Structural and Functional Studies of Regulatory Proteins in the Bacterial Exopolysaccharide Synthesis and Transport Pathways

by

Chelsy Caryn Chesterman

A thesis submitted to the Biochemistry graduate program in the Department of Biomedical and Molecular Sciences in conformity with the requirements for the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
September 2015

Copyright ©Chelsy Caryn Chesterman, 2015
Abstract

Extracellular sugar polymers, or exopolysaccharides, play crucial roles in gram-negative bacteria that enable their survival and proliferation in hostile environments. As a result, the proteins involved in the synthesis, transport, and regulation of these sugars are potential targets for antibiotic development.

*Escherichia coli* tyrosine kinase (Etk) is an inner membrane protein that is part of the capsular polysaccharide (CPS) export complex. The autophosphorylation of Etk is considered to be the regulatory signal that controls CPS export, but the mechanism for transmitting this signal to the other members of the protein complex has yet to be elucidated. It is expected that the periplasmic domain of Etk is involved. The first portion of this thesis discusses progress towards determining an x-ray crystal structure for Etk’s periplasmic region. Significant advancement occurred when the unstable full-length Etk protein was truncated to include only the periplasmic domain and the two transmembrane helices. Attempts to improve the reproducibility and quality of crystals obtained provided the realization that detergent concentration was a crucial factor. This led to the development of an improved method for detergent concentration measurement in small-volume, membrane protein samples using a colorimetric reaction with 2,6-dimethylphenol. Other traditional chemical reactions that can be applied to detergent measurement were also tested.

Another important exopolysaccharide produced in *E. coli* is lipopolysaccharide. The second part of this thesis presents the first crystal structure of a newly identified, key regulator of lipopolysaccharide synthesis, lipopolysaccharide assembly protein B (LapB). Zn-SAD phasing with high-redundancy data was used to determine the experimental phases required for structure solution. Retrospective analysis of the same data revealed that this structure could also have been solved with lower redundancy data. The protein contains nine tetratricopeptide repeats and a rubredoxin metal binding domain that fold together in an unexpectedly intimate arrangement. Proper functioning of LapB, including the binding of the rubredoxin domain to the TPR helices appears crucial for optimum cell growth.
Co-Authorship

Chapter 2: Purification, Characterization, and Crystallization of Membrane Bound *Escherichia coli* Tyrosine Kinase

This chapter was submitted for publication in the journal Protein Expression and Purification and was In Press at the time of thesis submission. This project was initiated by Dr. Zongchao Jia. Experiments were designed and carried out by Chelsey Chesterman. The manuscript was prepared by Chelsey Chesterman with editorial input from Dr. Zongchao Jia.

Chapter 3: Quantification of Detergent Using Colorimetric Methods in Membrane Protein Crystallography

This chapter was written in response to an invitation to contribute a book chapter to Methods in Enzymology Volume 557: Membrane Proteins -- Engineering, Purification, and Crystallization received by Dr. Zongchao Jia. Chelsey Chesterman was the primary author of this book chapter with editorial assistance from Dr. Zongchao Jia. The material in this chapter is largely based on two previous publications. A research paper promoting the 2,6-dimethylphenol approach (Prince, C., Jia, Z. Measurement of detergent concentration using 2,6-dimethylphenol in membrane-protein crystallization. Acta Cryst. D. 2012, 68, 1694-6) and a subsequent review paper (Prince, C., Jia, Z. Detergent quantification in membrane protein samples and its application to crystallization experiments. Amino Acids. 2013, 45, 6, 1293-302). Chelsey Chesterman was also the primary author of these previous publications and performed all of the experimental work required.
Chapter 4: An Unexpected Duo: Rubredoxin binds nine TPR motifs to form LapB, an essential regulator of lipopolysaccharide synthesis

This manuscript has been published (Prince, C. Jia, Z. Structure. 2015, 23, 1500-6). Chelsey Chesterman initiated this project, wrote the manuscript, and performed all of the experiments reported. Dr. Zongchao Jia provided advice on all stages of the project and edited the final manuscript.

Chapter 5: Zinc-SAD Phasing with Low Redundancy Data

This document summarizes the key technical points of an educational talk on Zinc-SAD phasing that was given at the American Crystallographic Association Annual Meeting in July 2015 by Chelsey Chesterman. Chelsey Chesterman performed the retrospective analysis that led to the observations reported and prepared the document with input from Dr. Zongchao Jia.

Appendix A: Structural Characterization of Legionella Effector Proteins

This appendix summarizes unfinished and unpublished work performed over the past 4 years. The document was prepared by Chelsey Chesterman who was involved in the majority of the experiments described. The significant contributions of Julie Campbell (high-throughput screening/RavL), Harrison Taylor (RavL), and David Langalaan (NMR data collection) are acknowledged and some of the data they collected is presented in this summary.
Acknowledgements

I wish to express my gratitude to my graduate supervisor, Dr. Zongchao Jia, for enabling the progression of my scientific studies. I have had many opportunities to grow and learn during my graduate program that would not have been possible without his support. I am also grateful to all of the wonderful people I worked with in the Jia Lab. In particular, I would like to recognize Dr. Mark Currie, who was my first supervisor in the Jia Lab. Mark was responsible for my first introduction to protein crystallography as an undergraduate volunteer. I would also like to thank Drs. Laura van Staalduinen and Mona Rahman for being patient sounding boards whenever I need to work through challenging problems. In addition, I appreciate the current efforts of Alexander Andrew, who is continuing my work on some of the Legionella targets. I am excited to see these stories continue to develop! I also need to thank Natalie Roy for providing technical assistance to the lab.

During my time at Queen’s, I have had the opportunity to work with numerous undergraduate and high-school students who assisted with experiments and the day-to-day prep work required. In no particular order, I would like to thank Yujia Zhu, Katherine Sedivy-Haley, Julie Campbell, Harrison Taylor, Cassandra Yoon, Daniel Schep, Kuba Zawadzki, Tyler Moore, and Luke Housell. I wish all of you the best of luck in your future careers.

I have relied on many skilled technicians throughout my graduate research including Kim Munro and David McLeod from the Protein Function Discovery Facility at Queen’s University, Richard Gillian at the Cornell High Energy Synchrotron Source, and the beamline operators that make data collection possible at the Advanced Photon Source. The Legionella project also benefited from the support of our collaborators in the Kingston-Montreal Bacterial Structural Genomics initiative who conducted the initial round of cloning. I would also like to acknowledge the Canadian Institutes of Health Research, the Ontario Government, and Queen’s University for funding my graduate research.

I have been very fortunate to have been supported throughout my degree by an amazing family. In particular, I would like to thank my loving parents, Kim and Phil Prince, for always having my best
interests at heart. I would also like to thank my Aunt, Kerry Heller, for providing a place for me to stay while commuting to Toronto for coursework.

Finally, and most importantly, I would like to thank my loving husband, Julian Chesterman, for being a constant source of love and encouragement. His unwavering dedication and generosity are a constant source of inspiration. I will be forever grateful for his patience as he carefully proofread every page of this thesis.
Statement of Originality

I hereby certify that all of the work described within this thesis is the original work of the author. Any published (or unpublished) ideas and/or techniques from the work of others are fully acknowledged in accordance with the standard referencing practices.

Chelsey Chesterman

September, 2015
# Table of Contents

Abstract .................................................................................................................................................. ii
Co-Authorship ........................................................................................................................................ iii
Acknowledgements ................................................................................................................................. v
Statement of Originality ........................................................................................................................... vii
List of Figures ......................................................................................................................................... xiii
List of Tables .......................................................................................................................................... xvi
List of Abbreviations ............................................................................................................................. xvii

Chapter 1 Introduction: Synthesis and Transport of Exopolysaccharides in Gram-Negative Bacteria...... 1
1.1 Overview ........................................................................................................................................... 1
1.2 Structure of the Gram-Negative Cell Envelope ................................................................................. 1
1.3 Exopolysaccharides in Gram-Negative Infections ............................................................................. 3
1.4 Capsular Polysaccharides .................................................................................................................. 4
  1.4.1 Wzy-dependent CPS Assembly and Export ................................................................................. 5
  1.4.2 Regulation of the Wzy Dependent Mechanism ........................................................................... 10
  1.4.3 ABC-dependent Transport ....................................................................................................... 12
1.5 Biofilm Polysaccharides .................................................................................................................... 14
  1.5.1 Cellulose and PGA .................................................................................................................. 14
  1.5.2 Colanic Acid ............................................................................................................................ 17
1.6 Lipopolysaccharide ............................................................................................................................ 18
  1.6.1 Structure of LPS ....................................................................................................................... 18
  1.6.2 Synthesis of Kdo-Lipid A ......................................................................................................... 20
  1.6.3 Assembly and Export of Lipopolysaccharide ........................................................................... 28
  1.6.4 Regulation of Lipopolysaccharide ............................................................................................ 34
1.7 Summary .......................................................................................................................................... 37
1.8 Hypothesis and Objectives ................................................................................................................ 38

Chapter 2 Purification, Characterization, and Crystallization of Membrane Bound *Escherichia coli*
Tyrosine Kinase ........................................................................................................................................ 39
  2.1 Foreword ........................................................................................................................................ 39
  2.2 Abstract ........................................................................................................................................ 39
  2.3 Introduction .................................................................................................................................... 40
Chapter 3 Quantification of Detergent Using Colorimetric Methods in Membrane Protein Crystallization

3.1 Foreword ............................................................................................................................. 65
3.2 Abstract ................................................................................................................................. 65
3.3 Detergent in Membrane Protein Crystallography ................................................................. 66
3.4 2,6-Dimethylphenol Assay for Sugar-Based Detergents .................................................... 67
  3.4.1 Overall Protocol ............................................................................................................... 68
  3.4.2 Assay Optimization ......................................................................................................... 69
  3.4.3 Interfering Compounds ..................................................................................................... 73
3.5 Molybdate Assay for Total Phosphate ................................................................................. 74
  3.5.1 Protocol .......................................................................................................................... 75
  3.5.2 Assay Optimization ......................................................................................................... 76
3.5.3 Interfering Compounds ................................................................. 77
3.6 Assay of Bile Salt with Sulfuric Acid ............................................. 78
  3.6.1 Assay Optimization & Interfering Compounds .......................... 79
3.7 Applications of Detergent Measurement ........................................ 81
  3.7.1 Detergent Content and Crystallizability .................................. 81
  3.7.2 Reducing Detergent Concentration in Centrifugal Concentrators .... 83
  3.7.3 Detergent Exchange ................................................................. 85
3.8 Concluding Remarks ..................................................................... 87

Chapter 4 An Unexpected Duo: Rubredoxin Binds Nine TPR Motifs to Form LapB, an Essential Regulator of Lipopolysaccharide Synthesis .................................................. 89
  4.1 Foreword ..................................................................................... 89
  4.2 Abstract ..................................................................................... 89
  4.3 Introduction ............................................................................... 90
  4.4 Results and Discussion ............................................................... 92
    4.4.1 Optimized LapB and Crystallization ..................................... 92
    4.4.2 LapB Structure ...................................................................... 94
    4.4.3 Mutations in the TPR-Rubredoxin Interface Inhibit Cell Growth in vivo .... 98
    4.4.4 LapB Mutants are Folded ..................................................... 101
    4.4.5 LapB Mutants Alter Local Structure ................................... 103
    4.4.6 Conclusion .......................................................................... 104
  4.5 Methods .................................................................................. 104
    4.5.1 Protein Expression and Purification of Soluble LapB ............... 104
    4.5.2 Competitive Metal Binding .................................................. 105
    4.5.3 Crystallization and X-Ray Diffraction ................................... 105
    4.5.4 Data Reduction and Model Building .................................... 105
    4.5.5 Plasmid Construction ......................................................... 106
    4.5.6 Bacterial Strains and Growth Conditions ............................. 106
    4.5.7 Plating Efficiency Assay ...................................................... 106
    4.5.8 Tryptophan Fluorescence ................................................... 107
    4.5.9 CD Spectroscopy ................................................................. 107
Chapter 5 Zinc-SAD Phasing with Low Redundancy Data .............................................................. 109
  5.1 Foreword .............................................................................................................................. 109
  5.2 Abstract ............................................................................................................................... 109
  5.3 Introduction ......................................................................................................................... 110
  5.4 Preparation of Protein Crystals ......................................................................................... 112
  5.5 Data Collection and Data Production in XDS ................................................................. 112
  5.6 Experimental Phasing ....................................................................................................... 112
  5.7 Automated Model Building in Phenix ............................................................................. 113
  5.8 Discussion ........................................................................................................................... 116
  5.9 Conclusions ....................................................................................................................... 117

Chapter 6 Discussion .................................................................................................................. 118
  6.1 Characterization of Proteins in the Post-Genomics Era ................................................ 118
  6.2 Predicting Protein Structure .............................................................................................. 118
    6.2.1 Phyre2 Prediction of LapB Structure ........................................................................ 118
    6.2.2 Phyre2 Prediction of Etk Periplasmic Domain Structure ........................................ 120
  6.3 Challenges in Structural Biology ....................................................................................... 122
    6.3.1 Spontaneous Proteolysis in Protein Biochemistry ..................................................... 122
    6.3.2 Membrane Proteins and the Detergent Conundrum ................................................. 123
    6.3.3 Difficult Crystals ......................................................................................................... 124
      6.3.3.1 Etk NM2 Prisms .................................................................................................. 124
      6.3.3.2 LapB Plate Clusters ......................................................................................... 125
  6.4 Determining Biological Assemblies .................................................................................. 125
    6.4.1 Debate Over the Correct Oligomeric State for Etk .................................................. 125
    6.4.2 Conflicting Evidence for the Oligomerization State of LapB .................................. 126
  6.5 Regulation of Exopolysaccharide Production and Transport in Gram-Negative Pathogens .... 129
    6.5.1 Etk in a Bi-membrane Complex ............................................................................. 129
    6.5.2 LapB as a Protein Scaffold ..................................................................................... 132
  6.6 Applying the Study of Cell Envelope Regulation: Antibiotic Research ....................... 135
    6.6.1 Potential Antibiotic Targets in Exopolysaccharide Regulons .................................. 136
      6.6.1.1 LapB in the Lipopolysaccharide Pathway ...................................................... 136
      6.6.1.2 E. coli Tyrosine Kinase ................................................................................. 137
  6.7 Summary and Future Perspective .................................................................................... 138
Appendix A Structural Characterization of *Legionella* Effector Proteins ................................................. 139
A.1 Foreword .......................................................................................................................................................... 140
A.2 Introduction .................................................................................................................................................. 140
A.3 Selection and Cloning of Targets for Structural Characterization .......................................................... 141
A.4 Expression and Solubility Trials .................................................................................................................. 141
A.5 Large Scale Investigations of Soluble Targets .......................................................................................... 144
A.6 RavL: A Putative Lipase with an Alpha-Beta Hydrolase Fold ................................................................. 144
    A.6.1 Protein Expression and Purification ..................................................................................................... 144
    A.6.2 Biophysical Characterization of RavL ................................................................................................. 146
    A.6.3 Modeling of RavL with Phyre2 ........................................................................................................... 147
    A.6.4 Enzymatic Activity of RavL ............................................................................................................... 149
    A.6.5 Re-Engineering RavL for Crystallization ......................................................................................... 150
    A.6.6 Crystallization of RavL ....................................................................................................................... 151
A.7 PieF, a Tiny Protein of Unknown Structure and Function ........................................................................... 153
    A.7.1 Protein Expression and Purification .................................................................................................. 153
    A.7.2 NMR Data Analysis ........................................................................................................................... 156

References .......................................................................................................................................................... 159
List of Figures

Figure 1-1 Extracellular sugars form layers of protection around gram-negative bacteria ........................................ 2
Figure 1-2 Schematic of Wzy-dependent assembly and transport of CPS .............................................................. 5
Figure 1-3 X-ray crystal structures for proteins from the Wzy-dependent mechanism for CPS export............. 7
Figure 1-4 X-ray crystal structures for two proteins of unknown function in group 4 CPS export .......... 9
Figure 1-5 Crystal structures of CPS regulators ................................................................................................. 10
Figure 1-6 Schematic of group 2 polysaccharide assembly and export .......................................................... 13
Figure 1-7 Structure of biofilm polysaccharides .............................................................................................. 14
Figure 1-8 Complex of BesA and BcsB determined by X-ray-crystallography ........................................... 15
Figure 1-9 Schematics of synthase dependent transport ................................................................................. 16
Figure 1-10 A single repeat unit of colanic acid ............................................................................................... 17
Figure 1-11 Overall structure of lipopolysaccharide ....................................................................................... 18
Figure 1-12 LptA catalyzes the first step of lipid A synthesis ........................................................................ 21
Figure 1-13 The second step in lipid A synthesis is performed by LpxC ......................................................... 22
Figure 1-14 LpxD catalyzes the third step of lipid A synthesis ................................................................. 22
Figure 1-15 The fourth step of lipid A synthesis is catalyzed by LpxH or LpxI .............................................. 24
Figure 1-16 The first disaccharide is synthesized by LpxB in the fifth step of the synthesis pathway .... 25
Figure 1-17 The sixth step is catalyzed by LpxK ........................................................................................... 26
Figure 1-18 WaaA adds the Kdo sugars required for final acylation in the seventh step of lipid A synthesis ................................................................................................................................. 27
Figure 1-19 The last two steps in the synthesis of lipid A are the addition of two additional acyl chains by LpxL and LpxM ............................................................................................................ 28
Figure 1-20 Structures of MsbA in different conformations ........................................................................ 29
Figure 1-21 The three mechanisms for assembling and transporting LPS across the inner membrane .... 30
Figure 1-22 Schematic of the attachment of sugar polymers to lipid A in the periplasm and its transfer to the exterior leaflet of the outer membrane ............................................................................................................. 32
Figure 1-23 Lpt proteins of known structure ................................................................................................. 34
Figure 1-24 Regulation of LpxC ................................................................................................................... 36
Figure 2-1 Optimization of detergent required for extraction of Etk FL from the cell membrane .......... 47
Figure 2-2 Radioactive phosphate incorporation into Etk FL ........................................................................ 48
Figure 2-3 Elution profiles for Etk FL solubilized in DDM, DM, FC-12 detergent ......................................... 49
Figure 2-4 Reducing aggregation of Etk FL .................................................................................................. 51
Figure 2-5 Size distribution of Etk FL determined from the sedimentation velocity experiment ........................................... 52
Figure 2-6 SDS-PAGE analysis of individual purification steps in the production of Etk FL .......................................................... 53
Figure 2-7 Mass spectroscopy of Etk FL degradation .................................................................................................................. 54
Figure 2-8 Etk schematics ............................................................................................................................................................. 54
Figure 2-9 SDS-PAGE analysis of the typical elution fractions obtained after nickel affinity purification of Etk NM2. .......................................................................................................................... 55
Figure 2-10 Tryptic fragments from the Etk NM2 construct and degradation product ................................................................. 56
Figure 2-11 Size exclusion of Etk NM2 solubilized in various common detergents used in membrane protein crystallography .................................................................................................................................. 57
Figure 2-12 Representative x-ray diffraction of Etk NM2-DDM crystals to 6.5 Å resolution ....................................................... 58
Figure 2-13 Crystals of Etk NM2 ................................................................................................................................................... 59
Figure 3-1 Colorimetric absorbance peaks for the quantification of sugars ........................................................................... 69
Figure 3-2 Comparison of reaction efficiency between phenol and 2,6-dimethylphenol ............................................................. 70
Figure 3-3 Summary of the phenol/2,6-dimethylphenol reactions ............................................................................................. 71
Figure 3-4 Optimization of the 2,6-dimethylphenol assay ......................................................................................................... 73
Figure 3-5 Reaction of 5 µL control samples with 65% sulfuric acid and 25 µL 2,6-dimethylphenol .............................................. 74
Figure 3-6 Summary of the molybdate assay ................................................................................................................................ 76
Figure 3-7 Optimization of the molybdate assay ......................................................................................................................... 77
Figure 3-8 Total phosphate assay with 5 µL of compatible control samples ...................................................................... 78
Figure 3-9 Summary of the bile salt measurement protocol .................................................................................................... 79
Figure 3-10 Optimization of the bile salt/sulfuric acid assay ...................................................................................................... 80
Figure 3-11 Effect of increasing concentrations of detergent on the phase behaviour of Etk NM2 crystallization drops ......................................................................................................................................................... 81
Figure 3-12 Reproducibility of Etk NM2 crystallization ............................................................................................................. 83
Figure 3-13 Detergent dilution using a centrifugal concentrator ................................................................................................. 85
Figure 3-14 Exchange of DDM to FC-12 ...................................................................................................................................... 87
Figure 4-1 LapB optimization and crystallization ......................................................................................................................... 93
Figure 4-2 LapB structure ............................................................................................................................................................. 95
Figure 4-3 Sequence alignments for key domains in LapB ............................................................................................................. 96
Figure 4-4 Rubredoxin-TPR binding interface ............................................................................................................................. 97
Figure 4-5 Temperature dependence of growth rate .................................................................................................................... 99
Figure 4-6 Plating efficiency assay ................................................................................................................................................ 100
Figure 4-7 In vitro biophysical studies of LapB’s secondary and tertiary folding ................................................................. 102
Figure 5-1 Anchors for successful manual model building ..................................................................................................... 114
Figure 6-1 Top 18 templates selected by Phyre2 in intensive mode for the LapB protein sequence …… 119
Figure 6-2 Phyre2 poorly models the structure of LapB …………………………………………………………… 120
Figure 6-3 Predicted structure of Etk’s N-terminal periplasmic domain ………………………………………… 121
Figure 6-4 The additional helix in one molecule of the LapB asymmetric unit clashes with where the same helix in the other molecule would be if it was present………………………………………………………… 127
Figure 6-5 Oligomerization state of LapB in solution is a tetramer ………………………………………………… 128
Figure 6-6 Potential interaction of Etk and YccZ ……………………………………………………………………… 132
Figure 6-7 Predicted binding sites for LpxC on the LapB protein ………………………………………………… 133
Figure 6-8 Electrostatic model of the LapB surface …………………………………………………………………… 135
Figure A-1 Example SDS-PAGE gels for Legionella small scale expression and purification trials …… 142
Figure A-2 Secondary structure prediction for RavL assembled by XtalPred ……………………………………… 145
Figure A-3 Nickel purification of His-tagged RavL …………………………………………………………………… 145
Figure A-4 RavL stability …………………………………………………………………………………………………… 146
Figure A-5 Models of RavL ………………………………………………………………………………………………… 148
Figure A-6 Intrinsic nickel binding of RavL ……………………………………………………………………………… 148
Figure A-7 Calcium dependent kinase activity of RavL ……………………………………………………………… 149
Figure A-8 Substrate preference of RavL evaluated using phenol red ……………………………………………… 150
Figure A-9 Example SDS-PAGE gels from purification of the RavL 18-369 no tag construct ………………… 151
Figure A-10 RavL crystals …………………………………………………………………………………………………… 152
Figure A-11 Representative diffraction images for RavL ……………………………………………………………… 152
Figure A-12 Secondary structure prediction for PieF assembled by XtalPred ……………………………………… 154
Figure A-13 NHSQC NMR spectrum collected from N-His-TEV-PieF ……………………………………………… 156
Figure A-14 A very preliminary model of PieF ………………………………………………………………………… 157
Figure A-15 Comparison of the secondary structure predicted by XtalPred with the secondary structure calculated from the backbone chemical shifts measured by NMR………………………………………………… 158
List of Tables

Table 1-1 Summary of enzymes involved in the 9-step synthesis of lipid A .................................................. 20
Table 2-1 Statistics for fitting of AUC sedimentation curves ................................................................. 43
Table 2-2 Aggregation state of Etk FL samples ....................................................................................... 50
Table 2-3 Analytical ultracentrifugation of Etk NM2 stabilized in two maltoside-based detergents .......... 58
Table 3-1 Stock solutions required for the total phosphate assay .......................................................... 75
Table 4-1 Data reduction, experimental phasing, model building, and refinement statistics for the structure determination of LapB ........................................................................................................... 94
Table 4-2 Strain construction for in vivo study of LapB function ............................................................ 108
Table 5-1 Recently published structures solved with Zn-SAD phasing .................................................. 111
Table 5-2 Structures deposited in the Protein Data Bank with a bound metal ion (as of July 2015) and the K edge of the respective metals that could have been used to collect anomalous signal .................. 111
Table A-1 Initial target list assembled for Legionella effector proteins ................................................. 143
Table A-2 Outcome of large scale protein expression, purification, and crystallization trials ............. 144
Table A-3 Breakdown of secondary structure components in RavL determined by CD spectroscopy and compared to multiple computational prediction algorithms ......................................................... 147
Table A-4 Full solubility screen of PieF after overnight incubation ..................................................... 155
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>ATPases associated with various cellular activities</td>
</tr>
<tr>
<td>ACP</td>
<td>acyl carrier protein</td>
</tr>
<tr>
<td>AUC</td>
<td>analytical ultracentrifugation</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>CMC</td>
<td>critical micelle concentration</td>
</tr>
<tr>
<td>CPS</td>
<td>capsular polysaccharide</td>
</tr>
<tr>
<td>DDM</td>
<td>n-dodecyl-β-D-maltoside</td>
</tr>
<tr>
<td>DM</td>
<td>n-decyl-β-D-maltoside</td>
</tr>
<tr>
<td>DMNG</td>
<td>decyl maltose neopentyl glycol</td>
</tr>
<tr>
<td>DLS</td>
<td>dynamic light scattering</td>
</tr>
<tr>
<td>DSC</td>
<td>differential scanning calorimetry</td>
</tr>
<tr>
<td>EPS</td>
<td>extracellular polymeric substance</td>
</tr>
<tr>
<td>Etk</td>
<td><em>E. coli</em> tyrosine kinase</td>
</tr>
<tr>
<td>FC-12</td>
<td>n-dodecylphosphocholine</td>
</tr>
<tr>
<td>FC-14</td>
<td>n-tetradecylphosphocholine</td>
</tr>
<tr>
<td>FOM</td>
<td>figure of merit</td>
</tr>
<tr>
<td>GT</td>
<td>glycosyl transferase</td>
</tr>
<tr>
<td>IMAC</td>
<td>immobilized metal affinity chromatography</td>
</tr>
<tr>
<td>Kdo</td>
<td>3-deoxy-D-manno-oct-2-ulosonic acid</td>
</tr>
<tr>
<td>LapB</td>
<td>lipopolysaccharide assembly protein B</td>
</tr>
<tr>
<td>LDAO</td>
<td>lauryldimethylamine-N-oxide</td>
</tr>
<tr>
<td>LMNG</td>
<td>lauryl maltose neopentyl glycol</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharide</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MAD</td>
<td>multi-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>MDR</td>
<td>multi-drug resistant</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>OG</td>
<td>n-octyl-β-D-glucoside</td>
</tr>
<tr>
<td>PCP</td>
<td>polysaccharide co-polymerase</td>
</tr>
<tr>
<td>PDC</td>
<td>protein-detergent complex</td>
</tr>
<tr>
<td>PGA</td>
<td>linear chain of N-acetyl-D-glucosamine residues in β(1,6) linkage</td>
</tr>
<tr>
<td>SAD</td>
<td>single wavelength anomalous dispersion</td>
</tr>
<tr>
<td>SAXS</td>
<td>small angle x-ray scattering</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>size exclusion chromatography</td>
</tr>
<tr>
<td>SeMet</td>
<td>selenomethionine</td>
</tr>
<tr>
<td>TEM</td>
<td>transmission electron microscopy</td>
</tr>
<tr>
<td>TPR</td>
<td>tetratricopeptide repeat</td>
</tr>
<tr>
<td>UDP</td>
<td>uridine-diphosphate</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction:

Synthesis and Transport of Exopolysaccharides in Gram-Negative Bacteria

1.1 Overview

Gram-negative bacteria synthesize and excrete a large array of diverse exopolysaccharides. The synthesis of these sugar moieties occurs primarily in the cytoplasm of the cell and involves dozens of enzymes. These sugars must then be polymerized and transported to the exterior of the cell by large protein complexes that span both cell membranes. The production of exopolysaccharides is regulated during both synthesis and export and can have a major impact on the viability of the bacterium. In this literature review, I will give an overview of the diversity of exopolysaccharides and describe their synthesis and export pathways. This will provide a framework for the discussion of the regulation of these pathways. In this thesis we will focus on the structural and functional characterizations of two exopolysaccharide regulators, Etk from the capsular polysaccharide (CPS) transport pathway and LapB from the lipopolysaccharide (LPS) synthesis pathway. Given the key role of exopolysaccharides in the spread and persistence of gram-negative pathogens, these pathways and their regulators are of potential interest for the development of new antibacterial compounds.

1.2 Structure of the Gram-Negative Cell Envelope

“Gram-negative bacteria” is a general label that refers to many species of bacteria that have a particular cell envelope structure. The term was originally coined by Hans Christian Gram and referred to bacteria that did not retain the stain crystal violet (Bartholomew and Mittwer, 1883). The defining feature of gram-negative bacteria is the structure of their cell envelope, which contains a thin layer
peptidoglycan sandwiched between two lipid bilayers. The inner membrane of the cell envelope is a bilayer structure composed primarily of phospholipids. In contrast, the outer membrane is asymmetric and contains 75% LPS in the outer leaflet (Whitfield and Trent, 2014). LPS contains a lipidic anchor bridged by two Kdo sugars (3-deoxy-d-manno-oct-2-ulosonic acid), which provide attachment points for LPS sugar polymers. Gram-negative bacteria are decorated by a large variety of extracellular sugars including CPS and excreted slime polysaccharides in the extracellular matrix, in addition to the sugar component of LPS (Figure 1-1).

Figure 1-1 Extracellular sugars form layers of protection around gram-negative bacteria. This schematic shows the relative positioning of these layers between the cell and its environment.
1.3 Exopolysaccharides in Gram-Negative Infections

Gram-negative bacteria have many key components that play crucial roles during gram-negative infections. These virulence factors can include biofilm formation, exopolysaccharides, cell motility, secretion systems, and effector proteins. Collectively, these properties promote survival of the bacteria in hostile environments and can perpetuate infections.

The gram-negative cell wall is an important virulence factor during infection. The outer leaflet of the outer membrane is composed primarily of LPS in gram-negative bacteria. The tight packing of LPS in the membrane provides a permeability barrier, which means that it is able to exclude many hydrophilic and hydrophobic small molecules from entering the cell including environmental toxins and antibiotics (Nikaido, 2003). Therefore, this structure is an essential component of the innate immune system in gram-negative bacteria. LPS is also a strong stimulant of immune responses in their hosts, which leads to the core component of LPS, lipid A also being known as endotoxin. Pico molar concentrations of LPS can be detected by the TLR4 receptor of innate immune cells including macrophages and endothelial cells (Raetz and Whitfield, 2002). The immune response to LPS can result in inflammation, fever, and nausea and is the source of many of the symptoms experienced during infection with gram-negative bacteria.

Extracellular sugars form the outer-most layer of the cell and are an important site of host-pathogen interactions. Protective layers on the cell surface can include exopolysaccharide and CPS. These polysaccharides can mediate interactions of the cell with its surroundings, prevent desiccation, and provide resistance to both non-specific and specific host immunity (Roberts, 1996). CPS is extremely variable with ~80 known serotypes from *Escherichia coli* alone (Orskov et al., 1977; Whitfield, 2006). Capsule diversity can correlate with the success of the bacteria during infection. For example, *Bacteroides fragilis* is only present in 1-2% of gastrointestinal flora in humans, but is the predominant isolate in soft-tissue infections (Pumbwe et al., 2006). *B. fragilis* has also been recently identified as having the highest known CPS diversity in gram-negative bacteria (Patrick et al., 2010). CPS can form a “shield” over the
cell that prevents the immune system from recognizing factors in the lipopolysaccharide and preventing phagocytosis and subsequent destruction of the bacteria (Whitfield, 2006).

A biofilm is a collection of bacteria that have adhered to a surface and are commonly encased by a self-produced layer of extracellular polymeric substances (EPS), also referred to as the biofilm matrix. The matrix contains 97% water in addition to polysaccharides, lipids, nucleic acids, and metabolites (Beloin et al., 2008). Biofilms are created to provide tolerance to environmental stressors and protection from toxic and host immunological defenses. There are several important polysaccharides that can play a role in the biofilm matrix including colanic acid, cellulose, and PGA (polymer of N-acetyl-D-glucosamine) (Japlan et al., 2004; Re and Ghigo, 2006; Sailer et al., 2003). Unlike capsular polysaccharides and lipopolysaccharides, biofilm polysaccharides are not serotype specific. Cell surface attached polysaccharides in CPS and LPS can also play indirect roles in biofilm formation or inhibition (Beloin et al., 2008).

1.4 Capsular Polysaccharides

Given the large variety of capsular polysaccharides (CPS), there is also a large range of biosynthesis and sugar modification pathways to facilitate the manufacture of specific sugar subunits in the cytoplasm. A thorough discussion of all of these enzymes is beyond the scope of this review. CPS is classified into four groups based on structural, biochemical, and genetic criteria (Whitfield, 2006). Despite the variety of surface polysaccharides and their synthesis mechanisms, only two mechanisms of assembly and transport have been described. The Wzy dependent mechanism is used for group 1 and 4 CPS and the ABC transporter dependent mechanism is used for groups 2 and 3 (Whitfield and Naismith, 2008). In addition, capsules from group 1 and 4 can be found in *Escherichia coli* species that cause intestinal infections, while groups 2 and 3 are primarily found in bacteria strains that cause extraintestinal infections (Whitfield, 2006).
1.4.1 Wzy-dependent CPS Assembly and Export

The Wzy dependent mechanism for the synthesis and transport of CPS is very similar between group 1 and group 4 polysaccharides, but homologous proteins with different names catalyze some of the steps. The following description is based on group 1 CPS synthesis and transport with characterized homologues in the group 4 machinery mentioned as appropriate. The machinery for the synthesis, assembly, and export of groups 1 and 4 CPS has been proposed to form a large multi-protein complex that spans both membranes (Figure 1-2).

![Figure 1-2 Schematic of Wzy-dependent assembly and transport of CPS. (A) Group 1 CPS. (B) Group 4 CPS. GT = glycosyltransferase.](image)

The process begins with the assembly of individual repeat units onto a carrier lipid in the inner membrane by the activity of glycosyltransferase enzymes (Roberts et al., 1999). In many species, group 1 CPS synthesis is initiated by WbaP glycosyltransferase, which transfers galactose-1-phosphate or glucose-1-phosphate to an undecaprenyl phosphate carrier lipid in the inner membrane. In contrast, group 4 CPS synthesis is initiated by the transfer of N-acetylglucosamine-1-phosphate by the WecA
glycosyltransferase to the same carrier lipid (Whitfield, 2006). Both these initiating enzymes contain transmembrane helices and are located in the inner membrane. The capsular polysaccharide repeat units are subsequently completed by additional glycosyltransferases that are peripheral membrane proteins. The lipid-linked sugar repeats are then flipped across the inner membrane by Wzx (Marolda et al., 2006; Roberts et al., 1999). Polymerization in the periplasm is carried out by Wzy (Drummelsmith and Whitfield, 1999; Roberts et al., 1999). Wzy is an integral membrane protein with ~12 transmembrane helices and two large periplasmic loops of comparable size (Islam et al., 2010). Wzy probably uses a catch and release polymerase mechanism that has been recently purposed (Islam et al., 2011). Wzy may also be responsible for releasing the completed polymer from the lipid carrier once it has reached an appropriate size (Schild et al., 2005; Whitfield, 2006). A third inner membrane protein, Wzz, acts as the primary chain length determinant and ensures that the characteristic modal lengths are produced (Raetz and Whitfield, 2002). Wzz proteins contain a large periplasmic domain flanked by two transmembrane helices and are part of the polysaccharide copolymerase-1 family (PCP-1) (Morona et al., 2000). Crystal structures of some of the PCP-1 family members have been determined and revealed oligomeric rings and protein folds with large coiled-coil domains (Kalynych et al., 2012; Tocilj et al., 2008).

Another essential member of the system, Wzc, interacts with the outer membrane export channel and may also function in chain length determination (Collins et al., 2007; Roberts et al., 1999). A low-resolution 3D tetrameric structure of Wzc has been determined using electron microscopy. The shape of the overall structure has been compared to a molar tooth, with the N-terminal domains responsible for oligimerization and the C-terminal kinase domains clearly separated (Collins et al., 2006). The low-resolution electron microscopy study has been followed-up by two crystal structures. The first, was of the monomeric kinase domain from Etk, the group 4 homologue of Wzc (Lee et al., 2008). The second crystal structure was of the same kinase domain from Wzc, this time in an octameric ring conformation (Bechet et al., 2010) (Figure 1-3AB).
Figure 1-3 X-ray crystal structures for proteins from the group 1 and 4 Wzy-dependent mechanism for CPS export. (A) Octamer of Wzc kinase domain – 3LA6, crystallized with ADP shown as spheres. (B) Monomer of Etk - 3C1O. (C) Octamer of Wza - 2J58. (D) Monomer of Wzi - 2YNK.

The polymerized capsular polymer is shuttled to the exterior of the cell through the outer membrane by Wza (Roberts et al., 1999). Wza (YccZ in group 4) is a member of the outer membrane auxiliary protein family. Early structural studies determined that Wza forms SDS-stable multimers and has a ring structure of 8 identical subunits when viewed by electron microscopy (Beis et al., 2004; Drummelsmith and Whitfield, 2000; Nesper et al., 2003). This allowed for the overall shape of the protein to be described
as mushroom-like with two rings stacked on top of each other, a four-fold symmetry, and a central open cavity (Beis et al., 2004). The high-resolution crystal structure of Wza (2.26Å) further describes this structure as 4 octameric rings with 8-fold symmetry (Dong et al., 2006) (Figure 1-3C). Interestingly, the crystal structure demonstrates that Wza uses an amphiphatic helical barrel to transverse the outer membrane in contrast to the beta barrel found in other outer membrane proteins. The cavity in the protein was found to be lined with polar residues that show little sequence conservation and water molecules bind simultaneously with the sugar to assist in the formation of hydrogen bonds (Dong et al., 2006). The ability of CPS to enter the channel of Wza has been recently confirmed using non-natural photo-cross-linkable amino acids to trap translocation intermediates (Nickerson et al., 2014). One aspect of the channel mechanism that has not been resolved is how the channel opens at the bottom to allow sugar polymers to enter. Wza does not conduct ions in isolation suggesting it is normally closed and it is logical that tight regulation of the export of polysaccharides through the opening and closing of the channel would be desirable (Dong et al., 2006).

Finally, Wzi is an outer membrane protein that is considered essential for the attachment of group 1 capsules to the exterior of the cell, but is not present in the group 4 pathway (Whitfield, 2006). A crystal structure of Wzi has been recently solved (Figure 1-3D) revealing an 18-strand β-barrel with long extracellular loops that block the entrance of the channel and are also required for the recognition of CPS (Bushell et al., 2010, 2013). The attachment mechanism for CPS to the exterior of the cell is not fully understood, although it has been recently proposed that ionic interactions with core sugar residues in LPS are responsible (Jiménez et al., 2012). Shorter chains of capsule sugar polymers can also be covalently attached to lipid A by WaaL to form K_LPS and be excreted by the Lpt protein system as LPS (Whitfield, 2006). A detailed discussion of LPS structure and excretion will be covered later.

So far, the working assumption is that the synthesis of group 4 CPS uses Wzx, Wzy, and Wzz for inner membrane transport, or similar proteins that have not yet been characterized. There is an additional
locus of genes expressed with Etk (Wzc), YccZ (Wza), and Etp (Wzb) proteins for group 4 CPS synthesis (Figure 1-2B). This locus also contains open reading frames for 4 additional proteins known as GfcA-D or YmcA-D dependent on the source (Peleg et al., 2005; Sathiyamoorthy et al., 2011). All of these proteins are currently of unknown function.

GfcA is a small inner membrane protein that is rich in threonine, alanine, and glycine residues (Sathiyamoorthy et al., 2011). GfcB is an outer membrane protein whose structure has been determined (PDB 2IN5), but it bares no homology to group 1 CPS export proteins and its function remains unknown (Figure 1-4). GfcD is also an outer membrane protein that is predicted to form a large outer-membrane β-barrel with 22 strands, but has no structure and no putative function has been reported (Sathiyamoorthy et al., 2011). Finally, a structure of the periplasmic GfcC protein was recently published which indicates that it has a very similar fold to the periplasmic domains of the outer membrane channel, Wza/YccZ (Sathiyamoorthy et al., 2011), but the role of this additional protein is also unclear (Figure 1-4). It has been determined that these poorly characterized proteins are essential for group 4 CPS production (Peleg et al., 2005) and represent a clear point of divergence between group 1 and group 4 polysaccharide assembly and export.

Figure 1-4 X-ray crystal structures for two proteins of unknown function in the group 4 CPS export pathway. (A) GfcC – 3P42 (B) GfcB – 2IN5.
1.4.2 Regulation of the Wzy Dependent Mechanism

The *in vivo* regulation of Wzy-dependent capsular transport is expected to involve autophosphorylation of the Wzc/Etk kinase in the inner membrane, although there have been no reports of this protein directly binding CPS. Kinase activation in Etk has been described as a two-step mechanism starting with the phosphorylation of a tyrosine near the active site (Lee et al., 2008). This phosphorylated tyrosine is then able to interact with a nearby arginine residue, opening the active site and leading to the phosphorylation of up to 7 tyrosines in a C-terminal cluster (Lee et al., 2008; Paiment et al., 2002) (Figure 1-5A). This activation mechanism is supported by mutational studies and molecular dynamic simulations that identified an additional arginine residue that modulates the opening of the active site (Lee et al., 2008; Lu et al., 2009).

![Figure 1-5 Crystal structures of CPS regulators. (A) Etk – 3CIO, with C-terminal tyrosine cluster in yellow (only 4 of 7 potential positions are visible in crystal structure), catalytic site tyrosine in cyan, and regulatory arginine residues in green. * denotes the arginine residue able to bind the phosphorylated tyrosine. # denotes the second arginine implicated by molecular dynamic simulations. (B) Wzb – 2WJA. Protein co-crystallized with a phosphate ion in red and orange, and a divalent cation, Ni, in green.](image)
The impact of phosphorylation on CPS export is central to current efforts to understand the mechanism of this enzyme. A mutational study of Wzc aimed to understand the role of phosphorylation by testing CPS export by a broad range of mutants. The results indicated that loss of any 1 or 2 tyrosine residues from the C-terminal cluster does not impair function where the loss of any 4 or more destroys function (Paiment et al., 2002). The final conclusion was that the overall level of phosphorylation in the tail is more important for the protein’s function than the specific residues phosphorylated. There is also evidence that cycling between phosphorylated and non-phosphorylated states may be essential (Nadler et al., 2012; Wugeditsch et al., 2001). De-phosphorylation of Wzc is accomplished by the cytoplasmic phosphatase Wzb, whose structure is also available (Hagelueken et al., 2009a) (Figure 1-5B).

It has been proposed that the level of phosphorylation could influence interaction between the C-terminal domains of the Wzc tetrameric complex leading to conformational change (Lee et al., 2008). This conformational change could then be transmitted across the inner membrane to the outer membrane channel protein to facilitate its open/closed state. Support for the existence of a periplasm-spanning complex starts with transmission electron microscopy (TEM) data that shows CPS exported where the inner and outer membrane are closer together (Collins et al., 2007). In addition, chemical cross-linking in vivo results in a large multimeric complex containing Wza and Wzc (Nesper et al., 2003), and complementation experiments also indicate that specific interactions occur between Wza and Wzc (Reid and Whitfield, 2005). Finally, the most convincing evidence of a trans-membrane complex is a 3D-structure of the Wza-Wzc complex solved using electron microscopy to ~12Å (Collins et al., 2007). Interaction between Wza and Wzc occurs in a 20Å band, but specific interactions are difficult to identify since the proteins appear to have undergone a large conformational change that has significantly widened the channel in this area. Mechanistically, the widening of this structure implies an opening of Wza at the bottom of the channel, which is in contrast to the closed structure of Wza in isolation and indicates that binding with Wzc could be responsible for channel opening (Collins et al., 2007). Wzc was also
determined to be in a heterogeneously phosphorylated state in the protein complex with 0-3 residues phosphorylated. The significance of this level of phosphorylation will not be known until there is a dissimilar case for comparison. Finally, it will be important to confirm the observations made in this structure using other techniques, because the conformational change seen could be a side effect of the cryo-negative staining process (Collins et al., 2007). Another question to be answered is how an 8-fold symmetrical channel interacts with the tetrameric kinase.

1.4.3 ABC-dependent Transport

Group 2 and 3 capsular polysaccharides are transported by ABC-dependent assembly and transport mechanisms (Figure 1-6). In both cases, the proteins involved appear similar, but there is little direct information for group 3 biogenesis and most of what is known is inferred from conservation between the two groups (Whitfield, 2006). As in the Wzy-dependent mechanism, capsule assembly begins with glycosyl transferases in the cytoplasm of the cell that assemble the serotype specific repeat unit. In the case of group 2 CPS these sugars are most commonly N-acetylmuraminic acid, N-acetylgalactosamine, and rhamnose, while group 3 CPS contains rhamnose, glucuronic acid, and 4,6-dideoxy-4-malonylaminoglucose (Whitfield, 2006). Assembly begins with an unclear initiator that is then extended by the activity of the NeuS family of glycosyltransferase enzymes (Whitfield, 2006). There are multiple proteins in the cytoplasm of E. coli that are also essential to preparing capsule sugar for transport including KspF, KpsU, KpsC, and KpsS. Mutations in these genes have been shown to accumulate products in the cytoplasm (Bronner et al., 1993; Cieslewicz and Vimr, 1996). The finished group 2 CPS is capped by diacylglycerophosphate at the reducing terminus and this lipid is likely directly linked to the last N-acetylmuraminic acid residue (Gotschlich et al., 1981).

The iconic feature of this pathway is the use of an ABC-transporter in the inner membrane. In E. coli, the nucleotide binding domains of this protein are the two chains of KpsT in the cytoplasm. Two chains of KpsM form the membrane-spanning module for sugar transport (Silver et al., 2001). This transport
couples the hydrolysis of ATP to the movement of the lipid linked group 2 capsule across the inner membrane.

Figure 1-6 Schematic of group 2 polysaccharide assembly and export. There are a large number of essential proteins in the cytoplasm that presumably interact with CPS, but their exact function is currently unknown. GT= glycosyltransferase

Once in the periplasm, group 2 polysaccharide is transported to the exterior of the cell by KpsD, an outer membrane channel protein that is coupled to the rest of the biosynthetic machinery by the periplasmic protein KpsE. Mutations in either of these proteins result in the accumulation of CPS in the periplasmic space (Bronner et al., 1993; Pazzani et al., 1993). The production of group 2 capsules is thermally regulated at the level of transcription in *E. coli*, while group 3 capsules are produced at all growth temperatures. In group two, thermoregulation appears to be the result of a complex regulatory network involving a number of global regulators in *E. coli* (Rowe et al., 2000).
1.5 Biofilm Polysaccharides

1.5.1 Cellulose and PGA

Cellulose is a linear polymer composed of glucose connected by β1-4 linkages (Nishiyama et al., 2003) and PGA is a β-1,6-linked N-acetyl-d-glucosamine homopolymer assembled from UDP-N-acetylglucosamine precursors (Whitney and Howell, 2013) (Figure 1-7). Cellulose is synthesized by a synthase dependent mechanism (Hashimoto et al., 2011) in the inner membrane by the membrane spanning complex BcsA/BcsB (Morgan et al., 2013). The structure of this complex was recently solved and demonstrates that BcsA forms both the glycosyltransferase in the cytoplasm and contains the transmembrane helices required to form a pore for cellulose transport across the inner membrane. The BcsB structure reveals that it contains two carbohydrate binding domains and two domains with folds similar to flavodoxin of unknown function (Morgan et al., 2013). This structure also caught a snapshot of the D-glycan chain being transported from the cytoplasm to the periplasm through the inner membrane channel portion of the synthase (Figure 1-8). In addition, UDP, which is the product after glucose is transferred from UDP-glucose to the growing glycan chain, was found in the glycosyltransferase active site. The full export of cellulose from the cell is still not fully understood, but involves multiple additional proteins across the two membranes (Figure 1-9A).

![Figure 1-7 Structure of biofilm polysaccharides. (A) Cellulose is a polymer of β1,4-glucose. (B) PGA is a polymer of β1,6-N-acetylglucosamine.](image)
Figure 1-8 Complex of BcsA and BcsB determined by X-ray-crystallography (4HG6). A polymer of cellulose has been trapped in the membrane-spanning channel (green). In addition, we see two detergent molecules in the transmembrane (TM) region and UDP in the glycosyltransferase active site. (B) Structure of BcsZ – 3QXQ with cellopentose in the active site.
BcsC is a large protein exported to the periplasm that contains an N-terminal tetratricopeptide repeat (TPR) domain and a large C-terminal β-barrel that likely embeds in the outer-membrane to form the exit channel for cellulose (Römling and Galperin, 2015). BcsC is required for the synthesis of cellulose in vivo, but not in vitro (Wong et al., 1990). The pathway for PGA assembly and export is synthase dependent and contains many of the same elements as the cellulose pathway (Whitney and Howell, 2013), but most have not been thoroughly characterized (Figure 1-9B).

Cellulose assembly may be mediated by another component of the machinery, BcsZ, a periplasmic endo-β-1,4-glucanase which degrades excess polymer (Mazur and Zimmer, 2011). Crystal structures have been determined for this enzyme with and without a 5-mer of cellulose trapped in the active site (Figure 1-8B). Cellulose production could also be regulated by BcsE, which binds c-di-GMP in the cytoplasm (Fang et al., 2014; Römling and Galperin, 2015). This is in addition to the PilZ domain at the C-terminus of BcsA, which also binds c-di-GMP to activate cellulose production (Whitney and Howell, 2013).
1.5.2 Colanic Acid

Colanic acid is a slime polysaccharide that can be an important for biofilm formation and prevention of desiccation, but does not appear to have a significant role in virulence (Whitfield and Paiment, 2003). Colanic acid consists of a repeating structure composed of fructose, glucose, galactose, and glucuronic acid (Stevenson et al., 1996) as shown in Figure 1-10. The synthesis and export of colanic acid uses the same pathway as group 1 CPS and strains that produce colanic acid do not co-express a CPS from group 1 (Jayaratne et al., 1993). Unlike CPS, a substantial amount of the colanic acid produced is excreted into the bacteria’s environment, and can contribute to the formation of biofilms. Why colanic acid is secreted while group 1 CPS is not, is unknown (Reid and Whitfield, 2005). As in group 1 CPS, the production of colanic acid is regulated by the phosphorylation state of Wzc and the presence of both phosphorylated and non-phosphorylated Wzc results in colanic acid with a wide size variation. In contrast, fully phosphorylated Wzc prevents production, while continuously dephosphorylated Wzc results in production of colanic acid with a small size distribution (Obadia et al., 2007). Colanic acid biosynthesis genes are part of a regulon that responds to the lifestyle of \textit{E. coli} and can respond to alterations/damage of the cell envelope, osmotic shock, or growth on surfaces in a biofilm (Whitfield, 2006).

![Figure 1-10 A single repeat unit of colanic acid.](image)

17
1.6 Lipopolysaccharide

1.6.1 Structure of LPS

There are three components of LPS, the lipid A anchor, core oligosaccharide, and the extended O-antigen sugars (Figure 1-11). There are $\sim 10^6$ molecules of lipid A in a single cell of *E. coli* and this component of the outer-membrane is essential for cell growth and proliferation (Emiola et al., 2015; Whitfield and Trent, 2014). The minimal lipid A structure consists of two sugar domains that arrange six hydrocarbon chains and its synthesis pathway is highly conserved (Galloway and Raetz, 1990). The main point of variation in lipid A is the length of the hydrocarbon chains, which is determined by the “hydrocarbon ruler” that is found in some lipid A synthesis enzymes for the selection of particular acyl donor chain lengths (Whitfield and Trent, 2014). The length of the fatty acids added by LpxA was originally characterized to be 14 carbons in *E. coli* (Anderson and Raetz, 1987), but *Pseudomonas aeruginosa* LpxA is specific for 10 carbon chains (Dotson et al., 1998), while *Helicobacter pylori* uses a chain length of 16 (Stead et al., 2008). It has also been shown that mutations in the binding pocket for the hydrocarbon ruler can change the chain length specificity (Whitfield and Trent, 2014).

![Lipid A structure](image)

**Figure 1-11 Overall structure of lipopolysaccharide.**
It is rare for bacteria to only produce lipid A. In most cases, lipid A is glycosylated with the core oligosaccharide, which provides the attachment sites for the O-antigen chains (Whitfield and Trent, 2014). The core oligosaccharide can be structurally divided into an inner and an outer region. The inner region typically consists of Kdo and l-glycero-d-manno-heptopyranose saccharides and the outer core is more variable and contains hexoses and hexosamines (Shang et al., 2015; Whitfield and Trent, 2014). There are five core oligosaccharide types that have been identified in *E. coli* and two in *Salmonella* (Raetz and Whitfield, 2002). Modifications to the core oligosaccharide can also occur during stress adaptation including the stress induced addition of rhamnose (Klein et al., 2011) or the addition of 4-amino-4-deoxy-l-arabinose to provide polymyxin resistance (Velkov et al., 2010). The modifications improve the bacteria’s survival in extreme environments and can mask gram-negative pathogens from the host immune-system (Needham and Trent, 2013).

While the lipid A and core oligosaccharide components are highly conserved across different strains, the O-antigens in LPS can be highly variable. There are over 180 recognized O-antigens in *E. coli* and a database has been created to collect and organize this data (Rojas-Macias et al., 2015; Stenutz et al., 2006). The O-antigens are generally not required for growth under permissive conditions (i.e. in strains raised in a laboratory). O-antigens are most commonly found in environmental bacteria where they help the organism resist environmental stresses, host immune systems such as the complement system, and man-made antibiotics (Raetz and Whitfield, 2002). O-antigen repeat units range from two to seven sugars arranged in a linear, branched, or double-branched architecture with four-sugar repeats being the most common (Stenutz et al., 2006). A wide variety of different sugar moieties is used in the assembly of these polymers including hexoses, pentoses, deoxyhexoses, lactyl substituted hexoses, heptoses, and nonuloses (Stenutz et al., 2006).
1.6.2 Synthesis of Kdo₂-Lipid A

Synthesis of the specialized lipid A component of lipopolysaccharide occurs in the cytoplasm of the cell using a 9-step enzyme catalyzed pathway that is summarized in Table 1-1. All of the enzymes involved are constitutively expressed (Raetz and Whitfield, 2002). The biosynthesis of the lipid A moiety is of particular interest because it is the only component of LPS that is essential for the viability of *E. coli* (Emiola et al., 2015). The Lpx enzymes that catalyze these steps are well conserved across LPS producing bacteria. Therefore, this is a major pathway of interest for pharmaceutical development and the regulation of LPS.

Table 1-1 Summary of enzymes involved in the 9-step synthesis of lipid A.

<table>
<thead>
<tr>
<th>Step</th>
<th>Enzyme</th>
<th>Location</th>
<th>Structures</th>
<th>Functional Unit</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LpxA</td>
<td>Cytoplasm</td>
<td>17</td>
<td>Trimer</td>
<td>Acyltransferase</td>
</tr>
<tr>
<td>2</td>
<td>LpxC</td>
<td>Cytoplasm</td>
<td>40</td>
<td>Monomer</td>
<td>Deacetylation</td>
</tr>
<tr>
<td>3</td>
<td>LpxD</td>
<td>Cytoplasm</td>
<td>10</td>
<td>Trimer</td>
<td>Acyltransferase</td>
</tr>
<tr>
<td>4</td>
<td>LpxH or LpxI</td>
<td>Peripheral/Integral Membrane</td>
<td>None</td>
<td>Unknown</td>
<td>Pyrophosphate bond cleavage</td>
</tr>
<tr>
<td></td>
<td>LpxI</td>
<td>Peripheral Membrane</td>
<td>2</td>
<td>Dimer</td>
<td>Inverting Glycosyltransferase</td>
</tr>
<tr>
<td>5</td>
<td>LpxB</td>
<td>Peripheral Membrane</td>
<td>None</td>
<td>Monomer or Dimer</td>
<td>Inverting Glycosyltransferase</td>
</tr>
<tr>
<td>6</td>
<td>LpxK</td>
<td>Peripheral/Integral Membrane</td>
<td>8</td>
<td>Monomer</td>
<td>Nucleotide Triphosphate Hydrolase</td>
</tr>
<tr>
<td>7</td>
<td>WaaA</td>
<td>Integral Membrane</td>
<td>3</td>
<td>Monomer</td>
<td>Glycosyltransferases</td>
</tr>
<tr>
<td>8</td>
<td>LpxL</td>
<td>Integral Membrane</td>
<td>No</td>
<td>Monomer</td>
<td>Acyltransferase</td>
</tr>
<tr>
<td>9</td>
<td>LpxM</td>
<td>Integral Membrane</td>
<td>No</td>
<td>Unknown</td>
<td>Acyltransferase</td>
</tr>
</tbody>
</table>

The first step in the synthesis of lipid A is the addition of the first acyl chain from an acyl-ACP (acyl carrier protein) donor to UDP-GlcNAc (uridine diphosphate N-acetylglucosamine) (Anderson and Raetz, 1987). This reaction is performed by LpxA (Figure 1-12), a trimeric protein that contains an active sites at each interface between the domains (Williams and Raetz, 2007). The acylation catalyzed by LpxA has an...
unfavorable equilibrium constant and therefore this first step does not commit the cell to lipid A biosynthesis (Emiola et al., 2015).

![Chemical reaction and structure diagram](image1.png)

**Figure 1-12** LptA catalyzes the first step of lipid A synthesis. Left: Chemical reaction. Right: Structure of the LpxA trimer (3I3A) with the product, UDP-3-N-(R-3-hydroxylauroyl)-GlcNAc3N bound in the three active sites shown as spheres.

The second step in lipid A synthesis is the deacetylation of the glucosamine sugar that is catalyzed by the metalloamidase LpxC (Figure 1-13). This represents the first committed step in the lipid A synthesis pathway. LpxC activity is dependent on a bound metal cofactor in the active site, which was originally determined to be zinc. In this initial study, it was also determined that the enzyme regained varying degrees of activity when the coordinated metal was substituted with Co\(^{2+}\), Ni\(^{2+}\) or Mn\(^{2+}\) (Jackman et al., 1999). More recently, there has been evidence published that this protein can also use Fe\(^{2+}\) as cofactor, and that when bound to reduced iron the enzyme is 6-8 times more efficient (Hernick et al., 2010). However, LpxC still has significantly greater affinity for Zn\(^{2+}\) relative to Fe\(^{2+}\) (Gattis et al., 2010). Despite this, there is evidence that LpxC does bind Fe\(^{2+}\) *in vivo* because the concentration of readily exchangeable iron can be significantly higher than zinc (Gattis et al., 2010; Hernick et al., 2010). The structure of LpxC
has been determined many times by NMR and crystallography, and it shares no homology with other deacetylases or amidases (Whitfield and Trent, 2014).

Figure 1-13 The second step in lipid A synthesis is performed by LpxC. Left: chemical reaction. Right: Crystal structure of LpxC (3U1Y) in complex with an inhibitor (green) which coordinates the Zn atom (grey) and blocks the active site.

Figure 1-14 LpxD catalyzes the third step of lipid A synthesis. Left: chemical reaction. Right: Trimeric structure of LpxD (2IU8) with palmitate in each of the three active sites. A single molecule of UDP-N-Acetylglucosamine has also been trapped in one of the active sites.

Deacetylation of the glucosamine sugar provides an open amino group that can then be used for addition of the second hydrocarbon chain by LpxD (Figure 1-14). Structure determination has shown that
LpxD has a similar trimeric structure and hydrocarbon ruler as found in LpxA (Badger et al., 2011; Bartling and Raetz, 2009; Buetow et al., 2007).

The fourth step involves the cleavage of the pyrophosphate bond in UDP-2,3-diacyl-GlcN to produce uridine monophosphate and lipid X (Figure 1-15). This reaction is most commonly catalyzed by a metallophosphoesterase LpxH (Babinski et al., 2002), but the α-proteobacteria family has evolved a distinct enzyme, LpxI, which employs a different mechanism to achieve the same result (Metzger and Raetz, 2010). LpxH and LpxI share no homology and use different catalytic mechanisms, with LpxI attacking the β-phosphate of UDP-2,3 diacyl-N-glucosamine and LpxH attacking the α-phosphate (Whitfield and Trent, 2014). Both LpxH and LpxI are dependent on divalent cations for catalysis, but LpxH has been shown to bind two Mn$^{2+}$ ions (Young et al., 2013), while LpxI binds a single cation, preferably Mg$^{2+}$, Mn$^{2+}$, or Co$^{2+}$ in vitro (Metzger and Raetz, 2010; Metzger et al., 2012). Studies in E. coli have shown that LpxI is able to rescue deficiencies in LpxH, confirming that these two enzymes are substitutes for one another (Metzger and Raetz, 2010). While there is no structure available for LpxH, the LpxI dimer has been co-crystallized with both its substrate and its product (Metzger et al., 2012). The LpxI enzyme consists of two distinct domains, the catalytic domain and a lipid-binding domain. The protein has an open conformation when the product is present, but the two domains are folded onto each other in a closed conformation when an inactive mutant is crystallized with the substrate (Metzger et al., 2012). This suggests that the catalytic mechanism of LpxI involves dynamic conformational shifts within LpxI’s structure to produce the lipid X product.
Figure 1-15 The fourth step of lipid A synthesis is catalyzed by LpxH or LpxI depending on the species. Right: chemical reaction. Left: Crystal structure of LpxI co-crystallized with the reaction substrate and a magnesium ion (4GGM).

The first disaccharide product is produced in the fifth step by LpxB, which is an inverting glycosyltransferase which combines one molecule of UDP-diacyl-N-glucosamine with one molecule of lipid X (Figure 1-16). This reaction results in the characteristic β-1’6-glycosidic linkage between these two subunits and releases uridine diphosphate (Crowell et al., 1986). The catalytic rate of LpxB is dependent upon the bulk surface concentration of its substrates in mixed micelle experiments which does suggest that catalysis occurs at the plasma membrane (Metzger IV and Raetz, 2009). LpxB is a peripheral membrane protein that has yet to be characterized.
Figure 1-16 The first disaccharide is synthesized by LpxB in the fifth step of the synthesis pathway.

In the 6th step, the disaccharide intermediate synthesized by LpxB is phosphorylated by LpxK at the 4’ position of the second sugar moiety (Figure 1-17). LpxK is a poorly conserved member of the P-loop nucleotide triphosphate hydrolase family, and the only member of this family that phosphorylates a lipid substrate (Whitfield and Trent, 2014). Phosphorylation by LpxK \textit{in vitro} is Mg2+ dependent in the presence of detergent micelles suggesting that phosphorylation also occurs at the membrane interface (Emptage et al., 2013). Multiple structures of LpxK have been solved including an open apo-conformation and a closed nucleotide-bound state (Emptage et al., 2012, 2013). In \textit{E. coli}, LpxK is predicted to have a N-terminal helix responsible for membrane integration. The LpxK structures available to date have been determined from \textit{Aquifex aeolicus}, which has an extended hydrophobic N-terminal helix that associates with the membrane, but is not embedded (Whitfield and Trent, 2014). This completes the synthesis of the lipid A precursor, lipid IV\textsubscript{A}, named for the four acyl-chains present.
The sixth step is catalyzed by LpxK. Right: chemical reaction. Left: Crystal structure of LpxK crystalized with AMP-PCP in the active site (4ITL).

The seventh step is the addition of two Kdo residues by WaaA (Figure 1-18). These additional sugar moieties are technically part of the inner core oligosaccharide, but their addition is generally required before the last two acyl chains can be added (Whitfield and Trent, 2014). In some species, additional phosphorylation of the new Kdo sugars is also required before synthesis can continue (Hankins and Trent, 2009). Therefore, the addition of Kdo sugars by WaaA is now synonymous with lipid A biosynthesis. WaaA is a glycosyltransferase with a N-terminal transmembrane domain and two distinct cytoplasmic domains with a catalytic site at their interface (Whitfield and Trent, 2014). Again, the structure of WaaA that is available is of a variant from Aquifex aeolicus, which contains a large hydrophobic domain for association to the inner membrane, instead of a transmembrane domain (Schmidt et al., 2012).
Figure 1-18 WaaA adds the Kdo sugars required for final acylation in the seventh step of lipid A synthesis.

The final stage of lipid A synthesis is the addition of secondary acyl chains which is catalyzed by the membrane bound acyltransferases, LpxL and LpxM (Figure 1-19). These enzymes are very similar to each other, but share no homology with LpxA or LpxD (Whitfield and Trent, 2014). These enzymes still use acyl-ACPs as their preferred hydrocarbon donors and also likely have a mechanism like the hydrocarbon ruler given their strict acyl-chain length preferences. As previously mentioned, Kdo2-lipid IV\textsubscript{A}, the product of WaaA glycosylation, is strongly preferred over lipid IVA alone with a difference as high as 6000-fold (Six et al., 2008). In \textit{E. coli}, LpxL performs the first acylation, attaching a lauryl group to the acyl chain in the 2’ position, followed by LpxM, which transfers a myristoyl group to the acyl chain in the 3’ position (Six et al., 2008; Whitfield and Trent, 2014; Xu et al., 2013).
Figure 1-19 The last two steps in the synthesis of lipid A are the addition of two additional acyl chains by LpxL and LpxM.

1.6.3 Assembly and Export of Lipopolysaccharide

Synthesis of the core oligosaccharide often begins during lipid A synthesis with addition of two Kdo-sugars by WaaA as described above. Additional sugars required for both the inner and outer regions of the core oligosaccharide are synthesized in the cytoplasm and transferred to new lipid A molecules by glycosyltransferases that are associated with the inner membrane (Whitfield and Trent, 2014). The Kdo₂-lipid A serves as the acceptor on which the core oligosaccharides are assembled via sequential glycosyltransferase reactions using nucleotide sugar precursors as the substrate and energy source (Raetz and Whitfield, 2002). The genes required to perform these glycosylations are clustered in three operons in the waa region of the chromosome in *E. coli* (David et al., 1998; Regué et al., 2001). Assembly of the core oligosaccharide is completed on the cytoplasmic face of the inner membrane before the lipid A is transported to the periplasmic face of the same membrane.
Fully synthesized lipid A is flipped across the inner membrane by the MsbA transporter. MsbA is an ABC transporter, which contains two nucleotide-binding domains and two transmembrane domains (Whitfield and Trent, 2014). Each polypeptide of MsbA provides a single nucleotide binding domain and transmembrane domain, therefore a dimer of MsbA is required to assemble the transporter (Ward et al., 2007). The ATPase-powered transport of lipid A can be directly stimulated by the presence of the complete, hexa-acyl-lipid A, but is not stimulated by precursors such as lipid X or lipid IVₐ (Doerrler and Raetz, 2002). Therefore, MsbA is a point of quality control that ensures that only complete lipid A is transported across the inner membrane. Crystal structures of MsbA from various points in the catalytic cycle have been solved (Ward et al., 2007) and demonstrate the opening and closing of this transporter to the two different sides of the membrane (Figure 1-20). Additional modifications to the core oligosaccharide for stress adaptation can occur in the periplasm. The synthesis and regulation pathways responsible for these changes are highly variable and were reviewed in 2013 (Chen and Groisman, 2013; Needham and Trent, 2013).

Figure 1-20 Structures of MsbA in different conformations. (A) Open. (B) Closed. (C) Nucleotide-bound, closed.
The O-antigen polysaccharide is assembled and transported across the inner membrane by one of three pathways, each of which has been recently reviewed; Wzy-dependent (Whitfield, 2006), ABC transporter dependent (Greenfield and Whitfield, 2012), or synthase dependent (Whitney and Howell, 2013). The overall architecture of the inner membrane components of these “machines” is similar to those described previously for CPS and cellulose and are summarized in Figure 1-21. In all three mechanisms, polymerization of the O-antigen polysaccharide is initiated by the transfer of N-acetylglucosamine-1-phosphate to undecaprenol phosphate by the membrane protein WecA (Greenfield and Whitfield, 2012).

![Figure 1-21](image)

**Figure 1-21** The three mechanisms for assembling and transporting LPS across the inner membrane. From left to right: Wzy-dependent, ABC transporter dependent, Synthase-dependent. NBD = nucleotide binding domain. TMD = transmembrane domain.

In the case of ABC-transporter dependent transport mechanism, this initial sugar is then extended by various cytoplasmic glycosyltransferases dependent on the serotype of the bacteria. O-antigen assembly is terminated in the cytoplasm by phosphorylation and/or methylation of the terminal sugars and these modifications are essential for transport of the polymer across the inner membrane (Greenfield and Whitfield, 2012). The ABC transporter responsible for moving the sugar polymer across the inner membrane contains the typical two-nucleotide binding domains on the cytoplasmic face of the membrane.
and two transmembrane domains. Specificity for O-antigens is provided by the nucleotide binding domain that is extended to also contain a sugar-binding domain which preferentially binds properly terminated O-antigens (Cuthbertson et al., 2007).

In the synthase dependent pathway, the complete polymerization of the O-antigen also occurs in the cytoplasm, and the sugar polymer is transported across the inner membrane by the synthesis protein embedded in the inner membrane. Examples of synthase-dependent mechanisms were given in section 1.5 on biofilm polysaccharides. The third pathway, the Wzy dependent pathway differs from the other two in that subunits of the O-antigen are assembled in the cytoplasm and final polymerization occurs in the periplasm. The Wzy-dependent mechanism was discussed in the context of capsular polysaccharide assembly and transport earlier in this chapter and will not be repeated here.

Regardless of the pathway used to assemble and transport the O-antigens across the inner membrane, attachment of the O-antigen to lipid A is catalyzed by WaaL in the periplasmic space (Figure 1-22) (Abeyrathne and Lam, 2007; McGrath and Osborn, 1991). It has been suggested that WaaL is a metal-independent inverting glycosylase (Ruan et al., 2012), although the exact mechanism and structure has yet to be elucidated. Systematic in vitro analysis has also suggested that WaaL has a relaxed specificity for the lipid and glycan moieties that it links (Han et al., 2012), which could explain why a single enzyme is able to complete the assembly of the diverse array of possible LPS structures.
Completed LPS still needs to be exported from the outer leaflet of the inner membrane to the outer leaflet of the outer membrane. Excess accumulation of LPS and LPS precursors in the inner membrane can lead to deformations of the membrane. Regardless of the method of assembly and transport across the inner membrane, LPS is transported to the outer-membrane by the Lpt transport system. This group of proteins forms a bi-membrane apparatus for the ATP-dependent movement of fully synthesized lipopolysaccharide (Figure 1-22). Four proteins (LptF, LptG, and two LptB subunits) come together to form an ABC protein complex in the inner membrane (Sperandeo et al., 2008). There is a molecular structure available for LptB, but not for the membrane spanning components (Figure 1-23B) (Sherman et al., 2014). An additional subunit, LptC, binds to the ABC protein complex in a 1:1 ratio. LptC forms the base of a protein bridge that transports LPS between the two membranes. LPS passes through a potentially continuous groove that spirals around the bridge formed by LptC-LptA-LptD (Whitfield and
The periplasm is spanned by multiple LptA proteins (Merton et al., 2012) and the bridge is anchored to the outer membrane by the N-terminal domain of LptD. All three involved polypeptides (LptC, LptA, and the N-terminal domain of LptD) form a twisted boat, or β-jellyroll structure with two β-sheets in opposition (Figure 1-23AC) as determined by x-ray crystallography (Dong et al., 2014; Suits et al., 2008; Tran et al., 2010). The unidirectional movement of LpxC to LpxA has also been experimentally observed (Tran et al., 2010).

LptD forms an extremely large outer membrane β-barrel channel that is 1.8 nm wide and is responsible for inserting LPS into the outer membrane (Haarmann et al., 2010; Wu et al., 2006). LptD forms a stable complex with LptE, a small protein which “plugs” the large LptD channel. LptE also binds LPS specifically and may operate as the point of substrate recognition for opening the channel (Chng et al., 2010). A crystal structure for the complete LptD-LptE complex has been recently published (Figure 1-23D) which confirms the tight relationship between these proteins and the blockage of the channel by LptE (Dong et al., 2014).
1.6.4 Regulation of Lipopolysaccharide

Given that LpxC is the first committed step of lipid A biosynthesis, it represents a key point in the regulation of the lipid A biosynthesis pathway. The regulation of LPS requires the degradation of LpxC by the essential membrane-bound protease FtsH (Führer et al., 2006). FtsH is an ATP-dependent metalloprotease that belongs to the AAA+ family (ATPases associated with various cellular activities) of enzymes and contains both AAA domain, and the protease domain. (Snider et al., 2008). A crystal structure of the AAA and protease domains for FtsH have been determined (Baumgartner and Bieniossek,
It has been demonstrated that the C-terminus of LpxC is essential for the recognition of LpxC by FtsH for degradation (Führer et al., 2006). WaaA, from later in the lipid A synthesis pathway, is also a substrate for degradation by FtsH (Katz and Ron, 2008). Additional substrates for FtsH in *E. coli* have also been identified (Westphal et al., 2012).

It has been shown that the degradation of LpxC is somehow mediated by the heat shock protein LapB (Klein et al., 2014; Mahalakshmi et al., 2014). The proposed regulation mechanism for the protein level of LpxC is shown in Figure 1-24. Direct binding between LapB and LpxC has not been observed, but both LapB and LpxC can bind the chaperones DnaJ and DnaK (Führer et al., 2006; Klein et al., 2014). In contrast, single deletions of DnaJ or DnaK do not appear to prevent LpxC degradation (Führer et al., 2006), leaving the molecular mechanism behind LapB mediated degradation of LpxC unclear. The LapB protein has also been pulled down with WaaC (a glycosyltransferase involved in core oligosaccharide synthesis), LptD, and FtsH suggesting that LapB may act as a central scaffold that coordinates the actions of various proteins in the LPS synthesis and export pathways (Klein et al., 2014). The overexpression of LpxC and resulting overexpression of LPS appear to be caused by reduced synthesis of normal glycerophospholipids since they share a common precursor, hydroxymyristoyl-ACP. Overproduction of the FabZ enzyme, which shuttles hydroxymyristoyl-ACP into the glycerophospholipid pathway, is a method for rescuing this phenotype (Mahalakshmi et al., 2014; Ogura et al., 1999; Putker et al., 2015).
Figure 1-24 Regulation of LpxC. Right: Schematic of the proposed regulation of LpxC by LapB. LapB facilitates the degradation of LpxC by an unknown mechanism. Left: Structure of the FtsH AAA and protease domains (2CEA) bound to ADP, zinc (grey), and magnesium (purple).

LpxC has also been central to scientific efforts to block LPS production. There have been numerous compounds reported to block the activity of LpxC (Brown et al., 2012; Jackman et al., 2000; Liang et al., 2013; Szermerski et al., 2014; Tomaras et al., 2014). Some of these compounds have also been tested in vivo and are active against *E. coli, Enterobacter, and Klebsiella*, at concentrations comparable to ampicillin (Onishi et al., 1996). In general, the Lpx enzymes have no mammalian counterparts which makes them promising targets for the development of novel antimicrobials (Whitfield and Trent, 2014). Acyltransferases LpxA and LpxD both act early in the lipid A synthesis pathway and have similar substrates, which has allowed the development of dual binding LpxA/LpxD inhibitory peptides, such as RJPXD33, that block the substrate binding pocket (Jenkins et al., 2014). Also, sulfonyl piperazine was recently identified as an inhibitor of LpxH and is the first compound reported to block this enzyme (Nayar et al., 2015).

The LPS production pathway can also be regulated at the point of LPS transport. The gram-negative bacteria have their own mechanisms for ensuring that LPS is only moved from the inner membrane when...
the complete Lpt-protein system has been built. For example, LptD is only able to bind LptA after it has been correctly integrated into the outer membrane and folded (Freinkman et al., 2012). LptA is degraded when the complete Lpt machinery cannot be assembled (Sperandeo et al., 2011). Artificial inhibitors have also been designed to target this essential machinery. The power for the initiation of transport comes from ATP hydrolysis by LptB as part of the LptBFGC complex in the inner membrane. The enzymatic cytoplasmic domain can be blocked both in isolation and as part of the LptBFGC complex. Two compounds were originally identified as inhibiting LptB in isolation and one of these compounds has been optimized and shown to kill *E. coli* that have a permeable outer membrane (Gronenberg and Kahne, 2010; Sherman et al., 2013).

1.7 Summary

Gram-negative bacteria produce an array of exopolysaccharides that are critical for the survival of the bacterium, especially during the environmental stress of infection of a host organism. This has made exopolysaccharide production pathways of interest for identifying novel pharmaceutical targets. By comparing the export pathways of different polysaccharides, it can be seen that there are only three different types of transport mechanisms: the Wzy-dependent mechanism, the ABC-transporter dependent mechanism, and the synthase dependent mechanism. The three bi-membrane machines are all composed of multiple parts that have been characterized to varying degrees. There is currently an effort to determine the structure of all of the various components, which has resulted in the recent publication of a large number of novel structures, despite the fact that many of the involved proteins are membrane bound. However, there are still many gaps in the high-resolution data that must be filled to get a full picture of the molecular mechanism behind these essential pathways.
1.8 Hypothesis and Objectives

High-resolution molecular structures for the exopolysaccharide regulators Etk and LapB will provide insight into their regulator mechanisms. Membrane proteins are difficult targets to analyze and experimental work is facilitated by improvements in the underlying methods of crystallography and structure determination. In order to obtain and exploit these structures this thesis aims to:

- Identify Etk truncations and conditions that can be used for the production of stable protein samples for crystallization experiments (Chapter 2).
- Improve the available methods for detergent measurement as it applies to the crystallization of membrane proteins (Chapter 3)
- Apply the structural information obtained for LapB to improve our understanding of protein function (Chapter 4)
- Explore the limits of experimental phasing by Zn-SAD for structure determination by x-ray crystallography (Chapter 5)
Chapter 2

Purification, Characterization, and Crystallization of Membrane Bound

*Escherichia coli* Tyrosine Kinase

2.1 Foreword

Repeated efforts have been made in the Jia lab and elsewhere to purify and crystallize constructs of *Escherichia coli* tyrosine kinase (Etk) that contain the periplasmic segment. At first glance, the simplest approach appears to be the truncation of the protein to produce a soluble periplasmic domain, but although this construct can be expressed and purified readily, it has resisted crystallization. In this manuscript, purification and biophysical characterization experiments which involved membrane-bound constructs of Etk are discussed, and the discovery of an approach that leads to crystallization is highlighted.

2.2 Abstract

*Escherichia coli* tyrosine kinase (Etk) is a membrane bound kinase in gram-negative bacteria that regulates the export of capsular polysaccharides (CPS). The molecular mechanism behind CPS regulation remains unclear, despite access to a crystal structure of the cytoplasmic kinase domain of Etk. In this study, an efficient protocol to produce full length Etk solubilized in n-dodecyl-β-D-maltoside has been established with high yield. We have determined that detergent solubilized Etk retains kinase activity, but the protein is prone to aggregation, degradation, and has been unsuccessful in protein crystallization trials. In response, we designed and characterized truncations of Etk that do not aggregate and have led to successful crystallization experiments. In this article, we discuss our optimized expression and purification protocol for membrane proteins, the design of Etk protein truncations, and the behavior of Etk during purification in a range of stabilizing detergents. These efforts have successfully produced
protein suitable for crystallization. Our results will be a useful guide for future structural and functional studies of the bacterial tyrosine kinase family.

2.3 Introduction

Although approximately 50% of current drug targets are membrane proteins (Drews, 2000), this class of protein remains severely under-represented in structural databases due to the significant challenges associated with expression and purification. *Escherichia coli* tyrosine kinase (Etk) is a 726 amino acid (81 kDa) protein imbedded in the inner membrane of *E. coli*, which contains a large periplasmic domain flanked by two transmembrane helices and a C-terminal cytoplasmic kinase domain (Doublet et al., 2002). Etk is also a representative member of the bacterial tyrosine kinase family, which is found across gram-negative bacteria. This family of tyrosine kinases in prokaryotes is structurally unique compared to eukaryotic kinases (Lee et al., 2008). Therefore, this family is an interesting target for the development of novel antibiotics.

Etk is a member of a protein complex that spans both the inner and outer membranes in *E. coli*, and is expected to bind the YccZ channel protein in the outer membrane (analogous to the Wzc-Wza complex reported in (Collins et al., 2007)). This complex is involved in the Wzy-dependent mechanism used for group 4 capsular polysaccharide (CPS) export from the cell (Cuthbertson et al., 2009) and polysaccharides exported by this mechanism play an important role in virulence (Ilan et al., 1999). Much effort has been exerted to functionally characterize the Wzy-dependent export mechanism, however structural information about the Wzy complex and its regulation remains limited (Cuthbertson et al., 2009). Based on studies of the homologous protein Wzc, Etk is expected to have a regulatory role in this complex which involves auto-phosphorylation of a C-terminal tyrosine cluster (Collins et al., 2007; Lee et al., 2008; Roberts et al., 1999). Two crystal structures of the kinase domain (Bechet et al., 2010; Lee et al., 2008) and a structure for the outer membrane channel protein are available (Dong et al., 2006), but the
transmission of regulatory signals from the kinase domain to the channel is unclear. Signal transduction likely involves conformational changes in the kinase domain (Lee et al., 2008; Olivares-Illana et al., 2008) that are transmitted across the inner membrane to Etk’s periplasmic domain. Etk’s periplasmic domain likely interacts directly with the outer membrane channel protein and could be responsible for the opening and closing of this structure (Collins et al., 2007; Dong et al., 2006). The study of signal transduction is hindered by a lack of high-resolution data for Etk’s periplasmic domain. Therefore, it is desirable to study the full-length Etk protein or truncations that contain the N-terminal domain. In order to carry out these experiments, the ability to purify membrane proteins reproducibly in large quantity is critical.

Herein we describe two constructs of Etk, the full length protein, and a truncation that contains the periplasmic domain flanked by the two transmembrane helices. We have developed a protocol for the production of large amounts of detergent solubilized membrane bound Etk by exploiting different expression temperatures, detergent options, and buffer additives for their effects on yield, purity, and stability. After biophysical characterization it was determined that the truncated version of Etk full length possessed superior stability that made it a better candidate for crystallization trials. In addition, the type of detergent used was a critical factor in crystallization success.

2.4 Materials and Methods

2.4.1 Plasmid Construction and Protein Expression

The gene for full length Etk (Etk FL) cloned into the vector pCA24N was kindly provided by the Genome Analysis Project in Japan (http://ecoli.aist-nara.ac.jp). The Y574N mutation of Etk FL was previously generated (Lee et al., 2008). The full-length plasmid was used as a PCR template for Etk NM truncations, which were cloned into pET21b using the NdeI and XhoI cut-sites to produce protein with a C-terminal His-tag. Protein was expressed using E. coli BL21 cells in either Terrific broth media and 0.5
mM IPTG (added at an OD of 0.6-0.8) or autoinduction media (Studier, 2005). Induced protein expression was carried out at 37 °C for 4 hours or 20 °C overnight. Cells were harvested by centrifugation at 3300 × g for 20 min.

**2.4.2 Protein Purification**

Cells were re-suspended in 20mM sodium phosphate (pH 7.0) and lysed using a combination of lysozyme, freeze/thaw in liquid nitrogen and sonication, or only sonication. Cell lysates were centrifuged at 21000 × g for 30 min to remove soluble material and the pellet was solubilized by incubation in 50 mM sodium phosphate (pH 7.0), 250 mM NaCl, and 0.2-3.0 % n-dodecyl-β-D-maltoside (DDM, solvation grade, Anatrace) overnight at 4°C. The amount of DDM detergent used in the solubilization buffer was optimized by purifying 1L of cell culture in 6 equal fractions with increasing amounts of DDM. The mass of each insoluble pellet was measured prior to solubilization. Remaining insoluble material was removed by a second 30 min centrifugation at 21000 × g. Purification was continued using immobilized metal affinity chromatography (IMAC) on Ni²⁺ resin (Qiagen) run with wash buffer (50 mM sodium phosphate pH 8.0 and 250 mM NaCl) that contained 0.01% DDM and increasing amounts of imidazole (25 mM wash, 250 mM elution). Protein yield was measured using OD (280 nm) and an extinction coefficient of 57300 for the full length protein or 29910 for Etk NM truncations. Standard purification of Etk NM truncations used the same protocol with buffers adjusted to pH 8.0 throughout.

During detergent screening, the extraction and purification protocol were modified as follows. The detergent concentration was kept at 2% during solubilization and lowered during purification to: 0.01% n-dodecyl-β-D-maloside (DDM), 0.1% n-decyl-β-D-maltoside (DM), 0.02% lauryldimethylamine-N-oxide (LDAO), 0.5% n-octyl-β-d-glucoside (OG), 0.0015% lauryl maltose neopentyl glycol (LMNG), 0.004% decyl maltose neopentyl glycol (DMNG), or 0.05% n-dodecylphosphocholine (FC-12). All detergents were purchased from Anatrace.
2.4.3 Size Exclusion Chromatography (SEC)

Size-exclusion chromatography was performed using a Sephacryl 300 column (GE Healthcare) and an AKTA purifier system (GE Healthcare). Running buffer for SEC was 50 mM Tris pH 8.0 and 250 mM NaCl supplemented with detergent at the concentrations listed under protein purification. The pH of the running buffer was altered in some experiments. For the additive screen presented in section 2.5.4, samples were dialyzed into buffer (50 mM Tris pH 8.0, 250 mM NaCl, and 0.01 % DDM) supplemented with combinations of arginine, glutamate, and/or imidazole prior to concentration in a centrifugal concentrator (30 or 50 kDa cutoff, Millipore) to 2 mL for sample loading. The void volume of the column was determined to be 34 mL using a blue dextran standard.

2.4.4 Analytical Ultracentrifugation

Analytical ultracentrifugation (AUC) experiments were performed using a Beckman Optima XL-I analytical ultracentrifuge equipped with an AN60Ti rotor (Beckman Coulter, Brea, CA). Sedimentation velocity runs employed absorbance optics, a rotor speed of 25000 rpm, and samples contained in a two-sector cell with a 12 mm optical path. Protein concentration was 1 mg/mL. Buffer density, buffer viscosity, and protein partial specific volume were estimated using SEDNTERP (http://www.rasmb.bbri.org/) and are listed in Table 1-1. The detergent component of the buffer was not included in these calculations and was assumed to have a minor effect on buffer density and viscosity.

Table 2-1 Statistics for fitting of AUC sedimentation curves.

<table>
<thead>
<tr>
<th></th>
<th>Etk Full Length</th>
<th>Etk NM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer Density</td>
<td>1.0113 g/cm³</td>
<td>1.03442 g/mL</td>
</tr>
<tr>
<td>Buffer Viscosity</td>
<td>0.01036 poise</td>
<td>0.012277 poise</td>
</tr>
<tr>
<td>Protein Partial Specific Volume</td>
<td>0.74040 cm³/g</td>
<td>0.7449 cm³/mL</td>
</tr>
</tbody>
</table>

During initial trials, scans were sequentially removed from the start of the sedimentation run during analysis in SEDFIT (Schuck, 2000) to determine the time and force required to sediment protein.
aggregates without eroding the oligomeric component of the sample. After aggregate removal by sedimentation, oligomeric samples were analyzed by fitting 200 sedimentation boundary scans with Lamm equation modeling performed by SEDFIT to calculate distributions by sedimentation coefficient. Assignment of the oligomeric states for peaks in the c(s) distribution was based on the molecular weight estimates determined by calculating weight-average frictional coefficient for the distribution. The percentage of individual oligomers within the overall c(s) distribution was determined using the c(s) integration function.

2.4.5 Sedimentation Assay

Protein from the non-aggregation peak in size exclusion was collected and centrifuged using a Type 100 Ti rotor and Optima L-100 XP Ultracentrifuge for 3 hours at 46000 × g and 4°C. Size exclusion chromatography was used to confirm the removal of aggregated protein and OD (280nm) readings taken before and after sedimentation were used to determine the amount of protein aggregated.

2.4.6 Kinase Activity Assay

A kinase domain truncation of Etk (Etk C) was prepared as previously reported (Lee et al., 2008). Kinase reactions were performed using a modified version of the protocol published by Crawley et al. (Crawley and Côté, 2009). Samples were prepared by removing non-specific aggregation using sedimentation followed by dilution to 0.5 mg/mL in kinase reaction buffer. This buffer contained 2 mM MgCl₂, 0.25 mM ATP, 20 mM Tris (pH 7.0), 300 mM NaCl, 0.01% DDM, 50 mM L-arginine, 50 mM L-glutamate for Etk FL and Etk FL Y574N and 2 mM MgCl₂, 0.25 mM ATP, 50 mM Tris (pH 9.5), 300 mM NaCl for Etk C. The autokinase reaction was performed using [γ-³²P] ATP (Perkin Elmer) at a specific activity of 400-500 cpm/pmol and a temperature of 25°C. Aliquots were removed from the reaction mixed with 2 × SDS sample buffer followed by boiling to stop the reaction. Following SDS-PAGE, Coomassie stained bands were excised and counted using a Beckman LS 9000 scintillation
counter and ScinitiVerse Universal LS cocktail (Fisher Scientific). Photodensitometry in the ImageJ software package was used to calculate the amount of the desired protein present in the sample. Data sets were calculated as the average of three replicates for each reaction.

2.4.7 Mass Spectrometry

MALDI mass spectrometry was performed on tryptic peptides of Etk by the Protein Function Discovery Facility at Queen’s University. Tryptic fragments recovered were compared to the sequence of Etk using the server at http://www.matrixscience.com.

2.4.8 Crystallization Screening and Optimization

High-throughput crystallization screens were set using a Phoenix liquid handling robot (Art Robbins) and 96-well low profile plates for sitting vapor diffusion. Drops were set with 400 nL of protein solution and 400 nL of well solution. A second drop with 400 nL protein, 400 nL well solution and 100 nL 0.1 M MgCl₂ was also set. Well solutions contained 100 µL of commercial screening kits including Classics I and II, JCSG I-IV, PACT (Qiagen) and MemGold I and II (Molecular Dimensions). Plates were monitored for crystal growth for ~4 months.

Successful lead conditions were expanded using hanging-drop vapor diffusion in 24-well plates (Hampton). Drop volumes were a minimum of 2 µL protein and 2 µL well solution and went up to 5 µL protein and 5 µL well solution. Successful crystallization of Etk NM2 occurred in 0.1 M MES pH 6.0, 5% PEG 300 and 25% PEG 200 and the condition was optimized to 0.1 M MES pH 6.0, 4% PEG 3000 and 20% PEG 200. Successful crystallization of Etk NM2 occurred with protein at concentrations above 6.9 mg/mL and DDM concentrations of 0.6-1.1 mg DDM/mg protein. Detergent concentrations were measured using a colorimetric chemical reaction previously reported (Prince and Jia, 2015a). Crystal leads were optimized using additive screening kits from Hampton Research. Additives were screened at 20% of the total drop volume (i.e. 2 µL protein, 2 µL precipitant, 1 µL additive).
2.5 Results

2.5.1 Protein Expression

Expression of Etk FL was performed using either 37 °C for 4 hours or 20 °C overnight. The difference between the two methods of expression was the final yield of protein. For expression at 20°C overnight, the total protein eluted from the Ni\(^{2+}\) column was 29 mg/L of culture compared to only 3.5 mg/L after protein expression at 37°C for 4 hours. In both cases the approximate amount of aggregated vs. non-aggregated protein was the same based on size exclusion chromatography. In these experiments, 2% DDM was used to extract the protein from the membrane. Expression at 20°C was used in future trials.

2.5.2 Protein Extraction and Purification

The amount of DDM detergent used in the overnight solubilization buffer was optimized in order to obtain maximal extraction of Etk FL from the membrane with minimal expense. A 1 L culture of cells overexpressing recombinant Etk FL was purified in 6 fractions with different amounts of DDM. After nickel-IMAC purification, SDS-PAGE and optical density at 280nm were evaluated and revealed that increasing the amount of detergent did increase the amount of protein solubilized until a plateau was reached (Figure 2-1 A&B). The first point on the plateau, i.e. 2% detergent, was chosen for future purification rounds. This corresponded to a ratio of 0.2g DDM/g wet cell pellet.
2.5.3 Kinase Activity Assay

Etk FL was tested along with a mutant protein, Etk FL Y574N for autophosphorylation activity. The isolated C-terminal kinase domain (Etk C), which had already been demonstrated to catalyze autophosphorylation, was used as a positive control (Lee et al., 2008). All three proteins were determined to be active to varying degrees. The Etk FL Y574N mutant had very low activity, which was consistent with previous experiments on Etk C (Lee et al., 2008). Etk C was twice as active as Etk FL solubilized in...
DDM (Figure 2-2). The lower kinase activity measured for Etk FL is likely due to the continuous aggregation of Etk FL under these conditions. In addition, the phosphorylation state (i.e. the number of phosphate groups in the C-terminal tyrosine cluster) of the purified protein in vitro may be different between Etk C and Etk FL. The phosphorylation state has been previously shown to be a regulator of Etk kinase activity (Paiment et al., 2002). Given that wild-type Etk FL showed significantly more activity than a known low-activity mutant version, we concluded that at least a portion of the Etk FL we are purifying is properly folded.

![Figure 2-2](image)

**Figure 2-2** Radioactive phosphate incorporation into Etk FL, Etk FL Y574N and Etk C due to autophosphorylation. Each point represents the average of 3 replicates and error bars represent the standard deviation of these measurements.

### 2.5.4 Protein Oligomerization and Aggregation

Protein aggregation was originally detected by size-exclusion analysis of the protein, where Etk FL elutes as two distinct peaks (Figure 2-3). The first peak is at the void volume of the column and is associated with protein aggregation. This peak is followed by a second broad peak containing Etk oligomers. The detergent used for membrane protein stabilization can affect protein integrity and
aggregation in vitro (Gan et al., 2011). Therefore, the dispersion of Etk FL in solution when solubilized by DDM, DM and FC-12 detergents was compared. None of these options were able to provide adequate stabilization, with FC-12 being the worst, resulting in less overall protein, and less non-aggregated protein (Figure 2-3).

Since the detergents tested did not prevent aggregation of Etk FL, the addition of small molecule additives to the basic buffer of 50 mM Tris pH 8.0, 250 mM NaCl and 0.01% DDM was also screened. The additives tested were 50 mM imidazole, a mixture of 50 mM L-arginine and 50 mM L-glutamate and a mixture of 50 mM imidazole, 50 mM L-arginine and 50 mM L-glutamate. All three combinations of small molecules reduce the amount of high molecular weight aggregates in the sample compared to Tris buffer containing sodium chloride and DDM alone. These results were observed during SEC (Figure 2-4A) and using an ultracentrifugation sedimentation assay (Table 2-2). In the sedimentation assay, Etk FL aggregates are removed from the sample by ultracentrifugation at 46000 × g for 3 hours, allowing the proportion of oligomeric protein to be measured. The removal of aggregates by sedimentation was
confirmed by immediately re-running the sample on SEC (Figure 2-4B). A sample stored for 48 hours was also re-tested using the sedimentation assay and showed that Etk FL continues to aggregate. Finally, it was determined that aggregation increased with concentration (Table 2-2). Protein used in the concentration-dependent aggregation experiment had not been previously run on a size-exclusion column, while the additive-dependent aggregation experiment used protein from the non-aggregate peaks isolated by SEC the previous day (Figure 2-4A). Therefore, it is not surprising that the overall percentage of aggregation observed was higher in the concentration-dependent aggregation experiment.

Table 2-2 Aggregation state of Etk FL samples. Etk FL was ultracentrifuged in 20mM Tris (pH 7.0) and 300mM NaCl, in the presence of additives and at different protein concentrations.

<table>
<thead>
<tr>
<th>Buffer Additive</th>
<th>OD (280) Before</th>
<th>OD (280) After</th>
<th>% Aggregated</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.765</td>
<td>0.436</td>
<td>43</td>
</tr>
<tr>
<td>50mM L-Arg, 50mM L-Glu 50mM Imidazole</td>
<td>0.875</td>
<td>0.628</td>
<td>28</td>
</tr>
<tr>
<td>50mM Imidazole</td>
<td>0.864</td>
<td>0.646</td>
<td>25</td>
</tr>
<tr>
<td>50mM L-Arg, 50mM L-Glu</td>
<td>0.807</td>
<td>0.652</td>
<td>19</td>
</tr>
<tr>
<td>*50mM L-Arg, 50mM L-Glu</td>
<td>0.8801</td>
<td>0.5629</td>
<td>36</td>
</tr>
<tr>
<td>**50mM L-Arg , 50mM L-Glu</td>
<td>0.5629</td>
<td>0.428</td>
<td>24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein Concentration (mg/mL)</th>
<th>OD (280) Before</th>
<th>OD (280) After</th>
<th>% Aggregated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.09</td>
<td>0.7688</td>
<td>0.2757</td>
<td>64</td>
</tr>
<tr>
<td>1.44</td>
<td>1.0188</td>
<td>0.3214</td>
<td>69</td>
</tr>
<tr>
<td>2.64</td>
<td>0.9312 (1/3 dilution)</td>
<td>0.646</td>
<td>75</td>
</tr>
</tbody>
</table>

* Without prior size exclusion chromatography  ** Second spin after 48 hours storage
Figure 2-4 Reducing aggregation of Etk FL. (A) Effect of additives on Etk FL. Size exclusion (S300) of Etk FL after dialysis into standard buffer plus one of three combinations of additives intended to increase the stability of the protein. A = 50 mM L-arginine, G = 50 mM L-glutamate, I = 50 mM imidazole. Control contains standard buffer of 50 mM Tris pH 8.0, 250 mM NaCl, 0.01 % DDM. (B) Effectiveness of sedimentation assay. Size Exclusion (S300) of Etk FL after sedimentation for 3 hours at 46000 g showing that the first peak has been almost eliminated.

Analytical ultracentrifugation of non-aggregated Etk FL revealed that it primarily forms monomers, dimers and tetramers when solubilized in 0.01 % DDM. When analyzed at a 95 % confidence level the peaks were very broad resulting in the monomer peak appearing as a shoulder to the dimer peak (Figure 2-5). The data was further analyzed at a lower confidence level (65 %), which allowed the peaks to be resolved for quantitation, yet still retained the same peak positions in the sedimentation coefficient distribution, c(s) (Figure 2-5). The relative proportion of these peaks was 19.6% monomer, 31.5 % dimer, and 31.1 % tetramer. Higher molecular weight peaks were also visible in this sample and included 7.4 % octamer, and three additional higher order species with peaks approximately 300 kDa apart.
Figure 2-5 Size distribution of Etk FL determined from the sedimentation velocity experiment. Fit of 200 boundary layer scans using the Lamm equation at the 65% confidence level is shown superimposed on the same fit determined at the 95% confidence level.

2.5.5 Protein Degradation

Throughout purification and characterization of Etk FL there was always a secondary band present around 50 kDa, which was presumed to be a His-tagged degradation product. To ensure that the cell lysis procedure and other purification choices were not aggravating degradation, impure protein after each initial purification step was analyzed by SDS-PAGE (Figure 2-6). The degradation band only becomes visible after removal of insoluble materials, and therefore, could be present and obscured by protein from the full cell during lysis. The amount of visible degradation in SDS-PAGE remains constant and there is no indication that a specific purification step is responsible. Furthermore, we found that cell lysis with sonication alone was as effective as the three stage lysozyme/freeze-thaw/sonication procedure. Given prior results, which indicate that the Etk kinase domain was most stable at pH 9.5 and above (Lee et al., 2008), we also tried varying pH as a means of controlling aggregation. Unfortunately this strategy was not successful and had an additional unintended side effect. Protein degradation in vitro noticeably
increased as the buffer became more basic (Figure 2-6JKL). Therefore, a buffer pH of 7.0 resulted in optimal protein purity. Overall, storage in a basic solution was the most significant cause of in vitro protein degradation. The identity of the degradation band was subsequently confirmed by mass spectroscopy to be approximately residues 1-527 of Etk. This protein fragment contains the N-terminal periplasmic domain and the two transmembrane helices (Figure 2-7).

Figure 2-6 SDS-PAGE analysis of individual purification steps in the production of Etk FL. (A) Initial suspension of cells (B) After lysozyme treatment (C) After freeze-thaw cycles (D) After sonication (E) Soluble protein removed by first centrifugation (F) Re-suspended pellet of insoluble materials (G) Immediately after addition of detergent (H) After 20 hour solubilization (I) Soluble protein after centrifugation (J) Eluted protein after Ni\textsuperscript{2+} column at pH 7.0. (K) Eluted protein after Ni\textsuperscript{2+} column at pH 8.0. (L) Eluted protein after Ni\textsuperscript{2+} column at pH 9.0.

2.5.6 Design and Implementation of Etk Truncations

In this study, the periplasmic domain of Etk is a major target for crystallization since its structure would allow us to connect the kinase domain and the outer membrane channel in a model of the Etk-channel complex. Therefore, it was decided to produce a truncation of this protein that contains the N-terminal domain flanked by the two transmembrane helices. A few different variations were designed with start points at amino acids 1 or 20 before the first transmembrane helix or ending at residues 450 or 527 after the second transmembrane helix and all of the truncations tested could be solubilized in
detergent. Ultimately it was Etk NM2, residues 20-450, that was chosen for in-depth study due to high expression yields, favorable *in vitro* behavior, and it was the most compact for crystallization (Figure 2-8).

```
1  MTTKMNTPP GSTQENEIDL LRLVGELWDL RKFIISVTAL FTLLIAVAYSL
51  LSTPIYQADT LVQVEQKQGN AILSGLSDMI PNSSPESAPE IQLLQSRMIL
101  GTKIAELNLRI DIVEQKYFPI VGRGARLTK EKPGELAISW MHIPQLNGQD
151  QQLTLTVGEN GHYTLEGEEF TVNGMVQQR KRDGVALTIA DIAKKGHTQF
201  VLSQRTELEA INALQETFTV SERSKESGML ELTMGDDPQ LITRILNSIA
251  NNYLQQNIAR QAAQDSQSL FLQRQLPEVR SELDQAEKL NVYRQQRDSV
301  DILNLEAKAVL EQIVNVDNQL NELTFREAAEI SGLYKKDDHT YRALLEKRT
351  LEQERKRLNK RVSAEMPSTQQ EVLR8LSRDVE AGRAVYQLLL NRQQELSISK
401  SSAIGNVRR DPATQPQPQV KPKKALNVVL GFILGFISHV GAVLARAMLR
451  RGVEAPEQLE EHGISVYAYTI PMSEWLEDKRT RLRKKNLFSN QQRHRTKNIP
501  FLAVDNPADS AVEAVRALRT SLHFMMETE NNINMTGAT PDSGKTFVSS
551  TLAAVIAQSD QKVLFIADDL RRYSHNLFIT VSNENHGLSEY LAGKDELNKV
601  IQHFGKGGFD VTRGQVPPN PSELMRDRRM RQLLEWANDH YDLVIVDTFP
651  MLAVSDDAVV GRSGVTSLLV ARFGILNATKE VSLSMQRLEQ AGVNIKGAIL
701  NGVIRKASTA YSYGYNNYYG SYSEKE
```

Figure 2-7 Mass spectroscopy of Etk FL degradation. Fragments identified for the degradation product are identified in red on the full protein sequence.

![Figure 2-7](image)

Figure 2-8 Etk schematics (A) Etk full length, (B) Etk degradation product identified by mass spectroscopy, and (C) Etk NM2.
Etk NM2 was primarily expressed using autoinduction media with a typical yield of 20-30 mg protein per liter of *E. coli* cell culture. It was found that autoinduction media provided higher and more reliable expression of Etk NM2 than Terrific broth, which was found to have typical yields of 10-15 mg protein/L of cell culture. The amount of DDM detergent used during extraction of Etk NM2 was optimized in the same manner as the full length protein and was also found to be 2% DDM. Etk NM2 was purified using the same protocol as the full length protein with all buffers at a pH of 8.0 (Figure 2-9). Although the tendency of membrane-bound Etk to degrade was not completely alleviated, the degradation product remains very close in size to the expressed construct and these two bands produce similar results in MALDI-mass spectroscopy of tryptic fragments (Figure 2-10). Etk NM2 was further characterized for its behavior in multiple detergents, oligomerization state, and crystallizability.

![Figure 2-9 SDS-PAGE analysis of the typical elution fractions obtained after nickel affinity purification of Etk NM2.](image)
Figure 2-10 Tryptic fragments from the Etk NM2 construct and degradation product. Fragments identified in mass spectroscopy are highlighted in red on the protein sequence. The top band, presumably residues 20-450 is numbered in blue and the slightly smaller band, which may have some further degradation is numbered in green. Both bands in SDS-PAGE have similar identified tryptic fragments, but the first fragment in the N-terminus is missing in the smaller band and may suggest further degradation from this end of the protein.

2.5.7 Detergent Screen for Etk NM2

The first analysis performed on the Etk NM2 truncation was to determine the behavior of this protein on size-exclusion chromatography as a crude estimation of the oligomerization state. Initial purification trials with DDM as the stabilizing detergent demonstrated that Etk NM2 was not aggregated. The behavior of Etk NM2 in 6 additional detergents was also evaluated and it was determined that four of these options (DM, FC-12, LMNG, and DMNG) were also able to solubilize the protein without aggregation in vitro (Figure 2-11). Both LDAO and OG detergents, which have the shortest hydrocarbon chains of the detergents tested, contained predominantly aggregated species that were not studied further.
Figure 2-11 Size exclusion of Etk NM2 solubilized in various common detergents used in membrane protein crystallography.

The behavior of Etk NM2 as an oligomer in solution can also be analyzed by analytical ultracentrifugation. In the case of Etk NM2, we tested the oligomerization behavior of this protein in both DDM and DM (Table 2-3). This truncation of Etk produces predominately dimers and monomers, with DM favoring monomers. The different proportions of monomers and dimers between DDM and DM suggest that the detergent is having a significant impact on the oligomerization state of the protein. Elution volume variation in the size-exclusion profile of the different Etk NM2-detergent complexes is probably a function of both changes in the size of the detergent micelle, and changes in the oligomerization state of the protein.
### Table 2-3 Analytical ultracentrifugation of Etk NM2 stabilized in two maltoside-based detergents.

<table>
<thead>
<tr>
<th>Sample</th>
<th>DDM</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monomer</td>
<td>24.6%</td>
<td>86.3%</td>
</tr>
<tr>
<td>Dimer</td>
<td>55.7%</td>
<td>5.9%</td>
</tr>
<tr>
<td>Bigger (indistinct peaks)</td>
<td>16.9%</td>
<td>2.7%</td>
</tr>
</tbody>
</table>

### 2.5.8 Crystallization of Etk NM2

In order to identify an appropriate protein detergent complex for the crystallization of Etk NM2, each non-aggregated protein-detergent complex was screened against several commercial 96-well screens. Only two leads were found (Figure 2-13AB) which crystallized Etk solubilized in DDM or DM detergents. Only the first condition, which crystallized the Etk NM2-DDM protein-detergent complex could be reproduced when attempting to grow larger crystals. The reproduction of these crystals required strict control of both the protein and the detergent concentration (Prince and Jia, 2012). Ultimately, the optimized crystals only diffracted to a resolution of 6.5 Å on the 23-IDB beamline at the Advanced Photon Source (Figure 2-12). Crystals consistently produced equivalent diffraction no matter the size tested (50-300 µM).

![Figure 2-12 Representative x-ray diffraction of Etk NM2-DDM crystals to 6.5 Å resolution. Data can be indexed to space group P3 with a unit cell of 113.9 Å, 113.9 Å, 583.0 Å, 90°, 90°, 120°.](image)
Figure 2-13 Crystals of Etk NM2 (A) Optimized crystals growth from Etk NM2-DDM protein detergent complexes. (B) Initial lead condition for Etk NM2-DM under UV light. (C) Additive screen hit with 0.1 M MgCl₂. (D) Additive screen hit for 0.1 M MnCl₂. (E) Crystallization of Etk NM2 without MgCl₂ for comparison. (F) Crystallization with MgCl₂ for comparison. Drops E and F were grown over the same reservoir solution and pictures were taken under polarized light. Drops G-J were hits from the Hampton detergent screen. Added detergent concentration is 2 times the critical micelle concentration for the specified detergent. (G) NDSB-201 (H) Triton X-100 (I) LDAO (J) Thesit. Scale bar applies to all images.
2.5.9 Crystallization of Etk NM2 using Additive Screening

Efforts were made to optimize the diffraction of the Etk NM2 crystals using small molecule additives in the Additive Screen and Silver Bullets kits (Hampton Research). The only conditions that had an effect on crystallization were those that introduced a divalent cation (Figure 2-13CD). In this case, the crystals appeared noticeably larger. The effect was most pronounced for the addition of MgCl$_2$ and we specifically confirmed this observation by comparing two drops staring a common reservoir solution in a vapor diffusion plate (Figure 2-13EF). The second round of screening was done with the high-throughput detergent screen (Hampton) and identified multiple conditions in which the morphology of the Etk NM2-DDM crystals changed when a second detergent was added to create a mixed micelle. Addition of LDAO, Thesit, NDSB-201 and Triton X-100 all resulted in new crystal morphologies, but unfortunately, we were not able to obtain diffraction from these crystals (Figure 2-13GHIJ).

2.6 Discussion

There are two contrasting approaches to the early stages of membrane protein purification. The first involves isolation of the membrane fraction by ultracentrifugation or density gradient centrifugation followed by solubilization (Collins et al., 2006). In this scenario, the isolation of the membrane represents a significant purification step. The second approach involves the direct solubilization of membrane proteins from a total cell homogenate including both soluble and membrane bound proteins (Prive, 2007). The faster speed of the second approach can be essential to the study of some proteins whose stability is lost over time. For the purification of Etk FL, an intermediary approach was applied in which the insoluble pellet containing the cell membrane was isolated and then solubilized. This was a response to the aggregation of Etk FL, which was shown to worsen over time. The same approach proved to be very effective in the isolation of truncated Etk NM2. Therefore, this compromised approach, which combines
the purification advantage of removing most soluble proteins, while remaining time efficient, may be applicable in other membrane protein studies.

Typical concentrations of DDM used for the solubilization of protein from lipid membranes range from 1-4% DDM (Prive, 2007). By optimizing of the amount of detergent used during the solubilization stage of purification, the protein yield can be maximized while avoiding unnecessary waste of an expensive reagent. With both the full length and truncated Etk, we saw a marked increase in the amount of protein solubilized when up to 2% detergent is used, and then a leveling off, indicating that additional DDM would not provide additional benefit. Another important variable to consider during solubilization is the ratio of detergent to lipid, which can assist batch to batch reproducibility (Prive, 2007). In this case, we used the mass of the cell pellet for standardization due to its ease of measurement.

Aggregation of membrane proteins over time is a common issue encountered by many researchers (Prive, 2007). During the optimization of Etk FL, the addition of additives to the final buffer was essential to reducing aggregation. A common starting point when screening stabilizing additives is the addition of a reducing agent such as DDT or beta-mercaptoethanol (Beresten et al., 1999; Yoshida et al., 2005), but this would not be expected to reduce Etk aggregation, since Etk contains no cysteine residues. The effectiveness of arginine and glutamate (both at 50 mM) for avoiding the aggregation of soluble proteins has been previously reported (Golovanov et al., 2004). Here, we demonstrate that arginine and glutamate are also beneficial during membrane protein purification. Additionally, the effect of imidazole on the long-term stability of Etk FL was investigated, and we saw a similar improvement in the aggregation state. Imidazole was evaluated because it is an amino acid mimic that Etk is exposed to during nickel purification. Although a combination of L-arginine, L-glutamate, and imidazole was not able to entirely eliminate the formation of aggregates, it represented a significant improvement over the original buffer of Tris, sodium chloride and DDM detergent alone. Complete removal of protein aggregates required
ultracentrifugation. This solution proved to be temporary and requires protein samples to be immediately used in further experiments.

The oligomerization state of Etk FL in our preparations was determined to consist primarily of a mixture of monomers, dimers and tetramers as shown by analytical ultracentrifugation. The regular pattern of approximately 300kDa intervals between the high molecular weight peaks (above an s-value of 10) suggests that these species may be caused by tetramers stacking together. The presence of residual large aggregation (s-value >29) in the sample (approx. 3%) after the initial sedimentation spin is probably a result of equilibration during the experimental setup and 11-hour data collection period. Our AUC results are consistent with published electron microscopy and native gel experiments for the Wzc homologue of Etk FL (Collins et al., 2006).

In contrast, AUC of the Etk NM2 truncation reveals primarily dimers and monomers in the same solubilizing detergent. This does suggest that the removed kinase domain could play a role in the formation of higher-order oligomers in Etk FL. Oligomerization mediated by the kinase domain was observed in the octameric ring structure of the Wzc kinase domain (homologous to Etk) (Bechet et al., 2010). We also observed that detergent is likely breaking up natural oligomers during solubilization by comparing the AUC results for Etk NM2 solubilized in DM or DDM. The Etk NM2-DM protein complex is primarily monomers instead of the mixture of dimers and monomers found in the Etk NM2-DDM protein. This supports the idea that Etk oligomers are being dissociated.

Previous work has suggested that Wzc (homologous to Etk) may be a tetramer (Collins et al., 2006) or an octamer natively (Bechet et al., 2010), and the channel protein Etk is expected to interact with also forms an octamer (Dong et al., 2006). From the perspective of crystallography, the smaller oligomers present in the detergent solubilized sample could be beneficial for crystallization. It has been shown that smaller, more compact proteins are far more likely to be crystallized, during large scale genomics studies of soluble proteins (Chandonia et al., 2006; Smialowski et al., 2006). In the case of Etk NM2, we were
able to obtain crystals in both DDM and DM detergents, but the crystals in DDM were far larger and grew more reliably. This is in contrast to the argument that smaller detergents, with smaller micelles are better for crystallization (Garavito et al., 1996; Kunji et al., 2008; Marone et al., 1999), since DDM detergent forms a larger micelle than DM. In addition, it could be argued that the DM protein is more monodisperse, which is another property of protein samples that can aid in crystallization (Geerlof et al., 2006). Although neither sample is perfect, the DM sample is 95% monomeric vs. a near 50/50 monomer/dimer split in the case of DDM. Although many “rules of thumb” that exist from soluble protein crystallization may be transferable to protein-detergent complexes, there is still not enough evidence to confidently predict the best detergent-protein combination based on these properties alone.

While studying full-length proteins is preferable, experimental considerations sometimes force truncation of the native sequence. For Etk, the most significant advancement made was the application of the Etk NM2 truncation and its improved in vitro properties. Recent work in improving the crystallizability of membrane proteins has focused on protein fusions or complexes with readily crystallizable soluble proteins such as lysozyme (Cherezov et al., 2007; Rasmussen et al., 2011; Rosenbaum et al., 2007), cytochrome b562RIL [32] and antibody fragments (Johnson et al., 2015). In these cases, it has been shown that these crystallization chaperons increase the soluble surface area of the protein available to form crystal contacts, and the proteins for fusion/complex formation have been specifically chosen for their own inherent crystallizability. In this context, the removal of the Etk kinase domain is counter-intuitive, since the kinase domain has been crystallized, and the presence of a large, crystallizable, soluble domain should be beneficial. Unfortunately the aggregation of this domain at physiological pH could not be overcome. Therefore, the ability of Etk NM2 protein to crystallize is attributed entirely to its improved behavior in vitro during protein expression, solubilization, and purification.
2.7 Conclusions

There are many membrane proteins of biological interest that have yet to be structurally characterized. Often, soluble domains of membrane proteins are more likely to be crystallized, but these truncations do not give a full perspective of the target. Therefore, technical advancements will be required to provide generalized protocols effective on a wide range of membrane proteins. In this article, we have looked at membrane bound constructs of Etk and applied many approaches for improving crystallization success. The biggest advancement was the need to truncate the full-length protein to contain only the periplasmic domain and membrane bound helices. Truncating the protein does not decrease the value of the potential structure given that crystals structures are already available for the third, cytoplasmic kinase domain. We observed that the Etk NM2 protein truncation can be crystallized, but it contradicts some current “rules of thumb”, including that more soluble surface area should be added to the protein and that smaller detergents are best for crystallization. Therefore, Etk is a cautionary tale that reminds us that we still have a lot to learn about membrane protein crystallization, and that all strategies for a given target must still be considered.
Chapter 3

Quantification of Detergent Using Colorimetric Methods in Membrane Protein Crystallography

3.1 Foreword

The solubilization of membrane proteins is challenging and currently it is most commonly accomplished with detergent micelles. After the initial hit for crystallization of Etk NM2 in DDM was identified, it quickly became apparent that the condition was difficult to reproduce without careful control of both the protein and detergent concentrations. The 2,6-dimethylphenol assay was developed to provide a quick measurement of sugar-based detergents that destroys a minimal amount of protein sample. The original research manuscript was published in Acta Crystallographica D and was followed by a review article in Amino Acids that summarizes the importance of detergent measurement in crystallography and compares the colorimetric methods with other available techniques. Finally, we were invited to write a book chapter for publication in Methods in Enzymology, which is the version included here. This chapter contains an abbreviated review of the detergent measurement literature, expanded methods and figures from the original research paper on the 2,6-dimethylphenol assay, and protocols for additional colorimetric techniques.

3.2 Abstract

Membrane protein crystallography has the potential to greatly aid our understanding of membrane protein biology. Yet, membrane protein crystals remain challenging to produce. Although robust methods for the expression and purification of membrane proteins continue to be developed, the detergent component of membrane protein samples is equally important to crystallization efforts. This chapter describes the development of three colorimetric assays for the quantitation of detergent in membrane
protein samples and provides detailed protocols. All of these techniques use small sample volumes and have potential applications in crystallography. The application of these techniques in crystallization pre-screening, detergent concentration modification, and detergent exchange experiments is demonstrated. It has been observed that the concentration of detergent in a membrane protein sample can be just as important as the protein concentration when attempting to reproduce crystallization lead conditions.

### 3.3 Detergent in Membrane Protein Crystallography

High-resolution structures of membrane proteins remain severely underrepresented in structural databases (Arinaminpathy et al., 2009). Purification of membrane proteins requires the addition of detergent for both extraction of the protein from the lipid membrane and stabilization of the protein in vitro. Unfortunately, the addition of detergent can be detrimental to crystallization efforts (Hitscherich et al., 2000). Protein solubilization results in the formation of protein-detergent complexes (PDC), in which detergent coats a large portion of the protein surface. The flexible and dynamic detergent belt surrounding the protein can dictate the properties of the PDC and impede the formation of well-ordered crystal lattices (Hitscherich et al., 2000; Prive, 2007; Sonoda et al., 2010).

Productive crystal contacts can only be formed on protein surfaces that are exposed to the solvent (Sonoda et al., 2010). Therefore, the protein to detergent ratio of the final PDC and the buried surface area of the protein are important parameters to consider (DaCosta and Baenziger, 2002; Gutmann et al., 2007). Practically, detergents with a smaller hydrocarbon chain are desirable because they reduce the overall PDC size (Kunji et al., 2008; Marone et al., 1999) and allow tighter crystal packing (Garavito et al., 1996), which can produce PDC crystals with better than average diffraction (Prive, 2007; Sonoda et al., 2010). In contrast, smaller detergents are more likely to denature and destabilize proteins or dissociate physiologically relevant oligomers (Gan et al., 2011; Prive, 2007). Typically, many detergents must be
screened to find an appropriate stabilizing detergent that allows crystallization of a particular membrane protein.

The level of detergent and/or lipid present in solution is also a crucial factor in the success of crystallization trials (DaCosta and Baenziger, 2002; Eriks et al., 2003; Prive, 2007; Strop and Brunger, 2005). In the past decade, many techniques have been proposed for the measurement of detergent concentrations. They include thin layer chromatography combined with densitometry (Eriks et al., 2003), gas chromatography coupled with flame ionization detection or mass spectrometry (Shi et al., 2008, 2009), nuclear magnetic resonance (Maslennikov et al., 2007), contact angle measurement (Kaufmann et al., 2006), refractive index measurement (Strop and Brunger, 2005), or Fourier-transform infrared spectrometry (DaCosta and Baenziger, 2002). Many of the most popular detergents used in membrane protein crystallography can also be quantified using specific chemical reactions that produce a colored product. These colorimetric assays include the phenol assay for sugars (Prince and Jia, 2012; Urbani and Warne, 2005), the molybdate assay for total phosphate (Ames, 1966), the sulfuric acid assay for bile salts (Urbani and Warne, 2005), and the methylene blue/chloroform assay for the detection of SDS (Prive, 2007). These colorimetric methods require limited equipment and provide a quick assessment that is particularly advantageous to crystallographers.

3.4 2,6-Dimethylphenol Assay for Sugar-Based Detergents

The measurement of sugar concentrations using a phenol/sulfuric acid colorimetric reaction was originally proposed in the 1950’s and has been used in a wide range of applications (Dubois et al., 1956). This technique has been shown to effectively quantify sugar-based detergents in membrane protein samples (Urbani and Warne, 2005). More recently, a modification of this technique using 2,6-dimethylphenol as a colorimetric reagent with higher sensitivity has been applied specifically to the quantification of sugar-based detergents in protein samples for crystallography (Mallya and Pattabiraman,
The three most popular sugar-based detergents in crystallography are n-dodecyl-β-D-maltoside (DDM), n-decyl-β-D-maltoside (DM), and n-octyl-β-D-glucoside (OG). Despite the wide variety of detergents available, these detergents are collectively used in approximately 45% of successful crystallizations (Membrane Protein Databank or MPDB - http://www.mpdb.tcd.ie). This assay can also be extended to an additional 12% of cases in which less common sugar-based detergents have been used (MPDB). Therefore, this assay addresses the most popular detergents for crystallography and could have broad applications.

3.4.1 Overall Protocol

These reactions typically consist of combining three components: the detergent sample, the colorimetric reagent (phenol or its derivatives), and sulfuric acid under a fume hood. Using a sample tube specifically designed for high-temperature reactions, such as a SafeSeal 1.5 mL tube (Starstedt, Germany) is crucial, because regular 1.5 mL plastic tubes cannot withstand the heat generated during exothermic mixing and will leak. Concentrations of sulfuric acid always refer to the final proportion after mixing. After the reaction cools to room temperature, the absorbance of the product is measured using a sealed disposable plastic cuvette. This reaction functionalizes phenol at the para-position resulting in a new peak in the visible spectrum at 490 nm for phenol derivatives and at 510 nm for 2,6-dimethylphenol derivatives (Figure 3-1). In the following sections, the optimization of this assay for the detection of sugar-based detergents in membrane protein samples is described. The objective was to minimize the quantity of sample consumed while maintaining measurement accuracy, which is critical given the small volume, concentrated, protein samples used in crystallography.
Figure 3-1 Colorimetric absorbance peaks for the quantification of sugars. Absorbance peaks resulting from the reaction of 25 µL 20% phenol/2,6-dimethylphenol, 75% sulfuric acid, and 0.2% DDM.

3.4.2 Assay Optimization

Previous work has shown 2,6-dimethylphenol to be a superior prochromogen to phenol and many other substituted phenols (Mallya and Pattabiraman, 1997). Specifically, 2,6-dimethylphenol has been shown to be three times more efficient than phenol as a colorimetric reagent. In this first stage of optimization, the performance of phenol and 2,6-dimethylphenol is directly compared using the same reaction protocol with a total reaction volume of 1 mL. Both reagents were prepared as a 20% solution in absolute ethanol. Phenol produces a colorless solution, compared to the 2,6-dimethylphenol stock solution which is slightly yellow. All reactions were performed in triplicate using a one-step protocol with 750 µL sulfuric acid added last and then mixed by inversion. Three scenarios were compared in which 25 µL 20% phenol, 75 µL 20% phenol, or 25 µL 20% 2,6-dimethylphenol were used as the colorimetric reagent to detect the presence of 0.1% DDM (Figure 3-2). The absorbance values reported are after subtraction of
a blank reaction with water. It was apparent that 2,6-dimethylphenol had a greater reaction efficiency and performed better, especially at low concentrations. For this reason, it was decided to perform further optimizations with 2,6-dimethylphenol, as this would allow measurement with the smallest possible sample volumes.

![Graph showing comparison of reaction efficiency between phenol and 2,6-dimethylphenol.](image)

**Figure 3-2** Comparison of reaction efficiency between phenol (490 nm) and 2,6-dimethylphenol (510 nm) as colorimetric reagents. Reactions were performed with 5 µL sample volumes and 75% sulfuric acid with a total reaction volume of 1 mL. Reported values are the average of three independent measurements and error bars represent one standard deviation.

There are two common methods that have been proposed for phenol-sulfuric acid reactions, the one-step and the two-step procedures (Figure 3-3A,B). In the one-step protocol, all reagents are mixed at the beginning, with sulfuric acid added last. A potential source of error in this method is the highly exothermic mixing of the reaction after sulfuric acid addition. In the two-step protocol, the sugar is reacted with sulfuric acid before the addition of 2,6-dimethylphenol. This protects the colorimetric reagent from the exothermic mixing of sulfuric acid with water, but increases the time required. Although the two-step procedure does increase the signal at most detergent concentrations, the increased signal at very low concentrations is minimal (due to increased absorbance from the blank), resulting in a minimal increase in sensitivity (Figure 3-3C).
Figure 3-3 Summary of the phenol/2,6-dimethylphenol reactions. Reaction schematics for the one-step (A) and two-step (B) colorimetric 2,6-dimethylphenol/sulfuric acid reaction for the measurement of sugars in aqueous solution. (C) Comparison of the resulting absorbance from the one-step and two-step reaction procedures using 75% sulfuric acid and 25 µL of 2,6-dimethylphenol. Reactions were performed in triplicate and error bars represent one standard deviation.

Previous work has also shown that the ratio of sulfuric acid and colorimetric reagent can affect the absorbance of the final product (Dubois et al., 1956; Mallya and Pattabiraman, 1997). The optimum sulfuric acid concentration is around 75% for the two-step procedure (Mallya and Pattabiraman, 1997). When sulfuric acid concentration in the one-step procedure is optimized to 65%, the “lost signal” is
recovered (compare Figure 3-3C and Figure 3-4A). A comparison of the one-step and two-step procedures reveals that both effectively quantify DDM with no significant difference in accuracy. Therefore, the more efficient and convenient one-step procedure is selected. The volume of 20% 2,6-dimethylphenol required was also optimized with 25 µL proving sufficient to obtain good signal (Figure 3-4B).

The length of the assay was also investigated to ensure the colorimetric reaction was complete before measurement. A reaction time course was performed to monitor the development and decay of absorbance at 510 nm. The standard deviation of the mean absorbance readings taken between 40 and 90 minutes, 0.013, is less than the average standard deviation in the triplicate measurements, 0.027 (Figure 3-4C), so readings taken within this time period are stable and should provide an accurate detergent concentration. The final optimized protocol for the 2,6-dimethylphenol assay employs the one-step procedure and requires the combination of 25 µL 20% 2,6-dimethylphenol, 650 µL sulfuric acid, 5 µL sample, and 320 µL water. Measurements are performed after 40 minutes. The optimized protocol has been used to generate a standard curve with linearity over a 20-fold range and a correlation coefficient of 0.99 (Figure 3-4D). This is now an extremely sensitive assay and samples should be diluted so that the resultant absorbance is within the standard curve and a minimum of 5 µL is always added to the reaction, allowing consistent pipetting. When very dilute samples are measured, the sample volume can be increased (by reducing the water added) to further increase the sensitivity of the reaction. Reactions are always performed in triplicate and a blank assay is performed in parallel with all experimental assays. The standard curve is re-generated each time the 2,6-dimethylphenol stock solution is re-made.
Figure 3-4 Optimization of the 2,6-dimethylphenol assay. (A) Optimization of absorbance signal from a one-step reaction with 5 µL 0.1% DDM, 25 µL 2,6-dimethylphenol, and varying concentrations of sulfuric acid. (B) Optimization of absorbance signal from a one-step reaction with 5 µL 0.1% DDM, 75% sulfuric acid, and varying amounts of 2,6-dimethylphenol. (C) Time course with 5 µL 0.1% DDM, 65% sulfuric acid, and 25 µL 2,6-dimethylphenol. (D) Standard curve generated by the reaction of 5 µL samples with 65% sulfuric acid and 25 µL 2,6-dimethylphenol using a one-step assay. Reactions were performed in triplicate and error bars represent one standard deviation.

3.4.3 Interfering Compounds

When tested with water alone, a resultant absorbance is observed. Therefore, subtraction of the absorbance from an appropriate blank is always required. To ensure that the 2,6-dimethylphenol assay can be used on complex protein samples, cross-reaction with common buffer components and protein was also tested (Figure 3-5). These tests used 5 µL samples of highly concentrated samples. The buffers, salt,
reducing agents, and protein tested were found to contribute negligible signal. The compound with the most interference, NaCl, gave only a small increase at 4 M, therefore the more typical concentrations in protein buffers (100-300 mM) would be expected to give negligible signal. More caution must be applied with additives like glycerol, which is sometimes added to protein buffers to improve solubility. In these experiments both glycerol and Brij 35 produced significant signal. When these compounds must be used, it is possible to subtract out their contribution using appropriate blank reactions.

![Absorbance (510 nm)](image)

**Figure 3-5** Reaction of 5 µL control samples with 65% sulfuric acid and 25 µL 2,6-dimethylphenol. Reactions were performed in triplicate. Error bars represent one standard deviation.

### 3.5 Molybdate Assay for Total Phosphate

This procedure was originally described by Bruce Ames in 1966 for the detection of organic phosphates (Ames, 1966). It has been suggested in multiple publications that this assay can be further applied to the fos-choline series of phospholipid derived detergents. These detergents combine a very
small head group with long aliphatic tail groups and are of interest to crystallographers because of their small micelle size. The common detergents in this class are fos-choline 12 and fos-choline 14, but thus far they represent less than 1% of the deposited structures in the membrane protein databank (MPDB). In the following sections, the effectiveness of a total-phosphate assay for the quantification of fos-choline 12 is discussed. This provides an additional advantage to the use of this detergent in membrane protein crystallization.

### 3.5.1 Protocol

This protocol is adapted from the original publication (Ames, 1966) with minor modifications to take advantage of modern equipment. This assay requires multiple stock solutions that should be prepared in advance (Table 3-1). The protocol consists of combining a 5 µL sample with magnesium nitrate and ethanol (30 µL) in a glass test tube followed by ashing over a strong flame. The resulting powder is then dissolved in 350 µL of 0.5 N hydrochloric acid by vortexing and 300 µL is transferred to a SafeSeal 1.5mL tube (Starstedt, Germany). The addition of excess acid allows exactly 300 µL of sample to be accurately transferred. The sample is then boiled on a hot plate for 15 minutes. After the tube cools, 700 µL of the ascorbic acid/molybdate solution is added. Incubation at 45 °C on a second hot plate results in a phosphomolybdate complex that is reduced by ascorbic acid to produce the colorimetric product (Figure 3-6A). Total phosphate in the sample can be quantified by measurement of the absorbance at its peak wavelength of 825 nm (Figure 3-6B).

### Table 3-1 Stock solutions required for the total phosphate assay.

<table>
<thead>
<tr>
<th>Solution</th>
<th>Composition</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% Magnesium Nitrate</td>
<td>1 g in 10 mL 95% ethanol</td>
<td>Room Temperature</td>
</tr>
<tr>
<td>0.5 N Hydrochloric Acid</td>
<td>1.8 mL concentrated HCl in 100 mL water</td>
<td>Room Temperature</td>
</tr>
<tr>
<td>10% Ascorbic Acid</td>
<td>1 g in 10 mL water</td>
<td>4°C up to a month</td>
</tr>
<tr>
<td>Ammonium Molybdate</td>
<td>0.42 g ammonium molybdate tetrahydrate 2.86 mL sulfuric acid in 100 mL water</td>
<td>Room Temperature</td>
</tr>
<tr>
<td>Colorimetric Mix</td>
<td>Mix 1 part 10% ascorbic acid to 6 parts ammonium molybdate</td>
<td>12 h on ice</td>
</tr>
</tbody>
</table>
Figure 3-6 Summary of the molybdate assay. (A) Procedure for the total organic phosphate assay. (B) Absorbance spectrum before and after reaction of fos-choline 12 in the total phosphate assay.

3.5.2 Assay Optimization

The main parameter optimized for this assay was the incubation time for the colorimetric reaction with ammonium molybdate and ascorbic acid. A 45°C incubation was performed using a heat block and it was found that the maximum absorbance signal was measured after 25 minutes (Figure 3-7A). Given the variation in absorbance with incubation time, this becomes an important variable to ensure the accuracy of the measurement.

Another potential source of error in this assay is the cleanliness of the glass test tubes used. Interference has been noted from silicates evolved from the glass during ashing and residue from trisodium phosphate detergents used during cleaning (Ames, 1966). Given the inexpensive nature of test
tubes, new glassware from the same manufacturer is recommended. The ammonium molybdate/ascorbic acid reagent is present in excess and will readily produce a dark blue solution with absorbance >3 AU when high concentrations of phosphate are tested. Therefore, it is important to dilute the sample prior to measurement to fit within the dynamic range of the assay as demonstrated by the standard curve (Figure 3-7B). The 5 µL sample volume of this assay is suitable for the small volume samples found in membrane protein crystallography.

Figure 3-7 Optimization of the molybdate assay. (A) Reaction time course of the ammonium molybdate/ascorbic acid reaction using 5 µL of 0.5% FC-12 as the substrate. (B) Standard curve with FC-12 and the total phosphate assay. Reactions were performed in triplicate. Error bars represent one standard deviation.

3.5.3 Interfering Compounds

Using a similar panel as the 2,6-dimethylphenol assay, the potential for interference from common protein buffer components as well as other detergents were assessed (Figure 3-8). HEPES buffer and glycerol were incompatible with the ashing process, resulting in black flakes that did not dissolve in the hydrochloric acid and interfered with the final absorbance measurement. As expected, phosphate buffers also resulted in a reaction with absorbance greater than 3 and obscured the measurement of phosphate from the detergent. There was minimal absorbance generated from highly concentrated salt, reducing agent, or protein and therefore, this assay is applicable to membrane protein samples as long as an
appropriate blank and a compatible buffer, such as Tris, are used. Caution must still be used when applying this technique to phosphorylated proteins as the additional phosphate will be detected.

![Absorbance (510 nm) chart](image)

Figure 3-8 Total phosphate assay with 5 µL of compatible control samples. Reactions were performed in triplicate. Error bars represent one standard deviation.

### 3.6 Assay of Bile Salt with Sulfuric Acid

The direct reaction of sulfuric acid with bile acid detergents such as CHAPS and cholate (Urbani and Warne, 2005) allows for the quantification of these detergents. The reaction of the sulfuric acid with the hydroxyl group in the cholate ring produces a colored product with an absorption peak at 390 nm that can be used for quantitation (Figure 3-9A). CHAPS and CHAPSO are present in approximately 2% of deposited membrane protein structures (MPDB). In this assay, a 5 µL sample is diluted with water and then sulfuric acid is added to a final volume of 1 mL in a SafeSeal 1.5 mL tube (Sarstedt, Germany). The reaction is allowed to cool and then the absorbance signal is measured (Figure 3-9B).
3.6.1 Assay Optimization & Interfering Compounds

Given the observed variation in absorbance with sulfuric acid concentration in the 2,6-dimethylphenol/sulfuric acid assay, a similar optimization strategy was applied here. In this case, the sulfuric acid was varied from 50-90% of the reaction volume and maximum absorbance occurred with 85% sulfuric acid (Figure 3-10A). The reaction time for color development was also considered and maximum absorbance was observed after 30 minutes (Figure 3-10B). Overall, this assay also performs well with small sample volumes of 5 µL, producing a standard curve with a linear range from 0.02 to 0.4% detergent (Figure 3-10C).
The panel of 12 control compounds was also tested with 85% sulfuric acid to assess potential side reactions (Figure 3-10D). The sulfuric acid reaction with bile salts appears valid for application to membrane protein samples, with the only cross-reaction being with DDM.

Figure 3-10 Optimization of the bile salt/sulfuric acid assay. (A) Absorbance generated by the reaction of sulfuric acid with 5 µL 0.3% CHAPS detergent and a total volume of 1 mL. (B) Reaction time course of 0.3% CHAPS with 85% sulfuric acid. (C) Standard curve for the CHAPS assay with 85% sulfuric acid. (D) Reaction of 5 µL of control samples with 85% sulfuric acid. All reactions were performed in triplicate. Error bars represent one standard deviation.
3.7 Applications of Detergent Measurement

3.7.1 Detergent Content and Crystallizability

During initial crystal screening of Etk NM2 (an inner membrane protein from *Escherichia coli*) solubilised in DM, a condition was identified which produced pseudo-crystals after phase separation. In an attempt to produce crystals from this condition, the effect of detergent concentration on the phase behaviour of the crystallization drops was explored. Etk NM2 samples with a protein concentration of 5.2-5.8 mg/mL were adjusted to contain three different concentrations of DM: 1.2, 2.0, and 4.9%. These samples were tested in parallel above a communal well solution of 0.1 M MES pH 5.5 and 30% PEG2000 MME. After 24 hours, divergent phase behaviour could be observed (Figure 3-11), although the original pseudo-crystals were not reproduced. This experiment gives a direct example of the significant effect detergent concentration can have on a crystallization experiment, even when all other conditions are kept constant.

![Figure 3-11](image)

**Figure 3-11** Effect of increasing concentrations of detergent on the phase behaviour of Etk NM2 crystallization drops. Trials were set up with Etk NM2 containing (A) 1.2%, (B) 2.0%, and (C) 4.9% DM. Drops contained 2 µL protein mixed with 2 µL well solution.

In order for crystal nucleation to occur, it is essential that super-saturation is achieved. Even in cases where seeding is employed, the crystallization drop must be maintained in the meta-stable zone of the phase diagram for crystal growth to continue. These conditions are achieved using precipitants (i.e. PEG,
alcohols, salts) mixed with the protein. It is commonly accepted that a minimum threshold for protein concentration exists in any specific crystallization experiment to ensure super-saturation and/or a metastable state are reached. The concentration of protein required is determined by the solubility of the protein, concentration of stabilizing additives, and the concentration of precipitant to be used. In membrane protein samples, the concentration of detergent is also an important variable. The Etk NM2 construct solubilized in DDM produces small-medium hexagonal prism crystals (Figure 12A). Crystallization of Etk NM2 was performed using the hanging drop method and a well solution containing a three-component mixture of 0.1 M MES (pH 6.0), 1-5% PEG3000 and 20-30% PEG200. A minimum of 24 wells were set up for each protein/detergent ratio tested. Without prior measurement of detergent concentration, the success rate for crystallization was less than 10%. Frozen samples from previous work, in addition to several new samples, were assayed for both detergent and protein concentration. By comparing detergent concentration to crystallization success, an optimal range of protein and detergent concentrations for crystallization was determined (Figure 12B). Based on this analysis, it was determined that crystallization requires a minimum protein concentration of 6.9 mg/mL and a maximum of 10.5 mg/mL. In addition the amount of detergent present in the sample has an important role, with crystallization most likely when the detergent concentration is between a detergent/protein ratio of 0.65 and 1.10 mg DDM/mg protein. Incorporation of this information has significantly increased the frequency of crystallization success.
3.7.2 Reducing Detergent Concentration in Centrifugal Concentrators

The usual school of thought is that variations in protein samples for crystallization can be minimized by careful application of a well-established and reproducible protocol. The difficulty with membrane protein samples is the protein concentration process, during which protein-free detergent micelles are also enriched (Prive, 2007; Shi et al., 2008; Urbani and Warne, 2005). The detergent concentration factor of centrifugal concentrators has been shown to vary between manufacturers and within units from the same...
manufacturer (Maslennikov et al., 2007). In order to minimize detergent variation, one could use the same concentrator repeatedly, but variation can still occur as the concentrator ages. Therefore, the concentration of detergent is often unknown after protein concentration and obtaining reproducible concentrations can be difficult. Despite this issue, ultrafiltration in centrifugal concentrators remains the most common technique for obtaining reproducible protein concentrations (Maslennikov et al., 2007).

The ability to quickly assess the detergent content and potential crystallizability of a sample is a valuable predictor of success. Even more compelling is the ability to correct the detergent content of samples that fail the criteria for crystallization. The addition of more detergent is relatively straightforward. More complicated is the removal of excess detergent from a pre-made sample. In this case there are limited options, though the use of detergent absorbent beads is being offered as a possible solution. These beads have been successfully applied in 2D crystallization experiments and complete detergent removal for protein exchange into lipids or other lipophilic alternatives (Arunmanee et al., 2014; Rigaud et al., 1997). In this case study, it is demonstrated that the concentrators themselves and serial dilution can also be used to make minor adjustments to the amount of detergent present. Given that the protein content can be measured non-destructively using the optical density of the solution at 280 nm and the limited loss of sample in the 2,6-dimethylphenol assay, one can confidently adjust the sample to be within the range of expected crystallization success (Figure 3-13). In this case, the starting sample of Etk NM2 solubilized in DDM contains 6.7 mg/mL protein and 1.0 mg DDM/mg protein, which is on the edge of the ideal zone for crystallization. Through four cycles of dilution and concentration the protein concentration is adjusted to 8.4 mg/mL and the detergent ratio is reduced to 0.81 mg DDM/mg protein, which makes this sample now a good candidate for crystallization success. Each cycle consists of diluting 2 mL of protein/detergent solution with 10 mL buffer without detergent and re-concentrating the protein in a 50 kDa cutoff concentrator (Millipore) back to 2 mL (about 25 minutes). When performing this
analysis the protein to detergent ratio is used as a more convenient guideline than absolute concentration, because slight variations in the volume after re-concentration will not negatively affect the outcome.

![Figure 3-13 Detergent dilution using a centrifugal concentrator. (A) Detergent concentration is monitored as the ratio of detergent to protein to account for small variations in sample volume after each round of concentration. (B) Position of the protein batch on the crystallization prediction graph at the start and the end of detergent dilution. Protein samples with confirmed crystals are in red and samples that did not crystallize are in blue.](image)

3.7.3 Detergent Exchange

Detergents with long aliphatic side chains can be more appropriate for extracting proteins from the membrane. During purification, some researchers find it valuable to extract membrane proteins using an inexpensive, low CMC (critical micelle concentration), detergent and later change to a potentially more expensive, specialized detergent that is more suitable for crystallization. A recent example of this application is crystallization of membrane-bound pyrophosphatases. In this case, crystals could be obtained from n-dodecyl-β-D-maltoside, but crystals with better diffraction were obtained after exchange into alternative detergents, the best being octyl glucose neopentyl glycol (Kellosalo et al., 2013). During these experiments, detergent measurement techniques that both quantitate and identify the detergent are especially valuable because they can be used to continually monitor the exchange for consistency. These
techniques include thin layer chromatography, gas chromatography, and nuclear magnetic resonance (Eriks et al., 2003; Maslennikov et al., 2007; Shi et al., 2009). NMR has been specifically used to monitor the exchange of fos-choline 14 for n-dodecyl-β-D-maltoside and n-decyl-β-D-maltoside in samples of Etk (Maslennikov et al., 2007). In addition, TLC has been applied to monitor the exchange of sarkosyl with n-dodecyl-β-D-maltoside in samples of the mitochondrial citrate transport protein (Eriks et al., 2003).

It is not uncommon for detergent exchange to be performed between detergents of different classes, in which case colorimetric assays could be applied to monitor the exchange and determine an appropriate endpoint when exchange is complete. For example, a protein is purified in DDM for its low CMC and high efficiency during extraction from the membrane, but then the protein is exchanged into a smaller micelle detergent that is more favorable in crystallization. There are desirable detergents that do not contain sugar groups. In any of these cases the 2,6-dimethylphenol/sulfuric acid assay can be applied to ensure that DDM in completely removed from the sample.

As an example, the exchange of DDM to FC-12 in a membrane protein sample bound to nickel resin is explored. By monitoring the detergent content of the wash fractions using colorimetric methods we can determine when DDM is no longer being eluted from the column and use this to determine the number of washes required. In the case of the exchange to FC-12, we see that after 7 washes (10 mL) with 0.05% FC-12 there is no detectable DDM being eluted from the column (Figure 3-14). In this experiment, 50 µL samples were tested with the 2,6-dimethylphenol assay to increase sensitivity. When the protein is subsequently eluted with imidazole, the eluted protein also does not contain a measurable quantity of DDM. In this example, the effect of equilibration time between washes is also evident. The average equilibration time between washes was 5 minutes, except between washes 4 and 5 when the column was equilibrated for 1 hour instead. In wash 5, we see an increased elution of DDM, presumably from the increased time for FC-12 and DDM to exchange. This would suggest that equilibration time is extremely important and that longer equilibration times could allow exchange to be performed with fewer washes.
Figure 3-14 Exchange of DDM to FC-12 by consecutive washes of Ni-agarose bound protein. Wash 0 represents the initial loading and elution of excess DDM. Washes 2-7 each represent 10 mL washes with 0.05% FC-12. Average equilibration time between washes was 5 minutes, with the exception of wash 5, which was equilibrated for 1 hour.

3.8 Concluding Remarks

The detergent concentration used during initial crystallization screens for membrane proteins remains an important consideration. It has been proposed that minimization of excess detergent and “empty” detergent micelles would be favourable to crystallization efforts (DaCosta and Baenziger, 2002; Wiener, 2004). It has also been suggested that concentrations near a detergent’s cloud point (i.e. the phase boundary in which inter-micellar attractive forces drive micelles into a separate phase) can show correlation with crystallization of the protein-detergent complex (Hitscherich et al., 2001; Wiener, 2004). In the specific case study examined here, it was found that detergent concentrations between 0.65 and 1.10 mg DDM/mg protein were most likely to produce crystals. Overall, the detergent required for the crystallization of a specific membrane protein is likely case dependent, and the tracking of detergent concentration during crystallization efforts may be critical to success.
Many factors are worth considering when measuring detergent concentration, including the availability of equipment, measurement precision, processing time, and sample volume required. This ensures smooth incorporation of the technique into the membrane protein preparation and crystallization pipeline. It has been found that working quickly with membrane proteins in vitro can be beneficial for crystallization (Prive, 2007; Screpanti et al., 2006). Therefore, the time-frame required for colorimetric assays (40-90 minutes) is practical for incorporation into the timeline required by membrane protein crystallization. As a group, the three colorimetric assays presented in this document cover a range of detergent options for crystallization experiments. Yet, there are still popular detergents like lauryldimethylamine-N-oxide that have no specific chemical group which can be targeted by these techniques. In these cases, the more general techniques listed in the introduction will be needed. Moving forward, the ability to measure detergent will be one more factor to consider when choosing a detergent for crystallography experiments.
Chapter 4

An Unexpected Duo: Rubredoxin Binds Nine TPR Motifs to Form LapB, an Essential Regulator of Lipopolysaccharide Synthesis

4.1 Foreword

This thesis also explores a second avenue of research involving LapB, a protein, which is now known to be a regulator of the lipopolysaccharide synthesis pathway. This published manuscript describes the crystal structure of LapB with a focus on the arrangement between the two folding domains. Functional studies are used to support the structural observations. Additional methods regarding the exploitation of Zn-SAD phasing and LapB are presented in Chapter 5.

4.2 Abstract

Lipopolysaccharide (LPS) synthesis and export are essential pathways for bacterial growth, proliferation, and virulence. The essential protein LapB from Escherichia coli has recently been identified as a regulator of LPS synthesis. We have determined the crystal structure of LapB (without the N-terminal transmembrane helix) at 2 Å resolution using zinc-SAD phasing derived from a single bound zinc atom. This structure demonstrates the presence of nine tetratrico peptide repeats (TPR) motifs, including two TPR folds that were not predicted from sequence, and a rubredoxin-type metal binding domain. The rubredoxin domain is bound intimately to the TPR motifs, which has not been previously observed or predicted. Mutations in the rubredoxin/TPR interface inhibit cell growth and *in vitro* studies indicate that these modifications cause local displacement of rubredoxin from its binding site without changing LapB’s secondary structure. LapB is the first reported structure to contain both a rubredoxin domain and TPR motifs.
4.3 Introduction

Interest in lipopolysaccharide assembly protein B (LapB) began with two studies of the Keio collection of E. coli single gene mutants where the lapB (previously known as yciM) knock-out strain showed defects in biofilm formation and increased sensitivity to 21 different antibiotics (Baba et al., 2006; Liu et al., 2010; Tabe et al., 2007). Further investigation has revealed that lapB is an essential gene and the Keio knock-out mutant also contains a compensatory mutation in the lpxC gene (Mahalakshmi et al., 2014). Study of a ΔlapB suppressor-free strain showed extreme growth defects under laboratory conditions and elevated lipopolysaccharide (LPS), heterogeneous LPS, and accumulation of LPS precursors (Klein et al., 2014). LPS is a key component of the outer-leaflet of the bacterial outer-membrane that is found in gram-negative bacteria (Whitfield and Trent, 2014). LPS forms an innate permeability barrier that protects the cells from antibiotics, detergents, and dyes due to the low fluidity of the LPS hydrocarbon domain and bridging of the negatively-charged components of lipid A and the inner core sugar by divalent cations (Nikaido, 2003).

The regulation of LPS synthesis requires the degradation of the LpxC enzyme by the essential membrane-bound protease FtsH (Führer et al., 2006). The LpxC enzyme catalyzes the first committed step of lipid A synthesis and tight regulation of this enzyme is essential to cell survival (Ogura et al., 1999). More recently it has been shown that the degradation of LpxC is somehow mediated by LapB (Klein et al., 2014; Mahalakshmi et al., 2014). Direct binding between LapB and LpxC has not been observed, leaving the molecular mechanism behind LapB mediated degradation of LpxC unclear. The LapB protein has been pulled down with WaaC and LptD, which are also implicated in LPS biogenesis. This suggests that LapB may act as a central scaffold that coordinates the actions of various proteins in the LPS synthesis and export pathways (Klein et al., 2014).

LapB contains three major structural motifs; the N-terminal transmembrane helix, several tetratricopeptide repeats (TPR), and a C-terminal rubredoxin metal-binding domain (Nicolaes et al.,
The N-terminal transmembrane helix anchors LapB to the inner membrane of gram-negative bacteria with the soluble domain in the cytoplasm (Nicolaes et al., 2013).

TPR proteins are members of the solenoid family and contain helix-turn-helix folds with 34 amino acids each and a consensus sequence defined by a pattern of small and large hydrophobic residues (Karpenahalli et al., 2007; Kobe and Kajava, 2000). TPR motifs are most commonly found in groups of three consecutive repeats (Andrea et al., 2003). The structures of many TPR-containing proteins have been recently solved, including complexes with small-molecules, peptides, or proteins bound to the concave face (Fodor et al., 2015; Pal et al., 2014; Wang et al., 2011).

Rubredoxin proteins form small non-heme iron binding sites that use 4 cysteine residues to coordinate a single metal ion in a tetrahedral environment. Rubredoxins are most commonly found in bacterial systems, but they have also been found in eukaryotes (Chen et al., 2006; Schweimer et al., 2000). Examples of rubredoxin domains within larger proteins have also been identified. The key features of these rubredoxin-like domains are the extended loops or “knuckles” and the tetra-cysteine mode of iron binding (Bitto et al., 2008). Rubredoxins are usually implicated in redox reactions/electron transfer. Specific roles are known in a few systems including n-alkane oxidation in P. aeruginosa (Hagelueken et al., 2007), and an electron acceptor for CO dehydrogenase in Acetobacterium woodii (Ragsdale et al., 1983), but the exact role of rubredoxin and rubredoxin domains remains poorly understood in most systems (Chen et al., 2006).

Although the mechanism of LapB function remains unclear, it has been shown that both the metal binding rubredoxin and inner membrane localization of LapB are essential to the proper functioning of this regulator (Nicolaes et al., 2013). We have determined the structure of the LapB cytoplasmic domain to 2.0 Å by X-ray crystallography. The direct functional implications of this structure are discussed using mutation studies performed both in vitro and in vivo. LapB is the first protein to be structurally characterized that contains both TPR motifs and a rubredoxin domain.
4.4 Results and Discussion

4.4.1 Optimized LapB and Crystallization

During initial trials, it was noted that LapB truncated to remove the N-terminal transmembrane helix (residues 19-389) could be expressed and purified as a soluble protein, but it was prone to degradation that could be minimized by further truncation of the protein to residues 50-389. In addition, LapB was pink-orange in color indicative of partial iron binding and consistent with parallel work that showed purified LapB bound to 20% iron and 80% zinc (Nicolaes et al., 2013). The metal-binding properties of LapB were further explored by expressing LapB in minimal media and altering availability of zinc ions. When extra zinc was provided, iron binding was significantly reduced and could not be detected by absorbance in the visible spectrum. In contrast, when zinc was removed, increased iron binding was observed suggesting metal binding is competitive (Figure 4-1A). Zinc-bound centers are known to have increased stability over iron-bound centers (Petros et al., 2006) and it was zinc-bound LapB that crystallized to form clusters of plates in space group P2₁ over a 3-8 week growth period (Figure 4-1B). The diffraction of crystals produced from LapB (50-389) was improved by approx. 1 Å in resolution over crystals grown from the LapB (19-389) construct. X-ray data reduction, phasing, model building, and refinement statistics can be found in Table 4-1.
Figure 4-1 LapB optimization and crystallization. (A) Purified LapB expressed with 10, 2, or 0 µM zinc sulfate available in minimal expression media. Iron bound protein produces two characteristic peaks at 370 and 485 nm in the visible absorbance spectrum marked by *. Inset: LapB samples in 96-well plate used for the collection of absorbance spectra. Iron bound protein is red-orange in color. (B) Crystals of LapB grow in layers of stacked plates in a starburst pattern. UV-image of the initial crystallization hit with LapB 19-389. (C) Polarized light image of optimized crystals from LapB 50-389. Scale bar is appropriate for both B and C.
Data reduction, experimental phasing, model building, and refinement statistics for the structure determination of LapB.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SAD Phasing</th>
<th>Refinement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁</td>
<td></td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td>a=38.31, b=152.25, c=65.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>α=90.0, β=93.91, γ=90.0</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.28292</td>
<td></td>
</tr>
<tr>
<td>Resolution Range (Å)</td>
<td>50-4.0 (4.1-4.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40.1-2.0 (2.1-2.0)</td>
<td></td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>12429</td>
<td>97844</td>
</tr>
<tr>
<td>I/Sigma</td>
<td>63.74(57.31)</td>
<td>11.01(2.03)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100.0(100.0)</td>
<td>98.1(96.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>21.2</td>
<td>3.15</td>
</tr>
<tr>
<td>Rmerge</td>
<td>5.0(5.5)</td>
<td>6.4(66.4)</td>
</tr>
<tr>
<td>CC*</td>
<td>99.9(99.9)</td>
<td>99.7(68.2)</td>
</tr>
<tr>
<td>HySS Figure of Merit</td>
<td>0.498</td>
<td></td>
</tr>
</tbody>
</table>

Refinement

| Rwork/Rfree      | 0.2056/0.2545 |
| r.m.s.d. Bond Length (Å) | 0.010 |
| r.m.s.d. Bond Angles (*) | 1.103 |
| Clashscore       | 4.2         |
| Ramachandran Favored | 96.37% |
| Ramachandran Outliers | 0.15% |
| Rotamer Outliers | 0.43%       |
| No. Protein Atoms | 5026        |
| No. Solvent Atoms | 265         |

Unique reflections = number of Friedel pairs for both anomalous and native data processing.

4.4.2 LapB Structure

Crystallized LapB is a dimer in the asymmetric unit with residues 67-389 in molecule A and residues 52-389 in molecule B (Figure 4-2A). Alignment of the two molecules demonstrates the same overall fold with a RMSD of 1.29 Å and backbone variation is mostly limited to the N-terminal TPR motifs. Analysis of this structure by PDBePISA (Krissinel and Henrick, 2007) suggests that this protein may form a biological dimer with an unusual ring shape (Figure 4-2B).
Figure 4-2 LapB structure. (A) Cartoon diagram of the LapB dimer crystallized in the asymmetric unit. Bound zinc atoms are indicated by black spheres. (B) Cartoon diagram of the LapB ring dimer predicted by PDBePISA (Krissinel and Henrick, 2007). (C) Monomer of LapB highlighting the position of the nine TPR motifs. (D) Alignment of the rubredoxin domain (red) with representative models of isolated rubredoxin (purple - 7RXN) and the rubredoxin-like domain in rubrerythrin (blue - 1B71).

Sequence predictors, such as TPRpred, have identified seven TPR motifs from the LapB sequence (Karpenahalli et al., 2007). The second helix of the first repeat is seen in molecule B of the crystal structure and repeats #2-7 are present in both molecules (Figure 4-2C). The structure reveals two additional TPR motifs that were not predicted from sequence. Based on the structure of these folds, the sequences of repeats 8 and 9 can be aligned (Figure 4-3A). Repeats 8 and 9 have 4 conserved positions
substituted and there is one residue missing between repeats 8 and 9. Overall, most conserved positions still contain a hydrophobic residue and these repeats form the same helix-turn-helix fold. Therefore, structure determination of LapB has revealed the presence of 2 additional TPR motifs with low sequence conservation.

### A

<table>
<thead>
<tr>
<th>Consensus</th>
<th>W LG Y A F A P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeat 1</td>
<td>LSRDYVAGVNFLSNOQDKAVDLFLDLMLKEDETGT</td>
</tr>
<tr>
<td>Repeat 2</td>
<td>VEAHLTLGNLRSRGGEVDAIRHQLMESAFLT</td>
</tr>
<tr>
<td>Repeat 3</td>
<td>LLAIQOLGDPYMAAGLYDARAEDMFQTLDELD</td>
</tr>
<tr>
<td>Repeat 4</td>
<td>IGALQQLQIYQATSEWQKIDVADERLVKLGDK</td>
</tr>
<tr>
<td>Repeat 5</td>
<td>AHFYECAQLQHMASDDLDRAMTLKKGAAAADKNS</td>
</tr>
<tr>
<td>Repeat 6</td>
<td>ARVSIIMGRVFMKGEYAKAVSLOQTVQISQREL</td>
</tr>
<tr>
<td>Repeat 7</td>
<td>SETLMLQTCYQLGKTAWEFLQRAVEENTGA</td>
</tr>
<tr>
<td>Repeat 8</td>
<td>DAELLMADILLIEARDGSEAQVITRQLQHRPT</td>
</tr>
<tr>
<td>Repeat 9</td>
<td>MRVFHKLMDYHLNEAEEGRKESLMLRDMVEEK</td>
</tr>
</tbody>
</table>

### B

E. coli LapB ----PRYRCQKCGFTAYTL--YWHCPSCRAWSTIKIPIRGLDLG
D. vulgaris Rbr -REQATKWRCRNCGYVHEGTAPELCACAHFKAHFELLGINW-
P. furiosus Rbr DIEIKKVYICPICGYTAVDE-APYCPVCAPKEKTVVFT----

Figure 4-3 Sequence alignments for key domains in LapB. (A) Alignment of the 9 tetratrico peptide repeats in LapB. Substitutions in repeats 8 and 9 are underlined. (B) Alignment of the rubredoxin domain with rubrerythrin (Rbr). Poorly conserved residues are underlined.

Phyre2 structure prediction for LapB indicated several TPR motifs followed by a loosely associated rubredoxin domain (Kelley and Sternberg, 2009; Klein et al., 2014). This is consistent with the isolated nature of most rubredoxin proteins. However, the crystal structure of LapB reveals an extremely tight and unexpected binding of rubredoxin to 9 highly organized TPR motifs. This interaction is mediated by two patches of contact, one dominated by large aromatic amino acids, and the second dominated by electrostatic amino acids (Figure 4-4A). These patches are located on the concave face of TPR motifs 5-7, in agreement with previously identified TPR complexes (Fodor et al., 2015; Pal et al., 2014; Wang et al., 2011). This leaves additional binding space for proteins to dock to motifs 1-4 and motifs 7-9, consistent
with the proposed role of LapB as a protein scaffold (Klein et al., 2014). Proteins docking to these locations on the concave face of the TPR motifs would also likely interact with the bound rubredoxin domain as a result of its central location (Figure 4-2A).

Figure 4-4 Rubredoxin-TPR binding interface (A) Binding of rubredoxin to the TPR motifs. Residues in the aromatic region (green) and the electrostatic region (yellow) are displayed as sticks. (B) Growth of *E. coli* at 42 °C during expression of wild type LpxC and during expression of both LpxC and LapB. Data plotted as the average and standard deviation of triplicate measurements. (C) Rescue of LpxC overexpression by LapB. ΔlapB Control is the lapB knock-out strain transformed with pDL804 and pMS604 for antibiotic resistance and expresses neither protein. LapB Control and LpxC Control contain one protein expression plasmid and pDL804 or pMS604 as required for antibiotic resistance. LpxC Control (orange) shows the expected growth defect. Rescue of LpxC overexpression is achieved by wild type LapB (purple) and some LapB mutants (blue). Mutant LapB proteins incapable of rescuing the growth defect are shown in green.
Isolated rubredoxin has a very well conserved fold, with 10 nearly identical structures available from different organisms. LapB’s rubredoxin domain is more similar to the rubrerythrin family’s rubredoxin domain than isolated rubredoxins (Figure 4-2D). Rubrerythrin family proteins have been identified in air-sensitive bacteria and contain a four-helix bundle with a diiron site and the rubredoxin domain (DeMaré et al., 1996). Rubrerythrins have been implicated in the reduction of hydrogen peroxide and the rubredoxin domain likely restores the reduced state of the diiron site post-catalysis (Iyer et al., 2005). When the structure of LapB is compared to rubrerythrin it is noted that side chains of W369 and W377 from the aromatic patch are in positions occupied by much smaller side chains. In addition, K381 and R384 from the electrostatic patch at the C-terminus align with hydrophobic residues in rubrerythrin (Figure 4-3B and Figure 4-4A). This opens the possibility that these positions may have evolved to bind the TPR motifs. Overall, the TPR motifs have significant contact with the rubredoxin domain, covering 31.4% of its surface area (Lee and Richards, 1971).

4.4.3 Mutations in the TPR-Rubredoxin Interface Inhibit Cell Growth in vivo

To enable the in vivo study of mutated LapB proteins in the ΔlapB strain from the Keio collection, wild type LpxC was overexpressed from a low copy number plasmid. The introduction of wild-type LpxC has a dominant effect over the mutated LpxC expressed from the bacterium’s genome (mutation characterized by Mahalakshmi et al., 2014). This results in a sharp decrease in the cell growth rate for at least 8 hours and is consistent with the growth defects observed in a suppressor free ΔlapB strain (Klein et al., 2014) (Figure 4-4B). In addition, cell growth experiments were performed at 42°C to maximize the observed phenotype (Figure 4-5). The increased phenotype for LapB knockouts at 42°C is consistent with LapB’s classification as a heat shock protein (Klein et al., 2014; Nicolaes et al., 2013).
Figure 4-5 Temperature dependence of growth rate shows a greater difference in \( \text{AlapB} \) cell lines at elevated temperatures. (A) 20 °C (B) 37 °C (C) 42 °C. The LpxC control cell line (blue) contains the LpxC plasmid and a control plasmid for tetracyclin resistance. The wild type cell line (red) contains the LpxC plasmid and a second plasmid for expression of wild type LapB.

To restore wild type conditions a second plasmid for the expression of LapB is introduced, which regulates the overexpression of LpxC and restores wild type growth conditions. Endpoint measurements were taken after 4 hours of growth in liquid culture, when both fast and slow growing strains are in optimal mid-log growth. A wide range of mutated \( \text{lapB} \) plasmids were introduced into this model system (Figure 4-4C). LapB mutants were also evaluated using a plating efficiency assay that provided comparable results (Figure 4-6).
Figure 4-6 Plating efficiency assay. Rescue of LpxC overexpression by LapB mutants evaluated using a plating efficiency assay. Each panel is labelled by the mutation in LapB (see Table 4-2 for full strain characteristics). Three dilutions of the culture are plated and the serial dilution is labelled at the top.
Seven mutant strains exhibited less than 90% relative growth due to the inability of LapB to rescue the toxic overexpression of LpxC. A LapB truncation lacking the N-terminal transmembrane helix was unable to rescue cell growth, consistent with previous work that showed membrane localization is essential for LapB function (Nicolaes et al., 2013). Deletion of the entire rubredoxin domain was also shown to prevent cell growth, which confirms an important function for this domain. It has previously been shown that mutation of the four metal coordinating cysteines also causes a loss of function in vivo (Nicolaes et al., 2013). Finally, side-chain mutations including H181A, S378P, W369S/W377S, D244A/E246A/E250A, and D244A/E246A/E250A/E253A were identified as impairing growth and are presumed to cause a loss-of-function in the LapB protein, specifically the ability of rubredoxin to dock to the TPR super-helix. Notably, the growth defect of D244A/E246A/E250A was partially alleviated by additional mutation of L247 to glutamate, suggesting that a minimum of two negative residues are required in the electrostatic patch. Given that loss-of-function mutations were found on both the rubredoxin and the TPR faces of the interaction surfaces we can propose that the specific binding of rubredoxin to the TPR helices is altered and the docked state of rubredoxin plays a role in function.

4.4.4 LapB Mutants are Folded

A wide selection of LapB mutations tested for cell growth in vivo were expressed and purified as soluble domains to assess the effect of the mutations on the structure of LapB. Mutated LapB showed equivalent recombinant expression in E. coli BL21 (DE3) to wild type LapB and could be purified in high yields using the same protocol. Notable exceptions were the W369S/W377S and D244A/E246A/E250A/E253A mutations which had a large propensity to aggregate and precipitate at high concentration. It was determined by circular dichroism (CD) spectroscopy that all single and double mutations do not affect the secondary structure of LapB (Figure 4-7A). The first mutant that showed a spectral change was the quadruple mutation of D244A/E246A/E250A/E253A, which potentially shows a slight loss of α-helical content. The concentration of this mutated protein was significantly decreased and
the apparent loss of secondary structure could also be an artifact of a decreased signal/noise ratio. Given the position of all four of these negative residues in the TPR repeats it is possible that mutating more than three negative residues from this patch starts to disrupt the super helical twist of the protein backbone (Figure 4-4A). Overall, mutation of 5 of these negative positions (D244A/E246A/E250A/E253A/N279A) was required before the protein became fully insoluble.

![Figure 4-7](image-url)

**Figure 4-7 In vitro biophysical studies of LapB’s secondary and tertiary folding.** (A) Circular dichroism spectra of LapB mutations overlay extremely well indicating consistent folding of the TPR α-helices. (B) Intrinsic tryptophan fluorescence of mutant LapB proteins. Purple – Cross-linked protein, peak wavelength of 331. Blue – Mutants that have a peak wavelength lower than 332 nm. Red – Wild type and mutants that have a peak wavelength of 332-333 nm. Green – Mutants that have a peak wavelength of 325 nm.

The rubredoxin deletion truncation was also expressed and purified to confirm that the TPR motifs could fold in the absence of rubredoxin. The CD spectrum of this truncation is consistent with the full-length protein, and although this truncation is more susceptible to degradation, it remains soluble. This
confirms an essential role for the rubredoxin domain in the function of LapB that is independent of the folding of the TPR motifs.

4.4.5 LapB Mutants Alter Local Structure

Given the presence of 3 tryptophan residues (of 4 in LapB) in the TPR-rubredoxin interface, intrinsic tryptophan fluorescence was used to determine the tertiary folding of LapB in solution (Figure 4-7B). Wild type LapB has an intrinsic tryptophan fluorescence peak of 332 nm suggesting that all tryptophan residues are in a hydrophobic environment. In addition, tryptophan fluorescence was observed after LapB was incubated with a cross-linker (glutaraldehyde). If the rubredoxin domain is mobile we would expect the cross-linker to “lock” the protein in the bound state and cause a significant change in fluorescence. Only a small 1 nm shift in fluorescence was observed for cross-linked LapB. Therefore, the tightly bound conformation of rubredoxin observed in the LapB crystal structure also occurs in solution (Figure 4-7B). Loss of function mutants H181A and S378P both cause a shift to a longer wavelength of tryptophan fluorescence consistent with a more hydrophilic tryptophan environment and weaker rubredoxin-TPR binding (Figure 4-7B). H181 is central to the aromatic patch on the TPR face and mutation of this single residue is enough to cause a growth defect. Mutation of both W369 and W377 in the rubredoxin domain was required to produce a similar cell growth defect. S378P also shows a significant fluorescence shift and CD spectroscopy confirms that there is no major loss of protein secondary structure. The mutation to proline more likely causes a local disruption in structure that moves the neighbouring tryptophan residues into less favorable positions for binding.

In contrast, tryptophan fluorescence indicates that mutation of K381A and K381A/R384A cause a shift to a more hydrophobic environment respectively (Figure 4-7B). This would suggest tighter packing of the rubredoxin into the TPR helices. It is possible that nearby R354 could bind the negative region of the TPR motifs resulting in a counter-clockwise rotation of rubredoxin that is consistent with tighter packing at the aromatic patch. Given that H181A and S378P both cause a reduction in cell growth, while
mutations of K381 and R384 do not have an effect, it can be proposed that tight packing of rubredoxin to the TPR motifs is necessary for the function of LapB.

4.4.6 Conclusion

The LapB regulator contains an unprecedented combination of TPR motifs and rubredoxin. Structure determination has revealed two additional TPR motifs that were not predicted due to poorly conserved sequences and an unexpected binding of rubredoxin to the concave face of the TPR super-helix. The rubredoxin domain has an important structural role and its association to the TPR motifs is essential to LPS regulation. The molecular mechanism of LapB’s down-regulation of LpxC remains unclear, but could involve both the scaffold-type binding typical of TPR motifs and the redox activity of the rubredoxin domain. A greater understanding of LPS regulation will aid in the development of future pharmaceuticals targeting Gram-negative bacterial pathogens.

4.5 Methods

4.5.1 Protein Expression and Purification of Soluble LapB

The lapB gene was amplified from E. coli K12 and inserted into a modified pET32a expression plasmid to express a soluble construct of HisTag-Thioredoxin-HisTag-TEV_Site-LapB. Protein was expressed in BL21 cells using auto-induction media (Studier, 2005). Cells were harvested by centrifugation at 3300 × g and re-suspended in 25 mL lysis buffer (50 mM Na₂HPO₄, 300 mM NaCl, 0.1% Triton X-100, 1 mg/mL lysozyme). Cells were lysed by sonication and lysate clarified by centrifugation (21000 × g) was mixed with 3 mL of Ni-agarose resin. Resin was washed and eluted by gravity with 0, 30, 60, 90, and 300 mM imidazole in wash buffer (50 mM Na₂HPO₄, 250 mM NaCl). Protein was cleaved with TEV protease overnight (1:10 protease/protein) at 4° C and run on a superdex 200 size exclusion column (GE). LapB is eluted as a single peak that is well separated from TEV/thioredoxin.
4.5.2 Competitive Metal Binding

Minimal autoinduction media (Studier 2005) with 0, 2, or 10 µM zinc sulphate was used to express protein with controlled trace metal availability. Protein was purified using the standard protocol and concentrated to 350 µM after size-exclusion chromatography. 300 µL of LapB protein was loaded in a 96-well UV-transparent plate and absorbance spectra were collected using a 96-well UV/Vis plate reader. Pictures of LapB in the 96-well plate were taken with a digital camera through the side of the plate.

4.5.3 Crystallization and X-Ray Diffraction

Soluble LapB truncations (5-10 mg/mL) were crystallized by hanging drop vapor diffusion at 20 °C with 0.2 M (NH₄)₂SO₄, 25% PEG 3350, and 0.1 M Bis-Tris pH 6.5 over a period of 3-8 weeks. Individual crystals were harvested after addition of 20-30% glycerol to drops covered with 15 µL mineral oil. X-ray diffraction data was collected at the 23-IDB GMAT beam line at the Advance Photon Source. Eight datasets with 200-360 frames each were collected using a micro-focused beam to collect five datasets from different areas of the first crystal and three datasets from a second crystal.

4.5.4 Data Reduction and Model Building

All datasets were integrated and merged using the XDS package (Kabsch, 2010). Zinc positions were determined using HySS (Grosse-Kunstleve and Adams, 2003) and the merged datasets truncated to 4.0 Å. Phases were extended using Phaser (McCoy et al., 2007). The correct hand was chosen by using RESOLVE density modification (Terwilliger, 2003). The initial model was built using the Autobuild and the Phase_and_build wizards in Phenix (Terwilliger et al., 2007). Model building was performed in Coot (Emsley and Cowtan, 2004). Constraints for the metal binding site were defined using ReadySet and phenix.refine was used for refinement (Afonine et al., 2012). Molecular figures were prepared using PyMOL (Schrödinger, LLC, 2010). Data has been deposited in PDB (4ZLH).
4.5.5 Plasmid Construction

Two plasmids with a lacUV5 promoter, pDL804 and pMS604 (Dmitrova et al., 1998) were provided by Dr. Keith Poole (Queen’s University) and are ampicillin and tetracycline resistant respectively. These vectors were modified to express full-length LapB and LpxC proteins with no additional residues. Single amino acid mutations were made in the LapB gene using Quick-Change site-directed mutagenesis (Stratagene) with 4% DMSO added to the PCR reactions. Multiple mutations were introduced simultaneously using overlap extension PCR (Ho et al., 1989). All mutations were confirmed by DNA sequencing.

4.5.6 Bacterial Strains and Growth Conditions

The ΔlapB E. coli strain (JW1272) was obtained from the Keio collection (Baba et al., 2006). The pDL804 and pMS604 vectors were transformed into the ΔlapB strain to generate the ΔlapB Control, which has resistance to ampicillin and tetracycline. The pDL804-lpxC plasmid was transformed into JW1272 with wild type or mutant LapB plasmids to create experimental strains (Table 4-2). In all cases, picked colonies were grown overnight in 5 mL of LB media supplemented with 100 µg/mL ampicillin and 10 µg/mL tetracycline. The optical density (OD) of each culture at 600 nm was adjusted to 0.5 and 300 µL of cells were used to start fresh 5 mL LB cultures supplemented with both antibiotics and 0.5 mM IPTG. 300 µL of cells was removed to measure OD in a 96 well plate at each time point. Cells were grown at 20 °C, 37 °C, or 42 °C with shaking at 225 rpm. Growth of the wild type strain (CCP104) was used as the reference for 100% cell growth after 4 hours. The remaining LapB mutants were evaluated by calculating their growth relative to this standard as a percent. An additional colony forming assay was performed and the methods/results are in the online supplemental information.

4.5.7 Plating Efficiency Assay

LapB mutant strains were picked from colonies into 4 mL of LB media with 100 µg/mL ampicillin and 10 µg/mL tetracyclin. Cultures were grown at 37 °C for 8 hours with shaking at 225 rpm. The optical density
of each culture was measured and adjusted by dilution to an absorbance of 0.1 with LB media. Serial dilutions were made and 20 µL of the $10^{-2}$, $10^{-4}$, and the $10^{-6}$ dilutions were plated on 3 mL of LB agar supplemented with 100 µg/mL ampicillin, 10 µg/mL tetracyclin, and 0.5 mM IPTG in 12-well cell culture plates. Plates were incubated overnight at 42 °C for colonies to grow.

4.5.1 Tryptophan Fluorescence

Intrinsic tryptophan fluorescence was measured using a Fluorolog Tau-3 Lifetime Fluorimeter (Jobin Yvon Horiba) from a 100 µL sample at 20 µM. Thirty scans were collected between 320 and 350 nm using an excitation wavelength of 295 nm and a band-pass of 1 nm. Curves are presented after averaging and normalization. Cross-linked protein was tested by separating a single 23 µM sample and incubating half the sample with 1% gluteraldehyde overnight.

4.5.2 CD Spectroscopy

Circular dichroism data was collected using a Chirascan spectrometer (Applied Photophysics) and a 200 µL protein sample at a concentration of 20 µM in 20 mM Tris pH 8.0 and 150 mM NaF. The path length of the cell was 0.1 mM and spectra were collected between 182 and 260 nm. Three scans were collected from each sample and averaged. Final spectra are presented after subtraction of an appropriate blank and normalization.
### Table 4-2 Strain construction for *in vivo* study of LapB function (related to “Plasmid Construction” in main experimental procedures).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Descriptive Name</th>
<th>Characteristics</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>JW1272</td>
<td>ΔlapB Strain</td>
<td>BW25113 ΔlapB::Kan</td>
<td>Original strain. Baba <em>et al.</em> 2006</td>
</tr>
<tr>
<td>CCP101</td>
<td>ΔlapB Control</td>
<td>BW25113 ΔlapB::Kan/pDL804, pMS604</td>
<td>Baseline growth in the presence of ampicillin, tetracyclin, and the respective resistance genes.</td>
</tr>
<tr>
<td>CCP102</td>
<td>LapB Control</td>
<td>BW25113 ΔlapB::Kan/pDL804, pMS604-ΔlapB</td>
<td>Baseline growth during overexpression of LapB only.</td>
</tr>
<tr>
<td>CCP103</td>
<td>LpxC Control</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604</td>
<td>Baseline growth during overexpression of wild type LpxC only.</td>
</tr>
<tr>
<td>CCP104</td>
<td>Wild Type</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapB</td>
<td>Baseline growth during overexpression of both wild type LpxC and wild type LapB.</td>
</tr>
<tr>
<td>CCP105</td>
<td>Rubredoxin Deletion</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔ354-389</td>
<td>Removal of the rubredoxin domain in LapB</td>
</tr>
<tr>
<td>CCP108</td>
<td>S378E</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔ378E</td>
<td></td>
</tr>
<tr>
<td>CCP109</td>
<td>S378R</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔ378R</td>
<td></td>
</tr>
<tr>
<td>CCP110</td>
<td>S378P</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔ378P</td>
<td></td>
</tr>
<tr>
<td>CCP111</td>
<td>W369S</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔW369S</td>
<td></td>
</tr>
<tr>
<td>CCP112</td>
<td>W377S</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔW377S</td>
<td></td>
</tr>
<tr>
<td>CCP114</td>
<td>H181A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔH181A</td>
<td></td>
</tr>
<tr>
<td>CCP115</td>
<td>R384A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔR384A</td>
<td></td>
</tr>
<tr>
<td>CCP116</td>
<td>K381A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔK381A</td>
<td></td>
</tr>
<tr>
<td>CCP117</td>
<td>K381A/R384A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔK381A,R384A</td>
<td></td>
</tr>
<tr>
<td>CCP118</td>
<td>D244A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔD244A</td>
<td></td>
</tr>
<tr>
<td>CCP119</td>
<td>E246A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔE246A</td>
<td></td>
</tr>
<tr>
<td>CCP120</td>
<td>L247D</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔL247D</td>
<td></td>
</tr>
<tr>
<td>CCP121</td>
<td>E250A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔE250A</td>
<td></td>
</tr>
<tr>
<td>CCP122</td>
<td>N279A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔN279A</td>
<td></td>
</tr>
<tr>
<td>CCP123</td>
<td>D244A/E246A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔD244A,E246A</td>
<td></td>
</tr>
<tr>
<td>CCP124</td>
<td>D244A/E246A/N279A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔD244A,E246A,N279A</td>
<td></td>
</tr>
<tr>
<td>CCP125</td>
<td>D244A/E246A/E253A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔD244A,E246A,E253A</td>
<td></td>
</tr>
<tr>
<td>CCP126</td>
<td>D244A/E246A/E253A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔD244A,E246A,E253A</td>
<td></td>
</tr>
<tr>
<td>CCP127</td>
<td>D244A/E246A/L247D/E250A</td>
<td>BW25113 ΔlapB::Kan/pDL804-ΔlpxC, pMS604-ΔlapBΔD244A,E246A,L247D,E250A</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5

Zinc-SAD Phasing with Low Redundancy Data

5.1 Foreword

During the structure determination for LapB, protein labeled with selenomethionine was successfully produced with the intention of using Se as a source of anomalous signal for phase determination. The SeMet derivative could not be crystallized for unknown reasons, so we explored the use of the intrinsically bound zinc as a source of anomalous signal to solve the structure of LapB. Originally, we used highly redundant Zn-SAD phasing data to determine the coordinates, but re-analysis of the data revealed that non-redundant data would also have been successful.

5.2 Abstract

The use of specifically bound metal ions for phasing is of growing interest and can be particularly useful when selenomethionine derivatives cannot be produced or do not crystallize. In general, structures determined by Zn-SAD phasing are derived from crystals with high symmetry, and diffraction data collected with high multiplicity. In the current case study, successful Zn-SAD phasing is demonstrated with low multiplicity data collected from crystals grown in a low-symmetry space group. This retrospective analysis suggests that Zn-SAD phasing is more powerful than the literature would currently suggest and confirms that collecting data at the zinc peak is a viable strategy for collecting anomalous data, even if crystals and/or data collection time are limited.
5.3 Introduction

The modern approach to experimental phasing is to collect anomalous scattering data using single wavelength anomalous dispersion (SAD) or multi-wavelength anomalous dispersion (MAD) experiments and a wide range of potential anomalous scatters. These options include selenomethionine labelling (Hendrickson, 1999), soaked halogens (Dauter et al., 2000; Nagem et al., 2001), other soaked ions including zinc (Cha et al., 2012), platinum (Kumar et al., 2013), rubidium (Korolev et al., 2001), cesium (Nagem et al., 2001; Rose et al., 2009), xenon (Evans, 2003; Prangé et al., 1998), and krypton (Prangé et al., 1998), or native sulfur or phosphorus atoms (Dauter and Adamiak, 2001; Debreczeni et al., 2003; Fukakusa et al., 2012; Hendrickson and Teeter, 1981; Kitago et al., 2005; Ramagopal et al., 2003; Watanabe et al., 2005; Yang et al., 2003).

Recently there has been growing interest in the use of natively bound ligand atoms such as zinc, copper, or cobalt as sources of anomalous signal. The use of Zn-SAD phasing is relatively rare and recent publications are summarized in Table 5-1. Anomalous signal from intrinsic metals is most commonly collected using synchrotron source where the K edges are available (Table 5-2). Zinc phasing at the Cu edge using a home X-ray source has also been recently published (Gessmann et al., 2015; Kim et al., 2013). Metal binding proteins are common and 5-10% of all proteins are predicted to bind zinc (Andreini et al., 2006). Additional metal bound proteins are also substituted with zinc during recombinant overexpression including rubredoxin domain proteins (Mahalakshmi et al., 2014; Petros et al., 2006), pseudoazurin (Gessmann et al., 2015), and iron-sulfur cluster proteins (Ramelot et al., 2004). The use of specific metal binding sites means that these sites are fully occupied, but large proteins often have only 1 or 2 metal atoms, and the phasing potential is lower in this case. For selenomethionine, the “rule of thumb” is that 1 anomalous scatterer is required for every 100 amino acids (Boggon and Shapiro, 2000). It is not uncommon to be below this threshold when using native metals. Therefore, high redundancy data is usually collected to solve the phase problem.
Table 5-1 Recently published structures solved with Zn-SAD phasing.

<table>
<thead>
<tr>
<th>Protein</th>
<th>No. of residues</th>
<th>No. of Zn</th>
<th>AA/Zn</th>
<th>Resolution</th>
<th>Multiplicity</th>
<th>Space Group</th>
<th>Wavelength (Å)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose isomerase</td>
<td>388</td>
<td>4</td>
<td>97</td>
<td>1.9 Å</td>
<td>9.0</td>
<td>I222</td>
<td>1.54178</td>
<td>(Kim et al., 2013)</td>
</tr>
<tr>
<td>SaFur</td>
<td>149</td>
<td>3</td>
<td>49</td>
<td>2.6 Å</td>
<td>8.4</td>
<td>P4_3</td>
<td>1.54178</td>
<td>(Kim et al., 2013)</td>
</tr>
<tr>
<td>Pseudoazurin</td>
<td>123</td>
<td>1</td>
<td>123</td>
<td>1.6 Å</td>
<td>7.9</td>
<td>P6_5</td>
<td>1.5418</td>
<td></td>
</tr>
<tr>
<td>BsFur</td>
<td>149</td>
<td>1</td>
<td>149</td>
<td>2.6 Å</td>
<td>20.3</td>
<td>P3_2_1</td>
<td>1.28230</td>
<td>(Lee et al., 2013)</td>
</tr>
<tr>
<td>TON_0340</td>
<td>270</td>
<td>53</td>
<td>5</td>
<td>2.3 Å</td>
<td>12.7</td>
<td>P4_2_2_1</td>
<td>1.2822</td>
<td>(Cha et al., 2012)</td>
</tr>
<tr>
<td>Lysozyme</td>
<td>147</td>
<td>3</td>
<td>49</td>
<td>1.8 Å</td>
<td>13.4</td>
<td>P4_2_2_1</td>
<td>1.2829</td>
<td>(Cha et al., 2012)</td>
</tr>
<tr>
<td>TM0665_HQ-Ala</td>
<td>291</td>
<td>1</td>
<td>291</td>
<td>2.15 Å</td>
<td>13.8</td>
<td>P4_2_2_2_1</td>
<td>1.2815</td>
<td>(Lee et al., 2013)</td>
</tr>
<tr>
<td>AHL Lactonase</td>
<td>250</td>
<td>2</td>
<td>125</td>
<td>2.5 Å</td>
<td>14</td>
<td>P2_1_2_1</td>
<td>1.2827</td>
<td>(Liu et al., 2005)</td>
</tr>
<tr>
<td>VWF CTCK</td>
<td>93</td>
<td>0.5</td>
<td>186</td>
<td>3.28 Å</td>
<td>13.7</td>
<td>P4_3_2</td>
<td>1.2823</td>
<td>(Zhou and Springer, 2014)</td>
</tr>
<tr>
<td>Csm3</td>
<td>351</td>
<td>1</td>
<td>351</td>
<td>2.37 Å</td>
<td>56.3</td>
<td>C222</td>
<td>1.28</td>
<td>(Hrle et al., 2013)</td>
</tr>
<tr>
<td>LapB</td>
<td>339</td>
<td>1</td>
<td>339</td>
<td>2.0 Å</td>
<td>21.2</td>
<td>P2_1_2_1</td>
<td>1.2892</td>
<td>(Prince and Jia, 2015b)</td>
</tr>
</tbody>
</table>

Table 5-2 Structures deposited in the Protein Data Bank with a bound metal ion (as of July 2015) and the K edge of the respective metals that could have been used to collect anomalous signal. The parameters for selenium are provided for comparison.

<table>
<thead>
<tr>
<th>Entries in PDB</th>
<th>K edge</th>
<th>f'</th>
<th>f''</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selenium</td>
<td>0.98</td>
<td>-8.31</td>
<td>3.85</td>
</tr>
<tr>
<td>Zinc</td>
<td>1.28</td>
<td>-6.43</td>
<td>3.89</td>
</tr>
<tr>
<td>Iron</td>
<td>1.74</td>
<td>-6.11</td>
<td>3.94</td>
</tr>
<tr>
<td>Cobalt</td>
<td>1.61</td>
<td>-8.94</td>
<td>3.94</td>
</tr>
<tr>
<td>Nickel</td>
<td>1.49</td>
<td>-8.13</td>
<td>3.92</td>
</tr>
<tr>
<td>Copper</td>
<td>1.38</td>
<td>-7.34</td>
<td>3.90</td>
</tr>
</tbody>
</table>

The structure of LapB, which binds a single zinc atom, was initially determined using high-redundancy Zn-SAD data and the structure and its functional implications have been published (Prince and Jia, 2015b). When protein crystallizes in a low symmetry space groups or there is limited beam time it can be difficult to collect the 10-15 fold redundancy that is commonly seen in the literature. In this retrospective analysis, the intention was to determine the limits of Zn-SAD phasing for the LapB structure, specifically the minimum amount of data required.
5.4 Preparation of Protein Crystals

The LapB protein from *Escherichia coli* is an iron binding protein with a single rubredoxin-type metal binding site (Mahalakshmi et al., 2014). The single transmembrane helix at the N-terminus of the protein was removed and LapB was expressed in the presence of excess zinc to displace the natively bound iron. Zinc-bound LapB is purified and crystallized as previously reported (Prince and Jia, 2015b).

5.5 Data Collection and Data Production in XDS

X-ray diffraction data was collected at the 23-IDB GM/CA beam line at the Advance Photon Source equipped with a CCD detector. A fluorescence scan of the Zn edge confirmed the presence of Zn and provided experimental values for peak absorbance, fʻ, and fʻʻ of 1.28292 Å, -8.88, and 5.45 respectively. Four case studies are considered with data collected from two different crystals (Table 5-3). The original published high-redundancy dataset includes merged data from both crystals and is provided as a reference. The 200 frame dataset was collected with an oscillation angle of 1.0° and is a single pass from the first crystal. The 400 frame merges in a second pass of 200 frames of data collected the same way from a different portion of the same crystal. The 720 frame dataset was collected with an oscillation angle of 0.5° from a second crystal and represents 360° of data collected in a single pass. Finally, the 360 frame dataset is the first half of the previous 720 frame dataset and represents 180° of data.

All datasets were indexed and integrated using the XDS software package (Kabsch, 2010). When needed, data is merged during scaling in XSCALE. R_meas, a merging residual corrected for measurement redundancy has been used instead of R_merge to minimize bias (Diederichs and Karplus, 1997).

5.6 Experimental Phasing

Matthew’s probabilities predict two molecules of LapB in the asymmetric unit (Kantardjieff and Rupp, 2003), so two zinc atoms are expected. Datasets were truncated to 3.5 Å during scaling for
determining the zinc atom positions. This provides a SigAno of 0.9-1.1 in the highest shell (3.6-3.5 Å) and minimizes the amount of data with SigAno<1.0 included. The minimum SigAno to have any signal is 0.8 (Sheldrick, 2010), and ideally one would like it to be greater than 1.0 (Evans, 2011). Experimental phasing was performed using the AutoSol pipeline (Terwilliger et al., 2009), which includes HySS (Grosse-Kunstleve and Adams, 2003), Phaser (McCoy et al., 2007), RESOLVE (Terwilliger, 2003), and a preliminary run of AutoBuild (Terwilliger et al., 2007). A second run of AutoBuild was performed with the same data truncated to 2.1 Å. This run was used to distinguish solutions that provide an appropriate starting point for model building.

Statistics for the input data and the AutoSol output are found in Table 5-3. The AutoSol FOM score is a numerical indicator of the expected accuracy of the phases calculated from the heavy atom substructure. When attempting SAD phasing, scores less than 0.35 are unlikely to be correct, while scores above 0.35 may be correct, and solutions with scores above 0.45 should be correct (Terwilliger, 2015).

5.7 Automated Model Building in Phenix

Ultimately, the ability to build a usable starting model is the most important factor required for solving the structure, and can be evaluated by the number of residues built and the Rwork/Rfree statistics. Initially, all of the correct LapB solutions have reasonable statistics and contain protein fragments that are not correctly positioned to form two complete molecules. Therefore, successful solutions are confirmed by the presence of an appropriate “anchor” for manual model building. In the high-redundancy reference data, the presence of a reasonable zinc binding site was used as the starting point for manual building (Figure 5-1A). This anchor is also present in the 400 frame dataset. In the case of the 720- and 360-frame datasets from crystal #2, a zinc binding site is not evident, but this protein contains an alternative anchor that also provides high confidence. The sequence His-Phe-Tyr-Cys contains four distinct side chains that are unlikely to be misplaced in density at a resolution of 2 Å. From a solid anchor point, the complete
structure can be manually built, and fragments can be re-positioned based on symmetry to build two complete molecules of LapB (Figure 5-1C). The only set of low-redundancy data that does not provide an adequate starting point for manual building was the 200-frame dataset from crystal #1, which still has an uninterpretable model after a full run of AutoBuild.

Figure 5-1 Anchors for successful manual model building. (A) Snapshot of a probable zinc bind site in the initial model generated by AutoBuild for the datasets with 2400 and 400 frames. Two of the cysteine residues are present and space for two additional cysteine residues is present in the unmodelled density. (B) Four unique residues including three ring side-chains and a cysteine define an alternative anchor present in all of the successful initial models. (C) Complete model of LapB.
Table 5-3 Summary of data integration, phasing, and initial model statistics with images of the initial models. Yellow: Previously published high-redundancy dataset. Purple: Data from crystal #1 collected at 1° oscillation. Blue: Data from crystal #2 collected at 0.5° oscillation.

<table>
<thead>
<tr>
<th></th>
<th>Amount of Data</th>
<th>Model #1</th>
<th>Model #2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2400 Frames</td>
<td>200 Frames</td>
<td>400 Frames</td>
</tr>
<tr>
<td></td>
<td>2040 Degrees</td>
<td>200 Degrees</td>
<td>400 Degrees</td>
</tr>
<tr>
<td>Multiplicity (3.6-3.5)</td>
<td>21.2 (21.2)</td>
<td>2.2 (2.2)</td>
<td>4.2 (4.2)</td>
</tr>
<tr>
<td>Completion (3.6-3.5)</td>
<td>100.0 % (100.0 %)</td>
<td>93.9 % (92.8 %)</td>
<td>98.6 % (98.7 %)</td>
</tr>
<tr>
<td>R_meas (3.6-3.5)</td>
<td>5.2 % (6.6 %)</td>
<td>2.8 % (3.5 %)</td>
<td>3.2 % (4.3 %)</td>
</tr>
<tr>
<td>SigAno (3.6-3.5)</td>
<td>2.155 (1.398)</td>
<td>1.491 (1.130)</td>
<td>1.651 (1.219)</td>
</tr>
<tr>
<td>I/Sigma (3.6-3.5)</td>
<td>57.67 (44.77)</td>
<td>33.55 (26.07)</td>
<td>40.19 (30.27)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.511</td>
<td>0.380</td>
</tr>
<tr>
<td>Multiplicity (5.4-5.3)</td>
<td></td>
<td>0.40</td>
<td>0.41</td>
</tr>
<tr>
<td>Completion (5.4-5.3)</td>
<td></td>
<td>59 %</td>
<td>49 %</td>
</tr>
<tr>
<td>R_work</td>
<td>0.47</td>
<td>0.51</td>
<td>0.48</td>
</tr>
</tbody>
</table>
5.8 Discussion

Interest in native metal phasing is likely to increase as improvements in the accuracy and speed of data collection improve the measurement of extremely small anomalous differences (González, 2003; Gorynia et al., 2006; Westermeier et al., 2009). Given that LapB is one of the largest proteins whose structure has been solved with Zn-SAD and P2₁ is a relatively low-symmetry space group, this was a good case study to determine the minimum data required for Zn-SAD phasing.

Four low redundancy datasets were considered which contained small amounts of data collected from two different crystals. There is clearly a difference in the quality of the data collected from the two crystals with the first crystal requiring measurement of 400° of data in two passes, while a single pass of 180° of the second crystal is sufficient to provide a solution. It should be noted that data from the first crystal was collected at an oscillation angle of 1.0° compared to 0.5° for the second crystal which may have impacted the data quality independent of the crystals.

In all solutions, fragments of the LapB structure are modelled in the unit cell and a large amount of manual building is required to re-arrange the fragments and complete the structure. This is true for both the high-redundancy and low-redundancy data and therefore is an independent property of this data that does not reduce the validity of the solutions obtained at low-redundancy.

There will always be debate regarding the choice of SAD or MAD data collection strategies, and both have their advantages. MAD phasing can be more powerful because both anomalous and dispersive differences can be included in the calculations and has been used numerous times with zinc (Meyer et al., 2006; Mitsuhashi, 2000; Qiu and Janson, 2004; Reiland et al., 2004; Strop et al., 2001; Sugahara et al., 2000). The main drawback of MAD phasing is the necessity to collect data at multiple wavelengths, but two-wavelength MAD may be a viable strategy for reducing the data collection time required (González, 2003). SAD has the advantage that it requires a simpler and faster data collection strategy that reduces the radiation damage of the crystal, which is important when crystals are especially radiation sensitive (Liu et
al., 2011). In the case of LapB, high-resolution diffraction was initially collected from four different crystals, but only 2 of the crystals produced diffraction with interpretable anomalous signal. Therefore, it was more time efficient to collect SAD data from all crystals available than longer MAD datasets on fewer samples.

5.9 Conclusions

When appropriate, Zn-SAD phasing is a viable substitute for SeMet labeling, even when large amounts of redundancy cannot be collected. We have demonstrated that the LapB structure could have been solved with only 2-fold multiplicity, or 180° of data in space group P2₁. Therefore, the use of intrinsic binding of a single zinc atom for SAD phasing is now a viable strategy given current tools, even in low-symmetry space groups with limited multiplicity.
Chapter 6

Discussion

6.1 Characterization of Proteins in the Post-Genomics Era

The structural genomics boom of the early 2000’s has provided an abundance of structures, but the data obtained from this massively high-throughput approach is skewed towards small, soluble, and well-behaved proteins. Many of the proteins in the exopolysaccharide synthesis and export pathways have at least one representative structure, but “missing links” in the high resolution data remain. A disproportionate number of the remaining unique proteins are membrane proteins, which present technical challenges. These proteins are not amenable to current high-throughput strategies and technical improvements for working with these challenging targets are required.

6.2 Predicting Protein Structure

With the large amount of structural data that has been accumulated by structural genomics initiatives, it was hoped that the structure of any protein could be predicted using sequence homology to map structure onto new sequences. Unfortunately, there are still sequences with few or no homologues that are poorly modeled by servers such as Phyre2 (Kelley et al., 2015). This is especially true for membrane proteins, where there is a smaller starting pool from which to draw structures. The value of experimentally determining new structures is evident in both the projects presented in this thesis.

6.2.1 Phyre2 Prediction of LapB Structure

TPR motifs are very common and have been crystallized in many different proteins, which lead to Phyre2 identifying 120 potential models for LapB with greater than 99% confidence. These predictions contain the conserved helix-turn-helix motifs in a vast number of arrangements (Figure 6-1).
The consensus model that Phyre2 predicts and the crystal structure are displayed in Figure 6-2. The rubredoxin domain is modeled in intensive mode as a series of loops in a mostly independent domain. This positioning is consistent with previous observations of isolated rubredoxin proteins (Bönisch et al., 2005; Chen et al., 2006; Schweimer et al., 2000) and rubredoxin-like domains (DeMaré et al., 1996; Iyer et al., 2005). In contrast, the LapB crystal structure revealed an intimate relationship between the TPR helices and rubredoxin, which is critical for function. Overall, the LapB model produced by Phyre2 shares little similarity with the crystal structure beyond the presence of TPR helices. Therefore, despite the large array of TPR proteins with available crystal structures, experimental structure determination is still valuable.
6.2.2 Phyre2 Prediction of Etk Periplasmic Domain Structure.

Phyre2 was also used to predict the structure of Etk’s periplasmic domain. In this case, the server was unable to identify a good template for modeling and instead used intensive mode to build loops for a large amount of the protein backbone, resulting in an uninterpretable model (Figure 6-3A). The closest protein family with known structure is the PCP-1 family of protein co-polymerases. In both Etk’s N-terminal domain and the PCP-1 proteins there is a mixed alpha helical/beta sheet domain that is followed by a large coiled-coil that accounts for about half of the domain (Figure 6-3). The major difference between Etk and the PCP-1 proteins is that the PCP-1 domains are significantly smaller, although the proportions of the secondary structure elements appear similar. For example, there are two beta strands in the PCP-1 proteins that form the edge of a sheet (Figure 6-3B-Black). One of these is the first secondary structure element when the protein enters the periplasm, and the second is the last. These sheets are clearly seen in all structures of the PCP-1 proteins and are also predicted to occur in Etk. Therefore, Etk is likely a “larger” version of the PCP protein structures that are already available. Unfortunately, automated servers
like Phyre2 and BLAST are not able to pick-up the relationship between these proteins, likely because of the differences in size.

Figure 6-3 Predicted structure of Etk’s N-terminal periplasmic domain. (A) Phyre2 structure prediction based on the protein sequence. (B) Known structure of WzzE, a protein from the related PCP-1 protein family. The first and last beta strand in this structure (discussed in text) are highlighted in black. (C) Secondary structure prediction based on the sequence of Etk’s N-terminal domain calculated by XtalPred (Slabinski et al., 2007).
6.3 Challenges in Structural Biology

6.3.1 Spontaneous Proteolysis in Protein Biochemistry

The degradation of proteins in vitro is generally considered a nuisance. Degradation is presumed to be caused by trace contamination of proteases from the expression system carried through into purified samples. Therefore, it is common practice to add protease inhibitors to some, or all stages of protein purification. For the full-length Etk membrane protein, both phenylmethanesulfonyl fluoride and the broader protease inhibitor tablet formulations were tested for their ability to control the degradation of Etk full-length, but neither was successful. In the case of LapB, the degradation of the purified protein is not immediately apparent and only occurs after a long period of storage at 4 °C or 20 °C. By the time the degradation of LapB was noted as a potential problem, crystals had already been obtained and likely grew from the degradation product. Therefore, no attempt was made to prevent the degradation of the protein using inhibitors.

Characterization of the protein degradation products can provide valuable insight into the properties of the protein in vitro. The degradation products for Etk and LapB were analyzed by MALDI-mass spectroscopy after tryptic fragmentation. This narrowed down the degradation point in both samples. This data was one of the main drivers behind the design of Etk NM2, the crystallizable truncation of Etk. The removal of a large soluble domain from a membrane protein is counter-intuitive based on the current membrane protein literature, which promotes the addition of fusion partners and antibodies to promote crystallization (Cherezov et al., 2007; Chun et al., 2012; Johnson et al., 2015; Rasmussen et al., 2011; Rosenbaum et al., 2007). A major driver behind the decision to remove the kinase domain was characterization of the protein degradation product and it is unlikely that this strategy would have been pursued otherwise.
Mass-spectroscopy also showed that LapB was suffering N-terminal degradation to approximately residues 65-68. Although the truncations at residues 65-68 were not able to form plate crystals, an intermediate truncation that was made at the same time, starting at residue 50, was surprisingly effective at improving the resolution of the crystals obtained. Limited proteolysis is a known technique for finding more crystallizable truncations of recalcitrant proteins (Dong et al., 2007; Wernimont and Edwards, 2009). In the case of LapB, not only did the degradation need to occur for crystals to form, it was also advantageous to re-clone the protein construct to avoid the drop heterogeneity caused by the removed fragment. Small amounts of protein degradation in vitro can be a powerful ally in the determination of novel structures. When stable degradation products are identified, efforts should be made to understand the nature of the degradation before determining whether it should be eliminated.

6.3.2 Membrane Proteins and the Detergent Conundrum

Fundamentally, the principles of membrane protein expression, purification, and crystallization are similar to the well-established methods optimized for soluble proteins. Practically, membrane protein researchers do encounter far more challenges. These challenges can begin during expression, with yields for membrane protein expression being notoriously low. Some strategies for improving protein insertion into the membrane have been developed (Eshaghi et al., 2005; Lee et al., 2015) or avoid this problem entirely using cell-free systems (Henrich et al., 2015). The next big hurdle is the solubilization and purification of membrane proteins in vitro, which currently requires the use of surfactants. In addition, membrane proteins for crystallography should be stabilized in detergents that are chemically distinct to provide a homogeneous environment for crystallization. The selection of optimal detergents is an important issue (Arnold and Linke, 2008; Linke, 2009; Sanders et al., 2004; Seddon et al., 2004). During the study of Etk NM2, it was also found that the concentration of detergent in crystallization samples was equally important and that crystallization could only be reproduced when both the protein and detergent were kept within a defined concentration range. In order to achieve this range, a classic phenol/sulfuric...
acid reaction for the quantitation of sugars was optimized so that it would be applicable to the small volume protein samples encountered in crystallography. This assay is applicable to any sugar-based detergent and therefore had broad applications for membrane protein crystallographers.

### 6.3.3 Difficult Crystals

#### 6.3.3.1 Etk NM2 Prisms

Although the reproducibility of Etk NM2 crystals was significantly improved, their diffraction was still limited to ~ 7 Å at synchrotron radiation sources. Enough data was collected to determine that the lattice symmetry of the crystal was hexagonal and that the unit cell dimensions were 113.9 Å, 113.9 Å, 583 Å, 90°, 90°, 120°. Extremely long unit cells are a common challenge in membrane protein crystallography that has to be overcome. The modest diffraction of membrane protein crystals in general could also be due to a relatively high solvent content of 75-80%, high B-factors, and limited protein-protein contacts in the crystal lattice (Rees et al., 2000). This is due to the presence of detergent, which takes up a significant portion of space in the crystal lattice. Detergent has the fluidic properties of a solvent and cannot make productive crystal contacts (Hitscherich et al., 2000; Prive, 2007; Sonoda et al., 2010). The unit cell content of Etk NM2 crystals is predicted to have 9 copies of the protein when there is 50 % solvent, but given that it is a membrane protein, it would not be surprising to find that the hexagonal crystals contain a hexamer of Etk NM2 and 67 % solvent (Kantardjieff and Rupp, 2003; Weichenberger and Rupp, 2014). Therefore, the Etk NM2 crystals produced in this study suffered from multiple fundamental problems, which ultimately limited diffraction and it was not possible to solve the structure. These are common problems and many membrane protein crystals are difficult to optimize beyond a resolution of 5-6 Å (Carpenter et al., 2008).
6.3.3.2 LapB Plate Clusters

The initial crystallization hit for LapB appeared after 2 months and took 5 months to reach the full size pictured in Chapter 4. During expansion and optimization, the growth time for these crystals was reduced to 3-8 weeks. Given the long growth times, it was suspected that the protein was degrading prior to crystallization, which is supported by visible degradation on SDS-PAGE when the protein is stored at 4 °C for the same length of time. Crystals were initially grown with the LapB protein truncated at residue 19 and diffracted to a maximum resolution of ~3 Å. Attempts to produce crystals with SeMet labeled protein were not successful and the quality of the diffraction was not sufficient to solve the structure using zinc or sulfur as a source of anomalous signal. After re-cloning of the protein to express a truncation starting at residue 50, crystals could be grown with the same appearance and behavior, but now diffracted to ~2 Å. Not only was the resolution improved, the quality of the diffraction was also improved, which allowed the measurement of accurate anomalous signal for solving the crystallographic phase problem (Chapter 5). The use of weak-signal phasing experiments is highly dependent on the availability of high quality data, but can often play a crucial role in the determination of structures not amenable to SeMet strategies. Due to the protein degradation, no new information could be obtained by using the 2 Å data for molecular replacement with the previously obtained 3 Å data (from crystals grown from drops set with the longer 19-389 residue construct).

6.4 Determining Biological Assemblies

6.4.1 Debate Over the Correct Oligomeric State for Etk

Another particular challenge is the determination of native oligomerization states for membrane proteins. Given that membrane proteins are most commonly purified and stabilized in an artificial detergent environment, there is always concern that the detergent is breaking up native oligomers into smaller non-functional units (Prive, 2007). This possibility appears to be true for Etk, where the full-length protein purifies in a range of oligomeric states that is primarily monomer, dimer, tetramer, and
octamer by analytical ultracentrifugation. When we look at the same data for the truncated protein, Etk NM2, this truncation is a mixture of monomer and dimer when solubilized in DDM, but is primarily monomeric in DM detergent. This result is consistent with the detergent artificially breaking up a native oligomer with the shorter-chain detergent having a more detrimental effect.

Therefore, my studies of Etk in vitro offer little insight into the continued debate over the oligomerization state of this protein. Currently, the most likely candidates for the biological unit are a tetramer or an octamer which have been observed in electron microscopy and crystallography experiments on the homologous Wzc protein (Bechet et al., 2010; Collins et al., 2006). In addition, the channel protein, Wza, that interacts with Wzc has a crystal structure with an octameric conformation, which further supports the octamer argument (Dong et al., 2006). Minimally, a high-resolution structure of Etk’s periplasmic domain that reveals a plausible mode of interaction between the kinase and the channel will be needed to resolve this debate. Ultimately, it may not be settled until high-resolution data for the complex between these proteins becomes available and the mechanism of action is elucidated.

### 6.4.2 Conflicting Evidence for the Oligomerization State of LapB

The crystal structure of LapB shows a dimer in the asymmetric unit. PDBePISA, a prediction server for biological assemblies (Krissinel and Henrick, 2007), also predicts a dimer based on the crystal structure. The predicted biological assembly is an unusually shaped ring dimer, with the two molecules of LapB in the plane and the rubredoxin domains in the center. Unfortunately, looking closer at this dimer, it appears that the crystal has not captured a biologically relevant oligomer. Both of the molecules in the ring dimer are truncations of the full-length protein, one starting at amino acid 52 and the second starting at 67. This difference in truncation is equivalent to one α-helix or half of a TPR repeat unit. When this missing helix is added to the second molecule (based on alignment with the larger molecule) these two helices occupy the same three-dimensional space (Figure 6-4). There also should be a further additional helix on both molecules to form the full first TPR repeat, which would be expected to continue in the
same direction as the previous knob-and-hole packing in the TPR super-helix (Andrea et al., 2003), continuing the conflict.

Figure 6-4 The additional helix in one molecule of the LapB asymmetric unit clashes with where the same helix in the other molecule would be if it was present. Upper left – LapB asymmetric unit as solved. Upper right – predicted ring dimer. Lower right – larger LapB molecule copied into the same place as the smaller molecule. Lower left – rotated view of the ring dimer showing that having the longer truncation in both molecules is incompatible.

The oligomerization state of LapB in solution was also investigated to determine if there was supporting evidence for a dimer. Truncated LapB in solution is consistently a tetramer as determined by size-exclusion chromatography, analytical ultracentrifugation, and SAXS (Figure 6-5). Modeling of four crystallographic monomers in the calculated SAXS envelope generates a range of similar structures, but none of these options appears more likely to be the oligomerization state in vivo. The precise modelling of the LapB tetramer in solution is hindered by the lack of features in the SAXS envelope. In order for a modelled tetramer in solution to be considered physiological, it would be expected that all four of the N-
termini would be pointing in the same direction since this protein is embedded in the inner membrane by a transmembrane helix attached to each unit. Given that a LapB dimer is not supported by multiple solution-based techniques, it is unlikely that the crystal structure represent a physiologically relevant oligomer. At this point I cannot propose a likely oligomer for the in vivo biological unit of LapB.

Figure 6-5 Oligomerization state of LapB in solution is a tetramer. (A) Size exclusion column using a Superdex200 column. Protein elutes as a single peak between 60 and 70 mL, which is consistent with an oligomer in the range of trimer – hexamer. Smaller trailing peak is the thioredoxin expression tag that is cleaved prior to chromatography. (B) Analytical ultracentrifugation runs with each truncation confirm a single peak with molecular mass (in brackets) consistent with a tetramer. (C) SAXS envelope for LapB residues 50-389 in solution. (D-E) Potential tetramers modelled based on the SAXS curve by Coral. Each modelling run produces a slightly different arrangement.
6.5 Regulation of Exopolysaccharide Production and Transport in Gram-Negative Pathogens

The important role of gram-negative exopolysaccharides in virulence is well known, and the full mechanism for the synthesis, assembly, and transport of these sugars is currently being investigated. These pathways are also highly regulated and the components of these regulons are also essential to the cell. This thesis looked at the structural components behind some of these regulatory devices.

6.5.1 Etk in a Bi-membrane Complex

Protein phosphorylation on tyrosine, long known to be an essential component of eukaryotic signaling pathways, has only recently gained recognition as an important mechanism in prokaryotes. The first example of a tyrosine kinase to be cloned and characterized was Ptk from Actinobacter johnsonii in the mid-1990s (Duclos et al., 1996; Grangeasse et al., 1997). It was quickly followed by two homologous proteins from E. coli, Etk and Wzc (Drummelsmith and Whitfield, 1999; Ilan et al., 1999; Vincent et al., 1999). These enzymes belong to a family of proteins called bacterial tyrosine (BY) kinases and are the most thoroughly characterized prokaryotic tyrosine kinases, both structurally and functionally (Lee et al., 2008). Wzc and Etk are capable of both autophosphorylation and phosphorylation of specific substrates in the cytoplasm (Obadia et al., 2007). The best-studied function of these proteins is their role in polysaccharide synthesis and export. In 1999, it was established that the exopolysaccharides exported by Etk and its homologues can play an important role in virulence (Ilan et al., 1999).

The recent structures determined for the Etk and Wzc kinase domains allow us to start to understand the mechanism of protein phosphorylation. These studies have revealed both the two-stage activation mechanism in the active site (Lu et al., 2009) and the importance of the C-terminal tyrosine cluster (Paiment et al., 2002). It has been suggested that the level of phosphorylation on the C-terminal tail is the crucial signal that results in a change in the oligomerization state of the inner membrane kinase domain. Specifically, it has been suggested that the unphosphorylated protein forms a tight ring structure, which
dissociates into monomers after phosphorylation of the C-terminal cluster (Olivares-Illana et al., 2008). Both the structure of the closed, isolated outer membrane channel and a low-resolution structure of the open channel in the bi-membrane complex are available (Beis et al., 2004; Collins et al., 2007). Based on this foundation, it is possible that Etk and the interaction between these two proteins is responsible for the regulated opening and closing of the outer membrane channel. In this model, the autophosphorylation of the kinase domain, and the resulting dissociation of the octamer would lead to a conformational change in the entire protein that is transmitted to the periplasmic domain. At this time, the state (phosphorylated or unphosphorylated) that is associated with the open channel is not clear. The phosphorylation state of the inner membrane kinase was heterogeneous in the EM structure of the Wzc-Wza complex, which is the only time that the channel has been observed open (Collins et al., 2007). It has been shown that cycling between phosphorylated and unphosphorylated forms of the inner membrane kinase is required (Hagelueken et al., 2009a), consistent with the concept that it would be undesirable to leave the outer membrane channel open indefinitely.

The outer membrane channel proteins contain a polysaccharide export sequence (PES) that comprises the smallest, inner-most domain of the channel protein. At this time, the exact function of this domain is unknown, but it is conceivable that it plays some role in the interaction of the channel protein with the regulatory kinase (Collins et al., 2007). Within the two inner-most rings, there is an outer-most helix that could provide a docking site for the kinase. In the periplasmic domain of Etk there is a large coiled-coil structure that is likely in an extended confirmation, based on the related PCP-1 family. Therefore, it is possible that the extended coiled-coil within Etk interacts with the exterior helix of the PES to form the kinase-channel complex (Figure 6-6). This model would need extensive validation, and will likely only be truly elucidated once a structure of the kinase’s periplasmic domain is determined, if not a complex structure between the two. To avoid attempting an unprecedented bi-membrane protein complex structure, truncation of the periplasmic domains to create a fully soluble complex was considered. It has already
been shown that the transmembrane domain of the outer membrane channel protein can be removed without impacting the folding of the periplasmic domain (Hagelueken et al., 2009b), and that the periplasmic domain of Etk can be produced in copious quantities. Unfortunately, attempts to perform pull-down experiments were unsuccessful. Therefore, a simple high-throughput approach to probe this interface with amino acid mutations is not currently available.

The structural gap between the kinase domain and the outer membrane channel remains (Figure 6-6). In Chapter 2, a thorough attempt was made to characterize membrane bound constructs of the Etk periplasmic domain to identify a crystallizable construct, which included the missing domain. Despite successful crystallization and improvements in the reproducibly of membrane protein crystallography experiments (Chapter 3), the diffraction of these crystals remains too weak for structure determination.
6.5.2 LapB as a Protein Scaffold

Enzymes have a defined function and importance to the cell for catalyzing chemical reactions, but they cannot always function independently in the cytoplasm. The regulation of enzyme activity can often be achieved by controlling the concentration of the enzyme or directly modifying the enzyme to make it more or less effective. Another important factor can be the localization of the enzyme within the cell to the area where the substrates or products are present, produced, or required.

The LapB protein appears to be an essential scaffold in gram-negative bacteria and the ability of LapB to localize to the inner membrane is crucial for the down-regulation of its target enzyme (LpxC)
(Nicolaes et al., 2013). It has already been shown that LapB is pulled down with the FtsH protease responsible for LpxC degradation (Klein et al., 2014), but how LapB is involved in targeting LpxC to this protease remains unknown. ZDOCK is an online protein docking server (Pierce et al., 2011) to which LapB and LpxC were submitted in order to predict potential binding modes (Figure 6-7). ZDOCK predicts that LpxC would bind to the concave face of the TPR helix, consistent with previous complex structures between TPR proteins and their binding partners (Fodor et al., 2015; Pal et al., 2014; Wang et al., 2011), but the lack of consistency in the docks generated could indicate that the predictions are not accurate. Attempts to confirm the direct binding between LpxC and soluble constructs of LapB in vitro using pull-down experiments have thus far been unsuccessful. Thus far, there is no evidence confirming that LapB and LpxC interact directly.

Figure 6-7 Predicted binding sites for LpxC on the LapB protein. LapB is shown in red and the potential binding modes of LpxC are in blue, cyan, and purple.

The question of how LapB targets LpxC to FtsH remains unknown. It is possible that there are additional proteins involved that have not been characterized which leads to many possible scenarios. For example, there could be an unknown adaptor protein that mediates binding between LpxC and LapB. Alternatively, LapB could be responsible for binding an enzyme and localizing it to the inner membrane where it can alter LpxC and mark it for degradation. LpxC could also be modified by an uncharacterized enzyme in the cytoplasm before it can bind LapB. There is currently no evidence to suggest that any of
these possibilities are more likely than the other options listed or additional scenarios that have not yet been considered.

The LapB structure does reveal a crucial and unexpected relationship between the TPR-protein scaffolding motifs and the rubredoxin metal binding domain. The rubredoxin-type metal binding domain is associated with the concave face of the TPR helices both in solution and in the crystal. Sequence alignments of the rubredoxin domain reveal four very poorly conserved locations that potentially evolved specifically to enable rubredoxin-TPR binding in LapB. When two of these locations (the two tryptophan residues) were simultaneously mutated, a loss of LapB function was observed, which was associated with a lack of binding between rubredoxin and the TPR. TPR helices are commonly involved in a wide range of interactions including complexes with small-molecules, peptides, and proteins (Fodor et al., 2015; Pal et al., 2014; Wang et al., 2011). In all of the available structures, the binding partner is located on the concave face of the TPR helical twist, similar to the position of rubredoxin. This also suggests that additional proteins binding to the TPR scaffold would neighbor the rubredoxin and both rubredoxin and the TPR helices could be part of the binding site.

Within the structure of LapB there is available space for other proteins to interact with TPR helices #1-4 or 7-9. Within these potential concave grooves, there are clusters of electrostatic patches, which may play a role in potential binding sites (Figure 6-8). In order to probe this hypothesis further, substitutions would need to be performed that eliminate and/or reverse the charges in these patches. This would provide a new series of mutations that could be screened in vivo for their ability to regulate the overexpression of LpxC.
Figure 6-8 Electrostatic model of the LapB surface. Blue indicates positively charged regions and red indicates negatively charged regions. Specific patches that might be involved in docking are indicated with purple arrows. Green circles mark the position of rubredoxin in these surface models.

6.6 Applying the Study of Cell Envelope Regulation: Antibiotic Research

Within healthcare settings, gram-negative bacteria are the causative agent of pneumonias, bloodstream infections, wound or surgical site infections, and meningitis. These infections are caused by virulent strains of *Klebsiella, Acinetobacter, Pseudomonas aeruginosa, Escherichia coli,* and many other less common species. The emergence of multi-drug resistant bacteria is of particular concern in a hospital setting where they can infect vulnerable patient groups, prolong recovery periods, and in extreme cases result in avoidable fatalities (Fraimow and Tsigrelis, 2011). Multi-drug resistant infectious bacteria can result in increased medical costs, longer hospital stays, and increased hospital mortality (Arnold et al., 2007). Gram-negative pathogens also have serious consequences in agriculture by infecting and destroying crop plants. An example is *Erwinia amylovora,* which is the causative agent of fire blight, a plant disease that can be devastating to commercial apple and pear production (Piqué et al., 2015).

The original classes of antibiotics that were discovered were found in nature by studying bacteria that produced these compounds to give themselves a competitive advantage in their habitat (Silver, 2011).
Therefore, the evolution of antibiotics was also paralleled by the evolution of antibiotic resistance genes that these original bacterial manufacturers used to protect themselves. The widespread use of antibiotics by humans has resulted in a general development of even more resistance genes. Bacterial pathogens have the ability to pass resistance genes horizontally on small DNA plasmids which accelerates the dissemination of these properties (Taylor, 2012). The horizontal transfer of antibiotic resistance is a major contributor to the evolution of multi-drug-resistant strains that have acquired multiple resistance genes. There is now a resistance gene in circulation for every known antibiotic, increasing the risk of strains with no viable treatments (Payne et al., 2007). The vast majority of antibiotics available are natural products or derivatives of natural products. Unfortunately, the space for developing new derivative compounds has now been extensively explored and novel antibiotic development on entirely new targets is required (Silver, 2011). In order for these new antibiotics to be successful they must have high potency and the ability to evade the numerous innate immunity mechanisms present in gram-negative bacteria.

6.6.1 Potential Antibiotic Targets in Exopolysaccharide Regulons

6.6.1.1 LapB in the Lipopolysaccharide Pathway

In general, the LPS synthesis enzymes have no mammalian counterparts which makes them promising targets for the development of novel antimicrobials (Whitfield and Trent, 2014). Recently, there has been great interest in targeting LpxC, the first committed step in LPS synthesis, with inhibitors. There are currently more crystal structures of LpxC-inhibitor complexes deposited in the protein databank than for any other protein from the synthesis pathway. Theoretically, proteins involved in the regulation of LpxC would also be valid targets for antimicrobial development.

This thesis supports the idea that LapB functions primarily as a scaffolding protein, and although its function is crucial, these proteins are extremely difficult to target with small-molecule pharmaceuticals. Since the most important feature of scaffolding proteins is the protein-protein binding interfaces, they are notoriously difficult to “drug” because of the large and open nature of the binding sites. It is difficult to
find small molecules that have thermodynamically favorable binding properties for these open sites. Most small molecules target smaller, enclosed binding sites where there is a greater enthalpic gain from binding and less of an entropic loss.

A different approach would be to target proteins that LapB binds. LapB can be produced as a soluble protein, which could make the characterization of protein-protein interactions much easier (i.e. you avoid the need for detergent to stabilize LapB). Several potential binding partners for LapB have already been identified by pull-down assays and many of these proteins are enzymes. Considering the enormous impact that LapB has on LpxC, it is possible that some of its other interactions are equally important. Therefore, by continuing to characterize LapB and its interacting partners, alternative enzymes involved in the regulation of LPS synthesis and transport that are also promising candidates for antibiotics development could be identified.

**6.6.1.2 E. coli Tyrosine Kinase**

Like the LPS synthesis proteins, the bacterial tyrosine kinases are also highly conserved and do not have eukaryotic counterparts. The structure of Etk/Wzc is significantly different than the conserved fold of eukaryotic kinases (Lee et al., 2008). Given that eukaryotic kinases have been successfully targeted by structure-based drug design multiple times (Cui et al., 2011; Grimminger et al., 2010; Lou et al., 2014), it is possible that the unique active site in bacterial tyrosine kinases may also be an appropriate target. Given the lack of homology between eukaryotic and prokaryotic kinases, the potential for unwanted side effects would be minimized by this strategy. The kinase domain of Etk is likely a better candidate for pharmaceutical design than the periplasmic domain, since it contains a defined, recessed binding pocket (Anderson, 2003). There is also academic interest in the design of small molecules against protein-protein interfaces, but it is practically more challenging due to the large, shallow, interfaces involved. Research continues to best understand how to best attack these challenging surfaces (Nero et al., 2014).
6.7 Summary and Future Perspective

This thesis describes my attempts to determine the structure of Etk and the successful structure determination for LapB. Multiple issues were encountered while attempting to crystallize Etk that lead to an in-depth look at protein preparation and crystallization techniques. Ultimately, truncation of Etk was required for successful crystallization, despite the removal of a large solvent exposed “crystallizable” domain (the kinase domain). This strategy is the opposite of what the literature currently recommends for membrane proteins. Reproducing Etk NM2 crystals was initially extremely difficult. Therefore, a fast, quantitative method for the measurement of detergent in protein samples was developed to predict protein samples suitable for crystallization based on both the protein and detergent concentration they contained. This assay could be widely applicable to all membrane protein researchers using sugar-based detergents for solubilization.

The second membrane protein studied, LapB, could be easily manipulated by truncating the N-terminal transmembrane helix to leave the bulk of the protein as a soluble protein. A crystal structure was successfully determined for this protein that revealed two common protein folds; TPR repeats and a rubredoxin metal binding domain. These motifs are intimately bound in an unexpected conformation that is essential to the function of this regulator of LPS biosynthesis in the cytoplasm.

Initial attempts to develop optimum inhibitors for LpxC were hindered by a lack of structural information (Raetz and Whitfield, 2002), and now that the structure is available, there has been an explosion of newly optimized inhibitors and protein-inhibitor complex structures. Therefore, a complete structural model of the LPS and CPS synthesis and export pathways would be advantageous for the development of novel small molecules that prevent the production of these essential extracellular sugars. Many of the unique enzymes in the LPS and CPS production pathways that currently have no structure available are membrane embedded or associated, which makes their structural characterization more challenging. In the long term, membrane proteins in the LPS and CPS pathways with functional domains
open to the periplasm or the extracellular space are more likely to make ideal pharmaceutical targets, because antimicrobial compounds would have fewer barriers to cross before they can bind to these proteins. This observation does not preclude the development of agents that target intercellular enzymes, but in this case, the antibiotics have to cross both membranes independently, or be co-administered with a membrane permeabilizer. Therefore, determining structures for the remaining components of these pathways will aid in our understanding of the exopolysaccharide production pathways and potentially reveal new targets for intervention.
Appendix A

Structural Characterization of *Legionella* Effector Proteins

A.1 Foreword

This appendix describes a structural characterization project that focused on *Legionella pneumophilia* effector proteins. This project was started near the end of my second year of graduate studies, as a “back-up” when the more difficult membrane protein targets were proving to be quite challenging. An initial screening of the effector proteins launched multiple avenues of research, but these small projects remain incomplete or have been passed on to other graduate students for completion. This appendix summarizes the initial screening campaign including the criteria for target selection, initial screening outcomes, and scale-up of selected targets. This is followed by a summary of the current progress for open targets.

A.2 Introduction

Specific gram-negative bacteria have a complex system of host-pathogen interactions that are mediated by secreted effector proteins. While most common species, such as *Escherichia coli*, have only a handful of effector proteins, specialized pathogens can have significantly more. An extreme example is *Legionella pneumophilia*, a gram-negative bacterium normally found in unicellular protozoan hosts. This bacterium is the causative agent of Legionnaire’s disease, a lung infection acquired by inhaling contaminated water that has been vaporized (i.e. from a shower, drinking fountain, or sprinkler) (Albert-Weissenberger et al., 2007).

*Legionella* contains over 300 identified effector proteins that allow the bacteria to be engulfed by and then reprogram a host cell to form a *Legionella* containing vacuole (LCV) that promotes the survival and replication of the bacteria using host-cell machinery. These effector proteins are secreted into the host cell by the Dot/Icm type IV secretion system, a multi-protein complex that spans both bacterial membranes (Hilbi et al., 2001). The signal sequence for translocation by the Dot/Icm secretion system is expected to
involve E block motifs, which contain groups of negatively charged residues that are found in the C-terminus of the target proteins (Huang et al., 2011). Effector proteins have commonly been identified by direct gene fusion assays, which attach the C-terminal 100 amino acids of potential effector proteins to a carrier protein. Protein fragments which are able to successfully enable secretion of the carrier protein are deemed to have come from effector proteins (Huang et al., 2011; Isberg et al., 2009).

**A.3 Selection and Cloning of Targets for Structural Characterization**

A list of *Legionella* effector proteins was pre-compiled for a grant application and was available to the research group. From this list of over 300 options, proteins were picked for cloning, expression, and purification trials. Proteins were chosen based on several criteria including their size, lack of transmembrane helices, and the crystallizability score calculated by XtalPred (Slabinski et al., 2007). Smaller, compact proteins are known to have a greater success rate in crystallization trials and therefore proteins larger than 500 amino acids were avoided. The difficulties associated with membrane proteins are also well known, and these targets are not amenable to high throughput approaches. Finally, most proteins chosen have XtalPred scores of 3 or lower, which represent average to optimal scores. The XtalPred score is a rank that predicts proteins likely to crystallize based on their similarity to proteins that have already been solved. The initial target list can be found in Table A-1. Proteins selected for screening were cloned into vectors with an N-terminal His-tag or GST-tag, both of which contained a TEV protease cleavage site between the protein and the expression/purification tag. Cloning was performed by Dr. Miroslaw Cygler’s research group.

**A.4 Expression and Solubility Trials**

Given the large number of initial constructs to be screened, samples were processed in a high throughput manner. Plasmids were transformed into *E. coli* BL21(DE3) expression cells and grown in 5 mL of LB media overnight. Cells were diluted (200 µL in 5 mL) in auto-induction media (Studier, 2005)
and grown at 37 °C for 3 hours followed by 20 °C for 24 hours to express protein. 100 µL of cell culture was removed and stored. The remaining culture was centrifuged at 4000 rpm for 10 minutes to pellet the cells. The media was decanted and the pellet re-suspended in 1 mL of lysis buffer (50 mM NaH₂PO₄ pH 7.5, 300 mM NaCl, 1 mg/mL lysozyme, 0.1 % Triton X-100). Re-suspended cells were pulsed with a sonicator 1-3 times to lyse cells. Lysate was centrifuged at maximum speed in a microcentrifuge at 4 °C for 30 min to pellet insoluble material. The supernatant was then transferred to a new tube and the pellet was re-suspended in 8 M urea. An SDS-PAGE gel was run to compare the samples. 10 µL aliquots of 2 × SDS-PAGE dye were mixed with 10 µL cell culture, 10 µL soluble fraction, and 3.5 µL pellet/urea sample. Examples of typical results for soluble, insoluble, and non-expressed proteins are shown in Figure A-1. The full classification of all of the constructs tested can be found in Table A-1.

![Figure A-1 Example SDS-PAGE gels for *Legionella* small scale expression and purification trials. The no expression example is His-MavJ, the soluble example is His-RavL, and the insoluble example is His-RavE. Blue boxes surround protein bands of the correct size for the target protein. T = total cell culture, S = supernatant, P = pellet.](image-url)
Table A-1 Initial target list assembled for Legionella effector proteins. Initial expression and solubility results are shown under the His-Tag and GST-tag columns for the two constructs that were tested for most targets. High = high protein expansion, Low = low protein expression, Not Det = protein expression not detected. Sol = soluble, InSol = insoluble.

<table>
<thead>
<tr>
<th>Name</th>
<th>Length</th>
<th>Size (kDa)</th>
<th>pI</th>
<th>Xtal Pred</th>
<th>His-Tag</th>
<th>GST-TAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>RavL</td>
<td>293</td>
<td>33</td>
<td>5.65</td>
<td>1</td>
<td>High Sol.</td>
<td>High Sol.</td>
</tr>
<tr>
<td>RavG</td>
<td>207</td>
<td>23</td>
<td>5.65</td>
<td>1</td>
<td>Low Sol.</td>
<td>Low Sol.</td>
</tr>
<tr>
<td>LirD</td>
<td>436</td>
<td>50</td>
<td>6.49</td>
<td>2</td>
<td>Low Sol.</td>
<td>Low Sol.</td>
</tr>
<tr>
<td>RavQ</td>
<td>360</td>
<td>41</td>
<td>9.19</td>
<td>4</td>
<td>Low Sol.</td>
<td>Low Sol.</td>
</tr>
<tr>
<td>MavG</td>
<td>448</td>
<td>51</td>
<td>5.29</td>
<td>4</td>
<td>Low Sol.</td>
<td>Low Sol.</td>
</tr>
<tr>
<td>Ceg29</td>
<td>272</td>
<td>32</td>
<td>8.72</td>
<td>4</td>
<td>Low Sol.</td>
<td>Low Sol.</td>
</tr>
<tr>
<td>MavC</td>
<td>482</td>
<td>55</td>
<td>5.22</td>
<td>4</td>
<td>Low Sol.</td>
<td>n/a</td>
</tr>
<tr>
<td>LegL2</td>
<td>428</td>
<td>48</td>
<td>6.03</td>
<td>1</td>
<td>Low Sol.</td>
<td>Not Det.</td>
</tr>
<tr>
<td>LegLC4</td>
<td>335</td>
<td>37</td>
<td>6.61</td>
<td>2</td>
<td>Low Sol.</td>
<td>Not Det.</td>
</tr>
<tr>
<td>LegP</td>
<td>226</td>
<td>30</td>
<td>5.72</td>
<td>3</td>
<td>Low Sol.</td>
<td>Not Det.</td>
</tr>
<tr>
<td>LegU2</td>
<td>246</td>
<td>28</td>
<td>8.33</td>
<td>4</td>
<td>Not Det.</td>
<td>High Sol.</td>
</tr>
<tr>
<td>RavK</td>
<td>246</td>
<td>28</td>
<td>6.18</td>
<td>1</td>
<td>High Insol.</td>
<td>High Insol.</td>
</tr>
<tr>
<td>RavC</td>
<td>268</td>
<td>31</td>
<td>6.17</td>
<td>2</td>
<td>High Insol.</td>
<td>High Insol.</td>
</tr>
<tr>
<td>RavE</td>
<td>334</td>
<td>38</td>
<td>5.92</td>
<td>2</td>
<td>High Insol.</td>
<td>High Insol.</td>
</tr>
<tr>
<td>Ceg3</td>
<td>255</td>
<td>29</td>
<td>6.25</td>
<td>2</td>
<td>High Insol.</td>
<td>High Insol.</td>
</tr>
<tr>
<td>VipF</td>
<td>286</td>
<td>33</td>
<td>7.11</td>
<td>2</td>
<td>High Insol.</td>
<td>Not Det.</td>
</tr>
<tr>
<td>Sde</td>
<td>301</td>
<td>34</td>
<td>5.56</td>
<td>3</td>
<td>High Insol.</td>
<td>High Insol.</td>
</tr>
<tr>
<td>MavI</td>
<td>204</td>
<td>23</td>
<td>4.61</td>
<td>3</td>
<td>High Insol.</td>
<td>n/a</td>
</tr>
<tr>
<td>RavW</td>
<td>295</td>
<td>33</td>
<td>6.98</td>
<td>4</td>
<td>High Insol.</td>
<td>High Insol.</td>
</tr>
<tr>
<td>MavD</td>
<td>210</td>
<td>25</td>
<td>7.01</td>
<td>4</td>
<td>High Insol.</td>
<td>High Insol.</td>
</tr>
<tr>
<td>MavK</td>
<td>444</td>
<td>52</td>
<td>5.39</td>
<td>4</td>
<td>Low Insol.</td>
<td>Not Det.</td>
</tr>
<tr>
<td>LegL1</td>
<td>296</td>
<td>32</td>
<td>5.26</td>
<td>4</td>
<td>Low Insol.</td>
<td>Not Det.</td>
</tr>
<tr>
<td>RavP</td>
<td>318</td>
<td>36</td>
<td>5.77</td>
<td>4</td>
<td>Low Insol.</td>
<td>Not Det.</td>
</tr>
<tr>
<td>RavN</td>
<td>212</td>
<td>24</td>
<td>4.98</td>
<td>4</td>
<td>Not Det.</td>
<td>High Insol.</td>
</tr>
<tr>
<td>Ceg11</td>
<td>224</td>
<td>26</td>
<td>5.88</td>
<td>4</td>
<td>n/a</td>
<td>High Insol.</td>
</tr>
<tr>
<td>Ceg18</td>
<td>243</td>
<td>55</td>
<td>5.35</td>
<td>4</td>
<td>Not Det.</td>
<td>Not Det.</td>
</tr>
<tr>
<td>LegL3</td>
<td>493</td>
<td>56</td>
<td>5.93</td>
<td>5</td>
<td>n/a</td>
<td>Low Insol.</td>
</tr>
</tbody>
</table>
A.5 Large Scale Investigations of Soluble Targets

The targets shaded in green in Table A-1 had a construct which produced soluble protein and were all scaled up by either myself, my undergraduate students, or other researchers within the lab. An exhaustive summary of all the large scale trials performed would be inefficient; therefore, I have grouped these proteins based on their ultimate outcome (Table A-2).

Table A-2 Outcome of large scale protein expression, purification, and crystallization trials.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystallized – Structure Determined</td>
<td>Lem23 (completed by Alexander Andrew)</td>
</tr>
<tr>
<td>NMR Data</td>
<td>PieF</td>
</tr>
<tr>
<td>Crystallized</td>
<td>RavL</td>
</tr>
<tr>
<td>Crystallized (Competitors completed</td>
<td>LegU2, MavC</td>
</tr>
<tr>
<td>structure first)</td>
<td></td>
</tr>
<tr>
<td>No Crystals / Degrades</td>
<td>RavD, LegP, LegLC4, LirD, MavL, MavJ, RavJ</td>
</tr>
<tr>
<td>Impure / No Crystal Screening</td>
<td>LegL2, RavQ, MavG</td>
</tr>
<tr>
<td>Precipitates / No Crystal Screening</td>
<td>RavG</td>
</tr>
<tr>
<td>No Large Scale Expression</td>
<td>Ceg29, Lem2</td>
</tr>
</tbody>
</table>

The following sections contain brief highlights of the initial expression, purification, and characterization data from two of the most successful targets, RavL and PieF.

A.6 RavL: A Putative Lipase with an Alpha-Beta Hydrolase Fold

A.6.1 Protein Expression and Purification

The RavL protein contains 293 amino acids and XtalPred predicts extensive secondary structure in the central portion of the sequence. Based on the secondary structure prediction, the first construct of RavL that was studied was a truncation from residues 18 – 260 which removes the predicted disorder at both ends of the protein sequence (Figure A-2).
Figure A-2 Secondary structure prediction for RavL assembled by XtalPred.

Large scale expression and purification of RavL was performed with *E. coli* grown in auto-induction media on a 500 mL scale, which yielded a minimum of 10 mg of protein. Overexpressed His-TEV-RavL (18-260) can be readily purified by nickel affinity chromatography using standard protocols. Surprisingly, after TEV cleavage to remove the His-Tag, RavL retains affinity for nickel resin (Figure A-3).

Figure A-3 Nickel purification of His-tagged RavL. (A) Initial nickel-affinity column eluted by gravity with washes of increasing imidazole concentration. (B) After TEV cleavage, the non-tagged RavL still binds to the nickel column, indicative of an intrinsic His-tag. (C) Size exclusion chromatography reveals that RavL is a monomer in solution.
A.6.2 Biophysical Characterization of RavL

Initial crystallization screening attempts were performed with RavL 18-260 without success, so biophysical characterization was carried out to gain a better understanding of this protein’s properties in vitro. Crystallization screening of RavL was carried out at room temperature. Circular dichroism (CD) spectroscopy and differential scanning calorimetry (DSC) were used to determine if the protein structure is stable at this temperatures (Figure A-4). The melting point of RavL is well above room temperature at 54 °C and the deconvolution of the secondary structure components from CD analysis at room temperature is consistent with predictions from multiple computational models, including PSIPRED (Buchan et al., 2013), SOPMA (Geourjon and Deléage, 1995), and Phyre2 (Kelley et al., 2015) (Table A-3).

Figure A-4 RavL stability. (A) Circular dichroism spectra for RavL. Protein was tested alone, in the presence of 38 µM EDTA and with 38 µM CaCl₂. In all three cases, the resulting spectra are the same. (B) Differential scanning calorimetry from 20 – 75 °C with a peak at 54.03 ± 0.02 °C.
Table A-3 Breakdown of secondary structure components in RavL determined by CD spectroscopy and compared to multiple computational prediction algorithms.

<table>
<thead>
<tr>
<th>Features</th>
<th>CD</th>
<th>SOPMA</th>
<th>PSIPRED</th>
<th>Phyre2</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Helix</td>
<td>33%</td>
<td>32%</td>
<td>31%</td>
<td>33%</td>
</tr>
<tr>
<td>β-Strand</td>
<td>15%</td>
<td>23%</td>
<td>14%</td>
<td>15%</td>
</tr>
<tr>
<td>β-Turn</td>
<td>21%</td>
<td>9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disorder</td>
<td>30%</td>
<td>36%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### A.6.3 Modeling of RavL with Phyre2

In order to design better constructs for crystallization, an *in silico* model of RavL was constructed using the Phyre2 protein structure prediction server (Kelley et al., 2015). This server uses homology to proteins of known structure to predict the structure of a provided protein sequence (Figure A-5). In the case of RavL, the structure predicted has similarity to an alpha-beta hydrolase, which is a large family of proteins that catalyze a range of reactions. The key feature of this family is a central parallel or mixed β-sheet that is surrounded by α-helices (Nardini and Dijkstra, 1999). RavL bears the most similarity to lipase enzymes from the alpha-beta hydrolase family, which are responsible for the hydrolysis of ester bonds in triglycerides (Nardini et al., 2000). Despite this similarity, the Phyre2 model of RavL contains only one of the three expected residues in the catalytic triad, with both the aspartate and the histidine absent from their conserved positions.

Given the intrinsic binding of RavL to nickel resin, it is also expected that there might be a clustering of histidine residues on the protein surface. There are two groups of three histidine residues near the C-terminus of the protein, but these are smaller groups than would normally be expected to grant affinity alone (Figure A-6). Although only two histidine residues are required to coordinate one nickel ion, it is usually expected that multiple nickel ions will be coordinated by a single His-tag to create a strong affinity. Given that the binding of RavL to a nickel column is weaker than normally observed for an engineered His-tag with 6 residues, it is possible that a smaller cluster is responsible for RavL’s affinity. At this point, the exact residues responsible for nickel affinity remain unclear.
Figure A-5 Models of RavL. (A) Example of a lipase protein structure for the alpha-beta hydrolase family (PDB: 1CLV). (B) Alignment of three lipase active sites demonstrating the arrangement of the conserved catalytic residues (PDB: 1CLV, 1EX9, 4GW3). (C) Phyre2 prediction of the RavL structure. Portions truncated by the original construct are highlighted in blue. The presumed catalytic serine is highlighted in red sticks.

Figure A-6 Intrinsic nickel binding of RavL. (A) Predicted model of RavL with the positions of the clustered histidine residues in the C-terminus highlighted with pink sticks. (B) Binding mode of a His-Tag to nickel-NTA resin.
A.6.4 Enzymatic Activity of RavL

Lipase enzymes commonly bind calcium which is required for their function. Therefore, the ability of RavL to hydrolyze triglycerides in the presence of calcium was determined. A proof of concept experiment using a commercially available kit confirmed the ability of RavL to perform a lipase reaction and does suggest that the reaction could be calcium dependent (Figure A-7). At minimum, RavL requires the presence of a cation, as indicated by the lack of activity in the presence of the metal chelator EDTA.

Figure A-7 Calcium dependent kinase activity of RavL measured using a commercial kit acquired from Sigma-Aldrich. This assay uses an enzyme coupled reaction to produce a colored product proportional to the hydrolysis of a proprietary mix of triglycerides.

RavL’s substrate preference was evaluated using a pH indicator assay that uses phenol red to track the release of free fatty acids from triglyceride substrates in the presence of calcium. This assay was based on a published protocol (Biver and Vandenbol, 2013) and indicates that the tributyrate substrate was more easily hydrolyzed by RavL (Figure A-8). Additional properties of this enzyme that could be investigated include other cations that can be used by RavL and the temperature dependence of the reaction.
Figure A-8 Substrate preference of RavL evaluated using phenol red as a pH indicator assay to track the release of fatty acids from triglyceride substrates after incubation at 37 °C for 16 hours. Volumes on the left indicate the amount of RavL enzyme added to each well. When required, protein buffer was added to bring the total added volume to 15 µL. Reactions are displayed in triplicate.

A.6.5 Re-Engineering RavL for Crystallization

The current Phyre2 model of RavL is unable to answer multiple questions about the protein generated during the *in vitro* characterization including the identity of the catalytic residues in the active site and the location of the intrinsic His-tag. A crystallographic structure of RavL could potentially address some of these issues. What the Phyre2 model does demonstrate is that secondary structure elements may have been removed by the initial protein truncation, which are indicated in blue (Figure A-5C). Therefore, although the initial truncation appears stable at room temperature it may not be adopting the ideal fold. Based on the knowledge gained thus far, three new tag-free constructs were generated, which all started at residue 18 and ended at residues 260, 269, and 293, for crystallization.

These tag-free constructs subsequently confirmed that RavL has intrinsic affinity for nickel resin allowing untagged protein to be purified directly from the lysate. In this case, the binding of RavL to the column is weaker, with protein starting to elute at 30 mM imidazole, and almost fully eluted at 90 mM
imidazole. Therefore, extensive washing of the column with wash buffer (with no imidazole) was performed and the slightly dirty RavL collected from the 60 - 90 mM imidazole fractions was subsequently further purified by size-exclusion chromatography (Figure A-9).

![Figure A-9 Example SDS-PAGE gels from purification of the RavL 18-369 no tag construct. Desired band is surrounded by a blue box.](image)

**A.6.6 Crystallization of RavL**

All three truncations of RavL have been extensively screened for crystallization, but only the truncation from residues 18-269 is able to successfully grow crystals. These initial crystals grow over a period of 10 days in ~ 0.2 M KCl and 18-26 % PEG 3350 (Figure A-10A). I have nicknamed this form of RavL crystals “cryo-phobic,” because after screening a range of cryo-protectants, a condition in which these crystals could be frozen and retain diffraction was not found. Diffraction of RavL crystals at room temperature confirms that these crystals contain protein and could diffract to sufficient resolution for molecular replacement (Figure A-11A). Unfortunately, without cryo-protection, the crystals are highly...
prone to radiation damage and only 5 images can be collected at room temperature before diffraction is completely lost.

Figure A-10 RavL crystals (A) Typical crystals grown from 18-26% PEG 3350 and 0.2 M KCl pictured under white light. (B) Typical crystals grown from 18-26% PEG 2000DME and 0.2 M KCl pictured under polarized light.

Figure A-11 Representative diffraction images (A) Original condition from home source to ~3.5 Å collected at room temperature. (B) Optimized crystals frozen in 25 % glycerol at synchrotron, ~ 7 Å.
In order to find alternative crystallization conditions for RavL that might produce more robust crystals, a manual PEG screen was performed, in which 5-30% of all available PEGs were tested with 0.2 M KCl for their ability to produce RavL crystals. This provided an alternative condition, with PEG 2000 DME. The crystals from the new condition appear to be over-nucleating which could be contributing to the multiple layers that can be seen in the larger crystals (Figure A-10B). In addition, during looping these crystals are highly prone to cracking indicating that they do not tolerate dehydration. After cryo-protection in glycerol or mineral oil, diffraction patterns to 6-7 Å resolution can be obtained at a synchrotron which confirms that the crystals are protein, but the patterns are smudged (Figure A-11B). Multiple factors could be contributing to the poor quality of the data including, multiple lattices in the crystal originating from the multiple layers apparent under the microscope, cracking of the crystal during looping, and non-optimal cryo-protection during freezing.

To start addressing these issues, attempts are being made to further optimize crystals using seeding. Crystal trays are also being set with HEPES as the protein buffer instead of Tris which would allow crystal cross-linking in situ which may make the crystals more tolerant to dehydration (Lusty, 1999). Finally, when enough high quality single crystals are available the cryo-protectant can be optimized.

**A.7 PieF, a Tiny Protein of Unknown Structure and Function**

**A.7.1 Protein Expression and Purification**

PieF is a tiny effector protein from *Legionella* that contains only 125 amino acids. This protein has no homology to other proteins outside of the *Legionella* genus. PieF comes from a region of genome plasticity, yet is well conserved within *Legionella*, suggesting that this protein could plays a significant functional role (Ninio et al., 2009). Currently, there are no other published clues to the *in vivo* function of PieF. Secondary structure prediction for PieF indicates that this protein may form a small compact folding domain with multiple secondary structure elements (Figure A-12).
Figure A-12 Secondary structure prediction for PieF assembled by XtalPred.

The PieF protein was originally expressed and purified using an N-His-TEV-PieF construct that can be purified by nickel-affinity chromatography. Given the small size of this protein, it was decided to pursue structural characterization of PieF by NMR. The His-tag and the linker/TEV site expressed with PieF contain 23 amino acids total, which is a significant addition to such a small protein. Therefore, attempts were made to cleave the His-tag, which failed. Even with a TEV/fusion protein ratio of 1:1 overnight, the tag cannot be removed. Therefore, this unstructured peptide was also present in the NMR sample.

To optimize the conditions for NMR data collection, three experiments were performed. To start, dynamic light scattering (DLS) measurements of PieF purified in a buffer of 20 mM Tris pH 8, 150 mM NaCl and stored at 4 °C provided a hydrodynamic diameter of 2.7 nm. This result contrasts with the same sample stored at room temperature, which aggregates to a particle size of 27 nm. Finally, if 10 % glycerol was added to the protein, the DLS reading after storage at room temperature was 3.0-3.5 nm, indicating that glycerol is improving the long-term oligomerization state of PieF. In the second experiment, a full solubility screen was performed on PieF samples mixed 1:1 with the test buffer solution (0.5 µL / 0.5 µL) in a vapor diffusion plate and stored overnight at room temperature. Drops are then observed to determine conditions in which the protein does not precipitate. Overall trends indicate that this protein is more stable at basic pH and lower salt concentrations and confirm that glycerol can improve solubility (Table A-4).
Table A-4  Full solubility screen of PieF after overnight incubation. H = heavy precipitation, L = low precipitation, clear = soluble.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>pH</th>
<th>ALONE</th>
<th>300mM NaCl</th>
<th>5% Glycerol</th>
<th>Salt+Glycerol</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 mM Sodium Acetate</td>
<td>4.6</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>100 mM Potassium Acetate</td>
<td>5</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>100 mM MES</td>
<td>5.5</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>100 mM MES</td>
<td>6</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>100 mM Bis-Tris</td>
<td>6.5</td>
<td>L</td>
<td></td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>100 mM Potassium Phosphate</td>
<td>6.5</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>100 mM ADA</td>
<td>6.5</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>100 mM PIPES</td>
<td>7</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>H</td>
</tr>
<tr>
<td>100 mM Imidazole</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>100 mM MOPS</td>
<td>7</td>
<td>H</td>
<td></td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>100 mM HEPES</td>
<td>7.5</td>
<td>L</td>
<td>H</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>100 mM Potassium Phosphate</td>
<td>7.5</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>100 mM Tris</td>
<td>7.5</td>
<td>L</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>100 mM Tricine</td>
<td>8</td>
<td></td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM HEPES</td>
<td>8</td>
<td>L</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM Bicine</td>
<td>8</td>
<td>H</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM Tris</td>
<td>8</td>
<td></td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM Tris</td>
<td>8.5</td>
<td></td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM Bis-Tris</td>
<td>8.5</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM TAPS</td>
<td>8.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM CAPSO</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM CHES</td>
<td>9</td>
<td></td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM Glycine</td>
<td>9.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 mM CAPS</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Nitrogen-based spectra in NMR are ideally collected at acidic pH (6.5 or below), which is not possible for PieF. Therefore, data was collected at pH 7.0, which was the lowest pH that could be achieved. The final buffer optimization was performed by comparing NHSQC spectra collected from N15-labelled protein with various additives. In these experiments, it was found that 20 mM DTT (high concentration reducing agent) was providing the best spectra and the presence of DTT removed the requirement for glycerol. Therefore, N-His-TEV-PieF data was collected in a buffer of 20 mM HEPES pH 7.0, 200 mM NaCl, and 20 mM DTT with the generous help of Dave Langelaan from the Smith Lab at Queen’s University.
A.7.2 NMR Data Analysis

The NHSQC NMR spectrum of PieF is shown in Figure A-13 with the assigned backbone peaks labeled. Given the higher than ideal pH used for data collection, some backbone N-H peaks are not visible due to the high rate of exchange between the amide proton and the solvent. Many of the missing peaks belong to the His-TEV tag that is expected to be flexible, and therefore would be subject to a higher rate of exchange. Assignment of all atom resonances has been attempted, and there are currently assignments determined for 86.5 % of the protein residues.

Figure A-13 NHSQC NMR spectrum collected from N-His-TEV-PieF.

Thus far, shift assignments have relied on through bond experiments, which specifically couple atoms connected by bonds within the protein. These assignments have been performed manually using the
ccpNMR Analysis package (Skinner et al., 2015). Automated methods for the assignment of through-space peaks in the NNOESY and CNOESY spectra was attempted, but has been hindered by low signal to noise. Initial models of PieF have been constructed based on the automated assignments using the ARIA software package (example: Figure A-13) (Linge et al., 2003). A large portion of the protein is still modeled as poorly defined loops, and thus far a converged ensemble has not been achieved. In order to proceed with this data, manual assignment of the through-space data will be required.

Figure A-14 A very preliminary model of PieF. The majority of the structure is still modelled by loops due to an insufficient number of unambiguous restraints.

At this point, we can extract the secondary structure components of PieF. These are predicted from experimental measurements of the backbone chemical shifts by the program DANGLE (Figure A-15) (Cheung et al., 2010). A comparison of the experimentally determined secondary structure components with the computationally predicted values indicates that PieF is less structured in the C-terminus than expected and that a predicted helix in the N-terminus appears to actually be a beta-strand. A more flexible C-terminus might be sensible for this protein given the presence of an E-block (EPEE) that could be involved in recognition by the Dot/Icm secretion system (Huang et al., 2011). Overall, the NMR structure needs to be completed before there will be a complete understanding of this protein’s structure.
Figure A-15 Comparison of the secondary structure predicted by XtalPred with the secondary structure calculated from the backbone chemical shifts measured by NMR.
References


