GENETIC AND PHENOTYPIC CHARACTERIZATION OF NOVEL BORDETELLA PERTUSSIS LIPID A MODIFICATIONS

by

Nita Reva Shah

B.Sc., The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Microbiology and Immunology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2014

Abstract

Lipopolysaccharide (LPS) is a component of the outer membrane in most Gram-negative bacteria. The lipid A region of LPS anchors the molecule to the outer membrane and forms the first barrier between Gram-negative bacteria and the extracellular environment. Lipid A is also important for bacterial interaction and activation of host immune cells through binding to Toll-like receptor 4 (TLR4), activation of which results in a downstream inflammation response. The work presented in this thesis explores the genetic basis for different lipid A structures and the effects of this structural variability on activation of TLR4 and resistance to cationic antimicrobial peptides (CAMPs). Penta-acyl lipid A from *B. pertussis* strain BP338 is modified with glucosamine (GlcN), and mutational analyses revealed that LgmA, LgmB, and LgmC are required for this modification. Bioinformatic analysis suggests the following hypothetical model: LgmA transfers *N*-acetylglucosamine (GlcNAc) from UDP-GlcNAc to the carrier lipid C55P, LgmC removes the acetyl group, and LgmB transfers GlcN from C55P to the lipid A phosphate. Glycosyltransferase assays with LgmA-expressing *E. coli* membranes show that LgmA transfers GlcNAc onto a lipid, therefore supporting the first step in this model. Site-directed mutagenesis has also identified a putative active site in LgmA and LgmC. Killing assays show the GlcN modification in *B. pertussis* increases resistance to numerous CAMPs and to outer membrane perturbation by EDTA. Lipid A from *B. pertussis* strain 18-323 exhibits low levels of TLR4 activation. Lipid A of strain 18-323 has no GlcN modification (due to an incomplete *lgm* locus) and a shorter C3' acyl chain compared to strain BP338 (due to a difference in LpxA). Complementation of 18-323 with BP338 lipid A-modifying genes *lpxA* and/or the *lgm* locus increase TLR4 activation, though the GlcN modification had a dominant effect. In hexa-acyl *E. coli* lipid A, shortening the C3 and C3' acyl chains had been shown to decrease TLR4 activation and resistance to polymyxin B, but increase activation of the limulus amebocyte lysate assay. Thus, varying the structure of lipid A, in both *B. pertussis* and *E. coli*, can affect TLR4 activation and resistance of the bacteria to CAMPs.

Preface

I designed, performed, and analyzed the data from all experiments detailed in this thesis, with the following exceptions:

- Mass spectrometry analysis and structural determination of all lipid A structures and isolation of LPS from the strains 18-323, 18-323 + pPtacLgmABCD, 18-323 + pPtacLpxA338, and 18-323 + pPtacLgmABCDLpxA338 was performed in the laboratory of Dr. Martine Caroff (Université de Paris-Sud, Orsay, France) by Sami Albitar-Nehme and Alexey Novikov.
- Whole genome sequencing of *Bordetella hinzii* strain ATCC 51730 and *Bordetella trematum* strain CCUG 13902 and assembly of the draft genomes was performed in the laboratory of Dr. Martin Hirst (University of British Columbia, Vancouver, Canada) by Michelle Moksa.
- I supervised Emma Kim in the cloning of the vectors: pET30LgmA and pPtacLgmAB, and in the analysis of the *B. pertussis* 18-323 sequence raw reads for the presence of the *lgm* locus genes.
- I supervised Andrew Low in the cloning of vectors: pBBR2LgmA D76G D77G, pBBR2LgmA D127G, pBBR2LgmA D129G, pBBR2LgmA D127G D129G, pBBR2LgmA D159N, pBBR2LgmC D80G D81G, pBBR2LgmC H130G, pBBR2LgmC D187G D189G, pBBR2LgmC E313G; and in the experiments shown in Figure 22 and Figure 25.

A version of Sections 3.2.1, 4.3, and 4.4 have been published. **Shah, N. R.**, S. Albitar-Nehme, E. Kim, N. Marr, A. Novikov, M. Caroff, and R. C. Fernandez. 2013. Minor modifications to the phosphate groups and the C3' acyl chain length of lipid A in two *Bordetella pertussis* strains, BP338 and 18-323, independently affect Toll-like receptor 4 protein activation. The Journal of Biological Chemistry 288:11751-11760.

A version of Section 3.6 has been published. Novikov, A., **N. R. Shah**, S. Albitar-Nehme, S. M. Basheer, I. Trento, A. Tirsoaga, M. Moksa, M. Hirst, M. B. Perry, A. E. Hamidi, R. C. Fernandez, and M. Caroff.

2013. Complete *Bordetella avium*, *Bordetella hinzii* and *Bordetella trematum* lipid A structures and genomic sequence analyses of the loci involved in their modifications. Innate Immunity: 10.1177/1753425913506950. I completed the genomic analysis to support the lipid A structures and wrote these sections of the paper.

The original draft genome sequences in Section 3.6 have been published. **Shah, N. R.**, M. Moksa, A. Novikov, M. B. Perry, M. Hirst, M. Caroff, and R. C. Fernandez. 2013. Draft genome sequences of *Bordetella hinzii* and *Bordetella trematum*. Genome announcements 1:e00838-13.

A version of Section 4.2 has been published. **Shah, N. R.**, R. E. W. Hancock, R. C. Fernandez. *Bordetella pertussis* lipid A glucosamine modification confers resistance to cationic antimicrobial peptides and increases resistance to outer membrane perturbation. Antimicrobial Agents and Chemotherapy 58(8):4931-4

Biohazard Approval Certificate numbers under which this research was conducted are: H06-0225, H06-0226, B06-0225, and B06-0226 issued by the Biosafety Committee, Office of Research Services at the University of British Columbia.

Radioisotope Approval Certificate number under which this research was conducted is MICB-3399-18 issued by the University of British Columbia Committee on Radioisotopes and Radiation Hazards.

Table of Contents

Abstract ..... **ii**

Preface ..... **iii**

Table of Contents ..... **v**

List of Tables ..... **x**

List of Figures ..... **xi**

List of Abbreviations ..... **xiv**

Acknowledgements ..... **xviii**

Dedication ..... **xx**

Chapter 1: Introduction **1**

1.1 *Bordetella pertussis* and whooping cough 1

1.1.1 Clinical manifestation 1

1.1.2 Molecular pathogenesis 2

1.1.2.1 BvgAS regulation system 2

1.1.2.2 Adhesins 3

1.1.2.3 Toxins 3

1.1.2.4 Other virulence factors 4

1.1.2.5 Immune evasion and modulation 4

1.1.3 Other *Bordetella* species 5

1.2 Lipopolysaccharide 6

1.2.1 LPS structure 8

1.2.1.1 Lipid A 10

1.2.1.2 Core sugars 10

1.2.1.3 O-antigen 10

1.2.2 LPS Biogenesis 11
1.2.2.1 Raetz lipid A biosynthesis pathway 11
1.2.2.2 Synthesis of the core sugars 12
1.2.2.3 O-antigen synthesis and ligation 12
1.2.2.4 Transport of LPS to the outer membrane 13
1.2.3 Biological role of LPS 13
1.2.3.1 LPS and membrane integrity 14
1.2.3.2 Host-pathogen interaction 14
1.2.4 LPS modification systems 16
1.2.4.1 Acyl chain modifications 17
1.2.4.2 Phosphate modifications 17
1.2.4.3 Core sugars and O-antigen modifications 19
1.3 *B. pertussis* LPS 19
1.3.1 *B. pertussis* LPS structure 20
1.3.2 Lipid A structure of strains BP338 and 18-323 20
1.3.3 GlcN modification and the *lgm* locus 23
1.3.4 Bioinformatic analysis of the *lgm* locus and a hypothetical model 25
1.4 Hypotheses 28
1.5 Thesis goals 29
Chapter 2: Materials and methods 31
2.1 Bacterial growth conditions 31
2.2 Strains, plasmids, and primers 31
2.2.1 Bacterial strains 31
2.2.2 Plasmids 33
2.2.3 Primers 35

2.3 Cloning of vectors and deletion strains 37
2.3.1 General cloning techniques 37
2.3.2 Generating markerless deletion mutants 39
2.3.3 Site-directed mutagenesis 42
2.3.4 Vectors to complement *B. pertussis* strains 44
2.3.5 Vectors to complement *E. coli* strain R0138 46
2.3.6 Vector for LgmA expression in *E. coli* strain BL-21 47
2.4 Preparation of bacterial cells for mass spectrometry analysis and TLR4-activation assays 47
2.5 Isolation of LPS and lipid A 48
2.5.1 Isolation of *B. pertussis* LPS for TLR4 activation assays 48
2.5.2 Isolation of *E. coli* LPS for TLR4 and LAL activation assays 48
2.5.3 Isolation of lipid A for mass spectrometry analysis 49
2.6 Mass spectrometry analysis 50
2.7 HEK-Blue hTLR4 activation assay 50
2.7.1 Maintenance of HEK-Blue cells 50
2.7.2 hTLR4 activation assay 51
2.8 *E. coli* growth curves 51
2.9 Bacterial survival assays 52
2.9.1 Polymyxin B growth curve assay 52
2.9.2 Killing assays 52
2.10 LAL activation assay 54
2.11 Glycosyltransferase assay 54
2.12 Western blot analysis 56
2.13 Whole genome sequencing 56
2.14 Reverse transcriptase PCR 57

2.15 Bioinformatic analysis tools 58
2.16 Statistical analysis 58
Chapter 3: Identification and characterization of the *lgm* locus in *Bordetella*..... 59
3.1 Introduction 59
3.2 The *lgm* locus is required for lipid A GlcN modification in *B. pertussis* 59
3.2.1 *lgmA*, *lgmB*, and *lgmC* are required for lipid A GlcN modification..... 60
3.2.2 Potential flippase replacements are not required for GlcN modification 66
3.3 LgmA functions as a GlcNAc transferase 75
3.4 Identification of the putative active site of LgmA 79
3.5 Identification of the putative active site of LgmC 84
3.6 The *lgm* locus in other *Bordetella* species 90
3.7 Discussion..... 98
Chapter 4: The biological effects of lipid A modifications in *B. pertussis*..... 102
4.1 Introduction 102
4.2 Effect of GlcN modification on resistance to CAMPs and membrane stability 102
4.2.1 GlcN modification increases resistance to CAMPs 103
4.2.2 GlcN modification increases resistance to OM perturbation 106
4.3 *lgm* locus and *lpxA* are involved in the differences in lipid A structure between strains BP338 and 18-323 108
4.3.1 Difference in GlcN modification is due an incomplete *lgm* locus 108
4.3.2 Difference in C3' acyl chain length is due to LpxA..... 112
4.4 C3' acyl chain length and GlcN modification individually affect hTLR4 activation..... 114
4.5 Discussion..... 116
Chapter 5: The biological effects of varying the C3 and C3' acyl chain lengths in *E. coli* hexa-acyl lipid A..... 121

5.1 Introduction 121
5.2 Generation of R0138 *E. coli* strains with different C3 and C3' acyl chain lengths 122
5.3 C3 and C3' acyl chain lengths affect bacterial growth 128
5.4 C3 and C3' acyl chain lengths affect resistance to polymyxin B 130
5.5 C3 and C3' acyl chain lengths affect hTLR4 activation 133
5.6 C3 and C3' acyl chain lengths affect LAL activation 135
5.7 Discussion 137
Chapter 6: Discussion 141
6.1 Discussion 141
6.2 Conclusions 149
6.3 Future directions 150
References 153
Appendices 161
Appendix A Mass spectra 161
Appendix B Sequence alignments 165

List of Tables

Table 1. List of bacterial strains 33

Table 2. List of plasmids 34

Table 3. List of primers 35

Table 4. Summary of markerless deletion mutant cloning 42

Table 5. Summary of site-directed mutant primer sets 44

Table 6. List of CAMPs and antibiotics used in killing assays 54

Table 7. GlcN modification of BP338 *lgm* locus mutants 64

List of Figures

Figure 1. Gram-negative bacterial membrane schematic. 7

Figure 2. General structure of LPS 9

Figure 3. Schematic for the TLR4-MD-2-lipid A interaction. 15

Figure 4. Lipid A structure of *B. pertussis* strains BP338 and 18-323 21

Figure 5. hTLR4 activation by BP338 and 18-323 *B. pertussis* strains 22

Figure 6. Initially identified *B. pertussis lgm* locus 24

Figure 7. Predicted topologies of the Lgm proteins 26

Figure 8. Hypothetical model for lipid A GlcN modification in *B. pertussis* 27

Figure 9. Schematic for the generation of markerless deletion mutants. 41

Figure 10. Transcription of *lgmABCD* in BP338 and 18-323 + pPtacLgmABCD 62

Figure 11. Mass spectrometry analysis of BP338 and BP338*lgmABCD*KO lipid A 63

Figure 12. hTLR4 activation by BP338 *lgm* locus mutants 65

Figure 13. Predicted topologies of BP1945 and LgmE 68

Figure 14. Current schematic of the *B. pertussis lgm* locus and transcription of *lgmE* in BP338 69

Figure 15. Schematic of the *lgmE* mutant 70

Figure 16. Transcription of *lgm* locus genes in BP338 *lgmE*KO strains 71

Figure 17. hTLR4 activation by BP338 *lgmE*KO strains 72

Figure 18. hTLR4 activation by BP338 BP1945KO strains 74

Figure 19. Expression of LgmA-His in *E. coli* 77

Figure 20. LgmA glycosyltransferase assay 78

Figure 21. Identifying conserved residues in LgmA 80

Figure 22. hTLR4 activation by LgmA mutants 81

Figure 23. Phyre predicted model of the structure of LgmA 83

Figure 24. Identifying conserved residues in LgmC 86
Figure 25. hTLR4 activation by LgmC mutants 87
Figure 26. Phyre predicted model of the structure of LgmC 89
Figure 27. *lgm* locus of the sequenced *Bordetella* species 93
Figure 28. Comparison of the start of LgmC between *Bordetella* species 94
Figure 29. Neighbour-joining trees of the *lgm* locus genes in *Bordetella* species 97
Figure 30. Influence of the *B. pertussis* lipid A GlcN modification on CAMP susceptibility 104
Figure 31. Complementation of BP338LgmABCDKO rescues resistance to polymyxin B 105
Figure 32. Influence of the *B. pertussis* lipid A GlcN modification on OM stabilization 107
Figure 33. Genetic analysis of the *lgm* locus of *B. pertussis* strains BP338 and 18-323 110
Figure 34. Lipid A structures of *B. pertussis* 18-323 strains complemented with BP338 lipid A-modifying genes 111
Figure 35. Alignment of LpxA from various Gram-negative species 113
Figure 36. hTLR4 activation by LPS from *B. pertussis* 18-323 strains with BP338 lipid A-modification genes 115
Figure 37. Model for how the *B. pertussis* GlcN modification facilitates CAMP resistance and OM stabilization 120
Figure 38. Viability of *E. coli* R0138 strains complemented with *lpxA* at 30°C and 42°C 123
Figure 39. Lipid A structures of *E. coli* R0138 strains complemented with exogenous *lpxA* 127
Figure 40. Growth of *E. coli* R0138 strains with varying acyl chain lengths 129
Figure 41. Polymyxin B resistance of *E. coli* R0138 strains with varying acyl chain lengths, assessed by growth curve assay 131
Figure 42. Polymyxin B resistance of *E. coli* R0138 strains with varying acyl chain lengths, assessed by percent survival assay 132
Figure 43. Activation of hTLR4 by LPS from *E. coli* R0138 strains with varying acyl chain lengths 134

Figure 44. Activation of the LAL assay by LPS from *E. coli* R0138 strains with varying acyl chain lengths 136

List of Abbreviations

ABC: adenosine triphosphate (ATP)-binding cassette

ADP: adenosine diphosphate

Amp: ampicillin

AmpR: ampicillin resistant

Ara4FN: *N*-formylated 4-amino-4-deoxy-L-arabinose

Ara4N: 4-amino-4-deoxy-L-arabinose

ATP: adenosine triphosphate

BG: Bordet-Gengou

bp: base pair

Bvg: *Bordetella* virulence gene

BvgA-P: activated BvgA

C10-OH: hydroxydecanoic acid

C12-OH: hydroxydodecanoic chain

C14-OH: hydroxymyristoyl chain

C16-OH: hydroxypalmitoyl chain

C55P: undecaprenyl-phosphate

C55PP: undecaprenyl-pyrophosphate

CAMP: cationic antimicrobial peptide

cAMP: cyclic adenosine monophosphate

cfu: colony forming units

CP: cytoplasm

CyaA: adenylate cyclase

dH_2O: deionized water

DMEM: Dulbecco's modified Eagle's medium

DNA: deoxyribonucleic acid

dpm: degradations per minute

EDTA: ethylenediaminetetraacetic acid

FHA: filamentous hemagglutinin

GalNAc: galactosamine

GlcN: glucosamine

GlcN[^{14}C]: carbon-14-labeled glucosamine

GlcNAc: N-acetylglucosamine

G-protein: guanine nucleotide-binding protein

h: hour(s)

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His: histidine

HRP: horseradish peroxidase

hTLR4: human Toll-like receptor 4

IFN-γ: interferon γ

IL-1β: interleukin-1β

IL-6: interleukin 6

IM: inner membrane

IPTG: isopropyl β-D-1-thiogalactopyranoside

IRF3: interferon regulatory transcription factor 3

Kan^{R}: kanamycin resistant

Kdo: 3-deoxy-D-*manno*-octo-2-ulosonic acid

L,D-Hep: L-*glycero*-D-*manno*heptose

LAL: limulus amebocyte lysate

LB: Luria-Bertani (lysogeny broth)

LPS: lipopolysaccharide

MALDI-MS: matrix-assisted laser desorption/ionization mass spectrometry

MD-2*: dimerizing MD-2

min: minute(s)

mRNA: messenger ribonucleic acid

MyD88: myeloid differentiation primary response gene 88

nal^R: nalidixic acid resistant

NFκB: nuclear factor kappa-B

nt: nucleotide

OD: optical density

OM: outer membrane

ORF: open reading frame

pBD1: porcine β-defensin 1

PBS: phosphate-buffered saline

PCR: polymerase chain reaction

Phyre: Protein Homology/analogy Recognition Engine

P_i: inorganic phosphate

PP: periplasm

PRN: pertactin

PT: pertussis toxin

RLU: relative light units

RNA: ribonucleic acid

rpm: rotations per minute

RT-PCR: reverse transcriptase polymerase chain reaction

s: second(s)

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SP: surfactant protein

sRNA: small ribonucleic acid

SS: Stainer-Scholte

TCT: tracheal cytotoxin

TetR: tetracycline resistant

TLR2: Toll-like receptor 2

TLR4*: dimerizing Toll-like receptor 4 (TLR4)

TLR4: Toll-like receptor 4

TNFα: tumor necrosis factor-α

UDP: uridine diphosphate

UMP: uridine monophosphate

vag: *vir*-activated genes

vrg: *vir*-repressed genes

w/v: weight per volume

Acknowledgements

First, and foremost, I would like to thank Dr. Rachel C. Fernandez for all her support and advice over the years. Her mentorship has helped transform me from a timid undergraduate student to a confident young researcher – I am truly indebted to her.

I am very grateful to my committee: Dr. Robert E.W. Hancock, Dr. Zakaria Hmama, and Dr. Natalie Strynadka, for their insightful comments and guidance.

Dr. Martine Caroff has been an invaluable collaborator, whose advice and discussions have been greatly appreciated.

I would especially like to express my gratitude to my fellow lab members over the years – their support and sharing in both laughs and commiserations will not be forgotten. Thank you to Dr. Nico Marr, for all your patience when guiding me in my first forays into research. Thank you to Dr. Nina Maeshima, for always taking the time to turn away from your own work to discuss my research. Thank you to the students I supervised: Emma Kim, Rose Lee, and Andrew Low, for your contributions to my thesis work and publications, and to Dr. Jackie Felberg, for setting up the HEK hTLR4 assay in our lab, and Leigh Hobbs, for testing the different LAL assay kits.

I have greatly benefited from the support of my floor and wing mates, with their willingness to share equipment, bacterial strains, and knowledge.

I would also like to thank the Natural Sciences and Engineering Research Council of Canada, the Canadian Institute for Health Research, and the University of British Columbia for their funding.

Finally, my heartfelt thanks go to my family and friends, for all their support and understanding.

For my parents, L.P.S. and P.C.S., who have always been there for me.

For my brother, N.P.S., who shares my love of science.

Chapter 1: Introduction

1.1 *Bordetella pertussis* and whooping cough

Bordetella pertussis is a Gram-negative, aerobic, coccobacillus that was first isolated in 1906 from a human patient, and was identified as the causative agent of whooping cough. This bacterium infects the human respiratory tract and is an obligate strictly human pathogen with no known animal or environmental reservoir (Fedele, Bianco, et al. 2013, Mattoo and Cherry 2005).

Whooping cough is an acute respiratory illness that is a significant cause of infant death worldwide (Mattoo and Cherry 2005). Despite extensive childhood vaccination programs, there are approximately 40 million cases of pertussis every year, 200 000 to 400 000 of which result in death, with most of these deaths occurring in infants (Crowcroft, Stein, et al. 2003, Leung, Robson, et al. 2007). Immunity gained via vaccination or infection is short-lived, thus adults and adolescents are susceptible to infection by *B. pertussis* and act as a reservoir for this human-restricted pathogen. These individuals are the primary source of transmission to partially vaccinated or unvaccinated infants (Gustafsson, Hessel, et al. 2006, Hallander, Gustafsson, et al. 2005). The ability of *B. pertussis* to evade and modulate the immune system likely contributes to the persistence of this pathogen in humans (Mattoo and Cherry 2005).

1.1.1 Clinical manifestation

The most severe symptoms of whooping cough are often observed in infants, with neonates having the highest risk of death, at 3%. As summarized by Mattoo and Cherry, individuals infected with *B. pertussis* can present with classical pertussis, mild respiratory symptoms, or be asymptomatic. Classic pertussis illness, known as whooping cough, lasts 6 to 12 weeks, and presents in three stages. After a 7 to 10 day incubation period, the first stage of the disease, the catarrhal stage, lasts 7 to 14 days and initially resembles rhinovirus infection with the patient developing a mild cough. Throughout this stage, the cough worsens until it develops into a paroxysmal cough, which is the defining characteristic of the second stage of disease: the paroxysmal stage. During this period, violent coughing fits are interrupted by sharp intakes

of breath, resulting in a whooping sound, for which the disease has been named. The force of these coughing fits often leads to vomiting and consequently weight loss. During this 2 to 6 week stage, there is a risk of complications, such as pneumonia, otitis media, seizures, and encephalopathy. The last stage of the disease (if not leading to fatality), the convalescent stage, presents as a gradual decrease in the severity of the cough, and may last for about 1 to 2 weeks (Mattoo and Cherry 2005).

1.1.2 Molecular pathogenesis

B. pertussis causes disease by means of a variety of virulence factors, which are under control of the two component *Bordetella* virulence gene (Bvg)AS regulatory system (Cummings, Bootsma, et al. 2006). Adhesins are surface proteins that adhere to the airway epithelial cells, thereby enabling the bacterium to remain in the respiratory tract, despite the expulsatory movement of airway cilia (de Gouw, Diavatopoulos, et al. 2011). Toxins are secreted proteins that are required for colonization and persistence in the host. In addition to this, many of these bacterial factors also have immune evasive and immunomodulatory properties (Mattoo and Cherry 2005).

1.1.2.1 BvgAS regulation system

The BvgAS two-component signal transduction system consists of BvgS, the sensor kinase, and BvgA, the response regulator. Activation of BvgS by unknown environmental signals triggers autophosphorylation of BvgS, which leads to phosphorylation and activation of BvgA. Activated BvgA (BvgA-P) can then bind to BvgA-specific promoters upstream of virulence genes to activate transcription of these *vir*-activated genes (*vag*) (Decker, James, et al. 2012). BvgA-P also activates BvgR, a repressor protein that suppresses expression of a subset of genes (*vir*-repressed genes, *vrg*) that are not expressed during the virulence-associated Bvg^+ phase (Merkel, Boucher, et al. 2003). In contrast, during the Bvg^- phase, BvgAS is not activated; consequently there is no up-regulation of the *vag* genes, and the *vrg* genes are expressed. During the Bvg^i intermediate phase, some Bvg^+ phase genes are expressed, in addition to a subset of exclusively intermediate phase genes, such as *bipA*. Modulating *in vitro* growth conditions, by growing *B. pertussis* at 25°C or in the presence of $MgSO_4$ or nicotinic acid, suppresses the Bvg system,

resulting in Bvg^i or Bvg^- phase growth. However, the Bvg^- phase *in vivo* is associated with survival in the environment by *Bordetella bronchiseptica*, and since *B. pertussis* is a human-restricted pathogen, the role of the Bvg^- phase in *B. pertussis* has long thought to be a relic from the ancestral *B. bronchiseptica* (Cummings, Bootsma, et al. 2006).

1.1.2.2 Adhesins

The two major *B. pertussis* adhesins that facilitate direct binding to human epithelial cells are filamentous hemagglutinin (FHA) and fimbriae (de Gouw, Diavatopoulos, et al. 2011). FHA is a secreted and surface-associated protein that is critical for early colonization. It has four binding domains that facilitate binding to glycolipids on the surface of ciliated and non-ciliated cells of the respiratory epithelium (Alonso, Reveneau, et al. 2002, Mattoo and Cherry 2005). Fimbriae are long, filamentous surface protein structures that consist of a major subunit (Fim2 or Fim3) that makes up the extended structure capped by the tip protein, FimD. The major subunit of fimbriae is known to bind sugars found throughout the mammalian respiratory tract, such as heparin sulfate, and FimD facilitates binding to monocytes (Geuijen, Willems, et al. 1996, Hazenbos, Geuijen, et al. 1995).

1.1.2.3 Toxins

B. pertussis possesses numerous toxins, including pertussis toxin (PT), adenylate cyclase (CyaA), and tracheal cytotoxin (TCT) (Mattoo and Cherry 2005). PT is a secreted, hexameric AB_5 toxin, in which the catalytic subunit (PtxA) is an adenosine diphosphate (ADP)-ribosylating toxin, and the B subunits bind the surface of eukaryotic cells. Entry of PT into the host cell results in activation of PtxA, leading to inactivation of guanine nucleotide-binding proteins (G-proteins) and downstream histamine sensitization and suppression of the immune response. CyaA functions as an adenylate cyclase and hemolysin. The N-terminal region of CyaA facilitates entry of the C-terminal region into the host cell, upon which the C-terminal region is activated by the host enzyme calmodulin. Activated CyaA proceeds to convert adenosine triphosphate (ATP) into cyclic adenosine monophosphate (cAMP), leading to an excessive level of the important secondary messenger cAMP in the host cell, resulting in an overall anti-

inflammatory and antiphagocytic effect (as reviewed in (Carbonetti 2010)). Unlike the protein toxins PT and CyaA, TCT is a disaccharide-tetrapeptide breakdown product of peptidoglycan (Rosenthal, Nogami, et al. 1987). In most Gram-negative bacteria, this product is recycled by the cell via transport back into the cytoplasm by AmpG. However, *B. pertussis* lacks a functional AmpG, and therefore constitutively releases TCT from the cell (Mielcarek, Debrie, et al. 2006). TCT has toxic effects on host cells, and can result in mitochondrial bloating, disruption of tight junctions, and damage to ciliated cells (Fedele, Bianco, et al. 2013, Mattoo and Cherry 2005).

1.1.2.4 Other virulence factors

An important class of *B. pertussis* virulence factors is the autotransporter proteins, which consists of a C-terminal beta-barrel domain that facilitates the transport of the N-terminal passenger domain across the outer membrane (OM) of the bacterium. The passenger domain is the functional region of the protein, and varies between different autotransporters. Pertactin (PRN), the first identified and characterized autotransporter in *Bordetella*, is suggested to be involved in adhesion to eukaryotic cells. Two other autotransporter proteins, BrkA and Vag8, are required for resistance against killing by complement (Fernandez and Weiss 1994, Marr, Shah, et al. 2011). Lipopolysaccharide (LPS) also plays a crucial role in *B. pertussis* infection, and this is discussed in section 1.3.

1.1.2.5 Immune evasion and modulation

Many of the virulence factors of *B. pertussis* have immunomodulatory properties, and can work in synergy with one another. For example, FHA, which functions as the primary adhesin of *B. pertussis*, also induces apoptosis of human phagocitic and epithelial cells. PT and CyaA assist in this effort to remove phagocytes by intoxicating alveolar macrophages via inactivation of G-proteins and inhibiting monocyte and neutrophil activities by increasing intracellular cAMP, respectively. Additionally, PT inhibits immune cell recruitment and suppresses antibody responses and CyaA promotes apoptosis of phagocytes (as reviewed in (de Gouw, Diavatopoulos, et al. 2011)). The autotransporter Vag8 also plays a role in immune evasion through binding of C1-inhibitor to the surface of the bacterium, thus protecting the

pathogen against killing by complement (Marr, Shah, et al. 2011). BrkA, another autotransporter, protects *B. pertussis* by inhibiting the classical pathway of complement, specifically by preventing deposition of the complement protein C4b on the surface of the bacterium, though the mechanism of this protection is still unknown (Barnes and Weiss 2001).

1.1.3 Other *Bordetella* species

Genus *Bordetella* is a member of Class Betaproteobacteria and is made up of many host-associated species. To date, there are eight sequenced *Bordetella* species: *B. pertussis*, *Bordetella parapertussis*, *B. bronchispetica*, *Bordetella avium*, *Bordetella hinzii*, *Bordetella homesii*, *Bordetella trematum*, and *Bordetella petrii* (Shah, Moksa, et al. 2013). *B. pertussis*, *B. bronchispetica*, and *B. parapertussis* have been reclassified as "subspecies", due to the similarity of their genomes (Gerlach, von Wintzingerode, et al. 2001). *B. parapertussis* is split into two lineages of strains: the human-adapted strains, which only infect humans, and the ovine-adapted strains, which only infect sheep. Human-adapted *B. parapertussis* is similar to *B. pertussis*, in that it is a strict human pathogen and causes the whooping cough disease, though *B. parapertussis* cases of whooping cough tend to be less severe. This may be due to the lack of PT, since *B. pertussis* is the only species that possesses this toxin. *B. bronchiseptica* causes kennel cough in small four-legged mammals, and is also able to survive in the environment. *B. avium* and *B. hinzii* both infect birds; *B. avium* is pathogenic whereas *B. hinzii* is a commensal in the respiratory tracts of fowls, but it can infect immunocompromised humans (Spears, Temple, et al. 2003, Vandamme, Hommez, et al. 1995). *B. homesii* is associated with septicemia in humans and *B. trematum* has been isolated from ear infections and skin wounds in humans (Mattoo and Cherry 2005, Weyant, Hollis, et al. 1995). Finally, *B. petrii* is the only identified environmental species of this genus, and is proposed to be the progenitor of the host-associated *Bordetella* species (von Wintzingerode, Schattke, et al. 2001). *Bordetella ansorpii* is the most recent addition to genus *Bordetella*; it was isolated in 2005 from an immunocompromised patient, though its genome has yet to be sequenced (Ko, Peck, et al. 2005).

1.2 Lipopolysaccharide

Gram-negative bacterial cells have two membranes that enclose the cytoplasm: the inner membrane (IM) and the outer membrane (OM), both of which contain integral or membrane-associated proteins (Figure 1). The IM surrounds the cytoplasm and consists of a symmetrical lipid bilayer of phospholipids. Peptidoglycan, a sugar-peptide mesh-like molecule that surrounds the cell, is found in the periplasm, the region between the IM and OM. The OM is an asymmetric lipid bilayer, the inner layer of which is composed of phospholipids, whereas the outer layer is made of LPS. Also known as endotoxin, LPS plays an important role in maintaining the integrity of the OM and in the host-pathogen interactions (Needham and Trent 2013, Raetz and Whitfield 2002).

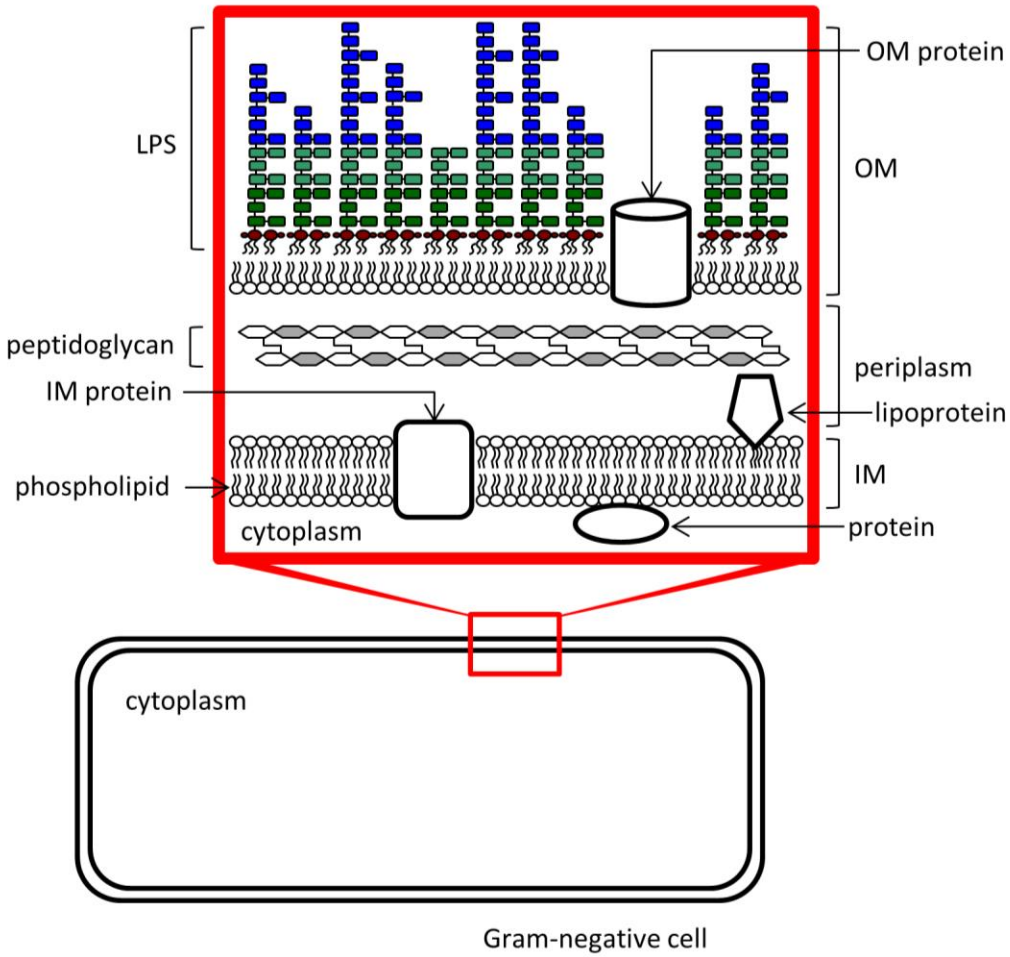


Figure 1. Gram-negative bacterial membrane schematic.

Gram-negative bacteria have two membranes: the inner membrane (IM) and the outer membrane (OM). Between these two membranes is the periplasm, which contains a thin layer of peptidoglycan. The IM is a symmetric bilayer, consisting primarily of phospholipids. The OM is an asymmetric membrane consisting of phospholipids in the inner leaflet and lipopolysaccharide (LPS) in the outer leaflet

1.2.1 LPS structure

LPS consists of three domains: lipid A, core sugars, and in some cases, the O-antigen (Figure 2 A). The lipid A region is amphipathic; the acyl chains are hydrophobic and anchor LPS to the OM, whereas the head group is hydrophilic, and interacts with the extracellular milieu. The lipid A domain is linked to the core sugars that extrude from the surface of the cell. Rough bacterial strains are characterized by the lack of O-antigen, and therefore the LPS of these strains consist of only the lipid A and core sugar regions. Smooth bacterial strains, however, have repeating O-antigen subunits linked to the core sugars, though the O-antigen is not added to every LPS molecule in the cell and the length of the O-antigen in most strains varies between different LPS molecules. This results in a heterogeneous mixture of LPS structures in a single bacterium (as reviewed by (Caroff and Karibian 2003)).

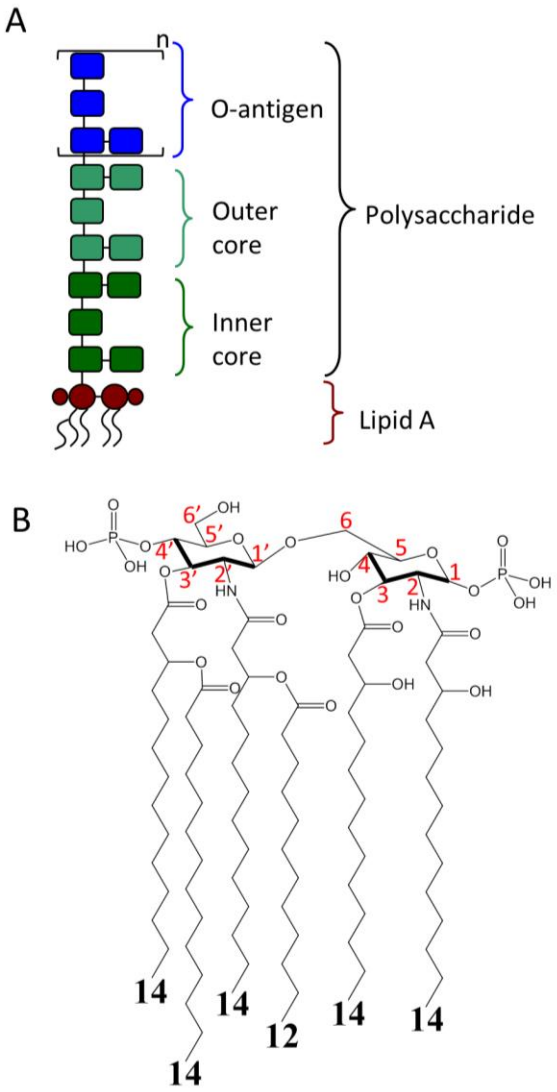


Figure 2. General structure of LPS
A) LPS consists of three main domains: lipid A, the core sugars (separated into the inner core and the outer core), and the O-antigen repeating polysaccharide. The core sugars and O-antigen make up the polysaccharide region. The 'n' denotes a number of repeats of the O-antigen subunit. B) Structure of *E. coli* lipid A, consisting of a di-glucosamine (GlcN) backbone decorated with phosphate groups and 6 acyl chains. Acyl chain length is denoted by the number directly below the chain. Acyl chains are either primary acyl chains, which are attached directly to the di-GlcN backbone, or secondary acyl chains, which are attached to a primary acyl chain. The numbering system of the di-GlcN backbone carbons is in red.

1.2.1.1 Lipid A

The structure of lipid A varies between different Gram-negative bacterial species. Generally, lipid A consists of a di-glucosamine (GlcN) backbone with four to seven primary and secondary acyl chains linked to the C2, C3, C2', or C3' carbons of the di-GlcN backbone (Figure 2B). Primary acyl chains are linked directly to the di-GlcN backbone and secondary acyl chains are connected to a primary acyl chain via an acyl-oxy-acyl linkage. The primary acyl chains at the C2 or C2' positions are connected to the di-GlcN backbone by amid linkages whereas the C3 or C3' position primary acyl chains are linked via ester bonds. The basic lipid A structure also has phosphate groups at the C1 and C4' positions, though these groups can be modified or removed by further modifications to lipid A. At the C6' position, one or two 3-deoxy-D-*manno*-octo-2-ulosonic acid (Kdo) sugars link the lipid A domain to the core sugar domain of LPS (as reviewed by (Caroff and Karibian 2003, Raetz and Whitfield 2002)). Some Gram-negative bacteria, such as *Aquifex pyrophilus*, deviate from this basic structure by using a di-galacturonic acid molecule backbone instead of the more common di-GlcN backbone (Plotz, Lindner, et al. 2000).

1.2.1.2 Core sugars

The core sugars of LPS can vary greatly between different bacteria, though the inner core region (the sugars proximal to the lipid A) tends to be conserved between members of a certain genus or family. The inner core often contains Kdo and L-*glycero*-D-*manno*heptose (L,D-Hep) sugars, whereas the sugars of outer core (distal to lipid A) are quite variable, likely due to selective pressures from the environment. The outer core region acts as a linker between the core polysaccharide and the O-antigen in smooth Gram-negative species (Raetz and Whitfield 2002).

1.2.1.3 O-antigen

In addition core sugars, smooth bacterial strains also have an O-antigen polysaccharide linked to the core polysaccharide of a fraction of the LPS molecules in the OM. The O-antigen is the most diverse region of LPS, and often varies between strains of a single species. It consists of a linear or branched repeating

pattern of sugars that varies in length between different LPS molecules in a single bacterium (Raetz and Whitfield 2002).

1.2.2 LPS Biogenesis

LPS is synthesized at the IM, and then transported to the outer leaflet of the OM. The first step is biosynthesis of lipid A-Kdo via the Raetz lipid A biosynthesis pathway, followed by a step-wise process of transfer and modification of the core sugars by individual glycosyltransferases directly onto lipid A-Kdo to produce lipid A-core. Synthesis of the lipid A-core molecule occurs at the cytoplasmic side of the IM, after which it is flipped to the periplasmic face of the IM. In smooth bacterial strains, the O-antigen is also built at the cytoplasmic side of the IM, then translocated to the periplasmic side of the IM, before ligation to produce the completed smooth LPS molecule, before transport to the surface (Raetz and Whitfield 2002).

1.2.2.1 Raetz lipid A biosynthesis pathway

The first stage of LPS biosynthesis is the Raetz lipid A biosynthesis pathway, which results in the production of lipid A-Kdo, take place in the cytoplasm and the inner leaflet of the IM, as reviewed by Raetz *et al*. (Raetz, Guan, et al. 2009). The enzymes of this pathway (Lpx proteins) are highly conserved across most Gram-negative bacterial species, though some species have variations in these Lpx proteins or additional modification enzymes that allow for atypical lipid A structures. In the general Raetz pathway, which was characterized in *Escherichia coli*, the first step is transfer of β-hydroxymyristoyl chain (C14-OH) from an acyl carrier protein to the C3 position of uridine diphosphate (UDP)-N-acetylglucosamine (GlcNAc) by LpxA (Anderson and Raetz 1987). LpxA is a homotrimer with an active site located between two monomeric subunits (Robins, Williams, et al. 2009). Orthologs of LpxA in other bacterial species can add acyl chains of different lengths to UDP-GlcNAc, based on the structure of the hydrocarbon ruler region of the LpxA active site (Shah, Albitar-Nehme, et al. 2013). LpxC irreversibly deacetylates the product of LpxA, thereby committing UDP-3-*O*-(acyl)-GlcN to the Raetz pathway, and this is followed by transfer of another C14-OH to the C2 position by LpxD, resulting in UDP-2,3-diacyl-

GlcN (Kelly, Stachula, et al. 1993, Raetz, Guan, et al. 2009). LpxH then cleaves UDP, to produce lipid X (2,3-diacyl-GlcN-1-phosphate), though some bacterial species lack LpxH, and instead the enzyme LpxI generates lipid X from UDP-2,3-diacyl-GlcN (Metzger and Raetz 2010). Then one molecule of UDP-2,3-diacyl-GlcN and one molecule of lipid X are joined by LpxB to form a β,1'-6 linkage (Raetz, Guan, et al. 2009). LpxK subsequently phosphorylates the 4' position of this molecule, to generate lipid IV_A, a tetra-acyl precursor of lipid A (Ray and Raetz 1987). In *E. coli*, WaaA (formerly KdtA) adds two Kdo sugars to the 6' position of lipid IV_A, followed by addition of the secondary acyl chains by LpxM and LpxL to produce hexa-acyl Kdo_2-lipid A. LpxM adds the C12-OH secondary acyl chain at the C2' position and LpxL adds the C14-OH secondary acyl chain at the C3' position. Bacteria lacking either LpxM or LpxL often generate penta-acyl lipid A, if they lack additional acyl-chain modifying enzymes (Raetz, Guan, et al. 2009).

1.2.2.2 Synthesis of the core sugars

Core sugars are added to Kdo_2-lipid A by a set of sequential glycosyltransferases at the cytosolic face of the IM. For example, in *E. coli* strains with an R1 LPS core structure, L,D-Hep is transferred to the Kdo proximal to the lipid A by the glycosyltransferase WaaC, followed by transfer of a second L,D-Hep by another glycosyltransferase WaaF onto the first L,D-Hep (Clementz and Raetz 1991, Gronow, Brabetz, et al. 2000). Successive glycosyltransferases and modifying enzymes build the remaining sugars in the R1 core, resulting in a molecule consisting of lipid A and the core sugars (Raetz and Whitfield 2002). Lipid A-core is then flipped to the periplasmic face of the IM by the ATP-binding cassette (ABC) transporter MsbA, upon which either the lipid A-core is directly transported to the OM via the Lpt system or the O-antigen is ligated to the outer core polysaccharide and then transported to the OM (Sperandeo, Deho, et al. 2009).

1.2.2.3 O-antigen synthesis and ligation

The repeating sugar subunits of the O-antigen polysaccharide are synthesized on the IM carrier lipid undecaprenyl-phosphate (C55P) before the fully synthesized O-antigen is ligated to the distal end of

rough LPS via WaaL on the periplasmic face of the IM. Most bacteria have one of three mechanisms for synthesizing the O-antigen: the Wzy-dependent pathway, the ABC-transporter-dependent pathway, or the synthase-dependent pathway. In all three pathways O-antigen synthesis is initiated by transfer of a sugar-1-phosphate to C55P to generate undecaprenyl-pyrophosphate (C55PP)-sugar by a transferase enzyme, such as WecA in *Salmonella enterica* serovar Typhimurium. This is followed by synthesis of the long, repeating sugar chain on C55PP, and flipping of C55PP-O-antigen to the periplasmic face of the IM, allowing WaaL to transfer the O-antigen from C55PP to the rough LPS to generate a smooth LPS molecule (as reviewed by (Raetz and Whitfield 2002)).

1.2.2.4 Transport of LPS to the outer membrane

Completion of LPS biosynthesis and MsbA activity results in LPS embedded in the periplasmic side of the IM by the hydrophobic acyl chains of lipid A. The Lpt system functions to transport this amphipathic molecule from the outer leaflet of the IM, across the periplasm and OM, to the outer leaflet of the OM. The current model for transport of LPS to the bacterial surface by the Lpt system is as follows: first, the ABC transporter complex, consisting of LptBCFG, extracts LPS from the periplasmic face of the IM and transfer it to LptA, which binds LPS (Chng, Gronenberg, et al. 2010, Ruiz, Gronenberg, et al. 2008). LptA is hypothesized to shield the hydrophobic acyl chains of LPS within a hydrophobic pocket during transport across the hydrophilic periplasmic environment, followed by transfer to the periplasmic OM lipoprotein LptE (Tran, Trent, et al. 2008). LptE is proposed to plug the pore of the β-barrel, integral OM protein LptD, thereby allowing the LptE-LptD complex to insert LPS directly into the outer leaflet of the OM through an opening in the LptD β-barrel wall (Freinkman, Chng, et al. 2011).

1.2.3 Biological role of LPS

LPS makes up the outer most layer of the bacterial membrane, and therefore acts as the first barrier between the bacterium and the extracellular environment (Needham and Trent 2013). Furthermore, LPS plays an important role in the interaction between the bacterium and the host: the polysaccharide region

extruding from the bacterial surface interacts with host factors and the lipid A region affects the immune system via activation of Toll-like receptor 4 (TLR4) (Caroff and Karibian 2003).

1.2.3.1 LPS and membrane integrity

The barrier created by LPS consists of tightly packed acyl chains inside the OM and di-GlcN backbone with negatively-charged phosphate groups making up the outer surface of the membrane. These negatively-charged phosphate groups are bridged by divalent cations, which also partially neutralize the overall negative charge of the OM surface, thereby increasing the integrity of the membrane. The barrier function of LPS also protects the bacterium from cationic antimicrobial peptides (CAMPs), which are found in both the environment, when produced by environmental bacteria, and in the host (Hancock 1997). CAMPs gain entry into bacterial cells by displacing the bridging cations and interacting with the negatively-charged phosphates of LPS, leading to self-promoted uptake of the CAMPs into the cell. Some bacteria are able to increase resistance to CAMPs by modifying the structure of lipid A (Hancock 1997, Raetz, Reynolds, et al. 2007).

1.2.3.2 Host-pathogen interaction

Both major domains of LPS, the polysaccharide region and the lipid A region, are involved in the host-pathogen interplay. The polysaccharide region, especially the O-antigen that extrudes from the cell in smooth bacterial strains, is highly antigenic, though these long chains also protect bacteria from numerous factors, such as complement and antibiotics (Caroff and Karibian 2003). Furthermore the O-antigen of some bacterial species aid infection by adhering to mammalian tissues, as found in the LPS of *Actinobacillus pleuropneumoniae* (Boekema, Stockhofe-Zurwieden, et al. 2003).

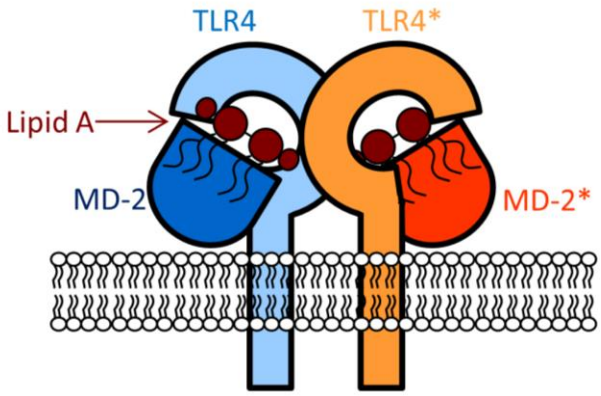


Figure 3. Schematic for the TLR4-MD-2-lipid A interaction.
Interaction with LPS (lipid A is shown in red for simplicity) promotes formation of the TLR4-MD-2 heterodimer (TLR4 and MD-2 shown in blue). Interaction between one TLR4-MD-2 heterodimer (blue) and another TLR4*-MD-2* heterodimer (orange) is mediated by interactions between lipid A bound to MD-2 (blue) and TLR4* (orange). This TLR4 dimerization leads to activation and downstream production of proinflammatory cytokines.

The lipid A region of LPS is recognized by the TLR4-MD-2 heterodimer that is found on many cell types, including macrophages and dendritic cells (Trent, Stead, et al. 2006). The acyl chains of lipid A sit in the hydrophobic pocket of MD-2 and the di-GlcN backbone sits at the top of the MD-2 pocket where the phosphate groups of lipid A interact with both TLR4 and TLR4* (the incoming, dimerizing TLR4 unit). This interaction results in dimerization and activation of TLR4 (Figure 3), which leads to the initiation of signaling cascades (Park, Song, et al. 2009). If TLR4 dimerizes at the surface of the host cell, signaling via the myeloid differentiation primary response gene 88 (MyD88)-dependent pathway results in early activation of the transcription factor nuclear factor kappa-B (NFκB) and production of pro-inflammatory cytokines, such as tumor necrosis factor-α (TNFα) and interleukin-1β (IL-1β). Internalization of TLR4 into an endosome results in signaling via the MyD88-independent pathway, which results in activation of interferon regulatory transcription factor 3 (IRF3) and the production of type I interferons, along with "late" activation of NFκB (Kagan, Su, et al. 2008, Maeshima and Fernandez 2013). Activation of TLR4 also leads to downstream production of costimulatory molecules that are required for the generation of an adaptive immune response (Raetz and Whitfield 2002).

1.2.4 LPS modification systems

Since the structure of LPS can affect its numerous biological roles, many Gram-negative bacteria have developed mechanisms to modify LPS structure to adapt to environmental stresses. Enzymes that facilitate LPS modification generally function at the periplasmic side of the IM, thereby modifying LPS before transport across the periplasm, or at the OM, thus modifying LPS that is already present at the surface (Needham and Trent 2013). Modification of the lipid A region, including the acyl chains and the phosphate groups, can affect CAMP resistance and TLR4 activation (Raetz, Reynolds, et al. 2007). Altering the structure of the polysaccharide region can also affect interactions with the immune system, such as increasing resistance to complement-mediated bacterial killing (Needham and Trent 2013). LPS modification systems do not modify every molecule of LPS within a bacterium; therefore, they generate heterogeneity in the LPS of a single bacterium (Caroff and Karibian 2003).

1.2.4.1 Acyl chain modifications

Though unmodified *E. coli* LPS is hexa-acylated (Figure 2), the lack of secondary acyltransferases (LpxL and LpxM) and the presence of acyl chain modifying enzymes in *E. coli* and other bacterial species can result in hepta-, penta-, or tetra-acyl LPS molecules (Trent, Stead, et al. 2006). Importantly, the number and position of the acyl chains in LPS affect the ability of LPS to activate TLR4, likely due to the position of these differently shaped LPS molecules in the MD-2-TLR4 binding pocket (Figure 3) (Needham, Carroll, et al. 2013). In *Salmonella*, PagP transfers C16-OH secondary acyl chain to the C2 position, resulting in hepta-acyl LPS molecules. This modification increases resistance to CAMPs, perhaps by increasing OM integrity due to more tightly packed LPS in the membrane (Guo, Lim, et al. 1998, Trent, Stead, et al. 2006). PagP-modification of *Salmonella* LPS also decreases activation of TLR4 (Kawasaki, Ernst, et al. 2004). However, PagP plays a slightly different role in *B. bronchiseptica*, since it incorporates the secondary C16-OH chain at the C3' position, resulting in hexa-acyl lipid A. This modification in *B. bronchiseptica* does not affect CAMP resistance, though it is required for persistent infection of the mouse respiratory tract (Preston, Maxim, et al. 2003). In contrast, PagL removes the C3 position primary acyl chain, and in *Salmonella*, this modification does not affect resistance to CAMPs, but does attenuate activation of TLR4 (Trent, Pabich, et al. 2001, Trent, Stead, et al. 2006). LpxO also modifies acyl chains, but instead of adding or removing them, LpxO hydroxylates the acyl chain and is hypothesized to be involved in coordination of the stress response (Needham and Trent 2013) .

1.2.4.2 Phosphate modifications

The phosphate groups of lipid A (Figure 2) are important for both susceptibility to CAMPs, since they lend to the overall negative charge of the OM, and for activation of TLR4, as they interact with key positive residues in TLR4 and TLR4* to promote dimerization. Thus, bacterial modification of these phosphate groups can act as a mechanism to evade the host immune system by decreasing susceptibility to CAMPs and modulating activation of TLR4 (Maeshima and Fernandez 2013, Needham and Trent 2013). For example, *Francisella tularensis* LpxE and LpxF remove the phosphate groups at the 1 and 4'

positions, respectively, although LpxF functions only in the absence of the C3' secondary acyl chain (Wang, Karbarz, et al. 2004, Wang, McGrath, et al. 2006). Removal of the 4'-phosphate group in *F. tularensis* results in increased resistance to CAMPs, but no difference in the ability of lipid A to activate TLR4. Interestingly, removal of the 1-phosphate in *Salmonella* that expresses LpxE results in a monophosphoryl lipid A species that has attenuated TLR4 activation (Raetz, Reynolds, et al. 2007).

In *Salmonella* and *E. coli*, enzymes encoded by the Arn locus function to modify the 4'-phosphate of lipid A with a 4-amino-4-deoxy-L-arabinose (Ara4N) moiety, resulting in greater resistance to the CAMP polymyxin B, though no reported difference in TLR4 activation (Gunn, Lim, et al. 1998, Trent, Stead, et al. 2006). The increase in CAMP resistance is thought to be due to a decrease in the overall negative charge of the OM by addition of the positively charged Ara4N group. This, hypothetically, results in less interaction of the LPS with the positively charged CAMPs, thereby increasing CAMP resistance (Needham and Trent 2013). The Arn pathway has been studied in *E. coli* and *Salmonella* to determine the mechanism by which the lipid A is modified with Ara4N (as reviewed by (Raetz, Reynolds, et al. 2007)). ArnA and ArnB initiate the pathway by synthesizing *N*-formylated UDP-Ara4N (UDP-Ara4FN) in the cytoplasm (Breazeale, Ribeiro, et al. 2005). This is followed by transfer of Ara4FN to C55P, the IM carrier lipid, by the GT2 family glycosyltransferase ArnC and then removal of the formate group by ArnD, to generate C55P-Ara4N (Breazeale, Ribeiro, et al. 2005). Next, ArnE and ArnF, two short proteins that each consist of 4 transmembrane helices, function as a flippase to transport C55P-Ara4N from the cytoplasmic face to the periplasmic face of the IM (Yan, Guan, et al. 2007). Finally ArnT, a member of the GT83 family of glycosyltransferases, adds Ara4N to the 4'-phosphate group of LPS (Trent, Ribeiro, et al. 2001). An ArnT homolog in *Pseudomonas aeruginosa* also modifies LPS with Ara4N, though in this bacterium both the 1- and 4'-phophates are generally modified and this modification increases both CAMP resistance and the inflammatory response, likely by increasing TLR4 activation (Gellatly, Needham, et al. 2012, Moskowitz, Ernst, et al. 2004). Thus, as seen with the PagP

acyl chain modification, homologous enzymes can modify lipid A in slightly different ways and result in different biological effects.

1.2.4.3 Core sugars and O-antigen modifications

The polysaccharide region of LPS can also be modified in a variety of manners, such as addition of sugar, phosphate, and phosphoethanolamine groups, by the bacterium to adapt to different environmental conditions (Needham and Trent 2013). One such example is the addition of sialic acid and glucose to the core sugars of *Neisseria gonorrhoeae* LPS. This increases the affinity of the LPS for the complement regulatory proteins Factor H and complement component 4b (C4b)-binding protein, thus leading to increased resistance to complement-mediated killing (Needham and Trent 2013, Ram, Ngampasutadol, et al. 2007).

1.3 *B. pertussis* LPS

LPS plays an important role in *B. pertussis* infection. As mentioned in Section 1.2.3, LPS is crucial to the structural integrity of the OM and interacts with the host immune system via activation of TLR4, which leads to downstream release of proinflammatory cytokines. Activation of human TLR4 (hTLR4) by penta-acyl *B. pertussis* LPS from strain Tohama I is not as robust as activation by hexa-acyl *E. coli* LPS, but still results in the production of proinflammatory cytokines such as TNFα, interleukin 6 (IL-6), and interferon γ (IFN-γ) (Marr, Novikov, et al. 2010). In addition, *B. pertussis* LPS acts as an antigen during natural infection, and protects the bacterium from surfactant proteins (SPs) A and D (Schaeffer, McCormack, et al. 2004, Trollfors, Lagergard, et al. 2001). SPA and SPD are lipid-binding lectins that are expressed in the human lower respiratory tract and damage bacterial membranes by binding to LPS (Chaby, Garcia-Verdugo, et al. 2005). The trisaccharide domain of *B. pertussis* LPS (see Section 1.3.1) prevents binding of SPA and SPD, likely through steric hindrance, therefore protecting the bacterium from this host clearance mechanism (Fedele, Bianco, et al. 2013, Schaeffer, McCormack, et al. 2004).

1.3.1 *B. pertussis* LPS structure

B. pertussis LPS consists of three regions: lipid A, the core polysaccharide, and a trisaccharide, which is attached to the distal end of the core polysaccharide in a fraction of the LPS molecules. Therefore, *B. pertussis* LPS separates into two bands by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis: a faster-migrating band B, which consists of only the lipid A and core sugars, and band A, which is band B plus the trisaccharide. The branched core polysaccharide contains 10 sugar subunits, including only one Kdo sugar that links the core sugars to the lipid A domain (Caroff, Brisson, et al. 2000). Though the lipid A structure can vary between different strains of *B. pertussis*, in general, it has a di-GlcN backbone decorated with a phosphate group at the 1 and 4' positions, and 5 acyl chains. A characteristic feature of *B. pertussis* lipid A is a short C10-OH primary acyl chain at the C3 position (Caroff, Brisson, et al. 2000, Marr, Novikov, et al. 2010).

1.3.2 Lipid A structure of strains BP338 and 18-323

The lipid A of *B. pertussis* wild-type strain BP338, a nalidixic acid-resistant derivative of strain Tohama I, consists of a di-GlcN backbone and generally 5 acyl chains: three C14-OH primary acyl chains at positions C2, C2', and C3', one C10-OH primary acyl chain at position C3, and a C14 secondary acyl chain at the C2' position (Figure 4, left). Furthermore, both phosphate groups may be modified with GlcN moieties, such that the bacterium contains LPS species with no GlcN modification, one GlcN modification at either the 1- or 4'-phosphate, or GlcN modification at both phosphate groups (Marr, Tirsoaga, et al. 2008). However, the lipid A structure of *B. pertussis* wild-type strain 18-323 differs from that of strain BP338 in two ways: 1) 18-323 does not have GlcN-modified lipid A, and 2)18-323 lipid A has C10-OH or C12-OH acyl chains at the C3' position, which are shorter than the C14-OH acyl chain at this position in BP338 (Figure 4) (Marr, Novikov, et al. 2010). These differences in lipid A structure are likely responsible for the greater ability of BP338 LPS to activate hTLR4 compared to LPS from the 18-323 strain (Figure 5) (Marr, Novikov, et al. 2010).

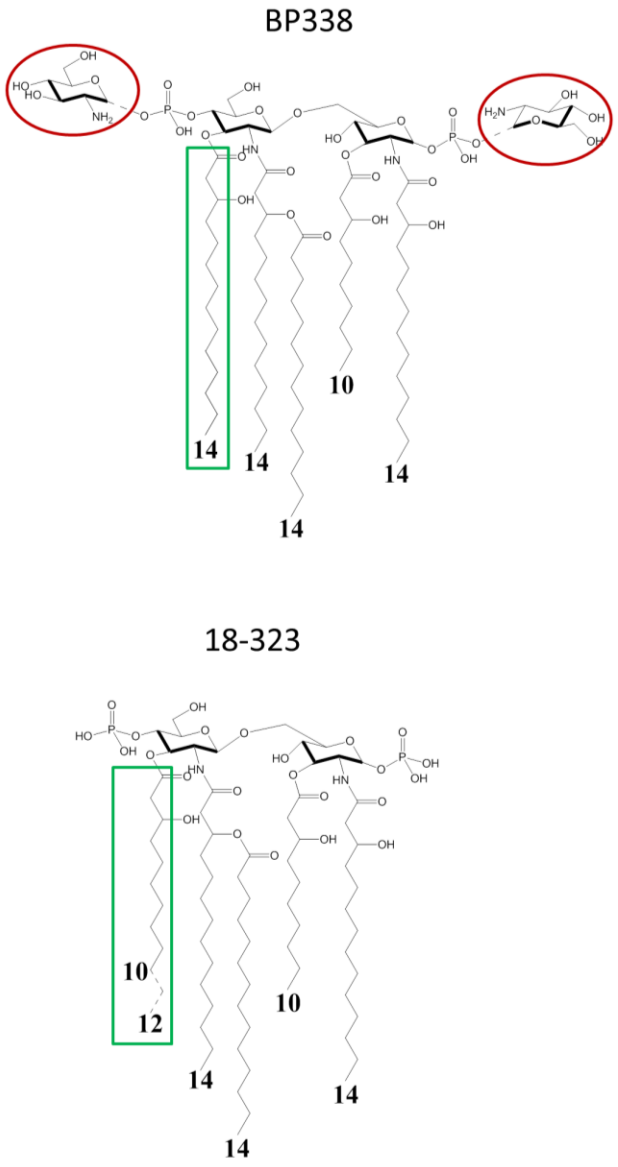


Figure 4. Lipid A structure of *B. pertussis* strains BP338 and 18-323
Both lipid A structures consists of a di-GlcN backbone decorated with 1- and 4'-phosphates and five acyl chains. BP338 lipid A (left) has GlcN modifications at the phosphate groups, circled in red, and a 14-carbon C3' acyl chain (green rectangle). 18-323 lipid A (right) has a 10- or 12-carbon C3' acyl chain (green rectangle). Acyl chain length is denoted by the number directly below or next to the chain. (Marr, Novikov, et al. 2010)

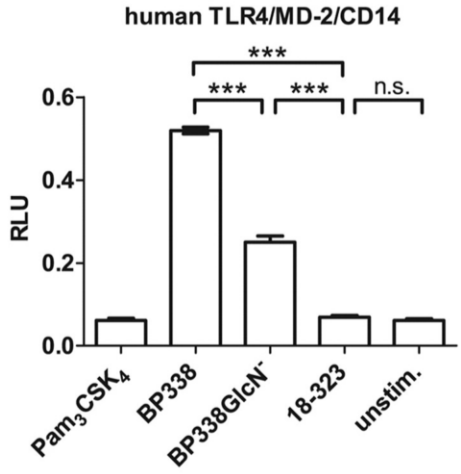


Figure 5. hTLR4 activation by BP338 and 18-323 *B. pertussis* strains
From Marr *et al*. 2010 (Marr, Novikov, et al. 2010), used with permission. Relative light units (RLUs; firefly luciferase/Renilla luciferase) as a readout of nuclear factor-kB activation by transfected HEK-293 cells after 4 h of stimulation with heat-killed *B. pertussis* BP338, its isogenic mutant BP338GlcN−, mouse challenge strain 18-323 (multiplicity of infection, 75), or the synthetic triacylated lipopeptide and Toll-like receptor 2 (TLR2) agonist Pam_3CSK_4 (1 mg/mL). n.s., not significant. ***, $P < 0.001$.

1.3.3 GlcN modification and the *lgm* locus

A transposon mutant of *B. pertussis* strain BP338 that lacks the lipid A GlcN modification has decreased levels of hTLR4 activation compared to the wild type BP338, though this GlcN-negative strain activates hTLR4 at a higher level than that of 18-323 (Figure 5) (Marr, Hajjar, et al. 2010). Interestingly, no difference is observed in the ability to activate murine TLR4 between *B. pertussis* BP338 strains that do or do not exhibit GlcN-modified lipid A (Marr, Hajjar, et al. 2010). The transposon in the BP338 mutant that lacks the GlcN modification disrupts the BP0398 locus tag (*lgmB*). Upstream of BP0398 is BP0399 (*lgmA*), and these two genes have overlapping start and stop codons, suggesting they may function together (Figure 6) (Marr, Tirsoaga, et al. 2008). Directly downstream of this gene pair is another set of two genes with overlapping start and stop codons: BP0397 (*lgmC*) and BP0396 (*lgmD*) (Marr, Tirsoaga, et al. 2008, Shah, Albitar-Nehme, et al. 2013). Thus, the *lgm* locus was originally predicted to encode the four genes: *lgmA*, *lgmB*, *lgmC*, and *lgmD*.

Figure 6. Initially identified *B. pertussis lgm* locus
The originally described *B. pertussis lgm* locus consists of *lgmA* (1059 bp), *lgmB* (1620 bp), *lgmC* (1104 bp), and *lgmD* (456 bp), formerly locus tags BP0399 to BP0396, respectively. The intergenic region between *lgmB* and *lgmC* is 188 bp.

1.3.4 Bioinformatic analysis of the *lgm* locus and a hypothetical model

The lipid A GlcN modification in *B. pertussis* strain BP338 is abrogated in the *lgmB* transposon mutant, suggesting LgmB is required for this GlcN modification. LgmB is a predicted homolog of *Salmonella* and *E. coli* glycosyltransferase ArnT, which modifies lipid A phosphate groups with Ara4N. The start codon of *lgmB* overlaps with stop codon of *lgmA*, a predicted homolog of ArnC, another glycosyltransferase in the lipid A Ara4N-modification pathway of *Salmonella* and *E. coli* (Marr, Tirsoaga, et al. 2008). The downstream gene *lgmC* encodes a member of the YdjC-like pfam protein family, and members of this family are involved in cleavage of cellobiose-phosphate. TTHB029, a member of the YdjC-like protein family, shows structural similarities with SpPgdA, a peptidoglycan GlcNAc deacetylase from *Streptococcus pneumonia* (Imagawa, Iino, et al. 2008). Lastly, *lgmD* is a small gene predicted to encode a protein composed primarily of four transmembrane helices (Figure 7D), similar to the structures of ArnE and ArnF, which form the C55P-Ara4N flippase in the *Salmonella* lipid A modification pathway (Shah, Albitar-Nehme, et al. 2013, Yan, Guan, et al. 2007).

Based on these observations, the lipid A GlcN modification (lgm) locus, consisting of *lgmA*, *lgmB*, *lgmC*, and *lgmD* (formerly locus tags BP0399 to BP0396, respectively), is proposed to be required for GlcN modification of lipid A in *B. pertussis* strain BP338 (Marr, Tirsoaga, et al. 2008). Furthermore, these Lgm proteins are predicted to function in the pathway summarized in Figure 8: LgmA transfers GlcNAc from UDP-GlcNAc to the IM carrier lipid C55P, then LgmC removes the acetyl group, followed by flipping of C55P-GlcN from the cytoplasmic to the periplasmic face of the IM by LgmD, and finally LgmB transfers GlcN from C55P to the phosphate of lipid A (Marr, Tirsoaga, et al. 2008). However, though previous data suggest *lgmB* is required for lipid A GlcN modification in *B. pertussis*, the involvement of the other *lgm* genes was still unknown at the beginning of this project, as was the function of the Lgm proteins.

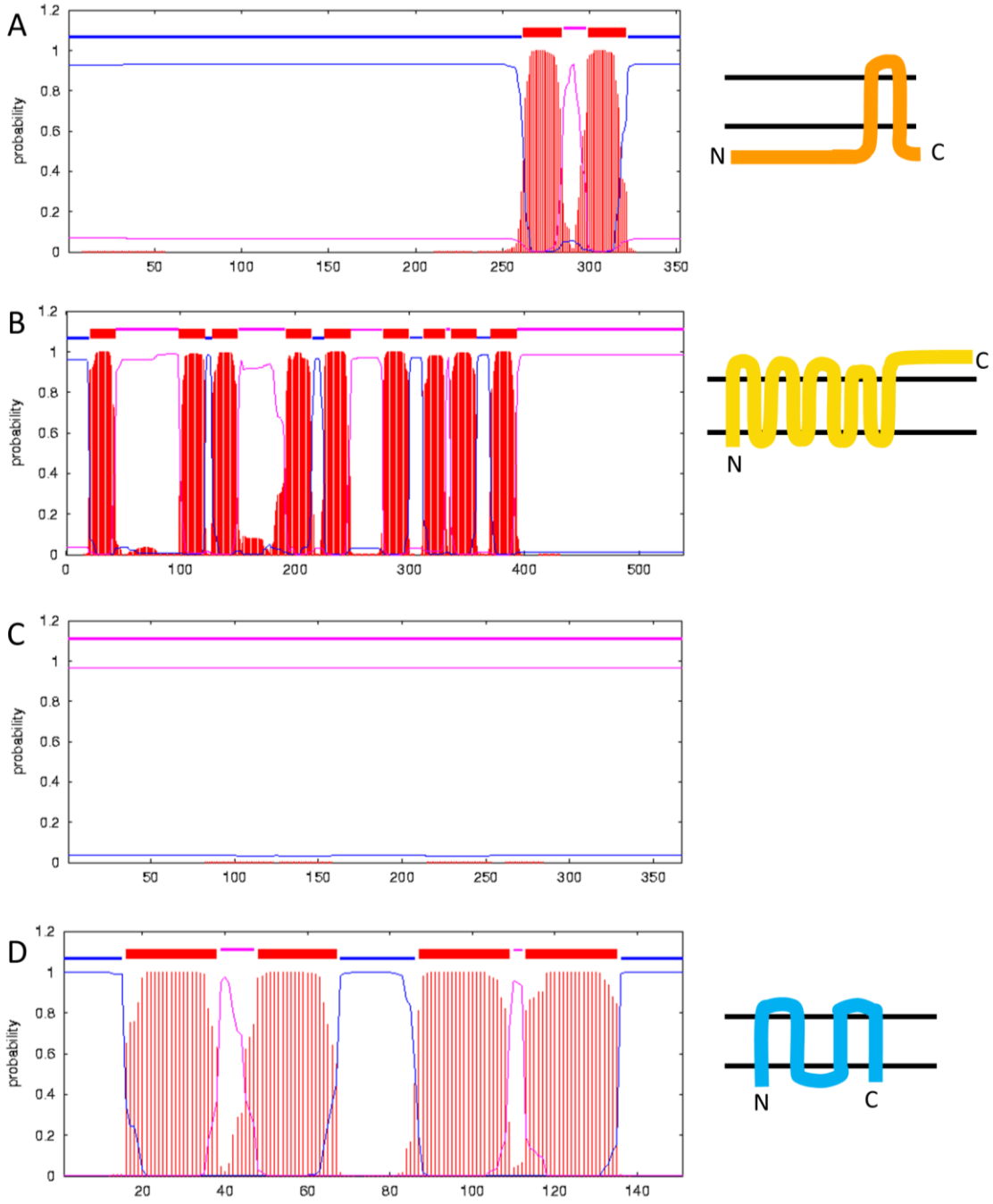


Figure 7. Predicted topologies of the Lgm proteins
A) LgmA, B) LgmB, C) LgmC, D) LgmD. Membrane topologies were predicted with TMHMM (Sonnhammer, von Heijne, et al. 1998). The amino acid position is along the x-axis, the probability of a transmembrane helix is along the y-axis (red vertical bars). For the horizontal lines: thick red line is predicted transmembrane region, blue is predicted cytoplasmic region, pink is predicted periplasmic region. The diagrams on the right depict a visual representation of the TMHMM prediction. The upper horizontal black line represents the periplasmic leaflet of the IM, the lower black line represents the cytoplasmic leaflet of the IM, and the coloured line represents the protein. For LgmC (C), since no transmembrane helices are predicted, the prediction by TMHMM of LgmC being placed in the periplasm or cytoplasm is not accurate.

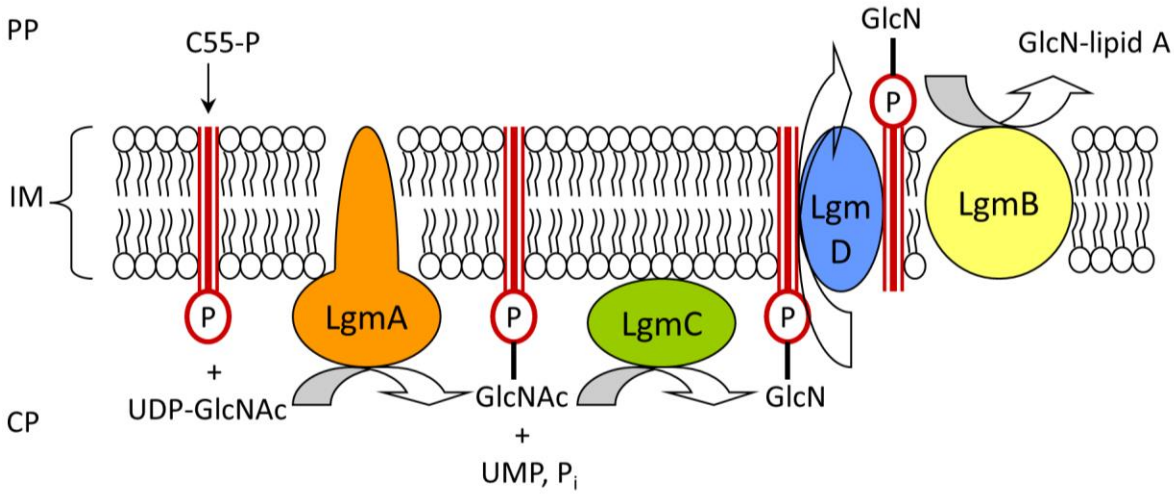


Figure 8. Hypothetical model for lipid A GlcN modification in *B. pertussis*
In this proposed model for the Lgm pathway in *B. pertussis*, LgmA transfers GlcNAc from UDP-GlcNAc to the IM carrier lipid C55P, resulting in the byproducts UMP and P_i. Then LgmC deacetylates C55P-GlcNAc, to generate C55P-GlcN, which is flipped to the periplasmic side of the IM by LgmD, followed by transfer of GlcN from C55P to lipid A by LgmB. PP, periplasm; IM, inner membrane; CP, cytoplasm; C55P, undecaprenyl phosphate; GlcNAc, N-acetyl-glucosamine; UMP, uridine monophosphate; P_i, inorganic phosphate; GlcN, glucosamine.

1.4 Hypotheses

Based on the bioinformatics analysis of the *lgm* locus, I hypothesize that all four genes of the *lgm* locus, *lgmA*, *lgmB*, *lgmC*, and *lgmD*, are required for GlcN modification of *B. pertussis* lipid A. Furthermore, since LgmA is a homolog of the GT2 glycosyltransferase ArnC, which transfers Ara4FN onto C55P, I predict that LgmA functions to transfer GlcNAc from UDP-GlcNAc to C55P (Figure 8) (Breazeale, Ribeiro, et al. 2005).

Modification of lipid A structures, especially with charged groups, often affect resistance to CAMPs (Needham and Trent 2013). Therefore, I hypothesize that the addition of the positively-charged GlcN moiety to the negatively-charged phosphate groups of *B. pertussis* lipid A will affect resistance to a variety of CAMPs. Since the negative charge of the phosphate groups is important for stabilizing the OM via coordination of cations (Needham and Trent 2013), I also predict these GlcN modification will affect OM stabilization in *B. pertussis*.

The lipid A structures of two *B. pertussis* strains, BP338 and 18-323, differ in two ways: the presence of the GlcN modification on the phosphate groups, and the length of the C3' acyl chain (Figure 4). Furthermore, the ability of these strains to activate hTLR4 greatly varies: BP338 induces moderate levels of hTLR4 activation whereas 18-323 activates hTLR4 at very low levels. A lack of GlcN modification in BP338 LPS results in a reduction in hTLR4 activation, however, this level of activation is still higher compared to 18-323 LPS (Marr, Novikov, et al. 2010). I therefore hypothesize that each difference in lipid A structure between *B. pertussis* BP338 and 18-323 strains individually affects hTLR4 activation.

Many Gram-negative bacteria, such as *Salmonella* species and *E. coli*, have hexa-acyl LPS. Removal and addition of acyl chains from hexa-acyl LPS can decrease TLR4 activation by these molecules and affect resistance to CAMPs (Trent, Stead, et al. 2006). However the effect of more subtle changes, such as

minor changes to the lengths of the acyl chains, have not been studied in species with hexa-acyl LPS. I hypothesize that gradually decreasing the length of the C3 and C3' C14-OH acyl chains in *E. coli* hexa-acyl LPS will affect many biological characteristics of the bacterium, including resistance to CAMPs and hTLR4 activation.

1.5 Thesis goals

A primary goal of this thesis is to characterize the *lgm* locus genes, which are the genes required for the modification of *B. pertussis* lipid A with GlcN. *lgmA*, *lgmB*, *lgmC*, and *lgmD* (locus tags BP0399 to BP0396, respectively) have been proposed to be required for this modification based on transposon mutagenesis and bioinformatics analysis (Marr, Tirsoaga, et al. 2008). Furthermore, these analyses suggest a hypothetical pathway for GlcN modification of lipid A (summarized in Figure 8) in which LgmA is proposed to transfer GlcNAc from UDP-GlcNAc to the carrier lipid C55P (Marr, Tirsoaga, et al. 2008). Therefore, demonstrating the function of LgmA could support or contradict this hypothetical model. To further characterize the first two enzymes of this model, mutational analysis of LgmA and LgmC, the proposed C55P-GlcNAc deacetylase, could identify functionally important amino acids in these two enzymes, and perhaps describe a putative active site region.

Analysis of the *lgm* locus in other *Bordetella* species may also highlight features of this locus, since many *Bordetella* species modify lipid A with GlcN. *B. hinzii* and *B. trematum*, however, have unmodified lipid A (Novikov, Shah, et al. 2013). Determining the genetic basis for the lack of GlcN modification in these two species may reveal clues as to the requirements for a functional Lgm pathway.

Another significant goal of this thesis is to characterize the biological effects of lipid A modifications in *B. pertussis*, such as the effect on CAMP resistance and membrane stability. Additionally, an important effect of lipid A modifications is the ability to influence activation of hTLR4 (Maeshima and Fernandez

2013, Needham and Trent 2013). The differences between the lipid A structures of the two *B. pertussis* strains BP338 and 18-323 provides a suitable model to dissect the relative effect of two different lipid A structural modifications on hTLR4 activation: the length of the C3' acyl chain and the GlcN modification.

The final goal of this thesis is to characterize the biological effects of gradually shortening the C3 and C3' acyl chain in hexa-acyl *E. coli*. Some attributes to explore are bacterial growth, resistance to polymyxin B, and activation of hTLR4 and the Limulus amebocyte lysate (LAL) assay by LPS that contains these different lipid A structures.

Chapter 2: Materials and methods

2.1 Bacterial growth conditions

B. pertussis strains were grown on BG agar supplemented with 15% defibrinated sheep's blood (Dalynn) at 37 °C or in Stainer-Scholte (SS) broth with 0.06% bovine serum albumin (Sigma-Aldrich) shaking at 180 rpm at 37°C (Stainer and Scholte 1970). All BP338 strain derivatives were grown in the presence of 30 µg/ml nalidixic acid and wild type strain 18-323 was grown in the absence of antibiotics, unless otherwise stated.

All *E. coli* strains were grown in LB broth shaking at 200 rpm or on LB agar with appropriate antibiotics. *E. coli* strain R0138 was grown on LB agar at 30°C or in LB broth at 30°C or 42°C, as specified, and *E. coli* strains DH5α, S17-1, and BL-21 were grown at 37°C. R0138 strains were grown in media supplemented with 12 µg/ml tetracycline and R0138 strains with pBlueScript II KS (-) derivatives were grown with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) when grown at 42°C.

B. pertussis and *E. coli* strains containing derivatives of pMMB67HE or pBlueScript II KS (-) were grown with 100 µg/ml ampicillin (Amp), strains containing derivatives of pBBR1MCS were grown with 34 µg/ml chloramphenicol, and strains containing derivatives of pSS4245, pET30b, or pBBR1MCS2 were grown with 50 µg/ml kanamycin.

2.2 Strains, plasmids, and primers

2.2.1 Bacterial strains

B. pertussis wild-type strains BP338 and 18-323, or mutants of these strains, were used in all experiments. *B. pertussis* strain BP338 is a nalidixic acid-resistant derivative of wild type strain Tohama I.

E. coli strain DH5α (Invitrogen) was used for cloning, plasmid maintenance, and transformation. *E. coli* strain DH5α has the following genotypic features: *endA* mutation eliminates non-specific endonuclease I activity, therefore allowing improved plasmid maintenance and preparations; *hsd* mutation which inactivates the methylation and restriction system that allows *E. coli* to recognize foreign deoxyribonucleic acid (DNA), therefore allowing transformation with polymerase chain reaction (PCR)-generated DNA. *E. coli* strain S17-1, which contains a chromosomally-integrated transfer (*tra*) locus, was used for conjugation. *E. coli* strain BL-21 was used for protein expression from pET30 vectors, since it contains a deletion of *lon* and a mutation in *ompT*, both of which result in decreased proteolysis, and contains the DE3 prophage, which encodes T7 ribonucleic acid (RNA) polymerase that is required for expression from T7 promoters, as found in pET30 vectors. The *E. coli* strain R0138 (Galloway and Raetz 1990) has *lpxA* under the control of a temperature sensitive promoter (*lpxA2*), such that the endogenous LpxA protein is expressed at low temperatures (i.e. 30°C), but is not expressed at higher temperatures (i.e. 42°C). The *recA* mutation in R1038 protects against reversion of *lpxA2* to wild type by preventing DNA recombination.

Table 1. List of bacterial strains

Strain	Description	Source
BP338	nal^R derivative of *B. pertussis* Tohama I wild-type strain	A. Weiss
BP338LgmABCDKO	Δ*lgmABCD*	This thesis
BP338LgmAKO	Δ*lgmA*	This thesis
BP338LgmBKO	Δ*lgmB*	This thesis
BP338LgmCKO	Δ*lgmC*	This thesis
BP338LgmDKO	Δ*lgmD*	This thesis
BP338BP1945KO	ΔBP1945	This thesis
BP338LgmDKOBP1945KO	Δ*lgmD* ΔBP1945	This thesis
18-323	*B. pertussis* wild type strain	A. Weiss
B. avium strain ATCC 5086	Wild type strain	M. Caroff
B. hinzii strain ATCC 51730	Wild type strain	M. Caroff
B. trematum strain CCUG 13902	Wild type strain	M. Caroff
E. coli R0138	TetR, *lpxA2 recA*	C. Raetz
E. coli BL-21 DE3	Δ*lon ompT*	Novagen
E. coli DH5α	*endA hsd*	Invitrogen
E. coli S17-1		(Simon, Priefer, et al. 1983)

2.2.2 Plasmids

The pSS4245 vector was a generous gift from Scott Stibitz. pSS4245 contains the following features that allow for the generation of markerless deletion mutants in *B. pertussis* strains: plasmid-encoded I-*Sce*I restriction endonuclease under the control of a Bvg-regulated promoter, such that I-*Sce*I is only expressed under Bvg^+ conditions (i.e. no supplementation of the media with $MgSO_4$), an I-*Sce*I enzyme cut site, kanamycin resistance cassette, and a streptomycin resistance cassette that provides resistance in *B. pertussis*, but not in *E. coli* strain S17-1.

Table 2. List of plasmids
Cloning details are in section 2.3.

Plasmids	Description	Source
pSS4245	Suicide vector (replicates in *E. coli*, but not *B. pertussis*), Kan^R	(Inatsuka, Xu, et al. 2010)
pSS4245LgmAKO	pSS4245 containing the BP338 *lgmA* upstream and downstream regions; for generation of BP338LgmAKO	This thesis
pSS4245LgmBKO	pSS4245 containing the BP338 *lgmB* upstream and downstream regions; for generation of BP338LgmBKO	This thesis
pSS4245LgmCKO	pSS4245 containing the BP338 *lgmC* upstream and downstream regions; for generation of BP338LgmCKO	This thesis
pSS4245LgmDKO	pSS4245 containing the BP338 *lgmD* upstream and downstream regions; for generation of BP338LgmDKO	This thesis
pSS4245LgmABCDKO	pSS4245 containing the BP338 *lgmA* upstream and the *lgmD* downstream regions; for generation of BP338LgmABCDKO	This thesis
pSS4245BP1945KO	pSS4245 containing the BP338 BP1945 locus tag upstream and downstream regions; for generation of BP338BP1945KO and BP338LgmDKOBP1945KO	This thesis
pBBR1MCS2	Broad-range vector, Kan^R	(Kovach, Elzer, et al. 1995)
pBBR2Pcpn	pBBR1MCS2 containing the Pcpn heat shock promoter (constitutive expression)	This thesis
pBBR2LgmA	pBBR2Pcpn containing *lgmA* of BP338	This thesis
pBBR2LgmA D76G D77G	D76G and D77G mutations in pBBR2LgmA	This thesis
pBBR2LgmA D127G	D127G mutation in pBBR2LgmA	This thesis
pBBR2LgmA D129G	D129G mutation in pBBR2LgmA	This thesis
pBBR2LgmA D127G D129G	D127G and D129G mutations in pBBR2LgmA	This thesis
pBBR2LgmA D159N	D159N mutation in pBBR2LgmA	This thesis
pBBR2LgmA W163R	W163R mutation in pBBR2LgmA	This thesis
pBBR2LgmA D159N W163R	D159N and W163R mutations in pBBR2LgmA	This thesis
pBBR2LgmC	pBBR2Pcpn containing *lgmC* of BP338	This thesis
pBBR2LgmC D80G D81G	D80G and D81G mutations in pBBR2LgmC	This thesis
pBBR2LgmC H130G	H130G mutation in pBBR2LgmC	This thesis
pBBR2LgmC D187G H189G	D187G and H189G mutations in pBBR2LgmC	This thesis
pBBR2LgmC E313G	E313G mutation in pBBR2LgmC	This thesis
pBBR2LgmC-E G15STOP	The *lgmE* G15STOP mutation in pBBR2LgmC	This thesis
pBBRLpxA338	pBBR1MCS2 containing *lpxA* of BP338	This thesis
pNMLgmAB	pBBR1MCS2 containing *lgmA* and *lgmB*	(Shah, Albitar-Nehme, et al. 2013)
pMMB67HE	Broad-range vector, AmpR, Ptac promoter (IPTG-inducible)	(Furste, Pansegrau, et al. 1986)
pPtacLgmABCD	pMMB67HE containing *lgmA*, *lgmB*, *lgmC*, and *lgmD* of BP338	(Shah, Albitar-Nehme, et al. 2013)
pPtacLgmABCD-E G15STOP	pPtacLgmABCD with a G15STOP mutation in *lgmE* (glycine 15 mutated to a stop codon)	This thesis
pPtacLgmABC	pMMB67HE containing *lgmA*, *lgmB*, and *lgmC* of BP338	This thesis
pPtacLgmABC-E G15STOP	pPtacLgmABC with a G15STOP mutation in *lgmE* (glycine 15 mutated to a stop codon)	This thesis
pPtacLgmAB	pMMB67HE containing *lgmA* and *lgmB* of BP338	This thesis
pPtacLpxA338	pMMB67HE containing *lpxA* of BP338	This thesis

Plasmids	Description	Source
pPtacLgmABCDLpxA338	pMMB67HE containing *lgmA*, *lgmB*, *lgmC*, *lgmD*, and *lpxA* of BP338	This thesis
pET30b	Expression vector, Kan^R	Novagen
pET30LgmA	pET30b containing *lgmA* with a C-terminal 3xGly linker connected to a 6xHis tag	This thesis
pBlueScript II KS (-)	AmpR, *lac* promoter (IPTG-inducible)	Stratagene
pBSlpxAEC	pBlueScript II with *lpxA* from *E. coli* strain DH5α, forward direction with respect to *lac* promoter	This thesis
pBSlpxA338	pBlueScript II with *lpxA* from *B. pertussis* strain BP338, forward direction with respect to *lac* promoter	This thesis
pBSlpxA338rev	pBlueScript II with *lpxA* from *B. pertussis* strain BP338, reverse direction with respect to *lac* promoter	This thesis
pBSlpxA18323	pBlueScript II with *lpxA* from *B. pertussis* strain 18-323, forward direction with respect to *lac* promoter	This thesis
pBSlpxA18323rev	pBlueScript II with *lpxA* from *B. pertussis* strain 18-323, reverse direction with respect to *lac* promoter	This thesis

2.2.3 Primers

Table 3. List of primers

Sequences are listed 5' to 3'. Underlined sequences represent restriction enzyme cut sites.

Primer	Sequence	Restriction enzyme sites
lgmAKO1fw	GGAATTCTGGTAGCCGTGCCGCAGCC	*Eco*RI
lgmAKO1rev	CCTTAATTAAACCACGGCGAAATTGACGGG	*Pac*I
lgmAKO2fw	CCTTAATTAATGACTCTCGCTACCCGATCC	*Pac*I
lgmAKO2rev	CCCCTAGGCAGCGTGGCCAGCACCAG	*Avr*II
lgmBKO1fw	CGAATTCGCGACTTCCGCCTGATGGAC	*Eco*RI
lgmBKO1rev	CCTTAATTAATCATTGGGAACGCGCCTTGGC	*Pac*I
lgmBKO2fw	GCTTAATTAACGACGAATGTCAGGAAGGCCA	*Pac*I
lgmBKO2rev	CCCCTAGGGACCGCGCTCAGGCGTC	*Avr*II
lgmCKO1fw	GGAATTCGCCGATGTGGGTGGTCGAC	*Eco*RI
lgmCKO1rev	CCTTAATTAATCAGACTCACTTTCCGCCAC	*Pac*I
lgmCKO2fw	CCTTAATTAAAATGAGTTCTTCCTCGAGACAAAC	*Pac*I
lgmCKO2rev	CCCCTAGGTCCGTGGCCGACTTCGTACG	*Avr*II
lgmDKO1fw	GGAATTCAGCGCATCTGGCTGCGG	*Eco*RI
lgmDKO1rev	CCTTAATTAATCATAAACGGCTTGCCAGGC	*Pac*I
lgmDKO2fw	CCTTAATTAAGGCCGCTCAGGTACCGGGC	*Pac*I
lgmDKO2rev	CCCCTAGGACTGCCCTCGGAGCAAAGCG	*Avr*II
BP1945KO1fw	CGAATTCACGAACGCCAGCCCCGCCACC	*Eco*RI
BP1945KO1rev	CCTTAATTAATCAGTGGATTTTTTTACAGCATAC	*Pac*I
BP1945KO2fw	CCTTAATTAAAAAGCCGGCTGGACAGTTG	*Pac*I
BP1945KO2rev	CCCCTAGGTGGTCGCGCAACGCGCACA	*Avr*II
lgmAD76GD77Gfw	CGTCGTCGGCGGCGGCAGCACCGACGATAC	
lgmAD76GD77Grev	TGCTGCCGCCGCCGACGACGATGATTTCCC	
lgmAD127Gfw	GTCTGGGCGCCGACATGCAGCATCCGCCCGAACT	
lgmAD127Grev	GATGCTGCATGTCGGCGCCCAGACAGATGACGGC	

Primer	Sequence	Restriction enzyme sites
lgmAD129Gfw	GTCTGGACGCCGGCATGCAGCATCCGCCCGAACT	
lgmAD129Grev	GATGCTGCATGCCGGCGTCCAGACAGATGACGGC	
lgmAD127GD129Gfw	CTCTGGGCGCCGGCATGCAGCATCCGCCCGAACT	
lgmAD127GD129Grev	GATGCTGCATGCCGGCGCCCAGACAGATGACGGC	
lgmAD159Nfw	GCGCAACGACGAGCCGTGGTTCAAGCGTGT	
lgmAD159Nrev	ACCACGGCTCGTCGTTGCGCTGGCGCCGCA	
lgmAW163Rfw	GCGCGACGACGAGCCGCGGTTCAAGCGTGT	
lgmAW163Rrev	ACCGCGGCTCGTCGTCGCGCTGGCGCCGCA	
lgmAD159NW163Rfw	GCGCAACGACGAGCCGCGGTTCAAGCGTGT	
lgmAD159NW163Rrev	ACCGCGGCTCGTCGTTGCGCTGGCGCCGCA	
lgmCD80GD81Gfw	GTGCGGCGGTGGTTTTGGCATGAACGAGGC	
lgmCD80GD81Grev	TGCCAAAACCACCGCCGCACACCGCGATGC	
lgmCH130Gfw	GACCTGGGCGTCGGTGTCGATTTCACCGAA	
lgmCH130Grev	TCGACACCGACGCCCAGGTCGACGTCCAGG	
lgmCD187GH189Gfw	ACGTCGGCGGCGGCCAGCACGTGCATCAGC	
lgmCD187GH189Grev	GTGCTGGCCGCCGCCGACGTAGTCGGGCGC	
lgmCE313Gfw	CGCCGCCGGATACGAGGTGCTGGCGCACCC	
lgmCE313Grev	GCACCTCGTATCCGGCGGCGCGCTGGGAGG	
BPlgmDlikeG15STOPfw	TTCATTGCCGTCTGATGCGCCGCGGCCGCC	
BPlgmDlikeG15STOPrev	GCGCATCAGACGGCAATGAACCAGGCAATC	
BP0399fw1	ATCTGTCTGGACGCCGACATGC	
BP0399rev1	CAGGTAACCGCCGTAGGACAGC	
BP0398fw1	TTCTTCGTCCACCAGCATTTCG	
BP0398rev1	AGCTGCAGGTCGAACGGATAGG	
BP0397-RTfw	CGGAACATTCCGGACCTTACCC	
BP0397-RTrev	CCACCAGCACGTGCATCAGC	
BP0396-RTfw	GCGCGCAGCTTGTCATAGGC	
BP0396-RTrev	CAAACCCCCAAAACCATCAAGG	
BPlgmDlike-RTfw	CCGTCGCGGTCGCCTGCGTG	
BPlgmDlike-RTrev	TGACGGAACGCCCACAGGCG	
vag8fw1	CCCCAAGCTTCGTCCGAGCACGGTATCAACG	*Hind*III
vag8rev1	CGCTCTAGACACATAGATCCCGGCGACTTCC	*Xba*I
lgmAfw4	CCATCGATTTCGCCGTGGTGTGTTCATG	*Cla*I
lgmArev3	GCTCTAGATCATTGGGAACGCGCCTTG	*Xba*I
lgmAfw3-NdeI	GGGAGTCATATGTGTTCATGTATACCGAATTCCG	*Nde*I
lgmArev3-XhoI	CCGCTCGAGACCACCACCTTGGGAACGCGCCTTGGC	*Xho*I
lgmCfw3	GGAATTCCATGACCAGTGAACGATACGA	*Eco*RI
lgmCrev3	GCTCTAGATCATAAACGGCTTGCCAG	*Xba*I
BPlpxAfw2	CCCCAAGCTTCTGCCGCATCGCTACCCGA	*Hind*III
BPlpxArev1	GCTCTAGACCGGCGACCATGCCTATG	*Xba*I
lpxAECfw1	GGAATTCGGCCTGATACGTGATTGATAAAT	*Eco*RI
lpxAECrev1	GCTCTAGACGCTGTTCAGTCATTAACGA	*Xba*I
BPlpxAfw3	GCTCTAGACGATCCGTAGCCTGGAAGA	*Xba*I
BPlpxArev3	GGAATTCCGGCGACCATGCCTATG	*Eco*RI
lpxAECfw1	GGAATTCGGCCTGATACGTGATTGATAAAT	*Eco*RI
lpxAECrev1	GCTCTAGACGCTGTTCAGTCATTAACGA	*Xba*I
BPlpxAfw3	GCTCTAGACGATCCGTAGCCTGGAAGA	*Xba*I
BPlpxArev3	GGAATTCCGGCGACCATGCCTATG	*Eco*RI

2.3 Cloning of vectors and deletion strains

2.3.1 General cloning techniques

All restriction endonuclease enzymes and other cloning-related enzymes were from New England Biolabs, unless otherwise indicated. The cloned genes in all vectors were confirmed by sequencing and PCR amplification. Plasmids were introduced into *E. coli* and *B. pertussis* strains using the following techniques.

A. $CaCl_2$-heat shock transformation

To prepare transformation-competent *E. coli* DH5α, S17-1, BL-21, or R0138 samples, 3 ml LB broth with no antibiotics was inoculated with *E. coli* and grown overnight while shaking at 200 rotations per minute (rpm) at 37°C (30°C for R0138). 1 ml of this overnight culture was used to inoculate 100 ml LB broth and grown to an OD_{600} of 0.4 to 0.5 at 37°C (30°C for R0138) and shaking at 200 rpm. These cells were harvested by centrifugation at 2100 g at 4°C for 10 min and immediately resuspended in 10 ml ice cold, sterile 50 mM $CaCl_2$ and incubated on ice for 10 min. The cells were harvested again and resuspended in 2 ml ice cold, sterile 50 mM $CaCl_2$ + 15% glycerol (w/v). Aliquotes of 50 µl were dispensed per Epindorf tubes and then stored at -80°C for storage until use for transformation.

To introduce a vector into *E. coli* DH5α, S17-1, BL-21, or R0138 cells: a 50 µl aliquot of transformation-competent cells was thawed on ice, then 1 µl of vector or a 20 µl ligation reaction was added to these cells, mixed and incubated on ice for 30 min. The mixture was then heat-shocked at 42°C for 45 s followed by a recovery period of 2 min on ice. 1 ml of LB broth was added to these cells and they were incubated at 37°C (30°C for R0138 strains) and shaken at 200 rpm for 1 h. Cells were then plated onto LB agar containing the required selection antibiotics and incubated at 37°C (30°C for R0138 strains) for approximately 16 h (24-30 h for R0138 strains) or until colonies were visible.

B. Electroporation

This protocol was used to introduce vectors into *B. pertussis* strains. To prepare electrocompetent bacterial samples: *B. pertussis* was first grown on BG agar containing the appropriate antibiotics for three days at 37°C. These bacteria were used to inoculate 300 ml SS broth supplemented with the appropriate antibiotics at an initial OD_{600} of 0.01 and grown at 37°C shaking at 180 rpm until an OD_{600} of 0.8 (approximately 48 hours). Bacteria were harvested by centrifugation at 11 350 g for 10 min at 4°C followed by two washes with 150 ml sterile deionized water (dH_2O) and one wash with 150 ml sterile 272 mM sucrose + 15% glycerol (w/v). The bacteria were then resuspended in 6 ml sterile 272 mM sucrose + 15% glycerol (w/v) and 400 µl aliquots were stored at -80°C for storage until use.

Introduction of a vector into an electrocompetent bacterial sample was accomplished as follows: 400 µl electrocompetent cells were thawed on ice and added to a 0.2 cm electrode gap electroporation cuvette (VWR). One µg of vector was added to the cuvette and the mixture was pulsed at 2.5 kV with a GenePulser Xcell (BioRad). The vector-bacteria mixture was immediately transferred into 1 ml SS broth and then incubated at 37°C shaking at 180 rpm for 1 h. The bacterial cells were then pelleted by centrifugation at 5 160 g for 5 min, then resuspended in 100 µl SS broth and plated onto BG agar containing the appropriate selection antibiotics. The BG agar plates were incubated at 37°C for 3-5 days, until colonies were visible.

C. Conjugation

A diparental mating protocol was used to introduce vectors into *B. pertussis* strains, as previously described (Marr, Hajjar, et al. 2010). First, the vector was introduced into *E. coli* S17-1 donor strain via $CaCl_2$-heatshock transformation. The *B. pertussis* acceptor strain was grown on BG agar containing the appropriate antibiotics for 3 days at 37°C and the *E. coli* S17-1 donor strain containing the vector was grown on LB agar supplemented with the appropriate antibiotics overnight at 37°C. *B. pertussis* cells

were then resuspended directly from the BG agar plate into SS broth and the *E. coli* cells were similarly resuspended in LB broth. The equivalent of 1 ml of *B. pertussis* at an OD_{600} of 1.0 and the equivalent of 100 µl of *E. coli* at an OD_{600} of 1.0 were mixed together and poured onto a mating plate (Marr, Hajjar, et al. 2010) and incubated at 37°C for 5 to 7 h. Bacterial cells were then removed from the mating plate with a sterile swab and streaked onto BG agar containing the appropriate selection antibiotics for both the parental *B. pertussis* strain and the vector. BG plates were incubated at 37°C for 3 to 5 days, until colonies were visible.

2.3.2 Generating markerless deletion mutants

The following general protocol was used to generate markerless deletion mutants in *B. pertussis* strains, as summarized in Figure 9. I will use the generation of BP338LgmAKO as an example, but a similar method was used for all other markerless deletion mutants. Approximately 300 to 500 base pairs of the up and downstream regions of the gene targeted for deletion, *lgmA*, were cloned into the vector pSS4245 to generate pSS4245LgmAKO, such that the upstream nucleotide (nt) region was directly joined to the downstream nt region by the cut site *Pac*I. This vector was then introduced into *E. coli* strain S17-1 via $CaCl_2$-heat shock transformation (Section 2.3.1). pSS4245LgmAKO was then introduced into the *B. pertussis* strain BP338 by conjugation (Section 2.3.1) and plated onto Bordet-Gengou (BG) agar supplemented with the appropriate selection antibiotics (i.e. naladixic acid, to select for BP338, and streptomycin, to select for pSS4245LgmAKO) and with 50 mM $MgSO_4$. The presence of 50 mM $MgSO_4$ in the growth medium suppresses the Bvg system, therefore the restriction endonuclease I-*Sce*I is not expressed from pSS4245LmgAKO. Since pSS4245 cannot replicate in *Bordetella* species, this vector must integrate into the chromosome via homologous recombination. It will likely integrate at either the upstream or downstream region of *lgmA*, thus introducing the entire vector into this region. The BG agar plates were incubated at 37°C for 3 to 5 days, until colonies were visible, and these colonies were then streaked onto a BG agar plate supplemented with only nalidixic acid. Since $MgSO_4$ is not present in this growth medium, the Bvg system of *B. pertussis* would be active, and I-*Sce*I, which is under the control of

a Bvg promoter, would now be expressed. I-*Sce*I endonuclease would cleave the I-*Sce*I restriction enzyme cut site present in pSS4245, therefore resulting in a double-stranded break in the chromosomal DNA at this site. The bacteria would attempt to repair this DNA break via homologous recombination, which would result in either removal of the originally inserted vector, pSS4245LgmAKO, or removal of the entire vector plus *lgmA*, resulting in a markerless deletion mutant of *lgmA*. Since the BG agar no longer contains streptomycin, there is no selection pressure to maintain the pSS4245 vector in the chromosome, and the resulting colonies were screened for kanamycin susceptibility, to indicate a loss of the pSS4245 vector, and by PCR for the generation of a markerless *lgmA* mutant. The lack of the targeted gene at the specific site in the chromosome was confirmed in all markerless deletion mutants by PCR amplification of the targeted chromosomal DNA region and sequencing of this product.

To generate the markerless deletion mutant, BP338LgmAKO, the *lgmA* intergenic upstream and downstream regions were PCR amplified from genomic DNA of *B. pertussis* strain BP338 using the primer sets (lgmAKO1fw and lgmAKO1rev) and (lgmAKO2fw and lgmAKO2rev), respectively. These regions were cloned into the suicide vector pSS4245 using *Eco*RI, *Pac*I, and *Avr*II, such that the *lgmA* intergenic upstream region is directly joined to the *lgmA* intergenic downstream region via a *Pac*I restriction enzyme cut site sequence. This resulted in the generation of pSS4245LgmAKO. As described above, pSS4245LgmAKO was used to generate the markerless deletion mutant BP338LgmAKO such that the *lgmA* gene was replaced by a *Pac*I restriction enzyme cut site sequence. A similar method was used to generate all markerless deletion mutants in this thesis, and the specific primer sets and parental strains are summarized in Table 4.

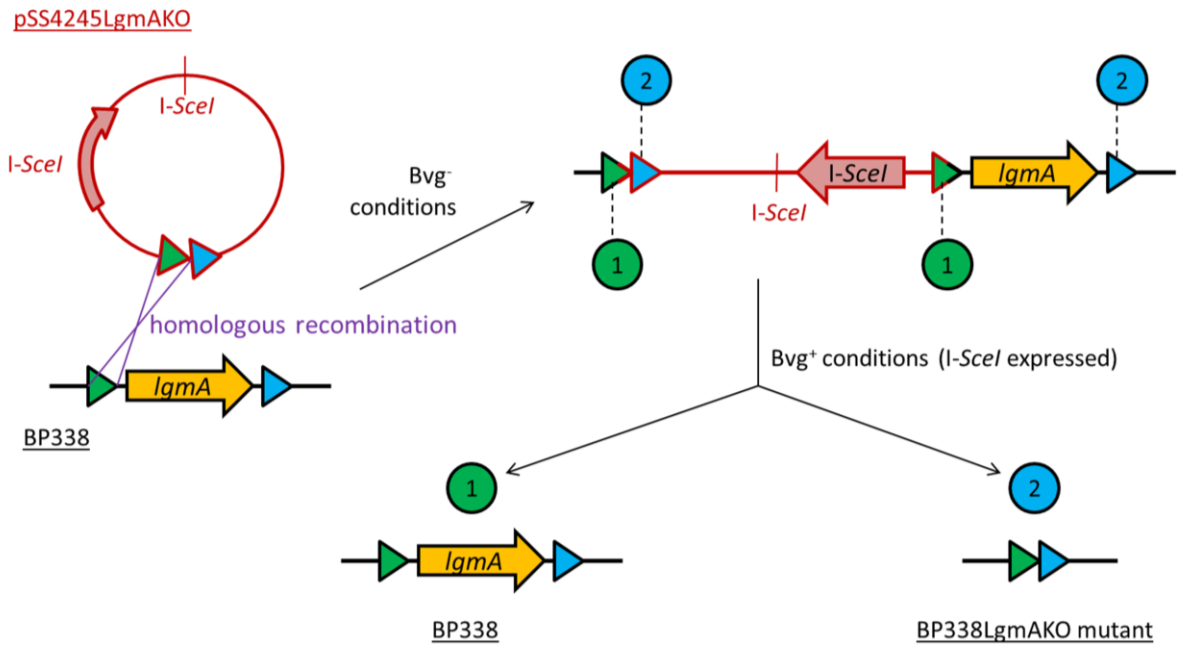


Figure 9. Schematic for the generation of markerless deletion mutants.

In this schematic, the generation of BP338LgmAKO, the *lgmA* clean deletion mutant of *B. pertussis* strain BP338 is used as an example. Vector and strain names are underlined. The pSS4245 vector contains an I-*Sce*I enzyme cleavage site and the I-*Sce*I gene (pink arrow) under the control of the pertussis toxin promoter, which induces expression under Bvg^+ conditions. The pSS4245LgmAKO vector (red lines) contains the region upstream of *lgmA* (green triangle) and the region downstream of *lgmA* (blue triangle). *B. pertussis* strain BP338 chromosomal DNA is depicted with a black line. In the first step, pSS4245LgmAKO was introduced into BP338 under Bvg^- conditions and, via homologous recombination, the vector inserted into the BP338 chromosome (in this case, homologous recombination via the region upstream of *lgmA* is depicted). This strain was then grown under Bvg^+ conditions, therefore inducing the expression of the I-*Sce*I restriction enzyme. I-*Sce*I would cleaved the chromosomal DNA at the I-*Sce*I cleavage site encoded by pSS4245LgmAKO, therefore promoting homologous recombination in this region in an effort to repair the double stranded break in the DNA. Two homologous combination events are probable: 1) homologous recombination via the upstream region, which leads to the removal of the vector, leaving the original BP338 *lgmA* gene intact, or 2) homologous recombination via the downstream region, which leads to the removal of the vector and the *lgmA* gene, resulting in the generation of a markerless *lgmA* mutant of BP338.

Table 4. Summary of markerless deletion mutant cloning

Deletion mutant	Parental strain	Suicide plasmid	Upstream primer set	Downstream primer set
BP338LgmAKO	BP338	pSS4245LgmAKO	lgmAKO1fw lgmAKO1rev	lgmAKO2fw lgmAKO2rev
BP338LgmBKO	BP338	pSS4245LgmBKO	lgmBKO1fw lgmBKO1rev	lgmBKO2fw lgmBKO2rev
BP338LgmCKO	BP338	pSS4245LgmCKO	lgmCKO1fw lgmCKO1rev	lgmCKO2fw lgmCKO2rev
BP338LgmDKO	BP338	pSS4245LgmDKO	lgmDKO1fw lgmDKO1rev	lgmDKO2fw lgmDKO2rev
BP338LgmABCDKO	BP338	pSS4245LgmABCDKO	lgmAKO1fw lgmAKO1rev	lgmDKO2fw lgmDKO2rev
BP338BP1945KO	BP338	pSS4245BP1945KO	BP1945KO1fw BP1945KO1rev	BP1945KO2fw BP1945KO2rev
BP338LgmDKOBP1945KO	BP338LgmDKO	pSS4245BP1945KO	BP1945KO1fw BP1945KO1rev	BP1945KO2fw BP1945KO2rev

2.3.3 Site-directed mutagenesis

The following general protocol was used to generate site-directed mutants. I will use pBBR2LgmA D127G as an example, though all site-directed mutants were generated using a similar procedure. The non-mutated original vector pBBR2LgmA was used as the template in a PCR reaction with the primers lgmAD127Gfw and lgmAD127Grev. The 60 µl PCR mixture contained: 12 µl 5X Q5® reaction buffer (New England Biolabs), 12 µl 5X GC enhancer (New England Biolabs), 1.2 µl 10 mM dNTP (BioBasic), 3 µl template, 3 µl of each primer at 10 uM, 0.6 µl Q5® High-Fidelity DNA Polymerase, dH_2O for the remaining reaction volume. The PCR reaction was as follows: 5 min at 95°C, 18 cycles of (50 s at 95°C, 50 s at 60 to 67°C, 8 min at 72°C), 10 min at 72°C. The reaction was treated with *Dpn*I (addition of 1 unit/µl *Dpn*I and incubation at 37°C for 1 h) to remove the original methylated pBBR2LgmA vector followed by inactivation of *Dpn*I by incubation at 80°C for 20 min. This reaction was introduced into *E. coli* DH5α cells by $CaCl_2$-heatshock transformation (Section 2.3.1) and plated onto LB agar containing kanamycin to select for the plasmid. Transformant colonies were screened for the presence of the site-directed mutation by sequencing. Table 5 summarizes the primer sets used to generate site-directed mutants with the previously described method.

To clone pPtacLgmABCD-E G15STOP, pPtacLgmABCD and pBBR2LgmC-E G15STOP vectors were digested with *NotI* and *AatII*, and the fragment containing the G15STOP mutation from pBBR2LgmC-E G15STOP was ligated to the pPtacLgmABCD vector backbone, to transfer the *lgmE* G15STOP mutation to this vector. Similarly, the *lgmE* G15STOP mutation was cut out of pPtacLgmABCD-E G15STOP and ligated into pPtacLgmABC with enzymes *DraIII* and *DrdI*, to generate pPtacLgmABC-E G15STOP.

Table 5. Summary of site-directed mutant primer sets

Site-directed mutant vector	Template vector	Primer set
pBBR2LgmA D76G D77G	pBBR2LgmA	lgmAD76GD77Gfw, lgmAD76GD77Grev
pBBR2LgmA D127G	pBBR2LgmA	lgmAD127Gfw, lgmAD127Grev
pBBR2LgmA D129G	pBBR2LgmA	lgmAD129Gfw, lgmAD129Grev
pBBR2LgmA D127G D129G	pBBR2LgmA	lgmAD127GD129Gfw, lgmAD127GD129Grev
pBBR2LgmA D159N	pBBR2LgmA	lgmAD159Nfw, lgmAD159Nrev
pBBR2LgmA W163R	pBBR2LgmA	lgmAW163Rfw, lgmAW163Rrev
pBBR2LgmA D159N W163R	pBBR2LgmA	lgmAD159NW163Rfw, lgmAD159NW163Rrev
pBBR2LgmC D80G D81G	pBBR2LgmC	lgmCD80GD81Gfw, lgmCD80GD81Grev
pBBR2LgmC H130G	pBBR2LgmC	lgmCH130Gfw, lgmCH130Grev
pBBR2LgmC D187G H189G	pBBR2LgmC	lgmCD187GD189Gfw, lgmCD187GD189Grev
pBBR2LgmC E313G	pBBR2LgmC	lgmCE313Gfw, lgmCE313Grev
pBBR2LgmC-E G15STOP	pBBR2LgmC	BPlgmDlikeG15STOPfw, BPlgmDlikeG15STOPrev

2.3.4 Vectors to complement *B. pertussis* strains

I complemented the *lgmA*, *lgmB*, and *lgmC* knockout mutants in *B. pertussis* strain BP338 (BP338LgmAKO, BP338LgmBKO, and BP338LgmCKO, respectively) with vectors expressing the respective wild-type gene.

To generate the *lgmA* and *lgmC* complementing vectors, I first generated pBBR2Pcpn, a vector that encoded a constitutively-expressing promoter, Pcpn. The Pcpn heat shock promoter was cut out of pBBRPcpnBrkA (Marr, Shah, et al. 2011) using the *Kpn*I sites, and cloned into the *Kpn*I site of pBBR1MCS2 in the same direction as the kanamycin resistance cassette. Then I cloned *lgmA* or *lgmC* into this vector such that the expression of these genes was driven by the Pcpn promoter. The *B. pertussis* BP338 *lgmA* gene was PCR amplified with primers lgmAfw4 and lgmArev3, and the product was cloned into the *Cla*I and *Xba*I sites of pBBR2Pcpn to generate the vector pBBR2LgmA. To construct pBBR2LgmC, I PCR amplified *lgmC* from BP338 with primers lgmCfw3 and lgmCrev3, and then cloned this PCR product into the *Eco*RI and *Xba*I sites of pBBR2Pcpn.

To complement BP338LgmBKO, the region containing *lgmA* and *lgmB* was cloned from pNMLgmAB into pMMB67HE using *Hind*III and *Xba*I to generate the vector pPtacLgmAB, in which *lgmA* and *lgmB* are under the control of the IPTG-inducible promoter, Ptac.

The vector pPtacLgmABC was used to complement the full *lgm* locus mutant BP338LgmABCDKO. The region containing *lgmA*, *lgmB*, and *lgmC* was cut out of pPtacLgmABCD with *Xho*I and cloned into pMMB67HE cut with *Sal*I, since *Xho*I and *Sal*I have identical overhanging regions.

I also complemented *B. pertussis* wild type strain 18-323 with vectors encoding lipid A-modifying genes from strain BP338: *lpxA*, the *lgm* locus, or both. First I cloned the vector pBBRLpxA338. *lpxA* and the intergenic upstream region of *lpxA*, which presumably contains the endogenous *lpxA* promoter, was PCR amplified from *B. pertussis* BP338 chromosomal DNA with primers BP338lpxAfw2-HindIII and BP338lpxArev1-XbaI. The PCR product was cloned into the *Xba*I and *Hind*III sites of pBBR1MCS2. Then, *lpxA* from *B. pertussis* BP338 and the intergenic upstream region of *lpxA* were cut out of pBBRLpxA338 with *Hind*III and *Xba*I, and the ends were blunted with Klenow. pPtacLgmABCD was cut with *Eco*RI to cut out the LgmABCD locus, and the ends of the vector backbone were blunted with Klenow, into which the DNA fragment containing *lpxA* from BP338 and the intergenic upstream region of *lpxA* were cloned. A clone was selected in which the *lpxA* gene is in the forward direction in comparison to the pPtac promoter, resulting in the vector pPtacLpxA338.

To complement 18-323 with both *lpxA* and the *lgm* locus from *B. pertussis* strain BP338, I generated the vector pPtacLgmABCDLpxA338. *lpxA* from BP338 and the intergenic upstream region of lpxA were cut out of pBBRLpxA338 with *Hind*III and *Xba*I, and the ends were blunted with Klenow. pPtacLgmABCD was cut with *Sap*I, resulting in a 13.04 kb linear vector. The ends of this vector were blunted with Kelnow, and pPtacLgmABCD was ligated to the blunt-ended fragment of *lpxA* from BP338 and the

intergenic upstream region. The resulting vector contains the *lgmABCD* locus under the control of the Ptac promoter, and the *lpxA* gene from BP338 under the control of its endogenous promoter.

2.3.5 Vectors to complement *E. coli* strain R0138

E. coli strain R0138 encodes a temperature-sensitive *lpxA* gene (*lpxA2*), and was complemented with vectors encoding *lpxA* from *E. coli* strain DH5α, *B. pertussis* strain BP338, or *B. pertussis* strain 18-323. I constructed the vector pBSLpxAEC, by PCR amplifying *lpxA* from *E. coli* DH5α chromosomal DNA with primers lpxAECfw1 and lpxAECrev1 and cloning this PCR product into pBlueScript II KS (-) using *Xba*I and *Eco*RI. This resulted in *lpxA* from *E. coli* oriented in the forward direction in reference to the *lac* promoter.

I constructed two vectors with *lpxA* from *B. pertussis* strain BP338: pBSLpxA338 and pBSLpxA338rev. To clone pBSLpxA338, the *lpxA* gene from *B. pertussis* strain BP338 was cut out of pPtacLpxA338 and cloned into pBlueScript II KS (-) using *Eco*RI and *Xba*I, resulting in *lpxA* from BP338 in the forward orientation in reference to the *lac* promoter. In pBSLpxA338rev, however, the BP338 *lpxA* gene was cloned in the reverse orientation in reference to the *lac* promoter by cutting *lpxA* out of pPtacLpxA338 and cloning it into pBlueScript II KS (-) using *Eco*RI and *Sal*I.

I also cloned the two vectors pBSLpxA18323rev and pBSLpxA18323, which both encode the *lpxA* gene from *B. pertussis* strain 18-323. First, I PCR amplified 18-323 *lpxA* from chromosomal DNA with primers BPlpxAfw3 and BPlpxArev3 and cloned this PCR product into pBlueScript II KS (-) with *Eco*RI and *Xba*I, such that *lpxA* from 18-323 is in the reverse orientation in reference to the *lac* promoter. This generated the vector pBSLpxA18323rev. Next, I cloned pBSLpxA18323 by excising an internal region of *lpxA* of 18-323 (that contains the single base pair difference when compared to *lpxA* of BP338) from pBSLpxA18323rev with *Asc*I and *Sph*I and cloning it into pBSLpxA338, which was also cut by *Asc*I and

*Sph*I remove the BP338 *lpxA* internal region. This resulted in *lpxA* of 18-323 in the forward orientation in reference to the *lac* promoter.

2.3.6 Vector for LgmA expression in *E. coli* strain BL-21

pET30LgmA was used to express LgmA in *E. coli* strain BL-21(DE3). First, *lgmA* was PCR amplified with primers lgmAfw3-NdeI and lgmArev3-XhoI to generate a product of *lgmA* with a C-terminal 3x glycine linker. Then, both this PCR product and the pET30b vector were digested with *Nde*I and *Xho*I and ligated together to generate a construct with *lgmA* connected to a C-terminal 6x histidine (His) tag by a 3x glycine linker.

2.4 Preparation of bacterial cells for mass spectrometry analysis and TLR4-activation assays

B. pertussis strains were grown on BG agar at 37 °C for 3 to 4 days, and these cells were used to inoculate SS broth, with the appropriate antibiotics, at an OD_{600} of 0.01. Cultures were grown at 180 rpm at 37 °C until an OD_{600} of 0.6 to 0.9, and samples from each culture were grown on BG agar to confirm a hemolytic phenotype. Bacterial cells were harvested into phosphate-buffered saline (PBS) pH = 7.4 to an OD_{600} of 5 and heat-inactivated by incubation at 56 °C for 1 h. To confirm heat-killing, 50 µl of each cell suspension was spotted onto BG agar supplemented with no antibiotics and incubated at 37 °C for 5 days. Heat-killed cells were stored at -20 °C. These cells were used for stimulation in a TLR4 activation assay, generation of highly purified LPS for use as the stimulus in the TLR4 activation assays, and direct lipid A isolation for use in matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) analysis.

For *E. coli* R0138 strains, a similar protocol was used to prepare the bacterial cells, with the following exceptions: *E. coli* R0138 strains were grown on LB agar at 30°C for 48 h, after which a single colony was used to inoculate LB broth supplemented with 1 mM IPTG. This culture was grown for 16-20 h, shaking at 200 rpm at 30°C for R0138 wild type strain and at 42°C for the complemented strains (R0138 + pBSlpxAEC, R0138 + pBSlpxA338, and R0138 + pBSlpxA18323). *E. coli* cells were heat-killed at

80°C for 1 h, and to confirm the bacteria were killed, 50 µl of each cell suspension was spotted onto LB agar supplemented with no antibiotics and incubated at 30 °C for 3 days. These *E. coli* R0138 cells were used for direct isolation of lipid A for use in MALDI-MS analysis.

2.5 Isolation of LPS and lipid A

2.5.1 Isolation of *B. pertussis* LPS for TLR4 activation assays

B. pertussis LPS preparations were extracted by an ammonium hydroxide-isobutyric acid method by members of the Caroff lab (M. Caroff Patent WO 2004/062690 A1, March 20, 2012). Primary extracts were subjected to a standard enzyme treatment (DNase, RNase, and proteinase K) and finally repurified with the acidified chloroform-methanol-water procedure as described (Tirsoaga, Novikov, et al. 2007). To be sure that no specific LPS molecular species were discriminated against during the process, all intermediate and final products were analyzed by MALDI-MS. High purity of the resulting LPS preparations was evidenced by three different methods. The absence of contaminating (non-LPS) peaks was attested by positive-ion MALDI mass spectra analysis; the absence of detectable protein contaminants was demonstrated by Tricine-SDS-PAGE (Schagger and von Jagow 1987) and silver staining (Tsai and Frasch 1982) loading up to 250 ng of LPS from each preparation; and the absence of lipoprotein content was further demonstrated based on the lack of detectable levels of cysteine by analysis with an amino acid analyzer (Hitachi L-8800, equipped with a 2620MSC-PS column, ScienceTec, Les Ulis, France).

2.5.2 Isolation of *E. coli* LPS for TLR4 and LAL activation assays

The protocol was adapted from Rezania *et al.* (Rezania, Amirmozaffari, et al. 2011). *E. coli* R0138 strains were grown on LB agar at 30°C for 48 hours. 50 ml LB broth supplemented with 1 mM IPTG was inoculated with freshly grown colonies and incubated at 42°C at 200 rpm for 16-20 h. Cells were harvested by centrifugation at 10 000 g for 5 min and washed twice with 0.15 M PBS (pH = 7.2) solution. Cells were then resuspened in 10 ml 0.15 M PBS (pH = 7.2), sonicated on ice for 10 min, and treated with

proteinase K (100 µg/ml proteinase K incubated at 65°C for 1 h) and then with RNase and DNase (40 µg/ml RNase A, 20 µg/ml DNase I, 0.02% MgSO4, and 0.004% chloroform incubated at 37°C for 16-20 h). LPS was extracted using a hot phenol extraction method. Briefly, 10 ml 90% phenol at 65-70°C was added to 10 ml of the cell mixture and incubated for 15 min at 65-70°C, then mixed in an ice bath for 15 min. The phases were separated by centrifugation at 8 500 g for 15 min and the aqueous layer was re-extracted using this hot phenol method. Sodium acetate was added to the final aqueous phase to a final concentration of 0.5 M and 10 volumes of ethanol were added to this mixture and stored at -20°C for 16-20 h to precipitate the LPS. This mixture was centrifuged at 2 000 g at 4°C for 10 min, and the remaining LPS pellet was resuspended in water and dialyzed against water at 4°C for 16-20 h to remove any residual phenol. The LPS sample was then lyophilized and resuspended in DMSO to a final concentration of 1 mg/ml. This purified *E. coli* LPS was used for HEK-Blue hTLR4 activation assays and LAL chromogenic assays.

2.5.3 Isolation of lipid A for mass spectrometry analysis

For MALDI-MS analysis, lipid A was isolated directly by hydrolysis of bacterial cells as described previously by members of the Caroff lab (El Hamidi, Tirsoaga, et al. 2005, Tirsoaga, El Hamidi, et al. 2007). Briefly, lyophilized bacterial cells (10 mg) were suspended in 200 µl of a isobutyric acid:1 M ammonium hydroxide (5:3, v/v) mixture and were kept for 2 h at 100°C in a screw-cap test tube under magnetic stirring. The suspension was cooled in ice water and centrifuged at 2000 g for 10 min. The recovered supernatant was diluted with 2 volumes of water and lyophilized. The sample was then washed once with 200 µl of methanol by centrifugation at 2000 g for 10 min. Finally, lipid A was extracted from the pellet in 100 µl of a chloroform:methanol:water (3:1.5:0.25, v/v) mixture. In some of the spectra, peaks corresponding to small contaminants were identified and marked with an "X" (peaks at *m/z* 1349 and 1377 in Figure 34).

2.6 Mass spectrometry analysis

Mass spectrometry was performed in the Caroff lab. LPS samples were dispersed in water at 1 µg/µl. Lipid A extracts in chloroform-methanol-water were used directly. In both cases, a few microliters of sample solution were desalted with a few grains of ion-exchange resin Dowex 50W-X8 (H^+). 0.5 to 1 µl aliquots of the solution were deposited on the target, covered with matrix solution, and allowed to dry. Dihydroxybenzoic acid (Sigma-Aldrich) was used as matrix. It was dissolved at 10 mg/ml in 0.1 M citric acid solution in the same solvents as those used for the analytes (Therisod, Labas, et al. 2001). Different analyte/matrix ratios (1:2, 1:1, 2:1, v/v) were tested to obtain the best spectra. Analyses were performed on a PerSeptive Voyager-DE STR time-of-flight mass spectrometer (Applied Biosystems) in linear mode, with delayed extraction. Negative- and positive-ion mass spectra were recorded. The ion-accelerating voltage was set at -20 kV, and the extraction delay time was adjusted to obtain the best resolution and signal-to-noise ratio.

2.7 HEK-Blue hTLR4 activation assay

2.7.1 Maintenance of HEK-Blue cells

HEK-Blue (InvivoGen) cell lines hTLR4 and Null2 cells were maintained in complete Dulbecco's modified Eagle's medium (DMEM) containing 10% heat-inactivated fetal calf serum, 100 units/ml penicillin, 100 µg/ml streptomycin (Gibco, Life Technologies), 2 mM GlutaMAX, 1 mM pyruvate (Life Technologies), and 100 µg/ml Normocin (InvivoGen). Both the Null2 and the hTLR4 cell lines express secreted alkaline phosphatase under the control of an NFκB promoter. In addition, the hTLR4 cell line is stably transfected with human TLR4, MD-2, and CD14 co-receptor genes, such that hTLR4 activation leads to transcription from the NFκB promoter and expression of secreted alkaline phosphatase. Null2 cells were grown in the presence of 100 µg/ml Zeocin (InvivoGen), and hTLR4 cells were grown in the presence of 100 µg/ml Zeocin, 200 µg/ml Hygrogold, and 30 µg/ml Blasticidin (InvivoGen). Cells were incubated at 37°C in humid air with 5% CO_2.

2.7.2 hTLR4 activation assay

The HEK-Blue (Invivo-Gen) manufacturer's guidelines were followed to assay hTLR4 activation by purified LPS or heat-killed bacterial samples. Briefly, HEK-Blue hTLR4 and Null2 were seeded in 96-well plates at 25 000 or 50 000 cells/well, 100 µl/well. All media used in this assay were pyrogen-free and consisted of complete DMEM, with the absence of the following antibiotics: Normocin, Zeocin, Hygrogold, and Blasticidin. HEK-Blue cells were incubated at 37°C (as described in Section 2.7.1) and at 24 h, 100 µl of fresh medium was added to each well, and the cells were incubated for another 24 h. HEK-Blue cells were then washed with 100 µl of medium and stimulated with 100 ng/ml purified *B. pertussis* LPS, 0.1 ng/ml purified *E. coli* LPS, or heat-killed *B. pertussis* cells (prepared as described in Section 2.4) at a 1/10 dilution. After 24 h of incubation, the supernatants were assayed for secreted alkaline phosphatase activity by mixing 20 µl of supernatant with 180 µl of QUANTI-Blue (InvivoGen) reagent in a 96-well plate and incubating this mixture at 37°C in the dark until the color of the mixture started turning blue. Absorbance at 650 nm was used as a read out of alkaline phosphatase activity, which indicates NFκB activation via hTLR4. In each assay, the Null2 cell line and unstimulated hTLR4 cells were used as negative controls.

2.8 *E. coli* growth curves

E. coli R0138 strains were grown on LB agar with appropriate antibiotics for 48 h at 30° C, and growth from these agar plates was used to inoculate LB broth with the appropriate antibiotics and grown shaking at 200 rpm at 30°C for 16 to 20 h. These cultures were used to inoculate fresh LB broth supplemented with the appropriate antibiotics at an initial OD_{600} of 0.1, and all complemented strains were also supplemented with 1 mM IPTG (R0138 + pBSlpxAEC, R0138 + pBSlpxA338, and R0138 + pBSlpxA18323). Cultures were grown at 30°C or 42°C shaking at 200 rpm, and growth was monitored by OD_{600} measurements over time.

2.9 Bacterial survival assays

2.9.1 Polymyxin B growth curve assay

E. coli R0138 strains were grown on LB agar with the appropriate antibiotics for 48 h at 30°C and these freshly grown bacterial were used to inoculate LB broth supplemented with the appropriate antibiotics and IPTG and grown at 42°C and shaking at 200 rpm for 16 to 20 h. Bacterial cultures were then diluted to an OD_{600} of 0.1 and incubated at 42°C and shaking at 200 rpm in the presence of polymyxin B concentrations: 0, 0.5, or 1.0 µg/ml. Growth was monitored via OD_{600} measurements over time.

2.9.2 Killing assays

Protocol was adapted from Wiegand *et al* (Wiegand, Hilpert, et al. 2008). *E. coli* R0138 strains were grown on LB agar supplemented with appropriate antibiotics for 48 h at 30°C. LB broth with the appropriate antibiotics and IPTG was inoculated with freshly grown R0138 colonies and incubated at 42°C at 200 rpm for 16 to 20 h. 1 ml of this culture was then used to inoculate 4 ml of LB broth supplemented with the appropriate antibiotics and IPTG, and this was grown at 42°C at 200 rpm for 1 h. Cells were then diluted in LB broth to an OD_{600} of 0.1, and mixed with an equal volume of polymyxin B diluted in LB broth for final concentrations of: 0, 0.5, 1.0, or 2.0 µg/ml polymyxin B in polypropylene 96-well plates. After incubation in polymyxin B for 5 min, 100 µl of 1/100 and 1/1000 diluted samples were plated on LB agar plates with no antibiotics and grown at 30°C for 48 h, after which the colony forming units (cfu) were determined to calculate percent survival. n = 3 for each condition (i.e. each strain at a specific polymyxin B concentration).

For *B. pertussis* assays, *B. pertussis* strains were grown on BG agar with no antibiotics at 37°C for 3 days, after which their hemolytic phenotype was confirmed. These cells were used to inoculate 20 ml SS broth supplemented with nalidixic acid at an initial OD_{600} of 0.01. Cultures were grown at 37°C while shaking at 180 rpm for 24 to 30 h. These cultures were then diluted into SS salts (10.72 g glutamic acid, 0.24 g proline, 2.50 g NaCl, 0.50 g KH_2PO_4, 0.20 g KCl, 0.10 g $MgCl_2$-$6H_2O$, 0.02 g $CaCl_2$, 3.175 g Tris-HCl,

0.59 g Tris base dissolved in 1L distilled deionized water, pH 7.6) to an OD_{600} of 0.002. The killing agent (peptide, antibiotic, lysozyme, and/or ethylenediaminetetraacetic acid (EDTA)) was diluted from the stock solutions into SS salts. 50 µl of the diluted bacterial culture was mixed with 50 µl of the diluted killing agent in polypropylene 96-well plates and incubated at 37°C for 2 h. A list of CAMP and antibiotics used as killing agents in this assay is in Table **6**.

For the *B. pertussis* spot assays, after the 2 h incubation, the bacteria and killing agent mixture was serially diluted in SS salts and 2 µl of each 10 fold dilution was spotted onto BG agar plates with no antibiotics and incubated at 37°C for 72 h to visualize growth.

For the *B. pertussis* percent survival assays, following the 2 h incubation, 100 µl of a 1/100 and 1/1000 dilution of the bacteria and killing agent mixture was plated onto BG agar plates with no antibiotics. After incubation at 37°C for 72 h, the percent survival was determine from the number of cfu for the samples with the killing agent compared to bacterial samples incubated with no killing agents. n = 3 for each condition (i.e. killing agent at a specific concentration).

Table 6. List of CAMPs and antibiotics used in killing assays
The estimated net charges are at pH 7.0 and expressed to the nearest integer value.

CAMP or Antibiotic	Sequence	Net Charge	Structure	Class
Polymyxin B (Sigma-Aldrich)		+5	cyclic, lipidated	polymyxin
Polymyxin E (Sigma-Aldrich)		+5	cyclic, lipidated	polymyxin
LL-37	LLGDFFRKSKEKIGK EFKRIVQRIKDFLRNL VPRTES	+6	α helical	cathelicidin
Indolicidin	ILPWKWPWWPWRR-NH_2	+3	extended	cathelicidin
HHC-10	KRWWKWIRW-NH_2	+3	α helical	synthetic
CP28	KWKLFKKIGIGAVLK VLTTGLPALKLTK-NH_2	+7	α helical	Insect hybrid
Gentamicin (Sigma-Aldrich)		+5	trisaccharide	aminoglycoside

2.10 LAL activation assay

The LAL assay detects the presence of endotoxin, also known as LPS, in a sample through the interaction of LPS with Factor C from the Horse Shoe crab, *Limulus polyphemus*. The Pyrochrome LAL assay (Associates of Cape Cod Inc.) manufacturer's guidelines were followed to assay LAL activation by 10 μg/ml purified LPS from *E. coli* strains. Briefly, LPS samples were diluted to 10 μg/ml in Glucashield Reconstitution Buffer using pyrogen-free instruments and a Pyroplate 96-well microplate (Associates of Cape Cod Inc.). An equal volume of freshly reconstituted Pyrochrome reagent was added to each sample and mixed by pipetting the samples up and down two times, carefully, to ensure no bubbles were formed. The presence of LPS was detected by monitoring absorbance at A_{405} over time.

2.11 Glycosyltransferase assay

The protocol was adapted from Song *et al.* (Song, Guan, et al. 2009) and the butanol extraction protocol was adapted from Ravishankar *et al*. (Ravishankar, Kumar, et al. 2005). To demonstrate glycosyltransferase activity of LgmA, *E. coli* membranes were used as the source of both the assayed protein, LgmA, and the acceptor lipid and UDP-GlcN[^{14}C]Ac was the donor molecule. An initial 10 ml

culture was inoculated with *E. coli* BL-21 (DE3) strains and grown for 16 to 20 h. This 10 ml culture was then added to 90 ml LB and grown for an additional 5 to 7 h, at which point cells were harvested from 500 µl of culture for Western blot analysis. The bacteria were induced with a final concentration of 1 mM IPTG and grown for another 16 h, upon which 500 µl of induced bacterial cells were harvested for Western blot analysis. The full 100 ml culture was then harvested by centrifugation at 6 000 g for 15 min at 4°C. From this point onwards, the cells were placed on ice, or centrifuged or stored at 4°C, to prevent degradation of the cell components. Cells were washed in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer and then resuspended in 10 ml 50 mM HEPES buffer. Cells were then disrupted via 5 passages through a cell homogenizer, EmulsiFlex-05 (Avestin), at an average of 9 000 psi, followed by removal of unbroken cells by centrifugation at 10 000 g for 20 min. The membranes were collected from the resulting supernatants via centrifugation at 100 000 g for 1 h. The membrane pellet was then resuspended in 50 mM HEPES buffer and the total protein concentration was determined by the Bradford assay (Bradford 1976) using Protein Assay reagent (BioRad). Membranes were then washed in 50 mM HEPES buffer and resuspended in 50 mM HEPES buffer at a total protein concentration of 5 mg/ml. The crude membrane preparation was stored at 4°C and used within a week as the enzyme source in the glycosyltransferase assay.

To assay transfer of GlcNAc from UDP-GlcNAc onto a lipid, the following reaction was mixed: 4 µl crude membrane preparation at 5 mg/ml, 2 µl 0.4 mM UDP-GlcN[^{14}C]Ac (specific activity of 250 mCi/mmol) (American Radiolabeled Chemicals), 2 µl 1% Triton X-100, 2 µl 500 mM HEPES buffer (pH 7.5), 2 µl 5 mM $MgCl_2$. The remaining reaction volume was filled with dH_2O to a total volume of 20 µl. The mixed reaction was incubated at 30°C for 1 h, followed by extraction of the lipids into butanol (Ravishankar, Kumar, et al. 2005) by the addition of 100 µl of butanol and 100 µl of water. This mixture was then agitated for 5 s and centrifuged at 2 000 g for 1 min to separate the aqueous and organic phases. Fifty µl of the top organic phase, which contains the extracted lipids, was added to 3 ml scintillation fluid

in a scintillation vial, and the degradations per minute (dpm) of each sample was determined by a LS 6000IC Scintillation counter (Beckman Coulter) with the ^{14}C settings and a 5 min count setting. The presence of ^{14}C in the organic phase measures transfer of GlcN[^{14}C]Ac onto a lipid.

2.12 Western blot analysis

E. coli cells were harvested from liquid cultures by centrifugation and the resultant cell pellet was resuspended in sample buffer and boiled for 5 min. Samples were run on SDS-PAGE at 180 V for 55 min to separate the proteins followed by transfer of the proteins onto Immobilon-P membrane (Millipore) at 100 V for 60 min on ice (Oliver and Fernandez 2001). The remaining SDS-PAGE gel, after transfer, was stained with Coomassie Brilliant Blue to confirm equal levels of protein were present in comparable samples (Oliver and Fernandez 2001). The Immobilon-P membrane was then probed with rabbit anti-HisTag antiserum, His-Probe (G-18) (Santa Cruz Biotechnology), at 1:10 000 dilution, followed by probing with secondary antiserum horseradish peroxidase (HRP)-conjugated goat anti-rabbit (ICN Biomedicals) at 1:20 000 dilution (Oliver and Fernandez 2001). The HRP-conjugated antibody was detected with Western Lightning® ECL (Perkin Elmer) followed by exposure to X-OMAT LS film (Kodak) for 30 s, as per the manufacturers' guidelines. PageRuler Prestained protein ladder (Fermentas Life Sciences) was used to estimate molecular mass.

2.13 Whole genome sequencing

Library preparation and genome sequencing was performed in the Hirst laboratory. The sequencing data for *B. hinzii* ATCC 51730 and *B. trematum* CCUG 13902 were obtained from PCR-free random fragment libraries sequenced on the MiSeq (Illumina, Hayward, CA) platform using indexed paired-end 250-nucleotide (nt) v2 chemistry and resulted in ~700-fold coverage for each genome. The nonindexed read length was 250 nt, with 84.4% of the postfilter paired-end reads having Q30 or greater. The sequence reads were subsampled (~2.2 M reads) and assembled into contigs using Velvet (Zerbino and Birney

2008) with a k-mer of 151. A total of 1,850,984/2,212,976 reads were assembled for *B. hinzii* and 1,878,624/2,212,458 reads were assembled for *B. trematum*, resulting in 98 contigs for *B. hinzii* and 83 contigs for *B. trematum*. These whole-genome shotgun projects have been deposited in GenBank under accession no. AWNM00000000 (*B. hinzii*) and no. AWNL00000000 (*B. trematum*).

2.14 Reverse transcriptase PCR

RNA was extracted from freshly grown bacterial cells using the Quick-RNA MiniPrep kit (Zymo Research), following the manufacturer's directions. Any contaminating DNA was removed using the DNA-free kit (Ambion), as per the manufacturer's instructions. The DNA-free RNA sample was then tested for DNA contamination by PCR using 0.1 µg of the RNA sample as a template and chromosome-specific primers. If the RNA sample was confirmed to be DNA-free by PCR, 1 µg RNA was converted to cDNA using SuperScript II RNase H- Reverse Transcriptase (Invitrogen), in accordance with the manufacturer's instructions for first-strand cDNA synthesis, with Random Primers (Invitrogen) and the RNaseOUT reagent (Invitrogen). PCR of cDNA was performed with the following specifications: cDNA from the equivalent of 0.1 µg RNA was the template for each reaction, 2 µl 10X ThermoPol® buffer (New England Biolabs), 4 µl 5X GC enhancer (New England Biolabs), 0.4 µl 10 mM dNTP (BioBasic), 1 µl each primer at 10 µM, 0.2 µl Taq DNA Polymerase. The PCR reaction was as follows: 5 min at 94°C, 30 cycles of (45 s at 94°C, 45 s at 63°C, 1 min at 72°C), 5 min at 72°C. For each reverse transcriptase (RT)-PCR, the successful conversion into cDNA was tested with positive control primers (*vag8* primers).

The following primer pairs were used to detect transcription of genes in RT-PCR experiments (Table 3): *lgmA*, BP0399fw1 and BP0399rev1; *lgmB*, BP0398fw1 and BP0398rev1; *lgmC*, BP0397-RTfw1 and BP0397-RTrev1; *lgmD*, BP0396-RTfw1 and BP0396rev1; *lgmE*, BPlgmDlike-RTfw1 and BPlgmDlike-RTrev1; *vag8*, vag8fw1 and vag8rev1.

2.15 Bioinformatic analysis tools

Bioinformatic tools were used to analyze individual sequences, or to compare multiple sequences. For all these web-based bioinformatics tools, I used the default settings. To predict the protein family of a single peptide sequence based on protein sequence similarity, I used Pfam 27.0 (http://pfam.sanger.ac.uk/) (Punta, Coggill, et al. 2012). To analyze a protein sequence for predicted membrane helices, and analyze the predicted membrane topologies of proteins, I used the TMHMM server v 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) (Sonnhammer, von Heijne, et al. 1998). To compare multiple protein sequences and generate a multiple sequence alignment, I used ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Thompson, Higgins, et al. 1994). Then, I generated neighbour-joining trees based on the ClustalW2 protein sequence alignments using The Methodes et Algorithmes pour la bio-informatique LIRMM website (http://www.phylogeny.fr/version2_cgi/one_task.cgi?task_type=bionj) (Gascuel 1997). Finally, to compare two nucleotide or protein sequences, I used NCBI Blast2 algorithms (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&BLAST_SPEC=blast2seq)

2.16 Statistical analysis

Data were analyzed using one-way ANOVA with a Bonferroni post-test to compare groups, unless otherwise specified. GraphPad Prism 5 or GraphPad Prism 6 software was used for all analyses.

Chapter 3: Identification and characterization of the *lgm* locus in *Bordetella*

3.1 Introduction

The lipid A region of *B. pertussis* LPS can be modified with GlcN moieties at the phosphate groups in the wild-type strain BP338 (Figure 4), which is a nalidixic acid resistant derivative of strain Tohama I. Abrogation of this GlcN modification was mapped to the gene locus tag BP0398 (*lgmB*) in a GlcN-negative transposon mutant, suggesting *lgmB* is required for this modification (Marr, Tirsoaga, et al. 2008). Furthermore, LgmB is a homolog of the glycosyltransferase ArnT, which transfers Ara4N onto lipid A in *Salmonella* (Trent, Ribeiro, et al. 2001). I hypothesized that the genes surrounding *lgmB* may also be involved in GlcN modification of lipid A in *B. pertussis* (Figure 6). BP0399 (*lgmA*), which is found directly upstream of *lgmB*, is a homolog of another enzyme in the Ara4N modification pathway: ArnC, which transfers Ara4FN onto the lipid carrier C55P (Breazeale, Ribeiro, et al. 2005). Additionally, bioinformatics analysis of BP0397 (*lgmC*) suggests deacetylase activity and BP0396 (*lgmD*) is a short protein that primarily consists of four predicted transmembrane helices, a similar structure to ArnE and ArnF, which act as C55P-Ara4N flippases (Imagawa, Iino, et al. 2008, Yan, Guan, et al. 2007). Based on these analyses, I hypothesized that LgmA, LgmB, LgmC, and LgmD are involved in a pathway that modifies the phosphate groups of LPS with GlcN, as summarized in Figure 8. Briefly, I predict LgmA transfer GlcNAc onto the carrier lipid C55P, followed by deacetylation by LgmC to generate C55P-GlcN. In this hypothetical model, C55P-GlcN is then flipped from the cytoplasmic to the periplasmic face of the IM by LgmD, and finally LgmB transfers GlcN onto the phosphate of lipid A.

3.2 The *lgm* locus is required for lipid A GlcN modification in *B. pertussis*

Locus tags BP0399 to BP0396 had been annotated in the *B. pertussis* Tohama I genome as putative open reading frames (ORFs) with unknown function (Parkhill, Sebaihia, et al. 2003). Based on transposon mutagenesis of BP0398 (*lgmB*), our lab originally postulated *lgmA*, *lgmB*, *lgmC*, and *lgmD* (BP0399 to BP0396, respectively) were part of the lipid A glucosamine modification locus (Marr, Tirsoaga, et al.

2008, Shah, Albitar-Nehme, et al. 2013). The first step in characterizing the *lgm* locus would be to determine whether all four lgm genes are required for GlcN modification of lipid A in *B. pertussis*.

3.2.1 *lgmA, lgmB,* and *lgmC* are required for lipid A GlcN modification

The ORFs for *lgmA, lgmB, lgmC,* and *lgmD* were all annotated as hypothetical proteins (Parkhill, Sebaihia, et al. 2003), therefore I first set out to confirm all four of these lgm genes are actually expressed in the BP338 strain. Previously, BP0399 and BP0398 (*lgmA* and *lgmB*, repectively) were shown to be transcribed in *B. pertussis* strain BP338 (Marr, Tirsoaga, et al. 2008), and I used RT-PCR to confirm these results, and to show *lgmC* and *lgmD* are also transcribed in BP338 (Figure 10).

Since *lgmB* was disrupted by the insertion of a transposon in *B. pertussis* BP338 to generate a strain lacking GlcN-modified lipid A, it is possible that insertion of the transposon affected expression of a downstream gene. Therefore, to confirm *lgmB* is required for GlcN modification of lipid A, and to determine if *lgmA*, *lgmC,* and *lgmD* are also needed for this modification, I generated markerless deletion mutants of each of the individual *lgm* locus genes and of the full *lgm* locus. This generated the following mutant strains: BP338LgmAKO, BP338LgmBKO, BP338LgmCKO, BP338LgmDKO, and BP338LgmABCDKO (full *lgm* locus deletion mutant). The lipid A from these strains was analyzed via negative-ion MALDI mass spectrometry. The presence of the GlcN modification is represented by peaks observed at *m*/*z* 1720 (one GlcN at either 1-phosphate or 4′-phosphate) and 1881 (GlcN at both 1-phosphate and 4′-phosphate), *i.e.* 161 and 322 mass units higher than the major unmodified penta-acyl lipid A species observed at *m*/*z* 1559 (Figure 11). Wild type BP338 and the *lgmD* mutant have GlcN-modified lipid A, whereas the *lgmA*, *lgmB*, *lgmC,* and the full *lgmABCD* locus mutants lack this modification (Figure 11, Table 7, Appendix A). I demonstrated these results were not due to downstream polar effects by complementing each mutant (BP338LgmAKO, BP338LgmBKO, and BP338LgmCKO) with a vector containing the deleted gene (pBBR2LgmA, pPtacLgmAB, and pBBR2LgmC, respectively), and analyzed the lipids A of these strains via negative-ion MALDI mass spectrometry. I found, in each

case, that complementation of the missing gene resulted in restoration of the lipid A GlcN modification phenotype (Table 7, Appendix A). This suggests that only *lgmA*, *lgmB*, and *lgmC* are required for GlcN modification of lipid A in *B. pertussis*, but *lgmD* is not. To further assess the dispensability of *lgmD*, I complemented the full *lgm* locus mutant either with *lgmABCD* or with only *lgmABC* (vectors pPtacLgmABCD and pPtacLgmABC, respectively) and tested these strains for lipid A GlcN modification in a similar manner. I found GlcN modification was restored in both strains, showing *lgmD* is not required for modification of lipid A with GlcN.

Previously, a transposon mutant of *B. pertussis* strain BP338 lacking the lipid A GlcN modification was shown to have decreased hTLR4 activation when compared to wild type BP338 (Marr, Hajjar, et al. 2010). Since I had generated several BP338 strains that lack GlcN-modified lipid A, I set out to confirm the link between GlcN modification and hTLR4 activation. I used the HEK-Blue hTLR4 activation assay to determine the ability of these strains to activate hTLR4 by stimulating HEK cells expressing hTLR4 with heat-killed bacteria (Figure 12). My results showed that strains with the GlcN-modified lipid A (BP338, BP338LgmDKO, and BP338LgmABCDKO complemented with either pPtacLgmABCD or pPtacLgmABC) have approximately 1.5 to 3.5 times higher levels of hTLR4 activity in this assay when compared with strains without the GlcN modification (BP338LgmAKO, BP338LgmBKO, BP338LgmCKO, and BP338LgmABCDKO). The *lgmB* transposon mutant, which lacks the GlcN modification, had previously been shown to also have decreased levels of hTLR4 activation compared to BP338.

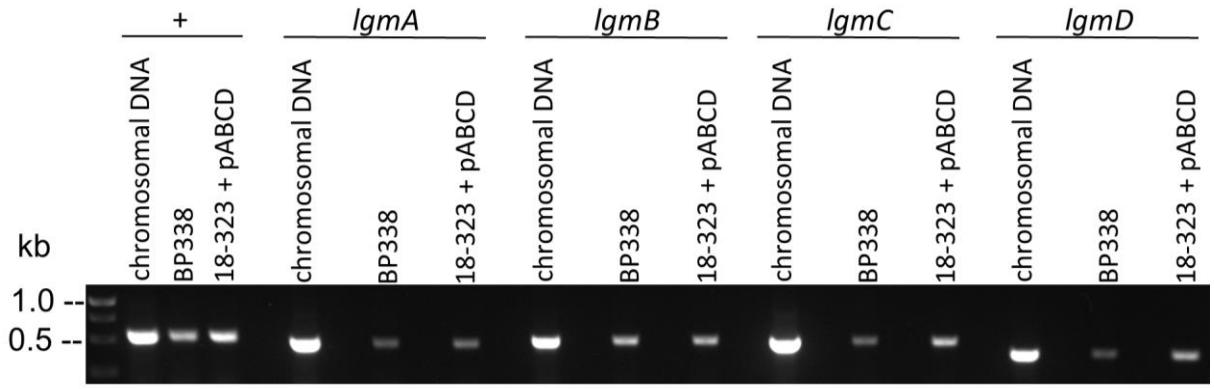


Figure 10. Transcription of *lgmABCD* in BP338 and 18-323 + pPtacLgmABCD
Reverse transcriptase PCR (RT-PCR) of *B. pertussis* BP338 and 18-323 + pPtacLgmABCD (18-323 + pABCD) with *vag8* (+), *lgmA*, *lgmB*, *lgmC*, or *lgmD* primers. Chromosomal DNA from BP338 was used as a positive control for each PCR reaction to indicate the size of the expected product. *vag8* is a Bvg^+ phase gene. Left-most lane is a DNA ladder in kilobases (kb).

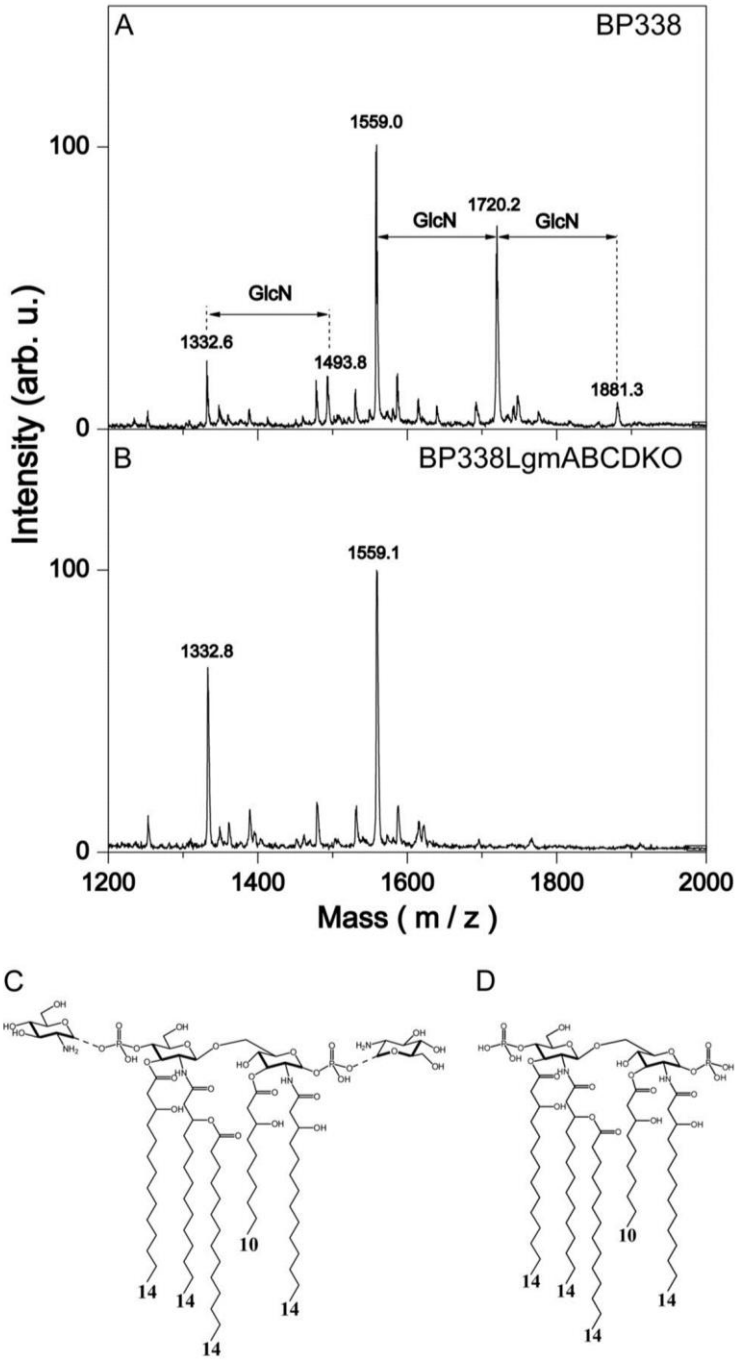


Figure 11. Mass spectrometry analysis of BP338 and BP338*lgmABCD*KO lipid A
Structural analysis of lipid A with negative-ion MALDI mass spectrum analysis. Mass spectra of BP338 (A) and BP338LgmABCDKO (B), the full *lgm* locus deletion mutant. Peaks at *m*/*z* 1559 represent penta-acyl lipid A that lack GlcN modification, peaks at *m*/*z* 1720 represent penta-acyl lipid A with one GlcN modification at either phosphate group, and peaks at *m*/*z* 1881 represent penta-acyl lipid A with a GlcN modification at both phosphate groups. The peaks at *m*/*z* 1333 and 1494 represent tetra-acyl species. arb. u., arbitrary units. C and D, the lipid A structures present in BP338 (C) and BP338LgmABCDKO (D) as determined by mass spectral analysis. The numbers at the bottom of the structures indicate the length of the acyl chains. From Shah *et al*. 2013 (Shah, Albitar-Nehme, et al. 2013), used with permission.

Table 7. GlcN modification of BP338 *lgm* locus mutants

Summary of lipid A glycosylation, as determined by negative-ion MALDI mass spectrum analysis. See Appendix A for mass spectra. pLgmA, pBBR2LgmA; pLgmAB, pPtacLgmAB; pLgmC, pBBR2LgmC; pLgmABCD, pPtacLgmABCD; pLgmABC, pPtacLgmABC. Modified from Shah *et al*. 2013 (Shah, Albitar-Nehme, et al. 2013), used with permission.

Strain	GlcN-modified lipid A
BP338	Yes
BP338LgmAKO	No
BP338LgmAKO + pLgmA	Yes
BP338LgmBKO	No
BP338LgmBKO + pLgmAB	Yes
BP338LgmCKO	No
BP338LgmCKO + pLgmC	Yes
BP338LgmDKO	Yes
BP338LgmABCDKO	No
BP338LgmABCDKO + pLgmABCD	Yes
BP338LgmABCDKO + pLgmABC	Yes

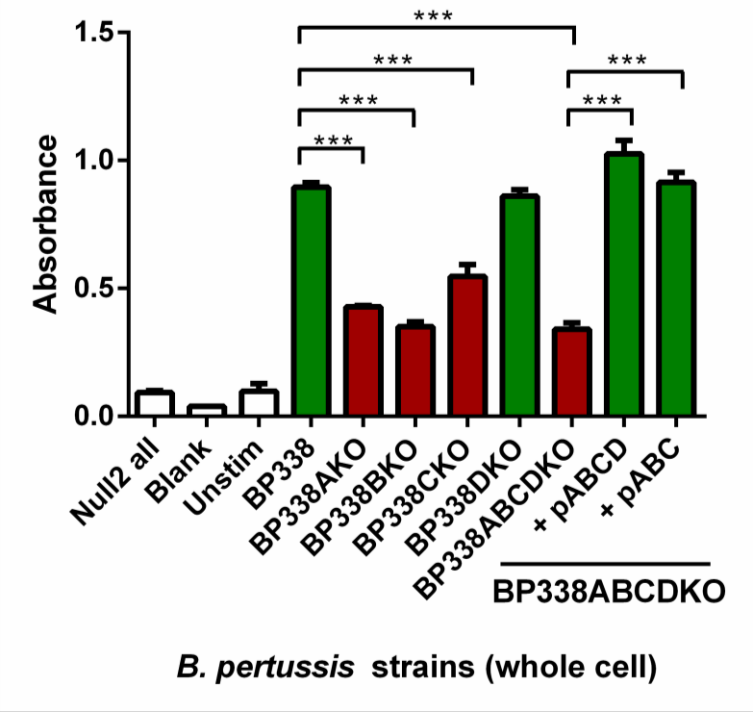


Figure 12. hTLR4 activation by BP338 *lgm* locus mutants
hTLR4 activation measured with HEK-Blue NFκB hTLR4 activity assay. Null2 all, stimulation of HEK-Blue Null2 cell line that lacks TLR4 expression with all LPS variants; Blank, medium only with no HEK-Blue cells; Unstim, HEK-Blue hTLR4 cells stimulated with medium only; BP338LgmABCDKO, the full *lgm* locus deletion mutant; BP338LgmABCDKO + pLgmABCD, full *lgm* locus deletion mutant complemented with pPtacLgmABCD*;* BP338LgmABCDKO + pLgmABC, full *lgm* locus deletion mutant complemented with pPtacLgmABC. Green bars represent strains with GlcN-modified lipid A and red bars represent strains with no GlcN modification (from Table 7). Absorbance at 650 nm (A_{650}). Graph shows the results of one representative experiment of three, $n = 6$ replicates per experiment. One-way ANOVA with a Bonferroni post-test used for statistical analysis. p values: < 0.001 (***). Modified from Shah *et al*. 2013 (Shah, Albitar-Nehme, et al. 2013), used with permission.

3.2.2 Potential flippase replacements are not required for GlcN modification

In my initial model (Figure 8), I hypothesized the need for a flippase to translocate C55P-GlcN from the cytoplasmic to the periplasmic face of the IM, since LgmA and LgmC are predicted to have globular domains in the cytoplasm, and LgmB, which primarily consists of predicted transmembrane helices, is predicted to have a short periplasmic C-terminal region (Figure 7). I had originally hypothesized that LgmD functions as this C55P-GlcN flippase, due to the similarities in predicted membrane topology between LgmD and ArnE and ArnF, which function to flip C55P-Ara4N from the cytoplasmic to the periplasmic face of the IM (Yan, Guan, et al. 2007). However, the *lgmD* mutant strain (BP338LgmDKO) still has GlcN-modified lipid A, suggesting LgmD is not required for the Lgm pathway. Alternatively, if LgmD does act as a flippase, another protein may compensate for the role of LgmD in the *lgmD*KO strain, therefore resulting in GlcN-modified lipid A, even in the absence of *lgmD*. ArnE and ArnF are small proteins that are made up of four transmembrane helices, just as *lgmD* is (Figure 7). Therefore, I searched for similar proteins in the *B. pertussis* Tohama I genome that may function as a C55P-GlcN flippase in the Lgm pathway.

Upon closer inspection of the *lgm* locus in *B. pertussis*, I identified a previously unannotated ORF, which I have named *lgmE*. *lgmE* is found within the original *lgm* locus, but in the opposite direction compared to *lgmABCD*, and it overlaps with *lgmC* and *lgmB* (Figure 14A). LgmE is predicted to contain four transmembrane helices (Figure 13A) and is a member of the GtrA-like Pfam family, members of which are predicted to function as flippases (Guan, Bastin, et al. 1999). Therefore, I hypothesized that LgmE may function either as a possible C55P-GlcN flippase replacement of LgmD in the Lgm pathway, or simply fulfills this role in the Lgm pathway, regardless of LgmD. Since *lgmE* had not previously been described, I first demonstrated that *lgmE* is transcribed in *B. pertussis* strain BP338 using RT-PCR with *lgmE*-specific primers (Figure 14B). Next, I tested whether LgmE is required for GlcN modification in *B. pertussis*. Since *lgmE* overlaps with both *lgmB* and *lgmC* (Figure 14A), deleting the *lgmE* gene would

also disrupt *lgmB* and *lgmC*. Therefore, instead of knocking out *lgmE*, I generated a single base pair mutant of *lgmE* on the plasmids pPtacLgmABCD and pPtacLgmABC such that the residue Gly15 of *lgmE* was mutated to a stop codon and this mutation in the overlapping *lgmC* was a silent mutation (Figure 15). This generated the plasmids pPtacLgmABCD-EG15STOP, which encodes for *lgmABCD* with a mutated *lgmE*, and pPtacLgmABC-EG15STOP, which contains *lgmABC* and a mutated *lgmE*. I then complemented the full *lgm* locus mutant, BP338LgmABCDKO, with these plasmids to generate a strain that either lacks only *lgmE* expression (BP338lgmABCDKO + pPtacLgmABCD-EG15STOP) or lacks both *lgmD* and *lgmE* expression (BP338lgmABCDKO + pPtacLgmABC-EG15STOP). The results showed these two strains transcribe the expected *lgm* genes by RT-PCR with primers specific to each lgm gene: BP338lgmABCDKO + pPtacLgmABCD-EG15STOP transcribes all five *lgm* genes whereas BP338lgmABCDKO + pPtacLgmABC-EG15STOP does not transcribe *lgmD* (Figure 16). Since the Gly15STOP mutation in *lgmE* stops production of LgmE at the translational level, *lgmE* is still expected to be transcribed. I tested these strains for the ability to activate hTLR4 using the HEK-Blue hTLR4 activation assay, and found both strains had wild-type levels of hTLR4 activation (Figure 17). Thus, both *lgmE* mutant strains have GlcN-modified lipid A, so LgmE is not required for the Lgm pathway.

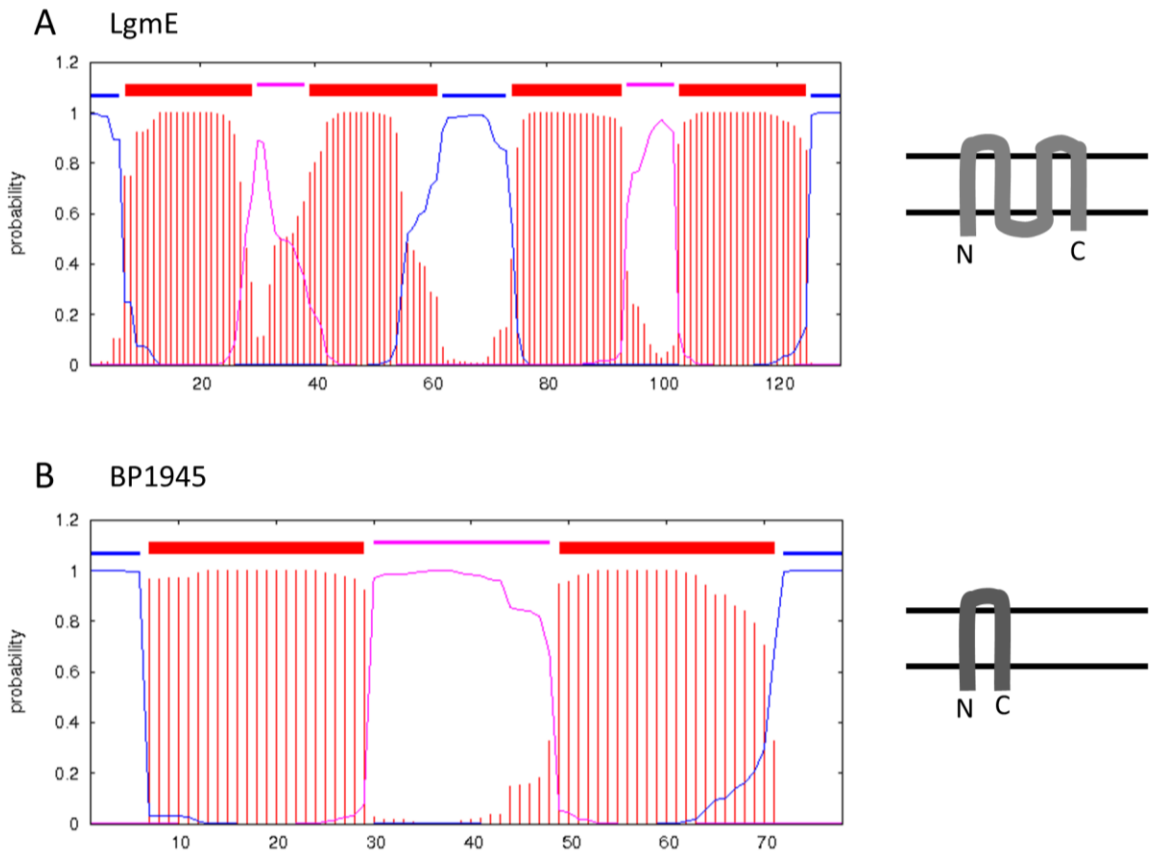


Figure 13. Predicted topologies of BP1945 and LgmE
A) LgmE, B) BP1945. Membrane topologies were predicted with TMHMM (Sonnhammer, von Heijne, et al. 1998). The amino acid position is along the x-axis, the probability of a tramsmembrane helix is along the y-axis (red vertical bars). For the horizontal lines: red is transmembrane region, blue is cytoplasmic region, pink is periplasmic region. The diagrams on the right depict a visual representation of the TMHMM prediction. The upper horizontal black line represents the periplasmic leaflet of the IM, the lower black line represents the cytoplasmic leaflet of the IM, and the thick grey line represents the protein.

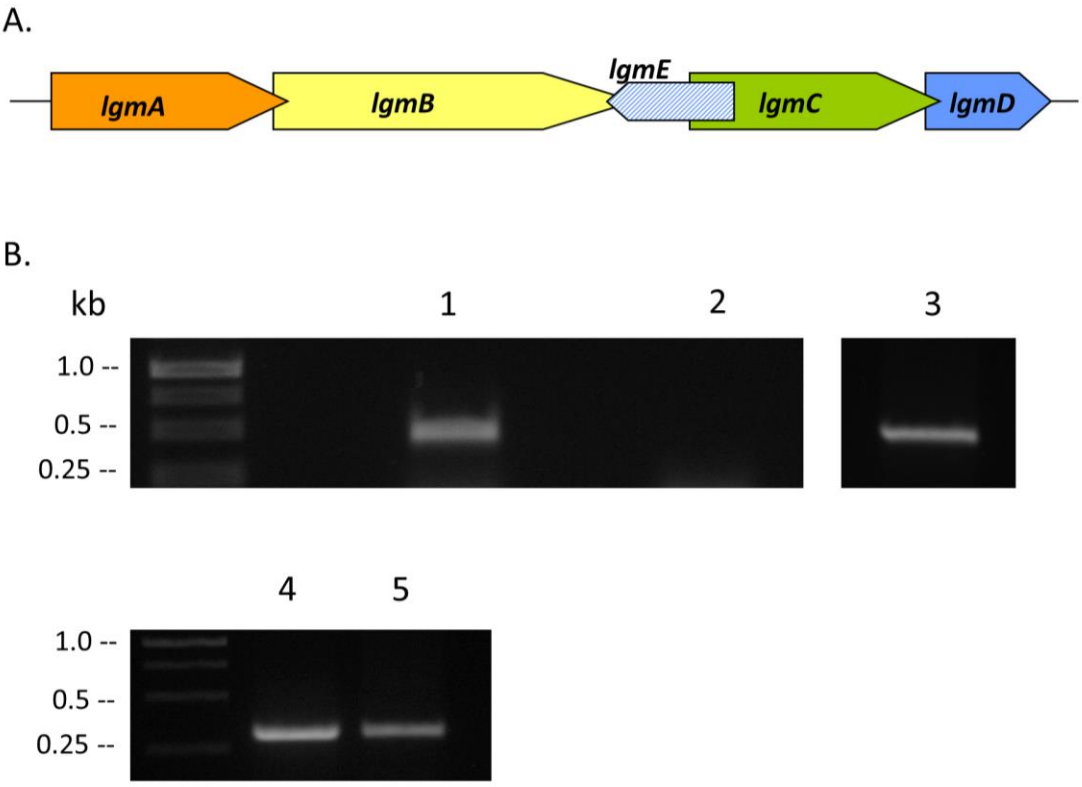


Figure 14. Current schematic of the *B. pertussis lgm* locus and transcription of *lgmE* in BP338
A) *B. pertussis lgm* locus consists of *lgmA*, *lgmB*, *lgmC*, *lgmD*, and *lgmE* (previously referred to as *lgmD-like* (Novikov, Shah, et al. 2013)). B) Reverse transcriptase PCR (RT-PCR) of *B. pertussis* BP338 with *lgmE* primers. Control to confirm no DNA contamination of total RNA preparation from BP338: lane 1 – PCR of BP338 chromosomal DNA with *lgmA* primers, lane 2 – PCR of RNA purification from BP338 with *lgmA* primers. Control to show successful conversion of RNA to cDNA: lane 3 – PCR of BP338 cDNA with *lgmA* primers. Test for *lgmE* transcription: lane 4 – PCR of PtaclgmABCD DNA with *lgmE* primers, lane 5 – PCR of BP338 cDNA with *lgmE* primers. Left-most lanes are DNA ladders in kilobases (kb). Band in lane 5 indicates *B. pertussis* strain BP338 transcribes *lgmE*.

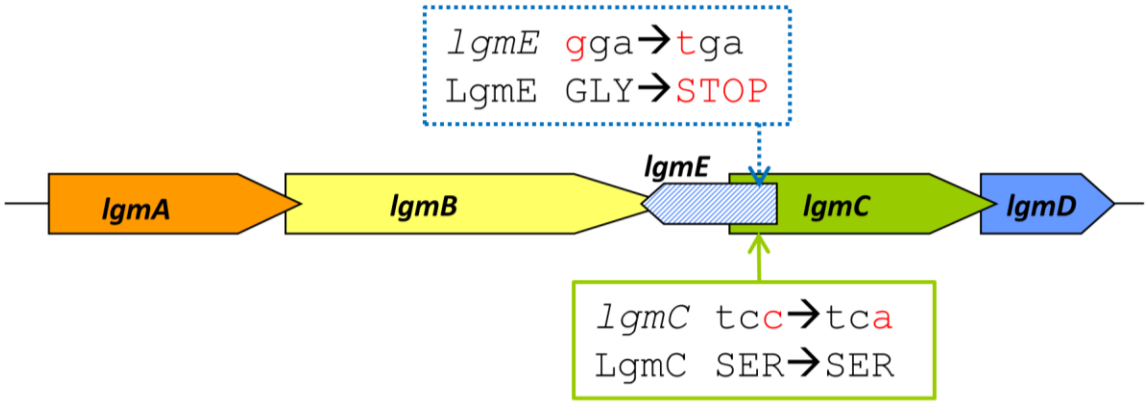


Figure 15. Schematic of the *lgmE* mutant
Gly15STOP mutation in *lgmE* results from changing 'g' to 't'. This same mutation in the overlapping *lgmC* results in a 'c' to 'a' mutation, which is silent (no change in the serine residue). Mutations are highlighted in red.

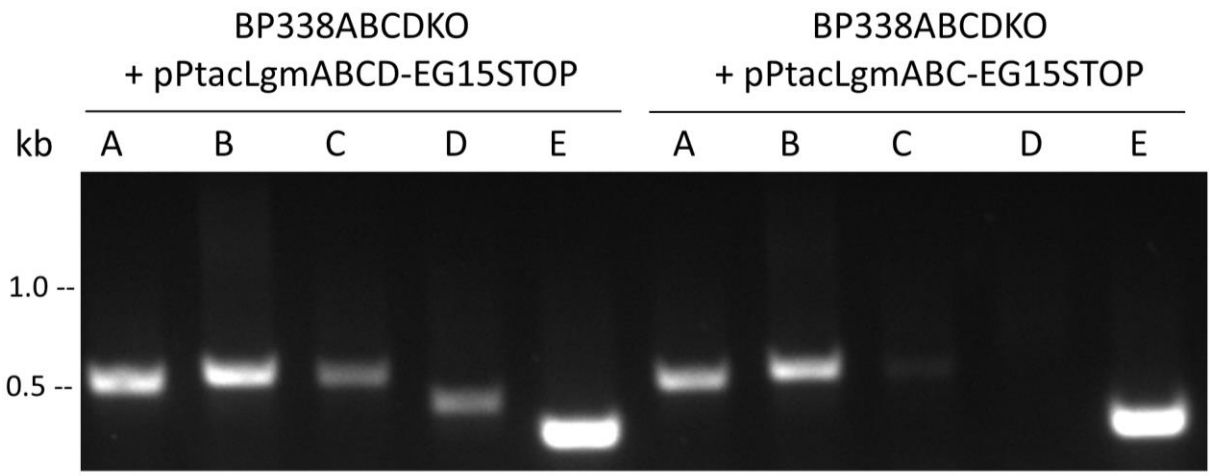


Figure 16. Transcription of *lgm* locus genes in BP338 *lgmE*KO strains
Reverse transcriptase PCR (RT-PCR) of *B. pertussis* BP338LgmABCDKO + pPtacLmgABCD-EG15STOP (*lgmABCD*KO strain complemented with *lgmABCD* and a mutated *lgmE*) and BP338LgmABCDKO + pPtacLgmABC-EG15STOP (*lgmABCD*KO strain complemented with *lgmABC* and a mutated *lgmE*) with *lgmA*, *lgmB*, *lgmC*, *lgmD*, and *lgmE* -specific primers (lanes labeled A, B, C, D, and E, respectively). Left-most markings (1.0 and 0.5) are size markers in kilobases (kb). BP338ABCDKO is BP338LgmABCDKO.

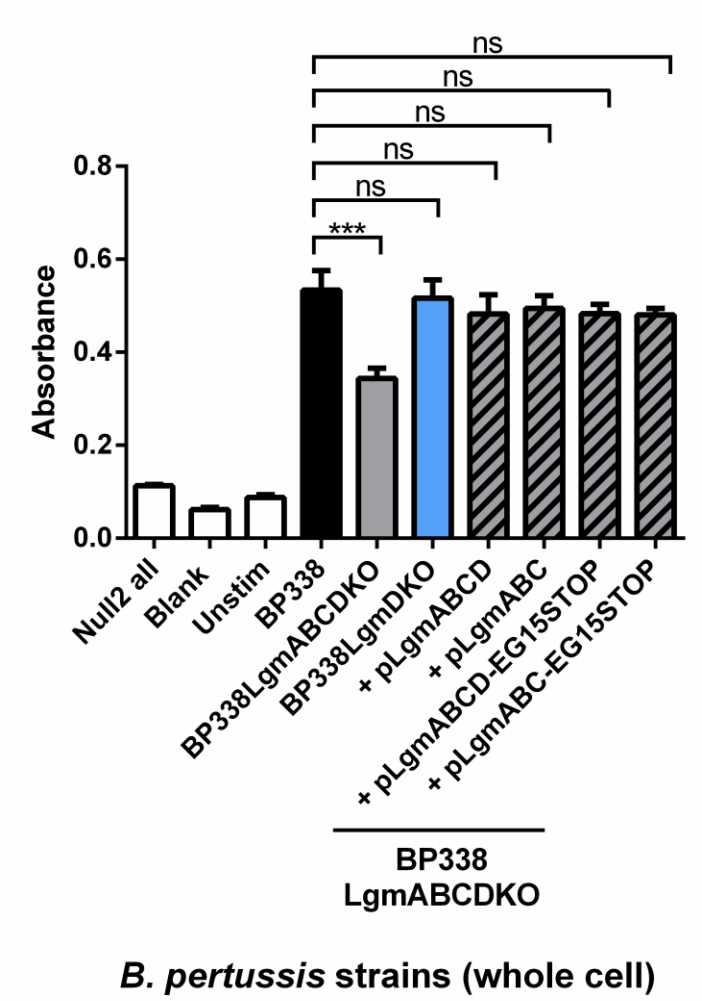


Figure 17. hTLR4 activation by BP338 *lgmE*KO strains
hTLR4 activation measured with HEK-Blue NFκB hTLR4 activity assay. Null2 all, stimulation of HEK-Blue Null2 cell line that lacks TLR4 expression with all LPS variants; Blank, medium only with no HEK-Blue cells; Unstim, HEK-Blue hTLR4 cells stimulated with medium only; BP338LgmABCDKO, the full *lgm* locus deletion mutant; BP338LgmDKO, full *lgmD* deletion mutant; + pLgmABCD, complemented with pPtacLgmABCD vector; + pLgmABC, complemented with pPtacLgmABC vector; + pLgmABCD-EG15STOP, complemented with pPtacLgmABCD-EG15STOP vector that has the *lgmE* G15STOP mutation; + pLgmABC-EG15STOP, complemented with pPtacLgmABC-EG15STOP vector that has the *lgmE* G15STOP mutation. Absorbance at 650 nm (A_{650}). Graph shows the results of one representative experiment of three, $n = 4$ replicates per experiment. One-way ANOVA with a Bonferroni post-test used for statistical analysis. p values: < 0.001 (***), no significant difference (ns).

A second possible candidate for a replacement flippase for *lgmD* is locus tag BP1945. BP1945 consists primarily of two predicted transmembrane helices (Figure 13A) and is found adjacent to the Bps locus, which is involved in poly-GlcNAc synthesis. To determine if BP1945 is involved in GlcN-modification of lipid A in *B. pertussis* strain BP338, I generated clean, markerless deletion mutants of BP1945 in both wild type BP338 and in the *lgmD* mutant to generate the strains BP338BP1945KO and BP338LgmDKOBP1945KO, respectively. I then tested the ability of these strains to activate hTLR4 using the HEK-Blue hTLR4 activation assay to determine if these strains had GlcN-modified lipid A, since strains lacking the modification have lower levels of hTLR4 activation. Both of these BP1945KO strains activated hTLR4 to the same level as wild type BP338 (Figure 18), demonstrating the BP1945 mutant strains have GlcN-modified lipid A. As such, BP1945 does not replace the activity of LgmD in the *lgmD* mutant and is not required for GlcN modification in *B. pertussis*.

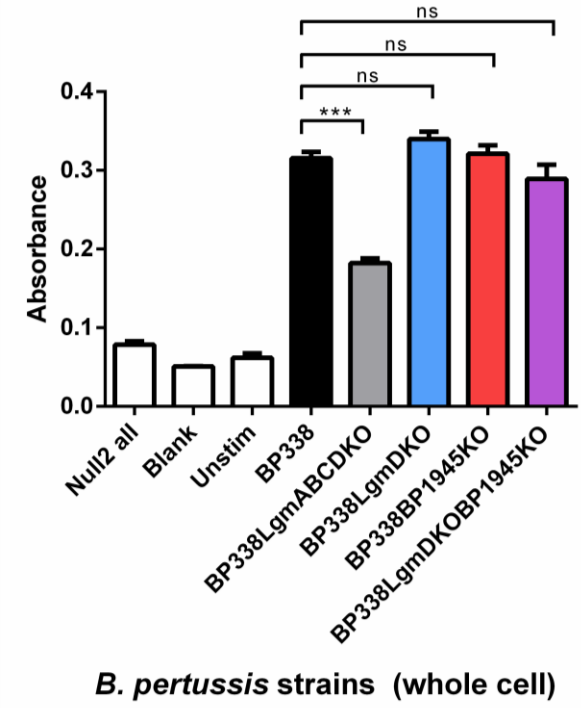


Figure 18. hTLR4 activation by BP338 BP1945KO strains
hTLR4 activation measured with HEK-Blue NFκB hTLR4 activity assay. Null2 all, stimulation of HEK-Blue Null2 cell line that lacks TLR4 expression with all LPS variants; Blank, medium only with no HEK-Blue cells; Unstim, HEK-Blue hTLR4 cells stimulated with medium only; BP338LgmABCDKO, the full *lgm* locus deletion mutant; BP338LgmDKO, full *lgmD* deletion mutant*;* BP338BP1945KO, BP1945 locus tag deletion mutant; BP338LgmDKOBP1945KO, *lgmD* and BP1945 locus tag double deletion mutant. Absorbance at 650 nm (A_{650}). Graph shows the results of one representative experiment of three, $n = 4$ replicates per experiment. One-way ANOVA with a Bonferroni post-test used for statistical analysis. p values: < 0.001 (***), no significant difference (ns).

3.3 LgmA functions as a GlcNAc transferase

Thus far, I had established LgmA, LgmB, and LgmC are required for GlcN modification of *B. pertussis* lipid A. LgmA is a homolog of the glycosyltransferase ArnC and is a predicted inner membrane protein, due to two predicted transmembrane helices (Figure 7). Since ArnC functions to transfer Ara4FN from UDP-Ara4FN to the inner membrane carrier lipid C55P in the *Salmonella* and *E. coli* Ara4N lipid A modification pathway (Raetz, Reynolds, et al. 2007), I predicted LgmA functions in a similar manner to transfer GlcNAc from UDP-GlcNAc to C55P in *B. pertussis* (Figure 8). To test the function of LgmA, I performed a glycosyltransferase assay using the radioactively-labelled substrate, UDP-GlcN[^{14}C]Ac, and crude membrane preparations from *E. coli* strains as the source of both LgmA and the acceptor lipid. The *E. coli* strains either expressed histidine-tagged LgmA (BL21 + pLgmA) or contained an empty vector control (BL21 + pET30). Immunoblot detection with an anti-HisTag antibody showed histidine-tagged LgmA (LgmA-His) was expressed in BL21 + pLgmA after induction with 1 mM IPTG (band at ~40 kDa) but not expressed in the empty-vector control (Figure 19). Furthermore, LgmA-His localized to the membrane preparation of the sample, which was used as the source of the enzyme in the glycosyltransferase assay (Figure 19). The lipids from each glycosyltransferase reaction were extracted using butanol, and this lipid fraction was measured for radioactivity, which would indicate the transfer of the radioactive substrate, UDP-GlcN[^{14}C]Ac, to a lipid. Negative control reactions, with either no UDP-GlcN[^{14}C]Ac or no membrane preparation, and therefore no enzyme source, had very low levels of radioactivity in the lipid fraction, therefore indicating no transfer of GlcN[^{14}C]Ac to a lipid acceptor in these reactions. The reaction with the membrane preparation from BL21 + pET30 as the enzyme source had higher than background levels of radioactivity in the lipid fraction, indicating background levels of GlcN[^{14}C]Ac transfer to lipids, likely from glycosyltransferases that naturally reside in *E. coli* membranes. However, reactions with LgmA-His-containing membrane preparations (BL21 + pLgmA) had a 40% to 60% increase in radioactivity of the lipid fraction when compared to BL21 + pET30 (Figure

20). This indicates the presence of LgmA increases transfer of GlcN[^{14}C]Ac to lipids in *E. coli*, and therefore, LgmA likely acts as a GlcNAc glycosyltransferase.

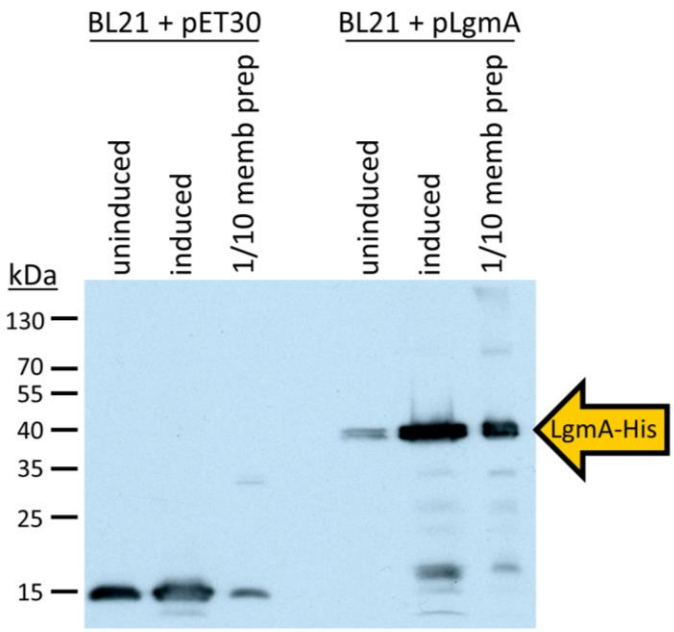


Figure 19. Expression of LgmA-His in *E. coli*
Western transfer and immunoblot detection with anti-HisTag antibody of *E. coli* BL-21 (DE3) cells harboring an empty vector plasmid (BL21 + pET30) or an LgmA-His expressing vector (BL21 + pLgmA). Induced cells were induced with 1 mM IPTG for 24 hours. 1/10 membr prep is a 1/10 dilution of the crude membrane preparation that is at 5 mg/ml total protein concentration. Orange arrow indicates LgmA-His protein band. Immunoblot shows one representative experiment of three, with similar results.

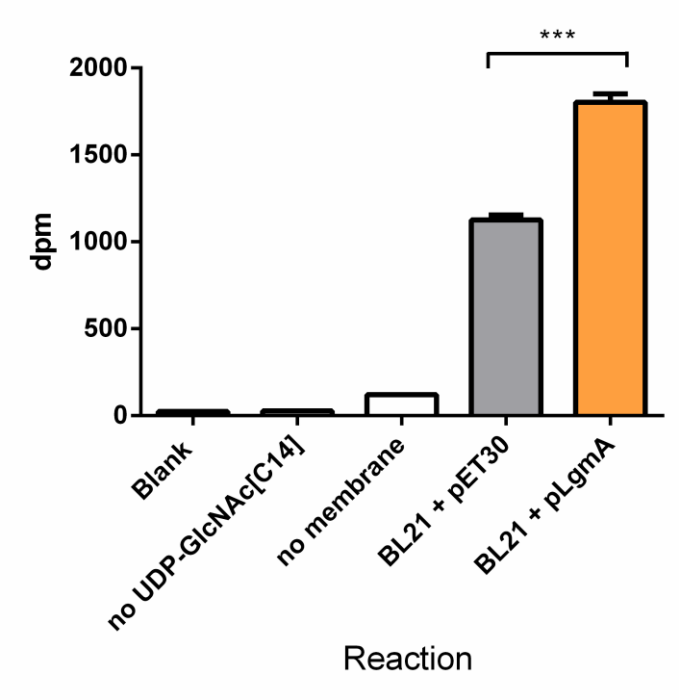


Figure 20. LgmA glycosyltransferase assay
Transfer of GlcN[^{14}C]Ac to the lipid fraction as measured by degradations per minute (dpm) of butanol-extracted lipids from each reaction. Blank, no reaction added; no UDP-GlcN[^{14}C]Ac, control reaction with no radioactive substrate (UDP-GlcN[^{14}C]Ac) but contains BL21 + pLgmA membrane preparation; no membrane, control reaction with no membrane preparation; BL21 + pET30, reaction with empty vector membrane preparation; BL21 + pLgmA, reaction with LgmA-His-containing membrane preparation. Graph shows the results of one representative experiment of three, $n = 3$ replicates for BL21 + pET30 and BL21 + pLgmA experiments. Student's t-test used for statistical analysis to compare BL21 + pET30 and BL21 + pLgmA reactions. p value < 0.001 (***).

3.4 Identification of the putative active site of LgmA

LgmA is predicted to be a member of the GT2 family of glycosyltransferases. Members of this family are predicted to have a conserved DXD motif, which may be involved in the coordination of divalent cations in the active site (Lairson, Henrissat, et al. 2008). To identify a putative active site of LgmA, I aligned LgmA with two GT2 family homologs: ArnC and Ftn_0545. ArnC functions in the *Salmonella* Ara4N lipid A modification pathway to transfer Ara4FN onto the carrier lipid C55P (Breazeale, Ribeiro, et al. 2005) and, similarly, Ftn_0545 functions to transfer N-acetyl-galactosamine (GalNAc) to C55P in the *Francisella novicida* lipid A-modification pathway (Song, Guan, et al. 2009). I identified the following conserved residues that I hypothesized to be involved in LgmA function: D76, D77, D127, and D129; the last two residues appear to be a possible conserved DXD motif (Figure 21A). These residues were mutated by Andrew Low to glycine in the construct pBBR2LgmA to generate the following mutants: D76G D77G, D127G, D129G, and D127G D129G. To test if these mutations affect the function of LgmA, the *lgmA* mutant, BP338LgmAKO, was complemented with either wild type *lgmA* (pBBR2LgmA) or mutated *lgmA* and RT-PCR was used to confirm these complemented strains all transcribed *lgmA* (Figure 22 insert). These complemented strains were shown to have GlcN modified-lipid A by stimulation of HEK-Blue hTLR4 cells with heat-killed bacteria to assess hTLR4 activation (Figure 22). The *lgmA*KO strain was found to have 28 to 51% lower levels of hTLR4 activation compared to wild type BP338, but this phenotype was partially rescued by complementation with wild type *lgmA* (+ plgmA), which had 36 to 73% higher levels of hTLR4 activation compared to *lgmA*KO. Complementation of the *lgmA*KO strain with the *lgmA* mutants D76G D77G, D127G, D129G, D127G D129G resulted in no recovery of hTLR4 activation, therefore these strains have unmodified lipid A, suggesting these amino acids may be required for LgmA activity.

Figure 21. Identifying conserved residues in LgmA
A) ClustalW2 (Thompson, Higgins, et al. 1994) alignment of *B. pertussis* Tohama I LgmA, *F. novicida* U112 Ftn-0545, and *E. coli* K-12 ArnC. Residues in red text and below a red arrow were targeted for mutational analysis. '*': Identical amino acid residues; ':': conserved amino acid residues; '.': semi-conserved amino acid residues.
B) ClustalW2 (Thompson, Higgins, et al. 1994) alignment of LgmA from *Bordetella* species: Bpe, *B. pertussis* Tohama I; Bbr, *B. bronchiseptica* RB50; Bpa, *B. parapertussis* 12822; Bav, *B. avium* 197N; Bhi, *B. hinzii* ATCC 51730; and Btr, *B. trematum* CCUG 13902. Bolded sequences (*B. hinzii* and *B. trematum*) have no lipid A GlcN-modification whereas all other species listed have the lipid A GlcN-modification. Residues in green text and below a green arrow are conserved in all GlcN-modification strains, but variable in non-GlcN-modifying strains and these residues were targeted for mutational analysis.

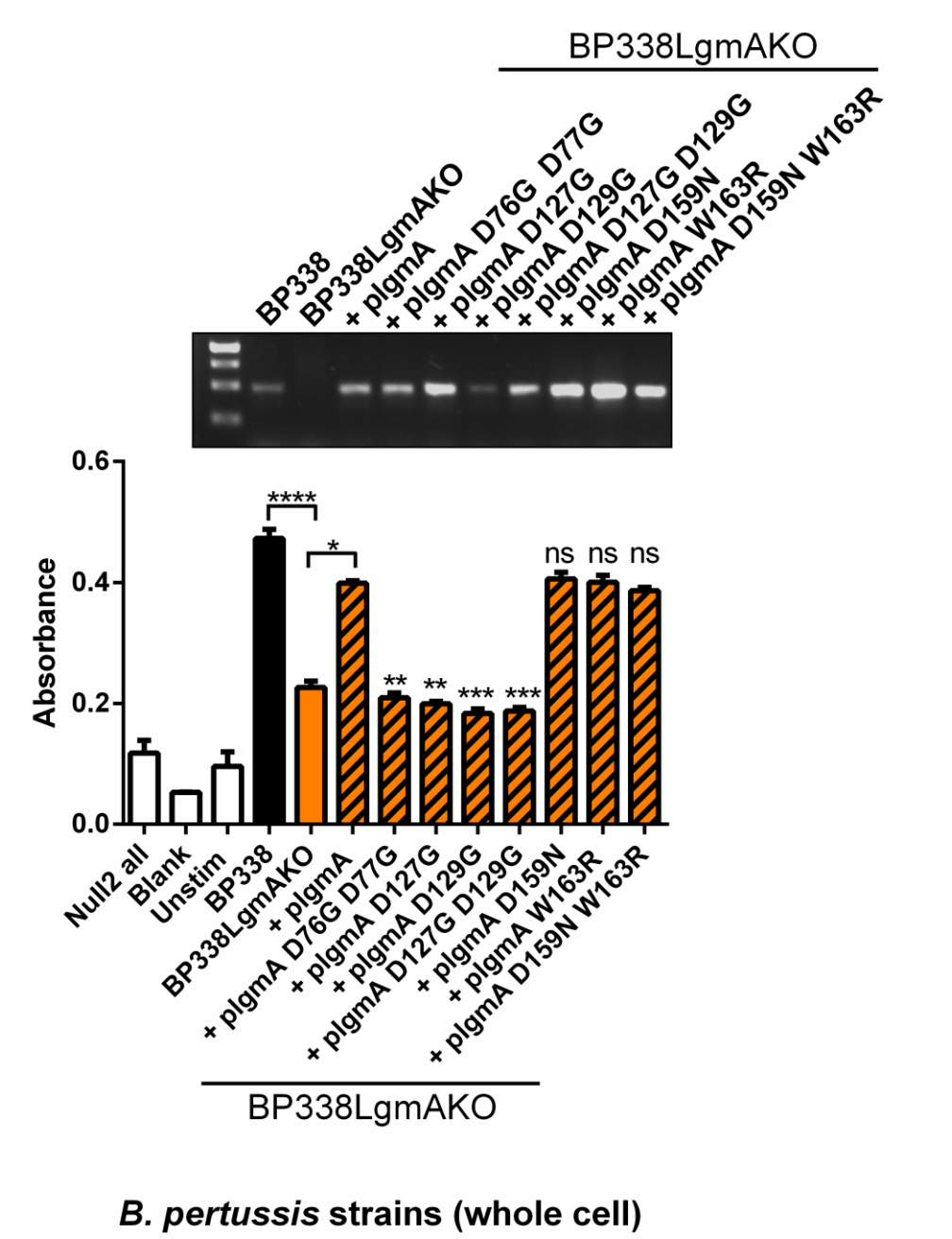


Figure 22. hTLR4 activation by LgmA mutants
hTLR4 activation measured with HEK-Blue NFκB hTLR4 activity assay. Null2 all, stimulation of HEK-Blue Null2 cell line that lacks TLR4 expression with all LPS variants; Blank, medium only with no HEK-Blue cells; Unstim, HEK-Blue hTLR4 cells stimulated with medium only; BP338LgmAKO, *lgmA* deletion mutant; + plgmA, complemented with pBBR2LgmA. Absorbance at 650 nm (A_{650}). Graph shows the results of one representative experiment of three, $n = 4$ replicates per experiment. One-way ANOVA with a Bonferroni post-test used for statistical analysis. p values: < 0.0001 (****), < 0.001 (***), no significant difference (ns). Bars with p value indicator directly above are in comparison with BP338LgmAKO + plgmA. Inset shows transcription of *lgmA* in each strain using RT-PCR and *lgmA*-specific primers. DNA ladder sizes (left-most lane), from top to bottom: 1.0, 0.75, 0.5, 0.25 kilobases (kb). Inset shows one representative experiment of three.

Since the structure of LgmA is unknown, I generated a hypothetical structure of LgmA using the Protein Homology/analogy Recognition Engine (Phyre) protein fold recognition server (Kelley and Sternberg 2009) to assess the position of these residues on a predicted 3-dimenstional model (accessed on December 2, 2013). The highest ranked Phyre-predicted model for LgmA covered amino acids 24 to 345 of the 352 residue long peptide, with an estimated precision of 100% and an E value of 1.3×10^{-21}. The model was based upon the structure of the *Bacteroides fragilis* strain NCTC 9343 protein BF2801 (PDB ID 3BCV), which is a predicted glycosyltransferase. I located the amino acid residues that, when mutated, result in a lack of LgmA activity on this structural model (Figure 23) and found these residues are clustered in one region of the predicted LgmA structure. I hypothesize the region of the predicted LgmA structure where D76, D77, D127, and D129 cluster may function as the active site.

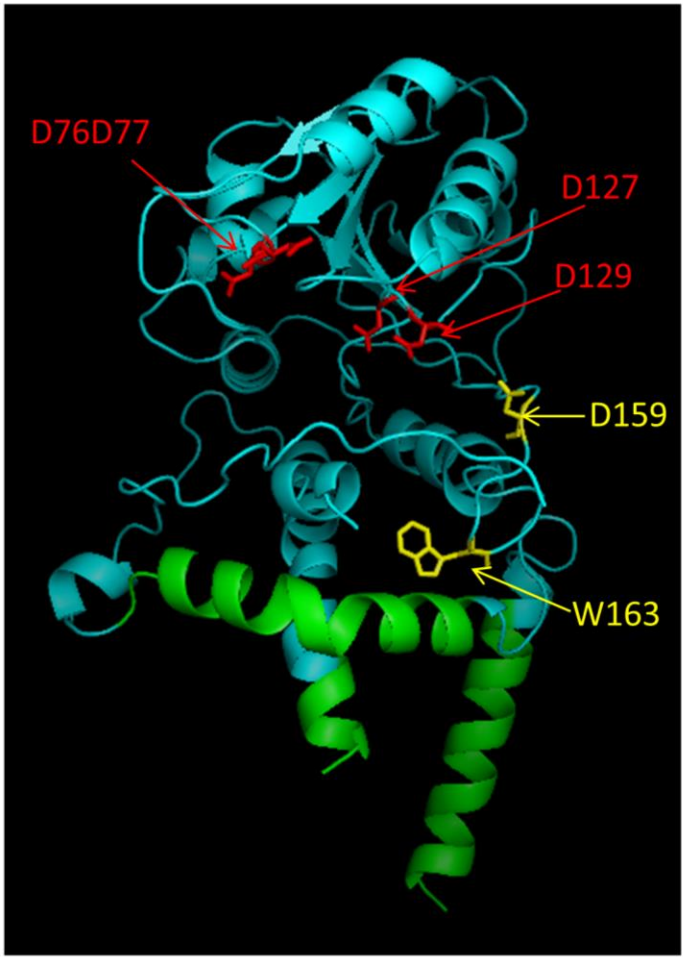


Figure 23. Phyre predicted model of the structure of LgmA
Model of LgmA structure by the Protein Homology/analogy Recognition Engine (Phyre) (Kelley and Sternberg 2009). The predicted transmembrane helix region (Figure 7) is green, the remainder of the protein is cyan. Residues that, when mutated, resulted in a lack of GlcN modification of lipid A are red, other mutated residues, that did not affect lipid A GlcN modification, are yellow. PyMOL was used to visualize this structure.

3.5 Identification of the putative active site of LgmC

In the hypothetical model (Figure 8), after LgmA transfers GlcNAc onto C55P, I predicted LgmC functioned to deacetylate C55P-GlcNAc to generate C55P-GlcN. Recently, LgmC was shown to function as I had originally predicted, to remove the acetyl group from C55P-GlcNAc in *B. bronchiseptica* (Llewellyn, Zhao, et al. 2012). LgmC is predicted to be a member of the YdjC-like protein family by Pfam. To identify putative active site residues, I aligned LgmC with the following homologs to find conserved residues: YdjC, Ftn_0544, and TTHB209, all three of which are also members of the YdjC-like family. YdjC is hypothesized to be involved in the cleavage of cellobiose-phosphate, Ftn_0544 functions to deacetylate C55P-GalNAc in the *F. novicida* lipid A modification system, and TTHB209 is structurally similar to *Sp*PgdA, a GlcNAc deacetylase in *Streptococcus pneumoniae* (Imagawa, Iino, et al. 2008, Llewellyn, Zhao, et al. 2012, Song, Guan, et al. 2009). Based on the ClustalW2 alignment of these proteins (Figure 24), I hypothesized the conserved residues D80, D81, H130, D187, H189, and E313 are required for LgmC activity, since these residues are conserved amongst the aforementioned YdjC-like family proteins. These residues were mutated in the plasmid pBBR2LgmC to generate the following mutants: D80G D81G, H130G, D187G H189G, and E313G. Then the *lgmC* mutant strain, BP338LgmCKO, was complemented with wild type *lgmC* (+ plgmC) or the *lgmC* mutant plasmids and these strains were shown to transcribed *lgmC* by RT-PCR with *lgmC*-specific primers (Figure 25 insert). To determine if these mutations in LgmC affect function, the ability of the complemented *lgmC*KO strains to activate hTLR4 was tested, thereby assessing the presence of the lipid A GlcN modification in these different strains. The *lgmC*KO strain had 27 to 54% lower levels of hTLR4 activation compared to the BP338 wild type (Figure 25), indicating a lack of GlcN-modified lipid A, as previously confirmed by mass spectrometry analysis (Section 3.2.1). Complementation with wild-type *lgmC* (+ *lgmC*) resulted in a partial rescue of hTLR4 activation, such that it was 10 to 40% greater than *lgmC*KO hTLR4 activation, but is was still lower than wild type levels of hTLR4 activation, suggesting lower-than-wild type-levels of GlcN modification (Figure 25). However, complementation with the *lgmC* mutant plasmids did not rescue

hTLR4 activation (Figure 25), showing these strains did not have GlcN-modified lipid A. This suggests the mutated amino acid residues are required for LgmC function, since LgmC is required for GlcN modification of lipid A.

Figure 24. Identifying conserved residues in LgmC
ClustalW2 (Thompson, Higgins, et al. 1994) alignment of *B. pertussis* Tohama I LgmC, *Enterobacter cloacae* EcWSU1 YdjC, *F. novicida* U112 Ftn_0544 , and *Thermus thermophilus* HB8 TTHB029. Residues in red text and below a red arrow were targeted for mutational analysis. '*': Identical amino acid residues; ':': conserved amino acid residues; '.': semi-conserved amino acid residues.

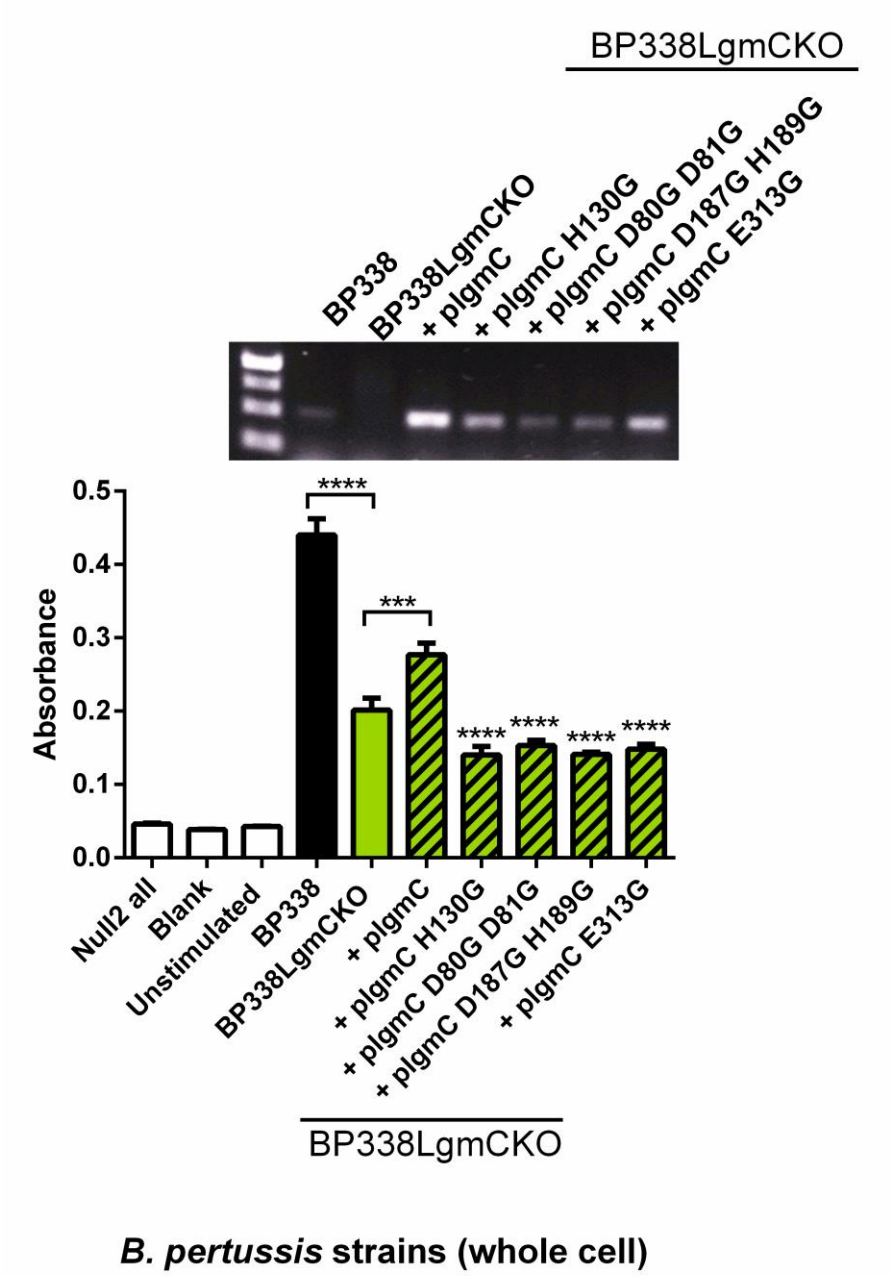


Figure 25. hTLR4 activation by LgmC mutants
hTLR4 activation measured with HEK-Blue NFκB hTLR4 activity assay. Null2 all, stimulation of HEK-Blue Null2 cell line that lacks TLR4 expression with all LPS variants; Blank, medium only with no HEK-Blue cells; Unstim, HEK-Blue hTLR4 cells stimulated with medium only; BP338LgmCKO, *lgmC* deletion mutant; + plgmC, complemented with pBBR2LgmC. Absorbance at 650 nm (A_{650}). Graph shows the results of one representative experiment of three, $n = 4$ replicates per experiment. One-way ANOVA with a Bonferroni post-test used for statistical analysis. p values: < 0.0001 (****), < 0.001 (***). Bars with p value indicator directly above are in comparison with BP338LgmCKO + plgmC. Inset shows transcription of *lgmC* in each strain using RT-PCR and *lgmC*-specific primers. DNA ladder sizes (left-most lane), from top to bottom: 1.0, 0.75, 0.5, 0.25 kilobases (kb). Inset shows one representative experiment of three.

Similarly to LgmA, the structure of LgmC is not yet solved, though LgmC is predicted to be a cytoplasmic protein (Figure 7). To analyze the possible positions of these residues, especially in relation to one another, I generated a hypothetical structure for LgmC via the Phyre protein fold recognition server. The structure with the highest estimated precision and lowest E value (100% and 9.1×10^{-26}, respectively), was based on the previously solved structure of *Eterococcus faecalis* strain V583 protein EF_3048, which is a putative enzyme involved in cellobiose-phosphate cleavage (PDB ID 2I5I). This hypothetical model covers 100 % of the LgmC peptide, and predicts a primarily alpha helical secondary structure along with some beta sheet structure (Figure 26) (Kelley and Sternberg 2009). Mutation of the conserved amino acids (D80G D81G, H130G, and E313G) resulted in a lack of LgmC function, and these residues are located in the vicinity of one another and around the same cavity on the predicted structure of LgmC. I hypothesize this region may function as the active site of LgmC.

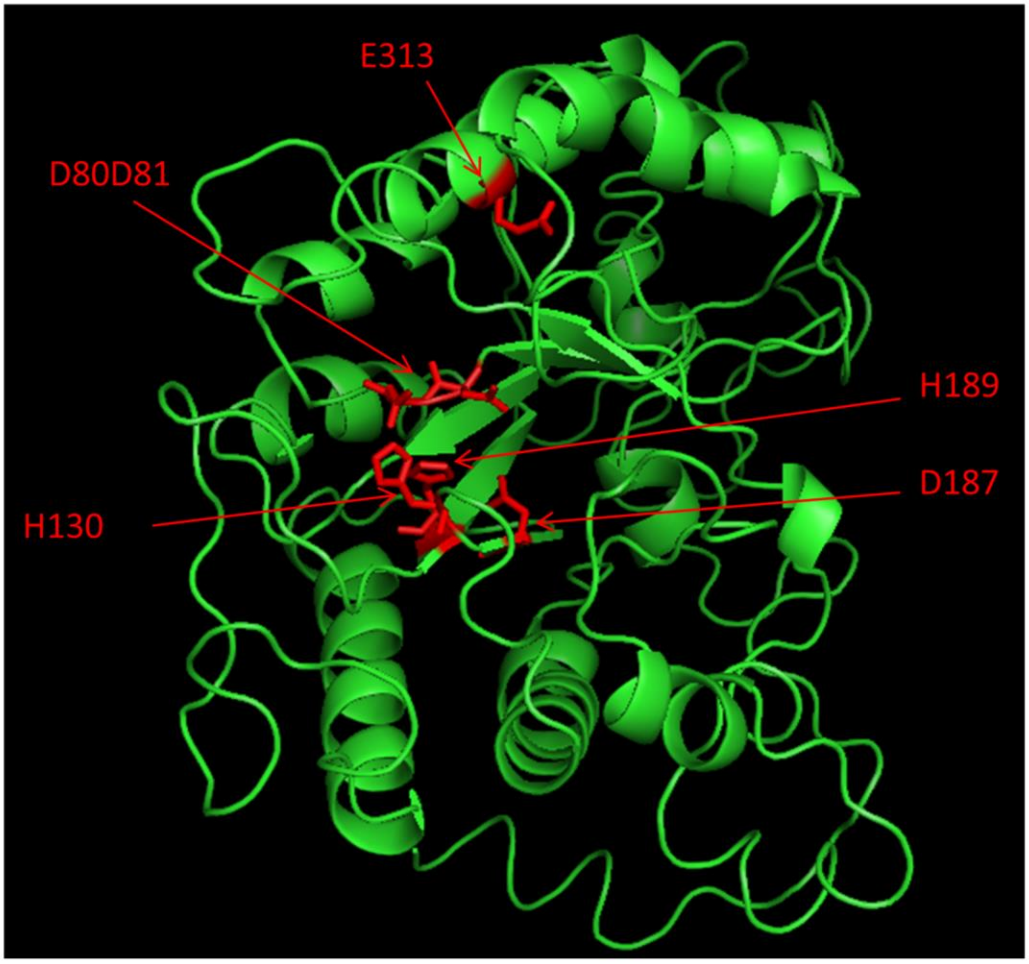


Figure 26. Phyre predicted model of the structure of LgmC
Model of LgmC structure by the Protein Homology/analogy Recognition Engine (Phyre) (Kelley and Sternberg 2009). LmgC protein is green. Residues that resulted in a lack of GlcN modification of lipid A when mutated are in red. PyMOL was used to visualize this structure.

3.6 The *lgm* locus in other *Bordetella* species

Thus far, I have analyzed the *lgm* locus of *B. pertussis* strain BP338, which is a derivative of the Tohama I strain, and demonstrated *lgmA*, *lgmB*, and *lgmC* are required for GlcN modification of lipid A in this strain. However, analysis of the lipid A structures of additional *Bordetella* species has revealed that other *Bordetella* also have GlcN-modified lipid A. Specifically, *B. bronchispetica*, *B. parapertussis*, and *B. avium* have been shown to have the lipid A GlcN modification, whereas *B. hinzii* and *B. trematum* do not (Novikov, Shah, et al. 2013). I hypothesized the lack of GlcN modification in *B. hinzii* and *B. trematum* was because of a clear genetic reason, such as the lack of the *lgm* locus in these species. However, the genome sequences of *B. hinzii* and *B. trematum* were not publically available at the time. I attempted to sequence the *lgm* locus in these species with primers specific to the *B. pertussis* or *B. avium* lgm loci. However, I was unsuccessful in my efforts, and it later became clear that this is likely due to the low degree of nucleotide sequence similarity of the *Bordetella* lgm loci, despite conserved amino acid residues in some regions of the Lgm proteins. Therefore we sequenced these genomes (strains *B. hinzii* ATCC 51730 and *B. trematum* CCUG 13902) on a MiSeq (Illumina) platform. Sequencing resulted in ~700-fold coverage for each genome and assembly with Velvet algorithms resulted in 98 contigs for *B. hinzii* and 83 contigs for *B. trematum* (Shah, Moksa, et al. 2013). I then used the tblastn algorithm, which searches a translated nucleotide sequence with a protein sequence query, to search the *B. hinzii* and *B. trematum* genomes, and the other *Bordetella* species, for the LgmABCDE proteins. Figure 27 summarizes the results of this analysis. I note that *lgmA*, *lgmB*, *lgmC*, *lgmD*, and *lgmE* were present in the other subspecies, *B. bronchiseptica* and *B. parapertussis*.

B. avium, alternatively, contains the three required lgm genes: *lgmABC*, along with *lgmE* (locus tag BAV2926), but no *lgmD*. *B. avium* has GlcN-modified lipid A, suggesting a fully functional set of *lgmABC* genes (Novikov, Shah, et al. 2013). Further analysis of the *B. avium lgm* locus revealed the LgmC homolog, BAV2925, is annotated to have a truncated N-terminus compared to the other *Bordetella*

LgmC proteins (Figure 28). This truncated LgmC would no longer encode conserved amino acids in the N-terminal region that are likely required for LgmC function, specifically the residues equivalent to *B. pertussis* LgmC D80 and D81 (see Section 3.5). However, there are methionine residues present upstream of the predicted start site of *B. avium lgmC* that may act as a start codon for this gene (Figure 28). If the *B. avium lgmC* gene begins translation at one of these upstream methionine residues, the aforementioned conserved D80 and D81 residues would be expressed at the N-terminus of the protein, which would agree with the presence of a functional LgmC, as required for the GlcN modification.

Analyzing the *B. hinzii* and *B. trematum* genomes revealed the presence of homologs of *lgmABC* in both species, which are the only genes that have thus-far been shown to be required for GlcN modification of lipid A. Therefore, this reflected the hypothesis that the unmodified lipid A produced by these species is not due to the lack of an *lgm* locus.. Furthermore, the residues I had previously identified as putative active sites for LgmA and LgmC were also conserved amongst the *Bordetella* species genomes I analyzed. *B. hinzii* has *lgmE* in a similar position as in the other *Bordetella* species, though *lgmD* is found elsewhere in the genome. *B trematum* does not possess *lgmE,* but like in *B. hinzii*, *lgmD* is found at a different location on the chromosome. Earlier in this chapter I have shown a *B. pertussis* strain BP338 that lacks LgmD and LgmE still had the GlcN modification, so the lack of *lgmE* in *B. trematum* is unlikely to be the cause of unmodified lipid A. Next, I aligned the LgmA proteins of the different *Bordetella* species to identify amino acids residues that are conserved amongst the species that have GlcN-modified lipid A (*B. pertussis*, *B. bronchiseptica*, *B. parapertussis*, and *B. avium*) but are more variable in *B. hinzii* and *B. trematum*, which do not have this modification. I hypothesized differences in these residues may be the reason why *B. hinzii* and *B. trematum* lack the GlcN modification. I identified the *B. pertussis* LgmA residues D159 and W163 (Figure 21B) and mutated these residues to the corresponding residues found in *B. hinzii* to generate the LgmA mutants D159N, W163R, and D159N W163R on a vector. I then complemented BP338LgmAKO with these mutated *lgmA* genes and tested the

ability of these complemented strains to activate hTLR4 to assess the presence or absence of the lipid A GlcN modification. The D159N, W163R, and D159N W163R mutant *lgmA* successfully complemented *lgmA*KO, as evident by the rescue of hTLR4 activation levels, showing these amino acids are not essential for LgmA activity (Figure 22). These results suggest the differences between *B. hinzii* and *B. trematum* LgmA at these residues are not the reason for the lack of GlcN-modified lipid A in these species. I also mapped the positions of these residues onto the Phyre-predicted model of LgmA, and found they are predicted to be located elsewhere on the structure compared to the previously identified putative active site residues (Figure 23). Therefore, I have not found a clear genetic reason for the lack of GlcN-modified lipid A in *B. hinzii* and *B. trematum*.

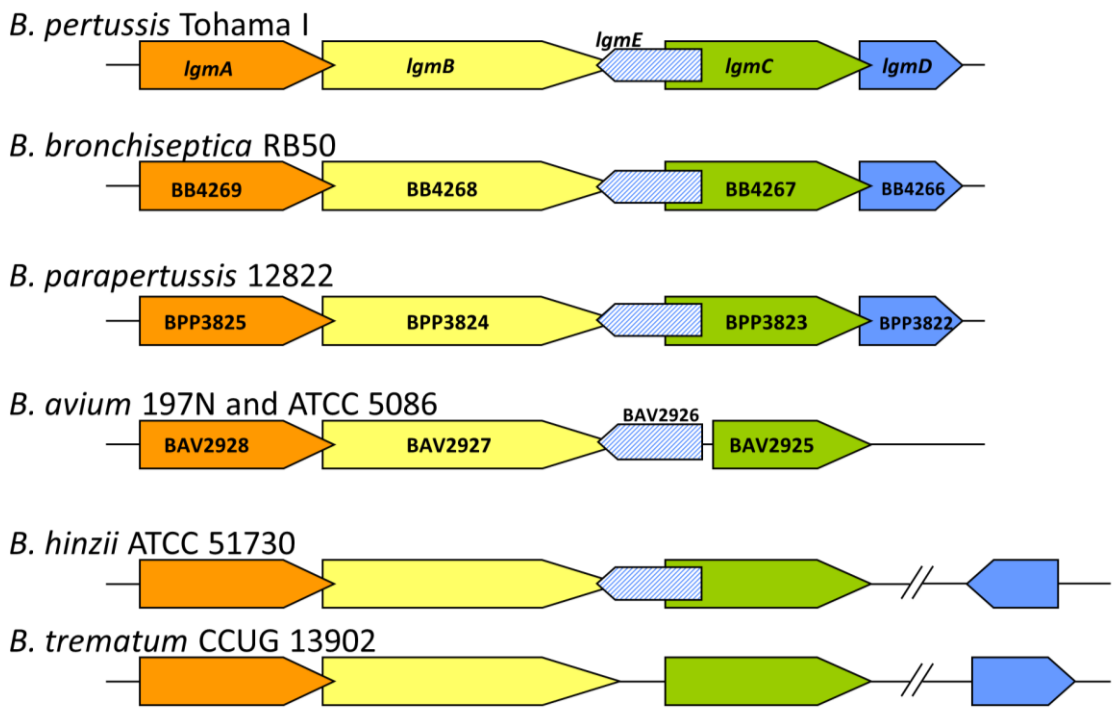


Figure 27. *lgm* locus of the sequenced *Bordetella* species
Schematic representation of the *lgm* locus in *Bordetella* species: *B. pertussis* Tohama I, *B. bronchiseptica* RB50, *B. parapertussis* 12822, *B. avium* 197N and ATCC 35086, *B. hinzii* ATCC 51730, and *B. trematum* CCUG 13902. *lgmA* is orange, *lgmB* is yellow, *lgmC* is green, *lgmD* is blue and *lgmE* is diagonally-striped in light blue. Double diagonal black line represents the following region is in a different location of the chromosome. The directionality of *lgmD* in *B. trematum* is unknown in reference to *lgmABC*. Locus tags are given for *B. bronchiseptica* RB50, *B. parapertussis* 12822 and *B. avium* 197N; the open reading frame for *lgmE* has not been assigned a locus tag and therefore remains unlabeled. Modified from Novikov *et al*. 2013(Novikov, Shah, et al. 2013), used with permission.

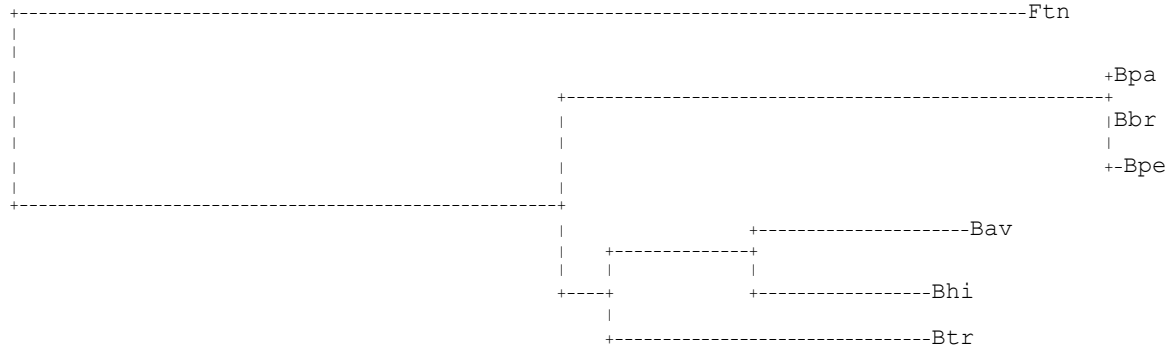
Figure 28. Comparison of the start of LgmC between *Bordetella* species
ClustalW2 (Thompson, Higgins, et al. 1994) alignment of LgmC protein sequence of *Bordetella* species: Bpe, *B. pertussis* Tohama I; Bbr, *B. bronchiseptica* RB50; Bpa, *B. parapertussis* 12822; Bav, *B. avium* 197N and ATCC 35086; Bhi, *B. hinzii* ATCC 51730; and Btr, *B. trematum* CCUG 13902. *B. avium* LgmC: annotated sequence is black, the upstream region (translated in frame) is red, X highlighted in red is a stop codon, red 'M's highlighted in blue are upstream methionine codons that could act as start codons. 'DD' highlighted in green are conserved aspartic acid residues. '*': Identical amino acid residues; ':': conserved amino acid residues; '.': semi-conserved amino acid residues.

To explore the relationship between LgmABCDE proteins from the *Bordetella* species I generated neighbour-joining trees based on ClustalW2 alignments (Figure 29). For the LgmC tree, I used the longest possible *B. avium* LgmC protein, as depicted in Figure 28, though this is still truncated compared to the other *Bordetella* LgmC proteins. The three subspecies, *B. pertussis*, *B. bronchiseptica*, and *B. parapertussis*, have very closely related Lgm proteins compared to the other *Bordetella* species. *B. avium*, *B. hinzii*, and *B. trematum* cluster in a branch separate from the three *Bordetella* subspecies. However, regarding LgmA, LgmB, and LgmC, the three Lgm proteins shared by *B. avium*, *B. hinzii*, and *B trematum*, the two bird-associated species, *B. hinzii* and *B. avium*, cluster separately compared to *B. trematum*. Therefore, even though *B. hinzii* Lgm proteins are more similar to *B. avium* compared to *B. trematum*, *B. avium* has GlcN-modified lipid A, and therefore a functional Lgm pathway, whereas *B. hinzii* and *B. trematum* do not (Novikov, Shah, et al. 2013). LgmD and LgmE are both four-transmembrane helix proteins (Figure 7, Figure 13), and have very low sequence identity between the different *Bordetella* species (Appendix B). The LgmD and LgmE proteins cluster with themselves (Figure 29F), supporting my annotation of the ORFs in the three subspecies as the *lgm* genes and in the sequenced species *B. hinzii* and *B. trematum*, even though *B. trematum* LgmD appears to be distantly related to the other LgmD proteins (Figure 29D).

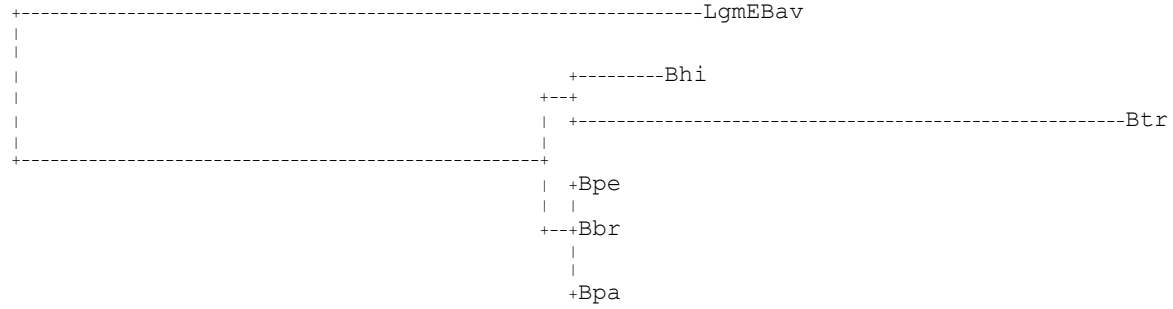
A. LgmA

B. LgmB

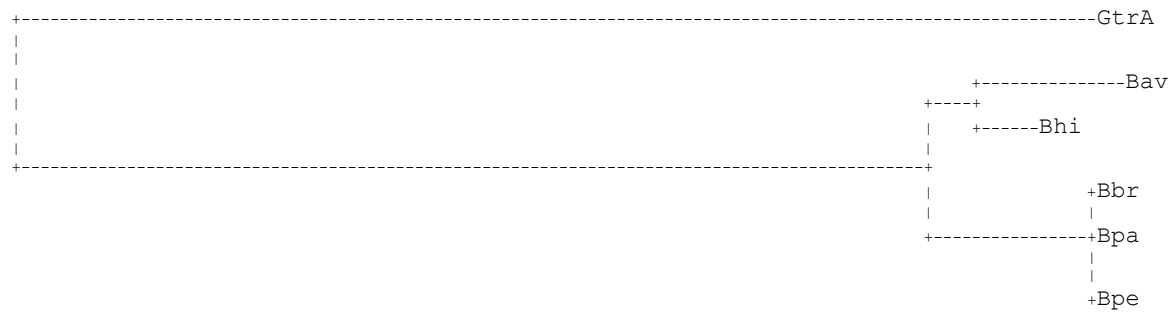
C. LgmC



D. LgmD



E. LgmE



F. LgmD and LgmE

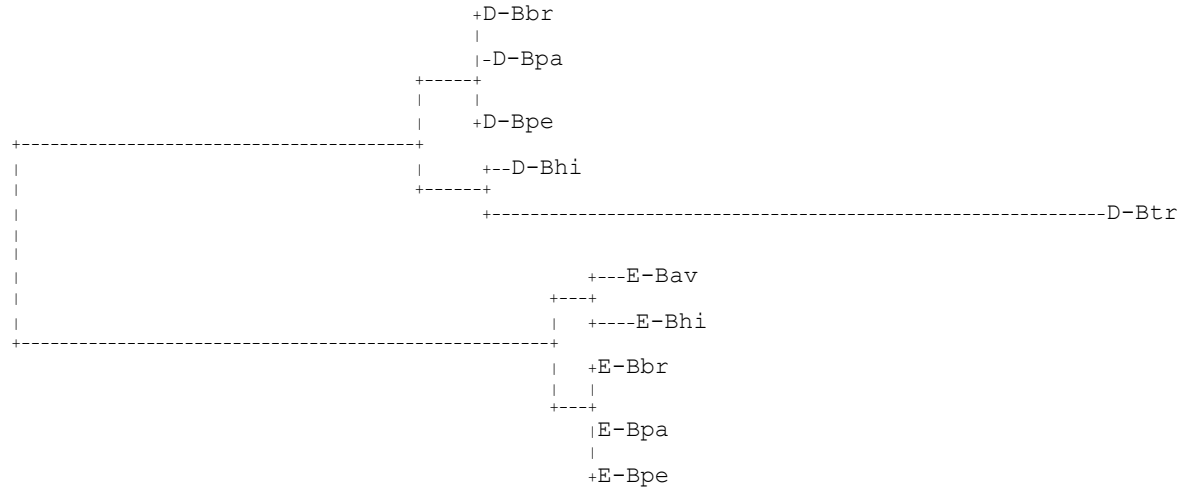


Figure 29. Neighbour-joining trees of the *lgm* locus genes in *Bordetella* species
Neighbour-joining trees of: A) LgmA, B) LgmB, C) LgmC, D) LgmD, E) LgmE, F) LgmD and LgmE. Bpe, *B. pertussis* Tohama I; Bbr, *B. bronchiseptica* RB50; Bpa, *B. parapertussis* 12822; Bav, *B. avium* 197N; Bhi, *B. hinzii* ATCC 51730; and Btr, *B. trematum* CCUG 13902. A) ArnC from *E. coli* K-12 rooted the LgmA tree. B) ArnT from *E. coli* K-12 rooted the LgmB tree. C) Ftn, Ftn_0544 from *F. novicida* U112 rooted the LgmC tree. D) LgmEBav, *B. avium* LgmE rooted the LgmD tree. E) GtrA from *E. coli* K-12 rooted the the LgmE tree. F) prefix 'D-' indicates LgmD, prefix 'E-' indicates LgmE. Neighbour-joining trees were built based on ClustalW2 (Thompson, Higgins, et al. 1994) alignments of the protein sequences (see Appendix B for ClustalW2 alignments) using the Methodes et Algorithmes pour la bio-informatique LIRMM website (http://www.phylogeny.fr/version2_cgi/one_task.cgi?task_type=bionj) (Gascuel 1997).

3.7 Discussion

I found that *lgmA*, *lgmB*, and *lgmC* are required for the modification of *B. pertussis* lipid A with GlcN, while *lgmD*, *lgmE*, and locus tag BP1945 are not. Mutational analysis of LgmA and LgmC identified residues D76, D77, D127, and D129 in LgmA and D80, D81, H130, D187, H189, and E313 in LgmC that may be part of a putative active site in these enzymes. Furthermore, I did not uncover any clear genetic reason for the lack of lipid A GlcN modification in *B. hinzii* and *B. trematum*.

I also demonstrated LgmA transfers GlcNAc onto a molecule that localizes to the lipid fraction. LgmC has recently been shown to function as a deacetylase by removing the acetyl group from C55P-GlcNAc to produce C55P-GlcN (Llewellyn, Zhao, et al. 2012), as I had originally hypothesized (Figure 7). Both LgmA and LgmC are required to modify the lipid A of *B. pertussis* with GlcN, and we predict these two enzymes function in the same pathway. Therefore, since the substrate for LgmC is C55P-GlcNAc, and LgmA transfers GlcNAc onto a lipid, the acceptor molecule for transfer by LgmA is probably the inner membrane carrier lipid C55P. ArnC, a homolog of LgmA in *E. coli* and *Salmonella*, also uses C55P as the acceptor when it transfers Ara4FN onto the carrier lipid in the Ara4N-lipid A modification pathway. This further supports the deduction that LgmA functions in the *B. pertussis* lipid A GlcN modification pathway by transferring GlcNAc onto C55P.

I had originally predicted that a flippase is required for the Lgm pathway, as ArnE and ArnF play this role in the Ara4N lipid A-modification pathway. Yet, my candidates, *lgmD*, *lgmE*, and BP1945, are not required for GlcN modification of lipid A, either individually or in combination with *lgmD*. It is possible that the presence of just one of these three genes may be sufficient to flip C55P-GlcN, since I did not concurrently knockout all three genes. If all three proteins, LgmD, LgmE, and BP1945, functioned in the same role as a C55P-GlcN flippase, I would predict knocking out two of these three proteins would decrease levels of lipid A GlcN modification in the bacterium. However, I did not observe any significant

decrease in hTLR4 activation by these double knock-out strains compared to wild type BP338, indicating these strains do not have a decrease in GlcN-modification of lipid A. Therefore, it is more likely that another protein functions as a flippase in this system, or no flippase is required, and spontaneous flipping of C55P-GlcN from the cytoplasmic to the periplasmic face of the IM is sufficient for the Lgm pathway.

Though translation of *lgmE* to produce the protein LgmE is not required for GlcN-modification of lipid A, generation of the *lgmE* transcript is not prevented by the mutations in the *lgmE*KO strains, in which a STOP codon was introduced at residue Gly15. I showed *lgmE* is transcribed in *B. pertussis* strain BP338, and since this transcript would be exactly complementary to, at the very least, the start of *lgmC* mRNA and to the end of *lgmB* mRNA, *lgmE* mRNA may be involved in the Lgm system, perhaps at a regulatory level.

The lack of GlcN-modification in *B. hinzii* and *B. trematum* lipid A remains unclear, since both strains have a complete copy of *lgmA*, *lgmB*, and *lgmC*, the required genes for the Lgm pathway. One possibility is that an as yet unidentified gene that is required for GlcN modification of lipid A is present in *B. pertussis*, and the other *Bordetella* species that possess this modification, but is absent from *B. hinzii* and *B. trematum*. Another possibility to consider is the presence of an inactivating mutation in *lgmA*, *lgmB*, or *lgmC* in these two species, though differences in LgmA at the residues equivalent to *B. pertussis* D159 and W163 are not responsible for this lack of modification. In *B. pertussis*, the *lgm* locus is part of the *Bordetella* virulence gene (*Bvg*) regulon (Marr, Tirsoaga, et al. 2008). Genome analysis indicates that the *bvgAS* two-component regulatory system genes are present in *B. hinzii* and *B. trematum*, suggesting that the lack of GlcN modification of lipid A is not due to the absence of the *bvgAS* genes. Culture conditions can also affect the expression of genes in the *lgm* locus (Marr, Tirsoaga, et al. 2008); however, all strains were grown under similar conditions when analyzed for lipid A structure (Novikov, Shah, et al. 2013).

Therefore, the lack of GlcN modification in *B. hinzii* and *B. trematum* may be due to the aforementioned differences in the *lgm* locus or to other variations in gene regulation and expression.

Mutational analysis of LgmA and LgmC identified key residues that may be involved in the function of these proteins. LgmA is a GT2 family glycosyltransferase with a GT-A fold, which is characterized by two adjacent β/α/β Rossmann domains (Lairson, Henrissat, et al. 2008). Glycosyltransferases with a GT-A fold often contain a conserved DXD motif that functions in the coordination of a divalent cation and/or ribose in the active site, and this has previously been considered a 'signature' of GT-A fold glycosyltransferases. However, a DXD motif is not found in all GT-A fold glycosyltransferases, and many non-glycosyltransferase enzymes possess conserved DXD motifs (Lairson, Henrissat, et al. 2008). In LgmA, I identified a conserved DXD motif that, when either one or both aspartates were mutated to glycine, resulted in no GlcN modification of *B. pertussis* lipid A. This suggests the aspartic acid residues of the conserved DXD motif in LgmA (D127 and D129) are required for activity. Whether these residues are involved in coordination of a cation remains to be seen.

LgmC is a member of the YdjC-like family of proteins. By aligning LgmC with other members of this protein family, including TTHB029 from *Thermus thermophiles* HB8, I identified several conserved residues: D80, D81, H130, D187, D189, and E313. Mutation of these residues resulted in a lack of GlcN-modified lipid A in *B. pertussis*, suggesting they are required for LgmC function. The structure of TTHB029 has been determined by crystallography (Imagawa, Iino, et al. 2008), and five hypothetical functionally important motifs have been identified. Motif 1 is DDXG and aligns with D80 and D81 of LgmC, which are part of a DDXG motif in LgmC. Motif 2 is GXH and the histidine of this motif in TTHB029 aligns closely with H130 in LgmC, which is also part of a GXH amino acid pattern. Motif 3 is a THXDXH motif in TTHB029, though only the DXH part of this motif is conserved in LgmC (D187 and H189). Motifs 1, 2, and 3 have been shown to bind Mg^{2+} in the TTHB029 crystal structure, suggesting

these residues may also be involved in coordinating divalent cations in LgmC. The last two motifs, 4 and 5, are presumed to be involved in the active site. Motif 4 contains H215 in TTHB029, which is equivalent to H295 in LgmC, and motif 5 contains R232 in TTHB029, which aligns with R310 in LgmC. I identified E313 of LgmC as a semi-conserved residue in the YdjC protein family, which is found close to the conserved residue R310, and mutational analysis suggests E313 may be involved in LgmC function. Motif 4 and 5 may be expanded to include additional conserved residues, following further mutational experiments.

In these mutational experiments of LgmA and LgmC, large amino acids, such as aspartic acid and histidine, were changed into glycine, which has a hydrogen atom as a side chain. This is a very drastic change in amino acid shape, and could result in deformation of the secondary structure of the protein in this region, which could result in abrogation of protein function, even if these specific residues are not directly involved in enzyme activity. Therefore, to confirm the requirement of the identified amino acids in protein function, further mutational analysis of these proteins is required. For example, mutating aspartic acid to leucine or asparagine, both of which have a similar shape to aspartic acid but lack the carboxylic acid group, may clarify the importance of this anionic carboxylic acid group.

lgmA and *lgmB* are regulated by the BvgAS virulence regulatory system in *B. pertussis*, and by an additional, unknown system, since *lgmA* and *lgmB* are optimally transcribed during growth in SS broth compared to growth on BG agar, which not observed for *vag8*, a known Bvg-regulated protein (Marr, Tirsoaga, et al. 2008). Despite the unknown component of *lgmA* and *lgmB* regulation, Bvg-regulation of these genes suggests involvement of the *lgm* locus in the virulence of *B. pertussis*. However, further analysis is required to elucidate the role of this GlcN modification, and other lipid A modifications, in pathogenesis and interaction with the host.

Chapter 4: The biological effects of lipid A modifications in *B. pertussis*

4.1 Introduction

Changes to lipid A structure can affect many factors, such as membrane stability, resistance to CAMPs, and activation of TLR4 (Caroff, Karibian, et al. 2002, Needham and Trent 2013). In the previous chapter I described how the *lgm* locus, specifically *lgmA*, *lgmB*, and *lgmC*, are required for modification of *B. pertussis* lipid A with GlcN at the phosphate groups. In this chapter, I will explore the relationship between *B. pertussis* lipid A structure and these previously mentioned factors. First I will assess the effect of the lipid A GlcN modification on the susceptibility of *B. pertussis* to a variety of CAMPs and on the stability of the bacterial membrane. Then, I will determine the individual and combined contribution of two structural features of *B. pertussis* lipid A to activation of hTLR4: the GlcN modification of the phosphate groups and the length of the C3' acyl chain.

4.2 Effect of GlcN modification on resistance to CAMPs and membrane stability

B. pertussis colonizes the human respiratory tract, and therefore must contend with bacterial-clearance mechanisms present in the nasal cavity, trachea, and lung, including several CAMPs (de Gouw, Diavatopoulos, et al. 2011, Laube, Yim, et al. 2006). One mechanism for resistance to CAMPs in some bacteria is modification of lipid A, such as the addition of Ara4N by ArnT or phosphoethanolamine by EptA to the phosphate groups in *Salmonella* lipid A (Needham and Trent 2013). Modification of the lipid A phosphates by GlcN in *B. pertussis* may also affect resistance to CAMPs, along with stabilization of the OM. We had previously suggested this modification did not affect resistance to the CAMP polymyxin B (Marr, Hajjar, et al. 2010). However, in these earlier polymyxin B susceptibility experiments, bacteria were grown on BG agar with a range of polymyxin B concentrations. Since *lgmA* and *lgmB* (BP0399 and BP0398, respectively) have reduced levels of transcription during growth on BG agar compared to SS broth, the lipids A from these bacteria were likely not optimally modified with GlcN (Marr, Tirsoaga, et

al. 2008). Thus, I hypothesized the GlcN modification may increase resistance to CAMPs, if the bacteria are grown in SS broth, and not on BG agar.

4.2.1 GlcN modification increases resistance to CAMPs

I tested the effect of *B. pertussis* lipid A GlcN modification on CAMP resistance by incubating wild-type BP338 bacteria and bacteria lacking the GlcN modification (BP338LgmABCDKO) with a range of CAMP concentrations. The bacteria were all initially grown in SS broth, to ensure optimal lipid A GlcN modification, before incubation with CAMPs or antibiotics for 2 h. I used this 2 h incubation method instead of growing the bacteria over 48 h in the presence of CAMPs to determine the minimum inhibitory concentration because recent studies have shown growth in the presence of sub-inhibitory concentrations of cationic peptides can upregulate LPS modifications involved in resistance to these very peptides (Fernandez, Jenssen, et al. 2012). I tested different structural classes of CAMPs, including: cyclic, lipidated CAMPs (polymyxin B and polymyxin E); alpha helical CAMPs (LL-37, HHC-10, and CP28); and an extended conformation CAMP (indolicidin) (Table 6). As shown in Figure 30, a greater percent of BP338 survived exposure to the different CAMPs when compared to its mutant, BP338LgmABCDKO, that lacks GlcN modification on lipid A. The greatest difference in survival between these two strains was observed when the bacteria were incubated with polymyxin B and polymyxin E. Complementation of the GlcN mutant with pPtacLgmABCD, which restored GlcN modification of LPS (Table 7), also restored wild-type levels of susceptibility to polymyxin B (Figure 31). Therefore, the lipid A GlcN modification in *B. pertussis* increases resistance to a variety of CAMPs. I observed no difference in susceptibility to another positively-charged antibiotic, the aminoglycoside gentamicin, between these two *B. pertussis* strains (Figure 30). This highlights the specificity of the GlcN modification resistance mechanism to CAMPs.

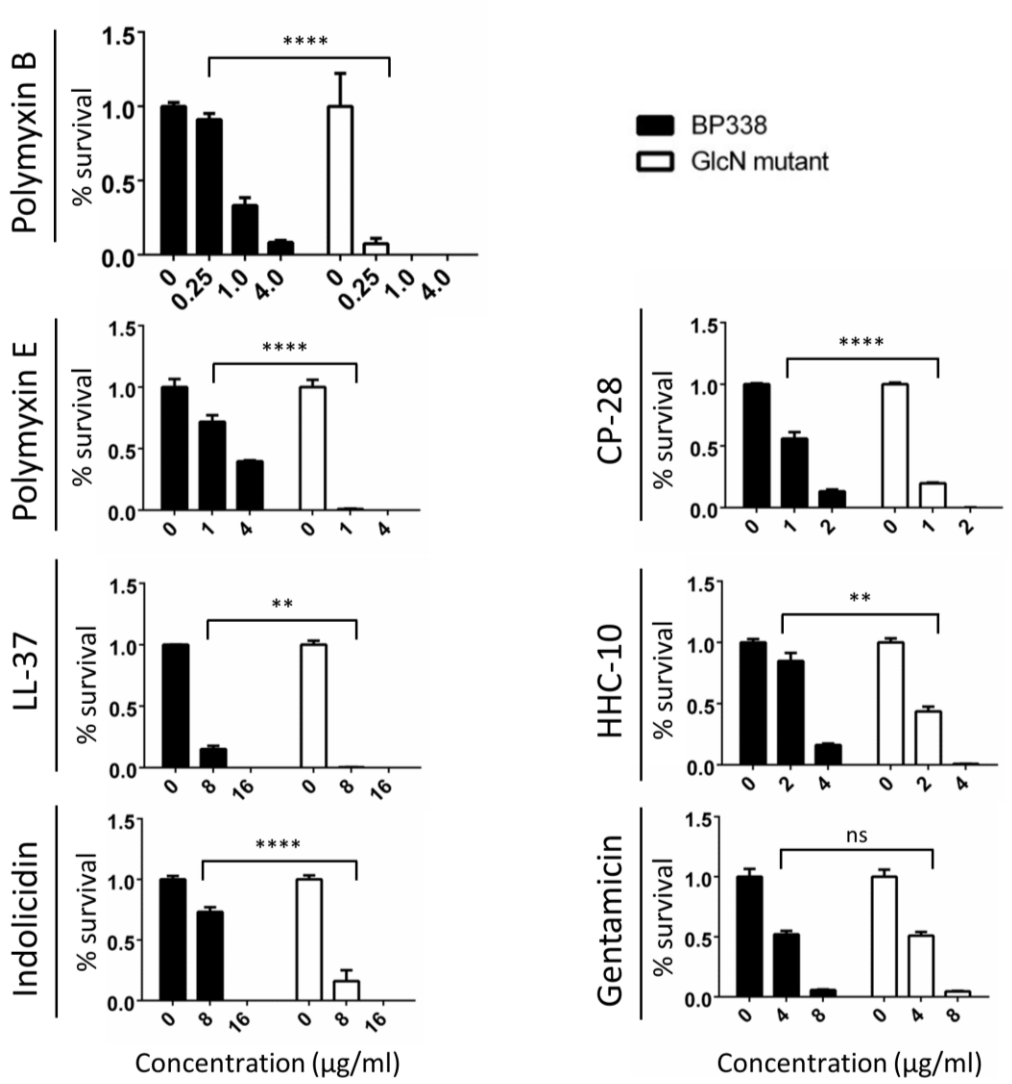


Figure 30. Influence of the *B. pertussis* lipid A GlcN modification on CAMP susceptibility
B. pertussis lipid A GlcN modification increased resistance to CAMPs. *B. pertussis* BP338 and BP338LgmABCDKO (GlcN mutant) strains were incubated with a range of killing agents including bacterial polymyxin B and polymyxin E (colistin), human LL-37, bovine indolicidin, insect CP28, and synthetic HHC-10. Gentamicin was used as a control. The concentrations of the killing agent are along the x-axis and percent survival (% survival) is along the y-axis. Graphs show the results of one representative experiment of three, n = 3 replicates per experiment. Statistical significance was determined by ANOVA, with a Bonferroni post-test to compare groups. **** $P < 0.0001$; ** $P < 0.01$; ns = no significant difference.

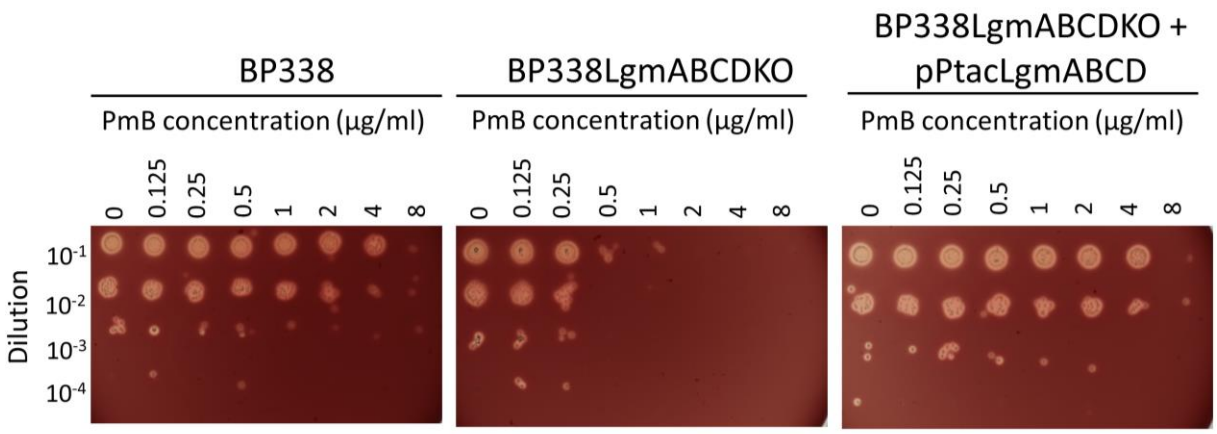


Figure 31. Complementation of BP338LgmABCDKO rescues resistance to polymyxin B
B. pertussis BP338, BP338LmgABCDKO, and BP338LgmABCDKO + pPtacLgmABCD (complemented) strains were incubated with a range of polymyxin B (PmB) concentrations for 2 hours, then diluted in a 1/10 dilution series before 2 µl of each sample at each dilution was spotted onto a BG agar plate. These are results of one representative experiment of three.

4.2.2 GlcN modification increases resistance to OM perturbation

The OM provides a protective barrier against many antimicrobial factors in the airway, such as lysozyme, which needs to permeate the OM to gain access to its substrate peptidoglycan (Laube, Yim, et al. 2006). CAMPs interact with the OM at sites where adjacent LPS molecules are bridged by divalent cations, causing perturbation of the OM and consequent self-promoted uptake of the CAMP (Hancock 1997). Any changes to the integrity of the OM, e.g. to engender increased resistance to CAMPs, have the potential to affect the ability of pathogens to survive in the respiratory tract. I tested the effect of lipid A GlcN modification on OM stabilization by incubating BP338 and the GlcN mutant, BP338LgmABCDKO, with EDTA, lysozyme, or both. EDTA perturbs the OM by chelating cations that bridge the phosphate groups of LPS molecules. I found that BP338 was more resistant to killing by EDTA alone compared to the GlcN mutant (Figure 32), showing that the GlcN modification stabilizes the OM, likely by decreasing the need for divalent cation stabilization. There was no significant difference in survival when the bacteria were incubated with lysozyme alone. However, membrane perturbation by 2 mg/ml EDTA rendered both bacterial strains more susceptible to killing by lysozyme, although the GlcN mutant had a larger decrease in survival when compared to BP338. This supports the conclusion that the GlcN modification on *B. pertussis* lipid A increases resistance to perturbation of the OM.

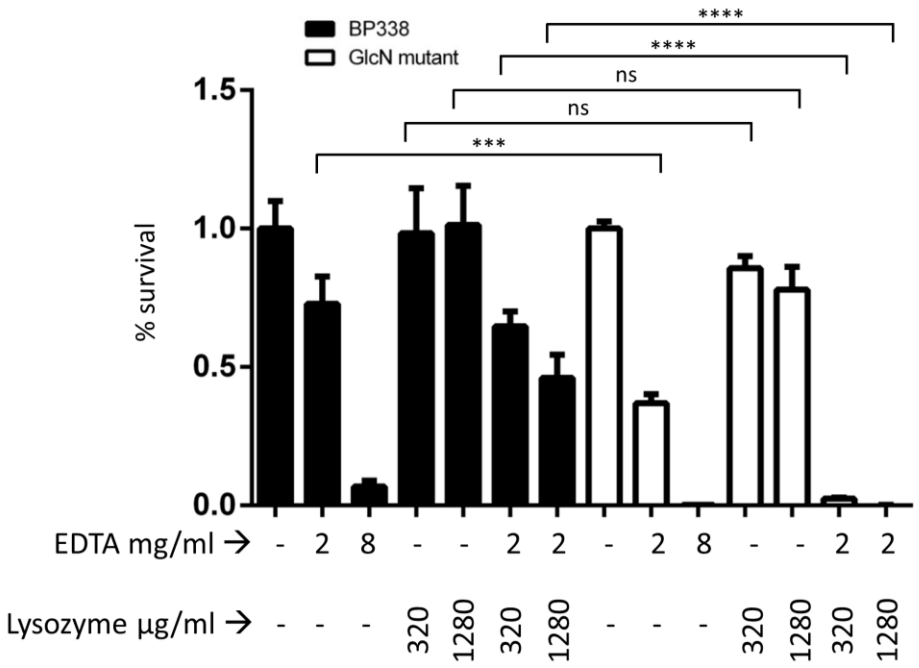


Figure 32. Influence of the *B. pertussis* lipid A GlcN modification on OM stabilization
B. pertussis lipid A GlcN modification increased resistance to OM perturbation. *B. pertussis* BP338 and BP338LgmABCDKO (GlcN mutant) strains were incubated with a range of killing agents including EDTA, lysozyme, and EDTA+lysozyme. The concentrations of the killing agent are along the x-axis and percent survival (% survival) is along the y-axis. Graph shows the results of one representative experiment of three, n = 3 replicates per experiment. Statistical significance was determined by ANOVA, with a Bonferroni post-test to compare groups. **** $P < 0.0001$; *** $P < 0.001$; ns = no significant difference.

4.3 *lgm* locus and *lpxA* are involved in the differences in lipid A structure between strains BP338 and 18-323

Previously, we have shown that the lack of GlcN-modified lipid A results in a decrease in activation of hTLR4 by *B. pertussis* BP338 LPS. However, the LPS of another *B. pertussis* wild-type strain, 18-323, exhibited even lower levels of hTLR4 activation (Figure 5) (Marr, Novikov, et al. 2010). Analysis of the structure of 18-323 lipid A revealed two differences compared to BP338 lipid A: 1) 18-323 lacks the GlcN modification, and 2) the C3' acyl chain of 18-323 lipid A is only 10 or 12 carbons long, compared to BP338 lipid A, which has a C14-OH acyl chain at this position (Figure 4) (Marr, Novikov, et al. 2010). Consequently, I hypothesized each structural difference between BP338 and 18-323 lipid A individually affect activation of hTLR4 by LPS. To test this hypothesis, I planned to gradually modify the lipid A of 18-323 to resemble that of BP338. In order to do this, I first needed to determine the genetic basis for the lack of GlcN and the production of shorter C3' acyl chain length in 18-323 lipid A.

4.3.1 Difference in GlcN modification is due an incomplete *lgm* locus

In *B. pertussis* strain BP338, the *lgm* locus, specifically *lgmA*, *lgmB*, and *lgmC*, are required for the lipid A GlcN modification. Prior to the completion of the *B. pertussis* 18-323 genome (Park, Zhang, et al. 2012), the individual raw sequence reads for this strain had been released, and these were analyzed for the presence of *lgmA*, *lgmB*, *lgmC*, and *lgmD* by blastn algorithms. This analysis revealed that 18-323 possesses a complete *lgmA*, but *lgmB* has a mutation at bp 981 where bases TT are deleted, resulting in a frameshift, and an early stop codon (Figure 33AB). No matches to *lgmC* or *lgmD* were found in the 18-323 genome sequence. The absence of *lgmC* and *lgmD* in the 18-323 genome was confirmed via PCR using primers to amplify internal fragments of each of the *lgm* locus genes (Figure 33C). In addition, the existence of the dinucleotide deletion in *lgmB* was verified by sequencing using primers spanning the deletion site. The absence of a complete *lgm* locus in 18-323 thus explains the lack of GlcN-modified lipid A in this strain.

To corroborate these results, I complemented 18-323 with pPtacLgmABCD, thereby introducing the *lgm* locus of BP338 into 18-323. Negative-ion MALDI-MS analysis of the lipid A from this strain shows an extra set of peaks at *m*/*z* 1664 and 1692, *i.e.* 161 mass units (the mass of GlcN) higher relatively to the corresponding peaks of major unmodified penta-acylated lipid A molecular species observed in the wild type strain spectrum at *m*/*z* 1503 and 1531 (Figure 34AB). These additional peaks indicate the presence of C10-OH or C12-OH C3′ acyl chains and the addition of a GlcN modification in 18-323 lipid A.

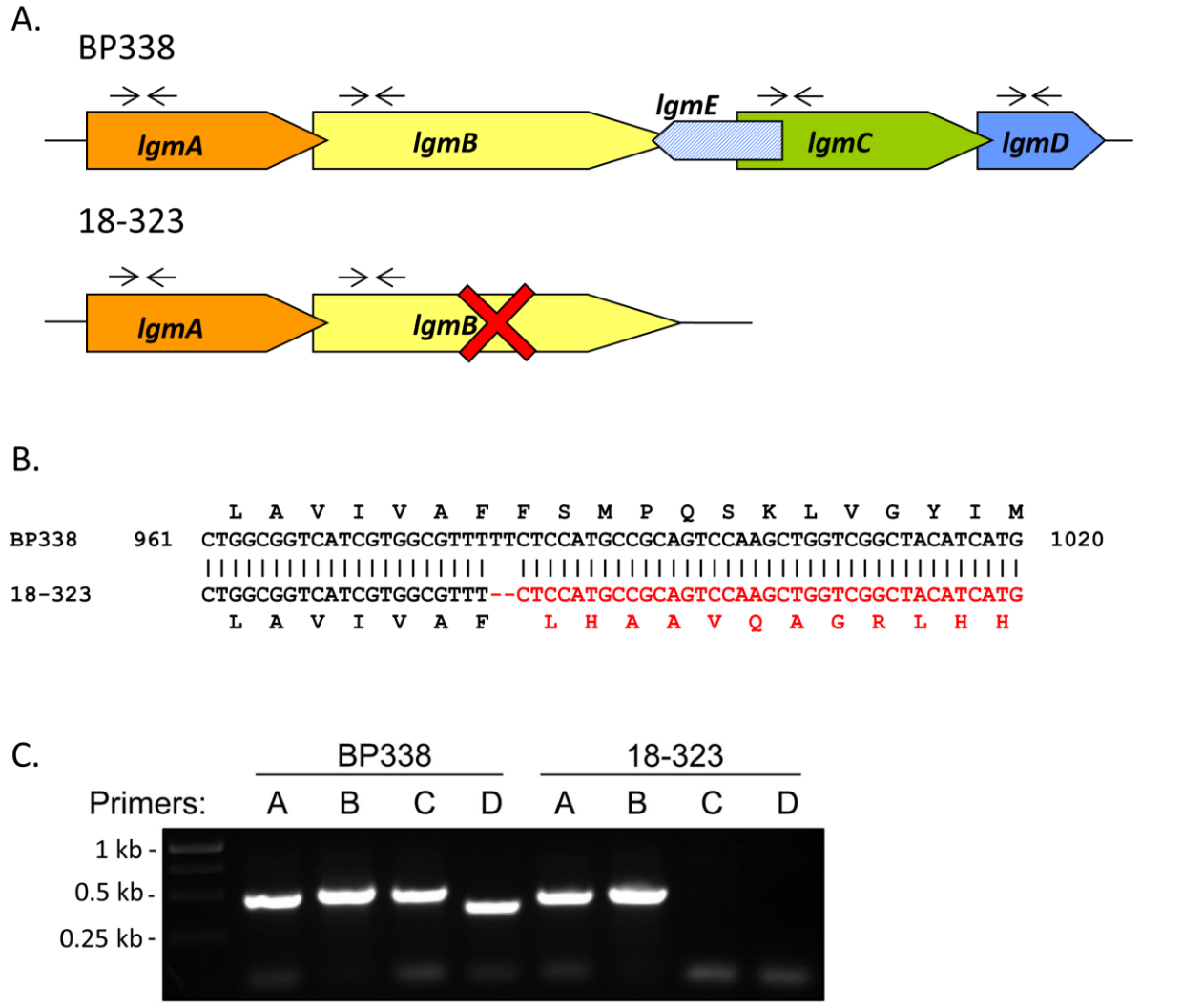


Figure 33. Genetic analysis of the *lgm* locus of *B. pertussis* strains BP338 and 18-323
A) Lgm loci of BP338 and 18-323, as determined by sequence analysis. These data were provided by the pathogen genomics group at the Wellcome Trust Sanger Institute and can be obtained from the Sanger Institute website. The red 'X' represents a TT deletion mutation at bp 981 of *lgmB* in 18-323; the forward arrows and the reverse arrows illustrate the annealing sites of the primers used for the PCR shown in (C). B) Comparison of *lgmB* sequence between BP338 (top nucleotide sequence) and 18-323 (bottom nucleotide sequence) in the region of the TT deletion in 18-323 *lgmB*. LgmB amino acid code is above (BP338) or below (18-323) the nucleotide sequence, such that the amino acid letter aligns with the third nucleotide in the corresponding codon. The mutation in 18-323 *lgmB*, and the downstream sequence is highlighted to point out the frameshift mutation. C) PCR of the *lgm* locus genes in BP338 and 18-323 using gene-specific primers. Expected positive bands: 0.48 kb (*lgmA* primers, A), 0.51 kb (*lgmB* primers, B), 0.50 kb (*lgmC* primers, C), and 0.40 kb (*lgmD* primers, D). Modified from Shah *et al*. 2013 (Shah, Albitar-Nehme, et al. 2013), used with permission.

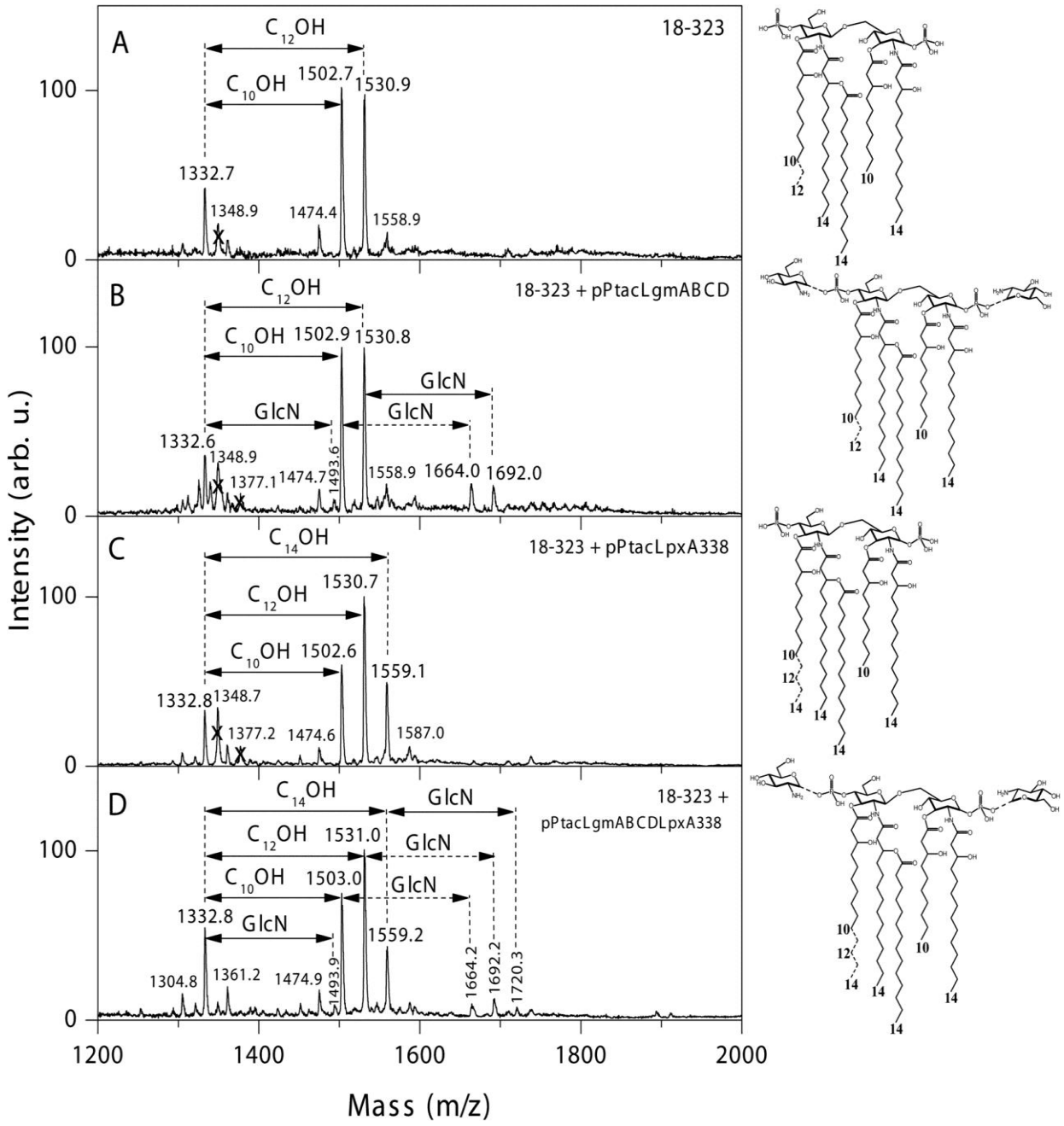


Figure 34. Lipid A structures of *B. pertussis* 18-323 strains complemented with BP338 lipid A-modifying genes

(A) 18-323 wild type strain, (B) 18-323 + pPtacLgmABCD (complemented with the *lgm* locus of BP338), (C) 18-323 + pPtacLpxA338 (complemented with *lpxA* of BP338), (D) and 18-323 + pPtacLgmABCDLpxA338 (complemented with the *lgm* locus of BP338 and *lpxA* of BP338). Arrows labeled with $C_{10}OH$, $C_{12}OH$, or $C_{14}OH$ indicate the 10-, 12-, or 14-carbon acyl chains absent in the tetra-acyl and present in the respective penta-acyl lipid A species. Arrows labeled with GlcN indicate the addition of GlcN at a phosphate group. The lipid A structures are summarized to the right of the mass spectra. Numbers at the bottom of the structures indicate the length of the acyl chains. Structures with GlcN modifications (B and D) have one GlcN added to either of the phosphate groups (peaks at *m*/*z* 1664, 1692, or 1720). arb. u., arbitrary units. From Shah *et al*. 2013 (Shah, Albitar-Nehme, et al. 2013), used with permission.

4.3.2 Difference in C3' acyl chain length is due to LpxA

Lipid A of BP338 has C14-OH C3′ acyl chains, whereas that of 18-323 has C10-OH and C12-OH C3′ acyl chains. LpxA is an essential enzyme that catalyzes the reaction that substitutes an acyl chain onto the C3 of UDP-GlcNAc, and this C3 acyl chain can then become the C3′ acyl chain, when two GlcN backbone subunits are joined by LpxB, to form the di-GlcN moiety of lipid A (Raetz, Guan, et al. 2009). I compared BP338 and 18-323 LpxA sequences and found a single amino acid difference: amino acid 173 is a serine in BP338 LpxA and a leucine in 18-323 LpxA. Residue 173 in *B. pertussis* LpxA is the equivalent of residue G176 in *E. coli* LpxA G176, which is positioned at the tip of the active site in close proximity to the C14 carbon of the 14-carbon long acyl chain that would be transfered onto GlcNAc by the activity of *E. coli* LpxA (Figure 35) (Williams and Raetz 2007). I hypothesized that the larger L173 in 18-323 LpxA occludes the tip of the active site, therefore allowing only C10-OH and C12-OH acyl chains into the active site, whereas the smaller S173 in BP338 LpxA allows C14-OH acyl chains.

To demonstrate that the difference at amino acid 173 of *B. pertussis* LpxA is the reason for the difference in C3′ acyl chain lengths between BP338 and 18-323, I introduced BP338 *lpxA* into 18-323 via the vector pPtacLpxA338. Analysis of the lipid A structural modifications of this strain via negative-ion MALDI-MS shows that introduction of BP338 *lpxA* into 18-323 generates a new major peak at *m*/*z* 1559. This peak corresponds to the addition of C3′ acyl chains with 14 carbons, as opposed to wild type 18-323, which only has peaks at *m*/*z* 1503 and 1531 (C10-OH and C12-OH C3′ acyl chains, respectively) (Figure 34AC). This shows that the LpxA from BP338 (with the single amino acid difference) alone is sufficient to introduce C14-OH C3′ acyl chains onto the lipid A of 18-323.

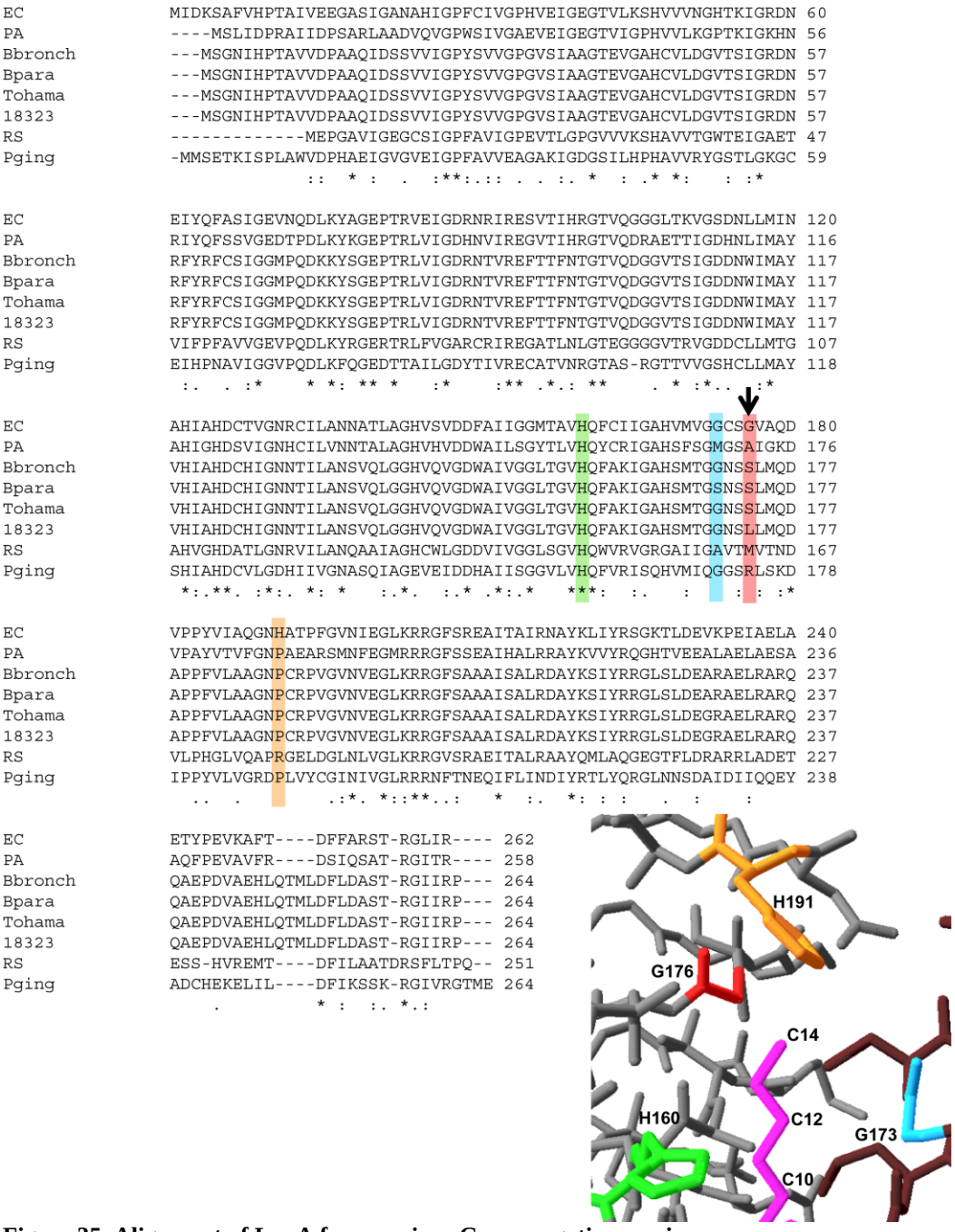


Figure 35. Alignment of LpxA from various Gram-negative species
Species (top to bottom) are: *E. coli* K-12 (EC), *P. aeruginosa* (PA), *B. bronchiseptica* (Bbronch), *Bordetella parapertussis* (Bpara), *B. pertussis* Tohama I BP338 (Tohama), *B. pertussis* 18-323 (18323), *R. sphaeroides* (RS), and *P. gingivalis* (Pging). The alignment was constructed with ClustalW2 (Thompson, Higgins, et al. 1994). The down arrow indicates the single amino acid difference between LpxA of BP338 and 18-323 (amino acid 173), which corresponds to G176 in *E. coli* LpxA. The bottom right inset shows the active site region between chains A (gray) and B (brown) of the *E. coli* LpxA homo-trimer and the amino acids surrounding the 14-carbon acyl chain (*p*ink). *E. coli* LpxA amino acids: H160 (green), G173 (blue), G176 (red), and H191 (orange). *E. coli* LpxA structure PDB ID 2QIA (Williams and Raetz 2007) was used to generate this figure in Swiss-Pdb Viewer (ExPASy) (Guex and Peitsch 1997). From Shah *et al*. 2013 (Shah, Albitar-Nehme, et al. 2013), used with permission.

4.4 C3’ acyl chain length and GlcN modification individually affect hTLR4 activation

I set out to explore the effect of these minute structural differences in *B. pertussis* lipid A (the presence of C14-OH acyl chains at the C3′ position and the GlcN modification) on hTLR4 activation. As such, I first generated a recombinant 18-323 strain that was complemented with both *lpxA* from BP338 and the full *lgm* locus (18-323 + pPtacLgmABCDLpxA338). The mass spectral analysis shows that lipid A from this strain displays a mixture of lipid A species, including species that have longer C14-OH C3′ acyl chains (peaks *m*/*z* 1559 and 1720) and species that contain the GlcN modification (peaks *m*/*z* 1664, 1692, and 1720) (Figure 34D). We tested the ability of purified LPS from these 18-323-derived strains to activate hTLR4 with the HEK-Blue hTLR4 activation assay (Figure 36). Our results show that 18-323 with BP338lpxA (18-323 + pPtacLpxA, which now has longer acyl chains at the C3′ position) was more effective in activating hTLR4 than wild type 18-323. When 18-323 lipid A was modified with GlcN at the phosphate groups (18-323 + pPtacLgmABCD), there was also a significant increase in hTLR4 activation over the wild type levels. By incorporating both these modifications into 18-323 to produce a longer C3′ acyl chain and the GlcN modification (18-323 + pPtacLpxA338LgmABCD), a significant increase in hTLR4 activationwas observed when compared to activation by LPS from the wild type (18-323) and from 18-323 with the longer acyl chain (18-323 + pPtacLpxA338). However, there was no difference in hTLR4 activation level when comparing 18-323 + pPtacLpxA338LgmABCD to 18-323 LPS that only has the GlcN modification (18-323 + pPtacLgmABCD). These results suggested that each modification of *B. pertussis* LPS alone was sufficient to cause an increase in hTLR4 activation. These results also suggest that when the GlcN modification is present, increasing the C3′ acyl chain length does not further increase hTLR4 activation. However, because there are varying levels of these modifications in each of the 18-323 strains, the relative contribution of each modification is difficult to determine.

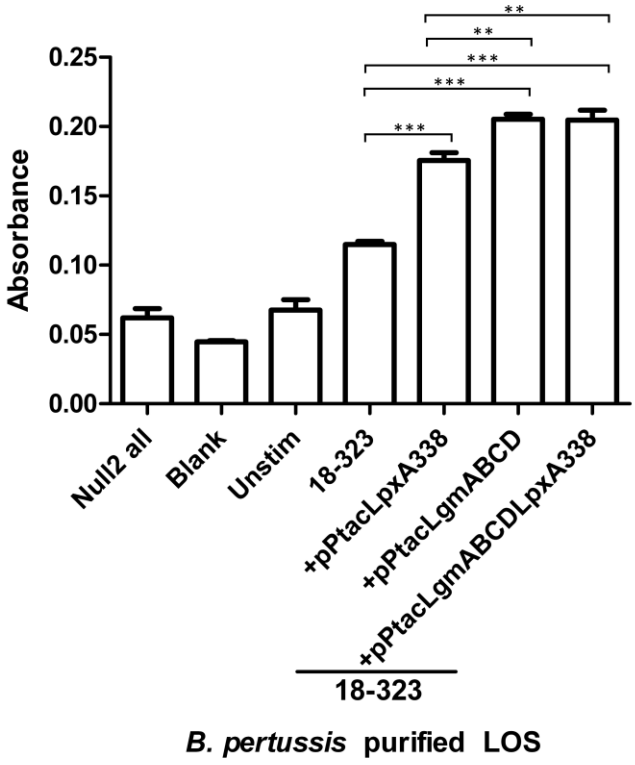


Figure 36. hTLR4 activation by LPS from *B. pertussis* 18-323 strains with BP338 lipid A-modification genes
hTLR4 activation measured with the HEK-Blue NFκB hTLR4 activation assay. Null2 all, stimulation of HEK-Blue Null2 cell line that lacks TLR4 expression with all LPS variants; Blank, medium only with no HEK-Blue cells; Unstim, HEK-Blue hTLR4 cells stimulated with medium only; 18-323 + pPtacLpxA338, 18-323 complemented with *lpxA* from BP338; 18-323 + pPtacLgmABCD, 18-323 complemented with *lgm* locus from BP338; 18-323 + pPtacLgmABCDLpxA338, 18-323 complemented with both *lpxA* and *lgm* locus from BP338. Absorbance at 650 nm (A_{650}). Graph shows the results of one representative experiment of three, $n = 6$ replicates per experiment. p values: <0.01 (**), < 0.001 (***). From Shah *et al*. 2013 (Shah, Albitar-Nehme, et al. 2013), used with permission.

4.5 Discussion

I obtained data to show that the *B. pertussis* lipid A GlcN modification increased resistance of the bacteria to CAMPs, but this modification did not appear to affect resistance to the positive aminoglycoside, gentamicin. Furthermore, the GlcN modification increased resistance to OM perturbation. I also demonstrated that the genetic basis for the 18-323 lipid A structure was due to the absence of a complete *lgm* locus, which explains the lack of the GlcN modification, and a single amino acid difference in LpxA, which explains the shorter C3' acyl chain when compared to BP338. This allowed me to test the effect of each individual structural difference between BP338 and 18-323 lipid A on LPS activation of hTLR4 by complementing 18-323 with BP338 lipid A-modifying genes. I found each modification (increasing the C3' acyl chain length or adding GlcN to the phosphates) alone increased the level of hTLR4 activation, but the addition of both modifications resulted in the same level of hTLR4 activation as when only the GlcN is added to lipid A.

Similar modifications of lipid A in different species do not always have the same effects. For example, Ara4N modification of *Salmonella* LPS does not affect activation of TLR4, whereas the same modification in *Pseudomonas aeruginosa* increases activation of TLR4 (Gellatly, Needham, et al. 2012). In this context, the recent findings that the *lgmB* gene in *B. bronchiseptica*, the causative agent of kennel cough in small mammals, is important for resistance to polymyxin B and porcine β-defensin 1 (pBD1) (Rolin, Muse, et al. 2014) does not necessarily mean the same is true for *B. pertussis*, especially considering the differences in LPS between these two species. *B. pertussis* LPS has 5 acyl chains and no O-antigen, whereas *B. bronchiseptica* LPS is hexa-acylated and has a long O-antigen, which has been shown to confer resistance to CAMPs (Banemann, Deppisch, et al. 1998, Marr, Tirsoaga, et al. 2008, Rolin, Muse, et al. 2014). Despite these considerable variations in LPS structure, GlcN modification still confers resistance to polymyxin B in both bacterial species. Furthermore, when comparing the effect on TLR4 activation in the natural hosts, the LgmB-mediated modification in *B. bronchiseptica* does not

affect activation of mouse TLR4, whereas the GlcN modification in *B. pertussis* increases activation of hTLR4 (Marr, Hajjar, et al. 2010, Rolin, Muse, et al. 2014). This adds a layer of complexity in the *B. pertussis* and human system that is not present in *B. bronchiseptica* and mice. The GlcN-modification in *B. bronchispetica* has also been linked to successful transmission between mice and colonization at lower infectious doses (Rolin, Muse, et al. 2014).

Previously, Ara4N modification to lipid A was proposed to protect *Salmonella* against CAMPs by decreasing the overall negative charge of the OM, thereby decreasing the affinity for positively charged CAMP molecules (Gunn, Lim, et al. 1998). I propose that in addition to this charge-masking mechanism of CAMP resistance, the modification of lipid A with a positively-charged sugar (GlcN in *B. pertussis*) may also function to stabilize the OM, thereby further excluding CAMPs from gaining access to the bacterium (Needham and Trent 2013). Bacteria with GlcN-modified lipid A were more resistant to lysozyme in the presence of the cation-chelator EDTA (Figure 32), which perturbs the OM by chelating the stabilizing cations that bridge the negatively charged phosphate groups of lipid A (Needham and Trent 2013). I propose that in wild-type *B. pertussis*, which modifies the phosphates of a proportion of its lipid A with GlcN, the positively-charged GlcN groups are able to coordinate negatively-charged phosphate groups on other unmodified lipid A molecules, thereby stabilizing the OM, even in the absence of cations, as described in Figure 37.

Stabilization of the OM is also critical for protection against numerous agents present in the airway that contribute to the clearance of infections, such as lysozyme, which is found at 1 mg/ml in sputum (Laube, Yim, et al. 2006, Needham and Trent 2013). To kill Gram-negative cells, lysozyme must traverse the OM to gain access to the peptidoglycan, which it degrades to promote cell death. Agents in the respiratory tract, such as lactoferrin, destabilize the OM, thus allowing lysozyme greater access to the peptidoglycan (Laube, Yim, et al. 2006). This type of synergy has been found between many antimicrobial factors in the

airway (Laube, Yim, et al. 2006). Therefore, the increase in resistance to OM perturbation afforded by the lipid A GlcN modification may also protect *B. pertussis* against lysozyme and other infection-clearing mechanisms.

Differences between enzymes in the Raetz lipid A biosynthesis pathway (Raetz, Guan, et al. 2009) can also be responsible for the variations observed in lipid A structures between different strains and species. Williams and Raetz suggest that the length of the acyl chain at the C3 and C3′ positions of lipid A is controlled by the hydrocarbon ruler region of LpxA, that is, the amino acids located near the active site in proximity to the acyl chain of the substrate, such as G173 and G176 in *E. coli* LpxA (Williams and Raetz 2007). Previous work, where mutating G173 of *E. coli* LpxA changed the acyl chain length specificity of the enzyme, supports this theory (Wyckoff, Lin, et al. 1998). My data also support this theory because amino acid 173 in *B. pertussis* LpxA is equivalent to G176 in *E. coli*, and if *B. pertussis* LpxA has a serine in this position (as seen in BP338), a C14-OH acyl chain can be added at the C3′ position, whereas if a leucine is found in position 173 (as seen in 18-323), only C10-OH and C12-OH acyl chains are found. Therefore, I propose that the larger leucine is occluding the tip of the active site of LpxA, thus only allowing 10- or 12-carbon acyl chains to be added. An alignment of LpxA sequences from several Gram-negative species suggests a similar correlation between strains with shorter acyl chains at the C3 and/or C3′ positions (*e.g. Pseudomonas aeruginosa* and *Rhodobacter sphaeroides*) and the presence of a larger amino acid at positions equivalent to *E. coli* LpxA G173, G176, or both (Figure 35). However, for *P. gingivalis* LpxA, this correlation, which is mostly based on sequence alignment, does not hold true and may be a reflection of the decreased sequence similarity of the *P. gingivalis* LpxA. Thus, although this analysis generally supports the role of both G173 and G176 of *E. coli* LpxA as a hydrocarbon ruler for the length of the acyl chain added to lipid A at the C3 and C3′ positions, there are likely more complexities involved in determining acyl chain length, especially in more distantly related LpxA species.

These results help to shed light on how the GlcN modification and C3' acyl chain length of penta-acyl LPS can affect various aspects of the bacterium, from CAMP resistance to activation of hTLR4. Yet, many bacteria have hexa-acyl LPS, such as *E. coli* and *Salmonella*. Studies looking at the removal and addition of entire acyl chains in *E. coli* hexa-acyl LPS demonstrate how such drastic changes to the acylation pattern can sometimes affect CAMP resistance and activation of TLR4. However, the effects of slight changes to the acyl chain length of hexa-acyl *E. coli* LPS have, thus far, not been studied.

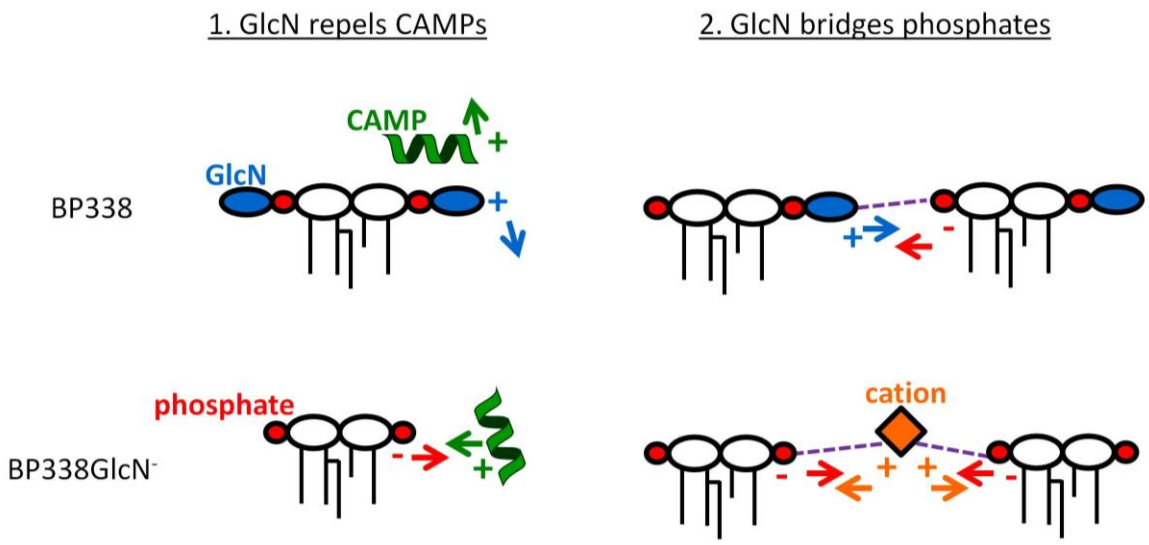


Figure 37. Model for how the *B. pertussis* GlcN modification facilitates CAMP resistance and OM stabilization

Top row depicts BP338 lipid A with the GlcN modification (blue) on some phosphate groups (red). Bottom row depicts BP338 lacking the GlcN modification (BP338GlcN$^-$). CAMP (green helix), cation (orange), coordination between oppositely charged elements (purple-dashed line). '+' and '-' indicate the charge of the correspondingly coloured factor. 1. In BP338, positively charged GlcN modification repels positively charged CAMPs; in BP338GlcN$^-$, the negatively charged, exposed phosphates groups attract the positively charged CAMPs. 2. In BP338, the positively charged GlcN that modifies some lipid A phosphate groups coordinates other negatively charged un-modified phosphates, thereby stabilizing the OM; in BP338GlcN$^-$, the negatively charged phosphate groups rely on coordinating positively charged divalent cations to stabilize the OM.

Chapter 5: The biological effects of varying the C3 and C3' acyl chain lengths in *E. coli* hexa-acyl lipid A

5.1 Introduction

Major changes to the acylation pattern of lipid A can have a wide range of effects on bacteria (Raetz, Reynolds, et al. 2007). For example, addition of an acyl chain at the C2 secondary position by PagP to generate hepta-acyl lipid A in *Salmonella* promotes resistance to CAMPs, and also attenuates TLR4 activation (Miller, Ernst, et al. 2005, Raetz, Reynolds, et al. 2007, Trent, Stead, et al. 2006). Thus, a single acyl chain modification can have multiple biological effects.

In regards to TLR4 activation, not only the number of acyl chains, but also their position on the lipid A di-GlcN backbone affects the degree of activation (Needham, Carroll, et al. 2013). Coats *et al.* (Coats, Berezow, et al. 2011) suggest that in penta-acyl lipid A species, longer fatty acid chains correlate with greater TLR4 activity. This hypothesis is based on the observation that penta-acyl *E. coli* LPS with acyl chain lengths of 12 and 14 carbons is a weaker TLR4 agonist in comparison with *Porphyromonas gingivalis* penta-acyl LPS, which contains C15, C16, and C17 acyl chains. Both these species have penta-acyl lipid A, and I have shown in *B. pertussis* penta-acyl lipid A, increasing the C3' acyl chain in isogenic strains increases hTLR4 activation (Section 4.4). This situation may be different in hexa-acyl LPS species, since hexa-acyl LPS tends to be a robust activator of TLR4, suggesting very strong interactions with TLR4 and MD-2 (Maeshima and Fernandez 2013). As such, changing the acyl chain length may or may not affect hTLR4 activation by hexa-acyl LPS.

Varying the length of the acyl chains in LPS may also affect other aspects, such as growth, resistance to CAMPs, and activation of the LAL assay, which is used in the biotechnology industry to detect endotoxin contamination.

5.2 Generation of R0138 *E. coli* strains with different C3 and C3' acyl chain lengths

To study the effects of lipid A acyl chain length variation in a hexa-acyl strain of *E. coli*, I generated isogenic strains with varying C3 and C3' acyl chains. The *E. coli* strain R0138 (Galloway and Raetz 1990) has *lpxA* under the control of a temperature sensitive promoter, such that endogenous LpxA is expressed at low temperatures (i.e. 30°C), but is not expressed at higher temperatures (i.e. 42°C). LpxA is the first enzyme in the Raetz LPS biosynthesis pathway that adds the acyl chain onto C3 of UDP-GlcNAc. Once the di-glucosamine backbone of lipid A is formed, this added acyl chain is substituted at either the C3 or C3' position in the final LPS molecule (Raetz, Reynolds, et al. 2007). LpxA from *B. pertussis* strain BP338 is able to add at least C10-OH and C14-OH acyl chains and *B. pertussis* strain 18-323 LpxA can add C10-OH and C12-OH acyl chains (Section 4.3.2). Therefore, to generate isogenic R0138 *E. coli* strains that differ only in the lengths of the C3 and C3' primary acyl chains, I transformed R0138 with plasmids containing *lpxA* from *E. coli* DH5α (R0138 + pBSlpxAEC), *B. pertussis* strain BP338 (R0138 + pBSlpxA338), or *B. pertussis* strain 18-323 (R0138 + pBSlpxA18323). In addition, I also complemented R0138 with *lpxA* from *B. pertussis* strains BP338 and 18-323 in the reverse orientation in the pBS vector (R0138 + pBSlpxA338rev and R0138 + pBSlpxA18323rev, respectively) as negative controls.

Since *lpxA* is an essential gene, R0138 is unable to grow at 42°C, as the endogenous LpxA is not expressed. However, if R0138 is complemented with exogenous LpxA it should grow at 42°C. I found all the R0138 strains can grow at 30°C, which is expected since the endogenous LpxA is expressed under these conditions (Figure 38). However, only the strains that are complemented with exogenous *lpxA* in the forward direction in the vector (R0138 + pBSlpxAEC, R0138 + pBSlpxA338, and R0138 + pBSlpxA18323) were able to grow at 42°C, suggesting successful complementation with functional LpxA in these strains (Figure 38).

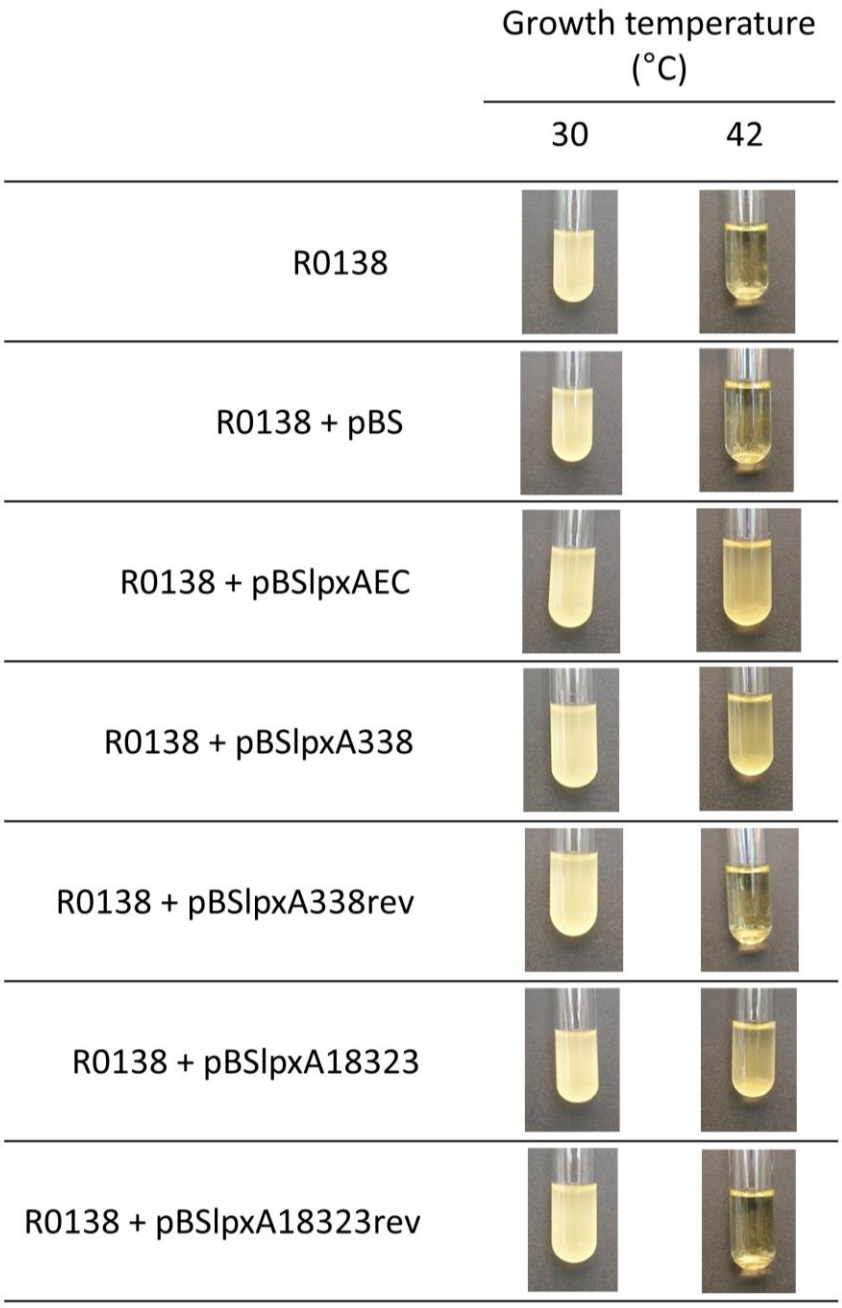


Figure 38. Viability of *E. coli* R0138 strains complemented with *lpxA* at 30°C and 42°C
LB broth with the appropriate antibiotics was inoculated with freshly grown R0138 colonies and incubated for 24 hours shaking at 200 rpm at 30°C or 42°C. Images were taken of 2 ml of culture with Olympus Camedia C-5060 camera. Images show the results of one representative experiment of three.

If the exogenous LpxA is fully functional in the Raetz lipid A biosynthesis pathway in these R0138 *E. coli* strains, I expect these strains to contain lipid A with different C3 and C3' acyl chain lengths. Analysis of the lipid A from these strains grown at 42°C (expression of the exogenous LpxA only) with mass spectrometry showed R0138 complemented with *E. coli lpxA* (R0138 + pBSlpxAEC) (Figure 39B) had primarily hexa-acyl lipid A species with C14-OH C3 and C3'acyl chains (1797 *m/z* peak), as seen in wild-type R0138 grown at 30°C (Figure 39A). However, both strains also had a minor peak at 1769 or 1770 *m/z*, indicating the presence of lipid A species with acyl chains shortened by two carbons, perhaps at the C3 or C3' position. Complementation of R0138 with *B. pertussis* strain BP338 *lpxA* (R0138 + pBSlpxA338) resulted in hexa-acyl lipid A with C3 and C3' acyl chains of C10-OH, C12-OH, and C14-OH (Figure 39C), as indicated by the following peaks: 1685 *m/z* represents lipid A with 10-carbon long C3 and C3' acyl chains; 1713 *m/z* represents lipid A with a 10-carbon C3 acyl chain and a 12-carbon C3' acyl chain or a C12-OH C3 acyl chain and a C10-OH C3' acyl chain; 1741 *m/z* represents lipid A with a C10-OH C3 acyl chain and a C14-OH C3' acyl chain, a C12-OH acyl chain at both the C3 and C3' positions, or a C14-OH C3 acyl chain and C10-OH C3' acyl chain; 1770 *m/z* represents lipid A with a C12-OH C3 acyl chain and a C14-OH C3' acyl chain or a C14-OH C3 acyl chain and a C12-OH C3' acyl chain; and finally 1779 *m/z* represents lipid A with a 14-carbon acyl chain at both the C3 and C3' positions. Therefore, compared to R0138 + pBSlpxAEC, the R0138 + pBSlpxA338 strain has lipid A with overall shorter C3 and C3' acyl chains. However, complementing R0138 with *B. pertussis* strain 18-323 *lpxA* (R0138 + pBSlpxA18323) results in a strain with the shortest acyl chains, with only C10-OH and C12-OH acyl chains at the C3 and C3' positions (Figure 39D). When grown at 42°C, all three complemented strains have slightly higher peaks representing lipid A with 7 fatty acyl chains, which could suggest a greater level of hepta-acyl lipid A in comparison to R0138 grown at 30°C. This could be due to the addition of an acyl chain by PagP to the hexa-acyl lipid A species in response to the temperature stress. Therefore, I confirmed the generation of isogenic *E. coli* strains with hexa-acyl lipid A

with gradually decreasing C3 and C3’ acyl chain lengths (R0138 + pBSlpxAEC to R0138 + pBSlpxA338 to R0138 + pBSlpxA18323) when grown at 42°C.

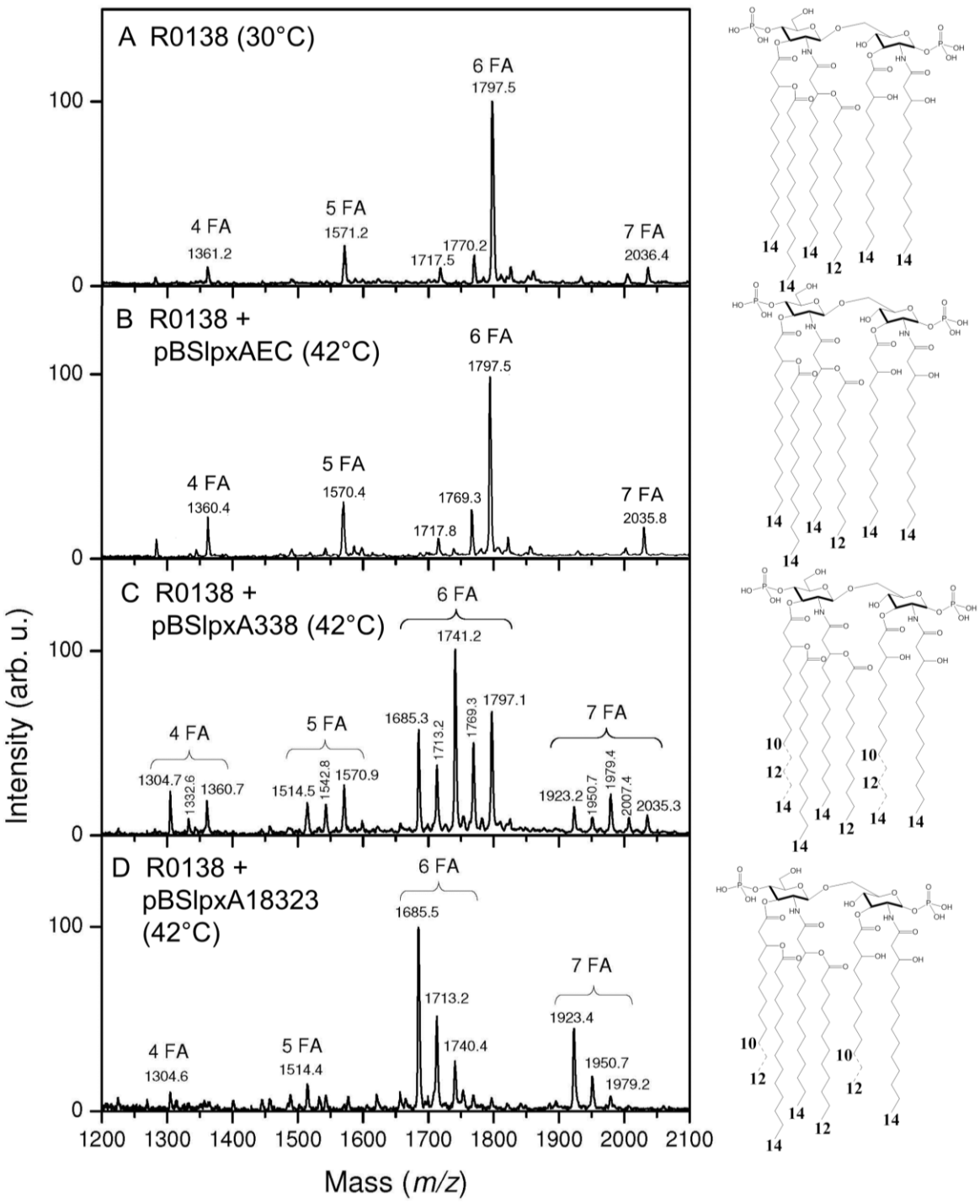

A R0138 (30°C)
6 FA
1797.5
4 FA
1361.2
5 FA
1571.2
1717.5
1770.2
7 FA
2036.4
B R0138 +
pBSlpxAEC (42°C)
6 FA
1797.5
4 FA
1360.4
5 FA
1570.4
1717.8
1769.3
7 FA
2035.8
C R0138 +
pBSlpxA338 (42°C)
6 FA
1741.2
4 FA
1304.7
1332.6
1360.7
5 FA
1514.5
1542.8
1570.9
1685.3
1713.2
1769.3
1797.1
7 FA
1923.2
1950.7
1979.4
2007.4
2035.3
D R0138 +
pBSlpxA18323
(42°C)
6 FA
1685.5
4 FA
1304.6
5 FA
1514.4
1713.2
1740.4
7 FA
1923.4
1950.7
1979.2
Intensity (arb. u.)
Mass (m/z)
0
100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
14
12
10

Figure 39. Lipid A structures of *E. coli* R0138 strains complemented with exogenous *lpxA*
A) R0138 (30°C), B) R0138 + pBSlpxAEC (42°C), C) R0138 + pBSlpxA338 (42°C), D) and R0138 + pBSlpxA18323 (42°C). Peaks at *m/z* 1685 represent hexa-acyl lipid A with acyl chains lengths (C3, C3') of (10, 10) carbons, *m/z* 1713 represent hexa-acyl lipid A with acyl chains lengths (C3, C3') of (10,12) and (12, 10) carbons, *m/z* 1741 represent hexa-acyl lipid A with acyl chains lengths (C3, C3') of (10,14), (12, 12) and (14, 10) carbons, *m/z* 1770 represent hexa-acyl lipid A with acyl chains lengths (C3, C3') of (12,14) and (14, 12) carbons, and *m/z* 1797 represent hexa-acyl lipid A with acyl chains lengths (C3, C3') of (14,14) carbons. 4FA, 4 fatty acid chains; 5FA, 5 fatty acid chains; 6FA, 6 fatty acid chains; 7FA, 7 fatty acid chains. The hexa-acyl lipid A structures are summarized to the right of the mass spectra. The numbers at the bottom of the structures indicate the length of the acyl chains

5.3 C3 and C3' acyl chain lengths affect bacterial growth

Though all three *lpxA*-complemented R0138 strains were able to grow at the non-permissive temperature, the growth rate could be affected by the difference in C3 and C3' primary acyl chain lengths of the lipid A. To test this assumption, I grew wild type R0138 and the three *lpxA*-complemented strains at 30°C and 42°C. I found all four strains grew at similar rates at 30°C, conditions under which the endogenous LpxA is expressed (Figure 40A). At 42°C, when only the complemented LpxA is functioning in the bacterium, wild-type R0138 is unable to grow (Figure 40B). R0138 + pBSlpxAEC and R0138 + pBSlpxA338, which have the longer C3 and C3' acyl chains, grow at similar rates at 42°C whereas the strain with the shortest acyl chains, R0138 + pBSlpxA18323, grows at a slower rate (Figure 40B). The association between shorter C3 and C3' acyl chains and a reduced growth rate could be the result of the enzymes in the lipid A biosynthesis pathway being less efficient with substrates with shorter acyl chains, or a decrease in fitness of the bacteria with shorter acyl chains due to factors such as membrane stability, or a combination of both these issues.

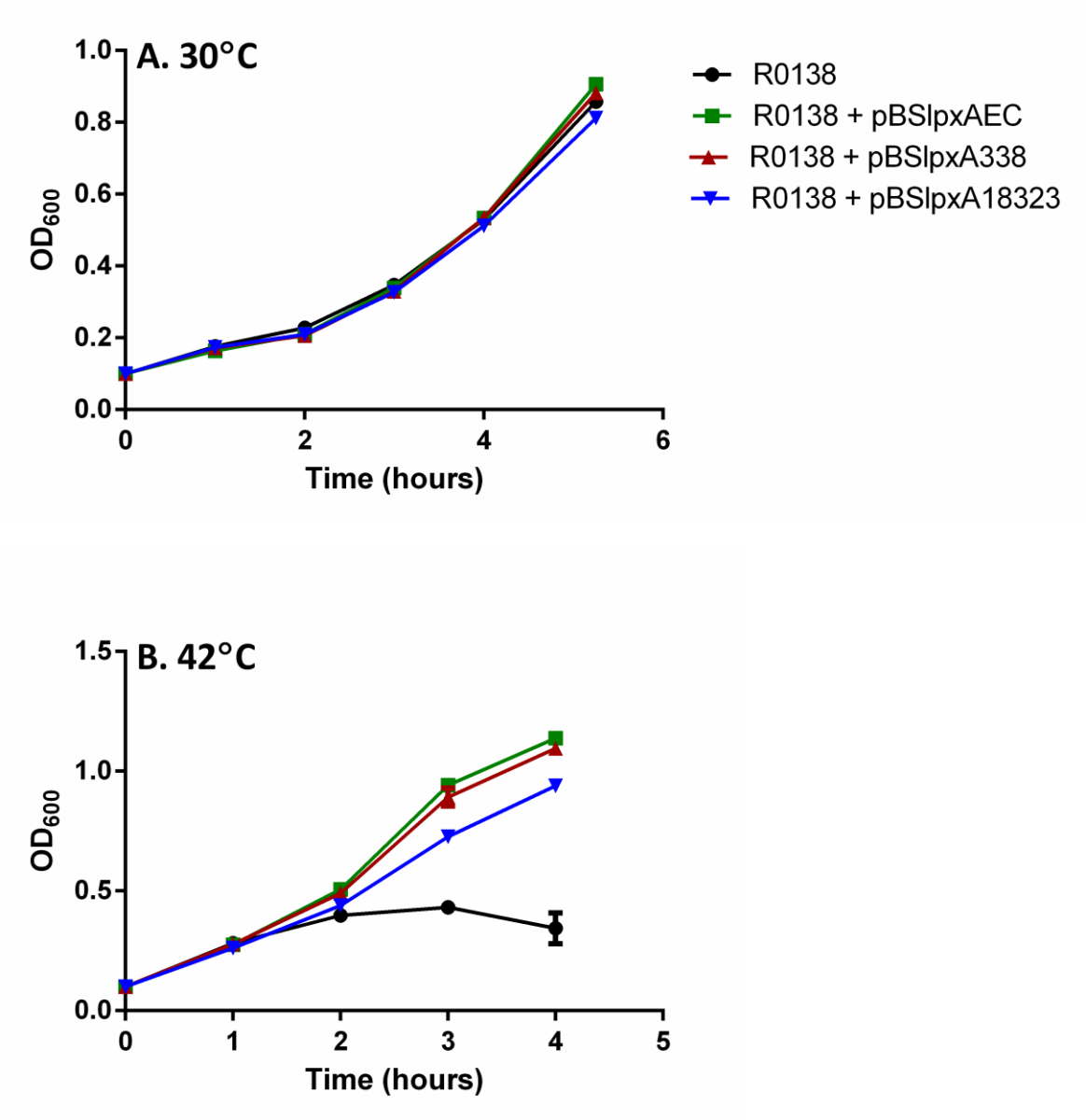


Figure 40. Growth of *E. coli* R0138 strains with varying acyl chain lengths
Growth curves of R0138 strains at 30°C (A) and 42°C (B). Graphs show the results of one representative experiment of three, n = 2 repeats per experiment.

5.4 C3 and C3' acyl chain lengths affect resistance to polymyxin B

Polymyxin B is a cationic peptide that is often used as a last resort for the treatment of multidrug-resistant Gram-negative pathogens (Zavascki, Goldani, et al. 2007). Polymyxin B, and other cationic peptides, are taken up by the bacterium via interaction with LPS and cause cell death through destabilization of the OM (Zavascki, Goldani, et al. 2007). In hexa-acyl *E. coli* lipid A, changes in the C3 and C3' acyl chain lengths may affect susceptibility to polymyxin B, since these changes could influence membrane stability or the interaction between polymyxin B and LPS. I used two methods to compare polymyxin B resistance between the *lpxA*-complemented strains with different acyl chain lengths. First, I grew the bacteria in the presence of different polymyxin B concentrations and monitored growth over time. Using this method, I found a marked difference in growth at 0.5 µg/ml polymyxin B (half the minimum inhibitory concentration of polymyxin B for *E. coli*): as C3 and C3' acyl chains decrease in length (R0138 + pBSlpxAEC to R0138 + pBSlpxA338 to R0138 + pBSlpxA18323), I observe a longer delay in growth when compared to growth in no polymyxin B (Figure 41). To confirm that these results were not due to other LPS modifications that might be induced by polymyxin B (Fernandez, Jenssen, et al. 2012), as I discussed in the previous chapter, I used a second method in which the bacteria were incubated with polymyxin B for 5 min followed by cfu determination to monitor killing of the bacteria. In this case, I also see an effect of C3 and C3' acyl chain length on polymyxin B resistance: at 0.5 µg/ml polymyxin B the strain with the shortest acyl chains (R0138 + pBSlpxA18323) has decreased survival compared to strains with longer acyl chains (R0138 + pBSlpxAEC and R0138 + pBSlpxA338) (Figure 42).

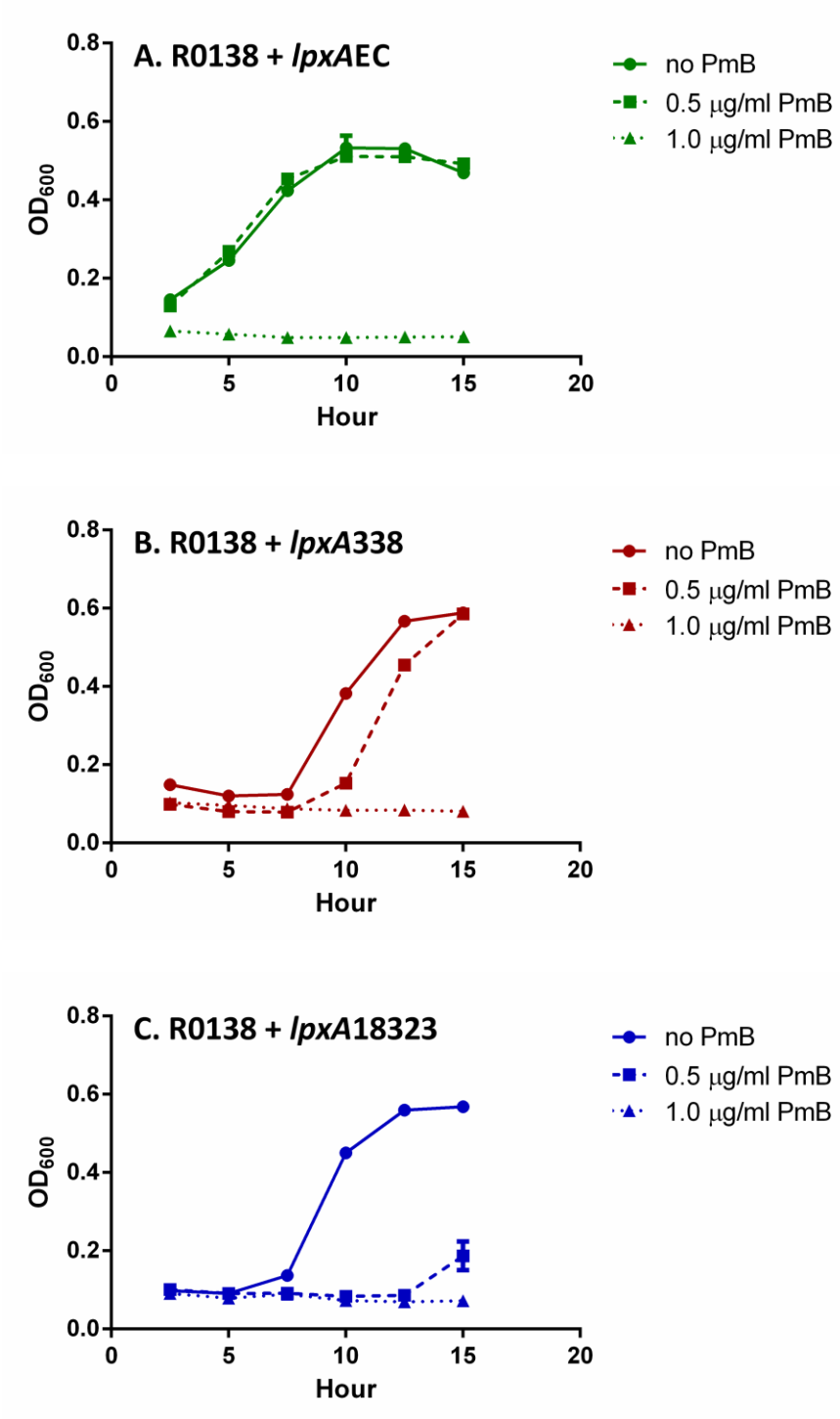


Figure 41. Polymyxin B resistance of *E. coli* R0138 strains with varying acyl chain lengths, assessed by growth curve assay

R0138 strains were grown in polymyxin B (PmB) (0, 0.5, and 1.0 μg/ml) at 42°C and growth was followed over time by OD_{600} readings of cultures. A) R0138 + pBSlpxAEC (R0138 + *lpxA*EC), B) R0138 + pBSlpxA338 (R0138 + *lpxA*338), C) R0138 + pBSlpxA18323 (R0138 + *lpxA*18323) Graphs show the results of one representative experiment of three, n = 2 replicates per experiment.

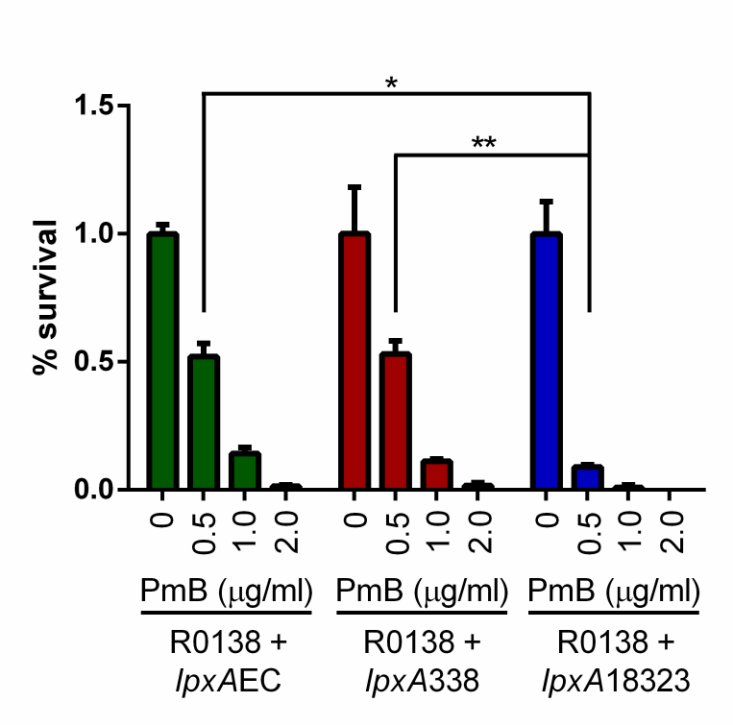


Figure 42. Polymyxin B resistance of *E. coli* R0138 strains with varying acyl chain lengths, assessed by percent survival assay

R0138 strains were incubated in polymyxin B (PmB) for 5 minutes and then plated on LB agar for cfu counting to monitor bacterial killing and determine percent survival (% survival). R0138 + pBSlpxAEC (R0138 + *lpxA*EC), R0138 + pBSlpxA338 (R0138 + *lpxA*338), R0138 + pBSlpxA18323 (R0138 + *lpxA*18323). Graphs show the results of one representative experiment of three, n = 3 replicates per experiment. *p*-values: <0.05 (*), <0.01 (**).

5.5 C3 and C3' acyl chain lengths affect hTLR4 activation

The C3' acyl chain length affected hTLR4 activation by penta-acyl *B. pertussis* LPS (Section 4.4). However hexa-acyl *E. coli* LPS has a higher level of hTLR4 activation than *B. pertussis* strain BP338 LPS (Marr, Novikov, et al. 2010), therefore minor changes in the length of acyl chains may not affect hTLR4 activation as they do in *B. pertussis* LPS. To determine if variation in C3 and C3' acyl chain lengths of *E. coli* lipid A affect hTLR4 activation, I stimulated HEK-Blue hTLR4 cells with purified LPS from the *lpxA*-complemented R0138 strains. I found a significant difference in hTLR4 activation between R0138 + pBSlpxAEC and R0138 + pBSlpxA18323 LPS (Figure 43), showing that as the C3 and C3' acyl chain lengths decrease, hTLR4 activation by LPS also decreases. This trend also holds true for R0138 + pBSlpxA338, which has intermediate C3 and C3' acyl chain lengths. LPS from this *E. coli* strain consistently has intermediate levels of hTLR4 activation compared to R0138 + pBSlpxAEC and R0138 + pBSlpxA18323, across all repeated experiments, though this difference is not always statistically significant.

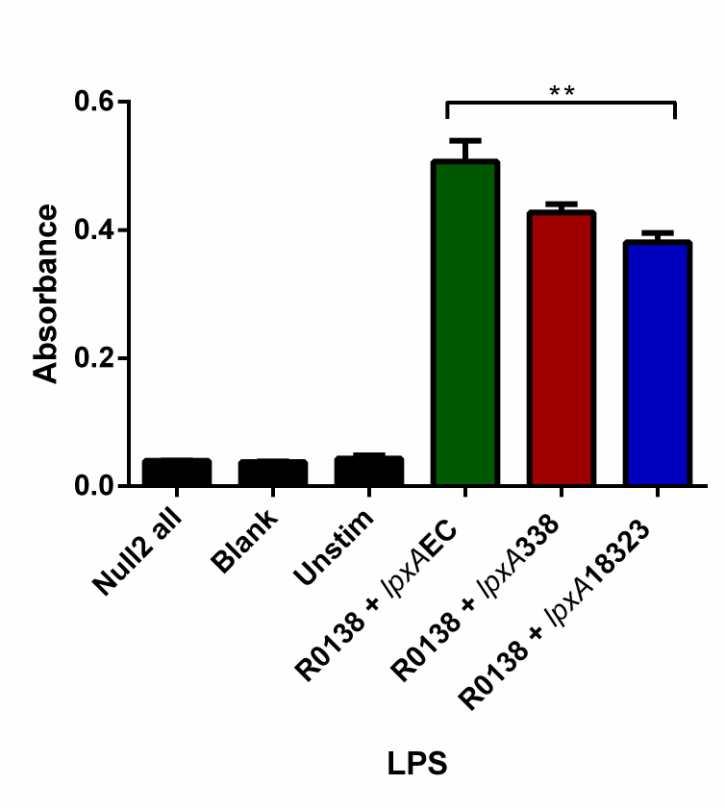


Figure 43. Activation of hTLR4 by LPS from *E. coli* R0138 strains with varying acyl chain lengths
hTLR4 activation measured with the HEK-Blue NFκB hTLR4 activation assay. Null2 all, stimulation of HEK-Blue Null2 cell line that lacks TLR4 expression with all LPS variants; Blank, medium only with no HEK-Blue cells; Unstim, HEK-Blue hTLR4 cells stimulated with medium only; R0138 + *lpxA*EC, R0138 complemented with *lpxA* from *E. coli*; R0138 + *lpxA*338, R0138 complemented with *lpxA* from *B. pertussis* strain BP338; R0138 + *lpxA*18323, R0138 complemented with *lpxA* from *B. pertussis* strain 18-323. Absorbance at 650 nm (A_{650}). Graph shows the results of one representative experiment of three, $n = 5$ replicates per experiment. p value: <0.01 (**).

5.6 C3 and C3’ acyl chain lengths affect LAL activation

The LAL assay detects LPS through interaction between LPS and the Factor C protein of the Horseshoe crab, which leads to a downstream enzymatic cascade that is linked to an experimental read out, such as a change in colour (Muta, Miyata, et al. 1991). We used the Pyrochrome LAL assay to assess the activation of the LAL cascade by LPS purified from the three *lpxA*-complemented *E. coli* R0138 strains. We found R0138 + pBSlpxAEC, the strain with the longest lipid A C3 and C3’ acyl chains, had the lowest levels of LAL activation when compared to R0138 + pBSlpxA338 and R0138 + pBSlpxA18323, which have gradually shorter C3 and C3’ acyl chains in comparison (Figure 44). Therefore, *E. coli* hexa-acyl LPS with longer C3 and C3’ acyl chains (14 carbons long) activates the LAL assay to a lesser degree compared with LPS with shorter C3 and C3’ acyl chains.

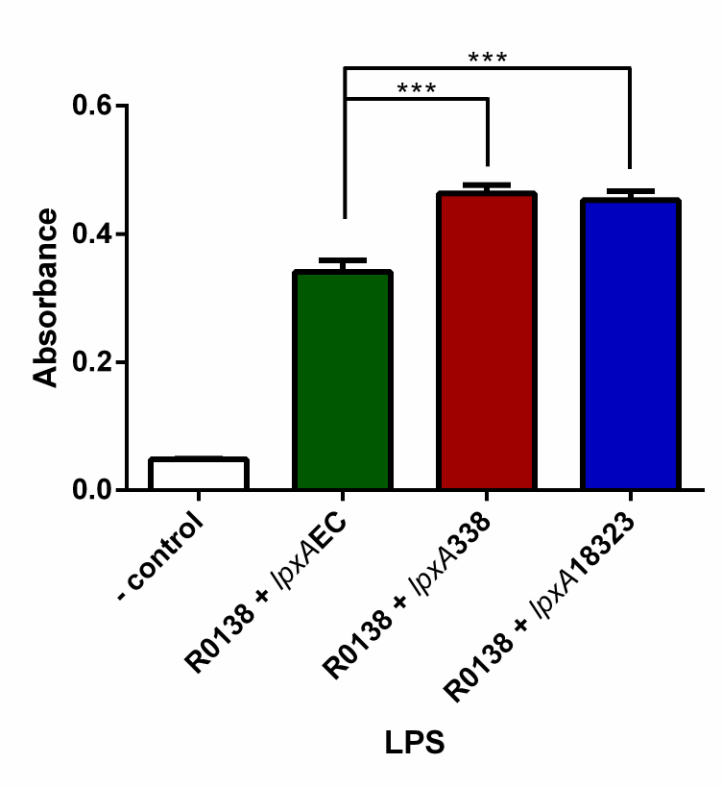


Figure 44. Activation of the LAL assay by LPS from *E. coli* R0138 strains with varying acyl chain lengths
Pyrochrome LAL assay used to assay purified LPS at 10 ug/ml. – control, endotoxin-free water. Absorbance at 405 nm (A_{405}). Graph shows the average of 4 individual experiments, n = 3 replicates per experiment. *p*-values: < 0.001 (***).

5.7 Discussion

Variation in the acyl chain lengths of LPS can affect many aspects of bacterial physiology, from membrane stability to interaction with the host immune system.

Complementing *E. coli* strain R0138 with *lpxA* from *B. pertussis* strains BP338 and 18-323 could provide further information regarding the lipid A biosynthesis pathway in *B. pertussis*, and also raises some queries regarding selectivity in this pathway. LPS from *B. pertussis* strain BP338 has C14-OH acyl chains at the C3' position and C10-OH acyl chains at the C3 position, suggesting LpxA from this strain is able to transfer both C10-OH and C14-OH acyl chains (Shah, Albitar-Nehme, et al. 2013). However, when R0138 is complemented with *lpxA* from BP338 and grown at the non-permissive temperature, C10-OH, C12-OH, and C14-OH acyl chains are added at the C3 and C3' positions. This indicates BP338 LpxA can also transfer C12-OH acyl chains, even though C12-OH chains are not observed in the final LPS structure in *B. pertussis* BP338 (Caroff, Brisson, et al. 2000, Marr, Tirsoaga, et al. 2008). Therefore, there is likely another level of control over the lengths of the C3 and C3' acyl chains in *B. pertussis* strain BP338, other than LpxA. Though LpxA controls the acyl chain length attached at the C3 position of UDP-GlcNAc during the initial step in LPS biosynthesis, another enzyme in this pathway likely plays a role to ensure the C3' position acyl chain is 10 carbons long and the C3' position acyl chain is 14 carbons long in the final lipid A structure. For example, LpxH could specifically cleave uridine-monophosphate from substrates with a C10-OH acyl chain at the C3 position to generate lipid X molecules with only a 10-carbon long chain at this position, therefore resulting in LPS molecules with only C10-OH acyl chains at the C3' position. Alternatively, LpxB in *B. pertussis* BP338 could specifically catalyze the joining of a lipid X molecule with a C10-OH acyl chain at the C3 position and a UDP-2,3-diacyl-GlcN molecule with C14-OH acyl chain at the C3 position. Since lipid A of *B. pertussis* strain 18-323 also consistently has a C10 acyl chain at the C3 position, the selective mechanism found in BP338 is probably also present in 18-323. However, the C3' position acyl chain in 18-323 LPS can be 10 or 12 carbons long, even when

complemented with BP338 *lpxA* which allows this strain to have 10-, 12-, or 14-carbon long acyl chains at this position. Therefore, the selectivity for the C3' acyl chain length in BP338, which ensures only 14-carbon acyl chains at this position, is not present in 18-323.

The presence of alternate C3 and C3' acyl chain lengths in the complemented *E. coli* R0138 strains demonstrated that the LPS biosynthesis pathway of *E. coli* has a certain level of substrate flexibility in the downstream enzymes that allows them to accommodate lipid A with different acyl chain lengths. However, we did observe a slight growth defect in the strain with the shortest acyl chains (R0138 + pBSlpxA18323), which could result from a decreased efficiency of the Raetz pathway with substrates with shorter acyl chains. This could also be a result of decreased membrane stability due to decreased acyl chain lengths, which is predicted to decrease the overall hydrophobic interactions of the LPS acyl chains in the OM (Guo, Lim, et al. 1998).

A decrease in membrane stability is sometimes associated with an increase in susceptibility to CAMPs (Guo, Lim, et al. 1998). As discussed in the previous chapter, CAMPs are an important part of the innate immune system of many organisms, and are secreted by epithelial and immune cells in humans (Miller, Ernst, et al. 2005). CAMPs are also found in the environment, such as polymyxin B, which is a CAMP produced by Gram-positive environmental bacteria. Polymyxin B interacts with surface-exposed regions of LPS at the bacterial OM and displaces the cations Ca^{2+} and Mg^{2+}, therefore destabilizing the OM and allowing polymyxin B to interact with the fatty acyl chains. This results in the insertion of polymyxin B into the membrane, which leads to OM leakiness (Zavascki, Goldani, et al. 2007). *Salmonella* that lacks the ability to generate hepta-acyl LPS (*pagP* mutant) has decreased resistance to CAMPS, suggesting a relationship between the level of acylation of LPS and CAMP resistance (Guo, Lim, et al. 1998). My findings also support this hypothesis, since decreasing the C3 and C3' acyl chain lengths of hexa-acyl *E. coli* LPS also decreases resistance to polymyxin B.

The acylation pattern of LPS can also be associated with different levels of TLR4 activation, though this is not always the case. In section 4.4, I have shown in *B. pertussis* penta-acyl LPS, increasing the length of the C3' acyl chain also increases hTLR4 activation. However, modifying the acyl chains of LPS does not always result in a difference in TLR4 activation. Removal of the primary C3 acyl chain in hexa-acyl *E. coli* LPS did not affect TLR4 activation whereas removal of the secondary C3' acyl chain decreased TLR4 activation by LPS (Needham, Carroll, et al. 2013). This suggests the location of the acyl chains is an important factor in the ability of LPS to activate TLR4. Since I modified the lengths of both the C3 and C3' acyl chains in *E. coli*, I am unable to determine if the decrease in hTLR4 activation as acyl chain lengths decrease is the effect of altering the C3 acyl chain, the C3' acyl chain, or both. In either case, the longer acyl chains in both penta-acyl and hexa-acyl LPS, in *B. pertussis* and *E. coli*, respectively, are associated with higher levels of hTLR4 activation.

Excessive release of proinflammatory cytokines in response to TLR4 activation can lead to in endotoxic shock, where these mediators damage small blood vessels, which can lead to septic shock and multiple organ failure (Raetz and Whitfield 2002). Therefore, when proteins are purified from bacteria for use in humans, it is important to ensure they are endotoxin-free. The LAL assay is used in the biotechnology industry to determine the endotoxicity of a sample, and this assay is based on the interaction between the Factor C enzyme of the Horseshoe crab and the di-GlcN backbone of lipid A (Chen and Mozier 2013, Muta, Miyata, et al. 1991). However, differences in lipid A structure can have different effects on LAL and TLR4 activation, since these two different methods of detecting lipid A depend on interaction with different proteins. For example, when comparing tetra-acyl and hexa-acyl lipid A: tetra acyl lipid A has higher levels of LAL activation whereas hexa-acyl lipid A has higher levels of TLR4 activation (Gutsmann, Howe, et al. 2010). My results also indicate a discrepancy between LAL activation and hTLR4 activation, as decreasing the C3 and C3' acyl chain lengths decreases hTLR4 activation but

increases LAL activation. This brings into question the effectiveness of the LAL assay for detecting differently shaped LPS molecules and casts further doubt on whether the LAL assay is a good reflection of TLR4 activation.

In conclusion, I found very small decreases in the C3 and C3' acyl chain lengths of *E. coli* hexa-acyl lipid A affect bacterial growth, resistance to polymyxin B, activation of hTLR4, and activation of the LAL cascade.

Chapter 6: Discussion

6.1 Discussion

The vast variety of lipid A structures amongst different Gram-negative bacterial species and strains can be attributed to both differences in Raetz lipid A biosynthetic pathway enzymes and also the presence of additional modification enzymes (Needham and Trent 2013, Raetz, Guan, et al. 2009). It is becoming increasingly evident that modifications to the structure of lipid A allow bacteria to adapt to different environments, as evident in the interaction between various human-adapted bacteria and the immune system (Needham and Trent 2013). Activation of hTRL4 by LPS can be significantly affected by alterations in lipid A structure, especially modification of the phosphate groups, which play a key role in the dimerization and subsequent activation of TLR4 (Maeshima and Fernandez 2013, Raetz, Reynolds, et al. 2007).

In *B. pertussis* strain BP338, the phosphate groups of lipid A are modified with GlcN moieties via LgmA, LgmB, and LgmC, and this modification resulted in an increase in CAMP resistance, greater OM stabilization, and an increase in hTLR4 activation by LPS. I have used ^{14}C-labelled substrate in an assay and shown LgmA functions to transfer GlcNAc onto a lipid, which is likely the inner membrane carrier lipid C55P. This is the proposed first step in lipid A GlcN-modification pathway of *B. pertussis*, and since the globular domain of LgmA was predicted to be cytoplasmic (Figure 7) and UDP-GlcNAc is found in the cytoplasm, I theorize that LgmA functions at the cytoplasmic side of the IM. I hypothesize the next step in this pathway is performed by the predicted cytoplasmic enzyme LgmC, which was proven to functions in the removal of the acetyl group from C55P-GlcNAc to produce C55P-GlcN (Llewellyn, Zhao, et al. 2012). Then, I predict LgmB functions to transfer GlcN from C55P onto lipid A in *B. pertussis* because the LgmB homolog ArnT functions in a similar manner to transfer Ara4N from C55P to lipid A in *E. coli* and *Salmonella* (Trent, Ribeiro, et al. 2001). I had originally postulated that a flippase enzyme translocates C55P-GlcN from the cytoplasmic side of the IM to the periplasmic side of the IM,

followed by transfer of GlcN onto lipid A by LgmB in the periplasm. However, the three flippase candidates, LgmD, BP1945, and LgmE, were shown not to be required for the GlcN modification of lipid A. An alternative enzyme may function as this flippase, such as the putative Gram-negative lipid II flippase of peptidoglycan synthesis, MurJ, which has 14 transmembrane helices (Butler, Davis, et al. 2013, Ruiz 2008). It is also possible that a flippase is not required for this pathway, either because spontaneous flipping of C55P-GlcN is sufficient, or because the transfer of GlcN to lipid A might occur in the cytoplasm, therefore negating the need for a C55P-GlcN flippase. However, the intrinsic rate of phospholipids flipping from one leaflet to another across a cytoplasmic membrane is relatively slow (half times in the order of hours to days) (Pomorski and Menon 2006), and this is likely insufficient for the GlcN-modification pathway in *B. pertussis*. Furthermore, though it is possible that GlcN-modification of lipid A occurs in the cytoplasm, LgmB is predicted to have very short cytoplasmic loops but longer periplasmic loops and a predicted periplasmic C-terminal region. In addition, a homolog of LgmB, ArnT, functions at the periplasmic side of the IM in the Ara4N lipid A-modification pathway (Casella and Mitchell 2008) and numerous residues along the putative periplasmic loop 5 of ArnT are required for function (Impellitteri, Merten, et al. 2010). This suggests the functional domains of LgmB are on the periplasmic face of the IM, and a flippase is probably required to transfer C55P-GlcN from the cytoplasmic to the periplasmic face of the IM, though this flippase has yet to be identified.

An interesting aspect of the *lgm* locus in *B. pertussis* is the presence of the previously unannotated ORF *lgmE*, a *cis*-encoded antisense RNA. Though LgmE was not required for the GlcN-modification of lipid A in *B. pertussis*, the RNA of *lgmE* may play a role in the regulation of the *lgm* locus. The full length of this RNA is not known, however, at the very least it is complementary to part of the *lgmB* and *lgmC* messenger RNAs (mRNA), and may act as a regulatory small RNA (sRNA). If *lgmABCD* is transcribed as a single transcript, *lgmE* RNA could anneal to the middle of this long transcript. Alternatively, if *lgmB* and *lgmC* are transcribed as separate mRNA, the *lgmE* RNA could bind to both mRNA molecules. Most

c*is*-encoded antisense sRNAs have been studied in plasmid and phage systems, where the sRNA is often involved in maintaining copy numbers. The role of *cis*-ecoded antisense sRNA in chromosomal systems is less clear. Some chromosomally-encoded sRNAs promote degradation of the mRNA they anneal to whereas others can repress expression. The *lgmE* transcript most resembles a group of *cis*-encoded antisense sRNAs that are encoded by an operon and the sRNA is complementary to the region in between ORFs on the opposite strand. In this case, the sRNA could function to promote cleavage of the full length operon mRNA between the ORFs to produce two mRNAs. In the case of the *lgm* locus, *lgmE* sRNA could anneal to the region between *lgmB* and *lgmC*, therefore promoting cleavage of the mRNA into an mRNA encoding *lgmA* and *lgmB* and a separate mRNA with *lgmC* and *lgmD*. Alternatively, this operon-sRNA architecture could lead to transcription termination. In this case, if *lgmABCD* was transcribed as a single mRNA, *lgmE* sRNA would anneal to the end of *lgmB* and the intergenic region between *lgmB* and *lgmC* to generate a quasi-hairpin structure, resulting in intrinsic (Rho-independent) transcription termination at this site, and therefore, lower levels of *lgmC* translation (Waters and Storz 2009). Thus, regardless of whether LgmE is expressed, *lgmE* RNA could play a role in the regulation of the lgm pathway, though the extent and function of such regulation remains to be seen. *lgmE* could be related to the Bvg-independent regulation of the *lgm* locus that is linked to growth on BG agar compared to growth in SS broth (Marr, Tirsoaga, et al. 2008).

B. pertussis shows a range of susceptibility to numerous CAMPs (Fernandez and Weiss 1996), and pBD1 is able to protect against *B. pertussis* infection in a newborn piglet model, demonstrating the protective capability of CAMPs against this pathogen (Elahi, Buchanan, et al. 2006). Recently, Taneja *et al*. have shown an LPS-independent mechanism in *B. pertussis* that increased CAMP resistance, namely the addition of a D-alanine group to an outer membrane component by the *dra* gene locus (Taneja, Ganguly, et al. 2013). *B. pertussis* has therefore evolved two separate mechanisms for CAMP resistance (modification of an outer membrane component with D-alanine and modification of lipid A with a

glucosamine group), suggesting this strategy of immune evasion plays an important role in the survival of this human-restricted pathogen.

The GlcN modification plays a dual role in *B. pertussis*, since it both increases hTLR4 activation (Marr, Hajjar, et al. 2010), leading to greater inflammation, which would tend to lead to increased LL-37 expression or release by degranulation from attracted neutrophils (De Smet and Contreras 2005), and it also increases resistance against CAMPs. I propose this lipid A modification may have originally arisen in *Bordetella* as an adaptation to the environment, to increase OM stabilization and protect against CAMPs and other environmental factors. In the human host, GlcN modification of *B. pertussis* lipid A continues to protect against antimicrobial factors, such as CAMPs, but this change in lipid A structure also affects hTLR4 activation. My results suggest GlcN-modified lipid A provides greater resistance to some CAMPs compared to others (Figure 30), though no pattern is observed based on the charge of the CAMPs at physiological pH (Table 6). We did, however, observe greater resistance to the polymyxins, which are cyclic, lipidated peptide antibiotics produced by environmental bacterial species, and this is likely due to different interactions between lipid A and CAMPs of different sequences and structures.

Other Gram-negative bacteria also modify the negatively-charged phosphates of lipid A with positively charged amino sugars. *Salmonella* and *E. coli* add Ara4N to lipid A phosphates, as does *P. aeruginosa*, and *F. turlarensis* can modify the lipid A phosphates with galactosamine (Needham and Trent 2013). In *Salmonella*, *E. coli*, and *P. aeruginosa*, the addition of Ara4N to lipid A increases resistance to polymyxins B, though this has yet to be reported in *F. tularensis* (Needham and Trent 2013). Modification of the phosphate groups of *Salmonella* and *E. coli* LPS with another positively-charged group, phosphoethanolamine, also increases resistance to CAMPs, such as polymyxins B, as does the removal of the negatively-charged phosphate groups themselves in *F. tularensis* LPS (Needham and Trent 2013). These observations suggest that a more positively-charged OM is related to greater resistance to

positively-charged CAMPs, which supports the hypothesis that the decrease in negative charge of the OM results in a decreased affinity for the positively charged CAMPs, thus endowing greater resistance to CAMPs. I propose another contributing factor to the CAMP resistance afforded by the addition of positively-charged groups to the phosphates of lipid A may be OM stabilization by the bridging of these positively-charged groups with unmodified negatively charged phosphates (Figure 37). OM stabilization would also protect these Gram-negative bacteria from membrane-perturbation agents, such as those encountered in the environment and in hosts. Therefore, I speculate modification of lipid A with positively charged groups provide an underappreciated protection against such membrane-destabilizing agents, especially in the context of host-pathogen interactions.

However, the role of lipid A acylation patterns and modification of these acyl chains in CAMP resistance and TLR4 activation is not as clear as the role of phosphate modification. Differences in lipid A acylation patterns can affect CAMP resistance, though this is not always the case (Trent, Stead, et al. 2006). *Salmonella* with mutations in *pagP*, the enzyme that adds a secondary palmitate acyl chain to the primary C2 carbon of *Salmonella* lipid A to produce hepta-acyl lipid A molecules, have decreased CAMP resistance (Guo, Lim, et al. 1998). This mutation also decreases OM permeability in *Salmonella*, and it is therefore hypothesized that lipid A with 7 acyl chains, as found in wild-type *Salmonella*, has more tightly packed LPS, thus increasing the barrier activity of the OM compared to the *pagP* mutant, which has only hexa-acyl lipid A (Guo, Lim, et al. 1998, Trent, Stead, et al. 2006). The activity of PagP in *B. bronchiseptica* also generates hepta-acyl lipid A species, however this does not increase resistance to CAMPs, though in *B. bronchiseptica* PagP adds a secondary palmitate chain at the C3' position (Preston, Maxim, et al. 2003). The removal of the C3 acyl chain by PagL in *Salmonella* would presumably result in a less-tightly packed LPS in the OM; however, this modification also does not affect CAMP resistance (Trent, Stead, et al. 2006). My results show decreasing the length of the C3 and C3' position acyl chains in *E. coli* results in decreased in resistance to the CAMP polymyxins B. Decreasing the acyl chain lengths

in *E. coli* may result in LPS in the OM that is not as tightly packed, and therefore may increase OM permeability, leading to increased susceptibility to CAMPs. Overall, whereas some data support the hypothesis that greater levels of LPS acylation results in greater CAMP resistance, there are likely other factors involved, since this does not hold true in the case of PagP in *Bordetella* or PagL in *Salmonella*. Lipid A acylation patterns also play a complex role in activation of TLR4. Needham *et al* found hTLR4 activation is affected not just by the number of acyl chains on an LPS molecule, but also the position of the acyl chains. For example, addition of a secondary palmitate group at the C2 position of hexa-acyl *E. coli* LPS, to generate hepta-acyl LPS, does not significantly affect hTLR4 activation, though this same modification in penta-acyl *E. coli* LPS decreases hTLR4 activation (Needham, Carroll, et al. 2013). My findings show the length of these acyl chains also affect hTLR4 activation by LPS. Increasing C3' acyl chain length in non-GlcN-modified penta-acyl *B. pertussis* lipid A increases activation of hTLR4, as does an increase in the C3 and C3' acyl chain lengths in hexa-acyl *E. coli*. However, in the case of *E. coli*, it is unclear whether the effect on hTLR4 activation is due to modification of the C3 acyl chain, the C3'acyl chain, or both. Furthermore, it is unknown whether changing the length of any other acyl chains in these lipid A structures would affect hTLR4 activation. In the case of *B. pertussis* LPS, I observed that increasing the C3' acyl chain in GlcN-modified LPS did not affect hTLR4 activation, therefore suggesting that the effect of some acyl chain modifications may depend on the presence or absence of other lipid A modifications.

Differential activation of hTLR4 by varying lipid A structures is due to the interaction of lipid A with the human TLR4-MD-2 receptor. The solved structures of hexa-acyl *E. coli* lipid A, with human TLR4-MD-2 and tetra-acyl lipid IV_A bound to human MD-2 has revealed some interactions that are important for hTLR4 activation. *E. coli* lipid A is a good agonist of hTLR4 whereas lipid IV_A is a poor activator of hTLR4. In general, these structures indicate the hydrophobic acyl chains of lipid A and lipid IV_A sit inside the pocket formed by MD-2, and the phosphate groups sit above the MD-2 pocket to interact with amino

acid residues in both TLR4 and TLR4*, thereby promoting dimerization and activation (Figure 3) (Ohto, Fukase, et al. 2007, Park, Song, et al. 2009). When comparing the solved structures for *E. coli* lipid A and lipid IV_A within MD-2, the two molecules are sitting in these co-crystals in opposite orientations, such that the 1-phosphate of *E. coli* lipid A is in the equivalent position as the 4'-phosphate of lipid IV_A (Park, Song, et al. 2009). The orientation of the lipid A or lipid IV_A inside MD-2 can have important implications for which phosphate group (1- or 4'-) and which acyl chain interacts with the dimerizing TLR4* (Figure 3) during activation of TLR4. It is possible that the specific orientation of the di-GlcN backbone in each of these structures is due to a crystallization artifact, and each molecule is able to interact with MD-2-TLR4 in either orientation *in vivo*. Furthermore, since there are fewer acyl chains to fit inside the MD-2 pocket in lipid IV_A, the entire molecule sits deeper inside MD-2 and the phosphate groups of lipid IV_A are positioned lower and at an angle compared to the phosphate groups of *E. coli* lipid A. This difference in phosphate group positioning is likely a main cause of the decreased hTLR4 activation by lipid IV_A compared to *E. coli* lipid A, since these negatively-charged phosphate groups interact with positively-charged amino acid residues in TLR4 and TLR4*. Another important difference in these structures is the position of the acyl chains: all four acyl chains of lipid IV_A sit inside the MD-2 pocket, however only five of the six acyl-chains of *E. coli* lipid A sit inside the MD-2 pocket, the sixth C2 acyl chain is positioned between the MD-2 molecule and TLR4*. The interaction of the C2 acyl chain with TLR4* is hypothesized to be important for TLR4 activation, though if hexa-acyl lipid A molecules can interact with TLR4-MD-2 in the flipped orientation, other acyl chains, such as the C3' acyl chain, may also be involved in this manner (Maeshima and Fernandez 2013, Park, Song, et al. 2009). I have shown that decreasing C3 and C3' acyl chain lengths in hexa-acyl *E. coli* lipid A results in a decrease in hTLR4 activation. Based on the structure of *E. coli* lipid A and TLR4-MD-2, I hypothesize that the decrease in the acyl chain lengths likely lowers the position of the lipid A molecule in the MD-2 pocket, therefore positioning the phosphate groups at a suboptimal level. This would result in decreased interaction between the phosphate groups with positively charged amino acids in TLR4 and TLR4*,

therefore resulting in a reduction of hTLR4 activation. If *E. coli* LPS can also interact with TLR4-MD-2 in the opposite orientation, such that the 1-phosphate interacts with TLR4*, it is also possible the C3' acyl chain may sit outside of the MD-2 pocket and interact with TLR4*, in which case a shorter C3' acyl chain would also decrease interaction with TLR4*.

Unlike *E. coli*, many Gram-negative species have penta-acyl lipid A, as found in *B. pertussis*. Though the structure of TLR4-MD-2 bound to penta-acyl lipid A has yet to be determined, the solved structures of tetra-acyl lipid IV_A and hexa-acyl lipid A within MD-2 can also shed light on how *B. pertussis* lipid A may interact with TLR4-MD-2 (Ohto, Fukase, et al. 2007, Park, Song, et al. 2009). However, it still remains unclear which orientation a penta-acyl lipid A molecule would have inside the MD-2 pocket; would it sit with the 4'-phosphate interacting with TLR4*, as lipid IV_A does, or would it sit with the 1-phosphate in this position, as seen in hexa-acyl lipid A? Furthermore, based on these structures alone, it is unclear whether all five acyl chains of *B. pertussis* lipid A would sit inside the MD-2 pocket or whether a single acyl chain would extrude from this pocket to interact with TLR4*. Recent work in our laboratory suggests that all 5 acyl chains of *B. pertussis* lipid A lie within the MD-2 pocket *in vivo* and the dimerization of TLR4 is instead mediated by a MD-2-TLR4* interaction (Maeshima N, Fernandez RC, unpublished data). Despite these uncertainties, the aforementioned solved structures still provide a basis for interaction between TLR4-MD-2, lipid A, and the incoming dimerizing TLR4*-MD-2*-lipid A (Figure 3). In the case of *B. pertussis* lipid A, which has only five acyl chains, I propose, in the absence of the GlcN modification on the phosphate groups and with shorter C3′ acyl chains, as seen in 18-323, the phosphate groups of lipid A are not able to interact with the key positively charged residues in TLR4 and/or TLR4* to cause dimerization and activation. However, when the C3′ acyl chain length is increased, this may position the lipid A molecule slightly more out of the MD-2 pocket and slightly closer to the positively charged residues in TLR4 and TLR4*, thus significantly increasing TLR4 dimerization and activation. Alternatively, when the phosphates are modified with positively charged GlcN, there is

greater hTLR4 activation when compared with unmodified lipid A. Based on ongoing work in our laboratory, the GlcN moieties on this penta-acyl lipid A likely interacts with negatively charged residues in TLR4 and TLR4* to promote dimerization (Maeshima N, Evans-Atkinson T, Hajjar AM, Fernandez RC, in revision). Adding both GlcN and longer acyl chain modifications to 18-323 lipid A also increases hTLR4 activation when compared with both wild type 18-323 and 18-323 with only the longer acyl chain modification, but this level is statistically equal to that of 18-323 with only the GlcN modification. This suggests that the slight increase in the C3′ acyl chain length in the GlcN-modified lipid A does not change the position within the TLR4-MD-2 complex enough to significantly affect the interaction between the negatively charged residues in TLR4 and TLR4* and the GlcNs.

Since activation of TLR4 by LPS leads to downstream NFκB and IRF3 activation and the generation of both an inflammatory response and a type I interferon response, respectively, LPS is a possible candidate for pharmacological uses, exemplified by the use of monophosphoryl lipid A as an adjuvant (Maeshima and Fernandez 2013). What makes LPS an even more attractive contender is the possibility of modulating the activation of hTLR4 by adjusting the structure of the lipid A region of LPS. Therefore, building a repertoire of tools to modify the structure of lipid A could aid in the design of superior adjuvants – even slight changes to the lipid A, such as changing the acyl chain lengths, may allow fine-tuning the hTLR4 response.

6.2 Conclusions

In conclusion, I have shown *lgmA*, *lgmB,* and *lgmC* are required for glycosylation of lipid A in *B. pertussis*, whereas *lgmD*, *lgmE,* and BP1945 are not. This leaves the identity or requirement for a C55P-GlcN flippase in the GlcN modification pathway unclear. However, I have demonstrated LgmA transfers GlcNAc from UDP-GlcNAc onto a lipid, which is likely the C55P, therefore supporting the first proposed step in this pathway. LgmC has been shown by others to function in the removal of the acetyl group from

C55P-GlcNAc to generate C55P-GlcN (Llewellyn, Zhao, et al. 2012), which also supports the proposed GlcN modification pathway. Furthermore I characterized a putative active site for LgmA (D76, D77, D127, and D129) and LgmC (D80, D81, H130, D187, H189, and E313).

I also elucidated the genetic basis for the structural differences in lipid A between *B. pertussis* strains BP338 and 18-323. The lack of GlcN modification in 18-323 is due to the absence of a complete *lgm* locus and the lack of 14-carbon long acyl chains at the C3' position of the lipid A in 18-323 is due to a single amino acid difference in the hydrocarbon ruler region of the enzyme LpxA between the two *B. pertussis* strains.

Furthermore, I demonstrated lipid A modifications in *B. pertussis* affect different biological interactions. The GlcN modification increases *B. pertussis* resistance to a variety of CAMPs, though not to the aminoglycoside gentamicin, and also increases resistance to OM perturbation. Additionally, the GlcN modification and longer C3' acyl chain lengths in *B. pertussis* lipid A individually increase activation of hTLR4. However, when both modifications are present, the same level of hTLR4 activation is observed as with only the GlcN modification, suggesting the GlcN modification may have a dominant effect on increasing activation.

In regards to *E. coli* lipid A structural differences, I demonstrated shortening the C3 and C3' acyl chains results in a minor growth defect, decreases resistance to polymyxin B, attenuates activation of hTLR4 by LPS, and increases LAL activation by LPS.

6.3 Future directions

There are many avenues for future research related to this project. Firstly, there are still questions regarding the regulation of the *lgm* locus, and consequently the GlcN modification of LPS, in *B. pertussis*.

lgmA and *lgmB* have a higher level of transcription when the bacteria are grown in SS broth compared to on BG agar (Marr, Tirsoaga, et al. 2008). The signal facilitating this difference in transcription is still unknown. Furthermore, there may be other Bvg-independent environmental signals, such as growth phase or the presence of CAMPs, which may further modulate regulation of the *lgm* locus.

Linked to these inquiries is the role of *lgmE*. I found *lgmE* mRNA overlaps, and would therefore anneal to, at least *lgmB* and *lgmC* mRNA. However, depending on the actual length of the *lgmE* transcript, which is still unknown, it may also overlap with *lgmA*. Additionally, the manner in which *lgmABCD* is transcribed, e.g. as one single transcript or as four individual transcripts, may affect how *lgmE* interacts and regulates expression of the *lgm* locus genes. If the *lgmE* transcript does anneal to *lgmABCD* mRNAs, then determining whether this association promotes the degradation of *lgmABCD* mRNAs or protects the *lgmABCD* mRNAs from degradation would shed light on how this locus is regulated. Also, *lgmE* may play a role in the non-Bvg-associated regulation of *lgmA* and *lgmB*.

Another avenue of research is explaining the lack of GlcN-modified lipid A in *B. hinzii* and *B. trematum*, since both species have a complete copy of *lgmA*, *lgmB*, and *lgmC*; thus far, the only genes that are required for this modification in *B. pertussis*. It is possible that another, as yet unidentified gene is required for the GlcN modification in *B. pertussis*, and this additional gene is not present in *B. hinzii* or *B. trematum*. Alternatively, an inactivating mutation in *lgmA*, *lgmB*, or *lgmC* could be responsible for the lack of modification in *B. hinzii* and *B. trematum*, or the lgm genes may not be expressed in these two species.

Though I have shown evidence that LgmA transfers GlcNAc onto a lipid, which is likely C55P, we have yet to prove the lipid acceptor of this reaction is the inner membrane carrier lipid C55P. Furthermore, the function of LgmB, the glycosyltransferase predicted to transfer GlcN from C55P to the phosphate group

of lipid A, has also not been proven. To complete the model for the Lgm pathway, determining whether a flippase is required, and if so, identifying this elusive flippase would also be important. Determining which face of the IM each Lgm enzyme functions at may reveal clues as to the requirement of a flippase in this system, because if all the Lgm enzymes function at the cytoplasmic face of the IM, there would be no need for a C55P-sugar flippase.

Further analysis of the LgmA and LgmC active sites is another topic to research. The amino acids I identified as possibly being part of the putative active sites of these enzymes need to be confirmed by further mutational analysis, perhaps by mutating the residues to amino acids that take up the same amount of space, but lack the functional group. Additional residues can also be targeted for mutational analysis, to broaden the characterization of the active sites of not only LgmA and LgmC, but also LgmB. GT83 glycosyltransferases are not well studied, so further study of LgmB, a member of this protein family, may lead to insights applicable to other GT83 protein family members.

References

1. **Alonso, S., N. Reveneau, K. Pethe, and C. Locht.** 2002. Eighty-kilodalton N-terminal moiety of *Bordetella pertussis* filamentous hemagglutinin: adherence, immunogenicity, and protective role. Infection and immunity **70:**4142-4147.
2. **Anderson, M. S., and C. R. Raetz.** 1987. Biosynthesis of lipid A precursors in *Escherichia coli*. A cytoplasmic acyltransferase that converts UDP-N-acetylglucosamine to UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine. The Journal of biological chemistry **262:**5159-5169.
3. **Banemann, A., H. Deppisch, and R. Gross.** 1998. The lipopolysaccharide of *Bordetella bronchiseptica* acts as a protective shield against antimicrobial peptides. Infection and immunity **66:**5607-5612.
4. **Barnes, M. G., and A. A. Weiss.** 2001. BrkA protein of *Bordetella pertussis* inhibits the classical pathway of complement after C1 deposition. Infection and immunity **69:**3067-3072.
5. **Boekema, B. K., N. Stockhofe-Zurwieden, H. E. Smith, E. M. Kamp, J. P. van Putten, and J. H. Verheijden.** 2003. Adherence of Actinobacillus pleuropneumoniae to primary cultures of porcine lung epithelial cells. Veterinary microbiology **93:**133-144.
6. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry **72:**248-254.
7. **Breazeale, S. D., A. A. Ribeiro, A. L. McClerren, and C. R. Raetz.** 2005. A formyltransferase required for polymyxin resistance in *Escherichia coli* and the modification of lipid A with 4-Amino-4-deoxy-L-arabinose. Identification and function oF UDP-4-deoxy-4-formamido-L-arabinose. The Journal of biological chemistry **280:**14154-14167.
8. **Butler, E. K., R. M. Davis, V. Bari, P. A. Nicholson, and N. Ruiz.** 2013. Structure-function analysis of MurJ reveals a solvent-exposed cavity containing residues essential for peptidoglycan biogenesis in *Escherichia coli*. Journal of bacteriology **195:**4639-4649.
9. **Carbonetti, N. H.** 2010. Pertussis toxin and adenylate cyclase toxin: key virulence factors of Bordetella pertussis and cell biology tools. Future microbiology **5:**455-469.
10. **Caroff, M., J. Brisson, A. Martin, and D. Karibian.** 2000. Structure of the *Bordetella pertussis* 1414 endotoxin. FEBS letters **477:**8-14.
11. **Caroff, M., and D. Karibian.** 2003. Structure of bacterial lipopolysaccharides. Carbohydrate research **338:**2431-2447.
12. **Caroff, M., D. Karibian, J. M. Cavaillon, and N. Haeffner-Cavaillon.** 2002. Structural and functional analyses of bacterial lipopolysaccharides. Microbes and infection / Institut Pasteur **4:**915-926.
13. **Casella, C. R., and T. C. Mitchell.** 2008. Putting endotoxin to work for us: monophosphoryl lipid A as a safe and effective vaccine adjuvant. Cellular and molecular life sciences : CMLS **65:**3231-3240.
14. **Chaby, R., I. Garcia-Verdugo, Q. Espinassous, and L. A. Augusto.** 2005. Interactions between LPS and lung surfactant proteins. Journal of endotoxin research **11:**181-185.
15. **Chen, L., and N. Mozier.** 2013. Comparison of Limulus amebocyte lysate test methods for endotoxin measurement in protein solutions. Journal of pharmaceutical and biomedical analysis **80:**180-185.
16. **Chng, S. S., L. S. Gronenberg, and D. Kahne.** 2010. Proteins required for lipopolysaccharide assembly in *Escherichia coli* form a transenvelope complex. Biochemistry **49:**4565-4567.
17. **Clementz, T., and C. R. Raetz.** 1991. A gene coding for 3-deoxy-D-manno-octulosonic-acid transferase in *Escherichia coli*. Identification, mapping, cloning, and sequencing. The Journal of biological chemistry **266:**9687-9696.

18. **Coats, S. R., A. B. Berezow, T. T. To, S. Jain, B. W. Bainbridge, K. P. Banani, and R. P. Darveau.** 2011. The lipid A phosphate position determines differential host Toll-like receptor 4 responses to phylogenetically related symbiotic and pathogenic bacteria. Infection and immunity **79:**203-210.
19. **Crowcroft, N. S., C. Stein, P. Duclos, and M. Birmingham.** 2003. How best to estimate the global burden of pertussis? Lancet Infect Dis **3:**413-418.
20. **Cummings, C. A., H. J. Bootsma, D. A. Relman, and J. F. Miller.** 2006. Species- and strain-specific control of a complex, flexible regulon by *Bordetella* BvgAS. Journal of bacteriology **188:**1775-1785.
21. **de Gouw, D., D. A. Diavatopoulos, H. J. Bootsma, P. W. Hermans, and F. R. Mooi.** 2011. Pertussis: a matter of immune modulation. FEMS microbiology reviews **35:**441-474.
22. **De Smet, K., and R. Contreras.** 2005. Human antimicrobial peptides: defensins, cathelicidins and histatins. Biotechnology letters **27:**1337-1347.
23. **Decker, K. B., T. D. James, S. Stibitz, and D. M. Hinton.** 2012. The *Bordetella pertussis* model of exquisite gene control by the global transcription factor BvgA. Microbiology **158:**1665-1676.
24. **El Hamidi, A., A. Tirsoaga, A. Novikov, A. Hussein, and M. Caroff.** 2005. Microextraction of bacterial lipid A: easy and rapid method for mass spectrometric characterization. Journal of lipid research **46:**1773-1778.
25. **Elahi, S., R. M. Buchanan, S. Attah-Poku, H. G. Townsend, L. A. Babiuk, and V. Gerdts.** 2006. The host defense peptide beta-defensin 1 confers protection against *Bordetella pertussis* in newborn piglets. Infection and immunity **74:**2338-2352.
26. **Fedele, G., M. Bianco, and C. M. Ausiello.** 2013. The Virulence Factors of *Bordetella pertussis*: Talented Modulators of Host Immune Response. Archivum immunologiae et therapiae experimentalis **61:**445-457.
27. **Fernandez, L., H. Jenssen, M. Bains, I. Wiegand, W. J. Gooderham, and R. E. Hancock.** 2012. The two-component system CprRS senses cationic peptides and triggers adaptive resistance in *Pseudomonas aeruginosa* independently of ParRS. Antimicrobial agents and chemotherapy **56:**6212-6222.
28. **Fernandez, R. C., and A. A. Weiss.** 1994. Cloning and sequencing of a *Bordetella pertussis* serum resistance locus. Infection and immunity **62:**4727-4738.
29. **Fernandez, R. C., and A. A. Weiss.** 1996. Susceptibilities of *Bordetella pertussis* strains to antimicrobial peptides. Antimicrobial agents and chemotherapy **40:**1041-1043.
30. **Freinkman, E., S. S. Chng, and D. Kahne.** 2011. The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. Proceedings of the National Academy of Sciences of the United States of America **108:**2486-2491.
31. **Furste, J. P., W. Pansegrau, R. Frank, H. Blocker, P. Scholz, M. Bagdasarian, and E. Lanka.** 1986. Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. Gene **48:**119-131.
32. **Galloway, S. M., and C. R. Raetz.** 1990. A mutant of *Escherichia coli* defective in the first step of endotoxin biosynthesis. The Journal of biological chemistry **265:**6394-6402.
33. **Gascuel, O.** 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Molecular biology and evolution **14:**685-695.
34. **Gellatly, S. L., B. Needham, L. Madera, M. S. Trent, and R. E. Hancock.** 2012. The *Pseudomonas aeruginosa* PhoP-PhoQ two-component regulatory system is induced upon interaction with epithelial cells and controls cytotoxicity and inflammation. Infection and immunity **80:**3122-3131.
35. **Gerlach, G., F. von Wintzingerode, B. Middendorf, and R. Gross.** 2001. Evolutionary trends in the genus *Bordetella*. Microbes and Infection **3:**61-72.
36. **Geuijen, C. A., R. J. Willems, and F. R. Mooi.** 1996. The major fimbrial subunit of *Bordetella pertussis* binds to sulfated sugars. Infection and immunity **64:**2657-2665.

37. **Gronow, S., W. Brabetz, and H. Brade.** 2000. Comparative functional characterization in vitro of heptosyltransferase I (WaaC) and II (WaaF) from *Escherichia coli*. European journal of biochemistry / FEBS **267:**6602-6611.
38. **Guan, S., D. A. Bastin, and N. K. Verma.** 1999. Functional analysis of the O antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. Microbiology **145 (Pt 5):**1263-1273.
39. **Guex, N., and M. C. Peitsch.** 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis **18:**2714-2723.
40. **Gunn, J. S., K. B. Lim, J. Krueger, K. Kim, L. Guo, M. Hackett, and S. I. Miller.** 1998. PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. Molecular microbiology **27:**1171-1182.
41. **Guo, L., K. B. Lim, C. M. Poduje, M. Daniel, J. S. Gunn, M. Hackett, and S. I. Miller.** 1998. Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. Cell **95:**189-198.
42. **Gustafsson, L., L. Hessel, J. Storsaeter, and P. Olin.** 2006. Long-term follow-up of Swedish children vaccinated with acellular pertussis vaccines at 3, 5, and 12 months of age indicates the need for a booster dose at 5 to 7 years of age. Pediatrics **118:**978-984.
43. **Gutsmann, T., J. Howe, U. Zahringer, P. Garidel, A. B. Schromm, M. H. Koch, Y. Fujimoto, K. Fukase, I. Moriyon, G. Martinez-de-Tejada, and K. Brandenburg.** 2010. Structural prerequisites for endotoxic activity in the Limulus test as compared to cytokine production in mononuclear cells. Innate immunity **16:**39-47.
44. **Hallander, H. O., L. Gustafsson, M. Ljungman, and J. Storsaeter.** 2005. Pertussis antitoxin decay after vaccination with DTPa Response to a first booster dose 31/2-61/2 years after the third vaccine dose. Vaccine **23:**5359-5364.
45. **Hancock, R. E.** 1997. The bacterial outer membrane as a drug barrier. Trends in microbiology **5:**37-42.
46. **Hazenbos, W. L., C. A. Geuijen, B. M. van den Berg, F. R. Mooi, and R. van Furth.** 1995. *Bordetella pertussis* fimbriae bind to human monocytes via the minor fimbrial subunit FimD. The Journal of infectious diseases **171:**924-929.
47. **Imagawa, T., H. Iino, M. Kanagawa, A. Ebihara, S. Kuramitsu, and H. Tsuge.** 2008. Crystal structure of the YdjC-family protein TTHB029 from *Thermus thermophilus* HB8: structural relationship with peptidoglycan N-acetylglucosamine deacetylase. Biochemical and biophysical research communications **367:**535-541.
48. **Impellitteri, N. A., J. A. Merten, L. E. Bretscher, and C. S. Klug.** 2010. Identification of a functionally important loop in *Salmonella typhimurium* ArnT. Biochemistry **49:**29-35.
49. **Inatsuka, C. S., Q. Xu, I. Vujkovic-Cvijin, S. Wong, S. Stibitz, J. F. Miller, and P. A. Cotter.** 2010. Pertactin is required for *Bordetella* species to resist neutrophil-mediated clearance. Infection and immunity **78:**2901-2909.
50. **Kagan, J. C., T. Su, T. Horng, A. Chow, S. Akira, and R. Medzhitov.** 2008. TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. Nature immunology **9:**361-368.
51. **Kawasaki, K., R. K. Ernst, and S. I. Miller.** 2004. Deacylation and palmitoylation of lipid A by *Salmonellae* outer membrane enzymes modulate host signaling through Toll-like receptor 4. Journal of endotoxin research **10:**439-444.
52. **Kelley, L. A., and M. J. Sternberg.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols **4:**363-371.
53. **Kelly, T. M., S. A. Stachula, C. R. Raetz, and M. S. Anderson.** 1993. The firA gene of Escherichia coli encodes UDP-3-O-(R-3-hydroxymyristoyl)-glucosamine N-acyltransferase. The third step of endotoxin biosynthesis. The Journal of biological chemistry **268:**19866-19874.

54. **Ko, K. S., K. R. Peck, W. S. Oh, N. Y. Lee, J. H. Lee, and J. H. Song.** 2005. New species of *Bordetella, Bordetella ansorpii* sp. nov., isolated from the purulent exudate of an epidermal cyst. Journal of clinical microbiology **43:**2516-2519.
55. **Kovach, M. E., P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop, 2nd, and K. M. Peterson.** 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene **166:**175-176.
56. **Lairson, L. L., B. Henrissat, G. J. Davies, and S. G. Withers.** 2008. Glycosyltransferases: structures, functions, and mechanisms. Annual review of biochemistry **77:**521-555.
57. **Laube, D. M., S. Yim, L. K. Ryan, K. O. Kisich, and G. Diamond.** 2006. Antimicrobial peptides in the airway. Current topics in microbiology and immunology **306:**153-182.
58. **Leung, A. K. C., W. L. M. Robson, and H. D. Davies.** 2007. Pertussis in adolescents. Adv Ther **24:**353-361.
59. **Llewellyn, A. C., J. Zhao, F. Song, J. Parvathareddy, Q. Xu, B. A. Napier, H. Laroui, D. Merlin, J. E. Bina, P. A. Cotter, M. A. Miller, C. R. Raetz, and D. S. Weiss.** 2012. NaxD is a deacetylase required for lipid A modification and *Francisella* pathogenesis. Molecular microbiology **86:**611-627.
60. **Maeshima, N., and R. C. Fernandez.** 2013. Recognition of lipid A variants by the TLR4-MD-2 receptor complex. Frontiers in cellular and infection microbiology **3:**3.
61. **Marr, N., A. M. Hajjar, N. R. Shah, A. Novikov, C. S. Yam, M. Caroff, and R. C. Fernandez.** 2010. Substitution of the *Bordetella pertussis* lipid A phosphate groups with glucosamine is required for robust NF-kappaB activation and release of proinflammatory cytokines in cells expressing human but not murine Toll-like receptor 4-MD-2-CD14. Infection and immunity **78:**2060-2069.
62. **Marr, N., A. Novikov, A. M. Hajjar, M. Caroff, and R. C. Fernandez.** 2010. Variability in the lipooligosaccharide structure and endotoxicity among *Bordetella pertussis* strains. The Journal of infectious diseases **202:**1897-1906.
63. **Marr, N., N. R. Shah, R. Lee, E. J. Kim, and R. C. Fernandez.** 2011. *Bordetella pertussis* autotransporter Vag8 binds human C1 esterase inhibitor and confers serum resistance. PloS one **6:**e20585.
64. **Marr, N., A. Tirsoaga, D. Blanot, R. Fernandez, and M. Caroff.** 2008. Glucosamine found as a substituent of both phosphate groups in *Bordetella* lipid A backbones: role of a BvgAS-activated ArnT ortholog. Journal of bacteriology **190:**4281-4290.
65. **Mattoo, S., and J. D. Cherry.** 2005. Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to *Bordetella pertussis* and other *Bordetella* subspecies. Clinical microbiology reviews **18:**326-382.
66. **Merkel, T. J., P. E. Boucher, S. Stibitz, and V. K. Grippe.** 2003. Analysis of *bvgR* expression in *Bordetella pertussis*. Journal of bacteriology **185:**6902-6912.
67. **Metzger, L. E. t., and C. R. Raetz.** 2010. An alternative route for UDP-diacylglucosamine hydrolysis in bacterial lipid A biosynthesis. Biochemistry **49:**6715-6726.
68. **Mielcarek, N., A. S. Debrie, D. Raze, J. Quatannens, J. Engle, W. E. Goldman, and C. Locht.** 2006. Attenuated *Bordetella pertussis*: new live vaccines for intranasal immunisation. Vaccine **24:**S54-S55.
69. **Miller, S. I., R. K. Ernst, and M. W. Bader.** 2005. LPS, TLR4 and infectious disease diversity. Nature reviews. Microbiology **3:**36-46.
70. **Moskowitz, S. M., R. K. Ernst, and S. I. Miller.** 2004. PmrAB, a two-component regulatory system of *Pseudomonas aeruginosa* that modulates resistance to cationic antimicrobial peptides and addition of aminoarabinose to lipid A. Journal of bacteriology **186:**575-579.
71. **Muta, T., T. Miyata, Y. Misumi, F. Tokunaga, T. Nakamura, Y. Toh, Y. Ikehara, and S. Iwanaga.** 1991. Limulus factor C. An endotoxin-sensitive serine protease zymogen with a mosaic

structure of complement-like, epidermal growth factor-like, and lectin-like domains. The Journal of biological chemistry **266:**6554-6561.

72. **Needham, B. D., S. M. Carroll, D. K. Giles, G. Georgiou, M. Whiteley, and M. S. Trent.** 2013. Modulating the innate immune response by combinatorial engineering of endotoxin. Proceedings of the National Academy of Sciences of the United States of America **110:**1464-1469.
73. **Needham, B. D., and M. S. Trent.** 2013. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. Nature reviews. Microbiology **11:**467-481.
74. **Novikov, A., N. R. Shah, S. Albitar-Nehme, S. M. Basheer, I. Trento, A. Tirsoaga, M. Moksa, M. Hirst, M. B. Perry, A. E. Hamidi, R. C. Fernandez, and M. Caroff.** 2013. Complete *Bordetella avium, Bordetella hinzii* and *Bordetella trematum* lipid A structures and genomic sequence analyses of the loci involved in their modifications. Innate immunity.
75. **Ohto, U., K. Fukase, K. Miyake, and Y. Satow.** 2007. Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. Science **316:**1632-1634.
76. **Oliver, D. C., and R. C. Fernandez.** 2001. Antibodies to BrkA augment killing of *Bordetella pertussis*. Vaccine **20:**235-241.
77. **Park, B. S., D. H. Song, H. M. Kim, B. S. Choi, H. Lee, and J. O. Lee.** 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature **458:**1191-1195.
78. **Park, J., Y. Zhang, A. M. Buboltz, X. Zhang, S. C. Schuster, U. Ahuja, M. Liu, J. F. Miller, M. Sebaihia, S. D. Bentley, J. Parkhill, and E. T. Harvill.** 2012. Comparative genomics of the classical *Bordetella* subspecies: the evolution and exchange of virulence-associated diversity amongst closely related pathogens. BMC genomics **13:**545.
79. **Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabbinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell.** 2003. Comparative analysis of the genome sequences of *Bordetella pertussis, Bordetella parapertussis* and *Bordetella bronchiseptica*. Nature genetics **35:**32-40.
80. **Plotz, B. M., B. Lindner, K. O. Stetter, and O. Holst.** 2000. Characterization of a novel lipid A containing D-galacturonic acid that replaces phosphate residues - The structure of the lipid A of the lipopolysaccharide from the hyperthermophilic bacterium *Aquifex pyrophilus*. Journal of Biological Chemistry **275:**11222-11228.
81. **Pomorski, T., and A. K. Menon.** 2006. Lipid flippases and their biological functions. Cellular and molecular life sciences : CMLS **63:**2908-2921.
82. **Preston, A., E. Maxim, E. Toland, E. J. Pishko, E. T. Harvill, M. Caroff, and D. J. Maskell.** 2003. *Bordetella bronchiseptica* PagP is a Bvg-regulated lipid A palmitoyl transferase that is required for persistent colonization of the mouse respiratory tract. Molecular microbiology **48:**725-736.
83. **Punta, M., P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. Sonnhammer, S. R. Eddy, A. Bateman, and R. D. Finn.** 2012. The Pfam protein families database. Nucleic acids research **40:**D290-301.
84. **Raetz, C. R., Z. Guan, B. O. Ingram, D. A. Six, F. Song, X. Wang, and J. Zhao.** 2009. Discovery of new biosynthetic pathways: the lipid A story. Journal of lipid research **50 Suppl:**S103-108.

85. **Raetz, C. R., C. M. Reynolds, M. S. Trent, and R. E. Bishop.** 2007. Lipid A modification systems in gram-negative bacteria. Annual review of biochemistry **76:**295-329.
86. **Raetz, C. R., and C. Whitfield.** 2002. Lipopolysaccharide endotoxins. Annual review of biochemistry **71:**635-700.
87. **Ram, S., J. Ngampasutadol, A. D. Cox, A. M. Blom, L. A. Lewis, F. St Michael, J. Stupak, S. Gulati, and P. A. Rice.** 2007. Heptose I glycan substitutions on *Neisseria gonorrhoeae* lipooligosaccharide influence C4b-binding protein binding and serum resistance. Infection and immunity **75:**4071-4081.
88. **Ravishankar, S., V. P. Kumar, B. Chandrakala, R. K. Jha, S. M. Solapure, and S. M. de Sousa.** 2005. Scintillation proximity assay for inhibitors of *Escherichia coli* MurG and, optionally, MraY. Antimicrobial agents and chemotherapy **49:**1410-1418.
89. **Ray, B. L., and C. R. Raetz.** 1987. The biosynthesis of gram-negative endotoxin. A novel kinase in *Escherichia coli* membranes that incorporates the 4'-phosphate of lipid A. The Journal of biological chemistry **262:**1122-1128.
90. **Rezania, S., N. Amirmozaffari, B. Tabarraei, M. Jeddi-Tehrani, O. Zarei, R. Alizadeh, F. Masjedian, and A. H. Zarnani.** 2011. Extraction, Purification and Characterization of Lipopolysaccharide from *Escherichia coli* and *Salmonella typhi*. Avicenna journal of medical biotechnology **3:**3-9.
91. **Robins, L. I., A. H. Williams, and C. R. Raetz.** 2009. Structural basis for the sugar nucleotide and acyl-chain selectivity of Leptospira interrogans LpxA. Biochemistry **48:**6191-6201.
92. **Rolin, O., S. J. Muse, C. Safi, S. Elahi, V. Gerdts, L. E. Hittle, R. K. Ernst, E. T. Harvill, and A. Preston.** 2014. Enzymatic Modification of Lipid A by ArnT Protects *Bordetella bronchiseptica* against Cationic Peptides and Is Required for Transmission. Infection and immunity **82:**491-499.
93. **Rosenthal, R. S., W. Nogami, B. T. Cookson, W. E. Goldman, and W. J. Folkening.** 1987. Major fragment of soluble peptidoglycan released from growing *Bordetella pertussis* is tracheal cytotoxin. Infection and immunity **55:**2117-2120.
94. **Ruiz, N.** 2008. Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II flippase in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **105:**15553-15557.
95. **Ruiz, N., L. S. Gronenberg, D. Kahne, and T. J. Silhavy.** 2008. Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **105:**5537-5542.
96. **Schaeffer, L. M., F. X. McCormack, H. Wu, and A. A. Weiss.** 2004. *Bordetella pertussis* lipopolysaccharide resists the bactericidal effects of pulmonary surfactant protein A. J Immunol **173:**1959-1965.
97. **Schagger, H., and G. von Jagow.** 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Analytical biochemistry **166:**368-379.
98. **Shah, N. R., S. Albitar-Nehme, E. Kim, N. Marr, A. Novikov, M. Caroff, and R. C. Fernandez.** 2013. Minor modifications to the phosphate groups and the C3' acyl chain length of lipid A in two *Bordetella pertussis* strains, BP338 and 18-323, independently affect Toll-like receptor 4 protein activation. The Journal of biological chemistry **288:**11751-11760.
99. **Shah, N. R., M. Moksa, A. Novikov, M. B. Perry, M. Hirst, M. Caroff, and R. C. Fernandez.** 2013. Draft Genome Sequences of *Bordetella hinzii* and *Bordetella trematum*. Genome announcements **1**.
100. **Simon, R., U. Priefer, and A. Puhler.** 1983. A broad host range mobilization system for invivo genetic-engineering - transposon mutagenesis in Gram-negative bacteria. Bio-Technol **1:**784-791.

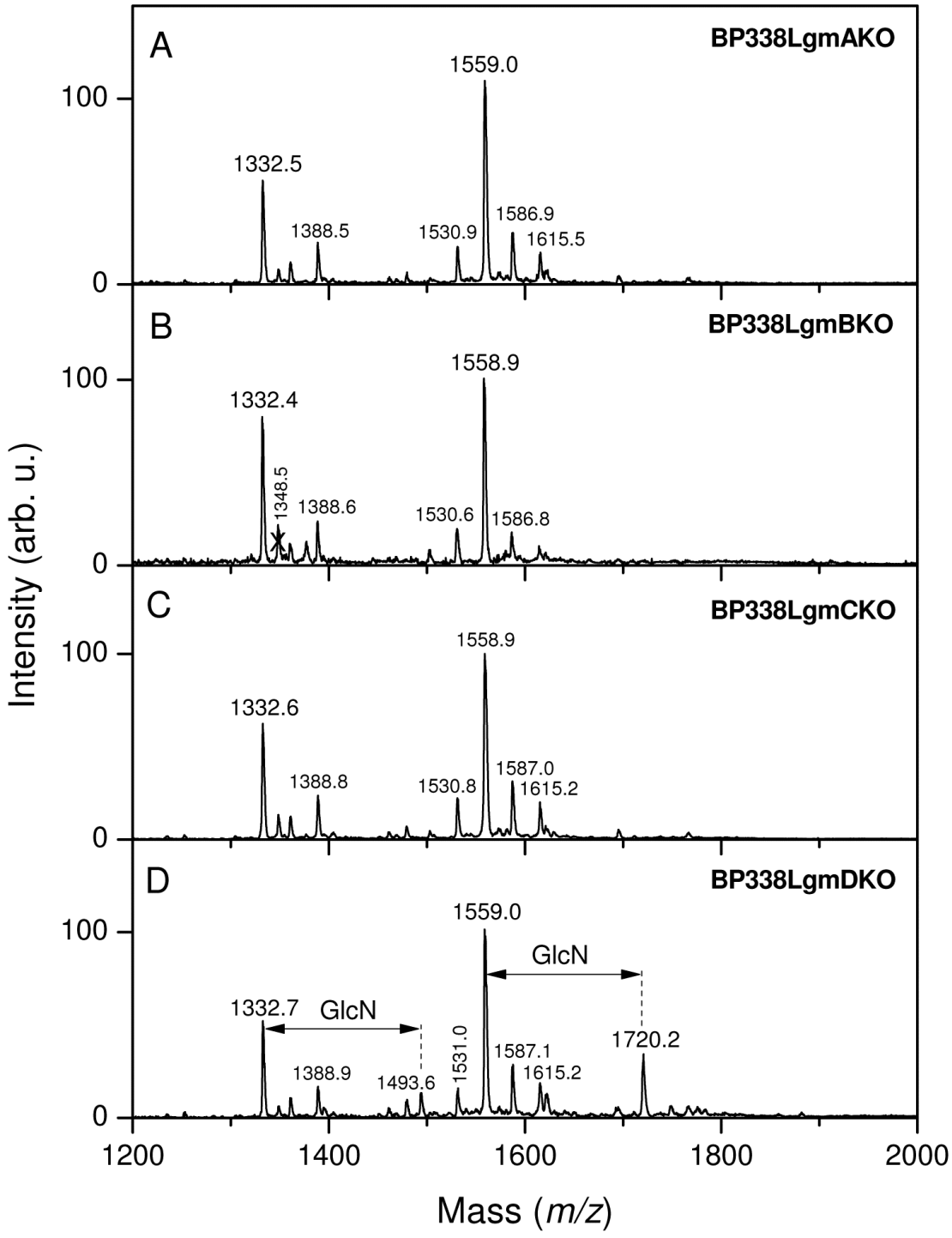
101. **Song, F., Z. Guan, and C. R. Raetz.** 2009. Biosynthesis of undecaprenyl phosphate-galactosamine and undecaprenyl phosphate-glucose in *Francisella novicida*. Biochemistry **48:**1173-1182.
102. **Sonnhammer, E. L., G. von Heijne, and A. Krogh.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology **6:**175-182.
103. **Spears, P. A., L. M. Temple, D. M. Miyamoto, D. J. Maskell, and P. E. Orndorff.** 2003. Unexpected similarities between *Bordetella avium* and other pathogenic bordetellae. Infection and immunity **71:**2591-2597.
104. **Sperandeo, P., G. Deho, and A. Polissi.** 2009. The lipopolysaccharide transport system of Gram-negative bacteria. Biochimica et biophysica acta **1791:**594-602.
105. **Stainer, D. W., and M. J. Scholte.** 1970. A simple chemically defined medium for the production of phase I *Bordetella pertussis*. Journal of general microbiology **63:**211-220.
106. **Taneja, N. K., T. Ganguly, L. O. Bakaletz, K. J. Nelson, P. Dubey, L. B. Poole, and R. Deora.** 2013. D-alanine modification of a protease-susceptible outer membrane component by the *Bordetella pertussis dra* locus promotes resistance to antimicrobial peptides and polymorphonuclear leukocyte-mediated killing. Journal of bacteriology **195:**5102-5111.
107. **Therisod, H., V. Labas, and M. Caroff.** 2001. Direct microextraction and analysis of rough-type lipopolysaccharides by combined thin-layer chromatography and MALDI mass spectrometry. Analytical chemistry **73:**3804-3807.
108. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research **22:**4673-4680.
109. **Tirsoaga, A., A. El Hamidi, M. B. Perry, M. Caroff, and A. Novikov.** 2007. A rapid, small-scale procedure for the structural characterization of lipid A applied to *Citrobacter* and *Bordetella* strains: discovery of a new structural element. Journal of lipid research **48:**2419-2427.
110. **Tirsoaga, A., A. Novikov, M. Adib-Conquy, C. Werts, C. Fitting, J. M. Cavaillon, and M. Caroff.** 2007. Simple method for repurification of endotoxins for biological use. Applied and environmental microbiology **73:**1803-1808.
111. **Tran, A. X., M. S. Trent, and C. Whitfield.** 2008. The LptA protein of *Escherichia coli* is a periplasmic lipid A-binding protein involved in the lipopolysaccharide export pathway. The Journal of biological chemistry **283:**20342-20349.
112. **Trent, M. S., W. Pabich, C. R. Raetz, and S. I. Miller.** 2001. A PhoP/PhoQ-induced Lipase (PagL) that catalyzes 3-O-deacylation of lipid A precursors in membranes of *Salmonella typhimurium*. The Journal of biological chemistry **276:**9083-9092.
113. **Trent, M. S., A. A. Ribeiro, S. Lin, R. J. Cotter, and C. R. Raetz.** 2001. An inner membrane enzyme in *Salmonella* and *Escherichia coli* that transfers 4-amino-4-deoxy-L-arabinose to lipid A: induction on polymyxin-resistant mutants and role of a novel lipid-linked donor. The Journal of biological chemistry **276:**43122-43131.
114. **Trent, M. S., C. M. Stead, A. X. Tran, and J. V. Hankins.** 2006. Diversity of endotoxin and its impact on pathogenesis. Journal of endotoxin research **12:**205-223.
115. **Trollfors, B., T. Lagergard, J. Taranger, E. Bergfors, R. Schneerson, and J. B. Robbins.** 2001. Serum immunoglobulin G antibody responses to *Bordetella pertussis* lipooligosaccharide and *B. parapertussis* lipopolysaccharide in children with pertussis and parapertussis. Clinical and diagnostic laboratory immunology **8:**1015-1017.
116. **Tsai, C. M., and C. E. Frasch.** 1982. A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. Analytical biochemistry **119:**115-119.

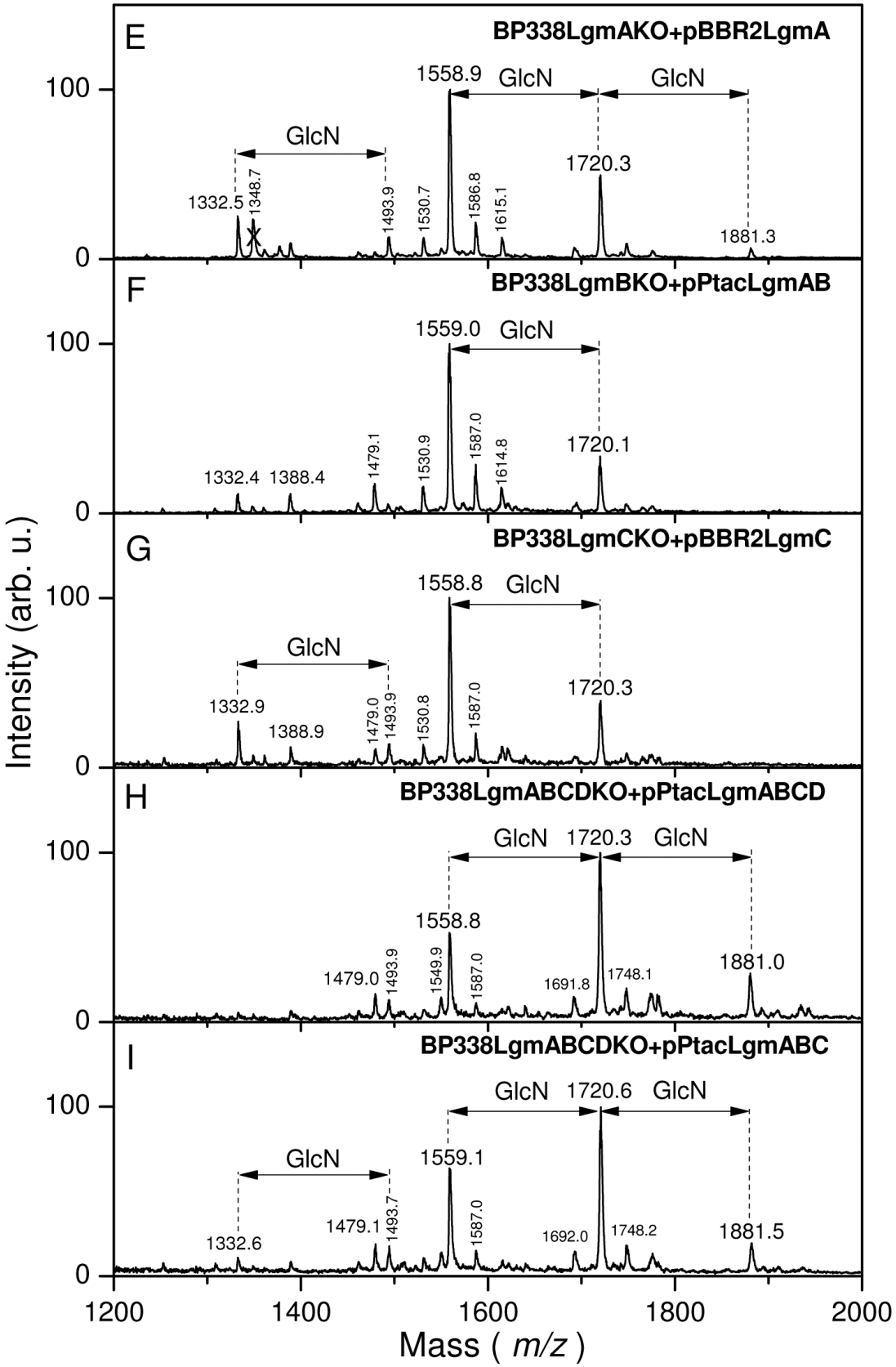
117. **Vandamme, P., J. Hommez, M. Vancanneyt, M. Monsieurs, B. Hoste, B. Cookson, C. H. W. Vonkonig, K. Kersters, and P. J. Blackall.** 1995. *Bordetella Hinzii* Sp-Nov, Isolated from Poultry and Humans. Int J Syst Bacteriol **45:**37-45.
118. **von Wintzingerode, F., A. Schattke, R. A. Siddiqui, U. Rosick, U. B. Gobel, and R. Gross.** 2001. *Bordetella petrii* sp nov., isolated from an anaerobic bioreactor, and amended description of the genus *Bordetella*. Int J Syst Evol Micr **51:**1257-1265.
119. **Wang, X., M. J. Karbarz, S. C. McGrath, R. J. Cotter, and C. R. Raetz.** 2004. MsbA transporter-dependent lipid A 1-dephosphorylation on the periplasmic surface of the inner membrane: topography of francisella novicida LpxE expressed in *Escherichia coli*. The Journal of biological chemistry **279:**49470-49478.
120. **Wang, X., S. C. McGrath, R. J. Cotter, and C. R. Raetz.** 2006. Expression cloning and periplasmic orientation of the *Francisella novicida* lipid A 4'-phosphatase LpxF. The Journal of biological chemistry **281:**9321-9330.
121. **Waters, L. S., and G. Storz.** 2009. Regulatory RNAs in bacteria. Cell **136:**615-628.
122. **Weyant, R. S., D. G. Hollis, R. E. Weaver, M. F. Amin, A. G. Steigerwalt, S. P. O'Connor, A. M. Whitney, M. I. Daneshvar, C. W. Moss, and D. J. Brenner.** 1995. *Bordetella holmesii* sp. nov., a new gram-negative species associated with septicemia. Journal of clinical microbiology **33:**1-7.
123. **Wiegand, I., K. Hilpert, and R. E. Hancock.** 2008. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nature protocols **3:**163-175.
124. **Williams, A. H., and C. R. Raetz.** 2007. Structural basis for the acyl chain selectivity and mechanism of UDP-N-acetylglucosamine acyltransferase. Proceedings of the National Academy of Sciences of the United States of America **104:**13543-13550.
125. **Wyckoff, T. J., S. Lin, R. J. Cotter, G. D. Dotson, and C. R. Raetz.** 1998. Hydrocarbon rulers in UDP-N-acetylglucosamine acyltransferases. The Journal of biological chemistry **273:**32369-32372.
126. **Yan, A., Z. Guan, and C. R. Raetz.** 2007. An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in *Escherichia coli*. The Journal of biological chemistry **282:**36077-36089.
127. **Zavascki, A. P., L. Z. Goldani, J. Li, and R. L. Nation.** 2007. Polymyxin B for the treatment of multidrug-resistant pathogens: a critical review. The Journal of antimicrobial chemotherapy **60:**1206-1215.
128. **Zerbino, D. R., and E. Birney.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research **18:**821-829.

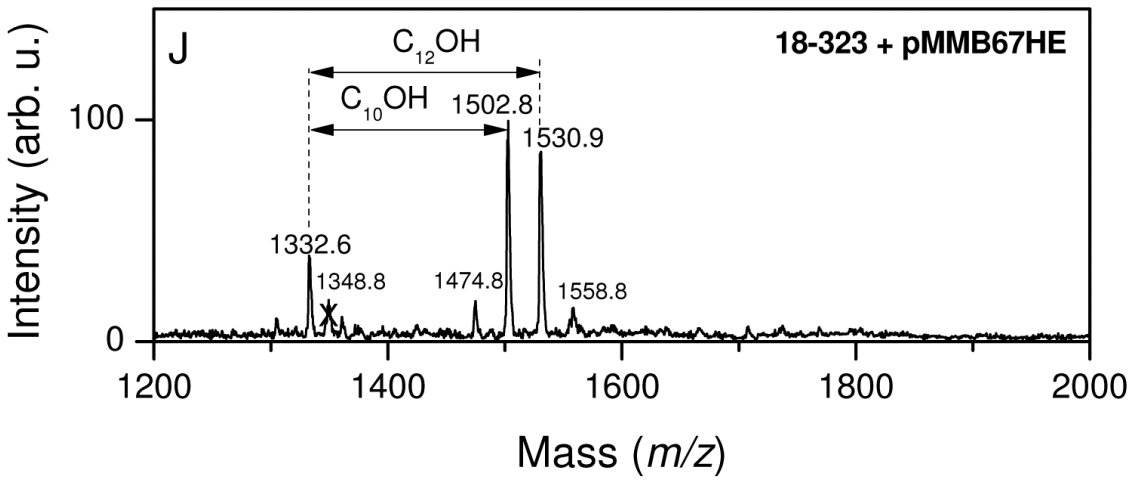
Appendices

Appendix A Mass spectra

Structural analysis of lipid A with negative-ion MALDI mass spectra of the following *B. pertussis* strains: A) BP338LgmAKO, B) BP338LgmBKO, C) BP338LgmCKO, D) BP338LmgDKO, E)BP338LgmAKO + pBBR2LgmA (complementation of BP338 *lgmA* mutant), F) BP338LgmBKO + pPtacLgmAB (complementation of BP338 *lgmB* mutant, G) BP338LgmCKO + pBBR2LgmC (complementation of BP338 *lgmC* mutant), H) BP338LgmABCDKO + pPtacLgmABCD (complementation of BP338 full *lgm* locus mutant *in trans* with full *lgm* locus, *lgmABCD*), I) BP338LgmABCDKO + pPtacLgmABC (complementation of BP338 full *lgm* locus mutant *in trans* with *lgmABC*), J) 18-323 + pMMB67HE (18-323 wild type complemented with an empty vector, a control for the pPtac vectors, as seen in Figure 34). For (A-I), peaks at *m/z* 1559 represent penta-acyl lipid A that lack GlcN modification, peaks at *m/z* 1720 represent penta-acyl lipid A with one GlcN modification at either phosphate group, and peaks at *m/z* 1881 represent penta-acyl lipid A with two GlcN modifications, one at each phosphate group. For (J), peaks at *m/z* 1520 or 1530 represent penta-acyl lipid A with C10-OH or C12-OH at the C3' position, respectively.


A
BP338LgmAKO
1559.0
1332.5
1388.5
1530.9
1586.9
1615.5
B
BP338LgmBKO
1558.9
1332.4
1348.5
1388.6
1530.6
1586.8
C
BP338LgmCKO
1558.9
1332.6
1388.8
1530.8
1587.0
1615.2
D
BP338LgmDKO
1559.0
GlcN
1332.7
GlcN
1720.2
1531.0
1587.1
1388.9
1493.6
1615.2
100
0
1200
1400
1600
1800
2000
Intensity (arb. u.)
Mass (m/z)


E
BP338LgmAKO+pBBR2LgmA
1558.9
GlcN
GlcN
GlcN
1720.3
1332.5
1348.7
1493.9
1530.7
1586.8
1615.1
1881.3
F
BP338LgmBKO+pPtacLgmAB
1559.0
GlcN
1720.1
1332.4
1388.4
1479.1
1530.9
1587.0
1614.8
G
BP338LgmCKO+pBBR2LgmC
1558.8
GlcN
GlcN
1720.3
1332.9
1388.9
1479.0
1493.9
1530.8
1587.0
H
BP338LgmABCDKO+pPtacLgmABCD
GlcN
1720.3
GlcN
1558.8
1479.0
1493.9
1549.9
1587.0
1691.8
1748.1
1881.0
I
BP338LgmABCDKO+pPtacLgmABC
GlcN
1720.6
GlcN
GlcN
1559.1
1479.1
1493.7
1587.0
1692.0
1748.2
1881.5
1332.6
Intensity (arb. u.)
0
100
1200
1400
1600
1800
2000
Mass (m/z)


J
18-323 + pMMB67HE
C12OH
C10OH
1502.8
1530.9
1332.6
1348.8
1474.8
1558.8
Intensity (arb. u.)
100
0
1200
1400
1600
1800
2000
Mass (m/z)

Appendix B Sequence alignments

ClustalW2 (Thompson, Higgins, et al. 1994) alignments of Bordetella LgmABCDE that are the basis for the trees in Figure 29. A) LgmA, B) LgmB, C) LgmC, D) LgmD, E) LgmE, F) LgmD and LgmE. Bpe, *B. pertussis* Tohama I; Bbr, *B. bronchiseptica* RB50; Bpa, *B. parapertussis* 12822; Bav, *B. avium* 197N; Bhi, *B. hinzii* ATCC 51730; and Btr, *B. trematum* CCUG 13902. A) ArnC from *E. coli* K-12. B) ArnT from *E. coli* K-12. C) Ftn, Ftn_0544 from *F. novicida* U112. D) LgmEBav, *B. avium* LgmE. E) GtrA from *E. coli* K-12. F) prefix 'D-' indicates LgmD, prefix 'E-' indicates LgmE. '*': Identical amino acid residues; ':': conserved amino acid residues; '.': semi-conserved amino acid residues.

A) LgmA

```
Bav             --MQAEFRSRALPAPLPAPLPTSEAAFEVAAETR-SFYVEAQISCIVPCLNEADNLCVLL 57
Bhi             --MQVEFRSQAFPASLPP----SEAPLEMAGEVR-SFYVEAQVSCIVPCLNEADNLRVLL 53
Btr             --MRYEFRSSSLPAALAG----TPASADIEAETRQAFFVPALVSCVVPCLNEADNLRLLL 54
Bbr             MFMYTEFRSQLLAGAGTS---AAGQSARMAVLAG-DGATGVQVSCVVPGLNEAANLRVLV 56
Bpa             MFMYTEFRSQLLAGAGTS---AAGQSARMAVLAG-DGATGVQVSCVVPGLNEAANLRVLV 56
Bpe             VFMYTEFRSQLLAGAGTS---AAGQSARMAVLAG-DGATGVQVSCVVPGLNEAANLRVLV 56
ArnC            -------------------------MFEIHPVKK--------VSVVIPVYNEQESLPELI 27
                                         :             :* ::*  **  .*  *:

Bav             PALRNRLESLCSAWEIIVIDDGSTDNTAELMANWSGLDG--FRYIQLARNFGKEAAISAG 115
Bhi             PALRSRLDSMCSAWEIIVIDDGSSDATPELMAEWTALDG--FRYIQLARNFGKEAAISAG 111
Btr             PALRNRLEAMCDRWEIIVVDDGSTDDTADLMAQWTLVDG--FRYVQLARNFGKEAALSAG 112
Bbr             PALRACLEQWCASWEIIVVDDGSTDDTAELMAQWSAVEG--IRYVQLSRNFGKEAALTAG 114
Bpa             PALRACLEQWCASWEIIVVDDGSTDDTAELMAQWSAVEG--IRYVQLSRNFGKEAALTAG 114
Bpe             PALRACLEQWCASWEIIVVDDGSTDDTAELMAQWSAVEG--IRYVQLSRNFGKEAALTAG 114
ArnC            RRTTTACESLGKEYEILLIDDGSSDNSAHMLVEASQAENSHIVSILLNRNYGQHSAIMAG 87
                       :     :**:::****:* :..::.: :  :.  :  : * **:*:.:*: **

Bav             LEAADGDVVICLDADMQHPPALIEEMLRRWQAGSEMVYAVRRNRDDEGYFKRLGSDWFYR 175
Bhi             LEAADGDVVICLDADMQHPPALIEEMLRRWRAGAEMVYALRRDRNDESRFKQLGSHWFYR 171
Btr             LEAADGEAVICMDADLQHPPALIEEMLRRWRAGAEMVYAVRRDRQDEGRFKRFGTACFYR 172
Bbr             LEAADGDAVICLDADMQHPPELIGDMLAAWRNGAEMVYAVRRQRDDEPWFKRVGARAFYR 174
Bpa             LEAADGDAVICLDADMQHPPELIGDMLAAWRNGADMVYAVRRQRDDEPWFKRVGARAFYR 174
Bpe             LEAADGDAVICLDADMQHPPELIGDMLAAWRNGADMVYAVRRQRDDEPWFKRVGARAFYR 174
ArnC            FSHVTGDLIITLDADLQNPPEEIPRLVAKADEGYDVVGTVRQNRQD-SWFRKTASKMINR 146
                :. . *: :* :***:*:**  *  ::     * ::* ::*::*:*   *:: .:  : *
```

```
Bav             LLSGSR-VDVPAGAGDFRLMDRRVVNALVALPERTRFMKGLYAWVGFKCEPLPYTPDSRM 234
Bhi             LLSGSR-VDVPAGAGDFRLMDRRVVEALVALPERTRFMKGLYAWVGFKSEALPYTPDSRM 230
Btr             LMSGAR-VQVPPGAGDFRLLDRRVVQALVSLPERTRFMKGLYAWVGFKSEALTYVPQARV 231
Bbr             LLSTARGVEVPPHAGDFRLMDRRVVEALVALPERTRFMKGLYAWVGFKSQAVPYTPQARR 234
Bpa             LLSTARGVEVPPHAGDFRLMDRRVVEALVALPERTRFMKGLYAWVGFKSQAVPYTPQARR 234
Bpe             LLSTARGVEVPPHAGDFRLMDRRVVEALVALPERTRFMKGLYAWVGFKSQAVPYTPQARR 234
ArnC            LIQRTTGKAMGDYGCMLRAYRRHIVDAMLHCHERSTFIPILANIFARRAIEIPVHHAERE 206
                *:. :    :   .  :*   *::*:*::   **: *:  *   .. :.  :.     *

Bav             HGQSHFRPMKLIGLAMDGLTAFTTWPLRVVSMLGLVFALMSMVYGAYLVGAYFLDG--NP 292
Bhi             HGRSHFRPMRLFGLALDGLTAFTTWPLRAVSLLGVLFALLALVYGAYLVGAYLLVG--NA 288
Btr             HGESHFRPWRLIRLGLDGLTAFTTWPLRMVSLLGLALAVLAMIYGGYLVAAYFAQG--NA 289
Bbr             HGASHFSAWKLFRLACDGLTAFTTWPLRLVSLIGVLFALLSLSYGGYLVADYLISG--NA 292
Bpa             HGASHFSAWKLFRLACDGLTAFTTWPLRLVSLIGVLFALLSLSYGGYLVADYLISG--NA 292
Bpe             HGASHFSAWKLFRLACDGLTAFTTWPLRLVSLIGVLFALLSLSYGGYLVADYLISG--NA 292
ArnC            FGESKYSFMRLINLMYDLVTCLTTTPLRMLSLLGSIIAIGGFSIAVLLVILRLTFGPQWA 266
                .* *::   :*: *  * :*.:** *** :*::*  :*: .:  .  **   :  *   .

Bav             VSGWTTIVTALLFFAGINLLSLGVVGEYVARIFDEVKGRPLYITRQRRGRARAARKANKK 352
Bhi             VSGWTTIMTALLFFAGVNLISLGVVGEYVARIFDEVKGRPLYIARQRRGRARAARKARAQ 348
Btr             VSGWTTIVTALLFFAGINLISLGVVGEYVARVFDEVKGRPLYIARQRRGRARGARQGEAR 349
Bbr             VSGWTTIVTALLFFAGINLISLGVVGEYVARIFDEVKGRPLFIARQRRGRAKRAAKARSQ 352
Bpa             VSGWTTIVTALLFFAGINLISLGVVGEYVARIFDEVKGRPLFIARQRRGRAKRAAKARSQ 352
Bpe             VSGWTTIVTALLFFAGINLISLGVVGEYVARIFDEVKGRPLFIARQRRGRAKRAAKARSQ 352
ArnC            AEGVFMLFAVLFTFIGAQFIGMGLLGEYIGRIYTDVRARPRYFVQQVIRPSSKENE---- 322
                ..*   :.:.*: * * ::::.:*::***:.*:: :*:.** ::.:*    :    :
```

B) LgmB

```
Bbr             MTLATRSMQPHAVSPGQARSWPLPAAGWLLLA-VGVWLAFLSWMRPLALPDEGRYAGVAW 59
Bpe             MTLATRSMQPHAVSPGQARSWPLPAAGWLLLA-VGVWLAFLSWMRPLALPDEGRYAGVAW 59
Bpa             MTLATRSMQPHAVSPGQARSWPLPAAGWLLLA-VGVWLAFLSWMRPLALPDEGRYAGVAW 59
Bav             -------MNTAVLSAR-SATVRWPA--WLVLAGVAVWLAFLAGIRPLTLPDEGRYGGVAW 50
Bhi             -------MSSVALPVSGVARHRIPA--WLVLAAVALWLMGLAWSRPLTLPDEGRYAGVAW 51
Btr             -------MKAVALSRAPADAPRLPA--WVVLAGIAVWLACLAWARPLTLPDEGRYAGVAW 51
ArnT            -------MKSVRYLIG-------------LFAFIACYYLLPISTRLLWQPDETRYAEISR 40
                       *..                  ::* :. :       * *  *** **. ::

Bbr             DMLRNGSFAVPLIDGMPYFHKPPLYYWLAELSFRLFGVNEWAARLPSALAAWASAVALYL 119
Bpe             DMLRNGSFAVPLIDGMPYFHKPPLYYWLAELSFRLFGVNEWAARLPSALAAWASAVALYL 119
Bpa             DMLRNGSFAVPLIDGMPYFHKPPLYYWLAELSFRLFGVNEWAVRLPSALAAWASAVALYL 119
Bav             EMLRSHSYLVPLMDGMPYFHKPPLYYWLAQASFAVFGLSEWSARLPSLLIAWMSIAGVYA 110
Bhi             EMLRSHSYLVPLMDGMPYFHKPPLYYWLAQLSFSVFGLSEWAARLPSLLIAWASIAGVYG 111
Btr             EMLRSDSPMVPLMNGMPYFHKPPLYYWLAQMSFAVFGLNEWAARLPSLLIAWASVAALYA 111
ArnT            EMLASGDWIVPHLLGLRYFEKPIAGYWINSIGQWLFGANNFGVRAGVIFATLLTAALVTW 100
                :** . .  ** : *: **.**   **: . .  :** .::..*    : :  : . :

Bbr             FVRR-HRDAASATLCVLVLATLPLFFGGAQYANMDMLVAGMITLCVLAGADTALRVRGGQ 178
Bpe             FVRR-HRDAASATLCVLVLATLPLFFGGAQYANMDMLVAGMITLCVLAGADTALRVRGGQ 178
Bpa             FVRR-HRDAASATLCVLVLATLPLFFGGAQYANMDMLVAGMITLCVLAGADTALRVRGGQ 178
Bav             FSRR-YRGEAFALCAVLVLSTMPFFYGGAQFANMDMSVAGLITLCVLAGADTIMRVSQGQ 169
Bhi             FARR-YRGERFALCAVLVLSSMPFFYGGAQFANMDMPVAGMITLCVLAGADTIMRVAAGL 170
Btr             FARR-YRGERFALTAAAVLSTMPFFYGGAQFANTDMSVAGLIALCVLAGVHTALCAAAGQ 170
ArnT            FTLRLWRDKRLALLATVIYLSLFIVYAIGTYAVLDP----FIAFWLVAGMCSFWLAMQAQ 156
                *  *  *.   *  .. :  :: :.:. . :*  *     :*:: ::**  :   .  .
```

```
Bbr             AWR---AMALATGVCAALAMLAKGLIGLVLPGAILLAWLAWRRDWRGLRALLWPPAILAF 235
Bpe             AWR---AMALATGVCAALAMLAKGLIGLVLPGAILLAWLAWRRDWRGLRALLWPPAILAF 235
Bpa             AWR---AMALATGVCAALAMLAKGLIGLVLPGAILLAWLAWRRDWRGLRALLWPPAILAF 235
Bav             PWR---YMSLATALAAALAVLAKGLIGVVLPAAILFFWLLMRRDWRGFKALIWPPAILLF 226
Bhi             PWR---RMSLATALLAALAVLAKGLIGIVLPGAILFFWLLMRRDFHGFKALVWPPAIALF 227
Btr             PWR---RWALTTAAAAGLAVLAKGLIGLVLPGAIVLGWLLLRRDWRGIKALLWPPAIGMF 227
ArnT            TWKGKSAGFLLLGITCGMGVMTKGFLALAVPVLSVLPWVATQKRWKDLFIYGWL-AVISC 215
                .*:      *  .  ..:.:::**::.:.:*   :: *:  :: ::.:    *  *:

Bbr             AVVAVPWFWLMQVRYPGFFQYFFVHQHFERFAQTGFNNVQPFWFYLPVIAGLALPWSLWA 295
Bpe             AVVAVPWFWLMQVRYPGFFQYFFVHQHFERFAQTGFNNVQPFWFYLPVIAGLALPWSLWA 295
Bpa             AVVAVPWFWLMQVRYPGFFQYFFVHQHFERFAQTGFNNVQPFWFYLPVIAGLALPWSLWA 295
Bav             LLVAVPWFVEMQLRYPSFFHYFFVYQHFERFALSGFNNVQPFWFYPPVLAGLALPWSLWL 286
Bhi             LLVALPWFVDMQLRYPGFFHYFFVYQHFERFALSGFNNVQPFWFYPPVLAGLTLPWSLWM 287
Btr             LLVAVPWFAYLQWRYPGFFHYFFIYQQFQRFTLTGFNNVQPFWFYPPVLIGLTLPWSLWL 287
ArnT            VLTVLPWGLAIAQREPNFWHYFFWVEHIQRFALDDAQHRAPFWYYVPVIIAGSLPWLGLL 275
                 :..:**   :  * *.*::***  ::::**:  . ::  ***:* **: . :***

Bbr             GGLLRKQFWAADADPDGLRRLALVWLAVIVAFFSMPQSKLVGYIMPVLPPLAFLLAEVVM 355
Bpe             GGLLRKQFWAADADPDGLRRLALVWLAVIVAFFSMPQSKLVGYIMPVLPPLAFLLAEVVM 355
Bpa             GGLLRKQFWAADADPDGLRRLALVWLAVIVAFFSMPQSKLVGYIMPVLPPLAFLLAEVVM 355
Bav             GGALRRGFWAAE-DTDGLRRLMLIWLLVILVFFSLPSSKLIGYILPAVPALAFLVAELVM 345
Bhi             GGTLRRGFWGAE-DVDGLRRLMLIWLVVVVGFFSLPSSKLIGYILPAVPALAFLIAELVL 346
Btr             GGALRRSFWRHE-DQDGLRGLMLLWLAVIVGFFSIPSSKLIGYVLPALPPLAFLVADLVL 346
ArnT            PGALYTGWKNRK---HSATVYLLSWTIMPLLFFSVAKGKLPTYILSCFASLAMLMAHYAL 332
                 * *   :   .   ..     * *  : : ***:...**  *::. ...**:*:*. .:

Bbr             GALRDPAVARATRRMARVSALVAVAICVTAVFVASFNARGSSRELALSLRGELRPDDTLV 415
Bpe             GALRDPAVARATRRMARVSALVAVAICVTAVFVASFNARGSSRELALSLRGELRPDDTLV 415
Bpa             GALRDPAVARATRRMARVSALVAVAICVTAVFVASFNARGSSRELALSLRGELRPDDTLV 415
Bav             GAWN-----QGMRRRALLSLGLGAVLCLLGIIIATLNPRGGSGPLGEQVRAEAQTGDTMV 400
Bhi             PAWE-----RGQRRRAQVCLGVAMALCVTGILVATFNPRGGSGPLGEQVRGEAGPYDTMV 401
Btr             PAWE-----QGRRARVWVSAGVAAALCVTGIAVATLKPRGGNGPLARQVITLMQPGDTTV 401
ArnT            LAAKNN--PLALRINGWININIAFGVTGIIATFVVSPWGPMNTPVWQTFESYKVFCAWSIFS 390
                 * .      . *    :   .. .  :  . ::.  . .       .      . .

Bbr             ALHTYPFDLQLYAHAARP--------MWVVDDWSNPEIPKRDNWRRELYDAVQFEPALGE 467
Bpe             ALHTYPFDLQLYAHAARP--------MWVVDDWSNPEIPKRDNWRRELYDAVQFEPALGE 467
Bpa             ALHTYPFDLQLYAHAARP--------MWVVDDWSNPEIPKRDNWRRELYDAVQFEPALGE 467
Bav             ALHHFPFDLGIYTASTEP--------IWVVDDWSNPEIPTRDNWRKELYDAAIFEPEVGR 452
Bhi             ALHHFPFDLGIYTASTEP--------LWVVDDWSNPEIPTRDNWRKELYDAAIFEPEVGK 453
Btr             ALHQFPFDLGVYGNLREP--------VWVVDDWRNPEIPTRDNWRKELYDAAQFEPEVGQ 453
ArnT            LWAFFGWYTLTNVEKTWPFAALCPLGLALLVGFSIPDRVMEGKHPQFFVEMTQESLQPSR 450
                    : :          *        : :: .:  *:   ..:  : : : .  .   ..

Bbr             RLLVSADTFQQRLCQAPEGSRYWVWGTAADEE---AYAPLRG------QAARFADARRS- 517
Bpe             RLLVSADTFQQRLCQAPEGSRYWVWGTAADEE---AYAPLRG------QAARFADARRS- 517
Bpa             RLLVSADTFQQRLCQAPEGSRYWVWGTAADEE---AYAPLRG------QAARFADARRS- 517
Bav             RVLVSNEVFNARLCAAPTGSRYWVWGQPSDND---AYVSIKG------EAPYFSDGRRL- 502
Bhi             RVLVSNAVFNERLCAAPTGSRYWVWGQTSDGD---AYAAIRG------EAPRFVDGRRQ- 503
Btr             GVLVSNEDFNARLCRADTGARFWIWGQPSDQD---AYPALRG------ESALVGDSRRR- 503
ArnT            YILTDSVGVAAGLAWSLQRDDIIMYRQTGELKYGLNYPDAKGRFVSGDEFANWLNQHRQE 510
                 :*..   .   *. :       ::  ..: .    *   :*      : .   : :*
```

```
Bbr             -----LWLVLVDEAFKGRVCDGTPTGG------------- 539
Bpe             -----LWLVLVDEAFKGRVCDGTPTGG------------- 539
Bpa             -----LWLVLVDEAFKGRVCDGTPTGG------------- 539
Bav             -----VWRVVVNDALRERVCAGKPIGGLPQK--------- 528
Bhi             -----VWRIDVNDAVRQRVCGGKPIGGSPQK--------- 529
Btr             -----VWRIEVNDAVRQRVCGQTPTGGSPRKSAPPEQAE- 537
ArnT            GIITLVLSVDRDEDINSLAIPPADAIDRQERLVLIQYRPK 550
                     :  :  :: ..  .       .
```

C) LgmC

```
Bav             ------------------------------------------------------------
Bhi             MTKSQSVVTTEGHHEGYQPAGHVRPGWRQAGEFDAQHGDQPVRCGGGAADKDEPGNLTKQ 60
Btr             MPECQQIVTAERHHEGDQPADHIGPCGRQAGQLYAEHDHQPVGGGRSGTDADEPGNLAKH 60
Bbr             -------MTSERYDEGQQPAHHVGPQRGQAGPFHAGDRDGPMGGGRGASDGNEPGNLSQQ 53
Bpa             -------MTSERYDEGQQPAHHVGPQRGQAGPFHAGDRDGPMGGGRGASDGNEPGNLSQQ 53
Bpe             -------MTSERYDEGQQPAHHVGPQRGQAGPFHAGDRDGPMGGGRGASDGNEPGNLSQQ 53
Ftn             ------------------------------------------------------------


Bav             ---MRMTVMGTTMKASAYSRCIAICGDDFGMDASIDRAIFQLLDAGRMSAVSCMSTGASF 57
Bhi             GGHAEVKVMDTTMKASAHGRRIAVCGDDFGMDASIDHAIFQLVDAGRLSAVSCMSTGASF 120
Btr             GGQAGKSVRDAGMDTLRQARRVAICGDDFGMDAGIDHAILRLIQARRLSAASCMSSAPGF 120
Bbr             AAHVEEVSPLRNQAGDVRCRRIAVCGDDFGMNEAIDGALIELAGAGRLSAVSCMPLAPAF 113
Bpa             AAHVEEVSPLRNQAGDVRCRRIAVCGDDFGMNEAIDGALIELAGAGRLSAVSCMPLAPAF 113
Bpe             AAHVEEVSPLRNQAGDVRCRRIAVCGDDFGMNEAIDGALIELAGAGRLSAVSCMPLAPAF 113
Ftn             -----------------MVKKIIICADDFGMSDNINSAIINLLEKKIINATSCMPNMPAF 43
                                  : : :*.*****.  *: *::.*     :.*.***.  ..*

Bav             AAHAQDLKTRA---VDIGLHLNFTEPLSPADG-------GMLPLRALLLRAYTGRLNANW 107
Bhi             ARHAADLKRRA---ADTGLHLNLTQALSPADA-------GMLPLKTLLLRAYAGRLDRTQ 170
Btr             ATRGAELLNSG---ADIGLHLNLTERLGLADT-------PRPALRRLLWRAYSRQLDLQW 170
Bbr             AADAPALARLD---VDLGVHVDFTEAFAGAAP-------AAPGLAALLWRAYAGQLDPDW 163
Bpa             AADAPALARLD---VDLGVHVDFTEAFAGAAP-------AAPGLAALLWRAYAGQLDPDW 163
Bpe             AADAPALARLD---VDLGVHVDFTEAFAGAAP-------AAPGLAALLWRAYAGQLDPDW 163
Ftn             KLGIAQLKKIYNDFSHVGIHLNLTEGNAFTNPKSITRNGKFLSLSKLLVKSKLRAINYDD 103
                      *        . *:*:::*:  . :             *  ** ::    ::

Bav             LRQEIARQLDAFEDRIGHAPHYVDGHQHVHQLPGVRQALLAELRQR-YAGQRPWLRLTPA 166
Bhi             VRQEIDRQLDAFEDQMGQAPHYVDGHQHVHQLPGVRGPLLEALRQR-YPRQRPWLRLTPA 229
Btr             VNQEICRQLDRFEDVLGLAPDYIDGHQHVHQLPGVRQLLLAELQRR-YAGRRPWLRLTSV 229
Bbr             IDARLASQFDAFERAFGRAPDYVDGHQHVHQLPGILPRLRALLKRR-YAGQRIWLRHTAP 222
Bpa             IDARLASQFDAFERAFGRAPDYVDGHQHVHQLPGILPRLRALLKRR-YAGQRIWLRHTAP 222
Bpe             IDARLASQFDAFERAFGRAPDYVDGHQHVHQLPGILPRLRALLKRR-YAGQRIWLRHTAP 222
Ftn             VYNELKAQINNFIEDWGALPDFIDGHQHVHHFPIIRKAVINLYKDFNMYTKQTYIRSTYK 163
                :  .:  *:: *    *  *.::*******::* :   :    :      :: ::* *

Bav             GTLQGMPWMPTLKAQLIAALGGHALAAQVRREAWPSNGRFFGVYGFQGGERVYAAFLHHW 226
Bhi             GAMQGLPLAAVFKAHAIAGLGGHALAAQARQDGWPGNRRFFGAYGFGGGRRAYASLLHHW 289
Btr             AALDGLPWAAHIKAHAIASLGGHALAARARALSWPSNRGFLGVYGFEGGRRGYARMLHHW 289
Bbr             GLQFGLPLAEAAKARLIGALGAGALARAAGQEGWQTNRRMLGVYGFTGGPRRYAGLLHHW 282
Bpa             GLQFGLPLAEAAKARLIGALGAGALARAAGQEGWQTNRRMLGVYGFTGGPRRYAGLLHHW 282
Bpe             GLQFGLPLAEAAKARLIGALGAGALARAAGQEGWQTNRRMLGVYGFTGGPRRYAGLLHHW 282
Ftn             MDKS------DFKSLIIYRSGAKKFYNMLIKNNIKHNSSFAGVYSLESDNQDFRKVILEA 217
                           *:  *   *.  :          *  : *.*.: .. : :  .: .
```

```
Bav             LSNALDGDLIMCHPALPG--PVEHAEQRVAELAVLSSAELGEWLVANGLLVQRLSLRRPA 284
Bhi             LFNALDGDLIMCHPALPG--PIEHAAQRVAEFEVLSSPELGEWLVANGLSVARLSQMVPA 347
Btr             LVHARDGDLLMCHPALAG--EIEHAAQRKAEYEVLADPDLGGWLAVNGLSVQRMSVILAS 347
Bbr             LMNARDGDLLMCHPGWPQVHGAAHASQRAAEYEVLAHPELGTWLARNGLRIVRLSQVRGR 342
Bpa             LMNARDGDLLMCHPGWPQVHGAAHASQRAAEYEVLAHPELGTWLARNGLRIVRLSQVRGR 342
Bpe             LMNARDGDLLMCHPGWPQVHGAAHASQRAAEYEVLAHPELGTWLARNGLRIVSLSQVRGR 342
Ftn             YTEIKDGGIIMCHPAADIDIKDPISQSRIKEFAYFNSKQALQDQKDHNIVL--------- 268
                  .  **.::****.         : .*  *   :   :       :.: :

Bav             LASETRQGSPLMR------TLQPLSGP- 305
Bhi             GSREASIGASLKRQGWGADANPPLISR- 374
Btr             AAWAAP-GAVGAR-------WRPATGR- 366
Bbr             QASQESGKVRNVPHSG---SFRRLASRL 367
Bpa             QASQESGKVRNVPHSG---SFRRLASRL 367
Bpe             QASQESGKVRNVPHFG---SFRRLASRL 367
Ftn             ----------------------------
```

D) LgmD

```
Bbr             MSSSSKQTPKTIKERFLHAFFFEIIAIGLSAPVAAWAMDQPLFDMGVLTAVIAWIALLWN 60
Bpa             MSSSSKQTPKTIKERFLHAFFFEIIAIGLSAPVAAWAMDQPLFDMGVLTAVIAWIALLWN 60
Bpe             MSSSSRQTPKTIKERFLHAFFFEIIAIGLSAPVAAWAMDQPLFDMGVLTAVIAWIALLWN 60
Bhi             MT----QAKKTLKERFFHAFLFEILAIGLCAPVAAWAMGKSLFEMGVLTAVIAWIALLWN 56
Btr             -MEVTLITLSPLKRRIVYVSLFELFAILLSTLILMALSDGSAQNSLPVAVIVSATAVLWN 59
LgmEBav         -------------MPALFRQIALFILVGCAAAATHWLAAVLCVEFGGMAPAWANVVG-WL 46
                             ..  :  :: :  .:           :   ::   :  .  *

Bbr             MVYNAGFERLERR-FGVVRTMPVRVAHAVGFELGLVLIIVPLAAWWLAISFWEAFMLDIG 119
Bpa             MVYNAGFERLERR-FGVVRTMPVRVAHAVGFELGLVLIIVLLAAWWLAISFWEAFMLDIG 119
Bpe             MVYNAGFERLERR-FGVVRTMPVRVAHAVGFELGLVLIIVPLAAWWLAISFWEAFMLDIG 119
Bhi             MVYNAGFDRLENH-MGWTRTLRLRVVHALGFETGLILIVIPLAAWWLDISLWQAFVLDIA 115
Btr             YLYNLGFEAWERRNHVMQRTLRVRCIHAVGFEGGLLLFCLPVYMLWYGVGPLVALGMELT 119
LgmEBav         LAFAVSFSGHYRLTFRHLALSWIVAARRFFLVSAAGFAVNELAFVWLLHTTRLPYELLLG 106
                  :  .*.   .          :   : . :  .  :    :   *       .  : :

Bbr             LLMFYLPYAFFYNLAYDKLR---ARWWG-RIEPAGA 151
Bpa             LLMFYLPYAFFYNLAYDKLR---ARWWG-RIEPAGA 151
Bpe             LLMFYLPYAFFYNLAYDKLR---ARWWG-RIEPAGA 151
Bhi             LVLFYLPYAFFYNLGYDKARGPVLRWLARRARYAGA 151
Btr             LMVFFLFYTFVFTLVFDKIFTLPQHYAKLTPAEQG- 154
LgmEBav         LILIVLACLTFVASRLWAFR-----HKPARATRH-- 135
                *::: *    .
```

E) LgmE

```
Bav             MPALFRQIALFILVGCAAAATHWLAAVLCVEFGGMAPAWANVVGWLLAFAVSFSGHYRLT 60
Bhi             MPALFRQIAWFVFVGCAAAATHWLVAVLCVEFAGLAPAWANVAGWLVAFVVSFSGHYRLT 60
Bpa             MRGLLRQIAWFIAVGCAAAATHWAVAVACVEWAGLPPLGANVVGWLLAFVVSFTGHFRLT 60
Bpe             MRGLLRQIAWFIAVGCAAAATHWAVAVACVEWAGLPPLGANVVGWLLAFVVSFTGHFRLT 60
Bbr             MRGLLRQIAWFIAVGCAAAATHWAVAVACVEWAGLPPLGANVVGWLLAFVVSFTGHFRLT 60
GtrA            ---MLKLFAKYTSIGVLNTLIHWVVFGVCIYVAHTNQALANFAGFVVAVSFSFFANAKFT 57
                   ::: :* :  :*   :  ** .   *:  .      **..*:::*. .** .: ::*
```

```
Bav             FRHLALSWIVAARRFFLVSAAGFAVNELAFVWLLHTTRLPYELLLGLILIVLACLTFVAS 120
Bhi             FRHLTLSWIIAARRFFLVSAAGFALNEAAYVWLLHATRLPYDLLLLALILVGLAFLTFVAS 120
Bpa             FRHLAASWTIAARRFFLVSALGFAINELSYAWLLHATSLPYDVLLALVLIGLAFLTFVAS 120
Bpe             FRHLAASWTIAARRFFLVSALGFAINELSYAWLLHATSLPYDVLLALVLIGLAFLTFVAS 120
Bbr             FRHLAASWTIAARRFFLVSALGFAINELSYAWLLHATSLPYDVLLALVLIGLAFLTFVAS 120
GtrA            FKASTT----TMRYMLYVGFMGTLS--ATVGWAADRCALPPMITLVTFSAISLVCGFVYS 111
                *:  :     : * :: *.  *      :  *  .   **  : *  .        ** *

Bav             RLWAFRHKPARATRH 135
Bhi             RLWAFRHKPAAGPHR 135
Bpa             RLWAFRHRHAP---- 131
Bpe             RLWAFRHRHAP---- 131
Bbr             RLWAFRHRHAP---- 131
GtrA            KFIVFRDAK------ 120
                :: .**.
```

F) LgmD and LgmE

```
E-Bpa           -------------MRGLLRQIAWFIAVGCAAAATHWAVAVACVEWAGLPPLGANVVGWLL 47
E-Bpe           -------------MRGLLRQIAWFIAVGCAAAATHWAVAVACVEWAGLPPLGANVVGWLL 47
E-Bbr           -------------MRGLLRQIAWFIAVGCAAAATHWAVAVACVEWAGLPPLGANVVGWLL 47
E-Bav           -------------MPALFRQIALFILVGCAAAATHWLAAVLCVEFGGMAPAWANVVGWLL 47
E-Bhi           -------------MPALFRQIAWFVFVGCAAAATHWLVAVLCVEFAGLAPAWANVAGWLV 47
D-Bbr           MSSSSKQTPKTIKERFLHAFFFEIIAIGLSAPVAAWAMDQPLFDMGVLTAVIAWIA---L 57
D-Bpa           MSSSSKQTPKTIKERFLHAFFFEIIAIGLSAPVAAWAMDQPLFDMGVLTAVIAWIA---L 57
D-Bpe           MSSSSRQTPKTIKERFLHAFFFEIIAIGLSAPVAAWAMDQPLFDMGVLTAVIAWIA---L 57
D-Bhi           MT----QAKKTLKERFFHAFLFEILAIGLCAPVAAWAMGKSLFEMGVLTAVIAWIA---L 53
D-Btr           -MEVTLITLSPLKRRIVYVSLFELFAILLSTLILMALSDGSAQNSLPVAVIVSATA---V 56
                               .   :  :. :  .:           :   :.   :  .   :

E-Bpa           AFVVSFTGHFRLTFRH-LAASWTIAARRFFLVSALG---FAINELSYAWLLHATSLPYDV 103
E-Bpe           AFVVSFTGHFRLTFRH-LAASWTIAARRFFLVSALG---FAINELSYAWLLHATSLPYDV 103
E-Bbr           AFVVSFTGHFRLTFRH-LAASWTIAARRFFLVSALG---FAINELSYAWLLHATSLPYDV 103
E-Bav           AFAVSFSGHYRLTFRH-LALSWIVAARRFFLVSAAG---FAVNELAFVWLLHTTRLPYEL 103
E-Bhi           AFVVSFSGHYRLTFRH-LTLSWIIAARRFFLVSAAG---FALNEAAYVWLLHATRLPYDL 103
D-Bbr           LWNMVYNAGFERLERR-FGVVRTMPVRVAHAVGFELGLVLIIVPLAAWWLAISFWEAFML 116
D-Bpa           LWNMVYNAGFERLERR-FGVVRTMPVRVAHAVGFELGLVLIIVLLAAWWLAISFWEAFML 116
D-Bpe           LWNMVYNAGFERLERR-FGVVRTMPVRVAHAVGFELGLVLIIVPLAAWWLAISFWEAFML 116
D-Bhi           LWNMVYNAGFDRLENH-MGWTRTLRLRVVHALGFETGLILIVIPLAAWWLDISLWQAFVL 112
D-Btr           LWNYLYNLGFEAWERRNHVMQRTLRVRCIHAVGFEGGLLLFCLPVYMLWYGVGPLVALGM 116
                 :   :.  :    .:       :  *  . :.      :        *       .  :

E-Bpa           LLALVLIGLAFLTFVAS----------RLWAFRHRHAP---- 131
E-Bpe           LLALVLIGLAFLTFVAS----------RLWAFRHRHAP---- 131
E-Bbr           LLALVLIGLAFLTFVAS----------RLWAFRHRHAP---- 131
E-Bav           LLGLILIVLACLTFVAS----------RLWAFRHKPARATRH 135
E-Bhi           LLALILVGLAFLTFVAS----------RLWAFRHKPAAGPHR 135
D-Bbr           DIGLLMFYLPYAFFYNLAYDKLR---ARWWG-RIEPAGA--- 151
D-Bpa           DIGLLMFYLPYAFFYNLAYDKLR---ARWWG-RIEPAGA--- 151
D-Bpe           DIGLLMFYLPYAFFYNLAYDKLR---ARWWG-RIEPAGA--- 151
D-Bhi           DIALVLFYLPYAFFYNLGYDKARGPVLRWLARRARYAGA--- 151
D-Btr           ELTLMVFFLFYTFVFTLVFDKIFTLPQHYAKLTPAEQG---- 154
                 : *::. *    .            :
```