polymerization\textsuperscript{174}. Both inner and outer membrane components are needed for the \textit{in vitro} production of polyM suggesting that Alg8 by itself is insufficient and that a transenvelope multiprotein complex is required for alginate polymerization, modification and secretion\textsuperscript{160}.

Alg44 is thought to act as co-polymerase for Alg8\textsuperscript{174}. Similarly to Alg8, Alg44 mutants do not synthesize or secrete alginate\textsuperscript{35,176}. Alg44 is an integral inner membrane protein with an N-terminal cytoplasmic PilZ domain that binds c-di-GMP followed by a single TM and a C-terminal periplasmic domain with structural similarity HlyD, AcrA and MexA. HlyD, AcrA and MexA are members of the membrane fusion protein family that couple an inner membrane transporter to an outer membrane factor, such as TolC, to form a channel to translocate substrates across the outer membrane\textsuperscript{35,174}. Interestingly, the periplasmic domain of Alg44 is required for alginate polymerization and Alg44 is needed for the stable expression of AlgE, two pieces of data that further support the hypothesis that alginate is produced by a transenvelope multiprotein complex\textsuperscript{174}. The binding of c-di-GMP to the PilZ domain of Alg44 is essential for alginate production suggesting an important role for c-di-GMP in the post-translational regulation of alginate biosynthesis\textsuperscript{35}.  

24
1.3.4.3 Alginate modification

PolyM is thought to be synthesized and transported across the inner membrane into the periplasm by Alg8 and Alg44 where some mannuronate residues are selectively acetylated or epimerized to guluronate, thus producing alginate.

Acetylation of mannuronate residues - Mannuronate residues in polyM can be modified by esterification with O-acetyl groups at the O2 and/or O3 positions\textsuperscript{177}. Acetylation affects the physicochemical properties of alginate such as its viscosity and ability to bind divalent cations, and appears to play a role in microcolony formation, biofilm architecture and immune evasion\textsuperscript{131,178-180}. The concerted action of three proteins, AlgI, AlgJ and AlgF is responsible for alginate acetylation. The deletion of any one of these proteins results in the secretion of non-acetylated alginate, therefore these proteins are not essential for the production of high-molecular-weight polymers\textsuperscript{181,182}. AlgF is a soluble periplasmic protein\textsuperscript{183}. AlgJ is a periplasmic protein that is tethered to the inner membrane by an uncleaved signal peptide lacking a typical signal peptidase cleavage site\textsuperscript{183}. AlgI is an integral inner membrane protein predicted to have seven TMs flanked by several long cytoplasmic and periplasmic domains\textsuperscript{183}. AlgI is thought to be responsible for transporting the O-acetyl groups across the inner membrane into the periplasm for the acetylation of alginate\textsuperscript{184}. AlgI homologs and type II membrane proteins, such as AlgJ, appear to form a gene cassette found in the biosynthetic operon of other bacterial polysaccharides and involved in the esterification of these polysaccharides\textsuperscript{184}. AlgI, AlgJ and AlgF are proposed to assemble into a complex at the
inner membrane and to interact with the polymerization/secretion complex (Figure 1.9).  

![Proposed alginate acetylation complex](image)

**Figure 1.9 Proposed alginate acetylation complex.** This complex consists of AlgI, AlgJ and AlgF and is believed to interact with the rest of the alginate biosynthetic complex. Figure reproduced from Franklin and Ohman (2002) with permission from ASM (DOI: 10.1128/JB.184.11.3000-3007.2002, copyright 2002).

**Epimerization of mannuronate residues** - AlgG is a C5-mannuronan epimerase that catalyzes the conversion of D-mannuronate to L-guluronate. AlgG is a soluble periplasmic enzyme that acts on polyM chains to produce the mixed MG blocks observed in *P. aeruginosa* alginate. Acetylated mannuronate residues are not substrates for AlgG epimerization. The AlgG protein is also needed for the secretion of high-molecular weight alginate, a function unrelated to epimerization. An AlgG point mutant defective in epimerization can still secrete polyM. However, an AlgG deletion mutant secretes degraded alginate. The structure of AlgG from *P. aeruginosa* has not yet been determined, however its C-terminus is predicted to fold into a right-handed β-helix, a
domain characteristic of enzymes that use polysaccharides as substrates. Site-directed mutagenesis studies suggest that a shallow groove on the surface of the \( \beta \)-helix corresponds to the catalytic face for epimerization\(^{187} \).

1.3.4.4 Periplasmic conduit

Three periplasmic proteins encoded by the \( \text{algD} \) operon, AlgG, AlgX and AlgK, have been proposed to assemble into a periplasmic conduit that guides alginate to the predicted outer membrane secretion pore, AlgE\(^{138,186,188} \). AlgK is a putative lipoprotein located in the periplasm\(^{189,190} \). As mentioned earlier, AlgG is a periplasmic C5 mannanuronan epimerase. AlgX is a soluble periplasmic protein of unknown function that exhibits 31\% sequence identity to AlgJ\(^{188,191} \). Deletion mutants of AlgG, AlgX and AlgK secrete low-molecular weight uronic acids which were demonstrated to be AlgL degradation products\(^{186,188,189} \). AlgL is a periplasmic lyase that catalyzes the cleavage of alginate into oligosaccharides. AlgL is also encoded in the \( \text{algD} \) operon and is thereby co-regulated and expressed with the other alginate biosynthetic genes\(^{192} \). AlgL has been implicated in bacterial dispersion from the biofilm and in regulating the length of alginate polymers\(^{52,193,194} \). Recently, AlgL was shown to depolymerize alginate that accumulates in the periplasmic space in mucoid bacteria. In the absence of AlgL, cells quickly accumulate alginate in the periplasm and burst\(^{138} \). Based on this information, Jain and Ohman have proposed a model where AlgG, AlgX, AlgK, AlgL and possibly Alg44 assemble into a conduit that transports alginate across the periplasm (\textbf{Figure 1.10a}); the enzymatic activity of AlgL is believed to be suppressed while associated with the secretion complex. The loss of AlgK, AlgX or AlgG interferes with assembly of the
conduit and hence leads to the degradation of alginate by AlgL and the secretion of low-
molecular weight uronic acids (Figure 1.10b)\textsuperscript{138}. In the absence of AlgL, the conduit fails
to assemble leading to the accumulation of periplasmic alginate and cell death (Figure
1.10c).

Figure 1.10 Proposed alginate biosynthetic complex. (a) Alginate scaffold present
under normal conditions. (b) Absence of AlgK, AlgG or AlgX leads to disruption of
scaffold and degradation of alginate by AlgL (lyase). (c) Absence of AlgL leads
disruption of scaffold and to the accumulation of alginate in the periplasm and cell lysis.
Figure adapted from Jain and Ohman (2005)\textsuperscript{138} with permission from ASM (DOI:

1.3.4.5 Translocation of alginate across the outer membrane
AlgE is an outer membrane protein thought to be responsible for alginate secretion across the outer membrane into the environment\textsuperscript{195,196}. AlgE is a β-barrel protein predicted to have 18 antiparallel strands interspersed by short periplasmic and longer extracellular loops; the recombinant protein has been shown to be exposed on the surface of \textit{E. coli} cells, to insert spontaneously into planar lipid bilayers and detergent micelles, and function as an anion-specific channel\textsuperscript{195,197}.

1.4 Thesis Objective

The focus of this project is to improve our understanding of biofilm formation and exopolysaccharide production in bacteria. We have chosen to study alginate production in \textit{P. aeruginosa} given the important role this exopolysaccharide plays in CF. Although alginate production in \textit{P. aeruginosa} is one of the best understood exopolysaccharide biosynthetic pathways, many questions remain unanswered regarding the polymerization, modification and secretion of this polysaccharide. The systematic structural and functional characterization of each protein implicated in alginate biosynthesis is needed to provide further mechanistic insight into the synthesis and secretion of alginate, as well as the assembly of the alginate biosynthetic complex. AlgK was chosen as a target given, at the beginning of the study, it was a protein of unknown function required for the secretion of high-molecular weight alginate. Thus, the objective of this study was to shed further light on the role of AlgK in alginate biosynthesis. To this end, several studies were conducted. The structure of AlgK was solved using X-ray crystallography. Biological studies were conducted to determine the subcellular localization of AlgK and
investigate how the absence of AlgK affects the localization of the porin AlgE. A bioinformatics analysis and literature search identified proteins involved in poly-β-1,6-GlcNAc and cellulose biosynthesis that may play the same role as AlgK in exopolysaccharide secretion.

1.5 Thesis Organization

The experiments and analyses described in this study are aimed at improving our understanding of the function of AlgK in alginate biosynthesis. CHAPTER 2 presents the crystal structure of AlgK from *P. Fluorescens* WCS374r and discusses several interesting features of the structure. The process and experimental methods to obtain a crystal structure of AlgK are also described in this chapter. Based on the data presented in CHAPTER 2 and APPENDIX 1, as well as what is currently known about AlgK, CHAPTER 3 discusses the role AlgK could be playing in alginate biosynthesis and the differences between the alginate and capsular polysaccharide biosynthetic systems. CHAPTER 3 also discusses interesting features shared between the cellulose, poly-β-1,6-GlcNAc and alginate biosynthetic systems. CHAPTER 4 presents our, at present, unsuccessful attempts to identify AlgK’s binding partners. CHAPTER 5 summarizes the results of the current study and proposes future experiments. APPENDIX 1 describes experiments conducted in our collaborator’s lab to determine the subcellular localization of AlgK, as well as investigate how the absence of AlgK affects the localization of the porin AlgE.
CHAPTER 2

STRUCTURE DETERMINATION OF ALGK BY PROTEIN X-RAY CRYSTALLOGRAPHY

2.1 Overview

Our collaborator demonstrated that AlgK is an outer membrane lipoprotein and it appears to play a role in the localization of AlgE (see APPENDIX 1). AlgK is predicted to be entirely \( \alpha \)-helical and carry four TPR or Sel1-like repeat (SLR) motifs\(^{198} \). In an attempt to gain further insight into the function of AlgK, its structure was determined using protein X-ray crystallography. The crystallization, structure determination and analysis of the structure are presented in this chapter. Initially, the \( P. \) aeruginosa protein was the main crystallization target. Three recombinant versions of the protein were expressed, and a distinct purification protocol was developed for each. \( P. \) aeruginosa AlgK crystals did not diffract beyond 8 Å despite various attempts at optimization. AlgK homologues from other Pseudomonads were therefore expressed, purified and screened for crystallization conditions. AlgK from \( P. \) fluorescens WCS374r yielded diffraction quality crystals that were used to determine the protein structure to a resolution of 2.5 Å. AlgK folds into a flexible solenoid protein carrying 9.5 TPR-like motifs.

Acknowledgements: Jason Koo, while a summer student in the Howell lab, and Patrick Yip, a technician in the Howell lab, generated the \( P. \) aeruginosa AlgK protease-resistant
core and *P. syringae* AlgK expression constructs, and helped with protein expression and crystallization screening. Patrick Yip built the *P. fluorescens* AlgK expression constructs and generated the SeMet triple mutant. Howard Robinson, a beamline scientist at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratories (BNL) measured the native and derivitized diffract data and determined the protein phases. Dr Julianne Kus, a former graduate student in Dr Lori Burrows lab, provided cultures of *P. putida* PP578, *P. syringae pv. maculicola* ES4326 and *P. fluorescens* WCS374r. Dr William J. Page provided a culture of *A. vinelandii*. Dr A. Mirela Neculai, a former postdoctoral fellow in the Howell lab, assisted in protein crystallization and refinement of structure. Dr G. David Smith, an Emeritus Scientist at The Hospital for Sick Children, and Dr. Shao-Yang Ku, a former graduate student in the Howell lab, made available various computer programs and scripts during the course of the structure determination and refinement.

2.2 Materials and Methods

2.2.1 Cloning

The amino acid sequence of AlgK proteins from *P. fluorescens* Pf-5, *P. aeruginosa* PAO1, *P. putida* (strain KT2440) and *P. syringae* pv. tomato obtained from the UniProtKB/TrEMBL database were aligned using the multiple sequence alignment program, T-Coffee (EMBnet Switzerland) and used to determine the signal peptidase II cleavage site in all AlgK homologues. The *algK* gene encoding the mature protein, *i.e.* without its signal sequence, was amplified from the genomic DNA of *P. aeruginosa* PAK, *P. fluorescens* WCS 374r and *P. syringae* pv. *maculicola* ES4326 using gene specific primers designed using the nucleotide sequence from *P. aeruginosa* PAO1, *Pseudomonas fluorescens* Pf-5 and *P. syringae* pv. tomato, respectively, obtained from the UniProtKB/TrEMBL database. PCR amplification products were typically cloned into the pSTBlue-1 acceptor vector, then subcloned into *NcoI*-*XhoI* or *NdeI*-*BamHI* sites of the pET26b or pET28a expression vectors (Table 2.1). The PCR products amplified from the *P. fluorescens* genomic DNA were digested and ligated directly into the expression vectors.

The pET26b-*algK* expression plasmids, referred to as 26K1 throughout the study, encode a modified version of the AlgK protein where its native signal sequence has been replaced by the PelB signal sequence to ensure efficient transport of the recombinant protein into the periplasm of *E. coli* cells. In addition, the N-terminal cysteine of the mature protein was mutated to methionine to prevent protein lipidation, and to facilitate purification it carries a non-cleavable C-terminal hexahistidine (His$_6$)-tag. The 26K2
expression plasmid was built similarly to the 26K1 construct, however it expresses an untagged version of AlgK that is secreted into the periplasm.

The pET28a-\textit{algK} expression plasmids, referred to as 28K1 throughout the study, encode a truncated version of AlgK lacking its signal sequence, therefore this protein remains in the cytoplasm of host cells. The 28K1 recombinant AlgK carries a thrombin-cleavable N-terminal His\textsubscript{6}-tag and similarly to the 26K1 constructs, the 28K1 recombinant AlgK protein carries a C1M mutation. The 26K1\textsubscript{ntag} expression plasmid was produced by cloning the \textit{NcoI-XhoI} fragment from the 28K1 expression plasmid into pET26b plasmid in-frame with the PelB signal sequence. The AlgK protein encoded in 26K1\textsubscript{ntag} plasmid is secreted into the periplasm and carries a thrombin-cleavable N-terminal His\textsubscript{6}-tag (Table 2.1).

**Table 2.1 AlgK expression vectors.** Description of the different expression vectors generated and used in this study. Abbreviations used in this table are defined as follows: SS (signal sequence); thr (thrombin cleavage site); Km\textsuperscript{R} (kanamycin resistance marker).

<table>
<thead>
<tr>
<th>Vector</th>
<th>Description</th>
<th>Recombinant protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET28a</td>
<td>\textit{E. coli} expression vector under control of T7 lac promoter; Km\textsuperscript{R} (Novagen)</td>
<td></td>
</tr>
<tr>
<td>pET26b</td>
<td>\textit{E. coli} expression vector under control of T7 lac promoter; Km\textsuperscript{R} (Novagen)</td>
<td></td>
</tr>
<tr>
<td>p26K1-PA</td>
<td>pET26b with \textit{P. aeruginosa algK}, without its signal sequence and carrying C1M mutation, cloned into the \textit{NcoI-XhoI} site and in-frame with PelB signal sequence and C-terminal His-tag, Km\textsuperscript{R}</td>
<td>PelBSS-AlgKC1M-His\textsubscript{6}</td>
</tr>
<tr>
<td>p26K2-PA</td>
<td>pET26b with \textit{P. aeruginosa algK}, without its signal sequence and carrying C1M mutation, cloned into the \textit{NcoI-XhoI} site and in-frame with PelB signal sequence, Km\textsuperscript{R}</td>
<td>PelBSS-AlgKC1M</td>
</tr>
<tr>
<td>p28K1-PA</td>
<td>pET28a with \textit{P. aeruginosa algK}, without its signal sequence and carrying C1M mutation, cloned into the \textit{NdeI-BamHI} and in-frame with N-terminal His tag, Km\textsuperscript{R}</td>
<td>His\textsubscript{6}-thr-AlgKC1M</td>
</tr>
<tr>
<td>p28K1-PA-Q27</td>
<td>pET28a with \textit{P. aeruginosa algK}, without its signal sequence and residues 1-26 of the mature protein, cloned into the \textit{NdeI-BamHI} and in-frame with N-terminal His tag, Km\textsuperscript{R}</td>
<td>His\textsubscript{6}-thr-AlgKA1−26</td>
</tr>
<tr>
<td>p28K1-PA-Q38</td>
<td>pET28a with \textit{P. aeruginosa algK}, without its signal sequence and residues 1-37 of the mature protein, cloned into the \textit{NdeI-BamHI} and in-frame with N-terminal His tag, Km\textsuperscript{R}</td>
<td>His\textsubscript{6}-thr-AlgKA1−37</td>
</tr>
<tr>
<td>Vector</td>
<td>Description</td>
<td>Expression Tag</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>p28K1-PA-A42</td>
<td>pET28a with <em>P. aeruginosa</em> algK, without its signal sequence and residues 1-41 of the mature protein, cloned into the NdeI-BamHI and in-frame with N-terminal His tag, Km&lt;sup&gt;r&lt;/sup&gt;</td>
<td>His&lt;sub&gt;r&lt;/sub&gt;-thr-AlgKA1–42</td>
</tr>
<tr>
<td>p26K1-PA-ntag</td>
<td>pET26b with Ncol/Xhol fragment from p28K1-PA, encoding the AlgK carrying an N-terminal thrombin-cleavable His-tag, cloned into the <em>NcoI</em>-<em>XhoI</em> site in-frame with the <em>PelB</em> signal sequence</td>
<td>PelBSS- His&lt;sub&gt;r&lt;/sub&gt;-thr-AlgKC1M</td>
</tr>
<tr>
<td>p26K1-PS</td>
<td>pET26b with <em>P. syringae</em> algK, without its signal sequence and carrying C1M mutation, cloned into the <em>Ncol</em>-Xhol site and in-frame with <em>PelB</em> signal sequence and C-terminal His-tag, Km&lt;sup&gt;r&lt;/sup&gt;</td>
<td>PelBSS-C1M-AlgK-His&lt;sub&gt;r&lt;/sub&gt;</td>
</tr>
<tr>
<td>p28K1-PS</td>
<td>pET28a with <em>P. syringae</em> algK, without its signal sequence and carrying C1M mutation, cloned into the NdeI-BamHI and in-frame with N-terminal His tag, Km&lt;sup&gt;r&lt;/sup&gt;</td>
<td>His&lt;sub&gt;r&lt;/sub&gt;-thr-AlgKC1M</td>
</tr>
<tr>
<td>p26K1-PF</td>
<td>pET26b with <em>P. fluorescens</em> algK, without its signal sequence and carrying C1M mutation, cloned into the <em>Ncol</em>-Xhol site and in-frame with <em>PelB</em> signal sequence and C-terminal His-tag, Km&lt;sup&gt;r&lt;/sup&gt;</td>
<td>PelBSS-C1M-AlgK-His&lt;sub&gt;r&lt;/sub&gt;</td>
</tr>
<tr>
<td>p28K1-PF</td>
<td>pET28a with <em>P. aeruginosa</em> algK, without its signal sequence and carrying C1M mutation, cloned into the <em>NdeI</em>-BamHI and in-frame with N-terminal His tag, Km&lt;sup&gt;r&lt;/sup&gt;</td>
<td>His&lt;sub&gt;r&lt;/sub&gt;-thr-AlgKC1M</td>
</tr>
<tr>
<td>p26K1-PF-M3</td>
<td>P26K1-PA encoding AlgK carrying the following mutations: L32M, I44M and L399M, Km&lt;sup&gt;r&lt;/sup&gt;</td>
<td>PelBSS-C1M-AlgK-L32M-I44M-L399M-His&lt;sub&gt;r&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

### 2.2.2 Protein expression

*E. coli* BL21 CodonPlus(DE3) cells (Stratagene) transformed with one of the AlgK expression vectors were grown at 37 °C in 1 l of Luria-Bertani medium (LB) containing kanamycin and chloramphenicol to a final concentration of 50 and 35 µM, respectively. The cells were cultured until their OD<sub>600</sub> attained 0.6, at which point AlgK expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.6 mM. After induction, the cells were grown for 15 hrs at 25 °C and subsequently harvested by centrifugation at 5,350 g for 10 min. The pellets were cooled on ice for 5 min prior to purification.

### 2.2.3 Protein purification
The purification protocols described in sections 2.2.3.1 and 2.2.3.2 were used for all 28K1 and 26K1 proteins, respectively. While developing these purification protocols, it became apparent that AlgK was highly susceptible to trace proteases. To prevent this degradation, β-mercaptoethanol, EDTA-free Roche Complete Protease Inhibitor (PI) cocktail and ethylenediaminetetraacetic acid (EDTA) were added to different steps in the purification process. In the presence of these proteases inhibitors a small amount of degraded AlgK could still be seen on an SDS-PAGE gel, however the degradation did not increase with time.

2.2.3.1 28K1 purification

All AlgK proteins expressed from the 28K1 plasmids were purified using the following protocol. The cell pellet was resuspended in buffer A (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol, 1 tablet of PI cocktail, 1 mM phenylmethylsulphonyl fluoride (PMSF)) and lysed by four sonication cycles of 30 sec pulses followed by 30 sec of cooling on ice. Cell lysates were centrifuged at 23,700 g for 30 min at 4 °C and the cleared cell lysates were run over nickel-nitrilotriacetic acid (Ni-NTA) resin pre-equilibrated with buffer B (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 mM imidazole). The His-tagged 28K1 protein bound to the Ni-NTA resin was washed with 100 ml of buffer C (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 20 mM imidazole, ½ tablet of PI cocktail, 5 mM β-mercaptoethanol, 1 mM PMSF). The 28K1 protein was eluted from the Ni-NTA resin with 25ml of buffer D (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 100 mM imidazole, 1/20 tablet of PI cocktail, 5 mM β-mercaptoethanol, 1 mM PMSF) and buffer-exchanged overnight at 4 °C in 2 l of buffer E (20 mM Tris-HCl pH
7.5, 150 mM NaCl, 10 mM imidazole). The following day, the N-terminal His-tag was cleaved by adding approximately one unit of thrombin (Sigma) for every 5-10 mg of protein in solution and incubating the digestion at 21 °C for ~6 hrs at which point the reaction was stopped by the addition of PMSF to a final concentration of 1 mM. The protein was incubated with Ni-NTA resin pre-equilibrated with buffer B then washed with buffer B to elute the untagged protein. β-mercaptoethanol and EDTA were added to the flow-through to a final concentration of 5 mM and 0.2 mM, respectively. The untagged protein was concentrated, and further purified and buffer exchanged into buffer F (20 mM Tris-HCl pH 8.5, 20 mM NaCl, 0.2 mM EDTA, 5 mM β-mercaptoethanol) by size-exclusion chromatography using a Superdex200 HR 10/30 gel filtration column (Pharmacia). The protein was concentrated using a concentrator with a 30 kDa molecular weight cut-off (Amicon) and screened for crystallization conditions.

2.2.3.2 26K1 purification

All AlgK proteins expressed from the 26K1 plasmids were purified using the following protocol. The first step of purification involves the release of the periplasmic contents by osmotic shock. The procedure detailed below was adapted from Sambrook and Russell201. Briefly, the cell pellet for 1 l of bacterial culture was washed with 100 ml of buffer G (10 mM Tris-HCl pH 7.5, 30 mM NaCl) before resuspending the pellet in 50 ml of buffer H (30 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 20% w/v sucrose, 1 mM PMSF) and mixed at 4 °C for 10 min. Cells were then centrifuged for 20 min at 11,300 g and the resulting pellets were resuspended in 75 ml of buffer I (0.1 mM MgCl₂, 1 mM PMSF, 1 tablet of PI cocktail and 15 mM β-mercaptoethanol) and mixed for 1 h at 4 °C.
Subsequently, the cells were centrifuged for 20 min at 11,300 \textit{g} after which the supernatant was collected and centrifuged for a further 30 min at 20,400 \textit{g}. The final supernatant, \textit{i.e.} the periplasmic content, was dialyzed overnight against 4 l of buffer B at 4 °C. The protein was purified by mixing the periplasmic contents with 10 ml Ni-NTA agarose (Qiagen) pre-equilibrated with buffer B for one hr at 4 °C. The protein bound-Ni-NTA resin was washed four times with 35 ml of buffer C using the batch method. During the final wash step, the resin was loaded into a column and the buffer was collected and discarded. The His-tagged protein was eluted by washing the resin with 30 ml of buffer J (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 150 mM imidazole, 1 mM PMSF, 5 mM β-mercaptoethanol, \(\frac{1}{2}\) tablet of PI cocktail). The eluted protein was diluted to 5-10 mg/ml with buffer F and further purified and buffer exchanged into buffer F by size-exclusion chromatography using a Superdex200 HR 10/30 gel filtration column (Pharmacia). The protein was concentrated using a concentrator with a 30 kDa molecular weight cut-off (Amicon) and screened for crystallization conditions.

\textit{2.2.3.3 Purification of 26K1ntag}

The osmotic shock and Ni-NTA protocols described in the previous section were also used as the first and second steps, respectively, in the purification of the 26K1ntag protein with the following modifications: the PI cocktail and PMSF were omitted from all buffers during nickel affinity purification. The protein eluted from the nickel column was dialyzed overnight at 4 °C against buffer B. The following day, the N-terminal His-tag was cleaved by adding approximately one unit of thrombin (Sigma) for every 5-10 mg of protein in solution and incubating the digestion at 21 °C for \(~15\) hrs at which point the
reaction was stopped by the addition of PMSF to a final concentration of 1 mM. The protein was incubated with Ni-NTA resin pre-equilibrated with buffer B then washed with buffer B to elute the untagged protein. β-mercaptoethanol and EDTA were added to the flow-through to a final concentration of 5 mM and 0.2 mM, respectively. The untagged protein was concentrated and further purified and buffer exchanged into buffer F by size-exclusion chromatography using a Superdex200 HR 10/30 gel filtration column (Pharmacia). The protein was concentrated using a concentrator with a 30 kDa molecular weight cut-off (Amicon) and screened for crystallization conditions.

2.2.4 Limited proteolysis studies

Purified *P. aeruginosa* 28K1 and 26K1 AlgK at a concentration of 1.5 mg/ml were incubated with chymotrypsin at a 1000:1 Molar ratio at 4 °C in buffer K (5% glycerol, 150 mM NaCl, 20 mM Tris-HCl pH 7.5) for 24 hrs. At various time points throughout the digestion, a small aliquot of protein was taken from the digestion, mixed with an equal volume of 2x SDS loading dye and stored at -20 °C. The following day, the frozen samples were thawed and separated on a 12% SDS-PAGE gel.

2.2.5 Western blots probing for His-tag on chymotrypsin-digested 26K1-PA and 28K1-PA

26K1-PA and 28K1-PA protease-resistant cores were separated on a 12% SDS-PAGE, then transferred to a PVDF membrane (Amersham Hybond-P, GE Healthcare). The membrane was blocked for 1 hr at 21 °C in Tris-buffered saline (TBS, 25 mM Tris-HCl pH 7.4, 2.7 mM KCl, 137 mM NaCl) containing 0.05% (v/v) Tween-20 and 5%
(w/v) skim milk powder, then incubated with a mouse anti-His antibody (Bio-Rad) overnight at 4 °C in TBS containing 0.05% (v/v) Tween-20 and 1% (w/v) skim milk powder at a dilution of 1:3000. The following day, the membranes were washed three times for 10 min in TBS containing 0.05% Tween-20 (TBST) and then incubated with a goat anti-mouse IgG-alkaline phosphatase conjugate (Bio-Rad) diluted 1:3000 in TBST containing 1% (w/v) skim milk powder for 1 hr at room temperature. The blots were subsequently washed twice in TBST for 10 min followed by a 5 min wash in TBS. The blots were incubated in Nitro-Blue Tetrazolium Chloride (NBT)/5-Bromo-4-Chloro-3'-Indolyphosphate p-Toluidine Salt (BCIP) substrate (Thermo Scientific) and developed for 3 min.

2.2.6 Purification of chymotrypsin-digested proteins

2.2.6.1 Purification of chymotrypsin-digested 26K1-PA

After nickel affinity purification, the 26K1 protein was digested for 7 hrs with chymotrypsin at a 1000:1 Molar ratio at 4 °C in buffer L (20 mM Tris-HCl pH 7.5, 150 mM NaCl). The digestion was stopped with PSMF added to a final concentration 1 mM. The 26K1 protease-resistant protein was further purified and buffer-exchanged using a Superdex200 HR 10/30 gel filtration column (Pharmacia) into buffer F. The peak fractions containing the least amount of undigested protein were pooled, concentrated using a concentrator with a 30 kDa molecular weight cut-off (Amicon) and screened for crystallization conditions.
2.2.6.2 Purification of chymotrypsin-digested 28K1-PA

After nickel affinity purification, the 28K1 protein was digested with chymotrypsin at a 1000:1 Molar ratio at 4 °C in buffer M (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM imidazole) and the reaction was stopped after 2 hrs with PMSF added to a final concentration of 1 mM. Chymotrypsin cleaved the N-terminus of the protein thereby removing the His-tag. The digestion mixture was run-over a Ni-NTA column pre-equilibrated with buffer B to remove the undigested tagged protein. The flow-through from the nickel column contained untagged protease-resistant 28K1 protein that was dialyzed into buffer N (20 mM Tris-HCl pH 7.5, 50 mM NaCl and 0.2 mM EDTA), then concentrated to 5 mg/ml and loaded onto a Q-sepharose column (Amersham Biosciences). The protein was eluted by applying a linear gradient from 0.05 to 1 M NaCl in 20 mM Tris-HCl pH 7.5. The peak fractions were pooled, concentrated using a concentrator with a 30 kDa molecular weight cut-off (Amicon), then further purified and buffer exchanged by gel filtration into buffer F. The purified protein was concentrated and screened for crystallization conditions.

2.2.7 Crystallization of 26K1-PF from P. fluorescens

The crystals used for diffraction studies were grown in condition 33 (10% w/v polyethylene glycol (PEG) 6000, 0.1 M 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.0) of the pHClear I Suite (Qiagen) using the hanging-drop vapor diffusion method. Crystals took approximately 3-6 weeks to grow. The 6 µl drop consisted of a 1:1 mixture of protein and well solution hanging above 1 ml of well solution. The trays were incubated at 21 °C. A detergent screen (Hampton Research) showed that the direct
addition of 0.4 µl of 5.6 mM CYMAL®-6 to a 3.6 µl drop containing a 1:1 ratio of protein to mother liquor to protein produced more singular crystals. Drops containing protein and detergent were hung above wells containing 0.6 ml of well solution without detergent.

2.2.8 Native data collection

Prior to data collection the crystals were cryo-protected using buffer O (10% v/v PEG6000, 0.1 M MES pH 6.0, 25% v/v glycerol and 0.1% w/v hyaluronate). The cryoprotected crystals were immediately vitrified in either liquid nitrogen or in a stream of nitrogen gas at -172 °C and either tested and/or stored. There was a considerable variation in the quality of the crystals and extensive testing of the crystals was conducted on our home source (R-AXIS IV++ image plate detector using Cu Kα X-ray radiation from an RU-H3R rotating anode generator) in order to find suitable candidates for data collection at the NSLS (BNL). A complete set of native data were collected on Station X29 (NSLS) at -172 °C. A total of 360 images of 1° ∆φ-oscillations were collected, i.e. 360 ° of data, on an ADSC Quantum-315 detector with a 300 mm crystal to detector distance and an exposure time of 4 sec per image. The data was integrated, reduced and scaled using XDS202.

2.2.9 Design and generation of selenomethionine (SeMet) triple mutant

The multiple sequence alignment of the AlgK orthologues was generated using T-Coffee200. The three mutations, L32M, I44M, and L399M were introduced sequentially
into the AlgK pET26b expression vector using the QuickChange® Site-Directed Mutagenesis strategy (Stratagene) and verified by DNA sequencing.

2.2.10 Expression, purification, crystallization and X-ray analysis of SeMet triple mutant

BL21 CodonPlus (DE3) cells (Stratagene) transformed with the p26K1-PF-M3 expression plasmid were grown at 37 °C in 1 l of LB containing kanamycin and chloramphenicol to final concentrations of 50 and 35 µM, respectively. When the OD600 of the culture reached 0.9, 140 ml of the culture was centrifuged at approximately 3,000 g for 15 min. The pellet was rinsed with M9 salts and resuspended in 1 l of M9 salts supplemented with 40 µg/ml L-amino acid (all amino acids except L-methionine), 0.4% w/v glucose, 2 mM MgSO4, 25 µg/ml FeSO4.7H2O, 1 µg/ml riboflavin, 1 µg/ml niacinamide, 1 µg/ml pyridoxine monohydrochloride, 1 µg/ml thiamine, 50 µg/ml kanamycin, 35 µg/ml chloramphenicol and 40 µg/ml L-SeMet203. The cells were grown to an OD600 of 0.6, then induced with IPTG added to a final concentration of 0.5 mM and grown overnight at 25 °C. The SeMet protein was purified using the same protocol as the native protein with the following modifications: 1) 2 mM dithiothreitol (DTT) was used instead of 15 mM β-mercaptoethanol in buffer I (0.1 mM MgCl2, 1 mM PMSF, 1 tablet of PI cocktail); 2) the protein-bound Ni-NTA resin was washed with 75 ml instead of 140 ml of buffer C; and 3) the SeMet protein eluted from the nickel column was concentrated to 20 mg/ml and exchanged into buffer F using a concentrator with a 30 kDa molecular weight cut-off (Amicon) prior to crystallization. Crystals of the triple mutant were grown using the same conditions as the native protein (10% (w/v) PEG6000, 0.1 mM MES pH
6.0) but with a protein concentration of 20 mg/ml and without detergent. Prior to data collection at –172 °C the crystals were cryo-protected using buffer O. Single anomalous dispersion (SAD) data were measured at a wavelength of 0.9792 Å at beam line X25, NSLS. To obtain higher resolution data, a second data set was collected on the same crystal at beam line X29, NSLS. The two datasets were processed and merged using the HKL data processing system v1.98.5\textsuperscript{204}.

2.2.11 Structure solution and model refinement

The positions of 22 seleniums in the asymmetric unit were found using Hkl2map v0.2 and were input into Solve/Resolve v2.11 to solve the protein phases\textsuperscript{205-207}. The initial model of AlgK built into the experimental map consisted of residues 19–34 and 69–314 and the non-crystallographic symmetry (NCS) operators were used to generate the complete contents of the asymmetric unit. Initial rounds of model refinement consisted of rigid body refinement, torsion angle simulated annealing at a starting temperature of 2500 K and applying a bulk solvent correction, and grouped B-factor refinement. Later rounds consisted of conjugate gradient minimization and individual B-factor refinement. NCS restraints were applied to all residues in the initial model and in subsequent rounds of refinement to residues 108–301. After several rounds of refinement when the R-factor attained 30%, the NCS restraints were released and waters, glycerols and chloride ions were added. The criteria used to define positive density in the $F_o - F_c$ maps as a water molecule were the following: the peak has a minimum sigma of 2 and is located between 2.8 to 3.4 Å away from a hydrogen bond donor or acceptor. The parameter and topology files for glycerol were obtained from the HIC-Up server\textsuperscript{208}. 
Procheck v3.5.4, ADIT, amide_check v1.5 (Dr G. David Smith, Hospital for Sick Children) and Molprobity were used to during the model building process and to validate the final structure²⁰⁹-²¹².

2.2.12 Superposition and comparison of the AlgK molecules present in the asymmetric unit

Each molecule in the asymmetric unit was superimposed in a pair wise fashion and Cα r.m.s. displacements for each residue were calculated using PROFIT v6.01 (Dr. G. D. Smith, Hospital for Sick Children).

2.2.13 Identification and characterization of TPRs in the AlgK structure

SMART was used to obtain the scores and boundaries of the four predicted TPR/SLRs in AlgK²¹³. Repeat 9 was identified as the highest scoring putative SLR. To identify additional TPR/SLRs within AlgK, besides those identified by SMART, DaliLite was used to superimpose repeat 9 onto the other pairs of anti-parallel α-helices to help identify the canonical TPR residues²¹⁴. PROMOTIF v3.01 (http://www.rubic.rdg.ac.uk/~gail/#Software) was used to calculate helical packing angles and define secondary structural elements²¹⁵. Using the primary sequence of AlgK, secondary structure elements were predicted using PSIPRED v2.6.

2.2.14 Identification of structural homologues and superposition with AlgK model

AlgK molecule C was input as the search query into the DALI webserver to identify structural homologues²¹⁶. Molecule C was used because it is the most well-
ordered and complete molecule in the asymmetric unit. DaliLite was used to superimpose AlgK onto the structural homologues.

2.3 Results

2.3.1 Crystallization of *P. aeruginosa* AlgK

At the beginning of this project, the primary crystallization target was AlgK from *P. aeruginosa*. AlgK is a periplasmic lipoprotein tethered to the outer membrane by a N-terminal N-acyl-diacylglycerolcysteine. The presence of two additional cysteines suggests AlgK may require the formation of a disulfide bond for proper folding in the periplasm. With this information in mind, three different constructs were built, all of which expressed a recombinant version of AlgK lacking its native signal sequence and with its putative lipidation site mutated to methionine. All expression vectors generated for this study as well as the proteins they express are listed and described in Table 2.1.

2.3.1.1 28K1-PA construct and recombinant protein

The p28K1-PA vector was generated by cloning *P. aeruginosa* *algK* into the pET28a expression vector. The AlgK protein expressed from the p28K1 plasmid, *i.e.* the 28K1-PA protein, lacks its native signal sequence, remains in the cytoplasm and carries a thrombin-cleavable N-terminal His$_6$-tag to facilitate purification. After induction, cells were lysed by sonication and the protein was purified from the whole cell lysate by nickel affinity chromatography. Subsequently, the His-tag was removed and the protein was
further purified by size exclusion chromatography (Figure 2.1). Up to 15 mg of purified protein could be obtained from each litre of bacterial culture.

Figure 2.1 Purification and digestion of 28K1-PA protein. Protein samples taken from different steps of the 28K1-PA purification and His-tag cleavage procedure were separated on a 12% SDS-PAGE gel: lane 1, molecular weight markers (kDa); lane 2, 28K1-PA protein eluted from the nickel column; lane 3, 28K1-PA protein incubated with thrombin for 1 hr; lane 4, 2 hrs; lane 5, 3 hrs; lane 6, 4 hrs; lane 7, 4.5 hrs; lane 8, untagged 28K1-PA protein after removal of residual tagged protein by nickel affinity chromatography; lanes 9-11, fractions of 28K1-PA protein eluted from gel filtration column.

2.3.1.2 26K1-PA construct and recombinant protein

For the 26K1-PA construct, *P. aeruginosa algK* was cloned in the pET26b expression vector, downstream and in-frame with the PelB signal sequence, thus producing a recombinant version of AlgK, referred to as the 26K1-PA protein, carrying the PelB signal sequence to ensure its secretion into the periplasm and an uncleavable 8-residue C-terminal His6-tag. The first step of purification was the release of the periplasmic contents by osmotic shock followed by nickel affinity and gel filtration.
chromatography to purify 26K1-PA from the periplasmic contents (Figure 2.2). Up to 15 mg of purified protein could be obtained from each litre of bacterial culture.

![Figure 2.2 Purified 26K1-PA protein.](image)

**Figure 2.2 Purified 26K1-PA protein.** Protein samples taken from the final step of purification and analyzed on a 12% SDS-PAGE gel: lane 1, molecular weight markers (kDa); lanes 2-4, peak elution fractions after gel filtration of the 26K1-PA protein.

### 2.3.1.3 26K1ntag-PA construct and recombinant protein

The 26K1ntag-PA construct was built by cloning the gene encoding the N-terminally His-tagged 28K1-PA protein into the pET26b expression vector in-frame with the PeLB signal sequence. The 26K1ntag-PA protein carries the PeLB signal sequence leading to its secretion into the periplasm and a thrombin cleavable N-terminal His-tag. The purification protocol is similar to that described for 26K1-PA protein, however after nickel affinity chromatography the His-tag is removed before being run on a gel filtration column, thereby producing an untagged periplasmic version of AlgK for crystallization.
screening (Figure 2.3). Up to 5 mg of purified cleaved protein was obtained from each litre of bacterial culture.

![Figure 2.3 Purified untagged 26K1ntag-PA protein.](image)

Protein samples taken from the final step of purification and analyzed on a 12% SDS-PAGE gel: lane 1, molecular weight markers (kDa); lanes 2-7, peak elution fractions after gel filtration of the 26K1ntag-PA protein.

Untagged cytoplasmic, *i.e.* 28K1-PA, C-tagged periplasmic, *i.e.* 26K1-PA, and untagged periplasmic, *i.e.* 26K1ntag-PA, versions of AlgK were concentrated to between 5-20 mg/ml and screened for crystallization conditions using commonly available commercial screens. A variety of crystals grew, the most common of which were stacks of thin plates in conditions where polyethylene glycol was the main precipitant and the pH ranged between 6.0 and 8.5 (Figure 2.5). None of the crystals diffracted further than 8 Å and were recalcitrant to optimization.
2.3.1.4 Production of a protease-resistant core of AlgK

The susceptibility of AlgK to trace proteases observed during the development of the purification protocols, discussed in section 2.2.3, suggested that the protein may harbour flexible regions that were interfering with crystal packing. Limited proteolysis studies were performed to identify a protease-resistant core of AlgK. When incubated with chymotrypsin for several hours at 4 °C, both the periplasmic 26K1-PA and tagged cytoplasmic 28K1-PA versions of AlgK gave a similar pattern of degradation (Figure 2.6). Western blots probing for the His-tag revealed that the N-terminus of the protein had been cleaved by chymotrypsin as the C-terminal His-tag on the 26K1-PA protein was detected, but the N-terminal His-tag on the 28K1-PA protein was not (data not shown). N-terminal sequencing of the 26K1-PA chymotrypsin-resistant fragments, which range in size from 45-47 kDa, suggested that cleavage was occurring at multiple sites, C-terminal
to residues Y26, Y34 and L41 of mature AlgK, where the lipidated cysteine corresponds to position 1.

Figure 2.6 Limited proteolysis studies of 28K1-PA and 26K1-PA AlgK. 12% SDS-PAGE gel of tagged 28K1 (upper panel) and 26K1 (lower panel) AlgK incubated with chymotrypsin for varying lengths of time. Lane 1 (M) contains the molecular weight markers (kDa).

The chymotrypsin-digested periplasmic and cytoplasmic versions of AlgK were purified and screened for crystallization conditions (Figure 2.7a and b), however no diffraction quality crystals were obtained. Three versions protease resistant core, starting at residue Q27, Q38 and A42 (Table 2.1), were expressed, purified and screened for
crystallization conditions (**Figure 2.7c, d and e**). None of the truncated proteins yielded diffraction quality crystals.

**Figure 2.7 Purification of digested and cloned protease resistant cores.** Each panel corresponds to a 12% SDS-PAGE gel onto which were loaded the peak fractions of: (a) chymotrypsin-digested 26K1-PA protein; (b) chymotrypsin-digested 28K1-PA; (c) 28K1-PA-Q27; (d) 28K1-PA-Q38 and (e) 28K1-PA-A42 after gel filtration chromatography. In all panels, lane 1 carries the molecular weight marker (kDa).

### 2.3.2 Crystallization of AlgK orthologues

AlgK orthologues were expressed, purified and screened for crystallization conditions. The sequence identity between AlgK from *P. aeruginosa* PAO1 and other Pseudomonads as well as *A. vinelandii* ranges between 55-60%. *P. putida* PP578, *P. syringae* pv. *maculicola* ES4326, *A. vinelandii* and *P. fluorescens* WCS374r cultures were obtained and their genomic DNA extracted. The *algK* genes from *P. syringae* and
*P. fluorescens* were successfully cloned into expression vectors. The *algK* gene lacking its signal sequence and lipidation site was cloned into the pET28a and pET26b vectors producing a cytoplasmic protein carrying a cleavable N-terminal His-tag equivalent to the 28K1-PA protein, as well as a periplasmic protein carrying a C-terminal His-tag equivalent to 26K1-PA protein, respectively (Table 2.1). These recombinant proteins were expressed and purified using the same protocols described above for the *P. aeruginosa* AlgK proteins.

2.3.2.1 *P. syringae* AlgK

The *P. syringae* proteins, i.e. 26K1-PS and 28K1-PS, expressed well as 1 l of bacterial culture yielded up to 30 mg of purified protein (Figure 2.8). The proteins could be concentrated up to 50 mg/ml without any visible precipitation and dynamic light scattering studies suggested the 26K1-PS protein was monodisperse in solution (data not shown). The purified 26K1-PS and 28K1-PS proteins were concentrated to 10-30 mg/ml and screened for crystallization conditions, however no protein crystals were obtained.

**Figure 2.8 Purified 26K1-PS protein.** Protein samples taken from the final step of purification and analyzed on a 12% SDS-PAGE gel: lane 1, molecular weight markers (kDa); lanes 2-5, peak fractions after gel filtration of the 26K1-PS protein.
2.3.2.2 *P. fluorescens* AlgK

The periplasmic C-terminally tagged protein 26K1-PF was the first of the two *P. fluorescens* recombinant proteins to be expressed and purified. After secretion into the periplasm, the protein has 452 residues including a 2 residue linker and a His$_6$-tag and is predicted to have a molecular weight of 49,561 kDa$^{217}$. The periplasmic *P. fluorescens* protein expressed well and up to 20 mg of purified protein could be purified from each litre of bacterial culture (Figure 2.9). The purified protein was concentrated to 30 mg/ml and screened for crystallization conditions.

![Figure 2.9 Purification of 26K1-PF protein](image)

**Figure 2.9 Purification of 26K1-PF protein.** Protein samples taken from different steps through-out the purification of 26K1-PF protein and analyzed by 12% SDS-PAGE gel: lane 1, molecular weight markers (kDa); lane 2, crude cell lysates after induction; lane 3, periplasmic extract; lane 4, purified AlgK after Ni-affinity column chromatography; lane 5, AlgK after gel filtration chromatography. Figure adapted from Keiski *et al.* (2007)$^{218}$ with permission from IUCr (copyright 2007).

Crystals were produced in several conditions, most of which contained PEG as the precipitant and grew at a pH between 6.0 and 7.5 (Figure 2.10). The most promising looking crystals were grown in condition 33 (10% PEG 6000 w/v, 0.1 M MES pH 6.0)
from the pHClear I Suite (Figure 2.10a). These crystals grew as flat plates in clumps and were very fragile; acupuncture needles were needed to carefully pry apart the crystals and manipulate them. Attempts to optimize the crystals using e.g. different buffers, different PEGs and different concentrations, additive and detergent screens, methylation, streak seeding (Figure 2.10c) and crystallization of the untagged cytoplasmic version showed that the direct addition of CYMAL®-6 to the crystallization drop produced more singular crystals (Figure 2.10b). During cryoprotection, the crystals, especially the larger ones, had a tendency to crack and break into pieces. Hyaluronate was added to help stabilize the crystals\textsuperscript{219}. To avoid damaging the crystals, they were cryoprotected in the drop in which they were grown by sequential addition of the glycerol cryoprotectant. Initially all but a few microlitres of the mother liquor in the drop were removed and replaced with 7 \( \mu l \) of 10\% v/v PEG6000, 0.1 M MES pH 6.0, 0.1\% w/v hyaluronate and 2.5\% v/v glycerol. This washing procedure was repeated several times, each time increasing the glycerol concentration by 2.5\% until a final concentration of 25\% v/v glycerol was attained. The cryoprotected crystals were immediately vitrified in liquid nitrogen or in a cold stream at -172 °C and either tested and/or stored. These crystals were sent to the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratories (BNL) where a complete set of native data was collected to 2.5 Å resolution on Station X29. The crystal belonged to space group P2\(_1\) with unit cell dimensions \( a = 79.01 \text{ Å}, b = 107.74 \text{ Å}, c = 118.91 \text{ Å}, \beta = 96.97 \text{ °} \). The data collection statistics are summarized in Table 2.2. On the basis of density calculations (\( V_m = 2.53 \text{ Å}^3\text{Da}^{-1} \)) we estimated that four molecules of the protein were present in the asymmetric unit\textsuperscript{220}.
2.3.3 Phasing

Recombinant AlgK from *P. fluorescens* WCS 374r contains 4 methionines in 452 residues, not including the N-terminal methionine, and therefore to increase the phasing
power obtainable in a SeMet SAD experiment, a methionine triple mutant was generated. A multiple sequence alignment of AlgK homologues from *P. fluorescens* WCS 374r, *P. aeruginosa* PAO1, *P. putida* KT2440, *P. syringae* pv. tomato and *Azotobacter vinelandii* was generated and used to determine which residues would be the best candidates for mutagenesis. Residues L32, I44 and L399 were chosen as these positions are methionines in at least one of the other homologues, and because leucine to methionine mutants have been shown to have the smallest isomorphous difference with the wild type protein\(^{221}\). A triple mutant was expressed, purified, crystallized under similar conditions as the native protein and sent to NSLS for data collection (Figure 2.10d). The crystal belonged to space group P2\(_1\) with unit cell dimensions \(a = 78.22\) Å, \(b = 107.03\) Å, \(c = 120.13\) Å, \(\beta = 96.097^\circ\). Two datasets were measured at a wavelength of 0.9792 Å, the selenium peak, one at beam line X25 (NSLS) and the second at beam line X29 (NSLS). The two datasets were processed and merged using the HKL data processing system v1.98.5\(^{204}\). The data collection statistics are summarized in Table 2.2

### 2.3.4 Structure solution and refinement

The crystal structure of AlgK was solved using the SAD method. The positions of 22 seleniums in the asymmetric unit were found using Hkl2map v0.2 and input into Solve/Resolve v2.11 to determine the protein phases, to improve them by density modification, as well as to generate a 2.8 Å non-crystallographic symmetry (NCS) - averaged electron density map\(^{205-207}\). The initial model of AlgK was built into the experimental map then refined against the native 2.5 Å resolution data. The model was refined by iterative rounds of manual rebuilding in Coot followed by refinement in CNS
v1.1\textsuperscript{222}. The progress of the refinement was monitored by reductions in $R_{\text{work}}$ and $R_{\text{free}}$, and $\sigma_A$-weighted $2F_o - F_c$ and $F_o - F_c$ maps as implemented in CNS were carefully examined at the end of each round of refinement\textsuperscript{223}. The refinement statistics are summarized in Table 2.2. The final $R_{\text{work}}$ and $R_{\text{free}}$ were 22 and 28%, respectively. According to Procheck, 91.8% of the residues lie in most favored region of Ramachandran plot with 8.1% in the additional allowed region, and 0.1% in generously allowed region. No residues lie in the disallowed regions\textsuperscript{210}. The quality of the electron density prevented several regions of residues from being built (Table 2.3).

Table 2.2 X-ray data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>SeMet-SAD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2\textsubscript{1}</td>
<td>P2\textsubscript{1}</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>79.09, 107.82, 119.03</td>
<td>78.22, 107.03 120.13</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (°)</td>
<td>90.00, 96.97, 90.00</td>
<td>90.00, 96.097, 90.00</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0 - 2.50 (2.60 – 2.50)</td>
<td>30.00 – 2.80 (2.90 - 2.80)</td>
</tr>
<tr>
<td>$I/\sigma I$</td>
<td>0.102 (0.546)</td>
<td>0.109 (0.694)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100)</td>
<td>99.2 (98.6)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.49 (7.53)</td>
<td>14.4 (10.9)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>49.05 – 2.50</td>
<td></td>
</tr>
<tr>
<td>No. reflections</td>
<td>68663</td>
<td></td>
</tr>
<tr>
<td>$R_{\text{work}}$ / $R_{\text{free}}$</td>
<td>0.22 / 0.28</td>
<td></td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteins</td>
<td>11461</td>
<td></td>
</tr>
<tr>
<td>Glycerol</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Chloride</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>853</td>
<td></td>
</tr>
<tr>
<td>Av. B-factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteins</td>
<td>48.7</td>
<td></td>
</tr>
<tr>
<td>Glycerol</td>
<td>62.4</td>
<td></td>
</tr>
<tr>
<td>Chloride</td>
<td>66.6</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>49.4</td>
<td></td>
</tr>
<tr>
<td>R.m.s deviations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bond lengths (Å) 0.010
Bond angles (°) 1.5

| Table 2.3 Summary of completeness of AlgK molecules in the asymmetric unit |
|---------------------------------|-------------|-------------|-------------|-------------|
|                                 | Molecule A | Molecule B | Molecule C | Molecule D |
| (a) Residues                   | 1–5        | 1–11       | 1–15       | 1–21       |
| with no or uninterpretable     | 43–70      | 45–53      | 46–53      | 42–128     |
| density and hence not modeled  | 83–84      | 388–389    | 402–408    | 189–190    |

2.3.5 General overview of structure

The crystallographic asymmetric unit contains four molecules of AlgK, arranged as a dimer of dimers. The small size of the dimer interface (~800–900 Å²/molecule) in comparison to true homodimers (~2,000 Å²/molecule) coupled with our analytical ultracentrifugation data (data not shown), which showed that AlgK is predominantly monomeric in solution suggests that the arrangement of AlgK molecules seen in the unit cell is not biologically relevant.\(^{224}\)

The most well-ordered and complete molecule of AlgK in the asymmetric unit, molecule C, contains 22 α-helices that pack into a right-handed superhelix that is approximately 70 Å long and 50 Å wide with an internal diameter of 20–30 Å (Figure 2.11 and Figure 2.12). At the N-terminus, α-helix H1 (residues 17–32) and residues 33–
45, which adopt an extended conformation, pack against the concave surface of the superhelix (Figure 2.11b). The body of the superhelix is composed of 21 anti-parallel \( \alpha \)-helices with residues 55–384 forming 9.5 TPR-like motifs (see below, repeats R1 to R10, Figure 2.11c and Figure 2.12). The exposed hydrophobic surface of the last TPR-like motif, R10, is shielded from the solvent by its interaction with helix H22 (residues 390–400). At the C-terminus, residues 410–422 form helix H23, which packs against the convex surface of the superhelix. A disulfide bond between adjacent \( \alpha \)-helices in repeats 4 and 5, helices R4B and R5A, is formed between C168 and C178.

2.3.6 Conformational Flexibility of AlgK

The four molecules of AlgK present in the crystallographic asymmetric unit exhibit significant conformational heterogeneity. This was apparent in the initial experimental electron density map and subsequently during the refinement. The maps revealed that the quality of the density varied both across a single molecule and between molecules, with the result being that only \(~90\%\) of the residues in molecules A, B and C, and 70\% of the residues in molecule D could be modeled. The major differences between molecules A and D, and B and C is the absence in molecule A of residues 54–70, and in molecule D of residues 54–128. In molecules B and C, residues 54–128 form \( \alpha \)-helices, R1B, R2A, R2B and R3A, which are stabilized by several crystal contacts. These crystal contacts are absent in molecules A and D causing this region to remain flexible in both molecules, as suggested by the \(~25 \text{ Å}^2\) higher B-factors of this region in molecule A and our inability to accurately model this region in molecule D.
Consequently, a significant amount of positive electron density remains visible in the unmodeled regions of molecule A and D.
Figure 2.11 AlgK structure. (a) Stereo diagram of residues 21–32 of molecule C with the final $2F_o - F_c$ electron density map. The density is contoured at $1\sigma$. (b) Ribbon representation of molecule C of AlgK shown in two orthogonal orientations. The N- and C-termini are labeled N and C, respectively. (c) Schematic diagram showing the organization of $\alpha$-helices (labeled H1, H22, H23) and TPR-like motifs (labeled R1 – R10) in AlgK. Discontinuities in the schematic diagram represent regions of molecule C that could not be modeled in the electron density. The disulfide bond between helix B of repeat 4 and helix A of repeat 5 is shown in red. In panels (b) and (c) both the A and B helices of each TPR-like motif are represented in the same color.
Figure 2.12 Multiple sequence alignment of AlgK homologues. The secondary structural elements as defined in monomer C are mapped onto the multiple sequence alignment. TPR-like motifs are labeled by 3 components: R for repeat, the number of the repeat and A/B denoting whether it is the A or B helix of the motif. Other helices are labeled H for helix. Residues in white and boxed in red are conserved in all homologues. Residues colored in red are conserved in all homologues when conservative substitutions are taken into account. Residues 35–45, predicted to fold into helix A of R1, are boxed in black. The abbreviations used are as follows: PfluorW, *P. fluorescens* WCS 374r; Avine, *Azotobacter vinelandii*; Paeru, *P. aeruginosa* PAO1; Malg, *Marinobacter Algicola*; and Aborku, *Alcanivorax borkumensis.*
To detect additional conformational differences between the molecules, an optimal structural alignment of all possible pairs of molecules was performed. The pairs of superimposed molecules yielded overall C$\alpha$ root mean square (r.m.s.) deviations of 1.0–2.5 Å, with molecules B and C being the most structurally similar. When the average local displacement for each residue was plotted as a function of residue number, it was apparent that the middle region of each molecule, residues ~160–350, aligned well (Figure 2.13); however this was not the case for the N- and C-termini of the molecules where average local displacements of up to 3 and 7 Å, respectively, were observed. The average local displacements for the N-terminus are an underestimate because, as mentioned above, residues 54–70 and 54–128 could not be modeled in molecules A and D, respectively.

![Figure 2.13 Local displacements](image)

**Figure 2.13 Local displacements.** The average C$\alpha$ local displacements for molecules A-D plotted as a function of residue number.

### 2.3.7 AlgK is a member of the TPR superfamily

The DALI server was used to identify structural homologues of AlgK$^{216}$. Although the order and score of the proteins identified depended on which AlgK
molecule was used as the query, not surprisingly given the arrangement of pairs of anti-parallel $\alpha$-helices found in AlgK, the list of homologues is dominated by proteins belonging to the TPR superfamily including: Tom70 from *Saccharomyces cerevisiae*, O-linked N-acetylglucosamine transferase (OGT) from *Homo sapiens*, PilF from *P. aeruginosa*, and HcpC from *Helicobacter pylori* (Figure 2.14)\textsuperscript{225-229}.

The TPR motif is a protein-protein interaction motif generally involved in the assembly of multiprotein complexes. A single repeat is 34 residues long and carries the following consensus sequence: W4-L7-G8-L11-A20-F24-A27-P32 (Figure 2.15a)\textsuperscript{230}. The consensus sequence is highly degenerate with no residue being invariant. Each repeat folds into two anti-parallel $\alpha$-helices, where the N- and C-terminal helices are referred to as the A- and B-helices, respectively (Figure 2.15b). TPRs are present in proteins as tandem arrays of 3–16 repeats that pack into a right-handed superhelix with helix A and B forming the concave and convex surface of the superhelix, respectively (Figure 2.15c)\textsuperscript{230}.

The structure of AlgK carries 9.5 pairs of anti-parallel $\alpha$-helical repeats, spanning helix B of repeat 1 (R1B) to helix B of repeat 10 (R10B, Figure 2.11c). By superposing AlgK’s individual pairs of helices, we found a conserved set of small and large hydrophobic residues strongly resembling the TPR motif (Figure 2.16).
Figure 2.14 Structural homologues of AlgK. Schematic diagram of molecule C of AlgK (magenta) superimposed with chain A of OGT (PDB code: 1W3B), r.m.s deviation of 8.9Å for 284 residues; chain A of PilF (PDB code: 2FI7), r.m.s. deviation of 2.5Å for 155 residues; and HcpC (PDB code: 1OUV), r.m.s. deviation of 2.9Å for 152 residues. In each panel the structural homologue is coloured in cyan. DaliLite was used to obtain the optimal alignment of the two structures by superimposing their Cα atoms.
Figure 2.15 TPR motif. (a) Cartoon representation of one TPR motif along with its consensus sequence. (b) Each TPR motif folds into two-antiparallel $\alpha$-helices referred to as Helix A and B. (c) Structure of OGT protein (PDB code: 1W3B) that carries 11.5 TPR motifs that stack into a right-handed superhelix.

Figure 2.16 Multiple sequence alignment of AlgK’s TPR-like repeats. The amino acid sequences of AlgK’s TPR-like repeats and the putative R1A helix were manually aligned using the TPR consensus sequence residues as a reference. Gaps were inserted arbitrarily within intra-repeat loops at the N-terminus of helix B. The helices, as found in molecule C, are boxed in black. Helix R10A is shorter in molecule C than other molecules in the asymmetric unit; the quality of the electron density in this region prevented the modeling of the entire helix. Conserved residues are colored in red and those belonging to the TPR consensus sequence are marked with stars.
The repeats vary in length from 34 to 40 residues (Figure 2.16 and Table 2.4). The inter-repeat packing angle between repeats R3–R4, R4–R5, and R5–R6 (i.e. -17° and 51°, and -16°, respectively, Table 2.5) are atypical for TPRs (i.e. ~11-32°) and are most likely the result of the highly conserved disulfide bond between helices R4B and R5A, as well as some non-conservative substitutions at key positions in the TPR consensus sequence in R3, R4, R5 and R6\textsuperscript{231}. In addition, while residues 35–45 are predicted to fold into a helix, in the structure they are predominantly in an extended conformation. Examination of the amino acid sequence shows that this putative helix carries most of the helix A consensus sequence found for AlgK (Figures 2.12 and 2.16) and suggests that under more biologically relevant conditions, residues 35–45 may fold into a helix that corresponds to R1A and that AlgK in fact carries 10 repeats. Interestingly, two of the cleavage sites identified in the limited-proteolysis studies of P. aeruginosa AlgK, i.e. Q38 and A42, overlap with the R1A helix, suggesting that the lack of stable structure in this region may be a common feature in AlgK orthologues.

**Table 2.4 TPR motif lengths.** Length of each TPR motif in each molecule in the asymmetric unit as well as the consensus for each repeat. For repeats R2-R8, the most commonly observed “length” was used to define the actual length of the repeat. For repeat R9 and R10, the lengths of these repeats in molecules A and D were used given the better quality of the electron density in these regions.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td>37</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>R3</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>R4</td>
<td>35</td>
<td>34</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>R5</td>
<td>35</td>
<td>36</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>R6</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>R7</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>R8</td>
<td>35</td>
<td>36</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>R9</td>
<td>38</td>
<td>37</td>
<td>40</td>
<td>38</td>
<td>38</td>
</tr>
</tbody>
</table>
Table 2.5 Inter-repeat packing angles. Table listing the average inter-repeat packing angle for pairs of consecutive of helices and its associated standard deviation. The inter-repeat packing angle is defined as the angle between the A helices of adjacent repeats. The angles are measured in degrees.

<table>
<thead>
<tr>
<th></th>
<th>Average angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2-R3</td>
<td>26.1 ± 1.8</td>
</tr>
<tr>
<td>R3-R4</td>
<td>-16.8 ± 2.3</td>
</tr>
<tr>
<td>R4-R5</td>
<td>50.6 ± 0.2</td>
</tr>
<tr>
<td>R5-R6</td>
<td>-16.0 ± 2.8</td>
</tr>
<tr>
<td>R6-R7</td>
<td>7.3 ± 2.0</td>
</tr>
<tr>
<td>R7-R8</td>
<td>25.5 ± 1.7</td>
</tr>
<tr>
<td>R8-R9</td>
<td>24.4 ± 3.4</td>
</tr>
<tr>
<td>R9-R10</td>
<td>12.7 ± 2.8</td>
</tr>
<tr>
<td>R10-H21</td>
<td>35.8 ± 4.9</td>
</tr>
</tbody>
</table>

2.4 Discussion

2.4.1 Conformational flexibility of AlgK

The r.m.s deviations and average local displacements obtained from the pairwise alignment of AlgK molecules in the asymmetric unit are too large to be attributed to crystal contacts alone as proteins crystallized in different space groups generally have C\(^\alpha\) r.m.s. deviations of less than 1Å\(^2\)\(^2\). Instead, we hypothesize that these differences are due to the inherent conformational flexibility of AlgK. Flexibility is a feature observed in other non-globular repeat proteins and large TPR proteins\(^3\)\(^3\). Three of the longest naturally-occurring TPR-containing proteins for which crystal structures are available,
OGT (11.5 repeats), Tom70 (11 repeats) and Pex5 (7 repeats) have average B-factors of 114 Å², 77 Å² and 58 Å², respectively. AlgK carries 9.5 TPR-like repeats and its residues have an average B-factor of 49 Å². The higher the average B-factor of a stretch of amino acids, the more flexible or motile the region. The average B-factor for OGT, Tom70, Pex5 and AlgK are high in comparison to globular proteins which typically ranges between 15-20 Å² suggesting that long TPR proteins have a more dynamic structure. AlgK’s flexibility could explain our inability to obtain crystals that diffract beyond 2.5Å resolution, the large number of statically disordered regions that could not be modeled and its susceptibility to trace proteases. At present, the biological significance of this conformational flexibility in AlgK remains to be determined.

2.4.2 AlgK’s TPR-like motifs

Our crystal structure revealed that AlgK has 9.5 pairs of antiparallel α-helices that resemble TPR motifs. AlgK’s repeats carry the TPR consensus sequence, yet they exhibit a broader range of helical packing angles and repeat lengths than is typically found in TPR domains and therefore, we refer to AlgK’s repeats as being “TPR-like”.

Since the structure of the first TPR-containing protein was determined, several variants of the TPR motif have been defined, such as the pentatricopeptide repeat (PPR), which contains 35 residues and the Sel1-like repeat (SLR), which is 36–44 residues long. The crystal structures of two SLR proteins are available, i.e. HcpC and HcpB, and are highly similar to that of TPR proteins. The superposition of individual TPR and SLR motifs from PP5 and HcpB yield r.m.s. deviations of less than 1Å, highlighting the similarity of the structure of the individual repeats. In both HcpC and HcpB, the
SLRs also stack into right-handed superhelical structures. Various superhelical parameters, such as handedness, pitch and helical packing angles have been used to describe TPR- and SLR-containing proteins. An examination of TPR structures suggests that these parameters can take on a range of values and there appears to be no clear demarcation of these parameters between the two motifs. The finding that AlgK is structurally homologous to both TPR- and SLR-containing proteins, i.e. HcpC was identified as an AlgK homologue by DALI, reinforces this idea. Alignment of the TPR, SLR, and AlgK consensus sequences reveals a broader, more general consensus sequence that could be used to define all three motifs. In TPR proteins, residues at these positions are thought to be responsible for determining how the helices pack and assemble into superhelical structures. These residues are used to define a broader, more general consensus sequence that could be used to define all three motifs (Figure 2.17). The number of amino acids in this new consensus sequence varies to account for differences in the lengths of intra- and inter-repeat loops as well as helices.

### Figure 2.17 TPR, SLR and AlgK consensus sequence

Comparison of AlgK's consensus sequence to the TPR and SLR consensus sequences obtained from the SMART database reveals a pattern of conserved residues and residue types present in all three motifs. The abbreviations are as follows: A, alanine; L, leucine; Y, tyrosine; G, glycine; sm, small side chain; ali, aliphatic side chain; hdp, hydrophobic side chain; aro, aromatic side chain.
2.5 Conclusions

The crystal structure revealed that AlgK folds into 22 α-helices that pack into a right-handed superhelix. AlgK is structurally homologous to TPR containing proteins and carries 9.5 TPR-like motifs. AlgK’s repeats are defined as being “TPR-like” as their length, consensus sequence and helical packing angles diverge somewhat from the canonical definition of a TPR. Given the similarities between the structures and consensus sequences of TPR, SLR and AlgK’s motifs, we propose a new consensus sequence to define all three. The structure exhibits some conformational flexibility, a feature commonly observed in other large TPR and non-globular repeat proteins.
CHAPTER 3

ALGK ACTS AS A SCAFFOLD FOR THE ASSEMBLY OF THE ALGINATE BIOSYNTHETIC COMPLEX

3.1 Overview

Given the role that TPR motifs typically play in protein-protein interactions and the assembly of multi-protein complexes, the presence of 9.5 TPR-like motifs in AlgK suggests that it binds to other proteins. The mapping of highly conserved residues onto the surface of AlgK revealed three putative binding sites, two near the N-terminus and one near the C-terminus of the protein. In light of the biological data suggesting the existence of a periplasmic channel containing AlgK, AlgG and AlgX, we propose that AlgK acts as a scaffold for the assembly of this channel thereby linking inner and outer membrane components of the alginate biosynthetic complex. We hypothesize that AlgK interacts with AlgE via an N-terminal binding site assembling into a novel polysaccharide secretin; this is supported by the finding that poly β-1-6-GlcNac and cellulose secretion involves a putative outer membrane protein predicted to carry both a porin and large TPR domain. Meanwhile, the C-terminal binding site in AlgK could bind to periplasmic and/or inner membrane Alg proteins. The alginate, poly β-1-6-GlcNac and cellulose biosynthetic systems exhibit similarities suggesting these three exopolysaccharides could be produced by a common biosynthetic complex which appears to be different from the Wzy- and
ATP-binding cassette (ABC)-transporter dependent systems described for capsular polysaccharide biosynthesis in *E. coli*.

The data presented in this chapter will be included in the following publication: Carrie-Lynn Keiski, Michael Harwich, Sumita Jain, A. Mirela Néculai, Patrick Yip, Howard Robinson, Lori L. Burrows, Dennis E. Ohman and P. Lynne Howell. AlgK is a TPR-containing outer membrane lipoprotein involved in alginate biosynthesis and the localization of the AlgE porin. Manuscript in preparation.

### 3.2 Materials and Methods

#### 3.2.1 Identification of putative AlgK binding sites

T-Coffee on the EMBL-EBI website was used to generate a multiple sequence alignment for the identification of highly conserved residues in the AlgK homologues. The sequences used in the multiple sequence alignment correspond to AlgK homologues from *Azotobacter vinelandii*, *P. aeruginosa* PAO1, *P. fluorescens* WCS 374r, *Marinobacter algicola* DG893 and *Alcanivorax borkumensis* (strain SK2/ATCC700651/DSM11573) and were obtained from the UniProt Knowledgebase (SwissProt/TrEMBL)\(^{199}\). CONTACTS (Dr. G. D. Smith, Hospital for Sick Children) and the Protein interfaces, surfaces and assemblies server, PISA, at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)\(^{240}\) were used to identify and study molecular interfaces in the asymmetric unit.
3.2.2 Bioinformatic characterization of proteins

The protein sequences of AlgK and AlgE from *P. aerugionsa* PAO1, PgaA from *E. coli* K12, HmsH from *Yersinia pestis* and BesC from *E. coli* K12 and *Salmonella typhimurium* were obtained from the UniProt Knowledgebase (SwissProt/TrEMBL) and the sequence of BpsA from *Bordetella bronchiseptica* RB50 was obtained from the NCBI Protein database. PSORTb V2.0 and the Signal P 3.0 server were used to predict the subcellular localization and detect potential signal peptides, respectively. PSIPRED v2.6 was used to predict the secondary structure of the proteins. InterProScan was used to identify protein motifs. The PRED-TMBB server was used to predict the porin domain. The full-length sequences of *E. coli* BcsC and *B. bronchiseptica* BpsA were submitted to the PRED-TMBB server, however, truncated versions of *E. coli* PgaA, *Y. pestis* HmsH (residues 500–822) and *S. typhimurium* BcsC (776–1143) were needed to obtain β-barrel predictions. For *E. coli* BcsC, the porin domain predicted by PRED-TMBB spanned residues 703–1140, which overlapped with TPR domain predicted by InterProScan (40–771). Secondary structure predictions suggest that region 703–771 folds into helices and therefore it is included in the TPR domain.

3.3 Results

3.3.1 Potential sites of protein-protein interaction

Given the role that TPR motifs play in mediating protein-protein interactions, we identified several potential binding sites on AlgK. We generated a multiple sequence alignment of closely and distantly related AlgK homologues that showed that the most
highly conserved residues mapped to the N- and C-terminal regions of the protein, suggesting that these areas are functionally important (Figure 3.1). We mapped the highly conserved residues, which are not part of the TPR consensus sequence, onto the surface of the model to identify more specific clusters of conserved residues. Interestingly, the concave surface of the protein does not carry many conserved residues; these are mostly located on the convex surface or in loop regions. Several clusters of highly conserved residues are visible when mapped onto the surface of the protein, which may correspond to potential sites of protein-protein interaction. The first potential binding site is on H1, the N-terminal helix. This helix has five highly conserved residues, A18, A22, N25, L29 and A30, which all lie on the same face of the helix and are buried when the helix interacts with the concave surface of the superhelix (Figure 3.2a).

The second potential protein interaction site is composed of several highly conserved charged residues, R70, R77 and E105, on repeats 2 and 3, which cluster on the concave surface of the superhelix (Figure 3.2b). Residues R70 and R77 overlap with “hypervariable” positions identified in TPRs which have been implicated in substrate binding in other systems\textsuperscript{246}. Residue R74 also clusters with these residues in the structure and is conserved in most homologues, suggesting that it could also be part of this binding site. If, as proposed in CHAPTER 2, residues 35–45 do fold to form helix R1A, then residues E36 and D43 might also contribute to this putative binding site. Both E36 and D43 are conserved in most homologues and are predicted to be on the concave surface of the superhelix.
Figure 3.1 Putative binding sites mapped onto multiple sequence alignment of AlgK homologues. Residues in putative binding sites 1, 2 and 3 are indicated by the orange, purple and green solid lines, respectively, at the bottom of the sequence alignment. The secondary structural elements, as defined in monomer C, are mapped onto the multiple sequence alignment. TPR-like motifs are labeled above helices. Residues in white and boxed in red are conserved in all homologues. Residues colored in red are conserved in all homologues when conservative substitutions are taken into account. Residues 35–45, predicted to fold into helix A of R1, are boxed in black. The abbreviations used are as follows: PfluorW, P. fluorescens WCS 374r; Avine, Azotobacter vinelandii; Paeru, P. aeruginosa PAO1; Malg, Marinobacter Algicola; and Aborku, Alcanivorax borkumensis.
Figure 3.2 Putative sites of protein-protein interaction in the AlgK model. (a) Schematic ribbon representation of helix H1 from molecule C, in isolation (N-terminus corresponds to bottom of helix) and in the context of the rest of the protein, highlighting the conserved residues of the first putative protein-protein interaction site shown in orange. The N- and C-termini of the helix are labeled N and C, respectively. (b) Schematic diagram of the surface of molecule C in three orthogonal orientations. For better visualization of the putative binding surfaces, helix H1, residues 33–45 and 428–438 have been omitted from the structure prior to rendering the surface. Conserved residues from putative binding sites 2 and 3 are colored in purple and green, respectively. The N- and C-termini are labeled N and C, respectively.
Repeats 9 and 10 contain a large number of highly conserved residues, suggesting that this region may also be functionally important. The third putative binding site consists of several highly conserved residues found in the loops that connect helices A and B of repeat 9 (G331, Y332, L333, G334, K335) and repeat 10 (K368, G369, D373). The loops are located next to each other, on the same surface of the superhelix (Figure 3.2b).

3.3.2 Outer membrane proteins containing a TPR and porin domain

Given that alginate secretion involves both a TPR protein tethered to the outer membrane and an outer membrane porin, we investigated whether similar pairs of proteins are implicated in other secretion systems. Interestingly, PgaA, a putative outer membrane protein required for the secretion of poly β-1-6-GlcNac in gram-negative bacteria, was recently predicted to have a large N-terminal TPR domain similar to OGT and C-terminal porin domain similar to FadL, a 14-stranded outer membrane β-barrel. In the absence of PgaA, cells accumulate poly β-1-6-GlcNac in the periplasmic space, suggesting the PgaA is responsible for transporting the polysaccharide across the outer membrane. The results of our bioinformatics analysis support the presence of a TPR and porin domain in PgaA. PgaA homologues such as BpsA in B. bronchiseptica and HmsH in Y. pestis are predicted to have a similar topology (Figure 3.3). HmsH has been shown to localize to the outer membrane, supporting the prediction that PgaA and its homologues are outer membrane proteins. Upon further analysis, we found that the bcs operons in E. coli K12 and S. typhimurium, responsible for the polymerization and export of cellulose, encode a putative outer membrane protein, BcsC, that we predict has an N-
terminal TPR domain followed by a C-terminal porin domain (Figure 3.3). Interestingly, PgaA, BpsA, HmsH, both BcsC homologues and AlgE all have a phenylalanine or tryptophan residue at the C-terminus of the protein, which is a defining characteristic of outer membrane β-barrel proteins\textsuperscript{249}. PgaA, its homologues and BcsC from \textit{S. typhimurium} are all predicted to fold into 16-stranded β-barrels and AlgE into an 18-stranded β-barrel. Unusually, BcsC from \textit{E. coli} is predicted to fold into a 15-stranded β-barrel (Figure 3.3). The porin domain of the \textit{E. coli} BcsC was predicted by Pred-TMBB to span residues 703–1140 corresponding to 18 β-strands, however residues 703-771 are predicted to carry TPR motifs and to fold into α-helices, suggesting the porin domain consists of residues 772-1140 and folds into a 15-stranded β-barrel; this seems unlikely given that β-barrels typically carry an even number of anti-parallel strands ranging from 8 to 22\textsuperscript{250}. To date, only one membrane β-barrel has been shown to have an odd number of strands, \textit{e.g.} the structure of the human voltage-dependent anion channel revealed that its β-barrel is built of 19 strands\textsuperscript{251}. Therefore, it seems more likely that BcsC folds into a 16-stranded β-barrel, as is predicted for its \textit{S. typhimurium} homologue, along with PgaA, HmsH and BpsA.

### 3.4 Discussion

In general, bacterial polysaccharide secretion appears to be mediated by a trans-envelope complex, consisting of at least an inner membrane transporter, an outer membrane transporter and an inner membrane component or components that link the two transporters\textsuperscript{252,253}. 
**Figure 3.3 Predicted TPR and porin domains.** Schematic representation of predicted TPR and porin domains in AlgK (P. aeruginosa PAO1), AlgE (P. aeruginosa PAO1), PgaA (E. coli K12), HmsH (Y. pestis), BpsC (B. bronchiseptica), BcsC E.coli (E. coli K12) and BcsC S.typh (S. typhimurium) from our bioinformatics analysis. The porin domain is coloured in yellow and the number of β-strands predicted in the porin is listed on the right. The predicted TPR domain is coloured in dark blue and the regions predicted to fold into helices are coloured in light blue. AlgK and AlgE are shown as they are organized in the algD operon in P. aeruginosa PAO1, highlighting their apparent translational coupling. The TPR and porin domain boundaries are listed beneath the proteins. Predicted signal sequences are coloured in red.

The best described polysaccharide secretion systems in bacteria are the Wzy-dependent and ABC transporter-dependent pathways responsible for capsular polysaccharide production in E. coli. Capsular polysaccharides are secreted by bacteria and remain tightly associated with their cell surface and are important virulence factors\(^{17}\). E. coli capsules have been classified into 4 groups, *i.e.* Group I-IV, according to the genetic and biosynthetic features of the polysaccharide\(^{254}\). Group I/IV and II/III capsular
polysaccharides are produced by the Wzy-dependent and ABC-transporter-dependent systems, respectively.

**3.4.1 Wzy-dependent biosynthesis of Group I/IV capsular polysaccharides**

Group I/IV capsular polysaccharides consist of long chains of a repeating oligosaccharide unit, which contains a unique combination of sugars, glycosidic bonds and, in some cases, covalently attached side groups\(^{17}\). The repeat units are assembled at the cytoplasmic face of the inner membrane on a carrier lipid, undecaprenyl phosphate, located in the inner membrane. The assembly of the repeat unit is initiated by a glycosyltransferase that transfers a sugar 1-phosphate from an activated sugar nucleotide precursor in the cytoplasm to undecaprenyl phosphate\(^{255,256}\). Subsequently, peripheral glycosyltransferases add the remaining sugars to complete the repeat unit\(^{17}\). Once finished, the membrane-bound repeat unit is flipped across the inner membrane by Wzx, an integral inner membrane protein\(^{257}\). The lipid-linked intermediates are polymerized in the periplasm by the integral inner membrane protein Wzy. The polysaccharide chains are elongated at their reducing end\(^{255}\). Three proteins have been implicated in the polymerization and translocation of Group Icapsular polysaccharides to the cell surface, namely Wza, Wzc and Wzb\(^{17}\). Wzc is an inner membrane protein involved in polymerization/chain length regulation and polysaccharide export\(^{258}\). Wzc is a member of the polysaccharide co-polymerase (PCP) 2 family, which is predicted to have a large periplasmic domain that folds into a coiled-coil structure which is flanked by 2 TMs\(^{259,260}\). The second TM is followed by a cytoplasmic tyrosine autokinase domain that targets the tyrosine-rich C-terminus of the protein\(^{261}\). Wzb is a cytoplasmic protein...
tyrosine phosphorylase that is believed to dephosphorylate Wzc and is required for the assembly of Group I capsular polysaccharides\textsuperscript{258}. Cryo-electron microscopy and other experimental data show that Wzc’s periplasmic domain interacts with the outer membrane secretin Wza forming a \textasciitilde 170 Å long transenvelope complex\textsuperscript{262-265}. Wza protomers assemble into a large octameric structure believed to be responsible for transporting Group I capsular polysaccharides across the outer membrane. The crystal structure of Wza revealed that it forms an \( \alpha \)-helical pore in the outer membrane that is large enough to accommodate capsular polysaccharide in an extended conformation and has a 100 Å long hydrophilic channel that protrudes into the periplasm\textsuperscript{266}. After polymerization by Wzy, the nascent polymer is thought to enter the Wza channel through holes in the Wza-Wzc complex, which then transports the polymer to the cell surface (Figure 3.4)\textsuperscript{252}. Polymerization and export are coupled in Group I capsular polysaccharide biosynthesis; in the absence of Wza, no polysaccharide accumulates intracellularly\textsuperscript{264}. How Group I capsular polysaccharides are attached to the cell surface remains unclear, however the outer membrane protein Wzi has been implicated in this process\textsuperscript{267}.

3.4.2 ABC transporter-dependent biosynthesis of Group II/III capsular polysaccharides

The \textit{E. coli} K5 capsule is a Group II capsular polysaccharide that has been extensively studied and will be used here as a model to illustrate the different steps in Group II/III capsular polysaccharide biosynthesis. The K5 polysaccharide is polymerized directly from its sugar nucleotide precursors at the cytoplasmic face of the inner
K5 biosynthesis is believed to be initiated by the transfer of a sugar from an activated nucleotide to an unidentified membrane-bound acceptor and elongated by the addition of individual sugars to the nonreducing end of the nascent polysaccharide by the glycosyltransferases KfiA and KfiC\textsuperscript{253,268,269}. At some point before, during or after synthesis of the polymer, a phosphatidyl diacylglycerol is added to the reducing end of the chain, which serves to tether the completed polysaccharide to the cell surface\textsuperscript{17}. KfiA and KfiC assemble into a complex at the inner membrane along with several other proteins implicated in polysaccharide biosynthesis, \textit{i.e.} KfiB, KfiD, KpsC, KpsS, KpsM and KpsT\textsuperscript{270,271}. KfiB is believed to play a structural role in the K5 biosynthetic complex\textsuperscript{270}. KfiD is involved in the production of the activated precursors\textsuperscript{269}. KpsS and KpsC are needed for polymerization and transport of the polymer across the inner membrane, however their exact role remains elusive and controversial\textsuperscript{17,271,272}. The nascent polysaccharide is transported across the inner membrane by KpsM and KpsT which assemble into an ABC-transporter. The channel of the ABC-transporter is thought to be formed by KpsM, which is predicted to carry 6 TMs, while KpsT is a cytosolic protein that associates with KpsM and corresponds to the nucleotide binding domain of the transporter. KpsM/T provides the energy for polysaccharide secretion\textsuperscript{271,273,274}. KpsD and KpsE are responsible for transporting the polysaccharide across the periplasm and outer membrane; in their absence, polysaccharide accumulates in the periplasm suggesting that synthesis and secretion are not coupled in this system\textsuperscript{275}. KpsD appears to correspond to the outer membrane secretion pore equivalent to Wza in the Wzy-dependent system\textsuperscript{17}. KpsE is a member of the PCP-3 family and is believed to act as an adaptor between KpsM/T and KpsD\textsuperscript{259,260}. KpsE is an inner membrane protein predicted
to carry a large periplasmic domain flanked by a TM at its N-terminus and an inner membrane associated amphipathic helix at its C-terminus\textsuperscript{276}. \textit{In vivo} chemical cross-linking experiments have shown that KfiA, KpsM, KpsT, KpsS, KpsE and KpsD assemble into complex that would span the cell envelope and be responsible for polysaccharide biosynthesis (\textbf{Figure 3.4})\textsuperscript{277}.

\begin{figure}[h]
\centering
\includegraphics[width=0.6\linewidth]{alginate_and_capsular_polysaccharide.png}
\caption{\textbf{Figure 3.4 Alginate and capsular polysaccharide secretion systems.} Cartoon representation of the multi-protein complexes thought to be responsible for synthesis and secretion in alginate, and Group I/IV, and II/III capsular polysaccharide production. The capsular polysaccharide systems have been simplified and only include those proteins thought to form the basic components of the trans-envelop complex and the enzymes responsible for polymerization. Glycosyltransferases are colored in orange and the Wzy polymerase is colored in yellow. Proteins responsible for the transport of capsular polysaccharides across the inner membrane are colored in green. Inner membrane proteins needed for polymerization that may bridge the inner and outer membrane biosynthetic machinery are colored in cyan. The outer membrane protein responsible for transporting the polysaccharide across the outer membrane is colored in dark pink. Capsular polysaccharide models were based on Figures 4 and 7 in Whitfield (2006)\textsuperscript{17}.}
\end{figure}

3.4.3 Exopolysaccharide production
Exopolysaccharide production has not been as extensively studied as capsular polysaccharide production. The genetic loci responsible for the production of many exopolysaccharides have been identified and a limited characterization of the genes has been conducted in some cases. Gene clusters responsible for the production of several exopolysaccharides encode signature homologues of the Wzy-dependent pathway suggesting they are produced by a similar mechanism. These exopolysaccharides include colanic acid, emulsan, xanthan, *Vibrio* polysaccharide, succinoglycan, amylovoran, psl-polysaccharide and the heteropolysaccharides produced by lactic acid bacteria\textsuperscript{20,278-282}. In contrast, the gene clusters responsible for the production of the exopolysaccharides poly β-1-6-GlcNAc, the Pel-polysaccharide, cellulose and alginate lack signature homologues of either of the known capsular polysaccharide biosynthetic pathways\textsuperscript{19,40,132,283}. Therefore, these exopolysaccharides may be synthesized and secreted by a novel mechanism distinct from capsule biosynthesis.

### 3.4.4 Alginate is secreted by novel secretion complex

The alginate operon appears to encode all three components of a trans-envelope secretion complex, each of which is distinct from their counterparts in the capsular biosynthetic systems. The integral inner membrane glycosyltransferase, Alg8, is believed to be responsible for synthesizing and transporting polyM across the inner membrane given it carries multiple TMs and the lack of other candidates encoded in the *algD* operon\textsuperscript{132,174}. Another family 2 glycosyltransferase, the streptococcal hyaluronan synthase is predicted to have a similar membrane topology as Alg8 and appears to be capable of synthesizing and secreting hyaluronan by itself when expressed in heterologous bacteria.
or acapsular mutants\textsuperscript{284-286}. The inner membrane protein, Alg44, could also participate in the translocation of polyM across the inner membrane. The integral outer membrane β-barrel protein, AlgE, is responsible for transporting alginate across the outer membrane into the extracellular milieu\textsuperscript{195}. We hypothesize that AlgK is involved in linking the inner and outer membrane transporters for alginate secretion, \textit{i.e.} Alg8/Alg44 and AlgE, respectively, (\textbf{Figure 3.4}).

\textbf{3.4.4.1 AlgK is a stable component of the periplasmic conduit}

Previous biological studies have suggested that AlgK is a stable component of the alginate biosynthetic complex. The FRD\textit{Δ}alg\textit{K} mutant appears non-mucoid when streaked on an agar plate given the cells secrete low molecular weight uronic acids instead of high molecular weight alginate: all of the alginate secreted by the mutant is dialyzable through a membrane with a 10 kDa molecular weight cut-off and 90% of it can diffuse through a 1 kDa membrane. In contrast, 90% of the alginate secreted by FRD1 cells is retained within a membrane with a 10 kDa cut-off\textsuperscript{186,189}. The AlgG and AlgX deletion mutants display a similar non-mucoid phenotype and secrete alginate fragments that correspond mostly to dimers and a mixture of dimers and trimers, respectively\textsuperscript{186,188}. In all three mutants, these low molecular weight fragments have been identified as AlgL degradation products\textsuperscript{186,188}. AlgL is a periplasmic lyase that plays an essential role in alginate production: in the absence of AlgL, mucoid cells rapidly accumulate alginate in their periplasm and burst\textsuperscript{138}. In light of these findings, AlgX, AlgG and AlgK have been proposed to assemble into a periplasmic channel for alginate secretion (\textbf{Figure 1.10a})\textsuperscript{138,186,188}. AlgL is also proposed to be a member of this scaffold,
however its enzymatic activity is believed to be inhibited while associated with the complex. The absence of AlgK, AlgG or AlgX has been proposed to destabilize the periplasmic channel thereby releasing AlgL and allowing it to degrade nascent alginate polymers (Figure 1.10b). In the absence of AlgL, the periplasmic conduit is destabilized, causing the accumulation of periplasmic alginate (Figure 1.10c)\textsuperscript{138}.

3.4.4.2 AlgK acts as a scaffold for the assembly of the complex

AlgE is predicted to be a $\beta$-barrel protein with 18 transmembrane strands interspersed by short periplasmic and longer extracellular loops\textsuperscript{195}. This topological prediction suggests that AlgE lacks a domain equivalent to the large periplasmic domain of Wza. This deficit would necessitate that other periplasmic proteins or domains form the bulk of the periplasmic conduit, a role we propose is fulfilled, at least in part, by the outer membrane lipoprotein AlgK. The large number of TPR-like motifs in AlgK make it an excellent candidate to bind to multiple proteins and act as a scaffold for the assembly of the periplasmic conduit thereby linking AlgE to Alg8 and Alg44 in the inner membrane. This idea is not without precedence as multiple proteins have been shown to bind the TPR-containing proteins Hop, PscG and YscG\textsuperscript{287-289}. In light of these findings, we hypothesize that AlgK, in addition to its role in the localization of AlgE, acts as a scaffold for the assembly of a multiprotein complex responsible for alginate polymerization and export. By mapping highly conserved residues onto the surface of the AlgK model, three potential sites of protein-protein interaction have been identified (Figure 3.2).
3.4.4.3 N-terminal binding sites

The putative binding sites found on the H1 helix and the concave surface of R2 and R3 are close to the N-terminus of AlgK which is attached to the outer membrane via its lipid moiety and thus positioned to interact directly with the periplasmic loops of AlgE (Figure 3.2 and 3.4). While the conservation of residues on one face of helix H1 at the N-terminus of AlgK may suggest that this region is involved in protein-protein interactions, we hypothesize that the association of helix H1 and residues 35–45 to the concave surface of the superhelix is not biologically relevant. As detailed in CHAPTER 3, residues 35–45 are predicted to fold into helix A of TPR-like repeat R1, i.e. R1A, in the structure of AlgK, and as such would be expected to bind to the hydrophobic face of helix R1B; this interaction would in turn alter the location and packing of helix H1. We believe that the conformation observed in the crystal structure is the consequence of the absence of AlgK’s binding partner(s), the lack of lipidation at the N-terminus of helix H1, and the crystallization process; presumably, the binding of H1 to the concave surface of the superhelix gave AlgK a more compact and rigid conformation thereby enabling it to pack into a well-ordered crystal. The crystal structure of AlgK from *P. fluorescens* and limited-proteolysis studies on *P. aeruginosa* AlgK show that R1A is disordered or partially-structured in both proteins suggesting the structural instability in this region may be a common feature in AlgK orthologues. We further hypothesize that residues in helix R1A participate in the putative binding site on the concave surface of R2 and R3, and may be disordered or partially-structured in the absence of its binding partner, as observed in other TPR proteins\textsuperscript{290}.  

---

89
3.4.4.4 C-terminal binding site

The third putative binding site in AlgK is located at the C-terminus of the protein and therefore could bind to periplasmic and/or inner membrane proteins. Potential periplasmic proteins that could bind to the C-terminus of AlgK include the following: 1) AlgG, the mannuronan epimerase; 2) AlgJ and F, proteins involved in alginate acetylation; 3) AlgX, a protein of unknown function and 4) AlgL, the alginate lyase. As mentioned previously, AlgG, AlgX and AlgL have been implicated along with AlgK in the formation of a periplasmic conduit, thus lending experimental support to the idea that these proteins might bind directly to AlgK\textsuperscript{138,186,188}. AlgK could also interact with the inner membrane protein Alg44. Examination of the \textit{algD} operon suggests that Alg44 may be the coupling protein linking the inner membrane polymerization and transport machinery, \textit{i.e.} Alg8 and Alg44, to the outer membrane transporter. Alg44 has a cytoplasmic N-terminal PilZ domain, a predicted single transmembrane region and a C-terminal periplasmic domain with structural similarity to HlyD, AcrA and MexA\textsuperscript{35,176}. HlyD, AcrA and MexA are members of the membrane fusion protein family that couple an inner membrane transporter to an outer membrane factor, such as TolC, to form a channel to translocate substrates across the outer membrane\textsuperscript{201,292}. The predicted structural similarity of Alg44 to HlyD, AcrA and MexA, and the observation that the \textit{in vivo} stability of AlgE appears to be dependent on the presence of Alg44 suggests that Alg44’s periplasmic domain may interact with AlgK, or another Alg protein in the periplasm given that structure predictions suggest Alg44’s periplasmic domain is not long enough to span the periplasm and interact directly with AlgE (A.M. Neculai, personal communication)\textsuperscript{174}. 

90
3.4.5 Similarities between alginate, cellulose and poly β-1-6-GlcNAc biosynthetic systems

Closer examination of the gram negative alginate, cellulose and poly β-1-6-GlcNAc biosynthetic systems reveals striking similarities, suggesting that these exopolysaccharides are produced by a common mechanism. The bcs and pga operons responsible for cellulose and poly β-1-6-GlcNAc polymerization and export in E. coli, each encodes four proteins, BcsABZC, and PgaABCD, respectively\(^40,283\). While these systems have fewer than the seven proteins of the algD operon that have been implicated in alginate polymerization and export, each appears to contain the necessary building blocks for a trans-envelope secretion complex.

**Inner membrane polymerase** - The alg, bcs, and pga operons all encode an integral inner membrane protein that is required for the polymerization of the homopolymer. Each of these proteins, namely Alg8, PgaC and BcsA, is predicted to have multiple transmembrane helices and a large soluble domain that exhibits sequence similarity to family 2 glycosyltransferases\(^40,174,247,283,293\).

**Inner membrane associated protein** - Each of the polymerases appears to be associated with an inner membrane protein that is needed for polymerization, namely Alg44, BcsB and PgaD, however these proteins do not share any sequence similarity\(^40,174,247,283,293\). These proteins could be involved in transport of the polysaccharide across the inner membrane and/or the regulation of polymerization, as is suggested by the regulation of alginate production by c-di-GMP binding to Alg44\(^35\).
**Outer membrane TPR/porin protein** - Both the *bcs* and *pga* operons encode an outer membrane component, BcsC and PgaA, respectively, that is predicted to contain both a large N-terminal TPR-containing domain and a C-terminal β-barrel porin domain. PgaA has previously been implicated in exopolysaccharide secretion\textsuperscript{247}. In alginate production, the TPR and porin domains are encoded by separate proteins, AlgK and AlgE, that appear to be translationally-coupled on the *algD* operon\textsuperscript{190}. This observation further strengthens our proposal that AlgK and AlgE interact directly and suggest that AlgK/AlgE, PgaA and BcsC define a new family of secretins where the porin domain mediates secretion of polysaccharide across the outer membrane and the TPR domain acts as scaffold for the assembly of the biosynthetic complex.

In light of these findings, we hypothesize that in cellulose and poly β-1-6-GlcNac production, like alginate production, the outer membrane TPR/porin protein forms a stable transenvelope complex with the inner membrane glycosyltransferase and/or its associated protein and that the large TPR-containing protein is responsible for binding or recruiting various proteins for the assembly of the complex (Figure 3.5).
Figure 3.5 Bacterial polysaccharide biosynthetic complexes. Cartoon representation of the multi-protein complexes thought to be responsible for synthesis and secretion in alginate, Poly β-1-6-GlcNAc and cellulose exopolysaccharide. The systems have been simplified and only include those proteins thought to form the basic components of the trans-envelop complex and the enzymes responsible for polymerization. Glycosyltransferases are colored in orange. Inner membrane proteins needed for polymerization that may bridge the inner and outer membrane biosynthetic machinery are colored in blue. The outer membrane secretin is colored in purple for all systems, except alginate where AlgE is purple and AlgK is light pink.

3.5 Conclusion

In light of the data presented in this chapter, we propose that AlgK acts as a scaffold that binds to multiple Alg proteins, such as AlgE and Alg44, to form a trans-periplasmic channel for the secretion of alginate. AlgK could bind to AlgE through its putative N-terminal binding sites and could bind to periplasmic and inner membrane proteins through its C-terminal TPR motifs. The biosynthetic machinery responsible for alginate production is different from the established capsular polysaccharide systems and
appears to represent a novel biosynthetic/secretion system. Other exopolysaccharides such as cellulose and poly β-1-6-GlcNac appear to be produced by a similar biosynthetic mechanism where AlgK/AlgE, PgaA, and BcsC define a new family of polysaccharide secretins.
CHAPTER 4

TOWARDS THE IDENTIFICATION OF ALGK’S BINDING PARTNERS

4.1 Overview

We have hypothesized that AlgK binds to AlgE assembling into a novel secretin and, in parallel, binds to other proteins implicated in alginate production thereby acting as a scaffold for the assembly of the alginate biosynthetic complex. To test these hypotheses, several experiments were performed and the preliminary results from these studies are presented in this chapter. Both purified AlgK and AlgE are currently available in the lab, thus two experiments were conducted to investigate whether these two proteins interact to form a stable complex. Nickel-affinity pulldown experiments using different His-tagged and untagged versions of purified AlgK and the AlgE proteins were performed. The two proteins were also mixed in different ratios and run over a gel filtration column in an attempt to isolate a stable complex. Neither of these experiments provided evidence that AlgK and AlgE interact. To identify other putative AlgK binding partners, we investigated how the absence of AlgK in the FRDΔalgK mutant affects the levels of other proteins implicated in alginate polymerization and secretion, i.e. Alg proteins. AlgK levels were also compared in FRD1 and other deletion strains. The preliminary results and observations from the stability studies are presented and
discussed, and suggestions are made on how to make these studies more quantitative and statistically rigorous when repeated in the future.

**Acknowledgements**: John Whitney, a PhD candidate in the Howell lab, produced the AlgE protein for these experiments and Patrick Yip produced the untagged AlgK protein.

Part of the text included in this chapter has been published in the following publication:

**4.2 Materials and Methods**

**4.2.1 Expression, purification and refolding of *P. aeruginosa* AlgE**

The expression, purification and refolding of AlgE were as described in Whitney *et al* (2009). After purification, the protein was detergent-exchanged into buffer A (20 mM Tris-HCl pH 8, 150 mM NaCl, 0.02% w/v dodecyl-β-D-maltopyranoside (DDM)) by running the protein over a HiLoad 16/60 Superdex 200 prep grade gel filtration column (Pharmacia). To remove the N-terminal His-tag, 12 mg AlgE in a volume of 15 ml was incubated with 10 units of thrombin (Novagen) overnight at 21 °C. The digestion was stopped the next day by the addition of PMSF to a final concentration of 1 mM. The digestion mixture was run over a Ni-NTA column (Qiagen) to remove any residual tagged protein. The cleaved protein was concentrated using a monoQ anion-exchange
column (Amersham Biosciences) and then dialyzed into buffer B (20 mM Tris-HCl pH 7.0, 300 mM NaCl, 0.02% w/v DDM and 10 mM imidazole).

4.2.2 Expression, purification and thrombin cleavage of P. aeruginosa 26K1ntag AlgK

The expression and purification of 26K1ntag AlgK were as described in CHAPTER 2, section 2.2.3.3. After purification and removal of the His6-tag, the protein was concentrated to approximately 1 mg/ml using a concentrator with a 30 kDa molecular weight cut-off (Amicon) and dialyzed against buffer B overnight at 4 °C.

4.2.3 Expression and purification of P. aeruginosa AlgK 26K1

The expression and purification of P. aeruginosa AlgK 26K1 were as described in CHAPTER 2, section 2.2.3.2. Prior to use the purified protein was dialyzed overnight at 4 °C against buffer B.

4.2.4 Nickel affinity co-elution studies

This experiment was conducted on an AKTAprime purification system at 4 °C. A 1 ml Ni-NTA Superflow cartridge (Qiagen) was pre-equilibrated with 20 ml of buffer B. Approximately 1 mg of His-tagged protein either (AlgK or AlgE) diluted in 20 ml of buffer B was loaded onto the Ni-NTA resin at a flow rate of 0.5 ml/min. The resin was then washed at a flow rate of 2 ml/min with 60 ml of buffer C (20 mM Tris-HCl pH 7.0, 300 mM NaCl, 0.02% w/v DDM, 19 mM imidazole) when using the untagged AlgK or buffer D (20 mM Tris-HCl pH 7.0, 300 mM NaCl, 0.02% w/v DDM, 25 mM imidazole)
when using untagged AlgE. After the first wash step, approximately 1 mg of untagged protein (either AlgK or AlgE) diluted in 20 ml of buffer B was loaded onto the Ni-NTA resin at a flow rate of 0.5 ml/min. The resin was then washed with 60 ml of buffer C or D at a flow rate of 2 ml/min. The tagged protein was eluted from the column with 30 ml of buffer E (20 mM Tris-HCl pH 7.0, 300 mM NaCl, 0.02% w/v DDM, 150 mM imidazole) at a flow rate of 2 ml/min. For each experimental run, two negative controls were run: one in the absence of tagged protein and the other in the absence of untagged protein. In both controls, the omitted protein was replaced by an equivalent volume of buffer B. 10 µl of the load, wash and elution fractions were mixed with an equal volume of 2X SDS loading buffer and boiled for 10 min in preparation for SDS-PAGE separation and Western blot analysis.

4.2.5 Western Blots for co-elution

20 µl fractions from the co-elution study were separated using a 12% SDS-PAGE gel then transferred to a PVDF membrane (GE Healthcare). Western blots were conducted using the protocol described in CHAPTER 2, section 2.2.5. Primary antibodies were all generated in rabbits and used at the following dilutions: α-AlgK at 1:6000 and α-AlgE at 1:10,000. A goat anti-rabbit IgG-alkaline phosphatase conjugate (Bio-Rad) was used as the secondary antibody at a dilution of 1:3000.

4.2.6 Gel filtration studies

Untagged AlgK and N-terminally His-tagged AlgE were incubated for approximately 3 hrs at 4 °C either by themselves or mixed in different ratios (Table 4.1)
in a total volume of 500 µl and then run over a Superdex200 HR 10/30 gel filtration column (Pharmacia).

<table>
<thead>
<tr>
<th>AlgK:AlgE ratios</th>
<th>[untagged AlgK] (µM)</th>
<th>[tagged AlgE] (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:1</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td>1:0</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>1:1</td>
<td>40</td>
<td>37</td>
</tr>
<tr>
<td>3:1</td>
<td>120</td>
<td>37</td>
</tr>
<tr>
<td>1:3</td>
<td>40</td>
<td>111</td>
</tr>
<tr>
<td>3:0</td>
<td>120</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.1 Ratios and concentrations of AlgK and AlgE.

4.2.7 Growth and harvesting of FRD bacterial cultures

For every bacterial culture, cells were freshly streaked from frozen stocks onto agar plates containing ½ LB and ½ Pseudomonas Isolation Agar (PIA) and incubated for 24 hrs at 37 °C. The following day, 30 ml of LB in a 250 ml Erlenmeyer flask was inoculated with a single colony and grown overnight at 30 °C. The next morning, 30 ml of LB in a 250 ml Erlenmeyer flask was inoculated with 0.3 ml of overnight culture and grown at 30 °C until the culture attained an OD$_{600}$ ~ 1.0 at which point ~ 1 ml of culture was harvested; the volume was adjusted to normalize for cell numbers in the different cultures. Cells were spundown at 6,785 g for 10 min. Cells were resuspended in 0.2 ml of 2x SDS loading dye containing 2 mM DTT then lysed by sonication (3 times for 2 sec). Whole cell lysates were boiled for 10 min then stored at -20 °C. Table 4.2 lists the FRD deletion mutants used for these experiments. FRD$_{algK}$C28S is a derivative of FRD1 that expresses a C28S mutant of AlgK and, similarly to the FRD$_{algK}$ mutant, it secretes degraded alginate (D. Ohman, personal communication).
4.2.8 Western blots for stability assays

5 µl of whole cell lysate was separated using a 12% SDS-PAGE gel then transferred to a PVDF membrane (GE Healthcare). Western blots were conducted as described in CHAPTER 2, section 2.2.5. Primary antibodies were all generated in rabbits and used at the following dilutions: \( \alpha \)-AlgJ 1:10,000; \( \alpha \)-AlgF 1:20,000; \( \alpha \)-AlgG 1:3000; \( \alpha \)-PilF 1:5000; \( \alpha \)-AlgL 1:1000; \( \alpha \)-AlgK 1:6000; \( \alpha \)-AlgE 1:10,000. A goat anti-rabbit IgG-alkaline phosphatase conjugate (Bio-Rad) was used as the secondary antibody at a dilution of 1:3000. PilF, a protein involved in type IV pilus biogenesis, was used as a loading control.

Table 4.2 FRD deletion mutants. List of FRD deletion mutants and their genotype.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRD(\Delta)algK (FRD1100)</td>
<td>algK::Gm</td>
<td>Reference(^{189})</td>
</tr>
<tr>
<td>FRD(\Delta)algG (FRD1200)</td>
<td>algG::Gm</td>
<td>Reference(^{186})</td>
</tr>
<tr>
<td>FRD(\Delta)alg8 (FRD1108)</td>
<td>Alg8::Gm</td>
<td>Reference(^{174})</td>
</tr>
<tr>
<td>FRD(\Delta)alg44 (FRD1144)</td>
<td>Alg44::Gm</td>
<td>Reference(^{174})</td>
</tr>
<tr>
<td>FRD(\Delta)algF(FRD1175)</td>
<td>(\Delta)algF</td>
<td>Reference(^{183})</td>
</tr>
<tr>
<td>FRD(\Delta)algJ(FRD1176)</td>
<td>(\Delta)algJ</td>
<td>Reference(^{183})</td>
</tr>
<tr>
<td>FRD(\Delta)algI(FRD1177)</td>
<td>(\Delta)algI</td>
<td>Reference(^{183})</td>
</tr>
<tr>
<td>FRD(\Delta)algX (FRD1algX::Gm)</td>
<td>algX::Gm</td>
<td>D. Ohman</td>
</tr>
<tr>
<td>FRDalgKC28S (FRD1150)</td>
<td>algKC28S</td>
<td>D. Ohman</td>
</tr>
</tbody>
</table>

4.2.9 PCR to verify insertion of Gm cassette in FRD\(\Delta\)alg44 and FRD\(\Delta\)algX

The nucleotide sequence for the Gm 3’-acetyltransferase AAC(3)-I gene (i.e. Gm cassette), \( algL \) and \( algK \) were obtained from NCBI nucleotide database and used to design gene specific primers (Table 4.3). The genomic DNA from FRD\(\Delta\)alg44 and
FRD\textit{algX} cells was extracted using the InstaGene matrix (Bio-Rad) following the manufacturer’s instructions. The primers, genomic DNA and the expected size of the PCR product are listed in Table 4.3. Each PCR was repeated twice.

Table 4.3 Summary of PCR reactions and expected size of products. Each row of the table corresponds to a PCR reaction, \textit{i.e.} the forward and reverse primers and genomic DNA used for the reaction along with the expected size of the PCR products.

<table>
<thead>
<tr>
<th>Lane in Figure 4.12</th>
<th>Primers</th>
<th>Genomic DNA</th>
<th>Expected size of PCR product (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Gm forward: GTTACGCAGCAGCAACGATGTACG Gm reverse: GGTGGCGGTACTTGGGTCGATATC</td>
<td>FRD\textit{alg44}</td>
<td>534 (Gm control)</td>
</tr>
<tr>
<td>3</td>
<td>Gm forward: GTTACGCAGCAGCAACGATGTACG Gm reverse: GGTGGCGGTACTTGGGTCGATATC</td>
<td>FRD\textit{algX}</td>
<td>534 (Gm control)</td>
</tr>
<tr>
<td>4</td>
<td>Gm forward: GTTACGCAGCAGCAACGATGTACG AlgK reverse: GGGAGGGAGGATGGGCATCTTC</td>
<td>FRD\textit{alg44}</td>
<td>1363 (Gm inserted properly)</td>
</tr>
<tr>
<td>5</td>
<td>Gm forward: GTTACGCAGCAGCAACGATGTACG AlgL reverse: GGCGATACGGATCAGGTGGGACG</td>
<td>FRD\textit{algX}</td>
<td>808 (Gm inserted properly)</td>
</tr>
<tr>
<td>6</td>
<td>Gm reverse: GGTGGCGGTACTTGGGTCGATTC AlgK reverse: GGGAGGGAGGATGGGCATCTTC</td>
<td>FRD\textit{alg44}</td>
<td>1363 (Gm inserted backwards)</td>
</tr>
<tr>
<td>7</td>
<td>Gm reverse: GGTGGCGGTACTTGGGTCGATTC AlgL reverse: GGCGATACGGATCAGGTGGGACG</td>
<td>FRD\textit{algX}</td>
<td>808 (Gm inserted backwards)</td>
</tr>
</tbody>
</table>

4.3 Results

4.3.1 NiNTA co-elution studies
To determine if AlgK interacts directly with AlgE, a nickel affinity co-elution study was conducted using purified AlgK and AlgE. The different combinations of tagged and untagged proteins assayed are listed in Table 4.4.

Table 4.4 Summary of experiments and controls for AlgK-AlgE co-elution studies. The figures containing the blots from these experiments and controls are also listed.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Control 1</th>
<th>Control 2</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untagged AlgK + N-terminally His-tagged AlgE</td>
<td>Untagged AlgK by itself</td>
<td>N-terminally His-tagged AlgE by itself</td>
<td>4.1</td>
</tr>
<tr>
<td>Untagged AlgE + C-terminally His-tagged AlgK</td>
<td>Untagged AlgE by itself</td>
<td>C-terminally His-tagged AlgK by itself</td>
<td>4.2</td>
</tr>
<tr>
<td>Untagged AlgE + N-terminally His-tagged AlgK</td>
<td>Untagged AlgE by itself</td>
<td>N-terminally His-tagged AlgK by itself</td>
<td>4.3</td>
</tr>
</tbody>
</table>

For each experiment, His-tagged protein was first bound to the Ni-NTA resin, and then subsequently untagged protein was run-over the charged resin. After both protein loading steps, the resin was washed extensively with buffer containing the minimal concentration of imidazole necessary to remove untagged AlgK or AlgE non-specifically bound to the column, i.e. 19 mM and 25 mM, respectively. After the final wash step, the tagged protein bound to the resin was eluted with 150 mM imidazole. If the proteins form a stable complex, the untagged protein should also be detected in the elution fractions. As controls, each experiment was conducted in the absence of either the tagged or untagged protein. For each of the three experiments and their corresponding controls, fractions from the protein loading, washing and elution steps were blotted for AlgK and AlgE. Faint protein bands originally thought to correspond to the untagged proteins of the proposed complex were observed in the elution fractions of all the experimental runs.
(Figures 4.1, 4.2 and 4.3, panel b). However, these bands were also observed in the elution fractions from the negative controls which lacked the untagged protein (Figures 4.1, 4.2 and 4.3, panel d) suggesting that these bands were due to the cross-reactivity of the antibodies against large quantities of purified protein, i.e. the AlgK antibody was cross-reacting with the tagged AlgE in the elution fractions.

**Figure 4.1 Co-elution studies with N-terminally His-tagged AlgE and untagged AlgK.** This figure contains Western blots of the protein loading, washing and elution fractions from the co-elution studies. The fractions collected from the experimental run were blotted for both AlgE and AlgK (panels a and b, respectively). The negative control runs containing only untagged AlgK or N-tagged AlgE were both blotted for AlgK (panels c and d, respectively).
Figure 4.2 Co-elution studies with C-terminally His-tagged AlgK and untagged AlgE. Western blots of the protein loading, washing and elution fractions from the co-elution studies. The fractions collected from the experimental run was blotted for both AlgK and AlgE (panels a and b, respectively). The negative control runs containing only untagged AlgE or C-tagged AlgK were both blotted for AlgE (panels c and d, respectively).

Figure 4.3 Co-elution studies with N-terminally His-tagged AlgK and untagged AlgE. Western blots of the protein loading, washing and elution fractions from the co-elution studies. The fractions collected from the experimental run was blotted for both AlgK and AlgE (panels a and b, respectively). The negative control runs containing only untagged AlgE or N-tagged AlgK were both blotted for AlgE (panels c and d, respectively).
4.3.2. Gel filtration studies

When AlgK and AlgE were incubated in a 1:1 or 3:1 ratio mixture then run over a gel filtration column, the elution profile gave two peaks that superimposed well with the proteins run individually on the column and no peak corresponding to a complex was observed (Figure 4.4 and 4.5). Interestingly, when run in a 1:3 ratio of AlgK to AlgE, 3 peaks were observed in the elution profile: the peak farthest to the right at 14.9 ml corresponded to AlgK, the middle peak at 12.9 ml corresponded to AlgE and, presumably, the peak on the left at 11.9 ml also corresponded to AlgE (Figure 4.6). Unfortunately, there was insufficient protein to run the 0:3 control of AlgE at the time. When the same amount of AlgK was run by itself on the column, i.e. 1:0 ratio AlgK to AlgE, both the UV280nm signal for the baseline and the AlgK peak in the chromatogram superimposed with their equivalents in the chromatogram of the 1:3 mixture of AlgK to AlgE, showing that the amount of unbound AlgK remained the same in the presence and absence of AlgE. This observation suggests that the peak at 11.9 ml in the AlgK to AlgE mixture did not contain any AlgK and instead may correspond to a different conformational state or an unfolded/misfolded form of AlgE. When run at a lower concentration (i.e. 37µM), the AlgE peak had a shoulder which appears to occur around the same volume as the 11.9 ml peak observed in the 1:3 mixture. This shoulder is typically observed when AlgE purified in DDM is run on a gel filtration column and contains a protease-sensitive misfolded/unfolded form of AlgE (J. Whitney, personal communication). It seems plausible that the change in the concentration favoured this unfolded/misfolded form producing the peak at 11.9 ml; running the peak elution fractions, without boiling, on an SDS-PAGE gel could have provided further insight into
the species of AlgE present in the different peaks. It seems unlikely that the 11.9 ml peak was caused by a change in the oligomeric state of AlgE as the shift appears to be too small.

**Figure 4.4 Chromatogram of AlgK and AlgE mixed in a 1:1 ratio.** The red curve corresponds to the elution profile of AlgK and AlgE mixed in a 1:1 ratio at a concentration of 40 and 37 µM, respectively. The blue and green curves correspond to the elution profiles of AlgE and AlgK, respectively, run individually on the size exclusion column.
Figure 4.5 Chromatogram of AlgK and AlgE mixed in a 3:1 ratio. The red curve corresponds to the elution profile of AlgK and AlgE mixed in a 3:1 ratio. The blue and green curves correspond to the elution profiles of AlgE and AlgK, respectively, run individually on the size exclusion column. The concentration of AlgK in the mixture and when run by itself is 120 µM.

![Chromatogram of AlgK and AlgE mixed in a 3:1 ratio](image1)

Figure 4.6 Chromatogram of AlgK and AlgE mixed in a 1:3 ratio. The red curve corresponds to the elution profile of AlgK and AlgE mixed in a 1:3 ratio. The blue and green curves correspond to the elution profiles of AlgE and AlgK, respectively, run individually on the size exclusion column. The concentration of AlgE in the mixture and when run by itself is 111 and 37 µM, respectively.

![Chromatogram of AlgK and AlgE mixed in a 1:3 ratio](image2)
4.3.3 Western blots comparing Alg protein levels in FRD mutants and FRD1

Cells were grown, harvested and lysed. The whole cell lysates were blotted for different Alg proteins (i.e. AlgG, AlgL, AlgK, AlgE, AlgF, AlgJ and Alg44) and the levels of these proteins were qualitatively compared in FRD1 and the FRD mutants (Table 4.5). The levels of Alg8, AlgX and AlgI could not be analyzed as we do not currently have antibodies against these proteins. At the time these studies were conducted, no ΔalgE mutant was available to investigate the effect of the loss of AlgE on the levels of other proteins. The Western blot analysis was performed once, with the exception of the FRDΔalg44 and FRDΔalgX mutants, which were run twice.

Table 4.5 Summary of stability assay results. Table listing mutant strains investigated and the observed changes in the level of protein present in the mutants compared to FRD1, as well as the figure which contains the corresponding Western blots. The strains probed using the Alg44 antibody can be found in Figure 4.12.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Changes in Protein Levels</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRDΔalgK</td>
<td>AlgL, AlgJ</td>
<td>4.7</td>
</tr>
<tr>
<td>FRDΔalgG</td>
<td>AlgL, AlgJ, AlgE</td>
<td>4.7</td>
</tr>
<tr>
<td>FRDΔalg8</td>
<td>Alg44</td>
<td>4.8</td>
</tr>
<tr>
<td>FRDΔalg44</td>
<td>AlgF, AlgJ</td>
<td>4.8</td>
</tr>
<tr>
<td>FRDΔalgF</td>
<td>Alg44</td>
<td>4.9</td>
</tr>
<tr>
<td>FRDΔalgJ</td>
<td>Alg44</td>
<td>4.9</td>
</tr>
<tr>
<td>FRDΔalgI</td>
<td>Alg44</td>
<td>4.9</td>
</tr>
<tr>
<td>FRDΔalgX</td>
<td>AlgL, AlgF</td>
<td>4.10</td>
</tr>
<tr>
<td>FRDalgKC28S</td>
<td>AlgL</td>
<td>4.11</td>
</tr>
</tbody>
</table>
Figure 4.7 Western blots of FRD1, FRDΔalgK and FRDΔalgG. Western blots of whole cell lysates from FRD1, FRDΔalgK (ΔK) and FRDΔalgG (ΔG) probed for different Alg proteins. The first lane of each blot contains a molecular weight ladder (kDa).
Figure 4.8 Western blots of FRD1, FRDΔalg8 and FRDΔalg44. Western blots of whole cell lysates from FRD1, FRDΔalg8 (Δ8) and FRDΔalg44 (Δ44) probed for different Alg proteins. The first lane of each blot contains a molecular weight ladder (kDa).
Figure 4.9 Western blots of FRD1, FRDΔalgF, FRDΔalgJ and FRDΔalgI. Western blots of whole cell lysates from FRD1, FRDΔalgF (ΔF), FRDΔalgJ (ΔJ) and FRDΔalgI (ΔI) probed for different Alg proteins. The first lane of each blot contains a molecular weight ladder (kDa). The contrast and brightness of the α-AlgL blot was modified to increase the visibility of the bands.
Figure 4.10 Western blots of FRD1 and FRD∆algX. Western blots of whole cell lysates from FRD1 and FRD1algX::Gm (∆X) probed for different Alg proteins. FRD2 is a derivative of FRD1 that carries a spontaneous mutation in the regulatory gene algT/U that turns off alginate production\(^{994}\). The first lane of each blot contains a molecular weight ladder (kDa).
Figure 4.11 Western blots of FRD1, FRDalgKC28S and FRD2. Western blots of whole cell lysates from FRD1, FRDalgKC28S (C28S) and FRD2 probed for different Alg proteins. FRD1 a and b are whole cell lysates originating from two different cultures. The alginate operon is not expressed in FRD2 and thus FRD2 is a negative control. The first lane of each blot contains a molecular weight ladder (kDa).
Figure 4.12 Western blots of FRD1 and FRD mutant strains probed for Alg44. Western blots of whole cell lysates from FRD1, FRD2 and the FRD deletion mutants probing for Alg44 and PilF. The contrast in the $\alpha$-Alg44 Western blot was increased to make the protein bands more visible. The molecular weight ladder is in kDa.

4.3.4 PCR for FRD$\Delta$alg44 and FRD$\Delta$algX mutant

Western blots of the FRD$\Delta$alg44 mutant showed that the Alg proteins appear to be either absent or present at reduced levels relative to those found in FRD1. In the
FRDΔalgX mutant, AlgJ appears to be absent and the levels of AlgL and AlgF are reduced in comparison to FRD1. The genes encoding the affected proteins in the FRDΔalg44 and FRDΔalgX mutants are all found downstream of the Gm cassette. To determine if the decrease in protein levels observed was due to the insertion of the Gm cassette in the wrong orientation, PCR was used to verify the orientation of the Gm cassette in the algD operon of FRDΔalg44 and FRDΔalgX mutants (Figure 4.13, lanes 2 to 7). Primers were designed to amplify the Gm cassette: the forward primer annealed to the 5’ end and the reverse primer to the 3’ end of the Gm gene (Table 4.3). The Gm cassette was amplified when using the genomic DNA for both mutants and produced a product of the expected size, i.e. ~500 bp (Figure 4.13, lanes 2 and 3). To verify that the Gm cassette was in the same orientation as the rest of the alginate operon, reverse primers that annealed to the 5’ end of algK and algL were designed (Table 4.3). The algK and algL genes are directly downstream of the Gm cassette in the FRDΔalg44 and FRDΔalgX mutants, respectively. The PCR reaction for FRDΔalg44 produced a product of expected size if the Gm cassette was inserted in the proper orientation, i.e. ~1400 bp (Figure 4.13, lane 4). The PCR with the FRDΔalgX mutant was attempted twice and never produced a product (Figure 4.13, lane 5). To verify that the absence of a PCR product was not due to the reverse algL primer not annealing to the genomic DNA, a positive control PCR was performed containing the reverse Gm and algL primers which should produce a product of approximately ~ 800 bp; this reaction produced a product (Figure 4.13, lane 7), albeit approximately 400 bp larger in size that what was expected given the information provided regarding the construction of the FRDΔalgX (Mike Harwich, personal communication). Given the FRDΔalgX mutant produces little degraded alginate and our
collaborators inability to fully complement the FRDΔalgX mutant, these results suggest that the Gm cassette is not inserted properly in the FRDΔalgX. A clean FRDΔalgX is currently being constructed (D. Ohman, personal communication).

Figure 4.13 Orientation of the Gm cassette in FRDΔalg44 and FRDΔalgX mutants. 1% agarose gel on which PCR products were separated: lane 1, molecular weight markers (bp); lane 2 and 3, PCR reaction amplifying the Gm cassette in the FRDΔalg44 and FRDΔalgX, respectively; lane 4, PCR reaction to show Gm cassette is inserted properly in the FRDΔalg44 mutant; lane 5, PCR reaction to show Gm cassette is inserted properly in the FRDΔalgX mutant; lane 6, PCR reaction to show Gm cassette inserted backwards in FRDΔalg44; lane 7, PCR reaction to show Gm cassette inserted backwards for FRDΔalgX.

4.4 Discussion

4.4.1 K-E interaction studies

Although two different experimental procedures were tried, no conclusive evidence was obtained to support a direct stable interaction between AlgK and AlgE. These negative results do not preclude the possibility that AlgK and AlgE form a
complex. The interaction between AlgK and AlgE could be weak causing the complex to fall-apart during the gel filtration and pulldown experiments. The absence of the lipid moiety at the N-terminus of AlgK may have caused the protein to adapt a conformation, perhaps similar to what was observed in the crystal structure, occluding AlgE’s binding site. The detergent micelles associated with AlgE could also be interfering with the AlgK-AlgE interaction. Additional experiments to detect an AlgK-AlgE complex are discussed in CHAPTER 5.

4.4.2 Protein stability assays

Upon examination of the Western blots for the stability assays it has become clear that the Alg proteins detected in the different mutants are present in a broad spectrum of levels which are not easily compared to those in FRD1 at a qualitative level alone. In the future a more quantitative approach needs to be used to compare the protein levels such as densitometry. The experiment should be repeated multiple times to get an estimate of the error associated with each measurement, and to determine the statistical significance of the differences between measurements for FRD1 and the mutants. A negative control such as FRD2 or PAO1 should be blotted along with FRD1 and the mutants. After the transfer of whole cell lysates from the SDS-PAGE gel to the PVDF membrane, the membrane should be stained with Ponceau S to show similar loading of whole cell lysates. Once statistically significant decreases are detected, Pulse-Chase studies and reverse transcriptase (rt)-PCR could be used to confirm that the observed changes are due to a loss of protein stability and not to a decrease in transcription levels. The differences
in protein levels in FRD1 and mutants detected by eye and other observations from the current stability assays are discussed in the following sections.

### 4.4.2.1 FRD\(^\Delta \)algK and FRDalgK\(^C28S\)

At a qualitative level, no decrease or loss of any Alg protein was detected in either of the two AlgK mutants, i.e. FRD\(^\Delta \)algK and FRDalgK\(^C28S\). An unprocessed version of AlgK was detected in the FRDalgK\(^C28S\) mutant as expected given it carries a mutation at its putative lipidation/cleavage site. Unexpectedly, an apparent increase in AlgL and AlgJ was observed in FRD\(^\Delta \)algK; an increase in the level of AlgL was also observed in the FRDalgK\(^C28S\) mutant. The increase in the levels of these proteins is not due to promoter of Gm cassette considering that AlgG, a protein downstream of the Gm cassette, appears to be present at similar levels in the mutants and FRD1. The mechanism causing the increase in these proteins remains unknown. AlgK was not detected in the FRD\(^\Delta \)alg44 mutant; it remains unclear whether the loss of AlgK present in the mutant is due to a decrease in its stability or transcription level; additional experiments are underway to clarify this issue. With the exception of the FRD\(^\Delta \)alg44, these results suggest that AlgK does not need to bind other proteins to maintain its stability in vivo and other the Alg proteins probed do not need to interact with AlgK to be stable.

### 4.4.2 Acetylation machinery mutants

Less AlgJ is detected in whole cell lysates of the FRD\(^\Delta \)algI and FRD\(^\Delta \)algF mutant than in FRD1, suggesting that AlgJ may be destabilized by the absence of AlgI and AlgF. Unfortunately, no AlgI antibody is available to test whether the absence of
AlgJ causes a decrease in the AlgI protein and AlgF appears unaffected by the absence of AlgJ or AlgI. All three proteins are required for the acetylation of alginate. AlgF is a periplasmic protein, AlgI is an integral membrane protein predicted to have 7 transmembrane helices and AlgJ is a type II membrane protein tethered to the periplasmic face of the inner membrane by an uncleaved signal peptide\(^\text{183}\). AlgI homologs and type II membrane proteins, such as AlgJ, appear to form a gene cassette which is found in the biosynthetic operon for some other bacterial polysaccharides where they are involved in the esterification of these polysaccharides\(^\text{184}\). Our lab has shown a direct interaction between AlgJ and AlgF: when co-expressed in \textit{E. coli}, AlgJ and AlgF can be co-purified. The decrease in the AlgJ protein in the absence of AlgI or AlgF further supports the idea that these proteins function together and assemble into a complex responsible for alginate acetylation. Interestingly, the levels of AlgL also appear to be decreased in the AlgF mutant and the levels of Alg44 appear to be increased in all acetylation mutants.

### 4.4.3 FRD\(\Delta\text{algX}\)

Western blots of FRD\(\Delta\text{algX}\) whole cell lysates revealed that AlgL and AlgF are present at reduced levels, and AlgJ is absent. The results from the PCR studies suggest that the Gm cassette is not inserted in the proper orientation and, as a result, appears to be interfering with the expression of the downstream genes, including AlgJ, AlgI, AlgL, AlgF and AlgA. This would explain why the AlgX mutant produces very little degraded alginate and cannot be fully complemented for alginate secretion (D. Ohman, personal communication). The \textit{algA} gene is downstream of \textit{algX} and is needed for the production of the alginate precursor, GDP-mannuronate\(^\text{161}\). In the FRD\(\Delta\text{algX}\) mutant, decreased
expression of AlgA limits the overall amount of alginate produced. A clean FRDΔalgX mutant will be generated by our collaborator.

### 4.4.4 FRDΔalg44

In the FRDΔalg44 mutant, AlgF, AlgJ are present at reduced levels and AlgG, AlgL, AlgE and AlgK are absent. One possible explanation is that the loss of Alg44 results in the destabilization of the other proteins. It was previously shown that AlgE is absent in FRDΔalg44 and could be partially complemented by providing Alg44 in trans\(^{174}\). Another possibility is that the loss of Alg44 or the presence of the Gm cassette results in a decreased expression of these proteins. PCR revealed that the Gm cassette is inserted in the proper orientation. Rt-PCR could be used to compare the levels of transcription of the alginate operon in FRD1 and FRDΔalg44, and could confirm that the Gm cassette is not having any polar affects on the downstream genes.

### 4.4.5 FRDΔalgG

AlgL, AlgJ, AlgE appear to be present at reduced levels in the FRDΔalgG mutant. Both AlgJ and AlgL are located downstream of the Gm cassette in the FRDΔalgG mutant, however it seems unlikely that the decrease in the levels of these proteins is due to polar affects of the cassette given that AlgF, a protein encoded by a gene downstream of the Gm cassette, appears to be present in the mutant at a similar level as FRD1. The reduced levels of AlgL are unexpected given the FRDΔalgG secretes low-molecular weight alginate degraded by AlgL. These results suggest the loss of AlgG results in the
destabilization AlgL, AlgJ and AlgE. The decrease in AlgJ by the loss of AlgG suggests that the acetylation machinery participates in the secretion complex.

4.5 Conclusions

The preliminary results presented in this chapter failed to conclusively show that AlgK forms a stable complex with other proteins. Additional experiments to identify AlgK binding partners and further characterize the alginate biosynthetic complex are proposed in CHAPTER 5. Interesting observations were made from the stability assays presented, however these experiments need to be repeated using a more quantitative and statistically rigorous method to measure and compare the protein levels in FRD1 and the mutants. Additional studies are needed to clarify whether the changes in protein levels observed in the FRDΔalg44 are due to a decrease in the stability or transcription of the algD operon.
CHAPTER 5

SUMMARY OF RESULTS AND FUTURE DIRECTIONS

5.1 Summary of Results

At the beginning of this study, AlgK was known to be processed in vivo to a 50 kDa protein and secreted into the periplasm. Its putative 27 residue signal sequence appeared to carry a lipobox consensus sequence suggesting that AlgK is a periplasmic lipoprotein. In an attempt to gain further insight into the function of AlgK, an AlgK deletion mutant was generated in the FRD1 background. FRDΔalgK cells do not secrete high-molecular weight alginate, instead they secrete low-molecular weight uronic acids which were later identified as AlgL degradation products. Given that the FRDΔalgG and FRDΔalgX mutants displayed a similar phenotype, these proteins along with AlgK and AlgL, were proposed to assemble into a periplasmic conduit for the secretion of alginate. AlgK was predicted to fold into an all α-helical protein and carry four TPR or SLR motifs.

5.1.1 AlgK is an outer membrane lipoprotein

Our collaborator used C\textsuperscript{14}-palmitic acid labeling and sucrose gradient fractionation to show that AlgK is a lipoprotein tethered to the outer membrane. Through studies conducted on other outer membrane lipoproteins in E. coli and P. aeruginosa, we
can infer that AlgK’s production follows a similar path: 1) the unfolded preprotein is translocated across the inner membrane by the Sec translocase; 2) at the periplasmic face of the inner membrane, AlgK is lipidated and its signal sequence cleaved; 3) the mature AlgK lipoprotein is released from the inner membrane, transported across the periplasm and inserted into the outer membrane by proteins in the Lol system\textsuperscript{295}. Our results show that despite its unusual lipoprotein signal sequence, AlgK is processed and sorted to the outer membrane. Lipoprotein sorting signals in \textit{Pseudomonas} are complex and not well understood, however they appear to be determined by residues at position +2, +3 and +4, where residue +1 is defined as the lipidated N-terminal cysteine of the mature lipoprotein\textsuperscript{296,297}. The findings of this study show that alanine, glycine and leucine at positions +2, +3 and +4 define an outer membrane sorting signal in Pseudomonads.

\textbf{5.1.2 AlgK promotes the proper localization of the porin AlgE}

Sucrose gradient fractionation of FRD1 and FRD\textit{ΔalgK} cells showed that in the absence of AlgK, AlgE localizes to both the inner and outer membrane. These results show that AlgK is involved in the localization of AlgE to the outer membrane and is consistent with the function of other outer membrane lipoproteins in outer membrane protein biogenesis. Through the study of other integral outer membrane β-barrels, we can put forth a model describing how AlgE gets to the outer membrane: the Sec machinery translocates AlgE across the inner membrane, periplasmic chaperones then shuttle AlgE to the outer membrane where the BAM complex folds and inserts it into the membrane\textsuperscript{298}. It is not clear where along this pathway and how AlgK is involved in AlgE
biogenesis, *i.e.* whether it is involved in translocation of the AlgE to the outer membrane or its folding and/or insertion into the membrane.

### 5.1.3 Structure of AlgK

The structure of AlgK was determined using X-ray crystallography. *P. aeruginosa* AlgK was the original crystallization target. Constructs were generated that expressed AlgK in the periplasm and cytoplasm requiring the development of two distinct purification protocols. No diffraction quality crystals of the full-length or a protease-resistant core of *P. aeruginosa* AlgK were obtained. AlgK orthologues from *P. syringae* and *P. fluorescens* were expressed, purified and screened for crystallization conditions. AlgK from *P. fluorescens* yielded crystals that diffracted to a resolution of 2.5 Å. To increase the phasing power, a SeMet triple mutant of AlgK was expressed, purified and crystallized. The crystal structure of AlgK was solved using the Se-SAD technique and revealed that AlgK folds into 22 α-helices that pack into a right-handed superhelix. AlgK is structurally-similar to TPR-containing proteins and its helices carry the TPR consensus sequence. Interestingly, AlgK’s TPR motifs differ somewhat from canonical TPR motifs: they vary in length between 34-40 residues and some display atypical packing angles. AlgK carries a total of 9.5 TPR-like motifs. Given the structural similarity of AlgK to both TPR and SLR-containing proteins, we propose that a new more general consensus sequence should be used to define both motifs. The conformational heterogeneity of the four molecules of AlgK present in the asymmetric unit, and the variability in the quality of the electron density across a single molecule and between molecules suggest that AlgK
is an inherently flexible protein, a feature observed in other long TPR-containing proteins\textsuperscript{227,228,235}.

5.1.4 AlgK acts as a scaffold for the assembly of the alginate biosynthetic complex

In addition to its role in the localization of AlgE, data suggests that AlgK is a component of the alginate biosynthetic complex: AlgK, along with AlgG, AlgL and AlgX, are implicated in the assembly of a periplasmic conduit for the secretion of alginate\textsuperscript{138,186,188}. In light of these findings along with the presence of 9.5 TPR-like motifs in AlgK, we hypothesize that AlgK acts as a scaffold for the assembly of this periplasmic conduit. Mapping highly conserved residues onto our model of AlgK revealed three putative sites of protein-protein interaction: two at the N-terminus and one at the C-terminus. We hypothesize that the C-terminus of AlgK binds to one or multiple periplasmic or inner membrane proteins, such as AlgX, AlgG, AlgL and/or Alg44, to form the periplasmic conduit. We further hypothesize that AlgE interacts with the N-terminus of AlgK resulting in the assembly of a novel exopolysaccharide secretin. In this secretin, AlgE would transport alginate across the outer membrane and AlgK would bind to multiple proteins for the assembly of the periplasmic conduit and act as a linker between the inner and outer membrane alginate transporters. That AlgE and AlgK correspond to the components of a secretin is supported by the identification of two putative outer membrane proteins, \textit{i.e.} PgaA and BcsC, implicated in the production of poly β-1-6-GlcNAc and cellulose, respectively, that are predicted to carry an N-terminal TPR domain followed by a C-terminal porin domain\textsuperscript{40,247}.
We propose that the biosynthetic complex responsible for alginate production is distinct from those responsible for capsular polysaccharide production where AlgE acts as the outer membrane transporter, Alg8 and Alg44 as the inner membrane transporter, and AlgK along with other periplasmic proteins act as the linker between the inner and outer membrane transporters. The biosynthetic systems responsible for the production of cellulose and poly \( \beta \)-1-6-GlcNAc in gram negative bacteria, display interesting similarities to the alginate biosynthetic complex suggesting these exopolysaccharides are produced by a related mechanism.

5.1.5 Identification of AlgK binding partners

Several *in vitro* experiments were conducted in attempts to identify AlgK binding partners. Protein stability assays were performed to identify AlgK binding partners by comparing Alg protein levels in the FRD\( \Delta \)algK mutant to those in FRD1. The level of AlgK present in different FRD1 Alg mutants was also investigated. Preliminary results suggest that, at a qualitative level, the absence of AlgK does not lead to a loss or decrease in the level of other Alg proteins. Interestingly, AlgK is not detected in the FRD\( \Delta \)alg44 mutant; additional studies are underway to determine if the loss of AlgK is due to a decrease in its stability or transcription. The stability studies are on-going and the preliminary results were presented and discussed. In addition to the stability assays, purified AlgK and AlgE from *P. aeruginosa* were tested for interaction by affinity and gel filtration chromatography; the presence of a stable AlgK-AlgE complex could not be demonstrated using these techniques.
5.2 Future directions

5.2.1 AlgK’s role in biogenesis of AlgE

Questions that could shed further light on the mechanism through which AlgK promotes the localization of AlgE in the outer membrane include the following:

1) Is AlgK involved in the folding and insertion of AlgE into the membrane?

There are multiple experiments that could help answer this question. One could determine if AlgE associated with the outer membrane of FRDΔalgK cells is properly folded and inserted into the membrane. There is evidence in the literature that unfolded β-barrel proteins can also associate with membranes. Given that the folded and unfolded forms of AlgE run at different molecular weights on an SDS-PAGE gel, i.e. ~37kDa and 55kDa, respectively, running outer membrane fractions from FRDΔalgK and FRD1 on an SDS-PAGE gel could help determine if the AlgE protein associated with the outer membrane of FRDΔalgK cells is folded and inserted.

The folding and insertion of AlgE into the membrane of FRDΔalgK cells could also be assessed by its surface exposure. Our AlgE antibody does not recognize the folded protein thus prohibiting the use of immunofluorescence or immunoelectron microscopy. Instead, extracellular modifying agents could be used to assess the surface accessibility of AlgE. To do this, whole cells would be incubated with protease, e.g.
proteinase K, or N-hydroxysuccinimide (NHS)-biotin and Western blots would be used to assess cleavage or chemical modification of AlgE, respectively\textsuperscript{300,301}. AlgE has been detected on the cell surface of mucoid \textit{P. aeruginosa} cells by NHS-biotinylation\textsuperscript{195}.

Given that unfolded recombinant AlgE can refold into detergent micelles and planar lipid bilayers\textsuperscript{195,197}, it would be interesting to see if the presence of AlgK during refolding into detergent or lipid bilayers causes an increase in the yield of the folded form of AlgE or modifies its folding kinetics. A gel shift assay could be used to monitor the folding state of AlgE.

2) Does AlgK interact with unfolded AlgE?

Methods to show an interaction between AlgK and unfolded AlgE include isolating a soluble AlgK-AlgE complex after rapid dilution of chemically unfolded AlgE in the presence of AlgK or after the co-expression of AlgK and AlgE in \textit{E. coli}; by itself, AlgE is expressed in inclusion bodies in \textit{E. coli}. Another method to test for an interaction between AlgK and AlgE would be to run a far Western using AlgK and unfolded AlgE.

3) Does AlgK need to be lipidated and tethered to the outer membrane to promote the proper localization of AlgE?

To explore the importance of lipidation and membrane localization in AlgK function, a \textit{P. aeruginosa} FRD1 derivative was built expressing AlgK carrying a C28S mutation, \textit{i.e.} FRD\textit{algKC28S}. AlgK was predominantly degraded in the FRD\textit{algKC28S}
cells, although a small amount of a higher molecular weight form of the protein, which was presumably the unprocessed protein, i.e. AlgK plus its signal sequence, was detected in whole cell lysates (D. Ohman, personal communication). FRDalgKC28S cells displayed a similar phenotype as FRDΔalgK, i.e. secretion of degraded alginate and mislocalization of AlgE. The lipidation of AlgK is not needed to maintain the stability of the protein given an unlipidated and stable form of mature AlgK was expressed and purified from the periplasm of *E. coli* cells for crystallographic studies. Presumably, the C28S mutation leads to the degradation of AlgK by interfering with its processing at the inner membrane. The results from this experiment failed to answer whether AlgK must be lipitated and localized at the outer membrane to promote the proper localization of AlgE. An alternative strategy would be to test the ability of a soluble non-lipidated periplasmic form of AlgK to complement alginate secretion and AlgE outer membrane localization in the AlgK deletion mutant. This could be accomplished by building a complementation vector that would express AlgK carrying a type I signal sequence, e.g. signal sequence from AlgF, AlgX or AlgG, enabling it to be secreted into the periplasm in a soluble form.

### 5.5.2 Identification and/or validation of AlgK binding partners

As discussed in CHAPTER 4, we hypothesize that AlgK acts as a scaffold for the assembly of the alginate biosynthetic complex. To do so, AlgK would need to bind to multiple proteins including AlgE in the outer membrane, and periplasmic and/or inner membrane Alg proteins. Several experiments, described in CHAPTER 5, were conducted however they failed to identify/confirm putative AlgK binding partners. Although
numerous experiments exist to detect protein-protein interactions and isolate protein complexes, affinity co-purification performed in combination with membrane-soluble crosslinking reagents have been successful in the identification of transenvelope complexes in both capsular polysaccharide systems\textsuperscript{264,277}. Similar experiments could help identify AlgK binding partners and establish the existence of a transenvelope multi-protein complex for alginate biosynthesis. This experiment would entail complementing the FRD\textDelta algK mutant with a tagged version of AlgK, incubating the cells with a membrane-permeable crosslinker and finally purifying the cross-linked complex via its affinity tag. Western blots and mass spectrometry could be used to identify proteins co-purified with AlgK. Showing that AlgK can bind to Alg proteins other than AlgE would further support that this protein plays a dual role in alginate production: as a chaperone involved in AlgE localization and as a stable component of the alginate biosynthetic complex.

Another method to identify AlgK binding partners involves comparing the localization of Alg proteins in FRD1 to FRD\textDelta algK using sucrose gradients. We hypothesize that AlgK may interact with one or more periplasmic proteins such as AlgX, AlgL and/or AlgG. Ideally, in the presence of AlgK, its binding partners would be found in the periplasm and associated with the outer membrane, while the absence of AlgK in the FRD\textDelta algK would abolish the outer membrane localization of these proteins. This technique was successfully used in the Group II capsular polysaccharide system to show the assembly of a complex at the inner membrane\textsuperscript{271}.

Once putative binding partners of AlgK are identified using either of these two methods, additional experiments such as isothermal titration calorimetry or surface
plasmon resonance using purified protein or bacterial two-hybrids will need to be conducted to determine which proteins interact directly with AlgK.

AlgK could also bind to alginate as it is being translocated through the periplasm. A procedure for the production of mannuronate di-, tri- and tetrasaccharides has been established in the lab and the binding of these fragments to AlgK could be tested using isothermal titration calorimetry.

5.5.3 PgaA, BcsC and AlgK/AlgE as a novel secretins

To validate our hypothesis that PgaA, BcsC and AlgK/E correspond to a novel secretin, the PgaA and BcsC proteins must be further characterized. The structural predictions suggesting that PgaA and BcsC have an N-terminal TPR domain and a C-terminal β-barrel domain need to be confirmed by solving their crystal structure. The outer membrane localization of both PgaA and BcsC needs to be established, especially given BcsC is not predicted by the SignalP program to carry a signal sequence. Also, a role for BcsC in the secretion of cellulose across the outer membrane needs to be shown; ideally, in the absence of BcsC, cellulose would accumulate in the periplasm of bacterial cells as is observed for poly β-1-6-GlcNAc in the absence of PgaA²⁴⁷.
APPENDIX 1

ALGK IS AN OUTER MEMBRANE LIPOPROTEIN THAT PROMOTES THE PROPER LOCALIZATION OF ALGE

A1.1 Overview

AlgK is a protein encoded in the algD operon that plays an important role in the secretion of high-molecular weight alginate. The FRDΔalgK mutant secretes low-molecular weight uronic acids identified as AlgL degradation products. AlgK is expressed in the cytoplasm as a 52.5 kilodaltons (kDa) preprotein that is processed within cells to a 50 kDa mature protein localized in the periplasm. AlgK’s signal sequence carries a putative lipobox where C28 in the preprotein corresponds to the site of lipidation and, after cleavage by signal peptidase II, becomes the N-terminus of the mature protein, i.e. C1. The experiments described in this chapter show that AlgK is a lipoprotein tethered to the outer membrane and that it is involved in the localization of the AlgE porin.

Acknowledgements: Dr. Sumita Jain, a former graduate student in Dr. Dennis E Ohman’s lab performed the lipidation study. Mike Harwich, a former graduate student in the Ohman lab conducted the sucrose gradients and the Western blots of the membrane.
fractionations. I calculated the succinate dehydrogenase activities, and interpreted the results and prepared the figures.

The data presented in this chapter will be included in the following publication: Carrie-Lynn Keiski, Michael Harwich, Sumita Jain, A. Mirela Neculai, Patrick Yip, Howard Robinson, Lori L. Burrows, Dennis E. Ohman and P. Lynne Howell. AlgK is a TPR-containing outer membrane lipoprotein involved in alginate biosynthesis and the localization of the AlgE porin. Manuscript in preparation.

**A1.2 Materials and Methods**

**A1.2.1 Palmitic acid labeling**

An overnight culture of FRD1 was subcultured in 10 ml modified alginate-producing (MAP) medium\(^{185}\) to an optical density at 600 nm (OD\(_{600}\)) of approximately 0.5, at which point 10 µCi of \(^{14}\)C-palmitic acid was added to the medium. Cells were cultured for an additional 90 min, then harvested and resuspended in 150 µl of Laemmli sample buffer. Samples were separated using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE, 12%), then transferred to a polyvinylidene fluoride (PVDF) membrane and exposed to film (Kodak). A similar protocol was used to radiolabel and detect AlgK expressed in *E. coli* JM109(DE3) cells transformed with pSJ172, a pET28 vector expressing AlgK.

**A1.2.2 Membrane fractionation**
Overnight cultures of FRD1 and the FRD$\Delta$algK mutant were subcultured in 4 l of L broth (10 g tryptone, 5 g yeast extract, 5 g NaCl, per liter) at 37 °C with vigorous shaking to an OD$_{600}$ of 1.0. The FRD$\Delta$algK mutant, also referred to in the literature as FRD1100, is an algK deletion mutant made in the FRD1 background where the algK gene was replaced by a gentamicin (Gm) resistance cassette$^{189}$. Cells were harvested and fractionated using sucrose gradient density fractionation following the protocol described by Hancock and Nikaido$^{302}$ with the following modifications: 1) tris(hydroxymethyl) aminomethane (Tris)-HCl pH 8.3 instead of Tris-HCl pH 8.0 was used to wash cells and resuspend cells prior to lysis; 2) cells were passed through a French press twice at 40,000 lb/in$^2$ instead of 15,000 lb/in$^2$; 3) for the second sucrose gradient, cells were spun for 18 hrs instead of 14. The four resulting fractions correspond to the inner membrane (IM), outer membrane (OM1 and OM2) and a mixture of the inner and outer membranes (M).

**A1.2.3 Succinate dehydrogenase activity**

Succinate dehydrogenase (SDH) activity in membrane fractions was measured using a modified version of the protocol described in Kasahara and Anraku$^{303}$. For each membrane fraction, 1 ml of reaction mixture contained 50 mM Tris-HCl pH 8.3, 4 mM sodium cyanide (NaCN), 40 mM disodium succinate, 0.04 mM 2, 6-dichloro phenolindophenol (DCPIP), 0.2 mM phenazine methosulfate (PMS) and 0.1 mg of total membrane protein. The components of the reaction mixture were added in the following order (Tris-HCl, disodium succinate, NaCN, water and protein) then incubated for 1 min at 21 °C before adding DCPIP and PMS to start the reaction. The decrease in absorbance at 600 nm was measured for 3 min. Data points in the linear range for each reaction were
used to calculate SDH activity. SDH activity was defined as nanomoles of DCPIP reduced per minute per milligram of protein.

**A1.2.4 Western blots for membrane fractionations**

The total protein in each membrane fraction was measured using the Bradford assay (Bio-Rad kit). Similar amounts of total protein for each membrane fraction were separated using a 12% SDS-PAGE gel, then transferred to PVDF membranes and blocked overnight in Tris-buffered saline (TBS, 100 mM Tris pH 7.0, 0.9% (w/v) NaCl,) containing 0.05% (v/v) Tween-20 and 1% (w/v) bovine serum albumin (BSA). The following day the membranes were incubated with anti-AlgE or anti-AlgK antibodies for 1 h in TBS containing 0.05% (v/v) Tween-20 and 1% (w/v) BSA at dilutions 1:40,000 or 1:20,000, respectively. The membranes were washed three times for 15 min in TBS containing 0.05% (v/v) Tween-20 and then incubated with a goat anti-rabbit antibody conjugated to horseradish peroxidase (HRP, Sigma) diluted 1:40,000 in TBS containing 0.05% (v/v) Tween-20 and 1% (w/v) BSA for 1 hr at 21 °C. The blots were subsequently washed three times in TBS containing 0.05% (v/v) Tween-20 then incubated with HRP substrate (SuperSignal West Pico Chemiluminescent Substrate, Pierce), dried and exposed to film (Kodak) for 5 min.

**A1.3 Results**

**A1.3.1 AlgK is a lipoprotein**
To confirm that AlgK is lipidated in vivo, we supplemented *P. aeruginosa* FRD1 cultures with C\(^{14}\)-labeled palmitic acid. However, this resulted in most of the expressed proteins being labeled because the bacteria catabolized the palmitic acid (data not shown). The experiment was repeated in *Escherichia coli* JM109 (DE3) carrying a plasmid expressing full-length AlgK, the autoradiogram revealed a band C\(^{14}\)-labeled by palmitic acid of similar size to AlgK that was absent in cells transformed with the vector alone (Figure A1.1a). This result confirms that AlgK is lipidated in vivo.

**A1.3.2 AlgK is tethered to the outer membrane**

Lipoproteins can be tethered to either the inner or outer membrane. To determine whether AlgK is retained in the inner membrane or sorted to the outer membrane, sucrose gradient fractionation was used to obtain enriched inner and outer membrane samples of FRD1 cells, which were then probed for AlgK using specific antisera. We measured the activity of the inner membrane enzyme succinate dehydrogenase (SDH) as a fractionation control. We observed AlgK only in the outer membrane-enriched fractions (Figure A1.1b). Our results, coupled with previous data, show that AlgK is a periplasmic lipoprotein that is sorted to the outer membrane.

**A1.3.3 AlgK promotes the proper localization of AlgE in the outer membrane**

Given the role that lipoproteins play in the insertion of proteins into the outer membrane, we investigated whether the absence of AlgK affected the localization of AlgE, the β-barrel porin responsible for transporting alginate across the outer membrane. Sucrose density fractionation of FRD1 cells revealed that AlgE is only present in the
outer membrane fractions (Figure A1.1b). This is in sharp contrast to FRDΔalgK cells where AlgE is found in both inner and outer membrane fractions. A comparison of the SDH activities of the inner and outer membrane-enriched fractions showed that this AlgE mislocalization was not due to the poor separation of membrane compartments.

Figure A1.1 Lipidation and subcellular localization of AlgK. (a) Autoradiogram of E. coli cells grown in the presence of C14-palmitic acid and transformed with either the empty vector or a vector expressing full-length AlgK. The positions of molecular weight markers (kDa) are shown on the left of the autoradiogram. (b) Membrane fractions of FRD1 cells (top and middle panels) blotted for AlgK and AlgE, and FRDΔalgK cells (lower panel) blotted for AlgE. The membrane fractions correspond to the inner membrane (IM), outer membrane 1 (OM1), outer membrane 2 (OM2) and a mixture of inner and outer membranes (M). Succinate dehydrogenase (SDH) activity is presented numerically beneath the corresponding membrane fractions.

A1.4 Discussion

A1.4.1 Palmitoylation experiment
Although our results strongly suggest that AlgK is lipidated, additional controls could have been conducted to show this unequivocally. A Western blot could have been performed to show that the labeled band corresponded to AlgK and not to another protein of similar molecular weight. Globomycin could have been used to further establish that AlgK is being lipidated. Globomycin is a specific inhibitor of signal peptidase II, the enzyme responsible for cleaving the signal peptide of prolipoproteins. Globomycin treatment would prevent the cleavage of AlgK’s signal peptide producing a slower-migrating band on an SDS-PAGE gel.

**A1.4.2 Outer membrane lipoprotein biogenesis**

The results presented in this chapter show that AlgK is an outer membrane lipoprotein. Lipoproteins are synthesized in the cytoplasm as preproteins carrying an N-terminal signal peptide that directs the unfolded protein to the SecYEG translocon for transport across the inner membrane (Figure A1.2). Once in the periplasm, the preprotein remains tethered to the inner membrane by the uncleaved signal peptide. Lipoprotein signal peptides carry a lipobox consensus sequence around the signal peptidase cleavage site consisting of L-(A/S)-(G/A)-C occupying positions -3 to +1 where the N-terminal cysteine of the mature protein is in position +1. AlgK’s signal sequence is 27 residues long and carries an unusual lipobox sequence, i.e. L-A-A-G-C in position -4 to +1. The prelipoprotein is processed by the sequential action of three inner membrane enzymes: 1) Lgt creates a thioether linkage between diacylglycerol and the cysteine at position +1; 2) LspA, also referred to as signal peptidase II, cleaves the signal peptide N-terminal to cysteine +1 and 3) Lnt aminoacylates the N-terminal cysteine.
The sorting signal which determines whether a lipoprotein is retained in the inner membrane or released and translocated to the outer membrane is located within the N-terminal residues of the mature lipoprotein and appears to be species-specific. In *E. coli*, aspartate at position +2 is an inner membrane retention signal\(^{295}\). The sorting signals in *P. aeruginosa* are more complex with residues at position +2, +3 and +4 determining the membrane localization of the protein. Aspartate at position +2 also acts as an inner membrane retention signal in *P. aeruginosa*, however this is modulated by the identity of the residues at positions +3 and +4. Serine and lysine at position +3 and +4, respectively, have been shown to act as inner membrane retention signals in *P. aeruginosa*\(^{296,297}\). AlgK has alanine, glycine and leucine residues at position +2, +3 and +4, respectively and was shown to localize to the outer membrane. These residues are highly conserved in all *Pseudomonas* homologues of AlgK for which sequences are available.

In both *E. coli* and *P. aeruginosa* it has been demonstrated that the release of mature lipoproteins from the inner membrane and their translocation to the outer membrane is accomplished by the Lol system comprised of 5 proteins, LolA through \(E\)\(^{295,309}\). LolCDE form a complex in the inner membrane that utilizes ATP to catalyze the transfer of the lipoprotein from the inner membrane to LolA, a periplasmic protein that forms a soluble complex with the lipoprotein and transports it to the outer membrane\(^{310,311}\). At the periplasmic face of the outer membrane, LolA transfers the lipoprotein to LolB, itself an outer membrane lipoprotein, which tethers the mature lipoprotein to the outer membrane\(^{310,312}\).
Figure A1.2 Model for biogenesis of AlgK and other outer membrane lipoproteins. Based on a figure from Tokuda (2009)\textsuperscript{307}. See text for description of process.

A1.4.3 Outer membrane β-barrel protein biogenesis

AlgE is an integral outer membrane β-barrel protein\textsuperscript{195,197}. Similarly to outer membrane lipoproteins, integral outer membrane proteins are synthesized in the cytoplasm and transported in an unfolded form across the inner membrane by the Sec machinery (Figure A1.3)\textsuperscript{306}. Periplasmic chaperones such as SurA, Skp and DegP bind to the proteins secreted by the Sec complex and transport them to the β-barrel assembly machinery (BAM) complex located in the outer membrane\textsuperscript{298,313}. The BAM complex is essential for the folding, assembly and insertion of outer membrane β-barrels\textsuperscript{314,315}. The BAM complex consists of BamA, a 12-stranded β-barrel with an N-terminal periplasmic domain carrying five POTRA (Polypeptide-TRansport-Associated) domains that is encoded in the genomes of all gram-negative bacteria sequenced to date\textsuperscript{315-317}. In \textit{E. coli},
BamA has been shown to form a complex with four outer membrane lipoproteins, *i.e.* BamB-E, referred to as BAM accessory components\(^{314,318}\).

**Figure A1.3 Biogenesis of AlgE and other outer membrane β-barrel proteins.** Based on Figure from Knowles *et al.* (2008)\(^{298}\). See text for description of process.

All Bam proteins play a role in outer membrane biogenesis, however only BamA and BamD are essential for viability and have been shown to interact directly with each other\(^{314,315,319}\). BamD is predicted to carry 6 tetratricopeptide repeat (TPR) motifs, a protein-protein interaction motif involved in the assembly of multi-protein complexes\(^{242}\). How BamA, the accessory lipoproteins and periplasmic chaperones coordinate outer membrane protein folding, assembly and insertion remains unclear\(^{298}\).

**A1.4.4 AlgK’s role in AlgE biogenesis**

The results presented in this chapter show that AlgK is involved in the localization of AlgE to the outer membrane. In the presence of AlgK, AlgE is exclusively
seen in the outer membrane. However, in the absence of AlgK, AlgE is associated with both the inner and outer membrane. AlgK’s role in AlgE localization correlates well with the function of other outer membrane lipoproteins in outer membrane biogenesis, e.g. the transport and/or insertion of proteins, capsular polysaccharides and lipopolysaccharides into the outer membrane\textsuperscript{264,312,314,320}. As discussed above, outer membrane lipoproteins such as LolB and the BAM accessory components have been implicated in outer membrane lipoprotein and β-barrel protein biogenesis, respectively. Outer membrane lipoproteins, referred to as pilotins, are involved in the assembly of Type II and III secretion systems, as well as Type IV pili biogenesis where they play an important role in the stabilization, multimerization and/or proper localization of the secretin. Secretins are integral outer membrane proteins that assemble into homomultimeric structures of 12-14 subunits that are up to 1 MDa in size\textsuperscript{320}. The exact role of pilotins in secretin biogenesis is complex and not well understood as it appears to vary depending on the secretion system and organism. Interestingly, a model has been proposed where the pilotin-secretin complex utilizes the Lol system to be shuttled to the outer membrane\textsuperscript{321}.

How AlgK promotes the localization of AlgE in the outer membrane is unknown. It is not clear where AlgK is involved in AlgE’s biosynthetic pathway. AlgK could be involved in the translocation of the protein across the periplasm as is proposed for some pilotins, however this seems unlikely given some AlgE is associated with the outer membrane of FRDΔalgK cells. Another possibility is that AlgK acts in conjunction with the BAM complex promoting the folding and/or insertion of AlgE into the outer membrane. In both scenarios, the absence of AlgK could cause a “back-up” in the AlgE biogenesis pathway, leading to the accumulation of AlgE at the inner membrane, as is
observed, either in a misfolded/aggregated form or tethered to the membrane by an uncleaved signal peptide\textsuperscript{322}. It is also unknown whether the AlgE associated with the outer membrane of FRD\textit{ΔalgK} is properly folded and inserted into the membrane; unfolded/misfolded transmembrane β-barrel proteins have been shown to associate with membranes\textsuperscript{299}.

Biological data suggests that AlgK, along with AlgG, AlgL and AlgX, assemble into a periplasmic channel for alginate secretion\textsuperscript{138,186,188}. These data, along with the findings presented in this chapter suggest that AlgK plays a dual role in alginate production: it is involved in AlgE biogenesis and is a stable component of the alginate biosynthetic complex, a role discussed in CHAPTER 3.

**A1.5 Conclusions**

In APPENDIX 1, we have shown that AlgK is an outer membrane lipoprotein that promotes the proper localization of AlgE to the outer membrane. Similarly to other integral outer membrane β-barrel proteins, AlgE is thought to be shuttled across the periplasm, folded and inserted into the outer membrane by the concerted action of periplasmic chaperones and the BAM complex. It is not clear where along this pathway AlgK could act to promote AlgE’s localization. Additional studies are needed to clarify this issue. These data suggest that AlgK plays two roles in alginate biosynthesis: it promotes the localization of AlgE to the outer membrane and it is a member of a periplasmic channel that translocates alginate across the periplasm.
REFERENCES


