

University of Nottingham
UK | CHINA | MALAYSIA

Cloning, Expression, Purification and Characterisation of the *Helicobacter pylori* Outer Membrane Proteins BabC and HomB

Mohammed Abdulaziz M Alissa

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

November 2021

“To know, is to know that you know nothing. That is the meaning of true knowledge”

- Socrates

Abstract

Helicobacter pylori (*H. pylori*) is a pathogenic bacterium that colonises the gastric mucosa of more than 50% of all humans and a leading cause of peptic ulceration and gastric cancer worldwide. Despite the harsh conditions in the stomach, the bacterial adherence to the gastric mucosa and epithelium plays a significant role in initial colonisation and in long-term persistence by using multitude of adhesins that facilitate the adhesion to the host cells. Studies are greatly needed to understand the dynamic interplay between these adhesins, some of which have not been fully characterised. This study is aimed at characterising the functional properties of the *H. pylori* outer membrane proteins BabC and HomB, which have not been studied well previously with unknown structure and function. Recombinant forms of BabC and HomB were expressed in the periplasm and cytoplasm of *Escherichia coli*, respectively. Carbohydrate microarray was used to identify host binding receptors for these putative adhesins, but no significant binding was detected. Structural analysis using Phyre^2 suggested that HomB could be a metzincin, and indeed we were able to demonstrate proteolytic activity of recombinant HomB. However, we were unable to corroborate metzincin activity using divalent metal chelating agents such as EDTA and 1,10-phenanthroline, possibly due to the much higher binding affinity of the enzyme for zinc than the inhibitors. The recombinant HomB also induced IL-8 production from the gastric epithelial cells, which has not been previously reported. This indicates that HomB has a potential role in the pathogenesis of *H. pylori* and seems to be a marker associated with peptic ulcer and gastric cancer. The study presented in this thesis lays a strong foundation for future work and guides future design of specific attachment and/or inflammation inhibitors as a novel therapeutic approach, to limit the increasing resistance of this pathogen to conventional antibiotic treatment.

Declaration

I declare that all the work presented in this dissertation is my own, except where otherwise stated.



Mohammed Abdulaziz M Alissa

July 2020

Acknowledgements

In the name of Allah, the Most Gracious and the Most Merciful. All praises to Allah for the strength He gave me and His blessing in completing this thesis.

During my studies, there are many people whom I owe gratitude. First of all, I would like to thank my supervisors: Prof. Franco H Falcone and Dr. Sebastiaan Winkler, for their direction, help, guidance and support. Although his move to another academic institution, I am especially grateful to Franco, the principal investigator of this project, who was following my progress closely and at the same time taught me how to be independent.

I would like to express my deep and sincere gratitude to my mother, Hayat and the soul of my father, Abdulaziz for my upbringing, for the endless opportunities they provided me with during my education, for their support and love over the years, for the values they taught me and have been the most important lesson I have ever taken in my life, for their encouragement and for everything really, because if it was not for them, I would not have been able to imagine the day when I submitted my doctoral thesis and made the biggest dream of my life come true. I am very thankful to my brothers and sisters who offered invaluable support and humour over the years.

I would also like to thank my other half and soulmate, my Reema, without whom I would not have been able to make the most of my PhD experience. Reema was always there cheering me up and stood by me throughout the good and bad times.

I am very thankful to those people who helped me: Dr. Mohammad Alfuhaily, Dr. Waleed Alharbi, Marina Khalli, Dr. Aymen Mohammed, Dr. Vasiliki Paraskevopoulou, Dr. Rob Delahay, Dr. Fahad Almughem, Dr. Abdulaziz Alouffi, Awaji Safhi, Dr. Waleed Rizg, Rayan Mushtaq, Tahani Alharbi, and all members of D19 room.

Last but not least, I acknowledge the University of Nottingham for providing me with a chance, reading resources and laboratory resources for my study and my project. I would also like to express my deepest appreciation to my sponsor 'Prince Sattam bin Abdulaziz University' for financial support and giving me an opportunity to study abroad and supporting me through my studies. I wish to express my sincere thanks to people from Prince Sattam bin Abdulaziz University who have supported me Prof Abdulaziz Alhamid (the previous Rector of Prince Sattam bin Abdulaziz University), Prof Maher Alenazi (the Head of College of Applied Medical Sciences), and Prof Ashraf T Abulhamd (the previous Head of the Department of Medical Laboratory Sciences). I would also like to thank the Saudi Arabian Cultural Bureau in London for their support and help during my studies.

List of abbreviations

Abbreviation	Expanded name
3D	Three-dimensional
AAS	Atomic absorption spectroscopy
AC	Affinity chromatography
ADAMs	A Disintegrin and Metalloproteases
AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride
AGS	Human gastric adenocarcinoma
AlpA	Adherence-associated lipoproteins A
AlpB	Adherence-associated lipoproteins B
AprA	*Pseudomonas aeruginosa* alkaline protease
ArsR	Acid-responsive signalling response regulator
ArsS	Acid-responsive signalling sensor
BabA	Blood group antigen-binding adhesin
BSA	Bovine serum albumin
$CaCl_2$	Calcium chloride
cag PAI	*cag* pathogenicity island
CagA	Cytotoxin-associated gene A
CEACAMs	Carcinoembryonic antigen-related cell adhesion molecules
CFG	Consortium for Functional Glycomics
CFU	Colony forming units
Co^{2+}	Cobalt
cspA	Cold-shock Protein A
DMSO	Dimethyl sulfoxide
DPBS	Dulbecco's phosphate buffered saline
DPP	Dipeptidyl peptidase
DTT	Dithiothreitol
E. coli	*Escherichia coli*

EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ERKs	Extracellular signal-regulated kinases
EtBr	Ethidium bromide
FBS	Fetal bovine serum
FITC	Fluorescein isothiocyanate
FXa	Factor Xa
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
gor	Glutathione reductase
GST	Glutathione S-transferase
H. pylori	*Helicobacter pylori*
H_2O_2	Hydrogen peroxide
HBS	HEPES-buffered saline
HomB	*Helicobacter* outer membrane B
Hop	*Helicobacter* outer membrane porins
HopQ	*Helicobacter* outer membrane protein Q
HopZ	*Helicobacter* outer membrane porin Z
Hor	Hop-related proteins
HRP	Horseradish peroxidase
HSA	Human serum albumin
Hte	High transformation efficiency
IARC	International Agency for Research on Cancer
IL-6	Interleukin-6
IL-8	Interleukin-8
IMAC	Immobilised metal ion affinity chromatography
IPA	Isopropyl alcohol
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
KC	Keratinocyte chemoattractant
K_D	Dissociation constant
LabA	LacdiNAc-specific binding adhesin

lacdiNAc	*N,N'*-diacetyllactosediamine or GalNAcβ1-4GlcNAc
LB	Lysogeny broth
LB_{ACG}	Ampicillin-, chloramphenicol-, and glucose-supplemented LB
LB_{amp}	Ampicillin-supplemented LB
LB_{AT}	Ampicillin- and tetracycline-supplemented LB
Le^b	Lewis b
LPS	Lipopolysaccharide
MALT	Mucosa-associated lymphoid tissue
MAPK	Mitogen-activated protein kinase
MBP	Maltose-binding protein
MCS	Multiple cloning site
MMPs	Matrix metalloproteinases
Mn^{2+}	Manganese
NCBI	National Centre for Biotechnology Information
NF	National Formulary
NF-kB	Nuclear factor-kB
Ni^{2+}	Nickel
NikR	Nickel-responsive regulator
NRSB	Non-reducing sample buffer
OD_{600}	Optical density at 600 nm
OipA	Outer inflammatory protein A
OMPs	Outer membrane proteins
PARs	Protease-activated receptors
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline supplemented with 0.05% Tween 20
PCR	Polymerase chain reaction
R10	Deca-arginine
RIA	Radioimmunoassay
RSB	Reducing sample buffer

RT-PCR	Reverse transcription polymerase chain reaction
SabA	Sialic acid-binding adhesin
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SEC	Size exclusion chromatography
SGF	Simulated gastric fluid
sLe^a	Sialyl-Lewis a
sLe^x	Sialyl-Lewis x
sLn	Sialyl-lactosamine
SNP	Single-nucleotide polymorphism
SOC	Super optimal broth with catabolite repression
SS1	Sydney strain 1
T4SS	Type IV secretion system
TBE	Tris/Borate/EDTA
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline containing 0.1% Tween 20
TCA	Trichloroacetic acid
TEE	Translation-enhancing element
TNF-α	Tumour necrosis factor alpha
tRNAs	Transfer ribonucleic acids
trp	Tryptophan promoter
TrxA	Thioredoxin A
trxB	Thioredoxin reductase
USP	US Pharmacopoeia
UV	Ultraviolet
VacA	Vacuolating cytotoxin
VacA-s1	Type s1 of VacA
WHO	World Health Organisation
WT	Wild type
Zn^{2+}	Zinc
$ZnCl_2$	Zinc chloride

β3GnT5	β1,3-*N*-acetylglucosaminyltransferase 5

List of tables and figures

List of tables

Table 1.1: Adhesins known to mediate attachment of *H. pylori* to the gastric epithelial surface 12
Table 3.1: Expression construct engineered in this study 69
Table 3.2: Primers for *babC* amplification 74
Table 3.3: Primers for amplifying sequences of the non-coding 5'- and 3'-regions (before and after) of *babC* gene 75
Table 3.4: Primers for amplification of the second *babC* segment 77
Table 3.5: Primers for the overlap extension PCR 77
Table 3.6: Primers for colony PCR and confirmatory sequencing 79
Table 3.7: Expression construct and expression host used in this study 81
Table 4.1: Expression constructs engineered in this study 129
Table 4.2: Primers for $homB_{9\text{-}511}$ amplification 132
Table 4.3: Primers used in the overlap extension PCR 135
Table 4.4: Primers for amplification of *homB* gene to be inserted into various vectors 136
Table 4.5: Primers for Colony PCR and Confirmatory Sequencing 140
Table 4.6: Expression constructs and expression host used in this study 141
Table 4.7: The antibodies used for immunodetection by Western blotting 151
Table 4.8: Tandem mass spectra data of periplasmic oligopeptide-binding protein from *E. coli* strain K12 169
Table 4.9: Optimisation of $HomB_{19\text{-}511}$ expression on a small-scale using MBP fusion protein and *E. coli* Rosetta-gami B (DE3) pLysS 183
Table 4.10: Tandem mass spectra data of MBP-$HomB_{19\text{-}511}$ 188
Table 4.11: Tandem mass spectra data of $HomB_{19\text{-}511}$ 193
Table 4.12: A brief summary showing the results of each combination of construct and cell type tested 194

Table 5.1: Protease inhibitors and group-specific reagent used in the inhibition studies 222
Table 5.2: HomB mutants and 5' phosphorylated primers used for site-directed mutagenesis..... 225
Table 5.3: 5'-phosphorylated primers used in site-directed mutagenesis to insert c-Myc and 6x His tags at C- and N-terminus of HomB 230
Table 5.4: Primers used to amplify IL-8, PAR2 and GAPDH genes..... 239
Table 5.5: Tandem mass spectra data of the two bands generated after cleavage of HomB mutant (E359A-M372A) from the MBP fusion partner 259

List of figures

Figure 1.1: Illustration of persistence strategies of *H. pylori* 7
Figure 1.2: Schematic of pathogenesis mediated by OMPs of *H. pylori* ... 16
Figure 1.3: Crystal structure of the extracellular region of BabA J99 18
Figure 1.4: The biosynthesis pathways of the dominant ABO/Le blood group antigens in healthy gastric mucosa 21
Figure 1.5: Crystal structure of the extracellular region of LabA J99..... 28
Figure 1.6: Diversity of *homA* and *homB* in the copy number and their genomic localisation in Western and East Asian *H. pylori* strains 35
Figure 2.1: Schematic description of the three stages of PCR..... 48
Figure 2.2: Suggested process of formation of inclusion bodies 52
Figure 2.3: Schematic diagram of pOPE101 expression vector..... 53
Figure 2.4: Schematic diagram of pET-22b(+) expression vector 54
Figure 2.5: Schematic diagram of pET22b(ΔpelB) expression cassette ... 55
Figure 2.6: Schematic diagram of the pCold-GST expression vector 56
Figure 2.7: Schematic diagram of pMAL-c5X expression vector..... 57
Figure 2.8: Schematic depiction of the ELISA principle (sandwich-type) used in the current study 62
Figure 3.1: Schematic representation of the adhesin BabA J99 65
Figure 3.2: Amino acid sequence of BabC 26695 65
Figure 3.3: Schematic representation of the gastric lumen and mucus layer 68
Figure 3.4: Schematic diagram of asymmetric PCR..... 71
Figure 3.5: Alignment of amino acid sequences of BabA J99 and BabC 26695..... 93
Figure 3.6: First round of asymmetric PCR 94
Figure 3.7: Second round of asymmetric PCR 96
Figure 3.8: Overlap extension PCR..... 98
Figure 3.9: Schematic diagram of overlap extension PCR..... 99
Figure 3.10: *babC* gene and pOPE101 after digestion by NcoI and BamHI 100

Figure 3.11: Schematic diagram of the pOPE101_*babC* expression cassette 103
Figure 3.12: Expression of recombinant BabC in *E. coli* XL10 Gold cells 103
Figure 3.13: IMAC purification of BabC 105
Figure 3.14: Purification of BabC by IMAC 105
Figure 3.15: c-Myc immunodetection of BabC after purification by IMAC 106
Figure 3.16: Binding of BabC to Le^a, Le^b, Le^y, and sLe^x in an ELISA 108
Figure 3.17: Comparison between the crowns of BabA J99 and BabC 26695 109
Figure 3.18: Predicted structure of BabC and crystal structure of β-lactoglobulin showing potential pepsin cleavage sites 111
Figure 3.19: Peptic digestion pattern of BabC and β-lactoglobulin 112
Figure 3.20: c-Myc immunodetection of BabC after peptic digestion 113
Figure 4.1: Amino acid sequence of HomB J99 128
Figure 4.2: Signal peptide prediction using SignalP 4.0 and 4.1 servers 131
Figure 4.3: Signal peptide prediction using SignalP 3.0 and 5.0 servers 131
Figure 4.4: Schematic diagram of overlap extension PCR 134
Figure 4.5: Overlap extension PCR for amplification of $homB_{9\text{-}511}$ gene . 160
Figure 4.6: $homB_{9\text{-}511}$ gene and pOPE101 after digestion by NcoI and NotI 161
Figure 4.7: Schematic diagram of the pOPE101_ $homB_{9\text{-}511}$ expression cassette 163
Figure 4.8: Expression of recombinant $HomB_{9\text{-}511}$ in *E. coli* XL10 Gold cells using pOPE101 vector at 24°C 163
Figure 4.9: Schematic diagram of the pET22b_$homB_{9\text{-}511}$ expression cassette 164
Figure 4.10: Expression of recombinant $HomB_{9\text{-}511}$ in *E. coli* BL21 cells using pET-22b(+) vector 165
Figure 4.11: Expression of recombinant $HomB_{9\text{-}511}$ in *E. coli* Origami B (DE3) cells using pET-22b(+) vector 166

Figure 4.12: Expression of recombinant HomB$_{9-511}$ in *E. coli* XL10 Gold cells using pOPE101 vector at 16°C 168
Figure 4.13: Schematic diagram of the pOPE101_*homB$_{19-511}$* expression cassette 171
Figure 4.14: Expression of recombinant HomB$_{19-511}$ in *E. coli* XL10 Gold cells using pOPE101 vector..... 172
Figure 4.15: Schematic diagram of the pOPE101_*homB$_{19-511R}$* expression cassette 173
Figure 4.16: Expression of recombinant HomB$_{19-511R}$ in *E. coli* XL10 Gold cells using pOPE101 vector 174
Figure 4.17: Schematic diagram of the pET22b(ΔpelB)_*homB$_{19-511}$* expression cassette..... 175
Figure 4.18: Expression of recombinant HomB$_{19-511}$ in *E. coli* Rosetta-gami B (DE3) pLysS cells using pET22b(ΔpelB) vector..... 176
Figure 4.19: Schematic diagram of the pCold-GST_*homB$_{19-511}$* expression cassette 177
Figure 4.20: Expression of recombinant HomB$_{19-511}$ in *E. coli* BL21 (DE3) cells using pCold-GST vector 178
Figure 4.21: Expression of recombinant HomB$_{19-511}$ in *E. coli* ArcticExpress (DE3) cells using pCold-GST vector..... 179
Figure 4.22: Schematic diagram of the pMAL-c5X_*homB$_{19-511}$* expression cassette 181
Figure 4.23: Expression of recombinant HomB$_{19-511}$ in *E. coli* Rosetta-gami B (DE3) pLysS cells using pMAL-c5X vector 182
Figure 4.24: AC purification of MBP-HomB$_{19-511}$ 185
Figure 4.25: Purification of MBP-HomB$_{19-511}$ by AC 186
Figure 4.26: An InstantBlue-staining gel of MBP-HomB$_{19-511}$ for Mass spectrometry following AC..... 187
Figure 4.27: Cleavage of HomB$_{19-511}$ from MBP using FXa protease 189
Figure 4.28: SEC purification of HomB$_{19-511}$ and MBP fusion protein 191
Figure 4.29: Purification of HomB$_{19-511}$ by SEC..... 191

Figure 4.30: An InstantBlue-staining gel of $HomB_{19-511}$ for Mass spectrometry and Immunodetection of MBP following SEC 192
Figure 5.1: The location of the conserved methionine “Met-tern” underneath the catalytic Zn ion..... 211
Figure 5.2: Classification of mononuclear metalloproteases 212
Figure 5.3: Discovery of Zn-binding motif in the predicted 3D structure of HomB..... 213
Figure 5.4: IL-8 production mediated via activation of PAR2 by HomB... 217
Figure 5.5: Principle of the proteolytic digestion of the FITC-labelled casein 220
Figure 5.6: Schematic diagram of the pMAL-c5X_*homB*$_{19-511}$ expression cassette with tag insertion 228
Figure 5.7: Amino acid and DNA sequences of HomB J99..... 242
Figure 5.8: Zn-binding motif and the conserved methionine in HomB sequence of some *H. pylori* strains 243
Figure 5.9: The pH effect on the activity of recombinant HomB 245
Figure 5.10: Proteolytic activity of the recombinant MBP-HomB..... 246
Figure 5.11: Effects of divalent metal cations Zn^{2+} and Ca^{2} on the activity of recombinant HomB..... 247
Figure 5.12: Effects of protease inhibitors and group-specific reagent on the activity of HomB..... 249
Figure 5.13: DNA sequencing confirming the desired mutation 251
Figure 5.14: AC purification of MBP-HomB mutants 252
Figure 5.15: Purification of MBP-HomB mutants by AC 253
Figure 5.16: SEC purifications of HomB mutants and MBP fusion protein 255
Figure 5.17: Purification of HomB mutants by SEC 256
Figure 5.18: An InstantBlue-staining gel of HomB mutant (E359A-M372A) for Mass spectrometry following FXa protease cleavage 258
Figure 5.19: The effects of Zn-binding motif, asparagine and conserved methionine mutations on the proteolytic activity of HomB. 261

Figure 5.20: Expression and purification of MBP-HomB containing c-Myc or 6x His tag at the C-terminus of HomB 264
Figure 5.21: Expression and purification of MBP-HomB containing c-Myc or 6x His tag at the N-terminus of HomB 265
Figure 5.22: SEC purification of HomB containing N-terminal c-Myc tag and MBP fusion protein 266
Figure 5.23: Purification of HomB containing c-Myc tag at the N-terminus 267
Figure 5.24: Histogram charts of carbohydrate microarray results..... 269
Figure 5.25: Predicted structure of HomB and crystal structure of β-lactoglobulin showing potential pepsin cleavage sites 271
Figure 5.26: Peptic digestion pattern of HomB and β-lactoglobulin 272
Figure 5.27: RT-PCR for expression detection of IL-8 and PAR2 mRNAs 273
Figure 5.28: IL-8 production from AGS cells treated with *H. pylori* HomB 275
Figure 5.29: The effect of excess amounts of Zn on the proteolytic activity of metalloproteases 279

List of appendix figures

Appendix Figure 1: Amino acid sequence alignment of HomA, HomB, HomC, and HomD 348
Appendix Figure 2: SnapGene map of pOPE101 349
Appendix Figure 3: SnapGene map of pET-22b(+) vector 350
Appendix Figure 4: SnapGene map of pCold-GST vector 351
Appendix Figure 5: SnapGene map of pMAL-c5X vector 352
Appendix Figure 6: SnapGene map of pOPE101_*babC* construct 353
Appendix Figure 7: DNA alignment of the N-terminal extracellular domain of *babA*, *babB*, and *babC* genes from H. pylori 26695 354
Appendix Figure 8: DNA Alignment of *homA* and *homB* genes from H. pylori J99 355
Appendix Figure 9: SnapGene map of pOPE101_$homB_{9\text{-}511}$ construct ... 356
Appendix Figure 10: SnapGene map of pET22b_$homB_{9\text{-}511}$ construct 357
Appendix Figure 11: SnapGene map of pOPE101_$homB_{19\text{-}511}$ construct 358
Appendix Figure 12: SnapGene map of pET22b(ΔpelB)_$homB_{19\text{-}511}$ construct 359
Appendix Figure 13: SnapGene map of pCold-GST_$homB_{19\text{-}511}$ construct 360
Appendix Figure 14: SnapGene map of pMAL-c5X_$homB_{19\text{-}511}$ construct 361

Table of contents

Abstract ..... **ii**
Declaration ..... **iii**
Acknowledgements ..... **iv**
List of abbreviations ..... **vi**
List of tables and figures ..... **xi**
List of tables xi
List of figures xiii
List of appendix figures xviii
Chapter 1: Introduction **1**
1.1 Summary 1
1.2 Discovery of *H. pylori* 2
1.3 *H. pylori* infection and chronic colonisation 5
1.4 Genetic diversity of *H. pylori* 8
1.5 *H. pylori* virulence factors 10
1.5.1 Adhesins 10
1.5.1.1 BabA, BabB, and BabC 14
1.5.1.2 SabA 22
1.5.1.3 HopQ 24
1.5.1.4 AlpA and AlpB 25
1.5.1.5 LabA 26
1.5.1.6 *Helicobacter* outer membrane B (HomB) 29
1.5.2 CagA 36
1.5.3 VacA 37
1.6 The potential clinical implications of this study 38
1.7 Study aims and objectives 41
Chapter 2: Overview of experimental methodology **43**
2.1 PCR 45
2.2 Recombinant protein expression in *E. coli* 49
2.2.1 Expression plasmids used in this research 53
2.2.1.1 Plasmids used for periplasmic expression 53
2.2.1.1.1 pOPE101 53
2.2.1.1.2 pET-22b(+) 54

2.2.1.2 Plasmids used for cytoplasmic expression 54
2.2.1.2.1 pET22b(ΔpelB) 54
2.2.1.2.2 pCold-GST 55
2.2.1.2.3 pMAL-c5X 56
2.2.2 *E. coli* strains used in this study 57
2.2.2.1 *E. coli* DH5α 57
2.2.2.2 *E. coli* XL10 Gold 58
2.2.2.3 *E. coli* BL21 (DE3) 58
2.2.2.4 *E. coli* Origami B (DE3) 58
2.2.2.5 *E. coli* ArcticExpress (DE3) 59
2.2.2.6 Rosetta-gami B (DE3) pLysS 59
2.3 ELISA 60
Chapter 3: Expression, purification and characterisation of BabC 63
3.1 Introduction 63
3.2 Materials and methods 69
3.2.1 Generation of BabC expression construct 69
3.2.1.1 Design of BabC expression construct 69
3.2.1.2 PCR amplification 70
3.2.1.2.1 Asymmetric PCR 71
3.2.1.2.1.1 First round of PCR amplification 72
3.2.1.2.1.2 Second round of PCR amplification 73
3.2.1.2.2 Amplification of sequences before and after *babC* gene 74
3.2.1.2.3 Overlap extension PCR 75
3.2.1.3 Cloning 77
3.2.2 Periplasmic expression in *E. coli* XL10 Gold 80
3.2.2.1 Preparation of cellular fractions for analyses 81
3.2.2.1.1 Total protein fraction 81
3.2.2.1.2 Periplasmic and spheroplast fractions 81
3.2.2.2 Analysis of cellular fractions 82
3.2.2.2.1 SDS-PAGE 82
3.2.2.2.2 Coomassie-staining 83
3.2.2.2.3 Western blot 83
3.2.3 Protein purification 85
3.2.3.1 Larger scale expression and periplasmic harvest 85

3.2.3.2 Immobilised metal ion affinity chromatography (IMAC) 86
3.2.3.3 Quantification of the recombinant BabC 86
3.2.4 ELISA for binding activity of BabC 88
3.2.5 Peptic digestion test 89
3.2.5.1 Prediction of pepsin cleavage sites 89
3.2.5.2 Pepsin activity 89
3.2.5.3 Peptic digestion 90
3.2.5.4 Sample analysis 91
3.3 Results 92
3.3.1 Cloning a *babC* gene fragment into expression vector 92
3.3.1.1 Amplification of *babC* gene 92
3.3.2 Periplasmic expression in *E. coli* XL10 Gold 102
3.3.3 Purification of recombinant BabC 104
3.3.4 Binding Activity of BabC 107
3.3.5 Resistance of BabC to peptic degradation 110
3.4 Discussion 114
Chapter 4: HomB recombinant expression and purification 127
4.1 Introduction 127
4.2 Materials and methods 129
4.2.1 Generation of HomB expression constructs 129
4.2.2 PCR amplification 130
4.2.2.1 Prediction of signal peptide 130
4.2.2.2 Ordinary PCR 132
4.2.2.3 Overlap extension PCR 133
4.2.2.4 Amplification of the other inserted genes 135
4.2.3 Cloning 138
4.2.4 Expression screening 141
4.2.4.1 Periplasmic expression 142
4.2.4.1.1 Periplasmic expression of pOPE101_*homB*$_{9\text{-}511}$ construct in *E. coli* XL10 Gold at 24°C 142
4.2.4.1.2 Periplasmic expression of pET22b_*homB*$_{9\text{-}511}$ construct in *E. coli* BL21 (DE3) 142
4.2.4.1.3 Periplasmic expression of pET22b_*homB*$_{9\text{-}511}$ construct in *E. coli* Origami B (DE3) 143

4.2.4.1.4 Periplasmic expression of pOPE101_*homB*$_{9-511}$ construct in *E. coli* XL10 Gold at 16°C 144
4.2.4.1.5 Periplasmic expression of pOPE101_*homB*$_{19-511}$ construct in *E. coli* XL10 Gold 144
4.2.4.1.6 Periplasmic expression of pOPE101_*homB*$_{19-511R}$ construct in *E. coli* XL10 Gold 144
4.2.4.2 Cytoplasmic expression 144
4.2.4.2.1 Cytoplasmic expression of pET22b(ΔpelB)_*homB*$_{19-511}$ construct in *E. coli* Rosetta-gami B (DE3) pLysS 144
4.2.4.2.2 Cytoplasmic expression of pCold-GST_*homB*$_{19-511}$ construct in *E. coli* BL21 (DE3) 146
4.2.4.2.3 Cytoplasmic expression of pCold-GST_*homB*$_{19-511}$ construct in *E. coli* ArcticExpress (DE3) 147
4.2.4.2.4 Cytoplasmic expression of pMAL-c5X_*homB*$_{19-511}$ construct in *E. coli* Rosetta-gami B (DE3) pLysS 147
4.2.4.3 Preparation of cellular fractions for analysis 147
4.2.4.3.1 Total protein fraction 147
4.2.4.3.2 Periplasmic fractions 148
4.2.4.3.3 Cytoplasmic soluble and insoluble protein fractions .. 148
4.2.4.3.3.1 Soluble and insoluble protein fractions in periplasmic expression 148
4.2.4.3.3.2 Soluble and insoluble protein fractions in cytoplasmic expression 149
4.2.4.4 Analysis of cellular fractions 150
4.2.4.4.1 SDS-PAGE 150
4.2.4.4.2 Coomassie-staining 150
4.2.4.4.3 Western blot 150
4.2.5 Protein purification 152
4.2.5.1 Larger scale expression and cytoplasmic harvest 152
4.2.5.2 Affinity chromatography (AC) 153
4.2.5.3 Quantification of the recombinant MBP-HomB$_{19-511}$ 153
4.2.5.4 Cleavage of HomB$_{19-511}$ from MBP using FXa protease ... 154
4.2.5.5 Purification of recombinant HomB$_{19-511}$ by SEC 154
4.2.5.6 Mass spectrometry 155
4.3 Results 158
4.3.1 Cloning *homB* gene fragments into expression vectors 158
4.3.2 Amplification of *homB*$_{9-511}$ gene by ordinary PCR 158

4.3.3 Amplification of *homB*$_{9-511}$ gene by overlap extension PCR 159
4.3.4 Expression screening 162
4.3.4.1 Periplasmic expression 162
4.3.4.1.1 Periplasmic expression of pOPE101_*homB*$_{9-511}$ construct in *E. coli* XL10 Gold at 24°C 162
4.3.4.1.2 Periplasmic expression of pET22b_*homB*$_{9-511}$ construct in *E. coli* BL21 (DE3) 164
4.3.4.1.3 Periplasmic expression of pET22b_*homB*$_{9-511}$ construct in *E. coli* Origami B (DE3) 165
4.3.4.1.4 Periplasmic expression of pOPE101_*homB*$_{9-511}$ construct in *E. coli* XL10 Gold at 16°C 167
4.3.4.1.5 Periplasmic expression of pOPE101_*homB*$_{19-511}$ construct in *E. coli* XL10 Gold 170
4.3.4.1.6 Periplasmic expression of pOPE101_*homB*$_{19-511R}$ construct in *E. coli* XL10 Gold 173
4.3.4.2 Cytoplasmic expression 174
4.3.4.2.1 Cytoplasmic expression of pET22b(ΔpelB)_*homB*$_{19-511}$ construct in *E. coli* Rosetta-gami B (DE3) pLysS 174
4.3.4.2.2 Cytoplasmic expression of pCold-GST_*homB*$_{19-511}$ construct in *E. coli* BL21 (DE3) 176
4.3.4.2.3 Cytoplasmic expression of pCold-GST_*homB*$_{19-511}$ construct in *E. coli* ArcticExpress (DE3) 178
4.3.4.2.4 Cytoplasmic expression of pMAL-c5X_*homB*$_{19-511}$ construct in *E. coli* Rosetta-gami B (DE3) pLysS 180
4.3.5 Purification of recombinant MBP-HomB$_{19-511}$ 184
4.3.5.1 Purification of recombinant MBP-HomB$_{19-511}$ by AC 184
4.3.5.2 Mass spectrometry for MBP-HomB$_{19-511}$ 187
4.3.5.3 Cleavage of HomB$_{19-511}$ from MBP using FXa protease 189
4.3.5.4 Purification of recombinant HomB$_{19-511}$ by SEC 190
4.3.5.5 Mass spectrometry for HomB$_{19-511}$ 192
4.4 Discussion 195
Chapter 5: Characterisation of HomB 208
5.1 Introduction 208
5.1.1 Metalloproteases 208
5.1.2 Protease-activated receptor 2 and IL-8 production 214
5.2 Materials and methods 219
5.2.1 Structural prediction of HomB by Phyre^2 219

5.2.2 Quantitative protease activity assay 219
5.2.2.1 Effect of pH on HomB activity 219
5.2.2.2 Proteolytic activity of the recombinant MBP-HomB 220
5.2.2.3 Effects of divalent metal cations Zn^{2+} and Ca^{2+} on HomB activity 221
5.2.2.4 Effects of protease inhibitors and group-specific reagent on HomB activity 221
5.2.3 Mutations in the Zn-binding motif in HomB 223
5.2.3.1 Generation of expression constructs of HomB mutants ... 223
5.2.3.1.1 Site-directed mutagenesis 223
5.2.3.1.2 Cloning and transformation 226
5.2.3.2 Protein purification 226
5.2.3.2.1 Larger scale expression and cytoplasmic harvest 226
5.2.3.3 Affinity chromatography 226
5.2.3.4 Quantification of the recombinant MBP-$HomB_{19-511}$ mutants 226
5.2.3.5 Cleavage of $HomB_{19-511}$ mutants from MBP using FXa protease 226
5.2.3.6 Purification of recombinant $HomB_{19-511}$ mutants by SEC .. 227
5.2.3.7 Mass spectrometry 227
5.2.3.8 Testing of HomB mutants for quantitative protease activity assay 227
5.2.4 Binding assay of HomB 228
5.2.4.1 Generation of expression constructs of HomB with a tag . 228
5.2.4.2 Site-directed mutagenesis 229
5.2.4.3 Cloning and transformation 229
5.2.4.4 Protein purification 231
5.2.4.4.1 Larger scale expression, cytoplasmic harvest, and purification by AC 231
5.2.4.4.2 Quantification of the recombinant proteins 231
5.2.4.5 Cleavage of HomB proteins containing the tags from MBP using FXa protease 231
5.2.4.6 Purification of recombinant $HomB_{19-511}$ containing c-Myc tag at the N-terminus by SEC 231
5.2.4.7 Carbohydrate microarray 232
5.2.5 Peptic digestion test 234

5.2.5.1 Prediction of pepsin cleavage sites 234
5.2.5.2 Pepsin activity 234
5.2.5.3 Peptic digestion 234
5.2.5.4 Sample analysis 235
5.2.6 Cell culture 236
5.2.6.1 Cell defrosting 236
5.2.6.2 Routine passaging 237
5.2.6.3 Cryopreservation 237
5.2.6.4 RT-PCR for detection of IL-8 and PAR2 mRNAs expressed by AGS cells 238
5.2.6.5 ELISA assay for IL-8 production from AGS cells treated with WT HomB, E359A mutant, and E359A-M372A mutant 240
5.2.7 Statistical analysis 240
5.3 Results 241
5.3.1 Discovery of Zn-binding motif 241
5.3.2 Quantitative protease activity assay 244
5.3.2.1 Effect of pH on HomB activity 244
5.3.2.2 Proteolytic activity of the recombinant MBP-HomB fusion 245
5.3.2.3 Effects of divalent metal cations Zn^{2+} and Ca^{2+} on HomB activity 246
5.3.2.4 Effects of protease inhibitors and group-specific reagent on HomB activity 248
5.3.3 Mutations in the Zn-binding motif in HomB 250
5.3.3.1 Successful cloning of HomB mutants 250
5.3.3.2 Expression and purification of MBP-HomB mutants 250
5.3.3.3 Purification of HomB mutants by SEC 254
5.3.3.4 Mass spectrometry 257
5.3.3.5 Testing of HomB mutants by quantitative protease activity assay 260
5.3.4 Binding assay of HomB 262
5.3.4.1 Successful insertion of c-Myc or 6x His either at the C- or N-terminus of HomB in the form of MBP-HomB 262
5.3.4.2 Expression and purification of MBP-HomB containing c-Myc or 6x His tag at the C-terminus 262
5.3.4.3 Expression and purification of MBP-HomB containing c-Myc or 6x His tag at the N-terminus of HomB 263

5.3.4.4 Purification of HomB containing N-terminal c-Myc tag 266
5.3.4.5 Carbohydrate microarray 268
5.3.5 Resistance of HomB to peptic degradation 270
5.3.6 Cell culture 273
5.3.6.1 RT-PCR for detection of IL-8 and PAR2 mRNAs expressed by AGS cells 273
5.3.6.2 IL-8 production from AGS cells treated with WT HomB, E359A mutant, and E359A-M372A mutant 274
5.4 Discussion 276
Chapter 6: General discussion and future works 293
6.1 The significance of the research and future work 293
6.1.1 A strategy for periplasmic expression for *H. pylori* OMPs 294
6.1.2 Development of a cytoplasmic expression strategy for *H. pylori* HomB 296
6.1.3 Binding activity of BabC and HomB 297
6.1.4 HomB is a protease 298
6.1.5 Recombinant HomB induces IL-8 production 299
6.1.6 Resistance of BabC and HomB to peptic degradation 301
6.2 Future work of BabC 304
6.3 New alternative hypothesis 306
References 307
Appendices 348

Chapter 1: Introduction

1.1 Summary

Helicobacter pylori (*H. pylori*) is the main infectious causative agent of gastric cancer and the first bacterium listed as a carcinogen. It is widespread, infecting about more than half of the world's population. One of the secrets for its success in establishing infection in an environment as hostile as the stomach is its ability to adhere to the surface of the gastric epithelium, where the pH is close to neutral, hiding underneath a protective mucus layer. This attachment is not only crucial to survival of *H. pylori*, it is also a good target for pharmacological treatment.

However due to the multitude of adhesins used by *H. pylori* for attachment (e.g. BabA, SabA, LabA, etc.) more work is needed to understand the dynamic interplay between these adhesins, some of which such as BabC and HomB have not been fully characterised, and the epithelium/gastric mucosa. Similarly to the recent elucidation of the atomic structure of BabA, we intended to systematically elucidate the function of other uncharacterised adhesins. In the present study, the *H. pylori* outer membrane proteins (OMPs) BabC and HomB were investigated through expressing them recombinantly and characterising their function in order to understand their roles in bacterial adhesion and gastric inflammation. This will enable the future design of specific attachment inhibitors as a strategy for treatment of *H. pylori* which has particular significance in the context of increasing resistance of this pathogen to conventional antibiotic treatment.

1.2 Discovery of *H. pylori*

Microorganisms have been seen in the stomachs of humans and animals more than a century ago. In 1881, spiral-shaped bacteria were observed in a dog's stomach by Rappin (1). This discovery was confirmed in 1892 by Bizzozero, who noted some spirilla present inside epithelial cells during studying the structure of gastric epithelium of a dog. This notice was considered as the first observation of microorganisms being associated with distinct cells of the stomach (1-3). However, the first extensive study of these bacteria living the gastric mucosa was undertaken by Salomon in 1896. He observed these bacteria in a wide variety of animals including dogs, cats, and brown Norway rats, but failed to find them in humans, monkeys, and pigs. He also failed to grow these bacteria on artificial media, but succeeded to proliferate them through feeding gastric mucus from infected animals to uninfected white mice (1,4,5).

Over the next few years, other researchers still observed these bacteria in many animals, especially Balfour who reported spiral bacteria around ulcers in 1906 (6), until 1909 when Carnot and Lelievre disagreed with the suggestions by Regaud and others stating that these spirilla were parietal cell secretions and not bacteria. However, Regaud's opinion did not last long as he convinced the doubters that the observed spirilla were bacteria and not secretions from parietal cells. Interestingly, this confusion was understandable, because it is currently known that these bacteria can invade the canaliculi of parietal cells (1). In 1920, Lime revealed that the stomach,

including the pylorus, fundus, and the cardia, was the only organ in which these bacteria were found except a few in the duodenum, close to the pyloric sphincter (7).

In 1920s, Luck discovered the urease enzyme in the gastric mucosa of cats and humans (2,4), as well as others detected vast volume of ammonia in the vomitus of patients with nephritis (8,9). Luck's work was then confirmed by Conway and Fitzgerald, who concluded that this enzyme was secreted endogenously by epithelial cells of the stomach in order to protect gastric mucosa from the gastric acid (10).

In the 1930s, gastric cancer was the top cause of cancer deaths in America but with no correlation with bacterial infection (11). Doenges conducted the first extensive study on humans and the macaque rhesus monkeys in 1939. He found spiral bacteria in 43% of 242 human gastric autopsy examined at post-mortem as well as in 100% of the macaque rhesus monkeys, but without any conclusions due to the significant autolysis (4,12). One year later, Freedberg and Barron conducted a study in order to see whether these bacteria could also be observed during life. They detected these spirochetes in 37% of 35 partial gastrectomy specimens usually associated with benign or malignant ulcers. Freedberg, Barron, and Gorham were the first researchers who used antibiotic therapy (intramuscular bismuth) for peptic ulceration (13). Steer published micrographs by scanning electron microscope for large numbers of S-shaped and kidney-shaped bacteria on the surface of the gastric epithelium and antrum, and in gastric metaplasia

in the duodenal bulb. He observed these bacteria in 73% patients with duodenal ulcer. However, he failed to culture them (14), and these bacteria were considered again to be clinically insignificant as they were on the mucosal lining cells but not intracellular. This was not until 1982, when Warren and Marshall succeeded to culture these bacteria on chocolate agar by using Campylobacter isolation method (15), but without addressing the potential pathogenic role of these bacteria.

In order to know whether the bacterium is the causative agent of gastritis or just colonise the damaged mucosa, Marshall (16), and Morris (17) decided to drink a solution containing 10^9 colony forming units (CFU) of the bacterium with antacid cimetidine. Both had signs of an acute gastritis, and with the use of endoscopy, they confirmed that this bacterium indeed was the cause of disease (18). As a result of this, the Nobel Prize in Physiology or Medicine was awarded to Barry Marshall and Robin Warren in 2005 for the discovery of *H. pylori* and its role in peptic ulcer and gastritis (19).

1.3 *H. pylori* infection and chronic colonisation

H. pylori is a motile, spiral-shaped, Gram-negative, microaerophilic bacterium discovered by Warren and Marshall in 1982 (15). The bacterium shows a strict tropism for human and primate gastric mucosa, especially in the less acidic gastric antrum region (20,21). It persistently colonises gastric mucosa of more than a half of all humans (22,23). Although the majority of infected persons is asymptomatic (24), chronic gastritis or peptic ulcer disease can be developed in 10-15% of the infected persons, while gastric cancer and gastric mucosa-associated lymphoid tissue (MALT) lymphoma can be developed in 1-3% of the infected individuals (25). More than 60% of all gastric cancer cases, which is the third leading cause of cancer death and the fifth most common cancer worldwide, has been found to be caused by *H. pylori* (26-29). Therefore, the World Health Organisation (WHO) and the International Agency for Research on Cancer (IARC) have classified this bacterium as a class I carcinogen (30).

Despite the harsh condition that microbes address in the lumen of the stomach, because of its acidic and proteolytic environment that destroy most pathogens ingested with food, *H. pylori* succeeds to establish chronic gastric infection. Furthermore, the bacterium can stay in the stomach for decades or even the lifespan duration of its host if the infection is not treated (31). Colonisation and infection are established via unique evolutionary adaptations, which are thought to have co-evolved with humans over thousands of years, since the bacterial presence can be traced back to

58,000 years ago (32). Once present in the gastric lumen, the bacterium depends on the expression of its urease that hydrolyses urea in gastric fluid into carbon dioxide and ammonia in order to resist the acidity of the stomach, where it can only survive for minutes. Consequently, generation of a higher pH microenvironment occurs and *H. pylori* maintains a buffered surface layer, periplasm and cytosol (33,34). The higher value of pH can also change the viscoelastic features of the mucus layer by decreasing the viscosity (35). With the help of its flagella that enhance motility (36), and its helical shape that provides a corkscrew-like movement (37), the bacterium can swim into the mucus layer of the stomach through the detection of chemotactic factors, including bicarbonate ions and urea (38). Here, *H. pylori* is protected from gastric acidity and proteases, such as pepsin, and is bound to the glycan residues of the mucin glycoproteins as both receptors and nutrients, which enable the proliferation of the bacterium (39). Although most bacterial population diffuse in the mucus layer, around 20% live directly attached to the gastric epithelial surface (40,41).

In order to establish chronic colonisation, *H. pylori* exploits a variety of mechanisms after its survival in the lumen of the stomach and localisation in the gastric mucosa. These include bacterial genome diversification, DNA and protein repair pathways, toxin secretion, proven to contribute to pathogenicity such as cytotoxin-associated gene A (CagA) and vacuolating cytotoxin (VacA), and circumvention of innate immune recognition and

immune response modulation. These mechanisms are shown in Figure 1.1 and have been discussed in detail in (42).

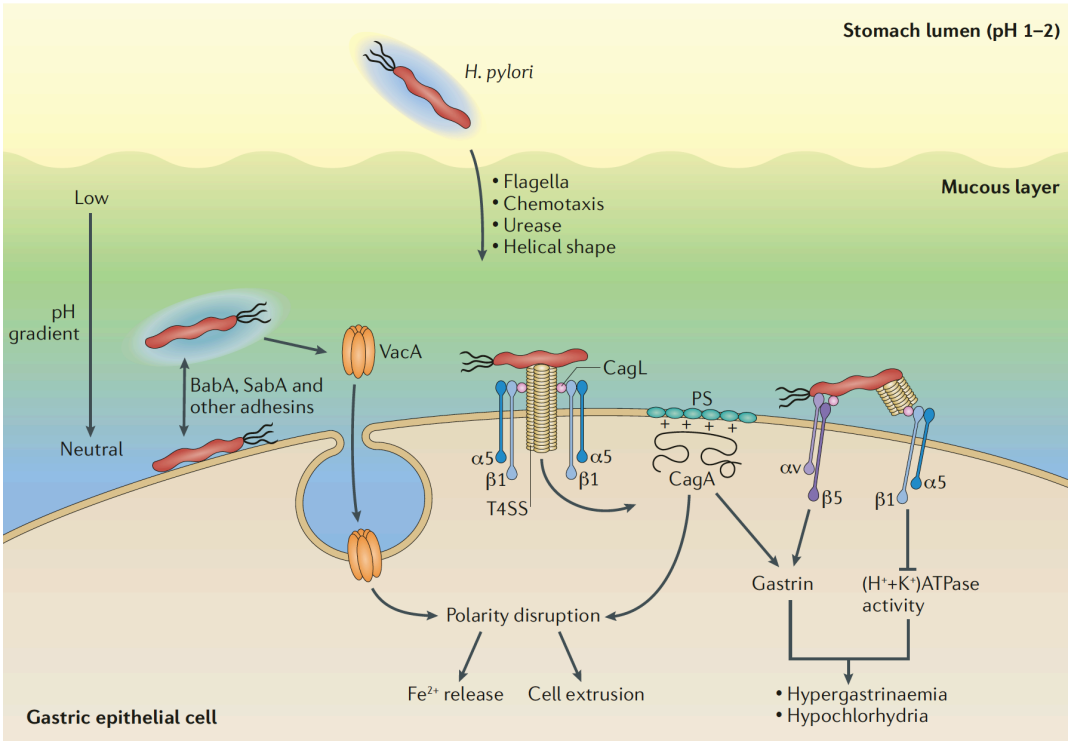


Figure 1.1: Illustration of persistence strategies of *H. pylori*
H. pylori penetration into the mucus layer during initial infection, assisted by the bacterial urease, the enhancement of motility because of the presence of flagella, its helical shape, and chemotactic factors. After initial infection, the bacterium uses other strategies for colonisation establishment, including toxin secretion, the actions of which are shown above. T4SS and PS stand for type IV secretion system and phosphatidylserine, respectively. Adapted from (42).

1.4 Genetic diversity of *H. pylori*

H. pylori's most interesting feature is its diversity. Even inside a single host, the number of different strains is impressive (43). *H. pylori* has around 1,600 genes and it is highly heterogeneous (44), with a high competence of picking up both DNA from other strains and environmental DNA (45). The bacterium was recognised as being heterogeneous by substantial evidence obtained from experimentations investigating (i) bacterial recombination; (ii) genome size; (iii) single-nucleotide polymorphism (SNP); and (iv) allelic variation (46-48). Genetic recombination is an interesting characteristic of *H. pylori* involved in its adaptation to the stomach of humans (49). CagA is an *H. pylori* virulence factor that has been isolated in various countries and has different polymorphisms which have direct impact on its clinical significance (45). The ability of *H. pylori* for the uptake of DNA results in this bacterium having one of the highest reported rates of recombination in pathogens, which cause high genomic variability among the *H. pylori* strains (48). As a result, the bacterium is more persistent in the stomach because of natural selection following these genetic variations. Further to the recombination, *H. pylori* can persist in the stomach through mutation by generating highly variable strains (50). Approximately 5-10% of the bacterial genome seems to be *H. pylori* specific (44). Comparative analysis of complete genome sequences of *H. pylori* strains 26695 and J99 provided new vision into *H. pylori* diversity and the bacterial ability to infect people long-term (51). Although the genome of the *H. pylori* is origin, several

alterations in the third base of codons have generated quasi-species, which are specific strains (45,47,52,53). Quasi-species in *H. pylori* have been generated due to the development over time of a pool of genetic variants from a single strain within a host caused by genetic drift, providing a bacterial population that can be selected by the altering stomach condition of the host. Several implications may have generated from this phenomenon, such as adaptation to new hosts that enable successful colonisation and adaptation to different gastric compartments which have different selective pressures, such as the non–acid-producing antrum and the acid-producing corpus (52). Both phase and antigenic variation participate in the bacterial persistence against immune system (54,55). Indeed, broad diversification in the genome of the bacterium increases the opportunity of the bacterial existence in the harsh condition of the stomach. It is necessary to explore the *H. pylori*'s biology to address a direct question about the potential technique of *H. pylori* diversity to establish an infection with prolonged persistence.

1.5 *H. pylori* virulence factors

Several studies focus on virulence factors that discriminate pathogenic bacteria from non-pathogens. In the case of *H. pylori*, the most important feature is its ability to persist in the stomach for many years despite the hostile environment of the stomach. *H. pylori* has various virulence factors that facilitate the adhesion process and interact with host cells in order to trigger the immune cells causing a disease.

1.5.1 Adhesins

The adherence of *H. pylori* to the gastric mucosa plays a significant role in initial colonisation as well as in long-term persistence in the stomach. There are many obstacles that *H. pylori* faces when colonising the gastric mucosa including the high viscosity of the mucus layer, continual shedding of mucin and epithelial cells, low pH, and relative deficiency of nutrients due to the peristalsis. To overcome some of these obstacles, *H. pylori* expresses a range of adhesins, which contribute to the establishment and persistence within the stomach through recognising and mediating interactions with carbohydrate receptors which are found on the gastric epithelium and in the mucus layer (56,57). The targeted receptors are fucosylated ABO blood group antigens, antigens (58), sialyl-Lewis x and sialyl-Lewis a (sLe^x and sLe^a) (59), and GalNAcβ1-4GlcNAc motif (also known as *N,N'*-diacetyllactosediamine [lacdiNAc]) (41). However, glycan-independent interactions of *H. pylori* adhesins with the gastric epithelium have also been

recently discovered, for example the adherence of HopQ adhesin to the carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) (60,61).

H. pylori is equipped with more than 60 OMPs (40), representing 4% of the bacterial genome (56), 33 members of which constitute a large family of paralogous OMPs comprising the *Helicobacter* outer membrane porins (Hop) and Hop-related proteins (Hor) (62). The Hop family of proteins includes 21 members, which share a carboxyl-terminal membrane spanning β-sheet motif, a conserved amino-terminal signal peptide sequence and a variable mid-region (56). Few of the Hop family are well-described adhesins participating in bacterial adherence to the stomach such as the adherence-associated lipoproteins A and B (AlpA and AlpB) (63), the blood group antigen-binding adhesin (BabA) (58), the sialic acid-binding adhesin (SabA) (59), the *Helicobacter* outer membrane porin Z (HopZ) (64), the *Helicobacter* outer membrane protein Q (HopQ) (60,61), and the lacdiNAc-specific binding adhesin (LabA) (41), suggesting that the entire Hop family such as BabB, BabC and SabB may have specific roles in adhesion. Table 1.1 shows adhesins known to mediate attachment of *H. pylori* to the gastric epithelial surface, but not all of these adhesins have known receptors. The possession of a variety of adhesins that recognise different moieties expressed on the gastric epithelial surface is believed to be helpful and useful for the bacterium by enabling it to adapt to alterations in the environment of the stomach during establishment of chronic

colonisation (65). Some of these OMPs can alter the number of dinucleotide repeats through slipped strand mispairing resulting in phase variation. This diversity in adhesins and phase variation of OMPs allow specificity of *H. pylori* attachment to specific host cell receptors *in vivo* under different conditions (57).

Table 1.1: Adhesins known to mediate attachment of *H. pylori* to the gastric epithelial surface

sLe^x, sLe^a, Le^x, lacdiNAc, and CEACAMs stand for sialyl-Lewis x, sialyl-Lewis a, *N,N'*-diacetyllactosediamine, and carcinoembryonic antigen-related cell adhesion molecules, respectively.

Adhesin	Abbreviation	Host Receptor	Reference
Blood group antigen-binding adhesin	BabA	Type 1 ABO/Lewis blood group antigens	Ilver *et al*. (58)
Adherence-associated lipoprotein A	AlpA	Unknown	Odenbreit *et al*. (63)
Adherence-associated lipoprotein B	AlpB	Unknown	Odenbreit *et al*. (63)
Helicobacter outer membrane porin Z	HopZ	Unknown	Peck *et al*. (64)
Sialic acid-binding adhesin	SabA	sLe^x, sLe^a, and Le^x	Mahdavi *et al*. (59), and Pang *et al*. (66)
LacdiNAc-binding adhesin	LabA	LacdiNAc	Rossez *et al*. (41)
Helicobacter outer membrane protein Q	HopQ	CEACAMs	Javaheri *et al*. (61)

Some studies have described mutant *H. pylori* strains and their binding capacity to cells and/or tissues *in vitro* (41,59,63,64). However, this approach does not give direct evidence that the adhesin targeted by the mutation is involved in binding. For example, if a second *H. pylori* protein is necessary to help BabA to bind to its ligand, and the second protein is mutated, then BabA would be unable to bind to its ligand. This does not mean that the second protein is also binding to the ligand, but it is involved in binding without binding directly itself. Furthermore, when one adhesin is targeted and mutated, leaving all the other adhesins intact some of which may bind to the same ligand, a partial or no effect at all could be seen with the mutation. Therefore, this approach is difficult to assess directly in *H. pylori* (57).

On the other hand, animal models are limited particularly in the case of *H. pylori* research in order to compare the ability of *H. pylori* strains, with and without the adhesin, to colonise and persist in the stomach of animal models. This is because of physiological differences between mouse and human gastric tissue and the fact that *H. pylori* is adapted to the human stomach. These animal model experiments study the adhesin in a molecular and physiological context which overlook the interspecies differences in, for instance, the structures of ligands present on the epithelium, and the gastric physiology, among others. Moreover, human *H. pylori* strains do not colonise mice very efficiently, and most mouse studies use mouse-adapted *H. pylori* Sydney strain 1 (SS1) (67).

Other studies have used recombinant adhesin proteins to examine their binding affinity to cells/ligands using a range of different experimental approaches (68). recombinant proteins give detailed information regarding the ligands, and the type of binding and binding affinity between adhesin and ligand can be measured. Furthermore, if a crystal structure is obtained, details of the binding at the atomic level can be determined. However, the weakness of this approach is that it is unknown how much of the adhesin is actually expressed on the surface of the bacterial cells, whether it is only expressed under specific circumstances, and how it behaves under different physiological conditions (e.g. different pH). Also, this approach only works if the protein can be expressed in a soluble, properly folded form, which needs to be demonstrated (e.g. by circular dichroism spectroscopy, or binding to ligands, etc.). Therefore, experimentally, this approach can be challenging. Each of the aforementioned experimental approaches has strengths and limitations, and thus the strongest evidence will come from a combination of these approaches (61).

1.5.1.1 BabA, BabB, and BabC

BabA was the first adhesin of *H. pylori* identified in 1998, and is responsible for bacterial adherence to Lewis b (Le^b) blood group antigen, which is expressed on gastric mucosa of most individuals (58). Le^b was discovered in 1993 as the main receptor involved in the *H. pylori* attachment to the gastric mucosa of healthy people (69). Binding of BabA to Le^b enhances

the ability of type IV secretion system (T4SS) to exert *cag* pathogenicity island (*cag* PAI)-dependent *H. pylori* pathogenicity through stimulating the release of proinflammatory cytokines and precancer-related factors in a response of the host to the disease (Figure 1.2A) (70).

It is reported that BabA has several chromosomal loci and alleles, in which *babA2* encodes BabA, while the *babA1* allele is non-functional and naturally very rare (71). Therefore, *babA1* strains are classified as strains not expressing BabA because they do not produce detectable BabA protein even if they do have a *babA* gene (72). It has been stated that *H. pylori* strains that produce CagA, type s1 of VacA (*VacA*-s1), and BabA can colonise a larger proportion of the gastric mucosa, stimulate severe gastric inflammation (73), and correlate with the highest risk of gastric cancer development (74).

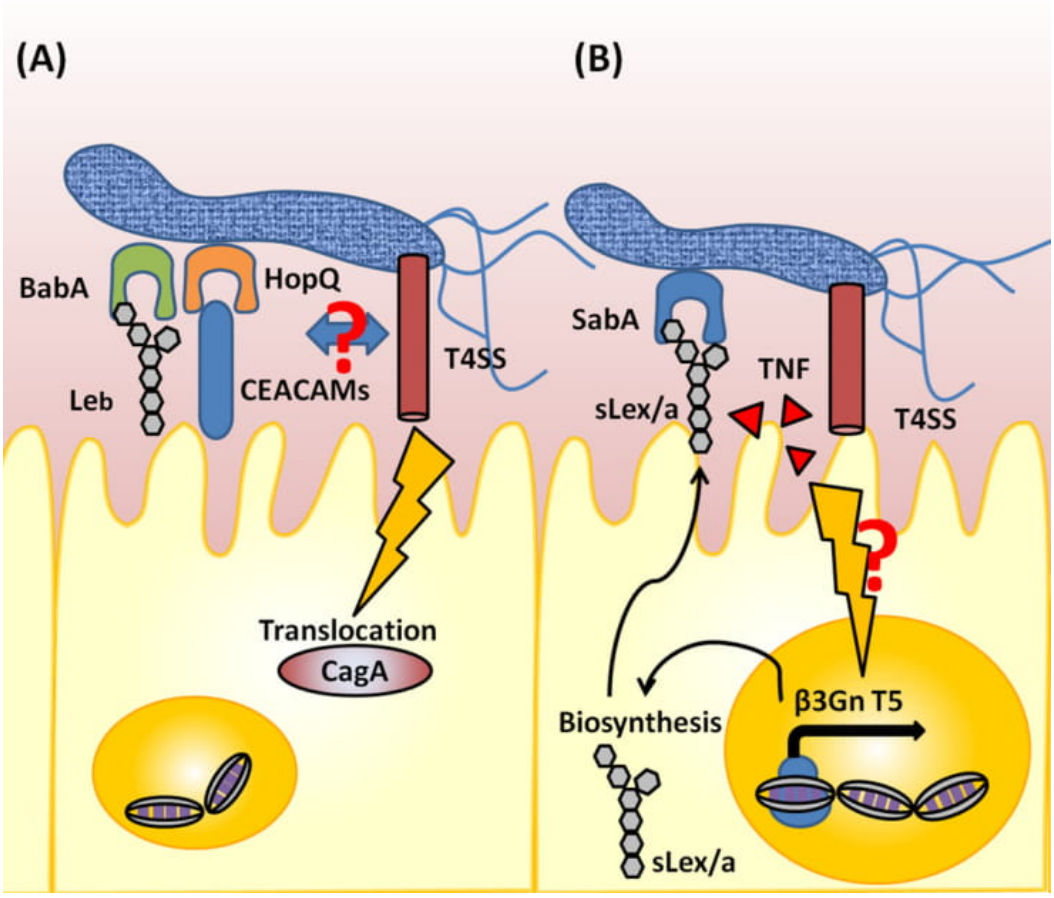


Figure 1.2: Schematic of pathogenesis mediated by OMPs of *H. pylori*
(A) BabA binds to Le^b antigen and enhances the translocation of CagA via the T4SS, whereas HopQ binds to CEACAMs and is important for the translocation of CagA. Although the mechanism is ambiguous, both BabA and HopQ are proposed to interact with the T4SS in the bacterial membrane area. (B) sLe^x is not expressed in the stomach of healthy individuals. β3GnT5 is induced in epithelial cells of the stomach by *cag* PAI-positive bacterium, and it is important for sLe^x biosynthesis; sLe^x antigen localises to the membrane area of the epithelial cells. The T4SS and TNF-α are proposed to stimulate the expression of β3GnT5, although the mechanism is unclear. Adapted from (75).

Hage and colleagues successfully expressed and purified a recombinant form of BabA in the periplasmic space of *Escherichia coli* (*E. coli*) cells with improvement of the expression and purification through the incorporation of a hexa-lysine tag in the C-terminal region of BabA (76). Furthermore, the crystal structure of the extracellular region of the protein was determined in presence and absence of Le^b in which the amino acids responsible for the binding were detected (Figure 1.3). Isothermal titration calorimetry (ITC) was used to measure the binding affinity of BabA to its ligand at pH 4.5 and 7.4 and the dissociation constant (K_D) was around 227 and 252 μM, respectively (68).

The initial bacterial adhesion to the gastric mucosa of healthy people is mediated by the adhesin BabA. This hypothesis is supported by evidence showing that only strains that express BabA maintained their binding to mucin MUC5AC at pH 7.4 *in vitro*, which correlates to the pH value on the gastric epithelial cells of the stomach, while the ones do not express BabA only bound to mucin MUC5AC at an acidic pH 3 *in vitro* (77). Since, the binding of BabA to Le^b has also been shown to be pH-dependent. The binding affinity is lowered at acidic values of pH, which may be a mechanism of defence established by the bacterium, in order to evade the effects of the constant turnover of epithelium and mucus. When the mucus and epithelial cells are shed into the lumen of the stomach, where the environment is acidic and bactericidal, the binding of BabA to Le^b is disrupted by the acidic condition. Therefore, *H. pylori* is not emptied with mucus (78).

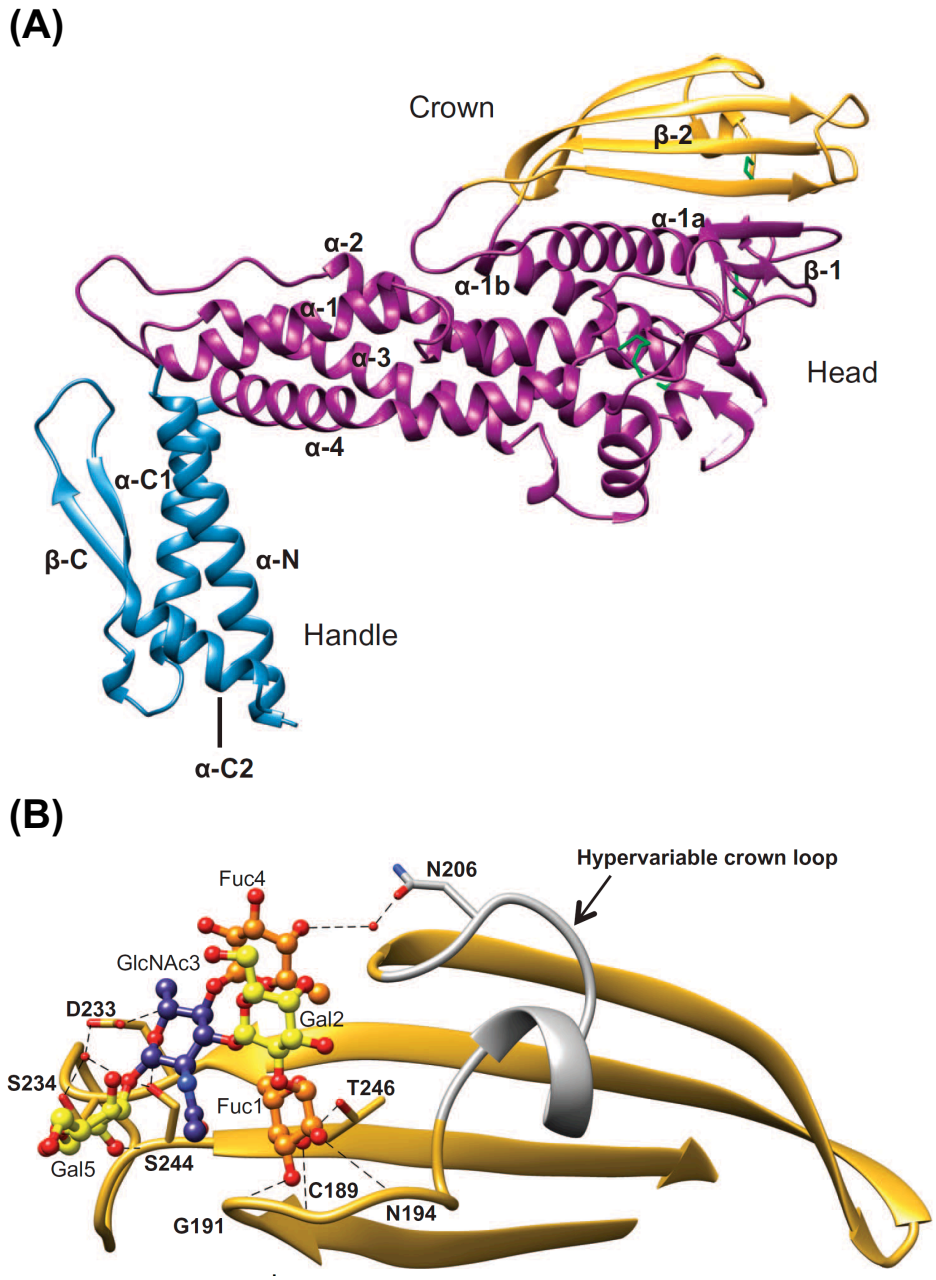
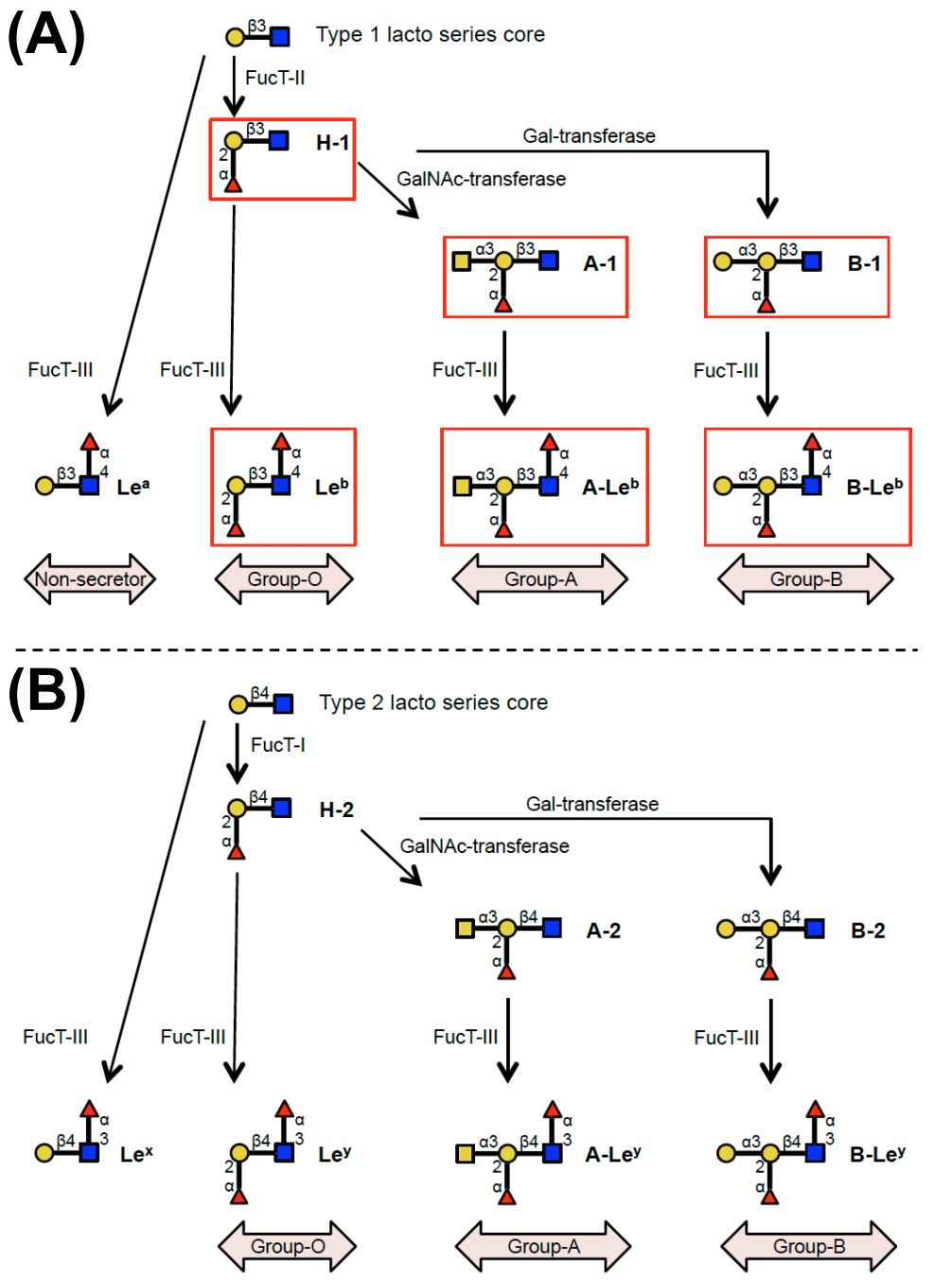


Figure 1.3: Crystal structure of the extracellular region of BabA J99
(A) Crystal structure of the extracellular region of BabA (PDB accession no. 4ZH0), consisting of a 'crown' (gold), a 'head' (purple) and a 'handle' (blue). The intramolecular disulphide bridges of the protein are shown in green. (B) Crystal structure of the binding site of BabA in presence of Le^b (PDB accession no. 4ZH7). The highlighted are the sugar moieties and the amino acids involved in hydrogen bonds, including a water-mediated bond between a fucose (Fuc4) and N206. The grey region of the crown is one of the two so-called hypervariable crown loops, which are now understood to fine-tune the specificity of BabA for the various blood group antigens. Adapted from (68).

BabA has specific affinity for ABO/Le blood group antigens which contain a type 1 lacto series core (69), while no affinity for ABO/Le blood group antigens with a type 2 core (58). The blood phenotype determines the presence of type 1 and type 2 ABO/Le blood group antigens (Figure 1.4). Although the majority of the bacterial strains (more than 95%) that bind to Le^b antigen can bind to A, B, and O antigens and are hence termed 'generalist' strains, there are strains which bind O antigens best further to Le^b antigen and therefore termed 'specialist' strains. The later strain are frequently found in the native population of South America, where only the Blood Group O phenotype (H Type 1 determinants) occurs (79). Recently, it was shown that the Globo H hexaglycosylceramide, e.g. the Blood Group O determinant on a Type 4 core chain can be bound by BabA of both strains (generalists and specialists), while the Globo A heptaglycosylceramide, e.g. the Blood Group A determinant on a Type 4 core chain can only be bound by the generalist strains. BabA binding to the two isoreceptors sets is because of conformational terminal disaccharides similarities of Globo H and H type 1 and the conformational terminal trisaccharides similarities of Globo A and A type 1 (80).

The sequence of *babB* and *babC* genes show high 5' and 3' homology to *babA* gene, but the central region differs (81-83). This indicates that the sequence divergence in the central region encodes a unique receptor binding region (84,85). BabB and BabC proteins have not been expressed

yet, their putative ligands, if any, are unknown, and their crystal structures remain to be determined.



Glycan symbolic representations can be interpreted with the following key:
fucose – ▲ , **galactose** – ● , **N-acetylgalactosamine** – □ , **Nacetylglucosamine** – ■

Figure 1.4: The biosynthesis pathways of the dominant ABO/Le blood group antigens in healthy gastric mucosa

The terminal glycan epitopes of the secreted type 1 (A) and type 2 (B) ABO/Le blood group antigens in the gastric mucosa are shown. Epitopes, which are outlined by a red box, act as receptors for BabA. Adapted from (86).

1.5.1.2 SabA

BabA is known to bind to Le^b receptors on the normal gastric mucosa of individuals, leading to the initial infection (58). However, upon inflammation of the stomach because of the persistent colonisation by *H. pylori*, the sLe^x and sLe^a glycosphingolipids are expressed on the gastric mucosa, resulting in inactivation of *babA* gene due to loss of Le^b. These molecules are upregulated only in the case of stomach inflammation and act as receptors for SabA (59). The process of SabA adherence is dynamic and collaborates with the tumour necrosis factor alpha (TNF-α) and T4SS, which induce the expression of sLe^x. It was shown that the expression of β1,3-*N*-acetylglucosaminyltransferase 5 (β3 GlcNAc T5 or β3GnT5), which is a glycosyltransferase and is important for the Lewis antigens biosynthesis, is induced in gastric epithelial cells by *cag* PAI-positive *H. pylori*. This induction is related to TNF-α, but not to interleukin-8 (IL-8). With respect to bacterial adherence through SabA, the expression of β3GnT5 was shown to increase the cell adherence for *H. pylori* strains with SabA positive (87). Furthermore, SabA transcription is regulated by acid-responsive signalling in *H. pylori* (88) (Figure 1.2B). These findings show that bacterial adaptation to the gastric condition is an adherence precursor through SabA; subsequently, *H. pylori* is able to persist in the stomach despite the inflammatory reaction.

It has also been suggested that SabA plays an important role in *H. pylori* colonisation by binding to sLe^x in individuals with no or weak Le^b antigen

expression (66). The strain J99 has a *sabA* gene (*jhp0662*) located at the *sabA* locus (89), while the strain 26695 has *sabA* at both *sabA* and *sabB* loci (90). It has also been shown that there are SabA variants in different clinical isolates of *H. pylori* which may exhibit different specificities and affinities for sLe^x, sLe^a, and sialyl-lactosamine (sLn). This variability in SabA ligand-binding specificity enhances *H. pylori* adaptation through rapidly modulating its adherence in order to allow optimal colonisation (66). It is suggested that the expression of SabA can quickly respond to environmental alterations in the stomach through the frequent on/off expression switch of SabA (83). Slipped strand mispairing, regulates the functional SabA status, where switch “on” indicates that the protein is functional, while it is non-functional when it is switched “off” (91-93). Pang *et al*. revealed the crystal structure of the SabA extracellular domain and discovered a predicted ligand-binding pocket for sLe^x and Le^x binding despite the crystallisation of protein-ligand complex was unsuccessful (66).

On the other hand, it has been shown that SabA binds to sialylated glycoconjugates on erythrocytes found in capillaries and post-capillary venules of the lamina propria of stomach mucosa, and thus identified as a haemagglutinin. Therefore, the small number of *H. pylori* found in that site could result in systemic dissemination; however, the bacterium has not been detected in blood of *H. pylori*-infected patients (94). It was reported that CagA-containing exosomes were found in serum of mice and patients with $CagA^+$ *H. pylori* infection (95,96). It was also demonstrated that the bacterial

infection significantly damages the function of endothelium in mice and patients through exosome-mediated mechanisms, and *H. pylori* infection might be a novel risk factor for cardiovascular diseases, particularly in young patients (96). SabA of non-opsonised *H. pylori* strains can also bind to sialylated glycoconjugates on neutrophils resulting in stimulation of neutrophils through the release of reactive oxygen species and stimulation of phosphatidylinositol 3-kinase causing oxidative damage of the gastric epithelium (97).

De Jonge *et al*. found that the *sabB* off-status correlated with duodenal ulcers (98). This might be related to gene conversion through intragenomic recombination of *sabB* with *sabA*, leading to expression of two copies of SabA, leading to increased binding to sLe antigens (99).

1.5.1.3 HopQ

Similarly to BabA, and SabA, HopQ is a member of the Hop family (89). HopQ is categorised into two different families of alleles, Type I and II alleles which are associated with other virulence factors. Cao and Cover stated that Type I alleles of HopQ are found more frequently in *cagA*-positive strains with *vacA*-s1 than *cagA*-negative strains with *vacA*-s2 (100). It was also shown that Type I alleles of HopQ are more prevalent in strains located in East Asia, based on an epidemiological study (101). This reveals the existence of geographical variations for the allele type of HopQ, similar to *cagA* and *vacA* genotypes (75).

HopQ was suggested to be important for the translocation of CagA, and its deletion decreases T4SS-dependent signalling activation (Mitogen-activated protein kinase (MAPK) signalling, nuclear factor-kB (NF-kB), and IL-8 production) (Figure 1.2A) (102). This finding was also reported for human granulocytes (103). It was recently demonstrated that HopQ binds to CEACAMs (CEACAM1, 3, 5, and 6) as a group of host cell receptors through binding their amino-terminal IgV-like domain (60,61). Sequence diversity has been found in the CEACAM-binding site of HopQ, and thus these variations may affect CagA translocation through T4SS (75).

1.5.1.4 AlpA and AlpB

The bacterium also uses the proteins AlpA and AlpB in its adherence to the gastric epithelial cells (63). It is suggested that both AlpA and AlpB bind to laminin and collagen IV, which are found in the extracellular matrix (104,105). Although clinical isolates of *H. pylori* strains were shown to produce other OMPs, such as BabA, SabA and outer inflammatory protein A (OipA), it was reported that they also expressed AlpA and AlpB proteins (62). This indicates that both proteins are important and play a significant role in bacterial colonisation. Subsequent research has shown that *H. pylori* strains with mutation in both proteins have poor colonisation of the stomachs of mice, Mongolian gerbils, and guinea pigs (105-107). Therefore, both proteins are clearly significant for *H. pylori* colonisation. However, their role in the pathogenesis of *H. pylori* infection is controversial.

It was suggested that AlpA and AlpB do not affect the CagA translocation through the T4SS, and thus both proteins are involved in the induction of signalling pathways, such as c-Jun, c-Fos, NF-kB, and MAPKs pathways, which can stimulate production of proinflammatory IL-8. Furthermore, a reduction in interleukin-6 (IL-6) and chemokine KC (for Keratinocyte chemoattractant) are associated with these factors in the gastric mucosa (106). However, it was reported that the wild type (WT) *H. pylori* induced a less severe inflammation in Mongolian gerbils than the AlpA- and AlpB-mutant *H. pylori* (105). This contradictory result might have occurred because of differences in *H. pylori* strains or host animal species (75).

1.5.1.5 LabA

Although BabA and SabA are the most prevalent adhesins used by *H. pylori* for adhesion process, not all strains express these adhesins functionally (62,108). In 2014, Rossez *et al.* identified and characterised the *H. pylori* adhesin LabA, which binds to lacdiNAc, a disaccharide found on MUC5AC gastric mucins and restricted to the gastric mucosa. Significantly, it was also shown that all the strains tested bind to lacdiNAc. Understanding the regulation of lacdiNAc expression in physiopathological cases will provide information about the role played by the glycan in gastric homeostasis and its evolutionary significance (41).

LabA was recombinantly expressed by our group with and without hexa-lysin tag at the C-terminus of the protein. Differential scanning fluorimetry and

circular dichroism spectroscopy revealed that the insertion of the hexa-lysin at the C-terminus of LabA enhanced its thermal stability of LabA without affecting the secondary structure (109). Unlike BabA (76), this tag did not protect the N-terminal of LabA from partial degradation in the periplasmic space of *E. coli* (109). In contrast to the data provided by Rossez *et al.* (41), Paraskevopoulou *et al.* in our group revealed no binding affinity of LabA from the *H. pylori* strains J99 and 26695 for the porcine gastric mucin or for any of the glycoconjugates investigated including lacdiNAc (110). This was confirmed by Mthembu *et al.* who showed no binding detection of *H. pylori* strains J99 and 26695 for lacdiNAc in a microtiter well-based adhesion assay. Furthermore, they did not find any binding of a global series of *H. pylori* strains to a soluble lacdiNAc-conjugate. Investigations are needed to show the importance of lacdiNAc in the binding of *H. pylori* to the gastric mucosa and mucins and to show how lacdiNAc relates to the LabA (111). Figure 1.5 shows the crystal structure of the extracellular region of LabA J99 produced by our group (112).

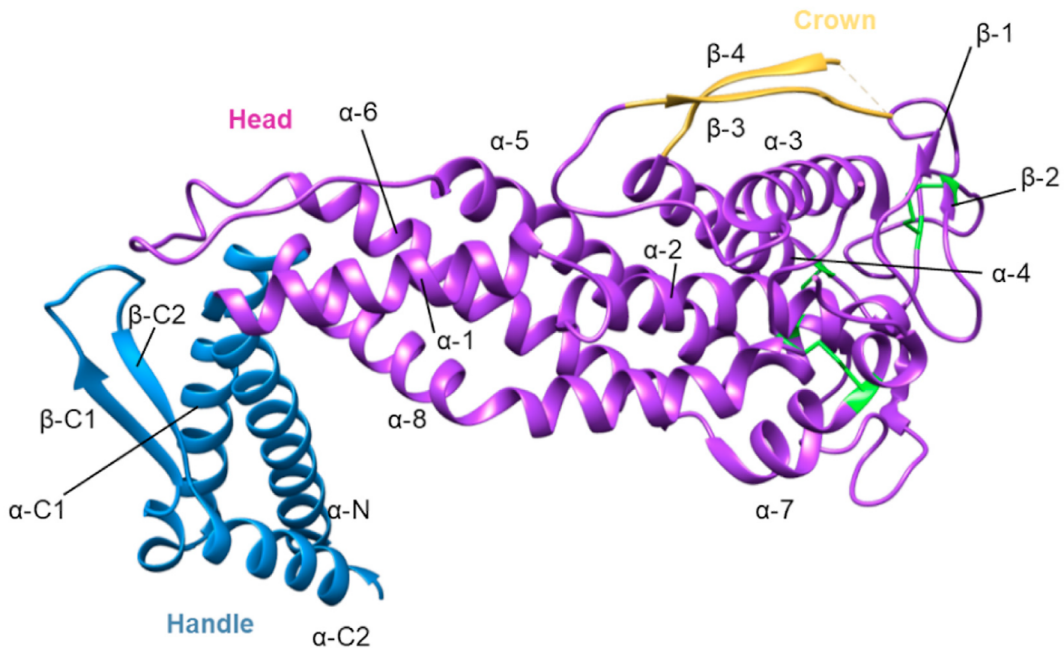


Figure 1.5: Crystal structure of the extracellular region of LabA J99
Crystal structure of the extracellular region of LabA J99 (PDB accession no. 6GMM), consisting of a 'crown' (gold), a 'head' (purple) and a 'handle' (blue). The intramolecular disulphide bridges of the protein are shown in green. Adapted from (112).

1.5.1.6 *Helicobacter* outer membrane B (HomB)

HomB, a putative OMP in locus *jhp0870*, was recently discovered in a strain of *H. pylori* isolated from a young patient with peptic ulceration (113). Studies have addressed that the existence of HomB in *H. pylori* is strongly associated with peptic ulcers (114). Further studies have linked the existence of HomB with increased severity of the disease resulting in gastric cancer (115,116). Therefore, *homB* gene has been suggested to be a discriminating marker to differentiate stomach cancer from peptic ulceration (115,116). *homB* gene is globally associated with the existence of other *H. pylori* virulence factors including *cagA*, *babA2*, *vacA* s1, *oipA*, and *hopQI* (113,114,117,118). The clinical outcome of *homB* positive strain is supported by the association of HomB with the above virulence factors, regardless of the patient age and the clinical origin of the strain (114). Therefore, HomB is generally seen in more virulent strains, suggesting that the status of *homB* could be used in determining the association of the genetic pattern with a higher pathogenicity degree of the bacterium (114).

Abadi *et al*. found that the existence of CagA in Iranian *H. pylori* strains has no significant influence on the type of the disease, suggesting that the development of gastric cancer is strongly associated with the presence of *homB* (116). On the other hand, Kang *et al*. demonstrated that the progression to gastric cancer is not associated with the presence of *homB* in South Korean *H. pylori* strains, suggesting that other virulence factors exist in the strain in a shape of hierarchy. These virulence factors have

different effects on the disease based on the existence of other virulence factor polymorphisms (119). However, the Iranian study (116) recruited 138 people undergoing endoscopic procedures for dyspepsia, whereas in the South Korean study (119), they recruited 260 people and used 115, 60, 55, and 30 isolates from patients diagnosed with gastritis, gastric ulcers, duodenal ulcers, and gastric cancer, respectively. Therefore, it is difficult to compare the two studies as they have different stages of disease. Generally, the smaller the number of participants, the more likely any confounding variables are to distort the results. Furthermore, a confounding variable, which is a factor that might have affected the result independently of the relationship between HomB and the disease status, could falsify the results. Possible confounding factors could be related to diet, specifically high salt diet, as there is a known link between high intake of salt and *H. pylori*-dependent gastric cancer (120). For example, East Asian diets (e.g. Japan, China, South Korea) are very high in salt intake (121), and if high salt dietary intake is a stronger factor determining progression to cancer, thus the fact that HomB does not correlate with cancer in the South Korean study may be because high salt diet is a stronger factor. In other words, high salt diet is more important in driving carcinogenesis in South Korea. This is compatible with the notion that gastric cancer mortality rates vary geographically; the highest rates are in East Asian countries like South Korea, Japan, and China (119), which have a high salt diet. On the other hand, it appears that Iran is also a high salt diet country (122), and thus may be that is not the reason for

the discrepancy between the two studies. None of the two studies (116,119) addresses high salt diet or any other confounding variables, which is a weakness of both studies. However, Azadnajafabad *et al*. show that males had a significantly higher prevalence of high salt intake, while among females, the participants with the lower economic status had a higher salt consumption than the participants with higher economic status (122). Therefore, two other potential confounding variables can be addressed; "dietary salt intake and gender" and "dietary salt intake and economic status" in which the discrepancy could be due to these other factors.

It was recently discovered that HomB can contribute to biofilm formation. Servetas *et al.* (123) revealed that the expression of *homB* is regulated by the ArsRS two component system, which suppresses *homB* expression in response to low pH. ArsRS is an acid responsive system which consists of two components including acid-responsive signalling sensor (ArsS) and acid-responsive signalling response regulator (ArsR). The former is a sensor kinase while the latter is a response regulator. Under conditions of reduced pH, ArsS autophosphorlyates and transfers the phosphate group to ArsR which, in turns, activates or represses the expression of downstream genes (123). Importantly, Servetas *et al.* showed that mutation of ArsRS enhances biofilm formation (124). The expression of *homB* was increased in Δ*arsS* strain, a non-phosphorylatable ArsR-D52N mutant strain and in Δ*nikR* strain. Therefore, *homB* expression is suggested to also be regulated by the nickel (Ni^{2+})-responsive regulator (NikR); a metallo-regulator sensitive

to Ni^{2+} availability. However, it was also demonstrated that the expression of *homB* was decreased in response to excess Ni^{2+} in both WT and Δ*nikR* strains, suggesting that the reason for the decreased expression of *homB* is NikR-independent. Moreover, It has been reported that no decrease in the expression of *homB* was seen in Δ*arsS* and ArsR-D52N mutant strains when exposed to pH 5 shock, while it exhibited a 10-fold decreased in WT (123).

It was shown that the early biofilm phenotype was negated in Δ*arsS* or ArsR-D52N strains with *homB* deletion in which significant biofilm formation was not observed until 48 hours in the double mutant strains. Furthermore, Δ*arsS* and ArsR-D52N strains formed significantly more biofilm than WT by 48 and 72 hours, respectively, while ArsR-D52NΔ*homB* and Δ*arsS* Δ*homB* did not. Generally, all of the Δ*homB* strains formed less biofilm compared to their cognate strains with *homB* expression, but these differences were not always statistically significant. In addition, *homB* expression *in trans* from a promoter that negated the ArsRS regulation, such as *ureA* promoter, led to enhanced formation of biofilm even in strains in which the native *homB* copy had been deleted (123). Despite the absence of a chromosomal copy of *homB*, the expression levels of *homB* in the G27 Δ*homB* pUreA-*homB* were similar to the expression levels of *homB* seen in G27 pUreA-*homB*. These findings suggest that there is a maximum expression level of *homB* that can be tolerated in *H. pylori* (123).

Although HomB is an important factor in biofilm formation by *H. pylori*, biofilm formation was not totally ablated after deletion of *homB* (123). It was

reported other OMPs play a role in biofilm formation in *H. pylori* strain TK1402 such as AlpB (125). It was also shown that ArsRS regulates numerous other OMPs including *labA*, *hopZ* and *sabA*, and their expression increases in Δ*arsS* strains (126). Therefore, it is reasonable that other OMPs regulated by ArsRS may contribute to the biofilm formation that remains in the Δ*homB* strains (123).

Biofilm formation by *H. pylori* may have a role in failure of antibiotic treatment (127). Haddadi *et al.* (128) found that the existence of *cagA*-*homB*$^{+}$ is associated with resistance to antibiotics, and *homB* can be used as a predictive marker of amoxicillin resistance in *H. pylori* isolated in Shiraz. Furthermore, a correlation between the presence of *homB* and acquisition of multi-resistance phenotypes particularly in peptic ulcer disease was found (128). This was attributed to the role of *cagA* and *homB* in biofilm formation and the established relation of these genes with antibiotic resistance phenotypes in upper gastrointestinal diseases (128). However, as mentioned earlier there are other OMPs that contribute to biofilm formation which may also be associated with antibiotic resistance.

Hom family is the smallest OMP family that includes HomA, HomB, HomC, and HomD (Appendix Figure 1) (83). Although *homB* gene shares 90% sequence identity with *homA*, a closely related paralog, at the 5' (segment 1) and 3' ends (segment 3), the middle region (segment 2), which is considered to be a defining allele region, is unique in the two genes (83,114,117). Considering the number of copies based on the sequenced strains of

H. pylori (51,93), *homA* and *homB* can be present in two conserved loci (locus A and B) within a genome as a single-copy (*homA* or *homB*) in both Western and East Asian strain, or double-copy (*homA*/*homA*, *homA*/*homB*, *homB*/*homA*, or *homB*/*homB*) in Western strains only (114,117). In the case of single-copy, a single *hom* gene occupies locus A in Western *H. pylori* strains, while locus B is occupied in East Asian strains (Figure 1.6) (117,119). These findings suggest that positions can be exchanged between *homB* and *homA*. Furthermore, occurrence of gene duplication, in response to environmental changes, by recombination events may results in *homB* existence (89,114,117).

Interestingly, it has been shown that HomB is expressed in the outer membrane of *H. pylori* as well as able to stimulate an *in vitro* proinflammatory response by inducing IL-8 secretion, suggesting that HomB, in presence of CagA, is involved in bacterial adherence and in inflammatory response based on the number of *homB* copies (83). This was confirmed by the reduced ability of *homB*-mutant strains to induce the secretion of IL-8 from gastric epithelial cells, and by the reduced capacity of the same mutant strain to bind to gastric epithelial cells. Because of knocking out the *homB* did not totally abolish bacterial adherence to gastric epithelial cells, HomB protein is involved in bacterial adherence but not the main adhesin (83,114). Furthermore, it was shown that *homB*-positive strains increase neutrophil infiltration and mucosal atrophy in the corpus mucosa (115). This would

appear to support the description of HomB as a virulence factor and/or marker for *H. pylori* strains associated with peptic ulcer and gastric cancer.

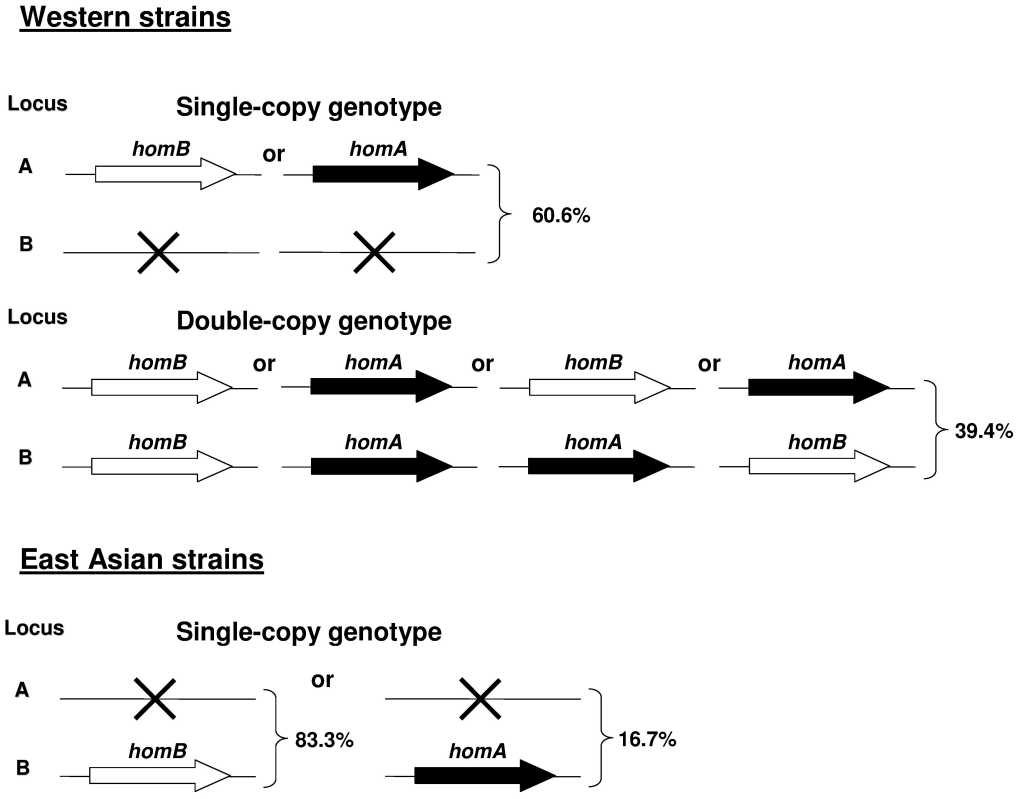


Figure 1.6: Diversity of *homA* and *homB* in the copy number and their genomic localisation in Western and East Asian *H. pylori* strains
The percentage denotes the frequency of each genotype (single-copy or double-copy) among Western and East Asian strains. X represents the unoccupied locus. Adapted from (117).

1.5.2 CagA

There are many virulence factors, other than adhesins, expressed by *H. pylori*. Some of these are translocated into gastric epithelial cells through T4SS, which is encoded by 6 genes located at the *cag* PAI which contains 31 genes (129). CagA protein is one of the virulence factors that is translocated into the cytoplasm of gastric epithelial cells through T4SS (130-134), and it is encoded by the gene *cagA*, which is one of the PAI genes (129). *cagA* gene is normally used as a marker for the whole locus of *cag* (135). *H. pylori* strains may either own the entire pathogenicity island (*cag* positive), miss it completely (*cag* negative), or possess parts of it which considered as cag^- strains (136). CagA protein is known as a major *H. pylori* virulence factor correlating with stomach cancer (90).

Once translocated into epithelial cells of the stomach, CagA is subsequently phosphorylated on tyrosine residues at the C-terminus by Src family kinases of the host cell (137,138) and Abl kinases (139), resulting in cell signalling stimulation. The most apparent influence *in vitro* is formation of cell-shape alteration termed the hummingbird phenotype (131,140). Interestingly, it has been reported that these changes in the cell are dependent upon the CagA-induced activation of the oncogenic hepatocyte growth factor receptor c-Met (141). Another effect of CagA is activation of c-Jun and c-Fos leading to cellular proliferation through stimulating transcription factors (142). Unphosphorylated CagA can result in loss of gastric epithelial barrier function through disruption of the epithelial apical junctional

complexes (143). This may be the way for allowing nutrient transfer to the bacteria and enhancing the pathogenicity through destroying the epithelial cell barrier (144).

1.5.3 VacA

VacA is an exotoxin produced by *H. pylori* which triggers many host immune responses, involving vacuolisation, apoptosis, and inhibition of T cell functions (145,146). It is resistant to the acid and pepsin digestion once activated by stomach acidity (147). Although all *H. pylori* strains have *vacA* gene and express VacA (148), the most active form of this protein is only produced by 40% of them (149). This is due to the genetic diversity in the *vacA* gene which varies in three main regions: a 5′ region which encodes the mature N-terminal protein (s1 or s2) following the signal sequence; a mid region which encodes part of the p55 epithelial cell–binding subunit (m1 or m2); and an intermediate region which encodes part of the p33 subunit (i1 or i2) (149,150). Strains that produce active forms of s1/i1/m2 and s1/i1/m1 VacA protein are highly correlated with peptic ulceration and gastric adenocarcinoma (151). These active forms are common in northern China and Japan and may contribute to the high gastric cancer rate in these countries (149).

1.6 The potential clinical implications of this study

A better understanding of binding specificities of adhesins may have clinical implications. For example, a precise knowledge of the totality of ligands recognised by *H. pylori* may enable the use of adhesion inhibitors. The use of adhesin inhibitors from natural sources has been suggested in a variety of settings (152), for example proanthocyanidins from cranberry juice against uropathogenic *E. coli* (153), or as suggested by our own lab, highly esterified rhamnogalacturonans from *Abelmoschus esculentus* (commonly known as Okra) extracts against *H. pylori* (154). While most of these studies have shown good to excellent antiadhesive properties *in vitro*, they tend to fail *in vivo* – a good example of this dichotomy is illustrated by the work of Lengsfeld and co-authors (155). Failure *in vivo* is mainly due to the fact that natural antiadhesion inhibitors are rapidly broken down by host- as well as microbiota-derived enzymes, resulting in a fast loss of action. As a result, even the use of cranberry juice as a common home remedy against urinary tract infection does not hold up against strict scrutiny, as shown in the systematic review of Gbinigie *et al*. (156). Perhaps these limitations could be overcome by using synthetic compounds, with similar blocking activities, but more resilient or even completely resistant to enzymatic degradation, perhaps in combination with antibiotic treatment. This is in my view an avenue well worth exploring, but requiring a detailed understanding of the molecular interactions between *H. pylori* adhesins and their natural ligands.

The difficulty in the case of *H. pylori* is its use of multiple outer membrane proteins for adhesion to the epithelium, e.g. BabA/B/C, SabA/B, HomA/B, LabA, HopQ, AlpA/B etc., with several putative adhesins still remaining to be characterised. The resulting picture is that of a complex network of adhesin-ligand interactions, not only based on protein-carbohydrate interactions, but also on protein-protein interactions, e.g. the AlpA and AlpB interaction with host laminin (105) or HopQ with host CEACAMs (60). The situation is made even more complicated by phase variation (55), a genetic mechanism by which several bacteria, including *H. pylori*, can rapidly modify the carbohydrate structure of their Lipopolysaccharide (LPS), thus enabling new 'inverted' interactions with host cells, in which the carbohydrate ligands are on the bacteria and the protein 'adhesins' are on host cells, in response to various stimuli. For example, the generation of DC-SIGN (Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin; a C-type lectin found on the surface of dendritic cells)-binding and DC-SIGN-non-binding pools of bacteria within one single strain, with opposite effects on the immune response, has been described (55). It may therefore take a few more decades to solve the *H. pylori* adhesion riddle.

Additionally, there are studies showing that *Helicobacter* infection also has some benefits, hence it may do not need to be treated in every single individual. For example, Hussain *et al*. show that adults infected with *H. pylori* have significantly higher numbers of circulating IL-10-secreting $CD4^{+}CD25^{hi}$ Tregs, which modulate the immune response and contribute in

tolerance to self-antigens and the prevention of autoimmune diseases (157). In the case of confirming that HomB could work as a protease, the other clinical implication would be the use of HomB protease inhibitors, which would perhaps reduce chronic inflammation. However, this can be difficult in the context of gastric physiology, the low pH, and the constant gastric emptying.

1.7 Study aims and objectives

H. pylori adhesins are known to facilitate the adhesion of the bacterium to receptors found on the gastric mucus layer and the surface of gastric epithelial cells. Intensive studies have been performed to understand the nature of *H. pylori*, and more interestingly to understand its mechanism of adhesion and its ability to persist in human stomach at very low pH. Although some of *H. pylori* adhesins have been well studied, other OMPs contribute to bacterial adherence, but their participation is unknown. On the other hand, it has been suggested that glycan structures are used as receptors by *H. pylori*, but the bacterial adhesins corresponding to these glycan structures have not been identified (158,159). Therefore, it is clear that the mechanisms governing bacterial adhesion needed for the prolonged colonisation and this mechanism is not yet entirely understood. More work is needed to understand the dynamic interplay between these adhesins, some of which have not been fully characterised. That is why we decided to do more research on other bacterial adhesins of *H. pylori* that have not been studied well previously, i.e. with unknown structure and function. This will enable the future design of specific attachment inhibitors as a strategy for treatment of *H. pylori* which has particular significance in the context of increasing resistance of this pathogen to conventional antibiotic treatment.

In summary, the aim of this study is to characterise the functional properties of the *H. pylori* adhesins BabC and HomB, which have not been studied well previously with unknown structure and function. The aims will be achieved

via three research objectives, which to the best of our knowledge have not been addressed in the literature to date. These objectives are:

1. Cloning, expression and purification of recombinant BabC. This is a fundamental step of the project to support further analysis and applications, such as determination of glycan receptors for the protein and assessment of BabC stability in acidic conditions in the presence of pepsin, in order to achieve the first aim of this project (Chapter 3).

2. Cloning and recombinant expression of HomB. Development of a recombinant HomB expression and purification method. This is another main step of the study in order to support subsequent characterisation and application (Chapter 4).

3. Characterisation of HomB biological activity. This provides essential, first insights of the function of HomB (Chapter 5).

Overall, this project is expected to give a better insight into the mechanism of *H. pylori* adherence, colonisation and perhaps pathogenicity. Such knowledge will inform new strategies of antimicrobial therapies.

Chapter 2: Overview of experimental methodology

A wide range of techniques was employed throughout the course of this project in order to provide answers to the various questions raised.

The use of standard molecular biology and biochemical techniques was required in order to produce the *H. pylori* adhesins BabC and HomB. The techniques included polymerase chain reaction (PCR) and expression of recombinant proteins. An explanation of the instrumentation used and the mechanisms behind processes used to achieve this objective is excluded from this Chapter. However, a brief description of the methods of PCR, expression of recombinant protein, and expression plasmids and *E. coli* strains used in the current study is provided in this Chapter.

For BabC protein, binding activity screening of the protein towards glycan receptors located on the gastric epithelium of the human, was performed using the methods of enzyme-linked immunosorbent assay (ELISA). Acid/pepsin stability of the protein was also assessed at pH 1.2 and pH 4.5 for various time points.

For HomB protein, proteolytic activity was assessed in order to confirm the results from the predicted structure of HomB protein by Phyre^2, which indicated that HomB is a metalloprotease. Binding activity of the protein was also assessed using carbohydrate microarray, which was performed alongside BabA J99, the latter used as a positive control. Furthermore, measuring IL-8 induction from adenocarcinoma gastric cell line (AGS) was

carried out using ELISA after treating the cells with HomB. Finally, acid/pepsin stability of the protein was also assessed at pH 1.2 and pH 4.5.

2.1 PCR

PCR was discovered in 1983 by Kary B. Mullis (160). It is a simple yet powerful method that allows a particular fragment of DNA to be copied billions of times in short time, owning its inventor the Nobel Prize for Chemistry in 1993 (161). Various fields such as forensic sciences and medical field have used PCR in order to identify criminals and diagnose genetic diseases, respectively (162). In the current research, PCR was commonly used in plasmid engineering in order to construct protein expression cassettes.

Amplification of a specific sequence of DNA, which is a gene or segment of gene transcribing and translating into an enzyme or protein, is the first step of constructing a recombinant plasmid. A genomic DNA used in the PCR amplification, called DNA template, can be extracted and purified from tissues, microorganisms and other different sources. Different strains exist as a result of DNA sequence variations, which have similar but not identical genomic DNA information (162). In order to perform a PCR, oligonucleotide primers, relatively short nucleotide sequences, are designed. A pair of primers, forward and reverse, is commonly required for DNA amplification in which the sequence of the forward primer is complementary to the antisense DNA strand, while the reverse primer is complementary to the sense DNA strand. The primer is usually eighteen to thirty nucleotides long, and it may contain start or stop codons, and a restriction enzyme site. For restriction cloning, the vector must contain the same restriction enzyme sites in the

primers in which the gene of interest will be ligated into the vector later on in the process (163).

PCR is performed using a machine that provides rapid changes in temperature for each of the heating stages of the multiple reaction cycles based on the Peltier effect. Those stages include denaturation, annealing and extension. The DNA template is initially heated to a high temperature during denaturation stage, resulting in breakage of the hydrogen bonds between the complementary nucleotides of the double-stranded DNA, resulting in separation of both DNA strands. Next, the temperature is dropped to a specified temperature, allowing annealing of the oligonucleotide primers to their complementary DNA template sequences. Lastly, the temperature is raised again, usually to 72°C, allowing a DNA polymerase to extend the annealed primers during the extension stage, making a copy of the double-stranded DNA that is required to be amplified (162). Based on the number of cycles and the exponential nature of the process, the copy number of the amplified DNA can be up to billions at the end of the PCR experiment (160).

Various DNA polymerases are available for PCR. The choice of a specific polymerase is based on the experimental needs. Using DNA polymerases that have the feature of proofreading, such as Q5 (from New England Biolabs) and KOD (from Merck) (164), is mostly recommended during DNA amplification in the case of protein expression or cloning, as these polymerases have the capability to replace any newly introduced

non-complementary nucleotides with the appropriate ones prior to further extension of the sequence of the primer. This avoids the development of mutations in the DNA sequence and expression of the incorrect protein (165). Figure 2.1 shows a schematic summary of the DNA amplification using PCR.

Agarose gel electrophoresis is the most popular technique for PCR product analysis by separating the products based on their size and charge. Ethidium bromide (EtBr) is usually used to stain the PCR products due to its capability to intercalate between the two DNA strands which subsequently fluoresces when the complex absorbs at ultraviolet (UV), and thus the products can be visualised (162). After DNA amplification by PCR and confirmation of its results on an agarose gel, the amplified DNA and the selected plasmid are both double digested at the restriction sites using suitable restriction enzymes, making both digested products ready to be ligated together using T4 DNA ligase (163). Consequently, an engineered vector containing the gene of interest is the outcome of this experiment.

For the purpose of protein expression and based on the expression system, the engineered vector containing the gene of interest requires to be used to transfect mammalian cells or transformed into bacterial host cells. In the current research, various strains of *E. coli* were used as host cells for expression of recombinant proteins BabC and HomB.

Various types of PCR were performed in the present study, including normal PCR, asymmetric PCR, overlap extension PCR, and reverse transcription

polymerase chain reaction (RT-PCR). All PCRs were performed using a PTC-200 cycler (MJ Research, USA).

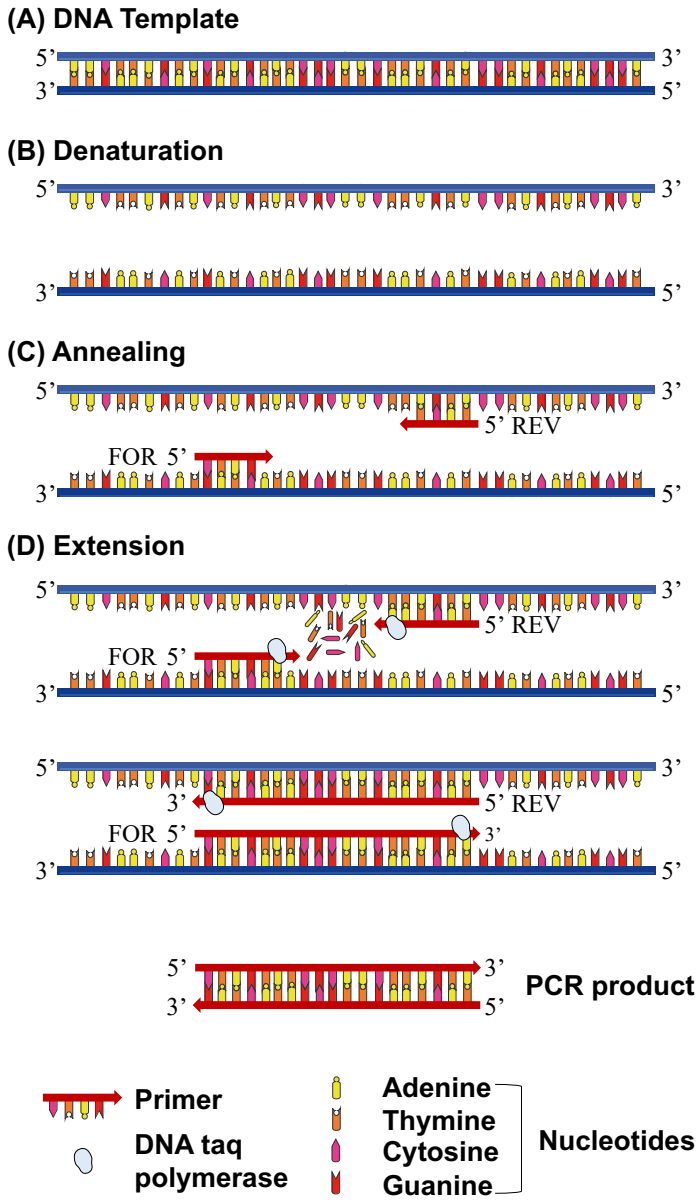


Figure 2.1: Schematic description of the three stages of PCR
(A) The two strands of the DNA template (B) are separated when heated to a high temperature during denaturation stage. (C) primers bind with specificity to the complementary DNA template strands, when the temperature is reduced to a specified temperature during annealing stage, and (D) then extend with the help of a DNA polymerase when the temperature rises again during extension stage.

2.2 Recombinant protein expression in *E. coli*

Since the first successful attempt to recombinantly express Somatostatin, a human peptide hormone, in *E. coli* in 1976 (166), major advances have been made. Till then, large amounts of biological materials had to be used to painfully extract and purify often small amounts of enzymes and proteins. Nowadays, biological macromolecules can be produced in large-scale following the development of recombinant protein production (167). Biochemical and structural characterisation of protein requires relatively large amounts of the protein to be studied, including studies of binding affinity and specificity, structure or thermal stability (168).

Recombinant proteins can be expressed using several different host systems, including mammalian, prokaryotic, plant, insect, and yeast cells (169), or even cell-free systems such as wheat germ, *E. coli* or rabbit reticulocyte lysates (170). Based on factors such as the aim, the physicochemical properties and required quality of the protein, the most suitable host system is chosen (171). Challenges can be faced when prokaryotic systems are used, as these organisms do not have the required mechanisms for post-translational modification, e.g. glycosylation. Furthermore, using prokaryotic cells for expressing complex protein folds containing formation of disulphide bond is considered a challenge (172). However, prokaryotic systems have many features which commonly render them the preferred host system for protein expression such as the use of

E. coli cells for protein expression, resulting in fast and high yield expression of the desired protein.

The most characteristic feature of *E. coli* cells is that they grow rapidly at a low price, with a cell division of each 20 minutes under optimal growth conditions (173). Further to the previous feature, compatible, safe strains and the range of commercially available vectors make *E. coli* cells an ideal host (174). Based on the choice *E. coli* strain to be used for protein expression, there are numerous plasmids available for the ligation of the gene of interest. There are factors required in order to choose a plasmid such as the preferred type of promoter and the multiple cloning site (MCS), which must have the same or compatible restriction sites present in the amplified gene introduced by the two primers (167).

In the current study, *E. coli* strains compatible with vectors where the protein expression is regulated with a *lac* operon were selected. The *lac* operon is composed of three different genes, *lacZ*, *Y* and *A*. β-galactosidase, encoded by *lacZ*, cleaves lactose into galactose and glucose and is responsible for allolactose production, which is the natural *lac* operon inducer. Lactose permease, encoded by *lacY*, is responsible for lactose uptake (175,176). Finally, transacetylase, encoded by *lacA*, appears to have a role in detoxification (176,177).

Protein expression is commonly induced by using isopropyl β-D-1-thiogalactopyranoside (IPTG), which binds the *lac* repressor and allows the three genes of *lac* operon to be transcribed (176).

Lysogeny broth (LB) medium is commonly used for *E. coli* growth because it does not contain fermentable sugars and *E. coli* cells thus use amino acids as a carbon source (173). This enables the transcription and translation of the cloned gene regulated with the *lac* operon only following IPTG addition to the LB medium.

Expression of protein occurs in the cytoplasm of *E. coli* cells; after expression, the protein can be located either in the cytoplasm or the periplasm of *E. coli* cells. Many proteins contain a specific N-terminal signal sequence which facilitates the translocate of the protein from the cytoplasm to the periplasm where the environment is oxidative which helps in the formation of the disulphide bonds allowing correct protein folding (178). Unlike periplasmic space, the reducing environment of the cytoplasm inhibits disulphide bond formation, resulting in formation of inclusion bodies (Figure 2.2) (178).

Factors, such as the rate of protein expression, can determine the localisation of the recombinant protein and in which compartment of the cell the protein will be found following the expression (167). However, endogenous proteins produced by the bacteria can also be found with the protein of interest in both compartments following IPTG induction. Therefore, it is necessary to purify the required protein from the endogenous proteins prior to any further investigations. In the current study, two types of protein purification were used, affinity chromatography (AC)/or immobilised metal

ion affinity chromatography (IMAC) and size exclusion chromatography (SEC).

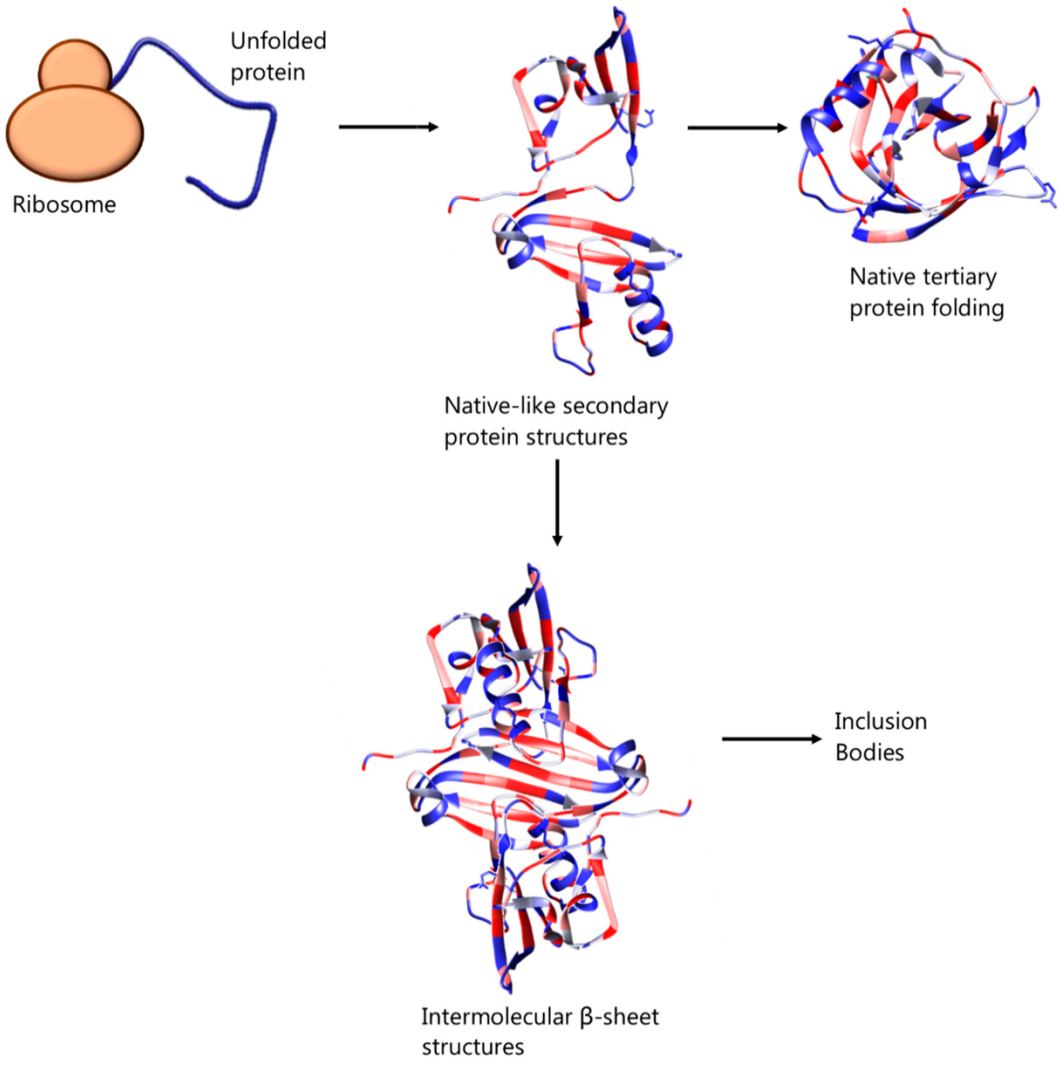


Figure 2.2: Suggested process of formation of inclusion bodies
Adapted from (179).

2.2.1 Expression plasmids used in this research

The functional characteristics of each plasmid used for the expression of recombinant BabC and HomB proteins are summarised below.

2.2.1.1 Plasmids used for periplasmic expression

2.2.1.1.1 pOPE101

pOPE101 plasmid contains a lac operator, which is induced by IPTG and located downstream of a synthetic P/A1/04/03 lac promoter, and c-Myc and hexa-histidine tags at the C-terminus for the purposes of immunodetection and protein purification, respectively. It also contains an ampicillin resistance marker for selection. Furthermore, it contains a pelB leader signal peptide (MKYLLPTAAAGLLLLAAQP) at the N-terminus to translocate the protein from the cytoplasm to the periplasm (Figure 2.3). This plasmid is suitable for expression of protein in *E. coli* (180). The map of pOPE101 vector is shown in Appendix Figure 2.

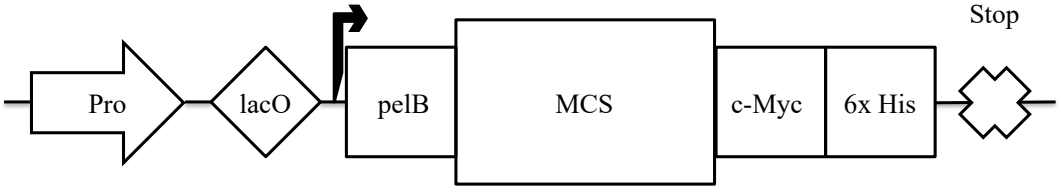


Figure 2.3: Schematic diagram of pOPE101 expression vector
The symbols and abbreviations denote the following: Pro – synthetic P/A1/04/03 lac promoter for initiation of expression; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence for transfer of expressed proteins to periplasmic space; MCS – multiple cloning site contains many restriction sites; c-Myc – c-Myc tag for immunodetection; 6x His – 6-histidine peptide tag for protein purification; and X/STOP – stop codon.

2.2.1.1.2 pET-22b(+)

pET-22b(+) plasmid contains a lac operator, which is induced by IPTG and located behind a T7 promoter sequence, and a hexa-histidine tag at the C-terminal region of the protein for the purposes of protein purification and immunodetection. It also contains an ampicillin resistance marker for selection. Furthermore, it contains a pelB leader signal peptide at the N-terminus to translocate the protein from the cytoplasm to the periplasm (Figure 2.4). This plasmid is suitable for expression of proteins in *E. coli* (181). The map of pET-22b(+) vector is shown in Appendix Figure 3.

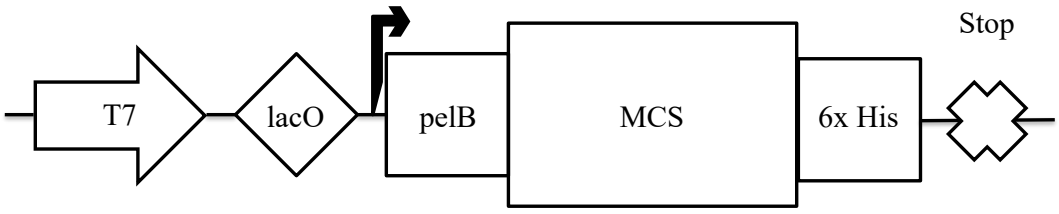


Figure 2.4: Schematic diagram of pET-22b(+) expression vector
The symbols and abbreviations denote the following: T7 – T7 promoter; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence for transfer of expressed proteins to periplasmic space; MCS – multiple cloning site contains many restriction sites; 6x His – 6-histidine peptide tag for immunodetection and protein purification; and X/STOP – stop codon.

2.2.1.2 Plasmids used for cytoplasmic expression

2.2.1.2.1 pET22b(ΔpelB)

The pelB leader sequence was deleted from the pET-22b(+) expression vector in order to express the protein in the cytoplasm of *E. coli*. NdeI should be used as a restriction enzyme to remove the pelB leader during digestion step (Figure 2.5).

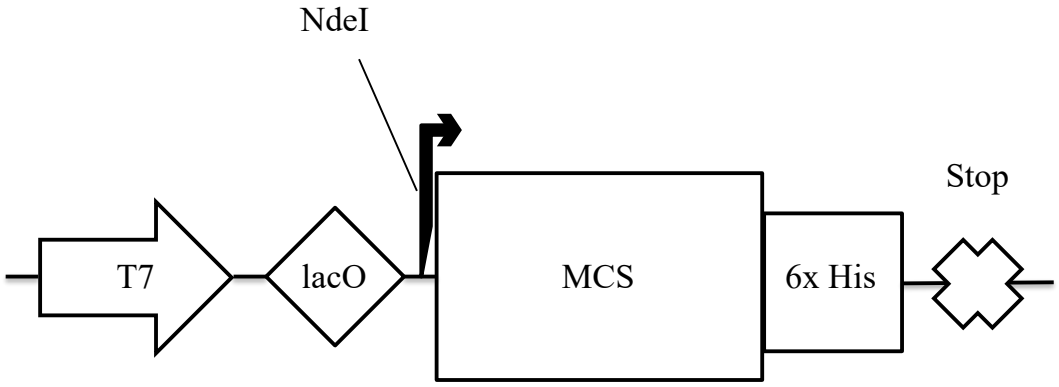


Figure 2.5: Schematic diagram of pET22b(ΔpelB) expression cassette
The symbols and abbreviations denote the following: T7 – T7 promoter; lacO – lac operator; arrow – start codon; MCS – multiple cloning site contains many restriction sites in which NdeI should be used to remove the pelB leader during digestion step; 6x His – 6-histidine peptide tag for immunodetection and protein purification; and X/STOP – stop codon.

2.2.1.2.2 pCold-GST

pCold-GST plasmid contains a lac operator, which is induced by IPTG and located downstream of a cold-shock Protein A (*cspA*) promoter, and upstream of a translation-enhancing element (TEE) consisting of five-codon sequence (MNHKV) to enhance translation initiation, a hexa-histidine tag for either protein purification or immunodetection, and GST, a 26 kDa protein produced by *Schistosoma japonicum*, which is used as a fusion protein for improvement of protein solubility, protein purification and immunodetection (182). It also contains an ampicillin resistance marker for selection. The vector also contains a sequence (IEGR), located upstream of the GST, which is recognised by factor Xa (FXa) protease which cleaves after this sequence, and a sequence (LEVLFQGP), located downstream of the GST, which is recognised by human rhinovirus 3C protease which cleaves after

glutamine (Figure 2.6). This plasmid is suitable for expression of protein in *E. coli* (183). The map of pCold-GST vector is shown in Appendix Figure 4.

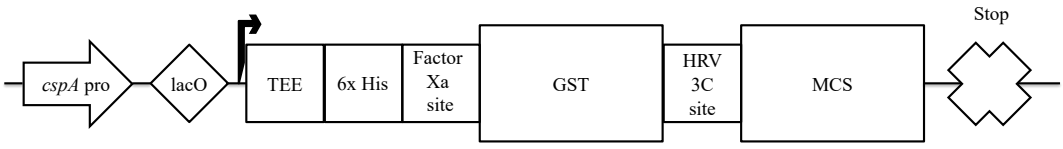


Figure 2.6: Schematic diagram of the pCold-GST expression vector
The symbols and abbreviations denote the following: *cspA* pro –*cspA* promoter; lacO – lac operator; arrow – start codon; TEE – a translation-enhancing element which is a five-codon sequence (MNHKV) to enhance translation initiation; 6x His – 6-histidine peptide tag can be used for immunodetection or protein purification; FXa site – a sequence (IEGR) which is recognised by FXa protease which cleaves after this sequence; GST – glutathione S-transferase used as a fusion protein for improvement of protein solubility, protein purification and immunodetection; HRV 3C site – a sequence (LEVLFQGP) which is recognised by human rhinovirus 3C protease that cleaves after glutamine; MCS – multiple cloning site contains many restriction sites; and X/STOP – stop codon.

2.2.1.2.3 pMAL-c5X

pMAL-c5X plasmid contains a lac operator, which is induced by IPTG and located downstream of a *tac* promoter, which is a strong *E. coli* promoter resulting from a hybrid between the tryptophan (*trp*) and *lac*UV5 promoters and gives high-level expression (184), and a maltose-binding protein (MBP), a 42 kDa protein encoded by *E. coli* K12 *malE* gene (185,186), used as a fusion protein for improvement of protein solubility (187-190), protein purification (191), and immunodetection (192). It also contains an ampicillin resistance marker for selection. The vector also contains FXa site downstream of the MBP fusion protein (Figure 2.7). This plasmid is suitable

for expression of protein in *E. coli* (193). The map of pMAL-c5X vector is shown in Appendix Figure 5.

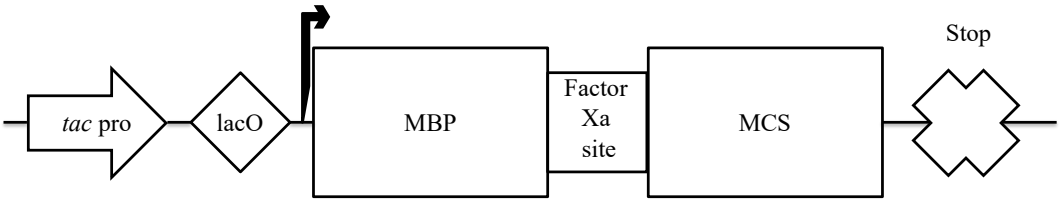


Figure 2.7: Schematic diagram of pMAL-c5X expression vector
The symbols and abbreviations denote the following: *tac* pro – *tac* promoter; lacO – lac operator; arrow – start codon; MBP – maltose-binding protein used as a fusion protein for improvement of protein solubility, protein purification and immunodetection; FXa site – a sequence (IEGR) which is recognised by FXa protease which cleaves after this sequence; MCS – multiple cloning site contains many restriction sites ; and X/STOP – stop codon.

2.2.2 *E. coli* strains used in this study

A brief summary of each *E. coli* strain used for both plasmid construction and protein expression is given below.

2.2.2.1 *E. coli* DH5α

This *E. coli* strain has mutations in *endA1*, which inhibits the degradation of plasmid DNA because of the inactivation of the intracellular endonuclease, and *recA*, which inactivates recombinases and inhibits the homologous recombination. Therefore, it is usually used for cloning (194).

2.2.2.2 *E. coli* XL10 Gold

This *E. coli* strain increases the efficiency of large DNA transformation because it exhibits the high transformation efficiency (Hte) phenotype (195). Furthermore, it has *endA1* and *recA* mutations. Therefore, it is usually used for cloning (195,196).

2.2.2.3 *E. coli* BL21 (DE3)

This *E. coli* strain is typically used for expression of protein in conjunction with expression vectors that contains T7 promoter. It contains a T7 RNA polymerase because it is a DE3 lysogen, which contains a lambda prophage in which the gene for T7 RNA polymerase is under control of the *lac*UV5 promoter (197). This polymerase binds to the expression vector through the T7 promoter in order to direct the cloned gene transcription. Furthermore, this strain is deficient in the proteases OmpT and Lon, which degrade outer and intracellular membrane proteins, respectively (198).

2.2.2.4 *E. coli* Origami B (DE3)

In addition to the characteristics of *E. coli* BL21 (DE3), *E. coli* Origami B (DE3) lacks the enzymes thioredoxin reductase (Δ*trxB*) and glutathione reductase (Δ*gor*), making an oxidising cytoplasmic environment which facilitates the formation of intramolecular disulphide bonds. Furthermore, this strain has mutation in *lacY1*, enabling a more precise control of expression rates by adjustment of the concentration of IPTG (198,199).

2.2.2.5 *E. coli* ArcticExpress (DE3)

This strain, which is derived from *E. coli* BL21-Gold cells, is characterised by improving the solubility of proteins that are difficult to express in soluble fractions and increase the yield of soluble protein. It can co-express Cpn10 and Cpn60, which are cold-adapted chaperonins, exhibit high activities of protein refolding at low temperatures during induction of protein expression, and thus they increase the yield of the soluble protein. The low temperature minimises the expression of endogenous bacterial proteins (200). Like *E. coli* BL21, this strain lacks the proteases OmpT and Lon (201).

2.2.2.6 Rosetta-gami B (DE3) pLysS

In addition to the characteristics of *E. coli* BL21 and *E. coli* Origami B, this strain has also the characteristic of the strain *E. coli* Rosetta which is enhancing protein expression of eukaryotes that contains six rare *E. coli* codons because it expresses rare transfer ribonucleic acids (tRNAs), which facilitate expression of gene which encodes these rare codons (202-205). pLysS is a plasmid that contains the LysS gene, which produce T7 lysozyme, a natural T7 RNA polymerase inhibitor (198).

By supplying rare codons, the Rosetta strains provide for “universal” translation, where translation would otherwise be limited by the codon usage of *E coli*.

2.3 ELISA

The ELISA was derived from radioimmunoassay (RIA), which was described around 1960, and developed to evade the health risks linked with the usage of radioactivity in RIA by replacing the radioisotope with an enzyme (206). Two independent groups described the ELISA in 1971 (207,208), and its principle is based on specific binding of antigen-antibody. Emitting fluorescence or generating colour is the final step of antigen detection, which can be either quantitative or qualitative (206).

ELISA is a rapid, simple, sensitive, versatile, and reliable test, with a detection limit as low as 10 pg/mL of protein antigens, depending on the affinity of the antibodies available (209). Generation of different categories of ELISAs have resulted when small alterations are applied to the major principle of ELISA, based on how the antigen is immobilised and detected. These categories include the direct, indirect, sandwich and competitive ELISA. The pros and cons of different kinds of ELISAs are described in detail in (210).

A sandwich-type ELISA technique was applied in the current study, where the protein antigen is trapped between a ligand and an antibody in the case of the ELISA for binding activity of BabC, which is different to the traditional sandwich ELISA that uses two antibodies performed for measuring IL-8. The high specificity for the detection of antigen is an advantage of this type of ELISA, where it does not need purification of the sample before the assay, as it includes washing steps after incubation with each component.

The initial step in this principle is the passive adsorption of the protein fragment of the neoglycoconjugates (ligand) or anti-IL-8 capture antibody on the 96-well plate surface through hydrophobic interactions between the specially treated plastic and the non-polar protein residues. Addition of a non-reacting protein, such as milk, rich in casein, or bovine serum albumin (BSA), is then performed in order to cover (or saturate) the surface of the plastic which has not been coated by the neoglycoconjugates or anti-IL-8 capture antibody; this prevents non-specific binding. Consequently, once the antigen-containing mixture is placed then, only the antigen binds specifically to its ligand (210).

For antigen detection, a specific primary antibody for the antigen and then a secondary antibody conjugated with an enzyme are sequentially incubated with the plate. Generation of colour or emission of fluorescence occurs when a substrate for the enzyme is added to the plate; this permits the detection of the desirable antigen (206).

After each step of incubation, phosphate buffered saline (PBS) supplemented with 0.05% Tween 20 (PBS-T) is used as washing buffer in order to eliminate any unbound antigen or antibody. This minimises the noise of background during detection and improves the specificity of the assay. Figure 2.8 shows a schematic representation of the assay. For quantitative purposes of ELISA assay, a standard curve must be simultaneously plotted using known concentrations of the analyte (210).

The assay can be used in medicine as a diagnostic tool, e.g. in the diagnosis of malaria, hepatitis B, and HIV or in pregnancy tests (210), but also in research as an analytical tool for the quantitative and qualitative detection of specific antibodies or proteins in a given sample (206).

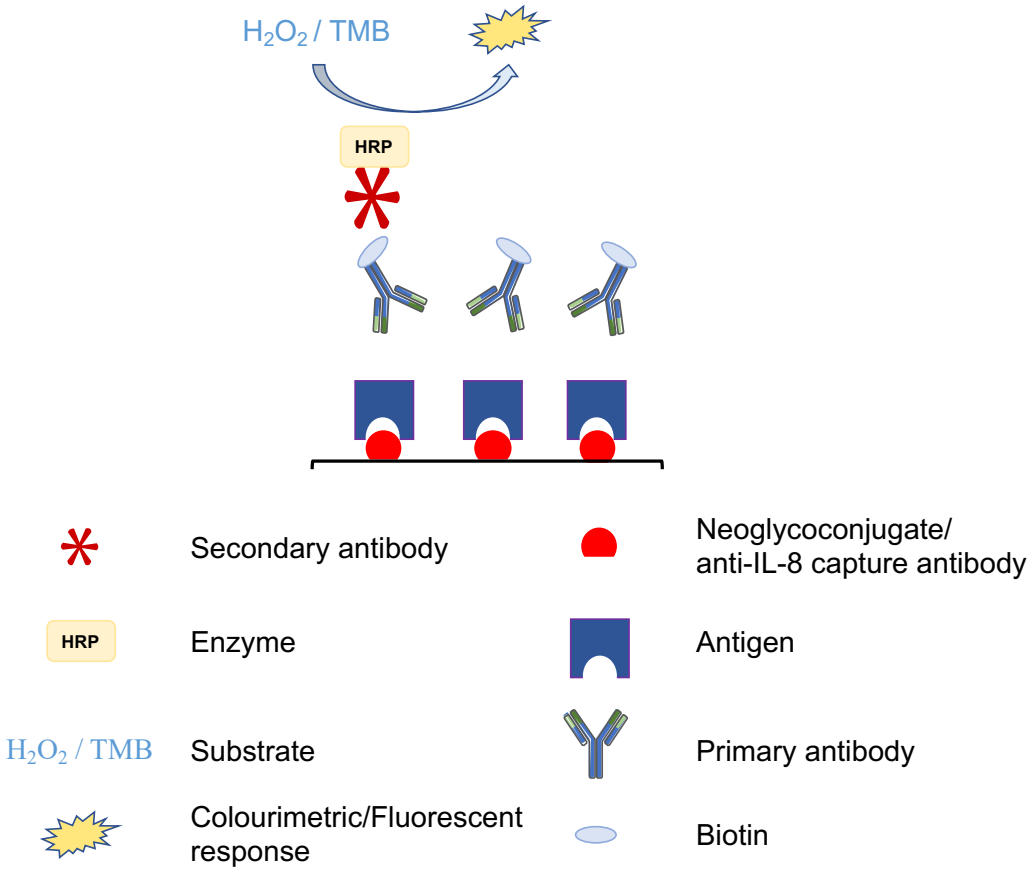


Figure 2.8: Schematic depiction of the ELISA principle (sandwich-type) used in the current study

A 96-well plate was used to immobilise the neoglycoconjugate or anti-IL-8 capture antibody that specifically binds the antigen available in a mixture of proteins. Sequential addition of a primary and a secondary antibody were performed for detection of antigen; a substrate emitting fluorescence or generating colour is needed for the enzyme to detect if antigen-antibody binding is available. HRP, H_2O_2, and TMB stand for horseradish peroxidase, hydrogen peroxide, and tetramethylbenzidine, respectively.

Chapter 3: Expression, purification and characterisation of BabC

3.1 Introduction

The first objective of this study was to recombinantly express and purify BabC, a putative adhesin used by *H. pylori* for colonising the gastric epithelium. Although *H. pylori* BabA had been successfully expressed in *E. coli* as a fusion protein using glutathione S-transferase (GST) in order to enhance the protein solubility (70), as a periplasmic protein (211), and as a protein refolded from an aggregation form (a 235 amino acid fragment of BabA) (212), all of these expression methods are not without disadvantages. A cleavage protease is required for GST-fusion removal prior to structural studies (213), low protein yield is associated with periplasmic expression (214), and the protein does not always regain its full function when the protein is refolded from the aggregation form using time-consuming methods (215,216).

On the other hand, Hage *et al*. (2015) tried to recombinantly express BabA using different expression vectors and either eukaryotic or prokaryotic host cells, in order to bypass the aforementioned disadvantages. They optimised the expression of BabA in the periplasmic space of *E. coli* (76), which was previously developed by Fei *et al*. (211) by inserting a polycationic hexa-lysine tag at the C-terminal region of BabA. The periplasmic expression is preferred over to the cytoplasmic expression because the environment of the periplasmic space is oxidative which helps in the formation of the disulphide bonds allowing correct protein folding, the

enhancement of protein solubility (178), and the ease of protein purification because periplasmic extracts contain less endogenous protein than cytosolic extracts (217). Thus, it would be of interest to express BabC adhesin in the periplasmic space of *E. coli* using the protocol of Hage *et al*., as BabC belongs to the same group of BabA in the Hop family.

The N-terminal extracellular domain of BabC was required for the purposes of the current study as it contains the putative ligand binding site of the protein. Similar to BabA, it is known that BabC belongs to the Hop family of *H. pylori* consisting of two domains: C-terminal transmembrane and N-terminal extracellular domains (Figure 3.1). The transmembrane domain, which is composed of antiparallel β-sheet structures that form a β-barrel unit, is highly conserved and considered to act as an autotransporter for the adhesin secretion (218,219). The hydrophobicity of the transmembrane domain, which may reduce the solubility of the protein during expression, was the reason behind designing suitable recombinant plasmids to express the N-terminal extracellular domain only. The predicted N-terminal extracellular domain of BabC contains eight cysteine residues, which may suggest that BabC contains up to four intramolecular disulphide bonds (Figure 3.2).

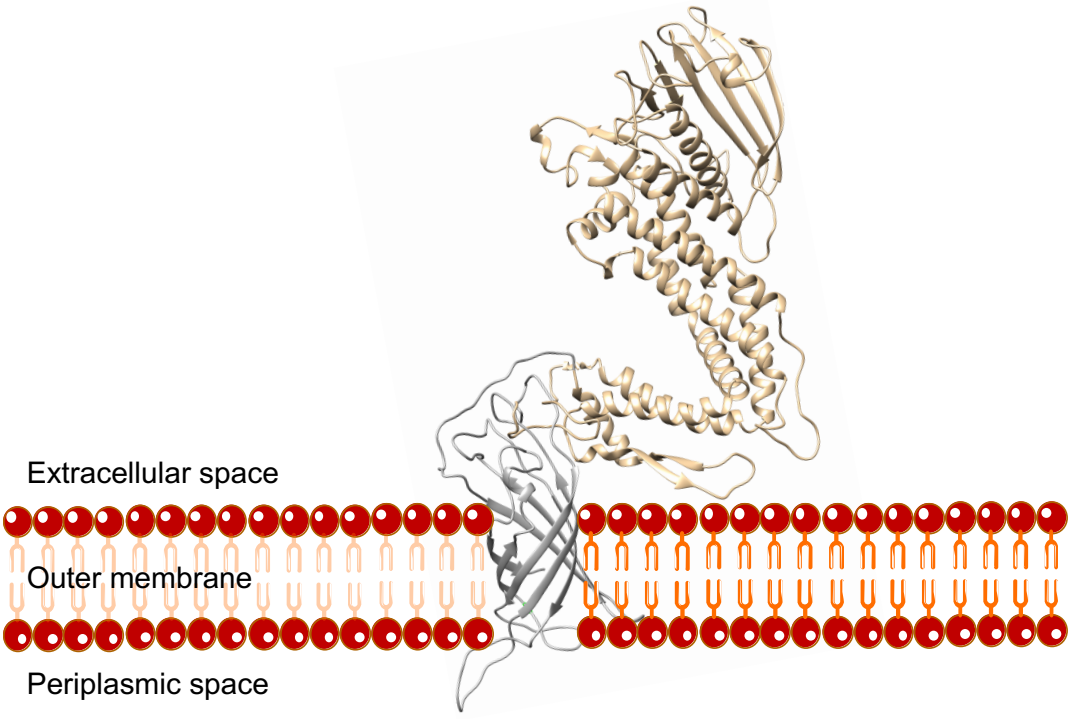


Figure 3.1: Schematic representation of the adhesin BabA J99
Phyre[2]-predicted crystal structure of BabA J99 (68) represents the N-terminal extracellular domain (gold) which is anchored to the outer membrane of *H. pylori* via the transmembrane domain (silver) of the protein.

```
  1 MKKHILSLAL GSLLVSTLSA EDDGFYTSVG YQIGEAAQMV TNTKGIQQLS DNYENLNNLL
 61 TRYSTLNTLI KLSADPSAIN AVRENLGAST KNLIGDKANS PAYQAVFLAI NAAVGLWNTI
121 GYAVMCGNGN GTESGPGSVI FNDQPGQDST QITCNRFEST GPGKSMSIDE FKKLNEAYQI
181 IQQALKNQSG FPELGGNGTK VSVNYNYECR QTADINGGVY QFCKAKNGSS SSSNGGNGSS
241 TQTTATTTQD GVTITTTYNN NKATVKFDIT NNAEQLLNQA ANIMQVLNTQ CPLVRSTNNE
301 NTPGGGQPWG LSTSGNACSI FQQEFSQVTS MIKNAQEIIA QSKIVSENAQ NQNNLDTGKP
361 FNPYTDASFA QSMLKNAQAQ AEMFNLSEQV KKNLEVMKNN NNVNEKLAGF GKEEVMTNFV
421 SAFLASCKDG GTLPNAGVTS NTWGAGCAYV GETISALTNS IAHFGTQEQQ IQQAENIADT
481 LVNFKSRYSE LGNTYNSITT ALSKVPNAQS LQNVVSKKNN PYSPQGIETN YYLNQNSYNQ
541 IQTINQELGR NPFRKVGIVN SQTNNGAMNG IGIQVGYKQF FGQKRKWGAR YYGFFDYNHA
601 FIKSSFFNSA SDVWTYGFGA DALYNFINDK ATNFLGKNNK LSLGLFGGIA LAGTSWLNSE
661 YVNLATVNNV YNAKMNVANF QFLFNMGVRM NLARSKKKGS DHAAQHGIEL GLKIPTINTN
721 YYSFMGAELK YRRLYSVYLN YVFAY
```

Figure 3.2: Amino acid sequence of BabC 26695
The amino acid sequence shows that the N-terminal extracellular domain of BabC contains six cysteine residues indicated by green circles, suggesting that BabC can have up to four intramolecular disulphide bonds. The amino acids in red colour are the signal peptide, while the underlined amino acids are the C-terminal transmembrane domain.

Following the successful expression of BabC, the next objective of this study was to recognise its putative receptor on the gastric epithelium, which has not been previously reported. It is known that BabA and SabA adhesins mediate the bacterial attachment to the epithelium of stomach by binding Le^b and $sLe^{a/x}$ glycan receptors, respectively. Most bacterial population diffuse in the mucous gel layer of the stomach, while only 20% live on the surface of the gastric epithelial cells (41). It has also been shown that LabA adhesin binds to lacdiNAc, a disaccharide found on MUC5AC gastric mucins and restricted to the gastric mucosa (41). Furthermore, HopQ adhesin binds to CEACAMs (60,61). As *H. pylori* has a variety of adhesins that recognise and adhere to different glycan and protein receptors expressed on the gastric epithelium and mucosa, it was therefore of interest to identify the binding receptor for BabC adhesin.

BabC was also characterised by assessing its stability against peptic digestion because of the acidity of the gastric fluid and the abundance of pepsin secreted in the stomach (220). The optimal activity of pepsin in order to hydrolyse the peptide bonds is within a range of pH 1-4, which matches the stomach's physiological environment. Around 90% of the maximum hydrolytic activity of pepsin is exhibited at pH 1.5, while the activity is decreased to around 35% at pH 4.5, and it is irreversibly inactivated at pH 8 (221,222). However, pepsin cannot hydrolyse all proteins (e.g. β-lactoglobulin) because the conformational folds of these proteins can make the peptic cleavage sites inaccessible (223). Unlike the harsh

environment of the stomach lumen, the more favourable microenvironment is the mucus layer which overlays the epithelium of the stomach. The pH gradient is the most characteristic feature of this area in which it is acidic at the interface of the lumen, and neutral at the surface of the epithelium because of the secretion of mucosal bicarbonate (224-226). The surface layer of the normal gastric mucus, which resists the activity of the luminal pepsin, mainly consists of the MUC5AC mucin (227,228), which may act as receptor for BabA because it contains type 1 ABO/Le blood group antigens (229-231). In the case of gastritis, SabA binds to sLe^x (59), which localises to the membrane region of epithelial cells of the stomach (75) (Figure 3.3). It is therefore important to assess the stability and susceptibility of BabC to the peptic digestion.

Three steps were required for the expression and purification of the recombinant BabC. First, the gene fragment, which encodes the N-terminal extracellular domain of BabC and obtained from PCR amplification, was cloned into the chosen vector. Next, small-scale expression was achieved following the protocol of BabA J99 expression (76). Finally, protein purification was carried out for larger scale expression for further investigations including identification of its binding receptor and evaluation of its susceptibility to peptic digestion.

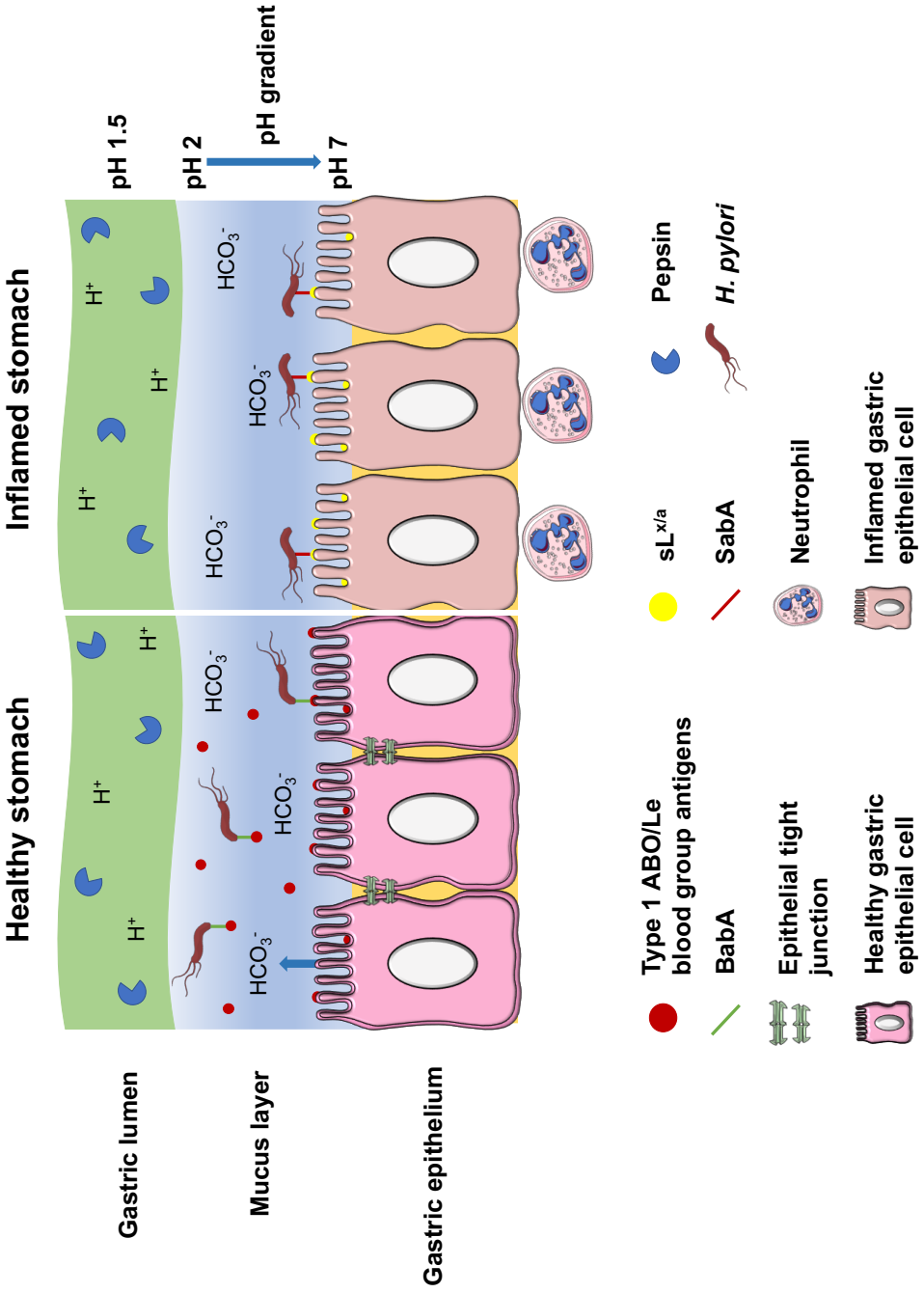


Figure 3.3: Schematic representation of the gastric lumen and mucus layer

In case of normal stomach, BabA adhesin binds to type 1 ABO/Le blood group antigens expressed in the epithelial cells and the mucus layer (MUC5AC), while SabA adhesin binds to sLe^x antigen, which localises to the membrane region of epithelial cells of the stomach, in the case of chronic gastritis. The gastric mucus layer is characterised by a pH gradient, due to the secretion of mucosal bicarbonate which acts as a barrier in order to prevent the activity of the luminal pepsin.

3.2 Materials and methods

3.2.1 Generation of BabC expression construct

The expression of more *H. pylori* adhesins was endeavoured in our group specially after the successful expression of *H. pylori* adhesins BabA and LabA by Naim Hage and Vasiliki Paraskevopoulou, respectively (76,109). In the current chapter, the N-terminal extracellular domain of BabC adhesin from *H. pylori* 26695 was chosen to be expressed and characterised, as the expression of this protein has not yet been reported. BabC expression construct engineered in this study is summarised in Table 3.1.

Table 3.1: Expression construct engineered in this study
babC$_{21\text{-}548}$ – *babC* gene fragment derived from *H. pylori* strain 26695 encoding amino acids 21-548.

Insert Gene	GenBank Accession No.	Vector	Cloning Cells
babC$_{21\text{-}548}$	AAD07380.1	pOPE101	*E. coli* XL10 Gold

3.2.1.1 Design of BabC expression construct

The initial step in designing BabC expression construct was the sequence determination of *babC* gene segment that would be amplified from the genomic DNA of *H. pylori* 26695 and cloned into the pOPE101 vector (Progen Biotechnik, Germany). This determination was performed based on the amino acid sequence alignment of BabC 26695 (GenBank accession no. AAD07380.1) with BabA J99 (GenBank accession no. AAD06409.1) (68) using CLC Main WorkBench 8 (Qiagen Bioinformatics, USA). The respective amino acid sequence was submitted to the structural prediction software

Phyre^2 (232) in order to obtain a predicted three-dimensional (3D) structure of BabC.

3.2.1.2 PCR amplification

The aim of this procedure was to amplify a segment of the *babC* gene (Genbank accession no. AAD07380.1"; Locus_tag = HP0317) from the genomic DNA of *H. pylori* 26695, kindly donated by Dr. Rob Delahay, the University of Nottingham (Genbank accession no. AE000511.1). The putative C-terminal transmembrane domain was excluded from the amplification based on the amino acid sequence alignment with BabA of *H. pylori* J99. Furthermore, *babC* gene encodes a twenty amino acid N-terminal signal peptide which is cleaved during BabC secretion to its outer membrane in *H. pylori* (219); this signal peptide was also excluded from the $babC_{548}$ gene fragment. Hexa-lysine tag was inserted directly at the C-terminus for protein solubility purposes.

Unfortunately, there was an internal restriction site in *babC* gene for NcoI, which is needed for cloning *babC* into pOPE101 (Appendix Figure 6). Therefore, a single nucleotide (adenine) was substituted with another (cytosine) without changing the amino acid (proline) produced.

3.2.1.2.1 Asymmetric PCR

In this method, three primers (Merck, UK) were used in two rounds of PCR amplification (Figure 3.4).

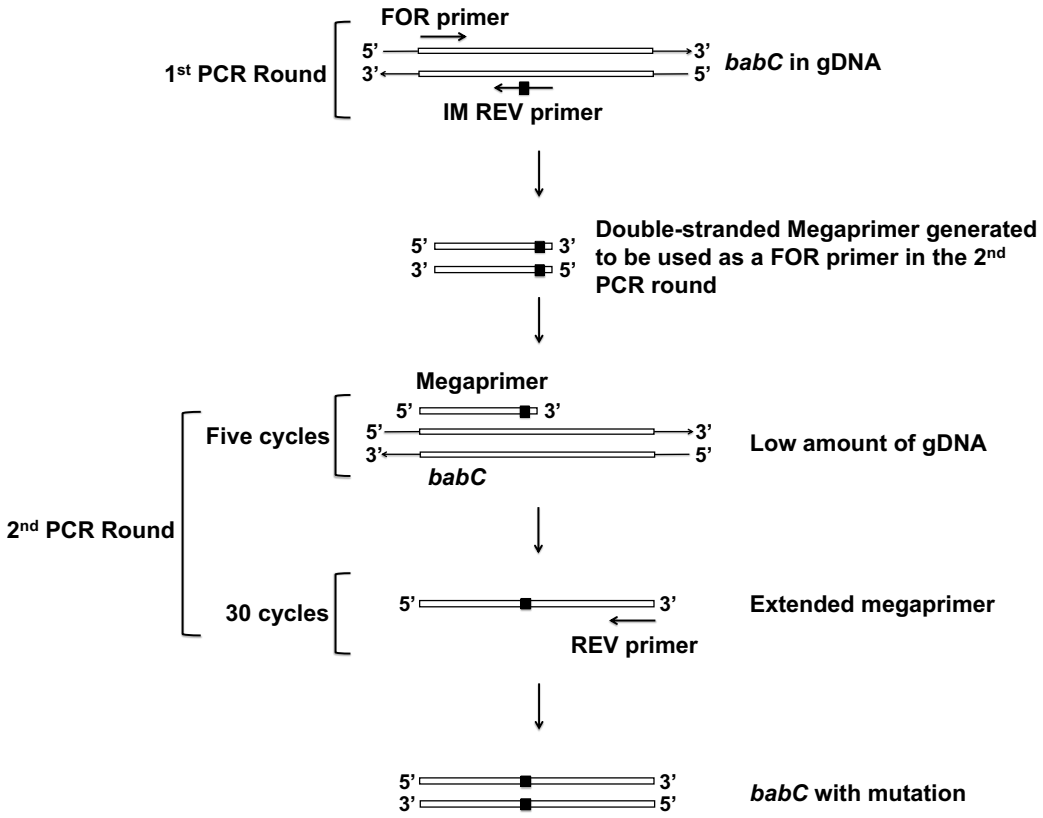


Figure 3.4: Schematic diagram of asymmetric PCR
The megaprimer product from the 1st PCR round was generated as double-stranded DNA using forward primer and internal reverse mutagenic primer removing the internal NcoI restriction site. Following the first round of PCR, the megaprimer was extracted and purified to be used as a forward primer in the 2nd PCR round. In the first five cycles of the 2nd PCR round, the megaprimer was extended in the presence of low template concentration. After the five cycles completed, a reverse primer was added to obtain the final *babC* gene with the desired mutation in the internal NcoI restriction site. FOR, REV, IM and gDNA denote forward, reverse, internal mutagenic and genomic DNA, respectively.

3.2.1.2.1.1 First round of PCR amplification

The purpose of this round was to generate a megaprimer (889 bp), which contains a single nucleotide mutation, in order to be used as a forward primer in the second PCR round. Two primers shown in Table 3.2A were used in this round, in which the reverse primer was internal mutagenic primer. The PCR reaction was done in a total volume of 50 µL. The reaction consisted of 25 µL Q5® High-Fidelity 2X Master Mix (New England Biolabs, USA) [which contains a proprietary broad-use buffer, 2 mM $MgCl_2$ and 0.2 mM dNTPs], 0.5 µM forward primer, 0.5 µM internal mutagenic reverse primer, and 3 ng genomic DNA of *H. pylori* 26695. The PCR was performed under the following conditions: 98°C for 30 sec of initial denaturation, followed by 35 cycles of template denaturation at 98°C for 10 sec, primer annealing at 65°C for 30 sec, and primer extension at 72°C for 30 sec. The reaction was finished with final extension at 72°C for 2 min.

The PCR product was loaded on 1% agarose gel (Merck, UK) and run in 0.5x Tris/Borate/EDTA (TBE) buffer. The gel was stained with SYBR Gold (Thermo Scientific, USA) after the electrophoresis run, and then visualised under blue-light transilluminator (Thermo Scientific, USA). The SYBR Gold was used instead of EtBr because it is safer and to avoid DNA mutation caused by the gel exposure to UV light. The band was then excised and purified using a QIAquick Gel Extraction Kit (Qiagen, USA), according to the manufacturer's protocol in order to be used in the second round of PCR as megaprimer. The DNA concentration was estimated by measuring the

absorbance at 260 nm using Nanodrop 2000 (Thermo Scientific, USA). Confirmatory sequencing of the megaprimer was performed (Source BioScience, Nottingham) using primers listed in Table 3.2A.

3.2.1.2.1.2 Second round of PCR amplification

In the second PCR round amplification, the reaction was also carried out in a total volume of 50 µL. It consisted of 10 µL megaprimer, 25 µL Q5® High-Fidelity 2X Master Mix, and low amount (0.5 ng/ µL) of genomic DNA of *H. pylori* 26695. Five cycles of asymmetric PCR were performed to extend the megaprimer under the following conditions: 98°C for 30 sec of initial denaturation, followed by denaturation at 98°C for 10 sec, and megaprimer annealing and extension in one step at 72°C for 1 min.

Following the five cycles, the reverse primer shown in Table 3.2B was added to the mixture to 0.5 µM for the next 30 cycles in which it would generate more copies of the extended megaprimer which contains the desired mutation, rather than amplifying the *babC* gene in the gDNA, which was only added in very low amount for that reason. During these cycles, denaturation of the extended megaprimer was at 98°C for 10 sec, primer annealing at 55°C for 30 sec, and primer extension at 72°C for 50 sec. The reaction finished with final extension at 72°C for 2 min. A band from the PCR product was then extracted from 1% agarose and purified by using a QIAquick Gel Extraction Kit. Confirmatory sequencing of the DNA was performed (Source BioScience, Nottingham) using primers shown in Table 3.2.

Table 3.2: Primers for *babC* amplification

FOR and REV denote forward and reverse primers, respectively. The number in the primer name indicates the 1st amino acid in BabC encoded by the nucleotides located downstream of the restriction site, which is underlined in the primer sequence.

[a] Primers used for 1st PCR round and confirmatory megaprimer sequencing.

[b] Primer used for confirmatory megaprimer sequencing only.

[c] Primers used for 2nd PCR round and confirmatory *babC* sequencing.

[d] Primer used for confirmatory *babC* sequencing only segment.

A. First round PCR

Gene	Primer Name	Primer Sequence (5'-3')
$babC_{21\text{-}548}$	BabC [21 FOR]_NcoI [a, b, d]	CCTGTCCCATGGAAGAC GACGGCTTTTAC
	BabC [internal mutagenic REV]_NcoI [a, b]	TGCTTAAACCCCAGGGTT GACCACCCCCTG
	BabC [FOR] [b]	GGCAAACATCATGCAAGT CC

B. Second round PCR

Gene	Primer Name	Primer Sequence (5'-3')
$babC_{21\text{-}548}$	Megaprimer [c]	889 bp fragment extracted from the first round of PCR
	BabC [548 REV]_BamHI [c, d]	TCTGCTGGATCCCTTCTT CTTCTTCTTCTTTAGTTCT TGGTTGATGGT

3.2.1.2.2 Amplification of sequences before and after *babC* gene

Because there are high homologies in the beginning and end of the sequence of *babA* and *babB* along with *babC*, another technique of PCR amplification was used in order to avoid amplifying both *babA* and *babB* with *babC*. This technique revolved around amplifying the sequences before and after *babC* gene in the non-coding 5'- and 3'- regions, in order to amplify *babC* without amplifying *babA* and *babB*. Although there are also similarities

between the sequences of the non-coding 5'- and 3'- regions of *babA*, *babB* and *babC,* primers were designed to amplify the desirable sequences only. The PCR reaction was performed in a total volume of 50 µL. The reaction consisted of 25 µL Q5® High-Fidelity 2X Master, 0.5 µM forward primer, 0.5 µM reverse primer shown in Table 3.3, and 3 ng genomic DNA of *H. pylori* 26695. The PCR was performed under the following conditions: 98°C for 30 sec of initial denaturation, followed by 35 cycles of template denaturation at 98°C for 10 sec, primer annealing at 52°C for 30 sec, and primer extension at 72°C for 100 sec. The reaction was finished with final extension at 72°C for 2 min.

Table 3.3: Primers for amplifying sequences of the non-coding 5'- and 3'- regions (before and after) of *babC* gene

FOR and REV denote forward and reverse primers, respectively. The number in the primer name denotes the position of the nucleotide in the genomic DNA of *H. pylori* 26695.

Gene	Primer Name	Primer Sequence (5'-3')
$babC_{21\text{-}548}$	BabC [334268] FOR	TATGGGCTAACACCACA
	BabC [330953] REV	CCTGAGTGAGATTAATTCCAT

3.2.1.2.3 Overlap extension PCR

Overlap extension PCR was the next step to get the full-length of *babC* gene. Two PCRs were carried out to generate a pair of DNA fragments with overlapping regions, while a third PCR was performed to obtain the full-length *babC*. The megaprimer was used as one of the DNA fragments. The second PCR generated the second segment using primers shown

in Table 3.4. The reaction consisted of 25 µL Q5® High-Fidelity 2X Master Mix, 0.5 µM forward primer, 0.5 µM reverse primer, and 5 ng of genomic DNA of *H. pylori 26695*. The PCR was performed under the following conditions: 98°C for 30 sec of initial denaturation, followed by 35 cycles of template denaturation at 98°C for 10 sec, primer annealing at 60°C for 30 sec, and primer extension at 72°C for 25 sec. The reaction completed after final extension at 72°C for 2 min. A band from the PCR product was then extracted from 1% agarose and purified by using Wizard® SV Gel and PCR Clean-Up System (Promega, USA), according to the manufacturer's protocol. The DNA concentration was estimated by measuring the absorbance at 260 nm using Nanodrop 2000.

The last PCR was performed in a total volume of 50 µL by mixing the megaprimer with the other segment that both share 98 bp in sequence (overlapping region). The reaction consisted of 25 µL Q5® High-Fidelity 2X Master Mix, 0.5 µM forward primer, 0.5 µM reverse primer shown in Table 3.5, and 30 ng of each segment. The PCR was performed under the following conditions: 98°C for 30 sec of initial denaturation, followed by 35 cycles of template denaturation at 98°C for 10 sec, primer annealing at 65°C for 30 sec, and primer extension at 72°C for 1 min. The reaction completed after final extension at 72°C for 2 min. A band from the PCR product was then extracted from 1% agarose and purified by using Wizard® SV Gel and PCR Clean-Up System, according to the

manufacturer's protocol. The DNA concentration was estimated by measuring the absorbance at 260 nm using Nanodrop 2000.

Table 3.4: Primers for amplification of the second *babC* segment
FOR and REV denote forward and reverse primers, respectively.

Gene	Primer Name	Primer Sequence (5'-3')
babC$_{280-548}$	BabC [280 FOR]	GGCAAACATCATGCAAGTCC
	BabC [548 REV]_BamHI	TCTGCTGGATCCCTTCTTCTT CTTCTTCTTTAGTTCTTGGTT GATGGT

Table 3.5: Primers for the overlap extension PCR
FOR and REV denote forward and reverse primers, respectively. The number in the primer name indicates the 1st amino acid in BabC encoded by the nucleotides located downstream of the restriction site, which is underlined in the primer sequence.

Gene	Primer Name	Primer Sequence (5'-3')
babC$_{21-548}$	BabC [21 FOR]_NcoI	CCTGTCCCATGGAAGACGAC GGCTTTTAC
	BabC [548 REV]_BamHI	TCTGCTGGATCCCTTCTTCTT CTTCTTCTTTAGTTCTTGGTT GATGGT

3.2.1.3 Cloning

Following PCR amplification, the product was extracted and purified after agarose gel electrophoresis, using Wizard® SV Gel and PCR Clean-Up System (Promega, USA). The DNA concentration was estimated by measuring the absorbance at 260 nm using Nanodrop 2000. The purified PCR product was then digested alongside the pOPE101 plasmid with restriction enzymes NcoI and BamHI (New England Biolabs, USA),

according to the manufacturer's protocol. After digestion, 3:1 insert to vector molar ratio was used to ligate *babC* gene into the pOPE101 expression vector using T4 DNA ligase (New England Biolabs, USA), under conditions specified by the manufacturer.

For plasmid propagation, 5 µL ligation mixture were transformed into 100 µL chemically competent *E. coli* XL10 Gold (Stratagene, USA), which is tetracycline resistant, through a heat shock method as follows: 30 minutes on ice, then 30 seconds at 42°C, followed by 2 minutes on ice. After the heat shock process, cells were grown in 900 µL super optimal broth with catabolite repression (SOC) at 37°C with shaking for one hour. For plasmid selection, transformants were plated on 100 µg/mL ampicillin-supplemented LB (LB_{amp}) (Merck, UK) agar and incubated at 37°C overnight.

Single colonies were picked up from the overnight growth plates for colony PCR through suspension in 20 µL LB_{amp}. Two primers listed in Table 3.6 were used in colony PCR, which spanned the MCS of pOPE101. REDTaq® ReadyMix™ (Merck, UK) was used for PCR amplification, according to the manufacturer's instructions; an annealing temperature of 58°C was used. The PCR products were screened for gene insertion via agarose gel electrophoresis stained with EtBr, and positive clones were subsequently cultured overnight in 5 mL LB_{amp} medium at 37°C. Plasmid extraction and purification was then carried out by using a plasmid miniprep kit (Qiagen, USA), according to the manufacturer's instructions. The concentration of purified plasmid was measured using Nanodrop 2000,

and confirmatory sequencing was obtained (Source BioScience, Nottingham) using primers shown in Table 3.6.

Table 3.6: Primers for colony PCR and confirmatory sequencing

FOR and REV denote forward and reverse primers, respectively. The numbers in the primer name denotes the corresponding base pairs (sense strand) in the pOPE101 or *babC* gene sequences.

[a] Primers used for colony PCR and confirmatory sequencing.

[b] Primers used for confirmatory sequencing only.

Vector	Primer Name	Primer Sequence (5'-3')
pOPE101	pOPE [27-46bp FOR] [a,b]	TTGACTTGTGAGCGGATAAC
	pOPE [1135-1154bp REV] [a,b]	ATGTGTCAGAGGTTTTCACC
	BabC [61-78 FOR] [b]	CCTGTCCCATGGAAGACGACG GCTTTTAC
	BabC [1627-1644REV] [b]	TCTGCTGGATCCCTTCTTCTTC TTCTTCTTTAGTTCTTGGTTGA TGGT
	BabC [840-859 FOR] [b]	GGCAAACATCATGCAAGTCC

3.2.2 Periplasmic expression in *E. coli* XL10 Gold

Small-scale expression was done following confirmation of the DNA sequences of the recombinant plasmid using the protocol for BabA expression (76) in order to assess its potential role. Like plasmid propagation, the *E. coli* XL10 Gold was also used for protein expression with pOPE101_*babC* vector (Table 3.7). 20 µL of a glycerol stock was inoculated into 10 mL of LB media supplemented with 100 µg/mL ampicillin (Merck, UK), and 12.5 µg/mL tetracycline (Merck, UK) (LB_{AT}), and incubated overnight at 37°C in a 50 mL centrifuge tube with shaking at 250 rpm. These antibiotics were added to the culture in order to inhibit the growth of non-transformed bacterial cells. 75 µL from the overnight culture was used to inoculate 15 mL LB_{AT} (1:200 dilution) in a 50 mL centrifuge tube and incubated at 37°C with with shaking at 250 rpm until the optical density of the culture at 600 nm (OD_{600}) reached around 0.6. Before inducing the culture with 0.1 mM IPTG at 24°C, a 1 mL was taken from the culture for a "total protein fraction" preparation (sample preparation method described in Section 3.2.2.1), for subsequent analysis. After 16 hours incubation with shaking, cellular fractions were prepared for analyses. In addition to a total protein fraction, periplasmic and spheroplast fractions were prepared (sample preparation method described in Section 3.2.2.1). Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) was done to analyse all protein fractions, and followed by Coomassie-staining and Western blotting, as described in Section 3.2.2.2.

Table 3.7: Expression construct and expression host used in this study

Construct Name	Vector Backbone	Expression Host
pOPE101_*babC*$_{21\text{-}548}$	pOPE101	*E. coli* XL10 Gold

3.2.2.1 Preparation of cellular fractions for analyses

3.2.2.1.1 Total protein fraction

In addition to the 1 mL that was taken from the culture before inducing the cells with 0.1 mM IPTG in order to prepare a "total protein fraction", another 1 mL was taken from the overnight culture after inducing the cells with the IPTG. The bacterial cells were harvested by centrifugation at 10,000 x *g* for 1 minute (25°C). The pellets were then resuspended in 100 µL of 100 mM Tris-Cl (pH 8.0), and 100 µL 2x non-reducing sample buffer (NRSB) containing 100 mM Tris-Cl (pH 6.8), 4% SDS, 20% glycerol, and 0.2% bromophenol blue. This mixture was heat-denatured at 100°C for 5 minutes before SDS-PAGE analysis.

3.2.2.1.2 Periplasmic and spheroplast fractions

A 10 mL sample of bacterial cells was taken from the overnight culture and harvested via centrifugation at 4,122 x *g* for 10 minutes (4°C). The pellet was shaken gently on ice for one hour after resuspension in 1 mL of cold spheroplast solution (1/10 culture volume), that is, 50 mM Tris-Cl (pH 8.0), 20% sucrose, 1 mM ethylenediaminetetraacetic acid (EDTA) and supplemented with cOmplete® EDTA-free protease inhibitor cocktail tablets (Roche, USA). This cell suspension was then centrifuged for 15 minutes

at 14,000 x *g* (4°C), and the supernatant was gathered as the first periplasmic extract. Thereafter, the pellet was resuspended, shaken on ice, and centrifuged again in 1 mL cold 5 mM $MgSO_4$ (1/10 culture volume). The supernatant was collected as the second periplasmic extract. Both extracts were then concentrated by adding 250 µL trichloroacetic acid solution (TCA) and centrifuged at 14,000 x *g* for 10 minutes (4°C). The supernatant was discarded, and the pellet was then washed twice with 200 µL cold acetone and centrifuged at 14,000 x *g* for 10 minutes at 4°C. After acetone evaporation, the pellet was then resuspended in 100 µL 100 mM Tris-Cl (pH 8) and 100 µL 2x NRSB. The sample was heat-denatured at 100°C for 5 minutes prior to SDS-PAGE analyses. The remaining pellet, which contains spheroplast that lack the periplasmic space, was resuspended in 100 µL PBS and then mixed with an equal volume of 2x NRSB. This fraction was also heat-denatured for SDS-PAGE.

3.2.2.2 Analysis of cellular fractions

3.2.2.2.1 SDS-PAGE

After sample preparation, 15 µL of cellular fraction samples were loaded into precast Tris-Glycine 4-20% TGX mini gel (Bio-Rad, USA), and run in a Mini-PROTEAN™ Tetra Cell (Bio-Rad, USA), according to the manufacturer's instructions. Two gels were run, one for Coomassie-staining and the other for Western blotting.

3.2.2.2.2 Coomassie-staining

After SDS-PAGE, the Tris-Glycine gel was stained with InstantBlue™ stain (Expedeon, USA) for 1 hour and destained with deionised water until perfect protein band resolution was visualised. Alternatively, in the case of unavailability of InstantBlue stain, which is safe, non-toxic and fast staining with a clear background, the gel was stained with Coomassie stain containing 0.1 % Coomassie Blue® in 10% acetic acid, 40% methanol, and 50% water for 1 hour, and then destained in 10% acetic acid, 40% methanol, and 50% water until the desired protein band resolution was visualised.

3.2.2.2.3 Western blot

Once the SDS-PAGE was completed, a Trans-Blot® Turbo™ Transfer System (Bio-Rad, USA) was used to blot the gel against a 0.2 µm nitrocellulose membrane using Trans-Blot Turbo™ Transfer Pack (Bio-Rad, USA). After blotting, the membrane was blocked and shaken with 5% non-fat milk powder diluted in blocking buffer, which is Tris-buffered saline (TBS) containing 0.1% Tween 20 (TBS-T), for 1 hour at room temperature. After the blocking, the membrane was incubated and shaken overnight at 4°C in blocking buffer containing a mouse anti-c-Myc clone 9E10 biotinylated antibody 'primary antibody' (AbD Serotec, USA) at a 1:5,000 dilution. After 3-4 times washing for 10 minutes with TBS-T in the next day, the membrane was incubated and shaken at room temperature for 1 hour with blocking buffer containing streptavidin conjugated with horseradish

peroxidase (Streptavidin-HRP) 'secondary antibody' (AbD Serotec, USA) at a 1:5,000 dilution. After 3-4 times washing for 10 minutes with TBS-T, the membrane was exposed to an enhanced chemiluminescent substrate Pierce™ ECL Western Blotting Substrate (Thermo Scientific, USA), which is peroxide solution and luminol enhancer solution. Imaging was taken under high binning mode by a FUJI-LAS 4000 charge coupled device camera (Fujifilm, Japan).

3.2.3 Protein purification

3.2.3.1 Larger scale expression and periplasmic harvest

After the small-scale expression, *E. coli* XL10 Gold cells containing pOPE101_*babC* construct were expressed on a larger scale. 40 µL of a glycerol stock was inoculated into 10 mL LB_{AT} and incubated with shaking at 37°C overnight. 2 mL of the overnight culture was used to inoculate 200 mL LB_{AT} grown in a 1 L notched glass Erlenmeyer flask in a total volume of 1000 mL. These inoculated media were incubated and shaken at 37°C until the OD_{600} reached around 0.6. The temperature was then decreased to 24°C, and the cultures were induced with the final concentration of 0.1 mM IPTG for 16 hours. Cells were harvested through centrifugation at 4,122 x *g* for 45 minutes (4°C). An osmotic shock method was applied to obtain the first periplasmic extract through resuspension of the pellet in cold spheroplast solution (1/10 culture volume) and shaking gently on ice for 1 hour. This suspension was centrifuged at 4,122 x *g* for 45 minutes. Following filtration of the supernatant through a 0.22 µm filter and obtaining the first periplasmic extracts, the pellet was resuspended again in cold 5 mM $MgSO_4$ (1/10 culture volume) and shaken gently on ice for 1 hour. After centrifugation at 4,122 x *g* for 45 minutes, the supernatant was passed through a 0.22 µm filter to obtain the second periplasmic extracts. cOmplete® EDTA-free protease inhibitor cocktail tablets were added to both extract solutions according to the manufacturer's instructions.

3.2.3.2 Immobilised metal ion affinity chromatography (IMAC)

High-resolution purification of histidine-tagged protein was performed by IMAC. The periplasmic extracts collected from the previous step were pooled together and loaded onto a 1 mL HisTrap™ HP column, containing precharged Ni Sepharose High Performance (GE Healthcare, USA) at a rate no greater than 1 mL/min. Using an ÄKTA™ start (GE Healthcare, USA), a gradient elution program was run at a flow rate of 1 mL/min. The wash/binding buffer consisted of 20 mM sodium phosphate (pH 7.4), 0.3 M NaCl, and 10 mM imidazole. The elution buffer consisted of 20 mM sodium phosphate (pH 7.4), 0.3 M NaCl, and 0.5 M imidazole. All fractions were then loaded onto a Tris-Glycine gel for SDS-PAGE analyses (Coomassie-staining and Western blotting). Following the analyses, the purified proteins were concentrated using Vivaspin 20 (Sartorius Stedim Biotech GmbH, Germany) with a membrane ultrafiltration of 10 kDa according to manufacturer's instructions. Therefore, the imidazole was removed from the buffer. The protein was then resuspended in 20 mM sodium phosphate (pH 7.4), and 0.3 M NaCl before storing at -80°C.

3.2.3.3 Quantification of the recombinant BabC

Following IMAC and imidazole removal, protein concentration was determined using a Pierce BCA Protein Assay Kit (Thermo Scientific, USA), following the manufacturer's protocol. Nanodrop 2000 was also used to determine the concentration of the recombinant BabC by measuring its

absorbance at 280 nm using an extinction coefficient of 41870 M^{-1} cm^{-1} (calculated from the amino acid sequence of BabC with the 6x lysine, c-Myc tag, and the 6x Histidine tag) (233).

3.2.4 ELISA for binding activity of BabC

A 100 µL of human serum albumin (HSA) (Merck, UK), HSA-Le^a, HSA-Le^b, and HSA-Le^y, and HSA-sLe^x glycoconjugates (Isosep AB, Sweden) were used to coat a 96-well Maxisorp® plate (Thermo Scientific, USA) and incubated at 4°C overnight at a concentration of 5 µg/mL in sodium carbonate buffer (pH 9.6). Once the plate was washed with PBS-T (0.05% Tween), it was blocked with 200 µL 2% non-fat dried skimmed milk in PBS at room temperature for 2 hours. Following washing again with PBS-T, sample wells were incubated at 4°C overnight with 100 µL of BabC at a concentration of 20 µg/mL in PBS-T. After washing, a mouse anti-c-Myc clone 9E10 biotinylated antibody diluted in PBS-T (1:2,000) was added to the sample wells and incubated at room temperature for 2 hours. The washing step was repeated again before incubating the sample wells with streptavidin-HRP (1:2,000 dilution) at room temperature for 2 hours. After washing again with PBS-T, wells were exposed 10 minutes to 100 µL of 0.1 mg/mL tetramethylbenzidine in sodium acetate buffer (pH 6.0) supplemented with hydrogen peroxide (H_2O_2) to a final concentration of 0.03% for chromogenic detection. The reaction was stopped by adding 50 µL 2.5 M H_2SO_4, and absorbance was then measured by using a Spark® 10M multimode microplate reader (Tecan, Switzerland) at 450 nm. BabA protein was used as positive control.

3.2.5 Peptic digestion test

Pepsin is the main protein-digesting enzyme in the stomach among the different gastric enzymes; hence, BabC's resistance to pepsin has been tested.

3.2.5.1 Prediction of pepsin cleavage sites

ExPASy PeptideCutter (Swiss Institute of Bioinformatics, (234)) was used to submit the amino acid sequences of BabC and β-lactoglobulin (Genbank accession no. AAI08214.1) in order to predict the potential pepsin cleavage sites at pH 1.3 and pH > 2. At pH 1.3, pepsin aims to cleave the peptide bonds with N-terminal leucine or phenylalanine residues. On the other hand, pepsin loses its strict specificity at pH > 2 and cleaves at sites with tryptophan, tyrosine, phenylalanine, or leucine residues at either N- or C-terminus (235). The predicted pepsin cleavage sites were mapped on the predicted structure of BabC and the crystal structures of β-lactoglobulin (PDB Accession code: 3BLG) (236).

3.2.5.2 Pepsin activity

The ratio of pepsin activity used in this study was 10 U per 1 µg of test protein. This ratio was chosen based on the recommendation of the US Pharmacopoeia (USP) and the National Formulary (NF) which evaluate the average activity of pepsin (237). This ratio was also used in other studies performed pepsin digestion (238,239). Pepsin (Merck, UK) obtained from

porcine gastric mucosa as a lyophilised powder with a specific activity ≥ 2,500 units/mg was used in this study using 10 U of pepsin activity/µg of BabC protein.

3.2.5.3 Peptic digestion

For peptic digestion test at pH 1.2, a single tube containing 20 µL of simulated gastric fluid (SGF) [0.084 N HCl (pH 1.2), 35 mM NaCl, and 50 U of pepsin] was warmed to 37°C before adding 1 µL of protein sample, at a concentration of 5 mg/mL. On the other hand, for peptic digestion test at pH 4.5, a single tube containing 20 µL of 50 mM KH_2PO_4 (pH 4.5) containing 50 U of pepsin was also preheated at 37°C before adding 1 µL of protein sample, at a concentration of 5 mg/mL. Once the protein was mixed with pepsin, the tubes were immediately incubated at 37°C for different durations of up to two hours. At required time points, the peptic digestion was quenched by adding 21 µL of 200 mM $NaHCO_3$ (pH 11). Zero time point sample was prepared by adding the quenching solution to the pepsin-containing buffer before the addition of the protein sample. The following proteins were tested:

- BabC
- β-lactoglobulin from bovine milk (Merck, UK)

3.2.5.4 Sample analysis

All the samples were mixed with 42 µL of 2x NRSB before heat denaturation at 100°C for 5 minutes. The samples were then loaded into precast Tris-Glycine 4-20% TGX mini gel (Bio-Rad, USA) and separated by SDS-PAGE to be analysed by using the following techniques:

- Gels containing BabC and β-lactoglobulin samples were stained with InstantBlue stain or Coomassie stain for the visualisation of the outcome of peptic digestion, as described in Section 3.2.2.2.
- The gel containing BabC samples was blotted onto a nitrocellulose membrane and analysed using the Western blotting method described in Section 3.2.2.2.

3.3 Results

3.3.1 Cloning a *babC* gene fragment into expression vector

PCR was required to amplify a fragment of *babC* gene from the genomic DNA of *H. pylori* 26695 in order to generate a BabC expression construct. This DNA fragment encoded only the amino acids 21-548, which represent the predicted N-terminal extracellular domain of BabC. The putative C-terminal transmembrane domain (amino acids 549-745) was excluded from the amplification based on the amino acid sequence alignment with BabA of *H. pylori* J99, which has structural information and an identified receptor (Figure 3.5).

3.3.1.1 Amplification of *babC* gene

Amplification of *babC* gene from the genomic DNA of *H. pylori* 26695 was the first step to generate BabC expression construct. Due to the presence of an internal NcoI restriction site in the coding region of the *babC* sequence, a single nucleotide (adenine) required to be substituted with cytosine without changing the amino acid (proline) encoded (CCA→CCC). Asymmetric PCR was performed for achieving site-directed mutagenesis from two rounds of PCR. To do this, two primers shown in Table 3.2A were used in the first PCR round to generate the megaprimer that contained a single nucleotide mutation in order to be used as a forward primer in the second PCR round.

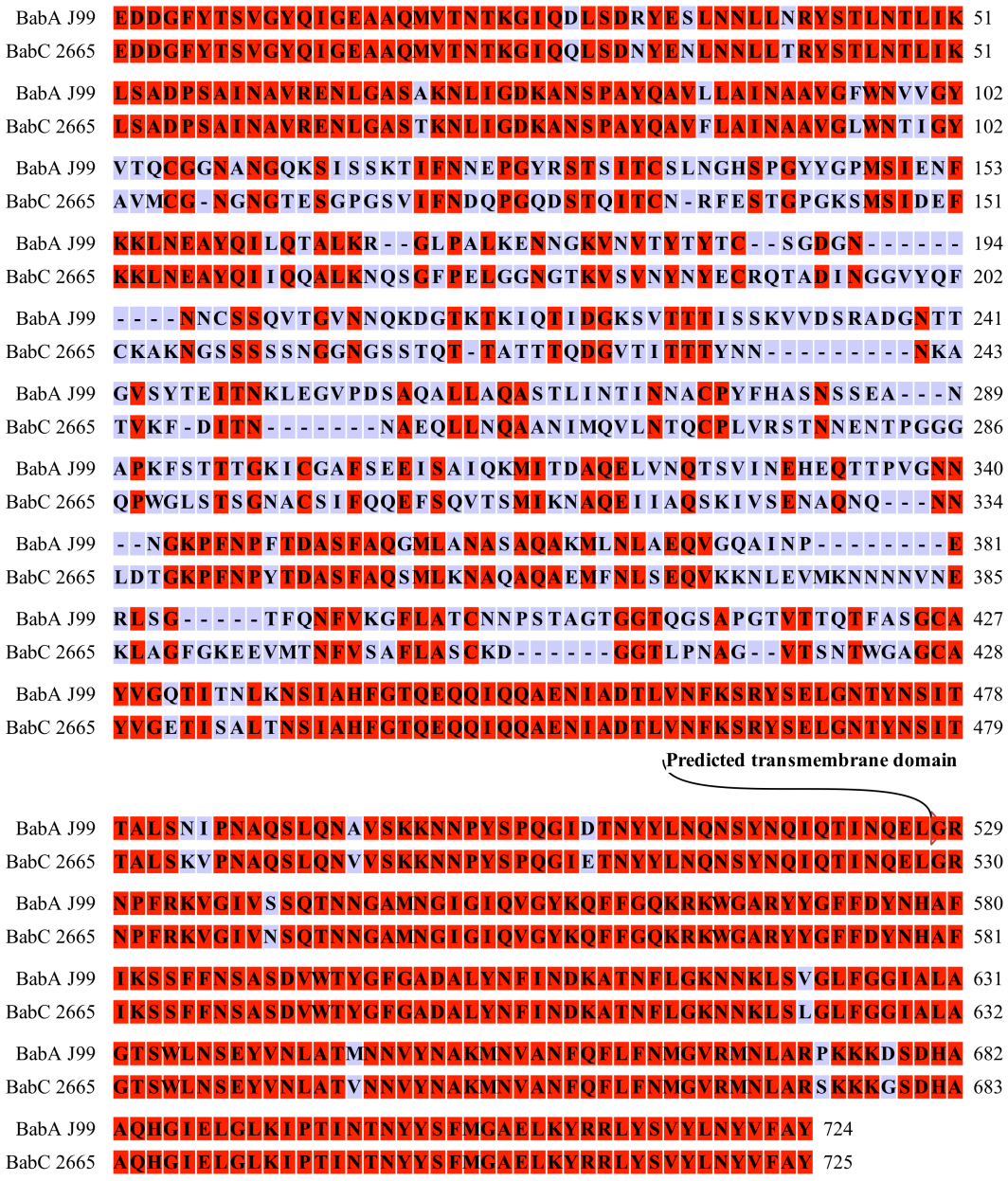


Figure 3.5: Alignment of amino acid sequences of BabA J99 and BabC 26695

BabC 26695 (GenBank accession no. AAD07380.1) and BabA J99 (AAD06409.1) share around 71% identity. The predicted C-terminal transmembrane domains of BabC 26695 and BabA J99 (indicated; starting at residue 548 of BabA) share around 97% sequence identity while the extracellular domains of both proteins share around 61% sequence identity. Red and blue boxshades highlight the identical and mismatched amino acids, respectively.

A single band at 889 bp, represented the megaprimer on agarose gel stained with SYBR Gold, was extracted and purified to produce pure DNA product, which was approximately 20.2 ng/μL (Figure 3.6A). DNA sequencing confirmed that the mutation had been successfully achieved (Figure 3.6B) and the megaprimer was ready for the second PCR round. The forward primer used to amplify *babC* gene introduced an NcoI restriction site, which facilitates its insertion into the MCS of the pOPE101 expression vector.

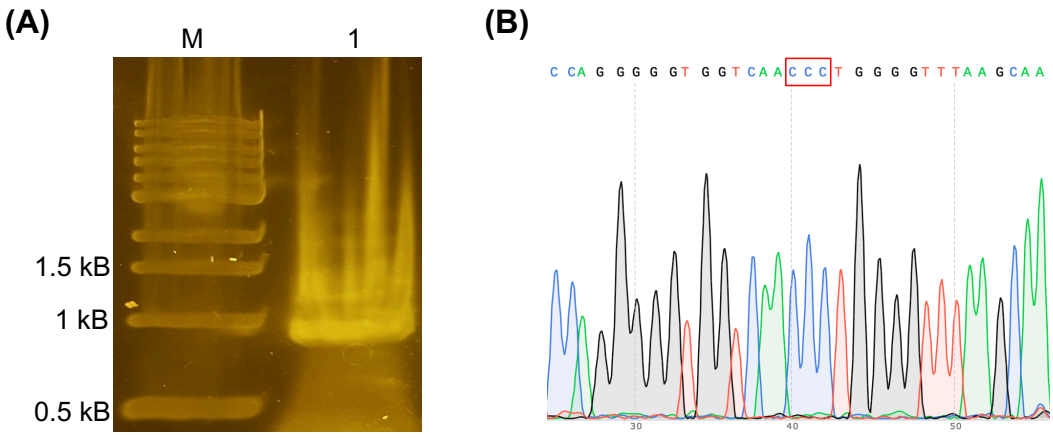


Figure 3.6: First round of asymmetric PCR
(A) Agarose gel stained with SYBR Gold representing the result of megaprimer following the first round of asymmetric PCR. The expected size of the megaprimer, which contains NcoI restriction site and six nucleotide overhangs at its 5' end, is 889 bp. Lanes: M – 1 kb DNA ladder (New England Biolabs, USA); 1 – megaprimer. (B) Representative DNA sequencing of the megaprimer and confirming the mutation in the red square by changing a single nucleotide without affecting the amino acid produced.

In the second PCR round of asymmetric PCR, particularly in the first five cycles, the megaprimer was used as a primer in order to anneal to the whole *babC* gene in the genomic DNA of *H. pylori* 26695, and then to extend to become a single strand (a full-length sense strand of *babC*) with mutation. After the five cycles, the addition of the reverse primer shown in Table 3.2B would amplify the extended megaprimer rather than amplifying the *babC* gene in genomic DNA that was added in very low amount for that reason. A band with an expected size at 1625 bp was visible on 1% agarose gel (Figure 3.7). However, DNA sequencing confirmed that there were many insertions in the second PCR product. Although asymmetric PCR is an easy and inexpensive method for achieving site-directed mutagenesis, it was nonetheless unsuitable for amplifying *babC* from genomic DNA because *babA*, *babB*, and *babC* show high homologies in the 5' and 3' regions causing the megaprimer to anneal to all of the three genes, resulting in many insertions (Appendix Figure 7).

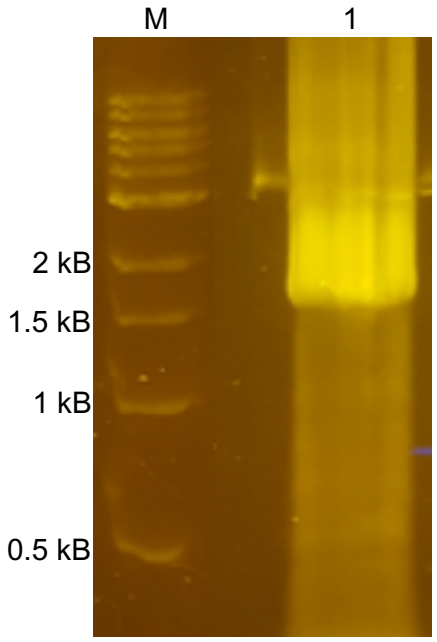


Figure 3.7: Second round of asymmetric PCR
Agarose gel stained with SYBR Gold representing the result of PCR product from the second round of asymmetric PCR. The PCR was unsuccessful as there were many insertions in the sequence resulting from annealing of the megaprimer to *babA*, *babB*, and *babC* that have high homologies in the 5' and 3' regions. Lanes: M – 1 kb DNA ladder (New England Biolabs, USA); 1 – a band at around 1625 bp as a result of the second round of asymmetric PCR.

Consequently, another strategy was used in order to avoid amplifying both *babA* and *babB* with *babC*. This strategy revolved around amplifying the non-coding untranslated sequences before 5'- and after 3'- ends of *babC* gene in order to amplify *babC* only, without amplifying *babA* and *babB*. Although there are similarities in non-coding untranslated sequences in the three genes, primers shown in Table 3.3 were designed to amplify the desired sequence only. However, this strategy failed with no PCR product obtained.

Therefore, overlap extension PCR was performed as a third strategy to amplify *babC* gene only. Three PCRs were carried out: two PCRs to generate two DNA fragments with overlapping regions, and the third PCR was performed to gain a full-length *babC* through mixing these two DNA fragments. The megaprimer, which was obtained from the asymmetric PCR, was used as one of the DNA fragments (1st PCR). The second PCR was done to amplify the other DNA fragment using the primers listed in Table 3.4. The forward primer was not mutant as there are also similarities between *babA*, *babB*, and *babC* in the sequences around the desired mutation. Thus, this primer was designed to amplify the other segment from *babC* gene only but without mutation. After successful amplification of the second fragment of 835 bp (Figure 3.8A), the third PCR was done by mixing the megaprimer and the other DNA fragment to obtain the full-length *babC* (1625 bp) (Figure 3.8B). Primers shown in Table 3.5 were used in the third PCR in which the forward primer was the forward primer in the first PCR (megaprimer amplification) while the reverse primer was the reverse primer in the second PCR (the second DNA fragment amplification). Both primers introduced restriction sites which facilitates its insertion into the MCS of pOPE101 expression vector. Interestingly, 50% of the third PCR products contained mutation as the megaprimer contained mutation while the other segment did not have mutation (Figure 3.9).

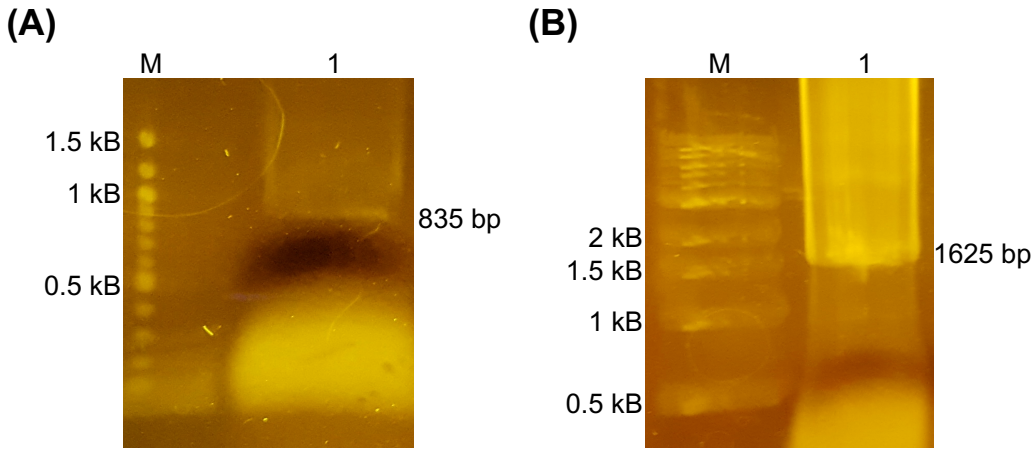


Figure 3.8: Overlap extension PCR

(A) agarose gel stained with SYBR Gold showing the second DNA fragment (853 bp) generated from the second PCR which shares 98 bp with the megaprimer (the first DNA fragment) but without mutation. Lanes: M – 100 bp DNA ladder (New England Biolabs, USA); 1 – the second DNA fragment (835 bp) for overlap extension PCR. (B) Full-length *babC* was obtained from the third PCR and seen on agarose gel stained with SYBR Gold. Overlap extension PCR resulted in 50% of *babC* with mutation. Lanes: M – 1 kb DNA ladder (New England Biolabs, USA); 1, full-length *babC*. The expected size of the full-length *babC*, which contains six nucleotide overhangs at its 5' and 3' ends and NcoI and BamHI restriction sites, is 1625 bp.

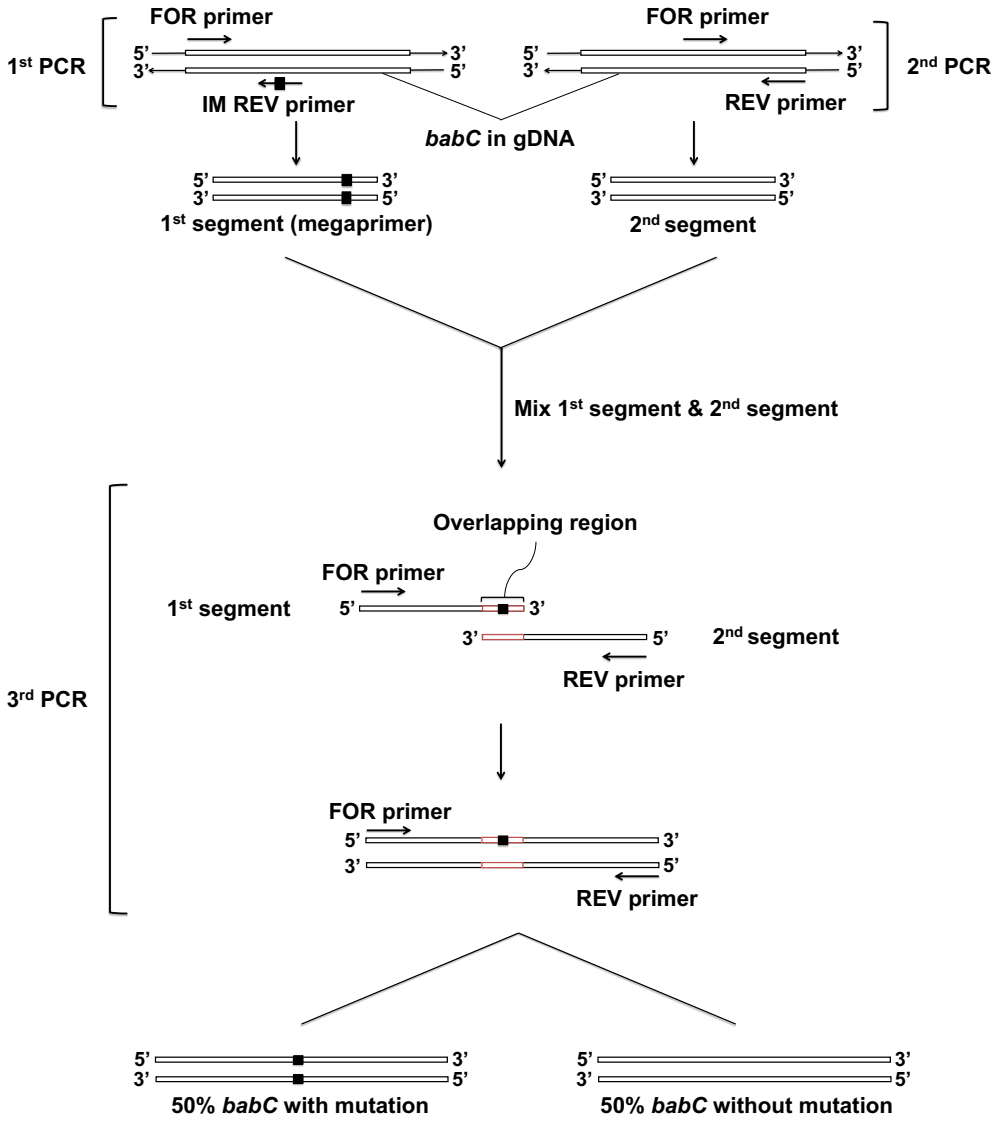


Figure 3.9: Schematic diagram of overlap extension PCR

Three PCRs were achieved to acquire a mutation within *babC* gene through overlap extension PCR in order to avoid the internal digestion with NcoI needed for cloning the gene into pOPE101. Two PCRs were carried out to generate two DNA fragments with overlapping regions, in which the megaprimer (produced in the first PCR) that contained mutation, was used as one of the DNA fragments. The second PCR was performed to amplify the other DNA fragment using primers that are unique for *babC*. However, this DNA fragment did not contain the desired mutation. The third PCR was done by mixing the megaprimer and the other DNA fragment, which share overlapping region, to obtain the full-length *babC*. Two primers were used in the third PCR in which the forward primer was used in the first PCR (megaprimer amplification) while the reverse primer was used in the second PCR (the second DNA fragment amplification). FOR, REV, IM and gDNA denote forward, reverse, internal mutagenic and genomic DNA, respectively.

After PCR amplification, purified *babC* and pOPE101 were then double-digested with NcoI and BamHI (Figure 3.10). In addition to the digestion of restriction sites added at both 5' and 3' ends by both NcoI and BamHI, NcoI internally digested the 50% of *babC* without the mutation, resulting in two bands (at the sizes of 866 bp and 741 bp) seen on agarose gel (Figure 3.10A). Furthermore, a band in agreement with the expected molecular size was seen on the agarose resulting from the digestion of the 50% of *babC* with mutation by NcoI and BamHI at the restriction sites at both 5' and 3' regions.

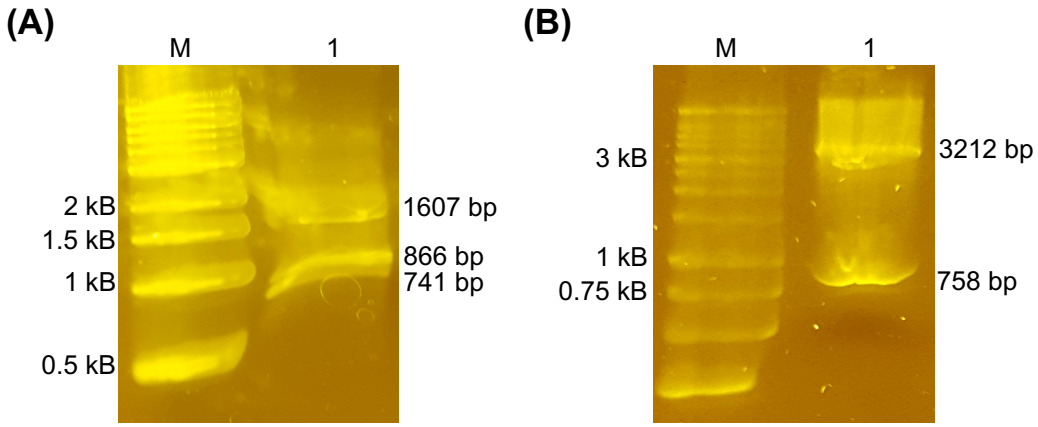


Figure 3.10: *babC* gene and pOPE101 after digestion by NcoI and BamHI

Agarose gel stained with SYBR Gold representing the results of digestion step for *babC* and pOPE101. (A) It shows that 50% of *babC* without mutation were digested internally with NcoI into two fragments 866 bp and 741 bp in addition to the digestion of restriction sites added at both the 5' and 3' ends by both NcoI and BamHI. Both NcoI and BamHI also digested 50% of *babC* with mutation at the restriction sites added at both the 5' and 3' ends to produce a band at 1607 bp. Lanes: M – 1 kb DNA ladder (New England Biolabs, USA); 1 – digested *babC* gene. (B) Two bands appear on agarose gel at the sizes of 3212 bp and 758 bp after digestion of pOPE101 by NcoI and BamHI at the restriction sites. Lanes: M – 1 kb DNA ladder (Promega, USA); 1 – digested pOPE101.

Following ligation and transformation into *E. coli* XL10 Gold for plasmid propagation, colony PCR was performed to confirm whether transformants harboured pOPE101_*babC* recombinant plasmid. The primers shown in Table 3.6 were designed for colony PCR in order to span pOPE101 MCS, therefore, the size of amplicon revealed whether colonies were positive or negative for *babC* insertion. It seems that the colonies were positive for *babC* insertion into pOPE101 as the expected band size was 1977 bp (Appendix Figure 6). Following plasmid purification, it was confirmed that the plasmid extracted from the positive transformant had 100% sequence identity to the desired recombinant plasmid and was cloned in frame with the pelB leader. Through this process, cloning of $babC_{21\text{-}548}$ gene fragment into pOPE101 vector was successful and the construct was transformed into *E. coli* X10 Gold. After sequence confirmation, the same *E. coli* strain was used for protein expression. Therefore, small-scale protein expression was required for further investigative purposes.

3.3.2 Periplasmic expression in *E. coli* XL10 Gold

After the successful expression of BabA in the periplasmic space of *E. coli* XL10 Gold cells using the pOPE101 vector, these cells were used for assessing the expression of the pOPE101_*babC* construct. Hexa-lysine tag was inserted directly after the inserted gene. The pOPE101_*babC* construct contains a synthetic P/A1/04/03 lac promoter and a pelB leader were present upstream of the inserted *babC* gene and are responsible for initiation of expression and transfer of the expressed proteins to the periplasmic space, respectively, while a c-Myc tag and a hexa-histidine tag were located downstream the inserted *babC* gene and they are useful for immunodetection and protein purification, respectively (Figure 3.11).

After inducing the cells with 0.1 mM IPTG and overnight incubation at 24°C, BabC was successfully expressed and secreted into the periplasmic space as demonstrated by the presence of the recombinant protein in the periplasmic fractions through SDS–PAGE and visualised with both InstantBlue staining and Western blotting (Figure 3.12). The expected molecular weight of BabC with the three tags (hexa-lysine tag, c-Myc tag, and hexa-histidine tag) was around 60 kDa. This successful small-scale expression was used as optimisation before moving forward for expression of BabC on a larger scale.

After the successful BabC expression on a small-scale, the protein was expressed on a larger scale.

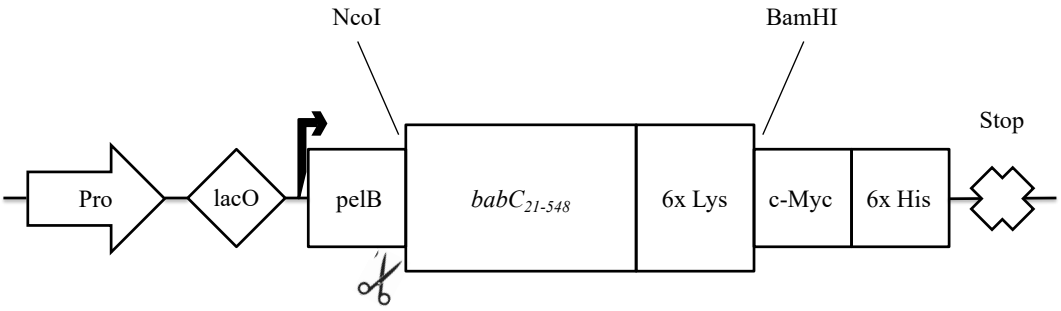


Figure 3.11: Schematic diagram of the pOPE101_*babC* expression cassette

pOPE101 was used for cloning *babC* through NcoI and BamHI restriction enzyme sites. The symbols and abbreviations denote the following: Pro – synthetic P/A1/04/03 lac promoter for initiation of expression; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence for transfer of expressed proteins to periplasmic space; *babC* – *babC* gene fragment; 6x Lys – hexa-lysine tag for improvement of protein expression and purification; c-Myc – c-Myc tag for immunodetection; 6x His – 6-histidine peptide tag for protein purification; and X/STOP – stop codon.

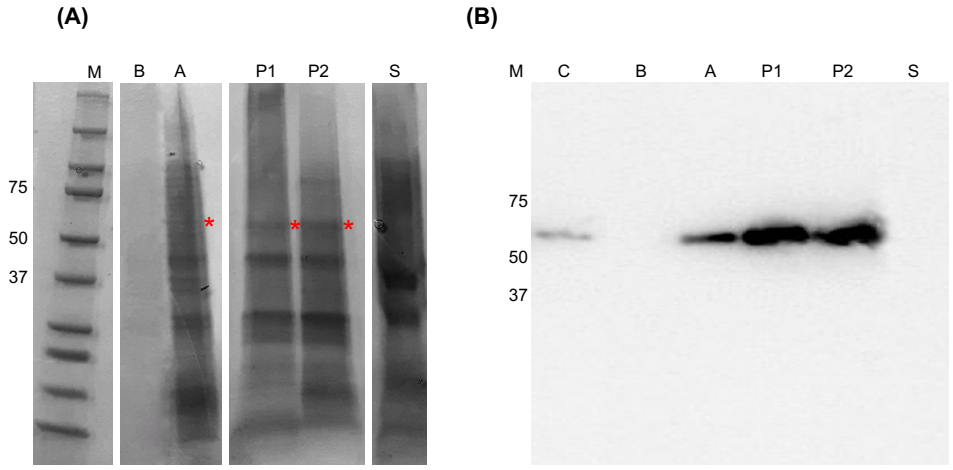


Figure 3.12: Expression of recombinant BabC in *E. coli* XL10 Gold cells

InstantBlue-staining gel (A) and Western blot using c-Myc immunodetection (B) showing the localisation of BabC in the periplasmic fractions following overnight expression in *E. coli* XL10 Gold cells at 24°C. The expected molecular weight of BabC is around 60 kDa and indicated by the red stars. The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 0.1 mM IPTG; A – total protein fraction after inducing the cells with 0.1 mM IPTG; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; S – spheroplast protein fraction; and C – BabA, which contains c-Myc tag, used as positive control (~59 kDa).

3.3.3 Purification of recombinant BabC

After expressing BabC on a small-scale, larger scale expression and purification were performed for further investigation. IMAC, an initial purification step, was used to purify BabC from the periplasmic extracts of *E. coli* XL10 Gold cells through the hexa-histidine tag present in the pOPE101_*babC* construct (Figure 3.13). The purity of BabC protein was relatively high, following this method, as determined by InstantBlue staining (Figure 3.14). However, at least one additional chromatographic step, e.g. SEC, is needed later for further purification, particularly for crystallisation purposes. After staining with InstantBlue, five IMAC-elution fractions were collected together, based on the clear bands in InstantBlue-staining gel, using Vivaspin 20 with a membrane ultrafiltration of 10 kDa in order to concentrate BabC protein and to eliminate imidazole. BabC protein was then resuspended in the binding buffer without imidazole before storing at -80°C for further investigations. The yield of the obtained protein was 2.5 mg per 1 L of bacterial culture. Western blot was then performed for c-Myc immunodetection and the band was at the expected molecular weight (Figure 3.15).

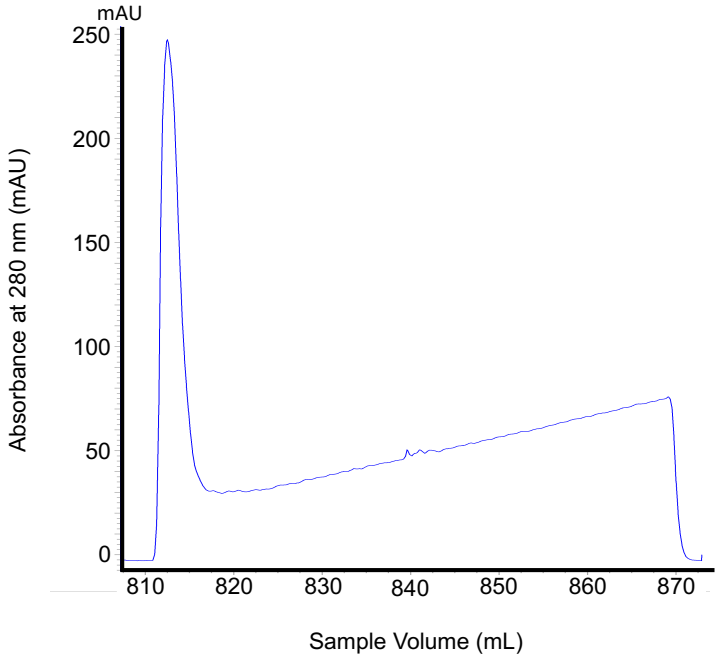


Figure 3.13: IMAC purification of BabC
Overlay of IMAC trace of BabC. The peak appeared after elution of the protein by a gradient of 500 mM imidazole.

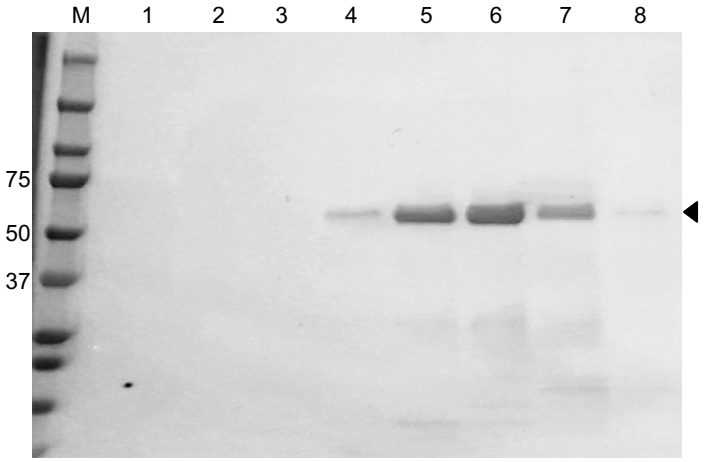


Figure 3.14: Purification of BabC by IMAC
8 IMAC-elution fractions, representing the peak from the IMAC purification of BabC (fractions 5-12), were collected after protein purification. Equal loading volume was used in all gel wells, and proteins were separated through SDS-PAGE and analysed by InstantBlue stain. The black arrowhead indicates the recombinant BabC (60 kDa). The abbreviations denote the following: M – molecular weight marker in kDa; 1-8 – IMAC-elution fractions in which BabC protein appeared clearly in 5 IMAC-elution fractions.

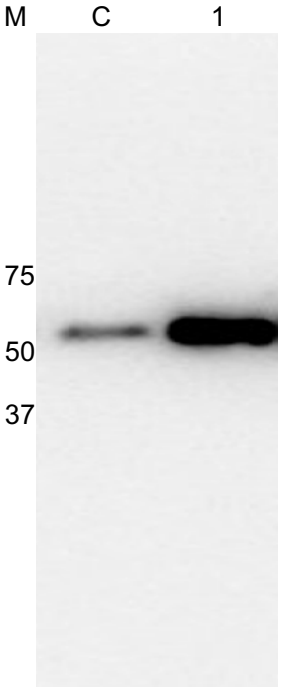


Figure 3.15: c-Myc immunodetection of BabC after purification by IMAC
The five IMAC-elution fractions with clear bands in the InstantBlue-staining gel were collected together, concentrated, and then analysed by Western blot after protein separation through SDS-PAGE. The abbreviations denote the following: M – molecular weight marker in kDa; C – BabA that contains c-Myc tag was used as positive control (~59 kDa); 1 – BabC (~60 kDa).

3.3.4 Binding Activity of BabC

As a primary test, BabC interactions with glycan epitopes commonly found in gastrointestinal mucins were tested in a more accessible glycoconjuaget format were assessed. HSA-Le^a, HSA-Le^b, HSA-Le^y, and HSA-sLe^x glycoconjugates, which were immobilised to a 96-well Maxisorp® plate, were assayed by using ELISA test. BabA protein, which is known to bind to Le^b, was used as positive control. Both BabA and BabC were assayed at a concentration of 20 μg/mL. The proteins binding to Lewis antigen glycoconjugates was detected by noticeable colour change (from colourless to blue) as well as by absorbance reading at 450 nm after targeting the c-Myc tag found in the C-terminus through using peroxidase-labelled antibodies. BabC protein did not show any binding activity towards the above glycoconjugates, while BabA bound very specifically to HSA-Le^b (Figure 3.16). This experiment was repeated twice for confirmation. Carbohydrate microarray analysis was considered the next necessary step in order to determine a glycan receptor for BabC.

The importance of the crown in BabA J99 is that it contains eight residues, which are C189, G191, N194, N206, D233, S234, S244 and T246, mediate hydrogen bonds with Fuc1, Fuc4, GlcNAc3, and Gal5 residues of Le^b (68). Gal5 residue interacts with the carboxyl and hydroxyl groups of D233 and S234 side chains respectively through water-mediated hydrogen bonds. Furthermore, GlcNAc3 residue interacts with the carboxyl group of D233 through hydrogen bond. It has also been shown that D233A/S244A

substitutions result in loss of binding affinity of BabA to Le^b. Differential scanning fluorimetry and circular dichroism spectroscopy revealed that D233A/S244A substitutions did not alter the global structure of BabA. Analysis of amino acid sequence of the crowns of BabA J99 and BabC 26695 shows that the residues D233 and S234 in BabA crown are missing in the amino acid sequence of the predicted BabC crown in which three hydrogen bonds will not be mediated with Le^b; two hydrogen bonds between D233 and Gal5/GlcNAc3 and one hydrogen bond between S234 and Gal5. Based on this analysis, binding of BabC to Le^b may not happen (Figure 3.17).

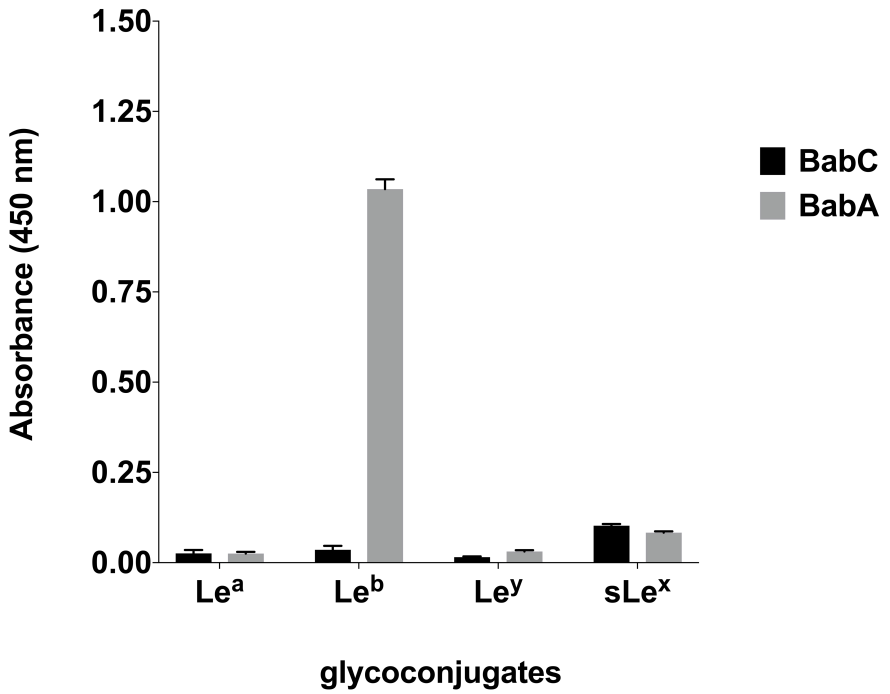


Figure 3.16: Binding of BabC to Le^a, Le^b, Le^y, and sLe^x in an ELISA
Binding of BabC to immobilised HSA-Le^a, HSA-Le^b, HSA-Le^y, and HSA-sLe^x glycoconjugates was assayed using an ELISA test. It shows that BabC do not have any binding to the previous glycoconjugates. BabA, which binds to Le^b, was used as positive control. Both BabA and BabC were assayed at a concentration of 20 µg/mL. Each value represents the mean ± SE of 3 independent experiments.

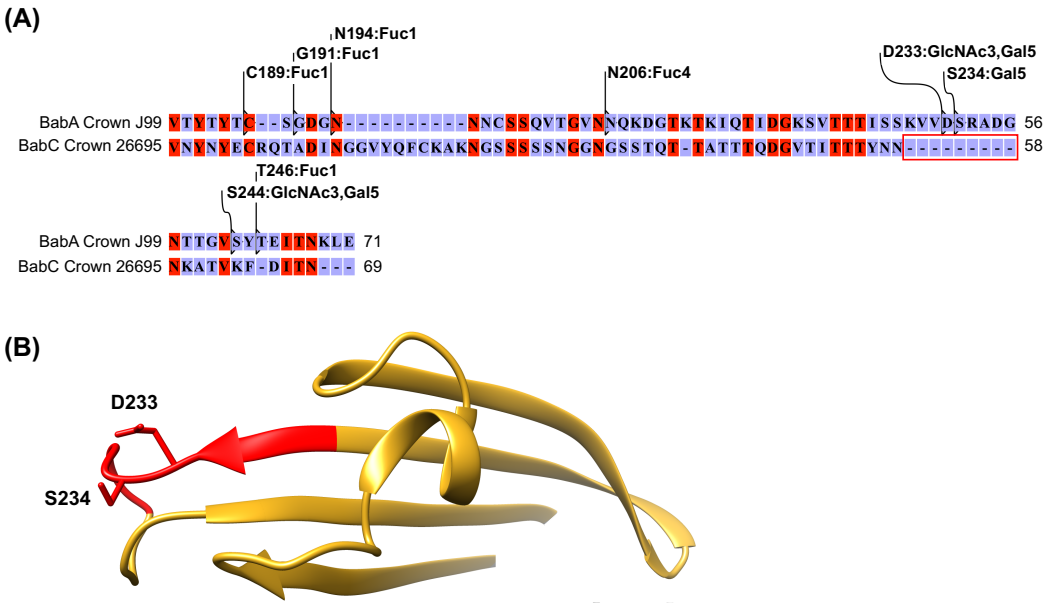


Figure 3.17: Comparison between the crowns of BabA J99 and BabC 26695

(A) Amino acid sequence alignment of the crowns of BabA J99 and BabC 26695. BabA shows significant binding with Le^b glycoconjugates through its crown. Amino acids of BabA crown involved in the formation of hydrogen bonds for each residue (Fucose, galactose, and N-acetylglucosamine residues) of Le^b are indicated. (B) The crystal structure of BabA crown with the missing amino acids (indicated in red colour) in BabC crown which contains two amino acid residues (D233 and S234) partially responsible for BabA binding to Le^b.

3.3.5 Resistance of BabC to peptic degradation

In order to examine its resistance to peptic digestion, BabC was assessed in parallel with bovine β-lactoglobulin, which is resistant to peptic digestion *in vitro* (223). Both proteins showed various predicted pepsin cleavage sites through analysis of amino acid sequences at pH 1.3 and pH > 2. These predicted pepsin cleavage sites were mapped on the predicted structure of BabC 26695 (Figure 3.18).

Despite the existence of several predicted pepsin cleavage sites, determined using ExPASy PeptideCutter tool (234), it was hypothesised that because of steric hindrance, many cleavage sites would be inaccessible once the protein was correctly folded. To determine this experimentally, peptic digestion for BabC and β-lactoglobulin was done at pH 1.2 and pH 4.5 in variable durations where samples were analysed by SDS-PAGE and visualised with InstantBlue following incubation with pepsin.

BabC was rapidly degraded by pepsin at pH 1.2 after 10 seconds and completely digested after 30 seconds (Figure 3.19A), while it was fully degraded at pH 4.5 after 5 minutes (Figure 3.19B). The degradation was characterised by BabC fragment generation. No intact BabC was observed after the time indicated above. On the other hand, as expected, β-lactoglobulin was completely resistant to peptic digestion throughout the two hours of incubation with pepsin (Figure 3.19C).

Immunodetection analysis was also performed for BabC peptic digestion at pH 4.5 for further investigation of the apparent C-terminal cleavage.

The analysis indicated that none of the observed fragments contained c-Myc tag at the C-terminal region. Furthermore, no intact BabC contained the C-terminal c-Myc tag after 15 seconds (Figure 3.20).

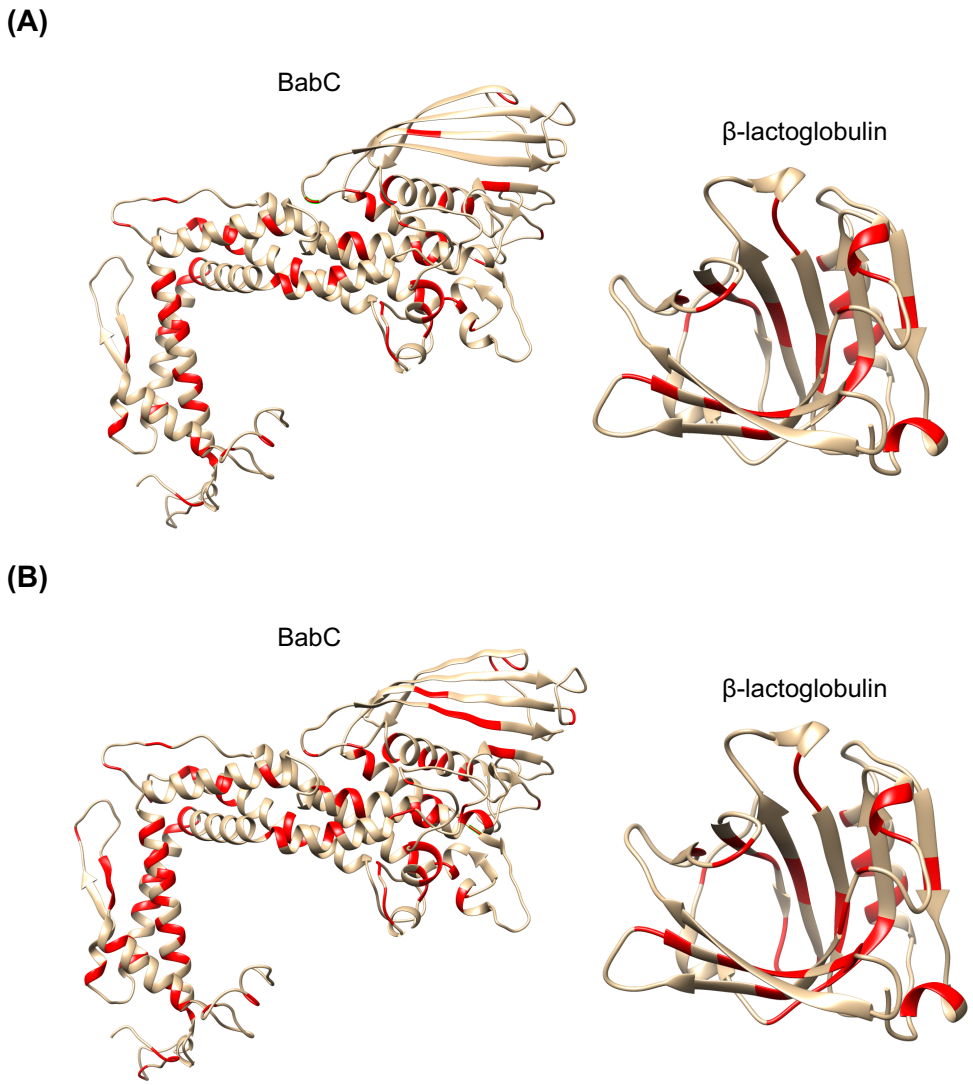


Figure 3.18: Predicted structure of BabC and crystal structure of β-lactoglobulin showing potential pepsin cleavage sites
The potential pepsin cleavage sites at pH 1.3 (A) and at pH > 2 (B) were mapped on the Phyre[2]-predicted structure of BabC (93% confidence) and the crystal structure of β-lactoglobulin and indicated in red colour. The cleavage sites were expected based on the possible cleavage of the peptide bond of the amino acid sequence.

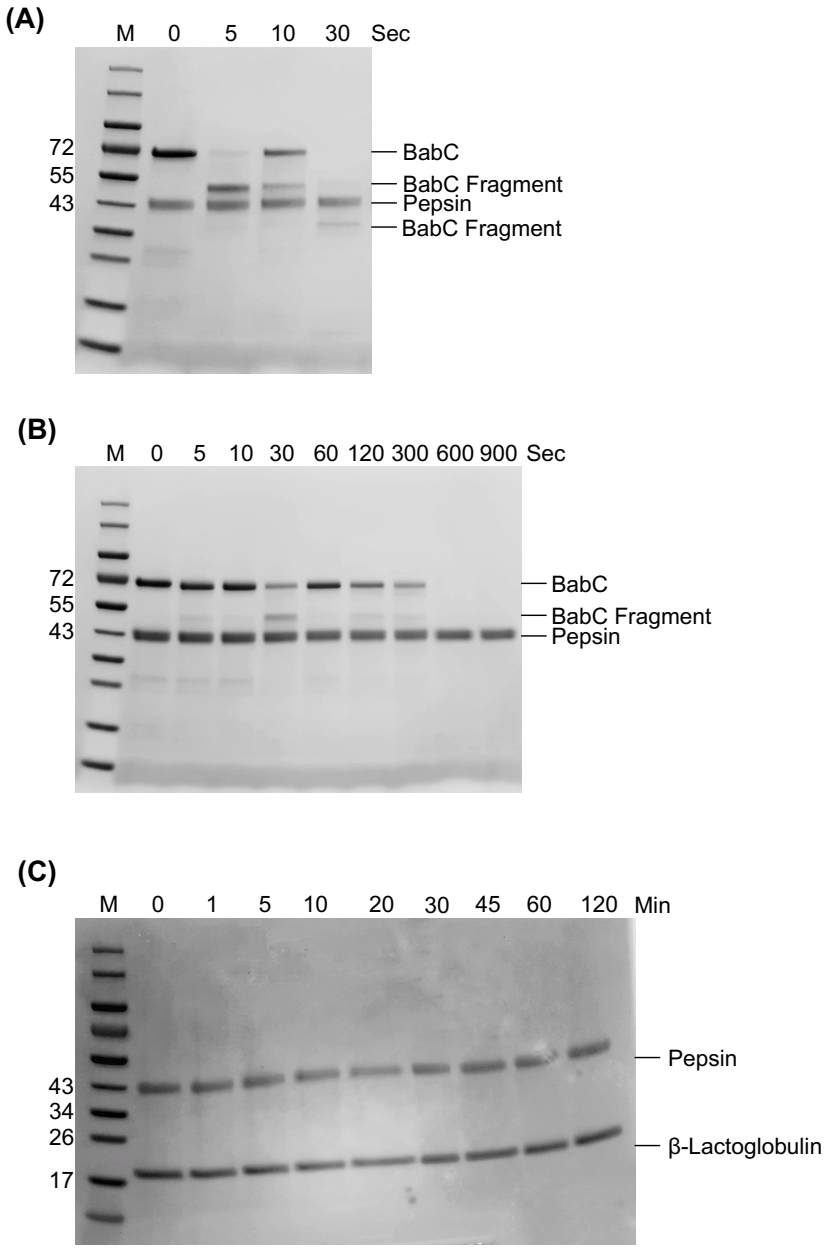


Figure 3.19: Peptic digestion pattern of BabC and β-lactoglobulin
InstantBlue-staining gels showing a time-course of peptic digestion for BabC at pH 1.2 (A) and pH 4.5 (B), and β-lactoglobulin at pH 4.5 (C) using the ratio recommended by the UPS and the NF which is 1 µg of test protein per 10 U of pepsin. BabC was rapidly degraded by pepsin at pH 1.2 after 10 seconds and fully digested after 30 seconds, while it was fully degraded at pH 4.5 after 5 minutes. BabC resistant fragments are indicated. β-lactoglobulin was resistant to the peptic digestion for two hours of incubation with pepsin. Time points indicated are in seconds (A) and (B) and in minutes (C). M – molecular weight marker in kDa (Fisher BioReagents™ EZ-Run™ Prestained Rec Protein Ladder, Fisher BioReagents (Thermo Scientific, USA) was used).

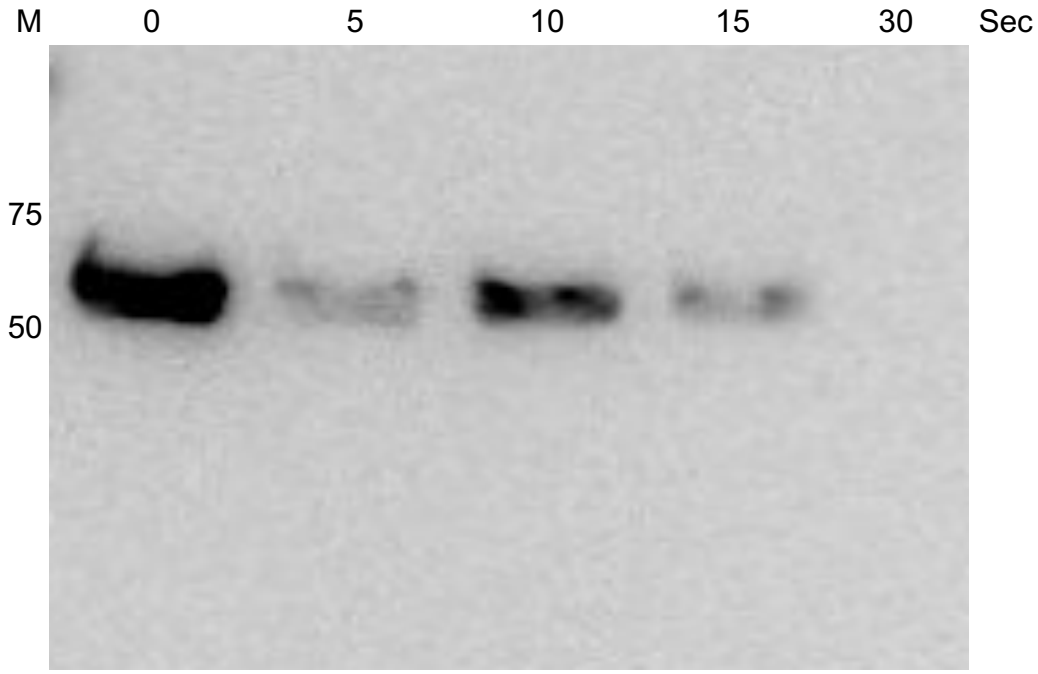


Figure 3.20: c-Myc immunodetection of BabC after peptic digestion
A time-course of peptic digestion for BabC at pH 4.5 analysed by Western blot. Time points indicated are in seconds. M – molecular weight marker in kDa.

3.4 Discussion

The first objective of the current study was to express recombinant BabC protein understood to mediate gastric adherence by the bacterium *H. pylori*. The design of the *babC* gene fragment used for screening of protein expression was considered in order to ensure that the transmembrane domain of BabC at the C-terminal region, which was unnecessary for the purpose of this research, was excluded from the expression cassette as previously demonstrated (66,76,109).

The Hop family of proteins of *H. pylori* share similarity in amino acid sequences at the N- and C-terminal regions with sequences of proteins displaying porin-like structures and their content in alternating amphipathic (hydrophilic and hydrophobic) amino acid sequences that form the antiparallel β-sheet structures of the membrane-spanning core region. This similarity supports that this part of proteins forms a transmembrane barrel (218), which permit the passive diffusion of extracellular nutrients and small hydrophilic molecules to the periplasm of the bacterium along a concentration gradient (240). The transmembrane domain is highly conserved and considered to act as an autotransporter (type V secretion) for secretion of the N-terminal domain beyond the outer membrane (218,219). The importance of exclusion of the hydrophobic transmembrane domain from the expression cassette was to avoid potential insolubility of the protein during expression and purification (241), and to decrease the metabolic burden on *E. coli* during heterologous

expression (241,242). Therefore, only the N-terminal extracellular domain of the protein, which is responsible for its binding characteristics, was required for the expression.

BabA of the *H. pylori* J99 is well-known to bind to the ABO/Le blood group antigens (243). Despite the long description of the transmembrane domain of BabA in the literature (218), there was no BabA structural information available in order to distinguish the precise amino acid sequence that corresponds to the transmembrane domain. Although several researchers have chosen various C-terminal truncation points for expression of BabA that has been successfully expressed (70,211,244), Hage *et al.* (2015) selected the residue 547 as C-terminal truncation point based on bioinformatic investigations. As a result of this, the putative corresponding C-terminal transmembrane domain of BabC *H. pylori* 26695 (amino acids 549-745) was excluded from the amplification based on the amino acid sequence alignment with BabA J99, which has structural information and an identified receptor (68,76).

The *babC* gene was recently discovered and originally identified as *hopU* (HP0317) in *H. pylori* strain 26695 (245), but does not exist in *H. pylori* strain J99 (246). Therefore, the genomic DNA of the *H. pylori* J99, which was used to express a functional BabA protein through cloning of the *babA J99* gene (243), cannot be used for the expression of BabC. For this reason, the genomic DNA of the *H. pylori* 26695 was used. Although the function of BabC is unknown, it is considered as a member of *H. pylori* adhesins since

its gene belongs to the Hop family (247). It was therefore of interest to examine the binding profile of BabC and to determine its ligand.

With regard to the restriction enzymes that would be used during the digestion process for the purpose of cloning of *babC* gene, NcoI was found to cut the *babC* gene sequence internally. As a result of this, site-directed mutagenesis was required in order to avoid the internal digestion of *babC* gene by NcoI. Asymmetric PCR was done for achieving the mutagenesis (248), however, it failed to amplify *babC* gene as there were many mutations detected by DNA sequencing. This might be because *babC* gene sequence shows high 5' and 3' homology to *babA* and *babB* (249), but the central region differs (245). Therefore, overlap extension PCR (250) was performed as the best option to obtain *babC* gene with the internal mutagenesis.

Naim Hage had previously evaluated various expression systems in order to choose the best expression system and purification protocol for *H. pylori* adhesins. After investigations with various insect, mammalian and bacterial cell lines, using suitable vectors for each, *E. coli* XL10 Gold® Ultracompetent cells were selected to be the host system for the BabA J99 expression. This *E. coli* strain was originally produced for the effective transformation of large molecules of DNA (251). pOPE101 was also used as an appropriate vector for this expression system. This vector contains a pelB leader sequence for translocating the protein from the cytoplasm, in which it was expressed, to the periplasm (180,252) via Sec pathway (253). Furthermore,

the vector contains a synthetic lac promoter associated with slower synthesis of protein than the pET system (211,254,255). The importance of the translocation of the proteins to the periplasm is its oxidising environment in which the proteins, with the help of enzymes such as DsbA and DsbC, can form disulphide bonds (256), which enables the production of functional proteins that are soluble and properly folded. Furthermore, the pelB leader is cleaved off in the periplasmic space because it contains signal peptidase or leader peptidase (256).

The protein solubility was improved by introducing a hexa-lysine tag into the expression cassette. This tag did enhance BabA J99 solubility (76), in congruence with earlier studies on the effect of polycationic tags on the solubility of proteins (242,257-260). One of the suggested mechanisms for the enhanced solubility is due to the repulsive electrostatic interactions, which occur between the polycationic tags with similar charges. This prevents the protein from aggregating and provides enough time for proper folding (242,259). Moreover, it has been shown that the insertion of the hexa-lysine tag improved the expression yield of BabA and prevented the proteolytic degradation of the C-terminal region of the protein (76). Unlike BabA, the N-terminal domain of LabA, did not show protection against proteolytic degradation and any improvement in the yield of the protein when the protein is expressed with a C-terminal hexa-lysin tag in the periplasm of host cells. However, the thermal stability of LabA was increased when the hexa-lysin tag was inserted without altering the folding of the protein (109).

It is also known that the low protein yield is one of the main limitations of the periplasmic expression because of the limited copies of the Sec-dependent signal peptide translocase complex and the actual surface of the periplasm that may limit the quantity of translocases that it can allocate (261,262). However, LabA was expressed in the periplasmic space of the *E. coli* with very high yield (109). As a result of this, the hexa-lysine tag was involved in the expression cassette of BabC as this protein belongs to the same clade as BabA in the Hop family and we assumed the same outcomes of protein expression in the periplasm of the host cells were likely.

Following IMAC, analysis of the elution fractions shows that other proteins, which are expressed in the periplasm of *E. coli* and do not contain hexa-histidine tag, non-specifically bind to the Ni^{2+}-NTA column. This could be because these proteins may contain a high level of histidine residues but their affinity to the Ni^{2+} column are lower compared to the proteins containing the hexa-histidine tag. In order to minimise the non-specifically bound proteins in IMAC elution fractions, a low concentration of imidazole was added to the washing buffer to help dissociate these proteins from the column in the wash-out step, for a BabC eluent with superior purity. This was confirmed with InstantBlue staining. After c-Myc immunodetection by Western blot, the first objective of this study was successfully achieved and BabC was expressed with relatively low yield, good purity and it was then suitable for potential applications and further investigations.

Following protein expression and purification steps, identification of the binding profile of BabC was one of the main aims of this study. As a primary experiment, BabC interactions with the common glycan epitopes in gastrointestinal mucins, which are Le^a, Le^b, Le^y, and sLe^x glycoconjugates (263), were investigated using an ELISA-like test. Although BabA was used as positive control for Le^b antigen, the main limitation of this experiment was that there were no positive controls used for the other antigens because either there are no proteins known to bind to these antigens or the protein is commercially unavailable i.e. SabA adhesin, which is known to bind to sLe^x. Consequently, the next step would be to analyse the protein using the glycan array in order to identify any putative binding profiles of the protein with various types of glycan.

Absence of binding to a specific ligand can have different reasons. The result can either indicate that the tested physiological ligand is not the right one, but in the worst-case scenario, it could be an indication that the protein is misfolded. Therefore, it is necessary to confirm that the obtained recombinant protein is properly folded. This can be done using a variety of techniques. The use of SEC for purification of BabC would confirm that the protein occurs as monomers of the expected size and is not an aggregate of denatured protein. Other techniques can be used to confirm folding. Circular dichroism can be used to assess the overall percentage of β-sheet and α-helices in the protein secondary structure, comparing the results with what is known for BabA (which has a very similar fold to BabB and BabC).

Secondly, Differential Scanning Fluorimetry could be used to confirm folding. In Differential Scanning Fluorimetry, the protein is heated stepwise (e.g. in a qPCR machine) in the presence of SYPRO Orange. This dye only binds to aromatic amino acid residues, which in a folded protein are not found on the surface, but are buried. Upon heating, the protein gradually unfolds, exposing hydrophobic amino acids which can bind to Sypro Orange, resulting in a measurable change in fluorescence. From this, a Tm for the protein can be defined, the temperature at which 50% of the protein is present in an unfolded state, while 50% is still folded.

However, we have reasons to believe that the BabC protein generated with our protocol is properly folded. First of all, the same periplasmic expression protocol resulted in functional, properly folded BabA (76) and LabA (109). The periplasmic space is an oxidising environment favouring formation of disulphide bonds, which BabA (and consequently BabC, which is predicted to have a very similar structure) needs for proper folding. Also, the obtained protein is fully soluble – a misfolded protein would have resulted in an insoluble fraction.

The role of BabC in the binding of *H. pylori* to gastric epithelial cells or gastric mucosa was assessed using glycan microarray by Max Planck Institute for Infection Biology in Germany. BabA that binds to type 1 ABO/Lewis blood group antigens was also assessed as a positive control. c-Myc-tagged BabA was analysed alongside BabC protein as a control because it is known to bind type 1 ABO/Le blood group antigens, and has been successfully used

on glycan arrays from the Consortium for Functional Glycomics (CFG) (Emory University, USA) (86). Furthermore, BabA, expressed by our group, was confirmed to be functional by showing its binding activity to Le^b using an ELISA assay before performing the carbohydrate microarray. Unfortunately, no significant binding with the c-Myc-tagged proteins by either BabC or BabA was observed. Although BabA binding was detected to five glycans that contained a unique determinant region of H-1, A-1 or B-1 blood group antigens by glycan array performed by the CFG (86), the current carbohydrate microarray did not detect any binding to BabA that was used as positive control. This can raise questions on the procedures they used for the detection of protein binding to the glycans. Therefore, the results are unreliable, and the array was planning to be performed by Alberta Glycomics Centre and Department of Chemistry (University of Alberta, Canada). However, the analysis was not performed because the sample was lost.

BabB and BabC proteins have not been expressed yet, their putative ligands, if any, are unknown, and their crystal structures remain to be determined. Although the *babB* gene was discovered as having significant similarity in the 5' and 3' to *babA*, the central region is different to *babA* (81), indicating that the sequence divergence in the central region encodes a unique receptor binding region (84,85). Based on this, studies suggested that gene conversion/recombination might occur between *babA* and *babB*, forming chimeric *babB/A* genes (264,265). This may occur as a result of selective pressure for adhesion in order to maintain the chronic

colonisation (266). On the other hand, it has also been suggested that BabB perhaps binds to an unidentified receptor on gastric mucosa or might modulate binding of other adhesins (82). However, recombinant BabB protein expression has not been described in the literature yet, and its crystal structure remains unknown.

Although BabC is considered as a member of *H. pylori* adhesins, its function is unknown (62). *babC* gene sequence shows high 5' and 3' homology to *babA* and *babB* (82), but the central region differs (83). These *bab* genes are located in the *H. pylori* chromosome at three different loci; Strain 26695 has *babC* gene at locus C, whereas strain HPAG-1 has *babC* and *babB* located at loci B and C, respectively (267). Gene recombination, therefore, seems to occur among these three loci, leading to hybrid genes with unknown binding properties (268). However, the molecular mechanism that regulates the gene conversion is poorly understood. Gene alterations in OMP genes do play significant roles in the environmental adaptation including the adhesion of the pathogen to the host cells (85). Therefore, BabB and BabC may (i) bind to some different glycans, (ii) bind to non-carbohydrate glycans, as is the case for HopQ adhesin which binds to CEACAMs, or (iii) only become active after genomic recombination resulting in hybrids of BabA/BabB and perhaps hybrids of BabA/BabC or BabB/BabC in order to maintain the chronic colonisation. Different studies have reported varying proportions of *bab* chimeras (269). According to the study conducted by Matteo *et al*., chimeric *babA*/*babB* and *babB*/*babA* genes were observed

in 20% and 15% of isolates, respectively (266). In another report conducted by Kim *et al*., *babA*/*babB* and *babB*/*babA* chimeras were observed in 6.25% and 1.25% of Korean *H. pylori* strains, respectively, while *babA*/*B*, *babB*/*babA* and *babB*/*babC* chimeras were identified in 28.75%, 2.5% and 1.25% of American *H. pylori* strains, respectively (270). Provided that chimeric adhesin genes still encode an open reading frame, one way to probe the binding activity of chimeric adhesins would be to recombinantly express the protein encoded by the chimeric sequences and screen for ligands, e.g. using carbohydrate arrays, or against gastric cell lines such as AGS. This could be accompanied by in silico analysis of predicted three dimensional structures in comparison with adhesins with known ligands, such as BabA. However, one should not discount the possibility that chimeric genes may simply function to switch off specific adhesins in response to a changing host environment, although this seems a complicated mechanism which could be solved in a different, easier way.

Consideration was given to the influence of the low pH of gastric fluid on the stability of the protein in the presence of pepsin. Although the environment of the gastric lumen is strongly acidic (pH ≤ 3), many researchers have reported a broad range of pH values in both fed (pH value between 3 to 6) and fasted (pH value between 1 to 8) states (271). Accordingly, a pH value of 4.5, which mimics the acidic conditions of post-prandial states, was used to achieve the peptic digestion experiments. On the other hand, a pH value of 1.2 was also suggested by the USP and the NF for analysis of the protein

digestibility by pepsin (237,272). Therefore, pH values of 4.5 and 1.2 were used in this experiment. It was found that no conformational change was seen in BabA using the circular dichroism spectroscopy at both neutral condition, which is similar to the pH of inner surface mucus layer where the bacterium's tropism is, and acidic (pH 4.5) condition. Furthermore, no variation was seen in the binding of BabA to Le^b under both neutral and acidic conditions (68). This reveals that BabA is stable under acidic conditions (pH 4.5). As no binding for BabC was detected, it was unnecessary to probe a possible conformational change of BabC by circular dichroism spectroscopy.

BabC was rapidly degraded by pepsin at pH 1.2 after 10 seconds and fully digested after 30 seconds, while it was also fully digested by pepsin at pH 4.5 after 5 minutes, although its fragments that were cut from their C-terminal regions could still be detected in SDS-PAGE. Previously, our group found that BabA was rapidly degraded by pepsin at pH 4.5 after 1 minute. Therefore, BabC appears to be unstable at acidic (pH 4.5) conditions in presence of pepsin. No C-terminal sequences containing c-Myc tag were detected by Western blot after 15-30 seconds incubation with pepsin, indicating that none of the observed fragments contained the C-terminus. This suggests that the C-terminus of BabC contains a highly susceptible protease cleavage site. It was also confirmed that incubation of BabC in the same acidic buffer, without pepsin, did not result in any detectable proteolysis.

Unlike β-lactoglobulin, which we used as the experimental control due to its known resistance to pepsin, no *H. pylori* adhesins have been reported using the same experiment in the literature, although *H. pylori* VacA, a secreted toxin, has been reported to be resistant to peptic digestion after 2 hrs of incubation (273). At the level of the entire level of the *H. pylori* bacterial cell, pepsin in combination with the gastric acid can destroy the bacterium as well as impair the bacterial motility in the gastric lumen, thus inhibiting the colonisation to the mucus layer of the stomach (274-276). Therefore, *H. pylori* is known to be pepsin-sensitive bacterium although it is being a specialised pathogen in the stomach. Although BabC is sensitive to peptic digestion, pepsin is not active in the neutral pH environment in the mucus layer of the stomach because of the bicarbonate barrier, hence bacteria close to the epithelium are safe from the effects of pepsin and high acidity. There are various reasons may explain the sensitivity to peptic digestion. First, *H. pylori* may be able to avoid the effects of HCl/pepsin by its known ability to penetrate the mucus layer and swim towards the epithelial surface, by sensing the urea gradient emanating from the gastric epithelium (277). Furthermore, the urease activity in the bacterium generates a buffered surface layer for the *H. pylori* in response to acidic environment (278), which modulates the enzymatic activity of pepsin. Second, our experiments have been performed with recombinant proteins, while BabA, BabC, and others in the whole bacterium are embedded in the outer leaflet of the LPS membrane, which may offer substantial protection against pepsin digestion,

by steric hindrance, with only the ligand-binding crown protruding from the membrane. In this context, it is interesting to note that while most of the extracellular domain of BabA has many predicted pepsin cleavage sites, the crown appears almost devoid of such cleavage sites. Third, adhesins in *H. pylori* might possess some additional posttranslational modifications, e.g. some glycosylation, which may increase stability to proteolytic degradation (279). Native BabA isolated from *H. pylori* was slightly more stable than recombinant BabA made in *E. coli* (86).

Chapter 4: HomB recombinant expression and purification

4.1 Introduction

The second objective of this study was to recombinantly express and purify HomB, a putative adhesin used by *H. pylori* for colonising the gastric epithelium, also able to induce IL-8 proinflammatory cytokine secretion *in vitro* (83,114). Although *H. pylori* HomB has previously been expressed in *E. coli* BL21 as a GST-fusion protein (114), the resulting protein needed to be refolded from an aggregated form. Generally, proteins do not always regain their full function when refolded from the aggregation form, despite using time-consuming methods (215,216). This is particularly problematic when the unknown enzymatic activity of the expressed protein is the target of the research, as is the case with HomB in this Chapter. As a result, it was essential to express HomB as a soluble recombinant protein.

Only the N-terminal extracellular domain of HomB was required for the purposes of the current study. It is known that HomB belongs to the Hom family of the Hop family of *H. pylori* (89) consisting of two domains: N-terminal extracellular and C-terminal transmembrane domains. The transmembrane domain, which is composed of antiparallel β-sheet structures that form a β-barrel unit, is highly conserved and considered to act as an autotransporter (type V secretion) for secretion of the N-terminal domain beyond the outer membrane (89,219). The hydrophobicity of the transmembrane domain, which may reduce the solubility of the protein during expression, was the reason behind designing suitable recombinant

plasmids to express the N-terminal extracellular domain only. The N-terminal extracellular domain of HomB contains six cysteine residues, which may suggest that HomB contains up to three intramolecular disulphide bonds (Figure 4.1).

```
  1 MRKLFIPLLL FSALEANEKN GFFIEAGFET GLLEGTQTQE KRHTTTKNTY ATYNYLPTDT
 61 ILKRAANLFT NAEAISKLKF SSLSPVRVLY MYNGQLTIEN FLPYNLNNVK LSFTDAQGNT
121 IDLGVIETIP KHSKIVLPGE AFDSLKEAFD KIDPYTFFFP KFEATSTSIS DTNTQRVFET
181 LNNIKTNLIM KYSNENPNNF NTCPYNNNGN TKNDCWQNFT PQTAEEFTNL MLNMIAVLDS
241 QSWGDAILNA PFEFTNSSTD CDSDPSKCVN PGVNGRVDTK VDQQYILNKQ GIINNFRKKI
301 EIDAVVLKNS GVVGLANGYG NDGEYGTLGV EAYALDPKKL FGNDLKTINL EDLRTILHEF
361 SHTKGYGHNG NMTYQRVPVT KDGQVEKDSN GKPKDSDGLP YNVCSLYGGS NQPAFPSNYP
421 NSIYHNCADV PAGFLGVTAA VWQQLINQNA LPINYANLGS QTNYNLNASL NTQDLANSML
481 STIQKTFVTS SVTNHHFSNA SQSFRSPILG VNAKIGYQNY FNDFIGLAYY GIIKYNYAKA
541 VNQKVQQLSY GGGIDLLLDF ITTYSNKNSP TGIQTKRNFS SSFGIFGGLR GLYNSYYVLN
601 KVKGSGNLDV ATGLNYRYKH SKYSVGISIP LIQRKASVVS SGGDYTNSFV FNEGASHFKV
661 FFNYGWVF
```

Figure 4.1: Amino acid sequence of HomB J99
The amino acid sequence shows that the N-terminal extracellular domain of HomB contains six cysteine residues indicated by green circles, suggesting that HomB can have up to three intramolecular disulphide bonds. The underlined amino acids are the C-terminal transmembrane domain.

Three steps were required for the expression and purification of the recombinant HomB. First, the gene fragment, which encodes the N-terminal extracellular domain of HomB, was cloned into different vectors designed for protein expression in a variety of *E. coli* strains. Next, small-scale expression was achieved for each expression system in order to assess the capability of generation of a soluble protein. Finally, protein purification was carried out on larger scale using the successful approach optimised during expression screening.

4.2 Materials and methods

4.2.1 Generation of HomB expression constructs

After the successful expression and purification of BabC, in addition to the expression of BabA and LabA adhesins by our group (76,109), the next target was the *H. pylori* HomB adhesin. In the current chapter, the N-terminal extracellular domain of HomB adhesin from *H. pylori* J99 was chosen to be expressed and characterised, as the expression of this protein has been reported once in the literature until the present day, but it was insoluble protein (114). HomB expression construct engineered in this study is summarised in Table 4.1.

Table 4.1: Expression constructs engineered in this study

The numbers used in the name of the inserted gene correspond to *homB* gene fragment derived from *H. pylori* strain J99 encoding amino acids 9-511 or 19 -511. The letter 'R' in the construct name stands for a deca-arginine (R10) designed to be inserted directly in the C-terminus of $HomB_{19-511}$.

* pelB leader, which translocates the protein from the cytoplasm to the periplasmic space of *E. coli*, was deleted.

Insert Gene	Vector	Cloning Cells
$homB_{9-511}$	pOPE101	*E. coli* XL10 Gold
$homB_{9-511}$	pET-22b(+)	*E. coli* DH5α
$homB_{19-511}$	pOPE101	*E. coli* XL10 Gold
$homB_{19-511R}$	pOPE101	*E. coli* XL10 Gold
$homB_{19-511}$	pET-22b(+)*	*E. coli* DH5α
$homB_{19-511}$	pCold-GST	*E. coli* DH5α
$homB_{19-511}$	pMAL-c5X	*E. coli* DH5α

4.2.2 PCR amplification

The aim of this procedure was to amplify a segment of the *homB* gene (Genbank accession no. WP_046655427.1) from the genomic DNA of *H. pylori* J99, kindly donated by Dr. Rob Delahay, the University of Nottingham (Genbank accession no. NZ_CP011330.1). Similar to *babC* gene amplification, the putative C-terminal transmembrane domain, representing amino acids 512-668 based on the information provided by the National Centre for Biotechnology Information (NCBI) (280), was excluded from the amplification.

4.2.2.1 Prediction of signal peptide

HomB is known to be a *H. pylori* OMP (89). Therefore, *homB* gene is predicted to encode an N-terminal signal peptide, which is cleaved during HomB secretion to its outer membrane in *H. pylori* (219), and this signal peptide was also excluded from the *homB* gene fragment using SignalP 4.0 Server (281), the updated version 4.1 (282), SignalP 5.0 server (283) and SignalP 3.0 server (284). However, both SignalP 4.0 and 4.1 servers predicted that *homB* gene does not encode an N-terminal signal peptide (Figure 4.2), but it is not reasonable because HomB is an OMP. As a result, $homB_{9\text{-}511}$ was amplified. On the other hand, SignalP 5.0 and 3.0 servers predicted that the gene encodes an N-terminal signal peptide with 16 amino acid sequence (Figure 4.3). As a result, $homB_{19\text{-}511}$ was amplified, where

18 amino acids (the predicted 16 amino acids + 2 amino acids) were excluded.

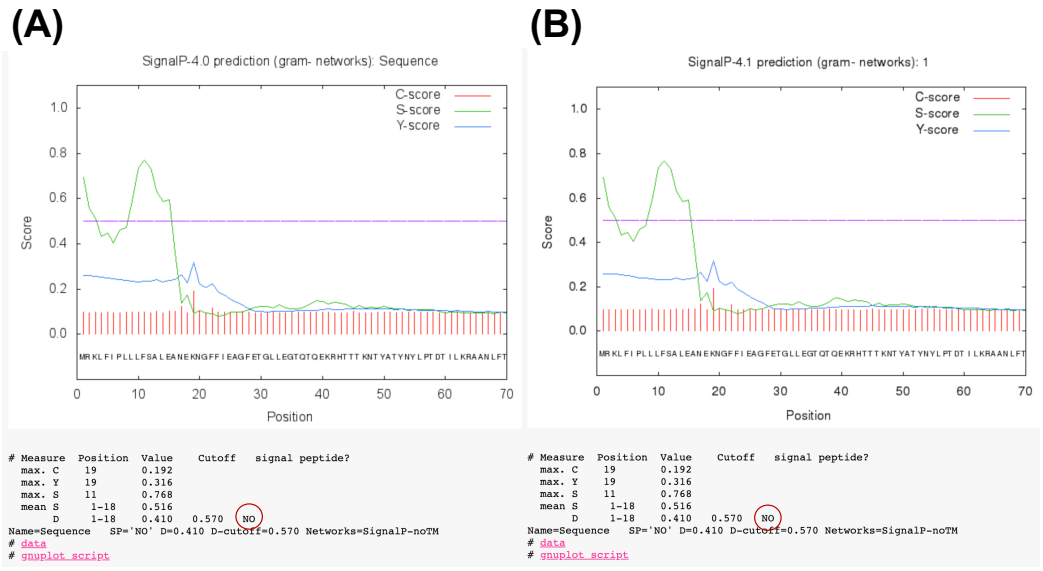


Figure 4.2: Signal peptide prediction using SignalP 4.0 and 4.1 servers
Both SignalP 4.0 (A) and 4.1 (B) servers predicted that *homB* gene does not encode an N-terminal signal peptide. However, HomB is a *H. pylori* OMP and it should contain a signal peptide. Therefore, eight amino acids were initially excluded from the amplification of HomB.

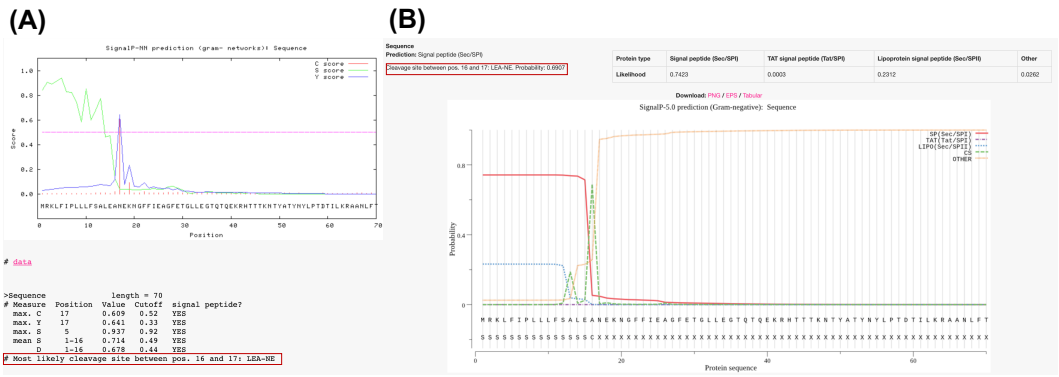


Figure 4.3: Signal peptide prediction using SignalP 3.0 and 5.0 servers
Both SignalP 3.0 (A) and 5.0 (B) servers predicted that *homB* gene encodes an N-terminal signal peptide (16 amino acids). As a result, 18 amino acids (the predicted 16 amino acids and 2 more amino acids) were excluded from the amplification of HomB.

4.2.2.2 Ordinary PCR

The first step was to build the construct pOPE101_*homB*$_{9\text{-}511}$. Q5® High-Fidelity 2X Master Mix was used to amplify *homB* gene fragment that encodes the amino acids 9-511 using two primers shown in Table 4.2. The PCR was performed under the following conditions: 98°C for 30 sec of initial denaturation, followed by 35 cycles of template denaturation at 98°C for 10 sec, primer annealing at 65°C for 30 sec, and primer extension at 72°C for 47 sec. The reaction was finished with final extension at 72°C for 2 min. The PCR product was loaded on 1% agarose gel and run in 0.5x TBE buffer. SYBR Gold was used to stain the gel after the electrophoresis run, and the gel was then visualised under blue-light transilluminator. The SYBR Gold was used instead of EtBr because it is safer and to avoid DNA mutation caused by the gel exposure to the UV light. The band was then extracted and purified using a QIAquick Gel Extraction Kit, according to the manufacturer's protocol. The DNA concentration was estimated by measuring the absorbance at 260 nm using Nanodrop 2000.

Table 4.2: Primers for *homB*$_{9\text{-}511}$ amplification

FOR and REV denote forward and reverse primers respectively. The number in the primer name indicates the 1st amino acid in HomB encoded by the nucleotides located downstream of the restriction site, which is underlined in the primer sequence.

Gene	Primer Name	Primer Sequence (5'-3')
homB$_{9\text{-}511}$	HomB [9 FOR]_NcoI	ATATATCCATGGTTTTATTATTCAGCGCTTTAG
	HomB [511 REV]_NotI	TATATAGCGGCCGCAACCCCTAAAATAGGGCTTC

4.2.2.3 Overlap extension PCR

Overlap extension PCR was performed using Q5® High-Fidelity 2X Master mix to amplify *homB*$_{9\text{-}511}$. Two PCRs were carried out to generate a pair of DNA fragments with overlapping regions, in which a third PCR was performed to get the full-length *homB* which designed to be inserted into pOPE101 (Figure 4.4). Two primers were designed to introduce the selected restriction sites at the 5' and 3' ends of the amplified gene in order to be incorporated into the pOPE101 expression vector, while the other primers were designed based on unique sequences in the mid region of *homB* in which they differ from that in *homA* gene. Table 4.3 shows the primers used in the overlap extension PCR. The annealing temperatures used in the three PCRs were 65°C, 55°C and 65°C respectively. A band from the PCR product was then extracted from 1% agarose and purified using a QIAquick Gel Extraction Kit, according to the manufacturer's protocol. The DNA concentration was estimated by measuring the absorbance at 260 nm using Nanodrop 2000.

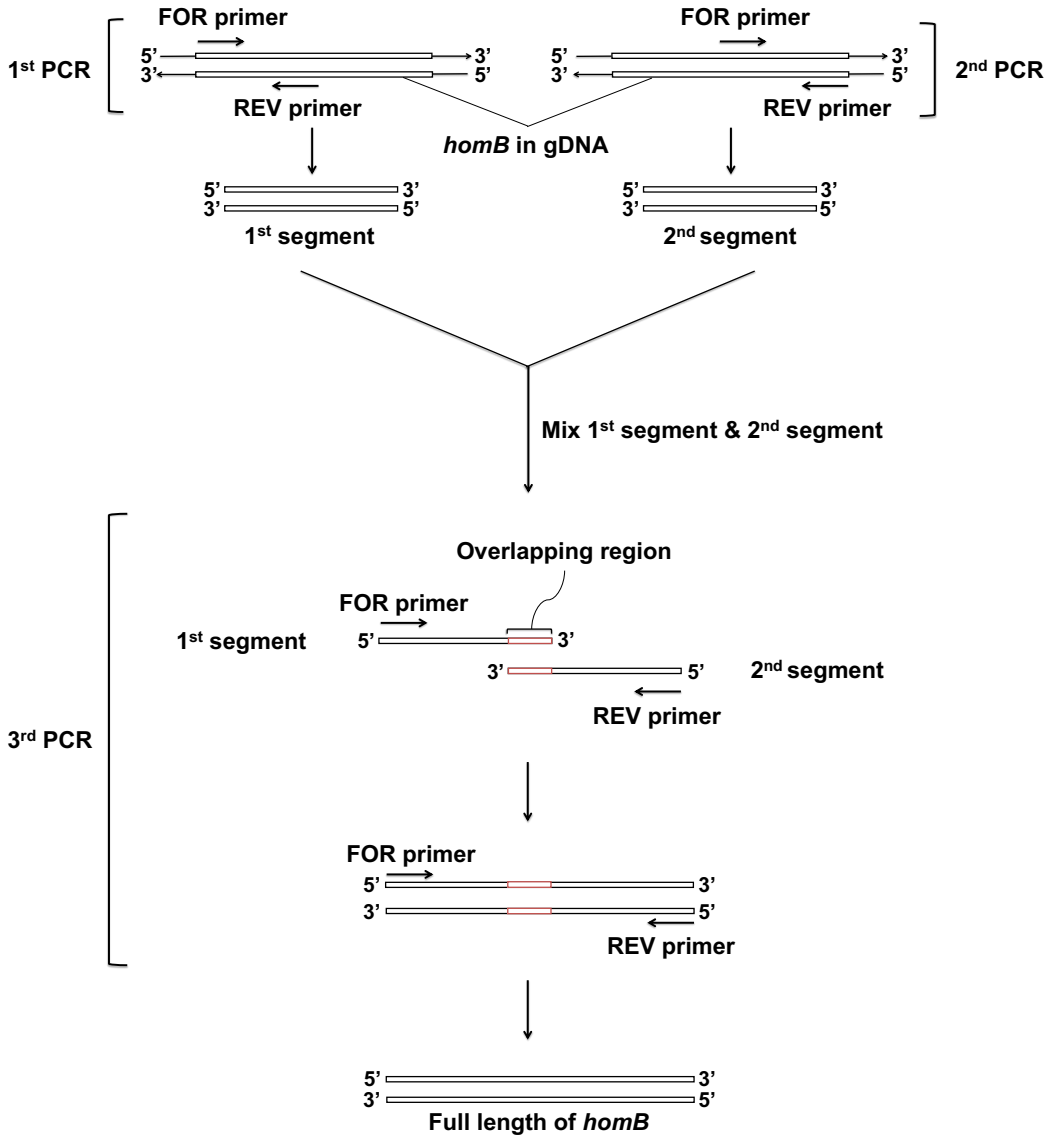


Figure 4.4: Schematic diagram of overlap extension PCR

Three PCRs were performed to obtain *homB*$_{9\text{-}511}$ gene from genomic DNA of *H. pylori* J99 using overlap extension PCR. Two PCRs were carried out to generate two DNA fragments with overlapping regions, in which the first PCR was achieved to amplify the first segment, while the second PCR was performed to amplify the second fragment using primers that contain unique sequences in the mid region of *homB* in which they differ from that in *homA* gene. The third PCR was performed to obtain the full length of *homB*$_{9\text{-}511}$ through mixing the first and second segments that have overlapping region. FOR, REV and gDNA denote forward, reverse and genomic DNA, respectively.

Table 4.3: Primers used in the overlap extension PCR

FOR and REV denote forward and reverse primers respectively. The number in the primer name indicates the 1st amino acid in HomB encoded by the nucleotides located downstream of the restriction site, which is underlined in the primer sequence, or by the nucleotides located in the beginning of the primer.

[a] Primers used for amplification of the first segment of *homB*$_{9\text{-}511}$.

[b] Primers used for amplification of the second segment of *homB*$_{9\text{-}511}$.

[c] Primers used for overlap extension PCR.

Gene	Primer Name	Primer Sequence (5'-3')
homB$_{9\text{-}511}$	HomB [9 FOR]_NcoI [a, c]	ATATATCCATGGTTTTATTATTCAGCGCTTTAG
	HomB [327 REV] [a]	TGTGCCATATTCACCATCATTGCC
	HomB [290 FOR] [b]	CAAGGTATTATTAATAATTTTAG
	HomB [511 REV]_NotI [b, c]	TATATAGCGGCCGCAACCCCTAAAATAGGGCTTC

4.2.2.4 Amplification of the other inserted genes

The primers used for amplification of the other *homB* gene to be inserted into various vectors are listed in Table 4.4.

Table 4.4: Primers for amplification of *homB* gene to be inserted into various vectors

Primer pairs used for amplification of *homB* gene designed to be inserted into different vector. FOR and REV denote forward and reverse primers respectively, and the underlined nucleotides are the restriction site for the corresponding restriction enzyme.

* primers designed to be inserted into the selected vector using Gibson Assembly in which the amplified genes do not need to be digested.

Insert Gene	Restriction Enzyme	Primer Sequence (5'-3')	Annealing Temperature
pET22b_*homB*$_{9\text{-}511}$	FOR-NcoI	CTGCTCCTCGCTGC CCAGCCGGCGATGG CCATGGTTTTATTAT TCAGC*	65°C
	REV-NotI	GGTGGTGGTGGTGG TGCTCGAGTGCGGC CGCAACCCCTAAAA TAGG*	
pOPE101_*homB*$_{19\text{-}511}$	FOR-NcoI	ATATATCCATGGCG AAAAACGGCTTTTTC ATAGAAGCCGGC	72°C
	REV-NotI	TATATAGCGGCCGC AACCCCTAAAATAG GGCTTC	
pOPE101_*homB*$_{19\text{-}511R}$	FOR-NcoI	ATATATCCATGGCG AAAAACGGCTTTTTC ATAGAAGCCGGC	72°C
	REV-NotI	TATATAGCGGCCGC ACGACGACGACGAC GACGACGACGACGA CGAACCCCTAAAAT AGGGCTTC	
pET22b(ΔpelB)_*homB*$_{19\text{-}511}$	FOR-NdeI	TTTTGTTTAACTTTA AGAAGGAGATATAC ATATGGATGAAAAAC GGCT*	55°C
	REV-XhoI	GATCTCAGTGGTGG TGGTGGTGGTGCTC GAGAACCCCTAAAA TAGGGC*	

pCold-GST_*homB*$_{19\text{-}511}$	FOR- XmaI	CGGATCTGGAAGTT CTGTTCCAGGGCCC GGGAAAAAACGGCT TTTTCATAG*	55°C
	REV- XbaI	GTGCTTTTAAGCAG AGATTACCTATCTAG AACCCCTAAAATAG GGC*	
pMALc5X_*homB*$_{19\text{-}511}$	FOR-XmnI	CAACAACCTCGGGA TCGAGGGAAGGATT TCAAAAAACGGCTTT TTCATA*	72°C
	REV-NcoI	ATCCGTCGACGATA TCGCGGCCGCCCAT GGTTAAACCCCTAA AATAGGG*	

4.2.3 Cloning

Following PCR amplification, the product was extracted and purified after agarose gel electrophoresis, using a QIAquick Gel Extraction Kit. The DNA concentration was estimated by measuring the absorbance at 260 nm using Nanodrop 2000. The purified PCR products designed to be inserted into pOPE101 vector were then digested alongside the vector with restriction enzymes, as shown in Table 4.4, according to the manufacturer's instructions. After digestion, 3:1 insert to vector molar ratio was used to ligate *homB* gene into the pOPE101 expression vector using T4 DNA ligase, under conditions specified by the manufacturer. The digestion process was not required for other purified PCR products, while it was needed only for the corresponding vectors using restriction enzymes listed in Table 4.4. After vector digestion, 2:1 insert to vector molar ratio was used to ligate *homB* gene into its expression vector using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, USA), as directed by the manufacturer.

For plasmid propagation, 5 µL ligation mixture were transformed into 100 µL chemically competent *E. coli* XL10 Gold, which is tetracycline resistant, in order to propagate pOPE101-based constructs introduced into bacteria through heat shock method as follows: 30 minutes on ice, then 30 seconds at 42°C, followed by 2 minutes on ice. On the other hand, pET-22b(+)-, pCold-GST- and pMAL-c5X-based constructs were transformed and propagated into *E. coli* DH5α cells (New England Biolabs, USA) through heat shock method as follows: 30 minutes on ice, then 90 seconds at 42°C,

followed by 2 minutes on ice. After the heat shock process, cells were grown in 900 µL S.O.C medium at 37°C with shaking for one hour, to allow for expression of antibiotic resistance-conferring proteins. For selection, transformants were plated on 100 µg/mL LB_{amp} agar and incubated at 37°C overnight.

Single colonies were picked up from the overnight growth plates for colony PCR through suspension in 20 µL LB_{amp}. Primers listed in Table 4.5 were used in colony PCR, which spanned the MCS of the hosting plasmid. REDTaq® ReadyMix™ was used for PCR amplification, according to the manufacturer's protocol; an annealing temperature of 58°C was used. The PCR products were screened for gene insertion via agarose gel electrophoresis stained with EtBr, and positive clones were subsequently cultured overnight in 5 mL LB_{amp} medium at 37°C. Plasmid extraction and purification were then carried out using a plasmid miniprep kit (Qiagen, USA), according to the manufacturer's instructions. The concentration of purified plasmid was estimated by measuring the absorbance at 260 nm using Nanodrop 2000, and confirmatory sequencing was obtained using primers shown in Table 4.5.

Table 4.5: Primers for Colony PCR and Confirmatory Sequencing

FOR and REV denote forward and reverse primers respectively. The numbers in the primer name denotes the corresponding base pairs (sense strand) in the pOPE101, pET-22b(+), pCold-GST, pMAL-c5X or *homB* gene sequences.

[a] Primers used for colony PCR and confirmatory sequencing.

[b] Primers used for confirmatory sequencing only.

[c] Primers used for confirmatory sequencing only for all constructs.

Construct Backbone	Primer Name	Primer Sequence (5'-3')
pOPE101	pOPE [27-46bp FOR] [a,b]	TTGACTTGTGAGCGGATAAC
	pOPE [1135-1154bp REV] [a,b]	ATGTGTCAGAGGTTTTCACC
pET-22b(+)	T7 Promoter [5117-5136 FOR] [a,b]	TAATACGACTCACTATAGGG
	T7 Terminator [5407-5425 REV] [a,b]	GCTAGTTATTGCTCAGCGG
pCold-GST	pCold-GST [208-226 FOR] [a,b]	ACGCCATATCGCCGAAAGG
	pCold-GST [1074-1093 REV] [a,b]	GGCAGGGATCTTAGATTCTG
pMAL-c5X	pMAL-c5X [2625-2643 FOR] [a,b]	GACTAATTCGAGCTCGAAC
	pMAL-c5X [2772-2790 REV] [a]	CTTTCGACTGAGCCTTTCG
	homB [868-890 FOR] [c]	CAAGGTATTATTAATAATTTTAG
	homB [958-981 REV] [c]	TGTGCCATATTCACCATCATTGCC

4.2.4 Expression screening

Table 4.6 shows the expression constructs and their expression hosts used in this study.

Table 4.6: Expression constructs and expression host used in this study

The numbers used in the name of the construct name correspond to *homB* gene fragment derived from *H. pylori* strain J99 encoding amino acids 9-511 or 19 -511. The letter 'R' in the construct name stands for a deca-arginine (R10) designed to be inserted directly in the C-terminus of $HomB_{19-511}$. ΔpelB means deletion mutation of pelB leader sequence from the vector.

Construct Name	Expression Type	Expression Host
pOPE101_*homB*$_{9-511}$	Periplasmic	*E. coli* XL10 Gold
pET22b_*homB*$_{9-511}$	Periplasmic	*E. coli* BL21 (DE3) *E. coli* Origami B (DE3)
pOPE101_*homB*$_{19-511}$	Periplasmic	*E. coli* XL10 Gold
pOPE101_*homB*$_{19-511R}$	Periplasmic	*E. coli* XL10 Gold
pET22b(ΔpelB) _*homB*$_{19-511}$	Cytoplasmic	*E. coli* Rosetta-gami B (DE3) pLysS
pCold-GST_*homB*$_{19-511}$	Cytoplasmic	*E. coli* BL21 (DE3) *E. coli* ArcticExpress (DE3)
pMAL-c5X_*homB*$_{19-511}$	Cytoplasmic	*E. coli* Rosetta-gami B (DE3) pLysS

4.2.4.1 Periplasmic expression

4.2.4.1.1 Periplasmic expression of pOPE101_*homB*$_{9\text{-}511}$ construct in *E. coli* XL10 Gold at 24°C

In addition to plasmid propagation, *E. coli* XL10 Gold was also used for protein expression. The same protocol described in Section 3.2.2 was used for the assessment of periplasmic expression of recombinant HomB$_{9\text{-}511}$ in *E. coli* XL10 Gold cells. In addition to a total protein fraction, periplasmic fractions were prepared (sample preparation method described in Section 3.2.2.1). Cytoplasmic soluble and insoluble protein fractions were also prepared instead of spheroplast fraction (sample preparation method described in Section 4.2.4.3.3.1). SDS-PAGE was performed to analyse all protein fractions, and followed by Coomassie-staining and Western blotting, as described in Section 3.2.2.2.

4.2.4.1.2 Periplasmic expression of pET22b_*homB*$_{9\text{-}511}$ construct in *E. coli* BL21 (DE3)

Following plasmid propagation into *E. coli* DH5α cells, pET22b_*homB*$_{9\text{-}511}$ expression construct was then transformed into chemically competent *E. coli* BL21 (DE3) (Merck Millipore, Germany) through a heat shock method as the following: 30 minutes on ice, then 90 seconds at 42°C, followed by 2 minutes on ice. After the heat shock process, cells were grown in 900 µL SOC at 37°C with shaking for one hour. For plasmid selection, transformants were plated on LB$_{amp}$ (100 µg/mL ampicillin) agar and incubated at 37°C

overnight. Single colony was picked up from the overnight growth plate in order to inoculate 5 mL of LB_{amp} (100 µg/mL ampicillin) media overnight at 37°C. 75 µL from the overnight culture was then used to inoculate 15 mL LB_{amp} (1:200 dilution) and incubated with shaking at 37°C until the OD_{600} reached around 0.6. Before inducing the culture with 0.1 mM and 0.5 mM IPTG at 24°C, a 1 mL was taken from the culture for a "total protein fraction" preparation (sample preparation method described in Section 3.2.2.1), for subsequent analysis. After 16 hours incubation with shaking, cellular fractions were prepared for analysis. In addition to a total protein fraction, periplasmic (sample preparation method described in Section 3.2.2.1) and cytoplasmic soluble and insoluble protein fractions (sample preparation method described in Section 4.2.4.3.3.1) were prepared. SDS-PAGE was performed to analyse all protein fractions, and followed by Western blotting, as described in Section 3.2.2.2.

4.2.4.1.3 Periplasmic expression of pET22b_*homB*$_{9-511}$ construct in *E. coli* Origami B (DE3)

E. coli Origami B (DE3) cells (Merck Millipore, Germany) was used to assess the expression of $HomB_{9-511}$ using the same protocol described in Section 4.2.4.1.2. However, the concentration of IPTG used was 0.1 mM. The heat shock method used to transform the construct into *E. coli* Origami B (DE3) was as the following: 5 minutes on ice, then 30 seconds at 42°C, followed by 2 minutes on ice.

4.2.4.1.4 Periplasmic expression of pOPE101_*homB*$_{9-511}$ construct in *E. coli* XL10 Gold at 16°C

The same protocol described in Section 4.2.4.1.1 was used except that the temperature and the shaking speed were reduced to 16°C and 90 rpm after inducing the cells with 0.1 mM IPTG.

4.2.4.1.5 Periplasmic expression of pOPE101_*homB*$_{19-511}$ construct in *E. coli* XL10 Gold

The same protocol described in Section 4.2.4.1.1 was used to assess the expression of HomB$_{19-511}$. The concentration of IPTG used were 0.1 mM and 1 mM.

4.2.4.1.6 Periplasmic expression of pOPE101_*homB*$_{19-511R}$ construct in *E. coli* XL10 Gold

The same protocol described in Section 4.2.4.1.1 was used to assess the expression of HomB$_{19-511}$ after adding deca-arginine (R10) at the C-terminal region of the protein. The IPTG concentrations used were 0.1 and 0.5 mM

4.2.4.2 Cytoplasmic expression

4.2.4.2.1 Cytoplasmic expression of pET22b(ΔpelB)_*homB*$_{19-511}$ construct in *E. coli* Rosetta-gami B (DE3) pLysS

Following plasmid propagation into *E. coli* DH5α cells, pET22b(ΔpelB)_*homB*$_{19-511}$ expression construct was then transformed into

E. coli Rosetta-gami B (DE3) pLysS (Merck Millipore, Germany) through a heat shock method as the following: 5 minutes on ice, then 30 seconds at 42°C, followed by 2 minutes on ice. After the heat shock process, cells were grown in 900 µL SOC at 37°C with shaking for one hour. For plasmid selection, transformants were plated on LB_{amp} (100 µg/mL ampicillin) agar and incubated at 37°C overnight. Single colony was picked up from the overnight growth plate in order to inoculate 5 mL of LB media supplemented with 100 µg/mL ampicillin, 34 µg/mL chloramphenicol (Merck, UK), and 0.1 M glucose (Thermo Scientific, USA) (LB_{ACG}), and the inoculation was then incubated overnight at 37°C. 5 mL from the overnight culture was then used to inoculate 400 mL LB_{amp} and incubated with shaking at 37°C until the OD_{600} reached around 0.6. Before inducing the culture with 0.5 mM IPTG at 24°C, a 1 mL was taken from the culture for a "total protein fraction" preparation (sample preparation method described in Section 3.2.2.1), for subsequent analysis. After 16 hours incubation with shaking, cellular fractions were prepared for analysis of 10 mL only. In addition to a total protein fraction, cytoplasmic soluble and insoluble protein fractions were prepared (sample preparation method described in Section 3.2.2.1 and Section 4.2.4.3.3.2). SDS-PAGE was performed to analyse all protein fractions, and followed by Western blotting, as described in Section 3.2.2.2.

4.2.4.2.2 Cytoplasmic expression of pCold-GST_*homB*$_{19\text{-}511}$ construct in *E. coli* BL21 (DE3)

pCold-GST_*homB*$_{19\text{-}511}$ construct transformation and plasmid selection were done as described in Section 4.2.4.1.2. Single colony was picked up from the overnight growth plate in order to inoculate 5 mL of LB_{amp} media (100 µg/mL ampicillin), and the inoculation was then incubated overnight at 37°C. 5 mL from the overnight culture was then used to inoculate 400 mL LB_{amp} and incubated with shaking at 37°C until the OD_{600} reached around 0.6. The culture was then incubated with shaking at 10°C for 10 minutes. Before inducing the culture with 0.5 mM IPTG at 10°C for 24 hours, a 1 mL was taken from the culture for a "total protein fraction" preparation (sample preparation method described in Section 3.2.2.1), for subsequent analysis. After 16 hours incubation with shaking, cellular fractions were prepared for analysis of 10 mL only. In addition to a total protein fraction, cytoplasmic soluble and insoluble protein fractions were prepared (sample preparation method described in Section 3.2.2.1 and Section 4.2.4.3.3.2). SDS-PAGE was performed to analyse all protein fractions, and followed by Coomassie-staining and/or Western blotting, as described in Section 3.2.2.2.

4.2.4.2.3 Cytoplasmic expression of pCold-GST_*homB*$_{19\text{-}511}$ construct in *E. coli* ArcticExpress (DE3)

Following plasmid propagation into *E. coli* DH5α cells, the expression construct pCold-GST_*homB*$_{19\text{-}511}$ was then transformed into *E. coli* ArcticExpress (DE3) (Agilent Technologies, USA) through a heat shock method as the following: 30 minutes on ice, then 30 seconds at 42°C, followed by 2 minutes on ice. The plasmid selection and the screening of protein expression were done as described in Section 4.2.4.2.2.

4.2.4.2.4 Cytoplasmic expression of pMAL-c5X_*homB*$_{19\text{-}511}$ construct in *E. coli* Rosetta-gami B (DE3) pLysS

pMAL-c5X_*homB*$_{19\text{-}511}$ construct transformation was performed as described in Section 4.2.4.2.1, and plasmid selection and screening of protein expression were done as described in Section 4.2.4.2.2. However, the temperature used after the OD_{600} reached around 0.6 was 12°C for 20 hours, and the concentration of IPTG used for inducing the cells was 15 µM. Furthermore, the shaking speed was slow (90 rpm).

4.2.4.3 Preparation of cellular fractions for analysis

4.2.4.3.1 Total protein fraction

The preparation of the total protein fraction described in Section 3.2.2.1.

4.2.4.3.2 Periplasmic fractions

The preparation of the periplasmic fractions described in Section 3.2.2.1.

4.2.4.3.3 Cytoplasmic soluble and insoluble protein fractions

4.2.4.3.3.1 Soluble and insoluble protein fractions in periplasmic expression

Soluble and insoluble protein fractions were prepared from the remaining pellet after preparing the periplasmic fractions. The pellet was resuspended in one tenth culture volume of 50 mM Tris-Cl (pH 8.0), 500 mM NaCl and 5% glycerol then probe sonication on ice (power 50%, 30 seconds ON, 60 seconds OFF) was used 5 times to lyse the cells in the suspension. Following centrifugation at 18,000 x *g* for 15 minutes (4°C), an equal amount of 2x reducing sample buffer (RSB) was mixed with the supernatant, which contained the cytoplasmic soluble protein fraction, and heat denatured for SDS-PAGE. The cell debris, which contained the cytoplasmic insoluble protein fraction, was resuspended in 100 µL PBS and then mixed with an equal volume of 2x RSB. This fraction was also heat-denatured for SDS-PAGE.

4.2.4.3.3.2 Soluble and insoluble protein fractions in cytoplasmic expression

A 10 mL sample of bacterial cells was taken from the overnight culture and harvested via centrifugation at 4,122 x *g* for 10 minutes (4°C). The pellet was then resuspended in either 1/10 culture volume of the following lysis buffer:

i. 50 mM Tris-Cl (pH 8.0), 500 mM NaCl and 5% glycerol and follow the same protocol described above. This protocol was applied for the vectors pET-22b(+) and pMAL-c5X.

ii. PBS contained 1 mg/mL lysozyme from chicken egg white (Merck, UK), which was used for bacterial cell wall lysis. After incubation with shaking at 37°C for 30 minutes, the suspension was subject to freezing in liquid nitrogen and then thawing at 37°C three times. Probe sonication on ice (power 50%, 30 seconds ON, 60 seconds OFF) was used 5 times to lyse the cells in the suspension. Following centrifugation at 18,000 x *g* for 15 minutes (4°C), an equal amount of 2x RSB was mixed with the supernatant, which contained the cytoplasmic soluble protein fraction, and heat denatured for SDS-PAGE. The cell debris, which contained the cytoplasmic insoluble protein fraction, was resuspended in 100 µL PBS and then mixed with an equal volume of 2x RSB. This fraction was also heat-denatured for SDS-PAGE. This protocol was selected for the vectors pCold-GST and pMAL-c5X.

4.2.4.4 Analysis of cellular fractions

4.2.4.4.1 SDS-PAGE

The protocol was carried out as described in Section 3.2.2.2.1.

4.2.4.4.2 Coomassie-staining

The protocol was carried out as described in Section 3.2.2.2.2.

4.2.4.4.3 Western blot

The protocol was carried out as described in Section 3.2.2.2.3. However, the antibodies used for immunodetection were as the following:

i. For the constructs of pOPE101, the primary antibody used was a biotinylated mouse anti-c-Myc at a 1:5,000 dilution, while the secondary antibody used was a Streptavidin-HRP at a 1:5,000 dilution.
ii. For the constructs of pET-22b(+), the primary antibody used was a mouse anti-His antibody (Merck, UK) at a 1:5,000 dilution, while the secondary antibody used was a goat anti-mouse IgG HRP conjugate (Merck, UK) at a 1:5,000 dilution.
iii. For the constructs of pCold-GST, the primary antibody used was a goat anti-GST antibody (Abcam, UK) at a 1:2,000 dilution, while the secondary antibody used was a goat anti-mouse IgG HRP conjugate at a 1:5,000 dilution.

iv. For the constructs of pMAL-c5X containing MBP fusion protein, the primary antibody used was an anti-MBP monoclonal antibody (New England Biolabs, USA) at a 1:10,000 dilution, while the secondary antibody used was a goat anti-mouse IgG HRP conjugate at a 1:5,000 dilution (Table 4.7).

Table 4.7: The antibodies used for immunodetection by Western blotting

Vector	Primary Antibody	Dilution	Secondary Antibody	Dilution
pOPE101	Biotinylated mouse anti-c-Myc	1:5,000	Streptavidin-HRP	1:5,000
pET-22b(+)	Mouse anti-His antibody	1,5000	a goat anti-mouse IgG HRP conjugate	1:5,000
pCold-GST	Goat anti-GST antibody	1:2,000	a goat anti-mouse IgG HRP conjugate	1:5,000
pMAL-c5X	Anti-MBP monoclonal antibody	1:10,000	a goat anti-mouse IgG HRP conjugate	1:5,000

4.2.5 Protein purification

4.2.5.1 Larger scale expression and cytoplasmic harvest

After the small-scale expression, *E. coli* Rosetta-gami B (DE3) pLysS cells, containing pMAL-c5X_*homB*$_{19\text{-}511}$ construct, were expressed on a larger scale. 90 µL of a glycerol stock was inoculated into 25 mL LB_{ACG} and incubated with shaking at 37°C overnight. 5 mL of the overnight culture was used to inoculate 400 mL LB_{amp} grown in a 2 L notched glass Erlenmeyer flask in a total volume of 2000 mL. These inoculated media were incubated and shaken at 37°C until the OD_{600} reached around 0.6. The temperature was then decreased to 12°C, and the cultures were shaken for 10 minutes at 90 rpm before inducing the cells with the final concentration of 15 µM IPTG for 20 hours. Cells were harvested through centrifugation at 4,122 x *g* for 45 minutes (4°C) and resuspended in 1/10 culture volume of PBS containing 1 mg/mL lysozyme. Following incubation with shaking at 37°C for 30 minutes, the suspension was subject to freezing in liquid nitrogen and then thawing at 37°C three times. Probe sonication on ice (power 50%, 30 seconds ON, 60 seconds OFF) was used 5 times to lyse the cells in the suspension. After centrifugation at 4,122 x *g* for 30 minutes, the supernatant was passed through a 0.22 µm filter to obtain the cytoplasmic soluble fraction, which was thereafter used for recombinant protein purification.

4.2.5.2 Affinity chromatography (AC)

AC was used to purify the recombinant $HomB_{19\text{-}511}$ proteins, synthesised from the pMAL-c5X_*homB*$_{19\text{-}511}$ construct, in the form of an MBP fusion. The extracts collected from the previous step were loaded onto a 5 mL MBPTrap™ HP column, containing Dextrin Sepharose High Performance medium (GE Healthcare, USA). Using the ÄKTA™ start system, a gradient elution program was run at a flow rate of 2.5 mL/min. The wash/binding buffer consisted of 20 mM Tris-HCl (pH 7.4) and 200 mM NaCl, while 10 mM maltose (Merck, UK) was used to elute MBP-$HomB_{19\text{-}511}$ proteins in a buffer consisting of 20 mM Tris-HCl (pH 7.4) and 200 mM NaCl. All fractions were then loaded onto a Tris-Glycine gel for SDS-PAGE analysis (Coomassie-staining). Following the analysis, the purified MBP-$HomB_{19\text{-}511}$ proteins were pooled and concentrated down to 1 mL by centrifugation at 4,122 x *g* (4°C) using Amicon Ultra-15 Centrifugal Filter Units (Merck Millipore, Germany) with a membrane ultrafiltration of 30 kDa. The protein was then diluted in the binding buffer to 1 mg/mL before the next step, which was cleavage of the MBP fusions from HomB proteins, or before storing the protein at -80°C.

4.2.5.3 Quantification of the recombinant MBP-$HomB_{19\text{-}511}$

The protocol was performed as described in Section 3.2.3.3. The extinction coefficient of 109030 M^{-1} cm^{-1} (calculated from the amino acid sequence of MBP-$HomB_{19\text{-}511}$) was used.

4.2.5.4 Cleavage of HomB$_{19-511}$ from MBP using FXa protease

FXa protease (New England Biolabs, USA) was used to cleave the MBP fusion protein from HomB by recognising the FXa cleavage site (I-E/D-G-R), cleaving after the arginine residue. 50 µL of 1 mg/mL FXa protease was mixed with 15 mL of 1 mg/mL HomB proteins. Furthermore, a final concentration of 2 mM $CaCl_2$ was added to the mixture which is required for optimal FXa protease activity. The mixture was incubated overnight at 23°C. FXa protease was then removed from the protein medium by Xarrest™ Agarose (Merck Millipore, Germany) according to the manufacturer's protocol.

4.2.5.5 Purification of recombinant HomB$_{19-511}$ by SEC

Following removal of FXa by Xarrest Agarose, the protein was concentrated down to 1 mL by centrifugation at 4,122 x *g* (4°C) using Amicon Ultra-15 Centrifugal Filter Units with a membrane ultrafiltration of 30 kDa. The 1 mL was then transferred to a 1.5 mL Eppendorf tube and centrifuged at 18,000 x *g* for 5 minutes (4°C). SEC was performed as a final step of protein purification by loading the supernatant of the concentrated sample onto a HiLoad® 16/600 Superdex® 75 pg gel filtration column (GE Healthcare, USA), which was pre-equilibrated with a 20 mM Tris-HCl (pH 7.4) and 200 mM NaCl, connected to the ÄKTA™ start system. The same buffer was also used for protein elution at a flow rate of 1 mL/min. All protein fractions and a volume from the mixture after FXa cleavage were then loaded onto

a Tris-Glycine gel for SDS-PAGE and Coomassie-staining after mixing with an equal amount of 2x RSB. Fractions that contained the cleaved HomB only were collected together and concentrated using Amicon Ultra-15 Centrifugal Filter Units with a membrane ultrafiltration of 30 kDa. The quantification of the recombinant HomB after cleavage was performed similar to the protocol described in Section 3.2.3.3. The extinction coefficient of 44310 M^{-1} cm^{-1} (calculated from the amino acid sequence of $HomB_{19-511}$) was used.

4.2.5.6 Mass spectrometry

The technique in-gel tryptic digestion was carried out in the Metabolomics and Proteomics Laboratory at University of York. Gel fragments were washed twice with 50% aqueous acetonitrile containing 25 mM ammonium bicarbonate (v:v), followed by one wash with acetonitrile before drying in a vacuum concentrator for 20 minutes. Sequencing-grade, modified porcine trypsin (Promega, USA) was dissolved in the 50 mM acetic acid supplied by the manufacturer, then diluted 5-fold with 25 mM ammonium bicarbonate to give a final trypsin concentration of 20 ng/μL. Ten μL of trypsin solution was added to the gel fragments to be rehydrated for 10 minutes, and then the gel fragments were covered by 25 mM ammonium bicarbonate solution and incubated overnight at 37°C.

A 1 μL aliquot of each peptide mixture was added to a ground steel MALDI target plate, followed immediately by an equal volume of a freshly-prepared

10 mg/mL solution of 4-hydroxy-α-cyano-cinnamic acid (Merck, UK) in 50% aqueous (v:v) acetonitrile containing 0.1%, trifluoroacetic acid (v:v).

A Bruker ultraflex III in reflectron mode, equipped with a Nd:YAG smart beam laser was used to obtain positive-ion MALDI mass spectra over a range of 800-4000 *m/z*. Final mass spectra were externally calibrated against an adjacent spot containing 6 peptides (des-Arg^1-Bradykinin, 904.681; Angiotensin I, 1296.685; Glu^1-Fibrinopeptide B, 1750.677; ACTH (1-17 clip), 2093.086; ACTH (18-39 clip), 2465.198; ACTH (7-38 clip), 3657.929.). A SNAP averaging algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) and a S/N threshold of 2 were used to obtain Monoisotopic masses.

For each spot the ten strongest precursors, with a S/N greater than 30, were chosen for MS/MS fragmentation. Fragmentation was done in LIFT mode without the introduction of a collision gas. The default calibration was used for MS/MS spectra, which were baseline-subtracted and smoothed (Savitsky-Golay, width 0.15 m/z, cycles 4); monoisotopic peak detection used a SNAP averaging algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) with a minimum S/N of 6. In order to perform spectral processing and peak list generation, Bruker flexAnalysis software (version 3.3) was used.

Tandem mass spectral data were submitted to database searching using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.6.1), through the Bruker ProteinScape interface (version 2.1). Search

criteria specified: Enzyme, Trypsin; Fixed modifications, Carbamidomethyl (C); Variable modifications, Oxidation (M) and Deamidated (NQ); Peptide tolerance, 60 ppm; MS/MS tolerance, 0.5 Da; Instrument, MALDI-TOF-TOF; Database, UniProt database (555594 sequences; 199016217 residues). Results were filtered to accept only peptides with an expected score of 0.05 or lower.

4.3 Results

4.3.1 Cloning *homB* gene fragments into expression vectors

PCR was required to amplify a fragment of *homB* gene from of *H. pylori* J99 genomic DNA in order to generate HomB expression constructs. These DNA fragments encoded the amino acids 9-511, which was thought to represent the predicted N-terminal extracellular domain of HomB, and the amino acids 19-511, which represent the predicted N-terminal extracellular domain. The putative C-terminal transmembrane domain (amino acids 512-668) was excluded from the amplifications based on the information provided by the NCBI.

4.3.2 Amplification of $homB_{9\text{-}511}$ gene by ordinary PCR

Amplification of *homB* gene from *H. pylori* J99 genomic DNA was the first step to generate HomB expression constructs. PCR was performed using the two primers listed in Table 4.2 in order to amplify *homB* gene. A band with an expected size at 1537 bp was visible on 1% agarose gel stained with SYBR Gold. Consequently, the corresponding gel slice was extracted and purified to obtain pure DNA product which was approximately 43.5 ng/μL. DNA sequencing confirmed that the DNA was mutated, as there were many insertions. *homB* and *homA* genes show high homologies in the 5' and 3' regions causing the primers to anneal to both genes, resulting in many insertions (Appendix Figure 8).

4.3.3 Amplification of *homB*$_{9-511}$ gene by overlap extension PCR

Overlap extension PCR was performed in order to avoid amplifying *homA* gene along with *homB* gene because there are unique sequences in the mid region of *homB* that are different from those in *homA*. Similar to the overlap extension PCR used for the amplification of *babC* gene, in the previous Chapter, three PCRs were carried out to amplify *homB*$_{9-511}$ gene: two PCRs to generate two DNA fragments with overlapping regions, and the third PCR was performed to gain a full-length *homB*$_{9-511}$ through mixing the two DNA fragments obtained from the first two PCRs. The first PCR was performed to amplify the first DNA fragment using the primers listed in Table 4.3, while the second PCR was performed to amplify the other DNA fragment, which shared overlapping region (114 bp) with the first DNA fragment, using the primers listed in Table 4.3. After successful amplification of the first and the second fragments of 971 bp and 680 bp respectively (Figure 4.5A), the third PCR was performed by mixing the first and the second DNA fragments to obtain the full-length *homB*$_{9-511}$ with a size at 1637 bp (Figure 4.5B), using two primers shown in Table 4.3, in which both primers introduced restriction sites which facilitates its insertion into the MCS of pOPE101 expression vector.

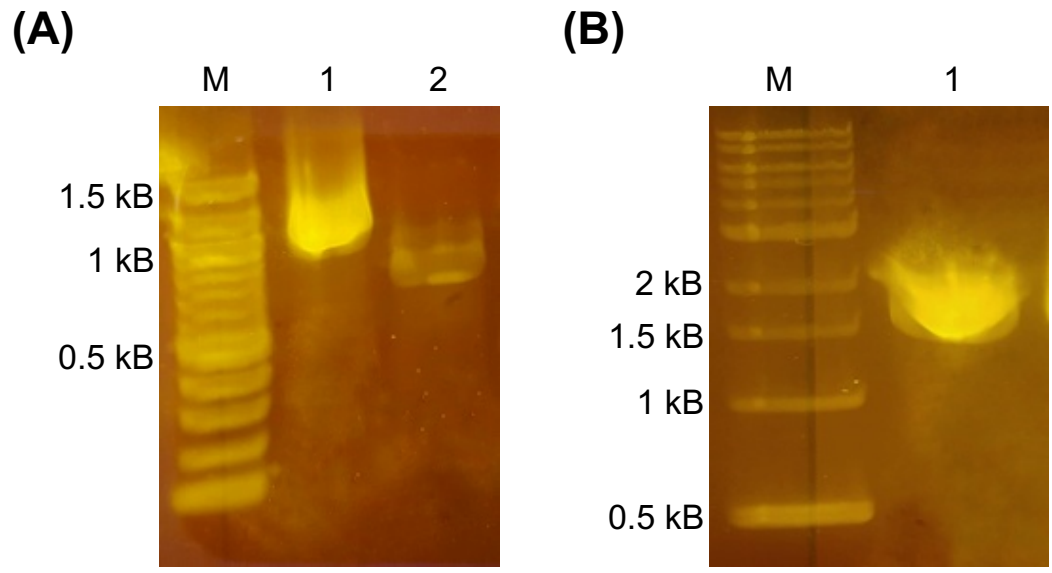


Figure 4.5: Overlap extension PCR for amplification of *homB*$_{9-511}$ gene
(A) Agarose gel stained with SYBR Gold showing the first and second *homB*$_{9-511}$ gene fragments generated from the first and the second PCRs. Lanes: M – 100 bp DNA ladder (New England Biolabs, USA); 1 – the first *homB*$_{9-511}$ fragment at the size of 971 bp that generated from the first PCR; 2 – the second *homB*$_{9-511}$ fragment at the size of 680 bp that generated from the second PCR. (B) SYBR Gold-stained agarose gel showing full-length *homB*$_{9-511}$ that was obtained from the third PCR by mixing the first and second *homB* fragments which share 114 bp overlap region. Lanes: M – 1 kb DNA ladder (New England Biolabs, USA); 1 – full-length *homB*$_{9-511}$ at the size of 1537 bp.

After PCR amplification, purified *homB* and pOPE101 were then double digested with NcoI and NotI (Figure 4.6). Following ligation and transformation into *E. coli* XL10 Gold for plasmid propagation, colony PCR was performed to confirm whether transformants harboured pOPE101_*homB*$_{9-511}$ recombinant plasmid (Appendix Figure 9). The primers shown in Table 4.5 were designed for colony PCR in order to span pOPE101 MCS, therefore, the size of amplicon revealed whether colonies were positive or negative for *homB*$_{9-511}$ insertion. It seems that the colonies were positive for *homB*$_{9-511}$ insertion as the band size was 1896 bp. Following

plasmid purification, it was confirmed that the plasmid extracted from the positive transformant had 100% sequence identity to *homB*$_{9-511}$. Furthermore, it was confirmed that *homB*$_{9-511}$ was cloned in frame with the pelB leader that translocates the protein.

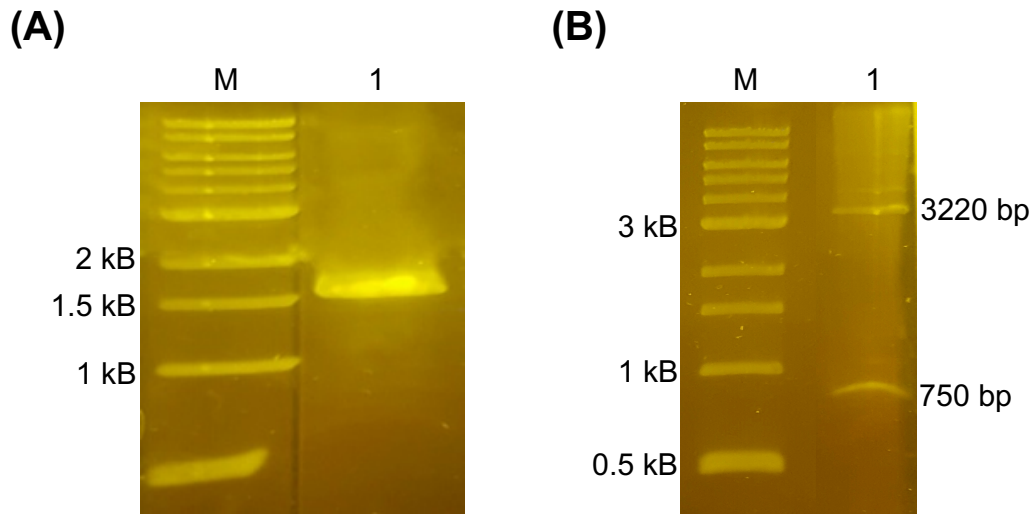


Figure 4.6: *homB*$_{9-511}$ gene and pOPE101 after digestion by NcoI and NotI

Agarose gel stained with SYBR Gold representing the results of digestion step for *homB*$_{9-511}$ and pOPE101 by NcoI and NotI. (A) It shows that *homB*$_{9-511}$ with an expected size at 1518 bp after digestion by NcoI and NotI at the restriction sites. Lanes: M – 1 kb DNA ladder (New England Biolabs, USA); 1 – digested *homB*$_{9-511}$ gene. (B) Two bands appear on agarose gel at the sizes of 3220 bp and 750 bp after digestion of pOPE101 by NcoI and NotI at the restriction sites. Lanes: M – 1 kb DNA ladder (Promega, USA); 1 – digested pOPE101.

Through this process, cloning of *homB*$_{9-511}$ and *homB*$_{19-511}$ genes into several vectors, which were designed for protein expression in *E. coli*, was successful using primers listed in Table 4.4. Each expression construct, if necessary, was then transformed into its selected expression host organism for protein expression screening.

4.3.4 Expression screening

Several expression host organisms harbouring $HomB_{9-511}$ or $HomB_{19-511}$ expression constructs were grown and screened on a small-scale culture in order to find an effective strategy for expressing HomB in a soluble form (Table 4.6).

4.3.4.1 Periplasmic expression

4.3.4.1.1 Periplasmic expression of pOPE101_*homB*$_{9-511}$ construct in *E. coli* XL10 Gold at 24°C

After the successful expression of BabA, BabC and LabA by our group using pOPE101 vector and *E. coli* XL10 Gold cells, the same cells and vector were used for assessing the expression of $HomB_{9-511}$. *homB*$_{9-511}$ is located downstream of a synthetic P/A1/04/03 lac promoter and a pelB leader, while it is located upstream of a c-Myc tag and a hexa-histidine tag (Figure 4.7). Similar to BabA and BabC expressions, 0.1 mM IPTG was used to induce *E. coli* XL10 Gold cells overnight at 24°C. Unfortunately, no $HomB_{9-511}$ was expressed in any cellular fraction, neither periplasmic nor cytoplasmic fractions after SDS-PAGE separation and analysis by InstantBlue staining and Western blot using c-Myc immunodetection (Figure 4.8).

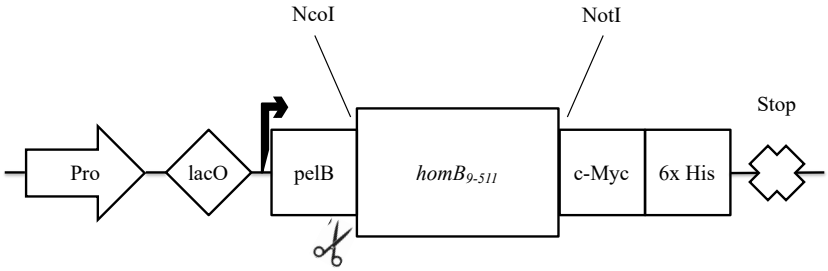


Figure 4.7: Schematic diagram of the pOPE101_ *homB*$_{9-511}$ expression cassette

pOPE101 was used for cloning *homB*$_{9-511}$ through NcoI and NotI restriction enzyme sites. The symbols and abbreviations denote the following: Pro – synthetic P/A1/04/03 lac promoter for initiation of expression; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence for transfer of expressed proteins to periplasmic space; *homB*$_{9-511}$ – *homB*$_{9-511}$ gene fragment that encodes the amino acids 9-511; c-Myc – c-Myc tag for immunodetection; 6x His – 6-histidine peptide tag for protein purification; and X/STOP – stop codon.

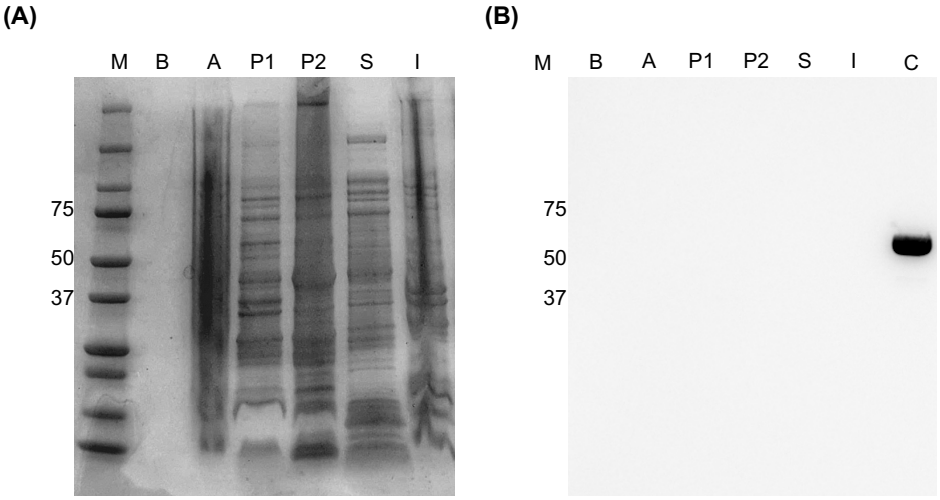


Figure 4.8: Expression of recombinant HomB$_{9-511}$ in *E. coli* XL10 Gold cells using pOPE101 vector at 24°C

InstantBlue-staining gel (A) and Western blot using c-Myc immunodetection (B) showing no expression of recombinant HomB$_{9-511}$ following overnight expression at 24°C in *E. coli* XL10 Gold cells using pOPE101 vector. The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 0.1 mM IPTG; A – total protein fraction after inducing the cells with 0.1 mM IPTG; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; S – cytoplasmic soluble fraction; I – cytoplasmic insoluble fraction; and C – BabC contains c-Myc tag was used as positive control (~60 kDa).

4.3.4.1.2 Periplasmic expression of pET22b_*homB*$_{9\text{-}511}$ construct in *E. coli* BL21 (DE3)

After the unsuccessful expression of HomB$_{9\text{-}511}$ in *E. coli* XL10 Gold cells using the pOPE101 vector, the next attempt was to assess the expression of HomB$_{9\text{-}511}$ using pET-22b(+) and *E. coli* BL21. *homB*$_{9\text{-}511}$ in the construct pET22b_ *homB*$_{9\text{-}511}$ (Appendix Figure 10) is located downstream of a T7 promoter, lac operator and a pelB leader, while it is located upstream of a hexa-histidine tag (Figure 4.9). *E. coli* BL21 cells were induced with 0.1 mM and 0.5 mM IPTG overnight at 24°C. Similar to pOPE101_*homB*$_{9\text{-}511}$ construct, no HomB$_{9\text{-}511}$ was expressed in any cellular fractions neither periplasmic nor cytoplasmic fractions after SDS-PAGE separation and analysis by Western blot using 6x His immunodetection (Figure 4.10).

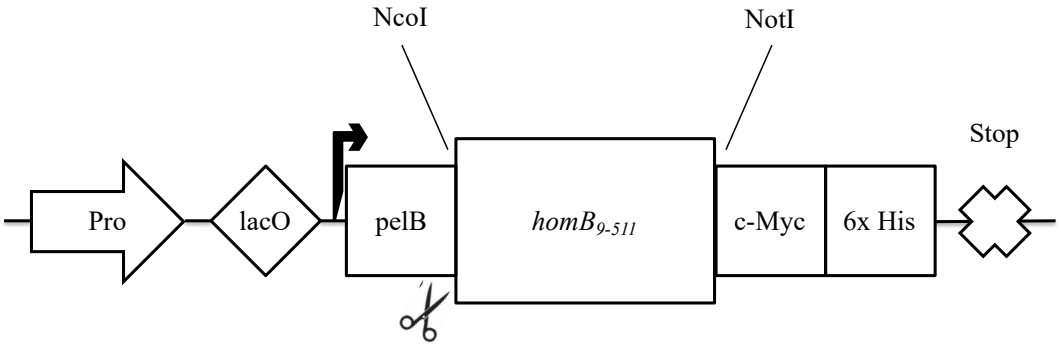


Figure 4.9: Schematic diagram of the pET22b_*homB*$_{9\text{-}511}$ expression cassette

pET-22b(+) was used for cloning *homB*$_{9\text{-}511}$ through NcoI and NotI restriction enzyme sites. The symbols and abbreviations denote the following: T7 – T7 promoter; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence for transfer of expressed proteins to periplasmic space; *homB*$_{9\text{-}511}$ – *homB*$_{9\text{-}511}$ gene fragment that encodes the amino acids 9-511; 6x His – 6-histidine peptide tag for immunodetection and protein purification; and X/STOP – stop codon.

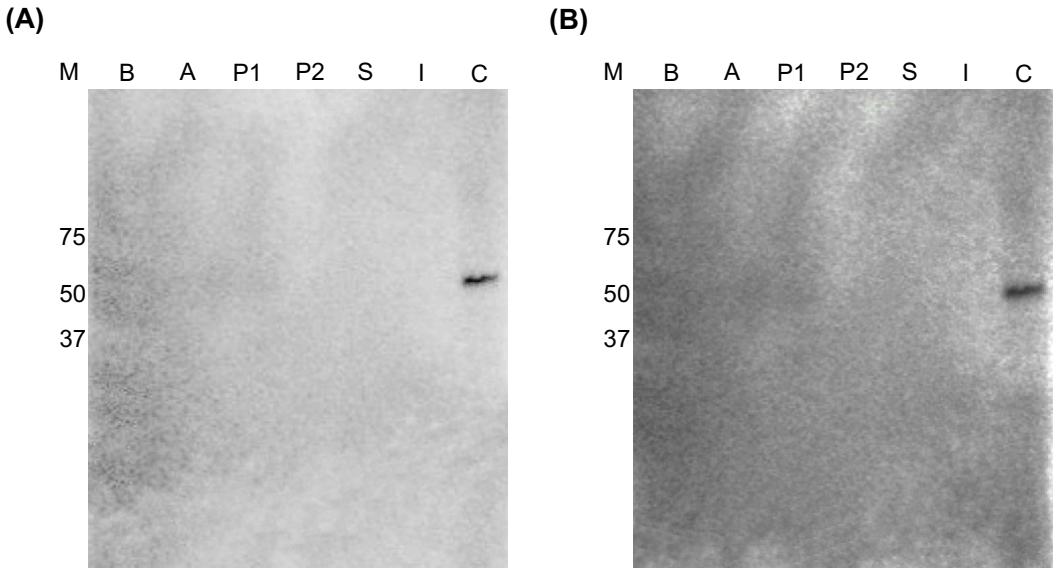


Figure 4.10: Expression of recombinant HomB$_{9-511}$ in *E. coli* BL21 cells using pET-22b(+) vector

Western blot using 6x His immunodetection showing no expression of recombinant HomB$_{9-511}$ following overnight expression at 24°C in *E. coli* BL21 cells using pET-22b(+) vector. The IPTG concentration used were 0.1 mM (A) and 0.5 mM (B). The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with IPTG; A – total protein fraction after inducing the cells with IPTG; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; S – cytoplasmic soluble fraction; I – cytoplasmic insoluble fraction; and C – BabC contains 6x His tag was used as positive control (~60 kDa).

4.3.4.1.3 Periplasmic expression of pET22b_*homB$_{9-511}$* construct in *E. coli* Origami B (DE3)

Following the unsuccessful expression of HomB$_{9-511}$ in BL21 cells using the vector pET-22b(+), the next step was to transform the construct pET22b_*homB$_{9-511}$* and assess its expression in *E. coli* Origami B cells. The bacterial cells were induced with 0.1 mM IPTG overnight at 24°C. Similarly to the first two attempts, no HomB$_{9-511}$ was expressed in any cellular fractions, as neither periplasmic nor cytoplasmic fractions after

SDS-PAGE separation and analysis by InstantBlue staining and Western blot using 6x His immunodetection demonstrated visible bands (Figure 4.11).

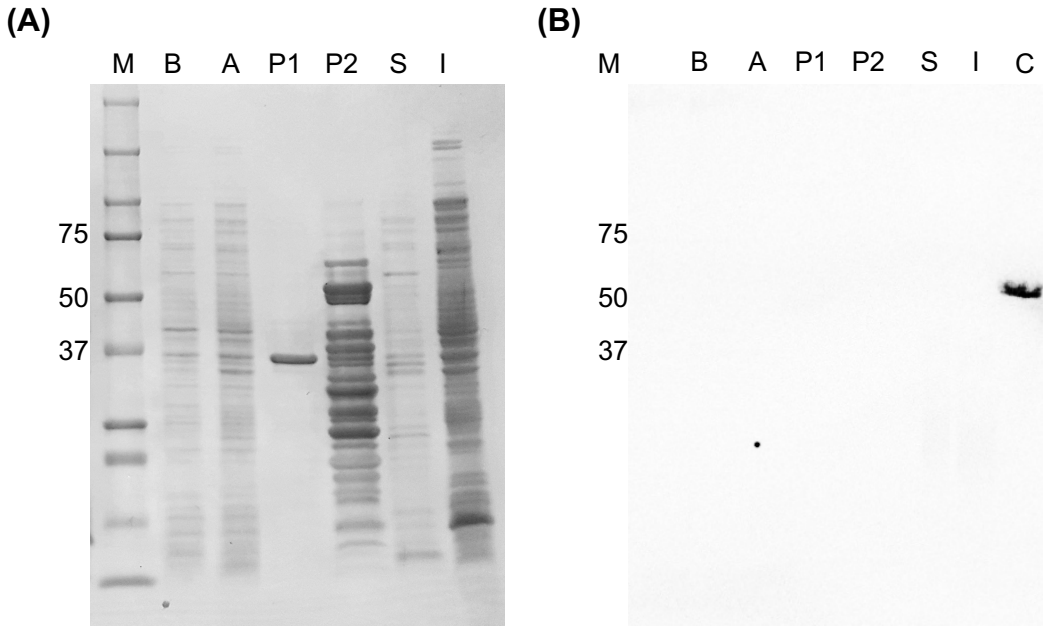


Figure 4.11: Expression of recombinant $HomB_{9-511}$ in *E. coli* Origami B (DE3) cells using pET-22b(+) vector

InstantBlue-staining gel (A) and Western blot using 6x His immunodetection (B) showing no expression of recombinant $HomB_{9-511}$ following overnight expression at 24°C in *E. coli* Origami B cells using pET-22b(+) vector. The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 0.1 mM IPTG; A – total protein fraction after inducing the cells with 0.1 mM IPTG; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; S – cytoplasmic soluble fraction; I – cytoplasmic insoluble fraction; and C – BabC contains c-Myc tag was used as positive control (~60 kDa).

4.3.4.1.4 Periplasmic expression of pOPE101_*homB*$_{9\text{-}511}$ construct in *E. coli* XL10 Gold at 16°C

The next attempt was to assess the expression of pOPE101_*homB*$_{9\text{-}511}$ again in *E. coli* XL10 Gold but with changing the overnight temperature from 24°C to 16°C and slowing down the overnight shaking speed to 90 rpm. Cellular fractions were separated by SDS-PAGE and then analysed by InstantBlue staining and Western blot using c-Myc immunodetection. The gel stained with InstantBlue suggested that there is some expression at the expected size, indicated by red stars (Figure 4.12A). However, the Western blot revealed that this expression was not for HomB$_{9\text{-}511}$ as the c-Myc immunodetection was negative (Figure 4.12B). Because of all the previous Western blot were negative, the thick band indicated by red star in the P2 fraction in Figure 4.12A was extracted for mass spectrometry analysis.

Mass spectrometry was used to confirm whether the protein is HomB$_{9\text{-}511}$. The excised band was subjected to tryptic digestion for peptide mapping. Unfortunately, the protein was matching with periplasmic oligopeptide-binding protein from *E. coli* strain K12 (Table 4.8).

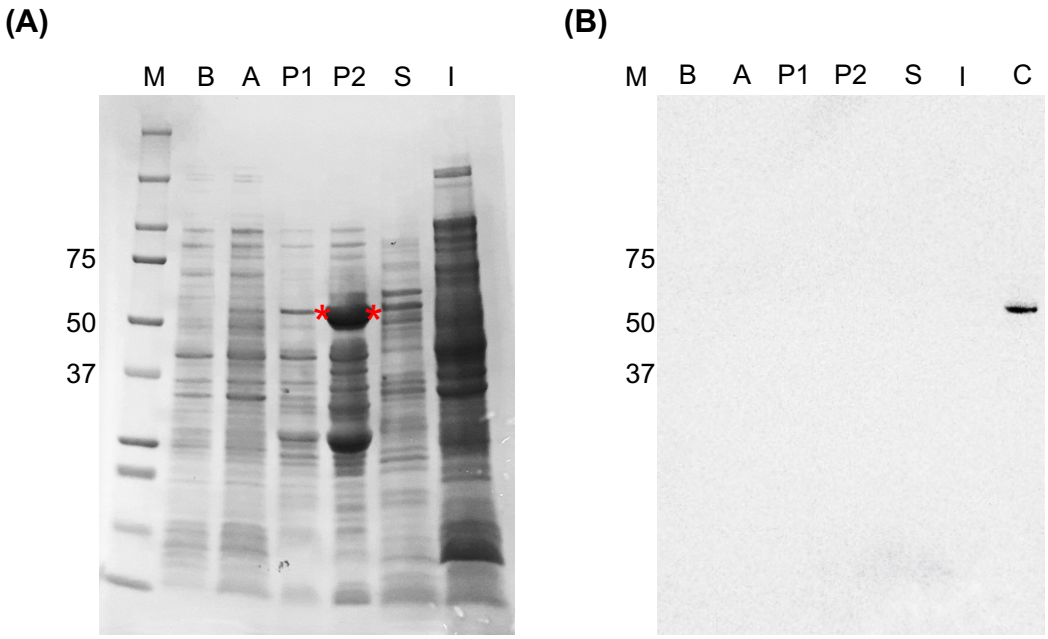


Figure 4.12: Expression of recombinant $HomB_{9\text{-}511}$ in *E. coli* XL10 Gold cells using pOPE101 vector at 16°C

InstantBlue-staining gel (A) and Western blot using c-Myc immunodetection (B) showing no expression of recombinant $HomB_{9\text{-}511}$ after overnight expression at 16°C and slowing down the overnight shaking speed. The band indicated by red stars thought to be $HomB_{9\text{-}511}$, but Western blot showed that the c-Myc immunodetection was negative. The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 0.1 mM IPTG; A – total protein fraction after inducing the cells with 0.1 mM IPTG; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; S – cytoplasmic soluble fraction; I – cytoplasmic insoluble fraction; and C – BabC contains c-Myc tag was used as positive control (~60 kDa).

Table 4.8: Tandem mass spectra data of periplasmic oligopeptide-binding protein from *E. coli* strain K12

Mass spectrometry confirmed that the excised band was not $HomB_{9-511}$.

Taxonomy ID: 83333

Periplasmic oligopeptide-binding protein [*E. coli* strain K12]

Mass: 60975 **Score:** 735 **Matches:** 6 (6) **Sequences:** 6 (6) **Calculated PI**: 6.05

Peptide match	Score	Expect	Rank	Unique
K.IEGVPESNISR.D	83	8.7e-06	1	U
K.WSDGTPVTAQDFVYSWQR.S	136	3.2e-11	1	U
K.TVINQVTYLPIASEVTDVNR.Y	189	1.5e-16	1	U
R.HQGTFDVAR.A	73	7.7e-05	1	U
K.AEQQLDKDSAIVPVYYYVNAR.L	165	4.4e-14	1	U
K.DSAIVPVYYYVNAR.L	88	2.4e-06	1	U

Protein sequence coverage: 14%. Matched peptides shown in bold red.

1	MTNITKRSLV	AAGVLAALMA	GNVALAADVP	AGVTLAEKQT	LVRNNGSEVQ
51	SLDPHK**IEGV**	**PESNISR**DLF	EGLLVSDLDG	HPAPGVAESW	DNKDAKVWTF
101	HLRKDAK**WSD**	**GTPVTAQDFV**	**YSWQR**SVDPN	TASPYASYLQ	YGHIAGIDEI
151	LEGKKPITDL	GVKAIDDHTL	EVTLSEPVPY	FYKLLVHPST	SPVPKAAIEK
201	FGEKWTQPGN	IVTNGAYTLK	DWVVNERIVL	ERSPTYWNNA	**KTVINQVTYL**
251	**PIASEVTDVN**	**R**YRSGEIDMT	NNSMPIELFQ	KLKKEIPDEV	HVDPYLCTYY
301	YEINNQKPPF	NDVRVRTALK	LGMDRDIIVN	KVKAQGNMPA	YGYTPPYTDG
351	AKLTQPEWFG	WSQEKRNEEA	KKLLAEAGYT	ADKPLTINLL	YNTSDLHKKL
401	AIAASSLWKK	NIGVNVKLVN	QEWKTFLDTR	**HQGTFDVARA**	GWCADYNEPT
451	SFLNTMLSNS	SMNTAHYKSP	AFDSIMAETL	KVTDEAQRTA	LYTK**AEQQLD**
501	**KDSAIVPVYY**	**YVNAR**LVKPW	VGGYTGKDPL	DNTYTRNMYI	VKH

4.3.4.1.5 Periplasmic expression of pOPE101_*homB*$_{19\text{-}511}$ construct in *E. coli* XL10 Gold

After excluding the full sequence of signal peptide consisting of the N-terminal 18 amino acids using SignalP 3.0 and 5.0 servers, the next attempt was to assess the expression of HomB$_{19\text{-}511}$ using pOPE101 vector and *E. coli* XL10 Gold cells (Figure 4.13) (Appendix Figure 11). The bacterial cells were induced with 0.1 mM and 1 mM IPTG overnight at 24°C, and the cellular fractions were separated by SDS-PAGE then analysed by InstantBlue staining and Western blot using c-Myc immunodetection. Finally, the protein was expressed after excluding the full sequence of signal peptide, but unfortunately HomB$_{19\text{-}511}$ was only seen in the cytoplasmic insoluble fraction of the cell lysates (Figure 4.14). Although the protein was successfully expressed, however, we concluded that this strategy had also failed, because the protein was expressed in insoluble form, which could cause problems with the planned downstream enzymatic experiments described in Chapter 5.

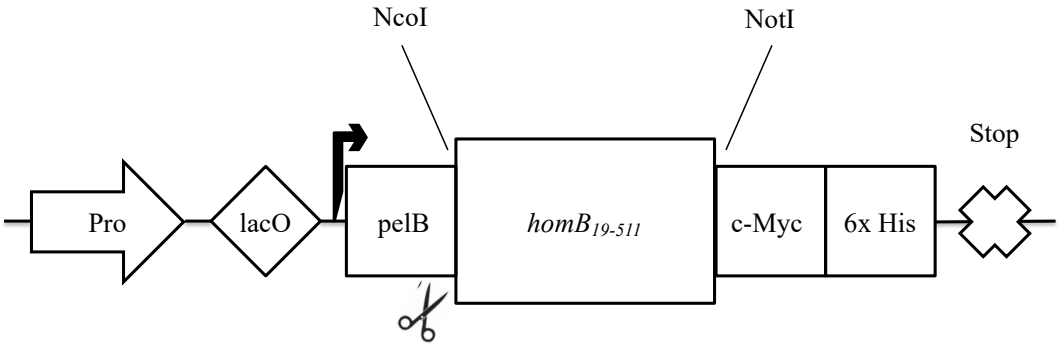


Figure 4.13: Schematic diagram of the pOPE101_*homB*$_{19\text{-}511}$ expression cassette

pOPE101 was used for cloning *homB*$_{19\text{-}511}$ through NcoI and NotI restriction enzyme sites. The symbols and abbreviations denote the following: Pro – synthetic P/A1/04/03 lac promoter for initiation of expression; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence for transfer of expressed proteins to periplasmic space; *homB*$_{19\text{-}511}$ – *homB*$_{19\text{-}511}$ gene fragment that encodes the amino acids 19-511; c-Myc – c-Myc tag for immunodetection; 6x His – 6-histidine peptide tag for protein purification; and X/STOP – stop codon.

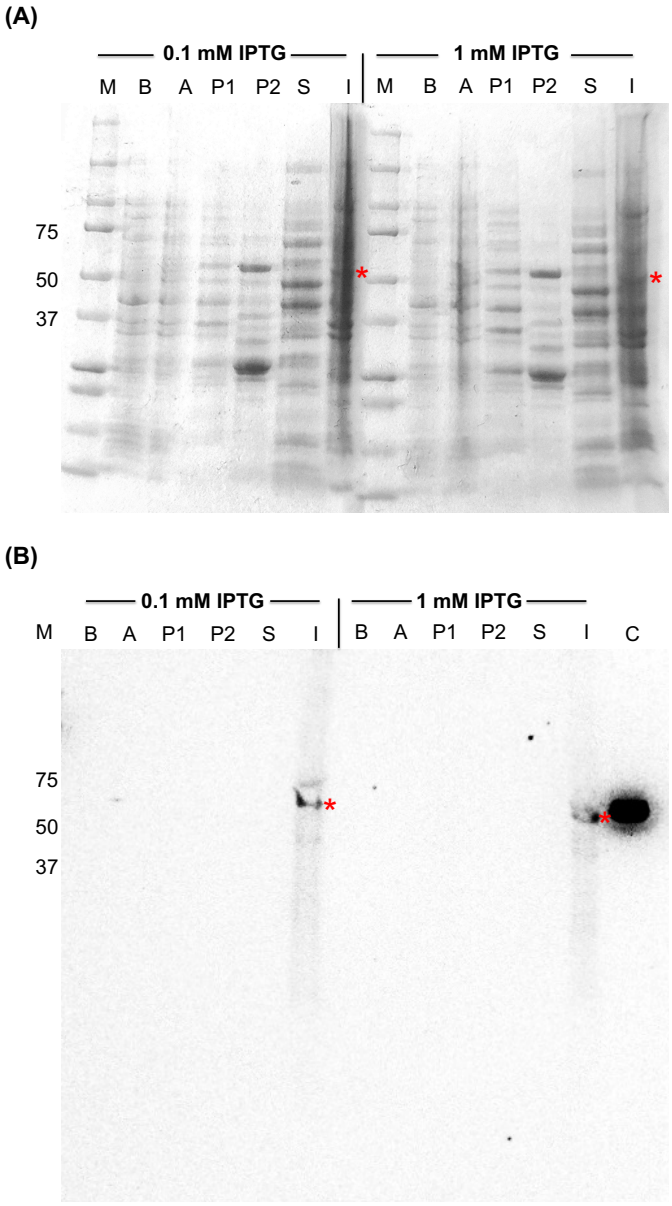


Figure 4.14: Expression of recombinant HomB$_{19\text{-}511}$ in *E. coli* XL10 Gold cells using pOPE101 vector

InstantBlue-staining gel (A) and Western blot using c-Myc immunodetection (B) showing the localisation of recombinant HomB$_{9\text{-}511}$, which is indicated by red stars at expected molecular weight ~57.3 kDa, in the cytoplasmic insoluble fraction of *E. coli* XL10 Gold cells after overnight expression at 24°C using pOPE101 vector. The cells were induced with 0.1 mM and 1mM IPTG. The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with IPTG; A – total protein fraction after inducing the cells with IPTG; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; S – cytoplasmic soluble fraction; I – cytoplasmic insoluble fraction; and C – BabC contains c-Myc tag was used as positive control (~60 kDa).

4.3.4.1.6 Periplasmic expression of pOPE101_*homB*$_{19\text{-}511R}$ construct in *E. coli* XL10 Gold

After the successful expression of HomB$_{19\text{-}511}$ but as insoluble form, improvement of protein solubility was required. To do this, a deca-arginine (242) was inserted into pOPE101_*homB*$_{19\text{-}511}$ between *homB*$_{19\text{-}511}$ and the c-Myc tag (Figure 4.15), and *E. coli* XL10 Gold cells were again used as expression host cells. The cells were induced with 0.1 mM IPTG overnight at 24°C, and the cellular fractions were separated by SDS-PAGE and then analysed by Western blot using c-Myc immunodetection. However, the Western blot showed that no HomB$_{19\text{-}511R}$ was found in any cellular fractions (Figure 4.16), and unfortunately the protein returned back as before without expression. Therefore, cytoplasmic expression was taken into consideration to be done as all the attempts for periplasmic expression had failed.

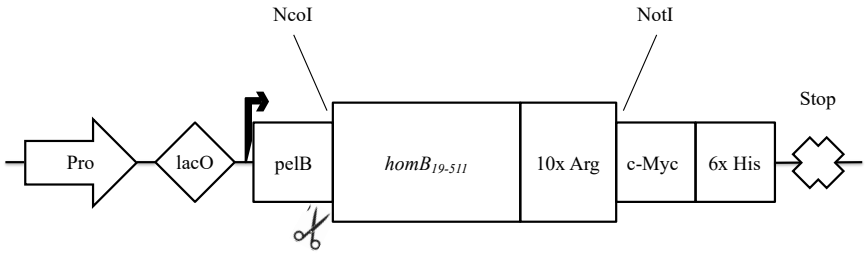


Figure 4.15: Schematic diagram of the pOPE101_*homB*$_{19\text{-}511R}$ expression cassette

pOPE101 was used for cloning *homB*$_{19\text{-}511R}$ through NcoI and NotI restriction enzyme sites. The symbols and abbreviations denote the following: Pro – synthetic P/A1/04/03 lac promoter for initiation of expression; lacO – lac operator; arrow – start codon; pelB – cleavable pelB leader sequence for transfer of expressed proteins to periplasmic space; *homB*$_{19\text{-}511}$ – *homB*$_{19\text{-}511}$ gene fragment that encodes the amino acids 19-511; 10x Arg – 10-arginine peptide tag for improvement of protein solubility; c-Myc – c-Myc tag for immunodetection; 6x His – 6-histidine peptide tag for protein purification; and X/STOP – stop codon.

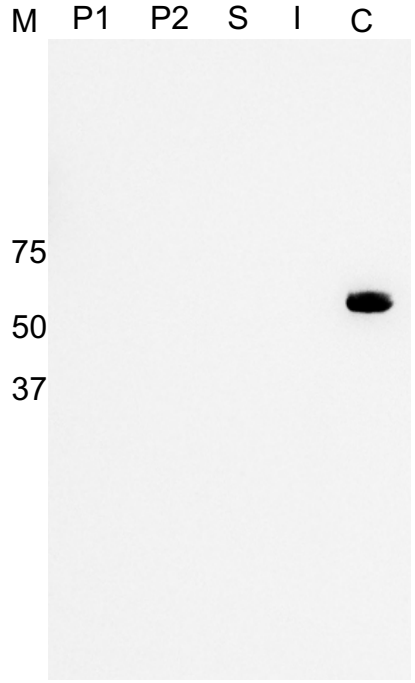


Figure 4.16: Expression of recombinant $HomB_{19\text{-}511R}$ in *E. coli* XL10 Gold cells using pOPE101 vector

Western blot using c-Myc immunodetection showing no expression of recombinant $HomB_{19\text{-}511R}$ following overnight expression at 24°C in *E. coli* XL10 Gold cells using pOPE101 vector. The abbreviations denote the following: M – molecular weight marker in kDa; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; S – cytoplasmic soluble fraction; I – cytoplasmic insoluble fraction; and C – BabC contains c-Myc tag was used as positive control (~60 kDa).

4.3.4.2 Cytoplasmic expression

4.3.4.2.1 Cytoplasmic expression of pET22b(ΔpelB)_*homB*$_{19\text{-}511}$ construct in *E. coli* Rosetta-gami B (DE3) pLysS

After the unsuccessful periplasmic expression, the next step was to express $HomB_{19\text{-}511}$ in the cytoplasm of *E. coli*. The first attempt of cytoplasmic expression was to express the protein using pET-22b(+) after removing the pelB leader from the vector (ΔpelB), which was then transformed into *E. coli* Rosetta-gami B. By removing the pelB leader, the expression of HomB from

pET22b(ΔpelB)_*homB*$_{19\text{-}511}$ construct (Figure 4.17) (Appendix Figure 12) would be confined to the cytoplasmic space. The bacterial cells were induced with 0.5 mM IPTG overnight at 24°C, and the cellular fractions were separated by SDS-PAGE then analysed by Western blot using 6x His immunodetection. Disappointingly, the Western blot showed that no HomB$_{19\text{-}511}$ was found in any cellular fractions (Figure 4.18). As a result, the next step was to express the protein with the presence of fusion protein.

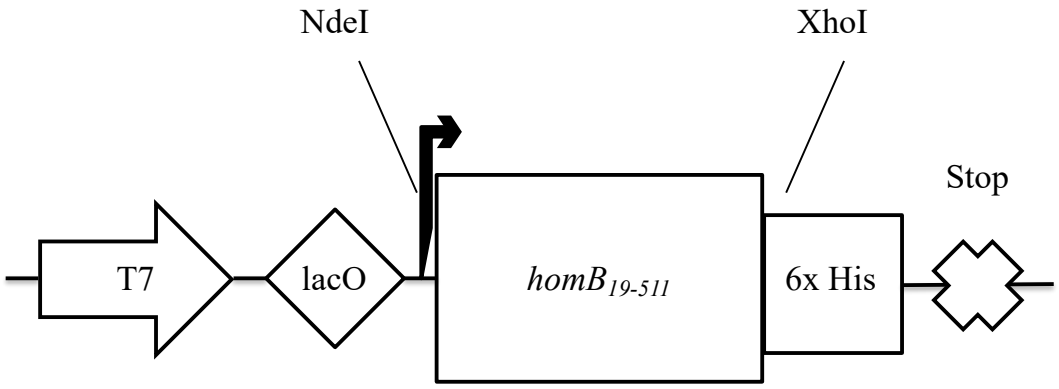


Figure 4.17: Schematic diagram of the pET22b(ΔpelB)_*homB*$_{19\text{-}511}$ expression cassette

pET-22b(+) was used for cloning *homB*$_{19\text{-}511}$ through NdeI and XhoI restriction enzyme sites. The symbols and abbreviations denote the following: T7 – T7 promoter; lacO – lac operator; arrow – start codon; *homB*$_{19\text{-}511}$ – *homB*$_{19\text{-}511}$ gene fragment that encodes the amino acids 19-511; 6x His – 6-histidine peptide tag for immunodetection and protein purification; and X/STOP – stop codon.

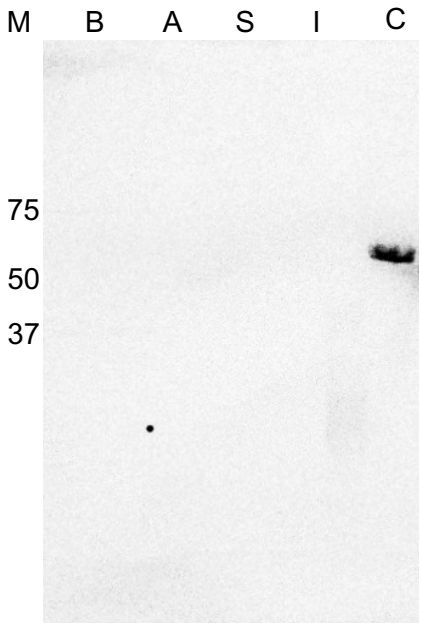


Figure 4.18: Expression of recombinant HomB$_{19\text{-}511}$ in *E. coli* Rosetta-gami B (DE3) pLysS cells using pET22b(ΔpelB) vector
Western blot using 6x His immunodetection showing no expression of recombinant HomB$_{19\text{-}511}$ following overnight expression at 24°C in *E. coli* Rosetta-gami B cells using pET22b(ΔpelB) vector. The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 0.5 mM IPTG; A – total protein fraction after inducing the cells with 0.5 mM IPTG; P1 – first periplasmic protein fraction; P2 – second periplasmic protein fraction; S – cytoplasmic soluble fraction; I – cytoplasmic insoluble fraction; and C – BabC contains 6x His tag was used as positive control (~60 kDa).

4.3.4.2.2 Cytoplasmic expression of pCold-GST_*homB$_{19\text{-}511}$* construct in *E. coli* BL21 (DE3)

The following attempt of cytoplasmic expression was to express HomB$_{19\text{-}511}$ protein after its incorporation with fusion protein. Thus, GST fusion protein was used as the first attempt to be incorporated into HomB$_{19\text{-}511}$ through choosing pCold-GST vector. *homB$_{19\text{-}511}$* in the construct pCold-GST_*homB$_{19\text{-}511}$* (Appendix Figure 13) is located downstream of the following: a *cspA* promoter for expression initiation, lac operator,

a TEE sequence (MNHKV) to enhance translation initiation, a hexa-histidine tag for either protein purification or immunodetection, and GST which used as a fusion protein for improvement of protein solubility, protein purification and immunodetection (Figure 4.19) (213).

The construct was transformed into *E. coli* BL21 cells, which were then induced with 0.5 mM IPTG overnight at 10°C. The cellular fractions were separated by SDS-PAGE then analysed by InstantBlue staining and Western blot using GST immunodetection. Unfortunately, $HomB_{19\text{-}511}$ was only seen in the cytoplasmic insoluble fraction of the cell lysates (Figure 4.20). Therefore, another *E. coli* strain was needed for the next attempt.

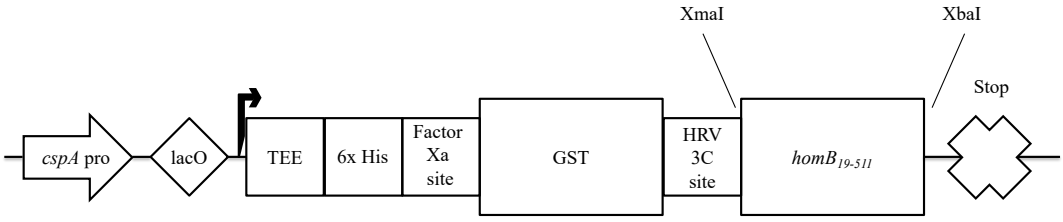


Figure 4.19: Schematic diagram of the pCold-GST_*homB*$_{19\text{-}511}$ expression cassette

pCold-GST was used for cloning *homB*$_{19\text{-}511}$ through XmaI and XbaI restriction enzyme sites. The symbols and abbreviations denote the following: *cspA* pro – *cspA* promoter; lacO – lac operator; arrow – start codon; TEE – a translation-enhancing element which is a five-codon sequence (MNHKV) to enhance translation initiation; 6x His – 6-histidine peptide tag can be used for immunodetection or protein purification; FXa site – a sequence (IEGR) which is recognised by FXa protease which cleaves after this sequence; GST – glutathione S-transferase used as a fusion protein for improvement of protein solubility, protein purification and immunodetection; HRV 3C site – a sequence (LEVLFQGP) which is recognised by human rhinovirus 3C protease that cleaves after glutamine; *homB*$_{19\text{-}511}$ – *homB*$_{19\text{-}511}$ gene fragment that encodes the amino acids 19-511; and X/STOP – stop codon.

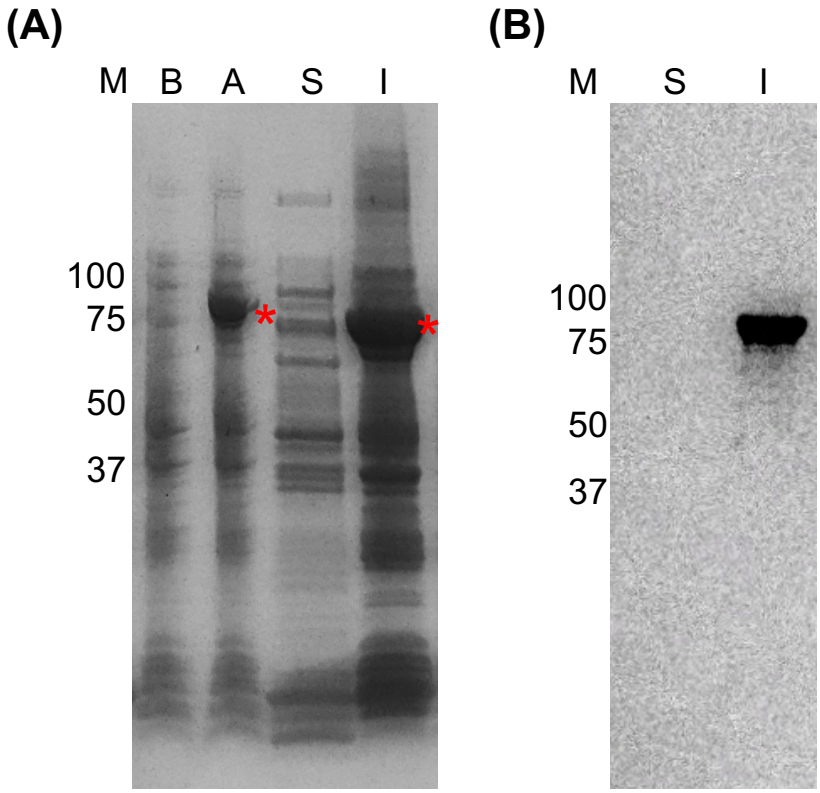


Figure 4.20: Expression of recombinant HomB$_{19-511}$ in *E. coli* BL21 (DE3) cells using pCold-GST vector

InstantBlue-staining gel (A) and Western blot using GST immunodetection (B) showing the localisation of recombinant HomB$_{19-511}$, which is indicated by red stars at expected molecular weight ~83.7 kDa, in the cytoplasmic insoluble fraction of *E. coli* BL21 cells after overnight expression at 10°C using pCold-GST vector. The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 0.5 mM IPTG; A – total protein fraction after inducing the cells with 0.5 mM IPTG; S – cytoplasmic soluble fraction; and I – cytoplasmic insoluble fraction.

4.3.4.2.3 Cytoplasmic expression of pCold-GST_*homB$_{19-511}$* construct in *E. coli* ArcticExpress (DE3)

Following the unsuccessful expression of pCold-GST_*homB$_{19-511}$* construct as a soluble protein using *E. coli* BL21 cells, this construct was then transformed into *E. coli* ArcticExpress (DE3) cells in order to assess its expression using these cells. The bacterial cells were induced with 0.5 mM

IPTG overnight at 10°C. The cellular fractions were separated by SDS-PAGE then analysed by InstantBlue staining and Western blot using GST immunodetection. Unluckily, $HomB_{19\text{-}511}$ was again only seen in the cytoplasmic insoluble fraction of the cell lysates (Figure 4.21). As a result, we decided to use another fusion protein for the next attempt.

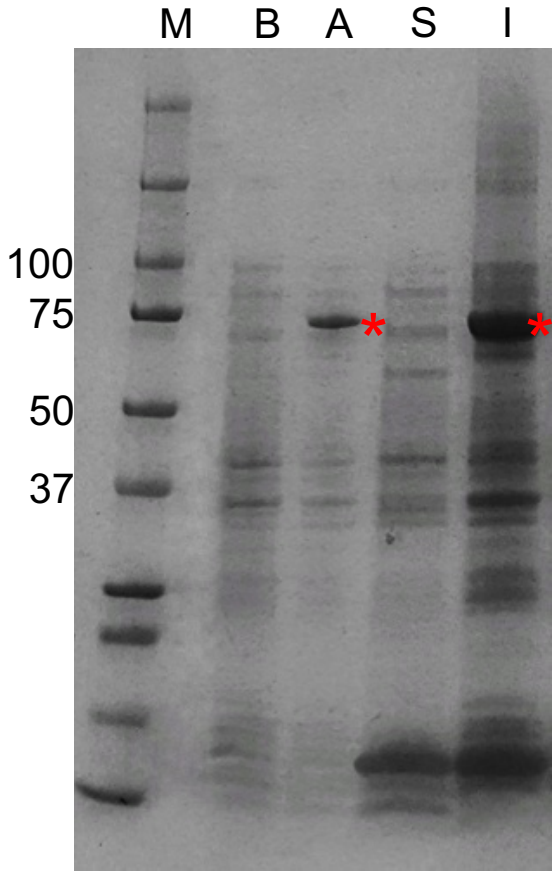


Figure 4.21: Expression of recombinant $HomB_{19\text{-}511}$ in *E. coli* ArcticExpress (DE3) cells using pCold-GST vector

InstantBlue-staining gel showing the localisation of recombinant $HomB_{19\text{-}511}$, which is indicated by red stars at expected molecular weight ~83.7 kDa, in the cytoplasmic insoluble fraction of *E. coli* ArcticExpress cells after overnight expression at 10°C using pCold-GST vector. The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 0.5 mM IPTG; A – total protein fraction after inducing the cells with 0.5 mM IPTG; S – cytoplasmic soluble fraction; and I – cytoplasmic insoluble fraction.

4.3.4.2.4 Cytoplasmic expression of pMAL-c5X_*homB*$_{19\text{-}511}$ construct in *E. coli* Rosetta-gami B (DE3) pLysS

The next option was to try to express HomB as MBP fusion protein. For this purpose, HomB$_{19\text{-}511}$ was subcloned into the pMAL-c5X vector. *homB*$_{19\text{-}511}$ in the construct pMAL-c5X_*homB*$_{19\text{-}511}$ (Appendix Figure 14) is located downstream of the following: a *tac* promoter, lac operator, MBP used as a fusion protein for improvement of protein solubility, protein purification and immunodetection, and FXa site which is a sequence (IEGR) recognised by FXa protease that cleaves after this sequence (Figure 4.22). The construct was transformed into *E. coli* Rosetta-gami B cells, which were then induced with 15 µM IPTG overnight at 12°C. The cellular fractions were separated by SDS-PAGE and then analysed by InstantBlue staining and Western blot using MBP immunodetection. Interestingly, HomB$_{19\text{-}511}$ was successfully expressed as a soluble protein at the expected molecular weight ~97.6 kDa. Although the bulk of the protein was found in the cytoplasmic insoluble fraction after analysing the cellular fractions, there was some protein in the cytoplasmic soluble fraction (Figure 4.23). It was also shown that the protein obtained using PBS (pH 7.4) containing 1 mg/mL lysozyme was more visible in InstantBlue-staining gel than the protein obtained using Tris-Cl (pH 8.0), 500 mM NaCl and 5% glycerol.

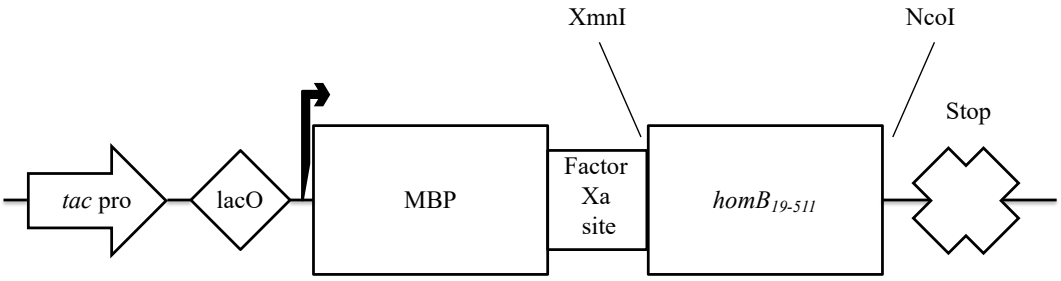


Figure 4.22: Schematic diagram of the pMAL-c5X_*homB*$_{19\text{-}511}$ expression cassette

pMAL-c5X was used for cloning *homB*$_{19\text{-}511}$ through XmnI and NcoI restriction enzyme sites. The symbols and abbreviations denote the following: *tac* pro – *tac* promoter; lacO – lac operator; arrow – start codon; MBP – maltose-binding protein used as a fusion protein for improvement of protein solubility, protein purification and immunodetection; FXa site – a sequence (IEGR) which is recognised by FXa protease which cleaves after this sequence; *homB*$_{19\text{-}511}$ – *homB*$_{19\text{-}511}$ gene fragment that encodes the amino acids 19-511; and X/STOP – stop codon.

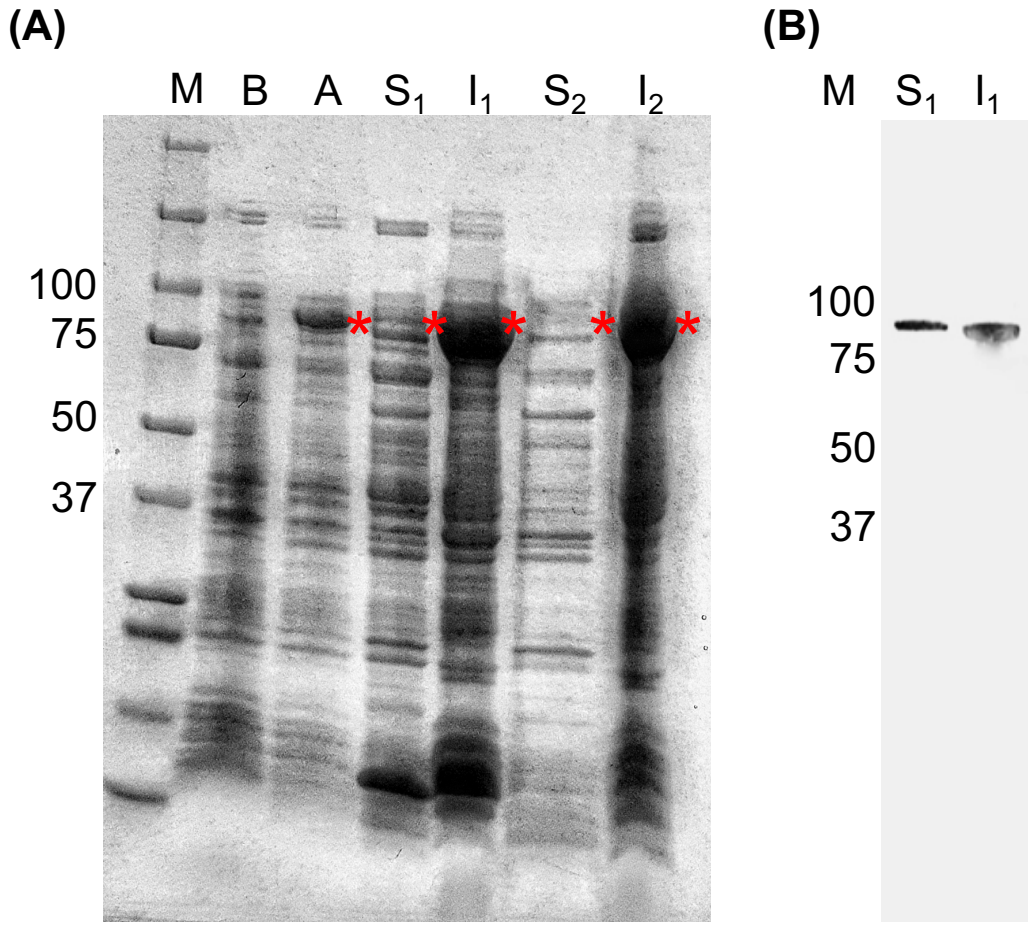


Figure 4.23: Expression of recombinant HomB$_{19\text{-}511}$ in *E. coli* Rosetta-gami B (DE3) pLysS cells using pMAL-c5X vector

InstantBlue-staining gel (A) and Western blot using MBP immunodetection (B) showing the localisation of recombinant HomB$_{19\text{-}511}$, which is indicated by red stars at expected molecular weight ~97,6 kDa, in the cytoplasmic soluble fraction of *E. coli* Rosetta-gami B cells after overnight expression at 12°C using pMAL-c5X vector. However, the bulk of HomB$_{19\text{-}511}$ protein was found in the cytoplasmic insoluble fraction. The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 15 µM IPTG; A – total protein fraction after inducing the cells with 15 µM IPTG; S_1 – cytoplasmic soluble fraction using PBS (pH 7.4) containing 1 mg/mL lysozyme; I_1 – cytoplasmic insoluble fraction using PBS (pH 7.4) containing 1 mg/mL lysozyme; S_2 – cytoplasmic soluble fraction using 50 mM Tris-Cl (pH 8.0), 500 mM NaCl and 5% glycerol; and I_2 – cytoplasmic insoluble fraction using 50 mM Tris-Cl (pH 8.0), 500 mM NaCl and 5% glycerol.

MBP-HomB$_{19\text{-}511}$ was obtained in soluble fraction following several attempts using different temperatures, IPTG concentrations, shaking speeds, and durations during induction of the cells as shown in Table 4.9.

Table 4.9: Optimisation of HomB$_{19\text{-}511}$ expression on a small-scale using MBP fusion protein and *E. coli* Rosetta-gami B (DE3) pLysS

Temperature	IPTG concentration	Shaking speed	Duration	Outcomes
18°C	0.5 mM	150 rpm	16 hr	Insoluble protein
18°C	0.1 mM	150 rpm	16 hr	Insoluble protein
12°C	15 µM	90 rpm	20 hr	Soluble and insoluble proteins

After the successful optimisation of HomB expression on a small-scale using MBP fusion protein, the protein was expressed on a larger scale.

4.3.5 Purification of recombinant MBP-HomB$_{19\text{-}511}$

4.3.5.1 Purification of recombinant MBP-HomB$_{19\text{-}511}$ by AC

After expressing HomB$_{19\text{-}511}$ on a small-scale using MBP fusion protein, larger scale expression and purification were performed for further investigation. The buffer used for cytoplasmic extraction was the PBS (pH 7.4) containing 1 mg/mL lysozyme. AC, an initial purification step, was used to purify MBP-HomB$_{19\text{-}511}$ proteins from the soluble cytoplasmic extract of *E. coli* Rosetta-gami B (DE3) pLysS cells through the MBP fusion proteins which bound to dextrin ligand in the column primarily by hydrogen bonds. Fractions containing MBP-HomB$_{19\text{-}511}$ were eluted by 10 mM maltose (Figure 4.24), separated through SDS-PAGE and visualised by InstantBlue. Their purity was relatively low after AC (Figure 4.25). After staining with InstantBlue, AC-elution fractions were pooled using Amicon Ultra-15 Centrifugal Filter Units with a membrane ultrafiltration of 30 kDa in order to concentrate MBP-HomB$_{19\text{-}511}$ proteins and to eliminate the maltose used in the elution buffer for elution. The yield of the obtained protein was 20 mg per 1 L of bacterial culture. MBP-HomB$_{19\text{-}511}$ proteins were then resuspended in the binding buffer to 1 mg/mL before cleavage of the MBP fusions from HomB proteins, or before storing the protein at -80°C.

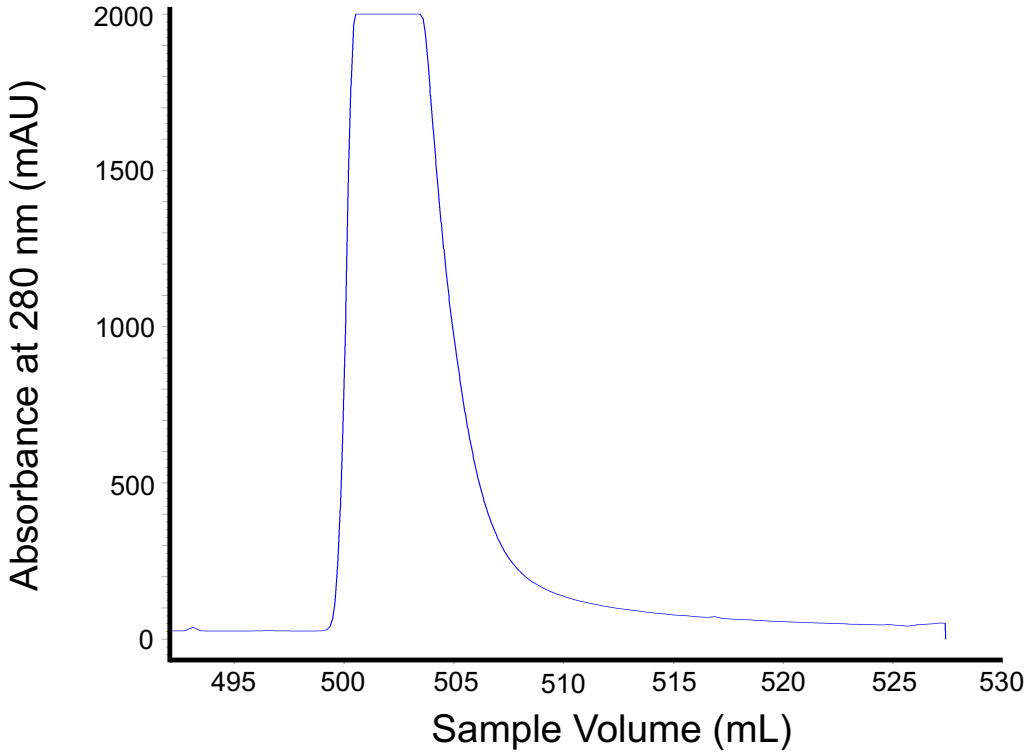


Figure 4.24: AC purification of MBP-HomB$_{19-511}$
Overlay of AC trace of MBP-HomB$_{19-511}$. The peak appeared after elution of the protein by a gradient of 10 mM maltose.

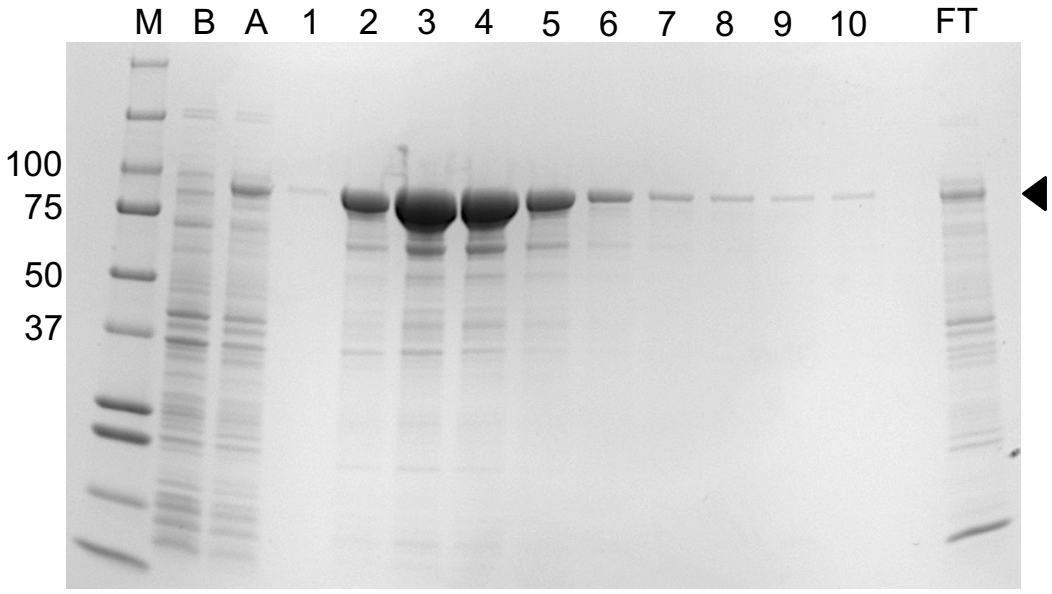


Figure 4.25: Purification of MBP-HomB$_{19-511}$ by AC

10 AC-elution fractions, representing the peak from the AC purification of MBP-HomB$_{19-511}$ (fractions 6-15), were collected after protein purification. Equal loading volume was used in all gel wells, and proteins were separated through SDS-PAGE and analysed by InstantBlue stain. The black arrowhead indicates the recombinant MBP-HomB$_{19-511}$ (~97.6 kDa). The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 15 µM IPTG; A – total protein fraction after inducing the cells with 15 µM IPTG; 1-10 – AC-elution fractions of MBP-HomB$_{19-511}$ protein; and FT – flow through.

4.3.5.2 Mass spectrometry for MBP-HomB$_{19\text{-}511}$

Upon the analysis of AC-elution fractions by InstantBlue staining, mass spectrometry was performed for the band at the expected molecular weight ~97.6 kDa (Figure 4.26). The band was excised and subjected to tryptic digestion for peptide mapping. The mass spectrometry analysis confirmed that the band was for MBP-HomB$_{19\text{-}511}$ based on the peptide match shown in Table 4.10.

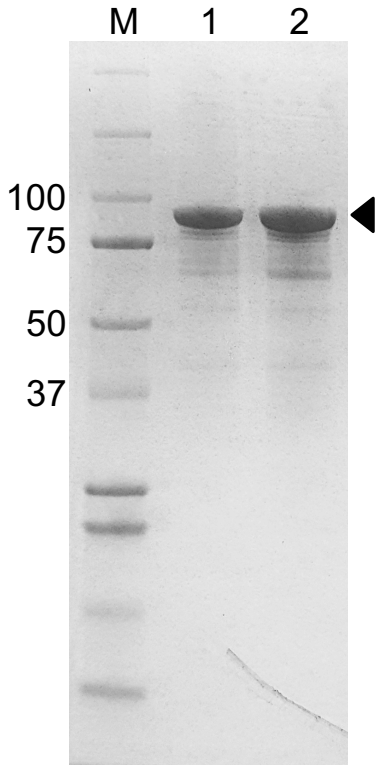


Figure 4.26: An InstantBlue-staining gel of MBP-HomB$_{19\text{-}511}$ for Mass spectrometry following AC

An InstantBlue-staining gel was done for the AC-elution fractions after concentration. The abbreviations denote the following: M – molecular weight marker in kDa; 1 – MBP-HomB$_{19\text{-}511}$ loaded with 2x RSB; and 2 – MBP-HomB$_{19\text{-}511}$ loaded with 2x NRSB. The band (lane 1), indicated by black arrowhead at the expected molecular weight ~97.6 kDa, was excised for mass spectrometry analysis.

Table 4.10: Tandem mass spectra data of MBP-HomB$_{19-511}$

Mass spectrometry confirmed that the excised band was MBP-HomB$_{19-511}$.

Taxonomy ID: 85963

Maltose-Binding Protein-Tagged Helicobacter Outer Membrane B

Mass: 97606 **Score:** 638 **Matches:** 9 (9) **Sequences:** 9 (9) **Calculated PI:** 5.29

Peptide match	Score	Expect	Rank	Unique
K.QGIINNFR.K	57	7.8e-06	1	U
K.TINLEDLR.T	54	2.2e-05	1	U
K.LYPFTWDAVR.Y	63	2.7e-06	1	U
K.SYEEELVKDPR.I	49	4.2e-05	1	U
K.GYGHNGNMTYQR.V	24	0.011	1	U
K.FEATSTSISDTNTQR.V	136	7.7e-14	1	U
K.NTYATYNYLPTDTILKR.A	105	6.3e-11	1	U
K.TFVTSSVTNHHFSNASQSFR.S	162	1.2e-16	1	U
K.NGFFIEAGFETGLLEGTQTQEKR.H	127	3.5e-13	1	U

Protein sequence coverage: 14%. Matched peptides shown in bold green (MBP) and bold red (HomB$_{19-511}$).

```
   1 MKIEEGKLVI WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ
  51 VAATGDGPDI IFWAHDRFGG YAQSGLLAEI TPDKAFQDKL YPFTWDAVRY
 101 NGKLIAYPIA VEALSLIYNK DLLPNPPKTW EEIPALDKEL KAKGKSALMF
 151 NLQEPYFTWP LIAADGGYAF KYENGKYDIK DVGVDNAGAK AGLTFLVDLI
 201 KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK VNYGVTVLPT
 251 FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG LEAVNKDKPL
 301 GAVALKSYEE ELVKDPRIAA TMENAQKGEI MPNIPQMSAF WYAVRTAVIN
 351 AASGRQTVDE ALKDAQTNSS SNNNNNNNNN NLGIEGRISK NGFFIEAGFE
 401 TGLLEGTQTQ EKRHTTTKNT YATYNYLPTD TILKRAANLF TNAEAISKLK
 451 FSSLSPVRVL YMYNGQLTIE NFLPYNLNNV KLSFTDAQGN TIDLGVIETI
 501 PKHSKIVLPG EAFDSLKEAF DKIDPYTFFF PKFEATSTSI SDTNTQRVFE
 551 TLNNIKTNLI MKYSNENPNN FNTCPYNNNG NTKNDCWQNF TPQTAEEFTN
 601 LMLNMIAVLD SQSWGDAILN APFEFTNSST DCDSDPSKCV NPGVNGRVDT
 651 KVDQQYILNK QGIINNFRKK IEIDAVVLKN SGVVGLANGY GNDGEYGTLG
 701 VEAYALDPKK LFGNDLKTIN LEDLRTILHE FSHTKGYGHN GNMTYQRVPV
 751 TKDGQVEKDS NGKPKDSDGL PYNVCSLYGG SNQPAFPSNY PNSIYHNCAD
 801 VPAGFLGVTA AVWQQLINQN ALPINYANLG SQTNYNLNAS LNTQDLANSM
 851 LSTIQKTFVT SSVTNHHFSN ASQSFRSPIL GV
```

4.3.5.3 Cleavage of HomB$_{19-511}$ from MBP using FXa protease

The FXa cleavage sequence (IEGR), located between the MBP and HomB, was exploited to cleave off HomB from the MBP fusion protein. FXa protease successfully cleaved off HomB from the MBP after overnight incubation at 23°C, and the protease was then removed from the mixture using Xarrest™ Agarose. ~97.6 kDa MBP-HomB$_{19-511}$ proteins completely disappeared when analysed by InstantBlue stain after SDS-PAGE separation, resulting in an appearance of two major bands at the expected molecular weights of ~55.14 kDa and ~42.5 kDa, representing HomB$_{19-511}$ and MBP fusion protein, respectively (Figure 4.27).

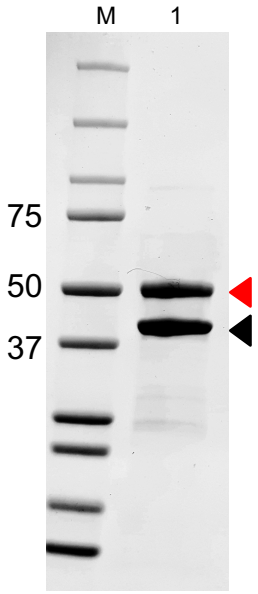


Figure 4.27: Cleavage of HomB$_{19-511}$ from MBP using FXa protease
InstantBlue-staining gel showing that HomB$_{19-511}$ was cleaved from MBP by FXa protease. Two bands appeared following the cleavage at the expected molecular weight ~55.14 kDa and ~42.5 kDa, representing HomB$_{19-511}$ and MBP fusion protein respectively. The abbreviations denote the following: M – molecular weight marker in kDa; 1 – HomB$_{19-511}$ (red arrowhead) and MBP (black arrowhead) after cleavage.

4.3.5.4 Purification of recombinant $HomB_{19-511}$ by SEC

Following FXa protease removal by the Xarrest™ Agarose, SEC was performed as a final step of protein purification. $HomB_{19-511}$ fractions were eluted between 53 and 58 mL, as illustrated by the peak of the chromatogram in Figure 4.28, while MBP fractions eluted thereafter (between around 58 and 65 mL). The peak structure suggested that no aggregation had taken place, as $HomB_{19-511}$ proteins were not eluted in the void volume suggesting that HomB is still soluble after its cleavage from MBP fusion proteins. All fractions were separated through SDS-PAGE then analysed by InstantBlue staining (Figure 4.29). Accordingly, fractions at the expected molecular weight ~55.14 kDa that contained less contaminants were collected together and concentrated using Amicon Ultra-15 Centrifugal Filter Units with a membrane ultrafiltration of 30 kDa. ~3 mg/mL were obtained from cleavage of 15 mL of 1 mg/mL MBP-$HomB_{19-511}$, which was then diluted to 1 mg/mL to be stored at -80°C for further investigations.

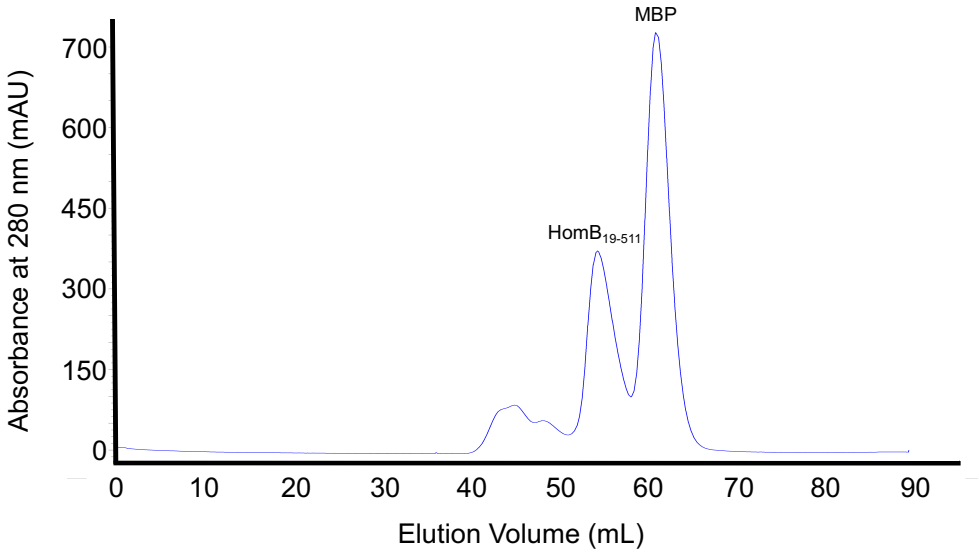


Figure 4.28: SEC purification of $HomB_{19-511}$ and MBP fusion protein
Overlay of SEC traces of $HomB_{19-511}$ and MBP proteins. $HomB_{19-511}$ proteins were eluted between 53 and 58 mL while MBP fusion proteins were eluted thereafter.

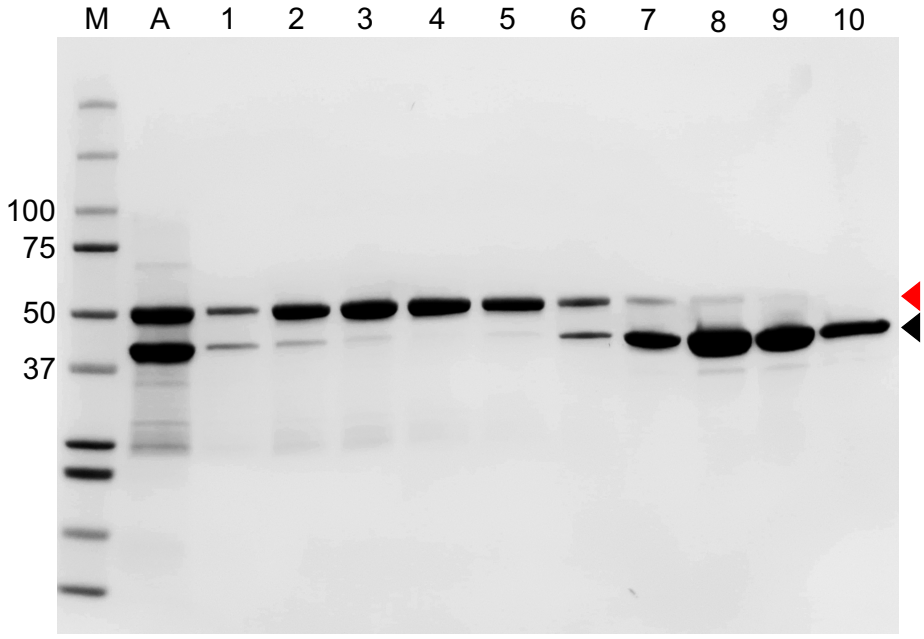


Figure 4.29: Purification of $HomB_{19-511}$ by SEC
InstantBlue staining of SEC-eluted fractions. The abbreviations denote the following: M – molecular weight marker in kDa; A – Following $HomB_{19-511}$ cleavage from MBP fusion proteins; 1-6 – correspond to the fractions from the $HomB_{19-511}$ (red arrowhead) peak; and 7-10 – correspond to the fractions from the MBP (black arrowhead) peak.

4.3.5.5 Mass spectrometry for $HomB_{19-511}$

Upon the analysis of SEC-elution fractions by InstantBlue staining, mass spectrometry was performed for the band at the expected molecular weight ~55.14 kDa (Figure 4.30). The band was excised and subjected to tryptic digestion for peptide mapping. The mass spectrometry analysis confirmed that the band was for $HomB_{19-511}$ based on the peptide match shown in Table 4.11.

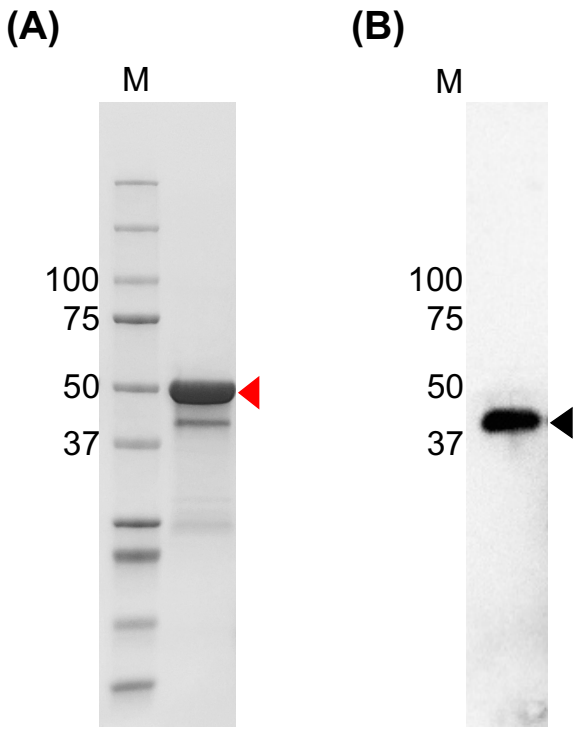


Figure 4.30: An InstantBlue-staining gel of $HomB_{19-511}$ for Mass spectrometry and Immunodetection of MBP following SEC
(A) An InstantBlue-staining gel was carried out for the SEC-elution fractions of $HomB_{19-511}$ after concentration. The band, indicated by the red arrowhead at the expected molecular weight ~55.14 kDa, was excised for mass spectrometry analysis. (B) Western blot using MBP immunodetection confirmed that the band indicated by black arrowhead, at the expected molecular weight ~42.5 kDa, was for MBP after cleavage of MBP-$HomB_{19-511}$ by FXa and then purification by SEC. M – molecular weight marker in kDa.

Table 4.11: Tandem mass spectra data of HomB$_{19-511}$

Mass spectrometry confirmed that the excised band was HomB$_{19-511}$.

Taxonomy ID: 85963
Helicobacter Outer Membrane B (HomB)
Mass: 55140 **Score:** 353 **Matches:** 6 (6) **Sequences:** 6 (6) **Calculated PI:** 5.57

Peptide match	Score	Expect	Rank	Unique
K.NTYATYNYLPTDTILKR.A	105	3.5e-11	1	U
K.EAFDKIDPYTFFFPK.F	69	1.6e-07	1	U
K.FEATSTSISDTNTQR.V	135	4.4e-14	1	U
K.QGIINNFR.K	34	0.00066	1	U
K.TINLEDLR.T	49	1.4e-05	1	U
K.TFVTSSVTNHHFSNASQSFR.S	29	0.0016	1	U

Protein sequence coverage: 16%. Matched peptides shown in bold red.

```
  1  ISKNGFFIEA  GFETGLLEGT  QTQEKRHTTT  KNTYATYNYL  PTDTILKRAA
 51  NLFTNAEAIS  KLKFSSLSPV  RVLYMYNGQL  TIENFLPYNL  NNVKLSFTDA
101  QGNTIDLGVI  ETIPKHSKIV  LPGEAFDSLK  EAFDKIDPYT  FFFPKFEATS
151  TSISDTNTQR  VFETLNNIKT  NLIMKYSNEN  PNNFNTCPYN  NNGNTKNDCW
201  QNFTPQTAEE  FTNLMLNMIA  VLDSQSWGDA  ILNAPFEFTN  SSTDCDSDPS
251  KCVNPGVNGR  VDTKVDQQYI  LNKQGIINNF  RKKIEIDAVV  LKNSGVVGLA
301  NGYGNDGEYG  TLGVEAYALD  PKKLFGNDLK  TINLEDLRTI  LHEFSHTKGY
351  GHNGNMTYQR  VPVTKDGQVE  KDSNGKPKDS  DGLPYNVCSL  YGGSNQPAFP
401  SNYPNSIYHN  CADVPAGFLG  VTAAVWQQLI  NQNALPINYA  NLGSQTNYNL
451  NASLNTQDLA  NSMLSTIQKT  FVTSSVTNHH  FSNASQSFRS  PILGV
```

Table 4.12 briefly summarises the results of each combination construct and expression host tested

Table 4.12: A brief summary showing the results of each combination of construct and cell type tested

The numbers used in the name of the construct name correspond to *homB* gene fragment derived from *H. pylori* strain J99 encoding amino acids 9-511 or 19 -511. The letter 'R' in the construct name stands for a deca-arginine (R10) designed to be inserted directly in the C-terminus of $HomB_{19-511}$. ΔpelB means deletion mutation of pelB leader sequence from the vector.

Construct Name	Expression Type	Expression Host	Induction Temp.	Result
pOPE101_*homB*$_{9-511}$	Periplasmic	*E. coli* XL10 Gold	24°C/16°C	No protein expression
pET22b_*homB*$_{9-511}$	Periplasmic	*E. coli* BL21 (DE3) *E. coli* Origami B (DE3)	24°C	No protein expression
pOPE101_*homB*$_{19-511}$	Periplasmic	*E. coli* XL10 Gold	24°C	Insoluble protein
pOPE101_*homB*$_{19-511R}$	Periplasmic	*E. coli* XL10 Gold	24°C	No protein expression
pET22b(ΔpelB) _*homB*$_{19-511}$	Cytoplasmic	*E. coli* Rosetta-gami B (DE3) pLysS	24°C	No protein expression
pCold-GST_*homB*$_{19-511}$	Cytoplasmic	*E. coli* BL21 (DE3) *E. coli* ArcticExpress (DE3)	10°C	Insoluble protein
pMAL-c5X_*homB*$_{19-511}$	Cytoplasmic	*E. coli* Rosetta-gami B (DE3) pLysS	12°C	Soluble and insoluble protein

4.4 Discussion

The second objective of the current study was to express recombinant HomB protein understood to be involved in mediating gastric adherence by the bacterium *H. pylori* but not the main adhesin and in inflammatory response by inducing the secretion of IL-8 from gastric epithelial cells (114). The design of the *homB* gene fragment used for screening of protein expression was considered in order to assure that the transmembrane domain of HomB at the C-terminal region, which was unnecessary for the purpose of this research, was excluded from the expression cassette as previously demonstrated in Chapter 3 about BabC. As a result of this, the putative C-terminal transmembrane domain of HomB *H. pylori* J99 (amino acids 512-668) was excluded from the amplification based on the information provided by the NCBI (280).

homB gene, a putative OMP in locus *jhp*0870, was recently discovered in a strain of *H. pylori* isolated from a young child with peptic ulcer (113). The gene is found in *H. pylori* strain J99, but does not exist in *H. pylori* strain 26695 (92,114,285), and thus the genomic DNA of the *H. pylori* J99 was used. Although the function of HomB is unknown, it is considered as a member of *H. pylori* OMPs since its gene belongs to the Hop family (89). Furthermore, HomB is known to induce secretion of IL-8 from gastric epithelial cells (114). As HomB contains the motif HEXXHXXGXXH, it might be considered as a metalloprotease, and it was therefore of interest to examine whether this protein works as a metalloprotease by proteolytic

hydrolysis of casein (Chapter 5). It was also of interest to examine the binding profile of HomB and to determine its ligand, as well as to study whether the protein induce the secretion of IL-8 from gastric epithelial cells *in-vitro*. To achieve the aforementioned aims, HomB was required to be expressed as a soluble protein.

PCR was performed for achieving the amplification of $homB_{9\text{-}511}$ gene. However, the PCR failed to amplify the gene as there were many mutations detected by DNA sequencing. This might be because *homB* gene sequence shares 90% sequence identity with *homA* gene especially at the 5' and 3' ends with the difference in the mid region (83,89,114). Therefore, overlap extension PCR was performed as the best option to obtain *homB* gene, exploiting the central region that is unique and different to *homA* gene.

Similar to the successful expressions of BabA (76), LabA (109) and BabC (Chapter 3), *E. coli* XL10 Gold cells were used as the first attempt to express recombinant $HomB_{9\text{-}511}$ in the periplasmic space using pOPE101 vector that contained pelB leader. The importance of periplasmic expression and the reason of choosing this type of expression were described in Section 3.4. As the hydrophobic domain of native HomB was excluded from the protein expression, the protein was expected to be observed in the periplasmic space of *E. coli* or at least as a soluble protein in the cytoplasmic space. Unfortunately, HomB was not expressed neither in periplasmic space nor in the cytoplasm as it was undetected by InstantBlue-staining and Western blot.

E. coli BL21 (DE3) cells were used as a second attempt to recombinantly express $HomB_{9-511}$ using pET-22b(+) vector that contained pelB leader to translocate the expressed protein to the periplasm of *E. coli* BL21 (181,252). pET-22b(+) vector contains T7 promoter which is designed for rapid overexpression in a DE3 lysogen strain because of the powerful T7 RNA polymerase action (197,261). The protein was not expressed again using two different concentrations of IPTG as the protein was undetected by Western blot. There were no obvious reasons why the protein could not be expressed at all. Therefore, expression of the protein using the vector pET-22b(+) were performed in another *E. coli* strain.

E. coli Origami B (DE3) cells were used as the next attempt of expression. This strain has an oxidising cytoplasmic environment because it lacks enzymes such as Δ*trxB* and Δ*gor* which help to form intramolecular disulphide bonds (286-290). However, the protein would be translocated to the periplasmic space of the bacteria as the pelB leader exist in the construct. Disappointingly, the protein was not expressed at all as the protein was undetected by Western blot. Although it was confirmed that $HomB_{9-511}$ was cloned in frame with the pelB leader, the protein was not even expressed as insoluble protein, in which case explanations and solutions could be considered.

Another attempt to express the protein again in *E. coli* XL10 Gold cells using pOPE101 vector by reducing the temperature from 24°C to 16°C. Although there was a thick band generated at the expect molecular weight in the

periplasmic 2 fraction, Western blot using c-Myc immunodetection revealed that this band was not for $HomB_{9-511}$ protein, as the result was negative. Mass spectrometry was also used as confirmatory analysis which confirmed that the excised thick band was for periplasmic oligopeptide-binding protein from *E. coli* strain K12 of which *E. coli* XL10 Gold is a derivative (291). Upon closer analysis of the construct using SignalP 3.0 (284) and 5.0 servers (283), it was discovered that the putative signal peptide of the protein had not been completely excluded during PCR amplification. The exclusion was for only 8 of 16 amino acids in the leader sequence. We believed that the remaining 8 amino acids might affect the protein expression. However, more investigations would be needed to confirm whether the signal peptide of the protein might affect the expression of HomB or not.

Following exclusion of the whole predicted signal peptide containing 16 amino acids plus two more amino acids, *E. coli* XL10 Gold cells were used as the first attempt to express recombinant $HomB_{19-511}$ in the periplasmic space using pOPE101 vector. The protein was successfully expressed but as insoluble protein. The IPTG concentration was then reduced from 1 mM to 0.1 mM in order facilitate periplasmic secretion in *E. coli* XL10 Gold cells (214), but this attempt was unsuccessful. It was unexpected to find the protein in the insoluble cytoplasmic fraction as the hydrophobic domain of $HomB_{19-511}$ was excluded from the construct design alongside the signal peptide that affected the protein expression of

$HomB_{9\text{-}511}$. This might have occurred when the synthesis of the protein was too rapid, making the Sec complex, which translocate the protein from cytoplasmic space to the periplasm of *E. coli*, to be jammed (292). This jamming might lead to the localisation of $HomB_{19\text{-}511}$ to the cytoplasmic space, which is a reducing environment, where it may not possible to form disulphide bonds. As a result, large volumes of protein would be expressed with improper folding and eventually aggregated into the cytoplasm to form inclusion bodies (293).

Therefore, the next strategy was to investigate whether the insertion of a polycationic amino acid tag would enhance the solubility of protein, resulting in translocation to the periplasmic space of *E. coli*. It has been shown that the addition of up to ten residues of lysine or arginine, at one of the protein termini, can significantly increase the protein solubility (76,242,258,259,294,295). The reason for the enhanced solubility is thought to reside in the repulsive electrostatic interactions, which occur between the polycationic tags with similar charges. This prevents the protein from aggregating and provides enough time for proper folding. It has also been shown that the insertion of the hexa-lysine tag improved the expression yield of BabA by preventing the proteolytic degradation of the C-terminus of the protein (76). Furthermore, the structure, function, and stability of the protein should not be altered or affected by inserting a polyionic tag (296), however, the stability of the protein was significantly improved when hexa-lysine tag was added to the C-terminus of BabA (76). No negative

influences have been observed when polyarginine tag (8-10 residues) was incorporated to antibody fragments or fused enzymes (297-299). For instance, the 3D structures of the maltodextrin-binding protein were identical with and without polyarginine tag (300). Consequently, ten arginine residues were incorporated into the C-terminal region of $HomB_{19\text{-}511}$ ($HomB_{19\text{-}511R}$) in order to investigate whether these incorporated arginine residues could solubilise the protein. However, the protein was not expressed at all, neither in the periplasmic space nor in the cytoplasm. There was no obvious reason why the protein expression was not successful. Although several periplasmic expression attempts were performed, all strategies for periplasmic expression of $HomB_{19\text{-}511}$ failed. As a result, the aim was changed to express $HomB_{19\text{-}511}$ in the cytoplasm of *E. coli* cells in order to save time instead of trying yet another periplasmic expression without successful results.

Cytoplasmic expression in *E. coli* Rosetta-gami B (DE3) pLysS cells was attempted in order to circumvent the possible jamming issue of Sec complex which may have occurred during the attempted periplasmic expression of $HomB_{19\text{-}511}$ in *E. coli* XL10 Gold. Similar to *E. coli* Origami B (DE3) cells, *E. coli* Rosetta-gami B (DE3) pLysS cells lack glutathione reductase (Δ*gor*) and thioredoxin reductase (Δ*trxB*) enzymes, making the cytoplasm to be an oxidising environment that promotes the formation of intramolecular disulphide bonds (286-290). Furthermore, the cells produce T7 lysozyme, which is a natural T7 RNA polymerase inhibitor, expressed by a

chloramphenicol-resistant plasmid containing the LysS gene, and thus the term pLysS is given to these cells (301,302). As a result, these cells are used to suppress basal T7 RNA polymerase expression before inducing them with IPTG, thereby stabilising pET recombinants that encode target proteins which influence cell viability and growth. The induced cells were grown at a low temperature, that was 24°C, in order to improve the soluble protein yield (303). Disappointingly, the protein was also not expressed using this strategy despite changing the expression type from periplasmic to cytoplasmic expression in order to avoid the potential jamming issue of Sec complex. Because all strategies performed to express $HomB_{19-511}$ had failed, with the exception of expressing $HomB_{19-511}$ in *E. coli* XL10 Gold that was insoluble, the next strategy was to express the protein with the presence of a fusion protein.

Using a fusion protein strategy is a common method of overcoming obstacles of expression. In this technique, one or more proteins is fused to a protein that is difficult to express in order to stabilise protein expression in the insoluble or soluble fraction, with the concept that the resulting expression will be driven by the fusion protein. For this reason, many fusion proteins such as GST (304), MBP (305), thioredoxin A (TrxA) (306), and others have been used (307). Although the fusion protein strategy provides potential benefits in protein expression and solubility, it has also disadvantages such as cleavage and removal of the protein of interest from the fusion protein using a protease and concerns about whether or not the

target protein keeps its structure and function. Finally, parallel investigation of a number of fusion proteins may be needed in order to produce the properties and yields of the target protein for the planned applications (307). The first attempt of a fusion protein strategy to express $HomB_{19-511}$ was done in *E.coli* BL21 (DE3) cells using pCold-GST vector, which is a cold-shock vector generated after incorporation of GST fusion protein into the pCold I vector (182). To improve the solubility and stability of a recombinant protein and to overcome its overexpression, *E. coli* cells are grown at low temperature (308,309). During the cold-shock technology, the temperature of the culture is reduced to a low incubation temperature, thereby halting the expression of most bacterial proteins and the growth of bacteria, giving the chance to induce the expression of cold-shock proteins (310). pCold I vector, which is one of a series of pCold expression vectors, is under the control of *cspA* promoter in which CspA is dramatically increased in its level after addition of 1 mM IPTG (183) and temperature downshift (182,183,311). Thus, the cold-shock system can significantly enhance protein expression, solubility, and purity (183). The vector also contains TEE, a five-codon sequence (MNHKV) that enhances translation initiation (312). The GST in the pCold-GST vector helps increasing the solubility of the protein at low temperature. It has been shown that nine out of ten proteins derived from plant, bacterial and mammalian sources, which could not be expressed in soluble fractions using conventional techniques, were successfully

expressed in soluble fractions using the novel cold-shock vector pCold-GST (182).

GST has been used to purify fusion proteins in one step (304), and it stabilises the recombinant protein in the soluble fraction and protects it against intracellular proteolysis (313). Specific protease cleavage sites are commonly available between the GST and the target proteins (307), such as HRV 3C protease recognition site in the pCold-GST in which HRV 3C protease recognises the site (182) that is (LEVLFQGP), and cleaves after glutamine (307,314,315). The induced *E. coli* BL21 cells were grown at a low temperature, that was 10°C, in order to improve the solubility of the protein. Unfortunately, large quantities of expressed protein would be incorrectly folded which aggregated in the cytoplasm to form inclusion bodies, despite the low overnight temperature and the low IPTG concentration used for inducing the cells. This result was similar to the outcomes of HomB which was expressed in *E. coli* BL21 cells using pGEX-4T-3 vector which contained GST tag (114). This may be because the cytoplasm of *E. coli* BL21 is a reducing environment which may affect the proper folding of the expressed protein (316). The other potential reason is that the GST tag is considered as a poor solubility enhancer (187,317). Furthermore, GST-fused proteins which are larger than 100 kDa or are enriched in either charged residues or hydrophobic regions can lead to the insolubility of the expressed protein (307), but this is not the case for the N-terminal extracellular domain of HomB that is approximately 55.14 kDa.

The next attempt was to re-express pCold-GST_HomB$_{19-511}$ in *E. coli* ArcticExpress (DE3) cells using the previous technique. This strain, which is derived from *E. coli* BL21-Gold cells, is engineered in order to improve the solubility of proteins that were difficult to be expressed in soluble fractions and increase the yield of soluble protein (201). *E. coli* express chaperonins that bind to and stabilise unfolded or partially folded proteins in order to facilitate correct protein folding. Unlike mesophilic host *E. coli* cells that have GroEL and GroES chaperonins which lose around 70% of their activity at low temperatures, ArcticExpress cells co-express Cpn10 and Cpn60, which are cold-adapted chaperonins, exhibit high activities of protein refolding at low temperatures during induction of protein expression, and thus they increase the yield of the soluble protein (318). The low temperature minimises the expression of endogenous bacterial proteins (200). Furthermore, this strain lacks Lon and OmpT proteases that can degrade recombinant proteins (201). Although the presence of GST tag expressed at a low temperature into *E. coli* ArcticExpress in which they would improve the solubility of the protein, large quantities of expressed protein would be incorrectly folded which aggregated in the cytoplasm to form inclusion bodies. This suggests that co-expression of the Cpn10 and Cpn60, in addition to the GST tag, are not enough to allow the proper formation of the disulphide bonds in HomB$_{19-511}$.

The next attempt of a fusion protein strategy to express HomB$_{19-511}$ was done in *E. coli* Rosetta-gami B (DE3) pLysS cells using the vector pMAL-c5X.

This vector is characterised by the presence of MBP fusion protein that can increase the folding and expression of eukaryotic proteins in *E. coli* cells (187,319-322). Furthermore, it has been shown that MBP tag improves the solubility of the partner protein (187-190). The vector uses the strong *tac* promoter, located upstream of the MBP tag, to express the fusion protein in large amounts (184), to be purified by an MBP-specific one-step affinity purification (191). However, not all MBP fusion proteins are highly soluble (323), and thus protein solubility can be increased by lowering the culture's temperature (308,309), after adding IPTG in the range of 0.01 to 0.02 mM with increasing in the induction time from 18 to 20 hours in order to gain a sensible yield of fusion protein (323). Therefore, *E. coli* Rosetta-gami B cells were induced with 15 μM IPTG for 20 hours at 12°C while lowering the flask shaking. Interestingly, MBP-$HomB_{19-511}$ was partially recovered in the soluble fraction, despite the bulk of the protein was expressed in the insoluble fraction. This suggests that IPTG concentration and the induction temperature and time, in addition to the MBP fusion protein, slightly improve the solubility of $HomB_{19-511}$. Furthermore, expression of the protein in *E. coli* cells such as Rosetta-gami B strain that has an oxidising cytoplasm could help in the solubility of $HomB_{19-511}$, because it lacks glutathione reductase and thioredoxin reductase enzymes, facilitating the formation of intramolecular disulphide bonds (286-290). It also suggests that the lysis buffer used in this study might affect the extraction of soluble proteins. Lysozyme dissolved in PBS lysis buffer was used to lyse

the bacterial cell wall with the help of freeze-thaw method and sonication in order to extract soluble proteins after centrifugation, while the other lysis buffer (50 mM Tris-Cl (pH 8.0), 500 mM NaCl and 5% glycerol) was not sufficient to lyse the bacterial cells in order to extract soluble proteins even after using sonication and centrifugation.

The successful approach identified during the screening of protein expression into *E. coli* Rosetta-gami B using the construct pMAL-c5X_HomB$_{19\text{-}511}$ was scaled up for protein purification. MBP fusion proteins were used to purify MBP-HomB$_{19\text{-}511}$ proteins from the soluble cytoplasmic extract of *E. coli* Rosetta-gami B (DE3) pLysS cells by AC. Proteins bound to dextrin ligand in the column primarily by hydrogen binding (322,324) and eluted by 10 mM maltose (325). Furthermore, mass spectrometry confirmed that the excised gel band at the expected size was MBP-HomB$_{19\text{-}511}$ proteins. Following removal of FXa protease, further protein purification was performed using SEC, showing that HomB$_{19\text{-}511}$ proteins were soluble in the buffer, which is required in order to characterise the protein. Mass spectrometry also confirmed that the excised gel band at the expected size was HomB$_{19\text{-}511}$ proteins.

After the confirmation by mass spectrometry that the excised gel band at the expected size was HomB$_{19\text{-}511}$ protein, the first objective of this study was successfully achieved after several attempts using different vectors designed for protein expression in a variety of *E. coli* strains. HomB$_{19\text{-}511}$ was

expressed at high yield using MBP tag, with good purity. This paved the way for the planned downstream investigations.

Chapter 5: Characterisation of HomB

5.1 Introduction

The third objective of this study was to characterise the recombinant HomB 26695 biological activity. This provides essential, first insights of HomB function.

5.1.1 Metalloproteases

Metalloproteases are proteases that perform their catalytic function in presence of divalent metal ions. Zinc (Zn^{2+}) is usually the catalytic metal ion occupying the active site of metalloprotease; however, Ni^{2+}, manganese (Mn^{2+}), or cobalt (Co^{2+}) can also be found in the active site (326,327). In one-step reaction including a mono- or dinuclear catalytic metal ion, a residue for a general base/acid, and a solvent molecule, metalloprotease cleaves its substrate without any covalent intermediate (328). The cleavage of peptide bonds by metalloprotease is important through participating in all main physiological activities, and thus their deregulation contributes to diseases, such as cardiovascular disorders, inflammation, cancer, etc (329-333).

Most metalloproteases are monometallic or mononuclear metalloproteases that require one metal ion for catalytic activity, while others, such as Mn^{2+} and Co^{2+} metalloproteases, have two metal ions that contribute in the catalysis. All mononuclear metalloproteases have a general acid/base residue, which is mostly a glutamic acid and required for catalysis, close to the catalytic metal. Mononuclear metalloproteases are divided into tribes,

clans and families based on their unique metal-binding sequence, which contains the residues responsible for general acid/base and coordination of the catalytic metal ion, and similarities of the overall folding (328).

Most mononuclear metalloproteases contain a short Zn-binding motif HEXXH, and thus they are known as zincins (326,334,335). Zincin is a tribe that has two histidine (H) residues for metal binding and a general base/acid glutamic acid (E; X means any amino acid). The glutamic acid activates a water molecule bound to Zn^{2+} at the active site (336), which is believed to carry out as nucleophilic attack in the carbonyl carbon of the substrate peptide bond (337). The zincin tribe is divided into the following clans: (i) aspzincins (338), (ii) S2P-zincins (339,340), (iii) FtsH-related enzymes (341,342), (iv) gluzincins (326), and (v) metzincins (334).

The characteristic of metzincin clan is that the Zn-binding motif is extended to be HEXXHXXGXXH/D+M, in which a histidine or aspartate is the third zinc-binding residue (X means any amino acid). Furthermore, a conserved glycine is found between the second histidine and the third metal-binding residue either histidine or aspartate. However, the glycine is replaced by an asparagine in pappalysins family. This clan is called metzincins because it contains a conserved methionine downstream of the Zn-binding motif and found below the catalytic ion at position 3 of 1,4-β-turn. Therefore, a hydrophobic pillow is formed by the so-called Met-turn under the catalytic ion, and this is structurally important as it stabilises and supports the active site (Figure 5.1) (334,343-345). Metzincin clan is divided into families based

on their structures including the serralysins (346), leishmanolysins (347), pappalysins (348), adamalysins/A Disintegrin and Metalloproteases (ADAMs)/reprolysins (349), astacins (350), snapalysins (351), matrixins or matrix metalloproteases (MMPs) (352), cholerilysins (353), archaemetzincins (354), fragilysins (355), toxilysins (356), igalysins (328), and thuringilysins (357). However, there are additional families reported from genome sequencing projects belonging to the metzincins that are as yet uncharacterised, temporarily called helicolysins, ascomycolysins, coelilysins, and gametolysins (344,345) (Figure 5.2).

The initial sequence analysis of HomB by Phyre^2 suggested that HomB could be a metallopeptidase since it contains the Zn-binding motif HEXXHXXGXXH. Furthermore, the conserved methionine is found downstream of this motif. Therefore, HomB can be *a priori* classified as a metzincin (Figure 5.3). It was reported that Hom family of *H. pylori* including HomB (JHP0870) could be hypothetically classified as new metzincins under the family helicolysins (344). However, no spatial structure has been identified for this family (358). Therefore, it would be of interest to examine the proteolytic activity of HomB in order to investigate whether this protein can act as a metalloprotease, and if so, whether it can be classified as a metzincin under the family helicolysins as mentioned in (344).

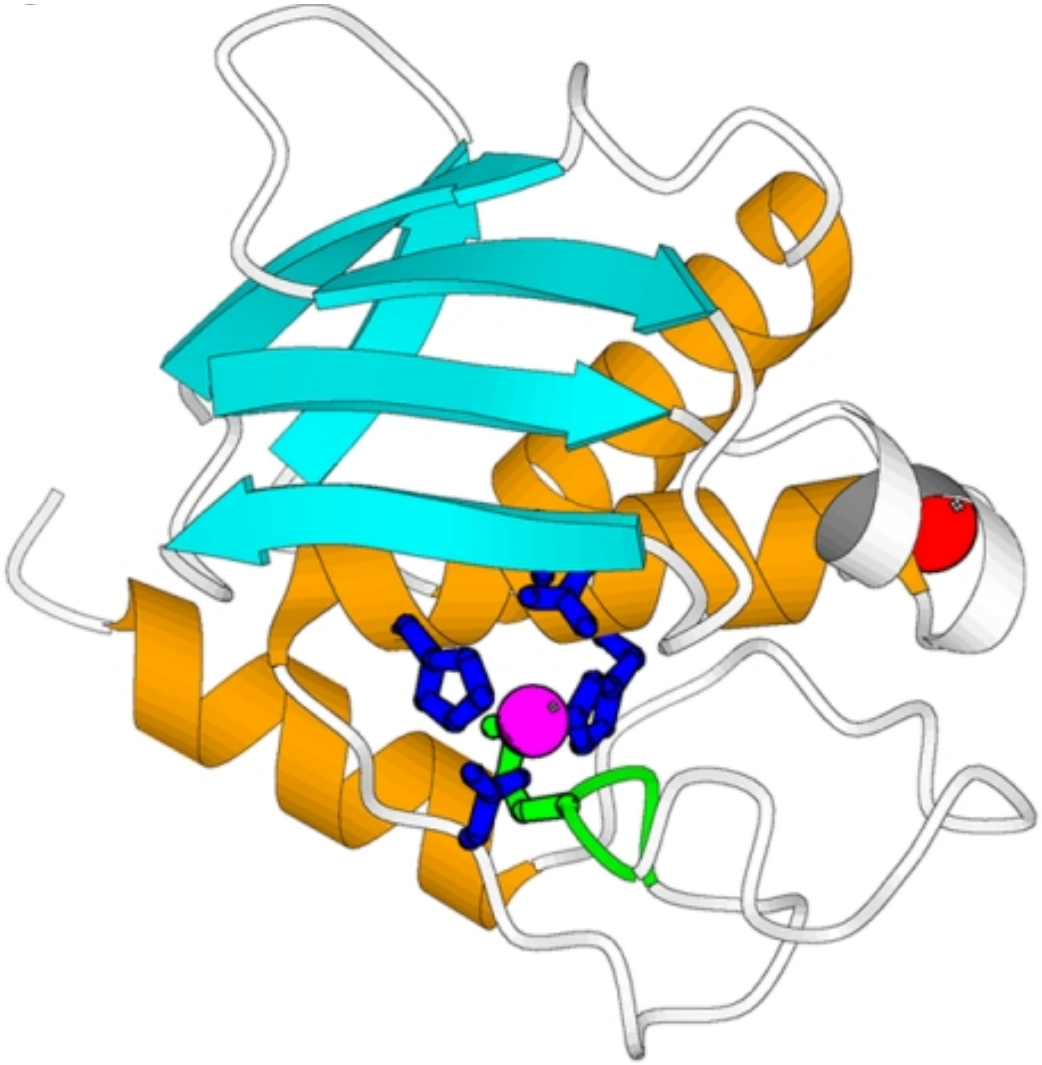


Figure 5.1: The location of the conserved methionine "Met-tern" underneath the catalytic Zn ion

A metzincin structure shows the location of the Met-turn, displayed in green, below the Zn ion and it stabilises and supports the active site. Adapted from (359).

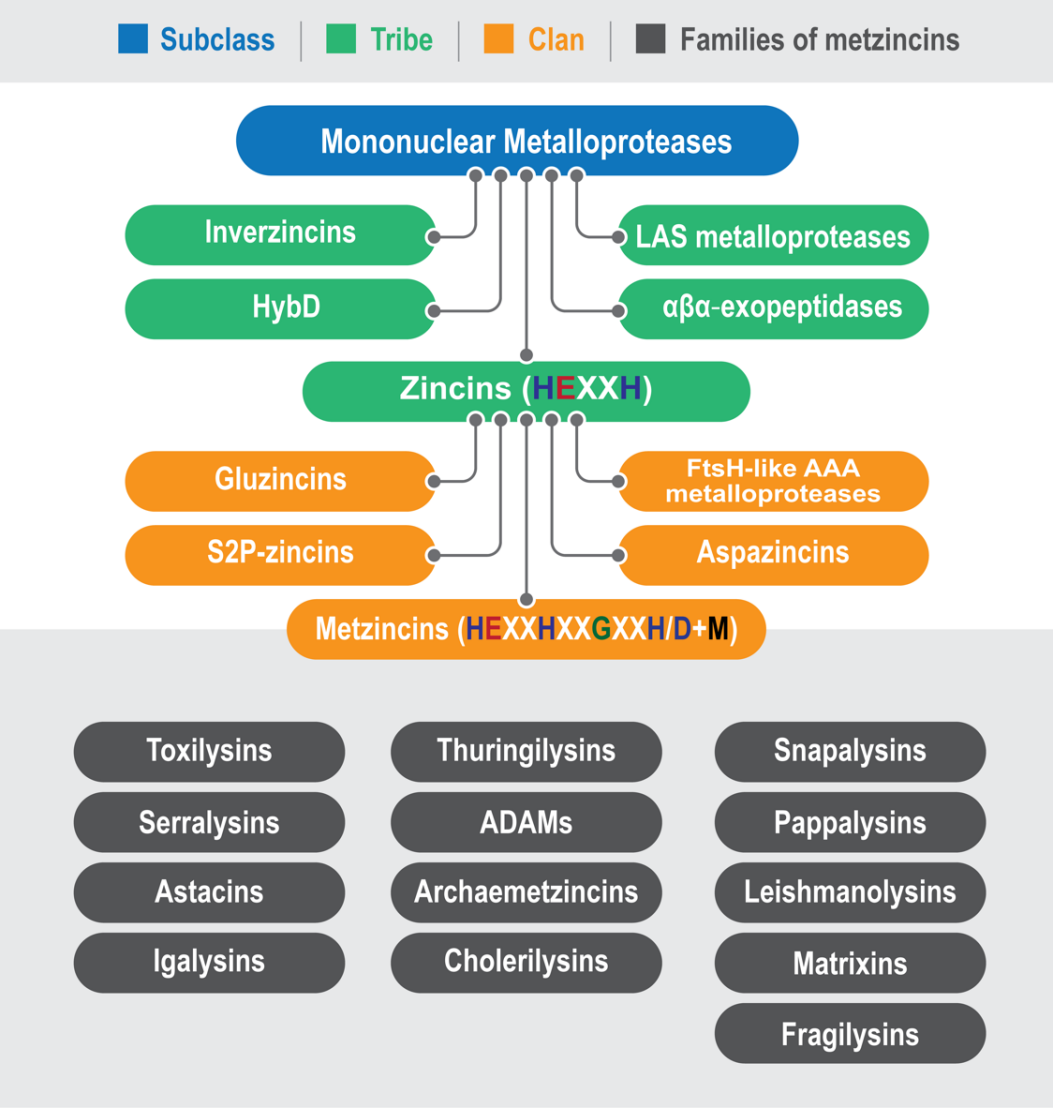


Figure 5.2: Classification of mononuclear metalloproteases
Zincin is a tribe in the subclass mononuclear metalloproteases which is distinguished from other tribes by containing a short metal-binding motif HEXXH. The two histidine residues (in blue colour) are for metal binding, a glutamic acid (in red colour) is for a general base/acid glutamic acid, and X means any amino acid. One of the most important clans in the tribe zincins is metzincin, which have an extension in the metal-binding motif to be HEXXHXXGXXH/D, in which the residues in blue colour are for metal binding, and the glutamic acid is for a general base/acid glutamic acid and shown in red colour. Furthermore, it contains a conserved methionine called “M-turn” (in black) found downstream of the Zn-binding motif and structurally located below the Zn ion as it stabilises and supports the active site. A conserved glycine (in green colour) is also found between the second histidine and the third metal-binding residue, which is either histidine or aspartate. Metzincins contain 13 families have been identified.

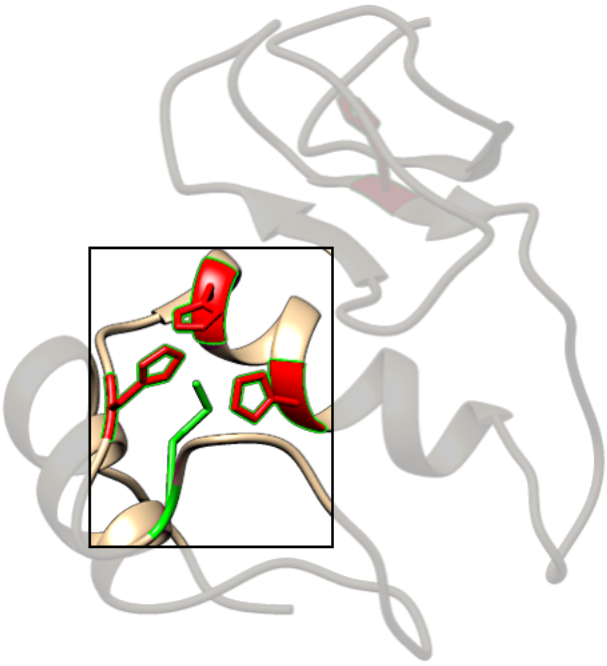


Figure 5.3: Discovery of Zn-binding motif in the predicted 3D structure of HomB

Phyre^2-predicted structure of the Zn-binding motif in HomB showing the three histidines (red colour) in the motif HEXXHXXGXXH which bind the zinc ion, and the methionine (green colour) which is structurally important as it stabilises and supports the active site. The presence of this motif in HomB suggests that HomB could be a metzincin.

5.1.2 Protease-activated receptor 2 and IL-8 production

Protease-activated receptors (PARs) are family of receptors consist of seven transmembrane helices, coupled with a G-protein (360). Four members of PARs family have been identified up to date, in which PAR2 is activated by trypsin, and trypsin-like serine proteases such as FXa, factor VIIa, mast cell tryptase, and neutrophil protease 3, while PAR1/3/4 are mainly activated by thrombin (361-363). It has been demonstrated that a variety of biological responses, such as inflammation, apoptosis, and proliferation of cells, can occur due to stimulation of intracellular signalling cascades initiated by serine proteases, such as thrombin and trypsin, via PAR activation (361). It was also shown that bacterial proteases, such as serralysin, which is a metzincin, activates PAR2 (364). The activation of PARs leads to various pathophysiological and physiological roles in many cells and tissues, such as gastrointestinal, central nervous, respiratory, and circulatory systems (360,365,366). Particularly, based on investigational findings, PAR2 activation contributes to inflammatory responses including many *in vivo* inflammation models in PAR2 deficient mice (367-373). Therefore, it has been shown that the activation of PAR by mast cell tryptase is a contributor in the development of colitis in both humans and animals, and tryptase inhibition by anti-PAR therapy is efficient for colitis treatment (374,375).

The activation of PAR2 is achieved by cleavage of the extracellular N-terminal domain of PAR2 by the serine protease after the amino acids 36.

This cleavage generates a new N-terminal sequence (SLIGKV) acting as a tethered ligand domain which then interacts with the extracellular loop II, which is a tethered ligand binding domain (QTIFIPALNITTCHDVLPEQLLVG), stimulating a signal transduction pathway through the activation of the associated G-protein. Furthermore, a synthetic agonist peptide (SLIGKV) similar to the tethered ligand domain can activate PAR2 (361,376). This results in induction of proinflammatory cytokines, such as IL-8 and IL-6, causing proinflammatory reactions in the epithelium (377,378), in which the IL-8 recruits and activates neutrophils to the infection site (379).

PAR2 is expressed in the gastrointestinal system including stomach (361,380). It was reported that *H. pylori* can induce IL-8 production through cleavage of PAR2 by its proteases, including a trypsin-like serine protease such as HtrA. This suggests that the pathogenesis of gastric diseases induced by *H. pylori* may be associated with the interaction between the bacterium and PAR2 (379).

It has been reported that HomB is able to stimulate an *in vitro* proinflammatory response by inducing IL-8 secretion, suggesting that HomB, in presence of CagA, is involved in the inflammatory response (83). Studies have also shown that the existence of HomB is associated with peptic ulcer disease (114). More recently, the existence of *homB* has been suggested to be a discriminating marker to differentiate stomach cancer from duodenal ulceration (115). Here, we hypothesise that *H. pylori* HomB can

induce proinflammatory response by cleavage of the extracellular N-terminal domain of PAR2 as shown in Figure 5.4. Therefore, it would be of interest to investigate the ability of the recombinant HomB in inducing IL-8 production, and if so, proving our hypothesis that the reason behind the IL-8 production is due to the cleavage of PAR2 by HomB.

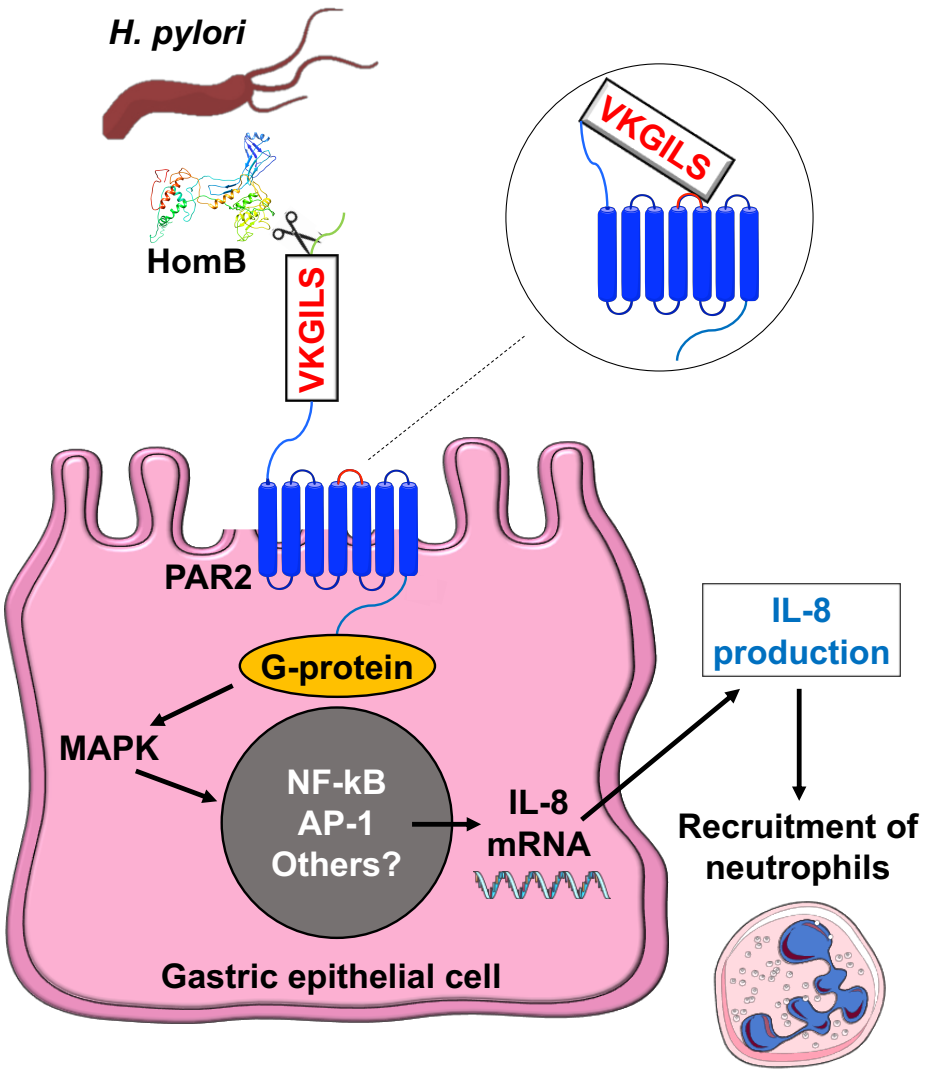


Figure 5.4: IL-8 production mediated via activation of PAR2 by HomB
Our hypothesis is that production of proinflammatory cytokines, such as IL-8, is associated with PAR2 activation on gastric epithelial cells. PAR2 is proposed to be activated by HomB, which cleaves the extracellular N-terminal domain of PAR2 (indicated in green colour), generating a new N-terminal sequence (SLIGKV) that acts as a tethered ligand domain. SLIGKV then interacts with the extracellular loop II (the red loop) that acts as a tethered ligand binding domain. This stimulates a signal transduction pathway through G-protein activation.
It is proposed that the activation of PAR2 on gastric epithelial cells by HomB can stimulate AP-1 and NF-kB-dependent IL-8 in association with activation of the extracellular signal-regulated kinases (ERKs) 1/2, which is the classical MAPK, and p38 MAPK. IL-8 recruits and activates neutrophils to the infection site, causing inflammation and gastric mucosal injury. This may suggest that gastric inflammation can be stimulated through the activation of PAR2 by *H. pylori* HomB, leading to gastric cancer in some cases.

It has been shown that HomB is expressed in the outer membrane of *H. pylori* and involved in bacterial adherence (83). Therefore, it is interesting to identify its binding receptor in order to expand our knowledge of how this protein works. Furthermore, HomB was characterised by assessing its stability against peptic digestion because of the acidity of the gastric fluid and the abundance of pepsin secreted in the stomach.

Seven steps were required for characterisation of HomB. First, investigation of the possibility of being HomB a metalloprotease through study its enzymatic activity using protease activity assay. Second, study the effect of divalent metal cations Zn^{2+} and Ca^{2+} on the proteolytic activity of HomB. Third, characterisation of proteolytic activity in the presence of various protease inhibitors. Fourth, generation of various mutations either single or double mutation including amino acids in the Zn-binding motif, asparagine next to the motif, and the conserved methionine in order to study the effect of these amino acids on the enzymatic activity of HomB. Fifth, identification of binding receptor for HomB by carbohydrate microarray. Sixth, determination of protein stability to peptic digestion in simulated gastric juice conditions. Lastly, assessment of the potential role of HomB in activation of PAR2 and induction of IL-8 production.

5.2 Materials and methods

5.2.1 Structural prediction of HomB by Phyre^2

The respective amino acid sequence was submitted to the structural prediction software Phyre^2 (232), in order to obtain a predicted 3D structure of HomB.

5.2.2 Quantitative protease activity assay

5.2.2.1 Effect of pH on HomB activity

Pierce Fluorescent Protease Assay Kit (Thermo Scientific, USA), quantitative protease activity test was carried out using a fluorescein isothiocyanate (FITC)-labelled casein, a substrate provided in the kit, following the manufacturer's protocol. Proteolytic digestion of the casein into smaller fluorescein-labelled fragments results in quantifiable alteration in fluorescence properties, representing an indication of proteolysis as shown in Figure 5.5. The reliance of the protease activity on pH of the reaction medium was assessed by measuring the hydrolysis of the substrate in different pH ranges of 50 mM Tris-HCl (pH 6, 7, 8 and 8.5). HomB was diluted in the buffer with different pH, to a final concentration of 11 µg/mL (200 nM) before mixing with the FITC-labelled casein in a 384-well white microplate with flat bottom (Greiner, USA). The plate was incubated at room temperature for 1 hour protected from light, and the fluorescence was then measured using a Spark® 10M multimode microplate reader at excitation/emission 485/538 nm. The assay was carried out three times

using 62.5 ng/mL trypsin as a general protease calibrator, and the results were expressed as fluorescence unit relative to the protease activity.

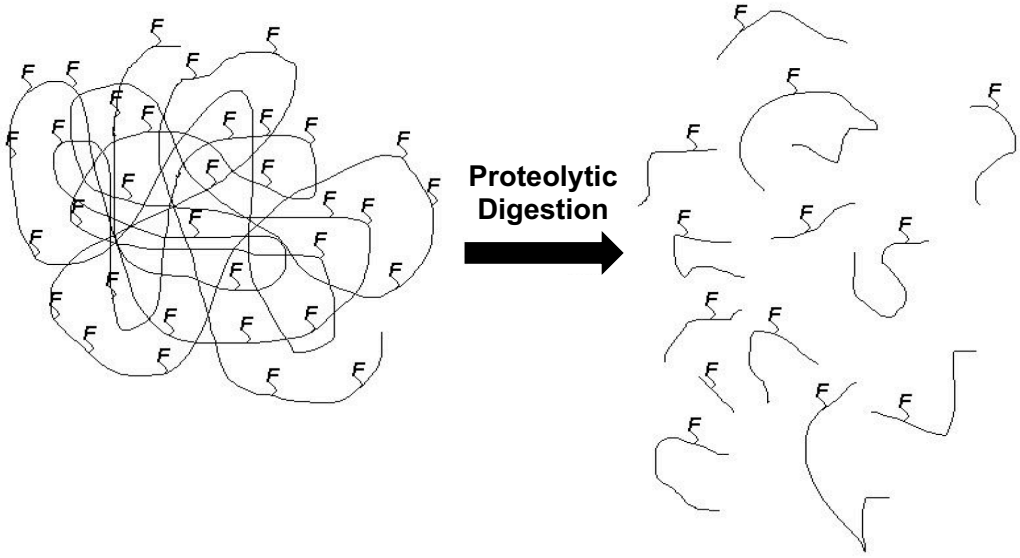


Figure 5.5: Principle of the proteolytic digestion of the FITC-labelled casein

Proteolytic digestion of the FITC-labelled casein into smaller fluorescein-labelled fragments results in quantifiable alteration in fluorescence properties (381).

5.2.2.2 Proteolytic activity of the recombinant MBP-HomB

Following the experiment, protease activity assay was performed for 19.52 µg/mL MBP-HomB (200 nM) using the optimal pH medium for HomB detected in the previous experiment. The assay was carried out three times using trypsin as a positive control.

5.2.2.3 Effects of divalent metal cations Zn^{2+} and Ca^{2+} on HomB activity

The effect of Zn^{2+} and Ca^{2+} on the activity of HomB was studied using zinc chloride ($ZnCl_2$) and calcium chloride ($CaCl_2$) at final concentrations varying from 0.01-5 mM and 1-20 mM in 50 mM Tris-HCl (pH = 8.5), respectively. HomB was incubated with different concentrations of $ZnCl_2$ and $CaCl_2$ at room temperature for 1 hour before mixing with the substrate. The experiment was carried out three times using trypsin as a positive control, and the hydrolysis of FITC-labelled casein was expressed as percentage compared to the activity of HomB without additives.

5.2.2.4 Effects of protease inhibitors and group-specific reagent on HomB activity

Inhibition studies were performed by adding different protease inhibitors and group-specific reagent, listed in Table 5.1, to HomB with incubation at room temperature for 1 hour prior to mixing with the substrate. The assay was repeated three times using trypsin as a positive control, and the results were expressed as percentage compared to the activity of HomB without additives.

Table 5.1: Protease inhibitors and group-specific reagent used in the inhibition studies

Abbreviations: EDTA, Ethylenediaminetetraacetic acid; AEBSF, 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride; DTT, dithiothreitol.

Protease Inhibitor	Type of Inhibition	Concentration
1,10-Phenanthroline	Zn^{2+} chelator	1 and 10 mM
EDTA	Metal chelator	10 mM
E-64	Cysteine protease inhibitor	10 μM
Leupeptin	Inhibitor of serine and cysteine proteases	100 μM
AEBSF	Serine protease inhibitor	1 mM
Pepstatin A	Aspartic protease inhibitor	10 μM
DTT	Reducing agent	10 mM

5.2.3 Mutations in the Zn-binding motif in HomB

5.2.3.1 Generation of expression constructs of HomB mutants

Five mutants of HomB (Table 5.2) were we generated, expressed, purified and characterised. The mutations were performed in the Zn-binding motif (**HE**XX**H**XXGXXH), the asparagine (**N369**), and the conserved methionine (**M372**) located downstream of the motif, in order to investigate the effect of the Zn-binding motif, asparagine, and methionine on the proteolytic activity. The mutations were a substitution of the amino acid of interest with alanine either as single or double mutation.

5.2.3.1.1 Site-directed mutagenesis

Q5® High-Fidelity DNA Polymerase (New England Biolabs, USA) was used to amplify pMAL-c5X_*homB*$_{19\text{-}511}$ with mutations using 5'-phosphorylated primers listed in Table 5.2. The PCR was performed under the following cycling conditions: 98°C for 30 sec of initial denaturation, followed by 35 cycles of template denaturation at 98°C for 10 sec, primer annealing at the temperature listed in Table 5.2 for 30 sec, and primer extension at 72°C for 215 sec. The reaction was finished with final extension at 72°C for 2 min. The PCR products were loaded on 1% agarose gel and run in 0.5x TBE buffer. SYBR Gold was used to stain the gel after the electrophoresis run, and the gel was then visualised under blue-light transilluminator. The bands were then extracted and purified using a QIAquick Gel Extraction Kit,

according to the manufacturer's protocol. The DNA concentration was estimated by measuring the absorbance at 260 nm using Nanodrop 2000.

Table 5.2: HomB mutants and 5' phosphorylated primers used for site-directed mutagenesis

The mutations were performed in the Zn-binding motif, the aspargine, and the conserved methionine (**HE**XX**H**XXGXXH**N**XX**M)** in which the amino acids in **bold** were substituted with alanine either as single or double mutation. FOR and REV denote forward and reverse primers, respectively. The underlined nucleotides are the substituted ones.

Mutant	DNA Template	Primer Sequence (5'- 3')	Annealing Temp.
H358A	pMAL-c5X_*homB*$_{19\text{-}511}$	[FOR] CACACTAAAGGCTATGGGCATAACG [REV] GCTGAATTCAGCCAAGATGGTTCT	72°C
E359A	pMAL-c5X_*homB*$_{19\text{-}511}$	[FOR] GCTATGGGCATAACGGGAATATGACCTATCAAAG [REV] CTTTAGTGTGGCTGAATGCATGCAAGATGGTTC	68°C
H358A and H362A	pMAL-c5X_*homB*$_{H358A}$	[FOR] CATAACGGGAATATGACCTATCAAAGAGTGCCG [REV] CCCATAGCCTTTAGTGGCGCTGAATTCAGCC	72°C
E359A and M372A	pMAL-c5X_*homB*$_{E359A}$	[FOR] GTAACGAAAGATGGTCAAGTGGAAAAGGATAGTAATGG [REV] CGGCACTCTTTGATAGGTCGCATTCCCGTTATG	71°C
E359A and N369A	pMAL-c5X_*homB*$_{E359A}$	[FOR] CTATCAAAGAGTGCCGGTAACGAAAGATGG [REV] GTCATATTCCCGGCATGCCCATAGCC	71°C

5.2.3.1.2 Cloning and transformation

The obtained linearised pMAL-c5X_*homB*$_{19\text{-}511}$ with mutations were ligated using T4 DNA ligase for 3 hours at room temperature. The transformations were performed in *E. coli* DH5α then re-transformed into *E. coli* Rosetta-gami B (DE3) pLysS after obtaining confirmatory sequencing using primers shown in Table 4.5.

5.2.3.2 Protein purification

5.2.3.2.1 Larger scale expression and cytoplasmic harvest

The protocol was carried out as described in Section 4.2.5.1.

5.2.3.3 Affinity chromatography

The protocol was carried out as described in Section 4.2.5.2.

5.2.3.4 Quantification of the recombinant MBP-HomB$_{19\text{-}511}$ mutants

The protocol was carried out as described in Section 4.2.5.3.

5.2.3.5 Cleavage of HomB$_{19\text{-}511}$ mutants from MBP using FXa protease

The protocol was carried out as described in Section 4.2.5.4.

5.2.3.6 Purification of recombinant HomB$_{19-511}$ mutants by SEC

The protocol was carried out as described in Section 4.2.5.5. The proteins were then resuspended in 20 mM Tris-Cl (pH 7.4), and 0.2 M NaCl, to a final concentration of 1 mg/mL, before storing at -80°C.

5.2.3.7 Mass spectrometry

The protocol was carried out as described in Section 4.2.5.6.

5.2.3.8 Testing of HomB mutants for quantitative protease activity assay

WT HomB and mutants were diluted in 50 mM Tris-HCl (pH 8), to a final concentration of 11 µg/mL (200 nM), prior to mixing with the substrate in a 384-well white microplate with flat bottom. The plate was incubated at room temperature for 1 hour protected from light, and the fluorescence was then measured at excitation/emission 485/538 nm. The assay was carried out three times using trypsin as a general protease calibrator, and the results were expressed as percentage compared to the activity of WT HomB.

5.2.4 Binding assay of HomB

5.2.4.1 Generation of expression constructs of HomB with a tag

Before carbohydrate microarray investigation, a tag such as c-Myc or 6x His tag needed to be inserted either at the C- or the N-terminus of HomB as shown in Figure 5.6. At the N-terminus of HomB, the tag located between the MBP fusion protein and HomB and after FXa cleavage site. The insertion of the tag will facilitate the detection of the protein by using a fluorescent antibody directed against the tag.

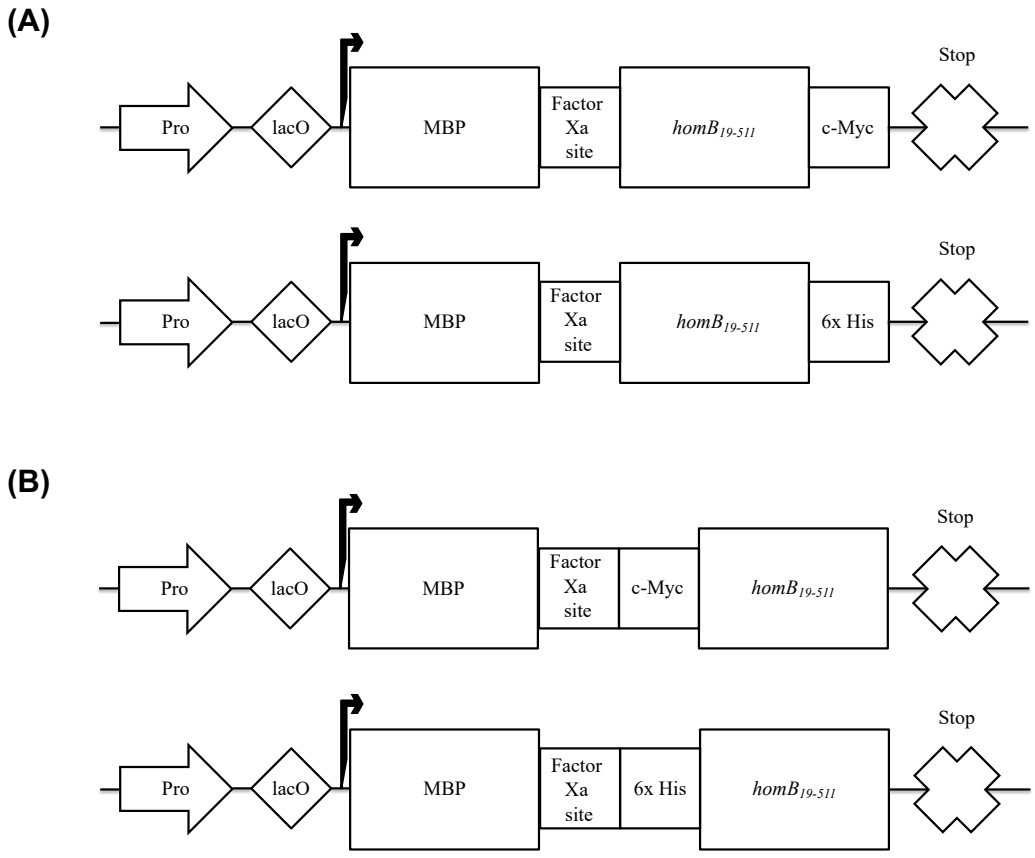


Figure 5.6: Schematic diagram of the pMAL-c5X_*homB*$_{19\text{-}511}$ expression cassette with tag insertion

The location of the inserted c-Myc or 6x His tag either at the C-terminus (A) or the N-terminus (B) of HomB.

5.2.4.2 Site-directed mutagenesis

Q5® High-Fidelity DNA Polymerase was used to amplify the construct pMAL-c5X_*homB*$_{19\text{-}511}$ with inserting c-Myc or 6x His tag at the C- or N-terminal region of HomB using 5' phosphorylated primers listed in Table 5.3. The PCR was performed under the following cycling conditions: 98°C for 30 sec of initial denaturation, followed by 35 cycles of template denaturation at 98°C for 10 sec, primer annealing at 72°C for 30 sec, and primer extension at 72°C for 215 sec. The reaction was finished with final extension at 72°C for 2 min. The PCR products were loaded on 1% agarose gel and run in 0.5x TBE buffer. SYBR Gold was used to stain the gel after the electrophoresis run, and the gel was then visualised under blue-light transilluminator. The bands were then extracted and purified using a QIAquick Gel Extraction Kit, according to the manufacturer's protocol. The DNA concentration was measured using Nanodrop 2000.

5.2.4.3 Cloning and transformation

The obtained pMAL-c5X_*homB*$_{19\text{-}511}$ with tags were ligated using T4 DNA ligase for 3 hours at room temperature. The transformations were performed in *E. coli* DH5α then re-transformed into *E. coli* Rosetta-gami B (DE3) pLysS after obtaining confirmatory sequencing using primers shown in Table 4.5.

Table 5.3: 5'-phosphorylated primers used in site-directed mutagenesis to insert c-Myc and 6x His tags at C- and N-terminus of HomB

FOR and REV denote forward and reverse primers respectively. The underlined nucleotides belong to the inserted tag.

Tag	Site of insertion	Primer Sequence (5'-3')
c-MYC	C-terminal	[FOR] TCAGAAGAAGATCTATAACCA TGGGCGGCCGCGATATC [REV] GATCAGCTTTTGTTCAACCCC TAAAATAGGGCTTCTAAAACT TTGC
6x His	C-terminal	[FOR] CATCATCACTAACCATGGGC GGCCGCGATATC [REV] GTGATGGTGAACCCCTAAAA TAGGGCTTCTAAAAC
c-MYC	N-terminal	[FOR] TCAGAAGAAGATCTAAAAAAC GGCTTTTTCATAGAAGCCG [REV] GATCAGCTTTTGTTCTGAAAT CCTTCCCTCGATCC
6x His	N-terminal	[FOR] CATCATCATAAAAACGGCTTT TTCATAGAAGCCGGC [REV] GTGATGATGTGAAATCCTTCC CTCGATCCC

5.2.4.4 Protein purification

5.2.4.4.1 Larger scale expression, cytoplasmic harvest, and purification by AC

The protocol of protein purification for the four pMAL-c5X_*homB*$_{19\text{-}511}$ constructs, containing c-Myc and 6x His tags at the C- and N-terminus of HomB, was carried out similar to the protein purification of pMAL-c5X_*homB*$_{19\text{-}511}$ construct described in Sections 4.2.5.1 and 4.2.5.2. Western blot using a biotinylated mouse anti-c-Myc or a mouse anti-His antibody, as described in Sections 3.2.2.2.3 and 4.2.4.4.3, was performed to confirm the insertion of the tags.

5.2.4.4.2 Quantification of the recombinant proteins

The protocol was carried out as described in Section 4.2.5.3.

5.2.4.5 Cleavage of HomB proteins containing the tags from MBP using FXa protease

The protocol was carried out as described in Section 4.2.5.4.

5.2.4.6 Purification of recombinant HomB$_{19\text{-}511}$ containing c-Myc tag at the N-terminus by SEC

The protocol was carried out as described in Section 4.2.5.5. Following SEC purification, SDS-PAGE separation and Western blot analysis using c-Myc immunodetection were carried out. The protein was then resuspended in

20 mM Tris-Cl (pH 7.4), and 0.2 M NaCl, to a final concentration of 1 mg/mL, before storing at -80°C.

5.2.4.7 Carbohydrate microarray

The *Helicobacter pylori* protein $HomB_{19-511}$, which contains N-terminal c-Myc tag, was analysed by the Glycosciences Laboratory (Imperial College London, UK) using their recently generated big screening arrays (in house designation 'Sets 42-56'), containing 885 lipid-linked oligosaccharide probes, and some arrayed as replicates (e.g. LNFP-III was included in each of array set of 64 probes for quality control purposes). c-Myc-tagged BabA was analysed alongside HomB protein as a control because it is known to bind to Le^b, and has been successfully used on glycan arrays from the CFG (86).

The analyses were performed at ambient temperature. In brief, microarray slides were blocked for 1 hour using the blocker solution HEPES-buffered saline (HBS; 10 mM HEPES, pH 7.4, 150 mM NaCl), containing 1% (w/v) BSA, 0.02% (v/v) casein blocker and 5 mM $CaCl_2$. The c-Myc-tagged proteins HomB and BabA were diluted in the blocker solution to 100 µg/mL and overlaid onto the arrays for 1.5 hour. The binding was detected with c-Myc Monoclonal Antibody (9E10) (Thermo Scientific, USA) and biotinylated anti-mouse IgG (Merck, UK); both overlaid at 10 µg/mL for 1 hour. The final detection was with 30 minutes overlay of streptavidin-Alexa Fluor 647 (Thermo Scientific, USA) at 1 µg/mL.

The microarray slides were scanned with GenePix 4300A scanner (PMT 350, Scan Power 100%) and the image analysis (quantification) was performed with GenePix® Pro 7 software. A dedicated in-house microarray software was used for data processing and presentation (http://www.beilstein-institut.de/en/publications/proceedings/glyco-2009).

5.2.5 Peptic digestion test

As mentioned early that pepsin is the main protein-digesting enzyme in the stomach among the different gastric enzymes; hence, HomB's resistance to pepsin has been tested.

5.2.5.1 Prediction of pepsin cleavage sites

The prediction of pepsin cleavage sites was similar to the prediction described in Section 3.2.5.1. The predicted pepsin cleavage sites were mapped on the predicted structure of HomB and the crystal structures of β-lactoglobulin (PDB Accession code: 3BLG) (236).

5.2.5.2 Pepsin activity

The ratio of pepsin activity used in this study was described in Section 3.2.5.2. Pepsin obtained from porcine gastric mucosa as a lyophilised powder with a specific activity ≥ 2,500 units/mg was used in this study using 10 U of pepsin activity/μg of HomB protein.

5.2.5.3 Peptic digestion

The protocol was carried out as described in Section 3.2.5.3. The following proteins were tested:

- HomB
- β-lactoglobulin from bovine milk

5.2.5.4 Sample analysis

The protocol was carried out as described in Section 3.2.5.4. Gels containing HomB and β-lactoglobulin samples were stained with InstantBlue stain for the visualisation of the outcome of peptic digestion.

5.2.6 Cell culture

A microbiological cabinet class II was used during the cell culture work. The main feature of preserving the cell culture was ensuring a sterile environment during experiments by regularly wiping the cabinet and equipment with wipes saturated with 70% isopropyl alcohol (IPA) prior and post experiments.

5.2.6.1 Cell defrosting

AGS cells (ATCC® CRL-1739™) with passage number 7, stored in vapor phase of liquid nitrogen, were placed on ice until they started defrosting. After that, 4 mL of warm RPMI-1640 medium (Merck, UK) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Gibco/ Thermo Scientific), 100 Units/mL penicillin and 0.1 mg/mL streptomycin (Merck, UK) and 2 mM L-glutamine (Merck, UK) were mixed with 1 mL of cells in a centrifuge tube, which was then centrifuged at 600 x *g* for 5 minutes. The supernatant, which contained 10% molecular biology dimethyl sulfoxide (DMSO) (Merck, UK) needed during cell preservation in the liquid nitrogen, was discarded and 5 mL of warmed RPMI-1640 medium were used to resuspend the cell pellet, which was transferred to 25 cm^2 cell culture flask with vent cap (Corning, USA). The flask was then placed in a humidified 5% CO_2 incubator at 37°C.

5.2.6.2 Routine passaging

Cell splitting (passaging) was performed when the cells become confluent by discarding the old medium, washing the cells in the flask with 5 mL of Ca^{2+}/Mg^{2+}-free Dulbecco's phosphate buffered saline (DPBS) (Merck, UK), and then trypsinisation of the cells by adding 1 mL of trypsin–EDTA solution (0.05% trypsin and 0.02% EDTA) (Merck, UK). The flask was then incubated for 10 minutes at 37°C until cells were detached. The flask was then tapped by fingers in order to detach the cells and 4 mL of warm medium were added to the cells to deactivate the trypsin. The whole solution in the flask was then transferred to a centrifuge tube and centrifuged at 600 x *g* for 5 minutes. The supernatant was discarded, and the cell pellet was resuspended in 5 mL of warm medium. 1 mL of the cells was transferred to 175 cm^2 cell culture flask with vent cap (Corning, UK), and 24 mL of warm medium were added in the flask, which was then placed in the incubator for further culture. The remaining resuspended pellet was used for further experiments or cryopreservation.

5.2.6.3 Cryopreservation

Following cell harvesting and counting, $5x10^6$ cells were resuspended in 900 µL medium. The cells were then transferred to cryovial and 100 µL of DMSO was added, to a final concentration of 10%. The cryovial was then placed in a Mr Frosty™ (Nalgene, USA) freezing container filled with IPA and stored

at -40°C for 1 hour and then overnight storage at -80°C. Then, the vial was stored in liquid nitrogen vapour phase until required.

5.2.6.4 RT-PCR for detection of IL-8 and PAR2 mRNAs expressed by AGS cells

The expression of IL-8 and PAR2 mRNAs was detected by RT-PCR, in which the harvested AGS cells were used for obtaining mRNA, which was reverse transcribed into cDNA using μMACS one-step cDNA kit (Miltenyi Biotec, Germany), according to the manufacturer's protocol.

Q5® High-Fidelity DNA Polymerase was used to amplify a fragment of IL-8 and PAR2 genes in the cDNA, in the case they were already transcribed, using primers listed in Table 5.4, which were designed to be intron spanning, in order to discriminate genomic DNA from cDNA. Specificity and amplicon size were verified using Primer Blast on the NCBI website. The PCRs were performed under the following conditions: 98°C for 30 sec of initial denaturation, followed by 35 cycles of template denaturation at 98°C for 10 sec, primer annealing at 65°C for 30 sec, and primer extension at 72°C for 7 sec and 28 sec for IL-8 and PAR2 fragments, respectively. Both reactions were finished with final extension at 72°C for 2 min. The PCR products were loaded on 1% agarose gel stained with EtBr and run in 0.5x TBE buffer. Following electrophoresis run, the gel was visualised under UV light. Amplification of glyceraldehyde 3-phosphate dehydrogenase

(GAPDH) gene was carried out alongside IL-8, as it is a well-known housekeeping gene (382), using primers listed in Table 5.4.

Table 5.4: Primers used to amplify IL-8, PAR2 and GAPDH genes
Primers below used to amplify IL-8, PAR2 and GAPDH genes in the cDNA after reverse transcription of mRNA obtained from harvested AGS cells. Primers were designed to be intron spanning, in order to discriminate genomic DNA from cDNA. Specificity and amplicon size were verified using Primer Blast on the NCBI website. FOR and REV denote forward and reverse primers respectively.

Gene	Primer Sequence (5'-3')	Product size (from cDNA/genomic DNA)
IL-8	[FOR] ACCGGAAGGAACCATCTCAC [REV] CCCAGTTTTCCTTGGGGTCC	284 bp/1374 bp
PAR2	[FOR] CAGCAGAGGCTCCGATTC [REV] CACATAGGCAGAGGCTGTGA	914 bp/none
GAPDH	[FOR] TGATGACATCAAGAAGGTGGTGAAG [REV] TCCTTGGAGGCCATGTGGGCCAT	240 bp/344 bp

5.2.6.5 ELISA assay for IL-8 production from AGS cells treated with WT HomB, E359A mutant, and E359A-M372A mutant

Production of IL-8 protein from the AGS cells was analysed using Human IL-8/CXCL8 DuoSet ELISA (R&D Systems, USA), according to the manufacturer's protocol. When the cells were grown confluently, AGS cells at a density of $25x10^4$/0.5 mL were seeded in 48-well plates with flat bottom (Corning, UK) and incubated in a humidified 5% CO_2 incubator at 37°C for 24 hours. The cells were then stimulated with 500 µM synthetic PAR2 agonist (SLIGKV amide; purity >95%) (Alpha Diagnostic International, USA), and 100 µg/mL (~1.8 µM) WT HomB, E359A mutant, and E359A-M372A mutant. Following incubation for 24 hours, the culture supernatant was obtained and analysed for IL-8 production. Cell viability was also monitored.

5.2.7 Statistical analysis

Statistical analysis of the experimental data was done using GraphPad Prism 7. Results are presented as means ± SEM. To describe and compare data parameters, 95% confidence limits for mean values were calculated using a one-way ANOVA with Tukey's correction for multiple comparisons.

5.3 Results

5.3.1 Discovery of Zn-binding motif

Phyre[2] was used to predict the 3D structure of the *H. pylori* HomB J99. Interestingly, the predicted 3D structure of HomB revealed that the protein contains a Zn-binding motif **HE**XX**H**XX**G**XX**H** followed by a conserved methionine "**M-turn**" as shown in Figure 5.7, which are the only part of HomB which had structural prediction with Phyre[2]. The Zn-binding motif contains three histidine residues His358, His362, and His368, glutamic acid residue Glu359 that is found next to the first histidine, glycine residue Gly365 that is located between the second and the third histidine residues, and Met372 that is located next to the motif. These residues are unique and characteristic for the metzincin clan, which can classify HomB as a metzincin. This novel discovery is very important in order to do more investigations on HomB related to its proteolytic activity, and if possible, to find an inhibitor against this protease. Figure 5.8 shows the Zn-binding motif and the conserved methionine in HomB sequence of some *H. pylori* strains, showing that the essential residues in the canonical sequence are fully conserved across several strains.

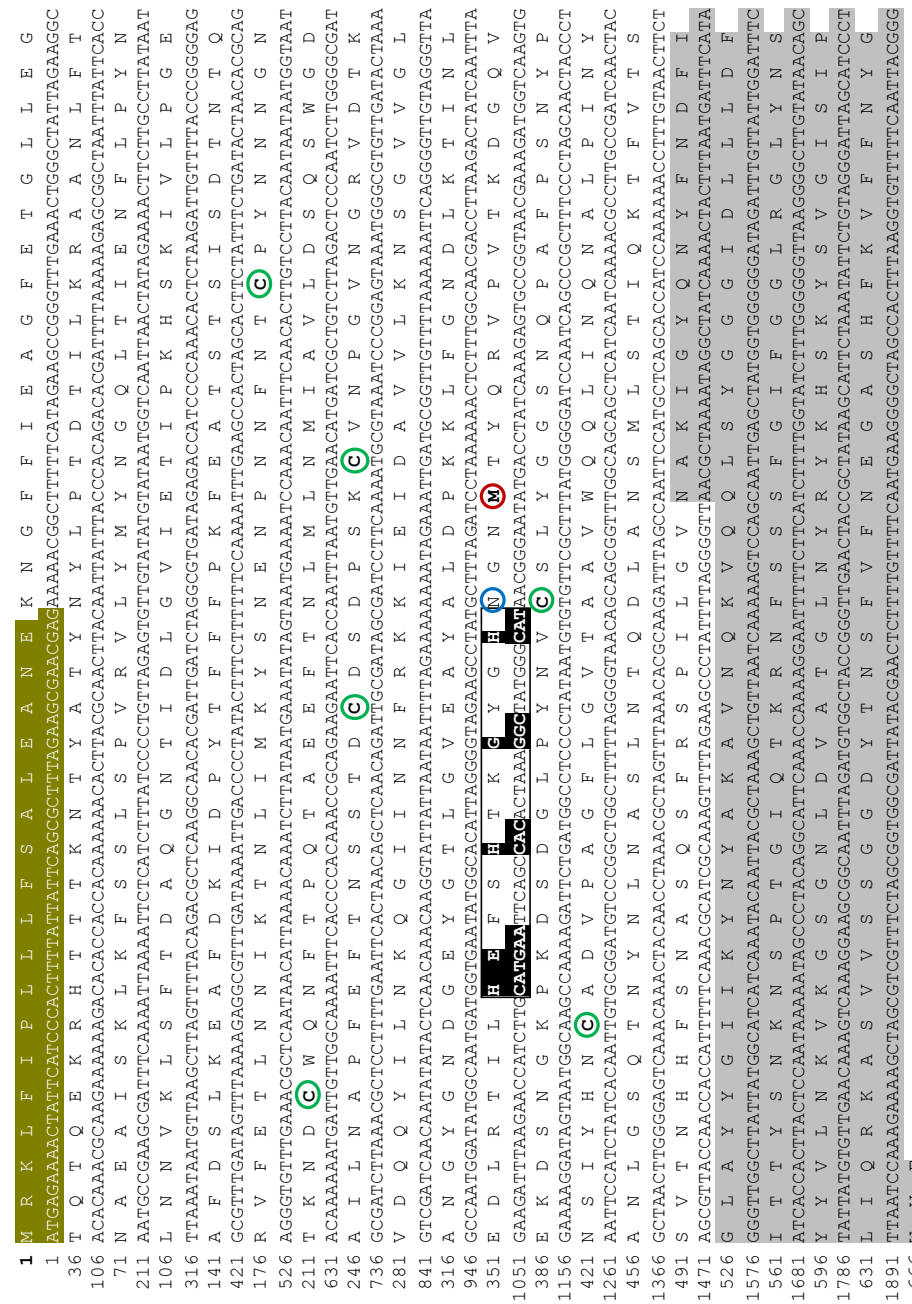


Figure 5.7: Amino acid and DNA sequences of HomB J99

The predicted structure of HomB shows that HomB contains a Zn-binding motif **HE**XX**H**XX**G**XX**H** highlighted in black colour. A conserved methionine "**M-turn**" is indicated by red circle, while the asparagine located after the Zn-binding motif is indicated by blue circle. HomB contain six cysteines indicated by green circles, which may suggest that HomB contains up to three disulphide bonds. The amino acids indicated by light yellow and grey are the signal peptide and the C-terminal domain, respectively, which were not included in our expressed recombinant protein.

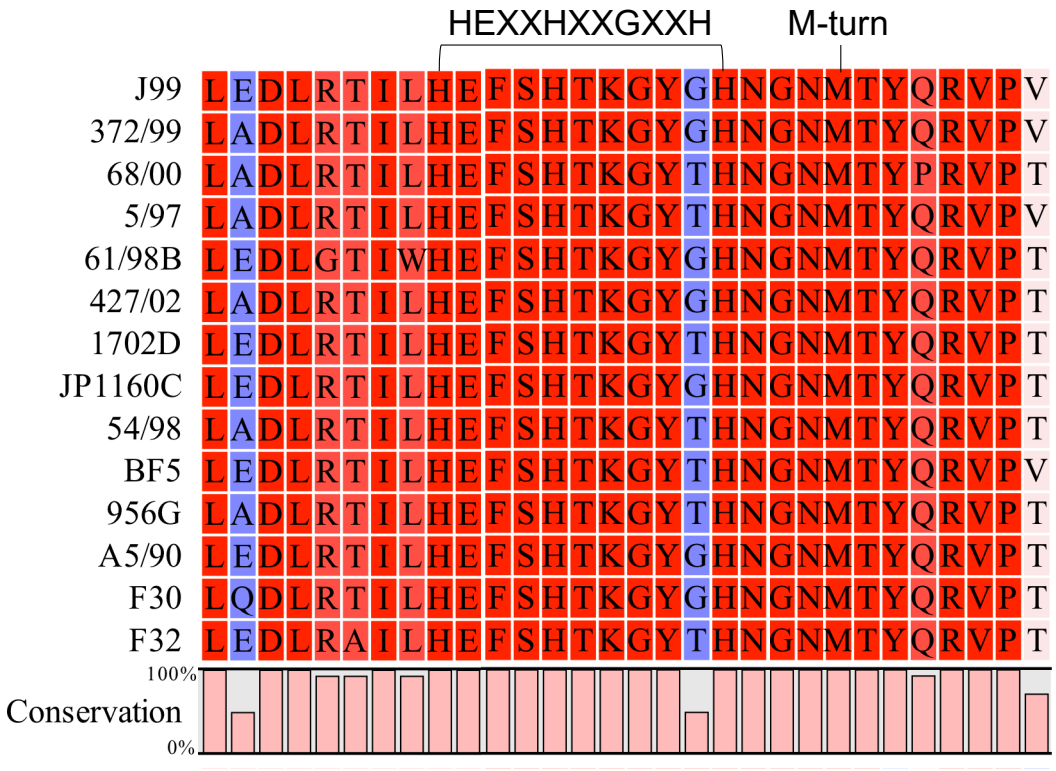


Figure 5.8: Zn-binding motif and the conserved methionine in HomB sequence of some *H. pylori* strains

5.3.2 Quantitative protease activity assay

5.3.2.1 Effect of pH on HomB activity

Following successful expression and purification of MBP_HomB proteins as well as complete cleavage of HomB from the MBP fusion partner by FXa, recombinant HomB was investigated to confirm whether this protein is a protease using protease activity assay. First, in order to determine the optimal pH range for HomB, different pH media of the standard Tris buffer ranging from 6.0 to 8.5 were used to evaluate the proteolytic activity of HomB. Figure 5.9 clearly shows that HomB is indeed a protease and the pH 8.5 gave the highest enzymatic activity of the recombinant HomB compared to the other pH values tested after applying blank correction. The blank correction was done by subtracting the average of blank readings, which do not contain HomB, from the corresponding average of sample readings, which contain HomB. Trypsin was used as a positive control at a concentration of 63 ng/µL.

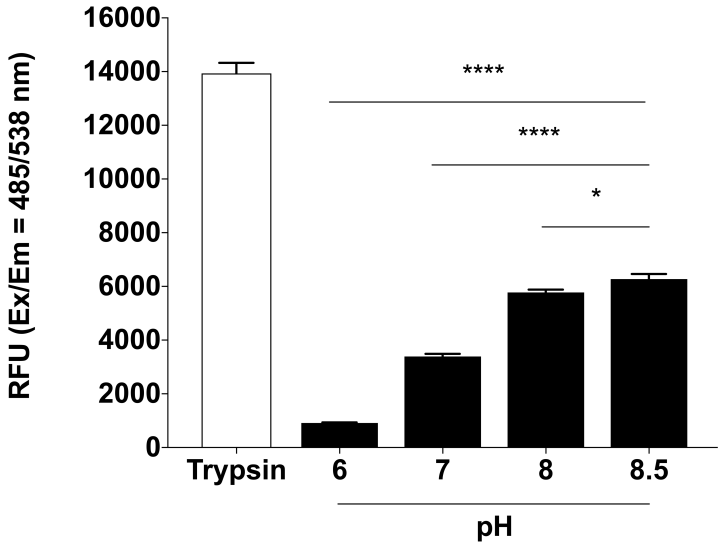


Figure 5.9: The pH effect on the activity of recombinant HomB
Each value represents the mean ± SE of 3 independent experiments. Data were determined by one-way ANOVA. * ($P < 0.05$); and **** ($p < 0.0001$) indicate significant difference. Trypsin was used as a positive control.

5.3.2.2 Proteolytic activity of the recombinant MBP-HomB fusion

Although FXa protease was removed after cleavage of HomB from the MBP fusion protein using the Xarrest Agarose described in Section 4.2.5.4, it was required to investigate whether FXa protease as a contamination could have caused the proteolytic activity in the previous experiment, as well as to check whether the MBP fusion protein with large molecular weight can affect the proteolytic activity of HomB by blocking its active site. To do so, protease activity assay was performed on the recombinant MBP-HomB protein at pH 8.5. This was carried out alongside trypsin which was used as a positive control. Blank correction was applied. Figure 5.10 shows that HomB was also active before the cleavage by FXa.

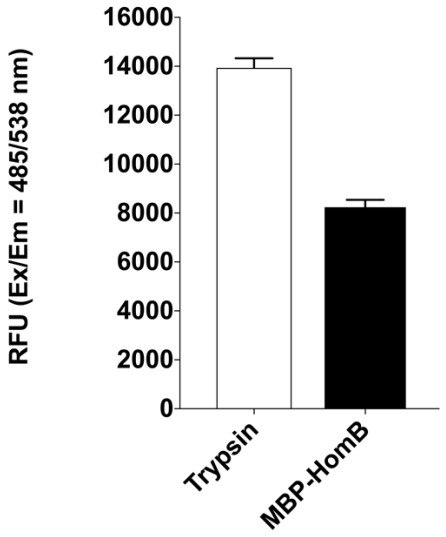


Figure 5.10: Proteolytic activity of the recombinant MBP-HomB
The proteolytic activity of 19.52 µg/mL MBP-HomB (200 nM) confirmed that (1) the FXa was not associated with the proteolytic activity after cleaving HomB from the MBP fusion protein, and (2) the large size of the MBP fusion partner did not block the active site of HomB. The value represents the mean ± SE of 3 independent experiments. Trypsin was used as a positive control.

5.3.2.3 Effects of divalent metal cations Zn^{2+} and Ca^{2+} on HomB activity

Further studies were performed to investigate the effects of divalent metal cations Zn^{2+} and Ca^{2+} on the activity of HomB. The investigation showed that Zn^{2+} at concentrations of 0.01 and 0.1 mM significantly decreased the proteolytic activity by approximately 35% and 51%, respectively compared to the one without any additives, while the activity was completely abolished in presence of 1 and 5 mM of Zn^{2+}. Furthermore, it was shown that Ca^{2+} at concentrations of 1 and 2 mM significantly increased the proteolytic activity by approximately 41% and 34%, respectively, compared to the enzymatic activity without any additives. The activity was insignificantly decreased by approximately 22% in presence of 5 mM of Ca^{2+}, while it was sharply

decreased by approximately 72% and 88% in presence of 10 and 20 mM of Ca^{2+}, respectively (Figure 5.11).

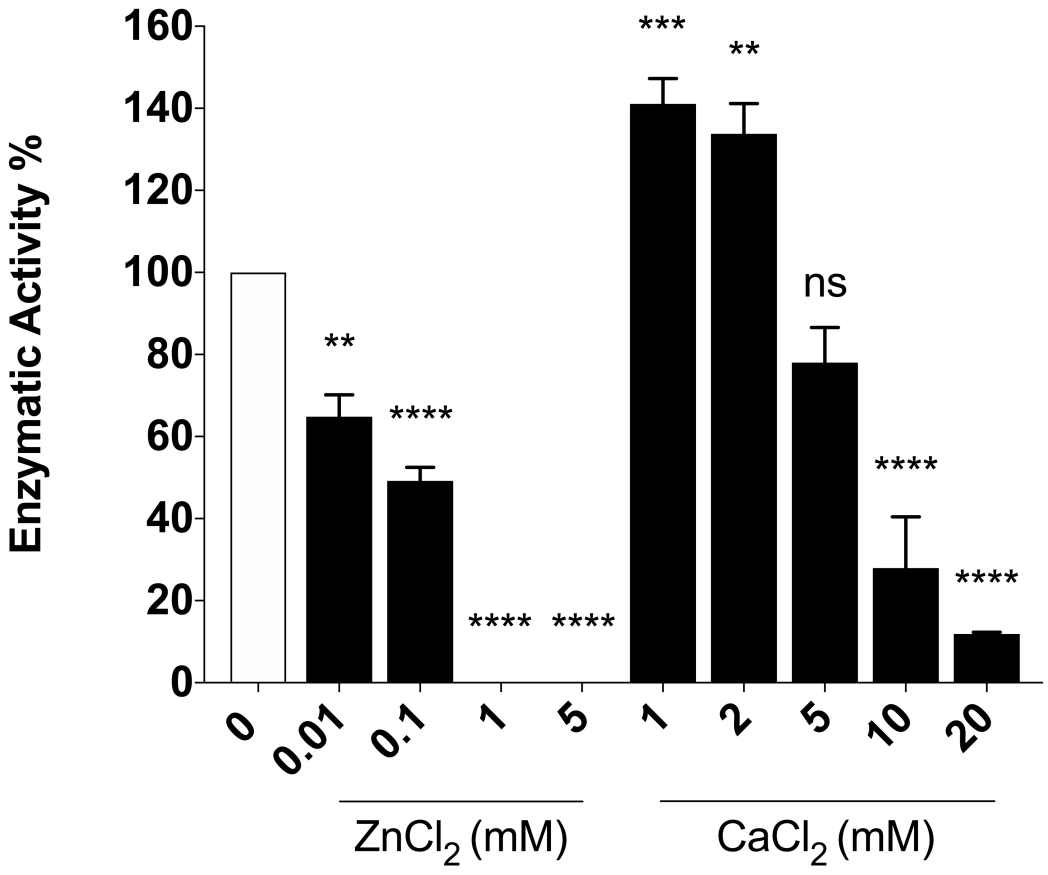


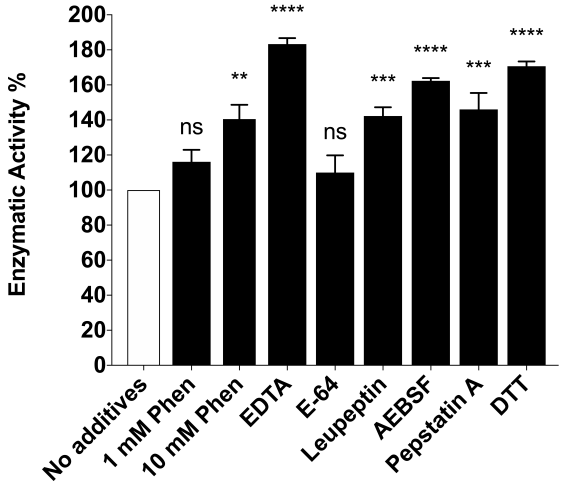
Figure 5.11: Effects of divalent metal cations Zn^{2+} and Ca^{2} on the activity of recombinant HomB

Each value represents the mean ± SE of 3 independent experiments. Data were determined by one-way ANOVA. ns indicates not significant; ** ($P < 0.01$); *** ($p < 0.001$); and **** ($p < 0.0001$ indicate significant difference from HomB without additives.

5.3.2.4 Effects of protease inhibitors and group-specific reagent on HomB activity

The effects of several protease inhibitors specific for the main classes of proteases on the proteolytic activity of HomB were tested by incubating each inhibitor with 11 µg/mL (200 nM) HomB for 1 hour before adding the HomB to the substrate. Interestingly, the enzymatic activity of HomB was not affected by 1,10-phenanthroline (Zn^{2+} chelator) at a concentration of 1 mM ($1x10^6$:200 nM, 5,000 molar ratio) despite that the activity was increased by approximately 16% compared to the one without any additives, while it was significantly increased by approximately 41% in presence of 10 mM of the inhibitor ($1x10^7$:200 nM, 50,000 molar ratio). In addition, EDTA (metal chelator) at a concentration of 10 mM sharply increased HomB activity by approximately 83%. 100 µM leupeptin (inhibitor of serine and cysteine proteases), 1mM 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) (serine proteases inhibitor), and 10 µM pepstatin A (Aspartic protease inhibitor) significantly increased the enzymatic activity of HomB by approximately 42%, 62%, and 46%, respectively. The activity was non-significantly increased by approximately 10% in presence of 10 µM of E-64 (cysteine protease inhibitor). It was also shown that 10 mM dithiothreitol (DTT), a reducing agent, markedly increased the activity by approximately 71% (Figure 5.12). Average protease inhibitors and DTT blank fluorescence was subtracted from sample (HomB with protease

inhibitor or DTT) readings to account for fluorescence changes caused by protease inhibitors and DTT.



Protease Inhibitor	Concn	Activity (%)[a]
Phen	1 mM	116%
Phen	10 mM	141%
EDTA	10 mM	183%
E-64	10 μM	110%
Leupeptin	100 μM	142%
AEBSF	1 mM	162%
Pepstatin A	10 μM	146%
DTT	10 mM	71%

Figure 5.12: Effects of protease inhibitors and group-specific reagent on the activity of HomB

Abbreviations: Phen, 1,10-Phenanthroline; EDTA, Ethylenediaminetetraacetic acid; AEBSF, 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride; DTT, dithiothreitol. Each value represents the mean ± SE of 3 independent experiments. Data were determined by one-way ANOVA. ns indicates not significant; ** ($P < 0.01$); *** ($p < 0.001$); and *** ($p < 0.0001$) indicate significant difference from HomB without additives.

[a] Percentage represents the mean of proteolytic activity in comparison with HomB activity without any additives (n=3).

5.3.3 Mutations in the Zn-binding motif in HomB

5.3.3.1 Successful cloning of HomB mutants

Side-directed mutagenesis using PCR technique was performed in order to generate the five HomB mutants (i) H358A, (ii) E359A, (iii) H358A-H362A, (iv) E359A-M372A, and (v) E359A-N369A using phosphorylated primers listed in Table 5.2. Following DNA successful amplification, ligation, transformation into *E. coli* DH5α, and plasmid purification, it was confirmed that the plasmids extracted from the positive transformants had 100% sequence identity to the recombinant plasmids with the desired mutations before re-transforming them into *E. coli* Rosetta-gami B (DE3) pLysS (Figure 5.13).

5.3.3.2 Expression and purification of MBP-HomB mutants

After DNA sequencing confirming the presence of the desired mutations in the plasmids, expression and purification of MBP-HomB mutants by AC were successfully performed following the protocols used for expression and purification of MBP-HomB$_{19\text{-}511}$. Fractions containing MBP-HomB mutants were eluted by 10 mM maltose (Figure 5.14), separated through SDS-PAGE and visualised by InstantBlue. Their purity was relatively low after AC (Figure 5.15). The concentrations of MBP-HomB mutants were measured by using the protocol in Section 4.2.5.3, and then diluted in the binding buffer to a final concentration of 1 mg/mL in order to be ready for the cleavage process.

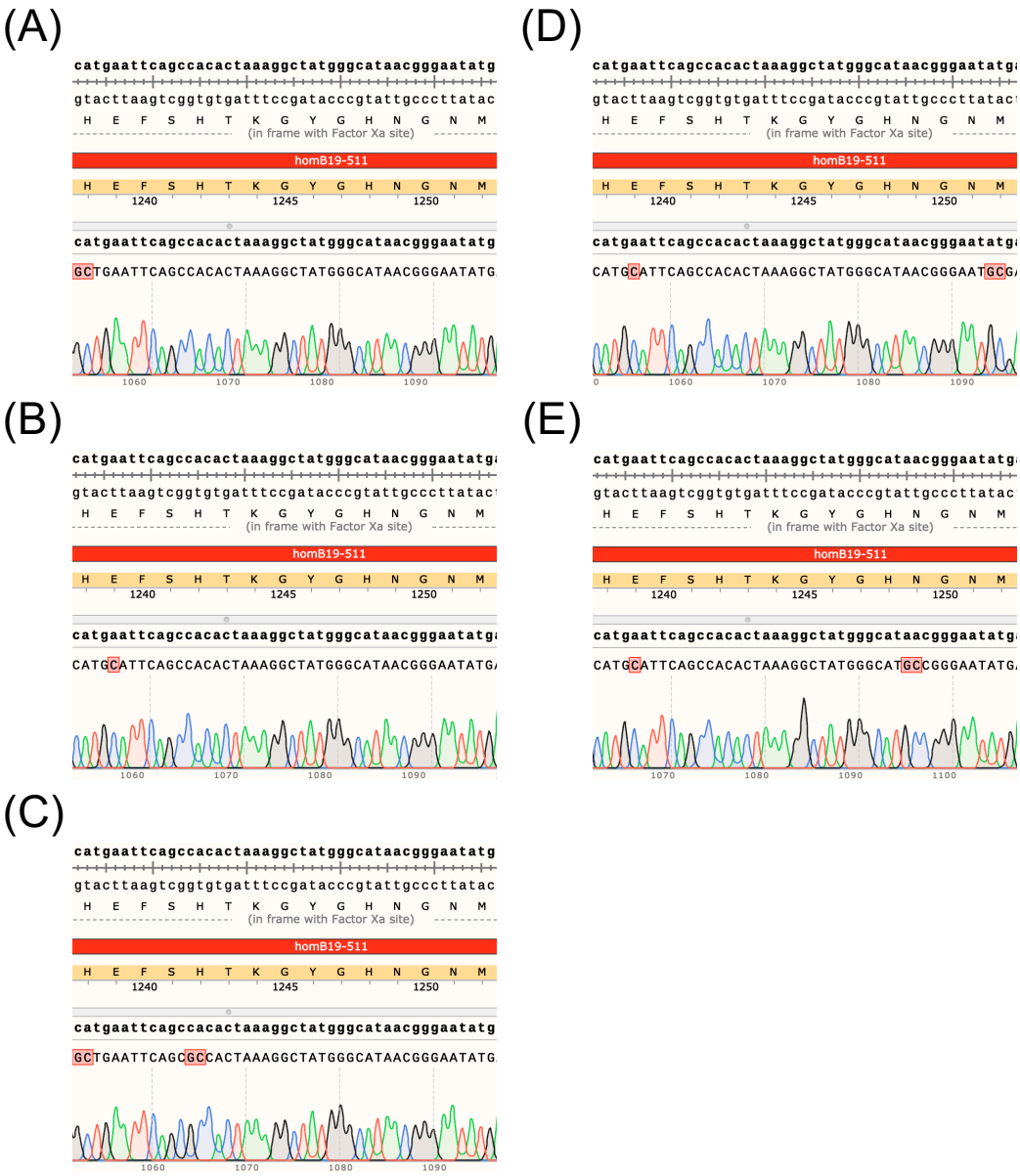


Figure 5.13: DNA sequencing confirming the desired mutation
SnapGene images confirming the desired mutation (highlighted in red colour) as described in Table 5.2. (A) H358A, (B) E359A, (C) H358A-H362A, (D) E359A-M372A, (E) E359A-N369A.

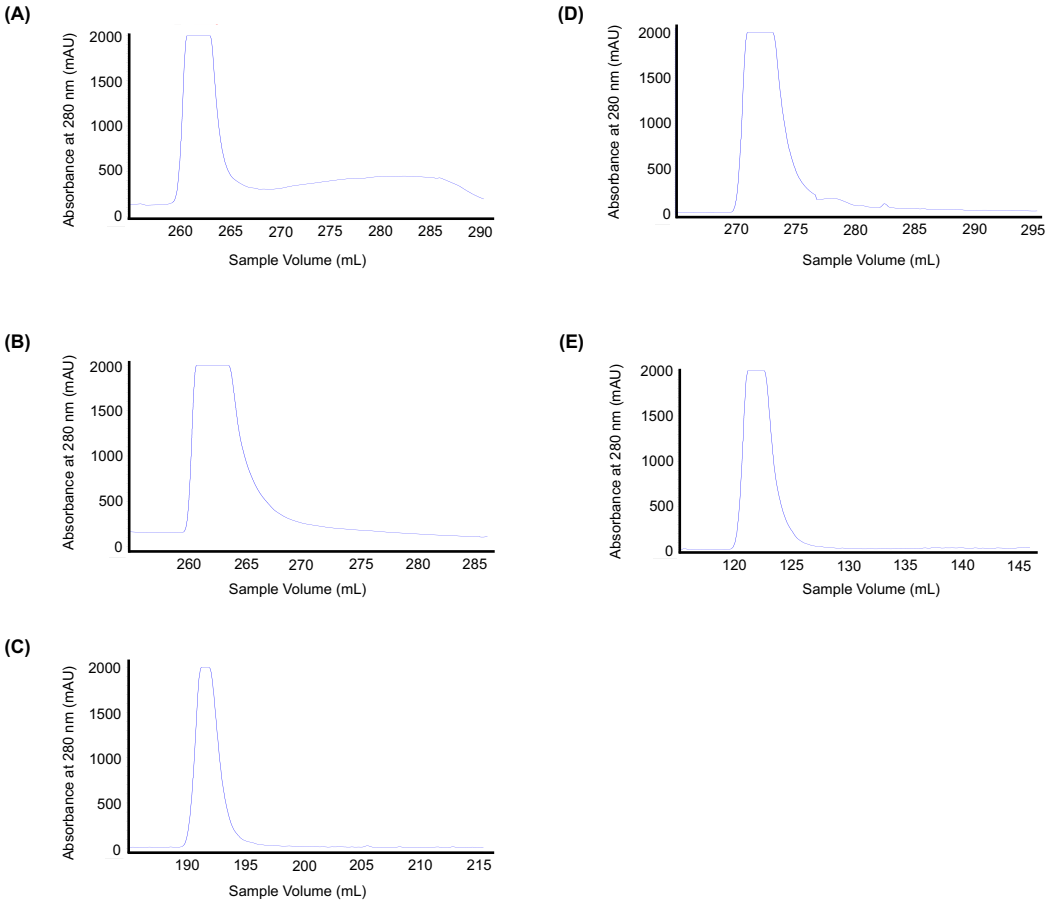


Figure 5.14: AC purification of MBP-HomB mutants
Overlay of AC trace of MBP-HomB mutants H358A (A), E359A (B), H358A-H362A (C), E359A-M372A (D), and E359A-N369A (E). The peak appeared after elution of the proteins by a gradient of 10 mM maltose.

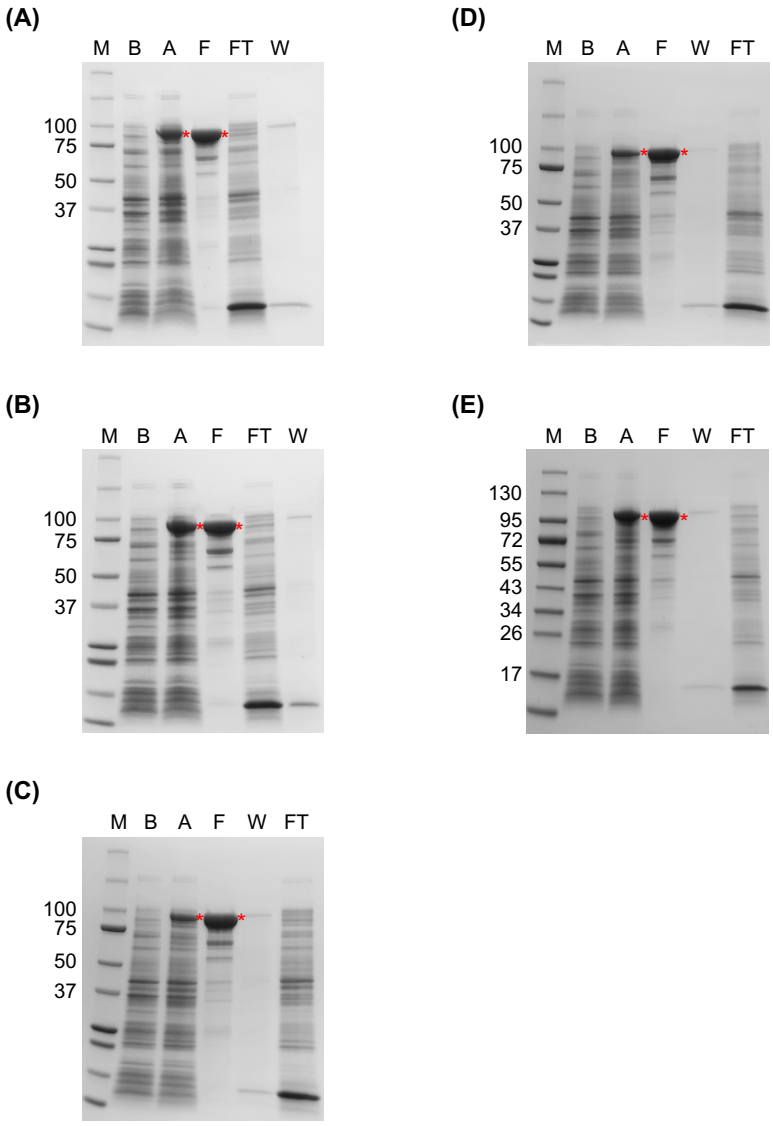


Figure 5.15: Purification of MBP-HomB mutants by AC

SDS-PAGE separation and InstantBlue-staining of a pool of AC-elution fractions, representing the peak from the AC purification of, were collected after protein purification of MBP-HomB mutants H358A (A), E359A (B), H358A-H362A (C), E359A-M372A (D), and E359A-N369A (E). The red star indicates MBP-HomB mutants (~97.6 kDa). The abbreviations denote the following: M – molecular weight marker in kDa; B – total protein fraction before inducing the cells with 15 μM IPTG; A – total protein fraction after inducing the cells with 15 μM IPTG; F – a pool of AC-elution fractions of MBP-HomB mutant; W – protein fraction during the washing out step; and FT – flow through. Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standards (Bio-Rad, USA) was used as a molecular weight marker in all experiments except (E) in which Fisher BioReagents™ EZ-Run™ Prestained Rec Protein Ladder, Fisher BioReagents (Thermo Scientific, USA) was used.

5.3.3.3 Purification of HomB mutants by SEC

After cleavage of HomB mutants from MBP using FXa protease, purifications by SEC were performed for further purification of HomB mutants. The protocol of SEC purification was similar to the one performed for $HomB_{19\text{-}511}$. Similarly to the WT $HomB_{19\text{-}511}$, Fractions of HomB mutants were eluted after 53 mL, as illustrated by the peak of the chromatogram in Figure 5.16, while MBP fractions eluted thereafter. The peak structure suggested that no aggregation had taken place, as HomB mutants were not eluted in the void volume suggesting that the proteins are still soluble after their cleavage from MBP fusion proteins. All fractions were separated through SDS-PAGE then analysed by InstantBlue staining (Figure 5.17).

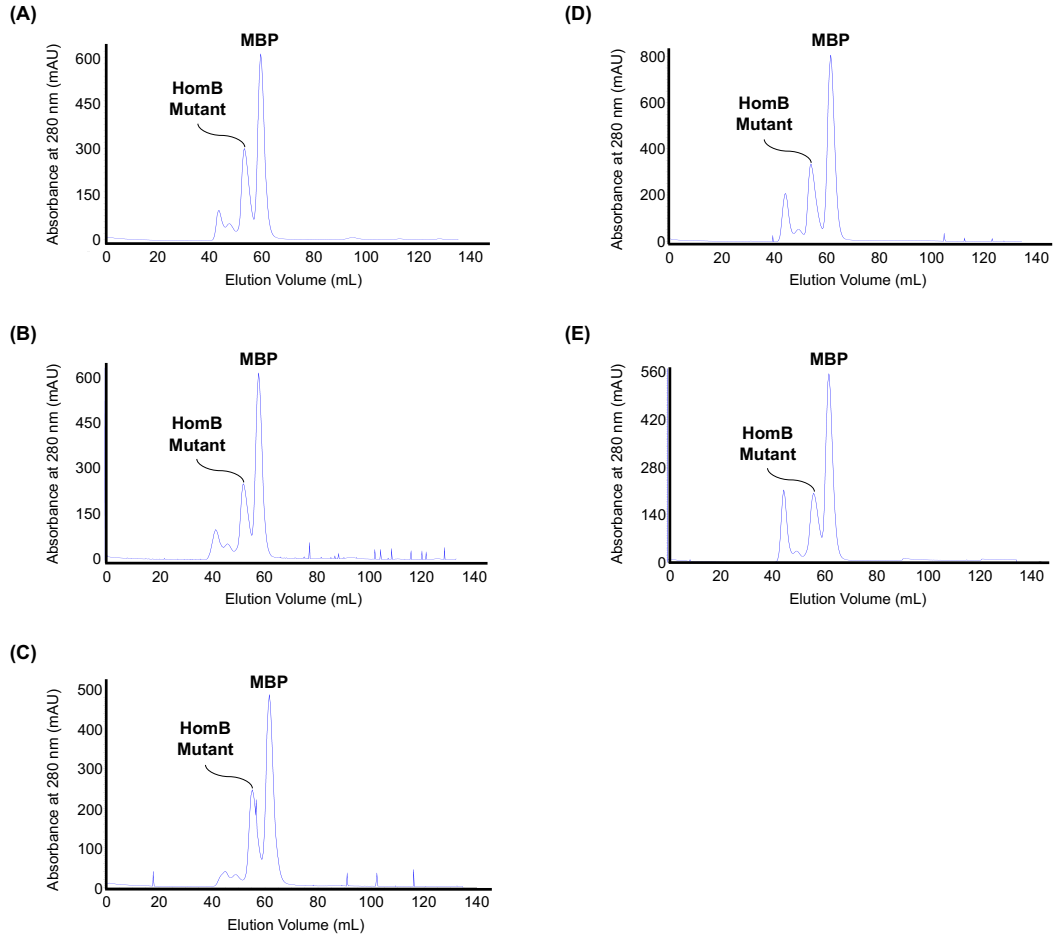


Figure 5.16: SEC purifications of HomB mutants and MBP fusion protein

Overlay of SEC traces of MBP proteins and HomB mutants which are H358A (A), E359A (B), H358A-H362A (C), E359A-M372A (D), and E359A-N369A (E). HomB mutants were eluted between 53 and 58 mL while MBP fusion proteins were eluted thereafter.

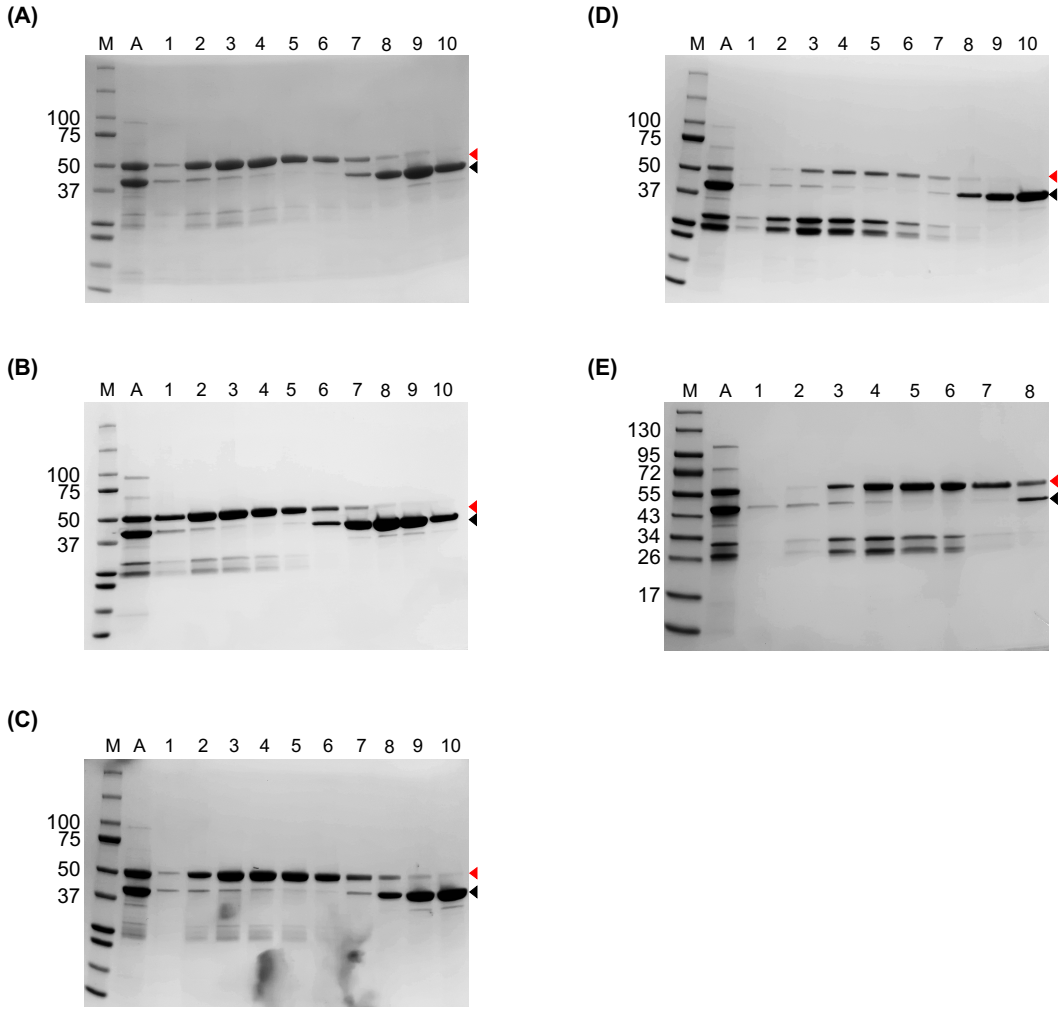


Figure 5.17: Purification of HomB mutants by SEC

InstantBlue staining of SEC-eluted fractions of HomB mutants which are H358A (A), E359A (B), H358A-H362A (C), E359A-M372A (D). The abbreviations denote the following: M – molecular weight marker in kDa; A – Following cleavage of HomB mutants from MBP fusion proteins; 1-10 – correspond to the fractions from the HomB mutants (red arrowhead) and the MBP (black arrowhead). Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standards was used as a molecular weight marker in all experiments except (E) in which Fisher BioReagents™ EZ-Run™ Prestained Rec Protein Ladder, Fisher BioReagents was used.

5.3.3.4 Mass spectrometry

Interestingly, Figure 5.17 shows that HomB becomes a slightly unstable protein by resulting in two bands at approximately 25 and 30 kDa in addition to the expected molecular weight of HomB, that is 54.14 kDa, in the E359A mutant. This instability was notably increased in E359-M372A or E359A-N369A double mutants, respectively. Thus, the conserved methionine and asparagine located downstream of the Zn-binding motif are significant in the stability of the protein in addition to the glutamic acid E359, but with lesser effect. The two bands highlighted by arrowheads (Figure 5.18) were identified by mass spectrometry, which confirmed that their amino acid sequences belong to HomB (Table 5.5). Although the instability of the protein was also seen with both HomB mutants H358A and H358A-H362A, the two bands generated as a result of protein instability were less clear (Figure 5.17).

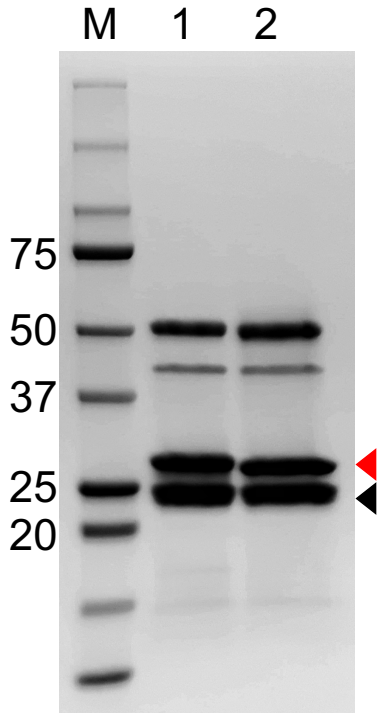


Figure 5.18: An InstantBlue-staining gel of HomB mutant (E359A-M372A) for Mass spectrometry following FXa protease cleavage

Two gel bands were generated after mutation of E359A-M372A. The abbreviations denote the following: M – molecular weight marker in kDa; 1 – HomB mutant (E359A-M372A) loaded with 2x RSB; and 2 – HomB mutant (E359A-M372A) loaded with 2x NRSB. The band in lane 1, indicated by red arrow at the expected molecular weight ~30 kDa, and the band in lane 1, indicated by black arrow at the expected molecular weight ~25 kDa, were excised for mass spectrometry analysis.

Table 5.5: Tandem mass spectra data of the two bands generated after cleavage of HomB mutant (E359A-M372A) from the MBP fusion partner

Mass spectrometry confirmed that the excised bands belong to HomB.

Taxonomy ID: 85963 **Helicobacter Outer Membrane B**

Sample 1 (the band at approximately 30 kDa)

Mass: 55140 **Score:** 353 **Matches:** 5 (5) **Sequences:** 5 (5) **Calculated PI:** 5.57

Peptide match	Score	Expect	Rank	Unique
K.NGFFIEAGFETGLLEGTQTQEK.R	197	6e-15	2	
K.NGFFIEAGFETGLLEGTQTQEKR.H	188	4.8e-14	2	
R.VLYMYNGQLTIENFLPYNLNNVK.L + Oxidation (M)	148	4.7e-10	1	
K.LSFTDAQGNTIDLGVIETIPK.H	188	5.6e-14	1	U
K.EAFDKIDPYTFFFPK.F	100	4e-05	1	

Taxonomy ID: 85963 **Helicobacter Outer Membrane B**

Sample 2 (the band at approximately 25 kDa)

Mass: 55140 **Score:** 353 **Matches:** 2 (2) **Sequences:** 2 (2) **Calculated PI:** 5.57

Peptide match	Score	Expect	Rank	Unique
K.NSGVVGLANGYGNDGEYGTLGVEAYALDP K.K	221	2.1e-17	1	
K.TFVTSSVTNHHFSNASQSFR.S	157	6.9e-11	1	U

Protein sequence coverage: 16% and 10% for samples 1 and 2, respectively. Matched peptides shown in bold green is for sample 1, while in bold red is for sample 2.

```
  1 ISKNGFFIEA GFETGLLEGT QTQEKRHTTT KNTYATYNYL PTDTILKRAA
 51 NLFTNAEAIS KLKFSSLSPV RVLYMYNGQL TIENFLPYNL NNVKLSFTDA
101 QGNTIDLGVI ETIPKHSKIV LPGEAFDSLK EAFDKIDPYT FFFPKFEATS
151 TSISDTNTQR VFETLNNIKT NLIMKYSNEN PNNFNTCPYN NNGNTKNDCW
201 QNFTPQTAEE FTNLMLNMIA VLDSQSWGDA ILNAPFEFTN SSTDCDSDPS
251 KCVNPGVNGR VDTKVDQQYI LNKQGIINNF RKKIEIDAVV LKNSGVVGLA
301 NGYGNDGEYG TLGVEAYALD PKKLFGNDLK TINLEDLRTI LHAFSHTKGY
351 GHNGNATYQR VPVTKDGQVE KDSNGKPKDS DGLPYNVCSL YGGSNQPAFP
401 SNYPNSIYHN CADVPAGFLG VTAAVWQQLI NQNALPINYA NLGSQTNYNL
451 NASLNTQDLA NSMLSTIQKT FVTSSVTNHH FSNASQSFRS PILGV
```

5.3.3.5 Testing of HomB mutants by quantitative protease activity assay

Further studies were performed by generation of HomB mutants in order to investigate the effect of the Zn-binding motif, the asparagine, and the conserved methionine, located downstream of the motif, on the proteolytic activity of HomB. WT HomB and the five HomB mutants were tested by incubating them with the substrate for 1 hour before fluorescence measurement. Unexpectedly, the investigation showed that the proteolytic activity was not affected after H358A mutation. Furthermore, the enzymatic activity was slightly increased by 3% with the E359A mutant compared to the WT HomB. Interestingly, the proteolytic activity was also unexpectedly increased by around 27% and 11% with the mutants E359A-M372A and H358A-H362A, respectively compared to the WT HomB. The activity was not affected with HomB mutant E359A-N369A (Figure 5.19). The implications of these results will be discussed below.

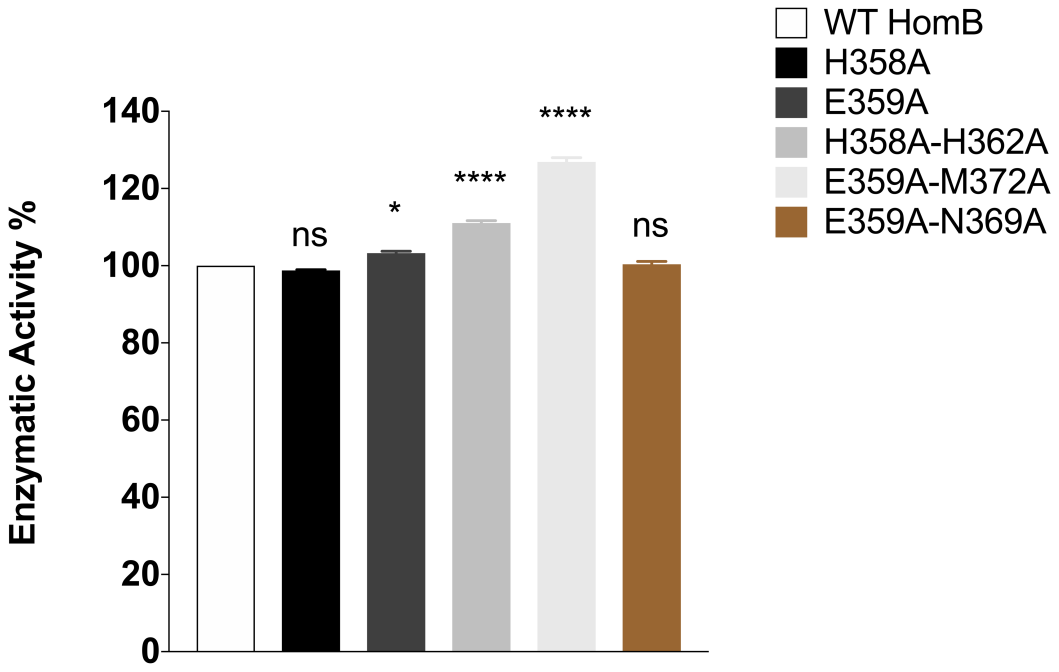


Figure 5.19: The effects of Zn-binding motif, asparagine and conserved methionine mutations on the proteolytic activity of HomB.

The proteolytic activity of HomB mutants were tested alongside WT HomB. It shows that HomB mutants (H358A) and (E359A-N369A) did not affect the proteolytic activity compared to the WT HomB, while the activity was significantly increased with the mutants E359A (~103%), E359A-M372A (~127%) and H358A-H362A (~111%) in comparison with WT HomB. Each value represents the mean ± SE of 3 independent experiments. Data were determined by one-way ANOVA. ns indicates not significant; * ($P < 0.05$); and **** ($p < 0.0001$) indicate significant difference from WT HomB.

5.3.4 Binding assay of HomB

5.3.4.1 Successful insertion of c-Myc or 6x His either at the C- or N-terminus of HomB in the form of MBP-HomB

Side-directed mutagenesis using PCR technique was performed in order to insert a c-Myc or 6x His tag at the C- or N-terminus of HomB before investigating the binding activity of HomB by carbohydrate microarray. This was carried out by using phosphorylated primers listed in Table 5.3. Following successful DNA amplification, ligation, transformation into *E. coli* DH5α, and plasmid purification, it was confirmed that the plasmids extracted from the positive transformants had 100% sequence identity to the recombinant plasmids with the desired insertions of the tags at the C- and N-terminal region of HomB before re-transforming them into *E. coli* Rosetta-gami B (DE3) pLysS.

5.3.4.2 Expression and purification of MBP-HomB containing c-Myc or 6x His tag at the C-terminus

After DNA sequencing confirming the insertion of c-Myc or 6x His tag at the C-terminus of HomB, both constructs containing these two tags were successfully expressed and purified by AC following the protocols used for expression and purification of HomB. After pooling and concentrating AC-elution fractions of both proteins and then diluting them in the binding buffer to a final concentration of 1 mg/mL, SDS-PAGE separation and Western blot analysis using c-Myc and 6x His immunodetection were

performed for the pooled AC-elution fractions within the peak. Unfortunately, no bands were detected for both proteins (Figure 5.20). As a result, the next step was to express the proteins containing both tags at the N-terminus of HomB.

5.3.4.3 Expression and purification of MBP-HomB containing c-Myc or 6x His tag at the N-terminus of HomB

The next step was to express and purify MBP-HomB containing c-Myc or 6x His tag at the N-terminus. Similarly to the previous steps performed, SDS-PAGE separation and Western blot analysis using c-Myc and 6x His immunodetection were performed for the pooled AC-elution fractions within the peak. Fortunately, Western blot showed that MBP-HomB with N-terminal c-Myc tag was positive with a band at ~98.5 kDa, while no bands were detected using 6x immunodetection (Figure 5.21).

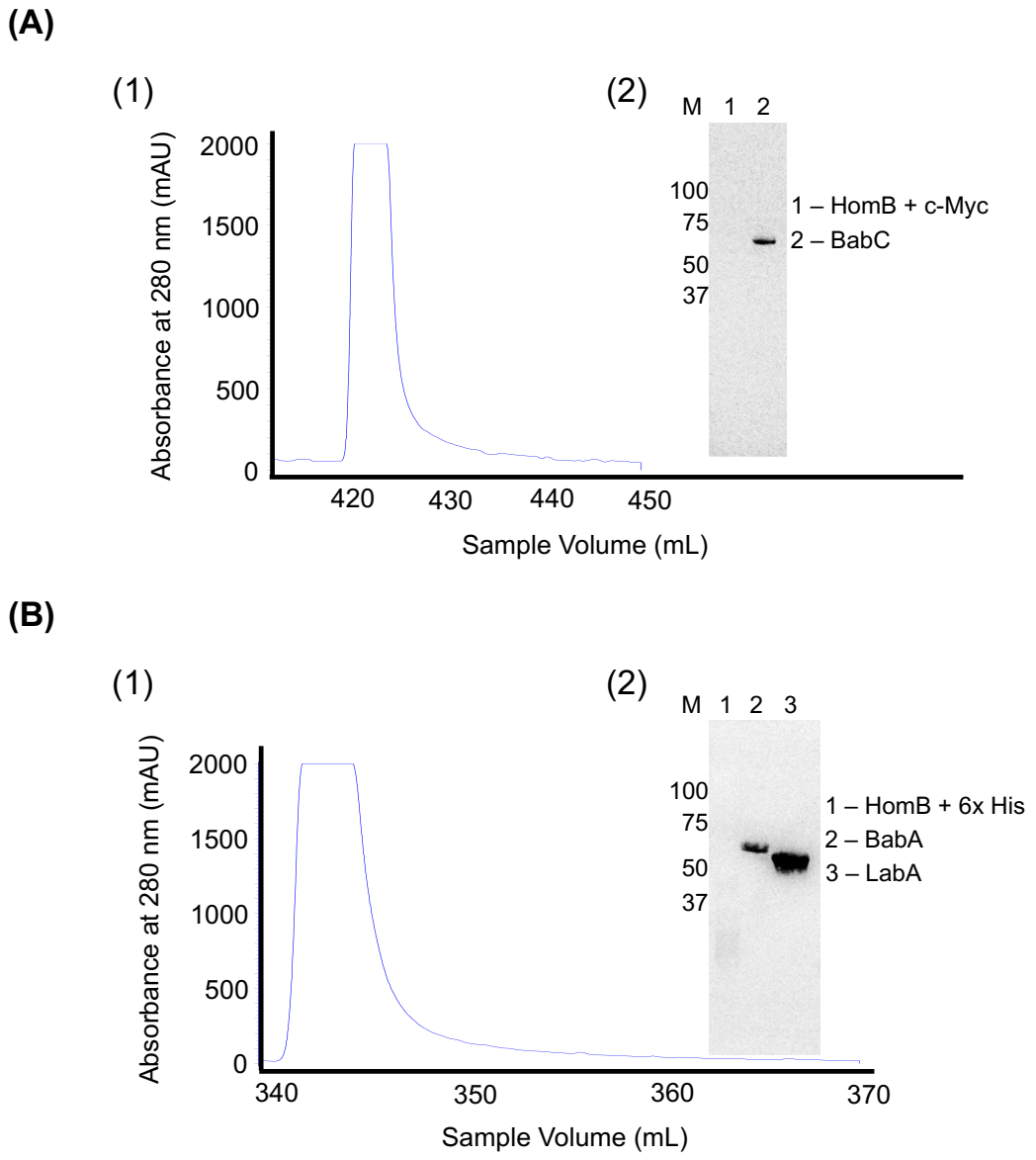


Figure 5.20: Expression and purification of MBP-HomB containing c-Myc or 6x His tag at the C-terminus of HomB

Expression and purification of HomB containing c-Myc (A) or 6x His (B) tag at the C-terminal region. (1) Overlay of AC trace of HomB containing either c-Myc or 6x His tag. The peak appeared after elution of the proteins by a gradient of 10 mM maltose. (2) Western blot using either c-Myc or 6x His immunodetection showing no bands were detected. BabC (~60 kDa), BabA (~59 kDa), and LabA (~54 kDa) were used as positive controls.

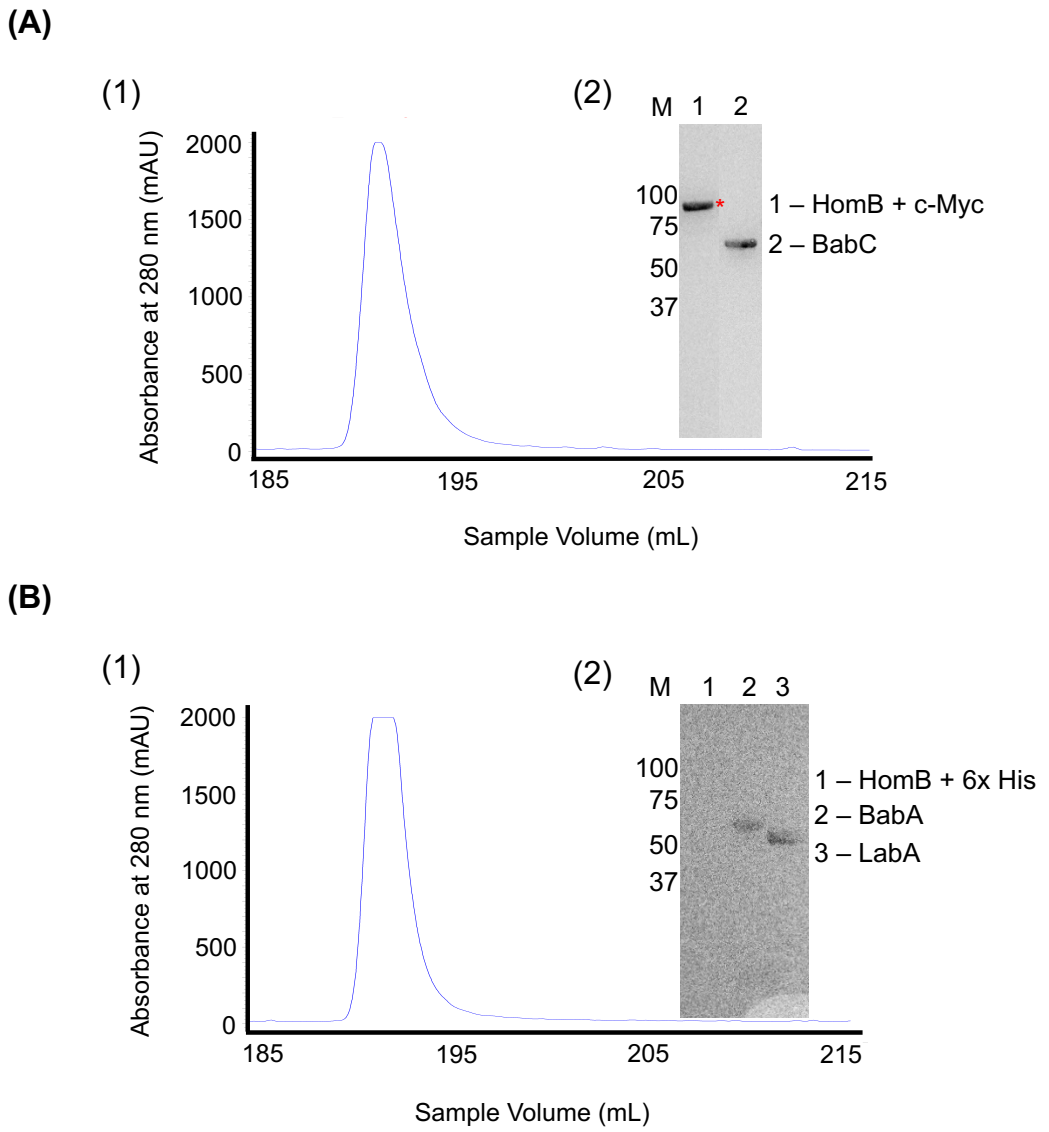


Figure 5.21: Expression and purification of MBP-HomB containing c-Myc or 6x His tag at the N-terminus of HomB

Expression and purification of HomB containing c-Myc (A) or 6x His (B) tag at the N-terminal region. (1) Overlay of AC trace of HomB containing either c-Myc or 6x His tag. The peak appeared after elution of the proteins by a gradient of 10 mM maltose. (2) Western blot using either c-Myc or 6x His immunodetection showing positive result (a band indicated by red star at ~98.5 kDa) with HomB containing c-Myc tag, while no bands were detected with HomB containing 6x His tag. BabC (~60 kDa), BabA (~59 kDa), and LabA (~54 kDa) were used as positive controls.

5.3.4.4 Purification of HomB containing N-terminal c-Myc tag

After cleavage of HomB containing N-terminal c-Myc tag from MBP using FXa protease, further purification of the protein was performed by SEC. Because the insertion of c-Myc tag slightly increases the molecular weight of HomB (~56 kDa), protein fractions were eluted after 50 mL, as illustrated by the peak of the chromatogram in Figure 5.22, while MBP fractions eluted thereafter. The peak structure suggested that no aggregation had taken place, as HomB mutants were not eluted in the void volume suggesting that the proteins are still soluble after their cleavage from MBP fusion proteins. Protein fractions were separated by SDS-PAGE then analysed by InstantBlue staining and Western blot using c-Myc immunodetection (Figure 5.23).

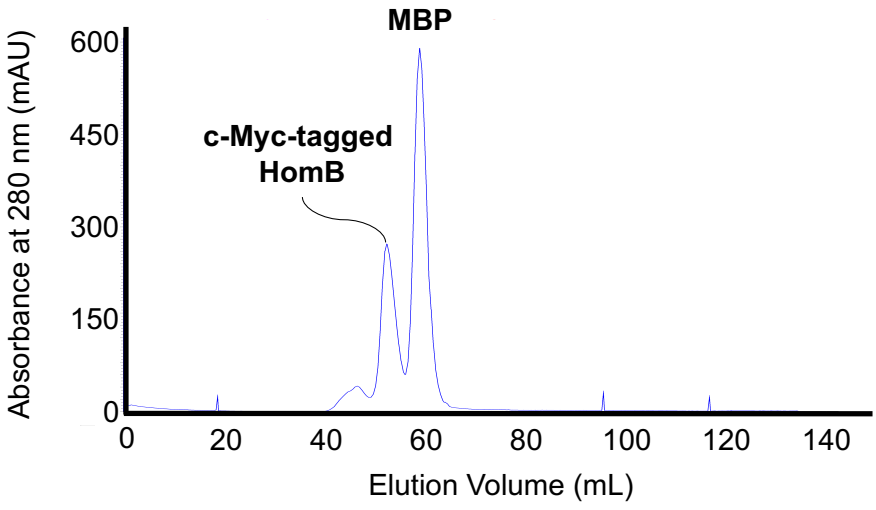


Figure 5.22: SEC purification of HomB containing N-terminal c-Myc tag and MBP fusion protein

Overlay of SEC traces of HomB and MBP proteins. HomB proteins were eluted after 50 mL while MBP fusion proteins were eluted thereafter.

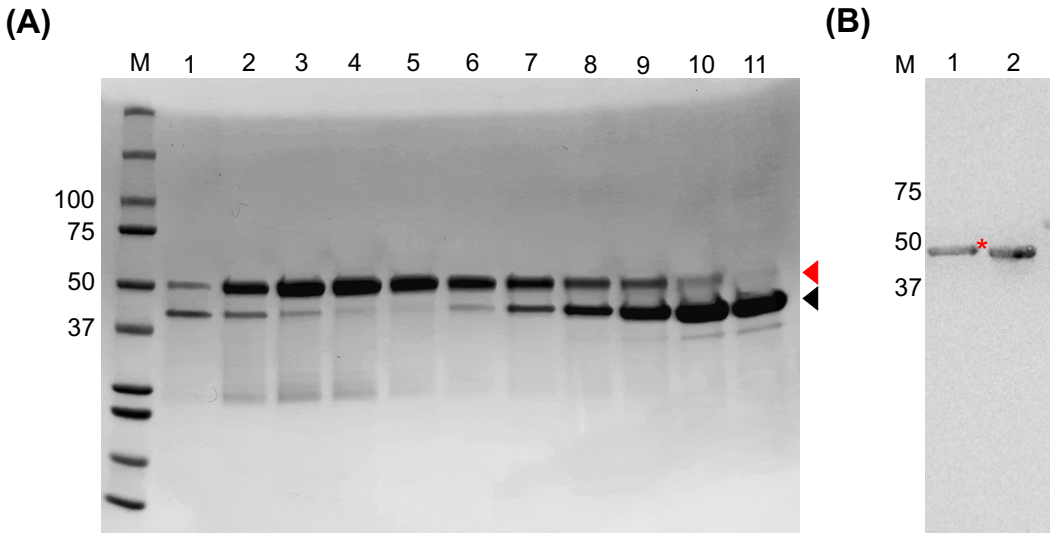


Figure 5.23: Purification of HomB containing c-Myc tag at the N-terminus

(A) InstantBlue staining of SEC-eluted fractions following FXa protease cleavage. The abbreviations denote the following: M – molecular weight marker in kDa; 1-7 – correspond to the fractions from the c-Myc-tagged HomB (red arrowhead at ~56 kDa) peak; and 8-11 – correspond to the fractions from the MBP (black arrowhead at ~42.5 kDa) peak. (B) Western blot using MBP immunodetection confirmed that the band indicated by red asterisk, at the expected molecular weight ~56 kDa, was for c-Myc-tagged HomB after cleavage of MBP-HomB containing c-Myc tag at the N-terminus of HomB by FXa protease and then purification by SEC. M – molecular weight marker in kDa; 1 – c-Myc-tagged HomB; 2 – LabA contains c-Myc tag was used as positive control (~54 kDa).

5.3.4.5 Carbohydrate microarray

Before performing carbohydrate microarray for HomB containing c-Myc tag, the protein was assessed by investigating its proteolytic activity using protease activity assay, in order to confirm whether the protein is functional. HomB was tested alongside BabA, which was used as positive control because it is known to bind type 1 ABO/Le blood group antigens. Furthermore, BabA, expressed by our group, was confirmed to be functional by showing its binding activity to Le^b using an ELISA assay before performing the carbohydrate microarray by the Glycosciences Laboratory at Imperial College London. Unfortunately, no significant binding with the c-Myc-tagged proteins by either BabA or HomB was observed. Figure 5.24 shows the microarray results plotted at histogram charts.

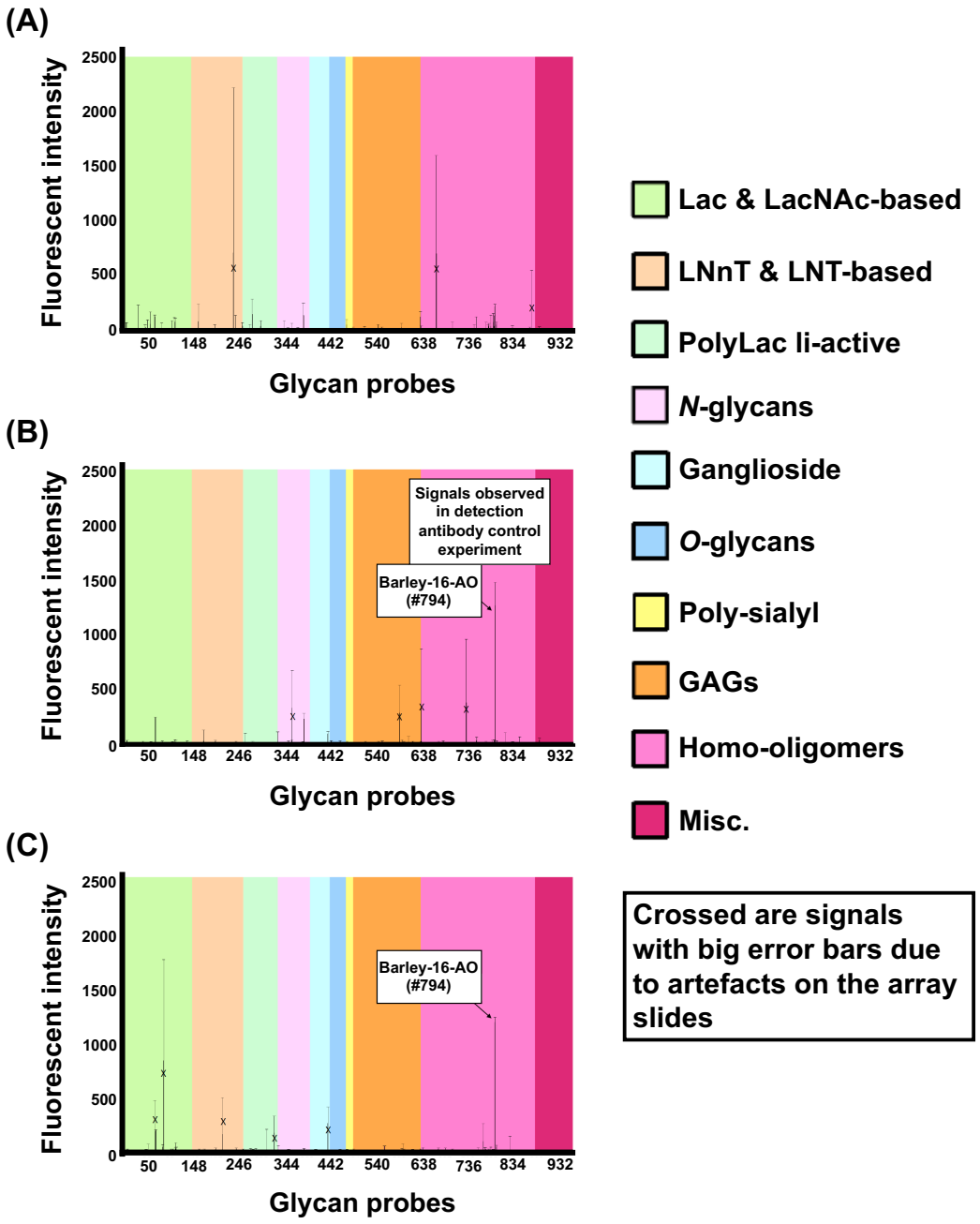


Figure 5.24: Histogram charts of carbohydrate microarray results
Carbohydrate microarray was performed on HomB (A), BabA (B), and c-Myc Monoclonal Antibody (9E10) (C) using screening array Sets 42-56. No binding was detected for both HomB and BabA (positive control), making the results unreliable.

5.3.5 Resistance of HomB to peptic degradation

In order to examine its resistance to peptic digestion, HomB was assessed in parallel with bovine β-lactoglobulin, which is resistant to peptic digestion *in vitro* as mentioned earlier (223). Both proteins showed various predicted pepsin cleavage sites through analysis of amino acid sequences at pH 1.3 and pH > 2. These predicted pepsin cleavage sites were mapped on the predicted structure of HomB J99 (Figure 5.25).

Despite the existence of several predicted pepsin cleavage sites, it was hypothesised that because of steric hindrance, many cleavage sites would be inaccessible once the protein was correctly folded. To determine this experimentally, peptic digestion for HomB and β-lactoglobulin was done at pH 1.2 and pH 4.5 in variable durations where samples were analysed by SDS-PAGE and visualised with InstantBlue following incubation with pepsin. Interestingly, the full-length HomB started to be degraded and fully digested by pepsin at pH 1.2 after 10 minutes incubation. The degradation was characterised by HomB fragment generation. Furthermore, the generated fragment by peptic digestion could resist peptic digestion even after 1 hour and disappeared after that (Figure 5.26A).

At pH 4.5, HomB started to be affected by peptic degradation after 10 minutes incubation until its fragment appeared after 1-hour incubation (Figure 5.26B). The most interesting feature is that HomB and its fragment were still observed even after 6 hours incubation with pepsin. On the other

hand, as expected, β-lactoglobulin was completely resistant to peptic digestion throughout the two hours of incubation with pepsin (Figure 5.26C).

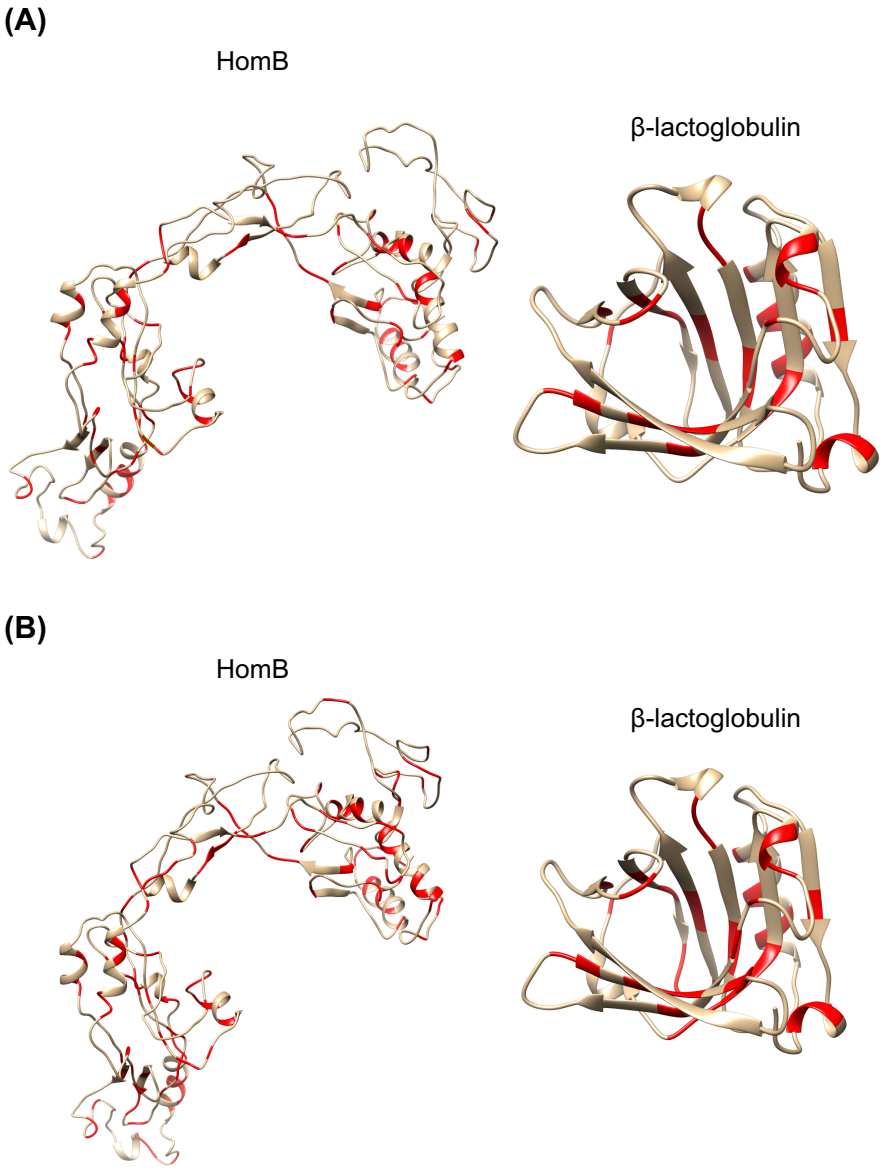


Figure 5.25: Predicted structure of HomB and crystal structure of β-lactoglobulin showing potential pepsin cleavage sites
The potential pepsin cleavage sites at pH 1.3 (A) and at pH > 2 (B) were mapped on the Phyre^2-predicted structure of HomB (20% confidence) and the crystal structure of β-lactoglobulin and indicated in red colour. The cleavage sites were expected based on the possible cleavage of the peptide bond of the amino acid sequence.

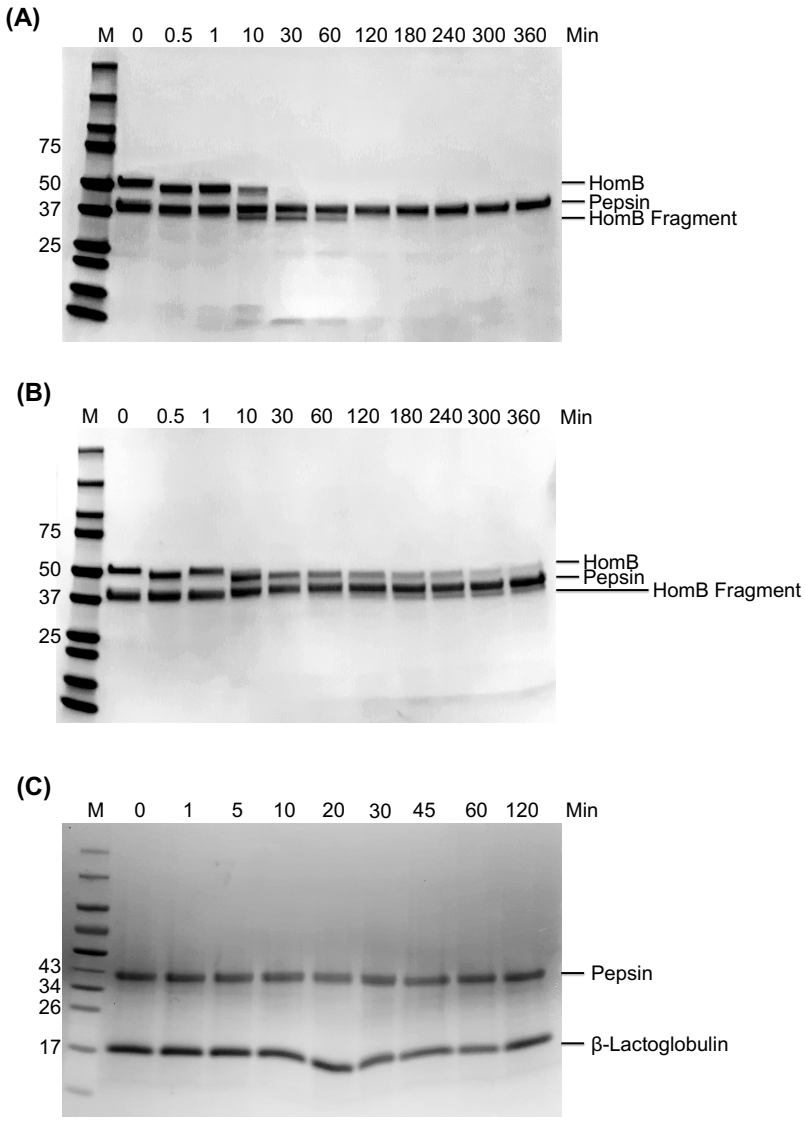


Figure 5.26: Peptic digestion pattern of HomB and β-lactoglobulin
InstantBlue-staining gels showing a time-course of peptic digestion for HomB at pH 1.2 (A) and pH 4.5 (B) and β-lactoglobulin at pH 4.5 (C) using the ratio recommended by the UPS and the NF which is 1 µg of test protein per 10 U of pepsin. HomB was fully degraded by pepsin at pH 1.2 between 10 and 30 minutes, while intact HomB was still observed even after 6 hours incubation with pepsin at pH 4.5 but it started affected by peptic digestion after 10 minutes until its fragmentation was observed after 1-hour incubation with pepsin. This suggests that HomB protein is resistant to the peptic digestion for long time. HomB resistant fragments are indicated. β-lactoglobulin was resistant to the peptic digestion for two hours of incubation with pepsin. Time points indicated are in minutes. M – molecular weight marker in kDa in which Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standards was used in (A) and (B), while Fisher BioReagents™ EZ-Run™ Prestained Rec Protein Ladder, Fisher BioReagents was used in (C).

5.3.6 Cell culture

5.3.6.1 RT-PCR for detection of IL-8 and PAR2 mRNAs expressed by AGS cells

The expression of IL-8 and PAR2 mRNAs was constitutively detected in untreated AGS cells. The detection was also performed for GAPDH mRNA as positive control because it is a housekeeping gene. Figure 5.27 confirms that mRNAs of IL-8, PAR2, and GAPDH were expressed in the untreated AGS cells using RT-PCR. This detection facilitated performing the next step, which was detection of IL-8 release after treating the AGS cells with HomB, and measurement of the IL-8 induction by ELISA.

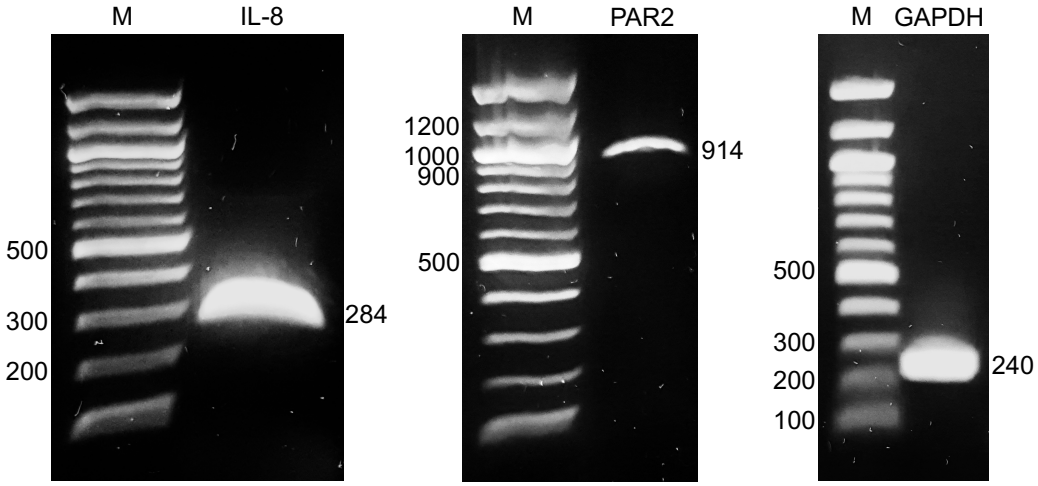


Figure 5.27: RT-PCR for expression detection of IL-8 and PAR2 mRNAs
Agarose gel stained with EtBr representing the results of RT-PCR for expression detection of IL-8 and PAR2 mRNAs. GAPDH was also amplified as this enzyme is a well-known housekeeping gene (positive control). M – 100 bp DNA ladder (New England Biolabs, USA) and the lengths indicated are in bp.

5.3.6.2 IL-8 production from AGS cells treated with WT HomB, E359A mutant, and E359A-M372A mutant

It was important to investigate whether the recombinant HomB induces IL-8 in AGS cells by using ELISA assay. Furthermore, the cells were stimulated by HomB mutants E359A and E359A-M372A in order to assess the effect of these mutants on IL-8 release. PAR2 agonist was used as a positive control because it induces IL-8 production through cleavage of PAR2 receptor as previously discussed in Section 5.1.2.

As shown in Figure 5.28, all forms of HomB significantly induced IL-8 production. Interestingly, the induction of IL-8 production was more significantly enhanced by HomB mutant E359A than WT HomB. Moreover, there was no significant difference in IL-8 induction between WT HomB and the mutant E359A-M372A or between both mutants.

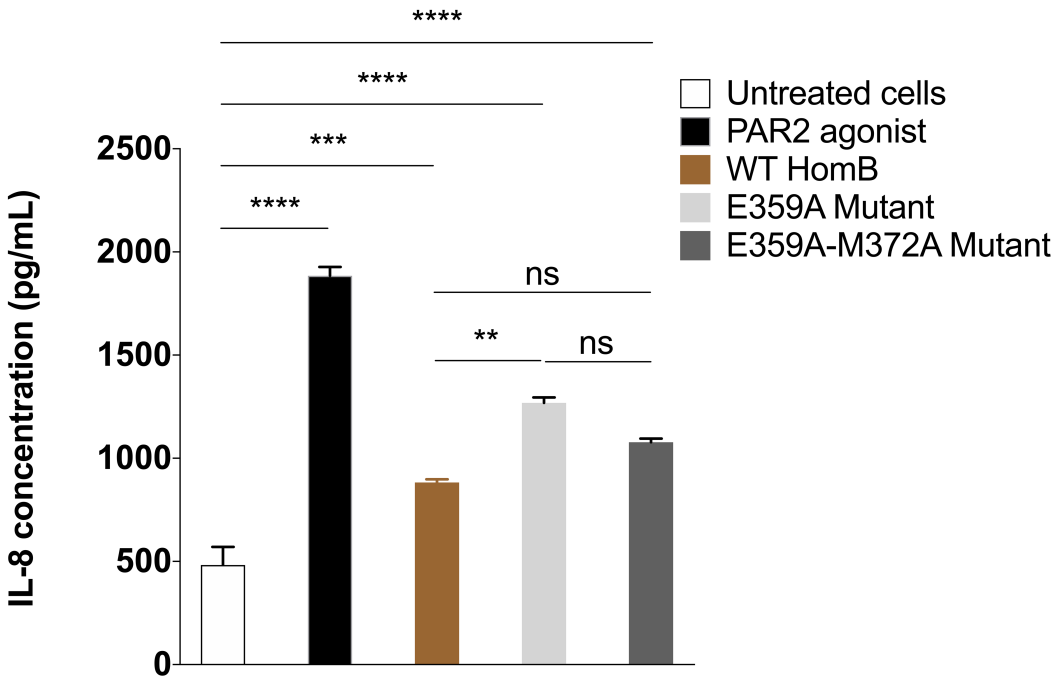


Figure 5.28: IL-8 production from AGS cells treated with *H. pylori* HomB
AGS cells were stimulated by WT HomB, E359A mutant, and E359A-M372A mutant. 24 hours later, ELISA was used to assess the IL-8 protein in the supernatants. Each value represents the mean ± SE of 3 independent experiments. Data were determined by one-way ANOVA. ns indicates not significant; ** ($P < 0.01$); *** ($p < 0.001$); **** ($p < 0.0001$) indicate significant difference.

5.4 Discussion

Following the successful expression and purification of HomB, functional characterisation of the protein was the final objective of this research. The 3D structure of HomB was predicted using Phyre^2. Interestingly, it was detected that HomB contains the Zn-binding motif **HE**XX**H**XX**G**XX**H**, followed by an M-turn. This motif was the only sequence identified by the web through discovering similarities between this motif and the ones in proteases belonging to metzincins. This discovery was very important as it pointed to its potential proteolytic activity; therefore, most of the work described in this chapter was directed at confirming that whether this protein is a metalloprotease.

The key factor for understanding the functionality of HomB, in particular in the context of its suggested role in exacerbation of gastric inflammation, is to prove its proteolytic activity. Therefore, HomB was subjected to protease activity assay. The activity of HomB was maximal at pH 8.5 (50 mM Tris-Cl), indicating that HomB is an alkaline protease, such as *Bacillus intermedius* MprBi (383), *Pseudomonas aeruginosa* alkaline protease (AprA) (384), and others. Therefore, the pH medium at 8.5 was used throughout the protease activity assays. Given that the pH in the colon is slightly basic, and in light of the maximal protease activity found in our study, it is comprehensible to suggest that HomB may be involved in infection of the large intestines. What advantage this gives to *H. pylori*, an organism known to reside in the acidic environment of the stomach still, however, eludes us. The DNA of *H. pylori*

has been identified in the colon (385). However, the association between *H. pylori* infection and inflammatory bowel disease is still debatable (386), and the colorectal cancer which is thought to be induced by *H. pylori* remains hypothetical and requires additional studies (387).

Although FXa protease was removed after cleavage of HomB from the MBP fusion protein using the Xarrest Agarose, the previous experiment would raise questions whether the proteolytic activity on the FITC-labelled casein could have been caused by contaminating traces of FXa. It is known that FXa is a serine protease (388) and can cleave PAR2 (361), and thus can interfere with our hypothesis required to be confirmed that is HomB can induce IL-8 production by cleavage of PAR2. There are two ways to confirm the proteolytic activity was not because of FXa. The first way was to examine the proteolytic activity of the protein MBP-HomB. However, this way was not the main way because the MBP fusion partner, which is around 42.5 kDa (322), could affect the proteolytic activity of HomB by blocking its active site, and thus the proteolytic activity of HomB would give a false negative result. The second way was to examine the proteolytic activity of HomB after treatment with the serine inhibitor AEBSF, which was performed together with other protease inhibitors and will be discussed later. Fortunately, the MBP-HomB showed proteolytic activity, confirming that HomB was the reason behind the proteolytic activity in the previous assay as well as the large molecular weight of MBP did not block the active site of HomB in order to work as a protease. As a result, FXa had no effects on the

proteolytic activity, and it was proved that FXa was completely removed before the SEC purification by the Xarrest Agarose.

After confirming that HomB is a protease, the effects of divalent metal cations Zn^{2+} and Ca^{2+} on the activity of HomB were investigated. It was reported that there is virtually no effect when some metalloproteases are pre-treated with low concentration of Zn^{2+} (10 μM), such as MprBi (383) and thermolysin which is a gluzincin (389). In fact, it has been shown that the proteolytic activity of some metalloproteases sharply decreases with higher concentrations of Zn^{2+}, as was the case for protease B (390), MprBi (383), angiotensin converting enzyme (391), human neutrophil collagenase (392), thermolysin (389), and carboxypeptidase A (393), among others. On the other hand, *Photorhabdus* PrtA and other serralysins are not inhibited by excess amounts of Zn^{2+} (390). The inhibition by excess Zn^{2+} is due to the formation of zinc monohydroxide by the inhibitory zinc, which binds to the glutamic acid carboxylate in the active site of the protease, generating a stabilising hydroxide bridge between the glutamic acid carboxylate and the catalytic Zn ion (394,395) (Figure 5.29). Unlike other metalloproteases, HomB was partially inhibited by low concentration of Zn^{2+} (10 μM), but this inhibition increased with the higher concentrations of Zn^{2+}, reaching complete inhibition at 1 mM. There is no obvious reason for the inhibition with low concentration of Zn^{2+}, but HomB may had Zn ion bound to the active site and any further addition of Zn^{2+} can inhibit its activity.

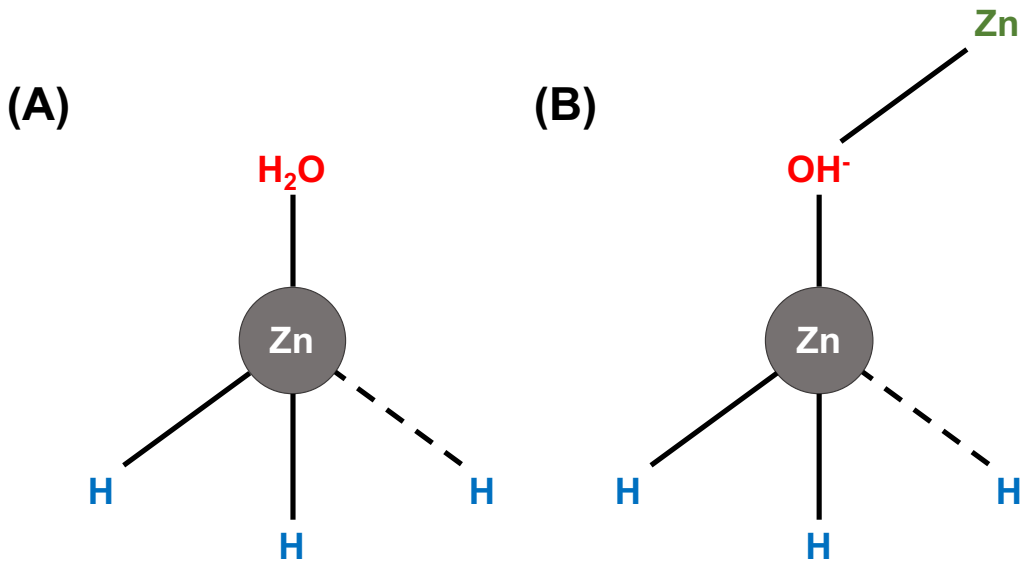


Figure 5.29: The effect of excess amounts of Zn on the proteolytic activity of metalloproteases
(A) Normal state of proteolytic activity of metalloproteases. (B) Inhibition of proteolytic activity of metalloproteases by zinc monohydroxide formed by the inhibitory zinc, generating a stabilising hydroxide bridge between the glutamic acid carboxylate and the catalytic Zn ion.

On the other hand, Ca^{2+} ion is important for some metalloproteases, such as MMPs and ADAMs which are calcium-dependent zinc-containing endopeptidases (328,345,396). Furthermore, it was reported the importance of Ca^{2+} ion for both thermal stability and activity of karilysin, which is an MMP-like enzyme, where 5 mM $CaCl_2$ increased the activity of the protease by 262% compare to the protease without additives (397). It is known that Ca^{2+} stabilises the structure of MMPs and increases the proteolytic activity by stimulating a conformational change. This will switch the protease from an inactive to an active conformation (398). MprBi also showed increasing in its activity by 30% with 10 mM $CaCl_2$, while the activity decreased with

20 mM $CaCl_2$ (383). HomB activity was increased by approximately 41% and 34% with 1 and 2 mM $CaCl_2$, while the activity started decreasing with higher concentrations of $CaCl_2$. The increased in HomB activity could be due to a conformational change caused by Ca^{2+}, while there is no obvious reason why the activity decreases with higher concentrations of Ca^{2+}.

After showing the effects of divalent metal cations Zn^{2+} and Ca^{2+} on the activity of HomB, it was important to show the effects of chosen protease inhibitors and the reducing agent DTT on the proteolytic activity of HomB. EDTA and 1,10-phenanthroline are chelating agents that have the ability to bind with high affinity and sequester the catalytic metal ion, inactivating the metalloprotease (399). EDTA can inactivate both Zn-dependent metalloproteases and some Ca-stabilised peptidases from different classes, while 1,10-phenanthroline has higher affinity for Zn^{2+} (2.5×10^6 M^{-1}) than for Ca^{2+} (3.2 M^{-1}), and low concentration of 1,10-phenanthroline can inactivate metalloprotease but does not have the capacity to chelate Ca^{2+} from Ca-binding proteins. Therefore, 1,10-phenanthroline is ideal as metalloprotease inhibitor (399). The inhibition by EDTA and 1,10-phenanthroline is commonly reversible in which the proteolytic activity can be recovered after removal of the chelating agent and addition of a metal (400). Furthermore, other protease inhibitors that inhibit the proteolytic activity of metalloproteases were used in order to make sure that the protein of interest is only a metalloprotease and not inhibited by these inhibitors. The reducing agent DTT was also used in order to cleave the disulphide

bonds (401) as HomB contains six cysteines and thus it can have up to three intramolecular disulphide bonds. Therefore, it was interesting to investigate the effect of cleavage of disulphide bond on the proteolytic activity of HomB. Unexpectedly, the activity of HomB was not inhibited or even reduced by 1,10-phenanthroline that generally inhibits metalloproteases. 1 mM 1,10-phenanthroline non-significantly increased the protein activity, while 10 mM from the same inhibitor significantly increased the activity. It would be reasonable to assume that 1,10-phenanthroline cannot inhibit or reduce the activity of HomB, if the affinity of HomB for the metal ion is much higher than the inhibitor's affinity, but the inhibitor increased the activity. For example, the binding affinity of dipeptidyl peptidase (DPP) III, a Zn-dependent aminopeptidase (402), for Zn is $4.5 \pm 0.1 \times 10^{-13}$ (403). This means that the binding affinity of the DPP III for Zn is 10^7 higher than that of 1,10-phenanthroline for the same ion. It is, therefore, suggested that a molar ratio of 10^7 molecules is necessary to induce appreciable inhibition. Investigating whether purified native HomB contains Zn^{2+} using atomic absorption spectroscopy (AAS) is, thus, warranted.

EDTA, on the other hand, increased the activity of HomB and did not show any inhibition or at least reduction in the activity of the protease. It was reported that the activity of both MMP-2 and MMP-9 increased in plasma-EDTA compared to plasma-citrate, plasma-heparin, and serum (404). Another observation reported that MMP-9 activity was increased in plasma-EDTA (405). The odd cysteine, found in the pro-peptide

domain of the MMPs, blocks the active site of these metalloproteases by forming an intramolecular complex with the Zn^{2+}, and the dissociation of this cysteine from the catalytic Zn^{2+} by “cysteine switch” can activate the metalloproteases (406). The possible explanation for the increased activity of both MMP-2 and MMP-9 is that EDTA displaces the cysteine during chelating the Zn^{2+}, generating a partial unfolding of the pro-peptide and an active protein (404,405,407).

With regard to the other protease inhibitors, leupeptin (inhibitor of serine and cysteine proteases), AEBSF (serine proteases inhibitor), and pepstatin A (Aspartic protease inhibitor) significantly increased the activity of HomB, while E-64 (cysteine protease inhibitor) had no influence on the activity of HomB. AEBSF result confirmed that FXa was completely removed by the Xarrest Agarose and it did not interfere with the proteolytic activity in the previous experiments. The most interesting feature is that the protein was active in the presence of metal ion chelators 1,10-phenanthroline and EDTA, which was expected to inhibit HomB similar to metzincins. Therefore, further studies are needed to investigate this abnormality in the world of metalloproteases. More investigations are also required in order to study the mechanism of the increased activity in the presence of other protease inhibitors and DTT.

After obtaining the findings regarding HomB activity in the presence of protease inhibitors and DTT, it would be of interest to investigate the effect of the Zn-binding motif, asparagine, and the conserved methionine on the

protein activity, especially after the interesting findings with metal ion chelators. Mutation of residues in the Zn-binding motif should affect the binding to the Zn ion and/or substrate hydrolysis. It was reported that three histidine residues in metzincins act as ligands for the catalytic Zn ion (344,345), and thus it is predicted that the Zn binding will be affected when these histidine residues are mutated (408). Therefore, two mutants H358A and H358A-H362A were designed to investigate the effect of these histidine residues on the enzymatic activity of HomB through affecting the binding of the catalytic Zn ion.

The glutamic acid located after the first histidine in the active site is important and thought to donate hydrogen atoms and to activate water molecule bound to Zn^{2+} at the active site, which acts in a nucleophilic attack of the carbonyl group of the substrate peptide bond. It is predicted that the proteolytic activity of the metalloprotease will be abolished when the glutamic acid residue is mutated (337,409). For example, it was reported that the substitution of the glutamic acid in the active site with alanine in pregnancy-associated plasma protein A and human matrilysin, abolished the enzymatic activity (410,411). Therefore, mutation of the glutamic acid (E359A) was designed to investigate the effect of this residue on the enzymatic activity of HomB through elimination of the general acid/base property and the nucleophilic attack caused by the Zn-bound water molecule.

The conserved M-turn is located under the catalytic ion, and it is structurally important as it stabilises and supports the active site (334,343-345).

Although this methionine is still debated for strict conservation, this residue has a role in the stability and folding of the catalytic domains (359,412,413) by increasing the hydrophobicity in the catalytic site, which improves the binding of Zn^{2+} to the histidines (414). Studies conducted mutational investigations of the conserved methionine on various metzincin families showed different findings (359). Studies stated that mutation of the methionine diminishes the catalytic activity (410,415,416), defects the tissue necrosis factor α-converting enzyme to activate ectodomain shedding (417). On the other hand, mutation of methionine to leucine or serine did not change the activity of MMP-2, while this metalloprotease was completely inactive and susceptible to proteolytic degradation when the methionine was substituted with cysteine (413). Furthermore, the activity and thermal stability of AprA was not significantly changed after mutation of the methionine to L-difluoromethionine (384). Therefore, mutation of the conserved methionine (M372A) was designed to investigate the effect of this residue on the catalytic activity and stability of HomB.

Through analysing some sequences of the Zn-binding motif in some metzincins families, the residue located after the third histidine is only conserved within families. This residue is characteristic for the different subfamilies, i.e. serralysins contain proline residue (418,419), astacins contain glutamic acid residue, MMPs contain serine residue (334), ADAMs contain aspartic acid residue (420), and toxilysins contain asparagine (356). HomB contains asparagine residue (N369) after the third histidine residue in

the Zn-binding motif, and it would be of interest to investigate its role in the catalytic activity and the stability of the protein.

Interestingly, our findings obtained from mutations in the Zn-binding motif and the conserved methionine of HomB were completely different to the results seen for other metzincins. H358A mutation did not alter the activity of HomB, while the activity was increased with the double mutant H358A-H362A. It is not clear why neither of the mutations abolished or at least reduced the activity of HomB. These mutations were expected to affect the binding of the catalytic Zn ion to the active site, and this affects the enzymatic activity of HomB. Furthermore, E359A mutation slightly increased the activity of HomB, instead of decreasing or abolishing the activity. There is no obvious reason for the increased activity, which is unusual for metzincins.

The double mutations E359A-M372A increased the activity of HomB, which is different to other metzincins, in which their activities are either not affected or reduced/abolished after mutation of the conserved methionine. However, this mutant appears to be unstable due to generating two bands at around 30 kDa and 25 kDa. Furthermore, the other double mutant E359A-N369A also appeared to be unstable due to generating the same two bands, but its enzymatic activity was not affected. FXa protease was used to cleave the MBP fusion partner from HomB mutants by recognising the FXa cleavage site (I-E/D-G-R), cleaving after the arginine residue. However, FXa sometimes cleaves at other sites based on the substrate conformation, and

the most common secondary site is glycine-arginine. It was observed that FXa can cleave at secondary sites in unstable proteins in *E. coli*, which may suggest that these proteins are in a partially unfolded state (421)[1]. Therefore, this could perhaps explain the degradation of both double mutants E359A-M372A and E359A-N369A, in which the sequence of HomB J99 contains glycine-arginine. This means that these two mutants may be in a partially unfolded state. Furthermore, this could indicate that the conserved methionine and the asparagine are important in the stability and folding of HomB protein. An alternative explanation could be that the substitutions created a new cleavage site for an autoproteolytic activity of HomB. X-ray crystallography are required to study the 3D structure of the protein and to understand its behaviour behind its catalytic activity.

Consideration was given to the role of HomB in the binding of *H. pylori* to gastric epithelial cells or gastric mucosa. It has been shown that HomB is expressed in the outer membrane of *H. pylori* and involved in bacterial adherence (83). It was reported that *H. pylori* strains bound to AGS cells, while it was shown significant reduction in the binding of the corresponding *homB* knockout mutant strains to AGS cells (114). However, because the knocking out *homB* did not totally abolish bacterial adherence to AGS cells, HomB protein is suggested to be involved in bacterial adherence but not the main adhesin (83,114). It would be of interest to investigate the binding with Le^b-deficient gastric cell lines, such as Kato III cells, for which the adhesion

[1] Walker, I., Riggs, P., Unpublished Observations

of *H. pylori* is independent of the binding of BabA to Le^b (114). Therefore, it was of interest to investigate the role of HomB in the bacterial adhesion by performing carbohydrate microarray in order to identify its binding receptor. Although BabA binding was detected to five glycans that contained a unique determinant region of H-1, A-1 or B-1 blood group antigens by glycan array performed by the CFG (86), the current carbohydrate microarray did not detect any binding to BabA that used as positive control. This can raise questions on the procedures they used for the detection of protein binding to the glycans. Therefore, the results are unreliable, and the array can be performed in another lab research centre, such as the CFG.

Consideration was also given to the influence of the low pH of gastric fluid on the stability of the protein in the presence of pepsin. The information about peptic digestion was discussed in Section 3.4. pH values of 4.5 and 1.2 were used in this experiment. Interestingly, the full-length HomB showed resistance to peptic digestion until 10 minutes of incubation with pepsin at pH 1.2, in which the protein started to degrade and be fully digested, while its fragment showed resistance to the peptic digestion after incubation of 1 hour, which is much longer time compared to the complete degradation of BabC that was fully degraded after 30 seconds at the same pH value. The most interesting feature is that HomB and its fragment, which appeared after 1-hour incubation, were observed even after 6 hours incubation with pepsin at pH 4.5, while BabA and BabC were fully digested after 1 and 5 minutes, respectively. It is known that pepsin cannot hydrolyse all proteins

(e.g. β-lactoglobulin) because the conformational folds of these proteins can make the peptic cleavage sites inaccessible (223). Despite the existence of several predicted pepsin cleavage sites in HomB protein, determined using ExPASy PeptideCutter tool (234), it was hypothesised that because of steric hindrance, many cleavage sites would be inaccessible once the protein was correctly folded. This suggests that HomB protein is resistant to the peptic digestion in the stomach for long time, and it therefore may have a more prominent role at the low pH value in the stomach. Further studies are needed to investigate the importance of this phenomenon.

After confirming the proteolytic activity of HomB using FITC-labelled casein as a substrate, it was interesting to confirm its proteolytic activity within the cells using AGS cells in order to prove our hypothesis that is HomB activates PAR2. PAR2 is expressed on many cell types including the gastrointestinal system (361,380), and inflammatory stimulants, such as LPS, IL-1α, or TNF-α, increase the expression of PAR2 (366,422). It has been reported that PAR2 plays a vital role in the pathogenesis of immunological and inflammatory disorders in many organs including gastric mucosa, and its activation induces proinflammatory cytokines, such as IL-8 production (361), which activates and recruits the neutrophils to the infection site (379). Furthermore, PAR2 activation promotes pepsin/pepsinogen secretion in gastric mucosa of rats (423). It has also been reported that induction of IL-8 production by gastric epithelial cells through the activation of PAR2 by serine proteases of *H. pylori* has been involved in the pathogenesis of gastric

inflammation (379). Therefore, it is evident that PAR2 is involved in the gastric inflammation.

The confirmation of the hypothesis would be through inducing the IL-8 production from the AGS cells in presence and absence of anti-PAR2. RT-PCR proved the expression of PAR2 and IL-8 mRNAs by AGS cells in order to proceed the experiments. Oleastro *et al.* showed that the IL-8 production obtained with *homB*-knockout *H. pylori* was lower than that obtained with the WT (114). Our findings support the previous data by showing that IL-8 production was induced when AGS cells treated with the recombinant HomB J99. This indicates that HomB are involved in the pathogenesis of gastric inflammation.

Interestingly, the mutants E359A and E359A-M372A induced IL-8 production. Furthermore, the IL-8 production obtained with the mutant E359A was higher than that obtained with the WT HomB, while there was no significant difference in IL-8 induction between WT HomB and the mutant E359A-M372A. As mentioned earlier, the proteolytic activity of E359A-M372A was higher than the WT HomB. Therefore, the IL-8 levels associated with the mutant E359A-M372A might reached the peak and started to decrease after 24 hr incubation with AGS cells, and this showed that no significant difference in IL-8 induction between WT HomB and the mutant E359A-M372A. Therefore, it would be better to measure the IL-8 levels with the mutant E359A-M372A at various durations. However, endotoxins of the expression host *E. coli* often contaminate recombinant

proteins (424), and IL-8 production could be induced by LPS of *E. coli* in AGS cells (425). The level of endotoxins can be measured using a Limulus amoebocyte lysate test or Endosafe-PTS system (426). Therefore, it would be of interest to eliminate the endotoxin using a 1% Triton X-114 phase separation method or polymyxin B columns (424) before cell culture investigations.

It has been shown that HomB is indeed a protease and the recombinant HomB induces IL-8 production from the gastric epithelial cells, which had not been previously reported. This confirms a potential role of HomB in the pathogenesis of *H. pylori* as it could initiate and/or promote chronic inflammation in the stomach. However, it was unable to corroborate that HomB is a metalloprotease, particularly a metzincin, based on the lack of inhibition by metalloprotease inhibitors and mutagenesis experiments despite that the structural analysis would suggest that it is a metzincin. This may indicate that the proteolytic activity is attributed to another region of the protein. However, even this couldn't be confirmed, as the proteolytic activity was not inhibited by any of the used protease inhibitors, which covered the four main mechanistic classes of proteases (serine-, cysteine- metallo- and aspartyl-proteases).

Although HomA and HomB share 90% sequence homology at the N- and C-terminal regions with the difference in the mid region (83,89,114) and both of them contain the Zn-binding motif in their sequences, they are nevertheless different in the contribution to bacterial pathogenesis. It has

been reported that HomB is associated with the presence of CagA and linked with peptic ulcer disease and gastric cancer, while HomA is not associated with the presence of CagA and it is linked with duodenal disease (114-116). There may be various reasons that explain the strong association of HomB with gastric cancer, but not with HomA as follows: (i) both proteins contain Zn-binding motif and their structural analysis would suggest that they are metzincins, but they are not active metzincins, and only HomB has proteolytic activity; or (ii) both HomA and HomB are proteolytically active, but the association with carcinogenicity originates from other properties, such as adhesin function. Therefore, we were interested in HomB because of its association with gastric cancer.

Our results show that the activity of HomB was maximal at pH 8.5, indicating that HomB is an alkaline protease. HomB also shows increased resistance to peptic digestion at pH 4.5, compared to BabA (86), LabA (110), and BabC. In light of the present observations, it can be hypothesised that if HomB is secreted by *H. pylori* via autoproteolytic cleavage (assuming it would, to some extent, survive gastric digestion, bile acids, and intestinal proteases), it may act remotely in the colon by inducing IL-8 production via PAR2 activation, eventually leading to colitis and colon cancer. Further studies are indeed required to test this hypothesis.

The limitation of this study was that the time is not sufficient in order to continue our work and prove our hypothesis by measuring the IL-8 production from the AGS cells treated with HomB in presence of anti-PAR2.

This is of significance for future studies. If the IL-8 levels produced by WT HomB in presence of anti-PAR2 were reduced or similar to those obtained from untreated cells, this would support our hypothesis that HomB induces IL-8 production through the activation of PAR2.

Chapter 6: General discussion and future works

6.1 The significance of the research and future work

H. pylori is the leading cause of gastric cancer and the first bacterium classified as a carcinogen. It is widespread, infecting about more than 50% of people. One of the most characteristic features of *H. pylori* in establishing infection in a hostile gastric environment is its ability to adhere to the surface of the gastric epithelium, where the pH is close to neutral, hiding underneath a protective mucus layer. Although the bacterium uses multitude of adhesins for attachment, such as BabA, SabA, and HopQ, more work is required to understand the dynamic interplay between these adhesins, some of which have not been fully characterised such as BabC and HomB, and the epithelium/gastric mucosa.

The *H. pylori* proteins BabA and LabA have previously been expressed and characterised by our group (76,109). However, BabA was the only adhesin that showed binding to its host receptor. Accordingly, this study was initiated to address the gap in knowledge regarding other *H. pylori* OMPs, such as BabC and HomB. Both proteins are putative adhesins used by *H. pylori* for colonising the gastric epithelium and/or mucosa, however, their host receptors are still unidentified. Furthermore, it has been reported that HomB is able to stimulate an *in vitro* proinflammatory response by inducing IL-8 secretion through knockout study, suggesting that HomB is involved in the inflammatory response (114). However, the mechanism through which

HomB induces proinflammatory cytokines and causes gastric inflammation and cancer remains unclear.

The aim of the current study was to express the *H. pylori* OMPs BabC and HomB in a recombinant manner, in order to characterise their functional properties, which have not been studied well previously, with unknown structure and function. The realisation of the properties of BabC as a *H. pylori* adhesin and HomB as a *H. pylori* adhesin and stimulant for proinflammatory cytokines can enable the future design of specific attachment and inflammation inhibitors for treatment of *H. pylori* as well as to avoid inflammation progress to gastric cancer. This realisation was achieved via experimentations aligned to specific study objectives presented in Section 1.6. The novel insights obtained from the findings 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 A strategy for periplasmic expression for *H. pylori* OMPs

The recombinant BabA protein was successfully expressed with high yield and purity after a modification in the protocol used for the expression of BabA in the periplasm of *E. coli* (211), by incorporation of a polycationic hexa-lysine tag at the C-terminus (76). It was also shown that this protocol was very effective for the expression of *H. pylori* LabA from different strains.

Furthermore, the protein was produced with very high amounts either with or without the hexa-lysine tag (109).

In addition to the improvement of protein yield and purity, it has been reported that the incorporation of the hexa-lysine tag at the C-terminus of BabA expression construct resulted in protecting the recombinant BabA from proteolytic degradation in the periplasm of *E. coli* (76), which contains more than twenty proteases with different roles (427). However, the mechanism of protection was not completely understood (76).

On the other hand, the incorporation of the hexa-lysine tag at the C-terminal region of LabA did not have an influence on the protein purity and yield. Furthermore, the tag did not protect LabA from partial N-terminal proteolysis in the periplasm of *E. coli*. However, it was shown that the tag enhanced both solubility and thermal stability of LabA, without changing the structure of the protein (109). Therefore, it was concluded that the insertion of the tag in the expression construct is not necessarily efficient for protein expression in the periplasmic space and each OMP must be investigated individually.

As a result of this, the hexa-lysine tag was incorporated into the expression cassette of BabC as this protein belongs to the same clade as BabA in the Hop family and we assumed the same outcomes of protein expression in the periplasm of the host cells were likely. Fortunately, the protocol was efficient for the expression of *H. pylori* BabC, which contains eight cysteine residues that may form up to four intramolecular disulphide bonds, in the periplasm of *E. coli* with high yield and purity after incorporation of

a C-terminal hexa-lysine tag. Further studies can be achieved in order to investigate the expression of BabC with and without the polycationic hexa-lysine tag.

6.1.2 Development of a cytoplasmic expression strategy for *H. pylori* HomB

The protocol mentioned earlier was suggested to be efficient for expression of other *H. pylori* adhesins because of their extended structural similarity in terms of the overall structural fold. However, despite similarity in the structure, the low level of identity of amino acid primary sequence suggested that this expression protocol may not be widely applicable to all *H. pylori* adhesins. It was shown that the use of this protocol for the expression of SabA and SabB was inappropriate due to their poor purity and their low expression yields (110). Furthermore, this protocol was not effective for the expression of HomB protein, which was aggregated (without polycationic tag) or failed to be expressed (with a polycationic deca-arginine tag). Several attempts using different vectors designed for either periplasmic or cytoplasmic expression were made, using a variety of *E. coli* strains. The N-terminal extracellular domain of HomB contains six cysteine residues, which may suggest that HomB contains up to three intramolecular disulphide bonds. We were able to express HomB in a soluble form in the cytoplasm of *E. coli* at high yield with good purity only after using the pMAL-c5X vector that contains MBP fusion protein, which improves the solubility of the partner

protein. However, the bulk of the expressed protein was aggregated. HomB was then cleaved from the MBP fusion protein using FXa protease, which was then completely removed by Xarrest™ Agarose, and SEC showed that HomB protein was soluble in the buffer, which paved the way for the planned downstream investigations.

6.1.3 Binding activity of BabC and HomB

Although the function of BabC is unknown, it is considered as a member of *H. pylori* adhesins since its gene belongs to the Hop family (247). As a primary experiment, BabC interactions with the common glycan epitopes in gastrointestinal mucins, which are Le^a, Le^b, Le^y, and sLe^x glycoconjugates (263), were investigated using an ELISA-like test. Although BabA was used as positive control for Le^b antigen, the main limitation of this experiment was that there were no positive controls used for the other glycan epitopes because either there are no proteins known to bind to these antigens or the protein is commercially unavailable i.e. SabA adhesin, which is known to bind to sLe^x.

HomB protein is also suggested to be involved in bacterial adherence but not the main adhesin, because knocking out *homB* gene did not totally abolish *H. pylori* adherence to AGS cells (83,114). However, it was suggested to investigate the binding with Le^b-deficient gastric cell lines, such as Kato III cells, for which the adhesion of *H. pylori* is independent of the binding of BabA to Le^b (114).

Therefore, it was of interest to investigate the role of BabC and HomB in the bacterial adhesion by performing carbohydrate microarray in order to identify their binding receptors. The carbohydrate microarray showed no binding activity was detected for both proteins. However, the carbohydrate microarray did not detect any binding to BabA that was used as positive control, which showed binding to Le^b antigen using ELISA before performing the carbohydrate microarray. This can raise questions on the procedures used for the detection of protein binding to the glycans. Therefore, the results are unreliable, and the array can be achieved in another lab research centre, such as the CFG.

With regard to BabC, gene recombination seems to occur among *bab* loci, leading to hybrid genes with unknown binding properties (268). However, the molecular mechanism that regulates the gene conversion is poorly understood. Gene alterations in OMP genes play significant roles in the environmental adaptation including the adhesion of the pathogen to the host cells (85). Therefore, BabC may (i) bind to some different glycans, (ii) bind to non-carbohydrate glycans, as is the case for HopQ, or (iii) only become active after genomic recombination resulting in hybrids of BabA/BabC or BabB/BabC in order to maintain the chronic colonisation.

6.1.4 HomB is a protease

The novelty of this research was in showing that HomB is indeed a protease, as suggested by sequence analysis. Although our structural analysis of

HomB had hinted that this protein could be a metzincin, as it contains the Zn-binding motif HEXXHXXGXXH followed by a conserved methionine, we were unable to corroborate that HomB is a metalloprotease, particularly a metzincin, based on the lack of inhibition by metalloprotease inhibitors and mutagenesis experiments. This may indicate that the proteolytic activity is attributed to another region of the protein. However, even this could not be confirmed, as the proteolytic activity was not inhibited by any of the used protease inhibitors, which covered the four main mechanistic classes of proteases (cysteine-, metallo-, serine-, and aspartyl-proteases). It would be of interest to examine another version of HomB protein from another *H. pylori* strain in order to investigate its proteolytic activity and to explore the effect of protease inhibitors and mutagenesis on its enzymatic activity. Furthermore, it would also be interesting to investigate whether purified native HomB contains any Zn^{2+} using AAS, and whether this Zn^{2+} can be removed from the active site by suitable chelators.

6.1.5 Recombinant HomB induces IL-8 production

It has been shown that HomB is able to stimulate an *in vitro* proinflammatory response by inducing IL-8 secretion. This was confirmed by the reduced ability of *homB*-mutant strains to induce the secretion of IL-8 from gastric epithelial cells (83,114). However, this has not been shown using recombinant purified protein. We were able to confirm that the recombinant HomB was indeed able to induce IL-8 production from gastric epithelial cells,

which had not been previously reported. This confirms a potential role of HomB in the pathogenesis of *H. pylori* as it could initiate and/or promote chronic inflammation in the stomach, ultimately leading to gastric cancer in a subset of patients. This would appear to support the description of HomB as a virulence factor and/or marker for *H. pylori* strains associated with peptic ulcer and gastric cancer.

However, the mechanism of induction of IL-8 production by HomB is unknown. We hypothesised that HomB can induce proinflammatory response by cleavage of the extracellular N-terminal domain of PAR2, which stimulates a signal transduction pathway through the activation of the associated G-protein. This results in induction of proinflammatory cytokines, such as IL-8 and IL-6, causing proinflammatory reactions in the epithelium (377,378), in which IL-8 recruits and activates neutrophils to the infection site (379). However, the time was not sufficient in order to prove this hypothesis. Therefore, it would be of interest to confirm our hypothesis by measuring the IL-8 production from the AGS cells treated with HomB in presence of anti-PAR2. This is of significance for future studies. If the IL-8 levels produced by WT HomB in presence of blocking anti-PAR2 antibodies were reduced or similar to those obtained from untreated cells, this would support our hypothesis that HomB induces IL-8 production through the activation of PAR2. Further evidence could come from using PAR2 negative cell lines stably transfected with PAR2, from cross-desensitisation

experiments with PAR2 agonists, or from knockdown of PAR2 expression by siRNA or CRISPR/Cas-9 experiments.

6.1.6 Resistance of BabC and HomB to peptic degradation

It is reasonable to assume that *H. pylori* has adopted several mechanisms to protect itself from the harmful effects of pepsin in the stomach. Consideration was given to the influence of the low pH of gastric fluid on the stability of BabC and HomB in the presence of pepsin. A pH value of 4.5, which mimics the acidic conditions of both pre- and post-prandial states, was used to achieve the peptic digestion experiments. On the other hand, a pH value of 1.2 was also suggested by the USP and the NF for analysis of the protein digestibility by pepsin (237,272). Therefore, pH values of 4.5 and 1.2 were used.

BabC was rapidly degraded by pepsin at pH 1.2 after 10 seconds and fully digested after 30 seconds, while it was also fully digested by pepsin at pH 4.5 after 5 minutes, although its fragments that were cut from their C-terminal regions could still be detected in SDS-PAGE. Previously, our group found that BabA was rapidly degraded by pepsin at pH 4.5 after 1 minute. Therefore, BabC appears to be unstable at acidic (pH 4.5) conditions in presence of pepsin. Furthermore, it was shown that the C-terminus of BabC contains a highly susceptible protease cleavage site. It was also confirmed that incubation of BabC in the same acidic buffer, without pepsin, did not result in any detectable proteolysis.

However, although BabC is sensitive to peptic digestion, pepsin is not active in the neutral pH environment in the mucus layer of the stomach because of the bicarbonate barrier, hence bacteria close to the epithelium are safe from the effects of pepsin and high acidity. There are various reasons which may explain the sensitivity to peptic digestion. First, *H. pylori* may be able to avoid the effects of HCl/pepsin by its known ability to penetrate the mucus layer and swim towards the epithelial surface, by sensing the urea gradient emanating from the gastric epithelium (277). Furthermore, the urease activity in the bacterium generates a buffered surface layer for the *H. pylori* in response to acidic environment (278), which modulates the enzymatic activity of pepsin. Second, our experiments have been performed with recombinant proteins, while BabC and others in the whole bacterium are embedded in the outer leaflet of the LPS membrane, which may offer substantial protection against pepsin digestion, by steric hindrance, with only the ligand-binding crown protruding from the membrane. In this context, it is interesting to note that while most of the extracellular domain of BabA has many predicted pepsin cleavage sites, the crown appears almost devoid of such cleavage sites. Third, adhesins in *H. pylori* might possess some additional posttranslational modifications, e.g. some glycosylation, which may increase stability to proteolytic degradation (279). Native BabA isolated from *H. pylori* was slightly more stable than recombinant BabA made in *E. coli* (86).

Interestingly, HomB showed resistance to peptic digestion until 10 minutes of incubation with pepsin at pH 1.2, in which the protein started to degrade and be fully digested, while its fragment showed resistance to the peptic digestion after incubation of 1 hour, which is much longer time compared to the complete degradation of BabC that was fully degraded after 30 seconds at the same pH value. The most interesting feature is that HomB and its fragment, which appeared after 1-hour incubation, were observed even after 6 hours incubation with pepsin at pH 4.5, while BabA and BabC were fully digested after 1 and 5 minutes, respectively. It is known that pepsin cannot hydrolyse all proteins (e.g. β-lactoglobulin) because the conformational folds of these proteins can make the peptic cleavage sites inaccessible (223). Despite the existence of several predicted pepsin cleavage sites in HomB protein, determined using ExPASy PeptideCutter tool (234), it was hypothesised that because of steric hindrance, many cleavage sites would be inaccessible once the protein was correctly folded. This suggests that HomB protein is resistant to the peptic digestion in the stomach for a longer time, and it therefore may have a more prominent role at the low pH value in the stomach. Further studies are needed to investigate the importance of this phenomenon. Interestingly, if a host receptor is identified, HomB could be used to generate model biomimetic particles that have gastric epithelium binding properties after removing the Zn-binding motif, provided that it shows resistance to peptic digestion.

6.2 Future work of BabC

It would be of interest to express and purify BabC protein from *H. pylori* strain other than *H. pylori* 26695. This would rule out the possibility that BabC is not functional in the chosen strain. Assuming that the ligand, in analogy to BabA, is a carbohydrate ligand, we could attempt to identify the ligand using carbohydrate arrays. For example, the CFG (http://www.functionalglycomics.org/static/consortium/resources/resourcecoreh.shtml) offers a mammalian array, which in its current version (5.2) contains 609 different glycan targets.

The simplest approach would be to create *H. pylori* BabC deletion mutants and assess the effect of binding, e.g. to mucin-expressing gastric cell lines *in vitro*, using fluorescently labelled bacteria. One could also envisage generating blocking antibodies, provided that BabA, BabB and BabC are sufficiently different to allow the generation of non-crossreactive antibodies. Such antibodies could then be tested for their effect on adhesion of *H. pylori* whole bacteria to cell lines *in vitro*.

Another more complicated approach is the one employed by John Klassen (University of Alberta, Canada) in which CsR-ESI-MS is used to screen a library of oligosaccharides for binding to BabC, as done previously for BabA (428). This method is semi-quantitative, providing a rough estimate of binding affinity.

Should any of these approaches lead to identification of a ligand, co-crystallization, as in the case of BabA (68), could be used to elucidate the interactions at the atomic level.

Once a structure has been obtained and validated experimentally (e.g. by mutagenesis of key interacting residues), the comparison between BabA and BabC binding specificities (or affinities?) may explain the coexistence of three adhesins (BabA, BabB and BabC) with a very similar sequence in one bacterial strain, perhaps shedding some further light on the dynamic pathogen-host interactions at the molecular level.

Carbohydrate microarray is the next plan to do after protein purification of BabC. In case of there is binding, it would be of interest to perform an ELISA-like test for BabC using the glycoconjugate detected in the carbohydrate microarray and to do BabC crystallography in presence and absence of its ligand.

6.3 New alternative hypothesis

Our results show that the activity of HomB was maximal at pH 8.5, indicating that HomB is an alkaline protease. Furthermore, HomB shows increased resistance to peptic digestion at pH 4.5, compared to BabA, BabC, and LabA. In light of the present observations, it can be hypothesised that if HomB is secreted by *H. pylori* through autoproteolytic cleavage (assuming it would, to some extent, survive gastric digestion, bile acids, and intestinal proteases), it may act remotely in the colon by inducing IL-8 production through the activation of PAR2, eventually leading to colitis and colon cancer. Further studies are indeed required to test this hypothesis.

References

1. Danon, S. J., and Lee, A. (2001) Other gastric helicobacters and spiral organisms. in *Helicobacter pylori: Physiology and Genetics*, ASM Press. pp

2. Marshall, B. J. (1986) Campylobacter pyloridis and gastritis. *The Journal of infectious diseases* **153**, 650-657

3. Mazzarello, P., Calligaro, A. L., and Calligaro, A. (2001) Giulio Bizzozero: a pioneer of cell biology. *Nature reviews Molecular cell biology* **2**, 776-781

4. zur Hausen, H. (2006) Infections causing human cancer. 2006. *Germany: Whiley-VCH Verlag GmbH & Co*

5. Salomon, H. (1896) *Über das Spirillum des Säugetiermagens und sein Verhalten zu den Belegzellen*, G. Fischer

6. Balfour, A. (1906) A haemogregarine of mammals and some notes on trypanosomiasis in the Anglo-Egyptian Sudan. *Jour. Trop. Med* **9**, 81-92

7. Lim, R. (1920) A parasitic spiral organism in the stomach of the cat. *Parasitology* **12**, 108-112

8. Bliss, S. (1926) The site of ammonia formation and the prominent role of vomiting in ammonia elimination. *Journal of Biological Chemistry* **67**, 109-140

9. Benedict, S. R., and Nash, T. P. (1926) The site of ammonia formation and the role of vomiting in ammonia elimination. *Journal of Biological Chemistry* **69**, 381-396

10. Modlin, I. M., and Sachs, G. (1998) *Acid Related Diseases-Biology and Treatment*, Schnetztor-Verlag GmbH, Germany

11. Callery, M. P. (2006) Helicobacter pylori in gastric cancer: Friend or foe? *Gastroenterology* **131**, 966-967

12. Doenges, J. L. (1939) Spirochaetes in the gastric glands of Macacus rhesus and of man without related disease. *Arch Path* **27**, 469-477

13. Freedberg, A. S., and Barron, L. E. (1940) The presence of spirochetes in human gastric mucosa. *The American Journal of Digestive Diseases* **7**, 443-445

14. Steer, H. (1984) Surface morphology of the gastroduodenal mucosa in duodenal ulceration. *Gut* **25**, 1203-1210

15. Warren, J. R., and Marshall, B. (1983) Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *The lancet* **321**, 1273-1275

16. Marshall, B. J., Armstrong, J. A., McGechie, D. B., and Clancy, R. J. (1985) Attempt to fulfil Koch's postulates for pyloric Campylobacter. *Medical Journal of Australia* **142**, 436-439

17. Morris, A., and Nicholson, G. (1987) Ingestion of Campylobacter pyloridis causes gastritis and raised fasting gastric pH. *American Journal of Gastroenterology* **82**

18. Hellström, P. M. (2006) This year's Nobel Prize to gastroenterology: Robin Warren and Barry Marshall awarded for their discovery of Helicobacter pylori as pathogen in the gastrointestinal tract. *World Journal of Gastroenterology: WJG* **12**, 3126

19. The Nobel Prize in Physiology or Medicine 2005. http://www.nobelprize.org/nobel_prizes/medicine/laureates/2005/illpres/

20. Cover, T. L., and Blaser, M. J. (1992) Helicobacter pylori and gastroduodenal disease. *Annual review of medicine* **43**, 135-145

21. Louw, J., Falck, V., Van Rensburg, C., Zak, J., Adams, G., and Marks, I. (1993) Distribution of Helicobacter pylori colonisation and associated gastric inflammatory changes: difference between patients with duodenal and gastric ulcers. *Journal of clinical pathology* **46**, 754-756

22. Cave, D. R. (1997) How is Helicobacter pylori transmitted? *Gastroenterology* **113**, S9-S14

23. Brown, L. M. (2000) Helicobacter pylori: epidemiology and routes of transmission. *Epidemiologic reviews* **22**, 283-297

24. Kusters, J. G., van Vliet, A. H., and Kuipers, E. J. (2006) Pathogenesis of Helicobacter pylori infection. *Clinical microbiology reviews* **19**, 449-490

25. Parreira, P., Magalhães, A., Reis, C., Borén, T., Leckband, D., and Martins, M. (2013) Bioengineered surfaces promote specific protein–glycan mediated binding of the gastric pathogen Helicobacter pylori. *Acta biomaterialia* **9**, 8885-8893

26. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **68**, 394-424

27. Barbosa-Jobim, G. S., Costa-Lira, E., Ralph, A. C. L., Gregorio, L., Lemos, T. L. G., Burbano, R. R., Calcagno, D. Q., Smith, M. A. C., Montenegro, R. C., and Vasconcellos, M. C. (2020) Biflorin inhibits the proliferation of gastric cancer cells by decreasing MYC expression. *Toxicol In Vitro* **63**, 104735

28. Chang, A. H., and Parsonnet, J. (2010) Role of bacteria in oncogenesis. *Clinical microbiology reviews* **23**, 837-857

29. Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer* **136**, E359-E386

30. IARC. (1994) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Schistosomes, Liver Flukes and Helicobacter pylori. IARC Working Group, Lyon

31. Polk, D. B., and Peek, R. M., Jr. (2010) Helicobacter pylori: gastric cancer and beyond. *Nature reviews. Cancer* **10**, 403-414

32. Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S. W., Yamaoka, Y., Graham, D. Y., Perez-Trallero, E., Wadstrom, T., Suerbaum, S., and Achtman, M. (2007) An African origin for the intimate association between humans and Helicobacter pylori. *Nature* **445**, 915-918

33. Weeks, D. L., Eskandari, S., Scott, D. R., and Sachs, G. (2000) A H+-gated urea channel: the link between Helicobacter pylori urease and gastric colonization. *Science (New York, N.Y.)* **287**, 482-485

34. Bauerfeind, P., Garner, R., Dunn, B., and Mobley, H. (1997) Synthesis and activity of Helicobacter pylori urease and catalase at low pH. *Gut* **40**, 25-30

35. Celli, J. P., Turner, B. S., Afdhal, N. H., Keates, S., Ghiran, I., Kelly, C. P., Ewoldt, R. H., McKinley, G. H., So, P., Erramilli, S., and Bansil, R. (2009) Helicobacter pylori moves through mucus by reducing mucin viscoelasticity. *Proc Natl Acad Sci U S A* **106**, 14321-14326

36. Gu, H. (2017) Role of Flagella in the Pathogenesis of Helicobacter pylori. *Current microbiology* **74**, 863-869

37. Montecucco, C., and Rappuoli, R. (2001) Living dangerously: how Helicobacter pylori survives in the human stomach. *Nature Reviews Molecular Cell Biology* **2**, 457-466

38. Yoshiyama, H., Nakamura, H., Kimoro, M., Okita, K., and Nakazawa, T. (1999) Chemotaxis and motility of Helicobacter pylori in a viscous environment. *Journal of gastroenterology* **34**

39. Juge, N. (2012) Microbial adhesins to gastrointestinal mucus. *Trends in microbiology* **20**, 30-39

40. Odenbreit, S. (2005) Adherence properties of Helicobacter pylori: impact on pathogenesis and adaptation to the host. *International journal of medical microbiology* **295**, 317-324

41. Rossez, Y., Gosset, P., Boneca, I. G., Magalhães, A., Ecobichon, C., Reis, C. A., Cieniewski-Bernard, C., Joncquel Chevalier Curt, M., Léonard, R., Maes, E., Sperandio, B., Slomianny, C., Sansonetti, P. J., Michalski, J., and Robbe-Masselot, C. (2014) The lacdiNAc-specific adhesin LabA mediates adhesion of Helicobacter pylori to human gastric mucosa. *The Journal of infectious diseases* **210**, 1286-1295

42. Salama, N. R., Hartung, M. L., and Muller, A. (2013) Life in the human stomach: persistence strategies of the bacterial pathogen Helicobacter pylori. *Nature reviews. Microbiology* **11**, 385-399

43. Salama, N., Guillemin, K., McDaniel, T. K., Sherlock, G., Tompkins, L., and Falkow, S. (2000) A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. *Proceedings of the national Academy of Sciences* **97**, 14668-14673

44. Yamaoka, Y. (2012) Pathogenesis of Helicobacter pylori-related gastroduodenal diseases from molecular epidemiological studies. *Gastroenterology research and practice* **2012**

45. Abadi, A. T. B. (2017) Strategies used by helicobacter pylori to establish persistent infection. *World journal of gastroenterology* **23**, 2870

46. Achtman, M., Azuma, T., Berg, D. E., Ito, Y., Morelli, G., Pan, Z. J., Suerbaum, S., Thompson, S. A., van der Ende, A., and van Doorn, L. J. (1999) Recombination and clonal groupings within Helicobacter pylori from different geographical regions. *Mol Microbiol* **32**, 459-470

47. Suerbaum, S., and Josenhans, C. (2007) Helicobacter pylori evolution and phenotypic diversification in a changing host. *Nature reviews microbiology* **5**, 441-452

48. Suerbaum, S., Smith, J. M., Bapumia, K., Morelli, G., Smith, N. H., Kunstmann, E., Dyrek, I., and Achtman, M. (1998) Free recombination within Helicobacter pylori. *Proc Natl Acad Sci U S A* **95**, 12619-12624

49. Falush, D., Kraft, C., Taylor, N. S., Correa, P., Fox, J. G., Achtman, M., and Suerbaum, S. (2001) Recombination and mutation during long-term gastric colonization by Helicobacter pylori: estimates of clock rates, recombination size, and minimal age. *Proc Natl Acad Sci U S A* **98**, 15056-15061

50. Wang, G., Humayun, M. Z., and Taylor, D. E. (1999) Mutation as an origin of genetic variability in Helicobacter pylori. *Trends in microbiology* **7**, 488-493

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

52. Kuipers, E. J., Israel, D. A., Kusters, J. G., Gerrits, M. M., Weel, J., van Der Ende, A., van Der Hulst, R. W., Wirth, H. P., Hook-Nikanne, J., Thompson, S. A., and Blaser, M. J. (2000) Quasispecies development of Helicobacter pylori observed in paired isolates obtained years apart from the same host. *The Journal of infectious diseases* **181**, 273-282

53. Covacci, A., and Rappuoli, R. (1998) Helicobacter pylori: molecular evolution of a bacterial quasi-species. *Current opinion in microbiology* **1**, 96-102

54. Portal-Celhay, C., and Perez-Perez, G. I. (2006) Immune responses to Helicobacter pylori colonization: mechanisms and clinical outcomes. *Clinical science* **110**, 305-314

55. Bergman, M., Del Prete, G., van Kooyk, Y., and Appelmelk, B. (2006) Helicobacter pylori phase variation, immune modulation and gastric autoimmunity. *Nature Reviews Microbiology* **4**, 151-159

56. Moore, M. E., Borén, T., and Solnick, J. V. (2011) Life at the margins: modulation of attachment proteins in Helicobacter pylori. *Gut microbes* **2**, 1593-1600

57. Testerman, T. L., McGee, D. J., and Mobley, H. L. (2001) Adherence and colonization. in *Helicobacter pylori: physiology and genetics*, ASM Press. pp

58. Ilver, D., Arnqvist, A., Ögren, J., Frick, I.-M., Kersulyte, D., Incecik, E. T., Berg, D. E., Covacci, A., Engstrand, L., and Borén, T. (1998) Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science (New York, N.Y.)* **279**, 373-377

59. Mahdavi, J., Sondén, B., Hurtig, M., Olfat, F. O., Forsberg, L., Roche, N., Ångström, J., Larsson, T., Teneberg, S., Karlsson, K.-A., Altraja, S., Wadström, T., Kersulyte, D., Berg, D. E., Dubois, A., Petersson, C., Magnusson, K., Norberg, T., Lindh, F., Lundskog, B. B., Arnqvist, A., Hammarström, L., and Borén, T. (2002) Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. *Science (New York, N.Y.)* **297**, 573-578

60. Königer, V., Holsten, L., Harrison, U., Busch, B., Loell, E., Zhao, Q., Bonsor, D. A., Roth, A., Kengmo-Tchoupa, A., Smith, S. I., Mueller, S., Sundberg, E. J., Zimmermann, W., Fischer, W., Hauck, C. R., and Haas, R. (2016) Helicobacter pylori exploits human CEACAMs via HopQ for adherence and translocation of CagA. *Nature microbiology* **2**, 1-12

61. Javaheri, A., Kruse, T., Moonens, K., Mejías-Luque, R., Debraekeleer, A., Asche, C. I., Tegtmeyer, N., Kalali, B., Bach, N. C., Sieber, S. A., Hill, D. J., Königer, V., Hauck, C. R., Moskalenko, R., Haas, R., Busch, D. H., Klaile, E., Slevogt, H., Schmidt, A., Backert, S., Remaut, H., Singer, B. B., and Gerhard, M. (2016) Helicobacter pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs. *Nature microbiology* **2**, 1-13

62. Odenbreit, S., Swoboda, K., Barwig, I., Ruhl, S., Borén, T., Koletzko, S., and Haas, R. (2009) Outer membrane protein expression profile in Helicobacter pylori clinical isolates. *Infection and immunity* **77**, 3782-3790

63. Odenbreit, S., Till, M., Hofreuter, D., Faller, G., and Haas, R. (1999) 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

64. Peck, B., Ortkamp, M., Diehl, K. D., Hundt, E., and Knapp, B. (1999) Conservation, localization and expression of HopZ, a protein involved in adhesion of Helicobacter pylori. *Nucleic acids research* **27**, 3325-3333

65. Etzold, S., and Juge, N. (2014) Structural insights into bacterial recognition of intestinal mucins. *Current opinion in structural biology* **28**, 23-31

66. Pang, S. S., Nguyen, S. T. S., Perry, A. J., Day, C. J., Panjikar, S., Tiralongo, J., Whisstock, J. C., and Kwok, T. (2014) The three-dimensional structure of the extracellular adhesion domain of the sialic acid-binding

adhesin SabA from Helicobacter pylori. *Journal of Biological Chemistry* **289**, 6332-6340

67. Lee, A., O'Rourke, J., De Ungria, M. C., Robertson, B., Daskalopoulos, G., and Dixon, M. F. (1997) A standardized mouse model of Helicobacter pylori infection: introducing the Sydney strain. *Gastroenterology* **112**, 1386-1397

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

69. Boren, T., Falk, P., Roth, K. A., Larson, G., and Normark, S. (1993) Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens. *Science (New York, N.Y.)* **262**, 1892-1895

70. Ishijima, N., Suzuki, M., Ashida, H., Ichikawa, Y., Kanegae, Y., Saito, I., Borén, T., Haas, R., Sasakawa, C., and Mimuro, H. (2011) BabA-mediated adherence is a potentiator of the Helicobacter pylori type IV secretion system activity. *Journal of Biological Chemistry* **286**, 25256-25264

71. Ghosh, P., Sarkar, A., Ganguly, M., Alam, J., De, R., and Mukhopadhyay, A. K. (2016) Helicobacter pylori strains harboring babA2 from Indian sub population are associated with increased virulence in ex vivo study. *Gut pathogens* **8**, 1

72. Fujimoto, S., Ojo, O. O., Arnqvist, A., Wu, J. Y., Odenbreit, S., Haas, R., Graham, D. Y., and Yamaoka, Y. (2007) Helicobacter pylori BabA expression, gastric mucosal injury, and clinical outcome. *Clinical Gastroenterology and Hepatology* **5**, 49-58

73. Rad, R., Gerhard, M., Lang, R., Schöniger, M., Rösch, T., Schepp, W., Becker, I., Wagner, H., and Prinz, C. (2002) The Helicobacter pylori blood group antigen-binding adhesin facilitates bacterial colonization and augments a nonspecific immune response. *The Journal of Immunology* **168**, 3033-3041

74. Gerhard, M., Lehn, N., Neumayer, N., Borén, T., Rad, R., Schepp, W., Miehlke, S., Classen, M., and Prinz, C. (1999) Clinical relevance of the Helicobacter pylori gene for blood-group antigen-binding adhesin. *Proceedings of the National Academy of Sciences* **96**, 12778-12783

75. Matsuo, Y., Kido, Y., and Yamaoka, Y. (2017) Helicobacter pylori outer membrane protein-related pathogenesis. *Toxins* **9**, 101

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

77. Lindén, S., Mahdavi, J., Hedenbro, J., Borén, T., and Carlstedt, I. (2004) Effects of pH on Helicobacter pylori binding to human gastric mucins: identification of binding to non-MUC5AC mucins. *Biochemical journal* **384**, 263-270

78. Bugaytsova, J. A., Bjornham, O., Chernov, Y. A., Gideonsson, P., Henriksson, S., Mendez, M., Sjostrom, R., Mahdavi, J., Shevtsova, A., Ilver, D., Moonens, K., Quintana-Hayashi, M. P., Moskalenko, R., Aisenbrey, C., Bylund, G., Schmidt, A., Aberg, A., Brannstrom, K., Koniger, V., Vikstrom, S., Rakhimova, L., Hofer, A., Ogren, J., Liu, H., Goldman, M. D., Whitmire, J. M., Aden, J., Younson, J., Kelly, C. G., Gilman, R. H., Chowdhury, A., Mukhopadhyay, A. K., Nair, G. B., Papadakos, K. S., Martinez-Gonzalez, B., Sgouras, D. N., Engstrand, L., Unemo, M., Danielsson, D., Suerbaum, S., Oscarson, S., Morozova-Roche, L. A., Olofsson, A., Grobner, G., Holgersson, J., Esberg, A., Stromberg, N., Landstrom, M., Eldridge, A. M., Chromy, B. A., Hansen, L. M., Solnick, J. V., Linden, S. K., Haas, R., Dubois, A., Merrell, D. S., Schedin, S., Remaut, H., Arnqvist, A., Berg, D. E., and Boren, T. (2017) Helicobacter pylori Adapts to Chronic Infection and Gastric Disease via pH-Responsive BabA-Mediated Adherence. *Cell Host Microbe* **21**, 376-389

79. Aspholm-Hurtig, M., Dailide, G., Lahmann, M., Kalia, A., Ilver, D., Roche, N., Vikström, S., Sjöström, R., Lindén, S., Bäckström, A., Lundberg, C., Arnqvist, A., Mahdavi, J., Nilsson, U. J., Velapatiño, B., Gilman, R. H., Gerhard, M., Alarcon, T., López-Brea, M., Nakazawa, T., Fox, J. G., Correa, P., Dominguez-Bello, M. G., Perez-Perez, G. I., Blaser, M. J., Normark, S., Carlstedt, I., Oscarson, S., Teneberg, S., Berg, D. E., and Borén, T. (2004) Functional adaptation of BabA, the H. pylori ABO blood group antigen binding adhesin. *Science (New York, N.Y.)* **305**, 519-522

80. Benktander, J., Ångström, J., Breimer, M. E., and Teneberg, S. (2012) Redefinition of the carbohydrate binding specificity of Helicobacter pylori BabA adhesin. *Journal of Biological Chemistry* **287**, 31712-31724

81. Pride, D. T., Meinersmann, R. J., and Blaser, M. J. (2001) Allelic variation within Helicobacter pylori babA and babB. *Infection and immunity* **69**, 1160-1171

82. Colbeck, J. C., Hansen, L. M., Fong, J. M., and Solnick, J. V. (2006) Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of Helicobacter pylori. *Infection and immunity* **74**, 4375-4378

83. Oleastro, M., and Ménard, A. (2013) The role of Helicobacter pylori outer membrane proteins in adherence and pathogenesis. *Biology* **2**, 1110-1134

84. Backert, S., Clyne, M., and Tegtmeyer, N. (2011) Molecular mechanisms of gastric epithelial cell adhesion and injection of CagA by Helicobacter pylori. *Cell Communication and Signaling* **9**, 28

85. Hanada, K., and Yamaoka, Y. (2014) Genetic battle between Helicobacter pylori and humans. The mechanism underlying homologous recombination in bacteria, which can infect human cells. *Microbes and infection* **16**, 833-839

86. Hage, N. (2015) *Helicobacter pylori Biomimics for Gastric-Targeted Drug Delivery.* Doctor of Philosophy, University of Nottingham

87. Marcos, N. T., Magalhães, A., Ferreira, B., Oliveira, M. J., Carvalho, A. S., Mendes, N., Gilmartin, T., Head, S. R., Figueiredo, C., David, L., Santos-Silva, F., and Reis, C. A. (2008) Helicobacter pylori induces β3GnT5 in human gastric cell lines, modulating expression of the SabA ligand sialyl–Lewis x. *The Journal of clinical investigation* **118**, 2325-2336

88. Goodwin, A. C., Weinberger, D. M., Ford, C. B., Nelson, J. C., Snider, J. D., Hall, J. D., Paules, C. I., Peek Jr, R. M., and Forsyth, M. H. (2008) Expression of the Helicobacter pylori adhesin SabA is controlled via phase variation and the ArsRS signal transduction system. *Microbiology (Reading, England)* **154**, 2231

89. Alm, R. A., Bina, J., Andrews, B. M., Doig, P., Hancock, R. E., and Trust, T. J. (2000) Comparative genomics of Helicobacter pylori: analysis of the outer membrane protein families. *Infection and immunity* **68**, 4155-4168

90. Kawai, M., Furuta, Y., Yahara, K., Tsuru, T., Oshima, K., Handa, N., Takahashi, N., Yoshida, M., Azuma, T., Hattori, M., Uchiyama, I., and Kobayashi, I. (2011) Evolution in an oncogenic bacterial species with extreme genome plasticity: Helicobacter pylori East Asian genomes. *BMC microbiology* **11**, 104

91. Exner, M. M., Doig, P., and Hancock, R. (1995) Isolation and characterization of a family of porin proteins from Helicobacter pylori. *Infection and immunity* **63**, 1567-1572

92. Alm, R. A., Ling, L.-S. L., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C., Dejonge, B. L., Carmel, G., Tummino, P. J., Caruso, A., Uria-Nickelsen, M., Mills, D. M., Ives, C., Gibson, R., Merberg, D., Mills, S. D., Jiang, Q., Taylor, D. E., Vovis, G. F.,

and Trust, T. J. (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. *Nature* **397**, 176

93. Oh, J. D., Kling-Backhed, H., Giannakis, M., Xu, J., Fulton, R. S., Fulton, L. A., Cordum, H. S., Wang, C., Elliott, G., Edwards, J., Mardis, E. R., Engstrand, L. G., and Gordon, J. I. (2006) The complete genome sequence of a chronic atrophic gastritis Helicobacter pylori strain: evolution during disease progression. *Proc Natl Acad Sci U S A* **103**, 9999-10004

94. Aspholm, M., Olfat, F. O., Norden, J., Sonden, B., Lundberg, C., Sjostrom, R., Altraja, S., Odenbreit, S., Haas, R., Wadstrom, T., Engstrand, L., Semino-Mora, C., Liu, H., Dubois, A., Teneberg, S., Arnqvist, A., and Boren, T. (2006) SabA is the H. pylori hemagglutinin and is polymorphic in binding to sialylated glycans. *PLoS pathogens* **2**, e110

95. Shimoda, A., Ueda, K., Nishiumi, S., Murata-Kamiya, N., Mukai, S.-a., Sawada, S.-i., Azuma, T., Hatakeyama, M., and Akiyoshi, K. (2016) Exosomes as nanocarriers for systemic delivery of the Helicobacter pylori virulence factor CagA. *Scientific Reports* **6**, 1-9

96. Xia, X., Zhang, L., Chi, J., Li, H., Liu, X., Hu, T., Li, R., Guo, Y., Zhang, X., Wang, H., Cai, J., Li, Y., Liu, D., Cui, Y., Zheng, X., Flaker, G. C., Liao, D., Hao, H., Liu, Z., and Xu, C. (2020) Helicobacter pylori Infection Impairs Endothelial Function Through an Exosome-Mediated Mechanism. *J Am Heart Assoc* **9**, e014120

97. Unemo, M., Aspholm-Hurtig, M., Ilver, D., Bergström, J., Borén, T., Danielsson, D., and Teneberg, S. (2005) The sialic acid binding SabA adhesin of Helicobacter pylori is essential for nonopsonic activation of human neutrophils. *Journal of Biological Chemistry* **280**, 15390-15397

98. De Jonge, R., Pot, R. G., Loffeld, R. J., Van Vliet, A. H., Kuipers, E. J., and Kusters, J. G. (2004) The functional status of the Helicobacter pylori sabB adhesin gene as a putative marker for disease outcome. *Helicobacter* **9**, 158-164

99. Talarico, S., Whitefield, S. E., Fero, J., Haas, R., and Salama, N. R. (2012) Regulation of Helicobacter pylori adherence by gene conversion. *Molecular microbiology* **84**, 1050-1061

100. Cao, P., and Cover, T. L. (2002) Two different families of hopQ alleles in Helicobacter pylori. *Journal of clinical microbiology* **40**, 4504-4511

101. Cao, P., Lee, K. J., Blaser, M. J., and Cover, T. L. (2005) Analysis of hopQ alleles in East Asian and Western strains of Helicobacter pylori. *FEMS microbiology letters* **251**, 37-43

102. Belogolova, E., Bauer, B., Pompaiah, M., Asakura, H., Brinkman, V., Ertl, C., Bartfeld, S., Nechitaylo, T. Y., Haas, R., Machuy, N., Salama, N., Churin, Y., and Meyer, T. F. (2013) Helicobacter pylori outer membrane protein HopQ identified as a novel T4SS-associated virulence factor. *Cellular microbiology* **15**, 1896-1912

103. Busch, B., Weimer, R., Woischke, C., Fischer, W., and Haas, R. (2015) Helicobacter pylori interferes with leukocyte migration via the outer membrane protein HopQ and via CagA translocation. *International Journal of Medical Microbiology* **305**, 355-364

104. Trust, T. J., Doig, P., Emody, L., Kienle, Z., Wadstrom, T., and O'Toole, P. (1991) High-affinity binding of the basement membrane proteins collagen type IV and laminin to the gastric pathogen Helicobacter pylori. *Infection and immunity* **59**, 4398-4404

105. Senkovich, O. A., Yin, J., Ekshyyan, V., Conant, C., Traylor, J., Adegboyega, P., McGee, D. J., Rhoads, R. E., Slepenkov, S., and Testerman, T. L. (2011) Helicobacter pylori AlpA and AlpB bind host laminin and influence gastric inflammation in gerbils. *Infection and immunity* **79**, 3106-3116

106. Lu, H., Wu, J. Y., Beswick, E. J., Ohno, T., Odenbreit, S., Haas, R., Reyes, V. E., Kita, M., Graham, D. Y., and Yamaoka, Y. (2007) Functional and intracellular signaling differences associated with the Helicobacter pylori AlpAB adhesin from Western and East Asian strains. *The Journal of biological chemistry* **282**, 6242-6254

107. de Jonge, R., Durrani, Z., Rijpkema, S. G., Kuipers, E. J., van Vliet, A. H. M., and Kusters, J. G. (2004) Role of the Helicobacter pylori outer-membrane proteins AlpA and AlpB in colonization of the guinea pig stomach. *Journal of medical microbiology* **53**, 375-379

108. Solnick, J. V., Hansen, L. M., Salama, N. R., Boonjakuakul, J. K., and Syvanen, M. (2004) Modification of Helicobacter pylori outer membrane protein expression during experimental infection of rhesus macaques. *Proceedings of the National Academy of Sciences* **101**, 2106-2111

109. Paraskevopoulou, V., Artiaga, V. G., Rowlinson, R., Winkler, G. S., Gellert, P., Stolnik, S., Overman, R., and Falcone, F. H. (2019) Introduction of a C-terminal hexa-lysine tag increases thermal stability of the LacDiNac binding adhesin (LabA) exodomain from Helicobacter pylori. *Protein Expression and Purification* **163**, 105446

110. Paraskevopoulou, V. (2018) *Gastric-Targeted Drug Delivery by Learning from Helicobacter pylori.* Doctor of Philosophy, University of Nottingham

111. Mthembu, Y. H., Jin, C., Padra, M., Liu, J., Edlund, J. O., Ma, H., Padra, J., Oscarson, S., Borén, T., Karlsson, N. G., Lindén, S. K., and Holgersson, J. (2020) Recombinant mucin-type proteins carrying LacdiNAc on different O-glycan core chains fail to support H. pylori binding. *Mol Omics* **16**, 243-257

112. Paraskevopoulou, V., Schimpl, M., Overman, R. C., Stolnik, S., Chen, Y., Nguyen, L., Winkler, G. S., Gellert, P., Klassen, J. S., and Falcone, F. H. (2021) Structural and binding characterization of the LacdiNAc-specific adhesin (LabA; HopD) exodomain from Helicobacter pylori. *Current Research in Structural Biology* **3**, 19-29

113. Oleastro, M., Monteiro, L., Lehours, P., Mégraud, F., and Ménard, A. (2006) Identification of markers for Helicobacter pylori strains isolated from children with peptic ulcer disease by suppressive subtractive hybridization. *Infection and immunity* **74**, 4064-4074

114. Oleastro, M., Cordeiro, R., Ferrand, J., Nunes, B., Lehours, P., Carvalho-Oliveira, I., Mendes, A. I., Penque, D., Monteiro, L., Mégraud, F., and Ménard, A. (2008) Evaluation of the clinical significance of homB a novel candidate marker of helicobacter pylori strains associated with peptic ulcer disease. *Journal of Infectious Diseases* **198**, 1379-1387

115. Jung, S. W., Sugimoto, M., Graham, D. Y., and Yamaoka, Y. (2009) homB status of Helicobacter pylori as a novel marker to distinguish gastric cancer from duodenal ulcer. *Journal of clinical microbiology* **47**, 3241-3245

116. Abadi, A. T. B., Rafiei, A., Ajami, A., Hosseini, V., Taghvaei, T., Jones, K. R., and Merrell, D. S. (2011) Helicobacter pylori homB, but not cagA, is associated with gastric cancer in Iran. *Journal of clinical microbiology* **49**, 3191-3197

117. Oleastro, M., Cordeiro, R., Ménard, A., Yamaoka, Y., Queiroz, D., Mégraud, F., and Monteiro, L. (2009) Allelic diversity and phylogeny of homB, a novel co-virulence marker of Helicobacter pylori. *BMC microbiology* **9**, 248

118. Oleastro, M., Cordeiro, R., Yamaoka, Y., Queiroz, D., Mégraud, F., Monteiro, L., and Ménard, A. (2009) Disease association with two Helicobacter pylori duplicate outer membrane protein genes, homB and homA. *Gut pathogens* **1**, 12

119. Kang, J., Jones, K. R., Jang, S., Olsen, C. H., Yoo, Y.-J., Merrell, D. S., and Cha, J.-H. (2012) The geographic origin of Helicobacter pylori influences the association of the homB gene with gastric cancer. *Journal of clinical microbiology* **50**, 1082-1085

120. Gaddy, J. A., Radin, J. N., Loh, J. T., Zhang, F., Washington, M. K., Peek Jr, R. M., Algood, H. M. S., and Cover, T. L. (2013) High dietary salt intake exacerbates Helicobacter pylori-induced gastric carcinogenesis. *Infection and immunity* **81**, 2258-2267

121. Tan, M., He, F. J., Wang, C., and MacGregor, G. A. (2019) Twenty-four-hour urinary sodium and potassium excretion in China: a systematic review and meta-analysis. *Journal of the American Heart Association* **8**, e012923

122. Azadnajafabad, S., Ebrahimi, N., Mohammadi, E., Ghasemi, E., Saeedi Moghaddam, S., Aminorroaya, A., Rezaei, N., Ghanbari, A., Masinaei, M., Mohammadi Fateh, S., Haghshenas, R., Gorgani, F., Kazemi, A., Dilmaghani-Marand, A., and Farzadfar, F. (2021) Disparities and Spatial Variations of High Salt Intake in Iran: A Subnational Study of Districts Based on the Small Area Estimation Method. *Public Health Nutr*, 1-28

123. Servetas, S. L., Doster, R. S., Kim, A., Windham, I. H., Cha, J. H., Gaddy, J. A., and Merrell, D. S. (2018) ArsRS-Dependent Regulation of homB Contributes to Helicobacter pylori Biofilm Formation. *Frontiers in microbiology* **9**, 1497

124. Servetas, S. L., Carpenter, B. M., Haley, K. P., Gilbreath, J. J., Gaddy, J. A., and Merrell, D. S. (2016) Characterization of key Helicobacter pylori regulators identifies a role for ArsRS in biofilm formation. *Journal of bacteriology* **198**, 2536-2548

125. Yonezawa, H., Osaki, T., Fukutomi, T., Hanawa, T., Kurata, S., Zaman, C., Hojo, F., and Kamiya, S. (2017) Diversification of the AlpB outer membrane protein of Helicobacter pylori affects biofilm formation and cellular adhesion. *Journal of bacteriology* **199**

126. Acio-Pizzarello, C. R., Acio, A. A., Choi, E. J., Bond, K., Kim, J., Kenan, A. C., Chen, J., and Forsyth, M. H. (2017) Determinants of the regulation of Helicobacter pylori adhesins include repeat sequences in both promoter and coding regions as well as the two-component system ArsRS. *Journal of medical microbiology* **66**, 798-807

127. Hathroubi, S., Servetas, S. L., Windham, I., Merrell, D. S., and Ottemann, K. M. (2018) Helicobacter pylori biofilm formation and its potential role in pathogenesis. *Microbiology and Molecular Biology Reviews* **82**

128. Haddadi, M. H., Negahdari, B., Asadolahi, R., and Bazargani, A. (2020) Helicobacter pylori antibiotic resistance and correlation with cagA motifs and homB gene. *Postgrad Med*, 1-9

129. Matysiak-Budnik, T., and Mégraud, F. (2006) Helicobacter pylori infection and gastric cancer. *European journal of cancer* **42**, 708-716

130. Bäckhed, F., Rokbi, B., Torstensson, E., Zhao, Y., Nilsson, C., Seguin, D., Normark, S., Buchan, A. M. J., and Richter-Dahlfors, A. (2003) Gastric mucosal recognition of Helicobacter pylori is independent of Toll-like receptor 4. *The Journal of infectious diseases* **187**, 829-836

131. Segal, E., Cha, J., Lo, J., Falkow, S., and Tompkins, L. (1999) Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by Helicobacter pylori. *Proceedings of the National Academy of Sciences* **96**, 14559-14564

132. Asahi, M., Azuma, T., Ito, S., Ito, Y., Suto, H., Nagai, Y., Tsubokawa, M., Tohyama, Y., Maeda, S., Omata, M., Suzuki, T., and Sasakawa, C. (2000) Helicobacter pylori CagA protein can be tyrosine phosphorylated in gastric epithelial cells. *The Journal of experimental medicine* **191**, 593-602

133. Odenbreit, S., Püls, J., Sedlmaier, B., Gerland, E., Fischer, W., and Haas, R. (2000) Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion. *Science (New York, N.Y.)* **287**, 1497-1500

134. Stein, M., Rappuoli, R., and Covacci, A. (2000) Tyrosine phosphorylation of the Helicobacter pylori CagA antigen after cag-driven host cell translocation. *Proceedings of the National Academy of Sciences* **97**, 1263-1268

135. Peek, R. M., and Blaser, M. J. (2002) Helicobacter pylori and gastrointestinal tract adenocarcinomas. *Nature Reviews Cancer* **2**, 28-37

136. Nilsson, C., Sillén, A., Eriksson, L., Strand, M.-L., Enroth, H., Normark, S., Falk, P., and Engstrand, L. (2003) Correlation between cag pathogenicity island composition and Helicobacter pylori-associated gastroduodenal disease. *Infection and immunity* **71**, 6573-6581

137. Stein, M., Bagnoli, F., Halenbeck, R., Rappuoli, R., Fantl, W. J., and Covacci, A. (2002) c-Src/Lyn kinases activate Helicobacter pylori CagA through tyrosine phosphorylation of the EPIYA motifs. *Molecular microbiology* **43**, 971-980

138. Selbach, M., Moese, S., Hauck, C. R., Meyer, T. F., and Backert, S. (2002) Src is the kinase of the Helicobacter pylori CagA protein in vitro and in vivo. *Journal of Biological Chemistry* **277**, 6775-6778

139. Tammer, I., Brandt, S., Hartig, R., König, W., and Backert, S. (2007) Activation of Abl by Helicobacter pylori: a novel kinase for CagA and crucial mediator of host cell scattering. *Gastroenterology* **132**, 1309-1319

140. Moese, S., Selbach, M., Kwok, T., Brinkmann, V., König, W., Meyer, T. F., and Backert, S. (2004) Helicobacter pylori induces AGS cell motility and elongation via independent signaling pathways. *Infection and immunity* **72**, 3646-3649

141. Churin, Y., Al-Ghoul, L., Kepp, O., Meyer, T. F., Birchmeier, W., and Naumann, M. (2003) Helicobacter pylori CagA protein targets the c-Met receptor and enhances the motogenic response. *The Journal of cell biology* **161**, 249-255

142. Meyer-ter-Vehn, T., Covacci, A., Kist, M., and Pahl, H. L. (2000) Helicobacter pylori activates mitogen-activated protein kinase cascades and induces expression of the proto-oncogenes c-fos and c-jun. *Journal of Biological Chemistry* **275**, 16064-16072

143. Amieva, M. R., Vogelmann, R., Covacci, A., Tompkins, L. S., Nelson, W. J., and Falkow, S. (2003) Disruption of the epithelial apical-junctional complex by Helicobacter pylori CagA. *Science (New York, N.Y.)* **300**, 1430-1434

144. Atherton, J. C. (2006) The pathogenesis of Helicobacter pylori–induced gastro-duodenal diseases. *Annu. Rev. Pathol. Mech. Dis.* **1**, 63-96

145. Gebert, B., Fischer, W., Weiss, E., Hoffmann, R., and Haas, R. (2003) Helicobacter pylori vacuolating cytotoxin inhibits T lymphocyte activation. *Science (New York, N.Y.)* **301**, 1099-1102

146. Schmees, C., Gerhard, M., Treptau, T., Voland, P., Schwendy, S., Rad, R., and Prinz, C. (2006) VacA-Associated Inhibition of T-cell Function: Reviewed and Reconsidered. *Helicobacter* **11**, 144-146

147. de Bernard, M., Papini, E., de Filippis, V., Gottardi, E., Telford, J., Manetti, R., Fontana, A., Rappuoli, R., and Montecucco, C. (1995) Low pH activates the vacuolating toxin of Helicobacter pylori, which becomes acid and pepsin resistant. *Journal of Biological Chemistry* **270**, 23937-23940

148. Cover, T. L., Tummuru, M., Cao, P., Thompson, S. A., and Blaser, M. J. (1994) Divergence of genetic sequences for the vacuolating cytotoxin among Helicobacter pylori strains. *Journal of Biological Chemistry* **269**, 10566-10573

149. Atherton, J. C., and Blaser, M. J. (2009) Coadaptation of Helicobacter pylori and humans: ancient history, modern implications. *The Journal of clinical investigation* **119**, 2475-2487

150. Rhead, J. L., Letley, D. P., Mohammadi, M., Hussein, N., Mohagheghi, M. A., Hosseini, M. E., and Atherton, J. C. (2007) A new Helicobacter pylori vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology* **133**, 926-936

151. Basso, D., Zambon, C. F., Letley, D. P., Stranges, A., Marchet, A., Rhead, J. L., Schiavon, S., Guariso, G., Ceroti, M., Nitti, D., Rugge, M., Plebani, M., and Atherton, J. C. (2008) Clinical relevance of Helicobacter pylori cagA and vacA gene polymorphisms. *Gastroenterology* **135**, 91-99

152. Wittschier, N., Lengsfeld, C., Vorthems, S., Stratmann, U., Ernst, J. F., Verspohl, E. J., and Hensel, A. (2007) Large molecules as anti-adhesive compounds against pathogens. *J Pharm Pharmacol* **59**, 777-786

153. Ermel, G., Georgeault, S., Inisan, C., and Besnard, M. (2012) Inhibition of adhesion of uropathogenic Escherichia coli bacteria to uroepithelial cells by extracts from cranberry. *J Med Food* **15**, 126-134

154. Gottesmann, M., Paraskevopoulou, V., Mohammed, A., Falcone, F. H., and Hensel, A. (2020) BabA and LPS inhibitors against Helicobacter pylori: pectins and pectin-like rhamnogalacturonans as adhesion blockers. *Applied microbiology and biotechnology* **104**, 351-363

155. Lengsfeld, C., Faller, G., and Hensel, A. (2007) Okra polysaccharides inhibit adhesion of Campylobacter jejuni to mucosa isolated from poultry in vitro but not in vivo. *Animal feed science and technology* **135**, 113-125

156. Gbinigie, O. A., Spencer, E. A., Heneghan, C. J., Lee, J. J., and Butler, C. C. (2020) Cranberry Extract for Symptoms of Acute, Uncomplicated Urinary Tract Infection: A Systematic Review. *Antibiotics (Basel)* **10**

157. Hussain, K., Letley, D. P., Greenaway, A. B., Kenefeck, R., Winter, J. A., Tomlinson, W., Rhead, J., Staples, E., Kaneko, K., Atherton, J. C., and Robinson, K. (2016) Helicobacter pylori-mediated protection from allergy is associated with IL-10-secreting peripheral blood regulatory T cells. *Frontiers in immunology* **7**, 71

158. Miller-Podraza, H., Lanne, B., Angstrom, J., Teneberg, S., Milh, M. A., Jovall, P. A., Karlsson, H., and Karlsson, K. A. (2005) Novel binding epitope for Helicobacter pylori found in neolacto carbohydrate chains: structure and cross-binding properties. *The Journal of biological chemistry* **280**, 19695-19703

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

160. Mullis, K. B. (1990) The unusual origin of the polymerase chain reaction. *Scientific American* **262**, 56-65

161. Shampo, M. A., and Kyle, R. A. (2002) Kary B. Mullis—Nobel Laureate for procedure to replicate DNA. in *Mayo Clinic Proceedings*, Elsevier

162. Garibyan, L., and Avashia, N. (2013) Research techniques made simple: polymerase chain reaction (PCR). *The Journal of investigative dermatology* **133**, e6

163. Hoseini, S. S., and Sauer, M. G. (2015) Molecular cloning using polymerase chain reaction, an educational guide for cellular engineering. *Journal of biological engineering* **9**, 2

164. Potapov, V., and Ong, J. L. (2017) Examining sources of error in PCR by single-molecule sequencing. *PloS one* **12**

165. Reha-Krantz, L. J. (2010) DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1804**, 1049-1063

166. Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., and Boyer, H. W. (1977) Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. *Science (New York, N.Y.)* **198**, 1056-1063

167. Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology* **5**, 172

168. Berkowitz, S. A., Engen, J. R., Mazzeo, J. R., and Jones, G. B. (2012) Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *Nature Reviews Drug Discovery* **11**, 527-540

169. Demain, A. L., and Vaishnav, P. (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnology advances* **27**, 297-306

170. Chong, S. (2014) Overview of cell-free protein synthesis: Historic landmarks, commercial systems, and expanding applications. *Current protocols in molecular biology* **108**, 16.30. 11-16.30. 11

171. Gomes, A. R., Byregowda, S. M., Veeregowda, B. M., and Balamurugan, V. (2016) An overview of heterologous expression host systems for the production of recombinant proteins.

172. Grangeasse, C., Stülke, J., and Mijakovic, I. (2015) Regulatory potential of post-translational modifications in bacteria. *Frontiers in microbiology* **6**, 500

173. Sezonov, G., Joseleau-Petit, D., and d'Ari, R. (2007) Escherichia coli physiology in Luria-Bertani broth. *Journal of bacteriology* **189**, 8746-8749

174. Sørensen, H. P., and Mortensen, K. K. (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of biotechnology* **115**, 113-128

175. Adhya, S. (1996) The lac and gal operons today. in *Regulation of gene expression in Escherichia coli*, Springer. pp 181-200

176. Marbach, A., and Bettenbrock, K. (2012) lac operon induction in Escherichia coli: systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *Journal of biotechnology* **157**, 82-88

177. Roderick, S. L. (2005) The lac operon galactoside acetyltransferase. *Comptes rendus biologies* **328**, 568-575

178. Latifi, A. M., Khajeh, K., Farnoosh, G., Hassanpour, K., and Khodi, S. (2015) The Cytoplasmic and periplasmic expression levels and folding of organophosphorus hydrolase enzyme in Escherichia coli. *Jundishapur journal of microbiology* **8**

179. Paraskevopoulou, V., and Falcone, F. H. (2018) Polyionic tags as enhancers of protein solubility in recombinant protein expression. *Microorganisms* **6**, 47

180. Progen E. coli Expression Vector pOPE101.

181. pET-22b(+) Vector Technical Bulletin TB038 (Merck Millipore, Darmstadt, Germany, 1998).

182. Hayashi, K., and Kojima, C. (2008) pCold-GST vector: a novel cold-shock vector containing GST tag for soluble protein production. *Protein expression and purification* **62**, 120-127

183. Qing, G., Ma, L.-C., Khorchid, A., Swapna, G., Mal, T. K., Takayama, M. M., Xia, B., Phadtare, S., Ke, H., Acton, T., Montelione, G. T., Ikura, M., and Inouye, M. (2004) Cold-shock induced high-yield protein production in Escherichia coli. *Nature biotechnology* **22**, 877

184. De Boer, H. A., Comstock, L. J., and Vasser, M. (1983) The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences* **80**, 21-25

185. Duplay, P., Szmelcman, S., Bedouelle, H., and Hofnung, M. (1987) Silent and functional changes in the periplasmic maltose-binding protein of Escherichia coli K12: I. transport of maltose. *Journal of molecular biology* **194**, 663-673

186. Duplay, P., and Hofnung, M. (1988) Two regions of mature periplasmic maltose-binding protein of Escherichia coli involved in secretion. *Journal of bacteriology* **170**, 4445-4450

187. Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L., and McCafferty, J. (2004) Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC biotechnology* **4**, 32

188. Kataeva, I., Chang, J., Xu, H., Luan, C.-H., Zhou, J., Uversky, V. N., Lin, D., Horanyi, P., Liu, Z., Ljungdahl, L. G., Rose, J., Luo, M., and Wang, B.-C. (2005) Improving Solubility of Shewanella Oneidensis MR-1 and Clostridium Thermocellum JW-20 Proteins Expressed Into Esherichia Coli. *Journal of proteome research* **4**, 1942-1951

189. Busso, D., Delagoutte-Busso, B., and Moras, D. (2005) Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in Escherichia coli. *Analytical biochemistry* **343**, 313-321

190. Braud, S., Moutiez, M., Belin, P., Abello, N., Drevet, P., Zinn-Justin, S., Courcon, M., Masson, C., Dassa, J., Charbonnier, J.-B., Boulain, J.-C., Ménez, A., Genet, R., and Gondry, M. (2005) Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. *Journal of proteome research* **4**, 2137-2147

191. Kellermann, O. K., and Ferenci, T. (1982) [75] Maltose-binding protein from Escherichia coli. in *Methods in enzymology*, Elsevier. pp 459-463

192. Gopal, G. J., Kumar, A., Pal, J., and Mukhopadhyay, G. (2014) Molecular characterization and polyclonal antibody generation against core component CagX protein of Helicobacter pylori type IV secretion system. *Bioengineered* **5**, 107-113

193. New England Biolabs E. coli Expression Vector pMAL-c5X.

194. New England Biolabs E. coli DH5α Competent Cells.

195. Stratagene XL10-Gold Ultracompetent Cells.

196. Lee, H., Park, S. J., Han, M.-J., Eom, G. T., Choi, M.-J., Kim, S. H., Oh, Y. H., Song, B. K., and Lee, S. H. (2013) Expression of a lipase on the cell-surface of Escherichia coli using the OmpW anchoring motif and its application to enantioselective reactions. *Biotechnology letters* **35**, 1677-1683

197. Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology* **189**, 113-130

198. Novagen. (2004) Competent Cells User Protocol TB009 Rev. F0104. 1-23

199. Minea, R. O., Swenson, S. D., and Markland Jr, F. S. (2014) Method of expressing proteins with disulfide bridges with enhanced yields and activity. Google Patents

200. Schein, C. H. (2004) A cool way to make proteins. *Nature biotechnology* **22**, 826-827

201. ArcticExpress Competent Cells and ArcticExpress (DE3) Competent Cells (Agilent Technologies, USA).

202. Brinkmann, U., Mattes, R. E., and Buckel, P. (1989) High-level expression of recombinant genes in Escherichia coli is dependent on the availability of the dnaY gene product. *Gene* **85**, 109-114

203. Seidel, H. M., Pompliano, D. L., and Knowles, J. R. (1992) Phosphonate biosynthesis: molecular cloning of the gene for phosphoenolpyruvate mutase from Tetrahymena pyriformis and overexpression of the gene product in Escherichia coli. *Biochemistry* **31**, 2598-2608

204. Del Tito, B., Ward, J. M., Hodgson, J., Gershater, C., Edwards, H., Wysocki, L. A., Watson, F. A., Sathe, G., and Kane, J. F. (1995) Effects of a minor isoleucyl tRNA on heterologous protein translation in Escherichia coli. *Journal of bacteriology* **177**, 7086-7091

205. Novinec, M., Pavšič, M., and Lenarčič, B. (2012) A simple and efficient protocol for the production of recombinant cathepsin V and other cysteine cathepsins in soluble form in Escherichia coli. *Protein expression and purification* **82**, 1-5

206. Gan, S. D., and Patel, K. R. (2013) Enzyme immunoassay and enzyme-linked immunosorbent assay. *J Invest Dermatol* **133**, e12

207. Engvall, E., and Perlmann, P. (1971) Enzyme-linked immunosorbent assay (ELISA). *Protides of the biological fluids*, 553-556

208. Van Weemen, B., and Schuurs, A. (1971) Immunoassay using antigen—enzyme conjugates. *FEBS letters* **15**, 232-236

209. Reen, D. J. (1994) Enzyme-linked immunosorbent assay (ELISA). in *Basic Protein and Peptide Protocols*, Springer. pp 461-466

210. Shah, K., and Maghsoudlou, P. (2016) Enzyme-linked immunosorbent assay (ELISA): the basics. *British Journal of Hospital Medicine* **77**, C98-C101

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

212. Maestre-Reyna, M. (2011) *Structural and functional studies of mucin-interacting adhesion domains from Candida glabrata and Helicobacter pylori.* PhD, Philipps-Universität Marburg

213. Harper, S., and Speicher, D. W. (2011) Purification of proteins fused to glutathione S-transferase. in *Protein Chromatography*, Springer. pp 259-280

214. De Marco, A. (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microbial cell factories* **8**, 26

215. Papaneophytou, C. P., and Kontopidis, G. (2014) Statistical approaches to maximize recombinant protein expression in Escherichia coli: a general review. *Protein expression and purification* **94**, 22-32

216. Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T. (2003) Practical considerations in refolding proteins from inclusion bodies. *Protein expression and purification* **28**, 1-8

217. Hand, K., Wilkinson, M. C., and Madine, J. (2018) Isolation and purification of recombinant immunoglobulin light chain variable domains from the periplasmic space of Escherichia coli. *PloS one* **13**

218. Alm, R. A., Bina, J., Andrews, B. M., Doig, P., Hancock, R. E., and Trust, T. J. (2000) Comparative genomics of Helicobacter pylori: analysis of the outer membrane protein families. *Infection and immunity* **68**, 4155-4168

219. Voss, B. J., Gaddy, J. A., McDonald, W. H., and Cover, T. L. (2014) Analysis of surface-exposed outer membrane proteins in Helicobacter pylori. *Journal of bacteriology* **196**, 2455-2471

220. Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) *Biochemistry, Fifth Edition: International Version*, 5th ed., W. H. Freeman, New York

221. Bohak, Z. (1969) Purification and characterization of chicken pepsinogen and chicken pepsin. *Journal of Biological Chemistry* **244**, 4638-4648

222. Ryle, A. (1970) The porcine pepsins and pepsinogens. in *Methods in Enzymology*, Academic Press, New York. pp 316-336

223. Reddy, I. M., Kella, N. K., and Kinsella, J. E. (1988) Structural and conformational basis of the resistance of beta-lactoglobulin to peptic and chymotryptic digestion. *Journal of Agricultural and Food Chemistry* **36**, 737-741

224. Bahari, H., Ross, I., and Turnberg, L. (1982) Demonstration of a pH gradient across the mucus layer on the surface of human gastric mucosa in vitro. *Gut* **23**, 513-516

225. Allen, A., and Flemström, G. (2005) Gastroduodenal mucus bicarbonate barrier: protection against acid and pepsin. *American Journal of Physiology-Cell Physiology* **288**, C1-C19

226. McColl, K. E. (2012) The elegance of the gastric mucosal barrier: designed by nature for nature. *Gut* **61**, 787-788

227. Ho, S. B., Roberton, A. M., Shekels, L. L., Lyftogt, C. T., Niehans, G. A., and Toribara, N. W. (1995) Expression cloning of gastric mucin complementary DNA and localization of mucin gene expression. *Gastroenterology* **109**, 735-747

228. McGuckin, M. A., Lindén, S. K., Sutton, P., and Florin, T. H. (2011) Mucin dynamics and enteric pathogens. *Nature Reviews Microbiology* **9**, 265

229. Atuma, C., Strugala, V., Allen, A., and Holm, L. (2001) The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **280**, G922-G929

230. Johansson, M. E., Sjövall, H., and Hansson, G. C. (2013) The gastrointestinal mucus system in health and disease. *Nature reviews Gastroenterology & hepatology* **10**, 352

231. Nordman, H., Davies, J. R., Lindell, G., De Bolos, C., Francisco, R., and Carlstedt, I. (2002) Gastric MUC5AC and MUC6 are large oligomeric mucins that differ in size, glycosylation and tissue distribution. *Biochemical Journal* **364**, 191-200

232. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845

233. Noble, J. E. (2014) Quantification of protein concentration using UV absorbance and Coomassie dyes. in *Methods in enzymology*, Elsevier. pp 17-26

234. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. in *The proteomics protocols handbook* (Walker, J. M. ed.), Humana Press, Totowa. pp 571-607

235. Keil, B. (2012) *Specificity of proteolysis*, Springer Science & Business Media

236. Qin, B. Y., Bewley, M. C., Creamer, L. K., Baker, H. M., Baker, E. N., and Jameson, G. B. (1998) Structural basis of the Tanford transition of bovine β-lactoglobulin. *Biochemistry* **37**, 14014-14023

237. US_Pharmacopoeia, and The_National_Formulary. (2000) USP XXIV, NF XIX. US Pharmacopoeia Convention, Easton, PA

238. Thomas, K., Aalbers, M., Bannon, G., Bartels, M., Dearman, R., Esdaile, D., Fu, T., Glatt, C., Hadfield, N., Hatzos, C., Hefle, S. L., Heylings, J. R., Goodman, R. E., Henry, B., Herouet, C., Holsapple, M., Ladics, G. S., Landry, T. D., MacIntosh, S. C., Rice, E. A., Privalle, L. S., Steiner, H. Y., Teshima, R., Van Ree, R., Woolhiser, M., and Zawodny, J. (2004) A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins. *Regulatory Toxicology and Pharmacology* **39**, 87-98

239. Wang, R., Edrington, T. C., Storrs, S. B., Crowley, K. S., Ward, J. M., Lee, T. C., Liu, Z. L., Li, B., and Glenn, K. C. (2017) Analyzing pepsin degradation assay conditions used for allergenicity assessments to ensure that pepsin susceptible and pepsin resistant dietary proteins are distinguishable. *PloS one* **12**, e0171926

240. Buchanan, S. K., Smith, B. S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., Van Der Helm, D., and Deisenhofer, J. (1999) Crystal structure of the outer membrane active transporter FepA from Escherichia coli. *Nature Structural & Molecular Biology* **6**, 56

241. Psakis, G., Nitschkowski, S., Holz, C., Kreß, D., Maestre-Reyna, M., Polaczek, J., Illing, G., and Essen, L. O. (2007) Expression screening of integral membrane proteins from Helicobacter pylori 26695. *Protein Science* **16**, 2667-2676

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

243. Hennig, E. E., Mernaugh, R., Edl, J., Cao, P., and Cover, T. L. (2004) Heterogeneity among Helicobacter pylori strains in expression of the outer membrane protein BabA. *Infection and immunity* **72**, 3429-3435

244. Subedi, S., Moonens, K., Romao, E., Lo, A., Vandenbussche, G., Bugaytsova, J., Muyldermans, S., Borén, T., and Remaut, H. (2014) Expression, purification and X-ray crystallographic analysis of the Helicobacter pylori blood group antigen-binding adhesin BabA. *Acta Crystallographica Section F: Structural Biology Communications* **70**, 1631-1635

245. Oleastro, M., and Menard, A. (2013) The Role of Helicobacter pylori Outer Membrane Proteins in Adherence and Pathogenesis. *Biology* **2**, 1110-1134

246. Shiota, S., Suzuki, R., and Yamaoka, Y. (2013) The significance of virulence factors in Helicobacter pylori. *Journal of digestive diseases* **14**, 341-349

247. Odenbreit, S., Swoboda, K., Barwig, I., Ruhl, S., Boren, T., Koletzko, S., and Haas, R. (2009) Outer membrane protein expression profile in Helicobacter pylori clinical isolates. *Infection and immunity* **77**, 3782-3790

248. Lai, R., Bekessy, A., Chen, C., Walsh, T., and Barnard, R. (2003) Megaprimer mutagenesis using very long primers. *Biotechniques* **34**, 52-56

249. Colbeck, J. C., Hansen, L. M., Fong, J. M., and Solnick, J. V. (2006) Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of Helicobacter pylori. *Infection and immunity* **74**, 4375-4378

250. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59

251. Stratagene XL10-Gold Ultracompetent Cells.

252. Zhou, Y., Zhou, Y., Li, J., Chen, J., Yao, Y., Yu, L., Peng, D., Wang, M., Su, D., He, Y., and Gou, L. (2015) Efficient expression, purification and

characterization of native human cystatin C in Escherichia coli periplasm. *Protein expression and purification* **111**, 18-22

253. Wickner, W., Driessen, A. J., and Haril, F.-U. (1991) The enzymology of protein translocation across the Escherichia coli plasma membrane. *Annual review of biochemistry* **60**, 101-124

254. Schmiedl, A., Breitling, F., Winter, C. H., Queitsch, I., and Dübel, S. (2000) Effects of unpaired cysteines on yield, solubility and activity of different recombinant antibody constructs expressed in E. coli. *Journal of immunological methods* **242**, 101-114

255. Casali, N., and Preston, A. (2003) *E. coli plasmid vectors: methods and applications*, Humana Press, Totowa, New Jersey

256. Ruiz, N., Chng, S.-S., Hiniker, A., Kahne, D., and Silhavy, T. J. (2010) Nonconsecutive disulfide bond formation in an essential integral outer membrane protein. *Proceedings of the National Academy of Sciences* **107**, 12245-12250

257. Bianchi, E., Venturini, S., Pessi, A., Tramontano, A., and Sollazzo, M. (1994) High level expression and rational mutagenesis of a designed protein, the minibody: from an insoluble to a soluble molecule. *Journal of molecular biology* **236**, 649-659

258. Park, S. H., Mrse, A. A., Nevzorov, A. A., Mesleh, M. F., Oblatt-Montal, M., Montal, M., and Opella, S. J. (2003) Three-dimensional structure of the channel-forming trans-membrane domain of virus protein “u”(Vpu) from HIV-1. *Journal of molecular biology* **333**, 409-424

259. Kato, A., Maki, K., Ebina, T., Kuwajima, K., Soda, K., and Kuroda, Y. (2007) Mutational analysis of protein solubility enhancement using short peptide tags. *Biopolymers: Original Research on Biomolecules* **85**, 12-18

260. Nautiyal, K., and Kuroda, Y. (2018) A SEP tag enhances the expression, solubility and yield of recombinant TEV protease without altering its activity. *New biotechnology* **42**, 77-84

261. Yoon, S. H., Kim, S. K., and Kim, J. F. (2010) Secretory production of recombinant proteins in Escherichia coli. *Recent patents on biotechnology* **4**, 23-29

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

263. Tailford, L. E., Crost, E. H., Kavanaugh, D., and Juge, N. (2015) Mucin glycan foraging in the human gut microbiome. *Frontiers in genetics* **6**, 81

264. Pride, D. T., and Blaser, M. J. (2002) Concerted evolution between duplicated genetic elements in Helicobacter pylori. *Journal of molecular biology* **316**, 629-642

265. Bäckström, A., Lundberg, C., Kersulyte, D., Berg, D. E., Borén, T., and Arnqvist, A. (2004) Metastability of Helicobacter pylori bab adhesin genes and dynamics in Lewis b antigen binding. *Proceedings of the National Academy of Sciences* **101**, 16923-16928

266. Matteo, M. J., Armitano, R. I., Romeo, M., Wonaga, A., Olmos, M., and Catalano, M. (2011) Helicobacter pylori bab genes during chronic colonization. *International journal of molecular epidemiology and genetics* **2**, 286

267. Shiota, S., Suzuki, R., and Yamaoka, Y. (2013) The significance of virulence factors in H elicobacter pylori. *Journal of digestive diseases* **14**, 341-349

268. Yamaoka, Y. (2008) Roles of Helicobacter pylori BabA in gastroduodenal pathogenesis. *World journal of gastroenterology: WJG* **14**, 4265

269. Ansari, S., and Yamaoka, Y. (2017) Helicobacter pylori BabA in adaptation for gastric colonization. *World Journal of Gastroenterology* **23**, 4158

270. Kim, A., Servetas, S. L., Kang, J., Kim, J., Jang, S., Cha, H. J., Lee, W. J., Kim, J., Romero-Gallo, J., Peek, R. M., Jr., Merrell, D. S., and Cha, J. H. (2015) Helicobacter pylori bab Paralog Distribution and Association with cagA, vacA, and homA/B Genotypes in American and South Korean Clinical Isolates. *PLoS One* **10**, e0137078

271. Mudie, D. M., Amidon, G. L., and Amidon, G. E. (2010) Physiological parameters for oral delivery and in vitro testing. *Molecular pharmaceutics* **7**, 1388-1405

272. US_Pharmacopoeia, and The_National_Formulary. (1995) USP XXIII, NF XVIII. US Pharmacopoeia Convention, Easton, PA

273. de Bernard, M., Papini, E., de Filippis, V., Gottardi, E., Telford, J., Manetti, R., Fontana, A., Rappuoli, R., and Montecucco, C. (1995) Low pH activates the vacuolating toxin of Helicobacter pylori, which becomes acid and pepsin resistant. *The Journal of biological chemistry* **270**, 23937-23940

274. Schreiber, S., Bücker, R., Groll, C., Azevedo-Vethacke, M., Garten, D., Scheid, P., Friedrich, S., Gatermann, S., Josenhans, C., and Suerbaum, S. (2005) Rapid loss of motility of Helicobacter pylori in the gastric lumen in vivo. *Infection and immunity* **73**, 1584-1589

275. Schreiber, S., Bücker, R., Groll, C., Azevedo-Vethacke, M., Scheid, P., Gatermann, S., Josenhans, C., and Suerbaum, S. (2006) Gastric antibacterial efficiency is different for pepsin A and C. *Archives of microbiology* **184**, 335-340

276. Bücker, R., Azevedo-Vethacke, M., Groll, C., Garten, D., Josenhans, C., Suerbaum, S., and Schreiber, S. (2012) Helicobacter pylori colonization critically depends on postprandial gastric conditions. *Scientific Reports* **2**, 994

277. Huang, J. Y., Sweeney, E. G., Sigal, M., Zhang, H. C., Remington, S. J., Cantrell, M. A., Kuo, C. J., Guillemin, K., and Amieva, M. R. (2015) Chemodetection and destruction of host urea allows Helicobacter pylori to locate the epithelium. *Cell host & microbe* **18**, 147-156

278. Krishnamurthy, P., Parlow, M., Zitzer, J. B., Vakil, N. B., Mobley, H. L., Levy, M., Phadnis, S. H., and Dunn, B. E. (1998) Helicobacter pylori containing only cytoplasmic urease is susceptible to acid. *Infection and immunity* **66**, 5060-5066

279. Russell, D., Oldham, N. J., and Davis, B. G. (2009) Site-selective chemical protein glycosylation protects from autolysis and proteolytic degradation. *Carbohydrate research* **344**, 1508-1514

280. The_National_Centre_for_Biotechnology_Information. (2019) Membrane Protein [Helicobacter pylori]: NCBI Reference Sequence: WP_046655427.1. https://www.ncbi.nlm.nih.gov/protein/WP_046655427.1?report=graph

281. Petersen, T. N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**, 785

282. Nielsen, H. (2017) Predicting secretory proteins with SignalP. in *Protein function prediction: methods and protocols* (Kihara, D. ed.), Humana Press, New York. pp 59-73

283. Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423

284. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *Journal of molecular biology* **340**, 783-795

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

286. Stewart, E. J., Åslund, F., and Beckwith, J. (1998) Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins. *The EMBO journal* **17**, 5543-5550

287. Bessette, P. H., Åslund, F., Beckwith, J., and Georgiou, G. (1999) Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proceedings of the National Academy of Sciences* **96**, 13703-13708

288. Ritz, D., Lim, J., Reynolds, C. M., Poole, L. B., and Beckwith, J. (2001) Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. *Science (New York, N.Y.)* **294**, 158-160

289. Liu, D., Schmid, R., and Rusnak, M. (2006) Functional expression of Candida antarctica lipase B in the Escherichia coli cytoplasm—a screening system for a frequently used biocatalyst. *Applied microbiology and biotechnology* **72**, 1024-1032

290. Derman, A. I., Prinz, W. A., Belin, D., and Beckwith, J. (1993) Mutations that allow disulfide bond formation in the cytoplasm of Escherichia coli. *Science (New York, N.Y.)* **262**, 1744-1747

291. University-of-California-San-Diego. (2019) Escherichia coli Strains and NIH Guidelines.

292. Dautin, N., and Bernstein, H. D. (2007) Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu. Rev. Microbiol.* **61**, 89-112

293. Sørensen, H. P., and Mortensen, K. K. (2005) Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial cell factories* **4**, 1

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

295. Johnson, E. C., Malito, E., Shen, Y., Rich, D., Tang, W.-J., and Kent, S. B. (2007) Modular total chemical synthesis of a human immunodeficiency virus type 1 protease. *Journal of the American Chemical Society* **129**, 11480-11490

296. Lilie, H., Richter, S., Bergelt, S., Frost, S., and Gehle, F. (2013) Polyionic and cysteine-containing fusion peptides as versatile protein tags. *Biological chemistry* **394**, 995-1004

297. Richter, S. A., Stubenrauch, K., Lilie, H., and Rudolph, R. (2001) Polyionic fusion peptides function as specific dimerization motifs. *Protein engineering* **14**, 775-783

298. Kleinschmidt, M., Rudolph, R., and Lilie, H. (2003) Design of a modular immunotoxin connected by polyionic adapter peptides. *Journal of molecular biology* **327**, 445-452

299. Bergelt, S., Frost, S., and Lilie, H. (2009) Listeriolysin O as cytotoxic component of an immunotoxin. *Protein Science* **18**, 1210-1220

300. Bucher, M. H., Evdokimov, A. G., and Waugh, D. S. (2002) Differential effects of short affinity tags on the crystallization of Pyrococcus furiosus maltodextrin-binding protein. *Acta Crystallographica Section D: Biological Crystallography* **58**, 392-397

301. Studier, F. W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of molecular biology* **219**, 37-44

302. Zhang, X., and Studier, F. W. (1997) Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. *Journal of molecular biology* **269**, 10-27

303. Schein, C. H., and Noteborn, M. H. (1988) Formation of soluble recombinant proteins in Escherichia coli is favored by lower growth temperature. *Bio/technology* **6**, 291

304. Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* **67**, 31-40

305. di Guana, C., Lib, P., Riggsa, P. D., and Inouyeb, H. (1988) Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* **67**, 21-30

306. Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T., and Ishii, S. (1995) Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin. *Journal of Biological Chemistry* **270**, 25328-25331

307. Young, C. L., Britton, Z. T., and Robinson, A. S. (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnology journal* **7**, 620-634

308. Schein, C. H. (1990) Solubility as a function of protein structure and solvent components. *Bio/technology* **8**, 308

309. Shirano, Y., and Shibata, D. (1990) Low temperature cultivation of Escherichia coli carrying a rice lipoxygenase L-2 cDNA produces a soluble and active enzyme at a high level. *FEBS letters* **271**, 128-130

310. Takara Bio Inc pCold™ GST DNA.

311. Goldenberg, D., Azar, I., Oppenheim, A., Brandi, A., Pon, C., and Gualerzi, C. (1997) Role of Escherichia coli cspA promoter sequences and adaptation of translational apparatus in the cold shock response. *Molecular and General Genetics MGG* **256**, 282-290

312. Etchegaray, J.-P., and Inouye, M. (1999) Translational enhancement by an element downstream of the initiation codon in Escherichia coli. *Journal of Biological Chemistry* **274**, 10079-10085

313. Kaplan, W., Erhardt, J., Sluis-Cremer, N., Dirr, H., Hüsler, P., and Klump, H. (1997) Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag. *Protein Science* **6**, 399-406

314. Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M., and Colonno, R. J. (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *Journal of Biological Chemistry* **265**, 9062-9065

315. Leong, L. E. (1999) The use of recombinant fusion proteases in the affinity purification of recombinant proteins. *Molecular biotechnology* **12**, 269-274

316. Kong, B., and Guo, G. L. (2014) Soluble expression of disulfide bond containing proteins FGF15 and FGF19 in the cytoplasm of Escherichia coli. *PLoS One* **9**, e85890

317. Vincentelli, R., Cimino, A., Geerlof, A., Kubo, A., Satou, Y., and Cambillau, C. (2011) High-throughput protein expression screening and purification in Escherichia coli. *Methods* **55**, 65-72

318. Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N., and Timmis, K. N. (2003) Chaperonins govern growth of Escherichia coli at low temperatures. *Nature biotechnology* **21**, 1266

319. Sachdev, D., and Chirgwin, J. M. (1999) Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein. *Journal of protein chemistry* **18**, 127-136

320. Kapust, R. B., and Waugh, D. S. (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science* **8**, 1668-1674

321. Sachdev, D., and Chirgwin, J. M. (2000) [20] Fusions to maltose-binding protein: Control of folding and solubility in protein purification. in *Methods in enzymology*, Elsevier. pp 312-321

322. Riggs, P. (2000) Expression and purification of recombinant proteins by fusion to maltose-binding protein. *Molecular biotechnology* **15**, 51-63

323. Tropea, J. E., Cherry, S., Nallamsetty, S., Bignon, C., and Waugh, D. S. (2007) A generic method for the production of recombinant proteins in Escherichia coli using a dual hexahistidine-maltose-binding protein affinity tag. in *Macromolecular Crystallography Protocols: Volume 1, Preparation and Crystallization of Macromolecules* (Doublié, S. ed.), Humana Press, Totowa, New Jersey. pp 1-19

324. Lebendiker, M., and Danieli, T. (2017) Purification of proteins fused to maltose-binding protein. in *Protein Chromatography* (Walls, D., and Loughran, S. T. eds.), Humana Press, New York. pp 257-273

325. Arya, R., Sabir, J., Bora, R. S., and Saini, K. S. (2015) Optimization of Culture Parameters and Novel Strategies to Improve Protein Solubility. in *Insoluble Proteins: Methods and Protocols* (García-Fruitós, E. ed.), Humana Press, New York. pp 45-63

326. Hooper, N. M. (1994) Families of zinc metalloproteases. *FEBS letters* **354**, 1-6

327. Auld, D. S. (2013) Catalytic mechanisms for metallopeptidases. in *Handbook of proteolytic enzymes* (Rawlings, N., and Salvesen, G. eds.), Academic Press, Oxford. pp 370–396

328. Cerdà-Costa, N., and Xavier Gomis-Rüth, F. (2014) Architecture and function of metallopeptidase catalytic domains. *Protein Science* **23**, 123-144

329. Dollery, C. M., and Libby, P. (2006) Atherosclerosis and proteinase activation. *Cardiovascular research* **69**, 625-635

330. Murphy, G., and Nagase, H. (2008) Reappraising metalloproteinases in rheumatoid arthritis and osteoarthritis: destruction or repair? *Nature clinical practice Rheumatology* **4**, 128-135

331. Nalivaeva, N. N., Fisk, L. R., Belyaev, N. D., and Turner, A. J. (2008) Amyloid-degrading enzymes as therapeutic targets in Alzheimer's disease. *Current Alzheimer Research* **5**, 212-224

332. Chang, C., and Werb, Z. (2001) The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis. *Trends in cell biology* **11**, S37-S43

333. Matrisian, L., and Lopez-Otin, C. (2007) Emerging roles of proteases in tumour suppression. *Nature Reviews Cancer* **7**, 800-808

334. Bode, W., Gomis-Rüth, F.-X., and Stöckler, W. (1993) Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the ‘metzincins’. *FEBS letters* **331**, 134-140

335. Jongeneel, C. V., Bouvier, J., and Bairoch, A. (1989) A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS letters* **242**, 211-214

336. Pannifer, A. D., Wong, T. Y., Schwarzenbacher, R., Renatus, M., Petosa, C., Bienkowska, J., Lacy, D. B., Collier, R. J., Park, S., Leppla, S. H., Hanna, P., and Liddington, R. C. (2001) Crystal structure of the anthrax lethal factor. *Nature* **414**, 229-233

337. Matthews, B. W. (1988) Structural basis of the action of thermolysin and related zinc peptidases. *Accounts of Chemical Research* **21**, 333-340

338. Fushimi, N., Nakajima, T., and Ichishima, E. (1999) Aspzincin, a family of metalloendopeptidases with a new zinc-binding motif identification of new zinc-binding sites (his128, his132, and asp164) and three catalytically crucial residues (glu129, asp143, and tyr106) of deuterolysin fromaspergillus oryzae by site-directed mutagenesis. *Journal of biological chemistry* **274**, 24195-24201

339. Gomis-Rüth, F. X., Botelho, T. O., and Bode, W. (2012) A standard orientation for metallopeptidases. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1824**, 157-163

340. Feng, L., Yan, H., Wu, Z., Yan, N., Wang, Z., Jeffrey, P. D., and Shi, Y. (2007) Structure of a site-2 protease family intramembrane metalloprotease. *Science (New York, N.Y.)* **318**, 1608-1612

341. Bieniossek, C., Schalch, T., Bumann, M., Meister, M., Meier, R., and Baumann, U. (2006) The molecular architecture of the metalloprotease FtsH. *Proceedings of the National Academy of Sciences* **103**, 3066-3071

342. Langklotz, S., Baumann, U., and Narberhaus, F. (2012) Structure and function of the bacterial AAA protease FtsH. *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* **1823**, 40-48

343. Stöcker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Rüth, F. X., McKay, D. B., and Bode, W. (1995) The metzincins--topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein science : a publication of the Protein Society* **4**, 823-840

344. Gomis-Rüth, F. X. (2003) Structural aspects of the metzincin clan of metalloendopeptidases. *Molecular biotechnology* **24**, 157-202

345. Gomis-Rüth, F. X. (2009) Catalytic domain architecture of metzincin metalloproteases. *Journal of biological chemistry* **284**, 15353-15357

346. Hamada, K., Hata, Y., Katsuya, Y., Hiramatsu, H., Fujiwara, T., and Katsube, Y. (1996) Crystal Structure of Serratia Protease, a Zinc-Dependent Proteinase from Serratia sp. E-15, Containing a β-Sheet Coil Motif at 2.0 AÅ Resolution. *The Journal of Biochemistry* **119**, 844-851

347. Schlagenhauf, E., Etges, R., and Metcalf, P. (1998) The crystal structure of the Leishmania major surface proteinase leishmanolysin (gp63). *Structure (London, England : 1993)* **6**, 1035-1046

348. Tallant, C., García-Castellanos, R., Seco, J., Baumann, U., and Gomis-Rüth, F. X. (2006) Molecular analysis of ulilysin, the structural prototype of a new family of metzincin metalloproteases. *Journal of biological chemistry* **281**, 17920-17928

349. Gomis-Rüth, F.-X., Kress, L. F., and Bode, W. (1993) First structure of a snake venom metalloproteinase: a prototype for matrix metalloproteinases/collagenases. *The EMBO journal* **12**, 4151-4157

350. Bode, W., Gomis-Rüth, F. X., Huber, R., Zwilling, R., and Stöcker, W. (1992) Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* **358**, 164-167

351. Kurisu, G., Kinoshita, T., Sugimoto, A., Nagara, A., Kai, Y., Kasai, N., and Harada, S. (1997) Structure of the zinc endoprotease from Streptomyces caespitosus. *The Journal of Biochemistry* **121**, 304-308

352. Tallant, C., Marrero, A., and Gomis-Rüth, F. X. (2010) Matrix metalloproteinases: fold and function of their catalytic domains. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1803**, 20-28

353. Yu, A. C., Worrall, L. J., and Strynadka, N. C. (2012) Structural insight into the bacterial mucinase StcE essential to adhesion and immune evasion during enterohemorrhagic E. coli infection. *Structure (London, England : 1993)* **20**, 707-717

354. Waltersperger, S., Widmer, C., Wang, M., and Baumann, U. (2010) Crystal structure of archaemetzincin AmzA from Methanopyrus kandleri at 1.5 Å resolution. *Proteins: Structure, Function, and Bioinformatics* **78**, 2720-2723

355. Goulas, T., Arolas, J. L., and Gomis-Rüth, F. X. (2011) Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. *Proceedings of the National Academy of Sciences* **108**, 1856-1861

356. Ng, N. M., Littler, D. R., Paton, A. W., Le Nours, J., Rossjohn, J., Paton, J. C., and Beddoe, T. (2013) EcxAB is a founding member of a new family of metalloprotease AB5 toxins with a hybrid cholera-like B subunit. *Structure (London, England : 1993)* **21**, 2003-2013

357. Arolas, J. L., Goulas, T., Pomerantsev, A. P., Leppla, S. H., and Gomis-Rüth, F. X. (2016) Structural basis for latency and function of immune inhibitor A metallopeptidase, a modulator of the Bacillus anthracis secretome. *Structure (London, England : 1993)* **24**, 25-36

358. Balaban, N., Rudakova, N., and Sharipova, M. (2012) Structural and functional characteristics and properties of metzincins. *Biochemistry (Moscow)* **77**, 119-127

359. Tallant, C., García-Castellanos, R., Baumann, U., and Gomis-Rüth, F. X. (2010) On the relevance of the Met-turn methionine in metzincins. *Journal of biological chemistry* **285**, 13951-13957

360. Macfarlane, S. R., Seatter, M. J., Kanke, T., Hunter, G. D., and Plevin, R. (2001) Proteinase-activated receptors. *Pharmacological reviews* **53**, 245-282

361. Ossovskaya, V. S., and Bunnett, N. W. (2004) Protease-activated receptors: contribution to physiology and disease. *Physiological reviews* **84**, 579-621

362. Reed, C. E., and Kita, H. (2004) The role of protease activation of inflammation in allergic respiratory diseases. *Journal of Allergy and Clinical Immunology* **114**, 997-1008

363. Uehara, A., Muramoto, K., Takada, H., and Sugawara, S. (2003) Neutrophil serine proteinases activate human nonepithelial cells to produce inflammatory cytokines through protease-activated receptor 2. *The Journal of Immunology* **170**, 5690-5696

364. Kida, Y., Inoue, H., Shimizu, T., and Kuwano, K. (2007) Serratia marcescens serralysin induces inflammatory responses through protease-activated receptor 2. *Infection and immunity* **75**, 164-174

365. Hollenberg, M. D., and Compton, S. J. (2002) International union of pharmacology. XXVIII. Proteinase-activated receptors. *Pharmacological reviews* **54**, 203-217

366. Nystedt, S., Ramakrishnan, V., and Sundelin, J. (1996) The proteinase-activated receptor 2 is induced by inflammatory mediators in human endothelial cells comparison with the thrombin receptor. *Journal of Biological Chemistry* **271**, 14910-14915

367. Cenac, N., Coelho, A.-M., Nguyen, C., Compton, S., Andrade-Gordon, P., MacNaughton, W. K., Wallace, J. L., Hollenberg, M. D., Bunnett, N. W., Garcia-Villar, R., Bueno, L., and Vergnolle, N. (2002) Induction of intestinal inflammation in mouse by activation of proteinase-activated receptor-2. *The American journal of pathology* **161**, 1903-1915

368. Ferrell, W. R., Lockhart, J. C., Kelso, E. B., Dunning, L., Plevin, R., Meek, S. E., Smith, A. J., Hunter, G. D., McLean, J. S., McGarry, F., Ramage, R., Jiang, L., Kanke, T., and Kawagoe, J. (2003) Essential role for proteinase-activated receptor-2 in arthritis. *The Journal of clinical investigation* **111**, 35-41

369. Hansen, K. K., Sherman, P. M., Cellars, L., Andrade-Gordon, P., Pan, Z., Baruch, A., Wallace, J. L., Hollenberg, M. D., and Vergnolle, N. (2005) A major role for proteolytic activity and proteinase-activated receptor-2 in the pathogenesis of infectious colitis. *Proceedings of the National Academy of Sciences* **102**, 8363-8368

370. Kawagoe, J., Takizawa, T., Matsumoto, J., Tamiya, M., Meek, S. E., Smith, A. J., Hunter, G. D., Plevin, R., Saito, N., Kanke, T., Fujii, M., and Wada, Y. (2002) Effect of protease-activated receptor-2 deficiency on

allergic dermatitis in the mouse ear. *Japanese journal of pharmacology* **88**, 77-84

371. Lindner, J. R., Kahn, M. L., Coughlin, S. R., Sambrano, G. R., Schauble, E., Bernstein, D., Foy, D., Hafezi-Moghadam, A., and Ley, K. (2000) Delayed onset of inflammation in protease-activated receptor-2-deficient mice. *The Journal of Immunology* **165**, 6504-6510

372. Schmidlin, F., Amadesi, S., Dabbagh, K., Lewis, D. E., Knott, P., Bunnett, N. W., Gater, P. R., Geppetti, P., Bertrand, C., and Stevens, M. E. (2002) Protease-activated receptor 2 mediates eosinophil infiltration and hyperreactivity in allergic inflammation of the airway. *The Journal of Immunology* **169**, 5315-5321

373. Seeliger, S., Derian, C. K., Vergnolle, N., Bunnett, N. W., Nawroth, R., Schmelz, M., VON DER WEID, P.-Y., BUDDENKOTTE, J. R., SUNDERKÖTTER, C., Metze, D., Andrade-Gordon, P., Harms, E., Vestweber, D., Luger, T. A., and Steinhoff, M. (2003) Proinflammatory role of proteinase-activated receptor-2 in humans and mice during cutaneous inflammation in vivo. *The FASEB Journal* **17**, 1871-1885

374. Isozaki, Y., Yoshida, N., Kuroda, M., Handa, O., Takagi, T., Kokura, S., Ichikawa, H., Naito, Y., Okanoue, T., and Yoshikawa, T. (2006) Anti-tryptase treatment using nafamostat mesilate has a therapeutic effect on experimental colitis. *Scand J Gastroenterol* **41**, 944-953

375. Yoshida, N., Isozaki, Y., Takagi, T., Takenaka, S., Uchikawa, R., Arizono, N., Yoshikawa, T., and Okanoue, T. (2006) anti-tryptase therapy in inflammatory bowel disease. in *Alimentary Pharmacology & Therapeutics Symposium Series*, Wiley Online Library

376. Gu, Q., and LEE, L.-Y. (2012) Protease-Activated Receptor 2. in *Therapeutic Targets: Modulation, Inhibition, and Activation* (Botana, L., and Loza, M. eds.), Wiley, Hoboken, New Jersey. pp 37-62

377. Asokananthan, N., Graham, P. T., Fink, J., Knight, D. A., Bakker, A. J., McWilliam, A. S., Thompson, P. J., and Stewart, G. A. (2002) Activation of protease-activated receptor (PAR)-1, PAR-2, and PAR-4 stimulates IL-6, IL-8, and prostaglandin E2 release from human respiratory epithelial cells. *The Journal of Immunology* **168**, 3577-3585

378. Johansson, U., Lawson, C., Dabare, M., Syndercombe-Court, D., Newland, A. C., Howells, G. L., and Macey, M. G. (2005) Human peripheral blood monocytes express protease receptor-2 and respond to receptor activation by production of IL-6, IL-8, and IL-1β. *Journal of leukocyte biology* **78**, 967-975

379. Kajikawa, H., Yoshida, N., Katada, K., Hirayama, F., Handa, O., Kokura, S., Naito, Y., and Yoshikawa, T. (2007) Helicobacter pylori activates gastric epithelial cells to produce interleukin-8 via protease-activated receptor 2. *Digestion* **76**, 248-255

380. Kawabata, A. (2003) Gastrointestinal functions of proteinase-activated receptors. *Life sciences* **74**, 247-254

381. Thermo Scientific Pierce Fluorescent Protease Assay Kit.

382. Barber, R. D., Harmer, D. W., Coleman, R. A., and Clark, B. J. (2005) GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiological genomics* **21**, 389-395

383. Rudakova, N., Balaban, N., Danilova, Y. V., Rudenskaya, G., and Sharipova, M. (2010) Characteristics of a novel secreted zinc-dependent endopeptidase of Bacillus intermedius. *Biochemistry (Moscow)* **75**, 1294-1301

384. Walasek, P., and Honek, J. F. (2005) Nonnatural amino acid incorporation into the methionine 214 position of the metzincin Pseudomonas aeruginosa alkaline protease. *BMC biochemistry* **6**, 21

385. Keenan, J., Beaugie, C., Jasmann, B., Potter, H., Collett, J., and Frizelle, F. (2010) Helicobacter species in the human colon. *Colorectal Disease* **12**, 48-53

386. Papamichael, K., Konstantopoulos, P., and Mantzaris, G. J. (2014) Helicobacter pylori infection and inflammatory bowel disease: is there a link? *World journal of gastroenterology: WJG* **20**, 6374

387. Dahmus, J. D., Kotler, D. L., Kastenberg, D. M., and Kistler, C. A. (2018) The gut microbiome and colorectal cancer: a review of bacterial pathogenesis. *Journal of gastrointestinal oncology* **9**, 769

388. Maignan, S., and Mikol, V. (2001) The use of 3D structural data in the design of specific factor Xa inhibitors. *Current topics in medicinal chemistry* **1**, 161-174

389. Holmquist, B., and Vallee, B. L. (1974) Metal substitutions and inhibition of thermolysin: spectra of the cobalt enzyme. *Journal of Biological Chemistry* **249**, 4601-4607

390. Massaoud, M. K., Marokházi, J., and Venekei, I. (2011) Enzymatic characterization of a serralysin-like metalloprotease from the entomopathogen bacterium, Xenorhabdus. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1814**, 1333-1339

391. Buenning, P., Holmquist, B., and Riordan, J. F. (1983) Substrate specificity and kinetic characteristics of angiotensin converting enzyme. *Biochemistry* **22**, 103-110

392. Mallya, S., and Van Wart, H. (1989) Mechanism of inhibition of human neutrophil collagenase by Gold (I) chrysotherapeutic compounds. Interaction at a heavy metal binding site. *Journal of Biological Chemistry* **264**, 1594-1601

393. Gomez-Ortiz, M., Gomis-Rüth, F. X., Huber, R., and Avilés, F. X. (1997) Inhibition of carboxypeptidase A by excess zinc: analysis of the structural determinants by X-ray crystallography. *FEBS letters* **400**, 336-340

394. Larsen, K. S., and Auld, D. S. (1989) Carboxypeptidase A: mechanism of zinc inhibition. *Biochemistry* **28**, 9620-9625

395. Larsen, K. S., and Auld, D. S. (1991) Characterization of an inhibitory metal binding site in carboxypeptidase A. *Biochemistry* **30**, 2613-2618

396. Meraz-Cruz, N., Ortega, A., Estrada-Gutierrez, G., Flores, A., Espejel, A., Hernandez-Guerrero, C., and Vadillo-Ortega, F. (2006) Identification of a calcium-dependent matrix metalloproteinase complex in rat chorioallantoid membranes during labour. *Molecular human reproduction* **12**, 633-641

397. Karim, A. Y., Kulczycka, M., Kantyka, T., Dubin, G., Jabaiah, A., Daugherty, P. S., Thogersen, I. B., Enghild, J. J., Nguyen, K.-A., and Potempa, J. (2010) A novel matrix metalloprotease-like enzyme (karilysin) of the periodontal pathogen Tannerella forsythia ATCC 43037. *Biological chemistry* **391**, 105-117

398. Gossas, T., and Danielson, U. H. (2006) Characterization of Ca2+ interactions with matrix metallopeptidase-12: implications for matrix metallopeptidase regulation. *Biochemical Journal* **398**, 393-398

399. López Pelegrín, M. d. M. (2015) *Mechanisms of proteolytic activity regulation exerted via a unique propeptide in matrix metalloproteinases and intra/intermolecular interactions in a novel family of minimal gluzincins.* Doctor of Philosophy, Universitat de Barcelona

400. Beynon, R. J., and Bond, J. S. (2001) *Proteolytic enzymes: a practical approach*, Oxford University Press Oxford

401. Konigsberg, W. (1972) Reduction of disulfide bonds in proteins with dithiothreitol. in *Methods in enzymology* (Hirs, C., and Timasheff, S. eds.), Academic Press, San Diego. pp 185-188

402. Prajapati, S. C., and Chauhan, S. S. (2011) Dipeptidyl peptidase III: a multifaceted oligopeptide N-end cutter. *The FEBS journal* **278**, 3256-3276

403. Fukasawa, K. M., Hata, T., Ono, Y., and Hirose, J. (2011) Metal preferences of zinc-binding motif on metalloproteases. *Journal of amino acids* **2011**

404. Castellazzi, M., Tamborino, C., Fainardi, E., Manfrinato, M. C., Granieri, E., Dallocchio, F., and Bellini, T. (2007) Effects of anticoagulants on the activity of gelatinases. *Clin Biochem* **40**, 1272-1276

405. Fainardi, E., Castellazzi, M., Bellini, T., and Dallocchio, F. (2007) Matrix metalloproteinases (MMP): determination of different forms by different techniques may require different preanalytical strategies. *Multiple Sclerosis Journal* **13**, 561

406. Van Wart, H. E., and Birkedal-Hansen, H. (1990) The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proceedings of the National Academy of Sciences* **87**, 5578-5582

407. Springman, E. B., Angleton, E. L., Birkedal-Hansen, H., and Van Wart, H. E. (1990) Multiple modes of activation of latent human fibroblast collagenase: evidence for the role of a Cys73 active-site zinc complex in latency and a" cysteine switch" mechanism for activation. *Proceedings of the National Academy of Sciences* **87**, 364-368

408. Franco, A. A., Buckwold, S. L., Shin, J. W., Ascon, M., and Sears, C. L. (2005) Mutation of the zinc-binding metalloprotease motif affects Bacteroides fragilis toxin activity but does not affect propeptide processing. *Infection and immunity* **73**, 5273-5277

409. Stöcker, W., and Bode, W. (1995) Structural features of a superfamily of zinc-endopeptidases: the metzincins. *Current opinion in structural biology* **5**, 383-390

410. Boldt, H. B., Overgaard, M. T., Laursen, L. S., Weyer, K., Sottrup-Jensen, L., and Oxvig, C. (2001) Mutational analysis of the proteolytic domain of pregnancy-associated plasma protein-A (PAPP-A): classification as a metzincin. *Biochem J* **358**, 359-367

411. Cha, J., and Auld, D. S. (1997) Site-directed mutagenesis of the active site glutamate in human matrilysin: investigation of its role in catalysis. *Biochemistry* **36**, 16019-16024

412. Pieper, M., Betz, M., Budisa, N., Gomis-Rüth, F.-X., Bode, W., and Tschesche, H. (1997) Expression, purification, characterization, and X-ray

analysis of selenomethionine 215 variant of leukocyte collagenase. *Journal of protein chemistry* **16**, 637-650

413. Butler, G. S., Tam, E. M., and Overall, C. M. (2004) The Canonical Methionine 392 of Matrix Metalloproteinase 2 (Gelatinase A) Is Not Required for Catalytic Efficiency or Structural Integrity. *Journal of Biological Chemistry* **279**, 15615-15620

414. Bode, W., Grams, F., Reinemer, P., Gomis-Rüth, F. X., Baumann, U., McKay, D. B., and Stöcker, W. (1996) The metzincin-superfamily of zinc-peptidases. *Adv Exp Med Biol* **389**, 1-11

415. Oberholzer, A. E., Bumann, M., Hege, T., Russo, S., and Baumann, U. (2009) Metzincin's canonical methionine is responsible for the structural integrity of the zinc-binding site. *Biological chemistry* **390**, 875-881

416. Hege, T., and Baumann, U. (2001) The conserved methionine residue of the metzincins: a site-directed mutagenesis study. *Journal of molecular biology* **314**, 181-186

417. Perez, L., Kerrigan, J. E., Li, X., and Fan, H. (2007) Substitution of methionine 435 with leucine, isoleucine, and serine in tumor necrosis factor alpha converting enzyme inactivates ectodomain shedding activity. *Biochemistry and cell biology* **85**, 141-149

418. Baumann, U. (1994) Crystal structure of the 50 kDa metallo protease from Serratia marcescens. *Journal of molecular biology* **242**, 244-251

419. Baumann, U. (2019) Structure–Function Relationships of the Repeat Domains of RTX Toxins. *Toxins* **11**, 657

420. Keller, K. E., Aga, M., Bradley, J. M., Kelley, M. J., and Acott, T. S. (2009) Extracellular matrix turnover and outflow resistance. *Experimental eye research* **88**, 676-682

421. Factor Xa Protease. New England Biolabs

422. Hamilton, J. R., Frauman, A. G., and Cocks, T. M. (2001) Increased expression of protease-activated receptor-2 (PAR2) and PAR4 in human coronary artery by inflammatory stimuli unveils endothelium-dependent relaxations to PAR2 and PAR4 agonists. *Circulation research* **89**, 92-98

423. Kawao, N., Sakaguchi, Y., Tagome, A., Kuroda, R., Nishida, S., Irimajiri, K., Nishikawa, H., Kawai, K., Hollenberg, M. D., and Kawabata, A. (2002) Protease-activated receptor-2 (PAR-2) in the rat gastric mucosa: immunolocalization and facilitation of pepsin/pepsinogen secretion. *British journal of pharmacology* **135**, 1292-1296

424. Liu, S., Tobias, R., McClure, S., Styba, G., Shi, Q., and Jackowski, G. (1997) Removal of endotoxin from recombinant protein preparations. *Clinical biochemistry* **30**, 455-463

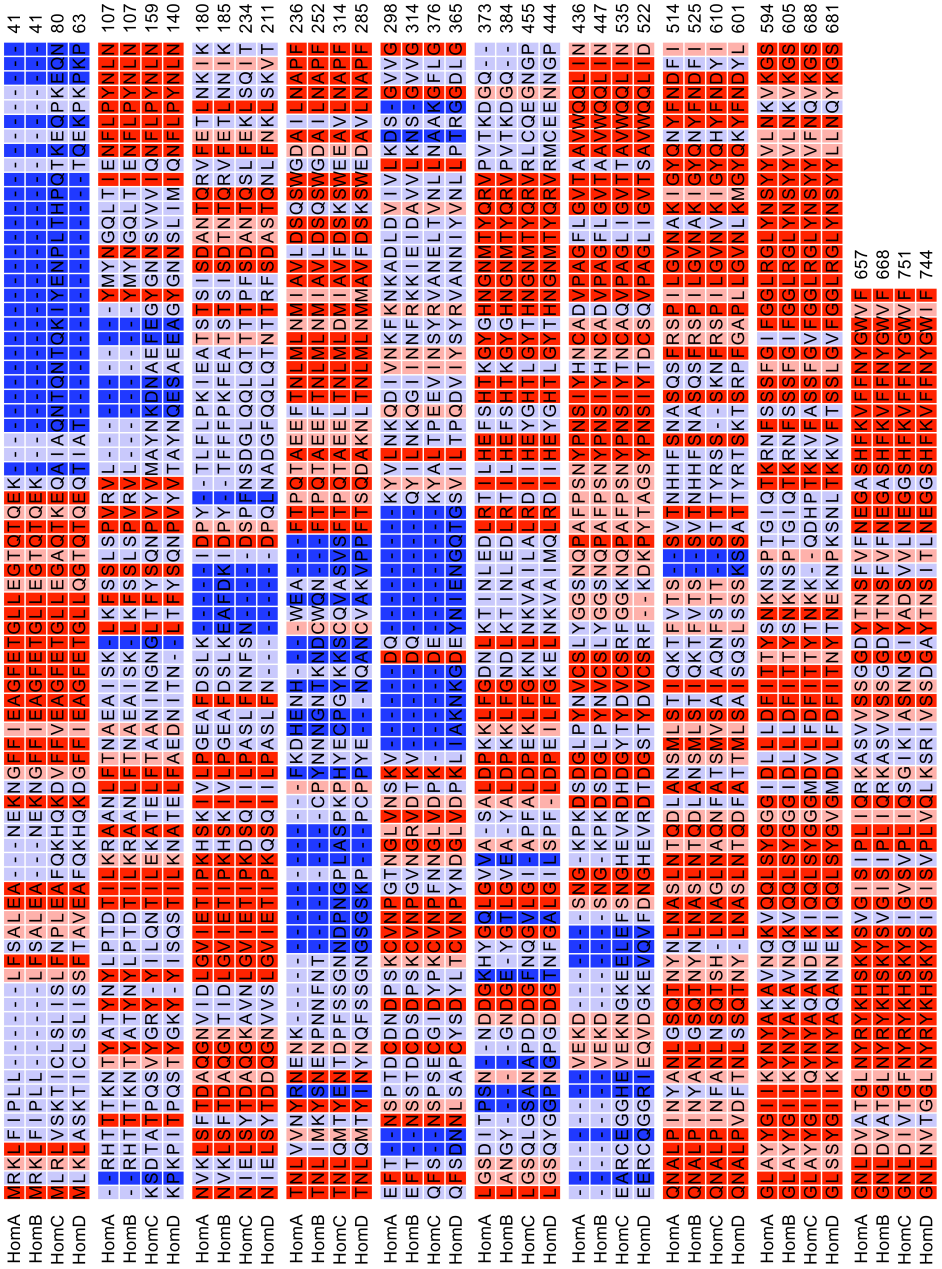
425. Su, B., Ceponis, P. J., Lebel, S., Huynh, H., and Sherman, P. M. (2003) Helicobacter pylori activates Toll-like receptor 4 expression in gastrointestinal epithelial cells. *Infection and immunity* **71**, 3496-3502

426. Minutolo, N. G., Sharma, P., Poussin, M., Shaw, L. C., Brown, D. P., Hollander, E. E., Smole, A., Rodriguez-Garcia, A., Hui, J. Z., Zappala, F., Tsourkas, A., and Powell, D. J., Jr. (2020) Quantitative Control of Gene-Engineered T-Cell Activity through the Covalent Attachment of Targeting Ligands to a Universal Immune Receptor. *J Am Chem Soc* **142**, 6554-6568

427. Merdanovic, M., Clausen, T., Kaiser, M., Huber, R., and Ehrmann, M. (2011) Protein quality control in the bacterial periplasm. *Annual review of microbiology* **65**, 149-168

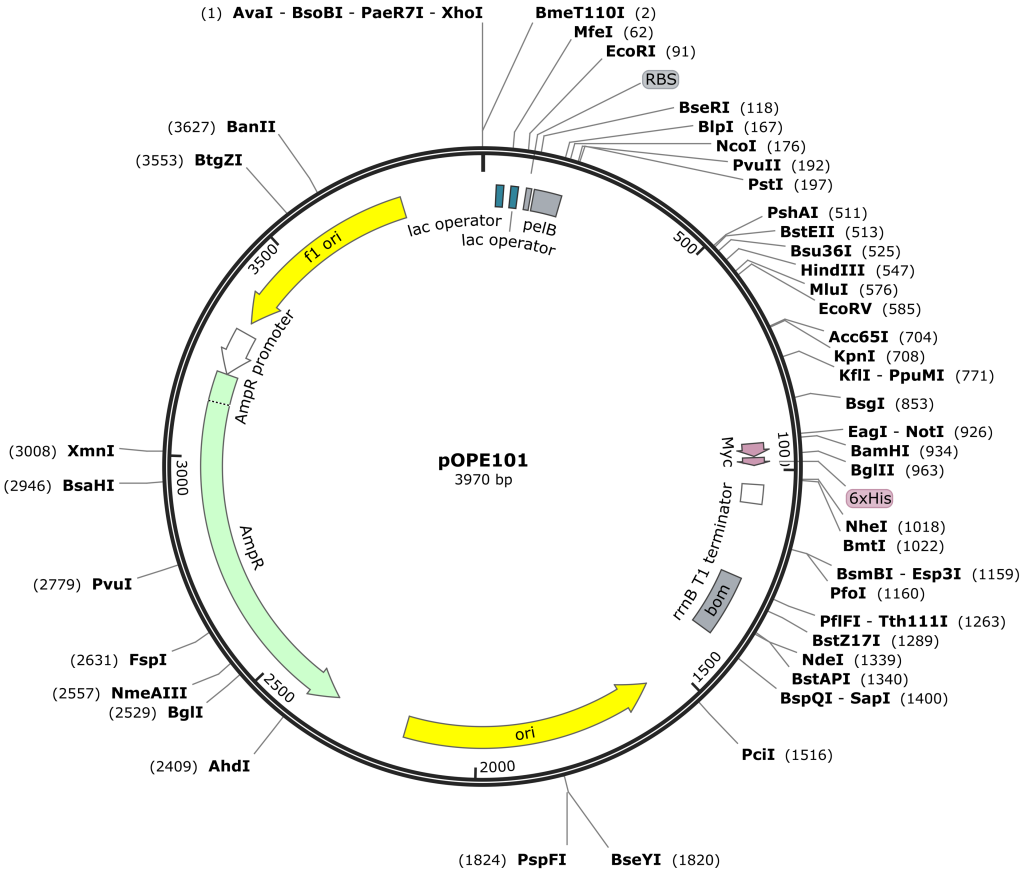
428. El-Hawiet, A., Chen, Y., Shams-Ud-Doha, K., Kitova, E. N., Kitov, P. I., Bode, L., Hage, N., Falcone, F. H., and Klassen, J. S. (2018) Screening natural libraries of human milk oligosaccharides against lectins using CaR-ESI-MS. *Analyst* **143**, 536-548

Appendices

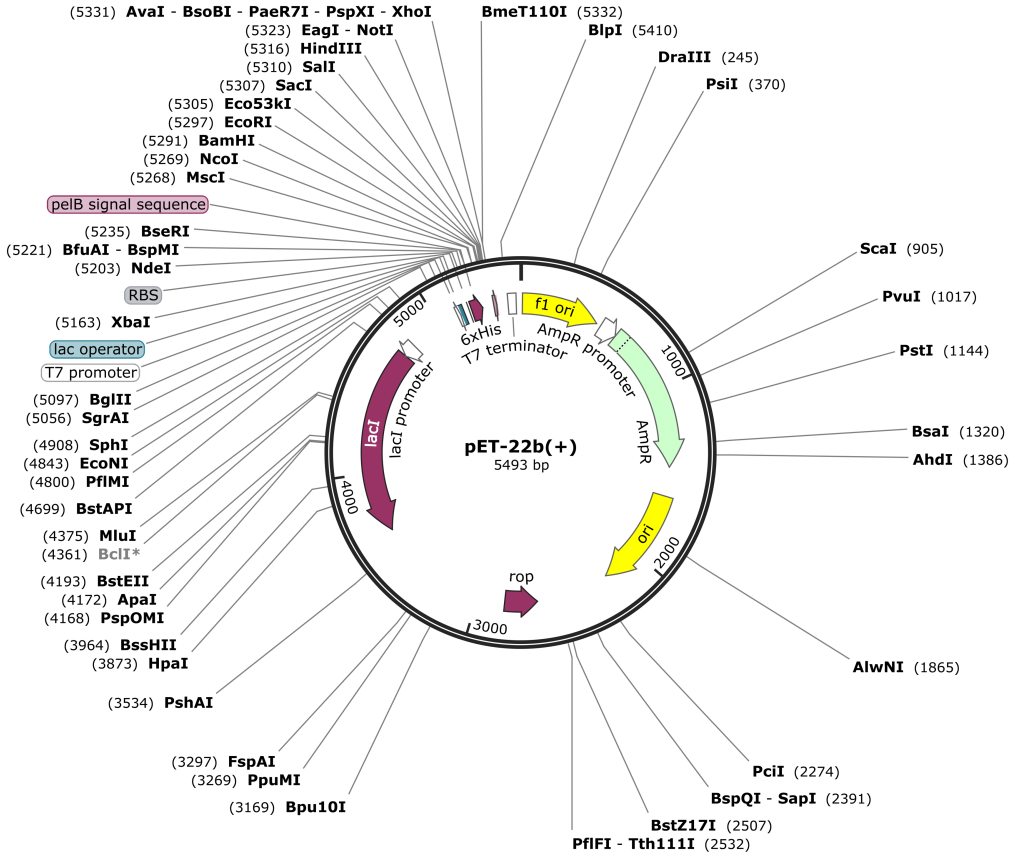


Appendix Figure 1: Amino acid sequence alignment of HomA, HomB, HomC, and HomD

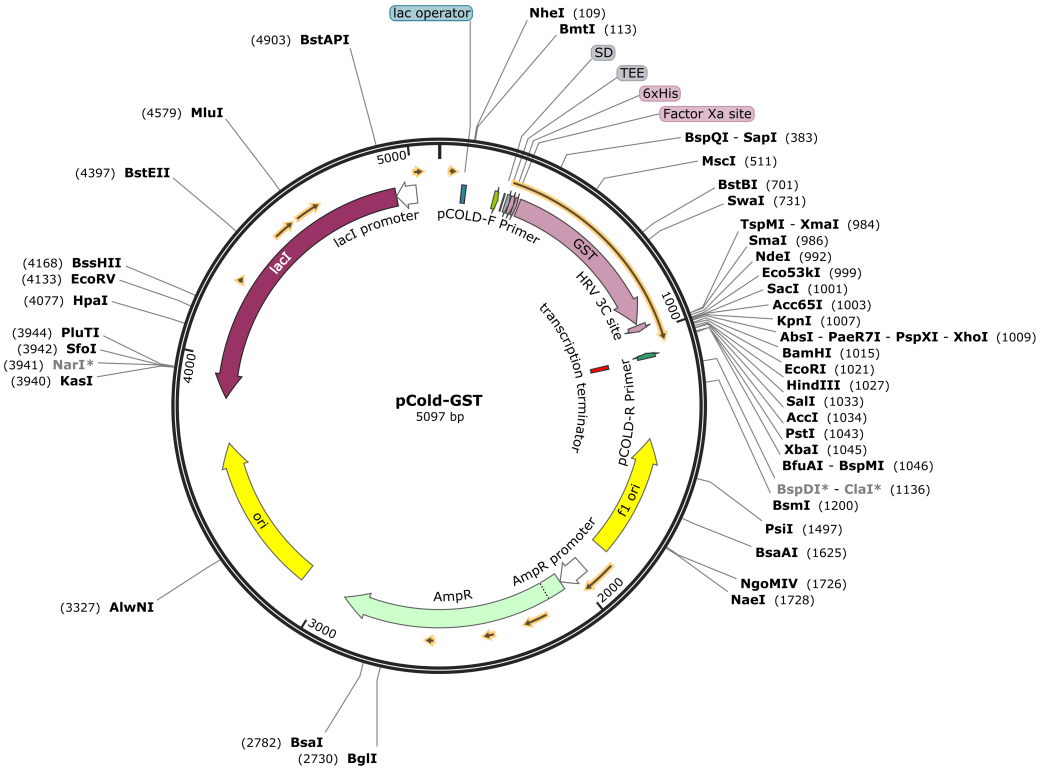
Amino acid sequence alignment of HomA (Genbank accession no. WP_046655475.1), HomB (WP_046655427.1), HomC (WP_000945291.1), and HomD (WP_000915421.1). Red, pink, and light blue boxshades highlight the identical amino acids in all Hom proteins, three Hom proteins, and two Hom proteins, respectively. Blue boxshades highlight the mismatched amino acids in all Hom proteins.



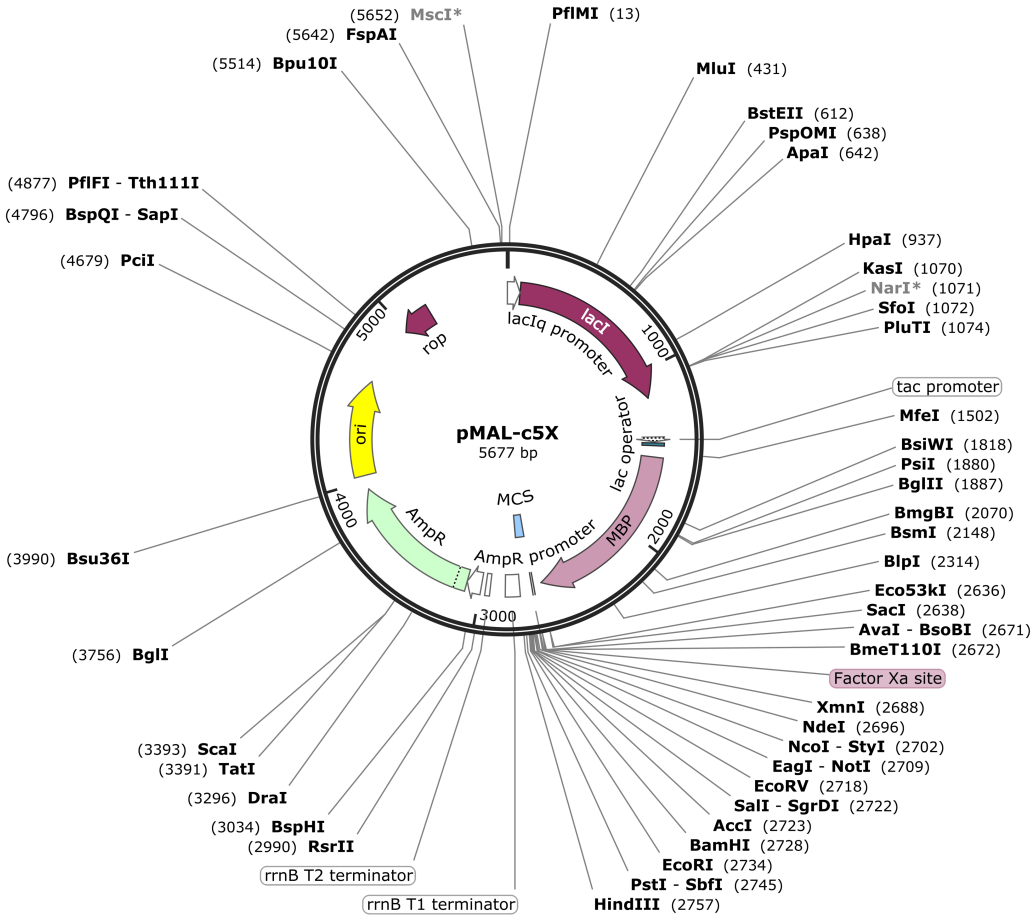
Appendix Figure 2: SnapGene map of pOPE101



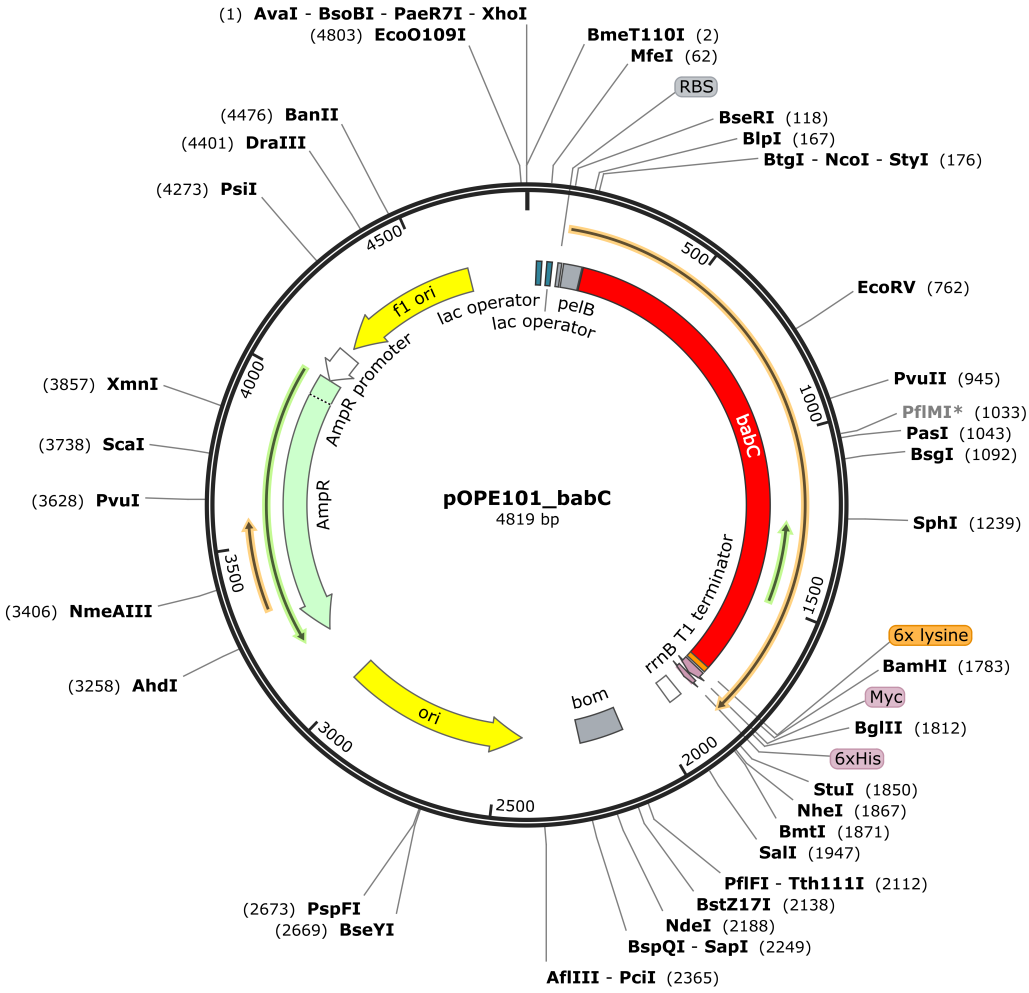
Appendix Figure 3: SnapGene map of pET-22b(+) vector



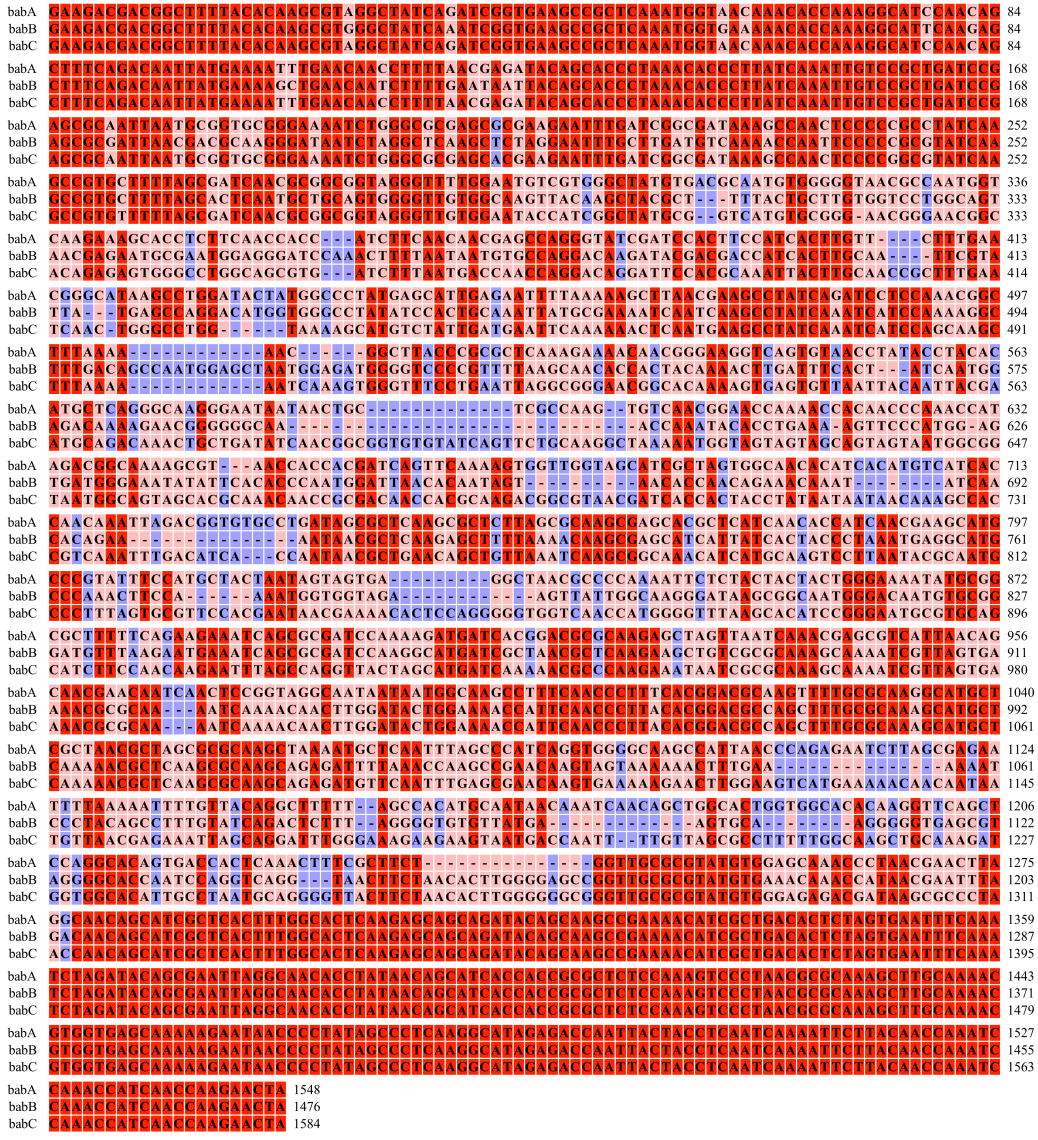
Appendix Figure 4: SnapGene map of pCold-GST vector



Appendix Figure 5: SnapGene map of pMAL-c5X vector

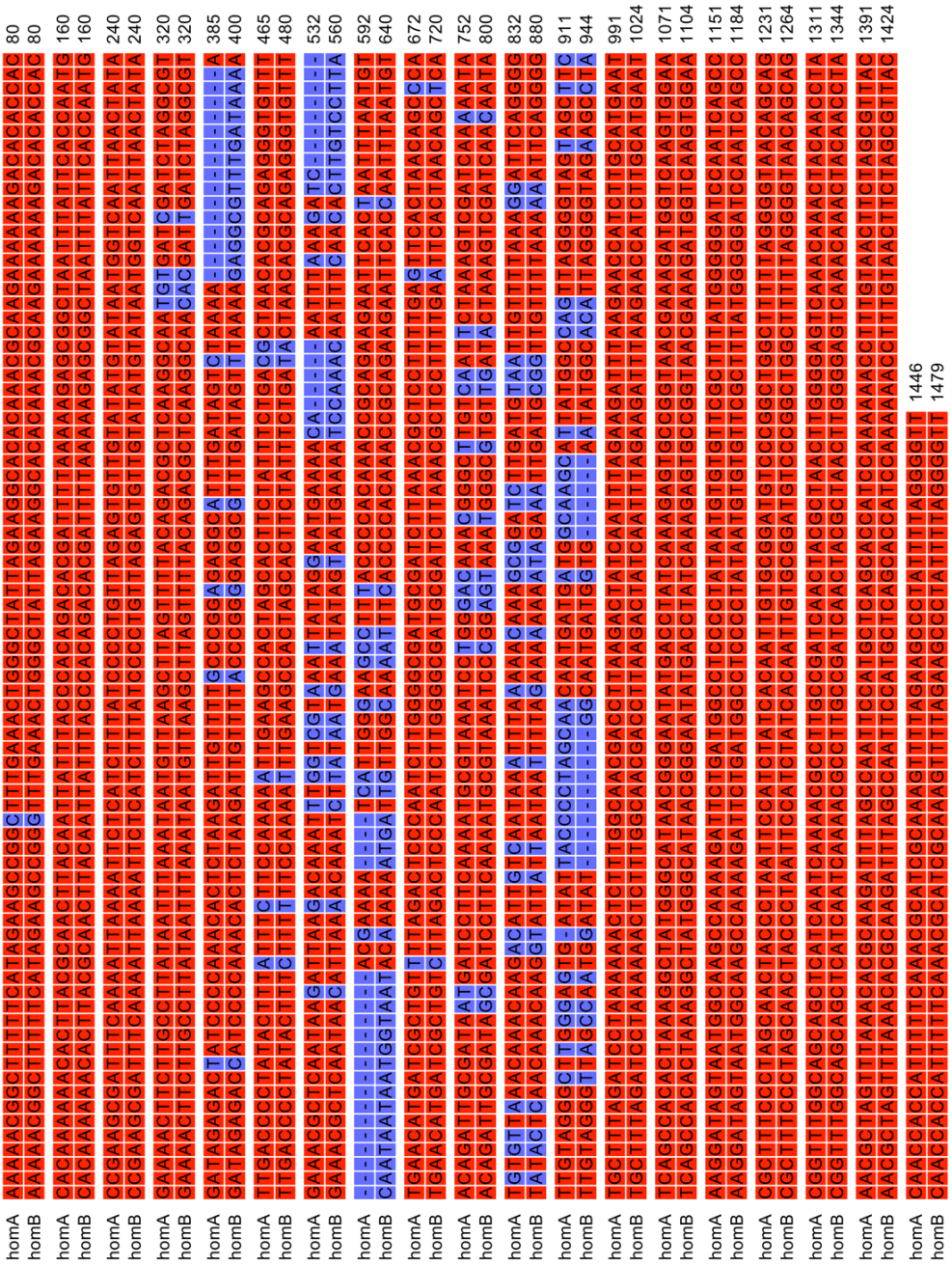


Appendix Figure 6: SnapGene map of pOPE101_*babC* construct



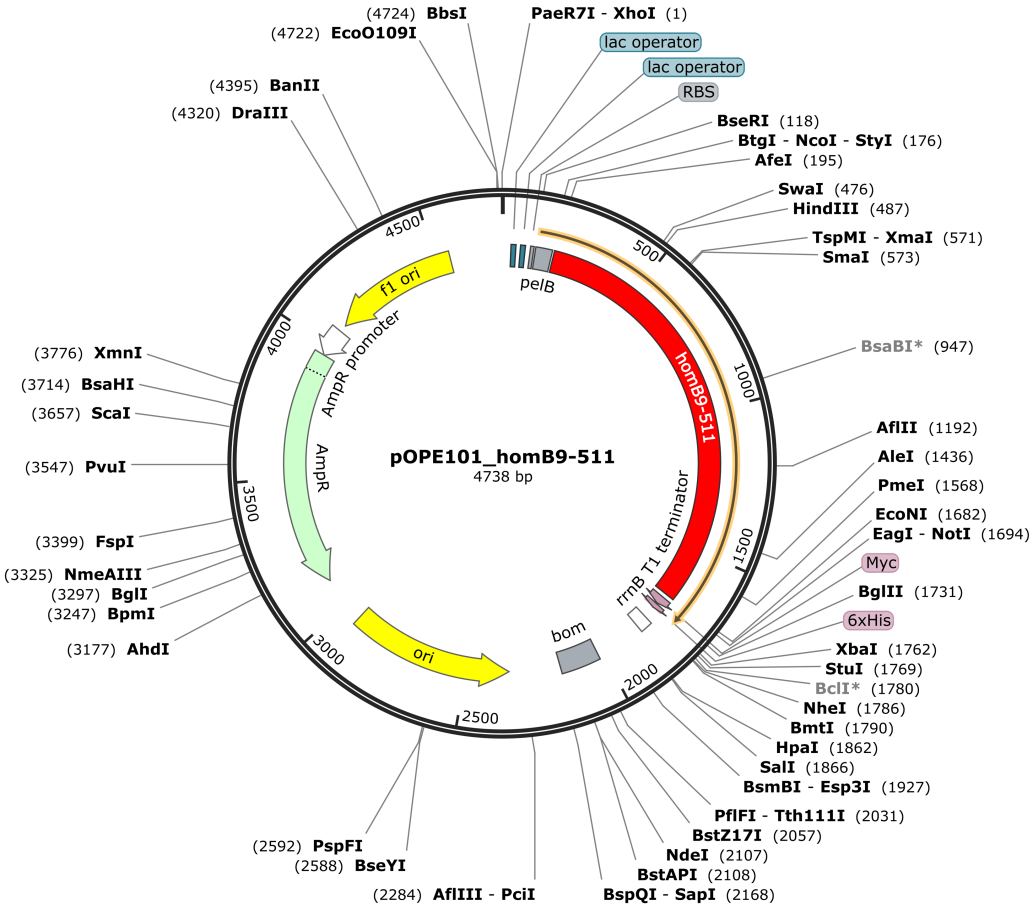
Appendix Figure 7: DNA alignment of the N-terminal extracellular domain of *babA*, *babB*, and *babC* genes from H. pylori 26695

DNA sequence alignment shows that *babA* (GenBank accession no. AAD08288.1), *babB* (AAD07943.1), and *babC* (AAD07380.1) have high homologies in the 5' and 3' regions. The sequence alignment shows the amplified region of *babC* gene only that encodes the extracellular domain. Identical nucleotides in all three genes are highlighted by red boxshade, while identical nucleotides in two genes only are highlighted by pink boxshade, and mismatches by blue boxshade.

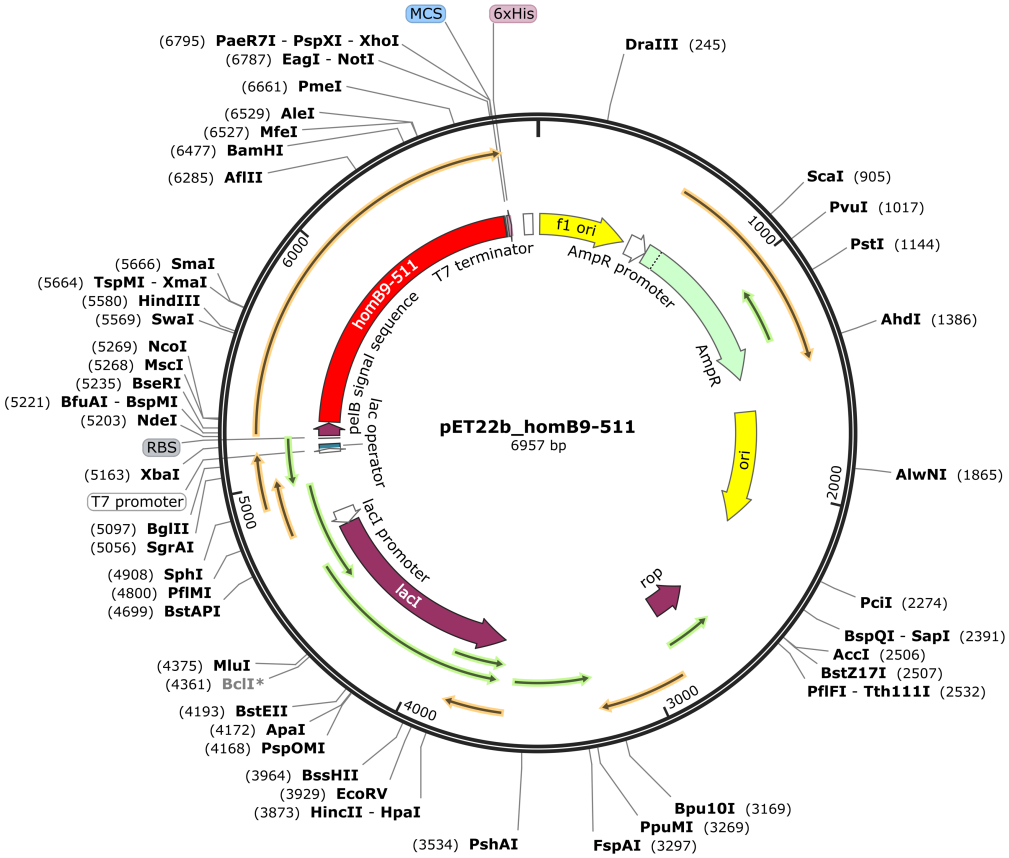


Appendix Figure 8: DNA Alignment of *homA* and *homB* genes from H. pylori J99

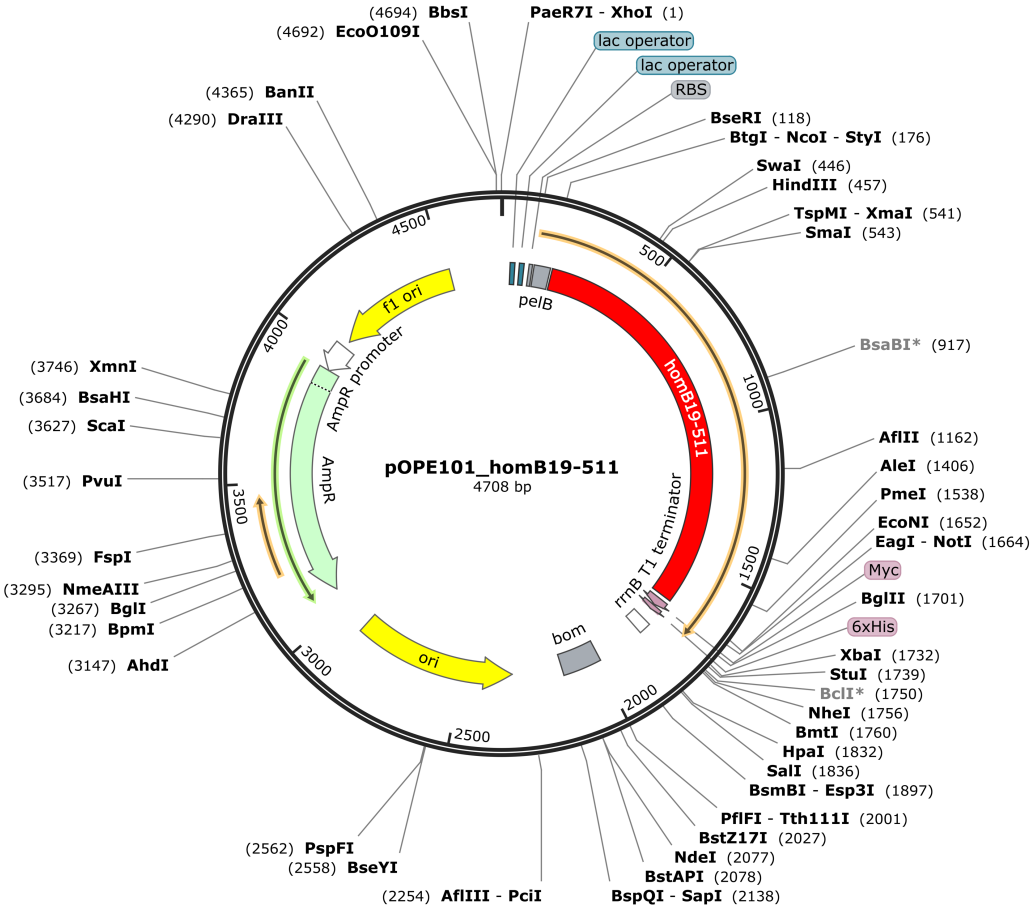
DNA sequence alignment shows that *homA* (Genbank accession no. WP_046655475.1) and *homB* (WP_046655427.1) have high homologies in the 5' and 3' regions. The sequence alignment shows the amplified region of *homB* gene only that encodes the extracellular domain. Identical nucleotides in both genes are highlighted by red boxshade, while mismatches by blue boxshade.



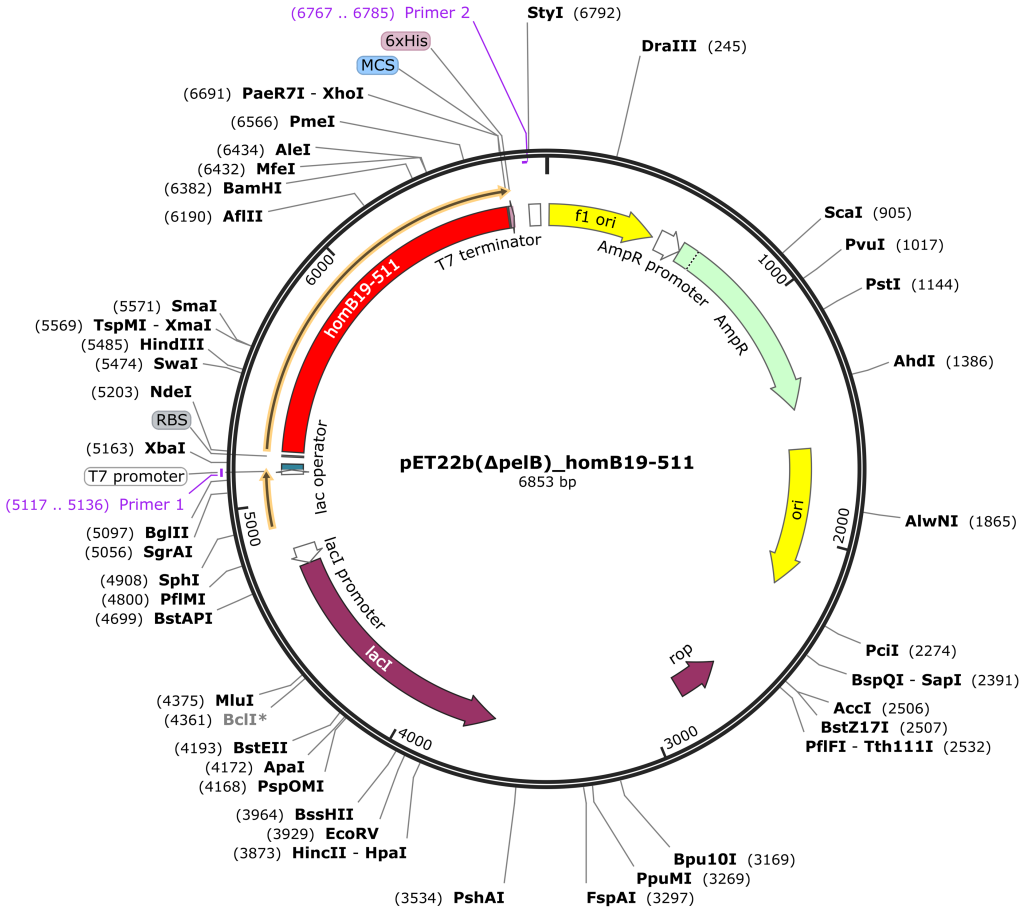
Appendix Figure 9: SnapGene map of pOPE101_*homB*$_{9\text{-}511}$ construct



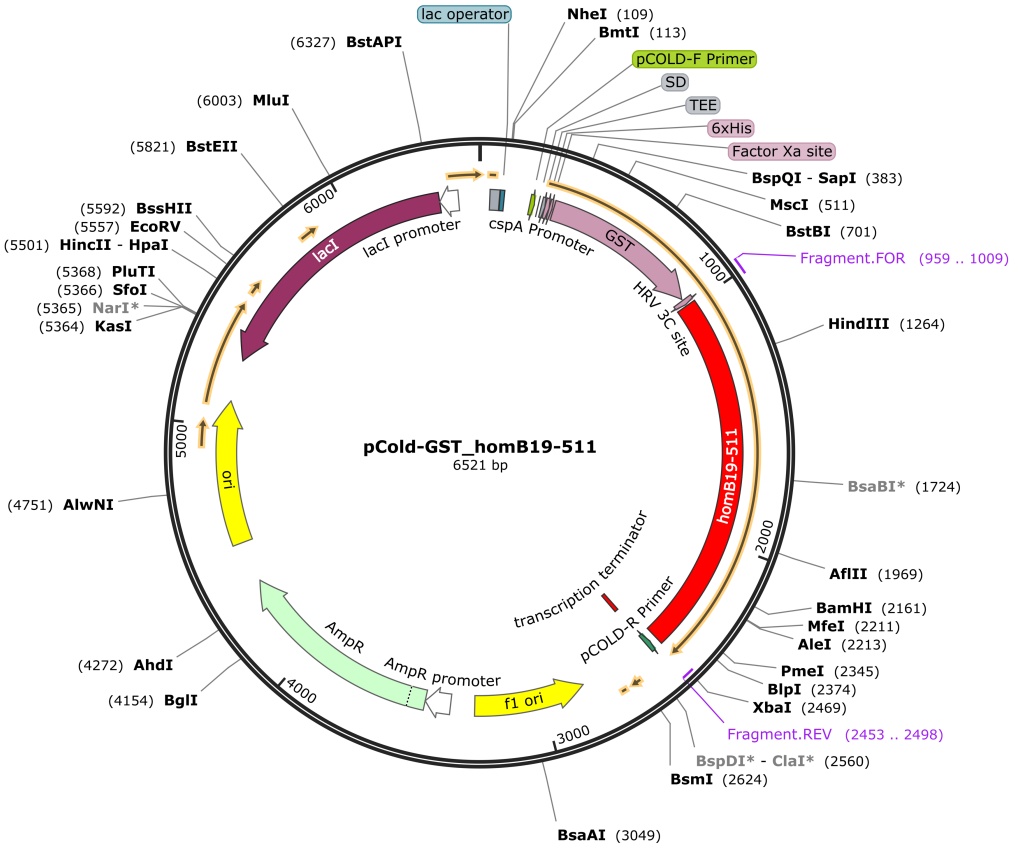
Appendix Figure 10: SnapGene map of pET22b_*homB*$_{9\text{-}511}$ construct



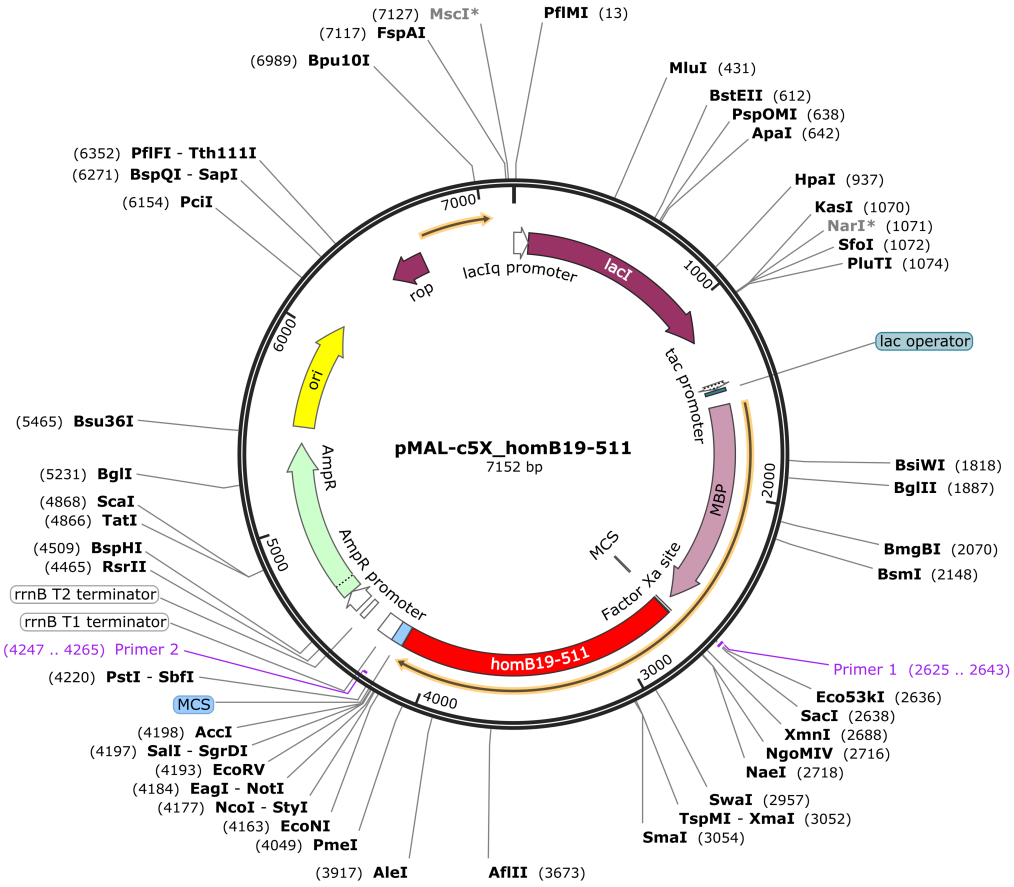
Appendix Figure 11: SnapGene map of pOPE101_*homB*$_{19\text{-}511}$ construct



Appendix Figure 12: SnapGene map of pET22b(ΔpelB)_*homB*$_{19\text{-}511}$ construct



Appendix Figure 13: SnapGene map of pCold-GST_*homB*$_{19\text{-}511}$ construct



Appendix Figure 14: SnapGene map of pMAL-c5X_*homB*$_{19\text{-}511}$ construct