



The University of
Nottingham

***Helicobacter pylori* biomimics for gastric-targeted drug delivery**

Naim Hage

Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy

July 2016

“Tout ce qui est dans la limite du possible doit être et sera accompli”

- Jules Verne

Abstract

Drugs that are preferentially absorbed through the stomach or the small intestine have a narrow time window for absorption since passage through this region of the gastrointestinal tract is rapid. A drug delivery system that can adhere to the gastric epithelium will substantially slow down drug transit and help overcome this problem. To achieve this, this study proposes the novel use of a glycan-binding adhesion protein from *Helicobacter pylori*, BabA, to create targeted drug delivery vectors that can mimic the attachment of this bacterium to the gastric epithelium.

In this work, a recombinant form of BabA was expressed in the periplasmic space of *Escherichia coli*; it was found that after the incorporation of a C-terminal hexa-lysine tag, the expression and purification of this protein was significantly improved to amounts that enabled its subsequent characterisation and application. Recombinant BabA retained the highly selective glycan-binding properties of *H. pylori* and next, its crystal structure was solved in the absence and presence of Lewis^b – a glycan well studied for its role in serving as a receptor for BabA. The structural models revealed that Lewis^b binding occurred through a network of hydrogen bonds within a single, shallow binding pocket at the tip of a β -unit in BabA. Binding studies then confirmed that this site was also responsible for the recognition of other glycan receptors. Using this insight, recombinant BabA was conjugated to model drug delivery vectors via a linkage that favoured exposure of its glycan-binding β -unit; the binding properties of BabA successfully translated to these model BabA-vectors.

The research presented in this thesis lays a strong foundation for future work to assess the *in vitro* and *in vivo* efficacy of biomimetic BabA drug carriers.

List of papers published

N. Hage, J. G. Renshaw, G. S. Winkler, P. Gellert, S. Stolnik, F. H. Falcone, Improved expression and purification of the *Helicobacter pylori* adhesin BabA through the incorporation of a hexa-lysine tag, *Protein Expression and Purification* **106**, 25-30 (2015).

N. Hage, T. Howard, C. Phillips, C. Brassington, R. Overman, J. Debreczeni, P. Gellert, S. Stolnik, G. S. Winkler, F. H. Falcone, Structural basis of Lewis^b antigen binding by the *Helicobacter pylori* adhesin BabA, *Science Advances* **1**, e1500315 (2015).

Declaration

I confirm that the work presented in this thesis is entirely my own, except in the case of collaborative work, which is clearly identified as such in the text. Excerpts of this thesis are found within the aforementioned publications; these passages have been personally authored and are a direct output of the experimental work of this study. Referrals to the results of other researchers are indicated with numerical references throughout the text and a list of these citations is found between pages 187 and 205.

A handwritten signature in black ink, appearing to read 'Naim Hage', is positioned to the left of a vertical line.

Naim Hage

Acknowledgements

I would firstly like to thank my supervisors – Franco Falcone, Snow Stolnik, Sebastiaan Winkler and Paul Gellert – for their abundant support and wise counsel over the past three and a half years. I am especially grateful to Franco, the principal investigator of this project, for his unfailing guidance and direct encouragement to advance this study according to the evolving directions presented by the science.

I would like to thank Ross Overman for his scientific insights and pivotal role in facilitating a new collaboration between the University of Nottingham and the Discovery Sciences division of AstraZeneca. I also extend my sincere gratitude to colleagues at AstraZeneca who made experimental contributions to this work – Eileen McCall, Mark Abbott, Jon Renshaw, Claire Brassington, Tina Howard, Judit Debreczeni and Melanie Snow. Additionally, I am thankful to Chris Phillips, Rick Davies, Gareth Davies and Geoff Holdgate for their useful discussions and roles in granting this project access to a wide array of equipment and resources.

Last but not least, I would like to thank my entire family, particularly my parents, and Amanda for their support during my time as a student.

This research has been made possible through funding from AstraZeneca and the Engineering and Physical Sciences Research Council (Grant EP/I01375X/1) as part of the Centre for Doctoral Training in Targeted Therapeutics, School of Pharmacy.

List of abbreviations

Abbreviation	Expanded name
BabA	Blood group antigen-binding adhesin
BLAST	Basic local alignment search tool
BLI	Bio-layer interferometry
BSA	Bovine serum albumin
CFG	Consortium for Functional Glycomics
CHO	Chinese hamster ovary
DLS	Dynamic light scattering
EBNA-1	Epstein Barr virus nuclear antigen 1
EBV	Epstein Barr virus
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Fuc	Fucose
FucT-I	Fucosyltransferase 1
FucT-II	Fucosyltransferase 2
FucT-III	Fucosyltransferase 3
Gal	Galactose
GalNAc	<i>N</i> -acetylgalactosamine
Glc	Glucose
GlcNAc	<i>N</i> -acetylglucosamine
GST	Glutathione S-transferase
HBS-ET	HEPES buffered saline supplemented with EDTA and Tween-20
HEK	Human embryonic kidney
Hop	<i>Helicobacter</i> outer membrane porin
HRP	Horseradish peroxidase
HSA	Human serum albumin
IMAC	Immobilised metal ion affinity chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry

Le	Lewis
LB	Lysogeny broth
LB _{amp}	Lysogeny broth supplemented with ampicillin
LB _{ATG}	Lysogeny broth supplemented with ampicillin, tetracycline and glucose
NCBI	National Centre for Biotechnology Information
Neu5Ac	<i>N</i> -acetylneuraminic acid
NHS	<i>N</i> -hydroxysuccinimide
NRSB	Non-reducing sample buffer
OD ₆₀₀	Optical density at 600 nm
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline supplemented with Tween-20
PCR	Polymerase chain reaction
PCTP	Sodium propionate, sodium cacodylate trihydrate, bis-tris propane
PDB	Protein Data Bank
PDI	Polydispersity index
PEG	Polyethylene glycol
RMSD	Root-mean-square deviation
RU	Response units
SabA	Sialic acid-binding adhesin
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEM	Standard error of the mean
SeMet	Selenomethionine
SOC	Super optimal broth with catabolite repression
SPR	Surface plasmon resonance
TBS	Tris buffered saline
TBS-T	Tris buffered saline supplemented with Tween-20
TCA	Trichloroacetic acid

List of tables and figures

Table 1.1: Examples of mucoadhesive gastroretentive formulations assessed in human trials	8
Table 1.2: Adhesins known to mediate <i>H. pylori</i> attachment to the gastric mucosa	17
Table 3.1: Expression constructs engineered in this study	55
Table 3.2: Primers for <i>babA</i> ₅₂₇ amplification	56
Table 3.3: Primers for colony PCR and confirmatory sequencing	58
Table 3.4: Expression constructs and corresponding hosts employed in this study	59
Table 3.5: Summary of yield from BabA ₅₂₇ and BabA _{527K} purification	89
Table 4.1: X-ray diffraction data collection and refinement statistics	108
Table 4.2: Primers for BabA _{527K} alanine point substitutions	110
Table 4.3: Binding affinity of BabA _{527K} to various ABO/Le blood group antigens	136
Table 5.1: Primers for AviTag™ peptide sequence insertion into BabA _{527K} expression constructs	153
Table 5.2: Microparticles used in this study	155
Appendix Table 1: Size calibration of the gel filtration column used in this study	206
Appendix Table 2: Neoglycoconjugates used in ELISA, SPR and BLI assays	206
Appendix Table 3: ABO/Le blood group antigens used in crystallographic determination and ITC assays	207
Appendix Table 4: Effect of type 1 ABO blood group antigen epitope presentation on BabA _{527K} binding	208
Appendix Table 5: Comparison of BabA _{527K} binding to type 1 and 2 ABO/Le blood group antigens in a glycan array	210
Appendix Table 6: Thermodynamic parameters of BabA _{527K} binding to various ABO/Le blood group antigens	212
Appendix Table 7: Thermodynamic parameters of BabA _{527K} -D233A/S244A binding to various ABO/Le blood group antigens	215

Figure 1.1: Schematic representation of the gastrointestinal tract	2
Figure 1.2: Schematic illustration of the most studied gastroretentive strategies	4
Figure 1.3: Schematic representation of the human gastric surface mucus layer.....	9
Figure 1.4: Examples of lectins used as gastrointestinal targeting moieties ..	13
Figure 1.5: Schematic representation of <i>H. pylori</i> gastric localisation	15
Figure 1.6: The dominant ABO/Le blood group antigens in the gastric mucosa and their relationship to blood group phenotype	20
Figure 1.7: Mucin composition of the gastric mucus layer.....	22
Figure 1.8: Sequence variability of BabA	24
Figure 1.9: Schematic representation of BabA-mediated drug delivery vectors	26
Figure 1.10: Drug release mechanisms from polymeric nanoparticles.....	27
Figure 2.1: The principle of SPR	33
Figure 2.2: A typical SPR sensogram.....	33
Figure 2.3: Schematic diagram of an ITC instrument	35
Figure 2.4: Representative output from a typical ITC experiment	36
Figure 2.5: Schematic representation of the sitting drop vapour diffusion method used in protein crystallisation	39
Figure 2.6: Schematic representation of the data collection step in X-ray crystallography	40
Figure 2.7: Far-UV circular dichroism spectra associated with the various protein secondary structures	43
Figure 2.8: A typical thermal-induced protein unfolding transition.....	45
Figure 2.9: A typical light scattering pattern	46
Figure 2.10: Schematic representation of typical particle intensity fluctuations	47
Figure 2.11: Schematic representation of biosensor tips used in BLI	49
Figure 2.12: Detection of molecule association to biosensor tips through optical interferometry	50
Figure 3.1: The predicted domain structure of BabA.....	53
Figure 3.2: Sequence alignment of BabA J99 and SabA 26695	71
Figure 3.3: Homology modelling of the C-terminal conserved domain of BabA	72
Figure 3.4: Representative amplification of the <i>babA</i> ₅₂₇ gene fragment.....	73

Figure 3.5: Representative verification of <i>babA</i> ₅₂₇ insertion into an expression vector.....	74
Figure 3.6: Representative DNA sequencing of a BabA expression construct	74
Figure 3.7: Schematic diagram of the pET22b_ <i>babA</i> ₅₂₇ expression cassette	76
Figure 3.8: Expression of recombinant BabA in <i>E. coli</i> BL21 (DE3).....	77
Figure 3.9: Schematic diagram of the pET22b(Δ pelB)_ <i>babA</i> ₅₂₇ expression cassette	78
Figure 3.10: Expression of recombinant BabA in <i>E. coli</i> Origami B (DE3).....	79
Figure 3.11: Schematic diagram of the pOPE101_ <i>babA</i> ₅₂₇ expression cassette	80
Figure 3.12: Expression of recombinant BabA in <i>E. coli</i> XL10 Gold	80
Figure 3.13: Comparison of BabA ₅₂₇ immunodetection methods	81
Figure 3.14: Schematic diagram of the pOPE101_ <i>babA</i> _{527K} expression cassette	82
Figure 3.15: Expression of recombinant BabA in <i>E. coli</i> XL10 Gold after the incorporation of a C-terminal hexa-lysine tag.....	82
Figure 3.16: Comparison of BabA ₅₂₇ and BabA _{527K} periplasmic expression in <i>E. coli</i> XL10 Gold.....	83
Figure 3.17: Schematic diagram of the AZ1_ <i>babA</i> _{527K} expression cassette ..	84
Figure 3.18: Expression of recombinant BabA in <i>S. frugiperda</i> Sf21	84
Figure 3.19: Schematic diagram of the AZ2_ <i>babA</i> _{527K} expression cassette ..	85
Figure 3.20: Expression of recombinant BabA in HEK 293-6E and CHO G22	86
Figure 3.21: IMAC purification of BabA ₅₂₇ and BabA _{527K}	87
Figure 3.22: SEC purification of BabA ₅₂₇ and BabA _{527K}	88
Figure 3.23: Analysis of BabA ₅₂₇ purity.....	90
Figure 3.24: BabA ₅₂₇ amino acid sequence showing peptides matched in additional lower molecular weight band	90
Figure 3.25: Analysis of BabA _{527K} purity	91
Figure 4.1: The biosynthesis pathways of the dominant ABO/Le blood group antigens in healthy gastric mucosa	99
Figure 4.2: Binding of BabA _{527K} to HSA-Le ^{b/y} in an ELISA.....	113
Figure 4.3: Immobilisation of HSA-Le ^b to a CM5 sensor chip via amine coupling	114
Figure 4.4: SPR analysis of BabA _{527K} binding to HSA-Le ^{b/y}	115
Figure 4.5: Detection of BabA _{527K} binding to a glycan array	118

Figure 4.6: Crystallisation of apo-BabA _{527K}	120
Figure 4.7: Crystallisation of SeMet apo-BabA _{527K}	121
Figure 4.8: Representative electron density map of the apo-BabA _{527K} model	122
Figure 4.9: Ribbon representation of the crystal structure of BabA _{527K}	123
Figure 4.10: Comparison of BabA _{527K} with the extracellular domain of SabA	124
Figure 4.11: Sequence alignment of BabA with other known <i>H. pylori</i> adhesins	125
Figure 4.12: Chemical structure of the hexasaccharide form of Le ^b	126
Figure 4.13: Crystallisation of BabA _{527K} :Le ^b	127
Figure 4.14: Superimposition of BabA _{527K} from apo- and co-crystal structures	127
Figure 4.15: The Le ^b binding site of BabA _{527K}	128
Figure 4.16: Hydrogen bonding between BabA _{527K} and Le ^b residues	129
Figure 4.17: Conservation of the Le ^b binding site.....	130
Figure 4.18: Binding affinity between wild-type BabA _{527K} and Le ^b	132
Figure 4.19: Binding affinity between the BabA _{527K} -N206A variant and Le ^b . 133	
Figure 4.20: Binding affinity between the BabA _{527K} -D233A/S244A variant and Le ^b	134
Figure 4.21: Type 1 and type 2 fucosylated ABO/Le blood group antigen molecular models	137
Figure 4.22: Secondary structure and thermal stability of wild-type and variant BabA _{527K} proteins.....	138
Figure 5.1: Schematic representation of the gastric lumen and mucus layer	148
Figure 5.2: Secondary structure and thermal stability of BabA _{527K} at neutral and acidic pH.....	157
Figure 5.3: Binding affinity between BabA _{527K} and Le ^b at neutral and acidic pH	158
Figure 5.4: Crystal structures of BabA _{527K} and β-lactoglobulin showing potential pepsin cleavage sites	159
Figure 5.5: Peptic digestion pattern of BabA _{527K} and β-lactoglobulin	160
Figure 5.6: BabA _{527K} amino acid sequence showing peptides matched in the largest peptic digestion fragment	161
Figure 5.7: c-Myc immunodetection of the BabA _{527K} peptic digestion pattern	162
Figure 5.8: Peptic digestion pattern of native BabA	163

Figure 5.9: Location of lysine residues on the surface of BabA _{527K}	164
Figure 5.10: Binding of biotinylated recombinant BabA proteins to HSA-Le ^{b/y} in an ELISA	165
Figure 5.11: Optimisation of protein:microparticle conjugation.....	166
Figure 5.12: Comparison of the hydrodynamic diameters of microparticles with various surface modifications, as determined by DLS	167
Figure 5.13: Immobilisation of HSA-Le ^b to an amine reactive biosensor tip via amine coupling	168
Figure 5.14: Binding of BabA _{527K} and its D233A/S244A-variant form to HSA-Le ^{b/y} in a BLI assay.....	169
Figure 5.15: Binding of various microparticles to HSA-Le ^{b/y} in a BLI assay .	170
Figure 6.1: Schematic representation of the glycan-binding BabA-particles created in this study	186
Appendix Figure 1: Amino acid sequence of the visible residues in the apo-BabA _{527K} atomic model.....	217
Appendix Figure 2: Amino acid sequence of the visible residues in the BabA _{527K} :Le ^b atomic model.....	217

Table of contents

Abstract	ii
List of papers published	iii
Declaration	iv
Acknowledgements	v
List of abbreviations	vi
List of tables and figures	viii
Chapter 1: Introduction	1
1.1 Summary	1
1.2 Gastric-targeted drug delivery	2
1.2.1 The narrow absorption window	2
1.2.2 Gastric retention	4
1.2.2.1 Floating systems	5
1.2.2.2 Expandable systems	5
1.2.2.3 Bioadhesive systems	5
1.2.3 Bioadhesive gastroretentive systems	6
1.2.3.1 Mucoadhesion	6
1.2.3.2 Cytoadhesion	11
1.3 Gastric targeting in nature – <i>Helicobacter pylori</i>	15
1.3.1 <i>H. pylori</i> infection	15
1.3.2 <i>H. pylori</i> adhesins	16
1.3.3 BabA-mediated <i>H. pylori</i> attachment	19
1.3.4 Variability in the BabA glycan-binding profile	22
1.4 Study Aim and Objectives	25
Chapter 2: Overview of experimental techniques	28
2.1 Recombinant expression	29
2.1.1 Expression vectors used in this study	29
2.1.2 Host organisms used in this study	30
2.2 Surface plasmon resonance	32
2.3 Isothermal titration calorimetry	35
2.4 X-ray crystallography	38
2.5 Circular dichroism spectroscopy	42
2.6 Differential scanning fluorimetry	44
2.7 Dynamic light scattering	46
2.8 Bio-layer interferometry	49

Chapter 3: Developing a recombinant BabA expression and purification method	52
3.1 Introduction	52
3.2 Experimental Procedures	55
3.2.1 Generation of BabA expression constructs	55
3.2.1.1 <i>babA</i> gene fragment design – sequence analysis	55
3.2.1.2 Polymerase chain reaction amplification	56
3.2.1.3 Cloning	57
3.2.2 Expression screening	59
3.2.2.1 Periplasmic expression in <i>E. coli</i> BL21 (DE3)	59
3.2.2.2 Cytoplasmic expression in <i>E. coli</i> Origami B (DE3).....	60
3.2.2.3 Periplasmic expression in <i>E. coli</i> XL10 Gold	60
3.2.2.4 Secretory expression in <i>S. frugiperda</i> Sf21	61
3.2.2.5 Secretory expression in HEK 293-6E	62
3.2.2.6 Secretory expression in CHO G22	63
3.2.2.7 Preparation of cellular fractions for analysis	63
3.2.2.8 Analysis of cellular fractions	65
3.2.3 Protein purification.....	67
3.2.3.1 Large-scale expression and periplasmic harvest	67
3.2.3.2 Immobilised metal ion affinity chromatography	67
3.2.3.3 Size-exclusion chromatography	68
3.2.3.4 Recombinant BabA quantification	68
3.2.3.5 Mass spectrometry	69
3.2.3.6 Tryptic peptide mapping	69
3.2.3.7 N-terminal protein sequencing.....	70
3.3 Results	71
3.3.1 Cloning a <i>babA</i> gene fragment into expression vectors	71
3.3.2 Recombinant expression in prokaryotic and eukaryotic hosts.....	76
3.3.2.1 Periplasmic expression in <i>E. coli</i> BL21 (DE3)	76
3.3.2.2 Cytoplasmic expression in <i>E. coli</i> Origami B (DE3).....	78
3.3.2.3 Periplasmic expression in <i>E. coli</i> XL10 Gold	80
3.3.2.4 Secretory expression in <i>S. frugiperda</i> SF21	84
3.3.2.5 Secretory expression in HEK 293-6E and CHO G22	85
3.3.3 Purification of recombinant BabA	87
3.4 Discussion	92
Chapter 4: Characterising the BabA glycan-binding site	98
4.1 Introduction	98
4.2 Experimental Procedures	102
4.2.1 Binding activity of BabA _{527K}	102
4.2.1.1 Enzyme-linked immunosorbent assay	102
4.2.1.2 Surface plasmon resonance	102

4.2.1.3	Mammalian glycan array screening	103
4.2.2	Crystallographic studies	104
4.2.2.1	Crystallisation of apo-BabA _{527K} and data collection	104
4.2.2.2	Generation of selenomethionine substituted BabA _{527K}	105
4.2.2.3	Crystallisation of SeMet apo-BabA _{527K} and data collection	105
4.2.2.4	Co-crystallisation of BabA _{527K} :Le ^b and data collection	106
4.2.2.5	Structure solution	107
4.2.2.6	Sequence analysis	109
4.2.3	Binding affinity of BabA _{527K} and BabA _{527K} variants	110
4.2.3.1	Expression and purification of BabA _{527K} variants	110
4.2.3.2	Isothermal titration calorimetry	111
4.2.3.3	Circular dichroism spectroscopy	112
4.2.3.4	Differential scanning fluorimetry	112
4.3	Results	113
4.3.1	Validating the functional activity of BabA _{527K}	113
4.3.1.1	Enzyme-linked immunosorbent assay	113
4.3.1.2	Surface plasmon resonance	114
4.3.1.3	Mammalian glycan array	116
4.3.2	Structural insight into BabA _{527K} glycan recognition	120
4.3.2.1	Determination of the crystal structure of apo-BabA _{527K}	120
4.3.2.2	Determination of the crystal structure of BabA _{527K} in complex with Le ^b	126
4.3.3	Glycan binding profile of BabA _{527K} and BabA _{527K} variants	132
4.4	Discussion	140
Chapter 5: Assessing BabA acid/pepsin stability and particle-conjugate		
	activity	147
5.1	Introduction	147
5.2	Experimental Procedures	150
5.2.1	Acidic pH conformation and functionality studies	150
5.2.1.1	Circular dichroism spectroscopy	150
5.2.1.2	Differential scanning fluorimetry	150
5.2.1.3	Isothermal titration calorimetry	150
5.2.2	Peptic digestion tests	151
5.2.2.1	Prediction of pepsin cleavage sites	151
5.2.2.2	Peptic digestion	151
5.2.2.3	Analysis of digested protein samples	152
5.2.3	Generation and assessment of BabA-particle conjugates	153
5.2.3.1	Biotinylation of BabA _{527K} and BabA _{527K} -D233A/S244A	153
5.2.3.2	Enzyme-linked immunosorbent assay	154
5.2.3.3	Conjugation of recombinant BabA to microparticles	154
5.2.3.4	Dynamic light scattering	155
5.2.3.5	Bio-layer interferometry	156

5.3 Results	157
5.3.1 Effect of acidic pH on BabA _{527K} conformation and functionality.	157
5.3.2 Resistance of BabA _{527K} to peptic degradation	159
5.3.3 Functionality of BabA-particle conjugates	164
5.4 Discussion	171
Chapter 6: General discussion and conclusions	176
6.1 Research significance and future work	176
6.1.1 Discovery of an improved periplasmic expression strategy.....	177
6.1.2 Revelation of the BabA glycan-binding mechanism	179
6.1.3 Identification of the Hypervariable Crown Loop.....	181
6.1.4 Observation of BabA susceptibility to peptic digestion	182
6.1.5 Development of functional glycan-binding BabA-vectors	183
6.2 Summary.....	186
References	187
Appendix	206

Chapter 1: Introduction

1.1 Summary

Gastroretentive drug delivery systems based on bioadhesion have long been researched, to limited avail, as strategies to increase the bioavailability of drugs that are preferentially absorbed through the stomach and small intestine. Lectin-mediated drug delivery systems may still represent a promising approach – *a priori*, these systems can attach to glycoproteins and glycolipids on the gastric epithelium and can thereby be retained in the stomach for prolonged durations. However, to date, and to the best of knowledge, no lectins have been reported for use in this context. In this thesis, the functional properties of the *Helicobacter pylori* adhesin BabA were studied with a view towards its potential use as such a gastric-targeting moiety. Indeed, while the role of this lectin in mediating *H. pylori* gastric colonisation is relatively well understood, its functional properties have not been fully characterised. Thus, this research contributes to the advancement of both drug delivery and bacteriology disciplines.

1.2 Gastric-targeted drug delivery

1.2.1 The narrow absorption window

The oral route is the most popular and convenient method of drug administration to the human body; it has recently been estimated that over 60% of marketed medicines are formulated as oral products (1). The absorption of drugs by the gastrointestinal tract is governed by several patient-specific and drug-specific variables (2). However, for any systemically acting drug to have its optimal therapeutic effect, it must have a suitable contact time with its preferred site of absorption (3). The gastrointestinal tract is characterised by having different transit times through its various regions (4). As shown in Figure 1.1, the residence time of drugs within the colon is markedly greater than transit time through the stomach and small intestine. Accordingly, if a drug displays preferential absorption through the stomach or small intestine (particularly through its proximal region), it is said to have a narrow time window for optimal absorption (5).

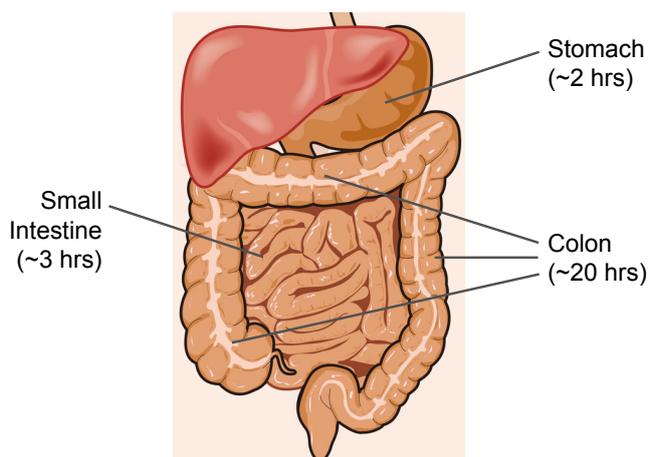


Figure 1.1: Schematic representation of the gastrointestinal tract

The estimated transit time of a drug through the different regions of the gastrointestinal tract in humans is indicated (3, 4). Note that the small intestine is divided into three parts not indicated in the diagram: i) the duodenum and ii) jejunum, which are collectively referred to as the proximal small intestine, and iii) the ileum, which is referred to as the distal small intestine (6). Reprinted from (7) with permission.

Several drugs display preferential absorption through the stomach or small intestine and this is attributable to a variety of factors. For example, some drug molecules better diffuse across the stomach membrane, rather than small or large intestine membranes, because they are unprotonated at acidic, gastric pH. The best studied example of such a case is furosemide, a diuretic used in the treatment of congestive heart failure (8). In other cases, such as the anti-Parkinson's drug, levodopa, specific active transporters that mediate drug absorption exist in the small intestine (9). However, in some instances, the colon simply presents as an unfavourable environment for drug absorption. A clinical study of the absorption properties of the recently marketed anti-platelet drug ticagrelor, which is used to prevent thrombotic events, revealed a marked decrease in bioavailability if its absorption was restricted to the colon (32%), compared to the small intestine and the colon (89%) (10). This was attributed to the greater surface area and larger gaps between cellular tight junctions found in the small intestine compared to the colon – these features facilitate the absorption of drugs with poor permeability properties (11). It has also been shown that the absorption of some drugs is limited to the stomach and/or small intestine because they are susceptible to degradation by colonic bacteria. An example of this is ranitidine, a widely prescribed drug used in the treatment of peptic ulcers (12).

These examples are representative of the most common causes of preferential drug absorption through the stomach or small intestine (13). For drugs with this absorption profile, the rapid transit through the stomach and small intestine can limit their bioavailability, and subsequent therapeutic effect in patients, if sufficient absorption does not occur. In the case of the discussed examples, which are all widely prescribed drugs, this has been successfully circumvented through the administration of a sufficiently high dose. However,

several drugs under development by the pharmaceutical industry are rendered completely ineffective due to this absorption profile (14, 15).

1.2.2 Gastric retention

In order to increase the absorption and subsequent bioavailability of such drugs, gastroretentive drug delivery strategies have been suggested. Unlike a conventional drug formulation, gastroretentive systems aim to retain a drug formulation in the stomach and also control the release of drug from the formulation. Through this, active therapeutic molecules are supplied to the stomach and small intestine at an appropriate, controlled rate. As such, contact time in these regions is maximised and futile drug release from the formulation in distal intestinal regions is prevented (16, 17).

To achieve gastric retention, several different formulation strategies have been employed over the years (18). The three most active approaches are shown below in Figure 1.2. These are i) floating systems, ii) expandable systems, and iii) bioadhesive systems (16).

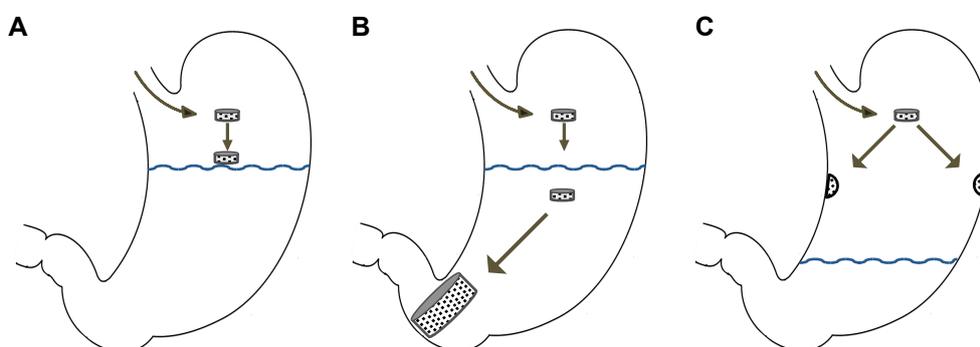


Figure 1.2: Schematic illustration of the most studied gastroretentive strategies Illustrated are (A) floating systems, (B) expandable systems, and (C) bioadhesive systems.

1.2.2.1 Floating systems

In a floating system, the drug formulation simply remains buoyant on the stomach's fluid contents until complete drug release is achieved. This strategy has been shown to significantly improve the delivery of levodopa, which is predominantly absorbed through the small intestine, and has been successfully commercialised as Madopar HBS[®] (19). Recent reviews that outline the mechanism of this strategy can be found in (13, 18).

1.2.2.2 Expandable systems

In an expandable system, the drug formulation swells upon contact with stomach fluid and is consequently too large to pass through the pyloric sphincter until complete drug release is achieved. This method has been shown to improve the delivery of the anti-diabetic drug metformin, which is also predominantly absorbed through the small intestine, and has been successfully commercialised as Glutmetza[®] (20). Recent reviews that outline the mechanism of this strategy can also be found in (13, 18).

1.2.2.3 Bioadhesive systems

An ideal bioadhesive system would mediate the prolonged attachment of a drug formulation to the stomach wall until complete drug release is attained. This strategy, which has not achieved clinical/commercial success so far, is the focus of this study because a successful bioadhesive approach could confer unique advantages over other gastroretentive systems. Floating and expandable systems clearly require the presence of fluid in the stomach to be functional. This represents a major disadvantage because it limits their use to the fed state only; in a fasted stomach, any liquid taken by a patient alongside the drug formulation is rapidly emptied into the intestine (4, 18, 20). On the

contrary, the bioadhesive approach does not depend on the presence of fluid in the stomach, thus, it is hoped that a successful system can be used in either fasted or fed states (21). Additionally, gastric adhesion specifically localises the drug in close proximity to the stomach epithelium, which generates a high concentration gradient between the luminal and serosal sides of the gastric membrane. This facilitates absorption through passive diffusion, which is highly beneficial for drugs preferentially absorbed through the stomach (22).

1.2.3 Bioadhesive gastroretentive systems

The term bioadhesion refers to the adherence of a material to a biological surface. Because the stomach is a mucosal membrane, bioadhesion to this organ actually encompasses i) mucoadhesion, which describes an attachment to the protective mucus gel layer; and ii) cytoadhesion, which describes an attachment to the underlying gastric epithelium.

1.2.3.1 Mucoadhesion

Mucus is a hydrogel found throughout the body that protects the epithelium of mucosal surfaces, including that of the gastrointestinal tract, from chemical, enzymatic, and mechanical damage; a review of its important, protective function can be found in (23). It is composed mainly of water (>95%) and mucins (1-5%), which are a family of 17 high molecular weight glycoproteins that give mucus its structure. In the stomach, there are three major mucins, which can be divided into two subfamilies: secreted mucins (MUC5AC and MUC6) and cell surface mucins (MUC1) – both of these contribute to the protective function of mucus (24-26) and are discussed in further detail in Section 1.3.3.

Research into mucoadhesive strategies began in the 1980s and is still very much active. Mucoadhesive materials have typically been chosen empirically based on their ability to bind mucus *in vitro* and *ex vivo*. The most studied mucoadhesive materials have been polymers and these can be divided into cationic or anionic subsets. Cationic polymers are thought to primarily impart their adhesive properties through electrostatic interactions with mucins that are negatively charged at physiological pH due to terminal sialic acid and sulphate groups. The best-studied cationic polymer is chitosan – a water-soluble, semi-synthetic polymer derived by the deacetylation of chitin, which itself is insoluble in aqueous systems (27). Anionic polymers, on the other hand, are thought to interact with mucins through other physicochemical processes such as hydrogen bonds and van der Waals forces. Typical examples of these polymers are carbomers and polycarbophil – these are synthetic polyacrylic acid derivatives well studied for their mucoadhesive properties (27). However, as reviewed in (28), mucoadhesion by both cationic and anionic polymers is likely to be imparted by several mechanisms, including entanglement within mucin strands, for example.

Initial mucoadhesive strategies were based on a simple idea: a material with mucoadhesive properties could be incorporated into a dosage form, such as a tablet or a granule-filled capsule, to mediate the attachment of the drug formulation to the gastric mucus layer. As surveyed by Waterman *et al.* in 2007, although such dosage forms display excellent mucoadhesive properties *in vitro* and in some animal models, no human trials showed significant gastric retention with this strategy [Table 1.1 (16)]. Cationic and anionic polymers formulated into microspheres, typically between 1 and 10 μm , have also been assessed. These microparticulates, which can in principle be used to encapsulate a drug of interest, were expected to have more intimate contacts

with the gastric mucus layer due to a higher surface area to volume ratio. Again, in animal models, several researchers have shown that *in vitro* mucoadhesion can correlate well with *in vivo* gastric localisation [reviewed in (29)]. However, success has not been reported through this approach either in human studies. Ultimately, this is because the prolonged gastric retention of mucoadhesive materials is governed by the rate of gastric mucus turnover (30, 31). As this is a frequent process, both mucoadhesive formulations and microspheres have generally been regarded as inefficient or suboptimal in mediating prolonged gastric retention.

Table 1.1: Examples of mucoadhesive gastroretentive formulations assessed in human trials

The mucoadhesive materials indicated were all formulated into oral dosage forms such as tablets or granule-filled capsules. In all studies, gastric retention time was compared to non-mucoadhesive formulations and analysed with γ -scintigraphy.

Material	Polymer Type	Gastric Retention	Reference
Chitosan	Cationic	No evidence	Säkkinen <i>et al.</i> (32, 33)
Polycarbophil	Anionic	No evidence	Khosla <i>et al.</i> (34) and Harris <i>et al.</i> (35)
Carbopol [®] 934P	Anionic	No evidence	Jackson <i>et al.</i> (36)
Cholestyramine	Anionic	No evidence	Jackson <i>et al.</i> (36)

The frequent turnover of the gastric mucus layer is one of the hallmarks of its protective function (37). Physiologically, as depicted in Figure 1.3, it is well known that the mucus that lines the human stomach, and the rest of the gastrointestinal tract, is a two-component system consisting of a loosely adherent layer that overlays a firmly adherent layer (38-40). In the stomach, the loosely adherent layer, which is estimated to be replaced every 1 to 4 hours, can be easily removed through physical suction *ex vivo*; it is this layer that is responsible for the trapping and rapid clearance of the aforementioned

mucoadhesive drug delivery systems. The firmly adherent layer acts as a thicker, unstirred layer that cannot be removed without disrupting the surface epithelium. It provides further physiological protection and is cleared less frequently (estimated to be between hours to days), though a well-defined timeframe is yet to be established (30, 38, 41).

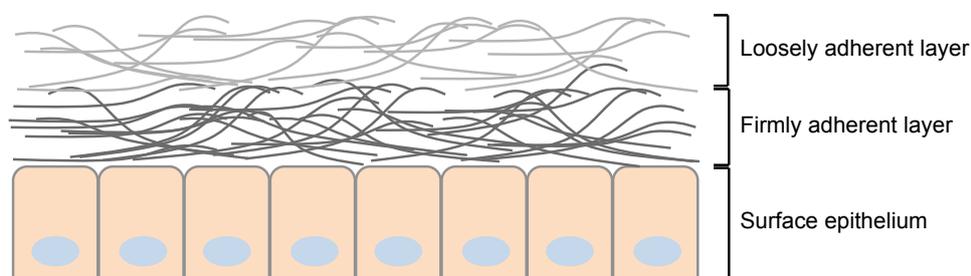


Figure 1.3: Schematic representation of the human gastric surface mucus layer

The loosely and firmly adherent mucus gel layers that overlay the gastric surface epithelium are shown; both layers are comprised of the secreted MUC5AC mucin (42). The thickness of these mucus layers in the human stomach has not yet been determined. In the antrum region of rat stomachs, Atuma *et al.* calculated the loosely adherent layer to be $\sim 120 \mu\text{m}$ and the firmly adherent layer to $\sim 150 \mu\text{m}$ (38). In the body region of mouse stomachs, Ermund *et al.* have calculated the firmly adherent layer to be $\sim 45 \mu\text{m}$ (40).

Given these features of the gastric mucus layer, a mucoadhesive strategy that successfully penetrates past the loosely adherent layer and preferentially accumulates in the firmly adherent layer may be successful in the relative prolongation of gastric retention. Indeed, mucus penetration by bacterial pathogens is observed in nature (43). However, it was the characteristics of two viruses that inspired researchers to develop drug delivery systems based on this strategy. In 2001, Olmsted *et al.* showed that the human papilloma virus ($\sim 55 \text{ nm}$) and Norwalk virus ($\sim 38 \text{ nm}$) can successfully partition through human cervical mucus (44). As such, it was thought that nanoscale drug carriers could provide a means to achieve mucus penetration. This was investigated by Lai *et al.* who found that polystyrene nanoparticles (ranging

from 100 to 500 nm), which are a particularly common model drug delivery vehicle, could not achieve similar penetration due to non-specific interactions with mucins. However, if densely coated with polyethylene glycol (PEG), polystyrene nanoparticles developed the ability to penetrate through the openings between mucin mesh fibres. This characteristic was attributed to the hydrophilic nature of PEG preventing hydrophobic interactions between the polystyrene core and mucins. Low molecular-weight PEG had to be used to attenuate possible entanglement between PEG chains and mucins (45). Indeed, low molecular-weight PEG has since been shown to have a similarly beneficial effect in mucus penetration with other relevant, biodegradable nanovehicles of a non-polystyrene core (46). These aforementioned studies were performed *in vitro* using human cervicovaginal mucus. However, in a recent *in vivo* study by Maisel *et al.*, PEGylated nanoparticles were shown to successfully penetrate through the loosely adherent layer of gastrointestinal mucus and localise within close proximity to the small and large intestine epithelium of mice after oral administration (47). Similar findings on the beneficial effects of PEGylation on gastrointestinal mucus penetration by nanoparticles have also been independently reported by Inchaurrega *et al.* (48).

It remains to be seen whether these so-called “mucus penetrating” nanoparticles are retained in the firmly adherent layer or whether the lack of mucin interaction simply results in transitory association (49). While their ability to penetrate into the firmly adherent mucus layer is a promising lead for drug delivery, because the entire gastrointestinal tract is a continuous mucosal membrane, a significant drawback of these systems is their inherent inability to selectively target the gastric mucosa.

1.2.3.2 Cytoadhesion

A step forward towards the selective gastric targeting of bioadhesive systems lies in the use of a specific cytoadhesive ligand that can enable the attachment of drug carriers to receptors on gastric epithelial surfaces. In principle, through this strategy, drug localisation is not only gastric-specific but the limiting effect of mucus turnover can also be evaded; the rate of gastric epithelial cell turnover is between 2 to 3 days and, as such, will not impact the acute clearance of cytoadhesive drug delivery vectors from the stomach (50).

It is believed that gastrointestinal epithelial cells express a wide array of glycans on their surfaces. However, an in-depth characterisation of the gastrointestinal glycosylation pattern has been a challenge because the glycoproteins and glycolipids expressed on epithelial cells can vary greatly due to individual factors, such as hormonal status, and microenvironmental factors, such as microbial colonisation and inflammation (26). Indeed, the complexity and heterogeneity of glycan structures also complicates their identification through glycoproteomic approaches (51).

One of the most successful approaches used to study gastrointestinal glycosylation patterns has been through lectin-histochemical staining. Though the name “lectin” originally described plant extracts capable of cross-linking red blood cells, the term is now widely used to describe a protein with carbohydrate binding properties; lectins can be obtained from various sources including plants, animals and microorganisms. In addition to having stringent glycan binding specificity, several lectins, particularly plant lectins, display good resistance to acid denaturation and enzymatic degradation (22). Because of these properties, lectins have also been considered to be good

ligands for gastrointestinal targeting – this is often referred to as the “second-generation” bioadhesive approach (52).

Naisbett and Woodley commenced research into lectin-mediated targeting in the early 1990s when they investigated the binding properties of *Lycopersicon esculentum* (tomato) lectin (53-56). They showed that this protein selectively bound to the *N*-acetylglucosamine residues of glycoproteins on rat intestinal epithelia *ex vivo*. Soon after, a number of complementary studies revealed that after conjugation to model microparticulate drug carriers, tomato lectin maintained its affinity for rat intestinal epithelia *ex vivo* (57, 58). In light of these promising results, several other lectins were studied in this capacity given that their glycan binding properties were already being exploited for histochemical staining. For example, *Ulex europaeus* (Gorse) agglutinin I (Figure 1.4) is a lectin that specifically recognises terminal $\alpha(1,2)$ linked fucose residues – these residues are abundantly found on intestinal M cells in mice (59). Foster *et al.* found that linking this protein to microspheres resulted in the significant and specific binding of particle-conjugates to these cells in a mouse gut loop model (60). Targeting effects have similarly been well studied with *Canavalia ensiformis* (Jack bean) agglutinin, also known as Concanavalin A. This lectin binds to mannose and glucose residues (Figure 1.4) and was shown by Russell-Jones *et al.* to mediate mannose-specific microparticle binding to the Caco-2 intestinal cell line (61). There are several other examples of such interactions included in a review by Bies *et al.*, where the use of lectins to target the gastrointestinal tract, and other mucosal membranes, is discussed (62).

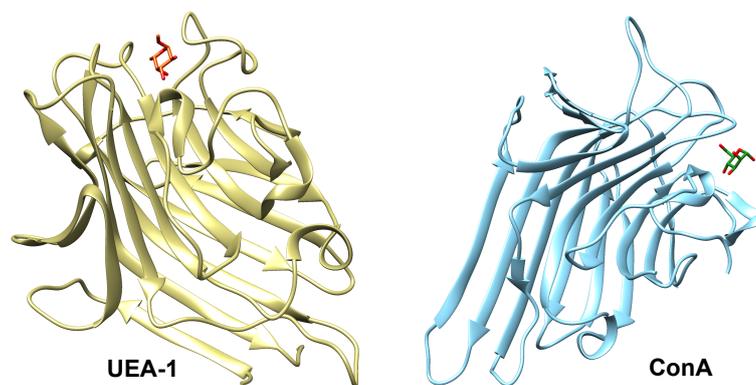


Figure 1.4: Examples of lectins used as gastrointestinal targeting moieties

Shown on the left is the ribbon representation of the crystal structure of *Ulex europaeus* agglutinin I (UEA-1) in complex with a fucose residue (red) [Protein Data Bank (PDB) Accession no: 1JXN]. On the right is the ribbon depiction of the crystal structure of Concanavalin A (ConA) in complex with a mannose residue (green) (PDB Accession no: 1I3H). For both lectins, only a single subunit is shown.

To some extent, the potential clinical applicability of lectin-mediated systems is already evident from the *in vivo* studies reported in the literature that have successfully used these delivery systems to target intestinal cells. For example, the aforementioned study by Foster *et al.* confirmed that UEA-1-conjugates retained targeting efficacy to mice intestinal M cells after intragastric administration (60). Recently, Liu *et al.* showed that wheat germ agglutinin grafted onto lipid nanoparticles not only mediated enhanced attachment to the Caco-2 cell line (63), but the conjugates were also able to improve the bioavailability of bufalin (a model drug molecule) ~2-fold in a rat study through association with the jejunum and ileum (64). Additionally, at least five *in vivo* small animal studies [reviewed in (65)] have reported that the efficacy of oral immunogens delivered through nanocarriers is significantly improved by intestinal M-cell targeting through lectin functionalisation.

Though these studies have reported enhanced targeting efficacy, it must be noted that the ability of the drug carriers employed to diffuse through mucus

was not studied in detail. Importantly however, as demonstrated in work by Montisci *et al.*, lectin-targeting *in vivo* can result in both cytoadhesion and mucoadhesion (66) because while lectins are capable of binding to epithelial tissue, they can also bind prematurely to the overlaying mucus layer – secreted mucins tend to carry the same glycans that are found on the underlying epithelium that they protect (67). This is an important factor that has not yet been studied in the context of mucus penetration by lectin-mediated drug delivery systems.

Although further work is needed to fully evaluate the potential of cytoadhesive systems, the literature suggests that lectin-mediated drug carriers are promising tools for intestinal delivery. However, no studies could be found that have reported using this approach for a gastric-targeted drug delivery system. Furthermore, to date and to the best of knowledge, no lectins have been evaluated in this context for their ability to target glycans on the gastric epithelium. Accordingly, identifying a gastric-targeting lectin can be considered the first step towards the development of a cytoadhesive, gastroretentive system. In search of such a lectin, this study adopted learnings from nature.

1.3 Gastric targeting in nature – *Helicobacter pylori*

1.3.1 *H. pylori* infection

An excellent example of gastric targeting in nature can be found in *H. pylori*. This Gram-negative, microaerophilic bacterium displays a strict tropism for human and primate gastric mucosa, particularly in the less acidic antrum region of the stomach (Figure 1.5) (68, 69). It is exceedingly successful at the population level, with more than one-half of the global population infected (70). Although this always leads to histologic gastritis, the majority (80-90%) of infected individuals do not develop clinical symptoms (71). However, in a subset of individuals, infection can lead to severe peptic ulceration and it is also the strongest known risk factor for gastric cancer (72).

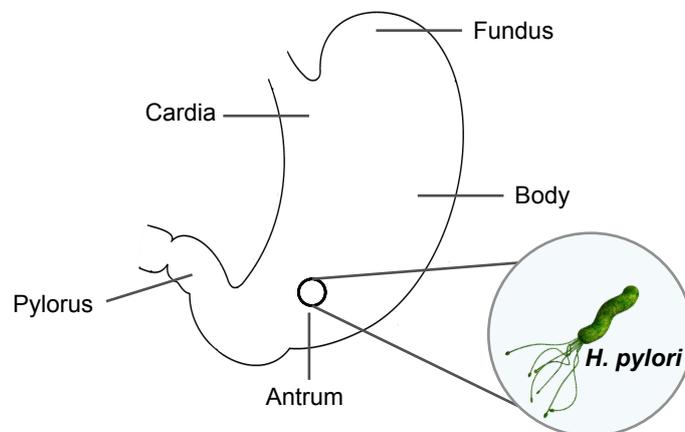


Figure 1.5: Schematic representation of *H. pylori* gastric localisation

H. pylori (inset) is a helicoidal-shaped bacterium, with unipolar flagella, typically found in the lower, less acidic antrum region of the human stomach.

Despite the acidic and proteolytic conditions of the gastric lumen, which creates a hostile ecological environment that kills most pathogens ingested with meals, *H. pylori* manages to establish chronic infection in the stomach. In fact, unless treated, the bacterium can persist in the stomach for decades or even the duration of its host's lifespan (72). Infection is established through unique evolutionary adaptations, which are believed to have co-evolved with

humans over millennia as the existence of the bacterium can be traced back to 58,000 years ago (73). First, while present in the stomach lumen, *H. pylori* withstands acidic conditions through the uptake and hydrolysis of urea in stomach fluid into ammonia and carbon dioxide using its cytosolic enzyme urease. Urease accounts for 10-15% of total protein in *H. pylori* cells and through the generation of ammonia, it enables the bacterium to maintain a buffered cytosol, periplasm and surface layer (74, 75). However, *H. pylori* is not an acidophile and must swim into the gastric mucus layer using its sheathed, unipolar flagella [Figure 1.5 (inset)]; this movement is guided by the detection of chemotactic factors, including urea and bicarbonate ions (76). Through this mechanism, the helicoidal-shaped bacterium penetrates the gastric mucus layer “like a screw into a cork” (77), and it subsequently colonises the firmly adherent mucus layer (Figure 1.3) (78). Here, it is protected from acid and proteases, and critically, the bacterium can establish contact with host cells. A comprehensive review of the infection-establishing mechanisms discussed here can be found in (77).

1.3.2 *H. pylori* adhesins

Host cell attachment is key to *H. pylori* infection. Without this mechanism, attempts to colonise the gastric mucosa will be futile because the bacteria will be cleared into the intestine during gastric mucus turnover. This mechanism is also necessary to enable the bacteria to gain access to vital nutrients from epithelial cells (78). Thus, while also able to remain motile within the gastric mucus layer, *H. pylori* possesses a chemotactic sensing mechanism that enables it to migrate towards the surface epithelium. This mechanism is based on the acidic to near-neutral pH gradient found from the luminal to apical regions of the mucus layer (79). Like other bacteria, to establish contact with

host cells, *H. pylori* employs adhesion proteins (adhesins) that recognise specific cell surface moieties. Even at the cellular level, *H. pylori* continues to display tropism. As Falk *et al.* showed in an *in situ* adherence assay, attachment only occurs to gastric surface mucous cells (also referred to as the gastric surface epithelium) but not to mucous neck, parietal or chief cells, which are all found deeper in the gastric glands (Figure 1.7) (80). To date, the adhesins listed in Table 1.2 have been identified as responsible for mediating *H. pylori* attachment to the gastric surface epithelium. However, specific receptors have not been identified for all adhesins.

Table 1.2: Adhesins known to mediate *H. pylori* attachment to the gastric mucosa

^aLacdiNAc is an abbreviation for *N,N'*-diacetyllactosediamine

Adhesin	Abbreviation	Host Receptor	Reference
Adherence-associated lipoprotein A	AlpA	Unknown	Odenbreit <i>et al.</i> (81)
Adherence-associated lipoprotein B	AlpB	Unknown	Odenbreit <i>et al.</i> (81)
Blood group antigen-binding adhesin	BabA	ABO/Lewis blood group antigens	Ilver <i>et al.</i> (82)
Helicobacter outer membrane porin Z	HopZ	Unknown	Peck <i>et al.</i> (83)
LacdiNAc-binding adhesin ^a	LabA	LacdiNAc moieties	Rossez <i>et al.</i> (84)
Sialic acid-binding adhesin	SabA	Sialylated glycans	Mahdavi <i>et al.</i> (85)

All of the abovementioned adhesins fall into the 21-member *Helicobacter* outer membrane porin (Hop) family. Members of this family have not been fully characterised with regards to pathological function, however, all members share a region of high sequence identity/similarity that is predicted to form a β -barrel unit for outer membrane insertion. The above listed adhesins within this family are characterised by putative N-terminal extracellular domains with their

predicted β -barrel units localised to their C-termini (86, 87). As a number of other Hop proteins also share this domain similarity, it is believed that other yet-to-be characterised adhesins may exist. Interestingly, some outer membrane proteins with apparently unrelated functions, such as the outer inflammatory protein OipA and the heat shock protein Hsp70, have also been implicated in bacterial attachment (88, 89). Furthermore, recent studies have proposed new glycan structures that act as *H. pylori* receptors, though corresponding adhesins have not yet been identified (90, 91). Taken together, this indicates that the full picture of *H. pylori* gastric colonisation is incomplete. It can be argued, however, that colonisation mediated by several different or complex mechanisms is expected. Possessing a wide range of adhesins that recognise various host moieties is certainly beneficial to *H. pylori* in sustaining lifelong colonisation as this enables the bacterium to adapt to changes in its ecological niche (92). This is already well exemplified in its two best-studied adhesins – BabA and SabA. BabA mediates specific attachment to ABO and Lewis (Le) blood group antigens. In addition to erythrocytes, these glycan structures are found abundantly expressed in the healthy gastric mucosa of the majority of the Western population. On other hand, SabA mediates equally specific attachment to sialylated glycans. These structures are rarely expressed in healthy stomachs but appear in response to inflammation, which is a typical result of chronic *H. pylori* infection (93). Thus, with regards to the use of lectins in gastric-targeted drug delivery systems, the well-characterised adhesin BabA presents as a favourable choice. While lacdiNAc moieties that serve as receptors for LabA are also found in healthy stomachs, this adhesin was only recently identified by Rossez *et al.* in 2014 (94), that is, after the commencement of this study. AlpA, AlpB and HopZ were considered less suitable choices because their host receptors are currently unknown.

1.3.3 BabA-mediated *H. pylori* attachment

In the gastric mucosa, surface mucous cells express ABO/Le blood group antigens that contain what is known as a type 1 lacto series core – this refers to their central Gal β 1-3GlcNAc linkage (Figure 1.6, Figure 1.7). Deeper within the gastric glands, mucous neck cells also express ABO/Le blood group antigens (Figure 1.7). However, though related, these glycans are inherently different as they contain a type 2 lacto series core, which denotes a central Gal β 1-4GlcNAc linkage (Figure 1.6) (95, 96). In line with preceding reports of strict tropism for surface mucous cells (80), Borén *et al.* showed in 1993 that *H. pylori* could only bind to gastric tissue that expressed ABO/Le blood group antigens with a type 1 core (97). Soon after, in 1998, Ilver *et al.* showed that this interaction is completely mediated by BabA, which displays no type 2 ABO/Le blood group antigen affinity (82). Through these two landmark studies, the role of BabA in mediating *H. pylori* attachment to gastric tissue was established. As suggested by their nomenclature, the composition of type 1 and type 2 ABO/Le blood group antigens in each individual is determined by blood phenotype. Figure 1.6 shows the relationship between blood phenotype and the aforementioned blood group antigens found in the gastric mucosa. Their full biosynthetic pathways are later described in Section 4.1.

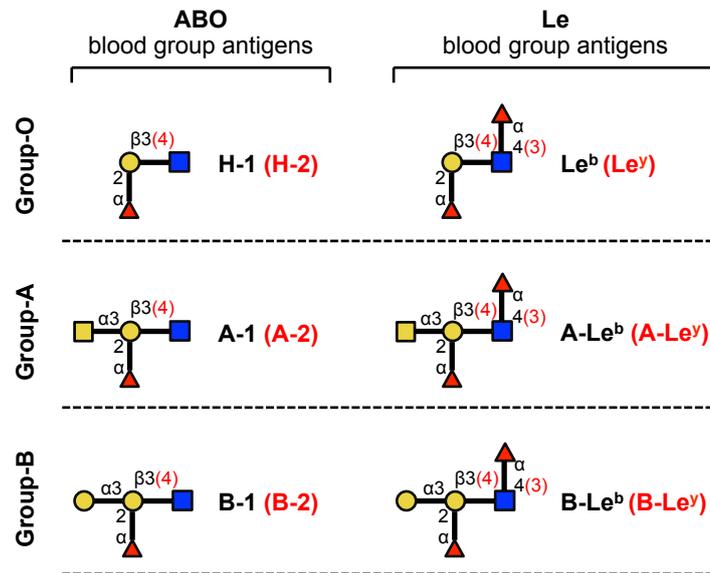


Figure 1.6: The dominant ABO/Le blood group antigens in the gastric mucosa and their relationship to blood group phenotype

Type 1 ABO/Le blood group antigens are characterised by a Gal β 1-3GlcNAc linkage (glycosidic bond and glycan name represented in black) while type 2 ABO/Le blood group antigens display a Gal β 1-4GlcNAc linkage (glycosidic bond and glycan name represented between parentheses and in red). Glycan symbolic representations can be interpreted with the following key: fucose – \blacktriangle , galactose – \bullet , N-acetylgalactosamine – \blacksquare , N-acetylglucosamine – \blacksquare .

As previously mentioned, *H. pylori* can attach to surface mucous cells but can also remain motile in the gastric mucus layer. A study in 1991 by Tzouveleki *et al.* was one of the first to show that *H. pylori* is actually able to attach to gastric mucus (98). It was later revealed that BabA plays a significant role in mediating this attachment, which was a rational finding given the composition of gastric mucus. The gastric mucus layer consists of two secreted mucins – MUC5AC and MUC6; MUC5AC originates from surface mucous cells whereas MUC6 originates from mucous neck cells, which, as previously mentioned, lie deeper within the gastric glands (Figure 1.7). ABO/Le blood group antigens are found as terminal glycan epitopes on both of these O-linked glycoproteins (43, 99). However, in agreement with the aforementioned cellular glycan

expression patterns of these mucus-secreting cells, MUC5AC strictly carries type 1 ABO/Le blood group antigens whereas MUC6 strictly carries the corresponding type 2 glycans. Importantly, the localisation of these mucins in the gastric mucus layer is further associated with cell type such that the surface mucus layer is exclusively comprised of MUC5AC while mucus found within the gastric glands is comprised of MUC6 (96, 100, 101). As shown in Figure 1.7, this results in a “laminated” gastric mucus layer with regard to mucin composition. So, after Tzouvelekis *et al.* showed that *H. pylori* can bind to gastric mucus, independent studies by Linden *et al.* and Van de Bovenkamp *et al.* showed that BabA mediates this attachment by binding to the type 1 ABO/Le blood group antigens on MUC5AC (102, 103). Importantly, this indicates that as with other lectins, there are anticipated benefits and obstacles associated with the use of BabA as a cytoadhesive-targeting moiety.

Another potential obstacle may lie in the protective role of the cell surface associated mucin found in the stomach, MUC1 (Figure 1.7). The extracellular domain of this glycoprotein can be released from the apical cell surface of the gastric epithelium via proteolytic cleavage (104) and/or the dissociation of its non-covalently linked mucin subunits in response mechanical stress (105, 106). This is of relevance because although the glycosylation pattern of the extracellular domain of this mucin has not been as well characterised as MUC5AC and MUC6, a recent study by Linden *et al.* showed that *H. pylori* binds to MUC1, which then acts a protective “decoy” receptor by being shed from the cell surface into mucus gel layer and gastric juice (107). Though the MUC1 binding mechanism of *H. pylori* is not yet fully understood, BabA was noted to contribute to attachment.

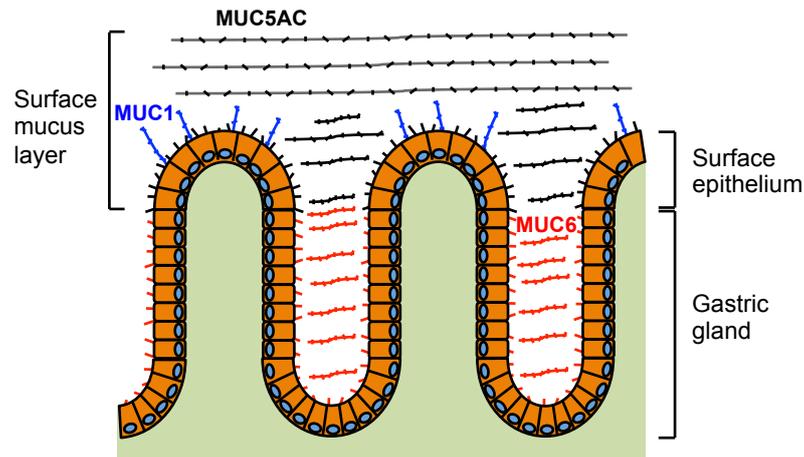


Figure 1.7: Mucin composition of the gastric mucus layer

Type 1 ABO/Le blood group antigens are found attached to both MUC5AC and surface epithelial cells (indicated in black) while type 2 ABO/Le blood group antigens are found attached to MUC6 and epithelial cells in the gastric glands (indicated in red). The glycosylation pattern of MUC1 (indicated in blue) has not been well characterised but Linden *et al.* have reported the presence of Le^b (107). MUC5AC and MUC6, which are the secreted mucins in gastric mucus, form a laminated mucus layer. Note that the surface mucus layer illustrated refers to both firmly and loosely adherent layers, which are both formed predominantly by MUC5AC (42). MUC1 is found attached to the surface epithelium, however, it may also be found attached to mucous neck cells within the gastric gland (not indicated) (108).

1.3.4 Variability in the BabA glycan-binding profile

Since the role of BabA in mediating *H. pylori* attachment to gastric surface epithelial cells and mucin was identified, its selective affinity for type 1 ABO/Le blood group antigens has been validated by several studies using various methodologies. This has included surface plasmon resonance (109), dynamic force spectroscopy (110), atomic force microscopy (111), oblique-incidence reflectivity difference microscopy (112), glycoconjugate arrays (113) and transgenic mice models (114). As such, despite no structural information detailing the molecular mechanism of glycan binding, the role of type 1 ABO/Le blood group antigens acting as BabA receptors has been well demonstrated (Figure 1.6).

However, a number of studies have shown that not all *H. pylori* strains are able to attach to Le^b (this Le blood group antigen is the most studied of all type 1 ABO/Le blood group antigens in *H. pylori* gastric-attachment) (82, 115, 116). *H. pylori* strains are considered to be highly genetically diverse (117), and in accordance with this, not all clinical isolates possess a functional *babA* gene so lack of Le^b attachment is expected in some cases (118-120). However, as shown by Hennig *et al.* in 2004 and again by Odenbreit *et al.* in 2009, not even the detection of the BabA protein on *H. pylori* cells guarantees Le^b binding affinity (121, 122). The study by Hennig *et al.* also showed, using optical density assays, that within functional BabA strains, attachment to Le^b glycoconjugates occurs to different extents (121). Variability in Le^b affinity was studied in closer detail, also in 2004, by Aspholm-Hurtig *et al.* (123). From a panel of 68 Swedish clinical isolates, a 1500-fold range in Le^b binding affinity was observed. Association constants (K_A) ranged from $2 \times 10^8 \text{ M}^{-1}$ to $3 \times 10^{11} \text{ M}^{-1}$, as determined through Scatchard analysis of suspended bacteria binding to radiolabelled Le^b glycoconjugates. These differences in binding affinity were believed to be due to the high sequence variability observed in the N-terminal extracellular domain of BabA. Indeed, as shown in Figure 1.8, a recent study by Nell *et al.* definitively illustrated that amino acid substitutions in this domain were highly common across a number of strains that displayed both variable and lack of Le^b affinity (124). The variability in the expression, functionality and affinity of BabA indicates that it plays a relative, rather than absolute, role in mediating *H. pylori* adherence.

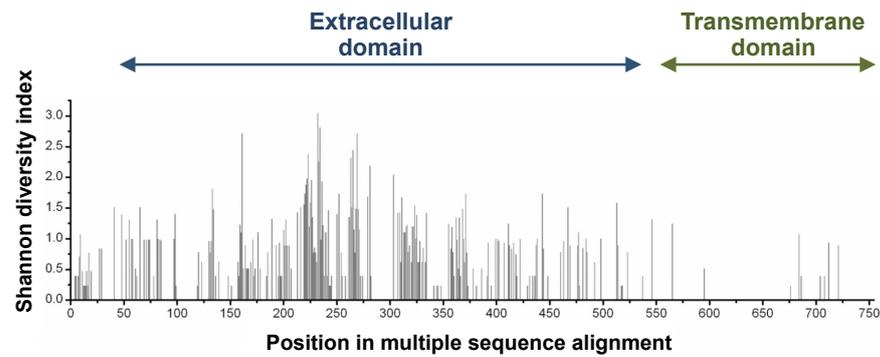


Figure 1.8: Sequence variability of BabA

Nell *et al.* showed that 25 *H. pylori* strains that expressed BabA but bound Le^b with different affinity (21 strains) or could not bind Le^b (4 strains) contained high sequence variability in their putative N-terminal extracellular domains (124). The Shannon Diversity Index used to illustrate variability can range from 0 (same amino acid present at given position across all sequences) to 4.322 (all 20 amino acids equally present at given position across all sequences). Reprinted from (124) with permission.

In further support of *H. pylori* being a genetically diverse species, BabA activity can be governed by an individual's blood type. While most strains with functional BabA proteins can recognise all of the aforementioned type 1 ABO/Le blood group antigens (generalist strains), others can only bind to the type 1 ABO/Le blood group antigens in individuals with the group-O phenotype, that is, Le^b and H-1 antigens (specialist strains) (Figure 1.6). This was reported by Aspholm-Hurtig *et al.* who also found that specialist strains are highly localised to South Amerindian populations that are composed almost entirely of the blood group-O phenotype (123). Though the observed differences in glycan recognition were attributed to sequence variability, no molecular mechanism was identified. Taken together, the literature convincingly informs that not all BabA proteins are created equal, and strain variability must be considered when employing BabA to avoid the creation of inert drug delivery systems.

1.4 Study Aim and Objectives

BabA plays a well-established role in the mediation of *H. pylori* attachment to type 1 ABO/Le blood group antigens in the gastric mucosa. This glycan binding profile makes BabA an excellent candidate for imparting cytoadhesive properties to particulate drug delivery vectors. However, to successfully create such a system, it is fundamentally imperative that the properties of BabA as an *H. pylori* adhesin are maintained *ex situ* and can be translated into a drug delivery system.

The aim of this study was to characterise the functional properties of BabA, in the context of its use as a gastric targeting moiety (Figure 1.9). This was achieved through three specific research objectives, which have been insufficiently addressed in the literature to date:

- 1. Development of a recombinant BabA expression and purification method.** This is an essential step of the study needed to support subsequent BabA characterisation and application.
- 2. Characterisation of the BabA glycan-binding site.** This is required to understand the mechanism through which BabA recognises type 1 ABO/Le blood group antigens in order to correctly exploit this interaction through a BabA-mediated drug delivery system.
- 3. Assessment of BabA acid/pepsin stability and particle-conjugate activity.** This provides crucial, first insights into the translational applicability of BabA as a targeting moiety.

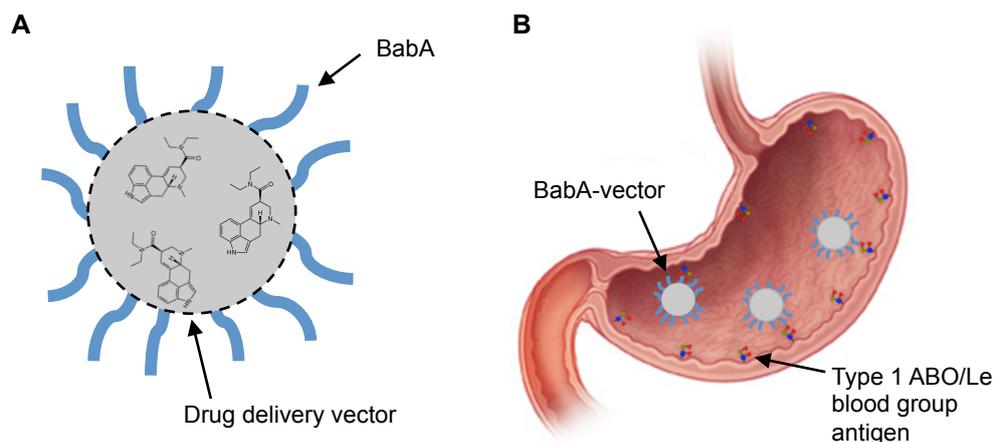


Figure 1.9: Schematic representation of BabA-mediated drug delivery vectors

This study wishes to propose **(A)** the use of BabA as a cytoadhesive moiety to direct drug delivery vectors to **(B)** type 1 ABO/Le blood group antigens in the gastric mucosa. The experimental work presented in this thesis focuses on characterising the functional properties of BabA in the context of its use as a gastric targeting moiety. **(B)** adapted from (125) with permission.

If successful, the translation of this biomimetic approach into a clinically applicable system will require several additional phases of development and assessment. These will include steps focused on drug delivery, such as the choice of delivery vehicle (e.g. polymeric nanoparticles, liposomes etc.), optimisation of drug encapsulation and, as shown in Figure 1.10, the incorporation of controlled drug release mechanisms. Product formulation will also have to be considered. Currently, it is envisaged that the drug delivery system will be orally administered via an encapsulating dosage form such that the gastric mucosa will be the first mucosal surface to which BabA-particles are exposed. Through such a formulation, the chances of effective attachment to the gastric mucosa are maximised because any potential premature or non-specific binding interactions to the oral cavity and oesophagus are avoided.

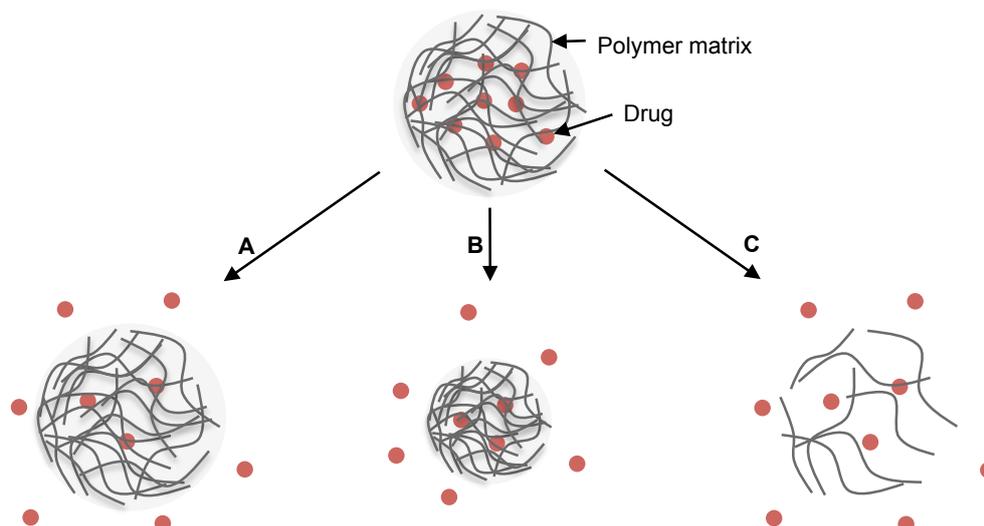


Figure 1.10: Drug release mechanisms from polymeric nanoparticles

Polymeric nanoparticles are shown here, as an example of a drug delivery vehicle, to illustrate how drug molecules can be released from drug delivery vehicles over time. The mechanisms illustrated are: **(A)** diffusion of the drug out of the polymer matrix, or **(B)** surface erosion of the polymer matrix, or **(C)** hydrolytic biodegradation of the polymer matrix (126).

While there are clearly longer-term translational milestones associated with this project, characterising the functional properties of BabA, in the context of its use as a gastric targeting moiety, underpins development of the entire system. In addition to progressing the development of a BabA-mediated drug delivery system, the insight gained from addressing the aforementioned research objectives of this study will provide relevant insight into a primary mechanism through which *H. pylori* colonise the gastric mucosa.

Chapter 2: Overview of experimental techniques

Throughout the course of this study, the use of a wide range of techniques was required to sufficiently encompass the breadth of each research objective.

The first objective of this study was to develop a method to express and purify recombinant BabA – this was performed using standard molecular biology and chromatographic techniques. As such, an explanation of the mechanism behind the processes and instruments employed to achieve this objective is precluded from this section. However, a brief description of the specific expression vectors and host organisms used is given in Section 2.1.

The second and third objectives of this study were to characterise the glycan-binding site of BabA and to assess its acid/pepsin stability and particle-conjugate activity. To achieve this, a number of techniques were employed for which the mechanism of the instrumentation is essential for interpretation of the data output. While a comprehensive overview of each of these techniques is beyond the scope of this thesis, the fundamental principles behind these biophysical, structural and biochemical methods are introduced below in Sections 2.2 to 2.8.

2.1 Recombinant expression

Recombinant expression was chosen over native protein extraction because it offers distinct advantages relevant to this study's downstream requirements: for characterisation of the BabA glycan-binding site, recombinant expression allows the relevance of specific amino acids required for binding affinity to be easily assessed through mutagenesis; for the generation of BabA-particle conjugates, this technique enables the incorporation of functional polypeptide tags into BabA to facilitate particle linkage.

2.1.1 Expression vectors used in this study

A brief summary of the functional features of each vector used for recombinant BabA expression is given below.

- ***pET-22b(+)* expression vector**

This vector confers ampicillin resistance, contains an isopropyl β -D-1-thiogalactopyranoside (IPTG)-inducible lac operator downstream of a T7 promoter sequence, and a C-terminal 6x His tag for immunodetection and purification. Additionally, there is an N-terminal pelB leader signal sequence (MKYLLPTAAAGLLLLAAQPAMA) for potential periplasmic localisation. This vector is suited to protein expression in *Escherichia coli* (127).

- ***pOPE101* expression vector**

This vector confers ampicillin resistance, contains an IPTG-inducible lac operator behind a synthetic P/A1/04/03 lac promoter, a C-terminal c-Myc (EQKLISEEDL) tag for immunodetection and a C-terminal 6x His tag for purification. There is also an N-terminal pelB leader signal sequence for potential periplasmic localisation. This vector is suited to protein expression in *E. coli* (128).

- **az_FB84_D132 expression vector**

This is an AstraZeneca R&D proprietary vector containing features optimised for *Spodoptera frugiperda* expression. It is based on the pFastBac™ vector, making it suitable for site-specific transposition into bacmid DNA. It contains no signal sequences, immunodetection tags, or purification tags, which must be added, as desired, through cloning.

- **pDEST12.2oriP_C718 expression vector**

This is an AstraZeneca R&D proprietary vector containing features optimised for mammalian expression. It contains no signal sequences, immunodetection tags, or purification tags, which must also be added through cloning.

2.1.2 Host organisms used in this study

A brief overview of the host organisms used for both plasmid construction and protein expression is given below.

- ***E. coli* DH5α**

This is a strain typically used for cloning as it has *recA* and *endA1* mutations, which prevent the homologous recombination and degradation of plasmid DNA, respectively (129).

- ***E. coli* XL10 Gold**

These cells are also typically used for cloning applications as they exhibit the Hte phenotype, which increases the transformation efficiency of large DNA. Like *E. coli* DH5α, they are also *recA* and *endA1* deficient (129).

- ***E. coli* BL21 (DE3)**

This *E. coli* strain is ideal for routine protein expression in conjunction with T7 promoter expression vectors. As it is a DE3 lysogen, it contains a

chromosomal copy of the T7 RNA polymerase gene. Once expressed, this polymerase binds to T7 promoters on expression vectors to direct transcription of the cloned gene. BL21 (DE3) is deficient in *lon* and *ompT* proteases, which degrade intracellular and outer membrane proteins, respectively (129).

- ***E. coli Origami B (DE3)***

This strain has all the features of *E. coli* BL21 (DE3) and additional mutations to enhance disulphide bond formation in the cytoplasm of *E. coli*. These mutations are in thioredoxin reductase and glutathione reductase enzymes, thereby creating an oxidising cytoplasmic environment. Additionally, this strain is a *lacY1* deletion mutant; this enables a more precise control of expression levels by adjusting IPTG concentration (129).

- ***Spodoptera frugiperda Sf21***

This is a lepidopteron insect cell line frequently used for protein expression in conjunction with a baculovirus expression vector system (130).

- ***Human Embryonic Kidney (HEK) 293-6E***

This is a derivative of the mammalian cell line HEK 293 optimised for transient protein expression. It stably expresses the Epstein Barr virus (EBV) nuclear antigen 1 (EBNA-1) and grows in suspension, as opposed to adherently. Presence of EBNA-1 has been shown to increase the semi-stable episomal propagation of plasmids containing an EBV origin of replication (131).

- ***Chinese Hamster Ovary (CHO) G22***

This is a derivative of the CHO mammalian cell line optimised and patented by AstraZeneca R&D for enhanced transient expression. It grows in suspension, stably expresses EBNA-1 and also glutamine synthetase – an enzyme that catalyses condensation of glutamate and ammonia to form glutamine (132).

2.2 Surface plasmon resonance

Surface plasmon resonance (SPR) was used as a direct method to observe binding between recombinant BabA and glycans. It is an extremely sensitive technique that detects the real-time association and dissociation of molecules and, as such, can also be used for affinity and kinetic determination (133).

SPR is a spectroscopic technique. First, a molecule of interest (ligand) is immobilised, through various possible chemistries, onto the dextran-coated gold surface of a glass-backed slide. A solution containing the ligand's binding partner (analyte) is then flown over the surface and interactions are assayed based on the SPR phenomenon. The principles of the SPR phenomenon are centred on the reflection of light. In the assay, the sensor chip is irradiated with a wedge-shaped beam of polarised light, from its glass side, at an angle range where only reflection occurs and no light passes through the glass slide. During the reflection of light, a phenomenon exists whereby an electromagnetic field component, known as the evanescent wave, traverses a short distance (~150 nm) across the glass slide into the gold/buffer layer of the chip. This energy is absorbed by free electrons in the gold layer (surface plasmons) but only from a specific angle within the range of incident light. Therefore, when the reflected light is analysed by a detector, a shadow is observed at the corresponding angle – the SPR angle. The SPR angle changes based on the refractive index of the surface layer buffer within the evanescent wave field (Figure 2.1). As this is largely determined by the surface concentration of molecules, association and dissociation binding events can be monitored through reversible changes to the SPR angle (133, 134).

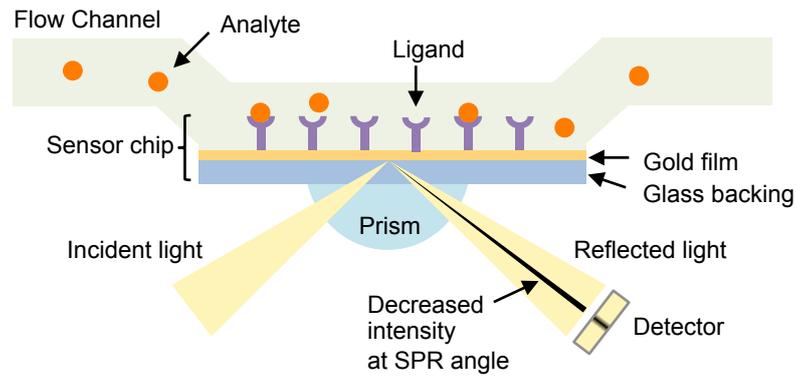


Figure 2.1: The principle of SPR

Changes to the surface concentration of molecules causes a change in the SPR angle, which can then be translated into a sensogram.

The main advantage of SPR is that binding events are monitored in real-time. Thus, the curvature of sensogram traces can be analysed to determine the kinetics of binding, that is, the association rate, k_{on} , and dissociation rate, k_{off} , constants (Figure 2.2).

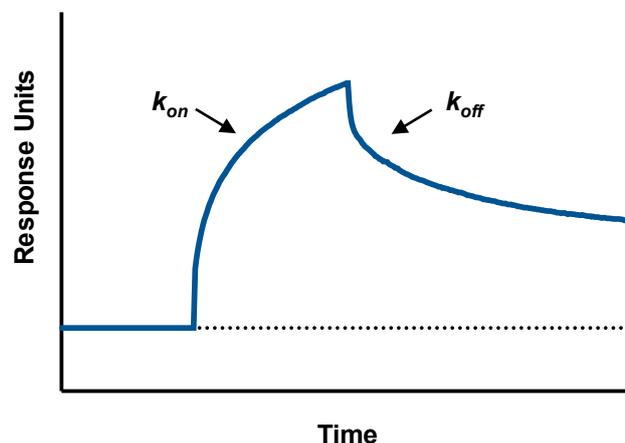


Figure 2.2: A typical SPR sensogram

A sensogram depicts changes in the SPR angle as Response Units. Distinct association and dissociation phases are evident if the analyte interacts with the ligand and kinetic rate constants can be determined by curve fitting.

These constants can also be used to determine binding affinity: the association constant, $K_A = k_{on}/k_{off}$; the inverse applies to the dissociation constant, $K_D = k_{off}/k_{on}$ (135). Alternatively, binding constants can be calculated

directly through equilibrium analysis. This is done by measuring the concentration of analyte required to saturate all ligand binding sites and fitting obtained data to a binding isotherm. Finally, because SPR is a highly sensitive technique, changes in the refractive index of buffers can also affect signal output, as a consequence, careful manipulation in combination with thorough results evaluation is needed to interpret experimental data (136, 137).

2.3 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was used to calculate the binding affinity and thermodynamic parameters of interactions between recombinant BabA and glycans. As its name suggests, this technique characterises binding interactions based on the amount of heat that is given out (exothermic processes) or taken up (endothermic processes) during binding events (138).

An isothermal titration calorimeter contains two identical reference and sample cells, which are both maintained at the same user-defined temperature through separate cell heaters. In an experimental run, the reference cell is filled with the same buffer as the protein-containing sample cell, which receives known amounts of ligand injections (Figure 2.3). During these injections, extremely sensitive thermocouple circuits detect the slight differences in temperature between the reference and sample cells. In the case of exothermic reactions, binding events cause an increase in the temperature of the sample cell. This temperature increase is fed back to the sample cell heater so that it reduces its heat/power input in order for the cell to be cooled back to the same temperature as the reference cell. The opposite occurs during endothermic reactions (139).

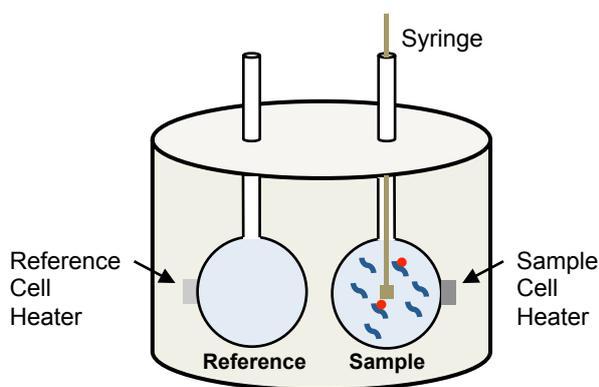


Figure 2.3: Schematic diagram of an ITC instrument

ITC determines binding heat change by comparing the temperatures of reference (buffer only) and sample (protein+ligand in buffer) cells during a binding experiment.

In other words, an isothermal titration calorimeter works by matching the temperatures of sample and reference cells over the course of a binding experiment. The signal output of these instruments is the positive or negative change in power ($\mu\text{cal}/\text{sec}$) applied to the sample cell in order to match the temperature of the reference cell over multiple titrations. A relatively high signal is observed upon initial titration of ligand into protein, this then decreases after repeated titrations as protein binding sites become saturated. The positive or negative power change applied to the sample cell equates to the heat absorbed or released during binding, respectively. Consequently, the heat change after each injection can be determined by calculating the area under the peak of the power change signal output (Figure 2.4) (140).

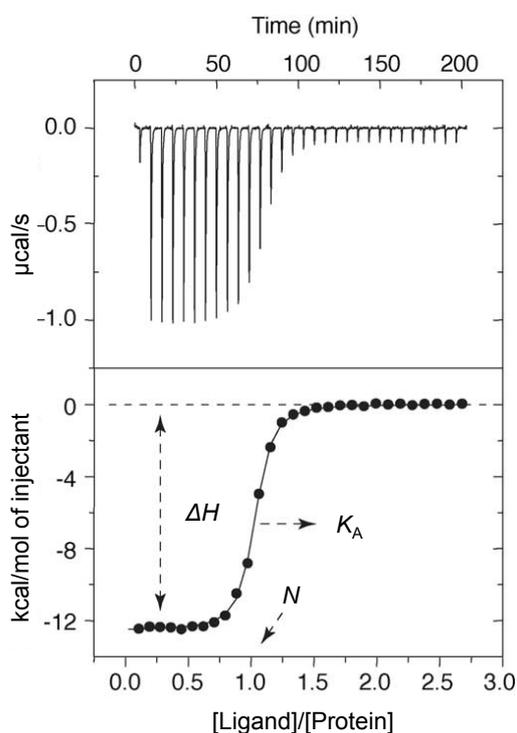


Figure 2.4: Representative output from a typical ITC experiment

The signal output (top) from an ITC experiment is transformed into a binding isotherm (bottom) by plotting the area under each peak, normalised per mole of injected ligand, against the ligand:protein molar ratio. K_A , ΔH and N can then be obtained after non-linear least squares analysis of the data. Note: low affinity interactions do not yield

sigmoidal curves but are similarly interpretable (141). Reprinted from (142) with permission.

Because a binding isotherm based on heat change can be generated, analysis of the data can yield the binding enthalpy change (ΔH), K_A and stoichiometry (N). The K_D can then be calculated from Equation 1:

$$K_A = 1/K_D \quad (1)$$

This can then be used to calculate the Gibbs free energy change (ΔG) of the complex using Equation 2, where R is the ideal gas constant ($= 8.31 \text{ J}^{-1} \text{ K}^{-1} \text{ mol}$) and T is temperature (in Kelvin):

$$\Delta G = RT \ln K_D \quad (2)$$

Finally, the binding entropy change (ΔS) can be calculated using Equation 3:

$$\Delta G = \Delta H - T\Delta S \quad (3)$$

As such, a full set of binding affinity constants and thermodynamic parameters can be determined through ITC. While the K_A (or K_D) and N for protein:ligand interactions are significant because they describe the strength and stoichiometry of the binding interaction, respectively, the relative enthalpic and entropic contributions of binding are similarly important in order to understand the mechanism of binding. Favourable enthalpy (where the value of ΔH is negative) is mediated by non-covalent interactions such as hydrogen bonds and van der Waal's contacts; favourable entropy (where the value of $-T\Delta S$ is negative) is effectively mediated by hydrophobic interactions (142, 143).

2.4 X-ray crystallography

X-ray crystallography is a structural method that was used to identify the recombinant BabA glycan-binding site. This technique enables the determination of the three-dimensional atomic structure of a protein, within a solid crystalline material, in the absence and presence of its binding partner.

X-ray crystallography is an elaborate technique based on complex physical phenomena and mathematical relationships that are readily calculable due to advances in computational software. Structural determination is achieved through a number of key steps. Briefly, proteins are crystallised and exposed to X-ray radiation where the electrons of each atom within the protein molecules diffract the X-rays. Diffracted X-rays are collected and interpreted using mathematical analyses to produce an electron density map of the entire protein, from which its atomic model can be directly inferred (144).

The initial crystallisation step is central to this technique: an X-ray signal diffracted by an individual molecule is not detectable, however, because a crystal lattice contains molecules identically packed in a repeated manner with three-dimensional, long-range order, the signal of diffracted X-rays is amplified to levels that enables their analysis. A popular method to crystallise proteins, which was used in this study, is the sitting drop vapour diffusion method. Here, a droplet containing a high concentration of protein is mixed with a well-defined solution containing a precipitant that effectively competes with the protein for water molecules. This droplet is placed in a sealed system adjacent to a significantly larger reservoir containing the same precipitant solution. In comparison to the droplet, this reservoir has a higher concentration of precipitant because it is undiluted. This causes water molecules to gradually diffuse from the droplet to the reservoir, through the vapour phase, so that

equilibrium may be reached (Figure 2.5). Through this process, the concentration of protein and precipitant in the droplet slowly increases and encourages the spontaneous formation of protein crystal nuclei and subsequent crystal growth, under the right conditions. Factors like pH and temperature also affect successful crystallisation, which may often end in protein precipitation if optimal conditions are not achieved. If available, small pieces of protein crystals can also be incorporated into the crystallisation droplet to encourage nucleation. This is known as crystal seeding (145).

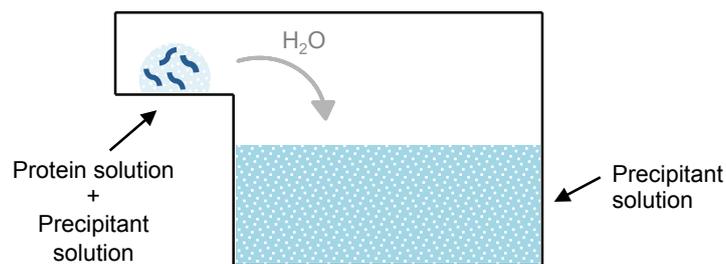


Figure 2.5: Schematic representation of the sitting drop vapour diffusion method used in protein crystallisation

Hundreds to thousands of precipitant solutions can be screened using this process to identify and optimise conditions that yield protein crystals.

After successful crystallisation, crystal irradiation is performed and a charged couple device detector collects the diffraction spots generated from X-ray photons shone at the protein crystal; crystals are gradually rotated during this process to maximise data collection (Figure 2.6). Because the structure of a material is mathematically related to its diffraction pattern through a function known as a Fourier transform, an inverse Fourier transform can be applied to create an electron density map of a protein from its diffractogram. Importantly, this requires a diffraction pattern to originate from a single crystal and consequently twinned crystals represent a significant hindrance to structure solution. To perform an inverse Fourier transform, both the amplitude and phases of the diffracted X-ray waves that generated the diffraction pattern

need to be known. However, an inverse Fourier transform cannot be readily applied in macromolecular crystallography. From a diffractogram, the amplitude of the waves is known because this relates to the intensity of the spots, however, all wave phase information is lost; this generates what is known as the phase problem (146).

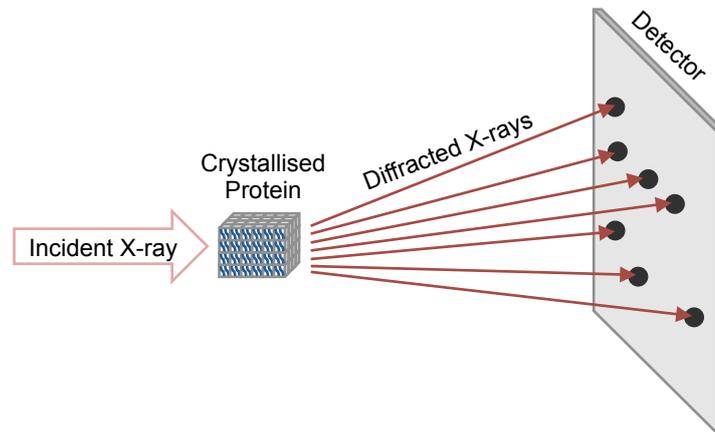


Figure 2.6: Schematic representation of the data collection step in X-ray crystallography

In order to obtain complete diffraction data, protein crystals are rotated during irradiation (not shown).

The phase problem can be solved using several different techniques. If the atomic coordinates are available for a protein with a similar amino acid composition to that of the protein of interest (greater than 20-40% sequence identity), the phase problem can be simply solved through a computational process known as molecular replacement. Here, the orientation of the structurally characterised molecule is theoretically modelled into the crystal lattice such that it would give a near-identical diffraction pattern to that of the unknown protein. If this is suitably similar, the wave phase information of the known molecule can effectively be borrowed to enable an inverse Fourier transform to be performed on the experimental diffraction pattern (146). Where no homologous protein exists, single-wavelength anomalous dispersion is one

of the most popular methods used to solve the phase problem. This method was used in this study. It is based on the fact that within the typical wavelengths of X-ray radiation employed in macromolecular crystallography (0.8-2.3 Å), after interaction with the atoms found in natural amino acids, the phase of incident X-ray waves is not changed during diffraction. However, if a type of atom known as an anomalous scatterer is present in the protein, at a specific wavelength that is still suitable for structural determination, the atom will partly absorb and partly diffract the X-rays. This unique event, known as anomalous scattering, can be detected in a diffraction pattern and enables the phase information for all waves to be gained through mathematical processes (147). An example of an anomalous scatterer is selenium, which is typically incorporated into a protein through the substitution of methionine for selenomethionine.

Typically, after the use of molecular replacement or single-wavelength anomalous dispersion to solve the phase problem, an electron density map can be built and after an appropriate number of rounds of iterative refinement, an atomic model can be generated. A detailed insight into the full mathematical processes used to translate experimental data into an atomic model can be found in (148).

2.5 Circular dichroism spectroscopy

To study protein conformational state, circular dichroism spectroscopy was used as an investigative tool. This technique enables the evaluation of a protein's secondary structure content. As such, it is effective for studying whether amino acid substitutions or microenvironmental factors affect recombinant BabA folding and stability (149).

Circular dichroism spectroscopy studies the absorbance of circularly polarised light by molecules. Circularly polarised light has two components – left- and right-handed circularly polarised light. If this form of light is passed through a chiral molecule with absorptive properties, left- and right-handed circularly polarised light will be absorbed to different extents because of the molecule's asymmetry. This effect is known as circular dichroism and the difference in left- and right-handed absorption can be measured by spectropolarimeters (150).

Proteins are well suited for investigations using circular dichroism spectroscopy. Between approximately 190 and 260 nm (the far-UV spectral region), their peptide bonds act as chromophores and the degree to which they absorb left- or right-handed circularly polarised light is dependent on the conformation of their asymmetric polypeptide chains. As such, α -helix, β -sheet and random coil structures each yield distinctive CD spectra (Figure 2.7) and the combination of these structural features in a protein can be characterised and studied in different contexts (151).

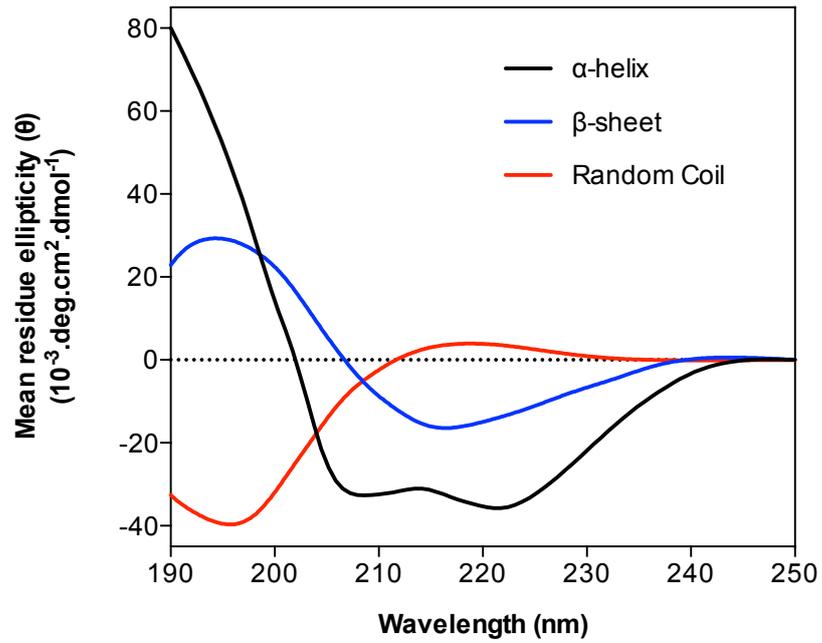


Figure 2.7: Far-UV circular dichroism spectra associated with the various protein secondary structures

The CD spectra shown are of poly-L-lysine in different conformations (due to changes in pH and temperature) (152). These spectra fingerprints are consistent amongst proteins dominated by the indicated structural features (153).

2.6 Differential scanning fluorimetry

Differential scanning fluorimetry was used to characterise the thermal-induced unfolding transition of recombinant BabA. This information is similarly useful for comparing the relative stabilities between protein variants, and to study the (de)stabilising effect of a protein's microenvironment (e.g. pH) (154).

In this experiment, the protein of interest is incubated with a unique fluorescent dye, the properties of which are essential to the assay, and subjected to an increasing heat gradient in a thermocycler with fluorescence excitation/emission filters. The fluorescent dye used, typically SYPRO Orange, non-specifically binds to hydrophobic surfaces and only emits a signal when in such local environments (its fluorescence is strongly quenched by water). Thus, during the thermal-induced unfolding transition of a protein, the dye emits fluorescence once bound to a protein's hydrophobic patches – these are typically hidden within the protein core in its folded conformational state. The more a protein unfolds, the more the fluorescent dye that binds. Accordingly, its thermal denaturation curve can be obtained by plotting fluorescence against temperature and its unfolding transition midpoint (T_m) can be calculated by non-linear regression fitting. Once completely unfolded, the protein's exposed hydrophobic core tends to lead to intermolecular aggregation and subsequently causes a displacement of the fluorescent dye, leading to re-quenching of its signal (Figure 2.8). Because the chemical structure of SYPRO Orange is proprietary, its exact mechanism of protein-binding is not known (155, 156).

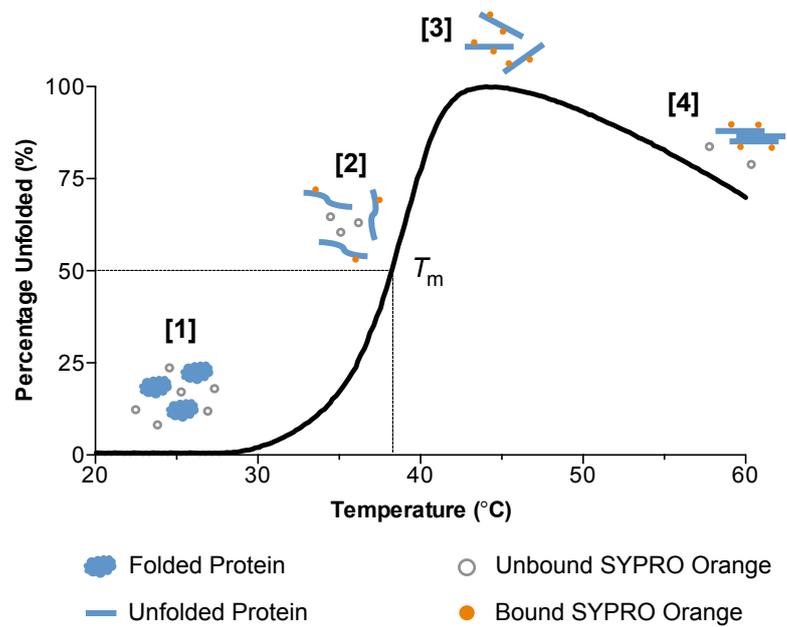


Figure 2.8: A typical thermal-induced protein unfolding transition

[1] Correctly folded proteins do not bind SYPRO Orange and the aqueous environment quenches its fluorescence, **[2]** Thermal denaturation allows the dye to bind to the internal hydrophobic core of the protein and fluorescence is emitted, **[3]** Maximal fluorescence intensity is reached at complete protein unfolding, **[4]** Protein aggregation results in dye-protein dissociation.

2.7 Dynamic light scattering

Dynamic light scattering (DLS) was used to assess the colloidal stability of model particles used or generated in this study. DLS measures the hydrodynamic diameter and polydispersity of particles in the sub-micron to micron range. Therefore, it can be used to study how the conjugation of BabA to particle surfaces impacts the ability of the system to remain suspended in solution, that is, to resist aggregation and/or sedimentation (157).

DLS calculates particle size by measuring the speed of Brownian motion, which is the random movement of particles in a suspension due to collision with surrounding solvent molecules. To determine this experimentally, the light scattering pattern of a suspension is analysed over time (Figure 2.9).

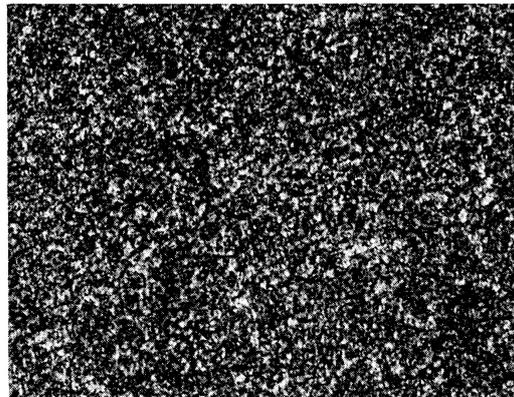


Figure 2.9: A typical light scattering pattern

A representative light scattering pattern, also known as a speckle pattern, at a single time point is shown. The mixture of bright and dark areas is due to diffracted light from a particle suspension arriving at the detector in phase (constructive interference) and out of phase (destructive interference), respectively. Reprinted from (158) with permission.

As diffusion occurs, particles are spatially rearranged and, as a result, their light scattering pattern changes. Smaller particles diffuse faster through a liquid than larger particles, hence, their light scattering pattern changes more

rapidly. The rate of changes to a particle's light scattering pattern can therefore be used to calculate the speed of Brownian motion. In practice, these changes are recorded by a detector that monitors fluctuations in light intensity over time (Figure 2.10). The translation of this data into the speed of Brownian motion is then achieved using a correlator, as described in (159).

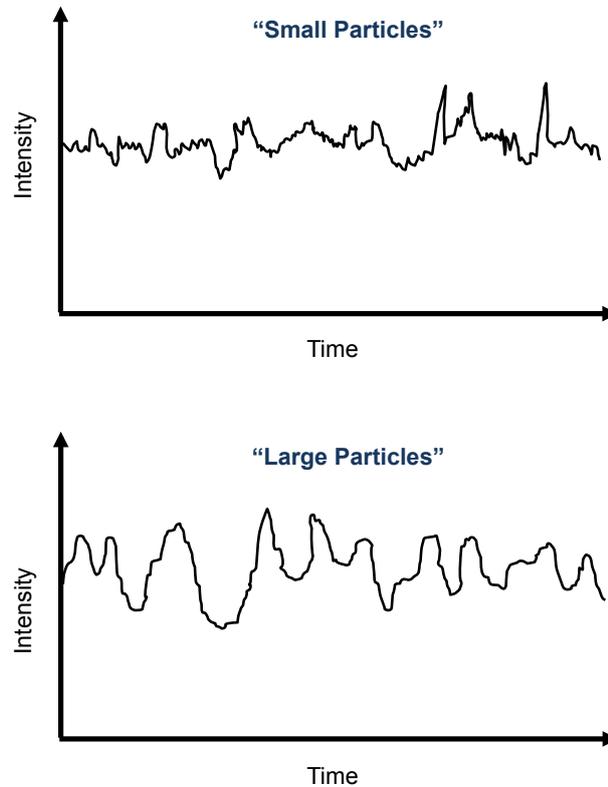


Figure 2.10: Schematic representation of typical particle intensity fluctuations

Fluctuations in light intensity occur more rapidly in smaller particles (top) than larger particles (bottom) because of their respective rates of diffusion within a suspension.

From the speed of Brownian motion, which is defined as the translational diffusion coefficient (D), particle size can be estimated using the Stokes-Einstein equation where $d(H)$ is the hydrodynamic diameter, k is the Boltzmann constant, T is the absolute temperature (in Kelvin) and η is the viscosity of the dispersant:

$$d(H) = \frac{kT}{3\pi\eta D}$$

As this equation converts the speed of particle diffusion through a fluid into particle size, size is reported as a hydrodynamic diameter. It is also important to note that this equation assumes that the particle is a perfect sphere. Through this method, an average hydrodynamic diameter can be calculated and the particle size distribution can also be analysed to reveal the polydispersity of the sample.

2.8 Bio-layer interferometry

To characterise the glycan binding activity of BabA-particle conjugates in this study, bio-layer interferometry (BLI) was used. Similar to SPR, this technique monitors real-time association and dissociation events so it can be used to simply assess whether binding takes place or to determine kinetic and affinity constants (160).

In this technique, molecules are immobilised onto the biocompatible surface of a glass fibre biosensor tip that is designed to support specific covalent or non-covalent linkages. Biosensor tips are then immersed into a buffer containing the binding partner of interest and only molecules that attach to the biosensor tip are detected, that is, buffer components, changes in refractive index etc. do not affect the detection of binding (161). BLI is a spectroscopic technique. In this assay, as depicted in Figure 2.11, white light of all wavelengths in the visible spectrum is sent down a biosensor tip, which contains two layers – an internal reference layer and a biocompatible surface layer. Incident light is reflected back up the tip from these two interfaces.

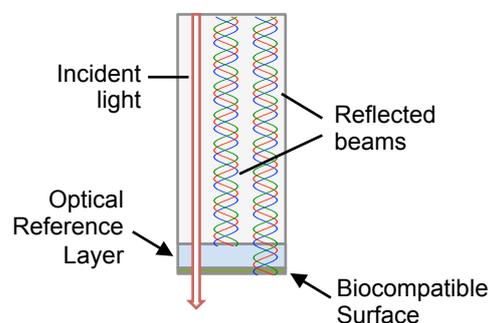


Figure 2.11: Schematic representation of biosensor tips used in BLI

White light is emitted down the biosensor tip and a portion is reflected back along an optic fibre, from the indicated surfaces, and collected by a spectrometer.

Because the reflections originate from two different distances, when the reference and surface layer light waves are analysed together, at each

wavelength, one of the following would be found: 1) some waves are perfectly out of phase (resulting in destructive interference and no wave amplitude); 2) some waves are perfectly in phase (resulting in constructive interference and a doubling of wave amplitude); and 3) some waves are partially in and out of phase. The detected amplitudes can be collated and translated into relative intensities, then plotted against their respective wavelengths to produce an interferometric profile (Figure 2.12A). An increase in the distance between the reference layer and the surface layer, due to molecule immobilisation to the biocompatible surface for example, again changes the relative phases of reflected waves. This causes a rightward shift to the interferometric profile (Figure 2.12B). This shift is a direct measure of molecule association to the biosensor tip. Further association to immobilised molecules will increase the height of the biocompatible surface layer and induce another rightward shift to the interferometric profile. As these shifts are reversible and occur in real-time, from this, a sensogram trace can be produced (162, 163)

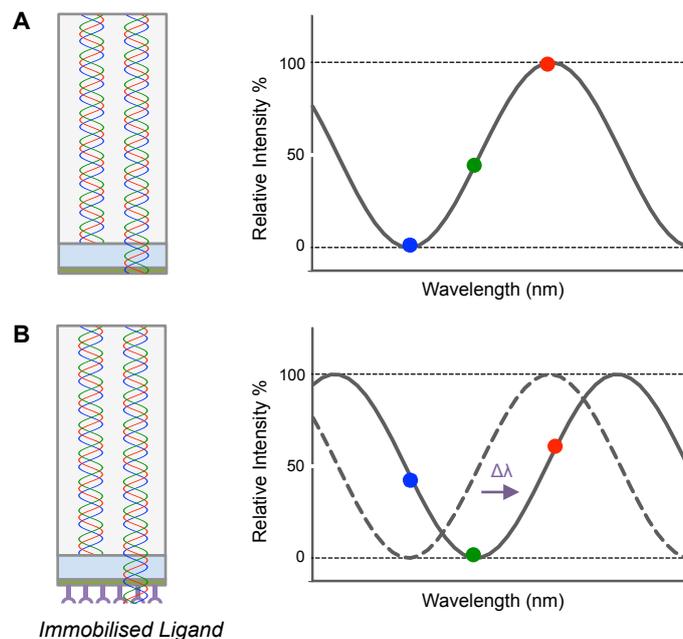


Figure 2.12: Detection of molecule association to biosensor tips through optical interferometry

Analysis of reflections from both optical and surface layers creates an interferometric

profile. In this example, initially, **(A)** blue light waves were completely out of phase (no wave amplitude), red light waves were completely in phase (doubling of wave amplitude); and green light waves were partially in phase (partial increase of wave amplitude). After molecule immobilisation, **(B)** blue light waves are no longer out of phase, green are now out of phase and red light waves are no longer perfectly in phase. This causes a rightward shift ($\Delta\lambda$) to the interferometric profile (this analysis is performed at all wavelengths in the visible light spectrum).

Chapter 3: Developing a recombinant BabA expression and purification method

3.1 Introduction

The first objective of this study was to develop a method to recombinantly express and purify BabA. Indeed, there have been reports of successful BabA expression in *E. coli* in the literature. These include secretion to the periplasmic space (112), expression as a fusion protein using glutathione S-transferase (GST) as a solubility enhancing partner (164), and refolding from inclusion bodies (165). However, each of these approaches has drawbacks. Periplasmic expression is associated with low yield (166), GST-fusion proteins require partner cleavage before structural studies (167), and refolding from inclusion bodies is time-consuming with proteins not always regaining full function (168, 169). Thus, it was deemed necessary to screen various expression vectors and host organisms in search of an efficient method to express recombinant BabA for use in this study.

In light of the desired characterisation and application of BabA, a recombinant protein containing only its glycan-binding region was necessary. At the commencement of this study, there was no structural insight into the molecular basis of glycan recognition by BabA. Thus, the location of its binding site was unknown. However, BabA is an outer membrane protein and like other members within the Hop family, it is expected to contain an N-terminal extracellular domain and a C-terminal domain predicted to form a β -barrel unit for outer membrane insertion (Figure 3.1) (86). Consequently, the C-terminal domain of BabA is not needed for glycan binding and may be excluded from the recombinant protein design. In fact, the expression of full-length outer membrane proteins is typically avoided, unless essential for study purposes,

as the hydrophobicity of their transmembrane domains usually leads to intracellular aggregation during routine expression (170).

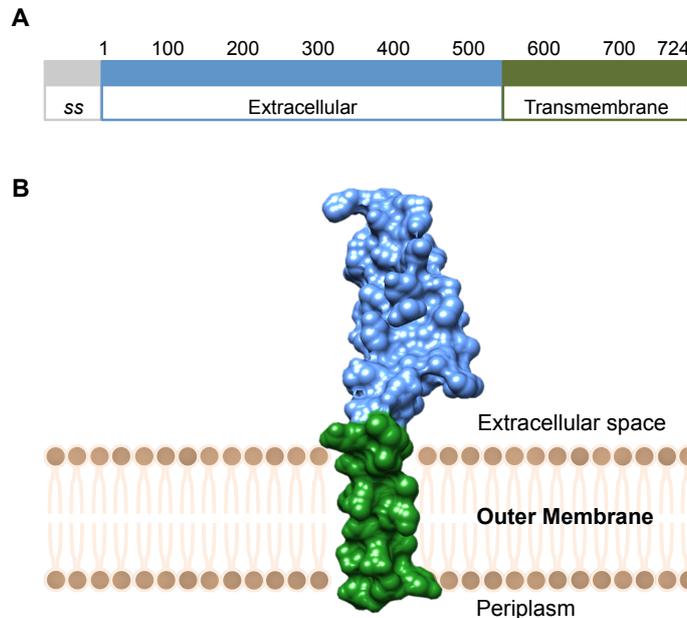


Figure 3.1: The predicted domain structure of BabA

BabA is predicted to have an N-terminal extracellular domain (blue) and a transmembrane C-terminal domain (green) that spans the *H. pylori* outer membrane. **(A)** Schematic representation of the BabA primary structure; the *babA* gene encodes a 20 amino acid N-terminal signal sequence (ss) that is cleaved during the transportation of BabA to the outer membrane (86); mature BabA is formed of 724 amino acids. **(B)** Schematic representation of the BabA N-terminal extracellular and C-terminal transmembrane domains orientated within the *H. pylori* outer membrane.

Recombinant BabA expression and purification was achieved in three steps. First, a fragment of the *babA* gene, corresponding to the expected extracellular domain of the protein, was cloned into a variety of vectors designed for protein expression in either prokaryotic or eukaryotic hosts. The genomic DNA of *H. pylori* J99 was chosen a source for cloning because this strain is known to express a generalist BabA protein, that is, one that binds to the type 1 ABO/Le blood group antigens in group-A, -B and -O individuals (121). Furthermore, BabA from *H. pylori* J99 has been confirmed to be functional and, when present on *H. pylori*, is involved in a high affinity glycan-

binding interaction – the association constant of *H. pylori* J99 attachment to Le^b glycoconjugates is reportedly $\sim 2 \times 10^{-12}$ M (123). These features were considered to be desirable for the proposed application of BabA in drug delivery. After the generation of expression constructs, a number of prokaryotic (*E. coli*) and eukaryotic (*S. frugiperda*, HEK and CHO) expression systems, described in Section 2.1.2, were assessed for their ability to generate a soluble recombinant protein. Given that the predicted extracellular domain of BabA contains eight cysteine residues, these expression systems were chosen because they supported disulphide bond formation, which is a common post translational modification in the outer membrane proteins of Gram-negative bacteria (171). Finally, the successful strategies identified during expression screening were scaled up for chromatographic protein purification.

Ultimately, the optimal method identified was an improvement to a previously reported strategy based on periplasmic expression (112).

3.2 Experimental Procedures

3.2.1 Generation of BabA expression constructs

The BabA expression constructs engineered in this study are summarised below in Table 3.1.

Table 3.1: Expression constructs engineered in this study

The abbreviations used in the naming of constructs represent the following: *babA*₅₂₇ – *babA* gene fragment derived from *H. pylori* strain J99 encoding amino acids 1-527; *babA*_{527K} – *babA* gene fragment derived from *H. pylori* strain J99, encoding amino acids 1-527, with a 3' hexa-lysine tag; Δ*pelB* – deletion mutation of *pelB* leader sequence from vector.

^aPlasmid construction was outsourced to Life Technologies, USA

Construct Name	Vector Backbone	Cloning Host
pET22b_ <i>babA</i> ₅₂₇	pET-22b(+)	<i>E. coli</i> DH5α
pET22b(Δ <i>pelB</i>)_ <i>babA</i> ₅₂₇	pET-22b(+)	<i>E. coli</i> DH5α
pOPE101_ <i>babA</i> ₅₂₇	pOPE101	<i>E. coli</i> XL10 Gold
pOPE101_ <i>babA</i> _{527K}	pOPE101	<i>E. coli</i> XL10 Gold
AZ1_ <i>babA</i> _{527K}	az_FB84_D132	n/a ^a
AZ2_ <i>babA</i> _{527K}	pDEST12.2_oriP	n/a ^a

These were generated using the steps described below.

3.2.1.1 *babA* gene fragment design – sequence analysis

To identify the region of BabA corresponding to its transmembrane domain, CLC Main WorkBench 7.6 (CLC bio, USA) was used to perform a sequence alignment (protein) of BabA J99 (GenBank accession no. AAD06409.1) with SabA 26695 (GenBank accession no. AFV41939.1). For a prediction of structural features within the putative C-terminal transmembrane domain of BabA J99, amino acids corresponding to residues 528-724 in mature BabA J99 were submitted to Phyre² (172). 95% of residues were modelled with >90% confidence; the model was based on 14 template structures – 12 of which were outer membrane proteins from Gram-negative bacteria.

3.2.1.2 Polymerase chain reaction amplification

Q5™ high-fidelity DNA polymerase (New England Biolabs, USA) was used to amplify a fragment of the *babA* gene (Genbank accession no. AE001439.1; Locus_tag = jhp_0833) from the genomic DNA of *H. pylori* J99 (kindly donated by John Atherton, University of Nottingham). This gene fragment, henceforth referred to as *babA*₅₂₇, encoded amino acids 1-527 of mature BabA J99 and was designed to exclude the putative C-terminal transmembrane domain. The *babA* gene also encodes a 20 amino acid N-terminal signal sequence that is cleaved in *H. pylori* during the secretion of BabA to its outer membrane (86); this sequence was also excluded from the *babA*₅₂₇ gene fragment. The polymerase chain reaction (PCR) was set up according to the manufacturer's protocol using the primers shown in Table 3.2. An annealing temperature of 60 °C was used for all reactions.

Table 3.2: Primers for *babA*₅₂₇ amplification

The incorporated restriction sites are designated in the primer name and underlined in the oligonucleotide sequence. FOR and REV indicate sense and antisense primers, respectively. The number in the primer name denotes the first BabA amino acid encoded by the nucleotides on the 3' end of the restriction site.

Construct	Primer Name	Sequence (5'-3')
pET22b_ <i>babA</i> ₅₂₇	BabA [1- FOR]_NcoI	TCGGATCCATGGAAGACGACGGC TTTTAC
	BabA [-527 REV]_XhoI	TATGTCCTCGAGGAGTTCTTGTT GATGGTTTGG
pET22b(ΔpeIB)_ <i>babA</i> ₅₂₇	BabA [1- FOR]_NdeI	TCGGATCATATGGAAGACGACGG CTTTTAC
	BabA [-527 REV]_XhoI	TATGTCCTCGAGGAGTTCTTGTT GATGGTTTGG
pOPE101_ <i>babA</i> ₅₂₇	BabA [1- FOR]_NcoI	TCGGATCCATGGAAGACGACGGC TTTTAC
	BabA [-527 REV]_BamHI	CGAGTTGGATCCGAGTTCTTGTT GATGG
pOPE101_ <i>babA</i> _{527K}	BabA [1- FOR]_NcoI	TCGGATCCATGGAAGACGACGGC TTTTAC
	BabA [-527K REV]_BamHI	TCTGCTGGATCCCTTCTTCTTCTT CTTCTTGAGTTCTTGTTGATGG

3.2.1.3 Cloning

PCR products were purified after agarose gel electrophoresis, using a QIAquick gel extraction kit (Qiagen, USA), then digested, alongside the desired expression vectors with restriction enzymes under conditions specified by the manufacturer (New England Biolabs, USA). Digested PCR products and expression vectors were mixed in a 3:1 insert to vector molar ratio and ligated with T4 DNA ligase (New England Biolabs, USA), according to the manufacturer's protocol. For plasmid propagation, ligation mixtures were transformed into chemically competent *E. coli* via the following heat-shock protocols. *E. coli* DH5 α cells were used to propagate pET-22b(+) based constructs using the following heat shock method for transformation: 30 minutes on ice, followed by 90 seconds at 42 °C, then 2 minutes on ice. Similarly, *E. coli* XL10 Gold cells were used to propagate pOPE101 based constructs using the following heat shock method for transformation: 30 minutes on ice, followed by 30 seconds at 42 °C, then 2 minutes on ice. After each heat shock process, cells were grown in super optimal broth with catabolite repression (SOC) medium for one hour at 37 °C (200 rpm) before being grown overnight on lysogeny broth (LB), containing ampicillin at 100 μ g/mL (LB_{amp}), agar plates for plasmid selection. Single colonies were selected from the overnight growth plates for colony PCR using primers, shown in Table 3.3, that spanned the multiple cloning site of the host vectors. GoTaq[®] DNA Polymerase (Promega, USA) was used for PCR amplification, according to the manufacturer's protocol; an annealing temperature of 56 °C was used in all reactions. Agarose gel electrophoresis was used to screen the PCR products for gene insertion. Subsequently, positive clones were grown overnight at 37 °C (200 rpm) in 5 mL LB_{amp} medium followed by plasmid extraction with a QIAprep miniprep kit (Qiagen, USA), according to the

manufacturer's protocol. Finally, confirmatory sequencing of the plasmids was performed (Source BioScience, Nottingham) using primers listed in Table 3.3.

Table 3.3: Primers for colony PCR and confirmatory sequencing

FOR and REV indicate sense and antisense primers respectively. T7 FOR and REV are universal primer sequences corresponding to the T7 Promoter and Terminator. The numbers in the primer name denotes the corresponding base pairs (sense strand) in the pOPE101 or *babA* DNA sequences.

^aPrimers used for colony PCR and confirmatory sequencing

^bPrimer used for confirmatory sequencing only

Construct Backbone	Primer Name	Sequence (5'-3')
pET22b(+)	T7 FOR ^a	TAATACGACTCACTATAGGG
	T7 REV ^a	GCTAGTTATTGCTCAGCGG
pOPE101	<i>babA</i> [904-923bp REV] ^b	GCCTCACTACTATTACTAGC
	pOPE[27-46bp FOR] ^a	TTGACTTGTGAGCGGATAAC
	pOPE[1135-1154bp REV] ^a	ATGTGTCAGAGGTTTTCCACC
	<i>babA</i> [904-923bp REV] ^b	GCCTCACTACTATTACTAGC

3.2.2 Expression screening

The expression constructs and corresponding hosts used in this study are summarised below in Table 3.4.

Table 3.4: Expression constructs and corresponding hosts employed in this study

The abbreviations used in the naming of constructs represented the following: *babA*₅₂₇ – *babA* gene fragment derived from *H. pylori* strain J99 encoding amino acids 1-527; *babA*_{527K} – *babA* gene fragment derived from *H. pylori* strain J99, encoding amino acids 1-527, with a 3' hexa-lysine tag; Δ*pelB* – deletion mutation of *pelB* leader sequence from vector.

Construct Name	Vector Backbone	Expression Host
pET22b_ <i>babA</i> ₅₂₇	pET-22b(+)	<i>E. coli</i> BL21 (DE3)
pET22b(Δ <i>pelB</i>)_ <i>babA</i> ₅₂₇	pET-22b(+)	<i>E. coli</i> Origami B (DE3)
pOPE101_ <i>babA</i> ₅₂₇	pOPE101	<i>E. coli</i> XL10 Gold
pOPE101_ <i>babA</i> _{527K}	pOPE101	<i>E. coli</i> XL10 Gold
AZ1_ <i>babA</i> _{527K}	az_FB84_D132	<i>S. frugiperda</i> SF21
AZ2_ <i>babA</i> _{527K}	pDEST12.2_oriP	HEK 293-6E and CHO G22

Recombinant BabA expression was screened using the following procedures.

3.2.2.1 Periplasmic expression in *E. coli* BL21 (DE3)

The pET22b_*babA*₅₂₇ expression construct was transformed into chemically competent *E. coli* BL21 (DE3) cells using a heat-shock protocol (5 minutes on ice, followed by 90 seconds at 42 °C, then 2 minutes on ice) and grown in SOC medium for one hour at 37 °C, then overnight on a LB_{amp}-agar plate. A single colony was used to inoculate 5 mL of LB_{amp} media overnight at 37 °C (200 rpm). From this starter culture, 100 μL was used to inoculate a 20 mL culture, which was grown at 37 °C (200 rpm) until the optical density of the culture at 600 nm (OD₆₀₀) reached 0.6. At this point, a 1 mL sample was taken and used to prepare a “total protein fraction” for subsequent analysis (sample preparation method described in Section 3.2.2.7). Various IPTG induction /

expression conditions were assessed:

- Growth at 24 °C (200 rpm) after induction with either 0.1 or 1 mM IPTG for 4 or 20 hours
- Growth at 37 °C (200 rpm) after induction with either 0.1 or 1 mM IPTG for 1 or 3 hours

“Soluble and insoluble fractions” were taken at each stated time point (sample preparation method described in Section 3.2.2.7). Recombinant BabA expression was analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie-staining, also described in Section 3.2.2.8.

3.2.2.2 Cytoplasmic expression in *E. coli* Origami B (DE3)

The same protocol as described above was used for the assessment of recombinant BabA expression in *E. coli* Origami B (DE3) cells. For this system, the pET22b(Δ pelB)_{*babA*_{527K}} expression construct was transformed into chemically competent cells using a slightly different heat-shock protocol: 5 minutes on ice, followed by 30 seconds at 42 °C, then 2 minutes on ice.

3.2.2.3 Periplasmic expression in *E. coli* XL10 Gold

The *E. coli* XL10 Gold strain was used for both plasmid propagation and protein expression with pOPE101_*babA*₅₂₇ and pOPE101_*babA*_{527K} vectors. Ten μ L of a glycerol stock was used to inoculate 5 mL of LB media containing 100 μ g/mL ampicillin, 12.5 μ g/mL tetracycline and 0.1 M glucose (LB_{ATG}) overnight at 37 °C (200 rpm). From this starter culture, 100 μ L was used to inoculate a 20 mL culture, which was grown at 37 °C (200 rpm) until the OD₆₀₀ reached 0.6. At this point, a 1 mL sample was taken and used to prepare a “total protein fraction”, (sample preparation method described in Section

3.2.2.7) for subsequent analysis. Cultures were then induced with 0.1 mM IPTG for 16 hours at 24 °C. In addition to a “total protein fraction”, “medium” and “periplasmic and spheroplast” fractions were prepared (sample preparation method described in Section 3.2.2.7). All protein fractions were analysed by SDS-PAGE followed by Coomassie-staining and Western blotting, as described in Section 3.2.2.8.

3.2.2.4 Secretory expression in *S. frugiperda* Sf21

Experimental procedures performed by Eileen McCall, AstraZeneca R&D.

The AZ1_*babA*_{527K} construct was transformed into *E. coli* DH10Bac™ (Life Technologies, USA) cells according to the manufacturer’s protocol, with modifications as detailed in (173). Transformants were grown at 37 °C for 48 hours on LB_{amp}-agar plates containing IPTG (0.1 mM) and X-gal (40 µg/mL) for blue/white screening. A single white colony was grown in LB_{amp} media followed by bacmid DNA isolation with a QIAprep miniprep kit (Qiagen, USA), according to the manufacturer’s protocol. *S. frugiperda* Sf21 cells, grown in suspension in a deep 24-well plate with SF900II serum-free medium (Life Technologies, USA), were transfected with bacmid DNA using Cellfectin® Transfection Reagent (Life Technologies, USA), according to the manufacturer’s protocol as per modifications in (173). After incubation for 7 days, cells were centrifuged at 2,800 g for 10 minutes (4 °C) allowing virus stocks to be obtained from medium, which were titrated with a standard plaque assay (BD Biosciences, USA), as directed by the manufacturer. Finally, to assess expression, as described in (173), 3 mL of Sf21 cells were grown in suspension in deep 24-well plates until the mid-log phase (3×10^6 cells/mL) then infected at a multiplicity of infection of 2.0. After 48 and 72 hours, cells were centrifuged at 2,800 g for 10 minutes (4 °C) and the supernatant saved

for the preparation of a “medium” fraction, as described in Section 3.2.2.7. This fraction was analysed by SDS-PAGE followed by Coomassie-staining and Western blotting, as described in Section 3.2.2.8. Sf21 cells were always grown at 27 °C, 700 rpm and 75% relative humidity using sterile techniques.

3.2.2.5 Secretory expression in HEK 293-6E

Experimental procedures assisted by Mark Abbott, AstraZeneca R&D.

The AZ2_*babA*_{527K} construct was transformed into *E. coli* DH5α cells, as described in Section 3.2.1.3, for plasmid propagation. Vector DNA was extracted from a 3 L culture using a Plasmid Giga Kit (Qiagen, USA), according to the manufacturer’s protocol. HEK 293-6E cells were grown using sterile techniques in vented shake flasks at 37 °C, 140 rpm and 5% CO₂. Cells were maintained in F17 medium (Invitrogen, USA) supplemented with 4mM L-glutamine, 25 µg/mL gentamicin and 0.1% Pluronic F68 (Sigma-Aldrich, USA). Once a cell density of 1.6-1.8x10⁶ cells/mL was reached, vector DNA was added directly to 10 mL of cells, to a final concentration of 0.75 µg/mL, in a 50 mL Tubespin tube (Sigma-Aldrich, USA) followed by polyethylenimine “Max” (Polysciences, USA) to a final concentration of 2.8 µg/mL. Cells were then grown at 37 °C, 700 rpm, 5% CO₂ and 90% relative humidity. 24 hours later, cells were fed with 6.7 mL of aforementioned media, supplemented with 0.8% HyPep1510 (Sheffield Bioscience, UK). Seven days later, cells were counted and harvested by centrifugation at 2,800 g for 10 minutes (4 °C). The supernatant was removed and saved for the preparation of a “medium” fraction, as described in Section 3.2.2.7. This fraction was analysed by SDS-PAGE followed by Coomassie-staining and Western blotting, as described in Section 3.2.2.8.

3.2.2.6 Secretory expression in CHO G22

Experimental procedures assisted by Mark Abbott, AstraZeneca R&D.

The AZ2_*babA*_{527K} construct was generated as described in Section 3.2.2.5 above. Cells were similarly grown in vented shake flasks at 37 °C, 140 rpm and 5% CO₂. However, CHO G22 cells were instead maintained in a MedImmune proprietary medium supplemented with 25 µM methionine sulphoximine and 100 µg/mL hygromycin B. Two passages prior to transfection, methionine sulphoximine and hygromycin B were omitted from the growth media. Once a cell density of 1.0-1.2x10⁶ cells/mL was reached, vector DNA was added directly to 10 mL of cells, to a final concentration of 0.5 µg/mL, in a 50 mL Tubespin tube (Sigma-Aldrich, USA) followed by polyethylenimine "Max" (Polysciences, USA) to a final concentration of 7 µg/mL. Cells were then grown at 37 °C, 700 rpm, 5% CO₂ and 90% relative humidity. 24 hours later, cells were fed with 6.7 mL of a modified MedImmune proprietary media. Seven days later, cells were counted and harvested by centrifugation at 2,800 g for 10 minutes (4 °C). The supernatant was removed and saved for the preparation of a "medium" fraction, as described in Section 3.2.2.7. This fraction was analysed by SDS-PAGE followed by Coomassie-staining and Western blotting, as described in Section 3.2.2.8.

3.2.2.7 Preparation of cellular fractions for analysis

- Total protein fraction

A 1 mL sample of bacterial cells was taken from culture and harvested through centrifugation at 10,000 g for 1 min (25 °C). The pellet was resuspended in 1/10 culture volume of phosphate buffered saline (PBS) and an equal amount of 2x non-reducing sample buffer (NRSB), that is, 0.1 M Tris-Cl (pH = 6.8), 20% glycerol, 4% SDS and 0.2% bromophenol blue, was added before heat

denaturation at 100 °C for 5 minutes ahead of SDS-PAGE.

- ***Soluble and insoluble protein fractions***

After cell harvesting, as described above, the obtained pellet was resuspended in 1/10 culture volume of 50 mM Tris-Cl (pH = 8.0), 500 mM NaCl and 5% glycerol then probe sonicated on ice (6 to 8 bursts at 10-20% duty). After centrifugation at 14,000 g for 10 minutes (25 °C), the supernatant was mixed with an equal amount of 2x NRSB and heat denatured for SDS-PAGE. This generated the soluble protein fraction. The pellet obtained after centrifugation was resuspended in 100 µL PBS and 100 µL 2x NRSB then heat denatured for SDS-PAGE. This generated the insoluble protein fraction.

- ***Supernatant protein fractions***

After cell harvesting, 1 mL of culture supernatant was concentrated by trichloroacetic acid (TCA) precipitation and resuspended in 100 µL PBS and 100 µL 2x NRSB. The resulting sample was then heat denatured for SDS-PAGE.

- ***Periplasmic and spheroplast fractions***

A 10 mL sample of bacterial cells was taken from culture and harvested through centrifugation at 10,000 g for 5 minutes (25 °C). The cell pellet was resuspended in 1/10 culture volume of cold Spheroplast solution [50 mM Tris-Cl (pH = 8.0), 20% sucrose, 1 mM ethylenediaminetetraacetic acid (EDTA)] and shaken gently for 30 minutes on ice. The cell suspension was then centrifuged for 10 minutes at 10,000 g (4 °C). The supernatant was saved for further processing while the pellet was resuspended in 1/10 culture volume of cold 5 mM MgSO₄, shaken and centrifuged again, as previously, to generate supernatant. One mL of the supernatant generated at each step was concentrated by TCA precipitation and resuspended in 100 µL PBS and 100

μL 2x NRSB. The resulting sample was then heat denatured for SDS-PAGE analysis. The remaining pellet containing spheroplasts, that is, cells lacking a periplasmic space, was resuspended in 100 μL PBS and 100 μL 2x NRSB then also heat denatured for SDS-PAGE.

3.2.2.8 Analysis of cellular fractions

- SDS-PAGE

After sample preparation, 15 to 25 μL of cellular fraction samples were loaded into precast Tris-Glycine 4-20% TGX mini gels (Bio-Rad, USA) and run in a Mini-PROTEAN™ Tetra Cell (Bio-Rad, USA), according to the manufacturer's protocol.

- Coomassie-staining

After SDS-PAGE, the Tris-Glycine gels were stained in 0.1 % Coomassie Blue® in 50% water, 40% methanol and 10% acetic acid overnight then destained the next day in 50% water, 40% methanol and 10% acetic acid until protein bands resolved appropriately.

- Western blotting

After SDS-PAGE, the Tris-Glycine gels were blotted against a 0.2 μm nitrocellulose membrane using a Trans-Blot® Turbo™ Blotting System (Bio-Rad, USA) as described by the manufacturer. After blotting, the membrane was incubated for 1 hour at room temperature in 5% non-fat milk powder diluted in Tris-buffered saline (TBS) containing 0.1% Tween 20 (TBS-T), henceforth referred to as blocking buffer. Next, the "primary antibody" was diluted into fresh blocking buffer and incubated with the membrane overnight at 4 °C. The membrane was then washed four times with TBS-T before incubation with a "secondary antibody" for 1 hour at room temperature in fresh

blocking buffer. After 4 washes, the membrane was exposed to chemiluminescent reagents [0.1 M Tris-Cl (pH = 8.0), 200 μ M p-coumaric acid, 1.25 mM luminol and 0.1% H_2O_2] and imaged with a FUJI-LAS 4000 charge coupled device camera (Fujifilm, Japan) under high binning mode. All blocking and wash steps were performed on a rocking platform. For c-Myc immunodetection, the “primary antibody” used was a mouse anti-c-Myc:Biotin clone 9E10 antibody (AbD Serotec, USA) at a 1:5000 dilution; the “secondary antibody” used was a Streptavidin-Horseradish peroxidase (HRP) conjugate (AbD Serotec, USA), also at a 1:5000 dilution. For 6x His immunodetection, the “primary antibody” used was a mouse anti-His antibody (GE Healthcare, USA) at a 1:5000 dilution; the “secondary antibody” used was a goat anti-mouse IgG HRP conjugate (GE Healthcare, USA), also at a 1:5000 dilution.

3.2.3 Protein purification

3.2.3.1 Large-scale expression and periplasmic harvest

E. coli XL10 Gold cells containing pOPE101_*babA*₅₂₇ and pOPE101_*babA*_{527K} constructs were grown in 75 mL of LB_{ATG} media overnight at 37 °C (200 rpm). Three mL of overnight culture was used to inoculate 600 mL of media grown in a 2 L Erlenmeyer flask. The culture was grown at 37 °C (200 rpm) until the OD₆₀₀ reached 0.6 when the temperature was then reduced to 24 °C and cells were induced with 0.1 mM IPTG for 12-16 hours. Cells were harvested through centrifugation (15 minutes at 10,000 g, 4 °C) and the pellet was resuspended in 1/10 culture volume of cold Spheroplast solution [50 mM Tris-Cl (pH = 8.0), 20% sucrose, 1 mM EDTA] and shaken gently for one hour on ice. After centrifugation (45 minutes at 15,000 g, 4 °C), the supernatant was removed and passed through a 0.22 µm filter. The remaining pellet was resuspended in 1/10 culture volume of cold 5 mM MgSO₄ and shaken gently for one hour on ice. After another centrifugation step (45 minutes at 15,000 g, 4 °C), the supernatant was again removed and passed through a 0.22 µm filter. Both solutions used for periplasmic extraction contained cOmplete[®] EDTA-free protease inhibitor cocktail tablets (Roche, USA), according to the manufacturer's instructions. At least 10 L of bacterial culture was grown each time recombinant BabA purification was performed.

3.2.3.2 Immobilised metal ion affinity chromatography

The periplasmic extracts obtained from the previous step were combined and loaded onto a column containing EDTA-resistant Ni²⁺ sepharose excel resin (GE Healthcare, USA) at a rate no greater than 2 mL/min. Using an ÄKTA purifier system (GE Healthcare, USA), a gradient elution program was run at a flow rate of 2 mL/min. The wash buffer consisted of 20 mM Tris-Cl (pH=7.4)

and 300 mM NaCl. The elution buffer consisted of 20 mM Tris-Cl (pH=7.4), 300 mM NaCl and 500 mM imidazole. Protein containing elution fractions were loaded onto a Tris-Glycine gel for SDS-PAGE and Coomassie-staining.

3.2.3.3 Size-exclusion chromatography

Immobilised metal ion affinity chromatography (IMAC) elution fractions were concentrated to no more than 5 mL using a Vivaspin sample concentrator (GE Healthcare, USA), according to the manufacturer's instructions. The concentrated samples were loaded onto a HiLoad 16/60 Superdex 75 (120 mL) gel filtration column (GE Healthcare, USA) connected to an ÄKTA purifier system (GE Healthcare, USA). The column was pre-equilibrated with a 20 mM Tris-Cl (pH=7.4) and 300 mM NaCl buffer, which was also used for elution at a flow rate of 1 mL/min. Protein containing elution fractions were loaded onto a Tris-Glycine gel for SDS-PAGE and Coomassie-staining. Further protein concentration using Vivaspin sample concentrators was performed if necessary. Molecular weight standards (Bio-Rad, USA) were used for size calibration of the column.

3.2.3.4 Recombinant BabA quantification

Protein concentration for periplasmic extracts, IMAC and size exclusion chromatography (SEC) elutions was determined with a Bradford assay reagent (Bio-Rad, USA), using bovine serum albumin (BSA) (Sigma-Aldrich, USA) as a standard, according to the manufacturer's instructions (174). SDS-PAGE followed by Coomassie-staining and Western blotting, as described above, was performed on IMAC- and SEC- purified fractions to determine purity. ImageJ software (National Institute of Health, USA) was used to quantify purity by analysing band density. The concentration of pure

recombinant BabA was subsequently determined by measuring absorbance at 280 nm using an extinction coefficient of $37290 \text{ M}^{-1}\text{cm}^{-1}$ (calculated from the actual amino acid sequence of BabA_{527K}) on a Nanodrop 2000 (Thermo Scientific, USA).

3.2.3.5 Mass spectrometry

Liquid chromatography–time-of-flight mass spectrometry was used to determine the molecular weights of proteins in IMAC-SEC purified BabA₅₂₇ and BabA_{527K} protein samples. Approximately 5 µg of protein samples were loaded onto an Agilent 1100 Series LC (Agilent Technologies, USA) coupled to a time-of-flight Q-ToF Premier mass spectrometer (Waters, USA), which was equipped with an electron spray ioniser for acquisition in a positive ionisation mode. The mass scan range was set to 500–2000 m/z and MassLynx (Waters, USA) was used to analyse the mass data.

3.2.3.6 Tryptic peptide mapping

Experimental procedures performed by Jon Renshaw, AstraZeneca R&D.

Tryptic peptide mapping was performed on gel slices of purified BabA protein bands after SEC. Bands were reduced with 10 mM dithiothreitol for 45 minutes at 65 °C, followed by alkylation of protein thiols by incubation with 50 mM iodoacetamide in the dark at room temperature for 20 minutes. Digestion of the proteins was performed using 250 ng of trypsin for 16 hours at 37 °C. Resulting fragments were analysed by liquid chromatography–quadrupole time-of-flight mass spectrometry. An UltiMate 3000 Nano LC system (Dionex, USA) coupled to a QSTAR Elite System (AB Sciex, USA) – a hybrid quadrupole time-of-flight mass spectrometer fitted with electrospray ionisation for acquisition in a positive ionisation mode – was used. Mascot (Matrix

Science, USA) was used to analyse data.

3.2.3.7 N-terminal protein sequencing

This was outsourced (AltaBiosciences, UK) for determination of the first five amino acids of BabA_{527K} through Edman degradation.

3.3 Results

3.3.1 Cloning a babA gene fragment into expression vectors

To generate BabA expression constructs, a PCR was first used to amplify a fragment of the *babA* gene from the genomic DNA of *H. pylori* J99. This gene fragment, termed *babA*₅₂₇, encoded only the expected N-terminal extracellular domain of BabA, that is, amino acids 1-527. As no structural information for BabA or any other *H. pylori* adhesin was available at this stage of the study (175), exclusion of the conserved C-terminal transmembrane domain was based on bioinformatic investigations. Sequence alignment of BabA with SabA, which is also from the Hop family and was then the only other *H. pylori* adhesin with an identified receptor, revealed a distinctly high similarity in their C-terminal regions between BabA amino acids 530-724 (Figure 3.2).

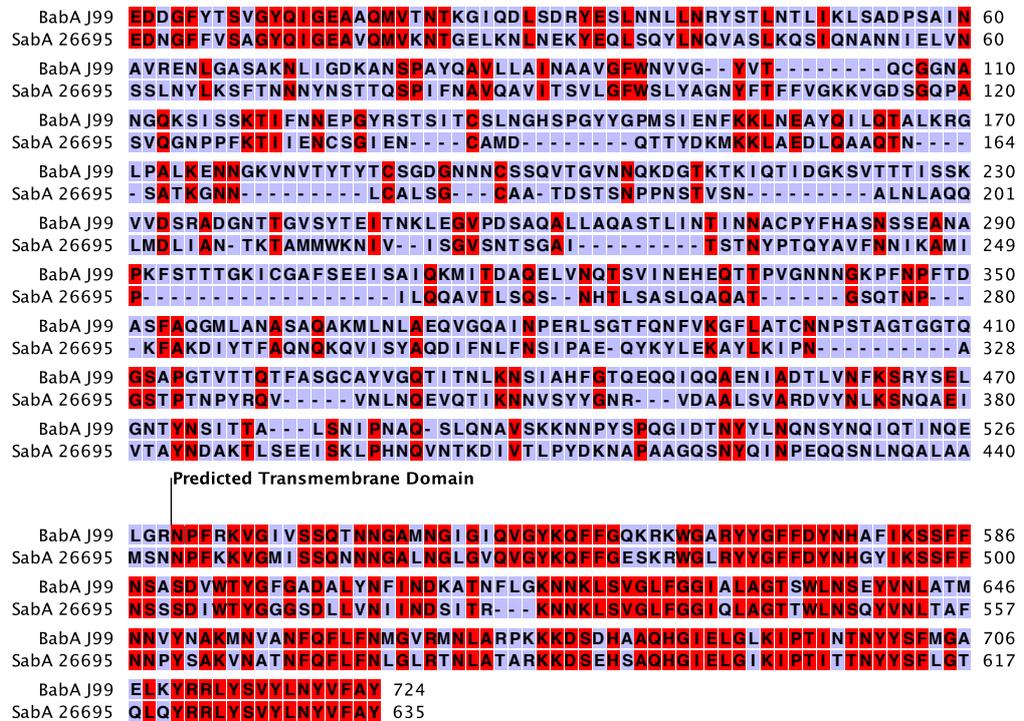


Figure 3.2: Sequence alignment of BabA J99 and SabA 26695

BabA and SabA share 40% sequence identity. The extracellular domains of BabA and SabA share 26% sequence identity and the predicted C-terminal transmembrane domains of BabA and SabA (indicated; commencing at residue 530 of BabA) share

73% sequence identity. Identical amino acids are highlighted by red boxshade and mismatches by blue boxshade.

Thus, the C-terminal truncation for recombinant BabA was arbitrarily chosen between residues 527-528, that is, two residues before what was anticipated as the start of the transmembrane domain. In support of the chosen truncation point, Phyre² predicted that the first out of seven β -strands in the proposed β -barrel transmembrane domain is formed between amino acids 541-560 (Figure 3.3). Consequently, it was assumed that the chosen truncation would not disrupt any ordered regions.

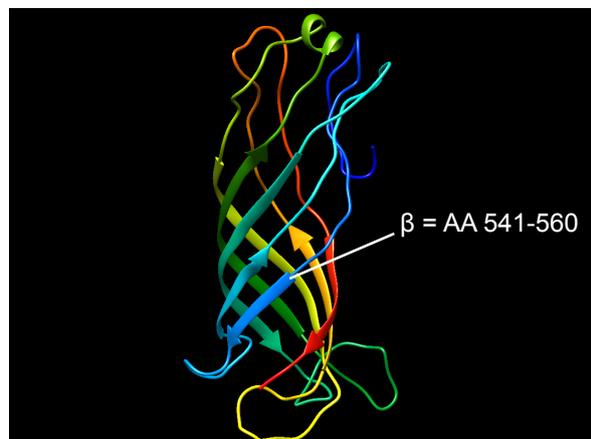


Figure 3.3: Homology modelling of the C-terminal conserved domain of BabA

BabA amino acids 528-724 were input into Phyre². The model structure is rainbow coloured from N- (blue) to C- (red) terminus. The first predicted β strand is indicated. The presented model structure is based on 14 template structures – 12 of which were outer membrane proteins from Gram-negative bacteria. The template protein with the highest confidence (92%) was outer membrane protein W (OmpW) of *E. coli* (PDB accession code: 2F1V).

Two DNA products were obtained after PCR amplification, with the larger product corresponding to the correct amplicon size; this was expected as the functionally uncharacterised, paralogous *H. pylori* outer membrane protein BabB contains identical sequences at its N- and C- termini to BabA (82, 86). Consequently, gel extraction was employed, before clean-up, to produce pure

*babA*₅₂₇ (Figure 3.4). Importantly, the primers used to amplify *babA*₅₂₇ introduced restriction sites that would enable its insertion into the multiple cloning site of a specific expression vector.

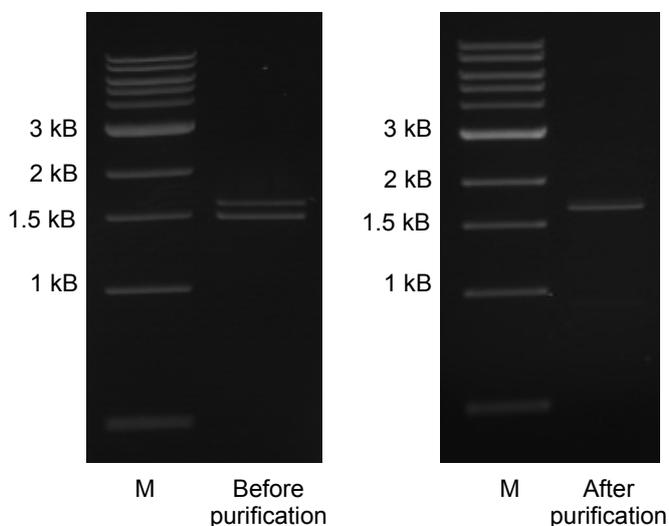


Figure 3.4: Representative amplification of the *babA*₅₂₇ gene fragment

The *babA*₅₂₇ gene fragment, amplified from *H. pylori* J99 genomic DNA, before (left) and after (right) purification on an agarose gel. Purification refers to band excision after agarose gel electrophoresis followed by a PCR clean-up step. The expected size of this PCR product, which contains NcoI and BamHI restriction sites and six-nucleotide overhangs at its 5' and 3' ends, is 1,605 bp. M – DNA ladder.

Purified *babA*₅₂₇ and recipient vectors were then restriction-digested according to the desired location of *babA*₅₂₇ in the expression cassette of each construct (the design of each construct is later presented in Section 3.2.2). After ligation and transformation into *E. coli* for plasmid propagation, colony PCR was used to detect whether transformants hosted plasmids containing *babA*₅₂₇. The primers used in these reactions were designed to span each recipient vector's multiple cloning site, thus, amplicon sizes revealed whether colonies were negative or positive for *babA*₅₂₇ insertion (Figure 3.5).

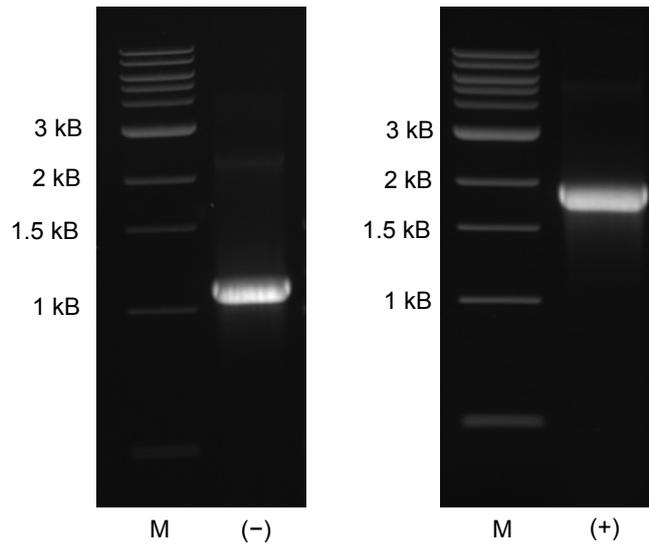


Figure 3.5: Representative verification of *babA*₅₂₇ insertion into an expression vector

Agarose gel of PCR-amplified multiple cloning sites from two separate bacterial colonies hosting either *babA*₅₂₇-negative [(-); left] or *babA*₅₂₇-positive [(+); right] pOPE101 constructs. The expected size of the pOPE101 multiple cloning site is 1128 bp, while successful *babA*₅₂₇ insertion results in a size of 1956 bp. M – DNA ladder.

Finally, DNA sequencing was used to confirm that plasmids extracted from positive transformants had 100% sequence identity to *babA*₅₂₇. DNA sequencing also confirmed *babA*₅₂₇ was cloned in frame with the *pelB* leader (Figure 3.6).

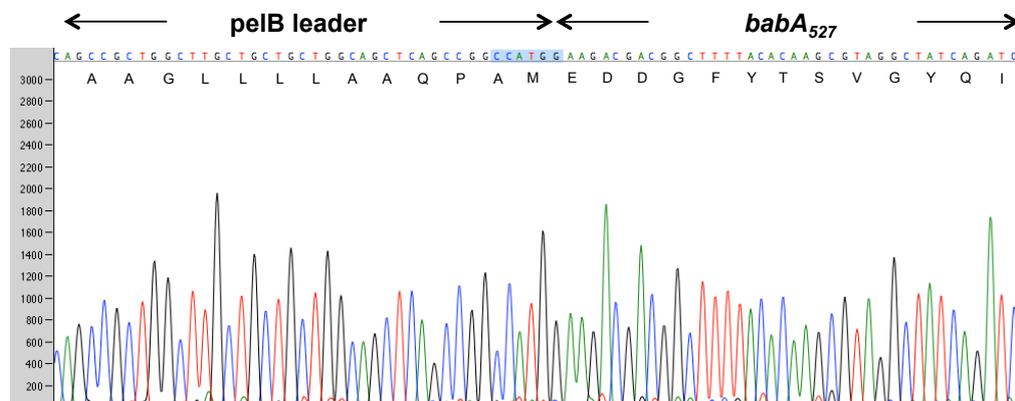


Figure 3.6: Representative DNA sequencing of a BabA expression construct

Chromatogram trace of the pOPE101_*babA*₅₂₇ construct sequenced after extraction from *E. coli* XL10 Gold cells. The six nucleotides highlighted in blue represent the NcoI restriction site. To its left is the *pelB* leader sequence and to its right is the

*babA*₅₂₇ gene fragment. Underneath the DNA sequence is the corresponding amino acid sequence, showing translation of the in-frame *babA*₅₂₇ gene fragment.

Through this process, the *babA*₅₂₇ gene fragment was successfully cloned into a number of commercial vectors designed for protein expression in *E. coli*. On the other hand, cloning of *babA*₅₂₇ into proprietary AstraZeneca R&D vectors for expression in eukaryotic hosts was outsourced. Where necessary, each expression construct was subsequently transformed or transfected into the expression host organism for which it had been engineered.

3.3.2 Recombinant expression in prokaryotic and eukaryotic hosts

In order to swiftly identify an efficient expression strategy that yielded recombinant BabA, various host organisms containing BabA expression constructs were grown on a small scale and screened for soluble protein expression.

3.3.2.1 Periplasmic expression in *E. coli* BL21 (DE3)

For the first attempt, the pET22b_*babA*₅₂₇ construct was transformed into *E. coli* BL21 (DE3) cells. This construct contained the *babA*₅₂₇ gene behind a cleavable pelB leader sequence for periplasmic secretion, as shown below in Figure 3.7.

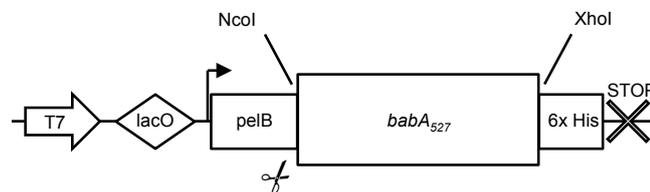


Figure 3.7: Schematic diagram of the pET22b_*babA*₅₂₇ expression cassette

*babA*₅₂₇ was cloned into the pET22b(+) vector using NcoI and XhoI restriction enzyme sites. The abbreviations and symbols represent the following: T7 – T7 promoter; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence; *babA*₅₂₇ – *babA*₅₂₇ gene fragment; 6x His – 6-histidine peptide tag; X/STOP – stop codon.

Transformed cells were induced with IPTG under a variety of conditions and cellular fractions were analysed via Coomassie-staining. Disappointingly, after expression at 24 °C, no recombinant BabA was found in the soluble fraction of cell lysates, which contain both soluble cytoplasmic and periplasmic proteins (Figure 3.8A). This also held true after expression at 37 °C (Figure 3.8B). As a result, recombinant BabA expression in *E. coli* BL21 (DE3) cells was regarded as unsuccessful.

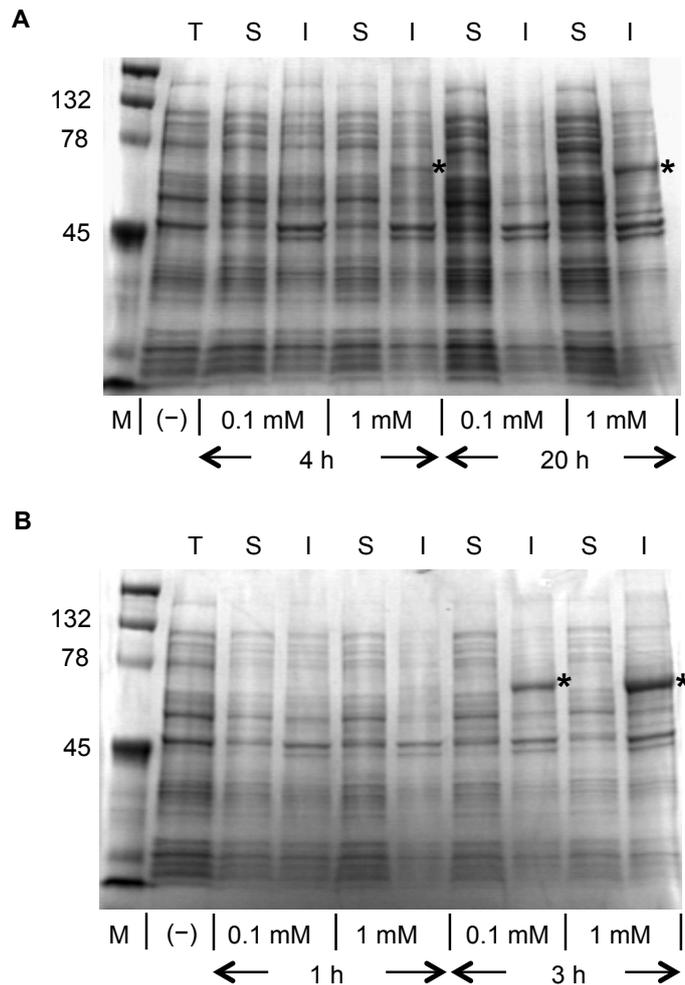


Figure 3.8: Expression of recombinant BabA in *E. coli* BL21 (DE3)

Coomassie-staining showing recombinant BabA localisation in the insoluble fraction of *E. coli* BL21 (DE3) cells after expression at **(A)** 24 °C and **(B)** 37 °C. Putative recombinant BabA proteins (expected molecular weight ~57 kDa) are indicated by black stars. The abbreviations and symbols represent the following: T – total protein fraction; S – soluble protein fraction; I – insoluble protein fraction; M – molecular weight marker; (-) – uninduced sample; 0.1 mM & 1 mM – IPTG induction concentration; 1 h, 3 h, 4 h & 20 h – induction duration in hours.

3.3.2.2 Cytoplasmic expression in *E. coli* Origami B (DE3)

Next, the pET22b(Δ pelB)_{*babA*}₅₂₇ construct was transformed into *E. coli* Origami B (DE3) cells to assess recombinant BabA expression. This construct contained *babA*₅₂₇ without a preceding periplasmic signal sequence, as shown below in Figure 3.9, thereby limiting expression to the cytoplasmic space.



Figure 3.9: Schematic diagram of the pET22b(Δ pelB)_{*babA*}₅₂₇ expression cassette

*babA*₅₂₇ was cloned into the pET22b(+) vector using NdeI and XhoI restriction enzyme sites. The abbreviations and symbols represent the following: T7 – T7 promoter; lacO – lac operator; arrow – start codon; *babA*₅₂₇ – *babA*₅₂₇ gene fragment; 6x His – 6-histidine peptide tag; X/STOP – stop codon.

Again, cells were induced with IPTG under a variety of conditions and cellular fractions were analysed through Coomassie-staining. Recombinant BabA was only found in the insoluble fraction of bacterial lysates, after expression at 24 °C (Figure 3.10A). This was also the case after expression at 37 °C (Figure 3.10B). Thus, this strategy was regarded as another unsuccessful attempt of the expression screen.

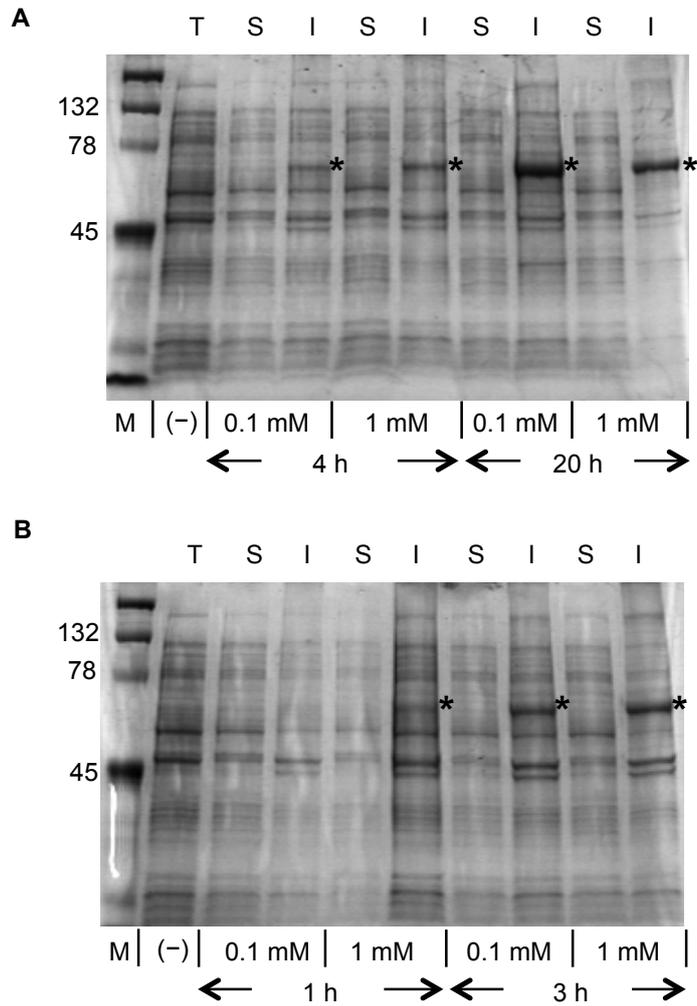


Figure 3.10: Expression of recombinant BabA in *E. coli* Origami B (DE3)

Coomassie-staining showing recombinant BabA localisation in the insoluble fraction of *E. coli* Origami B (DE3) cells after expression at **(A)** 24 °C and **(B)** 37 °C. Putative recombinant BabA proteins (expected molecular weight ~57 kDa) are indicated by black stars. The abbreviations and symbols represent the following: T – total protein fraction; S – soluble protein fraction; I – insoluble protein fraction; M – molecular weight marker; (-) – uninduced sample; 0.1 mM & 1 mM – IPTG induction concentration; 1 h, 3 h, 4 h & 20 h – induction duration in hours.

3.3.2.3 Periplasmic expression in *E. coli* XL10 Gold

Based on a successful expression strategy reported by Fei *et al.* (112), the pOPE101_ *babA*₅₂₇ construct was thereafter assessed for expression in *E. coli* XL10 Gold cells. This construct contains the *babA*₅₂₇ gene fragment behind a synthetic P/A1/04/03 lac promoter and a pelB leader (Figure 3.11).

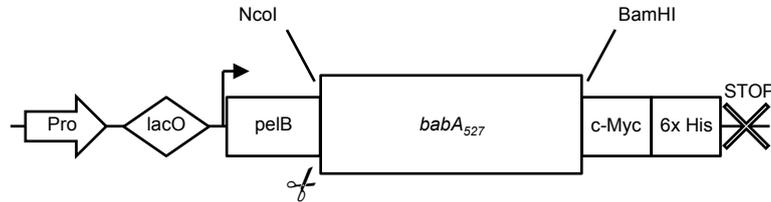


Figure 3.11: Schematic diagram of the pOPE101_ *babA*₅₂₇ expression cassette

*babA*₅₂₇ was cloned into the pOPE101 vector using NcoI and BamHI restriction enzyme sites. The abbreviations and symbols represent the following: Pro – synthetic P/A1/04/03 lac promoter; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence; *babA*₅₂₇ – *babA*₅₂₇ gene fragment; c-Myc – c-Myc tag; 6x His – 6-histidine peptide tag; X/STOP – stop codon.

Analysis of the cellular fractions, through Coomassie-staining and c-Myc immunodetection, indicated that recombinant BabA was successfully expressed and secreted to the periplasmic space after overnight induction with IPTG at 24 °C (Figure 3.12). This recombinant BabA protein is henceforth to as referred BabA₅₂₇.

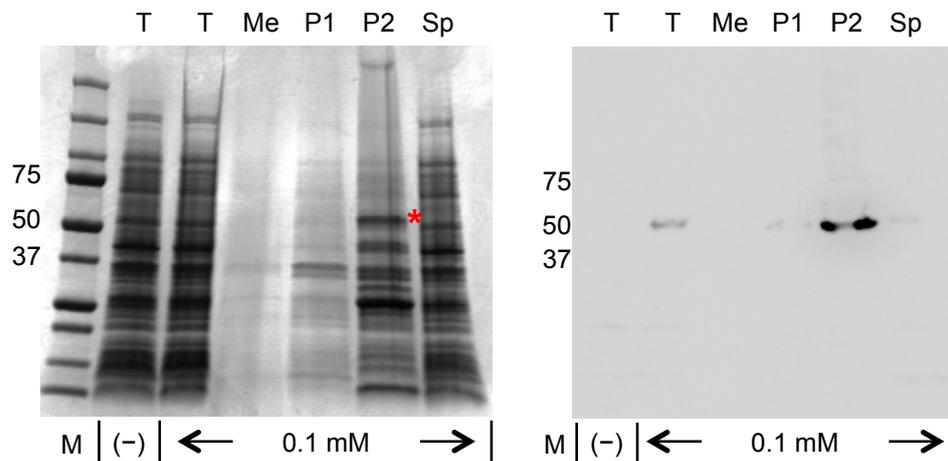


Figure 3.12: Expression of recombinant BabA in *E. coli* XL10 Gold

Coomassie-staining (left) and c-Myc immunodetection (right) showing BabA₅₂₇ localisation in periplasmic fractions after overnight expression in *E. coli* XL10 Gold cells at 24 °C. BabA₅₂₇ (expected molecular weight ~58 kDa) is indicated by the red star. The abbreviations and symbols represent the following: T – total protein fraction; Me – culture medium protein fraction; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; Sp – spheroplast protein fraction; M – molecular weight marker; (-) – uninduced sample; 0.1 mM – IPTG induction concentration.

In an attempt to further verify that the protein identified through c-Myc immunodetection was indeed recombinant BabA, separate Western blots were performed on the aforementioned periplasmic extracts using antibodies directed at both the c-Myc and the 6x His polypeptide tags of BabA₅₂₇. However, as shown in Figure 3.13, in contrast to c-Myc immunodetection, anti-His immunodetection resulted in non-specificity – the same bands, of an incorrect size, appeared both before and after IPTG induction. Thus, as anti-BabA antibodies are not commercially available, c-Myc recognition was the sole form of immunodetection used to verify successful recombinant BabA expression at this stage.

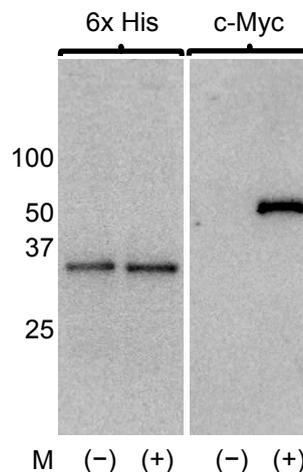


Figure 3.13: Comparison of BabA₅₂₇ immunodetection methods

(Left) 6x His and (Right) c-Myc immunodetection of BabA₅₂₇ (~58 kDa) in the periplasmic extracts of *E. coli* XL10 Gold cells before (-) and after (+) overnight 0.1 mM IPTG induction at 24 °C. M – molecular weight marker.

Because a successful expression strategy was obtained, consideration was given to the downstream applications of recombinant BabA. With a view towards potentially facilitating its conjugation to carboxylate particle surfaces, through carbodiimide coupling reactions, a hexa-lysine tag was incorporated into pOPE101_ *babA*₅₂₇ between *babA*₅₂₇ and the c-Myc tag (Figure 3.14).

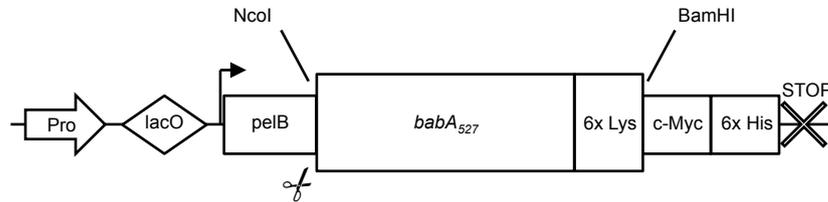


Figure 3.14: Schematic diagram of the pOPE101_ *babA*_{527K} expression cassette
*babA*_{527K} was cloned into the pOPE101 vector using NcoI and BamHI restriction enzyme sites. The abbreviations and symbols represent the following: Pro – synthetic P/A1/04/03 lac promoter; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence; *babA*₅₂₇ – *babA*₅₂₇ gene fragment; 6x Lys – 6-lysine peptide tag; c-Myc – c-Myc tag; 6x His – 6-histidine peptide tag; X/STOP – stop codon.

Again, both Coomassie-staining and c-Myc immunodetection showed that the hexa-lysine tagged, recombinant BabA protein was successfully expressed and secreted to the periplasmic space after overnight IPTG induction (Figure 3.15). This recombinant BabA protein is henceforth to as referred BabA_{527K}.

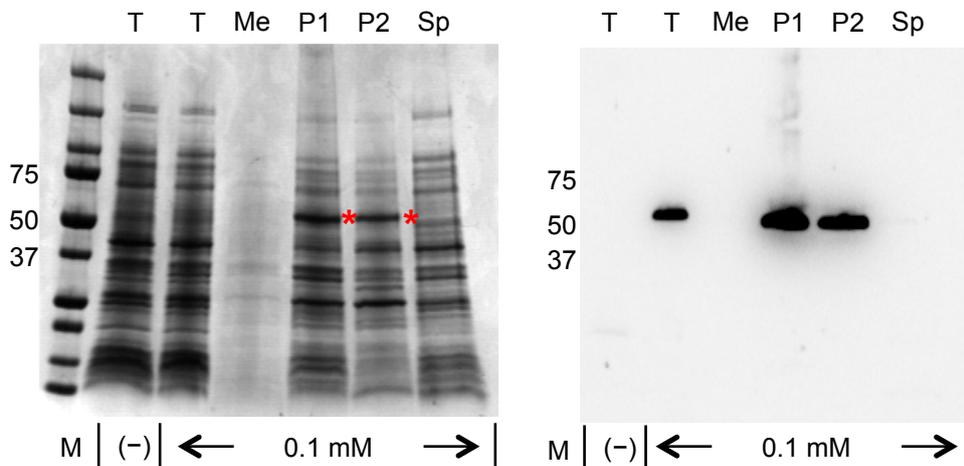


Figure 3.15: Expression of recombinant BabA in *E. coli* XL10 Gold after the

incorporation of a C-terminal hexa-lysine tag

Coomassie-staining (left) and c-Myc immunodetection (right) showing BabA_{527K} localisation in the periplasmic fraction after expression in *E. coli* XL10 Gold cells at 24 °C. BabA_{527K} is indicated by the red stars (expected molecular weight ~59 kDa). The abbreviations and symbols represent the following: T – total protein fraction; Me – culture medium protein fraction; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; Sp – spheroplast protein fraction; M – molecular weight marker; (-) – uninduced sample; 0.1 mM – IPTG induction concentration.

Interestingly, BabA_{527K} protein bands in Figure 3.15 appeared to be more pronounced than BabA₅₂₇ bands in Figure 3.12. Thus, to ascertain whether this was an artefact or reflected a difference in expression levels, BabA₅₂₇ and BabA_{527K} were expressed in parallel under identical conditions. As shown in Figure 3.16, after expression, greater amounts of BabA_{527K} were indeed found in periplasmic extracts than BabA₅₂₇. Nonetheless, both BabA₅₂₇ and BabA_{527K} were used for large-scale expression and purification, as later described in Section 3.3.3.

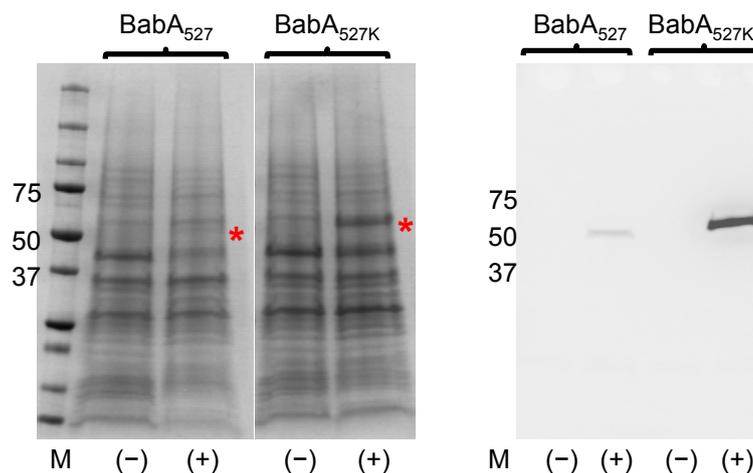


Figure 3.16: Comparison of BabA₅₂₇ and BabA_{527K} periplasmic expression in *E. coli* XL10 Gold

Coomassie-staining (left) and c-Myc immunodetection (right) of periplasmic extracts from *E. coli* XL10 Gold cells expressing BabA₅₂₇ and BabA_{527K} proteins before (-) and after (+) induction with 0.1 mM IPTG at 24 °C for 16 hours. The recombinant BabA proteins are indicated by red stars. M – molecular weight marker.

3.3.2.4 Secretory expression in *S. frugiperda* Sf21

In search of other potential expression hosts, *S. frugiperda* Sf21 cells were investigated using the baculovirus expression system. As shown in Figure 3.17, the expression construct employed, AZ1_*babA*_{527K}, contained a cleavable N-terminal extracellular secretory signal sequence preceding *babA*₅₂₇ in the expression cassette of a proprietary AstraZeneca R&D vector. It also contained the C-terminal polypeptide tags found in pOPE101_*babA*_{527K}.



Figure 3.17: Schematic diagram of the AZ1_*babA*_{527K} expression cassette

The abbreviations and symbols represent the following: Pro – Undisclosable promoter sequence; arrow – start codon; Secr – Undisclosable cleavable secretory signal sequence; *babA*₅₂₇ – *babA*₅₂₇ gene fragment; 6x Lys – 6-lysine peptide tag; c-Myc – c-Myc tag; 6x His – 6-histidine peptide tag; X/STOP – stop codon; A – SV40 polyadenylation signal.

Coomassie-staining of the culture medium of infected *S. frugiperda* cells revealed two proteins potentially correlating to BabA_{527K} (Figure 3.18). However, these proteins did not contain a detectable c-Myc epitope and, as such, were not confirmed to be recombinant BabA.

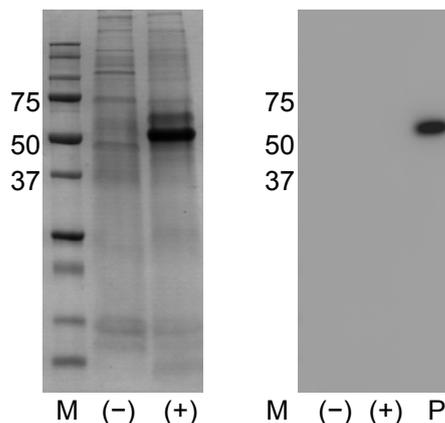


Figure 3.18: Expression of recombinant BabA in *S. frugiperda* Sf21

Coomassie-staining (left) and c-Myc immunodetection (right) showing recombinant BabA is not identifiable in the culture medium of *S. frugiperda* Sf21 cells. The (-) symbol represents non-infected cells while (+) represents cells infected with the AZ1_*babA*_{527K} construct. The expected molecular weight of recombinant BabA is ~59 kDa. M – molecular weight marker; P – periplasmic extract of *E. coli* XL10 Gold cells expressing BabA_{527K} as a Western blot positive control. Samples were taken 72 hours from point of infection.

3.3.2.5 Secretory expression in HEK 293-6E and CHO G22

For the final attempt of the expression screen, the AZ2_*babA*_{527K} construct was transfected into both HEK 293-6E and CHO G22 cells for transient recombinant BabA expression. The expression cassette of this construct closely resembled the AZ1_*babA*_{527K} vector, as shown in Figure 3.19.



Figure 3.19: Schematic diagram of the AZ2_*babA*_{527K} expression cassette

The abbreviations and symbols represent the following: Pro – Undisclosable promoter sequence; arrow – start codon; Igκ – Immunoglobulin k-chain secretory signal sequence; *babA*₅₂₇ – *babA*₅₂₇ gene fragment; 6x Lys – 6-lysine peptide tag; c-Myc – c-Myc tag; 6x His – 6-histidine peptide tag; X/STOP – stop codon; A – SV40 polyadenylation signal.

Frustratingly, as observed in *S. frugiperda* expression, Western blotting showed that none of the proteins visualised through Coomassie-staining of the culture medium from transfected cells contained a detectable c-Myc epitope to enable the positive identification of recombinant BabA (Figure 3.20).

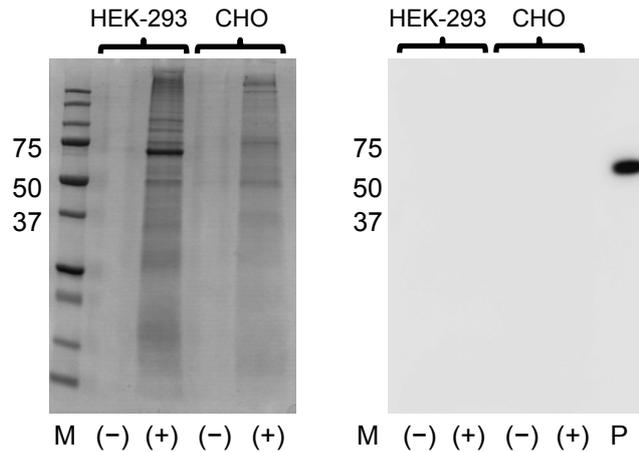


Figure 3.20: Expression of recombinant BabA in HEK 293-6E and CHO G22

Coomassie-staining (left) and c-Myc immunodetection (right) showing recombinant BabA is not identifiable in the culture medium of transfected HEK 293-6E or CHO G22 cells. The (-) symbol represents culture medium before transfection, while (+) represents culture medium after transfection. The expected molecular weight of recombinant BabA is ~59 kDa. M – molecular weight marker. P – periplasmic extract of *E. coli* XL10 Gold cells expressing BabA_{527K} as a Western blot positive control. Samples were taken eight days after transfection.

As the majority of expression strategies attempted were unsuccessful, only periplasmic expression in *E. coli* XL10 Gold cells was performed on a large scale to enable the purification of recombinant BabA.

3.3.3 Purification of recombinant BabA

BabA₅₂₇ and BabA_{527K} proteins were purified from the periplasmic extracts of *E. coli* XL10 Gold cells expressing the pOPE101 based constructs described in Section 3.3.2.3. This was achieved through their hexa-histidine tags via immobilised metal ion affinity chromatography (IMAC) as an initial purification step. Following imidazole elution, visible aggregates were observed in the BabA₅₂₇ but not BabA_{527K} fractions. These aggregates were removed via centrifugation and as such were excluded from subsequent analysis and processing. BabA₅₂₇ and BabA_{527K} IMAC fractions were then separated through SDS-PAGE and analysed via Coomassie-staining – a higher yield of BabA_{527K} was immediately apparent (Figure 3.21). Furthermore, an additional, lower molecular weight band was found in BabA₅₂₇ IMAC fractions, but not in BabA_{527K} IMAC fractions.

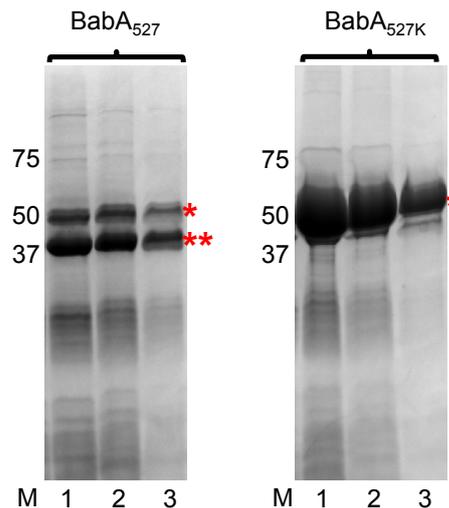


Figure 3.21: IMAC purification of BabA₅₂₇ and BabA_{527K}

BabA₅₂₇ (left) and BabA_{527K} (right) proteins from 3 consecutive IMAC-elution fractions were collected, separated via SDS-PAGE and analysed by Coomassie-staining (lanes 1-3). Gel loading volumes were equal, as were bacterial culture and periplasmic extract volumes used in IMAC purification. One red star indicates recombinant BabA. The two red stars indicate an additional, lower molecular weight band. M – molecular weight marker.

Size exclusion chromatography (SEC) was employed as a second and final purification step. A comparison of BabA₅₂₇ and BabA_{527K} SEC traces with molecular weight standards (Appendix Table 1) indicated that the elution peaks between 60 and 70 mL equate to a monomeric recombinant BabA protein (Figure 3.22). Greater relative amounts of aggregates (which would therefore elute before 60 mL) are observed in the BabA₅₂₇ trace than in the BabA_{527K} trace. The large peak shoulder in the BabA₅₂₇ trace (between 70 to 80 mL), which is largely absent in the BabA_{527K} trace, represents the lower molecular weight band observed through Coomassie-staining after IMAC. Accordingly, only BabA₅₂₇ and BabA_{527K} peak fractions between 60 and 70 mL were selected for further use and analysis.

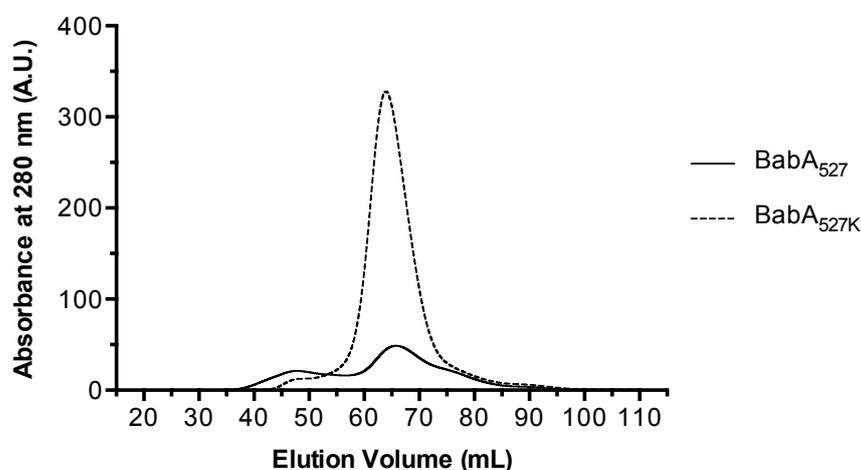


Figure 3.22: SEC purification of BabA₅₂₇ and BabA_{527K}

Overlay of SEC traces of BabA₅₂₇ (solid line) and BabA_{527K} (dashed line) after IMAC purification.

Table 3.5 below shows protein quantification after each purification step. It can be observed here, in comparison to BabA_{527K}, that the amount, or yield, of BabA₅₂₇ decreases markedly after each IMAC and SEC purification step; this poor recovery is attributed to the loss of BabA₅₂₇ due to protein aggregation, as described above. The final yield, from bacterial culture, of IMAC-SEC

purified BabA_{527K} (1.8 mg/L) was 22.5 times higher than that of BabA₅₂₇ (0.08 mg/L) after identical expression and purification conditions.

Table 3.5: Summary of yield from BabA₅₂₇ and BabA_{527K} purification

^aPeriplasmic extract from 10 L of *E. coli* culture after induction with 0.1 mM IPTG for 16 hours at 24 °C

^bTotal protein determined by Bradford assay (BSA as standard protein)

^cPurity estimated from ImageJ analysis of a scanned Coomassie stained SDS-PAGE gel by measuring target band as a fraction of all bands

Step	<i>BabA</i> ₅₂₇				<i>BabA</i> _{527K}			
	Total Protein (mg) ^b	Target Protein (mg) ^b	Yield (%)	Purity (%) ^c	Total Protein (mg) ^b	Target Protein (mg) ^b	Yield (%)	Purity (%) ^c
Extract ^a	276	18	100	6.4	284	26	100	9.3
IMAC	22	3.3	19	15	31	20	76	65
SEC	2.4	0.8	4.4	33	18	18	67	99

Coomassie-staining, as indicated in Table 3.5, and mass spectrometry were both used to analyse the purity of IMAC-SEC purified BabA₅₂₇ and BabA_{527K}. Coomassie-staining revealed that the additional lower molecular weight band previously seen in BabA₅₂₇ IMAC elutions was still found after SEC. The corresponding mass spectrometry trace further indicated that this band actually consisted of multiple proteins of similar molecular weight (Figure 3.23).

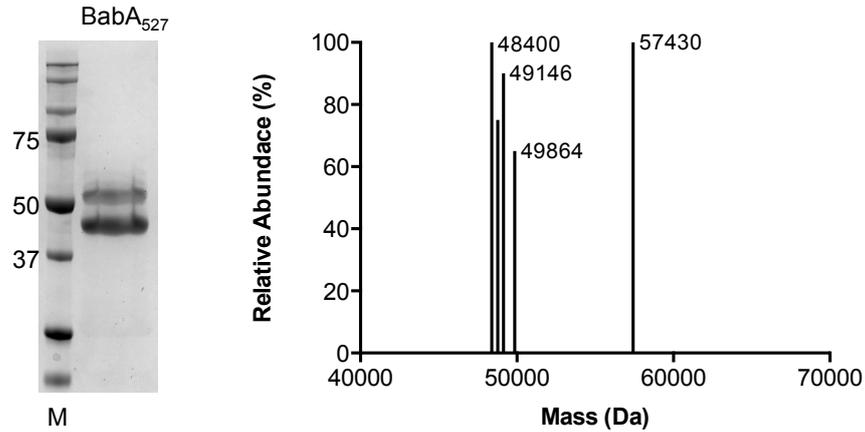


Figure 3.23: Analysis of BabA₅₂₇ purity

(Left) Coomassie stained SDS-PAGE gel and (right) liquid chromatography–time-of-flight mass spectroscopy trace of IMAC-SEC purified BabA₅₂₇. The expected molecular weight of BabA₅₂₇ is 58,444 Da. M – molecular weight marker.

To determine the identity of these lower molecular weight proteins, which resemble degradative fragments, peptide mapping of an excised gel slice was performed. Figure 3.24 shows the mapped peptides, generated through tryptic digestion, contained 45% of BabA₅₂₇ residues. No amino acids after K392 were matched.

```

1      . . . . .GYQIGEEAQMVTNTKGIQDLSDRYESLNLLNRYSTLNTLI
51     KLSADPSAINAVRENLGASAKNLIGDKANSPAYQAVLLAINAAVGFWNVV
101    GYVTQCGGNANGQKSISSKTIFNNEPGYRSTSITCSLNHGSPGYGPMSE
151    ENFKKLNEAYQILQTALKRGLPALKENNGKVNVTYTYTCSGDGNNNCSSQ
201    VTGVNNQKDGTKTKIQTIDGKSVTTTISSKVVDSSRADGNTTGVSYTEITN
251    KLEGVPDSAQALLAQASTLINTINNACPYFHASNSSEANAPKFSTTTGKI
301    CGAFSEEISAIQKMITDAQELVNQTSVINEHEQTPVGNNGKPFNPFTD
351    ASFAQGMLANASAQAKMLNLAEQVGQAINPERLSGTFQNFVKGFLATCNN
401    PSTAGTGGTQGSAPGTVTTQTFASGCAYVGQTITNLKNSIAHFGTQEQQI
451    QQAENIADTLVNFKSRYSELGNTYNSITTALSNIPNAQSLQNAVSKKNNP
501    YSPQGIDTNYLNQNSYNQIQTINQELGSEQKLISEEDLSHHHHH
    
```

Figure 3.24: BabA₅₂₇ amino acid sequence showing peptides matched in additional lower molecular weight band

Peptides shown in bold were identified through mass matching, after liquid chromatography–quadrupole time-of-flight mass spectrometry. Trypsin cleavage sites [C-terminal side of lysine (K) or arginine (R) unless next residue is proline (P)] are highlighted in red unless part of a matched peptide. The black dots at the N-terminus represent the first nine amino acids of BabA₅₂₇, which are cleaved during expression.

In stark contrast, IMAC-SEC purified BabA_{527K} was found to be highly pure (Figure 3.25). Interestingly, both BabA₅₂₇ and BabA_{527K} are 1,014 Da smaller than their expected molecular weights. This would correspond to a loss of the first 9 amino acids (EDDGFYTSV) from the N-terminus of their predicted amino acid sequences. Indeed, Edman degradation of pure BabA_{527K} confirmed that its first five N-terminal amino acids correspond to residues 10 to 14 of mature BabA (GYQIG).

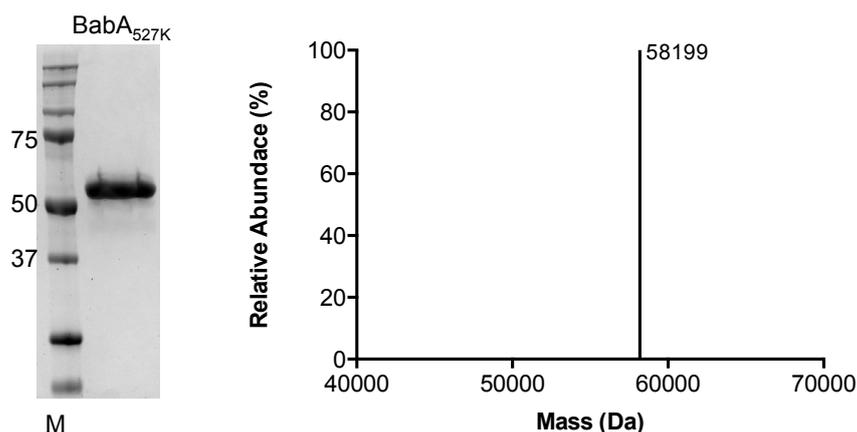


Figure 3.25: Analysis of BabA_{527K} purity

(Left) Coomassie stained SDS-PAGE gel and (right) liquid chromatography–time-of-flight mass spectrometry trace of IMAC-SEC purified BabA_{527K}. The expected molecular weight of BabA_{527K} is 59,213 Da. M – molecular weight marker.

Taken together, these results show that, somewhat surprisingly, the simple incorporation of a hexa-lysine tag to the C-terminus of recombinant BabA had a substantially beneficial effect on protein yield and purity. Both the yield and the purity of BabA_{527K} were reproducible over the entire course of this study.

3.4 Discussion

The first objective of this study was to establish an efficient method for the expression and purification of recombinant BabA. Consideration was given to the design of the *babA* gene fragment used for expression screening to ensure that the hydrophobic C-terminal transmembrane domain of BabA, which was superfluous with regards to the aim of this study, was excluded. This was also considered to be particularly important to avert potential recombinant protein insolubility during expression and purification (176). While the predicted β -barrel conformation of the BabA transmembrane domain has been long described in the literature (86), there has been no consensus on the exact amino acid sequence that corresponds to this domain because no structural information has been available. In this study, a single C-terminal truncation point at residue 527 was chosen based on bioinformatic investigations; however, various researchers have also reported successful expression after different C-terminal truncation points both upstream and downstream of residue 527 (112, 164, 177).

The first attempt to express recombinant BabA was in *E. coli* BL21 (DE3) cells using a vector containing a *pelB* leader signal sequence to direct synthesised protein to the periplasmic space. The periplasmic space is an aqueous oxidising environment containing enzymes that aid disulphide bond formation (e.g. DsbA and DsbC) and a signal peptidase that cleaves off the *pelB* leader (178). As *babA*₅₂₇ does not encode the hydrophobic domain of native BabA, an insoluble protein was not expected yet was observed. This may have been caused by saturation of the Sec complex, the machinery that mediates periplasmic secretion, if protein synthesis was too rapid (171). This is plausible as the expression strategy used, that is, a pET vector containing the T7 promoter in a DE3 lysogen strain, is designed for rapid overexpression due to

the powerful action of T7 RNA polymerase (179, 180). Saturation of the Sec complex would result in the localisation of recombinant BabA to the reducing environment of the cytoplasm where disulphide bond formation is not possible. Consequently, large amounts of expressed protein would be improperly folded and would ultimately aggregate to form inclusion bodies in the cytoplasm (181).

Attempts to facilitate periplasmic secretion in *E. coli* BL21 (DE3) cells, by reducing the rate of protein synthesis through changes to IPTG concentration and temperature, were unsuccessful (166). Thus, to circumvent the potential problem of Sec complex saturation, expression in *E. coli* Origami B (DE3) cells was attempted. This strain lacks thioredoxin reductase and glutathione reductase enzymes and thereby possesses an oxidising cytoplasmic environment, which reportedly aids intramolecular disulphide bonds formation (182-184). Disappointingly, inclusion body formation also occurred using this strategy despite attempts to support correct folding by similarly reducing the rate of protein synthesis. This suggests that oxidising conditions alone are not sufficient to enable the correct formation of the disulphide bridges in BabA. It must be noted that due to the poor specificity of anti-His immunodetection observed in this study, Western blotting could not be used to verify whether either strategy produced low levels of soluble protein that may have been undetectable through Coomassie-staining.

According to Fei *et al.*, a recombinant BabA protein could be expressed in the periplasmic space of *E. coli* XL10 Gold cells (a strain normally used for plasmid propagation) using the pOPE101 expression vector, which contains a pelB leader, for periplasmic secretion, behind a synthetic lac promoter associated with slower protein synthesis than the pET system (112, 185, 186). When employed in this study, this approach was successful for the expression

and periplasmic secretion of two soluble recombinant BabA proteins, which differed solely by the presence of a C-terminal hexa-lysine tag. Although a successful expression strategy was established, it is well known that one major limitation of periplasmic expression is low yield due to its restricted volume and the limited capacity of the Sec complex (179, 187). Secretory expression in eukaryotic cells, on the other hand, typically provides higher yields, in addition to enabling disulphide bond formation. Due to this advantage, the baculovirus expression system in insect cells [*S. frugiperda* Sf21 (130, 188)] and transient gene expression in mammalian cells [HEK 293-6E and CHO G22 (131, 132, 188-190)] were assessed. The expression constructs used in these eukaryotic systems contained cleavable secretory signal sequences to enable translocation of synthesised protein to the endoplasmic reticulum, for post-translational modifications, and secretion into the extracellular medium for subsequent purification (191, 192). However, after expression, recombinant BabA could not be identified in the extracellular medium of any of these cells through immunodetection of its c-Myc polypeptide tag, though Coomassie-staining revealed “expressed” proteins of the expected size. Immunodetection would have ideally been repeated using an anti-BabA antibody to discount potential issues of c-Myc tag recognition, however, no such antibodies are commercially available. It is possible though that successful recombinant protein secretion into the extracellular medium may have been impaired by glycosylation. While glycosylation in BabA has been suggested [Champasa *et al.* putatively identified BabA as one of 125 *H. pylori* proteins to undergo glycosylation based on the mass spectrometric identification of a glycosylated protein that possessed a 2.8% sequence match to BabA of strain 26695 (193)], such a post-translational modification has neither been verified nor characterised. Accordingly, glycosylation (or atypical glycosylation) during expression may have resulted in protein misfolding,

which typically results in intracellular retention and degradation (194, 195). This could be considered a likely occurrence as recombinant BabA contains seven potential *N*-linked glycosylation sites [based on the N-X-S/T consensus sequence, where X is not P (196)]. The same eukaryotic expression methods described in this thesis were subsequently used to successfully generate AstraZeneca R&D in-house mammalian proteins as positive controls for expression (results not shown). Thus, in the interests of progressing to the next objective of the study, the expression of recombinant BabA by these systems was not pursued further and troubleshooting options were not explored. This would have primarily included peptide mapping of the potential recombinant BabA bands identified through Coomassie-staining and attempting to detect the c-Myc tag of recombinant BabA after lysing cells.

From the expression screen, periplasmic expression in *E. coli* XL10 Gold cells was the only method capable of generating recombinant BabA proteins for purification – BabA₅₂₇ and BabA_{527K}. It was envisaged that the free amine groups in the hexa-lysine tag of BabA_{527K} could act as “hotspots” for carbodiimide coupling reactions during protein immobilisation onto model carboxylated polystyrene particles, as similarly reported by Allard *et al.* (197). Such immobilisation would be advantageous as chemical modification of an extraneous, non-native sequence is less likely to affect glycan binding activity than modification of free amine groups within BabA (198). The C-terminus of recombinant BabA was chosen for hexa-lysine tag insertion as it was hoped that linkage at this terminus could also provide a surface presentation of recombinant BabA similar to native BabA molecules, which are attached to the *H. pylori* outer membrane through their C-terminal domain. Meanwhile, it was also observed that after IMAC purification, BabA₅₂₇ lacked solubility and formed visible aggregates. As polycationic amino acid tags have been shown

to increase the solubility of several recombinant proteins (199-202), a further rationale to support the incorporation of the C-terminal hexa-lysine was to improve the solubility, and purification, of recombinant BabA. Indeed, BabA_{527K} visually appeared to be less prone to aggregation than BabA₅₂₇ after IMAC, and during subsequent SEC, fewer aggregates were also observed in BabA_{527K} traces than in BabA₅₂₇ traces. While this suggests an improvement to intrinsic protein solubility, a direct comparison based on these findings cannot be made because of differences in sample purity at both steps. Nonetheless, it can be said that, due to the reduction in protein aggregation effected, the C-terminal hexa-lysine tag resulted in an improvement to the purification and yield of recombinant BabA.

The greater yield of BabA_{527K} over BabA₅₂₇ may be also explained through the presence of proteolytic fragments, truncated from their C-termini, found during the purification of BabA₅₂₇. Such fragments were not found during BabA_{527K} purification, suggesting that the hexa-lysine tag alleviated proteolytic cleavage and consequently resulted in a higher yield of intact BabA_{527K}. It is important to note that the co-purification of BabA₅₂₇ fragments, which lacked their hexa-histidine tags, was unexpected as their binding to the Ni²⁺ column during IMAC is usually unlikely. However, the EDTA resistant Ni²⁺ sepharose resin used in purification required an imidazole-free equilibration buffer – this can result in the non-specific attachment of other proteins (203).

The mechanism and location of BabA₅₂₇ degradation was not confirmed. However, higher levels of intact BabA_{527K} than BabA₅₂₇ were also found in periplasmic extracts after identical expression conditions. This suggests that BabA₅₂₇ degradation occurred during expression, within the periplasmic space of *E. coli*, which is known to host at least 20 proteases with various

functionalities (204). For example, the protease DegP is known to degrade recombinant proteins that are misfolded upon overproduction in order to prevent protein aggregate formation (205). Another periplasmic protease, Tsp, is known to have a preference for unfolded substrates with hydrophobic amino acids at its C-terminus (206). Further experiments would be necessary to confirm whether BabA₅₂₇ is a substrate for these, or other, periplasmic proteases. Nevertheless, through an unknown mechanism, the C-terminal hexa-lysine tag clearly prevented proteolysis. One potential explanation is that the hexa-lysine tag may increase recombinant BabA solubility in the periplasmic space and thereby prevent aggregation and enzymatic attack. An alternative explanation could be that the C-terminal hexa-lysine tag interacts with other amino acids to protect susceptible cleavage sites. Ultimately, the protective effect of the hexa-lysine tag is considered an unexplained but fortuitous finding.

While the hexa-lysine tag protects against C-terminal proteolytic attack, it was consistently observed that both BabA₅₂₇ and BabA_{527K} lacked the first nine N-terminal amino acids of their predicted sequence. Interestingly, Subedi *et al.* have also recently reported the loss of the first nine N-terminal amino acids of a recombinant BabA protein expressed in the periplasmic space of *E. coli* (177). These combined findings suggest that these residues are prone to spontaneous cleavage by periplasmic proteases. Fortunately, the N-terminal truncation observed in this study was controlled and did not prevent reproducible BabA_{527K} expression and purification.

Thus, the first objective of this study was successfully met and albeit a relatively low yield, highly pure, hexa-lysine tagged recombinant BabA could consistently be generated for use in all further investigations and applications.

Chapter 4: Characterising the BabA glycan-binding site

4.1 Introduction

Having successfully expressed the extracellular domain of BabA, the second objective of this study was to identify and characterise its glycan-binding site. To date, the glycans known to act as BabA receptors are the type 1 ABO and Le blood group antigens, which are naturally found in healthy gastric mucosa, either expressed on surface mucous cells or attached to the MUC5AC mucin that comprises the surface mucus layer (Figure 1.7) (95, 207). As shown in Figure 4.1, their biosynthesis in the gastric mucosa commences with the initial fucosylation of a type 1 lacto series core chain by the fucosyltransferase 2 (FucT-II) enzyme. This generates the H-1 antigen, which is further modified by a GalNAc-transferase in blood group A individuals and a Gal-transferase in blood group B individuals, thereby producing the A-1 and B-1 antigens, respectively. H-1, A-1 and B-1 are ABO blood group antigens that are then further fucosylated to produce Le blood group antigens. This occurs through the action of fucosyltransferase 3 (FucT-III), which results in Le^b, A-Le^b and B-Le^b in blood group O, A and B individuals, respectively (208). However, a ~20% subset of the Western population, known as non-secretors, lack FucT-II and consequently possess none of the aforementioned ABO/Le blood group antigens. Rather, they synthesise the Le^a antigen from the type 1 lacto series core; this glycan is otherwise not typically found in the gastric mucosa of individuals with functional FucT-II enzymes (209). Importantly, the Le^a antigen is the only type 1 ABO/Le blood group antigen that does not act as a receptor for BabA (82). Closely related to type 1 ABO/Le blood group antigens are their type 2 counterparts. These glycans are produced through a similar biosynthetic pathway with the characteristic difference being an initial type 2 lacto series core chain, which is modified by the fucosyltransferase 1 (FucT-I)

enzyme. Type 2 ABO/Le blood group antigens differ from their type 1 counterparts through an alternate glycosidic bond in the Gal-GlcNAc core, that is, type 1 antigens have a Gal β 1-3GlcNAc linkage whereas type 2 antigens have a Gal β 1-4GlcNAc linkage (Figure 4.1). Type 2 ABO/Le blood group antigens are not found in the surface layer of the gastric mucosa; they are located in the deep glands, either attached to mucous neck cells or the MUC6 mucin (Figure 1.7). They do not act as receptors for BabA (82, 95, 101, 102).

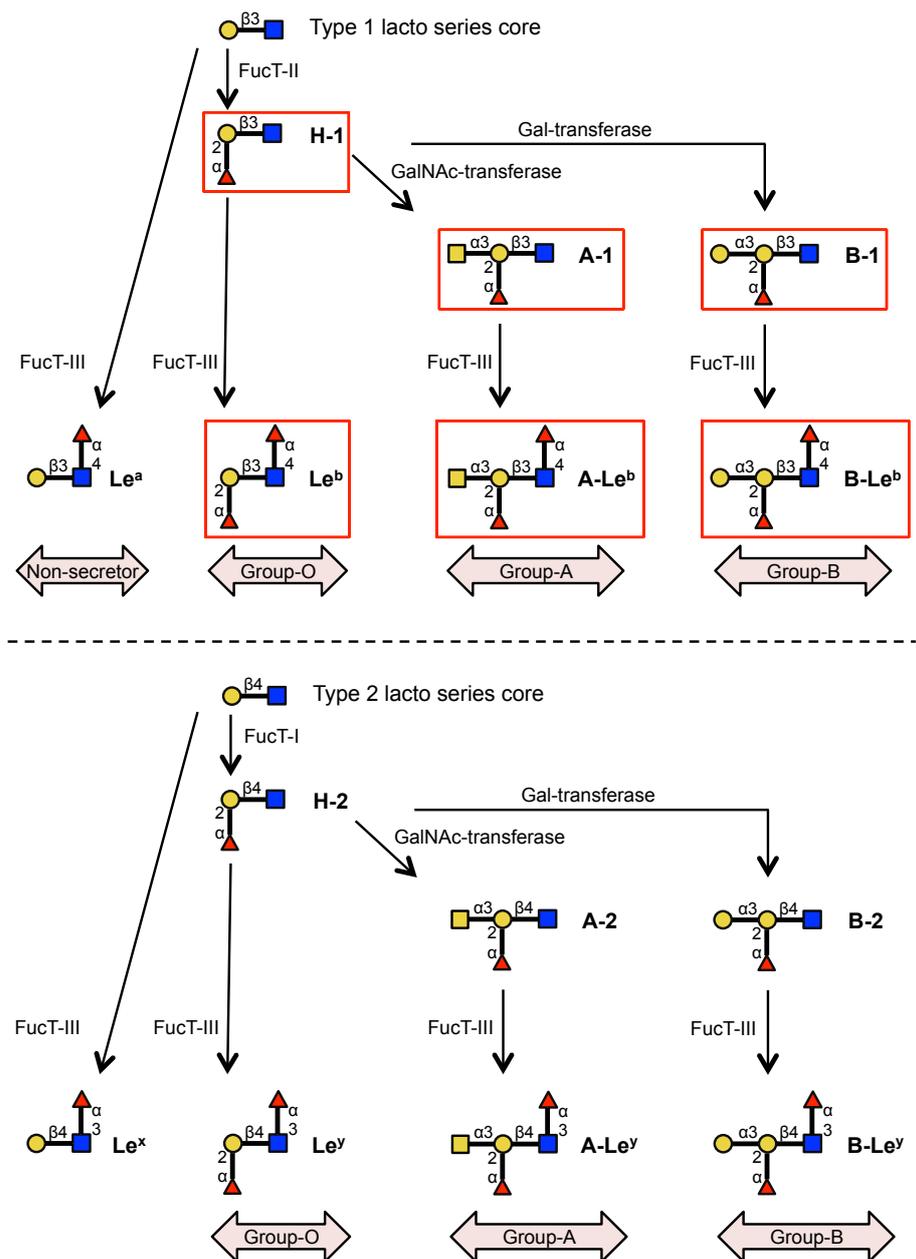


Figure 4.1: The biosynthesis pathways of the dominant ABO/Le blood group

antigens in healthy gastric mucosa

The terminal glycan epitopes of the type 1 (top) and type 2 (bottom) ABO/Le blood group antigens in the gastric mucosa are shown. Those that act as receptors for BabA are outlined by a red box. Glycan symbolic representations can be interpreted with the following key: fucose – , galactose – , N-acetylgalactosamine – , N-acetylglucosamine – .

The majority of published work to date has investigated the ability of BabA to interact with these host receptors through whole bacterium assays. Consequently, there has been a lack of information regarding the molecular basis for glycan recognition by BabA. Such information is, however, essential for utilising BabA as a targeting moiety. Importantly, knowledge of the location of its glycan-binding site will support or negate intentions to create BabA-particle conjugates through its C-terminal hexa-lysine tag, as previously suggested in Section 3.4. Furthermore, downstream, variant BabA proteins lacking binding properties are useful as an additional measure to confirm any glycan binding activity of BabA-vectors is specifically imparted by BabA, rather than non-specific effects. Relevant BabA-variants can only be generated with structural insight.

With these intentions in mind, after validating the binding activity of BabA_{527K}, its crystal structure was solved in the absence and presence of Le^b – the ABO/Le blood group antigen most studied in BabA-mediated *H. pylori* attachment. To validate the binding site revealed by the crystallographic model, and to support downstream studies, BabA_{527K} variants were generated through site directed mutagenesis. Subsequent binding studies indicated that the successful elimination of BabA glycan binding properties was achieved.

Importantly, BabA_{527K} displayed the same glycan recognition pattern as

H. pylori J99 cells. Therefore, in addition to aiding the design of BabA-microparticle conjugates, these findings explain the previously poorly understood molecular mechanism of BabA-mediated attachment by *H. pylori*.

4.2 Experimental Procedures

4.2.1 Binding activity of BabA_{527K}

4.2.1.1 Enzyme-linked immunosorbent assay

The composition of neoglycoconjugates used in this experiment is detailed in Appendix Table 2. 96-well Maxisorp[®] plates (Thermo Scientific, USA) were coated with 100 µL of human serum albumin (HSA) (Sigma Aldrich, USA), HSA-Le^y and HSA-Le^b glycoconjugates (Isosep AB, Sweden) at 5 µg/mL in 0.05 M sodium carbonate buffer (pH= 9.6) at 4 °C overnight. Plates were washed with PBS supplemented with 0.05% Tween 20 (PBS-T) and blocked with 200 µL 2% non-fat dried skimmed milk in PBS-T for 2 hours at room temperature. After washing again with PBS-T, sample wells were incubated with 100 µL of BabA_{527K} at different concentrations (1.0 – 20 µg/mL in PBS-T) at 4 °C overnight. A mouse anti-c-Myc:Biotin clone 9E10 antibody (AbD Serotec, USA), diluted 1:2000 in PBS-T, was added, after washing, for 2 hours at room temperature. This step was repeated with streptavidin-HRP (AbD Serotec, USA), also at a 1:2000 dilution. After a final series of washes, wells were exposed to 100 µL of 0.1 mg/mL tetramethylbenzidine in 0.1 M sodium acetate/acetic acid buffer (pH = 6.0) supplemented with H₂O₂ to a final concentration of 0.03%. The reaction was allowed to proceed for 10 minutes then stopped through the addition of 50 µL 2.5 M H₂SO₄. Absorbance was measured at 450 nm.

4.2.1.2 Surface plasmon resonance

The composition of neoglycoconjugates used in this experiment is detailed in Appendix Table 2. SPR assays were performed at 25 °C on a Biacore 3000 instrument (GE Healthcare, USA). HSA-Le^y and HSA-Le^b glycoconjugates (Isosep AB, Sweden) were immobilised onto two separate flow cells of a CM5

sensor chip (GE Healthcare, USA) via amine coupling. This was achieved by first converting the carboxymethyl groups of the CM5 surface layer to reactive *N*-hydroxysuccinimide esters via an injection of a 0.2 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and 0.05 M *N*-hydroxysuccinimide (NHS) at a rate of 5 μ L/min for 7 minutes. Next, each glycoconjugate, at a concentration of 25 μ g/mL in 10 mM sodium acetate (pH = 4.5), was coupled onto the CM5 surface until an increase of \sim 10,000 response units (RU) was achieved. To deactivate unreacted NHS esters, 0.1 M Tris-Cl (pH=8.0) was injected onto both flow cell surfaces at a rate of 5 μ L/min for 7 minutes. Each flow cell was equilibrated overnight in HEPES-buffered saline supplemented with EDTA and Tween-20 (HBS-ET). This consisted of 10 mM HEPES (pH = 7.4), 150 mM NaCl, 3 mM EDTA and 0.005% Tween-20. BabA_{527K} was then injected, in the same running buffer, over each flow cell surface in a serial flow manner, at concentrations up to 15 μ M at a rate of 30 μ L/min. Changes in RU were analysed during each injection, which had an association time of 120 seconds and a dissociation time of 300 seconds. Regeneration of the sensor chip surface was performed after each injection using 2 M potassium thiocyanate at a rate of 5 μ L/min for 1 minute.

4.2.1.3 Mammalian glycan array screening

Binding of BabA_{527K} to a mammalian glycan array was assessed by the Consortium for Functional Glycomics (CFG) (Emory University, USA) using their in-house glycan binding assay protocols and reagents. Binding of BabA_{527K} to the printed surface of the CFG version 5.2 array, containing 609 synthetic and mammalian glycans, was performed in a 20 mM Tris-Cl (pH = 7.4), 300 mM NaCl and 0.05% Tween-20 buffer. Bound BabA_{527K} was detected with a fluorescein isothiocyanate labelled anti-c-Myc antibody.

4.2.2 Crystallographic studies

4.2.2.1 Crystallisation of apo-BabA_{527K} and data collection

Experimental procedures assisted by Claire Brassington, AstraZeneca R&D.

Crystallisation was performed using the sitting drop vapour diffusion method in 96-well MRC crystallisation plates. Each droplet contained 200 nL of BabA_{527K} [at 20 mg/mL in 20 mM Tris-Cl (pH=7.4) and 300 mM NaCl] mixed with 200 nL of a precipitant solution, which was also present in an adjacent well (reservoir volume = 80 µL) for sample droplet equilibration. Screens were set up with BabA_{527K} at 4 °C and 20 °C using a variety of precipitant solutions as per the in-house AstraZeneca R&D crystallisation screening protocol. An initial hit was obtained after two weeks at 20 °C under the following conditions: 25% PEG 10,000, 0.1 M sodium propionate and sodium cacodylate trihydrate, bis-tris propane (PCTP) buffer (pH = 4.5). Crystallisation optimisation was performed by varying the pH and type of buffer, as well as the concentration and type of precipitant. An optimised hit was obtained three days after repeated dispensing trials and continued to grow for a further ten days at 20 °C under the following conditions: 27% PEG 3,350, 0.1 M sodium citrate (pH = 5.6) and 0.2 M ammonium acetate. Initial and optimised crystals were cryoprotected by transfer to a well solution including 20% glycerol for 30 seconds, plunged into liquid nitrogen and kept at 100 K during data collection. X-ray data was collected at the Diamond Light Source, Didcot, UK (Beamline i04, DECTRIS Pilates detector).

4.2.2.2 Generation of selenomethionine substituted BabA_{527K}

Selenomethionine (SeMet) substituted BabA_{527K} was expressed using the pOPE101_*babA*_{527K} expression construct in *E. coli* XL10 Gold cells. Cells were grown in M9 minimal media supplemented with 2 mg/L thiamine, 4 g/L glucose, 2 mM MgSO₄, 0.1 mM CaCl₂, 100 µg/mL ampicillin and 12.5 µg/mL tetracycline at 37 °C until the OD₆₀₀ reached ~0.6. The following amino acids were added for 15 minutes before IPTG induction at 24 °C for 16 hours: lysine, phenylalanine and threonine (100 mg/L each) and isoleucine, leucine, valine and DL-selenomethionine (50 mg/L each). Periplasmic extraction was performed using the same procedures, as described in Section 3.2.3.1. However, extracts were supplemented with EDTA to a final concentration of 1 mM. Purification was performed using the same IMAC-SEC protocol as for unlabelled BabA_{527K}, as described in Sections 3.2.3.2 and 3.2.3.3. EDTA was added to IMAC buffers only, to a final concentration of 1 mM. Incorporation levels of SeMet were assessed using liquid chromatography–time-of-flight mass spectrometry, as described in Section 3.2.3.5.

4.2.2.3 Crystallisation of SeMet apo-BabA_{527K} and data collection

Experimental procedures assisted by Claire Brassington, AstraZeneca R&D.

SeMet BabA_{527K} at a concentration of 20 mg/mL, in a 20 mM Tris-Cl (pH=7.4) and 300 mM NaCl buffer, was crystallised using the sitting drop vapour diffusion method as described in Section 4.2.2.1 above. Conditions yielding crystals were identified, at 20 °C, under the following conditions: 34% PEG 3350, 0.2 M ammonium acetate and 0.1 M PCTP (pH = 6.5). To optimise the hit before X-ray data collection, crystal seeding was used. Droplets were prepared by mixing 250 nL of SeMet BabA_{527K}, at a concentration of 20 mg/mL in a 20 mM Tris-Cl (pH=7.4) and 300 mM NaCl buffer, with 250 nL of the

aforementioned precipitant solution and 50 nL of seed stock. The seed stock was prepared by placing three large unlabelled BabA_{527K} crystals into a microcentrifuge tube with 50 µL of the precipitant solution. A seed bead was then added and the mixture was vortexed for 2 minutes immediately prior to dispensing trials. Crystals appeared overnight under the following conditions: 30% PEG 3350, 0.2 M magnesium acetate and 0.1 M PCTP (pH = 6.0), and continued to grow for a further 5 days at 20 °C. Finally, crystals were cryoprotected by transfer to a well solution including 20% glycerol for 30 seconds, plunged into liquid nitrogen and kept at 100 K during data collection. X-ray data was collected at the European Synchrotron Radiation Facility, Grenoble, France (Beamline ID23-1, DECTRIS Pilates detector).

4.2.2.4 Co-crystallisation of BabA_{527K}:Le^b and data collection

Experimental procedures assisted by Claire Brassington, AstraZeneca R&D.

BabA_{527K}:Le^b co-crystals were screened for at 4 °C and 20 °C using the sitting drop vapour diffusion method described in Section 4.2.2.1 above. The BabA_{527K}:Le^b complex was formed, before dispensing trials, by pre-incubating a four-fold molar excess of Le^b antigen hexasaccharide (Isosep AB, Sweden; Appendix Table 3) with 20 mg/mL unlabelled BabA_{527K} on ice. Both BabA_{527K} and Le^b were in a 20 mM Tris-Cl (pH=7.4) and 300 mM NaCl buffer. A suitable hit was obtained after three days at 20 °C and continued to grow for a further ten days under the following conditions: 22% PEG 3,350, 0.1 M PCTP buffer (pH = 6.0). Finally, crystals were cryoprotected by transfer to a well solution including 20% glycerol for 30 seconds, plunged into liquid nitrogen and kept at 100 K during data collection. X-ray data was collected at the Diamond Light Source, Didcot, UK (Beamline i04, DECTRIS Pilates detector).

4.2.2.5 Structure solution

Experimental procedures performed by Tina Howard and Judit Debreczeni, AstraZeneca R&D.

The structure of BabA_{527K} could not be solved with the unlabelled apo-BabA_{527K} dataset using the SabA 26695 model (PDB accession code: 4O5J) via molecular replacement [Phaser (210)] due to insufficient sequence identity. Thus, to solve the structure, highly redundant single-wavelength anomalous SeMet apo-BabA_{527K} diffraction data was processed using XDS, Truncate and Aimless (211). The anomalous completeness of the data was 98.3% (87.2% in outer shell), with an anomalous multiplicity of 12.0 overall (7.0 in outer shell). Anomalous correlation between half-sets was 0.760 in the inner shell and 0.280 overall. Crank2 (CCP4i) (212-218) was used to solve the structure. Programs used in pipeline were ShexIC (213), ShelxD (213), Refmac5 (214), Solomon (215), Multicomb (216), Parrot (217) and Buccaneer (218). ShelxC found four selenium atoms. 89% of the residues were built with four gaps in the chain. Coot (217) was used for model building; rebuilding of chains across gaps was necessary where they had crossed to symmetry related molecules in the initial model. The model was further refined as a 1.9 Å dataset (Refmac5) although the data was weak in the outer shell. The Ramachandran plot showed 100% of residues were allowed with 96.04% in the preferred region. This apo model was used in molecular replacement [Phaser (210)] to solve the structure of BabA_{527K} complexed with the Le^b. The sugar was fitted using Coot (0.8) (219) and refinement again carried out using Refmac5. X-ray data and refinement statistics are shown in Table 4.1.

Table 4.1: X-ray diffraction data collection and refinement statistics

Values in parentheses refer to the highest recorded resolution shell

§Excluding the partially visible galactose (Gal5) moiety

Components	SeMet apo-BabA_{527K} (SAD data collection)	BabA_{527K}:Le^b (Molecular Replacement)
PDB Deposition		
Accession number	4ZH0	4ZH7
Data collection		
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
<i>Cell dimensions</i>		
<i>a, b, c (Å)</i>	60.83, 93.04, 96.92	60.59, 91.77, 96.42
<i>α, β, γ (°)</i>	90, 90, 90	90, 90, 90
Wavelength (Å)	0.97925	0.920
Resolution (Å)	48.46-1.91 (1.98-1.91)	44.78-2.12 (2.18-2.12)
<i>R_{merge} (all I+ and I-)</i>	0.236 (3.427)	0.141 (0.643)
<i>I/σI</i>	13.2 (0.9)	11.8 (3.0)
Completeness (%)	98.7 (89.2)	99.1 (99.4)
Multiplicity	23.1 (13.8)	6.6 (6.5)
Refinement		
No. of reflections	39877	29319
<i>R_{work} / R_{free}</i>	0.189/0.231 (0.378/0.331)	0.171/0.223 (0.224/0.271)
<i>No. of atoms</i>		
Protein	3662	3654
Ligand		57
Water	132	156
<i>B-factors</i>		
Protein (Å ²)	25.327	20.97
Ligand (Å ²)		35.413 [31.9] [§]
Water (Å ²)	34.124	30.034
<i>R.m.s deviations</i>		
Bond lengths (Å)	0.0184	0.0165
Bond angles (°)	1.7756	1.726
Ramachandran plot		
Allowed residues (%)	100	100
Preferred residues (%)	96.04	96.45

4.2.2.6 Sequence analysis

For identification of proteins with structural similarity to BabA, the apo-BabA PDB file (Accession code: 4ZH0) was uploaded onto the Dali server (220) for an atomic coordinates similarity search within the PDB. For analysis of sequence similarity of BabA regions with other *H. pylori* adhesins, from the J99 strain, CLC Main WorkBench 7.6 (CLC bio, USA) was used for amino acid multiple sequence alignment of BabA (AAD06409.1); LabA (AAD05605.1); HopZ (AAD05591.1); SabA (AAD06240.1); AlpA (AAD06426.1); and AlpB (AAD06427.1). Indicated in brackets are Genbank accession numbers.

For analysis of *H. pylori* strain conservation of the Le^b binding site, available *babA* gene sequences corresponding to the Crown region (amino acids 183-253 in mature BabA J99) from Le^b binding strains identified in (120, 123, 124) were obtained from the European Nucleotide Archive and GenBank. CLC Main WorkBench 7.6 (CLC bio, USA) was used for the multiple sequence alignment (protein) of the aforementioned BabA J99 fragment with that of sequences obtained from a total of 28 *H. pylori* strains. To search for DNA sequences with similarity to the BabA Crown, the BabA gene fragment corresponding to the Crown [i.e. complement (914203-914421) in GenBank accession no. AE001439.1] was submitted to the National Centre for Biotechnology Information (NCBI) basic local alignment search tool (BLAST). The nucleotide (nr/nt) database was searched using megablast, discontinuous and blastn algorithms. To search for protein sequences with similarity to the BabA Crown, the BabA protein fragment corresponding to the Crown (i.e. residues 203-273 in GenBank accession no. AAD06409.1) was submitted to NCBI Protein-BLAST. The protein sequences (nr) database was searched using Protein-protein, Position-specific Iterated, Pattern Hit Initiated and Domain Enhanced Lookup Time Accelerated BLAST algorithms.

4.2.3 Binding affinity of BabA_{527K} and BabA_{527K} variants

4.2.3.1 Expression and purification of BabA_{527K} variants

Expression constructs encoding BabA_{527K} variants containing either N206A or D233A/S244A substitutions were generated with a Phusion site-directed mutagenesis kit (Thermo Scientific, USA), using the primers shown in Table 4.2 and the pOPE101_ *babA*_{527K} expression construct as the DNA template. The PCR reaction was set up according to the manufacturer's protocol using an annealing temperature of 72 °C for both reactions. Confirmatory sequencing of the plasmids was performed (Source BioScience, Nottingham) using primers previously listed in Table 3.3. Expression and purification was performed under the same conditions as BabA_{527K}, as described in Section 3.2.3.

Table 4.2: Primers for BabA_{527K} alanine point substitutions

[FOR] and [REV] denote sense and antisense primers, respectively. [Phos] denotes a 5' phosphorylation.

Primer Name	Sequence (5'-3')
BabA _{527K} -N206A [FOR]	[Phos]GAACCAAGACTAAAATCCAAACCA TAGAC
BabA _{527K} -N206A [REV]	[Phos]CGTCTTTTTGAGCATTACACCTGT GAC
BabA _{527K} -D233A [FOR]	[Phos]GTTCAAAGTGGTTGCTAGTCGTG CAGATG
BabA _{527K} -D233A [REV]	[Phos]TGATCGTGGTGGTTACGCTTTTGC CGTCTATG
BabA _{527K} -D233A/S244A [FOR]	[Phos]GTAATACAACAGGGGTGGCCTACA CCGAAATCAC
BabA _{527K} -D233A/S244A [REV]	[Phos]CATCTGCACGACTAGCAACCACTT TTGAAC

4.2.3.2 Isothermal titration calorimetry

The composition of oligosaccharides used in this experiment is detailed in Appendix Table 3. Calorimetric measurements were performed at 25 °C on a MicroCal iTC200 System (GE Healthcare, USA). All ABO/Le blood group antigens (purity > 90%) were obtained in oligosaccharide form from Elicityl SA, France except Le^b, which was obtained from Isosep AB, Sweden. The sample cell was filled with BabA_{527K} (at concentrations ranging between 0.035 mM and 0.25 mM) and stirred at 1000 rpm until the system reference power was equilibrated to 6 µCal/sec. The injection syringe contained the ABO/Le blood group antigen (at concentrations ranging between 0.35 mM and 10 mM) and nineteen repeated 2 µL injections were made allowing 120 seconds between each titration. In general, the optimal concentrations of BabA_{527K} and ABO/Le blood group antigens used were 0.1 mM and 2 mM, respectively. However, these conditions were modified when studying the interaction between BabA_{527K} and H-1, A-1 and B-1 antigens, where the protein and oligosaccharide concentration was increased to 0.25 mM and 10 mM, respectively. A-Le^b and B-Le^b antigens were not studied, as they were not readily commercially available. The optimal concentrations of BabA_{527K}-N206A and Le^b used were 0.25 mM and 10 mM, respectively. Binding of BabA_{527K}-D233A/S244A to ABO/Le blood group antigens was studied at protein and oligosaccharide concentrations of 0.1 mM and 2 mM, respectively. A buffer containing 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl was used for all calorimetric measurements. NITPIC (221) was used for baseline auto-determination and calorimetric data was analysed by peak integration using ORIGIN 7.0 software (OriginLab, USA), taking into account the heat of dilution for each glycan into buffer.

4.2.3.3 Circular dichroism spectroscopy

Circular dichroism spectra of BabA_{527K} and BabA_{527K} variants was measured using a J-810 Spectropolarimeter (Jasco, USA). Protein concentration was 1 µM in a buffer containing 20 mM Tris-Cl (pH = 7.4) and 3 mM NaCl. Measurements were made at 25 °C in a quartz cell with a 0.1 cm path length at a data pitch of 0.5 nm and scanning speed of 100 nm/min. Reported spectra are baseline-corrected for buffer alone and averaged from three independent scans.

4.2.3.4 Differential scanning fluorimetry

SYPRO Orange dye (20x final concentration) was added to 10 µM BabA_{527K} and BabA_{527K} variants in a buffer containing 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl. Changes in fluorescence were measured across an increasing temperature gradient from 25 °C to 60 °C using a LightCycler 480 II (Roche, Switzerland) at a ramp rate of 0.01 °C/s. Primary data points from three independent experiments were fitted to a 6-parameter unfolding equation (222) using the Prism analysis package (GraphPad Prism 6 Software, USA).

4.3 Results

4.3.1 Validating the functional activity of BabA_{527K}

To confirm that BabA_{527K} was functionally active after recombinant expression and purification out of the periplasmic space of *E. coli*, binding to various ABO/Le blood group antigens was assessed. This was achieved using three different techniques:

4.3.1.1 Enzyme-linked immunosorbent assay

As an initial test, BabA_{527K} interactions with HSA-Le^b and HSA-Le^y were assayed in a sandwich ELISA format. In this experiment, binding of BabA_{527K} to Le antigen neoglycoconjugates, which were immobilised to a Maxisorp[®] plate, was detected via an antibody directed to its C-terminal c-Myc tag. As shown in Figure 4.2, specific concentration-dependent binding to HSA-Le^b was observed.

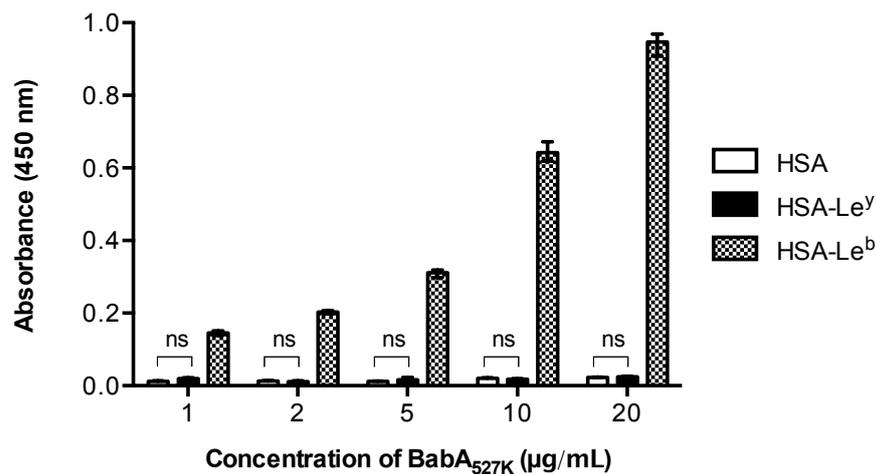


Figure 4.2: Binding of BabA_{527K} to HSA-Le^{b/y} in an ELISA

Binding of BabA_{527K} to immobilised HSA, HSA-Le^y and HSA-Le^b glycoconjugates was determined using a sandwich ELISA (n=3, error bars represent the standard error of the mean). ns = not significant (p > 0.05), unpaired two-tailed Welch's t-test.

4.3.1.2 Surface plasmon resonance

To further validate BabA_{527K} attachment to HSA-Le^b, SPR was employed to provide a direct approach to the detection of binding. Similar to the ELISA setup, HSA-Le^b and HSA-Le^y glycoconjugates were immobilised to the surface of CM5 sensor chip flow cells. This was achieved through amine coupling where an immobilisation level of ~10,000 RU was targeted (Figure 4.3).

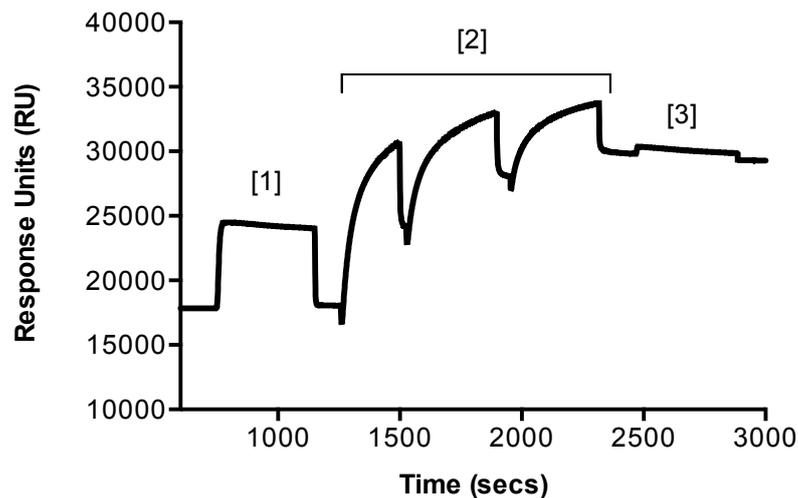


Figure 4.3: Immobilisation of HSA-Le^b to a CM5 sensor chip via amine coupling

[1] Activation of carboxymethyl groups on a CM5 sensor chip surface to NHS esters via injection of NHS/EDC, **[2]** Injections of HSA-Le^b, **[3]** Deactivation of unreacted NHS esters via injection of Tris-Cl. The amount of immobilised ligand (RU after deactivation minus from RU after activation) was ~10,000 RU. HSA-Le^y was similarly immobilised via amine coupling.

BabA_{527K} was injected in a serial flow manner over HSA-Le^y then HSA-Le^b surfaces. During injection over the HSA-Le^y flow cell, an increase in RU was observed due to the changes in the refractive index of the running and sample buffer. However, no association or dissociation phases are seen in this sensogram. On the other hand, BabA_{527K} clearly interacts with HSA-Le^b as an association phase is observed during sample injection. This is succeeded by a dissociation phase after injection termination at 120 seconds (Figure 4.4A). BabA_{527K} also displayed concentration-dependent binding towards HSA-Le^b,

but not HSA-Le^y, when studied in a multi-cycle experimental setup (Figure 4.4B).

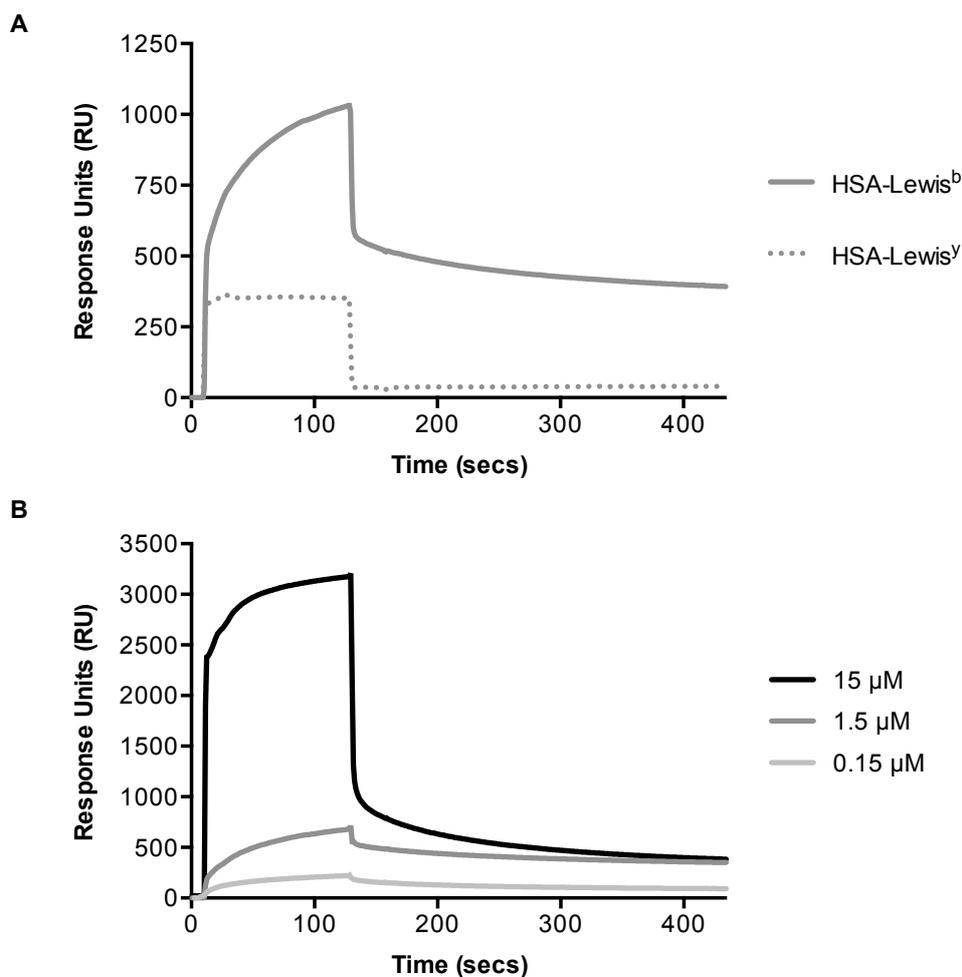


Figure 4.4: SPR analysis of BabA_{527K} binding to HSA-Le^{b/y}

(A) BabA_{527K}, at a concentration of 1.5 μM, was injected over HSA-Le^y then HSA-Le^b flow cells in a serial flow manner. (B) BabA_{527K} binding to HSA-Le^b was studied in a multi-cycle assay, that is, multiple analyte injections to the same ligand surface separated by regeneration steps. The RU shown have been reference subtracted for each sensorgram trace (HSA-Le^b minus HSA-Le^y).

SPR was not used to determine BabA_{527K} binding affinity to HSA-Le^b glycoconjugates in this study. This is because any apparent affinity determined is not reflective of a BabA_{527K}:Le^b interaction, but rather of BabA_{527K} binding to a multivalent complex.

4.3.1.3 Mammalian glycan array

Finally, binding of BabA_{527K} to a glycan array containing several natural and synthetic mammalian glycans was outsourced to the CFG. This was used to examine BabA_{527K} ABO/Le blood group antigen binding specificity as the array slide consisted of glycans containing the unique determinant regions of:

- H-1, A-1 and B-1 antigens, which are type 1 ABO blood group antigens
- Le^a, Le^b, A-Le^b and B-Le^b, which are type 1 Le blood group antigens
- H-2, A-2 and B-2 antigens, which are type 2 ABO blood group antigens
- Le^x, Le^y, A-Le^y and B-Le^y, which are type 2 Le blood group antigens

Out of 609 glycans, notable binding was detected to five glycans, each of which contained a unique determinant region of H-1, A-1 or B-1 (type 1 ABO) blood group antigens (Figure 4.5A/B). While the selective binding observed may be reflective of the BabA_{527K} glycan binding properties, it must be noted that epitope presentation on the array slide greatly affected BabA_{527K} binding: no binding to glycans containing the unique determinant regions of H-1, A-1 or B-1 blood group antigens was detected if these structures were part of larger sugar complexes, or if they were attached to the array slide directly through their central *N*-acetylglucosamine residue (Figure 4.5A/C, Appendix Table 4).

Of the five type 1 ABO glycan structures to which BabA_{527K} bound, three (corresponding to the H-1, A-1 and B-1 antigen) were attached to the array slide via the following glycan linkage and spacer arm: [Core (GlcNAc)]β1-3GalNAcα1-Sp14. Accordingly, to exclude the effect of epitope presentation, BabA_{527K} binding to type 1 Le blood group antigens and type 2 ABO/Le blood group antigens on the array slide was only analysed if their unique determinant regions were conjugated to the array slide through the aforementioned linkage. All of the type 2 ABO/Le blood group antigens listed

at the beginning of this section were immobilised to the array slide through this glycan linkage and spacer arm – no notable binding was observed to any of these glycans (Figure 4.5A/D, Appendix Table 5). However, no type 1 Le blood group antigens were conjugated to the array slide through the specified linkage (Figure 4.5A/E).

Low levels of binding, suggestive of non-specific binding either by BabA_{527K} or the detecting antibody, were detected to a large proportion of the 609 glycans tested – this indicates that the assay conditions could be optimised.

BabA_{527K} to blood group antigens containing type 1 and type 2 cores. The unique determinant regions of the H-1 and H-2 antigen are both linked to the array slide through the same glycan linkage and spacer arm, that is, [Core (GlcNAc)] β 1-3GalNAc α 1-Sp14. **(E)** The unique determinant regions of type 1 Le antigens were not attached to the array slide through this specified linkage, and BabA_{527K} did not display notable binding to the only form of the Le^b antigen present on the slide. Note that the Le^b antigen [60] contained the same glycan linkage and spacer arm as the H-1 antigen [67], to which BabA_{527K} did not display any notable binding. Unique determinant regions are surrounded by a light orange shade. Sp0, Sp8, Sp10, Sp14, Sp20 and Sp21 are CFG specific spacer arms. Glycan symbolic representations can be interpreted with the following key: fucose – , galactose – , glucose – , N-acetylgalactosamine – , N-acetylglucosamine – , mannose – .

The ELISA, SPR assay and mammalian glycan array screen collectively indicated that BabA_{527K} retained the glycan binding activity and specificity mediated by BabA on *H. pylori* J99 cells (82). Although not all type 1 ABO/Le blood group antigens that act as receptors for BabA-mediated *H. pylori* attachment could be tested, these assays irrefutably confirm that BabA_{527K} is functionally active and is therefore suitable for structural studies and downstream applications.

4.3.2 Structural insight into BabA_{527K} glycan recognition

4.3.2.1 Determination of the crystal structure of apo-BabA_{527K}

To identify the glycan-binding site of BabA_{527K}, determination of its crystal structure in the absence and presence of Le^b was pursued. First, apo-BabA_{527K} crystallisation conditions were screened for, at both 4 °C and 20 °C, using an in-house AstraZeneca R&D screen comprised of 192 conditions. A twinned crystal, which diffracted to ~3 Å, was obtained as an initial hit (Figure 4.6A). Its diffraction pattern revealed overlapping X-ray reflections – a defect that is manifested through the appearance of smeared spots – thereby rendering this dataset unsuitable for structural determination. Optimisation of the crystallisation conditions was consequently performed and this produced a larger, though still twinned, crystal. A single-crystal fragment was broken off for X-ray radiation and single diffraction data was collected to a resolution of 2.0 Å (Figure 4.6B).

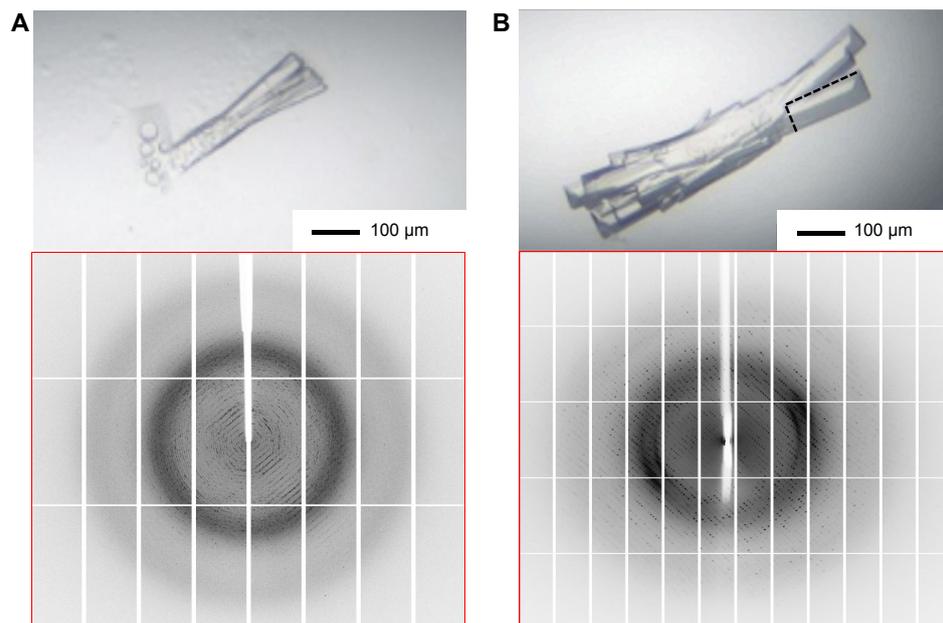


Figure 4.6: Crystallisation of apo-BabA_{527K}

Crystal image (top) and representative diffraction image (bottom) of (A) initial and (B) optimised apo-BabA_{527K} crystals. The broken black line in (B) indicates the single-crystal fragment broken off and subjected to X-ray radiation. Both crystals were obtained through the sitting drop vapour diffusion method.

Using this dataset, apo-BabA_{527K} structure solution was attempted via molecular replacement. This was unsuccessful due to the poor sequence identity (26%) between BabA_{527K} and the most related atomic model available, which was the extracellular domain of SabA from *H. pylori* 26695 (175) (Figure 3.2). Thus, to facilitate apo-BabA_{527K} structure solution, SeMet was substituted into BabA_{527K} in place of methionine to enable the use of anomalous dispersion to solve the phase problem. SeMet substitution was fulfilled during recombinant expression by introducing DL-selenomethionine into the *E. coli* XL10 Gold culture medium alongside amino acids known to inhibit methionine biosynthesis. Mass spectrometric analysis revealed SeMet substitution was both successful and efficient (incorporation efficiency = 94%). Once crystallised under similar conditions to unlabelled apo-BabA_{527K}, this protein also yielded a twinned crystal (results not shown). However, through crystal seeding, single crystals were obtained. A single crystal was subjected to X-ray radiation and produced a 1.9 Å dataset (Figure 4.7).

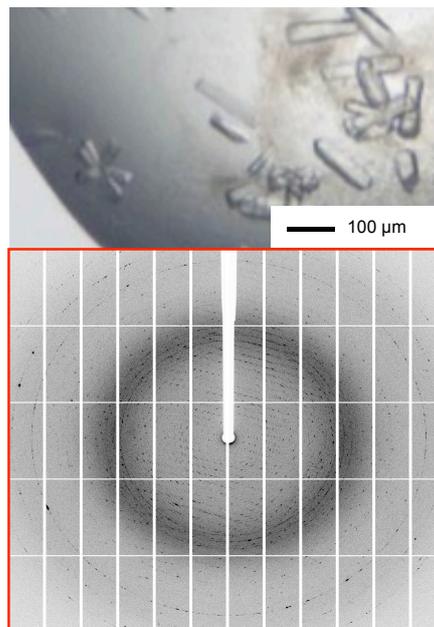


Figure 4.7: Crystallisation of SeMet apo-BabA_{527K}

Crystal image (top) and representative diffraction image (bottom) of optimised SeMet apo-BabA_{527K} crystals. Crystals were obtained by sitting drop vapour diffusion.

Single-wavelength anomalous dispersion was then successfully used to solve the phase problem of the SeMet apo-BabA_{527K} dataset. An electron density map was calculated, and after multiple rounds of iterative refinement, BabA_{527K} backbone and amino acid side chains were modelled (Figure 4.8). The final atomic model ran from residues Q27 to L527. In addition to the N- and C-terminal amino acids not visible in the electron density map, two disordered loops between A282 – P291 and S402 – Q410 were missing from the model (Appendix Figure 1).

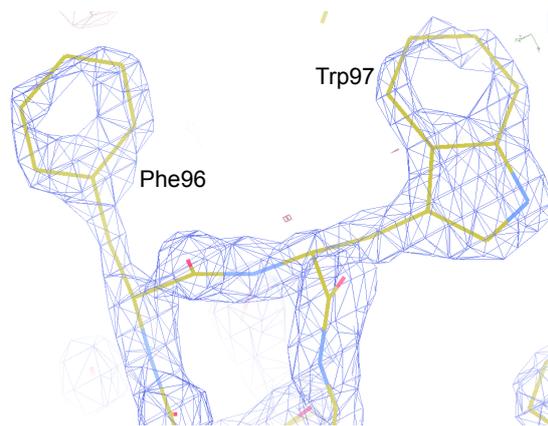


Figure 4.8: Representative electron density map of the apo-BabA_{527K} model

Two amino acids of BabA_{527K} modelled into the 1.9 Å-resolution electron density map [blue mesh ($2F_o - F_c$ map, contoured at 2.0σ)]. The visible hole in the centre of these aromatic residues indicates the high resolution of the calculated electron density map used to model BabA_{527K} backbone and amino acid side chains (sticks).

The crystallographic model (Figure 4.9) indicated that BabA_{527K} contains two predominantly α -helical regions, termed Handle and Head regions, and a third β -sheet motif located on top of the Head region. This β -strand unit was named the Crown. The Handle region, containing both the N- and C-termini of BabA_{527K}, forms an $\alpha+\beta$ unit. The N-terminal helix (α -N) forms a two-helix anti-parallel coiled coil bundle with a C-terminal helix (α -C1) of similar approximate length. This C-terminal helix is followed by a two-strand anti-parallel β -sheet (β -C) before ending with a short α -helix (α -C2), which packs against α -N in an

anti-parallel orientation. In the native BabA protein, the highly conserved putative β -barrel transmembrane domain sequence succeeds the α -C2 helix. The core of the Head region is comprised of a four-helix anti-parallel coiled coil bundle, similar to a tetratricopeptide repeat motif (α -1 to α -4), at a near perpendicular angle to the Handle region creating the markedly kinked tertiary structure. The connecting features between these four helices are: i) a loop containing a short α -helix between α -3 and α -4; ii) a 20 amino acid loop between α -2 and α -3; and iii) a \sim 200 amino acid segment between the α -1 and α -2 helices. This connecting segment, which extends out of the core of the Head region, contains a small β -sheet (β -1), a pair of interacting α -helices (α -1a and α -1b) and the Crown – the four-strand antiparallel β -sheet at the highest tip of the protein (β -2).

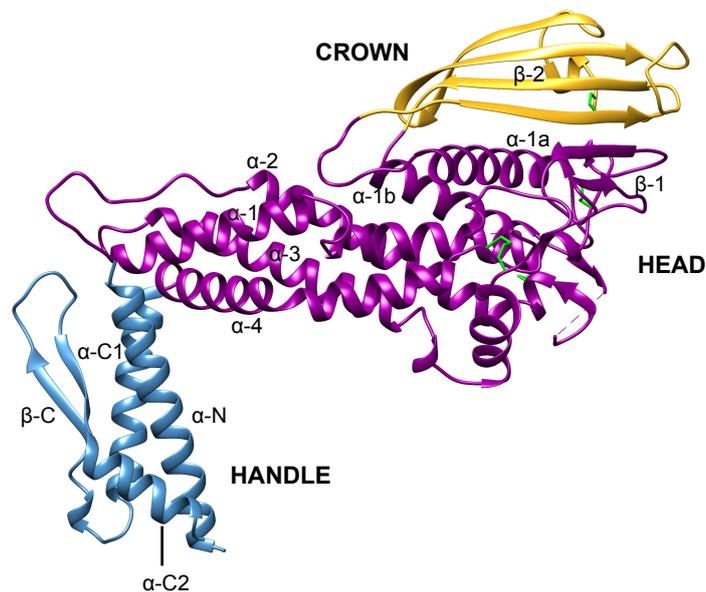


Figure 4.9: Ribbon representation of the crystal structure of BabA_{527K}

Indicated are the Handle (blue) and Head regions (dark magenta), and the Crown β -strand unit (gold). The four disulphide bridges are represented as green sticks.

To contextualise this novel crystal structure, BabA_{527K} was submitted to the Dali server (220) to identify structurally related molecules. This revealed the

extracellular domain of SabA 26695 (Figure 4.10A), which shares only 26% amino acid sequence identity with that of BabA_{527K}, as the single most related structure in the PDB database [Root-mean-square deviation (RMSD) = 3.7 Å for all C α atoms]. Superimposition of BabA and SabA shows that the α -1a helix is a common characteristic. This feature has been suggested to form part of the glycan-binding cavity of SabA, which recognises SLe^x antigens (Figure 4.10B). While both proteins share highly similar three-dimensional folds, the four-strand anti-parallel β -sheet Crown of BabA is altogether absent in SabA.

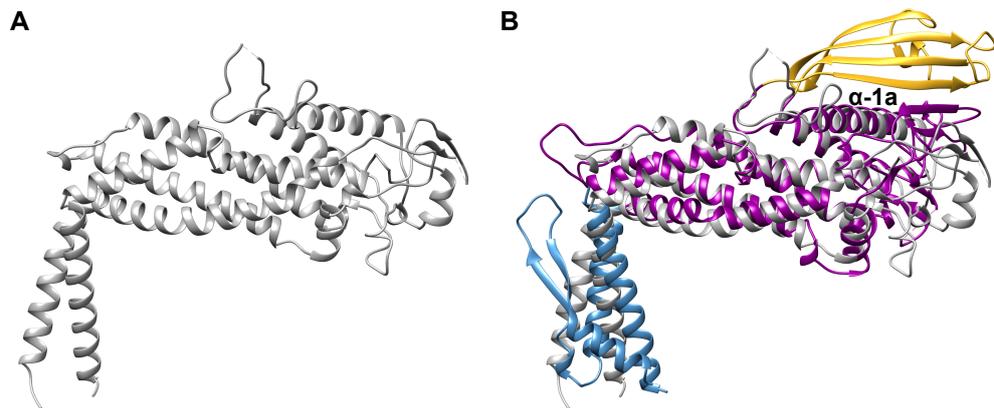


Figure 4.10: Comparison of BabA_{527K} with the extracellular domain of SabA

(A) Ribbon representation of the crystal structure of the extracellular domain of SabA 26695. (B) Superimposition of the extracellular domain of SabA 26695 and BabA_{527K}. Highlighted in colour are the Handle (blue) and Head regions (dark magenta), and the Crown β -strand unit (gold) of BabA_{527K}. The extracellular domain of SabA 26695 is indicated in grey.

Furthermore, at the protein level, a multiple sequence alignment indicated that the BabA Crown sequence is effectively absent in other known *H. pylori* adhesins (Figure 4.11). Moreover, apart from BabA, no DNA or protein sequences with similarity to the Crown were identified in *H. pylori* or any other organism in NCBI nucleotide sequence and protein sequence databases.

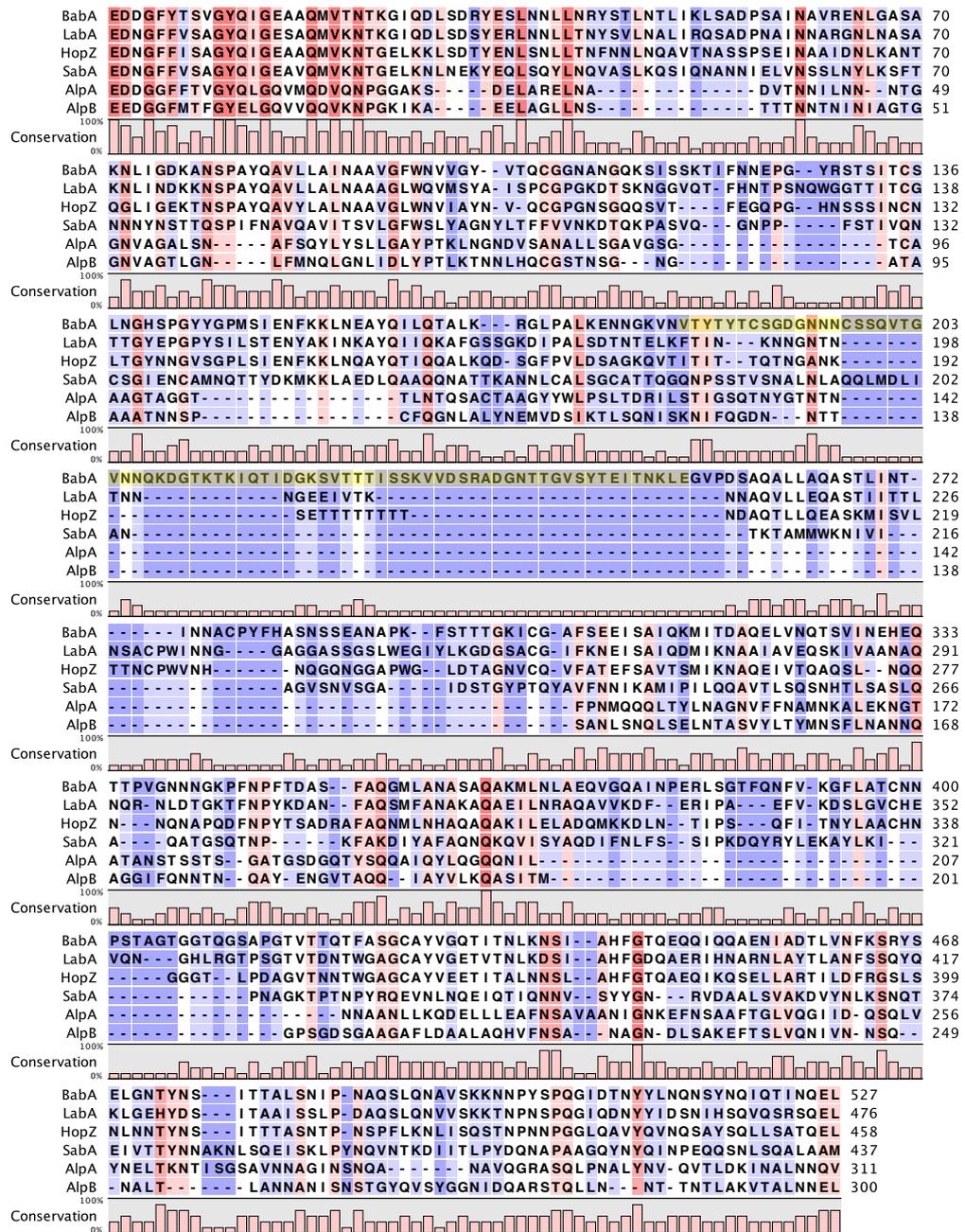


Figure 4.11: Sequence alignment of BabA with other known *H. pylori* adhesins
 Multiple sequence alignment of the putative extracellular domains of known adhesins, from *H. pylori* strain J99, indicates that the Crown sequence of BabA (highlighted in yellow, amino acids 183-253) is absent in other adhesins.

4.3.2.2 Determination of the crystal structure of BabA_{527K} in complex with Le^b

Having solved the structure of apo-BabA_{527K}, the crystal structure of BabA_{527K} bound to Le^b was pursued for glycan binding insight. To achieve this, the hexasaccharide form of the Le^b antigen was pre-incubated with BabA_{527K} prior to dispensing trials. As shown in Figure 4.12, this oligosaccharide contained two fucose residues (Fuc1 and Fuc4), two galactose residues (Gal2 and Gal5), a *N*-acetylglucosamine residue (GlcNAc3) and a glucose residue (Glc6).

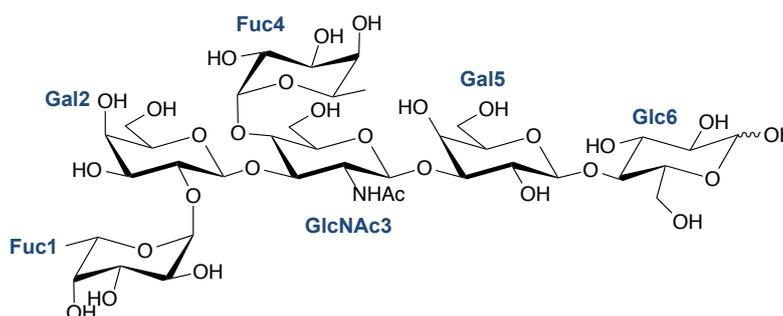


Figure 4.12: Chemical structure of the hexasaccharide form of Le^b

The monosaccharide units of Le^b are indicated.

Co-crystallisation screening yielded a large, twinned crystal. Nevertheless, after dislodging a single-crystal fragment, single X-ray diffraction data was successfully collected to a resolution of 2.1 Å (Figure 4.13). Molecular replacement was used to solve the phase problem of the BabA_{527K}:Le^b dataset using BabA_{527K} as a model. The BabA_{527K} chain built into the calculated electron density map run from residues Q27 to K528. Similar to the apo-model, two disordered loops (between A282 – F293 and S402 – Q410) were also missing from the BabA_{527K} chain (Appendix Figure 2).

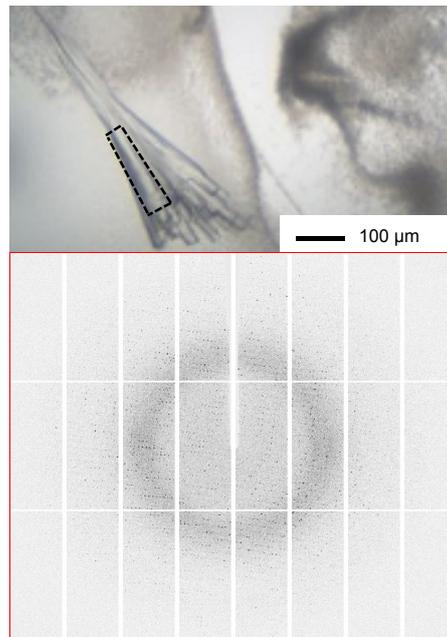


Figure 4.13: Crystallisation of BabA_{527K}:Le^b

Crystal image (top) and representative diffraction image (bottom) of BabA_{527K}:Le^b crystals. The broken black line in the crystal image indicates the single-crystal fragment broken off and subjected to X-ray radiation. Crystal obtained through the sitting drop vapour diffusion method.

As shown in Figure 4.14, BabA_{527K}:Le^b structural determination indicated that there was no conformational change to the BabA_{527K} chain after sugar complexation.

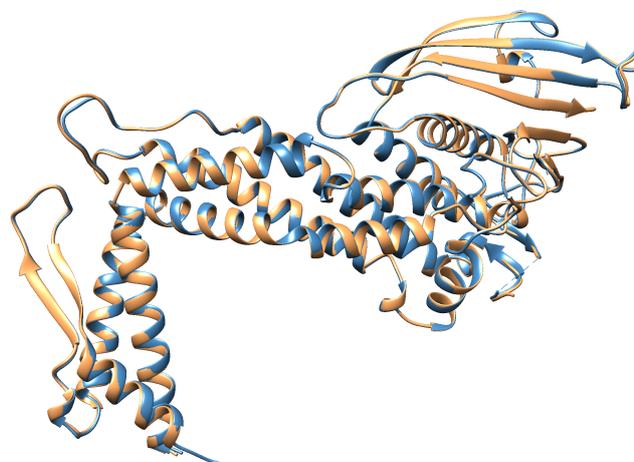


Figure 4.14: Superimposition of BabA_{527K} from apo- and co-crystal structures

No global conformational change occurs in BabA (sandy brown) after Le^b complex formation (steel blue) – RMSD = 0.25Å for all Cα atoms.

Furthermore, a single region of electron density, corresponding to Le^b, was identified adjacent to the Crown. Five of the six sugar residues could be modelled but the Glc6 unit was not seen and density for the adjoining Gal5 was partial (Figure 4.15A). Le^b bound to a shallow, solvent-exposed groove at the tip of the Crown (Figure 4.15B).

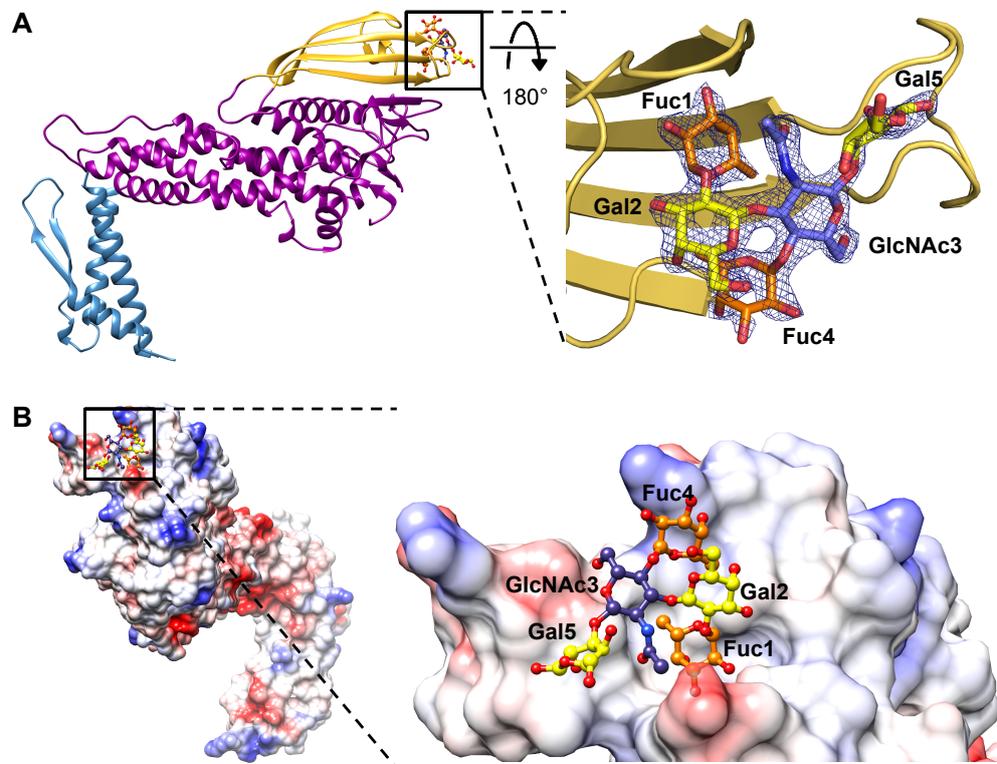


Figure 4.15: The Le^b binding site of BabA_{527K}

(A) The electron density map around Le^b ($2F_o - F_c$ map, contoured at 2.0σ) is shown. (B) An electrostatic surface representation of the Le^b binding pocket at the tip of BabA containing charged (red = negative; blue = positive) and neutral (white) patches. Fucose, galactose and *N*-acetylglucosamine residues are coloured orange, yellow and blue, respectively.

To investigate the binding mechanism, interactions between BabA_{527K}, Le^b and water atoms within bonding distance ($<4 \text{ \AA}$) were analysed. This indicated that binding is mediated solely by a network of hydrogen bonds between Le^b Fuc1, GlcNAc3, Fuc4 and Gal5 residues and a total of eight BabA amino acids. Fuc1 forms hydrogen bonds with the carbonyl backbone groups of C189, G191,

N194 and the hydroxyl group of the T246 side chain (Figure 4.16A). Fuc4 interacts with the hydroxyl group of the N206 side chain through a water-mediated hydrogen bond (Figure 4.16B). The GlcNAc3 residue forms two hydrogen bonds with the carboxyl and hydroxyl side chain groups of D233 and S244, respectively (Figure 4.16C). Lastly, Gal5 forms hydrogen bonds with both the S244 carbonyl backbone and hydroxyl side chain group. It also interacts with both the carboxyl group of the D233 side chain and the hydroxyl group of the S234 side chain through water-mediated hydrogen bonds (Figure 4.16D). No interactions were observed between Gal2 and BabA_{527K}.

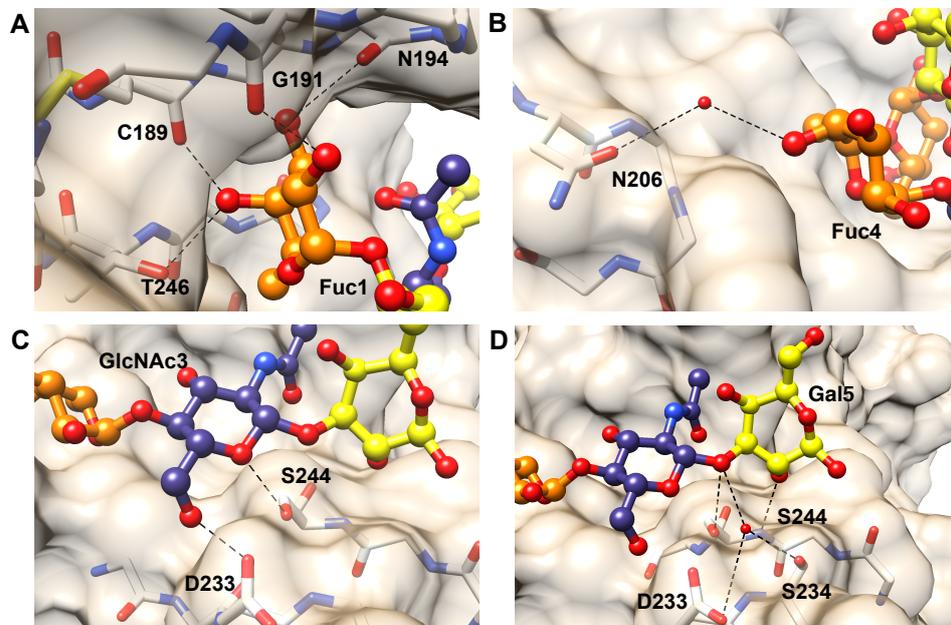


Figure 4.16: Hydrogen bonding between BabA_{527K} and Le^b residues

Fucose, galactose and *N*-acetylglucosamine residues are coloured orange, yellow and blue, respectively. Hydrogen bonds are represented by dotted black lines.

As this represents the first molecular insight into the structural basis of Le^b binding by BabA, the identified binding site could not be compared to relevant models. However, alignment of BabA_{527K} with BabA from 28 *H. pylori* strains, reported to bind Le^b glycoconjugates, revealed that the amino acids identified in mediating Le^b binding are highly conserved with the exception of N206

(Figure 4.17A). In fact, N206 is found within a region (residues 198 to 207 in BabA J99) with low amino acid conservation across the Le^b binding strains. Thus, this segment, which connects two antiparallel β -strands in the BabA_{527K} Crown, was termed the Hypervariable Crown Loop (Figure 4.17B). For the majority of analysed sequences, it was not known whether their corresponding BabA proteins displayed generalist or specialist binding patterns. However, Figure 4.17A shows that known specialist strains possess a conserved sequence within the Hypervariable Crown Loop that is markedly dissimilar, and longer, in comparison to known generalist strains.

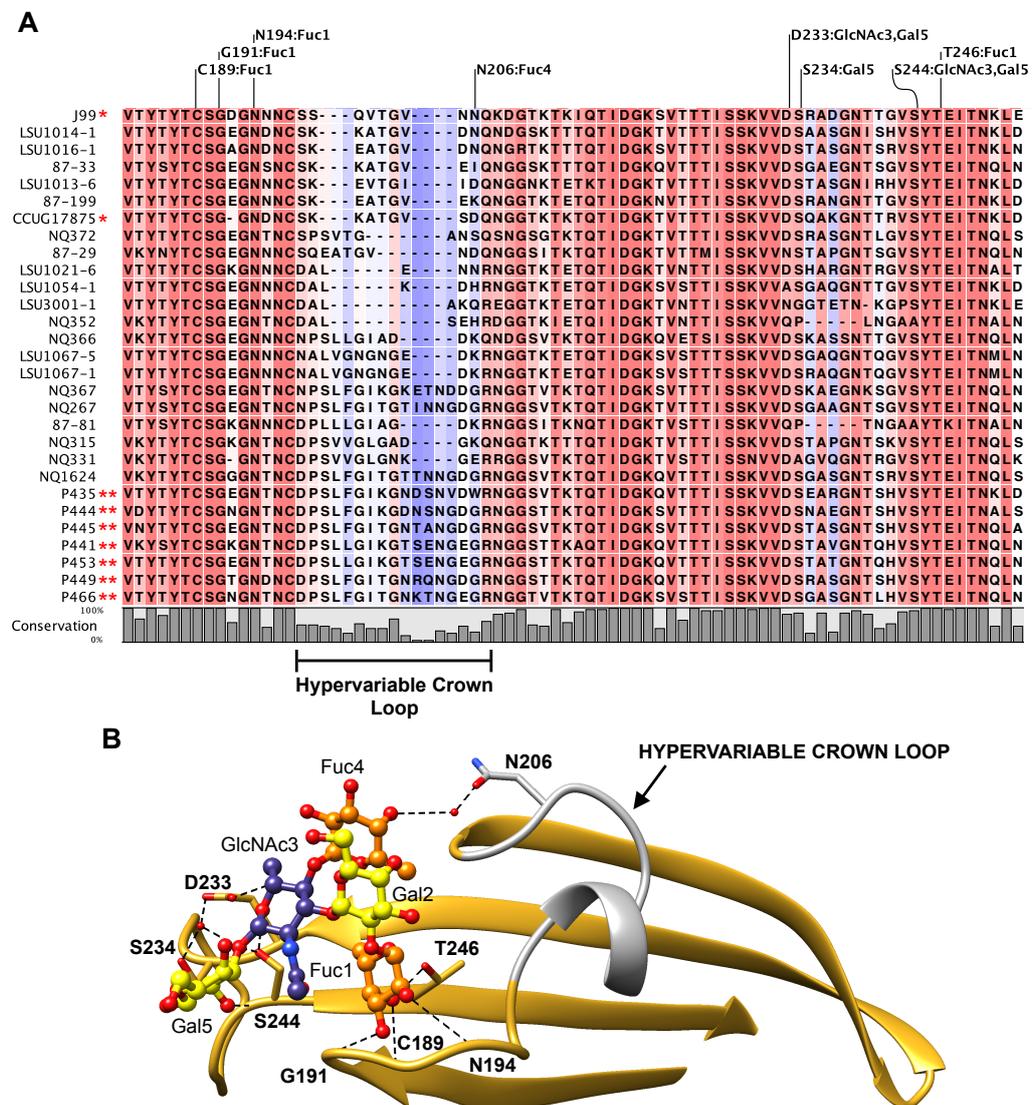


Figure 4.17: Conservation of the Le^b binding site

(A) Sequence alignment of the BabA_{527K} Crown (residues 183-253) with that of 28 Le^b binding *H. pylori* strains. Amino acids involved in hydrogen bond formation to each Le^b residue are indicated. (*) indicates generalist strains, which bind to the type 1 ABO/Le blood group antigens in group-O, -A and -B individuals. (**) indicates specialist strains, which bind to the type 1 ABO/Le blood group antigens in group-O individuals only (123). **(B)** The Hypervariable Crown Loop (residues 198-207 in BabA_{527K}) is indicated in grey. Fucose, galactose and N-acetylglucosamine residues are coloured orange, yellow and blue, respectively. Hydrogen bonds are represented by dotted black lines.

To build on the structural insight into Le^b antigen binding obtained from X-ray crystallography, the glycan binding characteristics of BabA_{527K} were next evaluated from a biophysical perspective.

4.3.3 Glycan binding profile of BabA_{527K} and BabA_{527K} variants

To further validate the binding mechanism of BabA_{527K}, ITC was used to characterise its interaction with the Le^b antigen hexasaccharide used in structural determination. ITC was specifically chosen over SPR as the biophysical technique to be used for glycan affinity determination so as to avoid the need for BabA_{527K} immobilisation and ligand surface regeneration, both of which can affect apparent affinity (223). In agreement with the structural model (Figure 4.18A), the thermodynamic parameters of binding confirmed a single-site interaction ($N \sim 1.07$) driven by non-covalent, that is, enthalpic contributions ($\Delta H \sim -10.9$ kcal/mole) rather than hydrophobic, that is, entropic contributions ($-T\Delta S \sim 6.0$ kcal/mole); binding was characterised by a weak dissociation constant (K_D) of ~ 252 μ M (Figure 4.18B).

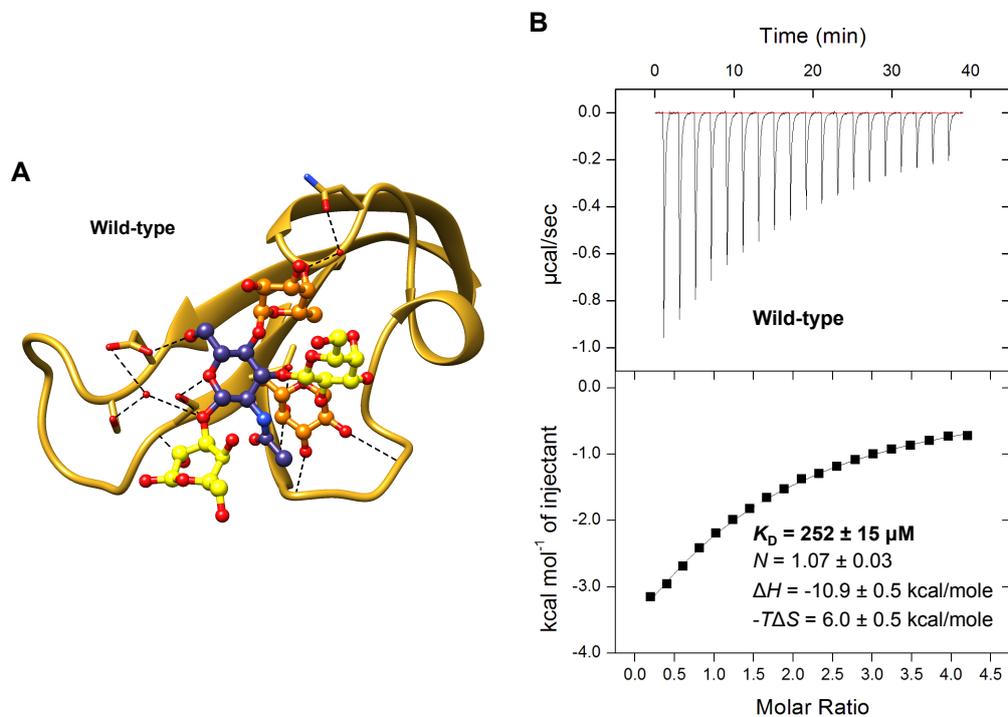


Figure 4.18: Binding affinity between wild-type BabA_{527K} and Le^b

(A) Observed interactions of wild-type BabA_{527K} with Le^b in the crystallographic model. The fucose, galactose and *N*-acetylglucosamine residues of Le^b are coloured orange, yellow and blue, respectively. Hydrogen bonds are represented by dotted black lines. (B) Calorimetric response (upper panel) and binding isotherm (lower panel) of wild-

type BabA_{527K} titrated with Le^b. The continuous line in the lower panel represents the least-squares fit of the data to a single-site binding model. The reported thermodynamic parameters are the average [\pm standard error of the mean (SEM)] of three independent experiments. The calorimetric titrations were performed at pH 7.4.

To validate and probe the significance of the interaction between Le^b and the Hypervariable Crown Loop, site-directed mutagenesis was used to perform an alanine point substitution at residue N206. A BabA_{527K}-N206A variant would lack the hydroxyl group of the asparagine side chain that was observed to form a single water-mediated hydrogen bond with the Fuc4 residue of Le^b (Figure 4.19A). Indeed, when studied through ITC, this variant had a lower affinity for Le^b ($K_D \sim 582 \mu\text{M}$) (Figure 4.19B). Again, the thermodynamic parameters of binding indicated a single-site interaction driven by enthalpic contributions rather than entropic contributions.

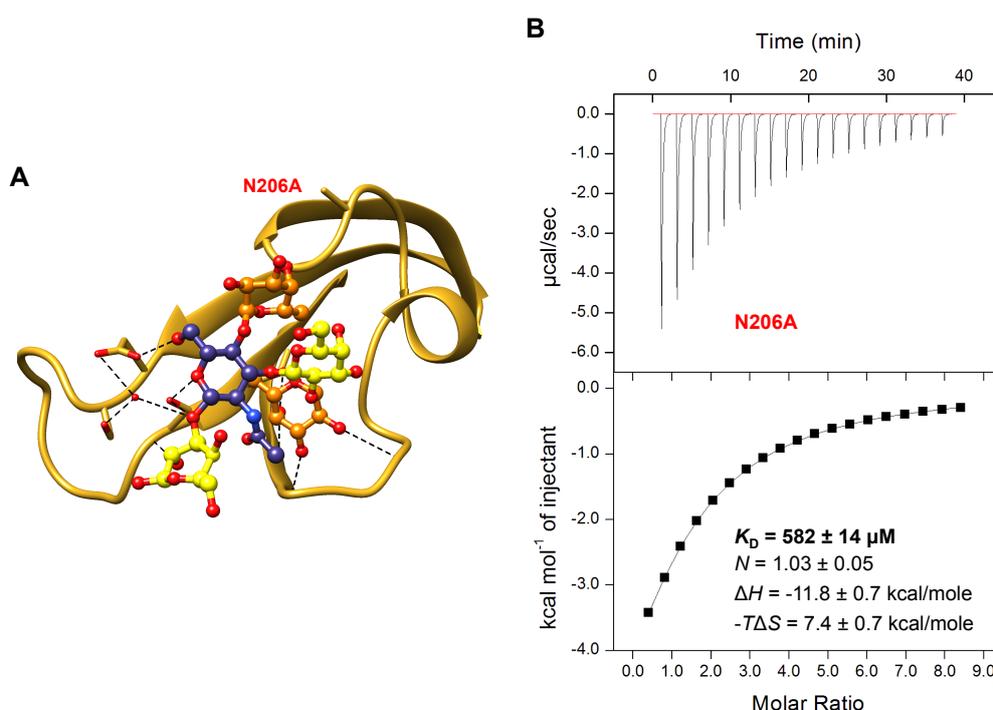


Figure 4.19: Binding affinity between the BabA_{527K}-N206A variant and Le^b

(A) Predicted interactions between BabA_{527K} with an N206A substitution (indicated) and Le^b, based on the crystallographic model. The fucose, galactose and *N*-acetylglucosamine residues of Le^b are coloured orange, yellow and blue, respectively.

Hydrogen bonds are represented by dotted black lines. **(B)** Calorimetric response (upper panel) and binding isotherm (lower panel) of BabA_{527K}-N206A titrated with Le^b. The continuous line in the lower panel represents the least-squares fit of the data to a single-site binding model. The reported thermodynamic parameters are the average (\pm SEM) of three independent experiments. The calorimetric titrations were performed at pH 7.4.

To validate and probe the significance of the interactions between Le^b and highly conserved amino acids in the Crown, site-directed mutagenesis was used to perform combined alanine point substitutions at residues D233 and S244. This variant would lack the carboxyl and hydroxyl side chains of D233 and S244, respectively, that form direct and water-mediated hydrogen bonds with GlcNAc3 and Gal5 of Le^b (Figure 4.20A). When studied through ITC, this variant displayed no detectable affinity for Le^b (Figure 4.20B).

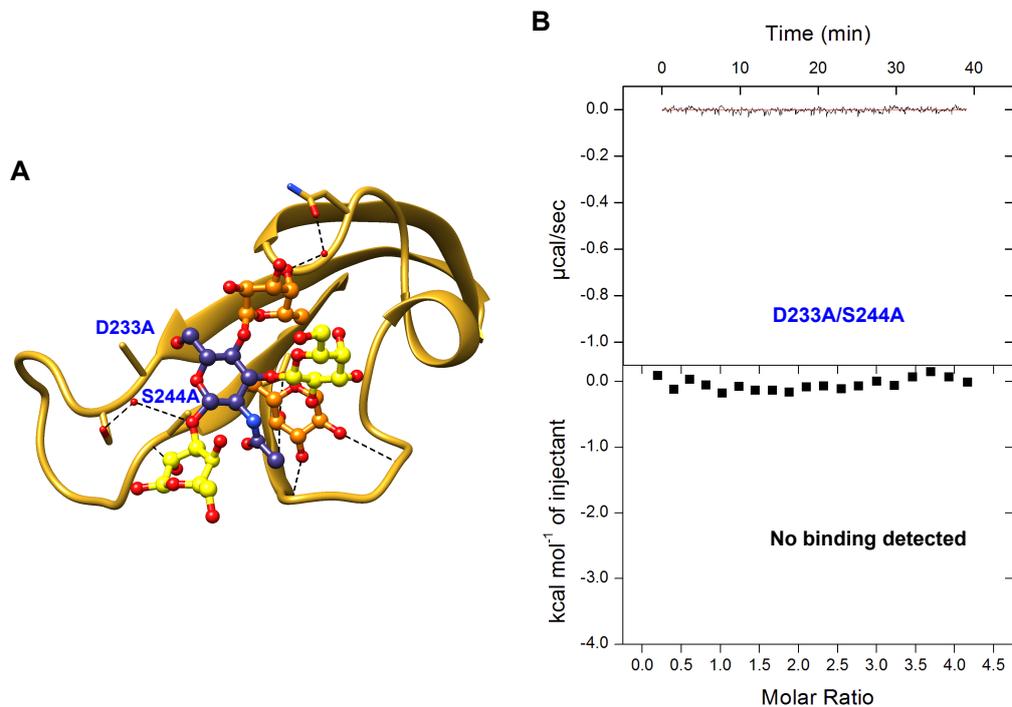


Figure 4.20: Binding affinity between the BabA_{527K}-D233A/S244A variant and Le^b
(A) Predicted interactions between BabA_{527K} with combined D233A/S244A substitutions and Le^b, based on the crystallographic model. The fucose, galactose and *N*-acetylglucosamine residues of Le^b are coloured orange, yellow and blue, respectively. Hydrogen bonds are represented by dotted black lines. **(B)** No

calorimetric response (upper panel) or binding isotherm (lower panel) was obtained by titrating BabA containing D233A/S244A substitutions with Le^b. The calorimetric titrations was performed at pH 7.4.

Having validated the BabA_{527K}:Le^b crystallographic model through BabA_{527K} variant binding studies, it was desirable to confirm that the identified binding site was responsible for BabA_{527K} recognition of the other type 1 ABO/Le blood group antigens. In support of this, no binding was detected, using ITC, between the BabA_{527K}-D233A/S244A variant and H-1, A-1 and B-1 antigens while wild-type BabA_{527K} bound these glycans, albeit with lower affinity than to Le^b (Table 4.3, Appendix Table 6 and 7). These structurally similar blood group antigens all lack the Le^b Fuc4 residue that forms a single water-mediated hydrogen bond with BabA_{527K} at N206. However, the A-1 and B-1 antigens have additional galactosamine and galactose residues, respectively, attached to Gal2.

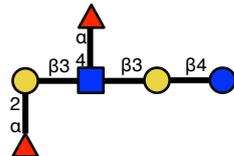
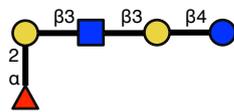
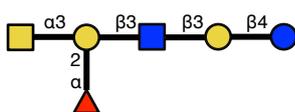
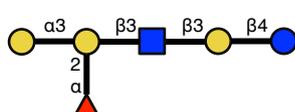
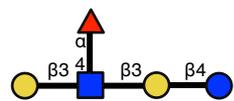
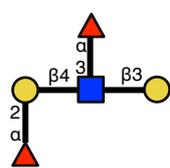
The type 1 ABO/Le blood group antigen binding specificity of BabA_{527K} was also confirmed through ITC. No binding was observed between BabA_{527K} and the Le^a antigen, which lacks the Le^b Fuc1 residue that forms direct hydrogen bonds with C189, G191, N194 and T246. Furthermore, no binding was observed between BabA and the closely related Le^y and H-2 antigens. Although Le^y and H-2 contain the same residues that were observed in our structural model to bind BabA as Le^b and H-1, respectively, they differ by having a Galβ1-4GlcNAc glycosidic linkage. Lastly, no binding was observed, using ITC, between BabA and the SLe^x antigen, which is recognised by SabA. This glycan lacks the Le^b Fuc1 residue and contains a terminal *N*-acetylneuraminic acid residue adjoined to a Galβ1-4GlcNAc core (Table 4.3, Appendix Table 6).

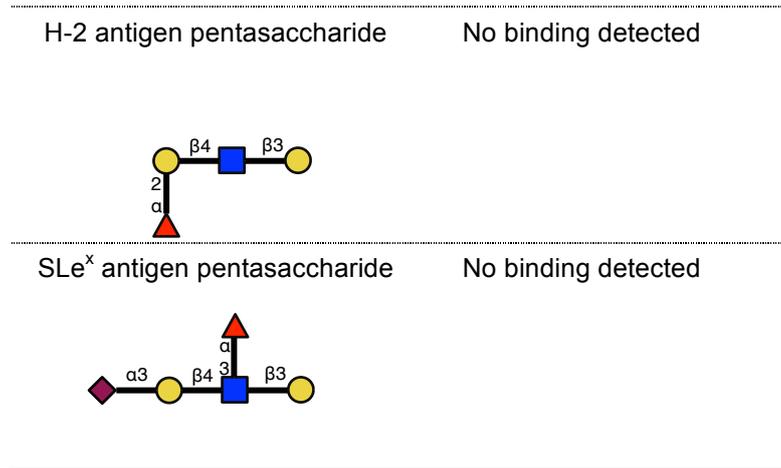
Table 4.3: Binding affinity of BabA_{527K} to various ABO/Le blood group antigens

Glycan symbolic representations can be interpreted with the following key: fucose – ▲, galactose – ●, glucose – ○, N-acetylgalactosamine – ■, N-acetylglucosamine – □, N-acetylneuraminic acid – ◆. Binding affinity was determined through calorimetric titrations performed at pH 7.4. Values are obtained from a single experiment, unless otherwise indicated.

^aAverage of three independent experiments.

^bAverage of two independent experiments.

ABO/Le blood group antigen	BabA _{527K} Binding Affinity
Le ^b antigen hexasaccharide	$K_D = 252 \mu\text{M}^a$
	
H-1 antigen pentasaccharide	$K_D = 617 \mu\text{M}^b$
	
A-1 antigen hexasaccharide	$K_D = 529 \mu\text{M}$
	
B-1 antigen hexasaccharide	$K_D = 416 \mu\text{M}$
	
Le ^a antigen pentasaccharide	No binding detected
	
Le ^y antigen tetrasaccharide	No binding detected
	



To better understand how Gal β 1-3GlcNAc (type 1) and Gal β 1-4GlcNAc (type 2) cores affect glycan recognition by BabA, three-dimensional Le blood group antigen models were calculated in minimum energy conformations with the SWEET-II system (224). As shown in Figure 4.21, this indicated that glycans with type 1 and type 2 cores markedly differ in their conformational orientation, and accordingly, in the positioning of their functional groups.

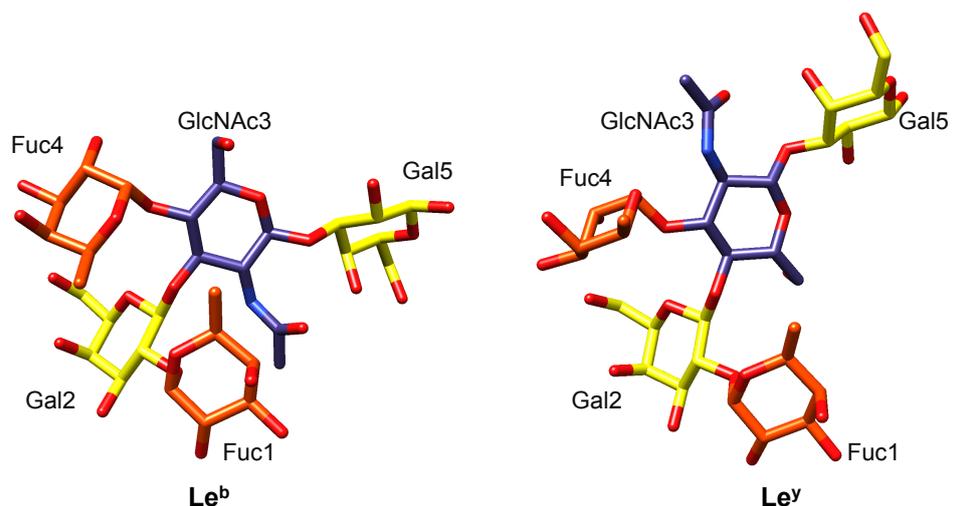


Figure 4.21: Type 1 and type 2 fucosylated ABO/Le blood group antigen molecular models

Stick models of Le^b (type 1) and Le^y (type 2) antigens show distinctly different three-dimensional orientations to due to Gal β 1-3GlcNAc and Gal β 1-4GlcNAc linkages, respectively. Fucose, galactose and *N*-acetylglucosamine residues are coloured orange, yellow and blue, respectively.

Finally, to confirm that the observed reduction and loss of BabA_{527K} binding affinity to the tested ABO/Le blood group antigens was a direct result of the alanine point substitutions, rather than a change to the global protein fold, circular dichroism spectroscopy was used to compare the secondary structure of BabA_{527K} variants against that of the wild-type. Neither the N206A nor the D233A/S244A mutations caused a change to the α -helical fold of BabA_{527K} (Figure 4.22A). In further support of this result, differential scanning fluorimetry revealed that neither mutation causes a major change to the conformational stability of BabA_{527K}, as indicated by the similarity in thermal unfolding temperatures between wild-type and variant proteins (Figure 4.22B).

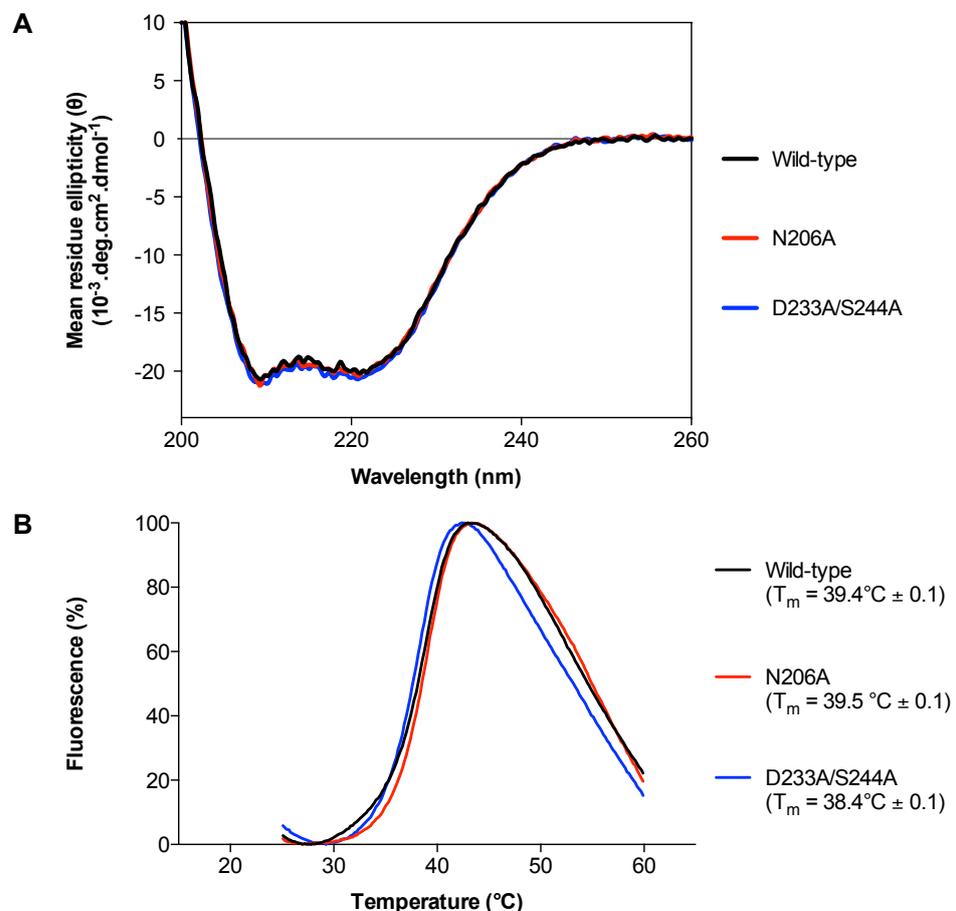


Figure 4.22: Secondary structure and thermal stability of wild-type and variant BabA_{527K} proteins

(A) Overlay of far-UV circular dichroism spectra averaged from three independent experiments. (B) Temperature-induced unfolding transition determined using

differential scanning fluorimetry. The reported midpoint temperature of each protein unfolding transition (T_m) is the average (\pm SEM) from three independent experiments.

Successfully generating a variant BabA_{527K} protein lacking ABO/Le blood group antigen affinity not only validates the crystallographic model but also offers a control to stringently identify BabA-specific targeting effects during the development of targeted drug delivery vectors.

4.4 Discussion

Following the successful expression and purification of recombinant BabA, the next objective of this study was to characterise its glycan-binding site. This information is crucial to ensure that recombinant BabA is correctly orientated when conjugated to the surface of drug delivery vectors such that it retains binding activity. However, before commencing structural studies, it was necessary to confirm that BabA_{527K} was functional and possessed its expected glycan-binding specificity. It has repeatedly been reported that *H. pylori* J99 cells recognise Le^b but not the closely related the Le^y antigen. Recent studies have also shown that both native and recombinant forms of BabA J99 maintain this characteristic (109, 112). Consequently, this binding selectivity was assessed in BabA_{527K} through an ELISA and SPR assay. Both techniques definitively confirmed that BabA_{527K} bound to HSA-Le^b but not to HSA-Le^y glycoconjugates, thereby suggesting that the protein was correctly folded and functional.

An important characteristic of *H. pylori* J99 cells is also the ability to attach to type 1 ABO/Le blood group antigens corresponding to blood group-A, -B and -O phenotypes (123). Through assessment of binding to a mammalian glycan array, this “generalist” glycan recognition profile was similarly found in BabA_{527K}. As also reported with *H. pylori* J99 cells, no binding was observed between BabA_{527K} and the wide range of type 2 ABO/Le blood group antigens present on the array. This further confirms the correct folding and binding activity of BabA_{527K}. Additionally, the retention of the characteristic features of *H. pylori* ABO/Le blood group antigen-binding specificity validates the efficacy of the recombinant strategy employed in generating a functional protein. The results of the mammalian glycan array screen also suggest that the BabA_{527K} glycan recognition site is highly specific in recognising type 1 ABO/Le blood

group antigens only, which greatly supports its use as a gastric- and glycan-specific targeting moiety. Unfortunately however, this assay does not provide definitive results regarding glycan specificity because immobilisation strategies can adversely affect the spatial orientation of glycans and thereby prevent their subsequent recognition by lectins in this assay (225); this was seen in BabA_{527K} binding to A-1, B-1, and H-1 antigens.

In contrast to the abundance of information in the literature regarding the BabA glycan binding profile, there was no published information describing BabA crystallisation or structure solution when this objective was first pursued. Nonetheless, despite difficulties in obtaining single crystals, which have since been similarly reported by Subedi *et al.* (177), the crystal structure of BabA_{527K} in the absence and presence of Le^b was successfully solved. With regards to expected features, these crystal structures confirm that recombinant BabA, which exists as a monomer, contains four intramolecular disulphide bonds. This provides structural evidence to support the choice of recombinant expression strategies in Chapter 3, which all facilitated this post-translational modification. With regards to unexpected features, the immediately apparent structural similarity between the extracellular domain of BabA J99 and that of SabA 26695 could be considered a surprise finding due to their low sequence identity. Perhaps this suggests that the three-dimensional folds of these proteins may be of functional significance. However, an in-depth analysis and understanding of the *H. pylori* outer leaflet is needed to aid the functional contextualisation and rationalisation of these adhesin structures. SabA does not contain the Crown β -sheet motif of BabA_{527K}, which was found in this study to possess its Le^b binding site. This structural difference fully explains the lack of SabA affinity to Le^b antigens (175). In line with previous studies, BabA_{527K} displayed no affinity for SLe^x antigens, which are conversely recognised by

SabA (226). Although a putative glycan-binding pocket in SabA has been suggested by Pang *et al.* (175), crystallographic insight into its molecular basis for glycan recognition is needed to understand the structural differences that distinguish binding of SLe^x antigens between these two important adhesins.

The Le^b binding interaction takes place in BabA_{527K} at the edge of its Crown region within a solvent-exposed binding site. The binding pocket is of a shallow topology, which is, in fact, typically observed in carbohydrate binding proteins and is known to result in few ligand contacts (227). As a result, lectins, including bacterial adhesins such as SabA, typically display low glycan binding affinity (175, 228). Low BabA_{527K}:Le^b affinity was similarly found in this study; these experiments were performed at pH 7.4 to reflect the near-neutral conditions found around surface mucous cells that host type 1 ABO/Le blood group antigens (229, 230). It must, however, be noted that any potential effects of the membrane-spanning domain of BabA on Le^b binding affinity are not captured. Furthermore, the multitude of BabA:Le^b interactions that occur between *H. pylori* and epithelium or mucin associated Le^b antigens, would exponentially increase the binding affinity of a single bacterium during colonisation. This effect of avidity on BabA:Le^b has been demonstrated in *H. pylori* J99 where bacteria bind multivalent Le^b glycoconjugates with substantially higher affinity ($K_D \sim 2$ pM) (123).

The structural model revealed that BabA, of the J99 strain, employs eight amino acids to bind to Le^b. Interestingly, these amino acids are highly conserved between strains, with the exception of N206, which interacts with Le^b Fuc4 and is located within the Hypervariable Crown Loop. The disparity in sequence identity and length observed in the Hypervariable Crown Loop may result in modified secondary structure folds in other *H. pylori* strains and

thereby affect the presentation of functional groups capable of interacting with Le^b Fuc4, or other residues. Accordingly, variability in this segment may be responsible, at least in part, for the differences seen in the affinity of *H. pylori* strains towards Le^b antigens (123, 124). Indeed, the interaction between the Hypervariable Crown Loop and Le^b, through N206 and Fuc4, did have an effect on binding affinity – an alanine point substitution at position 206 resulted in a 2.3-fold reduction in binding affinity. Furthermore, analysis of the binding interaction between BabA and the H-1 antigen supported this finding. H-1 lacks only the Fuc4 residue of Le^b and binds to BabA_{527K} with 2.4-fold lower affinity than Le^b. Thus, the data suggests that the interaction between the Hypervariable Crown Loop and Le^b Fuc4 can play a significant role in determining binding affinity.

Similar to the H-1 antigen, BabA_{527K} bound with lower affinity than Le^b to A-1 (2.1 fold) and B-1 (1.7 fold) antigens. It must be noted that the large sample amounts of BabA_{527K} required for binding analysis through ITC, in combination with the low protein yield after periplasmic expression, limited the ability to perform repeated A-1 and B-1 calorimetric titrations. As such, these reported dissociation constants would benefit from further validation. While this limits analysis of the data, the relatively weaker binding affinity observed, in comparison to Le^b, can still be considered rational because these glycans, like the H-1 antigen, lack the Le^b Fuc4 residue. However, A-1 and B-1 antigens do contain additional galactosamine and galactose residues, respectively; these extend out of Gal2 towards the Hypervariable Crown Loop. Accordingly, their relative affinities can only be explained with structural insight that reveals how these residues affect binding interactions with BabA_{527K}.

Further structural insight probing ABO/Le blood group A and B antigen binding could also aid in the understanding whether the Hypervariable Crown Loop has a role to play in the definition of “specialist” binding specificity. This characteristic of BabA binding was defined by Aspholm-Hurtig *et al.* upon the observation that certain South American Amerindian *H. pylori* strains only bound to the type 1 ABO/Le blood group antigens corresponding to blood group-O phenotypes but not those of group-A or group-B (123). Interestingly, an analysis of specialist strain Crown sequences, made publicly available by Aspholm-Hurtig *et al.*, revealed a markedly dissimilar and longer Hypervariable Crown Loop in comparison to that of the generalist J99 strain used in this study. It would be interesting to investigate whether the structural features formed by these sequences are able to modulate their BabA binding sites to prevent accommodation of the additional galactosamine and galactose residues found in ABO/Le blood group A and B antigens, respectively. Based on an extrapolation of the BabA_{527K}:Le^b structural model, these additional residues would extend out of Le^b Gal2 towards the Hypervariable Crown Loop. Given this proximity, the conformation of the Hypervariable Crown Loop in specialist strains could prove to be an important structural feature that creates the observed specificity in glycan binding.

No DNA or protein sequences with similarity to the BabA Crown were identified in *H. pylori* or any other organism. In this light, the Crown can be considered unique to BabA to enable specific attachment of *H. pylori* to type 1 ABO/Le blood group antigens in the human gastric mucosa. This is achieved through the network of hydrogen bonds presented, as partial deletions of these interactions can result in a loss of recognition:

- Combined alanine point substitutions to BabA_{527K} at D233 and S244, which form direct and water-mediated hydrogen bonds through their side chains to GlcNAc3 and Gal5, results in a complete loss of binding to Le^b. This occurs in spite of possible interactions between the Crown and Le^b through Fuc1 and Fuc4.
- These combined alanine point substitutions also result in a complete loss of BabA_{527K} binding affinity to H-1, A-1 and B-1 antigens, indicating a single glycan-binding site. This absence of affinity is observed despite possible direct hydrogen bonds BabA_{527K} and Fuc1 (based on an extrapolation of the BabA_{527K}:Le^b structural model), as these glycans do not contain Fuc4.
- However, BabA_{527K} contacts through D233 and S244, though necessary, are not sufficient to confer glycan recognition. The interaction between wild-type BabA_{527K} and Le^a shows that despite possible interactions between the Crown and GlcNAc3, Fuc4 and Gal5, the absence of only the Fuc1 residue results in a complete loss of binding.

Thus, no single sugar residue is responsible for glycan recognition by BabA_{527K}. Rather, it is the network of hydrogen bonds to multiple residues that forms the basis of molecular recognition. Consequently, it is no surprise that BabA_{527K} did not bind to the distantly related Le^y, H-2 and SLe^x antigens as they possess completely different three-dimensional conformations.

The structural and biophysical insight into the glycan binding mechanism of BabA_{527K} answers a previously elusive question of how BabA mediates *H. pylori* attachment to the gastric mucosa. While this is of substantial significance to researchers investigating the adaptations of *H. pylori* to the gastric environment, this insight immediately aids the rational design of BabA-

particle conjugates. For the development of proof-of-principle vectors, linkage of recombinant BabA to particle surfaces through its Handle region should enable glycan-binding activity to be retained given the relative location of its glycan-binding site at the tip of the Crown. Furthermore, the functional activity of BabA-vectors can now be stringently assessed by a comparison of its glycan binding properties with that of vectors hosting the variant BabA proteins developed through these experiments.

Chapter 5: Assessing BabA acid/pepsin stability and particle-conjugate activity

5.1 Introduction

Having successfully characterised the glycan binding mechanism of BabA, the final objective of this study focused on its use as a gastric targeting moiety through an assessment of its acid/pepsin stability and functional activity as a particle-conjugate. In order for BabA to direct drug delivery vectors to glycans on the gastric surface epithelium, the protein will first have to endure harsh conditions in the lumen of the stomach. This is mediated by gastric fluid, which is a denaturing acidic medium that, during digestion, also has an abundance of the proteolytic enzyme pepsin (231). Pepsin is an endopeptidase that has an optimal hydrolysis activity for peptide bonds within a pH range of 1-4. At pH 1.5 it exhibits ~90% of maximum activity but at pH 4.5 this is reduced to ~35% and at pH 8, it is irreversibly inactivated (232, 233). In comparison to other gastrointestinal proteases, pepsin has a broad specificity, which is discussed in detail in (234) and can be estimated using bioinformatics tools. Cleavage probability is uniquely determined by the identity of both amino acid residues on either side of the peptide bond. However, pepsin preferentially cleaves peptide bonds adjoined to bulky, hydrophobic residues, such as phenylalanine, tryptophan, tyrosine or leucine and it will not cleave at valine, alanine or glycine linkages (235). Although pepsin is the principal proteolytic enzyme of gastric fluid, not all proteins (e.g. β -lactoglobulin) are susceptible to peptic digestion because cleavage sites can be rendered inaccessible by conformational folds (236). In contrast to the harsh conditions of the lumen, the mucus layer that overlays the gastric epithelium is a more favourable microenvironment. It is characterised by a pH gradient from acidic at the luminal interface, to neutral at the epithelial surface, due to mucosal

bicarbonate secretions (229, 237, 238). The surface mucus layer, which also withstands luminal pepsin penetration, is principally composed of the MUC5AC mucin (100, 239). Because this glycoprotein carries type 1 ABO/Le blood group antigens, it may also act as receptor for BabA-vectors as it does for *H. pylori* (Figure 5.1) (38, 240, 241). In addition to these potential physiological barriers, BabA must be immobilised to drug delivery vectors in a manner that ensures exposure of its glycan-binding site. Several different methods can be employed to attach proteins to drug delivery vectors; these approaches are very much dependent on the material that constitutes the vector's core as conjugation strategies are based on particle surface chemistry. As reviewed in (242), while established methods are readily applicable, a poor choice of conjugation strategy can impair the functional activity of particle-conjugates not only due to inappropriate orientation but also due to loss of protein tertiary structure after immobilisation. As such, a robust assessment of functional activity is always needed to confirm the applicability of any protein in a particle-conjugate system.

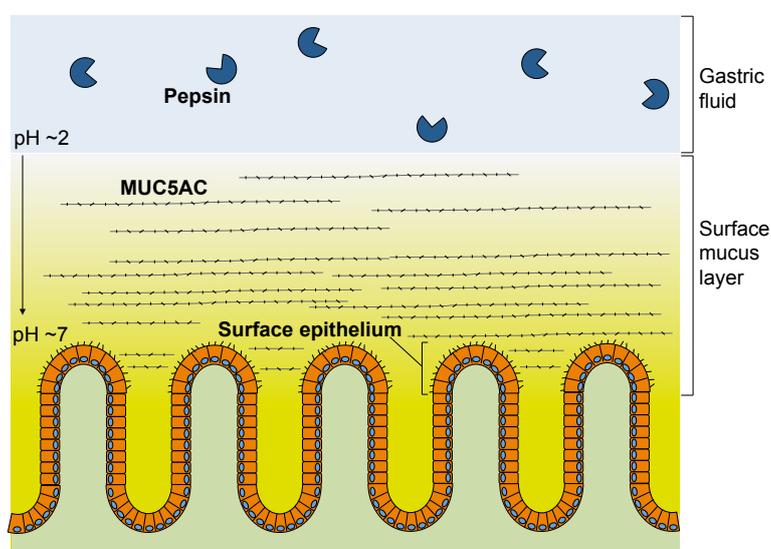


Figure 5.1: Schematic representation of the gastric lumen and mucus layer

MUC5AC is the major mucin constituent of the acid- and pepsin-resistant gastric surface mucus layer (207). BabA receptors, that is, type 1 ABO/Le blood group antigens, are found attached to both MUC5AC and surface epithelial cells (indicated).

Despite its important role in mediating gastric colonisation, the susceptibility of BabA to peptic digestion has not been published. Additionally, the effects of acidic conditions on its stability have not been described in the literature. It is also not known how pH affects the glycan binding affinity of BabA (92), however, this defines how BabA interacts with type 1 ABO/Le blood group antigens in the distal, acidic portion of the surface mucus layer compared to those found on the surface epithelial cells and on MUC5AC in the juxtamucosal, near-neutral region. Finally, no studies have described the immobilisation of BabA onto a particle surface such that it remains functional – this is clearly a key consideration for mediating the proposed biomimicry.

Therefore, to study the effect of the low pH found in the stomach lumen and mucus layer on recombinant BabA, its conformation and functionality was characterised at acidic conditions. Subsequently, its resistance to proteolytic cleavage by pepsin was evaluated in comparison to β -lactoglobulin – a protein known to possess intrinsic resistance to pepsin (243). Finally, BabA-particle conjugates (polystyrene-based microparticles were used as model particles, however, it must be noted that their non-biodegradability prevents their *in vivo* applicability) were created and studied in a binding assay. Importantly, this array of information provides essential insight needed to make a preliminary assessment of the suitability of BabA as a targeting moiety, and if applicable, lay a foundation for the future development of BabA-vectors.

Though the conformational stability and functionality of recombinant BabA was not adversely affected by acidic pH, the protein displayed limited resistance to pepsin. However, BabA-particle conjugates were successfully generated and found to be functional. This represents a significant step towards achieving the translational goal of this project.

5.2 Experimental Procedures

5.2.1 Acidic pH conformation and functionality studies

5.2.1.1 Circular dichroism spectroscopy

Circular dichroism spectra of BabA_{527K} was measured and analysed as detailed in Section 4.2.3.3. Protein concentration was 1 μ M in a buffer containing 50 mM KH₂PO₄ (pH = 4.5) or 20 mM Tris-Cl (pH = 7.4) and 3 mM NaCl.

5.2.1.2 Differential scanning fluorimetry

SYPRO Orange dye (20x final concentration) was added to 10 μ M BabA_{527K} in a buffer containing 50 mM KH₂PO₄ (pH = 4.5) or 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl. Changes in fluorescence were measured across an increasing temperature gradient from 20 °C to 80 °C using a iCycler iQ Real-Time PCR Detection System (Bio-Rad, USA) at a ramp rate of 0.01 °C/s. Primary data points from three independent experiments were fitted to a 6-parameter unfolding equation (222) using the Prism analysis package (GraphPad Prism 6 Software, USA), as detailed in Section 4.2.3.4.

5.2.1.3 Isothermal titration calorimetry

Calorimetric measurements were measured and analysed as detailed in Section 4.2.3.2. The sample cell was filled with BabA_{527K} at a concentration of 0.1 mM and the injection syringe contained the Le^b antigen hexasaccharide at a concentration of 2 mM. A buffer containing 50 mM KH₂PO₄ (pH = 4.5) was used for calorimetric measurements.

5.2.2 Peptic digestion tests

5.2.2.1 Prediction of pepsin cleavage sites

Amino acid sequences of BabA_{527K} and β -lactoglobulin (Genbank accession no. AAI08214.1) were submitted to ExPASy PeptideCutter [Swiss Institute of Bioinformatics, (244)] for a prediction of potential cleavage sites by pepsin at pH > 2 – at pH 1.3, pepsin reportedly cleaves peptide bonds where the N-terminal residue is a phenylalanine or leucine and displays negligible cleavage for other amino acids in this position; at pH >2, this specificity is lost (245). Predicted cleavage sites were mapped onto apo-BabA_{527K} and β -lactoglobulin (PDB Accession code: 3BLG) crystal structures.

5.2.2.2 Peptic digestion

One μ L of protein sample, at a concentration of 2 mg/mL, was pipette-mixed with 9 μ L of 50 mM KH₂PO₄ (pH = 4.5) containing pepsin from porcine gastric mucosa at a concentration of 750 units/mL (Sigma-Aldrich, USA), then immediately incubated for variable durations for up to two hours at 37 °C. At desired time points, samples were quenched through the addition of 10 μ L of 1 M Tris-Cl (pH = 10.0). Zero time point samples were prepared through the addition of the quenching solution to the pepsin-containing buffer before the addition of protein samples. The following protein samples were assessed:

- BabA_{527K}
- β -lactoglobulin from bovine milk (Sigma-Aldrich, USA)
- Native BabA from *H. pylori* strain CCUG 17875 (a kind gift from Thomas Borén, University of Umeå)

5.2.2.3 Analysis of digested protein samples

20 µL of 2x NRSB was added to each quenched protein sample, followed by heat denaturation at 100 °C for five minutes then separation by SDS-PAGE as described in Section 3.2.2.8. Digested protein samples were subsequently analysed using various techniques:

- **Coomassie-staining**

Gels containing BabA_{527K} and β-lactoglobulin samples were analysed with Instant Blue[®] stain (Expedeon, USA), according to the manufacturer's instructions.

- **Silver staining**

Gels containing native BabA samples were analysed using a PlusOne silver staining kit (GE Healthcare, USA), according to the manufacturer's instructions.

- **Western blotting**

Gels containing BabA_{527K} samples were blotted onto a nitrocellulose membrane and analysed via c-Myc immunodetection, as described in Section 3.2.2.8.

- **Tryptic peptide mapping**

Experimental procedures performed by Melanie Snow, AstraZeneca R&D

Following BabA_{527K} peptic digestion, an excised gel slice containing its uppermost digested fragment was subjected to tryptic peptide mapping, as described in Section 3.2.3.6.

5.2.3 Generation and assessment of BabA-particle conjugates

Wild-type and D233A/S244A-variant BabA_{527K} proteins were conjugated to microparticles by exploiting the avidin:biotin interaction, as described below.

5.2.3.1 Biotinylation of BabA_{527K} and BabA_{527K}-D233A/S244A

Protein biotinylation was performed through *in vitro* enzymatic biotinylation by the *E. coli* biotin ligase BirA, which specifically recognises and conjugates a single biotin molecule to the AviTag™ peptide sequence (GLNDIFEAQKIEWHE) (246). Using the primers shown in Table 5.1, a Phusion site-directed mutagenesis kit (Thermo Scientific, USA), was used to insert the AviTag™ peptide sequence, in between the c-Myc and hexahistidine tag of i) the pOPE101_*babA*_{527K} expression construct, and ii) the pOPE101_*babA*_{527K} expression construct containing combined D233A/S244A mutations. The PCR reaction was set up according to the manufacturer's protocol using an annealing temperature of 72 °C for both reactions. Confirmatory sequencing of the plasmids was performed (Source BioScience, Nottingham) using primers listed in Table 3.3.

Table 5.1: Primers for AviTag™ peptide sequence insertion into BabA_{527K} expression constructs

[FOR] and [REV] denote sense and antisense primers, respectively. [Phos] denotes a 5' phosphorylation.

Primer Name	Sequence (5'-3')
pOPE101-AviTag [FOR]	[Phos]AAAAGATCGAATGGCATGAATCCCATC ATCACCATCATC
pOPE101-AviTag [REV]	[Phos]GAGCTTCAAAAATATCATTTCAGTCCTA GATCTTCTTCTGAGATCAGC

Expression and purification was performed under the same conditions as BabA_{527K}, as described in Section 3.2.3. Subsequently, approximately 1 mg of both wild-type BabA_{527K} and BabA_{527K}-D233A/S244A, containing AviTag™

peptide sequences, were biotinylated using a BirA biotin-protein ligase kit (Avidity, USA), according to the manufacturer's protocol. Per 1 mg of protein, complete biotinylation was achieved after incubation at room temperature for two hours with 15 µg of BirA in a total reaction volume of 0.5 mL; complete biotinylation was verified using liquid chromatography–time-of-flight mass spectrometry, as described in Section 3.2.3.5. Biotinylated recombinant BabA proteins are henceforth referred to as BabA_{WT}[Biotin] and BabA_{D233A/S244A}[Biotin] for the wild-type form and D233A/S244A-variant form, respectively. Both proteins were separated from BirA via IMAC, as described in Section 3.2.3.2, using gravity columns instead of an ÄKTA purifier system. Protein-containing elutions were dialysed overnight into a buffer containing 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl. Finally, proteins were concentrated, as required, using Vivaspin sample concentrators (GE Healthcare, USA).

5.2.3.2 Enzyme-linked immunosorbent assay

ELISA quantification of BabA_{WT}[Biotin] and BabA_{D233A/S244A}[Biotin] binding to Le^{b/y} glycoconjugates was performed with modifications to the protocol detailed in Section 4.2.1.1, that is, the concentration range tested was 0.1-1.0 µg/mL and proteins were detected using streptavidin-HRP only, at a 1:10000 dilution.

5.2.3.3 Conjugation of recombinant BabA to microparticles

To optimise surface coating, a NeutrAvidin microparticle suspension, containing 50 µg solids in 5 µL of water, was mixed with ranging amounts (1.0, 2.5 and 5.0 µg) of BSA, BabA_{WT}[Biotin], and BabA_{D233A/S244A}[Biotin] in 5 µL of 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl. The protein-microparticle mixtures were incubated for 15 minutes at room temperature then centrifuged

at 15,000 g for 5 minutes. The resulting supernatant was then carefully aspirated, mixed with an equal volume of 2x NRSB and heat denatured. Finally, samples were separated by SDS-PAGE and the supernatant protein concentration was analysed by Coomassie-staining using Instant Blue[®] stain. Per 50 µg of NeutrAvidin microparticle solids, the optimum amount of both BabA_{WT}[Biotin] and BabA_{D233A/S244A}[Biotin] chosen for conjugation, via incubation, was 2.5 µg. The designations and descriptions for microparticles and protein-microparticle conjugates used or prepared in this study are summarised below in Table 5.2.

Table 5.2: Microparticles used in this study

The chosen designations indicate the outermost layer of the microparticles.

^aObtained from Life Technologies, USA

^bGenerated in this study

Designation	Description
Polystyrene ^a	Carboxylate-modified 1.0 µm microspheres with yellow-green fluorescence
NeutrAvidin ^a	NeutrAvidin [®] -labelled 1.0 µm microspheres with yellow-green fluorescence
BabA _{WT} ^b	NeutrAvidin [®] -labelled 1.0 µm microspheres with yellow-green fluorescence surface modified with BabA _{WT} [Biotin]
BabA _{D233A/S244A} ^b	NeutrAvidin [®] -labelled 1.0 µm microspheres with yellow-green fluorescence surface modified with BabA _{D233A/S244A} [Biotin]

5.2.3.4 Dynamic light scattering

Microparticle hydrodynamic diameters and polydispersity indices (PDI) were measured by DLS in a Zetasizer NanoZS instrument (Malvern, UK) at 25 °C. Microparticle suspensions, at a solids concentration of 25 µg/mL, were studied in a 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl buffer. The following microparticles were analysed: i) polystyrene-microparticles; ii) NeutrAvidin-

microparticles; iii) BabA_{WT}-microparticles and iv) BabA_{D233A/S244A}-microparticles. Each sample was measured in triplicate.

5.2.3.5 Bio-layer interferometry

The Octet[®] RED96 System (ForteBio, USA) was used for BLI studies; assays were performed in black 96-well plates (Thermo Scientific, USA) at 25 °C. At each assay step, the total sample or buffer volume in each well was 200 µL and a plate shaking speed of 1000 rpm was employed. The assay was designed to use amine reactive biosensor tips (ForteBio, USA). Amine reactive biosensor tips were first hydrated in water for 60 seconds followed by activation of their carboxymethyl groups to reactive NHS esters through a 300 second incubation with 0.2 M EDC / 0.05 M NHS. Le^b and Le^y HSA-glycoconjugates (Isosep AB, Sweden; Appendix Table 2), at a concentration of 5 µg/mL in 100 mM MES (pH = 4.3), were then immobilised to the tip surfaces through a 600 second incubation. Le^{b/y} antigen modified tips were washed for 120 seconds in a 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl buffer, followed by baseline equilibration for 120 seconds in the same buffer. Binding of test proteins and microparticles, to both Le^b and Le^y HSA-glycoconjugate modified amine reactive tips was assessed using association and dissociation times of 300 and 900 seconds, respectively. Test proteins – BabA_{527K} and BabA_{527K}-D233A/S244A – were studied at a concentration of 12.5 µg/mL whereas polystyrene-, NeutrAvidin-, BabA_{WT}- and BabA_{D233A/S244A}-microparticles were studied at a solids concentration of 250 µg/mL. Microparticle suspensions were directly prepared in the 96-well plates.

5.3 Results

5.3.1 Effect of acidic pH on BabA_{527K} conformation and functionality

To assess the effect of acidic pH on the folding of BabA_{527K}, circular dichroism spectroscopy was first used to study changes to its secondary structure. As shown in Figure 5.2A, acidic conditions do not disrupt the α -helical conformation that is characteristic of BabA_{527K} at neutral conditions. Though no effects on BabA_{527K} folding are apparent, differential scanning fluorimetry revealed that the lower pH buffer actually prolongs BabA_{527K} thermal-induced unfolding (Figure 5.2B).

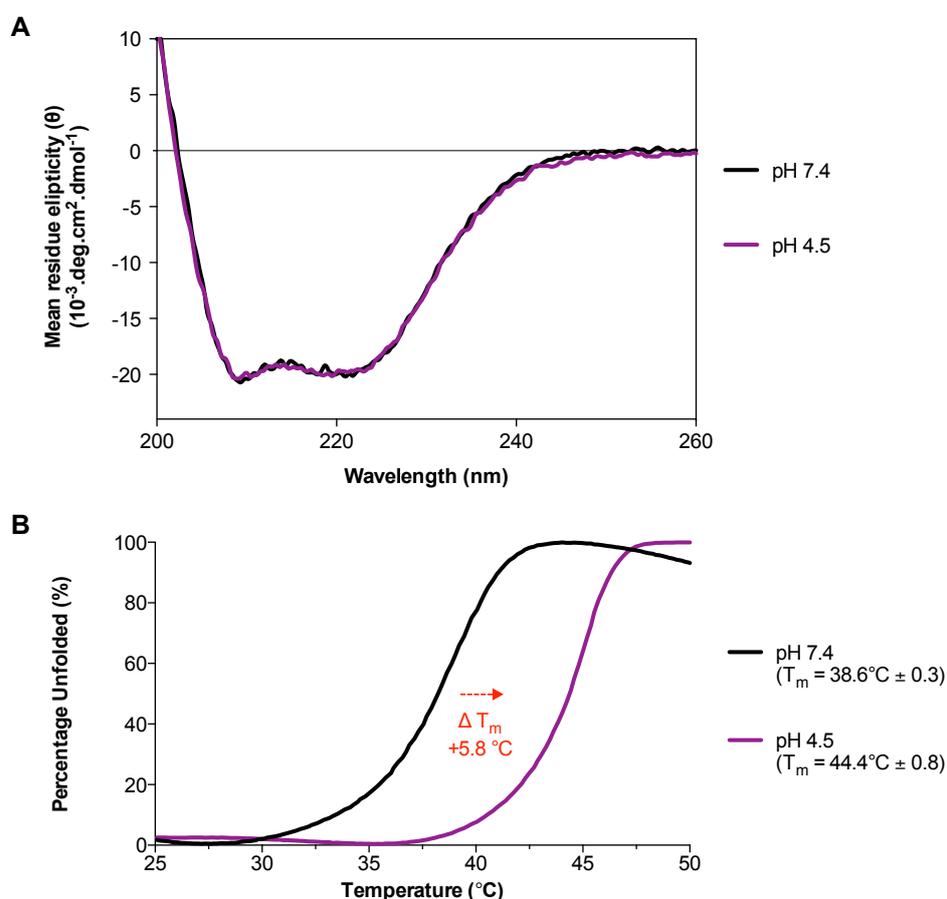


Figure 5.2: Secondary structure and thermal stability of BabA_{527K} at neutral and acidic pH

(A) Overlay of far-UV circular dichroism spectra averaged from three independent experiments. (B) Temperature-induced unfolding transition determined using differential scanning fluorimetry. The reported midpoint temperature of each protein unfolding transition (T_m) is the average (\pm SEM) from three independent experiments.

To further investigate whether acidic conditions adversely affected BabA_{527K} conformation, and thereby functionality, binding affinity to Le^b was measured through ITC. As determined by an unpaired two-tailed Welch's t-test, there are no significant differences ($p > 0.05$) in binding, neither between the dissociation constants nor the thermodynamic parameters of the interaction, at neutral or acidic pH (Figure 5.3).

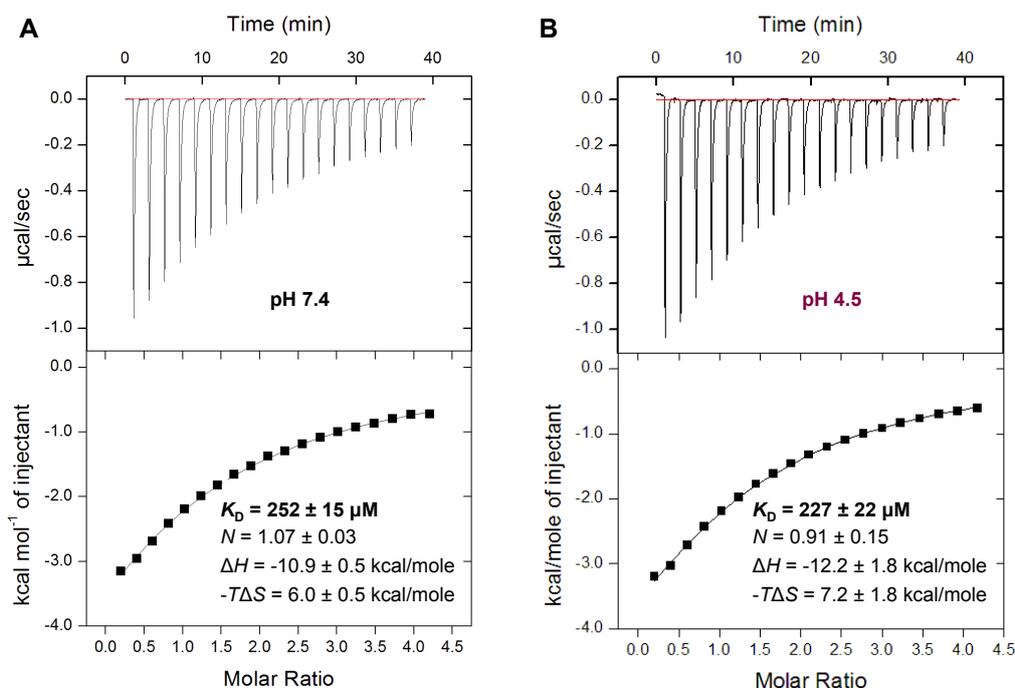


Figure 5.3: Binding affinity between BabA_{527K} and Le^b at neutral and acidic pH

Calorimetric response (upper panel) and binding isotherm (lower panel) of BabA_{527K} titrated with Le^b at **(A)** neutral and **(B)** acidic pH. The continuous line in the lower panels represents the least-squares fit of the data to a single-site binding model. The reported thermodynamic parameters are the average (\pm SEM) of three independent experiments.

To better understand the potential effect of the gastric environment on BabA_{527K}, its resistance to the proteolytic activity of pepsin was evaluated next.

5.3.2 Resistance of BabA_{527K} to peptic degradation

To characterise resistance to peptic digestion, BabA_{527K} was studied alongside bovine β -lactoglobulin, which has been shown *in vitro* to be intrinsically resistant to the proteolytic activity of pepsin. Analysis of the amino acid sequences of both proteins indicated that they both contain several predicted pepsin cleavage sites (Figure 5.4).

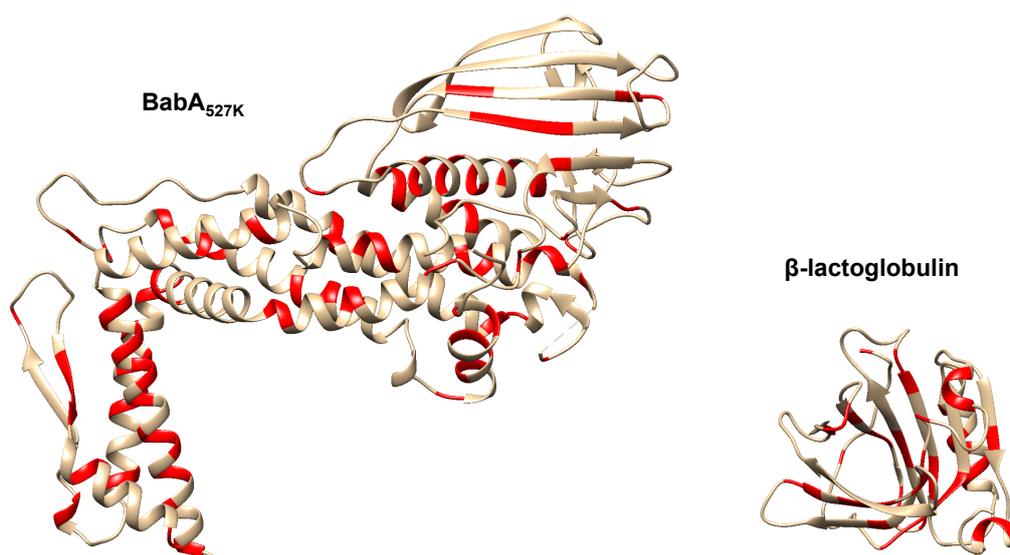


Figure 5.4: Crystal structures of BabA_{527K} and β -lactoglobulin showing potential pepsin cleavage sites

The potential pepsin cleavage sites (indicated in red) mapped onto BabA_{527K} (left) and β -lactoglobulin (right) crystal structures are based on the peptide bond cleavage probability of their amino acid sequences.

These predicted cleavage sites may be rendered inaccessible by the protein tertiary fold. To determine this empirically, peptic digestion for both proteins was performed, at pH 4.5, in a time course experiment where samples were analysed by SDS-PAGE and Coomassie-staining after incubation with pepsin. BabA_{527K} was rapidly degraded by pepsin; digestion was characterised by the immediate generation of three fragments, however, after incubation with pepsin for two minutes, no intact or fragmented BabA_{527K} is visible on the Coomassie-stained gel (Figure 5.5A). In stark contrast, no peptic digestion of

β -lactoglobulin was observed after incubation with pepsin for two hours, as expected (Figure 5.5B).

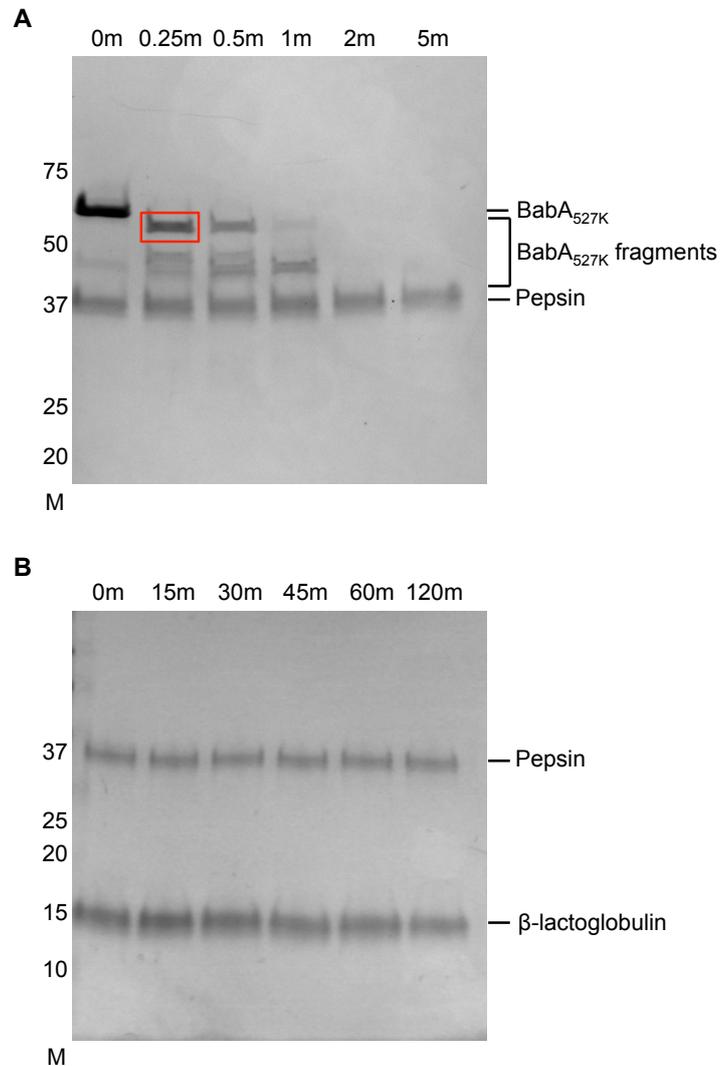


Figure 5.5: Peptic digestion pattern of BabA_{527K} and β -lactoglobulin

A time-course of **(A)** BabA_{527K} and **(B)** β -lactoglobulin peptic digestion at pH 4.5, analysed by Coomassie-staining. BabA_{527K} resistant fragments are indicated; the largest identifiable fragment, highlighted by the red box, was excised for tryptic peptide mapping. M – molecular weight marker. Time points indicated are in minutes.

Next, as also indicated in Figure 5.5A above, the largest BabA_{527K} peptic fragment was analysed by tryptic peptide mapping in an attempt to identify the cleaved regions of BabA_{527K}. Peptide masses were identified only between G25 and K392 – this corresponds to the start of the α -N helix in the Handle

region to the loop connecting the α -3 and α -4 helices in the Head region. No C-terminal sequences thereafter were detected (Figure 5.6).

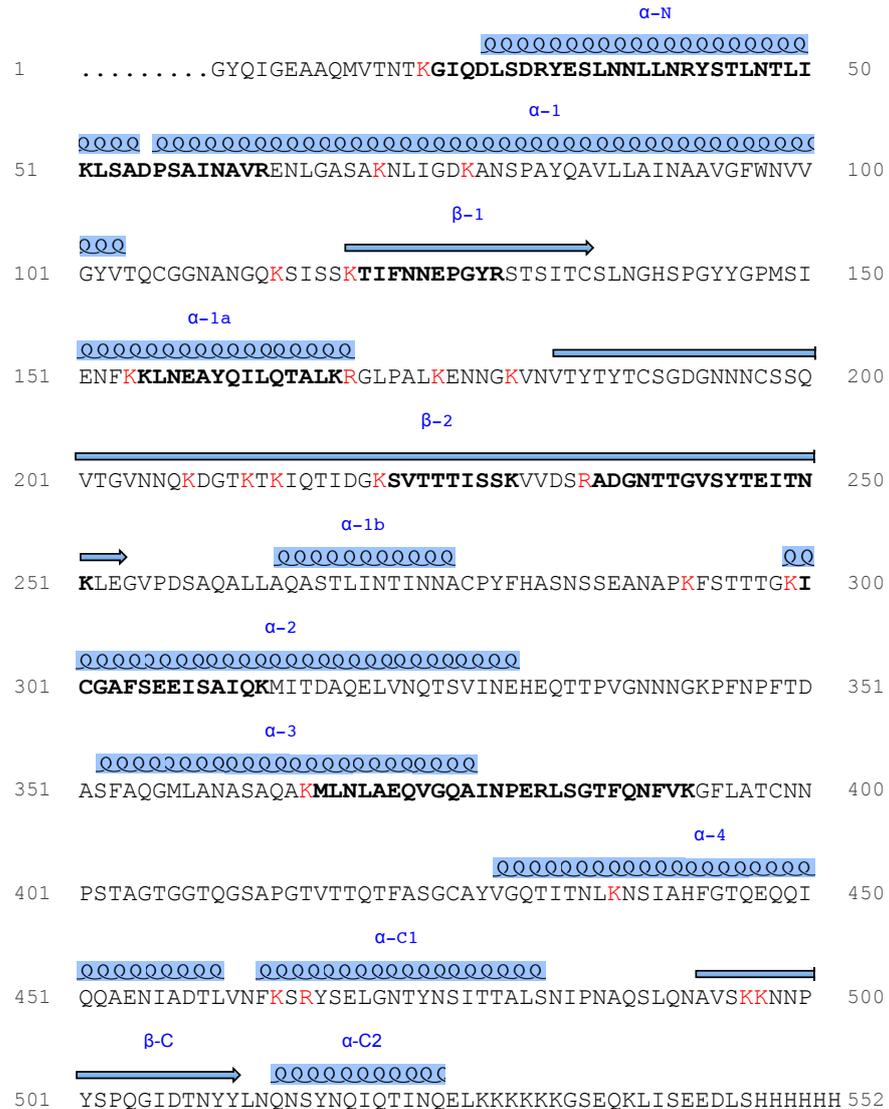


Figure 5.6: BabA_{527K} amino acid sequence showing peptides matched in the largest peptic digestion fragment

Secondary structure elements of BabA_{527K} are indicated. Peptides shown in bold were identified through mass matching, after liquid chromatography–quadrupole time-of-flight mass spectrometry. Trypsin cleavage sites [C-terminal side of lysine (K) or arginine (R) unless next residue is proline (P)] are highlighted in red unless part of a matched peptide. The black dots at the N-terminus represent the first nine amino acids of BabA_{527K}, which are cleaved during expression.

In further support of the apparent C-terminal cleavage, analysis of BabA_{527K} peptic digestion through immunodetection, rather than Coomassie-staining, indicated that none of the observed fragments contained the C-terminal c-Myc tag (Figure 5.7).

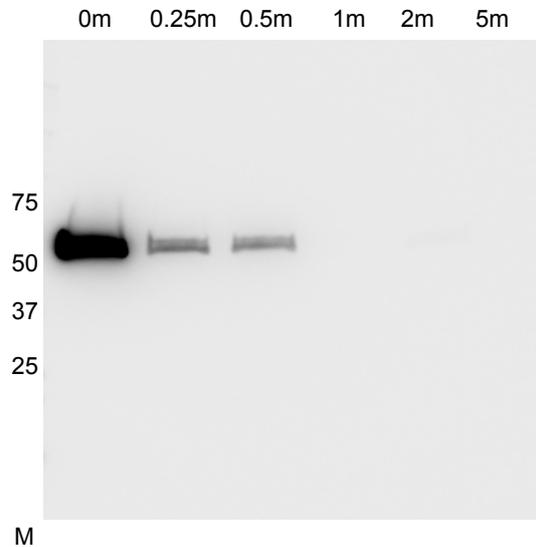


Figure 5.7: c-Myc immunodetection of the BabA_{527K} peptic digestion pattern

A time-course of BabA_{527K} and peptic digestion at pH 4.5 analysed by c-Myc immunodetection. M – Molecular weight marker. Time points indicated are in minutes.

To ascertain whether the lack of BabA_{527K} resistance to peptic degradation was an unexpected drawback of recombinant expression, native BabA, purified from the outer membrane of *H. pylori* CCUG 17875 (a kind gift from Thomas Borén, University of Umeå), was studied using the same *in vitro* assay. As with BabA_{527K}, native BabA was degraded by pepsin, though complete digestion required 30-60 minutes and no BabA fragments were detected (Figure 5.8). It was also verified that incubation of recombinant and native BabA in the same acidic buffer, without pepsin, did not result in any detectable proteolysis (results not shown).

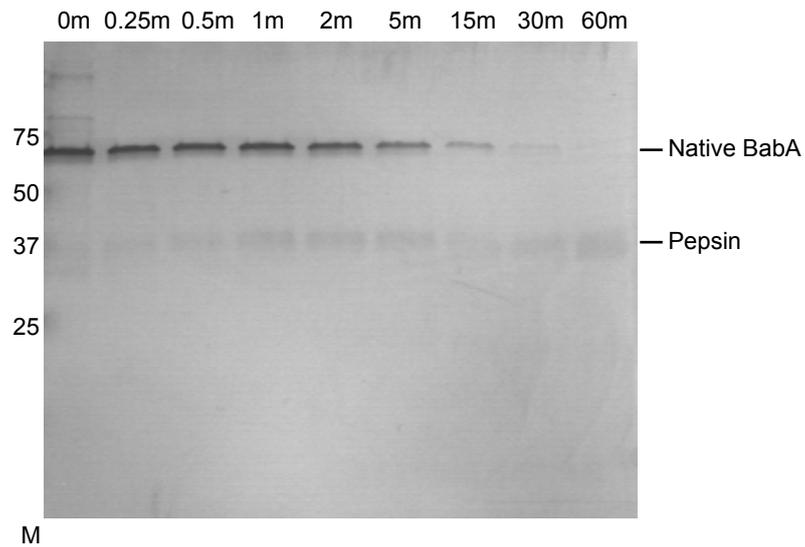


Figure 5.8: Peptic digestion pattern of native BabA

A time-course of native BabA (from *H. pylori* strain CCUG 17875) peptic digestion, at pH 4.5, analysed by silver staining. M – molecular weight marker. Time points indicated are in minutes.

In order to obtain further insight into the suitability of BabA as a gastric-targeting moiety, proof-of-principle BabA-microparticle conjugates were next generated and evaluated.

5.3.3 Functionality of BabA-particle conjugates

The initial strategy chosen for BabA_{527K} conjugation to microparticles was through the carbodiimide coupling method, where carboxylated particle surfaces are chemically converted into reactive NHS-esters to bind free protein amine groups. Indeed, the intended use of the C-terminal hexa-lysine tag in BabA_{527K} was to facilitate this reaction. However, preliminary conjugation attempts with this method resulted in irreversible microparticle aggregation (results not shown). This was attributed to the likely activation of multiple free amine groups found on the side chains of lysine residues, which are widely distributed over the BabA_{527K} surface (Figure 5.9) and could consequently act as attachment sites for more than one microparticle, effectively leading to cross-linking and aggregation.

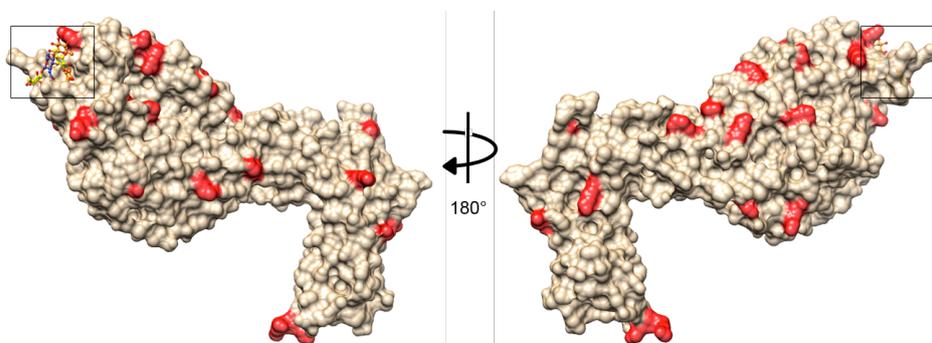


Figure 5.9: Location of lysine residues on the surface of BabA_{527K}

Surface representation of BabA_{527K} (tan) with lysine residues indicated (red). The BabA_{527K} glycan-binding site is indicated by a black square.

Thus, to circumvent aggregation, which hinders the handling and functional characterisation of particle suspensions, the avidin-biotin interaction was exploited. This interaction was to be used to mediate conjugation through the immobilisation of biotinylated, recombinant BabA onto the surface of 1 μm microparticles labelled with deglycosylated avidin (NeutrAvidin). Because with this strategy mono-biotinylated, recombinant BabA protein could be generated, it was hoped that each protein would only contain one microparticle

attachment site. The avidin-biotin interaction was chosen in order to facilitate the rapid generation of a colloiddally stable BabA-microparticle suspension and does not represent the intended strategy for future *in vivo* application.

To generate mono-biotinylated protein, the AviTag™ peptide sequence was first introduced into the C-terminus of BabA_{527K} and BabA_{527K}-D233A/S244A, in between their c-Myc and 6x His tags, to enable *in vitro* enzymatic biotinylation by BirA. An ELISA confirmed AviTag™ insertion and biotinylation did not disrupt the selective binding and non-binding properties of the wild-type and D233A/S244A-variant recombinant BabA proteins, respectively (Figure 5.10).

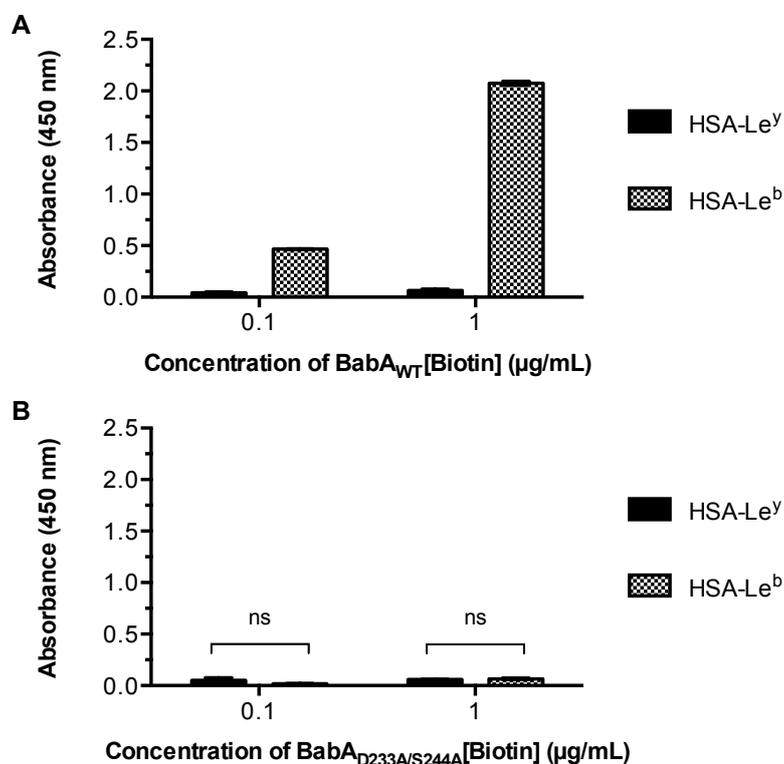


Figure 5.10: Binding of biotinylated recombinant BabA proteins to HSA-Le^{b/y} in an ELISA

Binding of biotinylated **(A)** wild-type and **(B)** D233A/S244A-variant recombinant BabA proteins to immobilised HSA-Le^y and HSA-Le^b glycoconjugates was determined using a sandwich ELISA (n=3, error bars represent the standard error of the mean). ns = not significant ($p > 0.05$), unpaired two-tailed Welch's t-test.

Next, the selectivity of protein capture by NeutrAvidin-microparticles was assessed. In support of the use of the chosen immobilisation strategy, Figure 5.11A shows BSA does not non-specifically bind to NeutrAvidin-microparticles. On the other hand, at the same range of concentrations tested, attachment of BabA_{WT}[Biotin] is clearly observed (Figure 5.11B). 2.5 µg of BabA_{WT}[Biotin] and BabA_{D233A/S244A}[Biotin] was used per 50 µg of NeutrAvidin-microparticles in all subsequent experiments to achieve saturation of the microparticle surfaces.

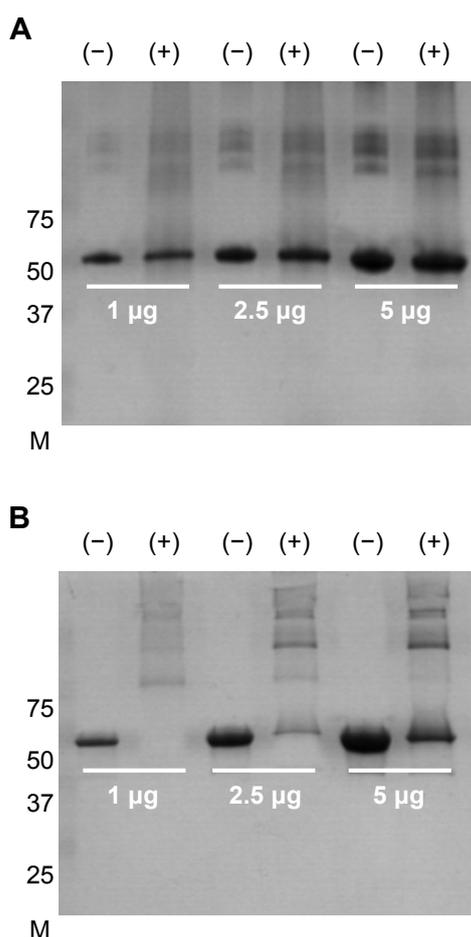


Figure 5.11: Optimisation of protein:microparticle conjugation

Various amounts of **(A)** BSA and **(B)** BabA_{WT}[Biotin] were incubated with NeutrAvidin-microparticles; unbound proteins were separated from protein-particle conjugates through centrifugation then analysed by SDS-PAGE and Coomassie-staining. (-) indicates the stated amount of protein loaded directly onto the gel, representing 100% unbound protein. (+) indicates the unbound protein retrieved after the stated amount of protein had been incubated with 50 µg of NeutrAvidin-microparticles.

To assess the effect of BabA_{WT}[Biotin] and BabA_{D233A/S244A}[Biotin] conjugation on the colloidal stability of NeutrAvidin-microparticles, DLS was used to measure changes in the hydrodynamic diameter of microparticles after surface modification. As shown in Figure 5.12 below, the hydrodynamic diameter of NeutrAvidin-microparticles is increased by ~67 nm and ~84 nm after conjugation of BabA_{WT}[Biotin] and BabA_{D233A/S244A}[Biotin], respectively. This modest change in hydrodynamic diameter is reflective of an increase in microparticle size due to an additional surface layer rather than agglomeration. As a reference, NeutrAvidin-labelled microparticles were compared to their unlabelled form, that is, polystyrene-microparticles. A larger hydrodynamic diameter due to this surface labelling was similarly observed; NeutrAvidin-microparticles were ~56 nm larger than polystyrene-microparticles. The low PDI of all the microparticle suspensions tested further indicates that all samples were monodisperse. As such, they were deemed suitable for functional studies.

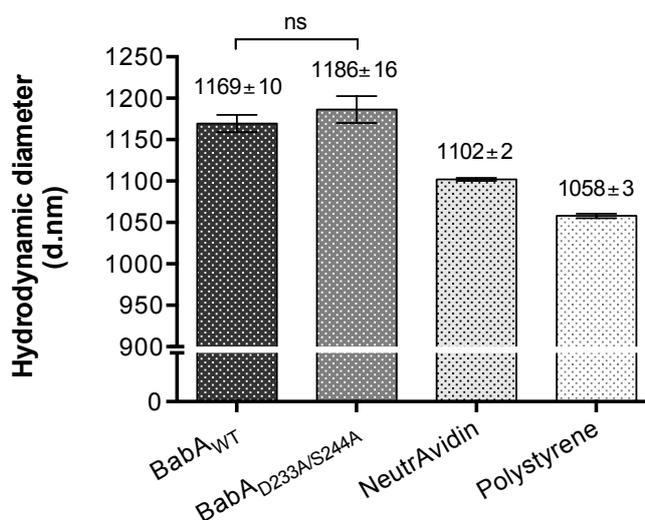


Figure 5.12: Comparison of the hydrodynamic diameters of microparticles with various surface modifications, as determined by DLS

DLS measurements were performed in a 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl buffer at 25 °C. The indicated hydrodynamic diameters are the average [± SEM (error bars)] of three independent experiments. All suspensions had a PDI < 0.15. ns = not significant ($p > 0.05$), unpaired two-tailed Welch's t-test.

To assess whether the BabA-microparticle conjugates retained the functional properties of recombinant BabA, binding to Le antigens was studied using BLI. BLI was the preferred biophysical technique to assay this interaction because of its low sample requirements, compared to ITC, and the absence of clogging effects, which is a risk associated with the use of SPR in the study of particulate analytes (223). In this setup, HSA-Le^{b/y} glycoconjugates were immobilised to a BLI biosensor tip via amine coupling (Figure 5.13).

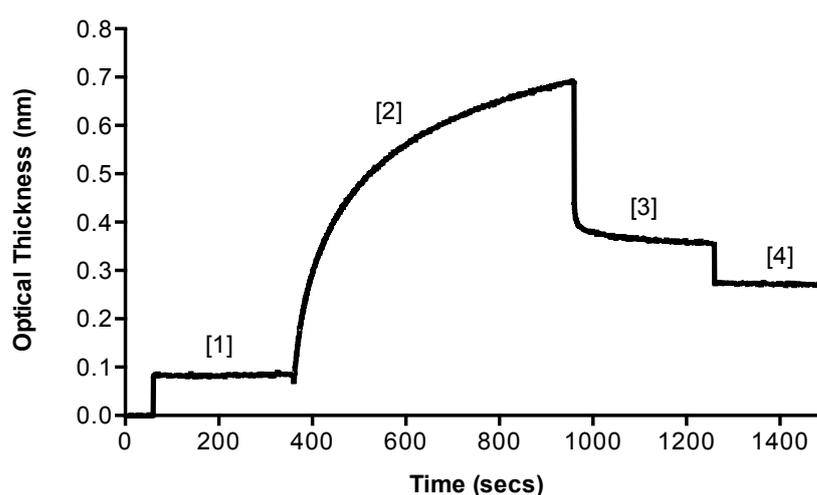


Figure 5.13: Immobilisation of HSA-Le^b to an amine reactive biosensor tip via amine coupling

[1] Activation of carboxymethyl groups on the amine reactive biosensor tip to NHS esters via tip submersion in NHS/EDC, [2] Loading of HSA-Le^b, [3] Deactivation of unreacted NHS esters via tip submersion in Tris-Cl, [4] Washing and baseline equilibration. HSA-Le^y was similarly immobilised via amine coupling.

The successful immobilisation of HSA-Le^b and HSA-Le^y glycoconjugates to the biosensor tips was further verified through a preliminary protein binding experiment, at pH 7.4, using BabA_{527K} and its D233A/S244A-variant form. As shown in Figure 5.14, BabA_{527K} elicits a ~3-fold higher increase in biosensor tip optical thickness than BabA_{527K}-D233A/S244A when probed with HSA-Le^b. Both proteins produce comparable response levels when probed with HSA-Le^y – response levels that are similar to the observed interaction between

BabA_{527K}-D233A/S244A and HSA-Le^b. This binding pattern indicates that the immobilised glycoconjugates act as functional receptors. However, a level of non-specific association is observed given the well-established lack of affinity of BabA_{527K} for Le^y and the non-binding characteristics of the D233A/S244A-variant.

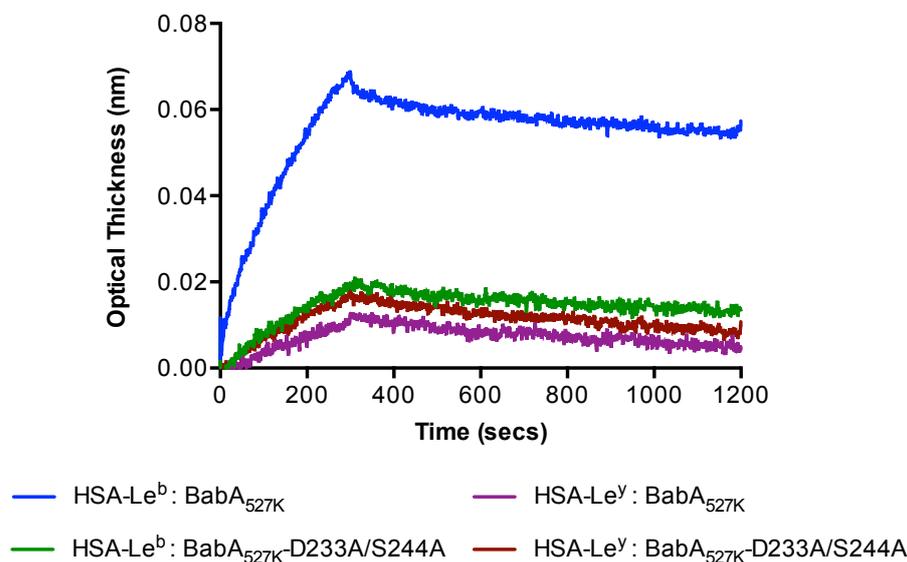


Figure 5.14: Binding of BabA_{527K} and its D233A/S244A-variant form to HSA-Le^{b/y} in a BLI assay

Binding was studied in a 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl buffer. Each well (200 μ L) contained 2.5 μ g of the stated recombinant BabA protein. Association and dissociation times were 300 and 900 seconds, respectively.

Given the validity of this approach to detect binding, BLI was used to investigate microparticle:HSA-Le^{b/y} binding, at pH 7.4. As shown in Figure 5.15, binding of BabA_{WT}-microparticles to HSA-Le^b on the biosensor tip is clearly seen while no interaction with HSA-Le^y is observed; BabA_{D233A/S244A}-microparticles do not interact with either HSA-Le^b or HSA-Le^y. Additionally, neither polystyrene- nor NeutrAvidin-microparticles interacted with HSA-Le^b or HSA-Le^y. Thus, the observed binding between BabA_{WT}-microparticles and HSA-Le^b is specifically mediated by the glycan binding activity of recombinant BabA. Importantly, the vast difference seen in the increase of the optical

thickness of HSA-Le^b biosensor tips during BabA_{527K} binding (max: ~0.07) compared to BabA_{WT}-microparticle binding (max: ~1.75) indicates that the response seen in the latter interaction is due to particle association with the biosensor tips, rather than binding of unconjugated protein in solution. The same total amount of recombinant BabA that was used for protein binding experiments was immobilised onto microparticles for protein-particle binding experiments.

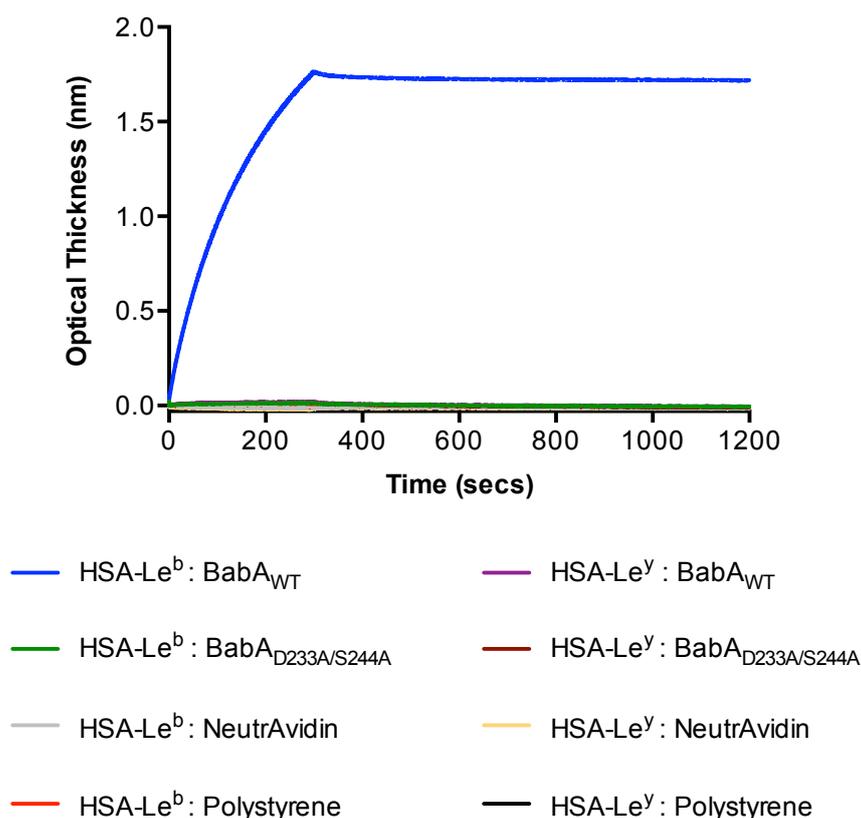


Figure 5.15: Binding of various microparticles to HSA-Le^{b/y} in a BLI assay

Binding was studied in a 20 mM Tris-Cl (pH = 7.4) and 300 mM NaCl buffer. Each well (200 μ L) contained up to 2.5 μ g of the stated recombinant BabA protein conjugated to 50 μ g microparticle solids. Association and dissociation times were 300 and 900 seconds, respectively.

The generation of functional BabA-microparticles completes the experimental objectives of this study and strongly supports the use of BabA as a targeting moiety.

5.4 Discussion

After developing an efficient, reproducible method to express recombinant BabA, and subsequently performing a thorough characterisation of its glycan-binding site, the final objective of this study focused on the prompt generation of essential data to provide relevant insights into the proposed use of BabA as a gastric-targeting moiety. First, the effect of low gastric fluid pH on protein stability was considered. Though it is widely accepted that the stomach lumen is frequently characterised by strongly acidic conditions ($\text{pH} \leq 3$), as recently collated by Mudie *et al.*, various researchers have recorded a wide range of pH values in both fasted (between 1 and 8) and fed (between 3 and 6) states (247). Consequently, in this study, a pH value of 4.5 was adopted to mimic the acidic conditions that may exist in both pre- and post-prandial states. At this pH, circular dichroism spectroscopy revealed that the secondary structure folds of recombinant BabA are not altered in comparison to that observed at neutral pH, which represents the conditions of the inner surface mucus layer where *H. pylori* prefer to thrive (79). This demonstrates the acid stability of BabA, however it would be of interest to further study conformation at more acidic pH levels.

In addition to its secondary structure being unaffected by the reduction of pH, no difference was observed in BabA_{527K} binding to Le^b antigens at acidic and neutral pH. This can be considered to be a rational finding given that the binding interaction does not involve charge-dependent electrostatic interactions, as shown through the BabA_{527K}:Le^b structural model. Though this further supports the acid-resistance of recombinant BabA, pH-independent binding affinity may actually compromise its intended use. As the strength of the interaction between recombinant BabA and Le^b, and perhaps other ABO/Le blood group antigens, is not affected by pH, secreted MUC5AC in the

gastric mucus layer, which carries these glycans, may compete with the surface epithelia for BabA-vectors (103, 229, 230, 248). As a consequence, similar to what is observed in *H. pylori*, it is expected that BabA *ex situ* can mediate both mucoadhesive and cytoadhesive attachment.

Interestingly, although pH reduction did not affect secondary conformation or Le^b binding activity, differential scanning fluorimetry showed that recombinant BabA had enhanced thermal stability at pH 4.5. This may be due to favourable changes to the protonation state of amino acids that enable the formation of structure-stabilising electrostatic interactions (249). Perhaps this may be an important feature of BabA to withstand the acidic, denaturing conditions of gastric fluid. However, electrostatic interactions between amino acid side chains can also be affected by the ionic strength of their environment – salt ions can interact with charged amino acids and consequently modulate inter-residue interactions (250). Because the acidic and neutral buffers used in this study inadvertently had different ionic strengths, the enhanced thermal stability of recombinant BabA solely due to a reduction in pH cannot be accurately suggested. Ultimately, due to the possible effects of microenvironment constituents on thermal stability, differential scanning fluorimetry will provide the most value when employed to study protein unfolding in physiological conditions that closely mimic the gastric milieu. Through this, calculated BabA unfolding transition temperatures will also be of biological significance.

Nonetheless, the differential scanning fluorimetry assay showed that in the pH 4.5 buffer, at 37 °C, the hydrophobic patches of BabA_{527K} are not available for SYPRO Orange dye binding. As this indicates that the three-dimensional fold of BabA_{527K} is intact and not disrupted at these conditions, this buffer system was used to assess BabA_{527K} peptic degradation in its native conformation.

BabA_{527K} was rapidly and completely digested by pepsin although intermediate fragments that were truncated from their C-termini were observed. In fact, in this respect, peptic digestion shares similarities with the previously observed degradation of BabA₅₂₇ during periplasmic expression. This suggests that BabA contains a highly susceptible C-terminal protease cleavage site; it would be of interest to identify this region and study the potential effect of sequence variations on both peptic and periplasmic proteolytic susceptibility. This can be achieved through combining a thorough mass spectrometric analysis of recombinant BabA degradation with exploration of potentially labile structural features revealed through the crystallographic model.

Unlike the experimental control β -lactoglobulin, which has been well characterised as a pepsin-resistant protein, no similar tests have been reported in the literature for *H. pylori* adhesins, although its secreted toxin VacA has been shown to be resistant to peptic digestion (243, 251). At the whole bacterium level, pepsin not only kills *H. pylori* but actually impairs its motility in the stomach lumen, thereby preventing its colonisation of the gastric mucus layer (252-254). Thus, despite being a specialised gastric pathogen, *H. pylori* is considered to be sensitive to the detrimental effects of pepsin. Consequently, it is perhaps not surprising that BabA is not intrinsically resistant to peptic digestion, particularly since pepsin cannot penetrate into the gastric mucus layer where BabA-mediated attachment occurs (248). With regards to its use as a targeting moiety, peptic digestion may be circumventable through the administration of BabA-vectors during fasted conditions as pepsin (pepsinogen) is only secreted into the stomach lumen in response to meals (255).

The physiological characteristics of the stomach clearly present considerable barriers for BabA-mediated drug delivery vectors. Nonetheless, representing a significant step towards achieving such systems, this study shows for the first time that the selective ABO/Le blood group antigen-binding properties of *H. pylori* can be mimicked by BabA-particle conjugates. The functional glycan binding activity of BabA_{WT}-microparticles clearly indicates that recombinant BabA is correctly orientated on the microparticle surfaces. This was achieved using insight from the crystallographic models as the C-terminal linkage of recombinant BabA to NeutrAvidin-microparticles was based on the relative positioning of the Handle region to the glycan-binding Crown unit. Indeed, this surface presentation mimics that of native BabA, which is inserted into the *H. pylori* outer membrane through its C-terminal transmembrane domain (86). Importantly, the variant form of BabA was used, as one of the experimental controls, to verify that the observed glycan binding of these model biomimetic vectors was specifically imparted by BabA. However, the complete lack of affinity of the BabA_{D233A/S244A}-microparticles for HSA-Le^b also further strengthens the validity of the BabA_{527K}:Le^b structural model and indicates the relevance of the observed hydrogen bond network in mediating *H. pylori* attachment. This is because the conditions under which binding was studied – multivalent BabA-particles binding to multivalent Le^b complexes, that is, avidity – are representative of the attachment of a single *H. pylori* cell to surface- or mucin-associated glycans during colonisation. It would be of interest to characterise the apparent affinity of BabA-vectors to HSA-Le^b, using a technique such as BLI, not only to see how this compares with reported *H. pylori* affinity, but also to understand how factors such as protein surface coating density and immobilisation strategy can affect glycan binding (161).

Though potential physiological barriers to using BabA as a targeting moiety have been identified, the successful generation of proof-of-principle BabA-vectors represents the first milestone towards the development of a gastric-targeted drug delivery system. Importantly, these results form the foundation for future studies investigating the *in vivo* applicability of this system.

Chapter 6: General discussion and conclusions

6.1 Research significance and future work

The motivation of this research was to address shortcomings in the development of bioadhesive drug delivery systems capable of mediating the gastric retention of drugs. While this field has yet to witness an effective approach, learnings from unsuccessful strategies have resulted in a sound understanding of the characteristics that these systems should possess in order to be effective. In brief, an ideal bioadhesive drug delivery vector would mediate gastric retention by evading gastric mucus turnover through attaching directly to the gastric epithelium. While such attachment may be achievable in drug delivery systems by using lectins as targeting moieties, no lectins that target the gastric epithelium have been suggested for this use in the literature. This study proposes a biomimetic approach based on BabA, a lectin that mediates the attachment of *H. pylori* to glycans naturally expressed on the gastric epithelium.

The aim of this work was to characterise the functional properties of BabA, in the context of its use as a gastric targeting moiety. In this regard, this study has shown that the glycan binding properties of BabA are fully maintained in a recombinant form of the protein; importantly, recombinant BabA can be successfully used to create a model, glycan-binding drug delivery vector. While it has not yet been assessed how this model system will fare in conditions mimicking the stomach, this study serves as the first demonstration of the potential use of BabA to direct drug delivery vectors to gastric glycans. It must be emphasised that a potential disadvantage of using lectins to target drug delivery vectors to glycans on epithelial cell surfaces *in vivo* is the possible mediation of premature, undesired interactions with mucins in the

overlying mucus layer. In a BabA mediated approach, this could similarly prove to be a hindrance given the presence of its glycan receptors on epithelial cells as well as attached to the secreted gastric mucin MUC5AC. Thus, it is imperative that this system is studied in a physiologically relevant microenvironment in order to make a true assessment of the viability of this biomimetic approach. Given that this is the first study concerning the development of BabA-vectors, such an assessment was beyond the scope of this research.

The realisation that the properties of BabA as an *H. pylori* adhesin can be successfully translated to a model targeted drug delivery system was achieved through experimentation aligned to specific research objectives. Importantly, through the design of this study, this resulted in the advancement of not only the use of BabA for drug delivery applications, but also in a mechanistic understanding of its role in mediating *H. pylori* gastric colonisation. The novel insights gained from the results of the experimental work presented in this thesis, in the context of their scientific relevance and opportunities for future work, are summarised and discussed below.

6.1.1 Discovery of an improved periplasmic expression strategy

While the most successful method of recombinant expression reported in this work was based on a published *E. coli* periplasmic expression strategy, it was the serendipitous incorporation of a C-terminal hexa-lysine tag into recombinant BabA that facilitated the progression of this research. Due to the significant advantages in expression yield and purity that this strategy conferred, it can be considered ideal to support future studies evaluating the use of BabA as a gastric-targeting moiety. Albeit a low yield in comparison to

the potential of other prokaryotic and eukaryotic expression strategies, periplasmic expression enabled rapid experimentation with a functional recombinant BabA protein. This is a significant advantage in comparison to other reported BabA expression methods that, depending on application, require additional, time-consuming interventions following expression, such as solubility-enhancing partner cleavage (164) and refolding from inclusion bodies (165). Thus, the efficient and reproducible periplasmic expression strategy described in this thesis will be of value to researchers requiring the swift generation of recombinant BabA for differing study needs. This strategy can also be employed to express recombinant BabA from strains other than J99 – the extracellular domain of BabA, with a C-terminal hexa-lysine tag, from strain CCUG 17875 was also successfully expressed and purified in our lab with a similar yield (results not shown in this thesis).

The incorporation of a C-terminal hexa-lysine tag was observed to improve recombinant BabA purification by enhancing solubility and conferring protection against proteolysis, which most likely occurred during periplasmic expression. Though not fully understood, given that this method significantly aided the generation of a functional recombinant protein, it will be of great interest to determine if this method also facilitates the recombinant expression of other paralogous proteins from the Hop family. For example, SabA and LabA, which are known to function as glycan-binding adhesins, are also suitable candidates for periplasmic expression; SabA contains three intramolecular disulphide bonds (175) while LabA contains six cysteine residues, which may also form intramolecular disulphide bonds. Thus, should the incorporation of a hexa-lysine tag also improve the periplasmic expression of these proteins, this strategy could prove to be a highly useful tool for bacteriologists looking to probe the biological role of members of the Hop

family at the protein level. In support of the applicability of this expression method, preliminary experiments by researchers in our lab investigating the unknown function of BabB found that this protein could not be at all expressed in the periplasm of *E. coli* without a C-terminal hexa-lysine tag (results not shown in this thesis).

From a drug delivery perspective, it is also of great interest to express and study other *H. pylori* adhesins as potential targeting ligands because they may confer particle-conjugate affinity towards other important gastric epithelial moieties. SabA, which recognises sialylated glycans in the gastric mucosa, deserves strong consideration. Sialylated glycans are not typically found in abundance in healthy stomachs but are significantly expressed under conditions of inflammation, such as the histologic gastritis caused by *H. pylori* infection in around half of the global population. Accordingly, a drug delivery system relying on both BabA and SabA may be more effective than BabA alone in imparting the gastric retention of drug delivery vectors to both healthy and inflamed stomachs.

6.1.2 Revelation of the BabA glycan-binding mechanism

Identification and characterisation of the BabA glycan-binding site was considered essential to ensure that the generation of functional BabA-vectors was not hindered by inappropriate protein immobilisation to particle surfaces. Of paramount importance, this led to the revelation of the mechanism through which BabA mediates *H. pylori* attachment to type 1 ABO/Le blood group antigens – this represents the first molecular insight into an *H. pylori* gastric colonisation mechanism. Understanding the structural basis for type 1 ABO/Le blood group antigen binding by BabA finally rationalises why other *H. pylori*

adhesins, which all lack the Crown β -unit sequence, do not bind to these glycans. Furthermore, knowledge of the location of the BabA glycan-binding site in *H. pylori* J99 may enable researchers to understand the amino acid differences that dictate why other *babA* expressing strains are incapable of attaching to type 1 ABO/Le blood group antigens. Such insight allows the role of BabA in *H. pylori* pathogenesis to be studied with a better fundamental understanding.

The structural basis of glycan recognition by BabA has long been sought to test the hypothesis that preventing the interaction between adhesins and host receptors can represent a new form of antimicrobial treatment by inhibiting bacterial adherence and subsequent colonisation. The potential of this approach has already been demonstrated in mice against uropathogenic *E. coli* where inhibition of its FimH adhesin, through the oral administration of novel anti-adhesion molecules, can successfully treat and prevent infection (256, 257). Consequently, the comprehension of the molecular interactions that constitute BabA glycan-binding is considered a promising lead for the development of new strategies for the treatment of *H. pylori* infections, which are globally gaining resistance to traditional antibiotics (258).

With regards to drug delivery applications, the structural identification of the BabA binding site also introduces the possibility that the proposed biomimetic vectors can be created using solely the Crown as a targeting moiety. While the activity of this fragment will need to be evaluated, the use of a smaller glycan-binding protein could be beneficial as more molecules can be packed onto particle surfaces; this could prove to be key as smaller particles ($< 0.2 \mu\text{m}$), with consequently smaller surface areas, achieve better penetration through mucin mesh fibres compared to larger particles ($> 0.5 \mu\text{m}$) (49).

6.1.3 Identification of the Hypervariable Crown Loop

The observation that a segment of the Crown region is involved in mediating BabA glycan binding, yet is poorly conserved between *H. pylori* strains, may explain the previously reported inter-strain variability in BabA glycan-binding affinity. As developing an understanding of BabA-mediated *H. pylori* glycan binding at the whole bacterium level was beyond the scope of this research, the proposed relevance of the Hypervariable Crown Loop must be validated through further investigation.

Future work could utilise the optimised recombinant expression techniques and biophysical binding assays described in this thesis to study the effect of strain variation on protein:glycan binding alongside bacterium:glycan binding investigations. Certainly, binding studies would benefit from crystallographic models because the effects of variability in regions other than the Crown cannot be excluded without structural insight. However, as this can be both challenging and resource-intensive, an innovative approach could be to develop a chimeric expression construct whereby recombinant BabA proteins have identical Handle and Head regions, corresponding to BabA J99 for example, but different Crown or Hypervariable Crown Loop sequences corresponding to the strain of interest. Perhaps through this strategy, the proposed effect of the Hypervariable Crown Loop on glycan-binding affinity can be more closely evaluated. In the absence of crystallographic information, such an approach could also aid in the evaluation of whether the Hypervariable Crown Loop has a role to play in the creation of “specialist” binding specificity.

Ultimately, the potential relevance of the Hypervariable Crown Loop is of greater interest to the study of *H. pylori* adherence mechanisms rather than

the development of drug delivery vectors. Nevertheless, a better understanding of the glycan binding properties in other *H. pylori* strains could reveal properties relevant to the use of BabA as a targeting moiety, such as the identification of enhanced adhesive characteristics for example.

6.1.4 Observation of BabA susceptibility to peptic digestion

Given that *H. pylori* is a pathogen believed to have evolved over millennia to adapt to the harsh conditions of the human stomach, it was somewhat unexpected to discover that one of its main colonisation mechanisms is susceptible to the proteolytic effects of pepsin. With regards to the proposed use of BabA, this is a significant drawback that may render BabA-vectors completely unsuitable for gastric targeting. Alternatively, their usage may be restricted to the fasted state, during which high levels of pepsin are not typically found in gastric fluid. Functional model vectors will have to be assessed in gastric fluid representative of both fasted and fed states in order to conclude on the use of BabA as a targeting moiety in a biorelevant context. Still, it is worth considering that the BabA Crown, which might retain glycan-binding function without the BabA Handle and Head, may not be readily digested by pepsin in its folded conformation; this must be determined empirically to assess whether Crown-vectors represent a better biomimetic approach than BabA-vectors.

While both recombinant and native forms of BabA were completely degraded by pepsin in this study, these observations do not directly translate to the susceptibility of this adhesin in the stomach lumen while attached to the *H. pylori* outer membrane. The activity of urease in *H. pylori* creates a buffered surface layer for the bacterium in response to acidic conditions (259) – this

may modulate the proteolytic activity of pepsin. *H. pylori* also possesses several outer membrane proteins [approximately 4% of its coding genome (87)] and a lipopolysaccharide layer – these could very well affect the accessibility of pepsin to BabA, as may potential yet-to-be characterised post-translational modifications, such as glycosylation. Thus, a study that compares the *in situ* and *ex situ* peptic susceptibility of BabA would greatly further the understanding of the findings observed in this work.

6.1.5 Development of functional glycan-binding BabA-vectors

As has been shown through this study, BabA can be used create functional glycan-binding particles and this greatly supports its use a gastric-targeting moiety. Performing an *in vitro* or *in vivo* study with these model particles is undoubtedly the most efficient way to assess the viability of this targeted drug delivery system. However, though desired, such as an assessment was beyond the scope of this study, mostly due to the focus of experimental work on characterising the BabA glycan-binding site to enable the rational design of particle-conjugates for this study, and for future research. Still, preliminary attempts were made to develop an *in vitro* gastric model using the MKN7 cell line. Unfortunately, the culture conditions required to support cell polarisation and the expression of ABO/Le blood group antigens have still not been identified by our lab. This remains under active investigation because a relevant model must be utilised to confirm that the findings observed in this study translate to an effective interaction between BabA-vectors and gastric epithelial cells.

The progression of this work to a clinically applicable system necessitates several further phases of evaluation and development. This will include

trailing biologically relevant, biodegradable drug delivery vehicles. As such, processes such as nano-/micro-particle formation, drug encapsulation/release and surface modification with BabA will need to be developed. The properties of the drug delivery system are important for the definition of the conjugation method of BabA to particle surfaces and the rate of controlled release for the drug. However, as a result of exciting, recent advancements, the properties of the delivery system can also impact targeting efficacy. Polymeric nanoparticles can be engineered such that targeting moieties are concealed in the particle core until conformational changes in the polymeric matrix occur in response to the pH of the microenvironment and leads to their exposure on particle surfaces (260). Given the pH gradient across the gastric mucus layer, such technology may be useful for asserting the cytoadhesive, rather than mucoadhesive, properties of BabA-particles. Indeed, this strategy may also be used to protect BabA from peptic degradation during exposure to gastric juice.

As already stressed, future work should also focus on studying the interactions between BabA-vectors and MUC5AC as part of an investigation into the properties that these biomimetic vectors must possess in order to penetrate through the frequently replaced gastric mucus layer. Further regarding undesirable attachment to mucins, the glycosylation pattern and the role of the gastric cell surface associated mucin MUC1 in acting as a releasable decoy receptor for *H. pylori* must be studied in greater detail to understand if this proposed mechanism, described by Linden *et al.* (107), could interfere with BabA-particle cytoadhesion.

While *H. pylori* exploits the abundance of type 1 ABO/Le blood group antigens in the gastric mucosa through BabA, it must be noted that these glycans can also be found in a variety of exocrine secretions, and accordingly, are also

found in different parts of the GI tract in varying amounts (95). For example, the salivary mucin MUC5B carries blood group antigens to which *H. pylori* have been shown to bind via BabA *in vitro* (113). However, gastric-specificity is not a prerequisite for BabA-particles *per se* because once administered, the gastric mucosa will be the first mucosal membrane to which BabA-particles are exposed after oral administration if delivered via an encapsulating dosage form. As such, a successful and rapid association of these particles to the gastric epithelium is expected to elicit effective gastric-specific retention. Nonetheless, the prevalence and relevance of blood group antigen expression throughout the GI tract should also be studied during the future development of BabA-vectors to understand how off-site binding may impact clinical applicability.

As has been emphasised in this thesis, no lectins have been studied in the literature for a role in gastric-targeted drug delivery. However, BabA and other *H. pylori* adhesins do not represent the only candidates. For example, *Ulex europaeus* lectin I and *Griffonia simplicifolia* isolectin 1B4 also attach to blood group antigens, albeit with less specificity by binding to terminal α -fucose and α -galactose residues, respectively, rather than to multiple sugar units (261); these plant lectins have previously been studied for use in intestinal-targeted drug delivery systems. To increase the chances of identifying a suitable gastric-targeting ligand, the efficacy of these plant lectins in binding to the gastric epithelium could also be assessed alongside BabA in well-defined *in vitro* and *in vivo* evaluation models. Indeed, the use of these lectins may confer specific advantages over BabA with regards to peptic susceptibility; plant lectins are typically resistant to gastrointestinal proteolytic enzymes (262).

6.2 Summary

Through the experimental work presented in this thesis, this study has demonstrated that a recombinant form of the *H. pylori* adhesin BabA can be used to create model biomimetic particles that have the same blood group antigen-binding properties found with the whole bacterium (Figure 6.1). As a result, this work represents the first step towards the potential use of BabA-vectors to achieve the gastric retention of drugs. Overall, though physiological barriers exist that may limit the use of BabA as a targeting moiety, the successful generation of proof-of-principle BabA-vectors represents the first milestone towards the potential use of *H. pylori* biomimics for the gastric targeting of drugs and lays the foundation for future translational research.

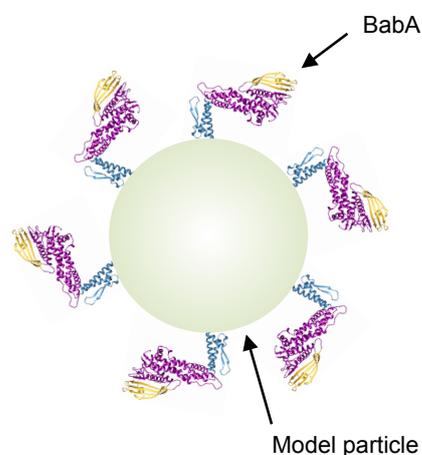


Figure 6.1: Schematic representation of the glycan-binding BabA-particles created in this study

Functional BabA-particle conjugates were created through three sequential steps in this study: 1) the determination of an efficient method to recombinantly express the extracellular domain of BabA as a soluble protein, 2) the identification and characterisation of its functionally active glycan-binding site, and finally 3) the development of particle-conjugates through a BabA site-specific linkage that ensured retention of its glycan-binding properties.

References

1. Y. Masaoka, Y. Tanaka, M. Kataoka, S. Sakuma, S. Yamashita, Site of drug absorption after oral administration: assessment of membrane permeability and luminal concentration of drugs in each segment of gastrointestinal tract, *Eur J Pharm Sci* **29**, 240–250 (2006).
2. M. N. Martinez, G. L. Amidon, A mechanistic approach to understanding the factors affecting drug absorption: a review of fundamentals, *J Clin Pharmacol* **42**, 620–643 (2002).
3. S. S. Davis, J. G. Hardy, J. W. Fara, Transit of pharmaceutical dosage forms through the small intestine, *Gut* **27**, 886–892 (1986).
4. S. S. Davis, Formulation strategies for absorption windows, *Drug Discov Today* **10**, 249–257 (2005).
5. A. Streubel, J. Siepmann, R. Bodmeier, Drug delivery to the upper small intestine window using gastroretentive technologies, *Curr Opin Pharmacol* **6**, 501–508 (2006).
6. C. G. Wilson, P. J. Crowley, *Controlled Release in Oral Drug Delivery* (Springer Science & Business Media, New York, 2011).
7. Gastrointestinal Tract and Liver - Image ID: 12001075 (available at <http://www.123rf.com>; Copyright: Peter Lamb; Accessed on 23/09/2015).
8. E. A. Klausner, E. Lavy, D. Stepensky, E. Cserepes, M. Barta, M. Friedman, A. Hoffman, Furosemide pharmacokinetics and pharmacodynamics following gastroretentive dosage form administration to healthy volunteers, *J Clin Pharmacol* **43**, 711–720 (2003).
9. S. M. R. Camargo, R. N. Vuille-dit-Bille, L. Mariotta, T. Ramadan, K. Huggel, D. Singer, O. Gotze, F. Verrey, The Molecular Mechanism of Intestinal Levodopa Absorption and Its Possible Implications for the Treatment of Parkinson's Disease, *J Pharmacol Exp Ther* **351**, 114–123 (2014).
10. R. Teng, J. Maya, Absolute bioavailability and regional absorption of ticagrelor in healthy volunteers, *J Drug Assess* **3**, 43–50 (2014).
11. A. Hoffman, Pharmacodynamic aspects of sustained release preparations, *Adv Drug Deliv Rev* **33**, 185–199 (1998).
12. A. W. Basit, L. F. Lacey, Colonic metabolism of ranitidine: implications for its delivery and absorption, *Int J Pharm* **227**, 157–165 (2001).
13. Y. Joshi, D. Mastropietro, H. Omidian, Enhanced Bioavailability via Extended Gastric Retention, *J Develop Drugs* **2**, 105 (2013).
14. D. Cook, D. Brown, R. Alexander, R. March, P. Morgan, G. Satterthwaite, M. N. Pangalos, Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework, *Nat Rev Drug Discov* **13**, 419–431 (2014).
15. F. E. Stuurman, B. Nuijen, J. H. Beijnen, J. H. M. Schellens, Oral Anticancer Drugs: Mechanisms of Low Bioavailability and Strategies for

- Improvement, *Clin Pharmacokinet* **52**, 399–414 (2013).
16. K. C. Waterman, A Critical Review of Gastric Retentive Controlled Drug Delivery, *Pharm Dev Technol* **12**, 1–10 (2007).
 17. B. N. Singh, K. H. Kim, Floating drug delivery systems: an approach to oral controlled drug delivery via gastric retention, *J Control Release* **63**, 235–259 (2000).
 18. P. Fasinu, V. Pillay, V. M. K. Ndesendo, L. C. du Toit, Y. E. Choonara, Diverse approaches for the enhancement of oral drug bioavailability, *Biopharm Drug Dispos* **32**, 185–209 (2011).
 19. W. Erni, K. Held, The hydrodynamically balanced system: a novel principle of controlled drug release, *Eur Neurol* **27**, 21–27 (1987).
 20. S. L. Schwartz, J. F. Wu, B. Berner, Metformin extended release for the treatment of type 2 diabetes mellitus, *Expert Opin Pharmacother* **7**, 803–809 (2006).
 21. J. Woodley, Bioadhesion, *Clin Pharmacokinet* **40**, 77–84 (2001).
 22. G. Ponchel, J. Irache, Specific and non-specific bioadhesive particulate systems for oral delivery to the gastrointestinal tract, *Adv Drug Deliv Rev* **34**, 191–219 (1998).
 23. A. P. Corfield, Mucins: A biologically relevant glycan barrier in mucosal protection, *BBA-Gen Subjects* **1850**, 236–252 (2015).
 24. J. Perez-Vilar, R. L. Hill, The structure and assembly of secreted mucins, *J Biol Chem* **274**, 31751–31754 (1999).
 25. N. Juge, Microbial adhesins to gastrointestinal mucus, *Trends Microbiol* **20**, 30–39 (2012).
 26. A. P. Moran, A. Gupta, L. Joshi, Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract, *Gut* **60**, 1412–1425 (2011).
 27. F. C. Carvalho, M. L. Bruschi, R. C. Evangelista, Mucoadhesive drug delivery systems, *Braz J Pharm Sci* **46**, 1–17 (2010).
 28. J. D. Smart, The basics and underlying mechanisms of mucoadhesion, *Adv Drug Deliv Rev* **57**, 1556–1568 (2005).
 29. J. K. Vasir, K. Tambwekar, S. Garg, Bioadhesive microspheres as a controlled drug delivery system, *Int J Pharm* **255**, 13–32 (2003).
 30. C. M. Lehr, F. G. J. Poelma, H. E. Junginger, J. J. Tukker, An estimate of turnover time of intestinal mucus gel layer in the rat in situ loop, *Int J Pharm* **70**, 235–240 (1991).
 31. G. P. Andrews, T. P. Lavery, D. S. Jones, Mucoadhesive polymeric platforms for controlled drug delivery, *Eur J Pharm Biopharm* **71**, 505–518 (2009).
 32. M. Säkkinen, T. Tuononen, H. Jürjenson, P. Veski, M. Marvola, Evaluation

of microcrystalline chitosans for gastro-retentive drug delivery, *Eur J Pharm Sci* **19**, 345–353 (2003).

33. M. Säkkinen, J. Marvola, H. Kanerva, K. Lindevall, M. Lipponen, T. Kekki, A. Ahonen, M. Marvola, Gamma scintigraphic evaluation of the fate of microcrystalline chitosan granules in human stomach, *Eur J Pharm Biopharm* **57**, 133–143 (2004).

34. R. Khosla, S. S. Davis, The effect of polycarbophil on the gastric emptying of pellets, *J Pharm Pharmacol* **39**, 47–49 (1987).

35. D. Harris, J. T. Fell, H. L. Sharma, D. C. Taylor, GI transit of potential bioadhesive formulations in man: A scintigraphic study, *J Control Release* **12**, 45–53 (1990).

36. S. J. Jackson, D. Bush, A. C. Perkins, Comparative scintigraphic assessment of the intragastric distribution and residence of cholestyramine, Carbopol 934P and sucralfate, *Int J Pharm* **212**, 55–62 (2001).

37. P. J. Sansonetti, War and peace at mucosal surfaces, *Nat Rev Immunol* **4**, 953–964 (2004).

38. C. Atuma, V. Strugala, A. Allen, L. Holm, The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo, *Am J Physiol Gastrointest Liver Physiol* **280**, G922–G929 (2001).

39. R. A. Cone, Barrier properties of mucus, *Adv Drug Deliv Rev* **61**, 75–85 (2009).

40. A. Ermund, A. Schutte, M. E. V. Johansson, J. K. Gustafsson, G. C. Hansson, Studies of mucus in mouse stomach, small intestine, and colon. I. Gastrointestinal mucus layers have different properties depending on location as well as over the Peyer's patches, *Am J Physiol Gastrointest Liver Physiol* **305**, G341–G347 (2013).

41. S. K. Lai, Y.-Y. Wang, J. Hanes, Mucus-penetrating nanoparticles for drug and gene delivery to mucosal tissues, *Adv Drug Deliv Rev* **61**, 158–171 (2009).

42. M. Phillipson, M. E. V. Johansson, J. Henriksnas, J. Petersson, S. J. Gendler, S. Sandler, A. E. G. Persson, G. C. Hansson, L. Holm, The gastric mucus layers: constituents and regulation of accumulation, *Am J Physiol Gastrointest Liver Physiol* **295**, G806–G812 (2008).

43. S. Etzold, N. Juge, Structural insights into bacterial recognition of intestinal mucins, *Curr Opin Struct Biol* **28**, 23–31 (2014).

44. S. S. Olmsted, J. L. Padgett, A. I. Yudin, K. J. Whaley, T. R. Moench, R. A. Cone, Diffusion of macromolecules and virus-like particles in human cervical mucus, *Biophys J* **81**, 1930–1937 (2001).

45. S. K. Lai, D. E. O'Hanlon, S. Harrold, S. T. Man, Y.-Y. Wang, R. Cone, J. Hanes, Rapid transport of large polymeric nanoparticles in fresh undiluted human mucus, *Proc Natl Acad Sci USA* **104**, 1482–1487 (2007).

46. B. C. Tang, M. Dawson, S. K. Lai, Y.-Y. Wang, J. S. Suk, M. Yang, P. Zeitlin, M. P. Boyle, J. Fu, J. Hanes, Biodegradable polymer nanoparticles that

rapidly penetrate the human mucus barrier, *Proc Natl Acad Sci USA* **106**, 19268–19273 (2009).

47. K. Maisel, L. Ensign, M. Reddy, R. Cone, J. Hanes, Effect of surface chemistry on nanoparticle interaction with gastrointestinal mucus and distribution in the gastrointestinal tract following oral and rectal administration in the mouse, *J Control Release* **197**, 48–57 (2015).

48. L. Inchaurreaga, N. Martín-Arbella, V. Zabaleta, G. Quincoces, I. Peñuelas, J. M. Irache, In vivo study of the mucus-permeating properties of PEG-coated nanoparticles following oral administration, *Eur J Pharm Biopharm*, doi:10.1016/j.ejpb.2014.12.021.

49. L. M. Ensign, R. Cone, J. Hanes, Oral drug delivery with polymeric nanoparticles: The gastrointestinal mucus barriers, *Adv Drug Deliv Rev* **64**, 557–570 (2012).

50. M. Kim, H. Ashida, M. Ogawa, Y. Yoshikawa, H. Mimuro, C. Sasakawa, Bacterial Interactions with the Host Epithelium, *Cell Host Microbe* **8**, 20–35 (2010).

51. P. H. Jensen, D. Kolarich, N. H. Packer, Mucin-type O-glycosylation--putting the pieces together, *FEBS J.* **277**, 81–94 (2010).

52. C. M. Lehr, Lectin-mediated drug delivery: the second generation of bioadhesives, *J Control Release* **65**, 19–29 (2000).

53. B. Naisbett, J. Woodley, The potential use of tomato lectin for oral drug delivery. 1. Lectin binding to rat small intestine in vitro, *Int J Pharm* **107**, 223–230 (1994).

54. B. Naisbett, J. Woodley, The potential use of tomato lectin for oral drug delivery: 2. Mechanism of uptake in vitro, *Int J Pharm* (1994).

55. B. Naisbett, J. Woodley, The potential use of tomato lectin for oral drug delivery: 3. Bioadhesion in vivo, *Int J Pharm* **114**, 227–236 (1995).

56. B. Naisbett, J. Woodley, The potential use of tomato lectin for oral drug delivery: 4. Immunological consequences, *Int J Pharm* **120**, 247–254 (1995).

57. B. Carreno-Gomez, J. F. Woodley, A. T. Florence, Studies on the uptake of tomato lectin nanoparticles in everted gut sacs, *Int J Pharm* **183**, 7–11 (1999).

58. H. Chen, V. Torchilin, R. Langer, Lectin-bearing polymerized liposomes as potential oral vaccine carriers, *Pharm Res* **13**, 1378–1383 (1996).

59. P. J. Giannasca, K. T. Giannasca, P. Falk, J. I. Gordon, M. R. Neutra, Regional differences in glycoconjugates of intestinal M cells in mice: potential targets for mucosal vaccines, *Am J Physiol Gastrointest Liver Physiol* **267**, G1108–G1121 (1994).

60. N. Foster, M. A. Clark, M. A. Jepson, B. H. Hirst, Ulex europaeus 1 lectin targets microspheres to mouse Peyer's patch M-cells in vivo, *Vaccine* **16**, 536–541 (1998).

61. G. J. Russell-Jones, H. Veitch, L. Arthur, Lectin-mediated transport of nanoparticles across Caco-2 and OK cells, *Int J Pharm* **190**, 165–174 (1999).
62. C. Bies, C. M. Lehr, J. F. Woodley, Lectin-mediated drug targeting: history and applications, *Adv Drug Deliv Rev* **56**, 425–435 (2004).
63. Y. Liu, P. Wang, C. Sun, N. Feng, W. Zhou, Y. Yang, R. Tan, Z. Chen, S. Wu, J. Zhao, Wheat germ agglutinin-grafted lipid nanoparticles: preparation and in vitro evaluation of the association with Caco-2 monolayers, *Int J Pharm* **397**, 155–163 (2010).
64. Y. Liu, P. Wang, C. Sun, J. Zhao, Y. Du, F. Shi, N. Feng, Bioadhesion and enhanced bioavailability by wheat germ agglutinin-grafted lipid nanoparticles for oral delivery of poorly water-soluble drug bufalin, *Int J Pharm* **419**, 260–265 (2011).
65. C. Gamazo, N. Martín-Arbella, A. Brotons, A. I. Camacho, J. M. Irache, Mimicking microbial strategies for the design of mucus-permeating nanoparticles for oral immunization, *Eur J Pharm Biopharm*, doi:10.1016/j.ejpb.2015.01.010.
66. M. J. Montisci, G. Giovannuci, D. Duchêne, G. Ponchel, Covalent coupling of asparagus pea and tomato lectins to poly(lactide) microspheres, *Int J Pharm* **215**, 153–161 (2001).
67. M. Wirth, K. Gerhardt, C. Wurm, F. Gabor, Lectin-mediated drug delivery: influence of mucin on cytoadhesion of plant lectins in vitro, *J Control Release* **79**, 183–191 (2002).
68. T. L. Cover, M. J. Blaser, Helicobacter Pylori and Gastrointestinal Disease, *Annu Rev Med* **43**, 135–145 (1992).
69. J. A. Louw, V. Falck, C. van Rensburg, J. Zak, G. Adams, I. N. Marks, Distribution of Helicobacter pylori colonisation and associated gastric inflammatory changes: difference between patients with duodenal and gastric ulcers, *J Clin Pathol* **46**, 754–756 (1993).
70. D. R. Cave, How is Helicobacter pylori transmitted? *Gastroenterology* **113**, S9–14 (1997).
71. J. G. Kusters, A. H. M. van Vliet, E. J. Kuipers, Pathogenesis of Helicobacter pylori infection, *Clin Microbiol Rev* **19**, 449–490 (2006).
72. D. B. Polk, R. M. Peek, Helicobacter pylori: gastric cancer and beyond, *Nat Rev Cancer* **10**, 403–414 (2010).
73. B. Linz, F. Balloux, Y. Moodley, A. Manica, H. Liu, P. Roumagnac, D. Falush, C. Stamer, F. Prugnolle, S. W. van der Merwe, Y. Yamaoka, D. Y. Graham, E. Perez-Trallero, T. Wadström, S. Suerbaum, M. Achtman, An African origin for the intimate association between humans and Helicobacter pylori, *Nature* **445**, 915–918 (2007).
74. D. L. Weeks, S. Eskandari, D. R. Scott, G. Sachs, A H⁺-gated urea channel: the link between Helicobacter pylori urease and gastric colonization, *Science* **287**, 482–485 (2000).

75. P. Bauerfeind, R. Garner, B. E. Dunn, H. L. Mobley, Synthesis and activity of Helicobacter pylori urease and catalase at low pH, *Gut* **40**, 25–30 (1997).
76. H. Yoshiyama, H. Nakamura, M. Kimoto, K. Okita, T. Nakazawa, Chemotaxis and motility of Helicobacter pylori in a viscous environment, *J. Gastroenterol.* **34 Suppl 11**, 18–23 (1999).
77. C. Montecucco, R. Rappuoli, Living dangerously: how Helicobacter pylori survives in the human stomach, *Nat Rev Mol Cell Biol* **2**, 457–466 (2001).
78. M. E. Moore, T. Borén, J. V. Solnick, Life at the margins: modulation of attachment proteins in Helicobacter pylori, *Gut Microbes* **2**, 42–46 (2011).
79. S. Schreiber, M. Konradt, C. Groll, P. Scheid, G. Hanauer, H.-O. Werling, C. Josenhans, S. Suerbaum, The spatial orientation of Helicobacter pylori in the gastric mucus, *Proc Natl Acad Sci USA* **101**, 5024–5029 (2004).
80. P. Falk, K. A. Roth, T. Borén, T. U. Westblom, J. I. Gordon, S. Normark, An in vitro adherence assay reveals that Helicobacter pylori exhibits cell lineage-specific tropism in the human gastric epithelium, *Proc Natl Acad Sci USA* **90**, 2035–2039 (1993).
81. S. Odenbreit, M. Till, D. Hofreuter, G. Faller, R. Haas, Genetic and functional characterization of the alpAB gene locus essential for the adhesion of Helicobacter pylori to human gastric tissue, *Mol Microbiol* **31**, 1537–1548 (1999).
82. D. Ilver, A. Arnqvist, J. Ogren, I. M. Frick, D. Kersulyte, E. T. Incecik, D. E. Berg, A. Covacci, L. Engstrand, T. Borén, Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging, *Science* **279**, 373–377 (1998).
83. B. Peck, M. Ortkamp, K. D. Diehl, E. Hundt, B. Knapp, Conservation, localization and expression of HopZ, a protein involved in adhesion of Helicobacter pylori, *Nucleic Acids Res* **27**, 3325–3333 (1999).
84. Y. Rossez, P. Gosset, I. G. Boneca, A. Magalhães, C. Ecobichon, C. A. Reis, C. Cieniewski-Bernard, M. Joncquel Chevalier Curt, R. Léonard, E. Maes, B. Sperandio, C. Slomianny, P. J. Sansonetti, J.-C. Michalski, C. Robbe-Masselot, The lacdiNAc-specific adhesin LabA mediates adhesion of Helicobacter pylori to human gastric mucosa, *J Infect Dis* **210**, 1286–1295 (2014).
85. J. Mahdavi, B. Sondén, M. Hurtig, F. O. Olfat, L. Forsberg, N. Roche, J. Ångström, T. Larsson, S. Teneberg, K.-A. Karlsson, S. Altraja, T. Wadström, D. Kersulyte, D. E. Berg, A. Dubois, C. Petersson, K.-E. Magnusson, T. Norberg, F. Lindh, B. B. Lundskog, A. Arnqvist, L. Hammarström, T. Borén, Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation, *Science* **297**, 573–578 (2002).
86. R. A. Alm, J. Bina, B. M. Andrews, P. Doig, R. E. Hancock, T. J. Trust, Comparative genomics of Helicobacter pylori: analysis of the outer membrane protein families, *Infect Immun* **68**, 4155–4168 (2000).
87. J. F. Tomb, O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D.

Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzegerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Wathley, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, J. C. Venter, The complete genome sequence of the gastric pathogen *Helicobacter pylori*, *Nature* **388**, 539–547 (1997).

88. A. Dossumbekova, C. Prinz, J. Mages, R. Lang, J. G. Kusters, A. H. M. van Vliet, W. Reindl, S. Backert, D. Saur, R. M. Schmid, R. Rad, *Helicobacter pylori* HopH (OipA) and bacterial pathogenicity: genetic and functional genomic analysis of hopH gene polymorphisms, *J Infect Dis* **194**, 1346–1355 (2006).

89. M. Huesca, A. Goodwin, A. Bhagwansingh, P. Hoffman, C. A. Lingwood, Characterization of an acidic-pH-inducible stress protein (hsp70), a putative sulfatide binding adhesin, from *Helicobacter pylori*, *Infect Immun* **66**, 4061–4067 (1998).

90. H. Miller-Podraza, K. Weikkolainen, T. Larsson, P. Linde, J. Helin, J. Natunen, K.-A. Karlsson, *Helicobacter pylori* binding to new glycans based on N-acetyllactosamine, *Glycobiology* **19**, 399–407 (2009).

91. H. Miller-Podraza, B. Lanne, J. Ångström, S. Teneberg, M. A. Milh, P.-A. Jovall, H. Karlsson, K.-A. Karlsson, Novel binding epitope for *Helicobacter pylori* found in neolacto carbohydrate chains: structure and cross-binding properties, *J Biol Chem* **280**, 19695–19703 (2005).

92. S. K. Lindén, C. Wickström, G. Lindell, K. Gilshenan, I. Carlstedt, Four modes of adhesion are used during *Helicobacter pylori* binding to human mucins in the oral and gastric niches, *Helicobacter* **13**, 81–93 (2008).

93. A. Magalhães, C. A. Reis, *Helicobacter pylori* adhesion to gastric epithelial cells is mediated by glycan receptors, *Braz J Med Biol Res* **43**, 611–618 (2010).

94. Y. Rossez, P. Gosset, I. G. Boneca, A. Magalhães, C. Ecobichon, C. A. Reis, C. Cieniewski-Bernard, M. Joncquel Chevalier Curt, R. Leonard, E. Maes, B. Sperandio, C. Slomianny, P. J. Sansonetti, J. C. Michalski, C. Robbe-Masselot, The LacdiNAc-Specific Adhesin LabA Mediates Adhesion of *Helicobacter pylori* to Human Gastric Mucosa, *J Infect Dis* **210**, 1286–1295 (2014).

95. V. Ravn, E. Dabelsteen, Tissue distribution of histo-blood group antigens, *APMIS* **108**, 1–28 (2000).

96. G. R. Van den Brink, K. M. Tytgat, R. W. Van der Hulst, C. M. Van der Loos, A. W. Einerhand, H. A. Büller, J. Dekker, *H pylori* colocalises with MUC5AC in the human stomach, *Gut* **46**, 601–607 (2000).

97. T. Borén, P. Falk, K. A. Roth, G. Larson, S. Normark, Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens, *Science* **262**, 1892–1895 (1993).

98. L. S. Tzouvelekis, A. F. Mentis, A. M. Makris, C. Spiliadis, C. Blackwell, D. M. Weir, In vitro binding of *Helicobacter pylori* to human gastric mucin, *Infect Immun* **59**, 4252–4254 (1991).
99. K. S. B. Bergstrom, L. Xia, Mucin-type O-glycans and their roles in intestinal homeostasis, *Glycobiology* **23**, 1026–1037 (2013).
100. S. B. Ho, A. M. Robertson, L. L. Shekels, C. T. Lyftogt, G. A. Niehans, N. W. Toribara, Expression cloning of gastric mucin complementary DNA and localization of mucin gene expression, *Gastroenterology* **109**, 735–747 (1995).
101. C. De Bolós, M. Garrido, F. X. Real, MUC6 apomucin shows a distinct normal tissue distribution that correlates with Lewis antigen expression in the human stomach, *Gastroenterology* **109**, 723–734 (1995).
102. S. Lindén, H. Nordman, J. Hedenbro, M. Hurtig, T. Borén, I. Carlstedt, Strain- and blood group-dependent binding of *Helicobacter pylori* to human gastric MUC5AC glycoforms, *Gastroenterology* **123**, 1923–1930 (2002).
103. J. H. B. Van de Bovenkamp, J. Mahdavi, A. M. Korteland-Van Male, H. A. Büller, A. W. C. Einerhand, T. Borén, J. Dekker, The MUC5AC glycoprotein is the primary receptor for *Helicobacter pylori* in the human stomach, *Helicobacter* **8**, 521–532 (2003).
104. M. A. McGuckin, A. L. Every, C. D. Skene, S. K. Lindén, Y. T. Chionh, A. Swierczak, J. McAuley, S. Harbour, M. Kaparakis, R. Ferrero, P. Sutton, Muc1 mucin limits both *Helicobacter pylori* colonization of the murine gastric mucosa and associated gastritis, *Gastroenterology* **133**, 1210–1218 (2007).
105. B. Macao, D. G. A. Johansson, G. C. Hansson, T. Härd, Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin, *Nat. Struct. Mol. Biol.* **13**, 71–76 (2006).
106. T. Pelaseyed, M. Zäch, A. C. Petersson, F. Svensson, D. G. A. Johansson, G. C. Hansson, Unfolding dynamics of the mucin SEA domain probed by force spectroscopy suggest that it acts as a cell-protective device, *FEBS J.* **280**, 1491–1501 (2013).
107. S. K. Lindén, Y. H. Sheng, A. L. Every, K. M. Miles, E. C. Skoog, T. H. J. Florin, P. Sutton, M. A. McGuckin, MUC1 Limits *Helicobacter pylori* Infection both by Steric Hindrance and by Acting as a Releasable Decoy, *PLoS Pathog* **5**, e1000617 (2009).
108. C. A. Reis, L. David, P. Correa, F. Carneiro, C. De Bolós, E. Garcia, U. Mandel, H. Clausen, M. Sobrinho-Simões, Intestinal metaplasia of human stomach displays distinct patterns of mucin (MUC1, MUC2, MUC5AC, and MUC6) expression, *Cancer Research* **59**, 1003–1007 (1999).
109. J. Younson, R. O'Mahony, H. Liu, S. Grant, C. Champion, L. Jennings, D. Vaira, C. G. Kelly, I. M. Roitt, J. Holton, A Human Domain Antibody and Lewisb Glycoconjugate That Inhibit Binding of *Helicobacter pylori* to Lewisb Receptor and Adhesion to Human Gastric Epithelium, *J Infect Dis* **200**, 1574–1582 (2009).
110. O. Björnham, J. Bugaytsova, T. Borén, S. Schedin, Dynamic force

spectroscopy of the *Helicobacter pylori* BabA–Lewis b Binding, *Biophys Chem* **143**, 102–105 (2009).

111. P. Parreira, Q. Shi, A. Magalhães, C. A. Reis, J. Bugaytsova, T. Borén, D. Leckband, M. C. L. Martins, Atomic force microscopy measurements reveal multiple bonds between *Helicobacter pylori* blood group antigen binding adhesin and Lewis b ligand, *J R Soc Interface* **11**, 20141040 (2014).

112. Y. Y. Fei, A. Schmidt, G. Bylund, D. X. Johansson, S. Henriksson, C. Lebrilla, J. V. Solnick, T. Borén, X. D. Zhu, Use of real-time, label-free analysis in revealing low-affinity binding to blood group antigens by *Helicobacter pylori*, *Anal Chem* **83**, 6336–6341 (2011).

113. A. Walz, S. Odenbreit, J. Mahdavi, T. Borén, S. Ruhl, Identification and characterization of binding properties of *Helicobacter pylori* by glycoconjugate arrays, *Glycobiology* **15**, 700–708 (2005).

114. A. Magalhães, J. Gomes, M. N. Ismail, S. M. Haslam, N. Mendes, H. Osório, L. David, J. Le Pendu, R. Haas, A. Dell, T. Borén, C. A. Reis, Fut2-null mice display an altered glycosylation profile and impaired BabA-mediated *Helicobacter pylori* adhesion to gastric mucosa, *Glycobiology* **19**, 1525–1536 (2009).

115. J. Celik, B. Su, U. Tirén, Y. Finkel, A. C. Thoresson, L. Engstrand, B. Sandstedt, S. Bernander, S. Normark, Virulence and colonization-associated properties of *Helicobacter pylori* isolated from children and adolescents, *J Infect Dis* **177**, 247–252 (1998).

116. D. T. Pride, R. J. Meinersmann, M. J. Blaser, Allelic Variation within *Helicobacter pylori* babA and babB, *Infect Immun* **69**, 1160–1171 (2001).

117. M. J. Blaser, D. E. Berg, *Helicobacter pylori* genetic diversity and risk of human disease, *J Clin Invest* **107**, 767–773 (2001).

118. M. Gerhard, N. Lehn, N. Neumayer, T. Borén, R. Rad, W. Schepp, S. Miehke, M. Classen, C. Prinz, Clinical relevance of the *Helicobacter pylori* gene for blood-group antigen-binding adhesin, *Proc Natl Acad Sci USA* **96**, 12778–12783 (1999).

119. J. Colbeck, L. Hansen, J. Fong, J. Solnick, Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of *Helicobacter pylori*, *Infect Immun* **74**, 4375–4378 (2006).

120. E. E. Hennig, J. M. Allen, T. L. Cover, Multiple chromosomal loci for the babA gene in *Helicobacter pylori*, *Infect Immun* **74**, 3046–3051 (2006).

121. E. E. Hennig, R. Mernaugh, J. Edl, P. Cao, T. L. Cover, Heterogeneity among *Helicobacter pylori* Strains in Expression of the Outer Membrane Protein BabA, *Infect Immun* **72**, 3429–3435 (2004).

122. S. Odenbreit, K. Swoboda, I. Barwig, S. Ruhl, T. Borén, S. Koletzko, R. Haas, Outer Membrane Protein Expression Profile in *Helicobacter pylori* Clinical Isolates, *Infect Immun* **77**, 3782–3790 (2009).

123. M. Aspholm-Hurtig, G. Dailide, M. Lahmann, A. Kalia, D. Ilver, N. Roche, S. Vikström, R. Sjöström, S. Lindén, A. Bäckström, C. Lundberg, A. Arnqvist, J.

- Mahdavi, U. J. Nilsson, B. Velapatiño, R. H. Gilman, M. Gerhard, T. Alarcon, M. López-Brea, T. Nakazawa, J. G. Fox, P. Correa, M. G. Dominguez-Bello, G. I. Perez-Perez, M. J. Blaser, S. Normark, I. Carlstedt, S. Oscarson, S. Teneberg, D. E. Berg, T. Borén, Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin, *Science* **305**, 519–522 (2004).
124. S. Nell, L. Kennemann, S. Schwarz, C. Josenhans, S. Suerbaum, Dynamics of Lewis b Binding and Sequence Variation of the babA Adhesin Gene during Chronic *Helicobacter pylori* Infection in Humans, *mBio* **5**, e02281–14 (2014).
125. Duodenitis (available at <http://www.fairview.org>; Copyright: Fairview; Accessed on 23/09/2015).
126. N. Kamaly, Z. Xiao, P. M. Valencia, A. F. Radovic-Moreno, O. C. Farokhzad, Targeted polymeric therapeutic nanoparticles: design, development and clinical translation, *Chem Soc Rev* **41**, 2971–3010 (2012).
127. *pET-22b(+)* Vector Technical Bulletin TB038 (Merck Millipore, Darmstadt, Germany, 1998).
128. *E. coli* Expression vector *pOPE101* Technical Bulletin PR3004/DS-170212 (Progen Biotechnik GmbH, Heidelberg, Germany, 2012).
129. K. Terpe, Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems, *Appl Microbiol Biotechnol* **72**, 211–222 (2006).
130. J.-C. Drugmand, Y.-J. Schneider, S. N. Agathos, Biotechnology Advances, *Biotech Adv* **30**, 1140–1157 (2012).
131. V. Jager, K. B. ssow, A. Wagner, S. Weber, M. Hust, A. Frenzel, T. Schirrmann, High level transient production of recombinant antibodies and antibody fusion proteins in HEK293 cells, *BMC Biotechnol.* **13**, 1–1 (2013).
132. O. Daramola, G. Dean, D. Hatton, J. Stevenson, Cells for transient expression and uses thereof, *US Patent 13/395,339* (2012).
133. M. A. Cooper, Optical biosensors in drug discovery, *Nat Rev Drug Discov* **1**, 515–528 (2002).
134. R. Bakhtiar, Surface plasmon resonance spectroscopy: a versatile technique in a biochemist's toolbox, *J Chem Educ* **90**, 203–209 (2013).
135. C. Hahnefeld, S. Drewianka, F. W. Herberg, Determination of kinetic data using surface plasmon resonance biosensors, *Methods Mol Med* **94**, 299–320 (2004).
136. *Biacore™ Assay Handbook 29-0194-00 Edition AA* (GE Healthcare Life Sciences, Uppsala, Sweden, 2012).
137. A. Onell, K. Andersson, Kinetic determinations of molecular interactions using Biacore-minimum data requirements for efficient experimental design, *J Mol Recognit* **18**, 307–317 (2005).
138. I. Jelesarov, H. R. Bosshard, Isothermal titration calorimetry and

differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition, *J Mol Recognit* **12**, 3–18 (1999).

139. M. M. Pierce, C. S. Raman, B. T. Nall, Isothermal titration calorimetry of protein-protein interactions, *Methods* **19**, 213–221 (1999).

140. S. Leavitt, E. Freire, Direct measurement of protein binding energetics by isothermal titration calorimetry, *Curr Opin Struc Biol* **11**, 560–566 (2001).

141. W. B. Turnbull, A. H. Daranas, On the value of *c*: can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc* **125**, 14859–14866 (2003).

142. E. Freire, Isothermal titration calorimetry: controlling binding forces in lead optimization, *Drug Discov Today Technol* **1**, 295–299 (2004).

143. E. Freire, Do Enthalpy and Entropy Distinguish First in Class From Best in Class? *Drug Discov Today* **13**, 1–11 (2008).

144. M. S. Smyth, J. Martin, x Ray crystallography, *Mol Pathol* **53**, 8–14 (2000).

145. M. A. Dessau, Y. Modis, Protein Crystallization for X-ray Crystallography, *JoVE* **47**, e2285 (2011).

146. G. L. Taylor, Introduction to phasing, *Acta Crystallogr D* **66**, 325–338 (2010).

147. L. M. Rice, T. N. Earnest, A. T. Brunger, Single-wavelength anomalous diffraction phasing revisited, *Acta Crystallogr D* **56**, 1413–1420 (2000).

148. G. Rhodes, *Crystallography Made Crystal Clear (Third Edition)* (Academic Press, London, 2006).

149. N. J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, *Nat Protoc* **1**, 2876–2890 (2007).

150. S. M. Kelly, T. J. Jess, N. C. Price, How to study proteins by circular dichroism, *BBA-Proteins Proteom* **1751**, 119–139 (2005).

151. S. M. Kelly, N. C. Price, The use of circular dichroism in the investigation of protein structure and function, *Curr Protein Pept Sci* **1**, 349–384 (2000).

152. N. Greenfield, G. D. Fasman, Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry* **8**, 4108–4116 (1969).

153. L. Whitmore, B. Woollett, A. J. Miles, D. P. Klose, R. W. Janes, B. A. Wallace, PCDDDB: the Protein Circular Dichroism Data Bank, a repository for circular dichroism spectral and metadata, *Nucleic Acids Res* **39**, D480–6 (2011).

154. R. J. Johnson, C. J. Savas, Z. Kartje, G. C. Hoops, Rapid and Adaptable Measurement of Protein Thermal Stability by Differential Scanning Fluorimetry: Updating a Common Biochemical Laboratory Experiment, *J Chem Educ* **91**, 1077–1080 (2014).

155. F. H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nat Protoc* **2**, 2212–2221 (2007).
156. S. Boivin, S. Kozak, R. Meijers, Optimization of protein purification and characterization using Thermofluor screens, *Protein Expr Purif* **91**, 192–206 (2013).
157. F. Gambinossi, S. E. Mylon, J. K. Ferri, Aggregation kinetics and colloidal stability of functionalized nanoparticles, *Adv Colloid Interface Sci* **222**, 332–349 (2015).
158. J. W. Goodman, Some fundamental properties of speckle, *JOSA* **66**, 1145–1150 (1976).
159. R. Pecora, Dynamic light scattering measurement of nanometer particles in liquids, *J Nanopart Res* (2000).
160. Y. Abdiche, D. Malashock, A. Pinkerton, J. Pons, Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet, *Anal Biochem* **377**, 209–217 (2008).
161. J. Wallner, G. Lhota, D. Jeschek, A. Mader, K. Vorauer-Uhl, Application of Bio-Layer Interferometry for the analysis of protein/liposome interactions, *J Pharmaceut Biomed* **72**, 150–154 (2013).
162. T. Do, F. Ho, B. Heidecker, K. Witte, L. Chang, L. Lerner, A rapid method for determining dynamic binding capacity of resins for the purification of proteins, *Protein Expr Purif* **60**, 147–150 (2008).
163. S. Kumaraswamy, R. Tobias, Label-free kinetic analysis of an antibody-antigen interaction using biolayer interferometry, *Methods Mol Biol* **1278**, 165–182 (2015).
164. N. Ishijima, M. Suzuki, H. Ashida, Y. Ichikawa, Y. Kanegae, I. Saito, T. Borén, R. Haas, C. Sasakawa, H. Mimuro, BabA-mediated Adherence Is a Potentiator of the Helicobacter pylori Type IV Secretion System Activity, *J Biol Chem* **286**, 25256–25264 (2011).
165. M. M. Reyna, Structural and functional studies of mucin-interacting adhesion domains from Candida glabrata & Helicobacter pylori, *PhD Thesis, Philipps-Universität Marburg* (2011).
166. A. de Marco, Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli, *Microb Cell Fact* **8**, 26 (2009).
167. S. Harper, D. W. Speicher, Purification of Proteins Fused to Glutathione S-Transferase, *Methods Mol Biol* **681**, 259–280 (2010).
168. C. P. Papaneophytou, G. Kontopidis, Statistical approaches to maximize recombinant protein expression in Escherichia coli: A general review, *Protein Expr Purif* **94**, 22–32 (2014).
169. K. Tsumoto, D. Ejima, I. Kumagai, T. Arakawa, Practical considerations in refolding proteins from inclusion bodies, *Protein Expr Purif* **28**, 1–8 (2003).

170. R. Grisshammer, Understanding recombinant expression of membrane proteins, *Curr Opin Biotechnol* **17**, 337–340 (2006).
171. N. Dautin, H. D. Bernstein, Protein Secretion in Gram-Negative Bacteria via the Autotransporter Pathway, *Annu Rev Microbiol* **61**, 89–112 (2007).
172. L. A. Kelley, M. J. E. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server, *Nat Protoc* **4**, 363–371 (2009).
173. E. J. McCall, A. Danielsson, I. M. Hardern, C. Dartsch, R. Hicks, J. M. Wahlberg, W. M. Abbott, Improvements to the throughput of recombinant protein expression in the baculovirus/insect cell system, *Protein Expr Purif* **42**, 29–36 (2005).
174. M. M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal Biochem* **72**, 248–254 (1976).
175. S. S. Pang, S. T. S. Nguyen, A. J. Perry, C. J. Day, S. Panjikar, J. Tiralongo, J. C. Whisstock, T. Kwok, The three-dimensional structure of the extracellular adhesion domain of the sialic acid-binding adhesin SabA from *Helicobacter pylori*, *J Biol Chem* **289**, 6332–6340 (2014).
176. G. Psakis, S. Nitschkowski, C. Holz, D. Kreß, M. Maestre-Reyna, J. Polaczek, G. Illing, L.-O. Essen, Expression screening of integral membrane proteins from *Helicobacter pylori* 26695, *Protein Sci* **16**, 2667–2676 (2007).
177. S. Subedi, K. Moonens, E. Romão, A. Lo, G. Vandebussche, J. Bugaytsova, S. Muyldermans, T. Borén, H. Remaut, Expression, purification and X-ray crystallographic analysis of the *Helicobacter pylori* blood group antigen-binding adhesin BabA, *Acta Crystallogr F* **70**, 1631–1635 (2014).
178. N. Ruiz, S.-S. Chng, A. Hiniker, D. Kahne, T. J. Silhavy, Nonconsecutive disulfide bond formation in an essential integral outer membrane protein, *Proc Natl Acad Sci USA* **107**, 12245–12250 (2010).
179. S. H. Yoon, S. K. Kim, J. F. Kim, Secretory production of recombinant proteins in *Escherichia coli*, *Recent Pat Biotechnol* **4**, 23–29 (2010).
180. F. W. Studier, B. A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *J Mol Biol* **189**, 113–130 (1986).
181. H. Sørensen, K. Mortensen, Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*, *Microb Cell Fact* **4**, 1 (2005).
182. D. Liu, R. D. Schmid, M. Rusnak, Functional expression of *Candida antarctica* lipase B in the *Escherichia coli* cytoplasm—a screening system for a frequently used biocatalyst, *Appl Microbiol Biotechnol* **72**, 1024–1032 (2006).
183. P. H. Bessette, F. Aslund, J. Beckwith, G. Georgiou, Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm, *Proc Natl Acad Sci USA* **96**, 13703–13708 (1999).
184. A. I. Derman, W. A. Prinz, D. Belin, J. Beckwith, Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*, *Science* **262**,

1744–1747 (1993).

185. A. Schmiedl, F. Breitling, C. H. Winter, I. Queitsch, S. Dübel, Effects of unpaired cysteines on yield, solubility and activity of different recombinant antibody constructs expressed in *E. coli*, *J Immunol Methods* **242**, 101–114 (2000).

186. N. Casali, A. Preston, *E. coli plasmid vectors: Methods and applications* (Humana Press, Tototwa, 2003).

187. G. Salinas, L. Pellizza, M. Margenat, M. Fló, C. Fernández, Tuned *Escherichia coli* as a host for the expression of disulfide-rich proteins, *Biotechnol J* **6**, 686–699 (2011).

188. J. E. Nettleship, R. Assenberg, J. M. Diprose, N. Rahman-Huq, R. J. Owens, Recent advances in the production of proteins in insect and mammalian cells for structural biology, *J Struct Biol* **172**, 55–65 (2010).

189. P. Meissner, H. Pick, A. Kulangara, P. Chatellard, K. Friedrich, F. M. Wurm, Transient gene expression: recombinant protein production with suspension-adapted HEK293-EBNA cells, *Biotechnol Bioeng* **75**, 197–203 (2001).

190. J. L. Yates, N. Warren, B. Sugden, Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells, *Nature* **313**, 812–815 (1985).

191. S. Kühn, P. F. Zipfel, The baculovirus expression vector pBSV-8His directs secretion of histidine-tagged proteins, *Gene* **162**, 225–229 (1995).

192. D. C. Tessier, D. Y. Thomas, H. E. Khouri, F. Laliberté, T. Vernet, Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide, *Gene* **98**, 177–183 (1991).

193. K. Champasa, S. A. Longwell, A. M. Eldridge, E. A. Stemmler, D. H. Dube, Targeted identification of glycosylated proteins in the gastric pathogen *Helicobacter pylori* (Hp), *Mol Cell Proteomics* **12**, 2568–2586 (2013).

194. B. Gasser, M. Saloheimo, U. Rinas, M. Dragosits, E. Rodríguez-Carmona, K. Baumann, M. Giuliani, E. Parrilli, P. Branduardi, C. Lang, D. Porro, P. Ferrer, M. L. Tutino, D. Mattanovich, A. Villaverde, Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview, *Microb Cell Fact* **7**, 11 (2008).

195. P. Whiteman, P. A. Handford, Defective secretion of recombinant fragments of fibrillin-1: implications of protein misfolding for the pathogenesis of Marfan syndrome and related disorders, *Hum Mol Genet* **12**, 727–737 (2003).

196. M. Nita-Lazar, M. Wacker, B. Schegg, S. Amber, M. Aebi, The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation, *Glycobiology* **15**, 361–367 (2005).

197. L. Allard, V. Cheynet, G. Oriol, B. Mandrand, T. Delair, F. Mallet, Versatile method for production and controlled polymer-immobilization of biologically active recombinant proteins, *Biotechnol Bioeng* **80**, 341–348

(2002).

198. T. K. Tiefenbrunn, P. E. Dawson, J. Schneider, Ed. Chemoselective ligation techniques: Modern applications of time-honored chemistry, *Biopolymers* **94**, 95–106 (2010).

199. H.-J. Jung, S.-K. Kim, W.-K. Min, S.-S. Lee, K. Park, Y.-C. Park, J.-H. Seo, Polycationic amino acid tags enhance soluble expression of *Candida antarctica* lipase B in recombinant *Escherichia coli*, *Bioprocess Biosyst Eng* **34**, 833–839 (2011).

200. A. Kato, K. Maki, T. Ebina, K. Kuwajima, K. Soda, Y. Kuroda, Mutational analysis of protein solubility enhancement using short peptide tags, *Biopolymers* **85**, 12–18 (2007).

201. M. A. Hossain, A. Belgi, F. Lin, S. Zhang, F. Shabanpoor, L. Chan, C. Belyea, H.-T. Truong, A. R. Blair, S. Andrikopoulos, G. W. Tregear, J. D. Wade, Use of a temporary “solubilizing” peptide tag for the Fmoc solid-phase synthesis of human insulin glargine via use of regioselective disulfide bond formation, *Bioconjug Chem* **20**, 1390–1396 (2009).

202. S. H. Park, A. A. Mrse, A. A. Nevzorov, M. F. Mesleh, M. Oblatt-Montal, M. Montal, S. J. Opella, Three-dimensional Structure of the Channel-forming Trans-membrane Domain of Virus Protein “u” (Vpu) from HIV-1, *J Mol Biol* **333**, 409–424 (2003).

203. *Ni Sepharose™ Excel Instructions 29-0138-67 AE* (GE Healthcare Life Sciences, Uppsala, Sweden, 2014).

204. M. Merdanovic, T. Clausen, M. Kaiser, R. Huber, M. Ehrmann, Protein Quality Control in the Bacterial Periplasm, *Annu Rev Microbiol* **65**, 149–168 (2011).

205. T. Clausen, C. Southan, M. Ehrmann, The HtrA Family of Proteases: Implications for Protein Composition and Cell Fate, *Molecular cell* **10**, 443–455 (2002).

206. K. R. Silber, K. C. Keiler, R. T. Sauer, Tsp: a tail-specific protease that selectively degrades proteins with nonpolar C termini, *Proc Natl Acad Sci USA* **89**, 295–299 (1992).

207. S. B. Ho, K. Takamura, R. Anway, L. L. Shekels, N. W. Toribara, H. Ota, The adherent gastric mucous layer is composed of alternating layers of MUC5AC and MUC6 mucin proteins, *Dig Dis Sci* **49**, 1598–1606 (2004).

208. S. Odenbreit, Adherence properties of *Helicobacter pylori*: Impact on pathogenesis and adaptation to the host, *Int J Med Microbiol* **295**, 317–324 (2005).

209. S. Lindén, J. Mahdavi, C. Semino Mora, C. Olsen, I. Carlstedt, T. Borén, A. Dubois, Role of ABO secretor status in mucosal innate immunity and *H. pylori* infection, *PLoS Pathog* **4**, e2 (2008).

210. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software, *J Appl Crystallogr* **40**, 658–674 (2007).

211. Collaborative Computational Project, Number 4, The CCP4 suite: programs for protein crystallography, *Acta Crystallogr D* **50**, 760–763 (1994).
212. P. Skubák, N. S. Pannu, Automatic protein structure solution from weak X-ray data, *Nat Commun* **4**, 2777 (2013).
213. G. M. Sheldrick, A short history of SHELX, *Acta Crystallogr A* **64**, 112–122 (2008).
214. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr D* **67**, 355–367 (2011).
215. J. P. Abrahams, A. G. W. Leslie, Methods used in the structure determination of bovine mitochondrial F1 ATPase, *Acta Crystallogr D* **52**, 30–42 (1996).
216. P. Skubák, W. J. Waterreus, N. S. Pannu, Multivariate phase combination improves automated crystallographic model building, *Acta Crystallogr D* **66**, 783–788 (2010).
217. K. Cowtan, General quadratic functions in real and reciprocal space and their application to likelihood phasing, *Acta Crystallogr D* **56**, 1612–1621 (2000).
218. K. Cowtan, The Buccaneersoftware for automated model building. 1. Tracing protein chains, *Acta Crystallogr D* **62**, 1002–1011 (2006).
219. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot, *Acta Crystallogr D* **66**, 486–501 (2010).
220. L. Holm, P. Rosenström, Dali server: conservation mapping in 3D, *Nucleic Acids Res* **38**, W545–9 (2010).
221. S. Keller, C. Vargas, H. Zhao, G. Piszczek, C. A. Brautigam, P. Schuck, High-Precision Isothermal Titration Calorimetry with Automated Peak-Shape Analysis, *Anal Chem* **84**, 5066–5073 (2012).
222. V. Consalvi, R. Chiaraluce, L. Giangiacomo, R. Scandurra, P. Christova, A. Karshikoff, S. Knapp, R. Ladenstein, Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile *Thermotoga maritima*, *Protein Eng* **13**, 501–507 (2000).
223. D. G. Myszka, M. D. Jonsen, B. J. Graves, Equilibrium analysis of high affinity interactions using BIACORE, *Anal Biochem* **265**, 326–330 (1998).
224. A. Bohne, E. Lang, C. W. von der Lieth, SWEET - WWW-based rapid 3D construction of oligo- and polysaccharides, *Bioinformatics* **15**, 767–768 (1999).
225. O. C. Grant, H. M. Smith, D. Firsova, E. Fadda, R. J. Woods, Presentation, presentation, presentation! Molecular-level insight into linker effects on glycan array screening data, *Glycobiology* **24**, 17–25 (2013).
226. M. Aspholm, F. O. Olfat, J. Nordén, B. Sondén, C. Lundberg, R. Sjöström, S. Altraja, S. Odenbreit, R. Haas, T. Wadström, L. Engstrand, C. Semino Mora, H. Liu, A. Dubois, S. Teneberg, A. Arnqvist, T. Borén, SabA is the H. pylori

- hemagglutinin and is polymorphic in binding to sialylated glycans, *PLoS Pathog* **2**, e110 (2006).
227. M. Ambrosi, N. R. Cameron, B. G. Davis, Lectins: tools for the molecular understanding of the glycode, *Org Biomol Chem* **3**, 1593 (2005).
228. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, M. E. Etzler, J. D. Esko, N. Sharon, *Microbial Lectins: Hemagglutinins, Adhesins, and Toxins* (Cold Spring Harbor Laboratory Press, New York, ed. 2, 2009).
229. H. M. Bahari, I. N. Ross, L. A. Turnberg, Demonstration of a pH gradient across the mucus layer on the surface of human gastric mucosa in vitro, *Gut* **23**, 513–516 (1982).
230. M. Phillipson, C. Atuma, J. Henriksnäs, L. Holm, The importance of mucus layers and bicarbonate transport in preservation of gastric juxtamucosal pH, *Am J Physiol Gastrointest Liver Physiol* **282**, G211–9 (2002).
231. J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry, Fifth Edition: International Version* (W. H. Freeman, New York, 2002).
232. Z. Bohak, Purification and characterization of chicken pepsinogen and chicken pepsin, *J Biol Chem* **244**, 4638–4648 (1969).
233. A. P. Ryle, The porcine pepsins and pepsinogens, *Methods in Enzymology* **19**, 316–336 (1970).
234. J. C. Powers, A. D. Harley, D. V. Myers, Subsite specificity of porcine pepsin, *Adv Exp Med Biol* **95**, 141–157 (1977).
235. J. Ahn, M.-J. Cao, Y. Q. Yu, J. R. Engen, Assessing the reproducibility and specificity of pepsin and other aspartic proteases, *Biochim Biophys Acta* **1834**, 1222–1229 (2013).
236. I. M. Reddy, N. K. D. Kella, J. E. Kinsella, Structural and conformational basis of the resistance of beta-lactoglobulin to peptic and chymotryptic digestion, *J Agric Food Chem* **36**, 737–741 (1988).
237. A. Allen, G. Flemström, Gastroduodenal mucus bicarbonate barrier: protection against acid and pepsin, *Am J Physiol Cell Physiol* **288**, C1–C19 (2004).
238. K. E. L. McColl, The elegance of the gastric mucosal barrier: designed by nature for nature, *Gut* **61**, 787–788 (2012).
239. M. A. McGuckin, S. K. Lindén, P. Sutton, T. H. Florin, Mucin dynamics and enteric pathogens, *Nat Rev Microbiol* **9**, 265–278 (2011).
240. M. E. V. Johansson, H. Sjövall, G. C. Hansson, The gastrointestinal mucus system in health and disease, *Nat Rev Gastroenterol Hepatol* **10**, 352–361 (2013).
241. H. Nordman, J. R. Davies, G. Lindell, C. de Bolós, F. Real, I. Carlstedt, Gastric MUC5AC and MUC6 are large oligomeric mucins that differ in size, glycosylation and tissue distribution, *Biochem J* **364**, 191–200 (2002).

242. C. Boyer, X. Huang, M. R. Whittaker, V. Bulmus, T. P. Davis, An overview of protein–polymer particles, *Soft Matter* **7**, 1599–1614 (2011).
243. K. Thomas, M. Aalbers, G. A. Bannon, M. Bartels, R. J. Dearman, D. J. Esdaile, T. J. Fu, C. M. Glatt, N. Hadfield, C. Hatzos, S. L. Hefle, J. R. Heylings, R. E. Goodman, B. Henry, C. Herouet, M. Holsapple, G. S. Ladics, T. D. Landry, S. C. MacIntosh, E. A. Rice, L. S. Privalle, H. Y. Steiner, R. Teshima, R. van Ree, M. Woolhiser, J. Zawodny, A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins, *Regul Toxicol Pharm* **39**, 87–98 (2004).
244. P. Artimo, M. Jonnalagedda, K. Arnold, D. Baratin, G. Csardi, E. de Castro, S. Duvaud, V. Flegel, A. Fortier, E. Gasteiger, A. Grosdidier, C. Hernandez, V. Ioannidis, D. Kuznetsov, R. Liechti, S. Moretti, K. Mostaguir, N. Redaschi, G. Rossier, I. Xenarios, H. Stockinger, ExPASy: SIB bioinformatics resource portal, *Nucleic Acids Res* **40**, W597–W603 (2012).
245. B. Keil, *Specificity of Proteolysis* (Springer Science & Business Media, Berlin Heidelberg, 2012).
246. D. Beckett, E. Kovaleva, P. J. Schatz, A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation, *Protein Sci* **8**, 921–929 (1999).
247. D. M. Mudie, G. L. Amidon, G. E. Amidon, Physiological Parameters for Oral Delivery and in Vitro Testing, *Mol Pharmaceutics* **7**, 1388–1405 (2010).
248. A. Allen, G. Flemström, Gastroduodenal mucus bicarbonate barrier: protection against acid and pepsin, *Am J Physiol Cell Physiol* **288**, C1–C19 (2004).
249. N. V. Di Russo, D. A. Estrin, M. A. Martí, A. E. Roitberg, J. M. Briggs, Ed. pH-Dependent Conformational Changes in Proteins and Their Effect on Experimental pKas: The Case of Nitrophorin 4, *PLoS Comput Biol* **8**, e1002761 (2012).
250. M. Roca, B. Messer, A. Warshel, Electrostatic contributions to protein stability and folding energy, *FEBS Lett* **581**, 2065–2071 (2007).
251. M. de Bernard, E. Papini, V. de Filippis, E. Gottardi, J. Telford, R. Manetti, A. Fontana, R. Rappuoli, C. Montecucco, Low pH activates the vacuolating toxin of *Helicobacter pylori*, which becomes acid and pepsin resistant, *J Biol Chem* **270**, 23937–23940 (1995).
252. S. Schreiber, R. Bucker, C. Groll, M. Azevedo-Vethacke, P. Scheid, S. Gatermann, C. Josenhans, S. Suerbaum, Gastric antibacterial efficiency is different for pepsin A and C, *Arch Microbiol* **184**, 335–340 (2005).
253. S. Schreiber, R. Bucker, C. Groll, M. Azevedo-Vethacke, D. Garten, P. Scheid, S. Friedrich, S. Gatermann, C. Josenhans, S. Suerbaum, Rapid Loss of Motility of *Helicobacter pylori* in the Gastric Lumen In Vivo, *Infect Immun* **73**, 1584–1589 (2005).
254. R. Bucker, M. Azevedo-Vethacke, C. Groll, D. Garten, C. Josenhans, S. Suerbaum, S. Schreiber, *Helicobacter pylori* colonization critically depends on

- postprandial gastric conditions, *Sci Rep* **2**, 994 (2012).
255. I. Gritti, G. Banfi, G. S. Roi, Pepsinogens: physiology, pharmacology pathophysiology and exercise, *Pharmacol Res* **41**, 265–281 (2000).
256. T. Klein, D. Abgottsporn, M. Wittwer, S. Rabbani, J. Herold, X. Jiang, S. Kleeb, C. Lüthi, M. Scharenberg, J. Bezençon, E. Gubler, L. Pang, M. Smiesko, B. Cutting, O. Schwaradt, B. Ernst, FimH Antagonists for the Oral Treatment of Urinary Tract Infections: From Design and Synthesis to in Vitro and in Vivo Evaluation, *J Med Chem* **53**, 8627–8641 (2010).
257. C. K. Cusumano, J. S. Pinkner, Z. Han, S. E. Greene, B. A. Ford, J. R. Crowley, J. P. Henderson, J. W. Janetka, S. J. Hultgren, Treatment and prevention of urinary tract infection with orally active FimH inhibitors, *Sci Transl Med* **3**, 109ra115 (2011).
258. P. Malfertheiner, F. Megraud, C. A. O'Morain, J. Atherton, A. T. R. Axon, F. Bazzoli, G. F. Gensini, J. P. Gisbert, D. Y. Graham, T. Rokkas, E. M. El-Omar, E. J. Kuipers, European Helicobacter Study Group, Management of Helicobacter pylori infection-the Maastricht IV/ Florence Consensus Report, *Gut* **61**, 646–664 (2012).
259. P. Krishnamurthy, M. Parlow, J. B. Zitzer, N. B. Vakil, H. L. Mobley, M. Levy, S. H. Phadnis, B. E. Dunn, Helicobacter pylori containing only cytoplasmic urease is susceptible to acid, *Infect Immun* **66**, 5060–5066 (1998).
260. S. Mura, J. Nicolas, P. Couvreur, Stimuli-responsive nanocarriers for drug delivery, *Nat Mater* **12**, 991–1003 (2013).
261. T. Matsui, J. Hamako, Y. Ozeki, K. Titani, Comparative study of blood group-recognizing lectins toward ABO blood group antigens on neoglycoproteins, glycoproteins and complex-type oligosaccharides, *Biochim Biophys Acta* **1525**, 50–57 (2001).
262. A. Pusztai, *Plant Lectins* (Cambridge University Press, Cambridge, 1991).

Appendix

Appendix Table 1: Size calibration of the gel filtration column used in this study

The HiLoad 16/60 Superdex 75 (120 mL) gel filtration column used for size exclusion chromatography was calibrated using the known molecular weight standards indicated.

Standard	Molecular weight (kDa)	Elution volume (mL)	
		Range	Median
Thyroglobulin (bovine)	670	42 - 50	46
γ -globulin (bovine)	158	50 - 62	56
Ovalbumin (chicken)	44	63 - 74	68.5
Myoglobin	17	76 - 91	83.5
Vitamin B ₁₂	1.35	105 - 121	113

Appendix Table 2: Neoglycoconjugates used in ELISA, SPR and BLI assays

Oligosaccharides are linked to HSA lysine residues via a spacer using isothiocyanate coupling. Glycan abbreviations can be interpreted as follows: Fuc – fucose, Gal – galactose, GlcNAc – *N*-acetylglucosamine, Glc – glucose. Spacer abbreviations can be interpreted as follows: APE – *p*-aminophenylethyl, APD – acetyl-phenylenediamine. HSA-Le^y and contained HSA-Le^b contained 13 and 22 moles of oligosaccharide per HSA, respectively.

Name	IUPAC nomenclature
HSA-Le ^y	Fuc α 1-2Gal β 1-4(Fuc α 1-3)GlcNAc β 1-APE-HSA
HSA-Le ^b	Fuc α 1-2Gal β 1-3(Fuc α 1-4)GlcNAc β 1-3Gal β 1-4(Glc)-APD-HSA

Appendix Table 3: ABO/Le blood group antigens used in crystallographic determination and ITC assays

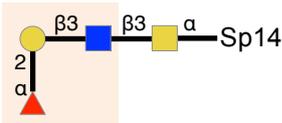
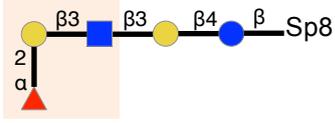
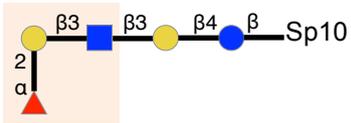
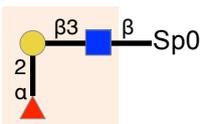
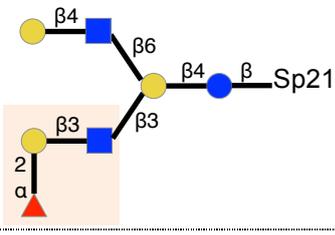
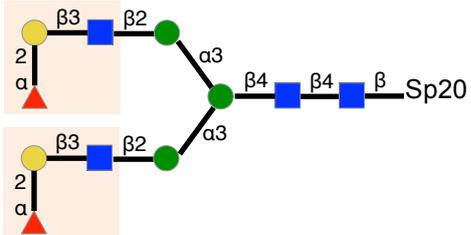
Glycan abbreviations can be interpreted as follows: Fuc – fucose, Gal – galactose, GlcNAc – *N*-acetylglucosamine, Glc – glucose, GalNAc – *N*-acetylgalactosamine, Neu5Ac – *N*-acetylneuraminic acid.

Name	IUPAC nomenclature
Le ^b antigen (hexasaccharide form)	Fuca1-2Galβ1-3(Fuca1-4)GlcNAcβ1-3Galβ1-4Glc
H-1 antigen (pentasaccharide form)	Fuca1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glc
A-1 antigen (hexasaccharide form)	GalNAca1-2Fuca1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glc
B-1 antigen (hexasaccharide form)	Galα1-2Fuca1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glc
Le ^a antigen (pentasaccharide form)	Galβ1-3(Fuca1-4)GlcNAcβ1-3Galβ1-4Glc
Le ^y antigen (pentasaccharide form)	Fuca1-2Galβ1-4(Fuca1-3)GlcNAcβ1-3Gal
H-2 antigen (tetrasaccharide form)	Fuca1-2Galβ1-4GlcNAcβ1-3Gal
SLe ^x antigen (pentasaccharide form)	Neu5Aca2-3Galβ1-4(Fuca1-3)GlcNAcβ1-3Gal

Appendix Table 4: Effect of type 1 ABO blood group antigen epitope presentation on BabA_{527K} binding

Glycan symbolic representations can be interpreted with the following key: fucose – , galactose – , N-acetylgalactosamine – , N-acetylglucosamine – , glucose – , mannose – . Unique determinant regions are surrounded by a light orange shade. Sp0, Sp8, Sp10, Sp14, Sp20 and Sp21 are CFG specific spacer arms.

^aReported values are average relative fluorescence units, standard deviation shown in brackets.

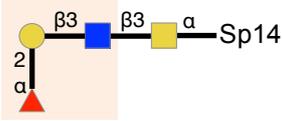
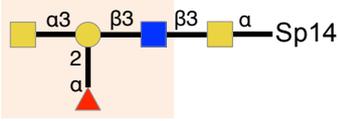
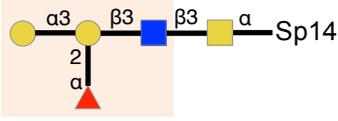
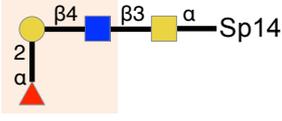
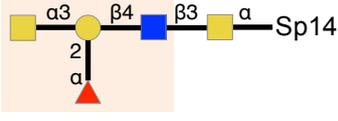
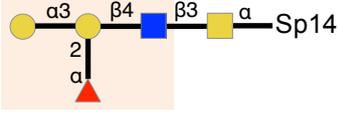
Chart Pos.	Glycan Structure	RFU ^a	H-1 antigen
421		5810 (2385)	
65		2656 (156)	
66		2025 (221)	
67		-119 (35)	
428		127 (67)	
359		389 (66)	

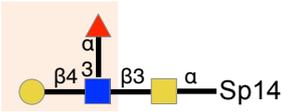
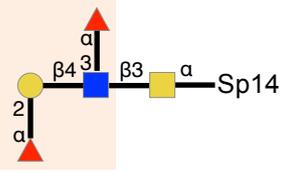
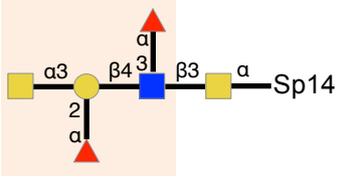
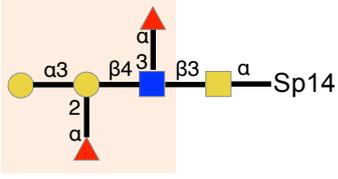
423		2522 (636)	A-1 antigen
82		-52 (149)	
372		-3 (16)	
422		4779 (827)	B-1 antigen
101		56 (57)	
373		-25 (94)	

Appendix Table 5: Comparison of BabA_{527K} binding to type 1 and 2 ABO/Le blood group antigens in a glycan array

For a direct comparison between type 1 and 2 ABO/Le blood group antigens, only those with unique determinant regions that were conjugated to CFG spacer arms via a galactosamine through an α -glycosidic linkage were selected. All other type 1 and 2 ABO/Le blood group antigens on the glycan array were not compared. Glycan symbolic representations can be interpreted with the following key: fucose – \blacktriangle , galactose – \bullet , N-acetylgalactosamine – \blacksquare , N-acetylglucosamine – \blacksquare , glucose – \bullet . Unique determinant regions are surrounded by a light orange shade. Sp14 is a CFG specific spacer arm.

^aReported values are average relative fluorescence units, standard deviation shown in brackets.

Chart Pos.	Glycan Structure	RFU ^a	Glycan	
421		5810 (2385)	H-1 antigen	Type 1
423		2522 (636)	A-1 antigen	
422		4779 (827)	B-1 antigen	
399		-53 (116)	H-2 antigen	Type 2
413		51 (156)	A-2 antigen	
412		-56 (30)	B-2 antigen	

400		83 (31)	Le ^x antigen
415		84 (57)	Le ^y antigen
417		83 (130)	A-Le ^y antigen
416		40 (19)	B-Le ^y antigen

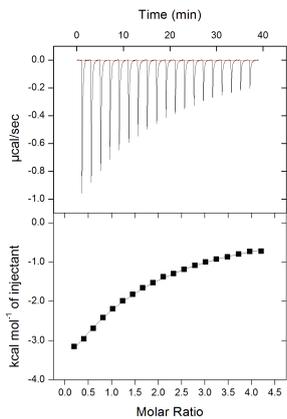
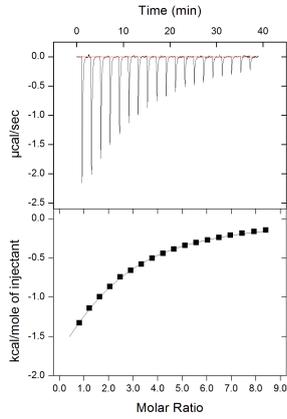
Type 2 (continued)

Appendix Table 6: Thermodynamic parameters of BabA_{527K} binding to various ABO/Le blood group antigens

The upper panels in each ITC trace show a representative calorimetric response obtained by titrating BabA_{527K} with the indicated ABO/Le blood group antigen. The lower panels depict the binding isotherm obtained where the continuous line represents the least-squares fit of the data to a single-site binding model (where applicable). Calorimetric titrations were performed at pH 7.4. The calculated thermodynamic parameters of each binding interaction are reported; values are obtained from a single experiment, unless otherwise indicated.

^aAverage \pm SEM of three independent experiments.

^bAverage \pm Range of two independent experiments.

ABO/Le blood group antigen and Thermodynamic parameters	ITC trace
<p>Le^b antigen hexasaccharide^a</p> <p>$K_D = 252 \pm 15 \mu\text{M}$</p> <p>$N = 1.07 \pm 0.03$</p> <p>$\Delta H = -10.9 \pm 0.5 \text{ kcal/mole}$</p> <p>$-T\Delta S = 6.0 \pm 0.5 \text{ kcal/mole}$</p>	
<p>H-1 antigen pentasaccharide^b</p> <p>$K_D = 617 \pm 45 \mu\text{M}$</p> <p>$N = 1.74 \pm 0.05$</p> <p>$\Delta H = -3.9 \pm 0.2 \text{ kcal/mole}$</p> <p>$-T\Delta S = 0.5 \pm 0.3 \text{ kcal/mole}$</p>	

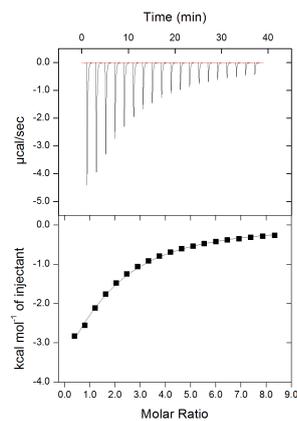
A-1 antigen hexasaccharide

$K_D = 529 \mu\text{M}$

$N = 1.37$

$\Delta H = -8.0 \text{ kcal/mole}$

$-T\Delta S = 3.5 \text{ kcal/mole}$



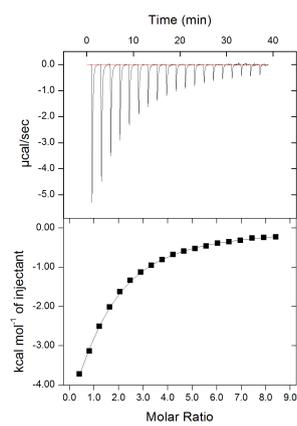
B-1 antigen hexasaccharide

$K_D = 417 \mu\text{M}$

$N = 1.08$

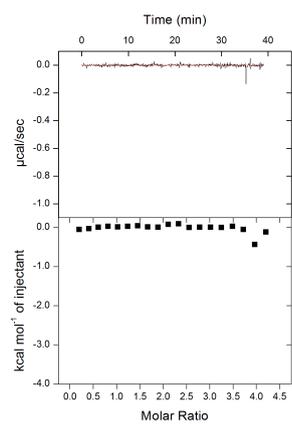
$\Delta H = -10.5 \text{ kcal/mole}$

$-T\Delta S = 5.9 \text{ kcal/mole}$



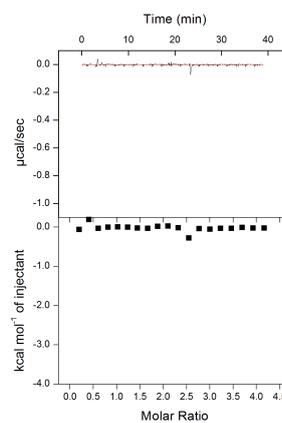
Le^a antigen pentasaccharide

No binding detected



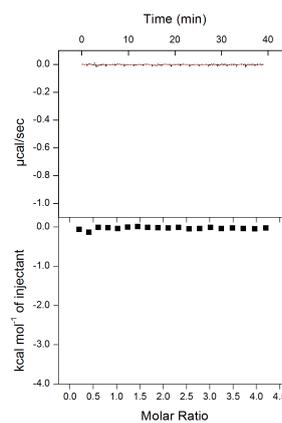
Le^y antigen pentasaccharide

No binding detected



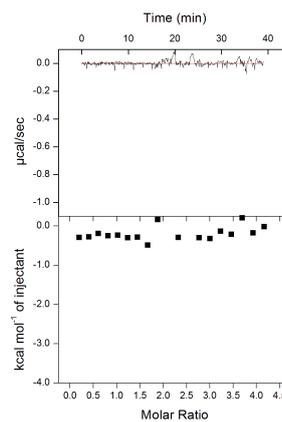
H-2 antigen tetrasaccharide

No binding detected



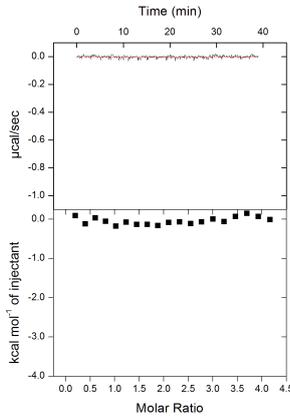
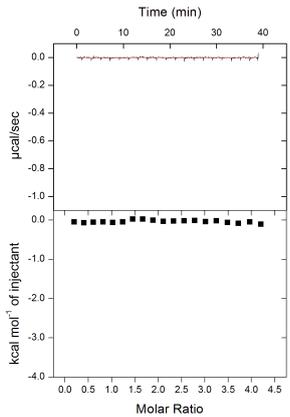
SLe^x antigen pentasaccharide

No binding detected



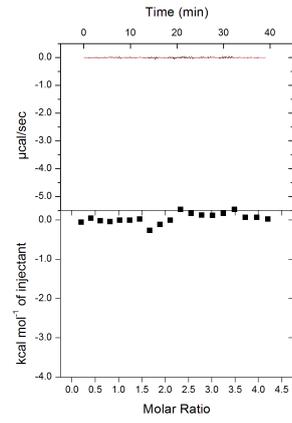
Appendix Table 7: Thermodynamic parameters of BabA_{527K}-D233A/S244A binding to various ABO/Le blood group antigens

Calorimetric titrations of BabA_{527K}-D233A/S244A with the indicated ABO/Le blood group antigens were performed at pH 7.4. No calorimetric response (upper panel) or binding isotherm (lower panel) indicative of a binding interaction was observed in any experiment.

ABO/Le blood group antigen and Thermodynamic parameters	ITC trace
Le^b antigen hexasaccharide No binding detected	
H-1 antigen pentasaccharide No binding detected	

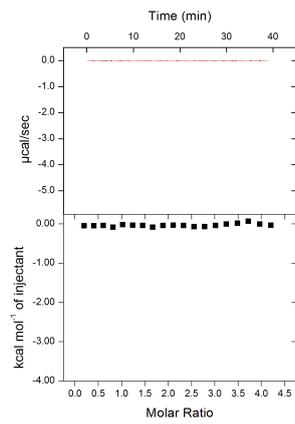
A-1 antigen hexasaccharide

No binding detected



B-1 antigen hexasaccharide

No binding detected



```

1      . . . . . GYQIGEEAQMVTNTKGIQDLSDRYESLNNLLNRYSTLNTLI
51     KLSADPSAINAVRENLGASAKNLIGDKANSPAYQAVLLAINAAVGFWNV
101    GYVTQCGGNANGQKSISSTIFNNEPGYRSTSIITCSLNHSPGYGPMSE
151    ENFKKLNEAYQILQTALKRGLPALKENNGKVNVTYTYTCSGDGNNCSSQ
201    VTGVNNQKDGTKTKIQTIDGKSVTTTISSEKVVDSRADGNTTGVSYTEITN
251    KLEGVPDSAQALLAQASTLINTINNACPYFHASNSSEANAPKFSTTTGKI
301    CGAFSEEISAIQKMITDAQELVNQTSVINEHEQTPVGNNGKPFNPFTD
351    ASFAQMLANASQAQKMLNLAEQVGQAINPERLSGTFQNFVKGFLATCEN
401    PSTAGTGGTQGSAPGTVTTQTFASGCAYVGQTTITNLKNSIAHFGTQEQQI
451    QQAENIADTLVNFKSRYSSELGNTYNSIT'TALSNI'PNAQSLQNAVSKKNNP
501    YSPQGIDTNYLQNSYNQIQ'TINQEL'KKKKKKGSEQKLI'SEEDLSHHHH
551    HH

```

Appendix Figure 1: Amino acid sequence of the visible residues in the apo-BabA_{527K} atomic model

Amino acid sequence of BabA_{527K}. Visible residues in the apo-BabA_{527K} atomic model are in boldface while amino acids not modelled are in red. Red dots at the protein N-terminus represent the first nine amino acids of BabA_{527K}, which are cleaved during periplasmic expression.

```

1      . . . . . GYQIGEEAQMVTNTKGIQDLSDRYESLNNLLNRYSTLNTLI
51     KLSADPSAINAVRENLGASAKNLIGDKANSPAYQAVLLAINAAVGFWNV
101    GYVTQCGGNANGQKSISSTIFNNEPGYRSTSIITCSLNHSPGYGPMSE
151    ENFKKLNEAYQILQTALKRGLPALKENNGKVNVTYTYTCSGDGNNCSSQ
201    VTGVNNQKDGTKTKIQTIDGKSVTTTISSEKVVDSRADGNTTGVSYTEITN
251    KLEGVPDSAQALLAQASTLINTINNACPYFHASNSSEANAPKFSTTTGKI
301    CGAFSEEISAIQKMITDAQELVNQTSVINEHEQTPVGNNGKPFNPFTD
351    ASFAQMLANASQAQKMLNLAEQVGQAINPERLSGTFQNFVKGFLATCEN
401    PSTAGTGGTQGSAPGTVTTQTFASGCAYVGQTTITNLKNSIAHFGTQEQQI
451    QQAENIADTLVNFKSRYSSELGNTYNSIT'TALSNI'PNAQSLQNAVSKKNNP
501    YSPQGIDTNYLQNSYNQIQ'TINQEL'KKKKKKGSEQKLI'SEEDLSHHHH
551    HH

```

Appendix Figure 2: Amino acid sequence of the visible residues in the BabA_{527K}:Le^b atomic model

Amino acid sequence of BabA_{527K}. Visible residues in the BabA_{527K} atomic model are in boldface while amino acids not modelled are in red. Red dots at the protein N-terminus represent the first nine amino acids of BabA_{527K}, which are cleaved during periplasmic expression.