



Development of Diagnostic Tools for the Detection  
of *Streptococcus agalactiae*

Hannah Swinburne

Thesis submitted for the degree of Doctor of Philosophy

Institute of Cellular Medicine

Faculty of Medical Sciences

Newcastle University

November 2017

## Abstract

*Streptococcus agalactiae* is a common opportunistic pathogen, which, when transmitted to the immunocompromised, elderly or newborns, can cause complications such as sepsis, pneumonia, meningitis and can also be fatal. In screening programmes expectant mothers are typically sampled at 35-37 weeks gestation to ensure microbial culture results are available prior to labour. As levels of *S. agalactiae* rapidly fluctuate, premature testing results in antibiotic prescription being informed by outdated information. This project aimed to develop tools for use in a sensitive, specific, point-of-care diagnostic test able to detect all strains of *S. agalactiae*. Three detection strategies were pursued based on peptide aptamers, antibodies and DNA amplification.

A bioluminescent protein, aequorin, was selected for use as a peptide aptamer scaffold in order to link epitope detection to reporter signal. Peptide insertion into aequorin eliminated expression of soluble protein, and therefore was deemed unsuitable for aptamer library construction.

A novel computer programme, IDRIS, identified peptides conserved throughout all currently sequenced *S. agalactiae* strains, which are also unique to the species. Through further bioinformatic analysis, three peptide sequences from SAG1474 protein were selected for in-house murine monoclonal antibody production. Antibodies targeted against one peptide (CVNLEENSQV) exhibited sensitivity to all *S. agalactiae* serotypes. Antibodies exhibited species specificity when tested against *Streptococcus pyogenes*; future work will focus on extending specificity studies, including analysis of bacteria present in the vaginal flora.

Recombinase polymerase amplification is a rapid, isothermal DNA amplification technique that may enable inexpensive point-of-care testing. Target-specific primers with single stranded overhang sequences allow amplified product to hybridise to a complementary capture probe and reporter-tagged detection probe. The plate-based system can detect  $3.8 \times 10^4$  copies of DNA and is able to specifically detect all *S. agalactiae* serotypes over *S. pyogenes* DNA. The test will be converted into a lateral flow assay in the future.

## Declaration

I hereby declare that this thesis has been composed by myself and has not been submitted in any previous application for a degree. The work presented has been performed by myself, unless otherwise stated. All sources of information have been appropriately acknowledged by means of reference.

## Acknowledgements

I am grateful to the EPSRC for funding this work.

I would like to thank Neil Keegan, Chris Johnson and Calum McNeil for their continual support throughout the project. Their knowledge, help, ideas, guidance and encouragement have pushed me to become a better scientist.

Thanks to Julia Spoons and Rob Bolt for countless reasons, most significantly in all their work in antibody development. Thanks to Beth Lawry for her advice in every aspect of the project. I'd like to thank Keith Flanagan, Elisa Anastasi and Anil Wipat for their help using the IDRIS system, Helen Glenwright and Wendy Smith for their support whilst working in the Centre for Bacterial Cell Biology and Nicholas Jakubovics for kindly donating *S. pyogenes* strains.

Special thanks to Cara, Emily, Gina, Katie, Suzie, Carl, Phil, Chen and all those mentioned above for their friendship; it has made the past three years wonderful.

## Contents

Chapter 1 <i>Streptococcus agalactiae</i> : Bacteria, Disease and Diagnostics .....	1
1.1 <i>Streptococcus agalactiae</i> bacterium .....	1
1.1.1 <i>Streptococcus agalactiae</i> .....	1
1.1.2 Serotypes .....	3
1.1.3 Related bacteria belonging to the genus streptococcus .....	3
1.1.4 Virulence factors.....	6
1.2 Group B streptococcal disease .....	8
1.2.1 Early onset disease .....	8
1.2.2 Late onset disease .....	10
1.2.3 Immunocompromised.....	11
1.2.4 Elderly .....	12
1.2.5 Worldwide disease .....	13
1.3 Treating and preventing GBS disease .....	13
1.3.1 GBS vaccine development.....	13
1.3.2 Antibiotic treatment.....	14
1.3.3 Maternal screening/risk- based treatment .....	16
1.4 Diagnosis of <i>S. agalactiae</i> infection .....	18
1.4.1 Current GBS diagnostic tests .....	18
1.4.2 Rapid test .....	24
1.5 Project aims .....	26
Chapter 2 Methods.....	28
2.1 Materials .....	28
2.2 General techniques.....	29
2.2.1 Bacterial cell growth .....	29
2.2.2 Bacterial cell storage .....	30
2.2.3 SDS-PAGE.....	30

2.2.4 Western blotting .....	31
2.2.5 Indirect enzyme linked immunosorbent assay (ELISA) .....	32
2.2.6 Protein concentration determination – BCA assay .....	33
2.2.7 Protein concentration determination – spectrophotometry .....	33
2.2.8 RNA extraction .....	33
2.2.9 cDNA synthesis .....	33
2.2.10 DNA extraction .....	34
2.2.11 Polymerase chain reaction .....	34
2.2.12 Agarose gel .....	34
2.2.13 Nucleic acid concentration determination - spectrophotometry .....	35
2.2.14 DNA purification .....	35
2.2.15 Statistics .....	35
2.3 Computing .....	36
2.3.1 IDRIS.....	36
2.3.2 PSORTb.....	37
2.3.3 SignalP .....	37
2.3.4 LipoP .....	37
2.3.5 TMHMM .....	38
2.3.6 I-TASSER.....	38
2.3.7 PSIPRED .....	38
2.3.8 Prodepth.....	39
2.3.9 Bcepred.....	39
2.3.10 Microarray data analysis .....	39
2.3.11 Protein BLAST (Basic local alignment tool) search .....	40
2.3.12 DNA alignment .....	40
2.3.13 Primer design: PCR and RPA .....	40
2.3.14 Band densitometry .....	41

2.4 Recombinant protein expression and purification .....	41
2.4.1 Overview .....	41
2.4.2 Competent cells.....	41
2.4.3 Transformation .....	42
2.4.4 Induction growth tests .....	42
2.4.5 Cell lysis .....	43
2.4.6 Periplasmic extraction .....	43
2.4.7 Protease cleavage.....	43
2.4.8 Protein purification - chromatography.....	44
2.4.9 Nickel-affinity chromatography .....	44
2.4.10 GST-affinity chromatography .....	44
2.4.11 Ion exchange chromatography.....	45
2.4.12 Size exclusion chromatography .....	45
2.4.13 Circular dichroism.....	46
2.4.14 Liquid chromatography mass spectrometry.....	46
2.5 Antibody production .....	47
2.5.1 Overview .....	47
2.5.2 Peptide - carrier protein conjugation.....	47
2.5.3 Ellman's test.....	47
2.5.4 Mouse immunisation .....	48
2.5.5 Cell culture .....	49
2.5.6 Viable cell count .....	49
2.5.7 Freezing cells .....	49
2.5.8 Thawing frozen cells.....	50
2.5.9 Fusion .....	50
2.5.10 Cloning .....	50
2.5.11 Antibody purification .....	51

2.6 Recombinase Polymerase Amplification (RPA) .....	51
2.6.1 RPA assay .....	51
2.6.2 Enzyme linked oligonucleotide assay .....	52
Chapter 3 Protein target selection .....	53
3.1 Introduction .....	53
3.2 GBS specific peptides .....	53
3.3 Protein localisation .....	58
3.3.1 PSORTb .....	58
3.3.2 SignalP .....	59
3.3.3 LipoP .....	59
3.3.4 TMHMM .....	60
3.3.5 Surface exposure .....	61
3.3.6 Protein localisation consensus .....	61
3.4 Protein selection .....	62
3.4.1 Experimental analysis of protein expression in 10 serotypes .....	62
3.4.2 Microarray data from previous literature .....	64
3.4.2.1 pH .....	64
3.4.2.2 Temperature .....	65
3.4.2.3 Oxygen concentration .....	66
3.4.3 Structure prediction .....	67
3.4.4 Evaluation of expression and structural modelling data .....	68
3.5 Token selection .....	70
3.5.1 Unique peptides .....	70
3.5.2 Polarity .....	71
3.5.3 Secondary structure .....	71
3.5.4 Surface exposure .....	72
3.5.5 Combination of predictive tools .....	72

3.5.6 Conclusion .....	75
Chapter 4 Antibody generation against <i>S. agalactiae</i> biomarkers .....	76
4.1 Introduction .....	76
4.1.1 Immunoassays in diagnostic tests.....	76
4.1.2 <i>In vivo</i> antibody production.....	76
4.2 Antibody production .....	76
4.3 Antibody characterisation against linear peptide.....	84
4.4 Production of recombinant SAG1474.....	86
4.4.1 Design of SAG1474 construct.....	86
4.4.2 Induction of SAG1474 .....	88
4.4.3 Design of a SAG1474 fusion protein .....	92
4.4.4 Induction of SAG1474-MBP .....	94
4.4.5 SAG1474-MBP purification .....	97
4.4.6 Structural characterisation of SAG1474-MBP .....	101
4.4.6.1 Far UV circular dichroism .....	101
4.4.6.2 Near UV circular dichroism .....	102
4.4.6.3 Melting temperature circular dichroism .....	103
4.5 Antibody characterisation against folded protein .....	104
4.6 Antibody characterisation against bacterial cells .....	107
4.6.1 Comparison of antibodies.....	109
4.6.2 Antibody characterisation against <i>S. agalactiae</i> serotypes .....	110
4.7 Conclusions and future work.....	112
Chapter 5 RPA DNA amplification.....	114
5.1 Introduction .....	114
5.1.1 DNA amplification in diagnostic tests .....	114
5.1.2 Isothermal DNA amplification .....	115
5.1.3 Recombinase polymerase amplification .....	116

5.1.4 Hybridisation-based detection of RPA amplicon .....	119
5.2 Recombinase polymerase amplification for the detection of <i>S. agalactiae</i> ....	121
5.2.1 Analysis of primers .....	122
5.2.2 Tailed-primer DNA amplification.....	130
5.2.3 Strain sensitivity and specificity .....	133
5.3 Conclusions and future work.....	136
Chapter 6 Peptide aptamer .....	137
6.1 Introduction .....	137
6.2 Aptamers .....	137
6.2.1 Peptide aptamers .....	137
6.2.2 Nucleic acid aptamers .....	138
6.2.3 Designing a scaffold .....	139
6.2.3.1 Peptide aptamer constraint.....	139
6.2.3.2 Functional or non-functional scaffolds .....	140
6.2.4 Aptamer scaffolds as diagnostic tools .....	141
6.2.4.1 Surface adsorption .....	141
6.2.4.2 Specificity .....	141
6.2.4.3 Compatible with label-free detection techniques .....	142
6.2.4.4 The need for additional antigen detection molecules in diagnostics ..	142
6.2.5 Production and screening of a peptide aptamer library .....	142
6.3 Design of a reporter-based peptide aptamer .....	145
6.3.1 Selection of reporter scaffold.....	145
6.3.2 Aequorin.....	145
6.4 Sequence design of peptide aptamer scaffold .....	148
6.4.1 Site of aptamer insertion .....	148
6.4.2 Sequence optimisation .....	151
6.5 Induction of aequorin WT* .....	154

6.6 Purification of aequorin WT* .....	158
6.7 Induction of aequorin FLAG .....	160
6.8 Conclusion .....	162
Chapter 7 Discussion .....	163
7.1 Recent research in GBS diagnostics .....	163
7.1.1 Optimisation of PCR .....	163
7.1.2 Other methods of detection .....	163
7.1.3 Alternative detection strategies .....	164
7.1.4 Sample preparation .....	164
7.1.5 Antibiotic susceptibility .....	165
7.1.6 Identification of multiple pathogens .....	165
7.1.7 Comparison to current work .....	165
7.2 IDRIS as a useful tool .....	166
7.3 ASSURED Diagnostics .....	166
7.3.1 Antibody for application to GBS diagnostics.....	167
7.3.2 RPA for application to GBS diagnostics .....	167
7.3.3 Aequorin as a peptide aptamer scaffold for application to GBS diagnostics .....	168
7.4 Conclusion .....	168
Chapter 8 References .....	170

## List of figures

Figure 1.1 Schematic of gram positive bacteria.....	2
Figure 1.2 Phylogenetic tree of streptococcal species .....	5
Figure 1.3 Worldwide approach to early onset GBS prevention.....	17
Figure 3.1 IDRIS selection of unique and conserved peptides.....	55
Figure 3.2 Transcript expression in 10 serotypes of GBS .....	63
Figure 3.3 The effect of pH on gene transcript expression.....	65
Figure 3.4 The effect of temperature on gene transcript expression .....	66
Figure 3.5 Structure of three candidate proteins with IDRIS tokens mapped on.....	68
Figure 3.6 Structure of SAG1474 with selected peptides mapped on .....	75
Figure 4.1 Binding of murine test bleeds to peptide .....	79
Figure 4.2 Mechanism of HAT selection.....	81
Figure 4.3 Peptide affinity screening programme of hybridoma fusions to cloned cells .....	83
Figure 4.4 Antibody naming convention .....	84
Figure 4.5 Antibody binding to peptide .....	85
Figure 4.6 Recombinant SAG1474 construct DNA sequence .....	87
Figure 4.7 Recombinant SAG1474 protein.....	88
Figure 4.8 Induction of SAG1474: western blots .....	89
Figure 4.9 Growth curve following induction of SAG1474 with IPTG.....	91
Figure 4.10 Recombinant SAG1474-MBP construct DNA sequence .....	93
Figure 4.11 Recombinant SAG1474-MBP construct protein .....	94
Figure 4.12 Induction of SAG1474-MBP: western blots .....	95
Figure 4.13 Growth curve following induction of SAG1474-MBP with IPTG.....	97
Figure 4.14 Purification of recombinant SAG1474-MBP .....	99
Figure 4.15 TEV protease cleavage of SAG1474-MBP.....	100
Figure 4.16 Circular dichroism study of SAG1474: Far UV .....	102
Figure 4.17 Circular dichroism study of SAG1474: Near UV.....	103
Figure 4.18 Circular dichroism study of SAG1474: Melting temperature.....	104
Figure 4.19 Antibody binding to recombinant SAG1474-MBP protein.....	106
Figure 4.20 Antibody binding to lysed and whole cell <i>S. agalactiae</i> and <i>S. pyogenes</i> .....	108
Figure 4.21 Assessment of antibody sensitivity and specificity .....	111
Figure 5.1 Diagrammatic representation of recombinase polymerase amplification.....	118

Figure 5.2 Visualisation of tailed RPA detection.....	121
Figure 5.3 RPA of GBS DNA using primer pair E.....	125
Figure 5.4 RPA of GBS DNA using primer pair F .....	126
Figure 5.5 RPA of GBS DNA using primer pair F in 10 serotypes at 37 °C.....	128
Figure 5.6 RPA of GBS DNA using primer pair F in all 10 serotypes at 25 °C .....	129
Figure 5.7 Concentration curve of RPA performed at 37 °C for 10 minutes.....	131
Figure 5.8 Concentration curve of RPA performed at 37 °C for 20 minutes.....	132
Figure 5.9 Concentration curve of RPA performed at 25 °C for 20 minutes.....	133
Figure 5.10 RPA of DNA extracted from bacterial cell cultures .....	134
Figure 5.11 RPA directly from lysed cell cultures .....	135
Figure 6.1 Nucleic acid structures .....	139
Figure 6.2 Selection of peptide aptamers exhibiting target affinity .....	144
Figure 6.3 Mechanism of aequorin light emission and regeneration .....	147
Figure 6.4 Structure of aequorin.....	151
Figure 6.5 Recombinant aequorin (WT* and FLAG) construct DNA and protein sequence.....	153
Figure 6.6 Induction and solubility test for Aequorin WT*: comparison of lysis protocols .....	155
Figure 6.7 Induction and solubility test for Aequorin WT*: comparison of time points .....	156
Figure 6.8 Induction and solubility test for Aequorin WT*: comparison of induction temperatures .....	157
Figure 6.9 Confirmation induced protein is aequorin.....	158
Figure 6.10 Purification of Aequorin WT* .....	159
Figure 6.11 Induction and solubility test for Aequorin FLAG .....	160
Figure 6.12 Growth curve of cells expressing aequorin FLAG .....	161

## List of tables

Table 1.1 Examples of Streptococcal species belonging to various Lancefield groups .....	4
Table 1.2 Early onset GBS disease.....	9
Table 1.3 Late onset GBS disease.....	10
Table 1.4 Types of diagnostic tests for <i>S. agalactiae</i> detection.....	20
Table 1.5 FDA approved NAAT-based GBS diagnostic tests.....	23
Table 1.6 The use of a rapid GBS test in various clinical scenarios.....	25
Table 2.1 Origin of <i>E. coli</i> and streptococcal strains.....	29
Table 2.2 Antibiotic concentrations.....	30
Table 2.3 Details on PCR experiments .....	34
Table 2.4 Recombinant proteins and associated parameters .....	41
Table 2.5 <i>E coli</i> strains for production of recombinant protein.....	42
Table 2.6 Experimental details for circular dichroism .....	46
Table 2.7 Sequences for use in RPA assay .....	52
Table 3.1 IDRIS generated report listing proteins containing peptides distinctive to GBS .....	57
Table 3.2 Cellular location as predicted by PSORTb .....	58
Table 3.3 Signal P results for SAG0771, SAG1473 and SAG1474.....	59
Table 3.4 LipoP results for SAG0771, SAG1473 and SAG1474.....	60
Table 3.5 TMHMM results for SAG0771, SAG1473 and SAG1474.....	61
Table 3.6 Summary of results, collated to inform protein target selection .....	69
Table 3.7 Analysis of token suitability for use as immunogenic peptides. ....	74
Table 4.1 Conjugation of peptide to carrier protein.....	77
Table 4.2 Comparison of antibody binding to <i>S. agalactiae</i> and <i>S. pyogenes</i> .....	109
Table 5.1 Properties of isothermal DNA amplification techniques.....	116
Table 5.2 Analysis of primer properties for amplification of GBS DNA .....	123
Table 5.3 Analysis of tailed primer properties.....	130
Table 6.1 Analysis of sites in aequorin in which to insert a peptide aptamer.....	149

## List of abbreviations

ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ASSURED	Affordable, sensitive, specific, user-friendly, rapid and robust, equipment free, deliverable to the people in need
ATP	Adenosine triphosphate
BM	Basic media
bp	Base pair
BSA	Bovine serum albumin
C3	3 carbon spacer
CBC	Comparative Biology Centre
CD	Circular dichroism
cDNA	Complementary deoxyribonucleic acid
CE	Conformité Européenne
cfu	Colony forming unit
dH <sub>2</sub> O	Highly purified water
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphate
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
ELONA	Enzyme linked oligonucleotide assay
ESDA	Exponential strand displacement amplification
FBS	Foetal bovine serum
FDA	US Food and Drug Administration
GAS	Group A streptococcus / Group A streptococcal
GBS	Group B streptococcus / Group B streptococcal
GEO	Gene Expression Omnibus
GFP	Green fluorescent protein
GMP	Guanosine monophosphate
GST	Glutathione S transferase

HDA	Helicase dependent amplification
HAT	Hypoxanthine, aminopterin, thymidine
HGRPT	Hypoxanthine guanine phosphoribosyltransferase
HIV	Human immunodeficiency virus
HRCA	Hyper branched rolling circle amplification
HRP	Horseradish peroxidase
HT	Hypoxanthine, thymidine (respect to antibody production)
HT	Feedback voltage (respect to circular dichroism)
IAP	Intrapartum antibiotic prophylaxis
Ig	Immunoglobulin
IL-6	Interleukin 6
IMP	Inosine monophosphate
IPTG	Isopropyl $\beta$ -D-1 thiogalactopyranoside
KLH	Keyhole limpet haemocyanin
LAMP	Loop mediated isothermal amplification
LB	Luria-Bertani broth
LRCA	Linear rolling circle amplification
LSDA	Linear strand displacement amplification
MBP	Maltose binding protein
NAAT	Nucleic acid amplification test
NASBA	Nucleic acid sequence based amplification
NC	Negative control
NICE	National Institute for Health and Care Excellence
NPV	Negative predictive value
NRAMP	Natural resistance-associated macrophage protein
OD	Optical density
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline, 0.05% tween 20
PC	Positive control
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PGRCA	Primer generation rolling circle amplification
pI	Isoelectric point

PMNL	Polymorphonuclear leukocytes
PPV	Positive predictive value
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
RPA	Recombinase polymerase amplification
RT-PCR	Reverse transcriptase polymerase chain reaction
<i>S. species</i>	Streptococcus species
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SELEX	Systematic evolution of ligands by exponential enrichment
SrtA	Sortase A
SSB	Single stranded binding protein
TBE	Tris-boric acid-EDTA buffer
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TEV protease	Protease from the tobacco etch virus
TK	Thymidine kinase
TM	Transmembrane
TMB	3, 3', 5, 5'-tetramethylbenzidine
TMP	Thymidine monophosphate
UMP	Uridine monophosphate
UTI	Urinary tract infection
UV	Ultraviolet

---

## Chapter 1 *Streptococcus agalactiae*: Bacteria, Disease and Diagnostics

---

### 1.1 *Streptococcus agalactiae* bacterium

#### 1.1.1 *Streptococcus agalactiae*

*Streptococcus agalactiae* is the causative agent of group B streptococcal (GBS) disease. In healthy individuals it acts as a commensal bacteria and a constituent of the normal vaginal and intestinal microbiota. However the bacterium causes severe symptoms, such as pneumonia, meningitis and sepsis, in infants, immunocompromised patients and the elderly and is a leading cause of neonatal death. (Heath *et al.*, 2004; Edwards and Baker, 2005; Skoff *et al.*, 2009) *S. agalactiae* is a facultative anaerobe, (Johri *et al.*, 2007) and is therefore able to survive in a variety of environments. *S. agalactiae* has been isolated from the rectum, (Platt *et al.*, 1995) throat, vagina, wounds, blood, cerebrospinal fluid and urine. (Mhalu, 1976)

As well as a human pathogen, *S. agalactiae* is also a leading cause of bovine mastitis (Keefe, 1997) and is known to be pathogenic in cats, dogs, monkeys, (Lammler *et al.*, 1998) fish, seals, dolphins, bullfrogs (Delannoy *et al.*, 2013) and crocodiles. (Bishop *et al.*, 2007)

Streptococci are gram positive bacteria; the plasma membrane is surrounded by a peptidoglycan cell wall which consists of glucosamine and muramic acid sugars. A polysaccharide capsule envelopes the bacterial cells, made up of repeating sugar groups. Although surrounded by the capsule, cell wall associated proteins may be exposed to the outer surface of the bacteria. Early experiments showed antisera produced in response to injection with a single serotype of GBS, protected mice against infection with a second serotype of GBS. As these strains of GBS did not share the same capsule type, the antigen must have been a non-capsular component, shared by the two strains and exposed to the bacterial surface. (Lancefield *et al.*, 1975) More recently antibodies raised against cell wall associated Sip protein were shown to be

immunoprotective against GBS challenge across multiple serotypes. (Brodeur *et al.*, 2000) This is significant as the capsule does not block antibody binding to surface proteins. This may be because the capsule is narrow *in vivo*, because there is a degree of flexibility and mobility or because it is only present at particular stages of bacterial cell growth. (Lindahl *et al.*, 2005) A schematic of a gram positive bacterial cell is shown in figure 1.1.

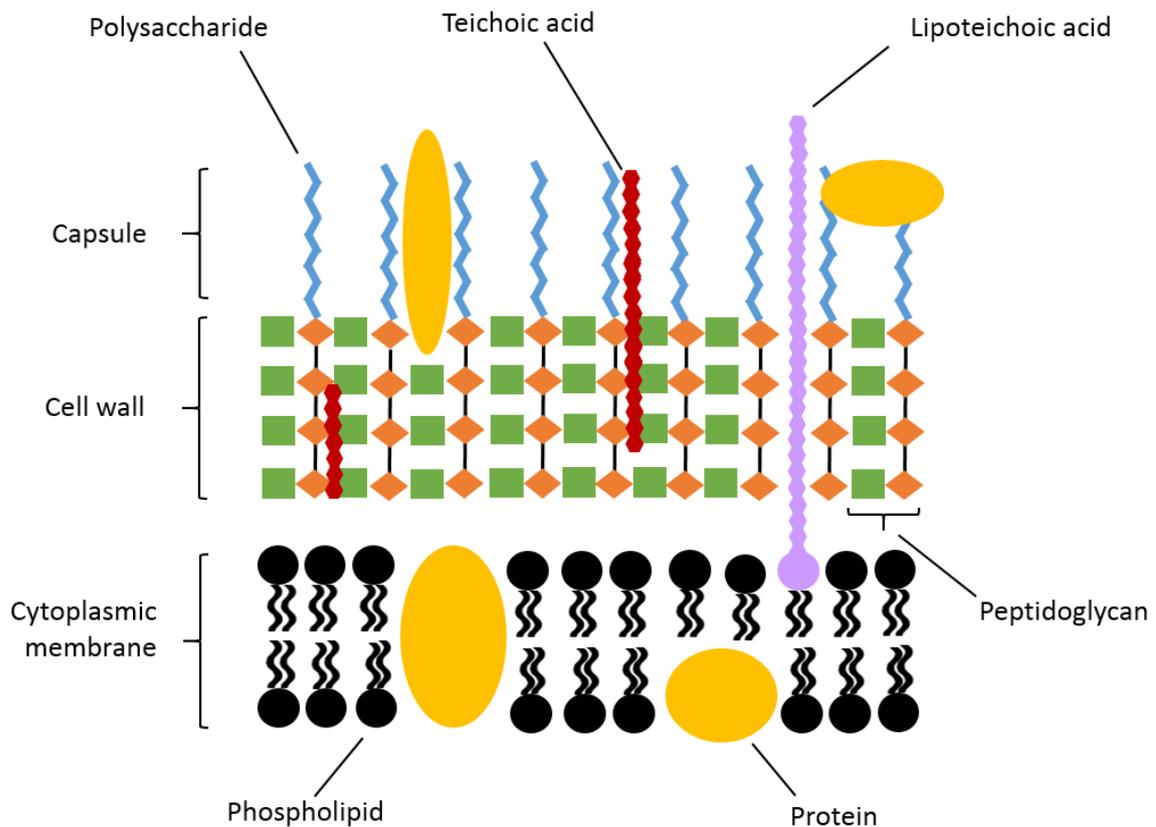


Figure 1.1 Schematic of gram positive bacteria

The figure shows the structure of a gram positive bacteria consisting of a cytoplasmic membrane, cell wall and capsule. Based on figures from (Fischetti, 2006) and (Murphy, 2011).

### 1.1.2 Serotypes

Most clinically isolated strains of *S. agalactiae* are encapsulated. (Benson *et al.*, 2002) *S. agalactiae* strains are classified into serotypes, the polysaccharide capsule of each serotype consists of a distinct monosaccharide. These are composed of galactose, glucose and N-acetylneuraminic acid in various conformations, with some monosaccharides containing N-acetylglucosamine or rhamnose. (Cieslewicz *et al.*, 2005) There are 10 variations, all of which have been isolated in human disease. (Cieslewicz *et al.*, 2005; Slotved *et al.*, 2007) The genetic variation between strains has been analysed through comparative genome hybridisation using 20 *S. agalactiae* strains. 18% of genes found in strain 2603 V/R, serotype V, were not found in all strains. This figure increased to 38% when only the genes unique to GBS were considered. (Tettelin *et al.*, 2002) There is evidence to suggest the expression of certain surface proteins is associated with particular serotypes. (Johnson and Ferrieri, 1984; Hauge *et al.*, 1996; Areschoug *et al.*, 1999; Lindahl *et al.*, 2005) This linkage can also be observed at the DNA (deoxyribonucleic acid) level. (Creti *et al.*, 2004; Shabayek *et al.*, 2014) In a study of the relationships between strains, profile clustering based on the sequences of 19 genes from 11 strains of GBS was performed. The results showed both serotype-dependent and -independent gene clustering. Three strains of serotype III were found to exhibit similar DNA sequences, whilst two strains within serotype Ia showed differential gene sequence patterns. The latter may be due to horizontal gene transfer. (Tettelin *et al.*, 2002) This partial correlation between serotype and DNA sequence was also observed in other studies. (Creti *et al.*, 2004; Shabayek *et al.*, 2014) The most common serotypes isolated from human infections are III, Ia, V, Ib and II, prevalence of each varies globally and over time. (Berg *et al.*, 2000; Ippolito *et al.*, 2010; Soares *et al.*, 2013; Vinnemeier *et al.*, 2015)

### 1.1.3 Related bacteria belonging to the genus streptococcus

Streptococci can be divided into pyogenic, and non-pyogenic streptococci. The former group mainly comprises  $\beta$ -haemolytic bacteria, whilst the latter consists of  $\alpha$ - and non-haemolytic bacteria. Pyogenic bacteria can be further classified by the presence of Lancefield antigens, see table 1.1. *Streptococcus agalactiae* is the only species to contain type B antigen, and this can therefore be used to identify the species. The group A antigen is found in *S. pyogenes*, *S. dysgalactiae* subspecies *equisimilis*, and

*S. anginosus* group. The group C antigen is predominantly found in zoonotic species but can also be found in *S. dysgalactiae* subspecies *equisimilis*. There are many more Lancefield antigens. (Spellerberg and Brandt, 2011)

Lancefield antigen	Species
<b>A</b>	<i>S. pyogenes</i> , <i>S. dysgalactiae</i> subspecies <i>equisimilis</i> , <i>S. anginosus</i> group
<b>B</b>	<i>S. agalactiae</i>
<b>C</b>	<i>S. dysgalactiae</i> subspecies <i>dysgalactiae</i> , <i>S. dysgalactiae</i> subspecies <i>equisimilis</i> , <i>S. equi</i> subspecies <i>equi</i> , <i>S. equi</i> subspecies <i>zooepidemicus</i> , <i>S. anginosus</i> group
<b>E</b>	<i>S. porcinus</i>
<b>F</b>	<i>S. anginosus</i> group
<b>G</b>	<i>S. dysgalactiae</i> subspecies <i>equisimilis</i> , <i>S. canis</i> , <i>S. anginosus</i> group
<b>L</b>	<i>S. dysgalactiae</i> subspecies <i>equisimilis</i>
<b>P</b>	<i>S. porcinus</i>
<b>U</b>	<i>S. porcinus</i>
<b>V</b>	<i>S. porcinus</i>
<b>None</b>	<i>S. anginosus</i> group, <i>S. porcinus</i>

**Table 1.1 Examples of Streptococcal species belonging to various Lancefield groups**  
Information obtained from (Spellerberg and Brandt, 2011).

There are several pathogenic species of significance within the Streptococcal genus. *S. pyogenes* causes group A streptococcal (GAS) disease which can lead to complications such as rheumatic fever (Bessen *et al.*, 1989) or acute glomerulonephritis. (Rodriguez-Iturbe and Batsford, 2007; Cunningham, 2008) Pneumonia, sepsis and meningitis can all result from infection with *S. pneumoniae*. Studies report 11-86% of the population to be infected with *S. pneumoniae* in the nasal cavity. (Bogaert *et al.*, 2004) *S. mutans* is a major cause of dental decay in humans, comprising 22-40% of the microbial flora found on carious teeth. (Loesche, 1986) *S. suis* is primarily a porcine pathogen, (Staats *et al.*, 1997) but human infections are rising following consumption of diseased meat. (Wertheim *et al.*, 2009)

138 streptococcal strains were analysed for the presence or absence of certain genes. A gene content matrix was used to calculate phylogenetic distances between each



#### 1.1.4 Virulence factors

*Streptococcus agalactiae* contains many virulence factors acting to evade the host's immune system, maximise dissemination, promote cell adhesion, adapt to alternative environments and function in several pathogenic pathways. Virulence factors include hyaluronidase,  $\beta$ -haemolysin, C5a peptidase, natural resistance-associated macrophage protein (NRAMP) homologues, sialic acid, D-alanylated lipoteichoic acids and CAMP factor, with some virulence factors performing multiple functions. (Landwehr-Kenzel and Henneke, 2014)

Hyaluronidase degrades the extracellular matrix of the host, clearing a path to enable the pathogen to spread. (Li and Jedrzejak, 2001) It also aids survival within phagocytes, which may be exploited as a transport system allowing rapid dissemination through the body. Hyaluronidase can act to reduce proinflammatory cytokine production. (Wang *et al.*, 2014b) A reduced inflammatory response in the vaginal tract of pregnant women, can allow the infection to ascend to the uterine tissue and increase the likelihood of preterm birth. (Vornhagen *et al.*, 2016)

$\beta$ -haemolysin is a pore-forming cytotoxin which causes damage to lung tissue. (Nizet *et al.*, 1996; Pritzlaff *et al.*, 2001) Septic shock, a condition in which an infection causes blood pressure to drop to a critical level, is dependent on nitric oxide production.  $\beta$ -haemolysin activates inducible nitric oxide synthase in murine macrophages, (Ring *et al.*, 2000) it can also initiate apoptosis in hepatocytes and may be responsible for liver failure. (Ring *et al.*, 2002)

C5a is a chemoattractant which draws polymorphonuclear leukocytes (PMNL) to the site of infection. C5a peptidase cleaves C5a at the PMNL binding site, reducing the chemotactic gradient. C5a peptidase is highly conserved between group A and group B streptococci, illustrating its essential role. In *S. pyogenes*, C5a peptidase deficient mutants were cleared from the air sac more efficiently than wild-type variants; indicating the role of C5a peptidase in immune evasion. (Ji *et al.*, 1996) C5a peptidase has also been proposed to act as an invasin. C5a peptidase deficient mutants were

able to invade an epithelial cell line to half the extent that wild-type GBS cells were capable of. (Cheng *et al.*, 2002b)

To survive the acidic environment of the vaginal tract, *S. agalactiae* contain pH-dependent NRAMP homologues which transport iron and manganese ions. Phagosomes employ reactive oxygen species to kill pathogenic bacteria; the removal of these toxic reactive oxygen species by superoxide dismutase is dependent on its cofactor, manganese. Iron is an essential cofactor in many enzymes. (Lisher and Giedroc, 2013) Therefore control of iron and manganese is essential for bacterial survival. A deletion mutant of the NRAMP *mntH*, resulted in reduced resistance to reactive oxygen species. This mutant also showed a reduced ability to adhere to the extracellular matrix at low pH. It is likely that this is dependent on the intracellular metal concentration, as adherence is recovered when the system is supplemented with manganese. (Shabayek *et al.*, 2016) Lactobacilli are the dominant genus found in the vaginal flora, they produce lactic acid, hydrogen peroxide and bacteriocin-like substances to restrict the growth of neighbouring bacteria. (Boris and Barbes, 2000) The NRAMP *mntH* mutant exhibited increased susceptibility to the anti-microbial agents produced by lactobacilli. This study has indicated the role of NRAMP homologues in pH-dependent host adhesion, resistance to oxidative stress and survival in the presence of bacteria found in the vaginal microbiome. (Shabayek *et al.*, 2016)

Sialic acid, D-alanylated lipoteichoic acid and CAMP factor function as virulence factors. Sialic acid is a component of the polysaccharide capsule, which inhibits the host's alternative complement pathway. (Edwards *et al.*, 1982) D-alanylation of lipoteichoic acids in the cell wall decreases cell wall flexibility and permeability, which in turn increases protection from the host cationic antimicrobial peptide response. (Saar-Dover *et al.*, 2012) CAMP factor is a toxin which causes the formation of transmembrane pores in erythrocyte membranes, (Lang and Palmer, 2003) via interaction with host glycosylphosphatidylinositol anchored proteins. (Lang *et al.*, 2007)

Several virulence factors, although not well studied in *S. agalactiae*, have homologues in other bacterial species. (Tettelin *et al.*, 2002)  $\alpha$ -enolase of *S. pyogenes* binds plasminogen, and is proposed to aid in tissue invasion. (Pancholi and Fischetti, 1998) In gram negative bacteria, such as *Pseudomonas aeruginosa*, secreted adenylate kinase induces macrophage cell death. Adenylate kinase catalyses the reversible formation of ADP (adenoside monophosphate) from ATP (adenosine triphosphate) and AMP (adenosine monophosphate). It is predicted that adenylate kinase binds to the macrophage cell surface and controls the levels of AMP, ADP and ATP to create a combination which is toxic to macrophages. (Markaryan *et al.*, 2001) Although streptococcal adenylate kinase functions as an ATPase, it is not known if it also acts as a virulence factor to induce macrophage cell death. (Thach *et al.*, 2014)

## 1.2 Group B streptococcal disease

19-28% of the population is colonised with *S. agalactiae* at any one time. (Benitz *et al.*, 1999) These commensal bacteria contain virulence genes (Otaguiri *et al.*, 2013) and therefore have the capability of causing severe or fatal infections within vulnerable patients. The large-scale colonisation of the population acts as a substantial reservoir of disease. This, coupled with the asymptomatic state of carriers, means it is difficult to prevent transmission. *S. agalactiae* can cause GBS disease in a variety of patient groups: neonatal infants, the elderly and the immunocompromised. Each is characterised by different methods of transmission, risk factors and clinical outcomes.

### 1.2.1 Early onset disease

Early onset GBS disease is defined as infection occurring within the first six days of life. Early onset GBS is caused by the transmission of bacteria from the mother to the infant prior to, or during delivery. During delivery neonates are colonised with maternal flora, the microbiota of an infant delivered vaginally and by caesarean differ considerably, with the dominant colonising bacteria being lactobacilli and staphylococci respectively. (Dominguez-Bello *et al.*, 2010) *Streptococcus agalactiae* is a resident of the vaginal flora, therefore vaginally born infants are at most risk of early onset GBS

disease. Risk factors for early onset disease include preterm birth, prolonged rupture of membranes, chorioamnionitis, maternal pyrexia during labour, maternal GBS vaginal or rectal colonisation or urinary tract infection (UTI) during pregnancy and the mother having had a previous child with neonatal GBS disease. (Hughes *et al.*, 2017) The mothers of 58% of infants who developed early onset GBS disease experienced prolonged membrane rupture, GBS carriage during pregnancy or preterm delivery. This figure increased to 73% of mother's whose infants died of GBS. (Heath *et al.*, 2004)

In recent years, incidence rates for early onset GBS disease of 0.37- 0.98 per 1000 live births have been reported. (Heath *et al.*, 2004; Puopolo *et al.*, 2005; Stoll *et al.*, 2011)) This is significantly reduced from rates observed prior to the recommendation of intrapartum antibiotic treatment, up to 5.46 per 1000 live births. (Benitz *et al.*, 1999) Early onset group B streptococcal disease can present in many ways, see table 1.2. When discharged from hospital, 7% of babies had a developed a disability or possible disability. (Heath *et al.*, 2004) The mortality rate associated with full term babies is 2%, (Schrag *et al.*, 2000) this increases to 23-30% for premature babies. (Schrag *et al.*, 2000; Brooks *et al.*, 2006)

	Prevalence
<b>Incidence</b>	0.37-0.98 per 1000 live births <sup>a, b, c</sup>
<b>Mortality</b>	2% (23-30% for premature babies) <sup>d, e</sup>
<b>Disability</b>	7% <sup>a</sup>
<b>Sepsis</b>	63% <sup>a</sup>
<b>Pneumonia</b>	26% <sup>a</sup>
<b>Meningitis</b>	11% <sup>a</sup>
<b>Focal infections</b>	0.6% <sup>a</sup>

**Table 1.2 Early onset GBS disease**

The table states the prevalence of early onset GBS disease and the likelihood of the conditions it causes. <sup>a</sup> (Heath *et al.*, 2004), <sup>b</sup> (Puopolo *et al.*, 2005), <sup>c</sup> (Stoll *et al.*, 2011), <sup>d</sup> (Schrag *et al.*, 2000) <sup>e</sup> (Brooks *et al.*, 2006)

### 1.2.2 Late onset disease

Infection of infants aged one week to three months is classified as late onset GBS. Late onset disease has a lower incidence rate, of 0.24 per 1000 live births, (Heath *et al.*, 2004) and the disease presents differently to early onset disease, with more cases of meningitis and focal infections, and fewer cases of sepsis and pneumonia, (Heath *et al.*, 2004) see table 1.3. A large nation-wide study of bacterial meningitis conducted in France was carried out over six years. Late onset GBS patients experienced a range of complications; 30% underwent shock, whilst 14% and 34% experienced seizures, prior to and during treatment. 32% required mechanical ventilation, 22% entered a coma and 17% of patients died from the infection. (Guilbert *et al.*, 2010) Other studies found late onset fatality to be 8% in the UK, (Heath *et al.*, 2004) and 2.8% in the US. (Schrag *et al.*, 2000) Bacteria were isolated from infected patients and analysed for serotype. 81% of isolated bacteria were found to be serotype III, (Guilbert *et al.*, 2010) in comparison to 38% for early onset disease. (Weisner *et al.*, 2004)

Prevalence	
<b>Incidence</b>	0.24 per 1000 live births <sup>a</sup>
<b>Mortality</b>	2.8-17% <sup>a, b, c</sup>
<b>Sepsis</b>	41% <sup>a</sup>
<b>Pneumonia</b>	8% <sup>a</sup>
<b>Meningitis</b>	43% <sup>a</sup>
<b>Focal infections</b>	7% <sup>a</sup>

**Table 1.3 Late onset GBS disease**

The table states the prevalence of late onset GBS disease and the likelihood of the conditions it causes. <sup>a</sup> (Heath *et al.*, 2004) <sup>b</sup> (Guilbert *et al.*, 2010) <sup>c</sup> (Schrag *et al.*, 2000)

In contrast to early onset disease which is vertically transmitted during delivery, late onset disease may derive from a number of sources. Environmental sources have been implicated in cases of nosocomial infections. (Al-Maani *et al.*, 2014) In around 50% of cases, the serotype of the neonatal infectious agent matches that of the intrapartum maternal coloniser. (Dillon *et al.*, 1987) However it is not definitive that the mother is the source of transmission.

In identifying methods of transmission that occur following birth, infants delivered by caesarean are particularly useful to study as colonisation with vaginal bacteria is unlikely. An association between late onset disease and breast milk infection has been suggested. *S. agalactiae* infection of breast milk was found in 25% of mothers with infants with late onset disease, both in those presenting with mastitis and those not. (Berardi *et al.*, 2013) Infants born to human immunodeficiency virus (HIV)-positive mothers have increased risk of contracting late onset disease. Reasons for this may include placental transfer of drugs to treat the maternal HIV infection, reduced placental transfer of maternal antibodies, reduced socioeconomic status or lack of breastfeeding. (Epalza *et al.*, 2010) Neonatal viral infection has been proposed to be associated with late onset disease. In one study, all GBS positive infants had viral symptoms, or confirmed viral infection, determined by virus isolation or detection of viral induced cytokines. (Raymond *et al.*, 2007)

Premature birth is a significant risk factor. Ultra-late onset GBS is the infection of infants over three months of age, and is extremely rare. 32% of infants who contracted ultra-late onset GBS were born under 32 weeks gestation, in comparison to 7% of those who developed late onset disease. It has been suggested that levels of maternal antibodies are responsible. (Guilbert *et al.*, 2010) Placental transfer of maternal antibodies that are capable of protecting against disease starts from 28-32 weeks gestation. Premature birth results in lower antibody transfer and reduced protection against disease. Antibody titres decrease following birth, according to their specific half-lives, (van den Berg *et al.*, 2011) therefore preterm infants with low GBS antibody levels at birth become increasingly more susceptible to GBS disease.

### 1.2.3 Immunocompromised

Prevalence of group B streptococcal disease in adults is increasing. In one study, the incidence rate increased from 3.6 to 7.3 cases per 100,000 people from 1990 to 2007. In patients under 65 years old, skin and soft tissue infections are the most common presentation. (Skoff *et al.*, 2009) Wounds and ulcers present easy routes of entry as the skin is broken, severe infections can lead to necrotizing fasciitis and toxic shock

syndrome. Other presentations include respiratory infections, genitourinary infections and joint and bone infections. (Sendi *et al.*, 2008)

Adult GBS infection is heavily associated with pre-existing conditions, such as diabetes mellitus, obesity, cancer, cardiovascular disease, renal disease, liver disease, chronic obstructive pulmonary disease, neurological diseases and immunosuppression. Previous studies of invasive GBS infection have found between 88% (Skoff *et al.*, 2009) and 100% (Edwards *et al.*, 2016) of patients to exhibit one or more of these conditions.

44-59% of GBS infected adults have diabetes. (Skoff *et al.*, 2009; Edwards *et al.*, 2016) Diabetic patients are more susceptible to microbial infections as a hyperglycaemic environment promotes disruption of the immune system. There is a reduction in inflammatory cytokine secretion, antioxidant response, leucocyte mobility and phagocytosis of infectious organisms. Treatment of diabetic complications adds to this as the number of medical interventions a patient experiences also increases the risk of acquiring infectious disease. (Casqueiro *et al.*, 2012) Commonly, diabetic patients with GBS disease present with skin, soft tissue or bone infection, with foot infections being particularly common. (Yanai *et al.*, 2012) Rare complications within diabetic patients include necrotising pneumonia, (Pacha *et al.*, 2017) skeletal muscle abscesses (Panikkath *et al.*, 2016) and abdominopelvic abscesses. (Ulett *et al.*, 2012)

#### 1.2.4 Elderly

Bacteremia is the most common presentation of GBS disease in patients over 65. (Skoff *et al.*, 2009) In comparison with infections of younger adults, elderly patients have an increased likelihood to present with urinary tract infections or pneumonia. The fatality rate of elderly patients diagnosed with GBS was 14-15% between 1997 and 2003. Risk factors include being a nursing or care home resident, being bedridden, or possessing a catheter. Infections are often diagnosed later in elderly patients partly due to a decreased ability to initiate a fever. Dementia decreases the capability of the individual to recognise symptoms of infectious diseases, and patients are therefore unable to report these to clinicians. In the elderly population there is also an increased

prevalence of underlying medical conditions in comparison to the general adult population. (Edwards and Baker, 2005) As previously discussed these play a significant role in susceptibility to GBS disease.

#### 1.2.5 Worldwide disease

Most GBS studies have been carried out in developed countries; GBS disease is three times as fatal in low income countries. (Edmond *et al.*, 2012) Similar rates of maternal *S. agalactiae* carriage are reported in developing countries such as North Africa, India, Pakistan and Saudi Arabia, (World Health Organization, 2006) however fewer cases of GBS disease are reported. This may be due to differences in transmission frequencies or a reduced ability to monitor infection on account of insufficient diagnostic techniques, undiagnosed causes of early death, or the larger number of births occurring away from healthcare sites. (Johri *et al.*, 2006) The true burden of global GBS disease is unknown.

### 1.3 Treating and preventing GBS disease

#### 1.3.1 GBS vaccine development

Early onset GBS disease can be prevented by administering antibiotics intravenously to the mother at the onset of labour, thus reducing the likelihood of transmission to the neonate. (Hughes *et al.*, 2017) However this is only capable of preventing early onset GBS disease. A vaccination programme may also be able to prevent late onset disease and disease in adults.

The outermost layer of the *S. agalactiae* bacterium is a polysaccharide capsule. Serotype-specific capsular polysaccharides, conjugated to carrier proteins for increased immunogenicity, have been used to confer protection to GBS challenge in animal models. (Wessels *et al.*, 1990; Paoletti *et al.*, 1992; Paoletti *et al.*, 1994; Paoletti and Kasper, 2002) In clinical trials, vaccination caused a rise in serum Ig (immunoglobulin) G levels and *in vitro* opsonophagocytosis of *S. agalactiae*. (Baker *et al.*, 1999; Baker *et al.*, 2000; Baker *et al.*, 2004; Baker *et al.*, 2007) This effect was

serotype specific, therefore a multivalent vaccine would be required for sufficient coverage of clinical isolates. A clinical trial performed in South Africa investigating the effect of multivalent immunisation (serotypes: Ia, Ib and III) in pregnant women resulted in raised antibody levels against all three serotypes in the mother and infant. (Madhi et al., 2016) In the UK 85% of colonising *S. agalactiae* strains are serotype III, Ia or V. (Weisner et al., 2004) As serotype distribution varies geographically, different combinations of vaccines may be required to cover the majority of strains found in different areas. (World Health Organization, 2006)

To eliminate serotype-specificity, an *S. agalactiae* protein could function as the immunogen. To be effective as an *S. agalactiae* specific vaccine target, the protein would need to be conserved in all strains, surface exposed and be immunogenic. (Nuccitelli et al., 2015) Proteins such as Sip and C5a peptidase have been investigated as vaccine targets. The Sip protein was conserved and expressed in all serotypes tested, and when injected into mice was protective against multiple serotypes of GBS. (Brodeur et al., 2000) C5a peptidase vaccination also gave protection against GBS challenge, independent of serotype. (Cheng et al., 2002a) Surface exposed proteins can be identified by surfome protease digestion followed by mass spectrometry. In the study of *S. agalactiae*, both Sip and C5a peptidase were found to be surface exposed. (Doro et al., 2009) It is likely that bioinformatic techniques able to screen multiple genomes will direct the discovery of vaccine targets. (Johri et al., 2006)

### 1.3.2 Antibiotic treatment

As there is no currently available vaccine, the strategy for early onset GBS disease prevention is maternal antibiotic treatment. Repeat doses are given during labour until the infant is delivered. (Hughes et al., 2017) Since intrapartum antibiotic prophylaxis (IAP) became the recommended treatment in the US in the 1990s, mortality due to GBS disease decreased at a faster rate (5% per year) than previously (3% per year). (Lukacs et al., 2004) The efficacy of IAP treatment, if given more than two hours prior to delivery, is 89%. (Lin et al., 2001) The advised minimal treatment period is four hours. Antibiotic treatment of mothers during pregnancy causes no reduction in GBS colonisation at the time of delivery. (Hughes et al., 2017)

If an infant presents with early onset symptoms, penicillin and gentamicin should be administered. (Hughes *et al.*, 2017) Antibiotics are also prescribed for treatment of adult GBS disease. Additional treatments such as surgery may be necessary in cases of skin or bone infection. (Farley, 2001)

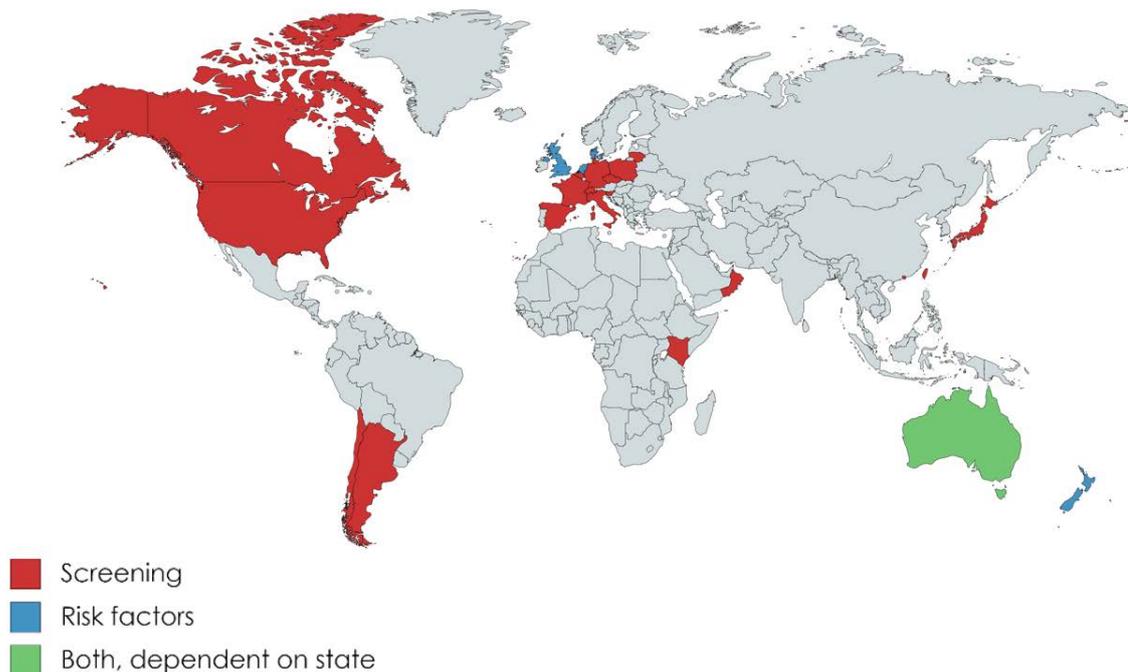
Benzylpenicillin is the first choice antibiotic. There are no reports of penicillin resistance in the UK, (Lamagni *et al.*, 2013) however decreased susceptibility has been reported in the US (Dahesh *et al.*, 2008) and East Asia. (Hsueh *et al.*, 2001; Chu *et al.*, 2007; Kimura *et al.*, 2008) 0.7-4% of penicillin treatments result in allergic reactions, with anaphylaxis being a rare but severe side effect. (Petri, 2011) In cases of penicillin allergy, clindamycin or erythromycin are often administered. Clindamycin resistance rose from a frequency of 3% to 9% from 1998 to 2010; erythromycin resistance increased from <3% to 15% from 1991 to 2010. (Lamagni *et al.*, 2013) An association between antibiotic treatment and infant colonisation with antibiotic resistant bacteria has been suggested, (Terrone *et al.*, 1999), however statistical significance was not observed by other groups. (Jaureguy *et al.*, 2004; Nogacka *et al.*, 2017)

There are some concerns over the danger of intrapartum antibiotic treatment to the long term health of the infant. A study concluded there was a marked increase in the percentage of children with functional impairments and cerebral palsy at seven years of age, born to mothers treated with erythromycin during labour, in comparison to those who were given a placebo. In this study mothers were treated with a low dosage of oral, broad spectrum antibiotics over 10 days (or until delivery). (Kenyon *et al.*, 2008) This contrasts the strategy to prevent GBS transmission where high dosage, targeted antibiotics are administered intravenously over a short period of time. The findings of the study may not be applicable to alternative antibiotic treatment programmes, and do not impact the recommendation for antibiotics to be used in GBS prevention. (McCartney *et al.*, 2002)

### 1.3.3 Maternal screening/risk- based treatment

In the US, Canada, Japan and many European countries, an antenatal screening programme is offered at 35-37 weeks gestation. (Burns and Plumb, 2013) However, countries such as the United Kingdom, Denmark, the Netherlands and New Zealand do not recommend universal screening, (Melin and Efstratiou, 2013) see figure 1.3. The UK National Screening Committee (2017) recently concluded that universal screening would not be appropriate for several reasons. There is no technique to identify which infants would be severely affected by GBS. The maternal GBS status fluctuates, therefore mothers who are colonised with *S. agalactiae* at the time of screening may not be positive during delivery and vice versa. Screening may increase unnecessary antibiotic treatment, which may be harmful to the infant, mother or to antibiotic stewardship. (UK National Screening Committee, 2012)

The UK follows a risk factor-based treatment plan; risk factors include preterm birth, maternal pyrexia in labour and prolonged membrane rupture. (UK National Screening Committee, 2017) However 33-42% of infants who present with early onset GBS disease are born to mothers who present no risk factors, (Heath *et al.*, 2004; Vergnano *et al.*, 2010) therefore the opportunity to prevent these infections is missed. In a screening trial conducted in the UK, cases of early onset disease fell by 84% in comparison to the pre-screening period, (Rao *et al.*, 2017) supporting the development of a universal screening programme.



**Figure 1.3 Worldwide approach to early onset GBS prevention**

Countries which conduct routine screening for GBS disease are shown in red, those which treat based on risk factors are shown in blue. In Australia prevention differs between states, shown in green. Information from: (Burns and Plumb, 2013; Melin and Efstratiou, 2013; Homer *et al.*, 2014; Cho *et al.*, 2017) Map template from: (mapchart.net, 2017).

The selection of treatment plans is heavily influenced by the cost of screening and preventative maternal treatment balanced with the potential savings associated with reduced time in hospital and lower costs of infant treatment. Costs outside of the health system should also be considered, such as loss of parental productivity, and therefore tax, and future schooling requirements associated with disability. Any scheme which reduced costs to a larger extent than the increase in diagnostic expenditure would be deemed economically viable, and may be suitable for use in the NHS. The current practice of treating women exhibiting risk factors is considered to not be cost effective. (Colbourn *et al.*, 2007; Kaambwa *et al.*, 2010) In one study, the most economically favourable strategy was determined to be universal treatment for all women. (Kaambwa *et al.*, 2010) This directly opposes a Public Health England campaign to decrease unnecessary antibiotic usage in order to combat antibiotic resistance. (Public Health England, 2017) Eliminating this option, the most cost viable strategy is the use

of enriched culture media for screening. (Kaambwa *et al.*, 2010) An alternative study suggested the treatment of women with preterm or high risk conditions coupled with screening of women at low risk of the disease to be the most cost effective approach. (Colbourn *et al.*, 2007)

#### 1.4 Diagnosis of *S. agalactiae* infection

The UK does not support specific GBS testing, however if there are symptoms of bacterial infection a standard direct plate test is often carried out. This non-specific test enables the identification of a number of bacterial species, however is associated with a high rate of false negative results, with 50% of infected samples reported as negative. (Centers for Disease Control and Prevention, 1999) Diagnostic tests with the aim of specific detection of *S. agalactiae* infections are described below.

##### 1.4.1 Current GBS diagnostic tests

*S. agalactiae* colonisation or infection is most commonly diagnosed by growth on blood agar plates, the CAMP test, or more recently with chromogenic media. As *S. agalactiae* strains are  $\beta$ -haemolytic, colonies form a small area of haemolysis on blood agar plates, characterised by a transparent zone of complete lysis. This differs from  $\alpha$ -haemolytic species which exhibit partial haemolysis, producing a green zone. (Wannamaker, 1965) Microbial culture techniques still dominates the field of infectious disease diagnostics, despite the lengthy process of 24-72 hours.

The sensitivity and specificity of these techniques can be enhanced by *S. agalactiae* enrichment. Incubation for 18-24 hours in selective medias such as TransVag (Todd-Hewitt broth supplemented with gentamicin and nalidixic acid) and Lim broth (Todd-Hewitt broth supplemented with colistin and nalidixic acid) promotes the growth of *S. agalactiae* in preference to other bacterial species. (Verani *et al.*, 2010) Subsequent tests are performed on specimens with an enhanced *S. agalactiae* content. In comparison to direct culture, prior enrichment in selective broth increased the number of positive samples detected in one study by 105%. (Platt *et al.*, 1995) The major

disadvantage of enrichment is the additional time required until results become available.

*S. agalactiae* haemolysis can be enhanced by incubation with *Staphylococcus aureus*  $\beta$ -haemolysin through interaction with *S. agalactiae* CAMP factor. The interaction between *S. aureus* and *S. agalactiae* was first discovered by Christie, Atkins and Munch-Petersen (Christie *et al.*, 1944) and the resulting test is named after them.  $\beta$ -haemolysin can be supplied in a disc or through incubation with *S. aureus*. (Wilkinson, 1977) This method is able to detect non-haemolytic strains; 5% of strains isolated from colonising bacteria are non-haemolytic. (Rodriguez-Granger *et al.*, 2015) Microbial culture of enriched samples, and the CAMP test are considered gold standard diagnostic tests to which new diagnostic tests are measured against to determine test accuracy. (de-Paris *et al.*, 2011)

Various diagnostic tests for GBS detection are described below, with key features summarised in table 1.4. Results were compared to a gold standard technique for calculation of test sensitivities and specificities. The sensitivity of a diagnostic test is the fraction of individuals with a disease that are identified by the test. The specificity of a diagnostic test is the probability that the diagnostic test will produce a negative result for individuals who do not have the disease. The positive predictive value (PPV) determines the likelihood of a patient having the disease if the test result is positive. The negative predictive value (NPV) states the likelihood of a negative result arising from a true lack of disease. Predictive values of diagnostic tests are dependent on disease prevalence in the sample population. In populations of high disease prevalence, the NPV is low and diagnostic tests with high specificities are required. Conversely, at low disease prevalence PPV is low, and diagnostic tests with increased sensitivities are needed. (Wong and Lim, 2011) As clinical validation tests were not performed in the same population, predictive values cannot be directly compared, and therefore are not reported below.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}}$$

$$\text{PPV} = \frac{\text{Sensitivity } \pi}{\text{Specificity } \pi + (1 - \text{Specificity})(1 - \pi)}$$

$$\text{NPV} = \frac{\text{Specificity } (1 - \pi)}{\text{Specificity } (1 - \pi) + (1 - \text{Sensitivity})\pi}$$

$\pi$  is defined as the prevalence in the population. Equations from (Wong and Lim, 2011)

Method	Enrichment	Time	Sensitivity / %	Specificity / %	Reference
<b>Pigmented agar / broth</b>	No	18-24 hours (if positive), 48 hours (if negative)	94.1 - 97.8 (agar), 85.5 - 96.6 (broth)	100	(Votava <i>et al.</i> , 2001; Tazi <i>et al.</i> , 2008)
<b>Latex agglutination</b>	No	15 minutes	13.3-100 (dependent on cfu number)	100	(Xie <i>et al.</i> , 2016)
<b>Optical immunoassay</b>	No	35 minutes	7.1 - 98.4 (dependent on colonisation levels)	41.5 - 98.1 (dependent on colonisation levels)	(Baker, 1996; Thinkhamrop <i>et al.</i> , 2003; Aziz <i>et al.</i> , 2005; Daniels <i>et al.</i> , 2009)
<b>Enzyme immunoassay</b>	No	10 minutes	16.1-20.7	98.8-99.8	(Baker, 1996)
<b>DNA probes</b>	Yes	120 minutes	90.1 - 97.4	97.5 - 100	(Williams-Bouyer <i>et al.</i> , 2000; Montague <i>et al.</i> , 2008; Peltroche-Llacsahuanga <i>et al.</i> , 2010; Wilson <i>et al.</i> , 2010)
<b>Nucleic acid amplification tests</b>	No (Yes)	80-210 minutes	90.3 (92.5 -100)	99 (92.5 -99.3)	(Goodrich and Miller, 2007; Block <i>et al.</i> , 2008; Ellem <i>et al.</i> , 2017)

**Table 1.4 Types of diagnostic tests for *S. agalactiae* detection**

This table states the sensitivities, specificities, diagnosis time and requirement for enrichment for a variety of types of GBS diagnostic test. Abbreviation: cfu (colony forming unit)

Chromogenic media or agar is used to facilitate easy identification of *S. agalactiae*. In a starch rich media under anaerobic conditions, *S. agalactiae* produces an orange coloured pigment, granadaene. (Votava *et al.*, 2001; Tazi *et al.*, 2008) A chromogenic media causes *S. agalactiae* to appear as pink colonies in an aerobic environment. Strains isolated from invasive disease were subjected to testing, 5.4% of these samples were found to be non-pigmented, however all strains were able to be detected by the observation of pink colonies in chromogenic media. (Votava *et al.*, 2001)

Latex agglutination is used in streptococcal serogrouping, and is based on detection of Lancefield carbohydrate antigens. (Davies *et al.*, 2003) It can be performed directly from patient samples and takes around 15 minutes. Latex particle-conjugated antibodies bind to *S. agalactiae*-specific antigen, causing latex particles to agglutinate. (Ghaddar *et al.*, 2014) This technique is highly sensitive in moderate-heavy colonisation (>50 colony forming units (cfu) per swab), however sensitivity drops to 66.7% and 13.3% for samples with 10-50 cfu per swab and <10 cfu per swab respectively. (Xie *et al.*, 2016) When implemented following enrichment, high sensitivities can be achieved. (Davies *et al.*, 2003) Other polysaccharide antigen detection diagnostics, including optical and enzyme immunoassays, have been developed, however they lack the sensitivity and specificity to be used clinically. (Baker, 1996; Thinkhamrop *et al.*, 2003; Aziz *et al.*, 2005; Daniels *et al.*, 2009) Carbohydrate-protein interactions are typically weak, (Fernandez-Alonso *et al.*, 2012; De Schutter and Van Damme, 2015) which may explain the low sensitivities observed for patient samples with light colonisation.

Further techniques include the use of nucleic acid probes which bind to specific regions within *S. agalactiae* ribosomal RNA (ribonucleic acid). The probes are labelled with fluorescent (Montague *et al.*, 2008; Peltroche-Llacsahuanga *et al.*, 2010; Wilson *et al.*, 2010) or chemiluminescent molecules. (Williams-Bouyer *et al.*, 2000) Although probe diagnostics exhibit promising sensitivities and specificities, they are limited by the requirement for enrichment and long processing time of around two hours.

Nucleic acid amplification tests (NAATs) have been used for GBS diagnosis, often in the form of quantitative polymerase chain reaction (qPCR). The amplification of target specific DNA is measured in real time by the increase in fluorescent signal. NAATs are often associated with high sensitivities and specificities. In the 2000s, the use of DNA amplification directly from patient samples with no pre-enrichment stage was found to not be accurate enough for use in clinical practice. (Aziz *et al.*, 2005; Daniels *et al.*, 2009) There have been several NAATs which have gained FDA (US Food and Drug Administration) approval, table 1.5 details those approved since 2002.

Device name (Company)	Type	Sensitivity / %	Specificity / %	Enrichment	Time to result / minutes	Document number
<b>IDI-Strep B Assay (Infectio Diagnostic)</b>	qPCR	89-97.2	94-97.3	No	Not specified	K022504
<b>Xpert GeneXpert Dx System (Cepheid)</b>	qPCR	94.6	94.8	No	75	K060540
<b>Smart GBS, SmartCycle Dx System (Cepheid)</b>	qPCR	98.7 (81.6)	90.5 (96.4)	Yes (No)	75	K062948
<b>HandyLab GBS Assay (HandyLab)</b>	qPCR	95	96.7	Yes	120	K090191
<b>Illumigene Group B Streptococcus (Meridian Bioscience)</b>	LAMP	97.4	92.3	Yes	60	K112125
<b>BD Max GBS Assay (Becton Dickinson)</b>	qPCR	Not specified	Not specified	Yes	Not specified	K111860
<b>Illumigene Group B Streptococcus (Meridian Bioscience)</b>	LAMP	98.6	93.2	Yes	60	K121044
<b>XPRT GBS LB GeneXpert Dx Systems (Cepheid)</b>	qPCR	99	92.4	Yes	55	K121539
<b>AmpliVue GBS Assay (Quidel)</b>	HDA	99.5	92.7	Yes	75-90	K133503
<b>Portrait GBS Assay (Great Basin)</b>	PCR, chip	97.9	96	Yes	90	K143312
<b>ARIES GBS Assay (Luminex)</b>	qPCR	96.1	91.4	Yes	Not specified	K162772
<b>GenePOC GBS LB (GenePOC)</b>	qPCR	95.9	95.5	Yes	120-180	K170557

**Table 1.5 FDA approved NAAT-based GBS diagnostic tests**

FDA approved NAATs for GBS detection were evaluated for reported sensitivity, specificity, detection time and requirement for enrichment. Types of NAAT: qPCR (quantitative PCR); LAMP (loop mediated isothermal amplification); HDA (helicase dependent amplification). Document numbers from FDA premarket notification files, accessed from (US Food and Drug Administration, 2017).

For marketing within Europe, diagnostic tests must comply by directive 98/79/EC (*in vitro* diagnostic medical devices) and carry a CE (Conformité Européenne) mark. (The European Parliament and the Council of the European Union, 1998) Several of the FDA approved tests detailed in table 1.5 are currently CE-marked, including the Illumigene Group B Streptococcus test, (Meridian Bioscience, 2018) GenePOC's GBS

LB test, (GenePOC, 2018) the Amplivue GBS Assay (Quidel, 2018) and the Xpert GBS test. (Cepheid, 2018) Currently there is no publically available database of CE-marked medical devices therefore the above list is non-exhaustive. However by 2022 EU directive 2017/746 will be instated, which requires certificates of devices to be registered on a publically accessible database.

Cepheid's GeneXpert GBS is an example of a state of the art NAAT for GBS diagnosis. Cells are captured on a filter and lysed by sonication. The DNA is eluted and this solution rehydrates the lyophilised PCR reagents. Real time PCR is performed and fluorescence is measured. (US Food and Drug Administration, 2017) In 2015, the National Institute for Health and Care Excellence (NICE) published an innovation briefing discussing the use of the Xpert GBS for intrapartum testing. (Jones *et al.*, 2015) It reported that Xpert GBS is an accurate test able to provide results in 50 minutes, a vast improvement over microbial culture. The recommended minimal length of IAP treatment prior to delivery is four hours. This was achieved in 68% of cases diagnosed by Xpert GBS, an increase over the 64% treated successfully in cases diagnosed by antenatal microbial culture. In preterm births the GBS status of the mother was known at least four hours before delivery in 23% of cases diagnosed by bacterial culture, and in 74% of cases diagnosed by Cepheid's Xpert GBS. (de Tejada *et al.*, 2011) However, 15% of diagnoses resulted in an invalid result, and required re-testing. (Hakansson *et al.*, 2014) For intrapartum testing to be viable, diagnosis will have to be carried out at or close to the point of care, as transportation to a centralised test centre will reduce the number of patients able to receive timely results. The state of the art cartridge-based system requires very little operator input therefore could be transitioned out of the laboratory and into a healthcare setting. However the high running cost and initial expense of the base station may prohibit use in public healthcare systems. There are no economic studies which show applicability for use across the NHS.

#### 1.4.2 Rapid test

Current US guidelines advise maternal screening at 35-37 weeks gestation. (Verani *et al.*, 2010) The GBS status of the vaginal flora rapidly fluctuates; 24-30% of women testing positive during pregnancy will not be during labour, and 5-6% of women testing

negative will be positive at the time of delivery. (Public Health England, 2015b) Outdated test results may misinform clinicians, leading to a lack of necessary treatment or inappropriate antibiotic usage. A rapid test would report the current state of maternal colonisation, and therefore accurate risk of transmission to the neonate.

Although a screening programme is not currently recommended in the UK, Public Health England (2015a) has stated that a rapid point of care diagnostic test able to provide results in under two hours may be of interest. NICE has indicated the role of rapid GBS testing in antibiotic stewardship, reducing IAP prescriptions. (Jones *et al.*, 2015) Table 1.6 summarises in which clinical scenarios a rapid test may be useful.

Clinical scenario	Rapid test useful?
<b>No risk factors</b>	In screening programme
<b>GBS colonisation in previous pregnancy</b>	In screening programme
<b>Preterm delivery</b>	Yes; preterm delivery is a risk factor for neonatal GBS disease, conventional tests may not have been completed by this time point
<b>Preterm pre-labour rupture of membranes</b>	Yes; preterm delivery is a risk factor for neonatal GBS disease, conventional tests may not have been completed by this time point
<b>Symptoms of vaginal/ urinary infection during pregnancy</b>	Yes; rapid testing would allow earlier maternal treatment
<b>Caesarean section</b>	No; GBS transmission unlikely
<b>Previous infant with GBS disease</b>	No; antibiotics would be advised irrespective of results
<b>Prolonged rupture of membranes</b>	No; if evidence of infection broad spectrum antibiotics should be administered
<b>Pyrexia in labour</b>	No; broad spectrum antibiotics should be administered

**Table 1.6 The use of a rapid GBS test in various clinical scenarios**

Statements concerning preterm delivery, premature pre-labour rupture of membranes, caesarean section, previous infant with GBS disease, prolonged rupture of membrane and pyrexia in labour obtained from (Public Health England, 2015a).

There are several scenarios in which testing would not influence treatment plans. In instances of maternal pyrexia, or when there is evidence of infection in cases of prolonged membrane rupture, broad spectrum antibiotics would be prescribed

regardless of GBS status. Targeted antibiotics would be prescribed in cases where a mother had previously had a baby infected with GBS infection. However a rapid test would be beneficial for preterm births, where results from a potential traditional screening programme at 35-37 weeks may not be available. In cases of preterm pre-labour membrane rupture there is an increased association with early onset disease. Rapid testing of GBS status at the time of rupture would enable appropriate antibiotics to be given. A rapid test could be used in a universal screening programme or as a diagnostic test for specified clinical presentations. In countries that currently conduct a screening programme, replacement of traditional tests with a point of care test may provide a more accurate infection status, resulting in more appropriate care.

Although a less strict timescale is applicable for detection of adult disease, point of care testing would allow diagnosis within a single clinic visit, therefore required treatment could start immediately. As *S. agalactiae* is a dominant cause of early onset sepsis and meningitis, neonates presenting with these symptoms are often prescribed a suitable treatment for GBS, ampicillin and gentamicin, prior to diagnosis. However, infants with late onset sepsis or meningitis are often prescribed gentamicin and cloxacillin. (Sivanandan *et al.*, 2011) Once a GBS diagnosis is confirmed, treatment is altered to penicillin or ampicillin. Rapid diagnosis would allow for correct treatment to be administered initially.

### 1.5 Project aims

The aim of the project is to progress towards developing a rapid diagnostic test primarily for use in intrapartum diagnosis of maternal *S. agalactiae* colonisation. This would best inform clinical treatment of the mother and child to reduce the prevalence of early onset group B streptococcal disease. The project will investigate three methods of detecting *S. agalactiae*: using antibodies, isothermal DNA amplification and peptide aptamers.

1. Current enzyme- and optical-immunoassays based on the detection of *S. agalactiae*-specific polysaccharide are not sensitive enough to be used for clinical detection of GBS. Development of an immunoassay which recognises protein antigen may overcome the limitations associated with weak carbohydrate-protein affinity, and theoretically could exhibit increased sensitivity. This project looks to develop antibodies against *S. agalactiae* utilising a novel computer programme to identify peptide sequences unique to, and conserved within, *S. agalactiae* strains. The eventual aim would be to incorporate these antibodies into a diagnostic test able to detect all strains of *S. agalactiae* whilst exhibiting no cross-reactivity with other bacterial species.
2. Cepheid's GeneXpert GBS shows promising sensitivity and specificity whilst operating at a relatively fast turn-around time. However its high cost for the base station and cartridge replacements restricts its widespread use in healthcare. The use of isothermal DNA amplification for the detection of *S. agalactiae* will be considered. These tests have rapid turnaround times and do not require an expensive thermocycler, making them more amenable to point of care testing.
3. Although antibodies are commonly used for antigen detection in diagnostics, there are several other types of detection molecule. A novel peptide aptamer scaffold which functions as a reporter protein whilst specifically recognising the target of interest will be considered. This would reduce the steps required in a detection assay, beneficial for a rapid point of care test for *S. agalactiae* diagnosis.

The World Health Organisation created the term ASSURED, to describe the desired characteristics of a diagnostic test. An optimal diagnostic test would be affordable, sensitive, specific, user friendly, robust and rapid, equipment-free and deliverable to the point of need. (Mabey *et al.*, 2004) These criteria will be considered in progressing towards a point of care GBS diagnostic tool.

---

## Chapter 2 Methods

---

### 2.1 Materials

*Escherichia coli* BL21 *pLysS DE3* and LMG194 strains were obtained from Novagen and ATCC respectively. *S. agalactiae* strains were obtained from the Statens Serum Institut, Copenhagen and *S. pyogenes* strains were kindly donated by Dr Nicholas Jakubovics. The NS-1 cell line was obtained from ATCC. BALB/c mice were obtained from Charles River. Custom ordered primers were purchased from Integrated DNA Technologies or Life Technologies. Capture and reporter nucleic acid probes were purchased from Biomers. Peptides were purchased from Biomatik, Aequorin WT\*, Aequorin FLAG, SAG1474 and SAG1474-MBP encoding plasmids were manufactured by GeneArt, Life Technologies. *pBADHisA* was supplied by Invitrogen, *pET28a* by Novagen and *pMAL p5x* plasmid was purchased from New England Biolabs. Luria-Bertani broth and Bacto Todd-Hewitt media were supplied by Melford and BD respectively. Bugbuster master mix for cell lysis was supplied by Novagen. Ellman's reagent was purchased from Thermo Scientific. Maltose binding protein was produced by Acris, OriGene Technologies. GelRed nucleic acid stain was supplied by Biotium. UltraPure agarose 1000 was purchased from Invitrogen. Gel loading dye was supplied by Biorad, the 100 bp DNA ladder was purchased from Promega and the PageRuler prestained protein ladder, 10 to 180 kDa from Thermo Scientific. Mouse anti-histidine tag antibody was purchased from Abcam (MCA1396GA), horseradish peroxidase (HRP)-conjugated goat anti-mouse heavy and light chain specific secondary antibody was purchased from Abcam (ab97023). ECL western blotting substrate was supplied by Pierce, and TMB microwell peroxidase substrate by SureBlue. Maleimide activated 96 well plates were supplied by Pierce. ISOLATE II RNA mini kit and Tetro cDNA synthesis kit were supplied by Bioline. Qiagen supplied the DNeasy blood & tissue kit and QIAquick PCR purification kit. Go TaqFlexi PCR kit was supplied by Promega. The Inject maleimide activated KLH and BSA conjugation kits were supplied by Thermo Scientific. Basic recombinase polymerase amplification kits were purchased from TwistDx. All other chemicals were provided by Sigma.

## 2.2 General techniques

### 2.2.1 Bacterial cell growth

Details of bacterial strains can be found in table 2.1. *E. coli* strains were cultured in Luria-Bertani broth (LB) (10 g l<sup>-1</sup> tryptone, 5 g l<sup>-1</sup> NaCl, 5 g l<sup>-1</sup> yeast extract) at 37 °C. *S. agalactiae* and *S. pyogenes* strains were cultured in Todd-Hewitt media (3.1 g l<sup>-1</sup> heart infusion, 20 g l<sup>-1</sup> neopeptone, 2 g l<sup>-1</sup> dextrose, 2 g l<sup>-1</sup> NaCl, 0.4 g l<sup>-1</sup> disodium phosphate, 2.5 g l<sup>-1</sup> sodium carbonate), and incubated at 37 °C, 5% CO<sub>2</sub>. Bacteria were grown on plates consisting of 7.5 g l<sup>-1</sup> agar diluted in their respective medias. Bacterial cultures were supplemented with antibiotics if appropriate as detailed in table 2.2.

Species	Serotype	Origin
<i>Escherichia coli</i>	BL21	Novagen
<i>Escherichia coli</i>	LMG194	ATCC, 47090
<i>Streptococcus agalactiae</i>	Ia	Statens Serum Institut, Copenhagen, 84902
<i>Streptococcus agalactiae</i>	Ib	Statens Serum Institut, Copenhagen, 84903
<i>Streptococcus agalactiae</i>	II	Statens Serum Institut, Copenhagen, 84904
<i>Streptococcus agalactiae</i>	III	Statens Serum Institut, Copenhagen, 84905
<i>Streptococcus agalactiae</i>	IV	Statens Serum Institut, Copenhagen, 84906
<i>Streptococcus agalactiae</i>	V	Statens Serum Institut, Copenhagen, 84907
<i>Streptococcus agalactiae</i>	VI	Statens Serum Institut, Copenhagen, 84913
<i>Streptococcus agalactiae</i>	VII	Statens Serum Institut, Copenhagen, 84908
<i>Streptococcus agalactiae</i>	VIII	Statens Serum Institut, Copenhagen, 84909
<i>Streptococcus agalactiae</i>	IX	Statens Serum Institut, Copenhagen, 84910
<i>Streptococcus pyogenes</i>	M1 SF370	ATCC, 700294. Donated by Dr Nicholas Jakubovics.
<i>Streptococcus pyogenes</i>	M2 75245	Clinical isolate, wound infection, Newcastle University. Donated by Dr Nicholas Jakubovics.
<i>Streptococcus pyogenes</i>	M5 Manfredo58	Donated by Dr Nicholas Jakubovics.

**Table 2.1** Origin of *E. coli* and streptococcal strains

Plasmid	Antibiotic	Working concentration / $\mu\text{g ml}^{-1}$	Solvent
<i>pLysS</i>	Chloramphenicol	25	Ethanol
<i>pBAD HisA</i>	Ampicillin	100	dH <sub>2</sub> O
<i>pET28a</i>	Kanamycin	50	dH <sub>2</sub> O
<i>pMAL p5x</i>	Ampicillin	100	dH <sub>2</sub> O

**Table 2.2 Antibiotic concentrations**

Antibiotics were reconstituted in ethanol or dH<sub>2</sub>O (highly purified water), at the specified concentrations.

### 2.2.2 Bacterial cell storage

Bacterial cell stocks were frozen from cultures grown for 16-21 hours in a solution of 50% cell culture, 25% glycerol, 25% dH<sub>2</sub>O, with stocks kept at -80 °C and in liquid nitrogen. Bacteria were resuscitated from these frozen stocks when required.

### 2.2.3 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins based on molecular weight, as originally described by Laemmli. (Laemmli, 1970) In this technique SDS is added to the sample, linearising proteins and coating them in a uniform negative charge; the sample is then loaded onto a gel. An electric current is applied, attracting the negatively charged proteins to the base of the gel, towards the anode. The stacking gel acts to concentrate proteins in a small volume. The buffer contains glycine which, when it enters the gel, becomes dominated by its zwitterionic state and therefore moves slowly. The chloride ions within the gel move quickly, creating a strongly negatively charged band which is able to pull along the glycine band. When the differentially charged bands pass through the sample well, proteins accumulate within the narrow gap as proteins have an electrophoretic mobility which is intermediate to the two fronts. The proteins then enter the resolving gel, which consists of a less porous structure. Small proteins navigate easily through the gel structure, with large proteins impeded to a greater extent.

SDS-PAGE buffer was added to protein samples (final concentration 3.1 mM Tris HCl, 0.4% SDS (w/v), 960 mM glycerol, 3.3  $\mu$ M EDTA (ethylenediaminetetraacetic acid), 300 nM bromophenol blue) and incubated at 95 °C for 5 minutes to disrupt the three-dimensional structure.  $\beta$ -mercaptoethanol was added to samples that required disulphide bond disruption, at a final concentration of 2.5%. Samples were run on discontinuous gels consisting of a stacking gel (3.5% acrylamide (w/v), buffered with 0.125 M Tris HCl, 0.1% SDS (w/v), pH 6.8) and a resolving gel (7.5-15% acrylamide, buffered with 0.375 M Tris HCl, 0.1% SDS (w/v), pH 8.8). The gels were catalysed with 0.125% ammonium persulphate (w/v) and 0.0025% TEMED (v/v). The acrylamide percentage of the resolving gel was varied dependent on the molecular weight of the protein to be studied. Electrophoresis was performed at 120-200 V in running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS (w/v), pH 8.3), until the dye front reached the end of the gel. Following gel electrophoresis, proteins were visualised by staining with Coomassie Brilliant Blue R for 1 hour, followed by background de-staining in 40% methanol, 10% acetic acid for 1 hour. Images were taken with a standard camera or LI-COR Odyssey Fc using the 700 nm channel. A protein ladder (PageRuler prestained protein ladder, 10 to 180 kDa) was used for protein size estimation.

#### 2.2.4 Western blotting

Protein samples separated by gel electrophoresis were transferred to a polyvinylidene difluoride membrane (pore size 0.45  $\mu$ m) via the wet electrophoretic transfer method. Membranes were activated by soaking in methanol for 15 seconds before incubation in transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol) for 15 minutes alongside the SDS PAGE gel. A sandwich of pad, filter paper, gel, membrane, filter paper, pad was constructed. The tank was filled with transfer buffer, and the system was kept at 4 °C. Proteins were transferred at a constant current, 200 mA, for 2 hours. The negatively charged proteins moved from the gel to the membrane towards the positively charged anode.

Membranes were incubated overnight at 4 °C in blocking buffer (5% milk solution or 2% bovine serum albumin (BSA) in phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4)). Membranes were washed

three times in wash buffer (blocking buffer, 0.05% tween 20). Membranes were incubated in  $1 \mu\text{g ml}^{-1}$  mouse anti-histidine tag primary antibody in wash buffer at room temperature for 2 hours. Membranes were washed in wash buffer three times before incubation with  $100 \text{ ng ml}^{-1}$  HRP-conjugated goat anti-mouse heavy and light chain specific secondary antibody in wash buffer at room temperature for 1 hour. Membranes were washed four times in wash buffer, three times in PBST (PBS, 0.05% tween 20) and twice in PBS. ECL western blotting substrate was added to the membrane, and the chemiluminescent signal detected using a LI-COR Odyssey Fc. ECL western blotting substrate contains luminol and a peroxide reagent. Luminol turnover is catalysed by HRP, producing a chemiluminescent signal.

#### 2.2.5 Indirect enzyme linked immunosorbent assay (ELISA)

Antigen solutions were prepared in carbonate buffer (50 mM sodium carbonate, pH 9.6). Protein solutions were prepared at  $2 \mu\text{g ml}^{-1}$ . For whole cell antigen, bacterial cells were pelleted by centrifugation and re-suspended in carbonate buffer in the original volume. For lysed cell antigen, bacterial cells were pelleted by centrifugation and re-suspended in 1/10 volume of lysis buffer (Tris-EDTA,  $2 \text{ mg ml}^{-1}$  lysozyme,  $50 \text{ U ml}^{-1}$  mutanolysin). The solutions were incubated at  $37 \text{ }^\circ\text{C}$  for 1 hour. Carbonate buffer was added to the original bacterial cell culture volume.

96 well plates were coated with  $100 \mu\text{l}$  antigen, and incubated overnight at  $4 \text{ }^\circ\text{C}$ . Wells were blocked with  $200 \mu\text{l}$  2% milk in carbonate buffer for 1 hour at room temperature or at  $4 \text{ }^\circ\text{C}$  overnight. Wash steps consisted of a 5 minute incubation in PBST. Plates were washed 3 times and incubated with  $100 \mu\text{l}$  primary antibody for 1 hour at room temperature, either in the form of purified antibody diluted in PBS, cell culture solution or murine sera. The plates were washed three times and incubated with  $100 \mu\text{l}$   $100 \text{ ng ml}^{-1}$  HRP-conjugated goat anti-mouse IgG heavy and light chain specific secondary antibody diluted in PBS for 1 hour at room temperature. Plates were washed 3 times and incubated with  $100 \mu\text{l}$  TMB (3, 3', 5, 5'-tetramethylbenzidine) for 30 minutes. The absorbance at 630 nm was read on a Tecan Infinite 200.

#### 2.2.6 Protein concentration determination – BCA assay

A BCA (bicinchoninic acid) assay was performed to determine protein concentration. A standard curve using known concentrations of BSA was produced. 10 µl sample or BSA standard and 200 µl of BCA reagent (0.08% (w/v) cupric sulphate in BCA) was added to each well, and the plate was incubated at 37 °C for 30 minutes. Absorbance was read at 562 nm on a Tecan Infinite 200.

#### 2.2.7 Protein concentration determination – spectrophotometry

Protein concentration was determined by spectrophotometry (NanoVue Plus) through the absorbance of UV (ultraviolet) light by tryptophan, tyrosine and cysteine residues. The extinction coefficient values and molecular weights were estimated through use of ExpASy software (Gasteiger *et al.*, 2005) see table 2.4 in section 2.4.1. Results were evaluated for the presence of a distinct peak at 280 nm.

#### 2.2.8 RNA extraction

RNA was extracted from *S. agalactiae* strains using ISOLATE II RNA mini kit. The protocol for gram positive bacteria was followed, 3 mg ml<sup>-1</sup> lysozyme and 60 U ml<sup>-1</sup> mutanolysin were utilised for enhanced cell lysis. RNA was eluted in 50 µl dH<sub>2</sub>O.

#### 2.2.9 cDNA synthesis

Reverse transcription was performed to convert the RNA extracted from *S. agalactiae* to cDNA (complementary DNA) for use in PCR reactions. Tetro cDNA synthesis kit was utilised. Briefly, 5 µg RNA was added to 4 µl buffer, 1 µl RiboSafe RNase inhibitor, 1 µl random hexamer primers, and 1 µl Tetro reverse transcriptase, the solution was made up to 20 µl with dH<sub>2</sub>O. The reaction mix was incubated for 10 minutes at 25 °C followed by 30 minutes at 45 °C. The reaction was terminated by incubation at 85 °C for 5 minutes.

### 2.2.10 DNA extraction

DNA was extracted from *S. agalactiae* strains using Qiagen's DNeasy blood & tissue kit. DNA extraction was performed according to the manufacturer's protocol, 60 U ml<sup>-1</sup> mutanolysin was added to the lysis buffer immediately before use. DNA was eluted in 200 µl elution buffer.

### 2.2.11 Polymerase chain reaction

DNA was amplified using GoTaq Flexi PCR kit. The reaction mix consisted of 20% (v/v) buffer, 2 mM MgCl<sub>2</sub>, 0.2 mM dNTPs (deoxynucleotide triphosphates), 1 µM forward primer, 1 µM reverse primer, 1.25 U GoTaq DNA polymerase and 0.5 µg DNA. PCR was performed using a MyCycler Biorad thermocycler. Samples were initially denatured at 95 °C for 2 minutes, followed by 25 cycles of denaturation (95 °C for 1 minute), annealing (see table 2.3 for annealing temperature, for 1 minute) and extension (72 °C for 1 minute (if amplicon ~500 bp) or 3 minutes (if amplicon ~2500 bp)) and final extension at 72 °C for 5 minutes. The primers used and resulting amplicon lengths are detailed in table 2.3.

Target	Primer sequences	Amplicon length / base pairs	Annealing temperature / °C
<b>SAG0771</b>	F:GTGTGAAAACGAATGATGCC R:GTGTGAAAACGAATGATGCC	499	51
<b>SAG1473</b>	F:GCACTTACTGTCTTAACATGTG R:ATAATCCCTCCCGTAGATGC	500	51
<b>SAG1474</b>	F:AGCCAACATGGTGAGAAGTG R:GGAATACGGATAGAACCACCAG	498	50
<b>RPA template</b>	F:ATGGAGACGGCTAACAATAA R:TACCATCAGGCTTGTTCTCT	2583	48

**Table 2.3 Details on PCR experiments**

Annealing temperatures determined by Primer3Plus

### 2.2.12 Agarose gel

DNA samples were analysed by electrophoretic separation on agarose gels. The percentage of agarose used was dependent on the size of DNA fragment to be analysed (3% for samples 100-200 bp, 2% for samples ~500 bp). Agarose was melted in 0.5X TBE (tris-boric acid-EDTA buffer) (47 mM Tris, 1.3 mM EDTA, 45 mM boric

acid). 0.01% (v/v) GelRed nucleic acid stain was added, and gels were cast. Gel loading dye was added to samples prior to loading onto the gel. DNA standards (PCR low ladder, 20 bp and 100bp DNA ladder) were used for molecular weight estimation. Electrophoresis was performed at 80 V in 0.5x TBE, until the dye front approached the end of the gel. Gels were imaged using a LI-COR Odyssey Fc.

#### 2.2.13 Nucleic acid concentration determination - spectrophotometry

A spectrophotometer (NanoVue plus) was used to calculate nucleic acid concentration by exploiting its absorbance of UV light at 260 nm. An absorbance reading of 1 A (with a 1 cm pathlength) equates to 50  $\mu\text{g ml}^{-1}$  DNA and 40  $\mu\text{g ml}^{-1}$  RNA. An absorbance scan of wavelengths 220 nm to 325 nm was completed, allowing assessment of sample purity.

DNA copy number was calculated using the equation below:

$$\text{Copy number (molecules)} = \frac{\text{mass (g)} \times 6.0221 \times 10^{23} (\text{molecules mol}^{-1})}{\text{molecular weight (g mol}^{-1})}$$

#### 2.2.14 DNA purification

PCR product was purified for use in subsequent RPA reactions using Qiagen's QIAquick PCR purification kit, following the centrifugation protocol. All centrifugations were performed at 13,000 g for 1 minute. DNA was eluted in 50  $\mu\text{l}$  dH<sub>2</sub>O.

#### 2.2.15 Statistics

Non-parametric Mann-Whitney U tests were performed to assess statistical significance between two groups, comparing median values. One assumption the Mann-Whitney U test requires is that groups have equal variances, this was validated using a Lavene's test. A Kruskal-Wallis test was performed to extend the comparison to multiple groups. For analysis of data extracted by microarray studies, unpaired student t tests were performed, replicating the strategies followed by the authors of the

studies. (Klinzing *et al.*, 2009; Santi *et al.*, 2009) These tests were performed on Minitab. (Minitab Inc, 2014)

## 2.3 Computing

### 2.3.1 IDRIS

IDRIS is a bespoke in-house computer system developed at Newcastle University within the i-sense project ([www.i-sense.org.uk/](http://www.i-sense.org.uk/)). (Flanagan *et al.*, 2014) IDRIS compares amino acid sequences to identify peptides which appear in a user defined group of bacteria, and in no other bacteria. The reported amino acid sequences are therefore conserved within and unique to the group of interest. IDRIS consists of two workflows: ApID1 and ApID2. Within ApID1, the translated protein sequences from all complete bacterial genomes in the RefSeq database are captured by IDRIS. IDRIS creates a database of 15-mer amino acid “tokens”, capturing each 15-amino acid run via a sliding window which moves across the sequence in single-step increments. Each peptide is annotated with the protein and species it belongs to, along with its predicted cellular localisation. Protein localisation is determined via PSORTb, SignalP and LipoP (discussed further in following sections). Within ApID2 a group of interest is designated, tokens which appear in each proteome in this group are selected. This group of tokens is compared to tokens from proteomes external to the group; tokens are eliminated if they also belong to any other proteome. The resulting IDRIS report lists group-distinctive peptides, the protein they are found within and their predicted cellular localisations.

IDRIS v1.0 was utilised for initial work within this project. Subsequent iterations (v1.5 and v2.0) include updated RefSeq data and the capability to analyse protein localisation in gram negative bacteria. Currently the IDRIS token database comprises of 7.65 billion tokens from 7,059 organisms.

### 2.3.2 PSORTb

PSORTb is a protein localisation prediction tool. (Yu *et al.*, 2010) It was created on a dataset of 2652 gram positive proteins with known localisations. PSORTb v3.0 has a higher accuracy than competitor tools, but is less sensitive, therefore the frequency of unknown localisations is higher. PSORTb is able to identify 25 sequence motifs and 6 profiles (sequence motifs with position specific weighting) for gram positive bacteria. The programme returns a localisation prediction score out of 10 for 4 locations: extracellular, cell wall, cytoplasmic membrane and cytoplasmic. PSORTb was used to confirm IDRIS localisation predictions of SAG0771, SAG1473 and SAG1474.

### 2.3.3 SignalP

SignalP predicts the presence of and cleavage site of a signal peptide within a protein sequence. (Petersen *et al.*, 2011) A large C-score is found in the first residue of the mature protein following signal peptide cleavage. The S-score discriminates residues in a signal peptide from those in a mature protein or protein with no signal peptide. The Y-score is derived from the C-score and the gradient of the S-score to report a more precise prediction of the cleavage site. The mean S-score is an average of the S-scores of all amino acids up to the residue before the maximum Y-score. The D-score is determined by a weighted combination of the mean S-score and maximum Y-score. A D-score over 0.45 (range: 0-1) represents the presence of a signal peptide. The probability of a false positive result is 0.001. SignalP was used to confirm IDRIS localisation predictions of SAG0771, SAG1473 and SAG1474.

### 2.3.4 LipoP

LipoP is a prediction tool which distinguishes lipoprotein signal peptides from alternative signal peptides, N-terminal membrane helices and other proteins. (Juncker *et al.*, 2003) LipoP predicts cleavage sites for signal peptides and lipoprotein signal peptides if appropriate. The highest scoring protein class is returned as the best fitting prediction. Although the server was created for use in gram negative bacteria, it has been found to be applicable to gram positive bacteria also. (Rahman *et al.*, 2008) LipoP was used to confirm IDRIS localisation predictions of SAG0771, SAG1473 and SAG1474.

### 2.3.5 TMHMM

TMHMM predicts the presence, length and orientation of transmembrane helices within a protein sequence. (Krogh *et al.*, 2001) A probability plot is created using an average score, from all potential structure models, of the probability of each amino acid to exist in a transmembrane helix, intracellular or extracellular environment. TMHMM was used to confirm IDRIS localisation predictions of SAG0771, SAG1473 and SAG1474.

### 2.3.6 I-TASSER

I-TASSER is a server which predicts structural models based on a protein sequence. (Roy *et al.*, 2010) Secondary structural templates are assembled into full length models that are iteratively threaded through homology models to improve the prediction. Models are refined to minimise steric hindrance and optimise hydrogen bonding. A confidence score (C-score) is calculated from the convergence parameters of models and the significance of alignment with templates. In validation studies TM-scores were determined by analysing structural similarity between a predicted structure and its empirically determined structure. A strong correlation was found between C-score and TM-score. A predicted TM-score is calculated for each predicted structure. A TM-score (range: 0-1) of  $\geq 0.5$  is considered a correct fold. I-TASSER was used to visualise the structure of SAG0771, SAG1473 and SAG1474. I-TASSER was also used in the analysis of peptide secondary structure and surface exposure with the SAG1474 protein.

### 2.3.7 PSIPRED

PSIPRED v3.3 is a tool for the prediction of secondary structure. (Jones, 1999) The server utilises evolutionarily related sequences, determined by position specific iterated BLAST, (Altschul *et al.*, 1997) to train a machine learning algorithm to predict the secondary structure of each amino acid residue. PSIPRED was used in analysing the secondary structure of SAG1474 for antigen selection.

### 2.3.8 Prodepth

Prodepth is used to predict surface accessible amino acids within a protein sequence. (Song *et al.*, 2009) The server predicts the distance between each residue and its nearest water molecule. The larger the length, the deeper the residue is buried. The threshold for exposure is set to 3.0 Å. Prodepth combines data on solvent accessibility (SCRATCH), secondary structure (PSIPRED), disordered regions (DISOPRED) and sequence alignments (PSI-BLAST) to generate its output. Input programmes shown in brackets. Prodepth was used for analysis of SAG1474 peptide surface exposure for antigen selection.

### 2.3.9 Bcepred

Bcepred is a tool to predict linear B-cell epitopes from a protein sequence. (Saha and Raghava, 2004) The server allows threshold values for various parameters (hydrophilicity, flexibility/mobility, accessibility, polarity, surface exposure and turns) to be set as desired. A study into optimal epitope prediction determined that using hydrophilicity, flexibility, polarity and surface exposure with threshold values 2.0, 1.9, 2.3 and 2.4 respectively, and an overall threshold of 2.38 enabled 58.7% accuracy. These parameters and threshold values were utilised for the study of SAG1474 tokens.

### 2.3.10 Microarray data analysis

The Gene Expression Omnibus (GEO) and ArrayExpress are public repository databases consisting of gene expression studies including microarray studies. These databases were searched to investigate changes in gene expression in response to temperature (GEO: GSE14573) and pH (ArrayExpress: E-TABM-708). Data from the two microarray studies were extracted and expressed as the average normalised fluorescent signal of the test condition as a ratio to the control condition. Values were considered to be statistically different from the control if ratios were <0.5 or >2 and had a p value  $\leq 0.01$ .

#### 2.3.11 Protein BLAST (Basic local alignment tool) search

Protein BLAST searches (Altschul *et al.*, 1990) were performed to analyse similarity between query proteins and sequenced proteins within the specified dataset. Specific organisms can be included or excluded from the search criteria. Results showed peptide sequences with the highest similarity to the query, the alignment between these sequences, and the percentage of amino acid similarity and identity.

#### 2.3.12 DNA alignment

DNA encoding for the *S. agalactiae* SAG1474 protein was inputted to a standard nucleotide BLAST to find all variants of the sequence within the NCBI GenBank. The aligned sequences (Zhang *et al.*, 2000) were exported and visualised using Jalview. (Waterhouse *et al.*, 2009) DNA alignments were manually adjusted to allow analysis of the number of base mismatches found in each prospective RPA primer.

#### 2.3.13 Primer design: PCR and RPA

Primers were designed using Primer3Plus. (Untergasser *et al.*, 2007) For primers for use in PCR, the following parameters were set: primer size: 18-22 bp, primer T<sub>m</sub>: 52-63 °C, maximum primer T<sub>m</sub> difference: 10 °C, primer GC content: 40-60%, maximum 3' self-complementarity: 3, maximum self-complementarity: 8, optimal amplicon length: 500 bp (base pairs) or 2500 bp, and all penalties: 1. For use in RPA, primers were designed using the following parameters: primer size: 18-35 bp, optimal amplicon length: 80-275 bp and all penalties: 1. Potential primers were evaluated using IDT OligoAnalyzer 3.1, to predict melting temperatures of hairpin structures and the change in Gibbs' free energy ( $\Delta G$ ) associated with primer self-dimers and hetero-dimers. For PCR primers thermal properties were calculated using the following conditions: 0.5  $\mu$ M oligo, 50 mM Na<sup>+</sup>, 2 mM Mg<sup>2+</sup>, 0.2 mM dNTPs and at the primers' annealing temperature. For RPA primers thermal properties were calculated using the following conditions: 0.5  $\mu$ M oligo, 0 mM Na<sup>+</sup>, 14 mM Mg<sup>2+</sup>, 0.24 mM dNTPs and at 25 °C.

#### 2.3.14 Band densitometry

LI-COR Image Studio was used to analyse SDS-PAGE and agarose gel images. Band densities were calculated for comparison of protein levels. Background levels (median values surrounding the bands, with a border width of 3) were subtracted from protein signals.

### 2.4 Recombinant protein expression and purification

#### 2.4.1 Overview

Table 2.4 shows the recombinant proteins made in this work and several predicted parameters.

Protein	Molecular weight / kDa	Extinction coefficient /M <sup>-1</sup> cm <sup>-1</sup>	Isoelectric point (pI)
<b>Aequorin- WT*</b>	22.6	43.4	5.2
<b>Aequorin-FLAG</b>	23.6	44.9	5.0
<b>SAG1474</b>	69.4	65.3	8.5
<b>SAG1474-MBP</b>	115	133	6.5

**Table 2.4 Recombinant proteins and associated parameters**

Protein parameters were calculated using ExPASy. (Gasteiger *et al.*, 2005)

#### 2.4.2 Competent cells

Competent cells are permeable to accept horizontally transferred DNA. Chemically competent *E. coli* cells were created by the addition of positively charged cations to minimise electrostatic repulsion between the bacterial cell and external DNA. 500 µl cell suspension from an overnight culture was added to 50 ml LB broth and grown to an OD (optical density)<sub>600 nm</sub> of 0.35-0.45. Further steps were carried out at 4 °C or on ice. Cells were centrifuged at 700 g for 10 minutes. The pellet was re-suspended in 5 ml filter-sterilised (0.22 µm pore size) transformation and storage buffer (12.5 mM PEG (polyethylene glycol) 8000, 30 mM MgCl<sub>2</sub>, 5% dimethyl sulphoxide (DMSO) in LB media). 100 µl aliquots were frozen at -80 °C.

### 2.4.3 Transformation

Plasmid DNA was transformed into competent *E. coli* cells. 50 ng of plasmid DNA was added to 500  $\mu$ l of thawed competent cells and incubated on ice for 30 minutes. Cells were heatshocked at 42 °C for 45 seconds and transferred to an ice bath for 1-2 minutes. 800  $\mu$ l of LB media was added and cells were incubated at 37 °C for 90 minutes. Cells were centrifuged at 14,000 g for 1 minute, and 800  $\mu$ l of supernatant was removed. Cell suspensions were cultured on antibiotic selective media to select for transformants that had received the plasmid-derived antibiotic resistance. Table 2.5 details the *E. coli* strains used in the production of recombinant proteins.

Name	Genotype	Recombinant protein
<b>BL21</b>	BL21 <i>pLysS DE3</i>	N/A
<b>LMG194</b>	LMG194	N/A
<b>Aeq-WT*(BL21)</b>	BL21 <i>pLysS DE3</i> <i>pBADHisA:Aeq-WT*</i>	Aequorin WT*
<b>Aeq-FLAG(BL21)</b>	BL21 <i>pLysS DE3</i> <i>pBADHisA:Aeq-FLAG</i>	Aequorin FLAG
<b>Aeq-WT*(LMG194)</b>	LMG194 <i>pBADHisA:Aeq-WT*</i>	Aequorin WT*
<b>Aeq-FLAG(LMG194)</b>	LMG194 <i>pBADHisA:Aeq-FLAG</i>	Aequorin FLAG
<b>SAG1474(BL21)</b>	BL21 <i>pLysS DE3 pET:SAG1474</i>	SAG1474
<b>SAG1474-MBP(BL21)</b>	BL21 <i>pLysS DE3</i> <i>pMALp5x:SAG1474-MBP</i>	SAG1474-MBP

**Table 2.5 *E. coli* strains for production of recombinant protein**

### 2.4.4 Induction growth tests

Growth tests were performed to assess the optimum conditions for protein expression. *E. coli* cells were grown to an OD<sub>600 nm</sub> of 0.5 and induced with the appropriate compound: isopropyl  $\beta$ -D-1 thiogalactopyranoside (IPTG) for genes located in *pET28a* and *pMAL p5x* plasmids and arabinose for *pBADHisA* plasmids. Several parameters were investigated: cell type (BL21, LMG194), temperature (37 °C,

30 °C, 25 °C and 20 °C), concentration of inducer (IPTG: 0 mM – 2 mM, arabinose: 0%-0.2%) and time of induction (1-16 hours). Cells were pelleted by centrifugation.

#### 2.4.5 Cell lysis

Cells were lysed by sonication or chemical lysis. For sonication, cells were re-suspended in 20 mM sodium phosphate buffer, pH 7.4 with 150 or 500 mM NaCl. Cell suspensions were sonicated on ice with six 15-second bursts. Cells were chemically lysed by re-suspension in Bugbuster protein extraction reagent followed by incubation at room temperature for 20 minutes whilst shaking. Lysed cells were centrifuged to separate the soluble and insoluble fractions.

#### 2.4.6 Periplasmic extraction

The periplasmic fraction was extracted as per *pMAL* guidelines. (New England BioLabs, 2017b) Cells were re-suspended in 400 µl 30 mM Tris-HCl, pH 8.0, 20% sucrose. EDTA was added to 1 mM, and the solution was shaken for 10 minutes at room temperature. All further steps were completed at 4 °C. Cells were pelleted at 8000 g for 10 minutes, and re-suspended in 400 µl ice cold 5 mM MgSO<sub>4</sub>. The solution was shaken for 10 minutes and centrifuged at 8000 g for 10 minutes. The periplasmic fraction was found in the supernatant.

#### 2.4.7 Protease cleavage

SAG1474-MBP was cleaved at the TEV protease (protease from the Tobacco Etch Virus) cleavage site to separate SAG1474 from the MBP binding protein. The protein sample was dialysed overnight at 4 °C into 25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 14 mM β-mercaptoethanol, with one buffer change. TEV protease was added to 1% (w/w). The sample was incubated at 4 °C for 16 hours. TEV protease was removed via chromatography.

#### 2.4.8 Protein purification - chromatography

Large scale bacterial culture was performed using the induction, cell lysis and protein extraction conditions previously determined. Proteins were purified from the crude cellular protein mixture using nickel-affinity, ion-exchange, GST (glutathione S transferase)-affinity and size exclusion chromatography using an Äkta start system. All buffers were filtered (0.22 µm) and degassed prior to use. Samples were filtered (0.22 µm) prior to addition to the column. All purification steps were performed at a speed of 1 ml min<sup>-1</sup>.

#### 2.4.9 Nickel-affinity chromatography

Proteins containing a polyhistidine tag can be purified by exploiting the affinity for nickel. Proteins are eluted by raising the imidazole concentration of the buffer, which has a higher affinity for nickel and therefore outcompetes the protein. (Bornhorst and Falke, 2000) His-tagged proteins were purified using a HisTrap HP column. 20 mM sodium phosphate buffer, pH 7.6, 500 mM NaCl was used for column buffers. Imidazole was added to the loading, wash and elution buffers at final concentrations of 10 mM, 20 mM and 250 mM respectively. The system was equilibrated with loading buffer. Sample was added to an equal volume of wash buffer and loaded onto the column. The column was flushed with wash buffer until there was no change in absorbance spectra. Elution buffer was added to the system at a concentration gradient of 5% per 1 ml eluted. 1 ml fractions were collected.

#### 2.4.10 GST-affinity chromatography

The GST-tagged protein TEV protease was purified using a GSTrap FF column. GST binds to its substrate glutathione, immobilised onto a column. The tagged protein is eluted by addition of an excess of free glutathione to the system. Briefly, the sample was dialysed in PBS overnight at 4 °C with one buffer change. The column was equilibrated with PBS, the sample was added and the column was washed with 5 column volumes of PBS. GST-affinity chromatography was performed to remove a GST-tagged contaminant (TEV protease), therefore flow through was collected. The column was cleaned by addition of elution buffer (50 mM Tris HCl, pH 8.0, 10 mM reduced glutathione).

#### 2.4.11 Ion exchange chromatography

Proteins were purified by ion exchange chromatography using HiTrap CM FF (cationic exchanger) and HiTrap DEAE FF (anionic exchanger) columns. Proteins with a higher pI than the pH of the buffer are positively charged, and are able to bind to a cationic exchange column. Proteins with a lower pI than the pH of the buffer are negatively charged, and therefore bind to an anionic exchange column. Proteins are eluted by increasing the salt concentration, with the higher concentration of ions outcompeting the affinity of the protein for the column. (Selkirk, 2004) For cationic exchange chromatography, the protein sample was dialysed overnight at 4 °C into loading buffer (50 mM sodium phosphate, pH 7.5, 50 mM NaCl), with one buffer change. The column was equilibrated with loading buffer and the sample was loaded onto the column. The column was washed with 5 column volumes of loading buffer. Protein was eluted with 10 column volumes of elution buffer (50 mM sodium phosphate, pH 7.5, 1 M NaCl). 1 ml fractions were collected. For anionic exchange chromatography, the protein sample was dialysed overnight at 4 °C into loading buffer (50 mM sodium phosphate buffer, pH 7.2, 50 mM NaCl) with one buffer change. The column was equilibrated with loading buffer and the sample was loaded onto the column. The column was washed with 5 column volumes of loading buffer. Elution buffer (50 mM sodium phosphate buffer, pH 7.2, 1 M NaCl) was added to the system at a concentration gradient of 2% per 1 ml eluted, until elution buffer was at 50%. 100% elution buffer was used to wash remaining protein from the column. 1 ml fractions were collected.

#### 2.4.12 Size exclusion chromatography

Proteins were separated by size using a HiLoad 16/600 Superdex 200 prep grade column. Large proteins cannot enter the pores within the beads which form the chromatography column, and therefore flow through the column quickly. Smaller proteins enter the pores and their progress through the column is slowed. Thus proteins are separated based on their molecular size. The column was equilibrated with PBS, and the sample was loaded onto the column in 1 or 2 ml volumes. Samples were concentrated using a vivaspin column to reduce buffer volume if necessary. The column was washed with PBS until the desired protein was eluted. 1 ml fractions were collected.

#### 2.4.13 Circular dichroism

Circular dichroism (CD) is a technique utilised to study protein structure, based on the differential absorbance of right handed, and left handed, circularly polarised light. Further explanation can be found in section 4.4.6. Circular dichroism (CD) was performed to analyse the secondary and tertiary structure of proteins, using far and near UV respectively. The melting temperature of protein secondary structure was also analysed. These experiments were performed using a JASCO J-810. Protein samples were dialysed into 20 mM potassium phosphate, pH 7.5, 50 mM NaCl. Background scans were performed using buffer and subtracted from far and near UV scans. The following table, table 2.6, shows experiments details. HT (feedback voltage) values were monitored to ensure they did not exceed 800 V.

	Far UV	Near UV	Thermal melt
<b>Varied parameter</b>	190-250 nm	250-320 nm	10-95 °C
<b>Measurement increment</b>	0.2 nm	0.2 nm	0.2 °C
<b>Protein concentration / <math>\mu\text{M}</math></b>	9.9	9.9	2.4
<b>Cuvette path length</b>	0.2 mm	1 cm	1 cm
<b>Number of scans</b>	9	9	1
<b>Number of background scans</b>	3	3	0

**Table 2.6 Experimental details for circular dichroism**

#### 2.4.14 Liquid chromatography mass spectrometry

Liquid chromatography mass spectrometry was performed by Newcastle University Protein and Proteome Analysis. The protein sample was extracted from an SDS-PAGE gel, protease digested and peptides were separated by liquid chromatography. The mass:charge ratios of ionised peptides were determined by mass spectrometry. A reference data set comprising of proteins from *E. coli*, *S. agalactiae* and the recombinant SAG1474-MBP sequence was utilised to determine protein identity.

## 2.5 Antibody production

### 2.5.1 Overview

Murine antibodies were raised in house against three peptides CVQTNDSNPT, CVNLEENSQV and CNMENLSQEERI. This work was performed in combination with Julia Spoons and Robert Bolt.

### 2.5.2 Peptide - carrier protein conjugation

Peptides were conjugated to maleimide activated keyhole limpet haemocyanin (KLH) for murine immunisation. Each peptide contained an N-terminal cysteine residue to enable directed conjugation. Peptides were also conjugated to maleimide activated BSA for use in antibody screening for target affinity. Peptides were conjugated to their carrier proteins using Imject maleimide activated KLH and BSA conjugation kits, according to the manufacturer's protocol. Briefly, 2 mg carrier protein was reconstituted in dH<sub>2</sub>O to 10 mg ml<sup>-1</sup> and added to 2 mg peptide dissolved in conjugation buffer. Conjugation mix was incubated at room temperature for 2 hours. Conjugated peptide was purified using a desalting column to remove EDTA prior to injection into mice.

### 2.5.3 Ellman's test

An Ellman's test was carried out to estimate the coupling efficiency of peptide to KLH and BSA and to estimate the average number of peptide molecules attached to the carrier protein. Cysteine standards (0-5 mM) were used to create a standard curve, from this, the concentration of free cysteine residues in the unconjugated peptide and the peptide after conjugation were interpolated.

200 µl Ellman's buffer (0.2 M sodium phosphate buffer, pH 8.0, 0.15 M NaCl, 0.1 M EDTA) was loaded into each well. 20 µl of Ellman's reagent (5,5'-dithio-bis-(2-nitrobenzoic acid) at 1 mg ml<sup>-1</sup> diluted in Ellman's buffer, and 10 µl of cysteine standards or sample was added to wells. The plate was incubated at room temperature for 15 minutes, and the absorbance was read at 412 nm on a Tecan Infinite 200.

The following calculations were performed:

$$\text{Conjugated cysteine} = \text{total cysteine (moles)} * -\text{unconjugated cysteine (moles)} **$$

$$\text{Conjugation efficiency} = \frac{\text{conjugated cysteine (moles)}}{\text{total cysteine (moles)}}$$

$$\text{Number of peptides per protein unit} = \frac{\text{peptide (moles)} \times \text{conjugation efficiency}}{\text{carrier protein (moles)}}$$

\* Determined by the number of free cysteine residues in the peptide sample prior to conjugation.

\*\* Determined by the number of free cysteine residues in the protein-peptide sample following conjugation.

#### 2.5.4 Mouse immunisation

6-8 week old female BALB/c mice were immunised with KLH-conjugated peptide. Two week intervals were scheduled between the first four immunisations. The time between the penultimate immunisation and pre-immune boost varied from 14 days to 98 days. The final pre-immune boost was performed four days prior to spleen harvest. Each mouse received a total of five immunisations. For each immunisation mice received two 50 µl subcutaneous injections of 1 mg ml<sup>-1</sup> total protein, one into each flank. Peptide solutions were mixed 1:1 with Freund's adjuvant to create an emulsion in order to minimise dispersion of solution following injection. This decreases degradation and elimination from the body and promotes prolonged antibody generation. (Stills, 2005) Initial immunisation was carried out with Freund's complete adjuvant, the following three immunisations contained Freund's incomplete adjuvant. The final pre-immune boost was diluted in PBS. 50-60 µl tail bleeds were taken from each mouse prior to the immunisation programme and again following the 4<sup>th</sup> immunisation. Serum was isolated from whole blood samples via centrifugation at 12,000 g for 10 minutes at

room temperature. All murine work was carried out at the Comparative Biology Centre (CBC), Newcastle University under licence number: PCEAA8B1B.

#### 2.5.5 Cell culture

NS-1 cells (P3/NSI/1-Ag4-1), which produce but do not secrete kappa chain antibodies, were utilised as the myeloma fusion partner. NS-1 and hybridoma cells were grown in a 5% CO<sub>2</sub> incubator at 37 °C. Cells were grown in various medias. Basic media (BM) consisted of 10% foetal bovine serum (FBS), 2 mM L-glutamine, 100 U ml<sup>-1</sup> penicillin and 100 µg ml<sup>-1</sup> streptomycin in RPMI 1640 medium. HAT-selective media consisted of 20% FBS, 2 mM L-glutamine, 100 U ml<sup>-1</sup> penicillin, 100 µg ml<sup>-1</sup> streptomycin and HAT supplement (final concentration 0.1 mM sodium hypoxanthine, 0.4 µM aminopterin, 1.6 µM thymidine) in RPMI 1640 medium. HT media consisted of BM with HT supplement (final concentration 0.1 mM sodium hypoxanthine, 1.6 µM thymidine). BM condensed media consisted of 10% condensed in BM. Serum-free media consisted of 100 U ml<sup>-1</sup> penicillin and 100 µg ml<sup>-1</sup> streptomycin in EX-CELL 610-HSF serum-free medium for hybridoma cells. Freezing media contained 10% FBS, 10% DMSO in BM. All media was pre-warmed before addition to cells, unless otherwise specified. All centrifugation steps were performed at 400 g for 5 minutes unless otherwise specified.

#### 2.5.6 Viable cell count

A viable cell count was completed by adding equal volume of 0.4% trypan blue to a cell culture sample. Cells were transferred to a haemocytometer and visualised microscopically at 100x magnification. Viable cells resisted staining, showing as clear spherical cells. Non-viable cells stained blue and often had rough edges. Viable cells within four squares of the haemocytometer were counted and the average value was taken. This was multiplied by 10<sup>4</sup> and doubled (to account for the 1:1 dilution in trypan blue) to calculate the number of viable cells per ml.

#### 2.5.7 Freezing cells

Cells were grown to confluence and collected by centrifugation. The cell pellet was re-suspended gently in freezing media at approximately 10<sup>6</sup> cells ml<sup>-1</sup> and transferred to

a cryovial. Vials were placed in a freezing container, which controlled the rate of cooling at 1 °C per minute until -80 °C was reached. Cells were stored in liquid nitrogen.

#### 2.5.8 Thawing frozen cells

Cells were thawed rapidly by incubation of cryovials in a 37 °C water bath. Media was added to cell solutions to dilute the DMSO and cell solutions were centrifuged. The cell pellet was re-suspended in BM and transferred to the appropriately sized well or flask, dependent on cell number and viability.

#### 2.5.9 Fusion

$2 \times 10^7$  NS-1 cells were pelleted and re-suspended in 50 ml RPMI 1640 medium in preparation for fusion. Mice were sacrificed by the CBC and spleens were harvested and kept in RPMI 1640 medium on ice. A small incision was made in the spleen and cells were gently pressed out into 50 ml RPMI 1640 medium. The culture fluid was pipetted up and down several times using a 10 ml serological pipette to create a single cell suspension, and passed through a sieve. Spleen and NS-1 cells were pelleted and re-suspended in 50 ml RPMI 1640 medium and pelleted again. Each pellet was re-suspended in 25 ml RPMI 1640 medium and the cell solutions were combined and pelleted. 1 ml PEG-DMSO was added slowly over one minute and this was stirred for an additional one minute. 1 ml RPMI 1640 medium was added over one minute, and 9 ml added over the following two minutes. Cells were pelleted and re-suspended in HAT selective media. 50  $\mu$ l of this cell suspension was added to each well of eight 96 well plates pre-charged with 150  $\mu$ l HAT selective media. The plates were incubated until sufficient growth had occurred before screening.

#### 2.5.10 Cloning

Fusion wells were screened for antibody binding to the BSA-conjugated peptide. ELISA screens were completed using the protocol described in section 2.2.5. Cells with positive binding were cultured and expanded into 24-well and 12-well plates in HT media. Following sufficient growth, these cells were used to seed 96-well subcloning plates. A series of six-fold dilutions was set up with wells seeded at a range of 600-0.1

cells per well. Viable cell counts were performed as described in section 2.5.6. Cells were diluted in BM-conditioned media to support the growth of low cell numbers. Target-binding wells which derived from one cell per well were deemed to be monoclonal cell lines. Those wells which exhibited target binding and were derived from more than one cell per well were used to seed further subcloning plates.

#### 2.5.11 Antibody purification

Hybridoma cell lines were grown in serum free media for 10-14 days until cells were overgrown and 50% of cells were visibly dead. Cell culture fluid was centrifuged at 700 g for 10 minutes to remove hybridoma cells. Antibodies were purified through addition of ammonium sulphate. Interactions between antibodies and water molecules are disrupted as water molecules show affinity for the added ammonium sulphate. This effectively reduces the availability of water, causing antibodies to precipitate. (Grodzki *et al.*, 2010) Ammonium sulphate was slowly added to the cell culture supernatant at a final concentration of 2.05 M (50% saturation). The solution was incubated overnight at 4 °C whilst stirring constantly. The precipitated protein collected in the pellet following centrifugation at 3000 g for 30 minutes. The pellet was re-suspended in PBS and dialysed overnight at 4 °C. The solution was centrifuged at 14000 g for 5 minutes to remove any insoluble fraction and the supernatant was collected.

## 2.6 Recombinase Polymerase Amplification (RPA)

### 2.6.1 RPA assay

RPA is an isothermal DNA amplification technique. RPA kits were used according to manufacturer instructions. Briefly, 2.4 µl 10 µM forward primer, 2.4 µl 10 µM reverse primer, 29.5 µl rehydration buffer and template DNA was combined. dH<sub>2</sub>O was added to give a total volume of 47.5 µl. The solution was vortexed, briefly centrifuged, added to the RPA enzyme pellet. The mixture was pipetted up and down 20 times until fully reconstituted. 2.5 µl 280 mM magnesium acetate was added to the curved lids. The reaction was initiated by centrifuging the RPA tube, causing magnesium to be drawn into the RPA reaction solution. Tubes were vortexed and re-centrifuged before

incubating in a PCR block at the desired temperature. Reactions were stopped by transferring samples to a -80°C freezer. Negative controls which contained no template DNA or with an incubation time of 0 minutes were used. A positive control was provided by the manufacturers.

#### 2.6.2 Enzyme linked oligonucleotide assay

An enzyme linked oligonucleotide assay (ELONA) was performed to analyse the amplification using tailed-primers, as previously described by Jauset-Rubio *et al.*, 2016.

50 µM capture probe was reduced in an equal volume of 10 mM TCEP HCl (Tris(2-carboxyethyl)phosphine hydrochloride) by incubation for 1 hour at room temperature. Reduction solution was added to PBS in a 0.4% solution (v/v). Maleimide plates were washed three times with PBST. 100 µl capture probe solution was added to each well and the plate was incubated overnight at 4 °C. The plate was washed three times in PBST and blocked with 200 µl 100 µM 6-mercaptohexanol. This was incubated at room temperature for 40 minutes. The plate was washed three times with PBST. 35 µl of RPA product was added to each well, and the plate was incubated for 30 minutes at room temperature whilst shaking. The plate was washed three times in PBST and 100 µl 10 nM reporter probe was added. This was incubated for 30 minutes at room temperature whilst shaking. The plate was washed three times in PBST prior to the addition of 100 µl TMB. The plate was read at 630 nm every 5 minutes on a Tecan Infinite 200. See table 2.7 for primer and probe sequences.

Name	Sequence
<b>Capture probe</b>	GGCATGTCTGTGATTACTTTTTTTTTTTTTTTT-C6 Thiol
<b>Reporter probe</b>	HRP-AGTGGTCGTCGTTTTACA
<b>Forward primer</b>	GTAATCACAGACATGCC-C3-ACTACTAGACGCCAAGAAGCTATTGAAGAGG
<b>Reverse primer</b>	TGTAACAACGACGACCACT-C3-TCACCACCTTAATACTGTGCCCTAACC

**Table 2.7 Sequences for use in RPA assay**

---

## Chapter 3 Protein target selection

---

### 3.1 Introduction

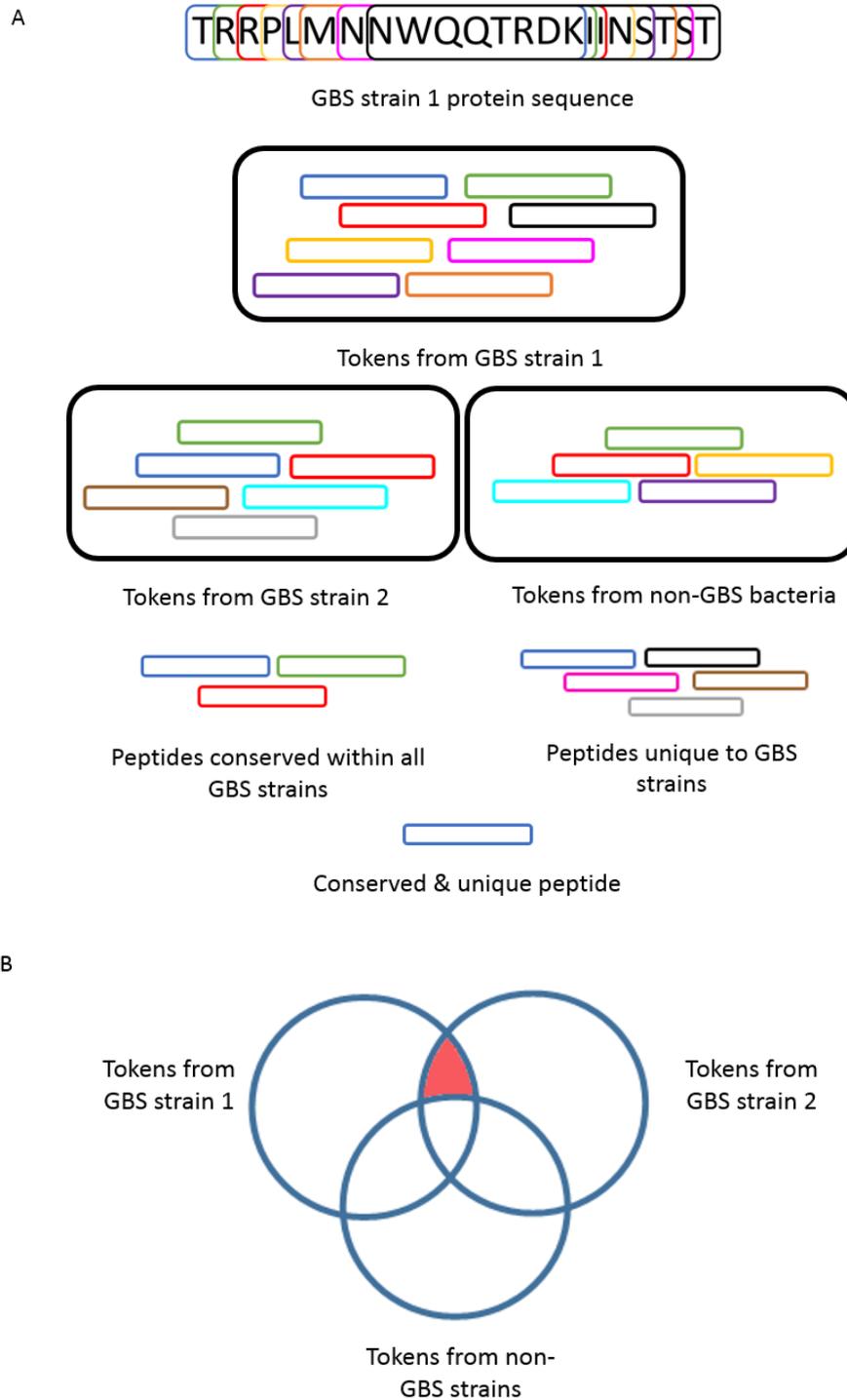
Diagnostic molecules for the detection of infectious disease are often generated against proteins or bacterial cells and may exhibit binding affinity to any epitope within the target. Many of these epitopes will not be unique to the desired organism, therefore cross-reactivity is likely. By raising an antibody against a specified peptide, proteins with similar amino acid sequences can be identified bioinformatically and cross-reactivity can be studied. This targeted approach can result in an increased confidence of specificity. Within a bacterial species there are often multiple serotypes and strains, which may contain significant sequence variability within their proteomes. Antibodies generated against regions of diversity would have a decreased ability to detect certain strains. Peptides found in all strains of interest can be selected as antibody targets to maximise sensitivity.

One limitation of peptide-raised antibodies is that for the generation of a successful antibody, the peptide sequence must hold the same shape in the complex three-dimensional structure of the native protein as it does in solution. Therefore the antibody epitope must be linear rather than discontinuous.

### 3.2 GBS specific peptides

IDRIS v1.0 is a bespoke in-house computer system developed at Newcastle University within the i-sense project ([www.i-sense.org.uk](http://www.i-sense.org.uk)), created to identify sequences of amino acids, which are unique to, and conserved throughout, a user-defined group of interest. (Flanagan *et al.*, 2014) The programme has been utilised to identify potential protein targets for diagnostic molecules. The IDRIS system translates every bacterial genome sequence found in NCBI RefSeq into all possible reading frames, and then splits the protein sequences into tokens, which are sequences of 15 amino acids. Tokens are

analysed for their uniqueness to a selected group of bacterial strains and their conservation within the bacterial group. The proteins that these tokens belong to are analysed to predict protein location in the cell. A diagrammatic representation can be seen in figure 3.1. A diagnostic target would ideally be surface exposed to enable molecular detection with minimal sample processing required, possibly utilising whole cell detection. IDRIS uses SignalP, (Petersen *et al.*, 2011) LipoP (Juncker *et al.*, 2003) and PSORTb (Yu *et al.*, 2010) to assess protein location.



**Figure 3.1 IDRIS selection of unique and conserved peptides**

A: Protein sequences are broken up into tokens and compared to tokens from all strains within the bacterial group of interest. Conserved peptides are selected. Tokens are also compared with tokens from other bacteria. Unique peptides are selected. Peptides which appear in both of these groups are reported as tokens distinctive to the bacterial group of interest. B: Tokens found in all GBS strains and no other bacterial strains are reported.

All *Streptococcus agalactiae* strains found in the NCBI RefSeq database (as of 18/06/2015) were used as the group of interest. The resulting IDRIS report listed all tokens found in every GBS proteome in the database that were also not found in any proteomes outside of this group. Table 3.1 shows the proteins from this report which are predicted to contain a signal peptide, as determined by SignalP. PSORTb categorised three of the returned proteins as cell wall associated proteins, these have the highest likelihood of all returned proteins of being surface exposed, as the cell wall is the most exterior layer of gram positive cells, excluding the polysaccharide capsule. (Shockman and Barrett, 1983) Cell wall associated proteins of *S. agalactiae* have been reported to be exposed to surface (Lancefield *et al.*, 1975; Brodeur *et al.*, 2000) possibly due to the capsule being narrow or flexible, or not being present at certain phases of bacterial growth. (Lindahl *et al.*, 2005) The proteins considered further were SAG0771 (NP\_687786.1), SAG1473 (NP\_688467.1) and SAG1474 (NP\_688468.1) (locus tag (protein ID)).

Locus tag	Protein ID	Protein name / function	Localisation
SAG0032	NP_687068.1	Group B streptococcal surface immunogenic protein (Sip)	Ex
SAG0017	NP_687053.1	Peptidoglycan hydrolase (PcsB)	Ex
SAG1197	NP_688206.1	Hyaluronate lyase	Ex
SAG0031	NP_687067.1	M24/M37 family peptidase	Ex
SAG0771	NP_687786.1	Cell wall surface anchor family protein	CW
SAG1473	NP_688467.1	Cell wall surface anchor family protein	CW
SAG1474	NP_688468.1	Amidase	CW
SAG0147	NP_687183.1	D-alanyl-D-alanine carboxypeptidase	CM
SAG0146	NP_687182.1	Penicillin-binding protein 4	CM
SAG1466	NP_688460.1	Glutamine ABC transporter	CM
SAG0136	NP_687172.1	Amino acid ABC transporter	CM
SAG0535	NP_687564.1	Zinc ABC transporter substrate-binding protein	CM
SAG1007	NP_688018.1	Iron-compound ABC transporter	U
SAG2174	NP_689159.1	Serine protease	U
SAG0034	NP_687070.1	Sugar ABC transporter sugar-binding protein	U
SAG1441	NP_688438.1	Maltose/maltodextrin ABC transporter protein	U
SAG2148	NP_689133.1	LysM domain-containing protein	U
SAG0290	NP_687325.1	ABC transporter substrate-binding protein	U
SAG0963	NP_687975.1	Hypothetical protein	U
SAG0886	NP_687900.1	Hypothetical protein	U
SAG0833	NP_687848.1	Hypothetical protein	U
SAG0451	NP_687482.1	Hypothetical protein	CM
SAG2056	NP_689042.1	Hypothetical protein	U
SAG1491	NP_688485.1	Hypothetical protein	U
SAG1762	NP_688752.1	Hypothetical protein	Ex
SAG0013	NP_687049.1	Hypothetical protein	U
SAG1227	NP_688234.1	Hypothetical protein	U
SAG1214	NP_688223.1	Hypothetical protein	CM

**Table 3.1 IDRIS generated report listing proteins containing peptides distinctive to GBS**

This report, generated by IDRIS, lists all the proteins containing peptides specific to *Streptococcus agalactiae* that are conserved across all sequenced strains, and are also predicted to have a signal peptide (SignalP score of over 0.45). The table details the protein name or putative function and the PSORTb predicted localisation (Ex=extracellular, CW=cell wall, CM=cytoplasmic membrane, U=unknown).

The IDRIS report is limited by the quality and quantity of data and annotations inputted into RefSeq. The IDRIS report is subject to change in later versions as the database is populated with updated RefSeq data. IDRIS v1.0 comprised of 2.5 billion tokens from 2,764 organisms. In the most recent update the dataset consisted of 7.65 billion tokens from 7,059 organisms.

### 3.3 Protein localisation

Minimal research has been conducted on SAG0771, SAG1473 and SAG1474 proteins, therefore bioinformatic tools are essential to determine protein localisation. IDRIS utilises the programmes PSORTb, SignalP and LipoP to provide information on the proteins it returns. These tools, alongside TMHMM, were repeated manually on several proteins to gain a clear insight into the protein localisation.

#### 3.3.1 PSORTb

PSORTb v3.0 recognises 25 protein sequence motifs and six profiles for gram positive bacteria to categorise proteins into four locations: extracellular, cell wall, cytoplasmic membrane and cytoplasmic. Profiles are similar to sequence motifs but also incorporate position specific weighting into the criteria. (Yu *et al.*, 2010) PSORTb classifies SAG0071, SAG1473 and SAG1474 as cell wall associated proteins, see table 3.2, based on the identification of the LPXTG profile; there were no sequence motifs detected.

	SAG0771	SAG1473	SAG1474
<b>Extracellular</b>	0.01	0.01	0.01
<b>Cell wall</b>	9.98	9.98	9.99
<b>Cytoplasmic membrane</b>	0.01	0.01	0.00
<b>Cytoplasmic</b>	0.00	0.00	0.00

**Table 3.2 Cellular location as predicted by PSORTb**

Localisation scores out of a total value of 10.00. SAG0771, SAG1473 and SAG1474 are predicted to be cell wall associated proteins.

### 3.3.2 SignalP

Signal peptides direct proteins to their destination within the cell or to the secretory pathway. Signal peptides share several common features: an N-terminal positively charged region, a hydrophobic region and a C-terminal protease cleavage site. (Jain *et al.*, 1994) There are several computer programmes designed to predict the presence of, and the site of, signal peptides from an amino acid sequence; SignalP 4.0 is reported to be the best prediction tool. (Petersen *et al.*, 2011) The SignalP D-score is calculated from a combination of scores representing the prediction of cleavage sites and the likelihood of residues forming a signal peptide. Positive results are reported if the D-score is over 0.45 (range: 0-1). Using this cut off point, the probability of a false positive result is 0.001. (Petersen *et al.*, 2011) The proteins investigated were all found to have a signal peptide, see figure 3.3.

Protein	D-score	Signal peptide?	Cleavage site - sequence position	Cleavage site - amino acids
<b>SAG0771</b>	0.750	Yes	32-33	AKA-ET
<b>SAG1473</b>	0.731	Yes	27-28	GVA-DT
<b>SAG1474</b>	0.696	Yes	27-28	IYA-NS

**Table 3.3 Signal P results for SAG0771, SAG1473 and SAG1474.**

Proteins with D-scores above 0.45 are reported as containing a signal peptide. It is predicted that SAG0771, SAG1473 and SAG1474 have signal peptides with cleavage following residue 32, 27 and 27 respectively.

### 3.3.3 LipoP

Similarly to signal peptides, lipoprotein signal peptides have recognisable N-terminal motifs. These are composed of an N-terminal positively charged region, a short hydrophobic region and a C-terminal polar region adjacent to the protease cleavage site, called a lipobox. (Juncker *et al.*, 2003) LipoP distinguishes between four classes of protein sequences: signal peptides, lipoprotein signal peptides, N-terminal transmembrane helices and cytoplasmic regions. Scores for each classification are reported and represented graphically. Cleavage sites for signal peptides and lipoprotein signal peptides are also predicted. (Juncker *et al.*, 2003) Although developed for predicting lipoprotein sequences in gram-negative bacteria, there have also been reports of success in gram-positive bacteria. The probability of a false result

is 0.048. (Rahman *et al.*, 2008) LipoP did not classify any of the proteins tested as lipoproteins, see figure 3.4.

Protein	Signal peptide score	Lipoprotein signal peptide score	N-terminal transmembrane helices score	Cytoplasmic region score	Result	Site of signal peptide cleavage Sequence position
<b>SAG0771</b>	10.65	< -3	-0.39	-0.20	Signal peptide	30-31
<b>SAG1473</b>	17.96	13.32	< -3	-0.20	Signal peptide	27-28
<b>SAG1474</b>	16.24	< -3	-2.57	-0.20	Signal peptide	27-28

**Table 3.4 LipoP results for SAG0771, SAG1473 and SAG1474.**

The highest scoring class of protein type is reported as the predicted result. It is predicted that SAG0771, SAG1473 and SAG1474 have signal peptides with cleavage following residue 30, 27 and 27 respectively.

### 3.3.4 TMHMM

Cellular membranes are lipophilic environments where charged or polar amino acids are not tolerated well, therefore transmembrane (TM) helices generally consist of non-polar amino acid sequences. They can also be characterised by their length, around 17 amino acids, to span the membrane. (Hildebrand *et al.*, 2004) TMHMM 2.0 is reported to be the best computational programme to predict TM helices, with 82.7% of membrane spanning regions predicted correctly. (Moller *et al.*, 2001) This decreases when signal peptides are considered, with 60% of these peptides incorrectly identified as TM helices. (Krogh *et al.*, 2001) This statistic is reflected in the TMHMM output of the three selected proteins, SAG0771, SAG1473 and SAG1474. Two out of three signal peptides previously identified by SignalP and LipoP were predicted to be transmembrane helices by TMHMM. It is predicted that SAG0771 and SAG1473 have C-terminal TM helices, however as a cleavage event is likely to occur at the LPXTG motifs during anchorage to the cell wall, (Navarre and Schneewind, 1994) this region would not be present in the mature protein, see figure 3.5.

Protein	Inside		TM helix		Outside		TM helix		Inside	
	Start	End	Start	End	Start	End	Start	End	Start	End
<b>SAG0771</b>					1	483	484	506	507	512
<b>SAG1473</b>	1	6	7	26	27	168	169	186	187	192
<b>SAG1474</b>	1	4	5	27	28	680				

**Table 3.5 TMHMM results for SAG0771, SAG1473 and SAG1474.**

TMHMM predicts which regions within a protein sequence are internal, external and transmembrane. The N- and C-terminal regions are likely not to be present in the processed protein due to removal of the signal sequence and cleavage at the LPXTG motif. The remainder of peptide sequences in the three proteins, SAG0771, SAG1473 and SAG1474, are predicted to be external to the membrane.

### 3.3.5 Surface exposure

Cell wall association of proteins does not guarantee surface exposure. A study investigating cell surface exposed proteins incubated *S. agalactiae* cells with proteases trypsin and protease K, and the resulting peptide sequences were analysed by mass spectroscopy. To identify the origin of each peptide, sequences were compared to genome data. Of the 22 proteins predicted by PSORTb to be cell wall associated, nine were found to be surface exposed. SAG1473 and SAG1474 were established as being surface exposed; SAG0771 was not. (Doro *et al.*, 2009) Additionally the protease-released peptide from SAG1474 was identified as an IDRIS token, supporting the selection of this peptide as a promising target for antibody production.

### 3.3.6 Protein localisation consensus

In a previous study SAG0771 expression was found to be dependent on sortase A (SrtA), an enzyme which anchors proteins to the cell wall via cleavage of the LPXTG motif. Western blotting experiments showed a reduction in SAG0771 concentration in SrtA knockout strains compared to wild type. (Nobbs *et al.*, 2008) SAG1473 and SAG1474 also contain LPXTG motifs, and are therefore likely to be under similar control. Manual inspection of IDRIS data and previous literature supported the three candidates for further work as they are predicted to be cell wall associated proteins.

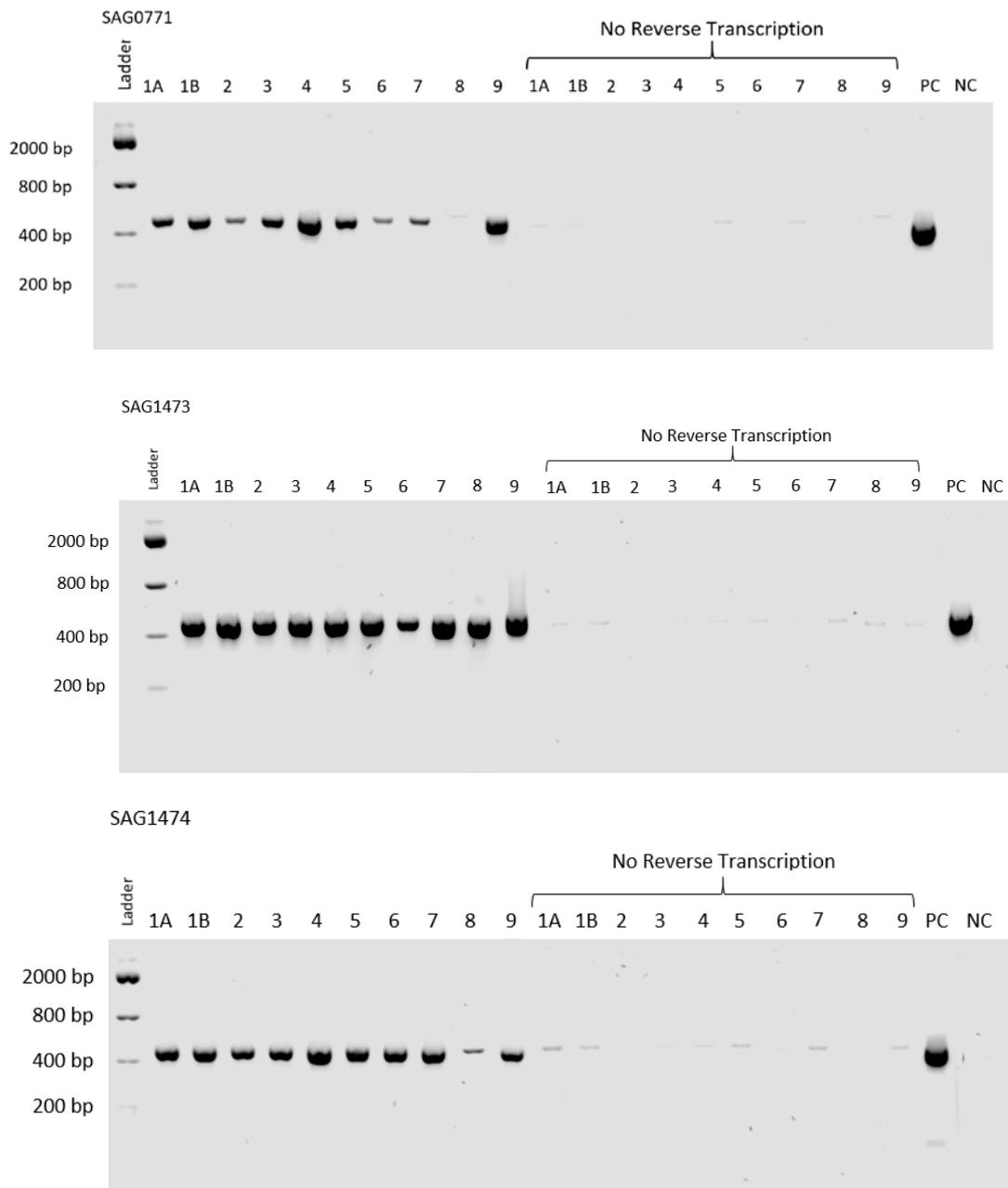
### 3.4 Protein selection

IDRIS identifies promising targets using sequencing data. As these proteins have not necessarily been empirically observed previously, there is no assurance that proteins are expressed at all, or at levels feasible for diagnostic detection. Proteins should be studied experimentally prior to selection. The optimal protein would be highly expressed in all strains under varying conditions.

#### 3.4.1 Experimental analysis of protein expression in 10 serotypes

As protein expression profiles vary across serotypes, (Lindahl *et al.*, 2005) it is important to select diagnostic targets contained within all strains. Although the link between protein expression and serotype is not complete, (Tettelin *et al.*, 2002; Creti *et al.*, 2004; Shabayek *et al.*, 2014) if a protein target is found within an example of each of the 10 different serotypes, there is a considerable likelihood that it will be found in all strains.

To estimate protein expression levels, reverse transcription PCR (RT-PCR) was performed in 10 strains of GBS representing different serotypes: 1A, 1B, 2-9. RNA was purified from cultures grown at 37 °C in 5% CO<sub>2</sub> to a mid-log phase (OD<sub>600 nm</sub> 0.5). Contaminating DNA was removed via DNase treatment and reverse transcription was performed to produce cDNA. PCR amplification of 500 bp fragments using primers specific to SAG0771, SAG1473 and SAG1474 was completed. PCR products were imaged following electrophoresis on an agarose gel, see figure 3.2.



**Figure 3.2 Transcript expression in 10 serotypes of GBS**

Extracted RNA was reverse transcribed to cDNA and amplified by PCR to analyse transcript presence in the 10 serotypes of GBS (1A, 1B, 2-9). Negative control PCR experiments containing RNA without a reverse transcription step were used to validate each sample. An additional negative control (NC) contained no nucleic acid. Extracted DNA was amplified as a positive control (PC). Representative agarose gels are shown. Presence of SAG0771, SAG1473 and SAG1474 transcripts was studied. Results show SAG1473 and SAG1474 transcripts are expressed across all 10 serotypes, SAG0771 transcript expression is reduced to a large extent in serotype 8.

In summary, SAG1473 transcripts were consistently expressed across all strains tested. SAG1474 transcripts were expressed across all strains with serotype 8 showing reduced expression. Levels of SAG0771 expression varied across strains with serotype 8 showing minimal or no expression. As RT-PCR is a crude method of analysing protein levels, the reduced levels of SAG1474 transcript may or may not be significant.

Contaminating DNA is a common problem associated with RT-PCR assays in which you cannot differentially amplify cDNA. In some cases primers can be designed which span exon-exon boundaries, however this is not possible when studying bacterial samples. Contaminating DNA was examined by amplifying extracted RNA without prior conversion to cDNA. As contaminating DNA was found to contribute between 0.1-10.6% of the signal of the test samples, it was considered acceptable. This was determined by calculating and comparing the band density of the images produced.

RNA transcript analysis is a crude indicator of protein expression as it is not influenced by RNA stability, ribosome binding site accessibility, translation rate or protein degradation. (Cham *et al.*, 2003; Van Assche *et al.*, 2015) Previous studies have identified peptides specific to SAG1473 and SAG1474, (Doro *et al.*, 2009) indicating that transcripts are then translated.

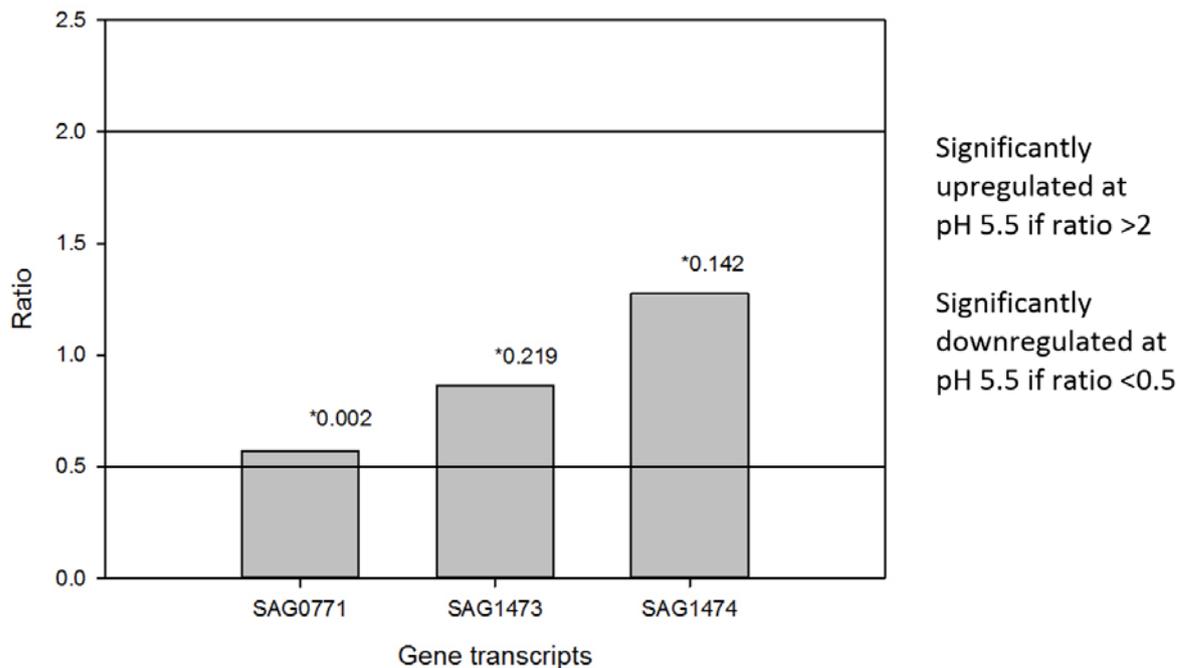
#### 3.4.2 Microarray data from previous literature

*Streptococcus agalactiae* is present at many locations within the human body, including the vagina, rectum, blood, brain and lungs. (Heath *et al.*, 2004) The optimum protein target would be expressed at comparable levels under the diverse conditions found at these sites, thus increasing the chances that a diagnostic test could be adapted for multiple sample types for multiple patient groups.

##### 3.4.2.1 pH

Vaginal and rectal swabs are commonly used in combination for GBS diagnostics, (Public Health England, 2015b) these sites vary in pH significantly from pH 3.8-4.0 to

7.0-8.0 respectively. (Jantzen *et al.*, 1989; Riedewald *et al.*, 1990) The study of low pH values is problematic due to slow growth under laboratory conditions (Yang, 2011) therefore the effect of pH on the GBS transcriptome has been studied by culturing cells at pH 5.5 and pH 7.0. (Santi *et al.*, 2009) There were no significant changes in transcriptome expression across the three proteins that were identified as potential targets, see figure 3.3.



**Figure 3.3 The effect of pH on gene transcript expression**

Changes in GBS gene expression in response to changes in pH were analysed by microarray. (Santi *et al.*, 2009) P values were calculated by an unpaired 2 tailed student's t test. A two-fold difference of expression with a p value of  $\leq 0.01$  is reported as a statistically significant change of expression. No transcripts exhibit significant differential expression at pH 5.0 and pH 7.0. Data accessible at ArrayExpress database at EMBL-EBI ([www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress)) under accession number E-TABM-708.

#### 3.4.2.2 Temperature

GBS infection is often associated with a fever. The effect on the transcriptome of incubation at 42 °C, to emulate fever, was compared to *S. agalactiae* incubated at 37 °C, simulating normal body temperature. (Klinzing *et al.*, 2009) SAG0771 showed a significant rise in transcript expression at increased temperature, however there was no significant change observed in SAG1473 or SAG1474, see figure 3.4.

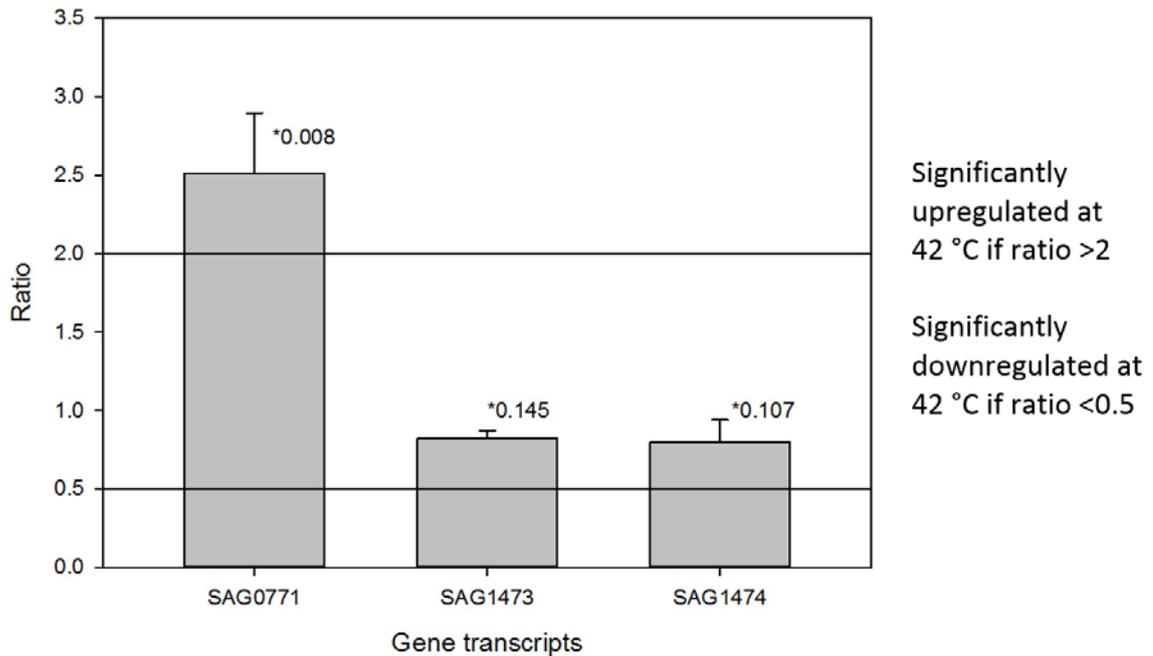


Figure 3.4 The effect of temperature on gene transcript expression

Ratio of gene transcript expression at 42 °C compared to 37 °C (Klinzing *et al.*, 2009) P values were calculated by an unpaired 2 tailed student's t test. A two-fold difference of expression with a p value of  $\leq 0.01$  is reported as a statistically significant change of expression. SAG0771 is significantly upregulated at 42 °C in comparison to 37 °C. Data accessible at NCBI GEO, (Edgar *et al.*, 2002) accession GSE14573.

#### 3.4.2.3 Oxygen concentration

*Streptococcus agalactiae* is a facultative anaerobe, therefore can exist in varying concentrations of oxygen. To evaluate the effect of oxygen concentration on the transcriptome, a previous study compared cultures after incubation at 0% and 12% oxygen. No microarray data has been published, however the paper lists all surface exposed proteins which were downregulated at 0% oxygen. SAG0771, SAG1473 and SAG1474 do not feature on this list. (Johri *et al.*, 2007)

The data collected both experimentally and through study of microarray data supports the use of SAG1473 or SAG1474 as tools to generate antibodies against, as they are transcribed consistently under varying conditions and across all GBS serotypes.

### 3.4.3 Structure prediction

Knowing the correct structure of each protein would enable epitopes to be selected based on their exposure to the protein surface, and their position relative to other potential epitopes and the cell wall binding region. Antibodies against two spatially distinct sites would enable the generation of a capture and detection-based diagnostic test against a single protein. As the protein structures are unknown, predicted models are the most suitable alternative, but must be considered with their associated confidence levels. I-TASSER is a computational tool that uses amino acid sequences to predict protein structure. (Roy *et al.*, 2010) I-TASSER was used to provide five predicted structures for each protein, figure 3.5 shows the structures with the highest C-scores, and therefore the models with highest confidence. In validation studies, known protein structures were compared to their I-TASSER prediction. TM-scores were assigned to these comparisons with a score of 1 representing a perfect match. A TM-score of  $\geq 0.5$  is considered a correct fold, and a TM-score of  $< 0.17$  is considered a random pairing. A correlation between C-score and TM-score was observed; therefore C-score is able to predict matching to the true structure. A C-score of  $\geq -1.5$  equates to a likely correct fold. (Roy *et al.*, 2010) SAG0771 was the only sequence to return a reliable structure, however the structure for SAG1474 does approach reliability (TM=0.44). Using SAG0771 as the antibody target would enable informed epitope selection. The model confidence is decreased in SAG1474 and further in SAG1473.

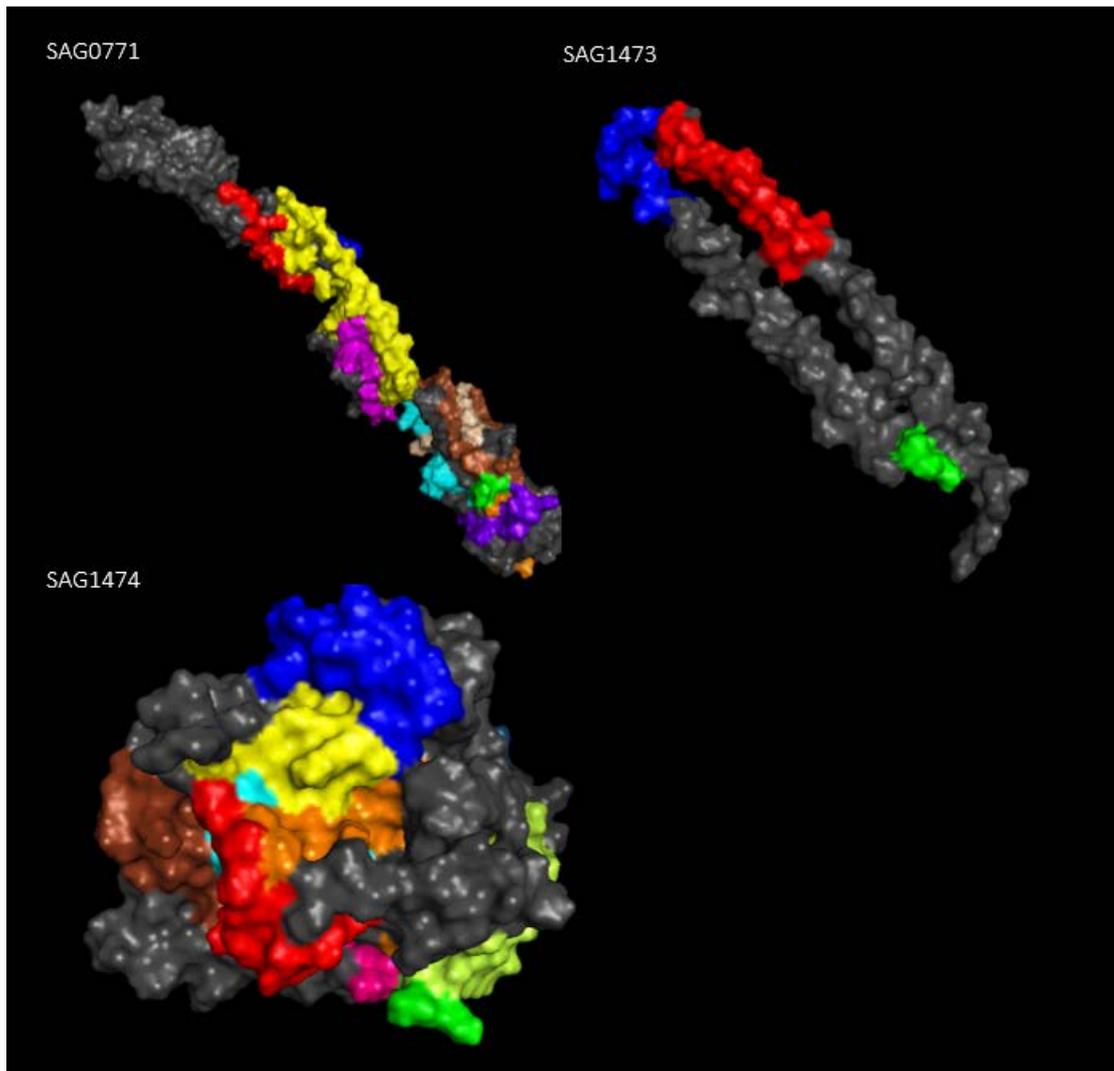


Figure 3.5 Structure of three candidate proteins with IDRIS tokens mapped on. The structures of SAG0771, SAG1473 and SAG1474 were predicted by I-TASSER, and visualised by PyMOL. (Schrodinger, 2015) IDRIS tokens are coloured. The cell wall anchoring motif is coloured in green. Structures are reported as reliable if the TM score is  $\geq 0.5$ , a TM score  $< 0.17$  is considered a random prediction.

SAG0771: C-score: -1.28, TM-score: 0.544, 8 tokens covering 218 amino acids.

SAG1473: C-score: -3.29, TM-score: 0.289, 2 tokens covering 42 amino acids.

SAG1474: C-score: -2.08, TM-score 0.440, 13 tokens covering 335 amino acids.

#### 3.4.4 Evaluation of expression and structural modelling data

Table 3.6 outlines data collected to inform the decision of which protein to raise antibodies against. Despite the reduced reliability in the proposed structure in comparison to SAG0771, SAG1474 was selected for antibody generation, largely due to its surface exposure. SAG1473 was not selected as it only contains two IDRIS

tokens. If antibodies were not able to be created against either token, the project would be limited to whole-cell diagnostics. Selecting SAG1474 with multiple tokens increases the likelihood of developing a system which can detect proteins.

SAG0771	SAG1473	SAG1474
Cell wall associated	Cell wall associated	Cell wall associated
Large reduction in transcript expression in serotype 8	Transcript expressed across all serotypes	Transcript expressed across all serotypes, small reduction in transcript expression in serotype 8
No significant change in transcript expression at decreased pH	No significant change in transcript expression at decreased pH	No significant change in transcript expression at decreased pH
Significant increase in transcript expression at elevated temperature	No significant change in transcript expression at elevated temperature	No significant change in transcript expression at elevated temperature
Protein: Not proven to be surface exposed	Protein: Surface exposed	Protein: Surface exposed
Token: Not proven to be surface exposed	Token: Not proven to be surface exposed	Token: Surface exposed
High confidence in predicted structure	Low confidence in predicted structure	Intermediate confidence in predicted structure
Multiple tokens	2 tokens; would need both regions to be antigenic to develop non-whole cell test	Multiple tokens

**Table 3.6 Summary of results, collated to inform protein target selection**

A series of experiments and computational predictions are summarised in the table. These form the basis of the discussion of which protein target to select for antibody generation. SAG1474 was selected as it has given the most favourable results.

In a previous study, the genomes of *Streptococcal agalactiae*, *Streptococcus pyogenes* and *Streptococcus pneumoniae* were compared and several cell-wall associated proteins were identified as being unique to *S. agalactiae*, these being SAG0677, SAG0771, SAG1052, SAG1331, SAG1473 and SAG1996. SAG1474 was

not identified as unique. (Tettelin *et al.*, 2002) Identifying peptides within SAG1474 as unique, indicates the significance of IDRIS.

### 3.5 Token selection

There are 13 IDRIS tokens within SAG1474, covering 328 amino acids of the sequence. Peptides for antibody production are typically between 10 and 20 amino acids in length, (Hancock and O'Reilly, 2004) therefore sequences of this length were selected for antibody production. Residue polarity and predicted secondary structure, surface exposure and a combined prediction of antigenicity were used to estimate which peptides would generate successful antibodies.

#### 3.5.1 Unique peptides

IDRIS identifies stretches of 15-mer amino acids which are unique to *Streptococcus agalactiae*. If there are peptide sequences within other organisms which differ by a single amino acid IDRIS will not report these, however the specificity of any resulting antibodies may be compromised. Using BLAST to check for similar sequences allows uniqueness to be evaluated.

Similarity matches to proteins found in organisms which would not be present in a patient sample can be disregarded. For whole-cell diagnostics only cell surface proteins need to be considered; for lysed cell diagnostics, proteins would have to contain sequences matching both the capture and detection epitopes to pose a specificity problem. The likelihood of cross-reactivity of the resulting diagnostic is limited, and for the final peptides selected for antibody production, no substantial matches have been identified.

An additional use for BLAST results is to confirm that peptides do not match proteins found within mice. For antibody production, the selected peptides must be capable of evoking a murine immune response. Cells expressing antibodies which exhibit binding to a self-protein would be eliminated during clonal deletion. (Rose, 2015) If there is

similarity between a selected sequence and a self-protein, the ability to produce antibodies against it may be compromised. No murine proteins with high similarity to the selected peptides were identified.

### 3.5.2 Polarity

Antibody-antigen interactions are dominated by Van der Waals forces and polar attractions. When designing an antigenic peptide, regions with high charge should be avoided. (Hancock and O'Reilly, 2004) An additional consideration is that peptides must be soluble in order to conjugate to a carrier protein prior to immunisation. The inclusion of polar and charged residues facilitates solubility. Amino acids were classified into three groups: non-polar (neutral (G, A, V, I, L, M, F, Y, W)), polar (neutral with electron dense regions (S, T, N, Q, C, P)), and charged amino acids (R, H, K, D, E), shown in table 3.7. Within the tokens no regions of high charge were identified. Peptide sequences with a distribution of non-polar, polar and charged residues were selected for antibody production.

### 3.5.3 Secondary structure

For a peptide-generated antibody to be functional in binding a protein, the three-dimensional conformation of the epitope must be identical in both formats. Secondary structure can be used to determine whether this will be likely.  $\beta$ -sheets are often formed with distant regions of protein sequence, (Baldi *et al.*, 2000) therefore a peptide that resides in this structure cannot be held in the same conformation without the support of the rest of the protein.  $\alpha$ -helical peptides which are long enough to form the complete helix may display the same shape as in the protein. (Hancock and O'Reilly, 2004) Unstructured loops may not be influenced by adjacent sequences within the protein, therefore holding the same conformation in its free and constrained form. Alternatively loops may lack a consistent structure, and therefore be unable to bind antibodies. Secondary structure was analysed by combining results from the prediction tools I-TASSER and PSIPRED, results can be found in table 3.7. The position specific iterated BLAST tool identifies sequences similar to the input. PSIPRED is a machine learning programme which uses this data to predict secondary structure. (Jones, 1999) PSIPRED 3.2 is able to correctly predict the secondary structure for 81.6% of residues.

(Bloomsbury Centre for Bioinformatics, n.d.) In peptide selection, regions of  $\beta$ -strand structure were avoided.

#### 3.5.4 Surface exposure

Exposure of peptides to the protein surface is essential for antibody binding, the computing tools I-TASSER and prodepth were used to predict the likelihood of this. Prodepth estimates the distance from each residue to the nearest water molecule, with distances over 3.0 Å suggesting that the amino acid is buried within the protein, and distances of under 3.0 Å suggesting that residues are exposed. Prodepth has an accuracy of 78.2%. (Song *et al.*, 2009) I-TASSER results were visualised with PyMOL (Schrodinger, 2015) and manually inspected for surface exposure, the results can be seen in table 3.7.

#### 3.5.5 Combination of predictive tools

Polarity, surface exposure and secondary structure are useful tools to predict whether a peptide sequence will be antigenic. In previous studies, flexibility, hydrophilicity and the presence of certain amino acids at numbered positions within turns have also been used to inform antigenic predictions. (Karplus and Schulz, 1985; Parker *et al.*, 1986; Pellequer *et al.*, 1993) The programme Bcepred predicts whether peptides are likely B cell epitopes by consolidating a selection of these results and reporting an overall positive or negative prediction. A study has found the optimal combination of features to be hydrophilicity, flexibility, surface accessibility and polarity. Under these settings Bcepred was able to correctly identify antigenic regions with an efficiency of 58.70%. (Saha and Raghava, 2004) These properties were used in the study of SAG1474 IDRIS tokens and data can be found in table 3.7.

Residue	Charge	Exposure	Structure	Antigenic
P	N	E	L	N
T	P	E	L	NN
T	P	U	L	NN
N	P	E	L	NN
T	P	E	LH	NN
I	N	U	LH	NN
V	N	E	LH	Y
Q	P	E	LH	Y
T	P	E	L	Y
N	P	E	L	Y
D	C	E	L	Y
S	P	E	L	Y
N	P	E	L	Y
P	N	E	L	Y
T	P	E	LH	Y
A	N	U	LH	Y
K	C	U	LH	Y
F	N	U	LH	N
Token 2				
L	N	B	LH	NN
F	N	U	LH	NN
K	C	U	LH	NN
A	N	U	LH	NN
H	C	B	LH	NN
S	P	E	LH	NN
S	P	U	LH	NN
L	N	U	LH	NN
V	N	U	L	Y
N	P	U	L	Y
L	N	U	L	Y
E	C	U	L	Y
E	C	E	L	Y
N	P	E	L	Y
S	P	U	L	Y
Q	P	E	L	Y
V	N	U	L	Y
Token 3				
D	C	E	LH	Y
K	C	E	LH	Y
R	C	U	LH	Y
A	N	E	LH	Y
I	N	E	LH	N
Y	P	U	LH	NN
N	P	E	LH	NN
M	N	E	LH	NN
E	C	U	L	Y
N	P	E	L	Y
L	N	E	L	Y
S	P	E	LH	Y
Q	P	E	LH	Y
E	C	E	LH	Y
R	C	E	LH	Y
I	N	E	LH	N
A	N	B	LH	NN
L	N	U	LH	NN
F	N	B	LH	NN
N	P	B	LH	NN
R	C	E	LH	NN
Q	P	U	LH	NN
W	P	B	H	NN
E	C	B	H	NN
P	C	U	H	NN
M	N	B	H	NN
L	N	B	LH	NN
R	C	B	LH	NN
R	C	B	LH	NN
T	P	B	LH	N

Residue	Charge	Exposure	Structure	Antigenic
V	N	U	H	N
N	P	U	H	NN
M	N	E	H	NN
A	N	U	H	NN
Y	P	U	H	NN
D	C	U	H	NN
I	N	U	H	NN
I	N	U	H	NN
A	N	U	H	NN
K	C	E	H	NN
E	C	E	H	NN
N	P	U	LH	NN
P	N	U	LH	NN
S	P	E	LH	NN
L	N	E	LH	NN
N	P	U	LH	NN
A	N	B	LS	NN
V	N	U	LS	NN
I	N	U	S	N
T	P	U	LS	Y
T	P	U	LS	Y
R	C	U	L	Y
Q	C	E	H	Y
R	P	U	H	Y
E	C	E	H	Y
A	N	E	H	Y
I	N	U	H	N
E	C	U	H	NN
A	C	E	H	NN
R	C	U	H	NN
K	C	U	LH	NN
L	N	E	L	NN
K	C	U	L	NN
D	C	E	L	NN
T	P	E	L	NN
Q	P	E	L	NN
N	P	E	L	NN
P	N	B	L	NN
F	N	B	L	NN
L	N	B	L	NN
Token 5				
F	N	U	L	N
D	C	U	L	NN
S	C	U	LH	NN
S	P	U	H	NN
Y	P	U	H	NN
V	N	U	H	NN
K	C	U	H	NN
K	C	U	H	NN
Y	P	B	H	NN
K	C	U	H	NN
D	C	E	H	NN
L	N	U	LH	NN
G	N	B	L	NN
F	N	B	L	NN
I	N	B	LS	NN
L	N	B	LS	NN
I	N	B	LS	NN
L	N	B	LS	NN
G	N	B	LS	NN
Q	P	B	LS	N
T	P	B	LS	N

Residue	Charge	Exposure	Structure	Antigenic
M	N	B	LS	N
T	P	B	LS	NN
P	N	B	LS	NN
I	N	B	LS	NN
A	N	B	LS	Y
S	P	B	LS	Y
G	N	B	L	Y
S	P	B	L	Y
D	C	B	L	Y
A	N	B	L	Y
G	N	B	L	Y
G	N	B	LH	Y
S	P	B	LH	Y
I	N	B	LH	N
R	C	B	LH	NN
I	N	B	LH	NN
P	N	B	LH	NN
S	P	B	LH	NN
S	P	B	LH	NN
W	P	B	LH	NN
T	P	B	LH	NN
G	N	B	L	NN
L	N	U	L	NN
V	N	B	L	NN
G	N	B	LS	NN
L	N	B	LS	NN
K	C	B	LS	NN
P	N	B	L	NN
T	P	B	L	NN
Token 7				
V	N	B	L	Y
S	P	B	L	Y
N	P	B	L	Y
E	C	B	L	Y
K	C	B	L	Y
P	N	B	L	Y
D	C	B	L	Y
S	P	B	L	Y
Y	P	B	L	Y
S	P	U	LH	Y
T	P	U	LH	Y
Token 8				
F	N	U	L	N
L	N	B	L	NN
P	N	B	L	NN
T	P	B	L	Y
K	C	B	L	Y
S	P	U	L	Y
S	P	B	H	Y
R	C	B	H	Y
D	C	B	H	Y
A	N	B	H	Y
E	C	B	H	Y
T	P	B	H	Y
L	N	B	H	N
L	N	B	H	NN
T	P	B	H	NN
Y	P	U	LH	NN
L	N	B	LH	NN
K	C	B	L	NN
K	C	U	L	NN
S	P	U	L	NN
D	C	U	L	NN
Q	P	E	L	NN
T	P	B	L	NN
L	N	B	L	NN
V	N	B	L	NN
S	P	B	L	NN
V	N	B	L	NN
N	P	B	L	NN
D	C	U	L	NN
L	N	U	L	NN
K	C	B	L	N

Residue	Charge	Exposure	Structure	Antigenic	Token 9				
					Residue	Charge	Exposure	Structure	Antigenic
L	N	U	L	N	F	N	F	H	N
P	N	U	LS	N	L	N	U	H	N
I	N	B	LS	N	R	C	U	LH	N
A	N	B	LS	N	K	C	E	LH	N
Y	P	B	LS	N	Q	P	U	L	N
T	P	B	LS	N	G	N	E	L	N
L	N	U	L	N	F	N	E	L	N
K	C	U	L	N	K	C	U	LS	N
S	P	U	LH	N	V	N	E	LS	N
P	N	E	LH	N	T	C	U	LS	N
M	N	E	LH	N	E	C	U	LS	N
G	N	U	L	Y	I	N	U	LS	N
T	P	U	L	Y	D	C	U	L	N
E	C	E	L	Y	L	N	U	L	N
V	N	B	L	Y	P	N	U	LH	N
S	P	B	L	Y	I	N	U	LH	N
Q	P	E	H	Y	D	C	U	LH	N
D	C	E	H	Y	G	N	U	H	N
A	N	U	H	Y	R	C	U	H	N
K	C	U	H	Y	A	N	U	H	N
N	P	U	H	Y	L	N	U	H	N
A	N	E	H	Y	M	N	U	H	N
I	N	B	H	N	R	C	U	H	N
M	N	B	H	N	D	C	B	LH	N
D	C	E	H	N	Y	P	B	LH	N
N	P	E	H	N	S	P	B	LH	N
N	N	B	H	N	T	P	B	LH	N
V	N	B	H	N	L	N	B	LH	N
					A	N	B	LH	N
					I	N	B	H	N
					G	N	B	H	N
					M	N	B	H	N
					G	N	B	H	N
					G	N	B	H	N
					A	N	B	H	N
					F	N	U	H	N
					S	P	U	H	N
					T	P	U	H	N
					I	N	U	H	N
					E	C	U	H	N
					K	C	U	H	N
					D	C	U	H	N
					L	N	U	H	N
					K	C	U	H	N
					K	C	U	LH	N

Residue	Charge	Exposure	Structure	Antigenic	Token 10				
					Residue	Charge	Exposure	Structure	Antigenic
L	N	U	H	N	L	N	F	H	N
H	C	U	H	N	L	N	U	H	N
K	C	U	LH	N	R	C	U	LH	N
Q	P	E	LH	N	K	C	E	LH	N
F	N	E	L	N	Q	P	U	L	N
P	N	E	L	N	G	N	E	L	N
I	N	E	L	N	F	N	E	L	N
F	N	E	L	N	K	C	U	LS	N
L	N	E	LS	N	V	N	E	LS	N
S	P	U	LS	N	T	C	U	LS	N
P	N	B	LS	N	E	C	U	LS	N
L	N	B	LS	N	I	N	U	LS	N
S	N	B	LS	N	D	C	U	L	N
P	N	B	LS	N	L	N	U	L	N
T	P	U	L	N	P	N	U	LH	N
A	N	B	L	N	I	N	U	LH	N
S	P	B	L	N	D	C	U	LH	N
L	N	B	L	N	G	N	U	H	N
A	N	B	L	N	R	C	U	H	N
S	P	B	L	N	A	N	U	H	N
L	N	B	L	N	L	N	U	H	N
A	N	B	L	N	M	N	U	H	N
I	N	B	L	N	R	C	U	H	N
M	N	B	L	N	D	C	B	LH	N
T	P	B	LH	N	Y	P	B	LH	N
G	N	B	LH	N	S	P	B	LH	N
L	N	B	LH	N	T	P	B	LH	N
S	N	B	LH	N	L	N	B	LH	N
E	C	E	L	N	A	N	B	LH	N
S	N	E	L	N	I	N	B	LH	N
G	N	U	LS	N	G	N	B	H	N
L	N	U	LS	N	M	N	B	H	N
P	N	U	LS	N	G	N	B	H	N
I	N	U	LS	N	A	N	B	H	N
J	N	B	LH	N	F	N	U	H	N
A	N	B	LH	N	S	P	U	H	N
N	P	B	LH	N	T	P	U	H	N
M	N	B	L	N	I	N	U	H	N
T	P	B	L	N	E	C	U	H	N
G	N	U	L	N	K	C	U	H	N
L	N	U	L	N	D	C	U	H	N
P	N	U	L	N	L	N	U	H	N
A	N	B	LS	N	K	C	U	H	N
N	B	LH	L	N	K	C	U	LH	N
S	P	B	LS	N	S	P	B	LH	N
I	N	B	LS	N	S	P	U	LH	N
P	N	B	LS	N	V	N	E	LH	N
T	P	B	LS	N	K	C	U	LH	N
Y	P	B	LS	N	N	P	B	LH	N
L	N	B	LS	N	K	C	E	LH	N
S	P	B	LS	N	P	N	E	LH	N
E	C	E	L	N	S	P	U	LH	N
S	N	E	L	N	V	N	B	H	N
G	N	U	LS	N	M	N	U	H	N
L	N	U	LS	N	A	N	E	H	N
P	N	B	LS	N	Y	P	U	LH	N
I	N	B	LS	N	Q	C	U	LH	N
J	N	B	U	N	K	P	B	LH	N
G	N	B	U	N	A	N	E	LH	N
					L	N	U	LH	N

Residue	Charge	Exposure	Structure	Antigenic	Token 11				
					Residue	Charge	Exposure	Structure	Antigenic
L	N	U	H	N	L	N	F	H	N
H	C	U	H	N	L	N	U	H	N
K	C	U	LH	N	R	C	U	LH	N
Q	P	E	LH	N	K	C	E	LH	N
F	N	E	LH	N	Q	P	U	L	N
P	N	B	LH	N	G	N	E	L	N
I	N	U	LS	N	F	N	E	L	N
F	N	B	S	N	K	C	U	LS	N
L	N	B	S	N	V	N	E	LS	N
S	N	B	S	N	T	C	U	LS	N
P	N	B	LS	N	E	C	U	LS	N
T	P	U	L	N	I	N	U	LS	N
A	N	B	L	N	D	C	U	L	N
S	P	B	L	N	L	N	U	L	N
L	N	B	L	N	P	N	U	LH	N
L	N	B	L	N	I	N	U	LH	N
A	N	B	L	N	D	C	U	LH	N
S	P	B	L	N	G	N	U	H	N
L	N	B	L	N	R	C	U	H	N
A	N	B	L	N	A	N	U	H	N
S	P	B	L	N	L	N	U	H	N
L	N	B	L	N	M	N	U	H	N
A	N	B	L	N	R	C	U	H	N
S	P	U	L	N	D	C	B	LH	N
L	N	U	L	N	Y	P	B	LH	N
A	N	B	L	N	S	P	B	LH	N
S	N	B	L	N	T	P	B	LH	N
E	C	E	L	N	L	N	B	LH	N
S	N	E	L	N	A	N	B	LH	N
G	N	U	LS	N	I	N	B	LH	N
L	N	U	LS	N	G	N	B	H	N
P	N	U	LS	N	M	N	B	H	N
I	N	U	LS	N	G	N	B	H	N
J	N	B	U	N	A	N	B	H	N
A	N	B	U	N	F	N	U	H	N
L	N	U	LH	N	S	P	U	H	N

Residue	Charge	Exposure	Structure	Antigenic	Token 12				
					Residue	Charge	Exposure	Structure	Antigenic
K	C	U	LH	Y	L	N	F	H	N
S	P	B	LH	Y	L	N	U	H	N
S	P	U	LH	Y	R	C	U	LH	N
V	N	E	LH	Y	K	C	E	LH	N
K	C	U	LH	Y	Q	P	U	L	N
N	P	B	LH	Y	G	N	E	L	N
K	C	E	LH	Y	F	N	E	L	N
P	N	E	LH	N	K	C	U	LS	N
S	P	U	LH	N	V	N	E	LS	N
V	N	B	H	N	T	C	U	LS	N
M	N	U	H	N	E	C	U	LS	N
A	N	E	H	N	I	N	U	LS	N
Y	P	U	LH	N	D	C	U	L	N
Q	C	U	LH	N	L	N	U	L	N
K	P	U	LH	N	P	N	U	LH	N
A	N	E	LH	N	I	N	U	LH	N
L	N	U	LH	N	D	C	U	LH	N

**Table 3.7 Analysis of token suitability for use as immunogenic peptides.**

The table assesses the suitability of each residue in each token. Column 1: amino acid code for each residue; column 2: residues are classified as charged (C), polar (P) or non-polar/hydrophobic (N); column 3: residues are classified as exposed (E), buried (B) or unknown (U) i.e. methods give opposing predictions, based on predictions from I-TASSER and prodepth; column 4: residues are classified as having loop (L), helix (H), sheet (S), or unknown structure (U) i.e. methods give opposing predictions, based on predictions from I-TASSER and PSIPRED; column 5: Residues are classified as being antigenic (Y) or non-antigenic (N), based on Bcepred prediction. Green denotes favourable, orange intermediate, and red undesirable qualities for selection as an immunogenic peptide. Peptides selected to raise antibodies against are coloured grey.

### 3.5.6 Conclusion

Three peptides were chosen for antibody production: Peptide A (VQTNDSNPT), Peptide B (VNLEENSQV) and Peptide C (NMENLSQEERI). Each peptide derives from an IDRIS token, found in all *Streptococcus agalactiae* strains, and in no other bacteria. They all contain a combination of polar, charged and hydrophobic residues, exist as a loop or helix structure, are predicted to be surface exposed by at least one method, and give a positive antigenic combination score, as judged by Bcepred. Peptide A was previously identified as being exposed to the bacterial surface by protease degradation. (Doro *et al.*, 2009) The selected tokens have been mapped onto the predicted structure of SAG1474, see figure 3.6. The spatially distinct sites may allow the binding of two antibodies simultaneously, leaving open the possibility for a capture and detection antibody pair to recognise the same protein, which would be required for lysed-cell or protein detection.

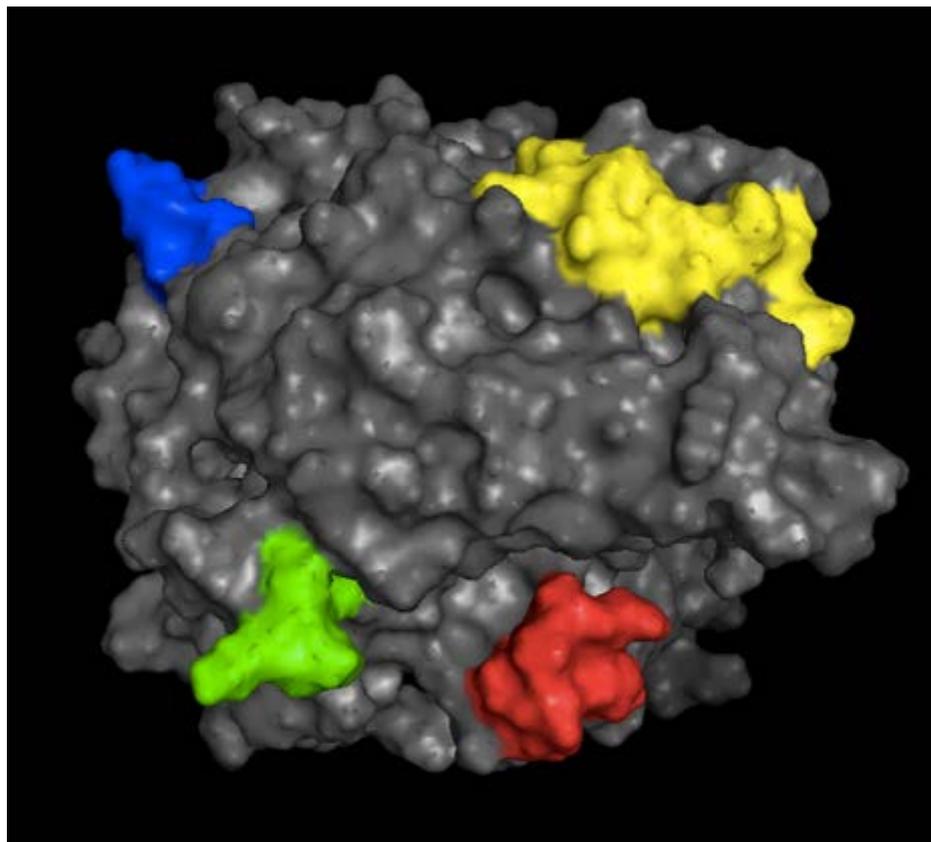


Figure 3.6 Structure of SAG1474 with selected peptides mapped on. The structure of SAG1474 was predicted by I-TASSER (C-score: -2.08; TM score: 0.440), and visualised by PyMOL. (Schrodinger, 2015) The three peptides selected to raise antibodies against are coloured (peptide A red, peptide B blue, peptide C yellow). The cell wall anchoring motif is coloured in green.

---

## Chapter 4 Antibody generation against *S. agalactiae* biomarkers

---

### 4.1 Introduction

#### 4.1.1 Immunoassays in diagnostic tests

Since the development of hybridoma technology, (Kohler and Milstein, 1975) monoclonal antibodies have become a significant tool in disease detection. Antibodies have been employed to detect streptococcal infections using various techniques such as latex agglutination, (Ascher *et al.*, 1991; Greenberg *et al.*, 1995) lateral flow tests, (Jorgensen *et al.*, 2015) ELISAs, (Egger *et al.*, 1990) and electrochemical tests. (Wang *et al.*, 2017)

#### 4.1.2 *In vivo* antibody production

Antibodies are made by exploiting an animal's immune system. Naïve antibody-producing B lymphocytes, expressing IgM and IgD, are produced in the bone marrow. (Chaplin, 2010) During initial antigen exposure, the antibody immune response is dominated by IgM molecules, which exhibit polyspecificity and weak affinity to the target. (Boes, 2000) Naïve B cells migrate to the spleen and lymph nodes. Upon pathogen exposure, antigen-specific B cells are activated by clonal selection and these cells multiply. During proliferation, B cells undergo isotype switching to IgG, IgA and IgE, and experience somatic hyper-mutation. Errors are introduced at an increased rate in comparison to typical cell replication, thus enabling the production of antibodies exhibiting stronger antigen binding. Mature B lymphocytes become memory B cells or short-lived plasma cells which attack and eliminate the threat. (Chaplin, 2010)

### 4.2 Antibody production

Short peptide sequences are not able to evoke an immune response in mice, (Hancock and O'Reilly, 2004) and are susceptible to proteolytic cleavage *in vivo*. (Vlieghe *et al.*,

2010) Peptides were conjugated to a carrier protein as larger structures are more immunogenic. BSA and KLH are commonly used as conjugation proteins. (Lateef *et al.*, 2007) KLH was selected, as use of BSA as an immunogen can cause problems in cross-reactivity with FBS whilst culturing hybridoma cells (Hancock and O'Reilly, 2004) and in downstream applications as BSA is commonly used in protein blocking. A cysteine residue was added to the N-terminus of each peptide to facilitate maleimide conjugation. This ensured all peptides were displayed in the same orientation. The immune system was presented with limited immunogen conformations, therefore the effective concentrations were high. Peptides were also conjugated to BSA to assess peptide affinity in antibody screening assays. Peptide-BSA conjugates were utilised, opposed to free peptide, so peptides were presented in a similar orientation as in antibody production *in vivo*. BSA was utilised as the carrier protein to ensure selected antibodies showed affinity to the peptide rather than the KLH moiety.

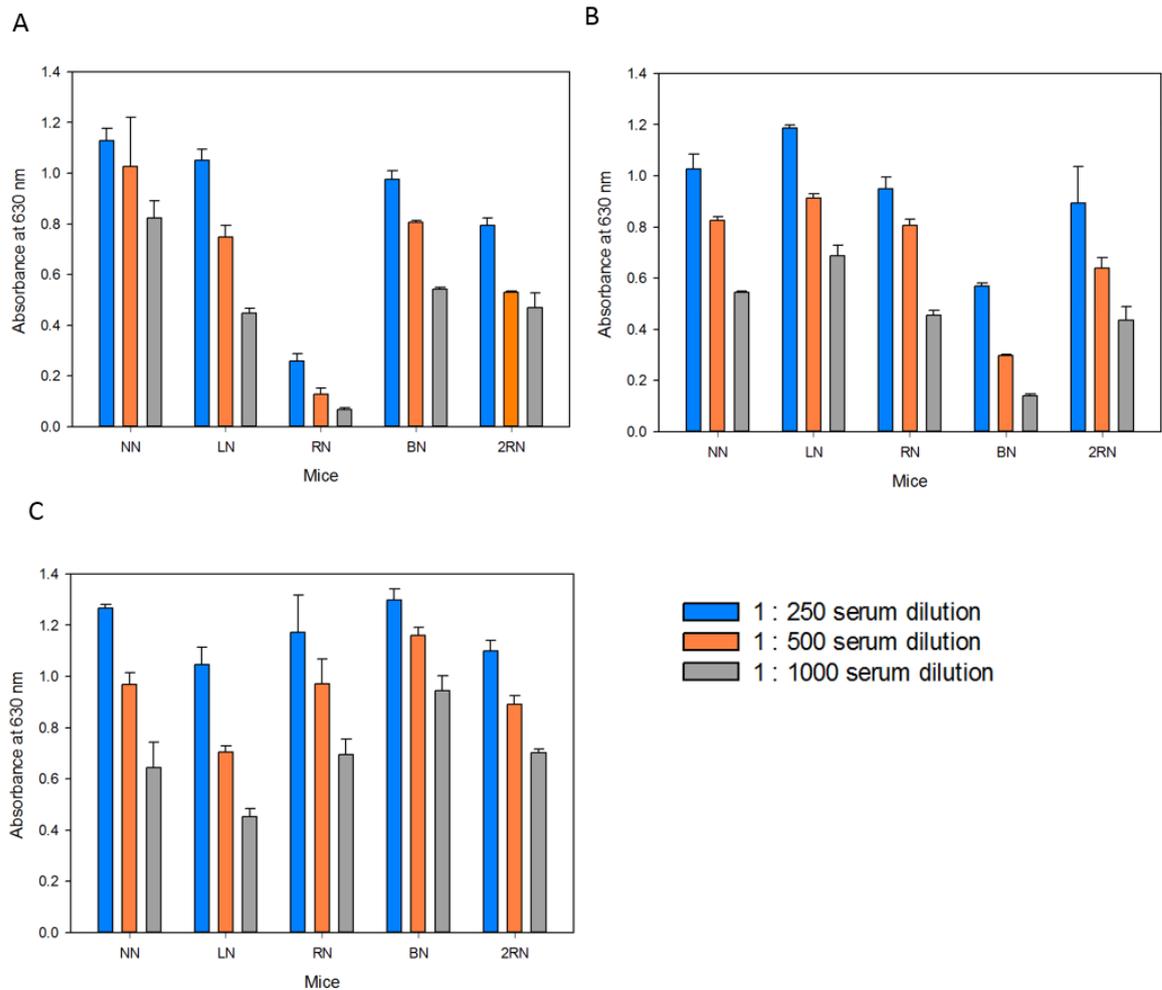
The number of free cysteine residues was determined before and after conjugation to calculate the number of peptides which had successfully bound to the carrier protein. The reduction in free cysteines represents the fraction of total peptide bound to the carrier protein. Table 4.1 shows example data from peptide conjugation to KLH.

Peptide	Conjugation efficiency / %	Number of peptides per KLH subunit	Concentration / mg ml <sup>-1</sup>
<b>A</b>	75	250-275	2.95
<b>B</b>	94	290-330	2.45
<b>C</b>	83	200-220	4.13

**Table 4.1 Conjugation of peptide to carrier protein.**

An Ellman's assay was performed to assess conjugation efficiency to the carrier protein KLH. A standard curve using known concentrations of reduced cysteine was created. The numbers of free cysteine residues within the peptide sample before and after conjugation were interpolated from the standard curve. The conjugation efficiency of the Ellman's assay was calculated and used to estimate the number of peptide molecules attached to each KLH subunit. The peptide number per KLH subunit is expressed as a range as KLH is formed of 350-390 kDa subunits. A BCA assay was performed using known BSA concentrations to create a standard curve; conjugated peptide-protein concentration was interpolated from the data, which was used to determine the dilution required for murine immunisations and screening assays. The data shown is a representative example from one round of peptide conjugation to KLH; peptide conjugation was performed several times, to KLH and BSA.

To create monoclonal murine antibodies, 15 mice were injected with immunogen, five mice for each peptide. On repeated exposure to the same antigen, memory B cells are converted to plasma cells and these circulate around the body, lowering the concentration of antigen available to induce the immune response. In mice, levels of circulating plasma cells decrease after two weeks, (Greenfield, 2014) therefore immunisations were separated by this time period. Four immunisations were performed, followed by a test bleed to assess if target-binding antibodies were being produced, see figure 4.1. 13 of the 15 mice established a strong immune response to their immunogen and were therefore suitable for antibody production. A pre-immune antigen boost was performed to cause a large amount of B lymphocytes to accumulate in the lymph nodes and spleen. (Greenfield, 2014)



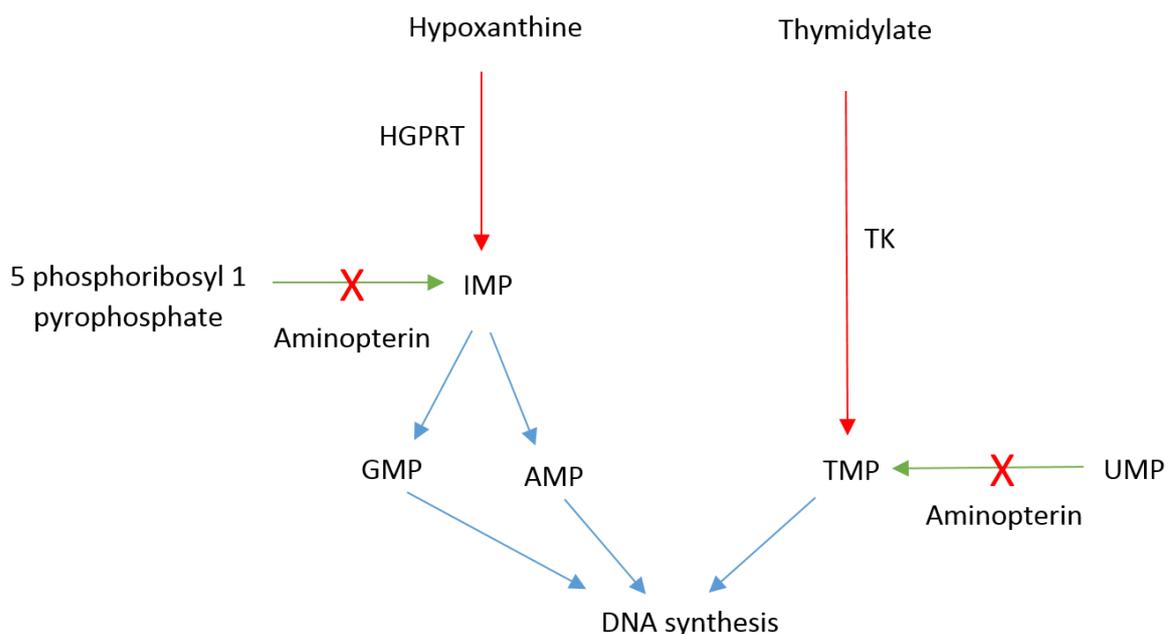
#### Figure 4.1 Binding of murine test bleeds to peptide

Affinity of murine serum to the respective peptide antigens was assessed via ELISA. A serum sample was extracted prior to immunisation to assess any background signal, these results have been subtracted from the test bleed data post immunisation. All background signals < 0.06. Error bars show standard deviation of the data, n=3. Mice are named NN, LN, RN, BN and 2RN in each cage. A: Test bleeds from mice injected with peptide A. B: Test bleeds from mice injected with peptide B. C: Test bleeds from mice injected with peptide C. All mice responded well to the immunisation except mouse A-RN and B-BN; these were not used for antibody production.

Harvested spleen cells were fused with a myeloma cell line, NS-1; (Kohler *et al.*, 1976) this process combines the immortal properties from the myeloma cell line to the antibody-producing properties of the spleen cell. Hybridomas were then selectively cultured. Cells extracted from the spleen cannot survive indefinitely, so died over time. Growth of the myeloma cell line was prohibited as it contains the deletion of a protein which is essential when cultured under certain conditions. The growth of hybridomas

was permitted as they contain the essential protein from the spleen cell, and immortality from the myeloma cell.

In hypoxanthine-aminopterin-thymidine media, cells are unable to utilise *de novo* DNA synthesis as it is blocked by aminopterin. Cells must utilise the salvage pathway, which is enabled as the precursors hypoxanthine and thymidine are present in the media. The enzyme thymidine kinase (TK) phosphorylates thymidine, forming thymidine monophosphate (TMP), a precursor for one of the four nucleotides required for DNA and RNA formation. The enzyme hypoxanthine guanine phosphoribosyltransferase (HGRPT) converts hypoxanthine to inosine monophosphate (IMP), which is a precursor for guanosine triphosphate and adenosine triphosphate, a further two nucleotides essential for DNA and RNA synthesis. Myeloma cell lines used for hybridoma creation contain deletions in one of two proteins, HGRPT or TK, and are therefore unable to utilise the salvage pathway to synthesise DNA, so are unable to survive in HAT (hypoxanthine, aminopterin, thymidine) media. (Shay, 1987) NS-1 cells are deficient in HGPRT. (Cammack *et al.*, 2006) The mechanism of HAT selection is shown in figure 4.2. The wildtype phenotype is recovered through fusion of the myeloma cell with a spleen cell, and therefore hybridomas are capable of proliferating. (Greenfield, 2014)

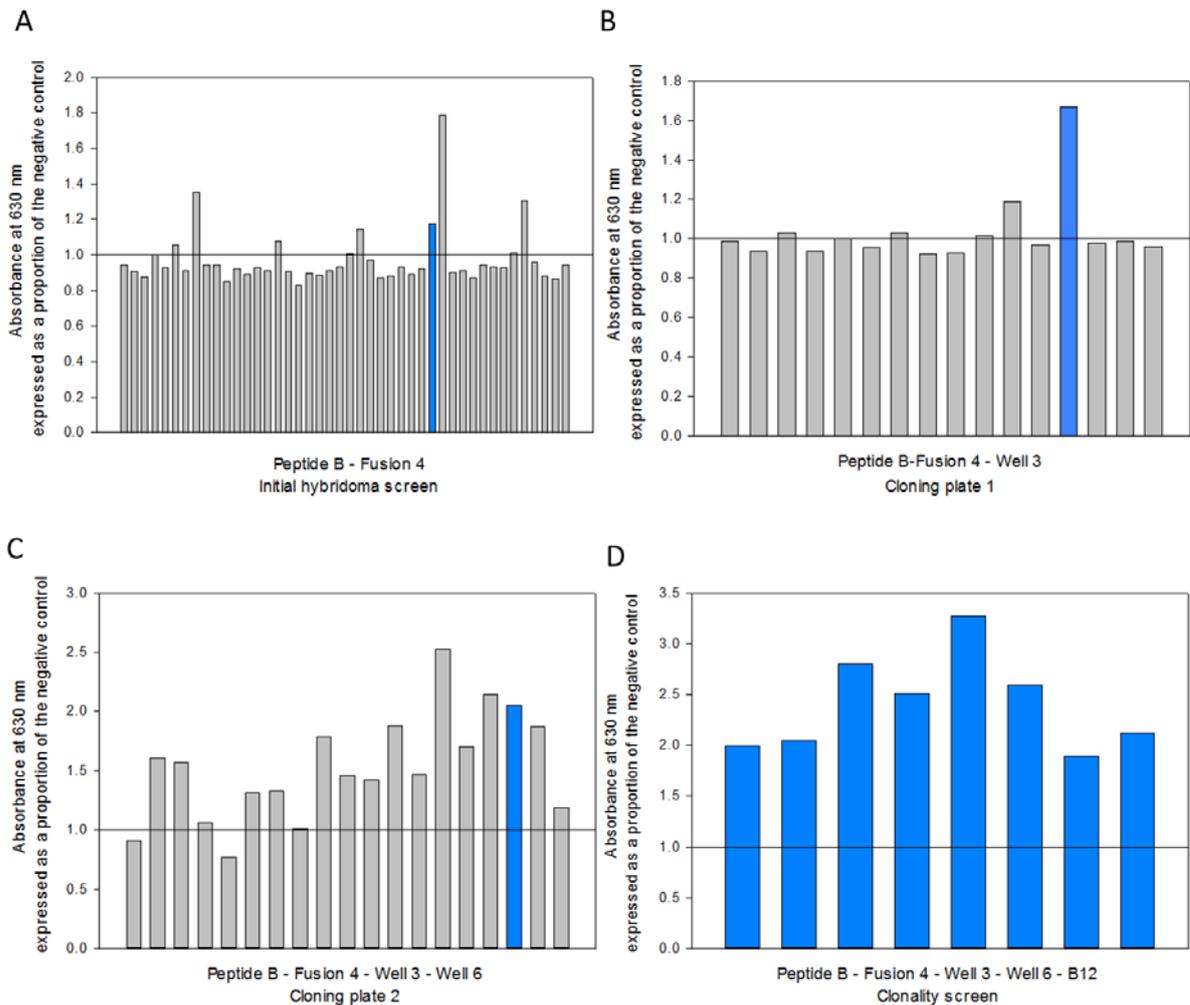


**Figure 4.2 Mechanism of HAT selection**

*De novo* synthesis of DNA is blocked in the presence of aminopterin (green pathways). Cells must utilise the salvage pathways (red) using the enzymes HGPRT and TK to synthesise precursors for DNA synthesis (common pathways shown in blue). Myeloma cell lines containing HGPRT or TK deletions are unable to utilise these salvage pathways, and in the presence of HAT media are unable to proliferate, and therefore die. Information of the *de novo* and salvage pathways from (Mofatt and Ashiharab, 2002). Abbreviations: HGPRT (hypoxanthine guanine phosphoribosyltransferase), IMP (inosine monophosphate), GMP (guanosine monophosphate), AMP (adenosine monophosphate), TK (thymidine kinase), UMP (uridine monophosphate), TMP (thymidine monophosphate).

Following cell fusion, hybridomas were seeded at low density. Cells grow optimally in response to growth factors expressed by neighbouring cells. To create a more favourable environment, hybridomas can be cultured onto a feeder layer of cells, macrophages are often used, however this can restrict the hybridoma yield (Greenfield, 2014) as cells must compete for nutrients, oxygen and space. Alternative strategies include incubation with conditioned media, the addition of interleukin 6 (IL-6), oxaloacetate pyruvate insulin, or hybridoma cloning factors. IL-6 and oxaloacetate pyruvate insulin promote cell growth, whilst IL-6 also increases antibody production. (Greenfield, 2014) Condimed, a hybridoma cloning supplement containing growth factors and cytokines, was added to low cell density cultures.

At this stage cultures were likely to comprise of a mixed population of cells. Limiting dilution screening was carried out in order to create monoclonal cell lines, in which all cells were statistically likely to be derived from a single clone, and therefore produce a uniform antibody population. Hybridomas secrete antibody into the supernatant. Once cell growth reached a sufficient level, the culture media was screened for antigen-binding antibodies. Cell cultures associated with positive binding were selected for expansion. These cells were re-seeded at a range of dilutions, spanning 600-0.1 cells per well. Wells that exhibited positive target binding which originated from more than one cell per well were re-seeded in further rounds of subcloning, until a monoclonal cell line was obtained. Figure 4.3 shows the path taken to select and clone cell line B4-3-6-B12. The naming convention for the initial term is derived from the immunising peptide (A, B or C) and which spleen was used for fusion (1-5). The subsequent naming pattern follows the number associated with the positively binding well for each round of subcloning. In this example three rounds of sub cloning were completed (3, 6, B12), see figure 4.4 for details. The cell line B4-3-6-C10 follows the same pathway of cloning, deviating only in the final positively binding well selected.



**Figure 4.3** Peptide affinity screening programme of hybridoma fusions to cloned cells

A: All cell growth positive wells from hybridoma fusion plates were assessed for binding to their peptide antigen via ELISA. B: The target-affinity positive wells were seeded into a cloning plate. Cloning plates were set up at a cell density of 600 cells per well in the first 16 wells, with 6-fold serial dilutions across the plate. The final wells statistically held less than one cell each so any growth positive wells were likely to be cloned. C: The cloning process was continued until positive ELISA results were obtained from cloned cells. D: The cells were then checked for clonality by seeding a plate at one cell per well, and assessing target binding via ELISA. Graphs show the signal obtained for each well tested as a proportion to that obtained for the negative control (incubation with cell culture fluid in comparison to PBS). From each round of screening, multiple wells which produced a higher signal than the negative control were selected for subcloning. In this figure the process to obtain the clone B4-3-6-B12 is shown in blue. The final graph shows that all growth positive wells, which were statistically derived from a single cell, had affinity for the peptide, therefore a monoclonal cell line with target affinity was produced.

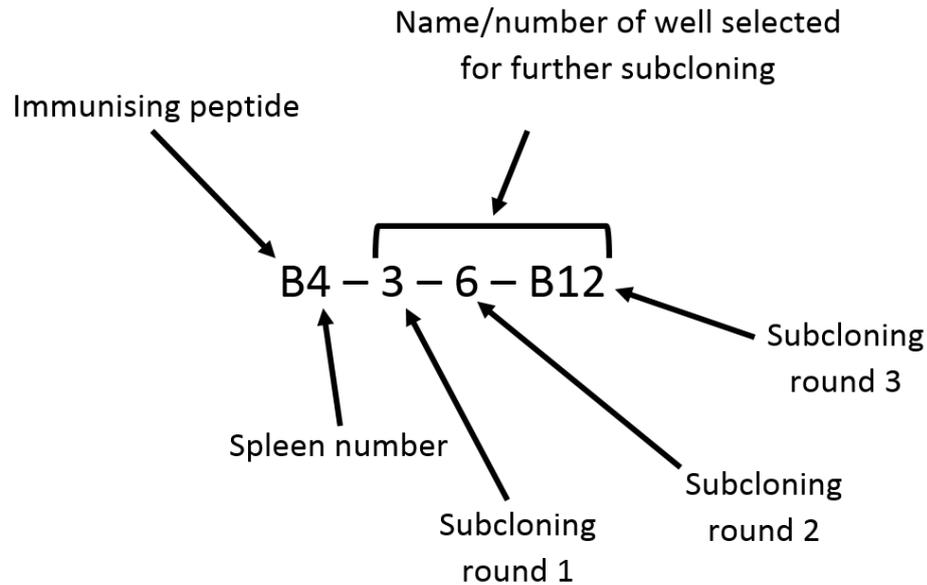


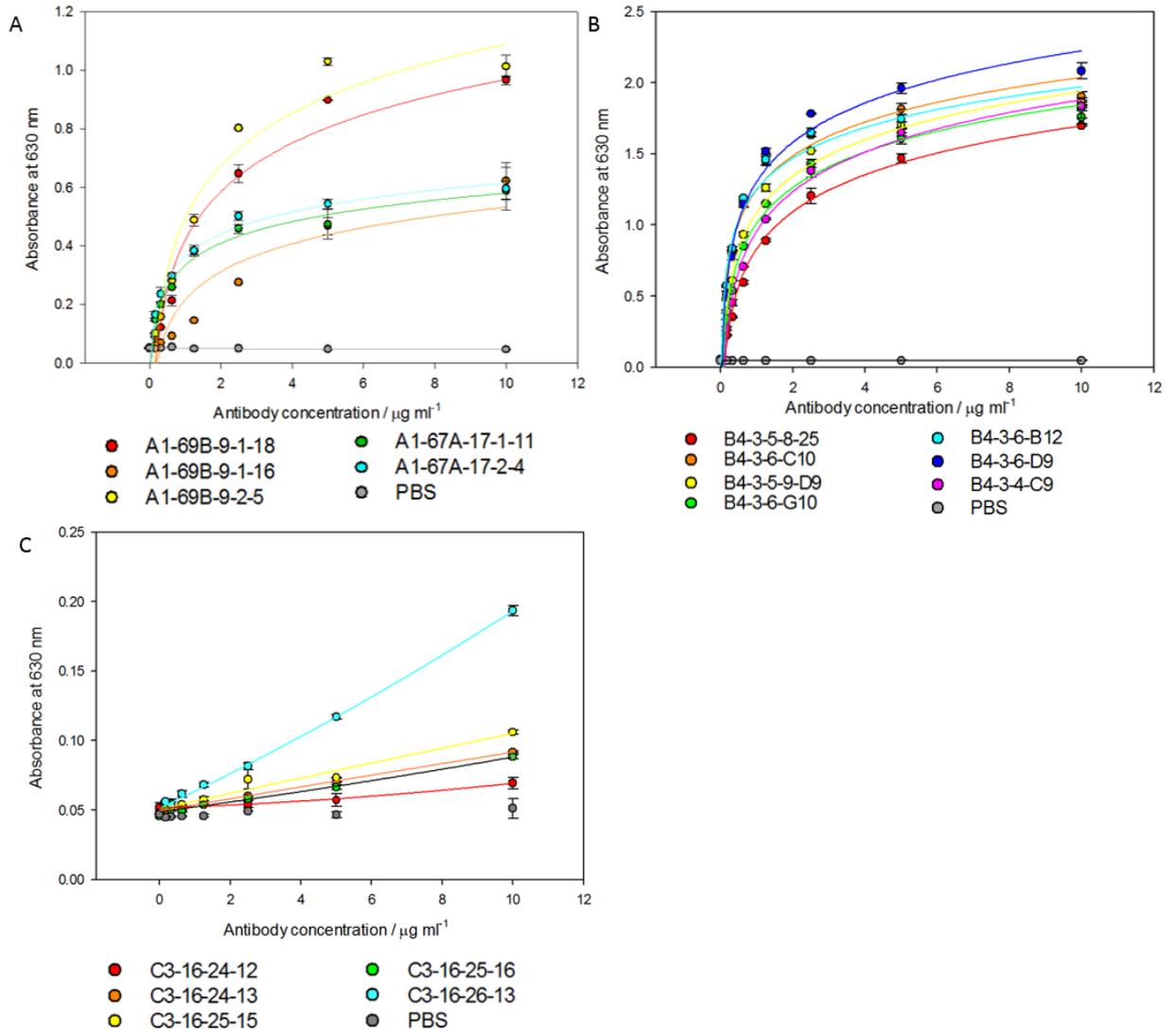
Figure 4.4 Antibody naming convention

The antibody naming convention is shown using the cell line B4-3-6-B12 as an exemplar.

A final catalogue of 17 antibodies was created, consisting of five antibodies binding peptide A, seven antibodies against peptide B and five antibodies raised against peptide C.

#### 4.3 Antibody characterisation against linear peptide

ELISAs were performed to assess the ability of each antibody to bind their respective peptide across a range of concentrations (0-10  $\mu\text{g ml}^{-1}$ ), see figure 4.5. Antibodies raised against peptide B exhibited the strongest binding, producing an absorbance signal at 630 nm around twice as high as that of the two strongest binding antibodies generated against peptide A. The remaining peptide A antibodies and all peptide C antibodies displayed minimal binding. During clonal selection these antibodies did exhibit target-binding properties, it is possible that the cultures were not fully cloned, and during the scale up process cultures became dominated by a clone which did not produce antigen-binding antibodies. Results support the further investigation of two peptide A antibodies (A1-69B-9-1-18 and A1-69B-9-2-5) and seven peptide B antibodies.



**Figure 4.5 Antibody binding to peptide**

Antibody binding to peptide was analysed via ELISA. Concentrations of antibody in the range  $0\text{-}10 \mu\text{g ml}^{-1}$  were used for primary incubation onto peptide-coated ELISA plates. Plates were incubated with TMB for 30 minutes. Peptide A and B antibodies were fitted to logarithmic lines of best fit with  $R^2$  values ranging from 0.87-0.97 for peptide A antibodies and 0.97-0.99 for peptide B antibodies. Antibodies raised against peptide C fitted to a quadratic equation with  $R^2$  values ranging from 0.87 – 1.00. Lines of best fit were determined by SigmaPlot. Error bars show standard deviation,  $n=2$ .

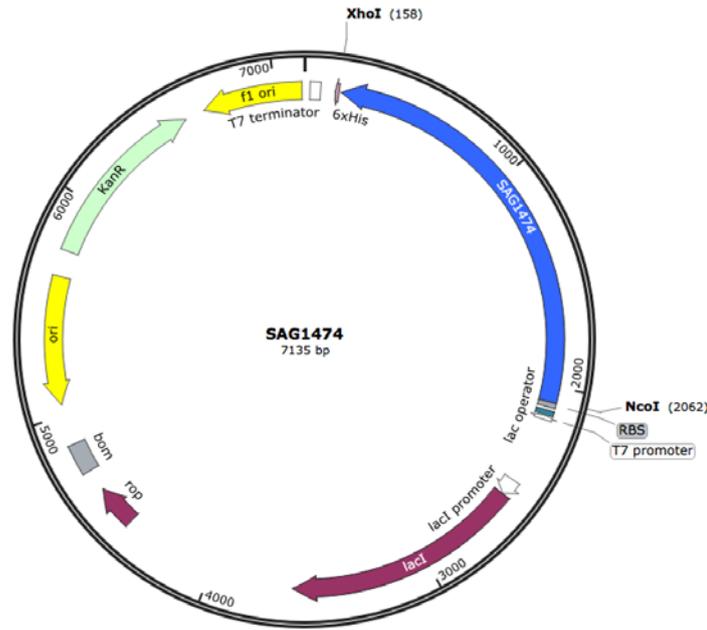
## 4.4 Production of recombinant SAG1474

Antibodies were raised, and initially screened against peptides. For use in a diagnostic test, the antibodies will have to bind the folded-protein form, in which the peptide sequence may hold an alternative conformation, or may be buried within the protein. A recombinant version of the protein SAG1474 was made to assess binding in the native form.

### 4.4.1 Design of SAG1474 construct

For recombinant protein production, the amino acid sequence of SAG1474 was edited to remove the signal sequence, identified by SignalP, to enable cytoplasmic expression. As cell wall associated proteins are cleaved at the LPXTG motif, (Navarre and Schneewind, 1994) the C-terminal sequence was removed. A polyhistidine tag was added to the recombinant protein to enable nickel affinity purification. The C-terminus was selected for tag insertion, as in the native protein it is this site which tethers to the cell wall, therefore modification is less likely to cause perturbations to the folded structure. The DNA sequence was optimised for high efficiency expression in *E. coli* by GeneOptimizer® software (GeneArt, Life Technologies). The DNA sequence was edited to insert NcoI and XhoI restriction enzyme sites to enable insertion into a *pET28a* plasmid, causing the recombinant protein to be under control of IPTG induction. Further editing removed two internal NcoI sites, modifying the sequence to contain alternative codons that would not alter the translated amino acid sequence. The plasmid map and construct DNA sequence are shown in figure 4.6, the protein sequence is found in figure 4.7. Details on protein parameters can be found in table 2.4, section 2.4.1.

A – Plasmid map



B – Construct DNA sequence

NcoI restriction site

CCATGGGCAATAGCACCgAAACCAGCGCAAGCGTTGTTCCGACCACCAATACCATTGTTcAGACCAATGATAGCAATCCGACCG  
 CAAAATTTGTTAGCGAAAGCGGTCAGAGCGTTATTGGTCAGGTTAAACCGGATAATAGCGCAGCACTGACCACCGTTGATACAC  
 CGCATCATATTAGCGCACCGGATGCACTGAAAACCACCCAGAGCAGTCCGGTTGTTGAAAGCACCAGCACAAAACCTGACCGAAG  
 AAACCTATAAACAGAAAGATGGTCAGGATCTGGCAAATATGGTTCGTAGCGGTCAGGTGACCAGCGAAGAAGTGGTGAATATG  
 GCATATGATATTATCGCCAAAAGAAAACCCGAGCCTGAATGCAGTTATTACCACCCGTCGCAAGAAGCAATTGAAGAAGCACGT  
 AAACCTGAAAGATACCAATCAGCCGTTTCTGGGTGTTCCGCTGCTGGTTAAAGGTCTGGGTcATAGCATTAAAGGTGGCGAAACC  
 AATAATGGTCTGATTTATGCCGATGGCAAATTTCCACCTTTGATAGCAGCTATGTGAAAAAATACAAGATCTGGGCTTTATCA  
 TTCTGGGCCAGACCAATTTTCCGGAATATGTTGGCGCAATATTACCGATAGCAAACCTGATGGTCTGACCCATAATCCGTGGGA  
 TCTGGCGCATAATGCCGGTGGTAGCAGCGGTGGTAGTGCAGCAGCAATTGCAAGCGGTATGACCCCGATTGCCAGCGGTAGTG  
 ATGCAGGCGGTAGCATTcGTATTCCGAGCAGCTGGACCGGTCTGGTGGGCTGAAACCgACCCGTGGTCTGGTTAGCAATGAAA  
 AACCGGATTCATATAGCACCGCAGTTCATTTCCGCTGACCAAAGCAGCCGTGATGCAGAAACCCTGCTGACCTATCTGAAAAA  
 AAGCGATCAGACCTGGTTAGCGTTAATGATCTGAAAAGCCTGCCGATTGCATATACCCTGAAAAGTCCGATGGGCACCGAAGT  
 TAGCCAGGATGCAAAAATGCAATTATGGATAACGTGACCTTTCTCGCAAACAGGGTTTTAAAGTTACCGAAATTGATCTGCC  
 GATCGATGGTcGTGCACTGATGCGTGATTATAGCACCTGGCAATTGGTATGGGTGGTGCATTTAGCACCATTGAGAAAGACCT  
 GAAAAACACGGCTTTACCAAAGAAGATGTTGATCCGATTACCTGGCAGTGCATGTTATTATCAGAATAGCGATAAAGCCGA  
 ACTGAAAAAATCTATCATGGAAGCCAGAAACACATGGATGATTATCGTAAAGCGATGGAAAAACTGCACAAAACAGTTTCCGAT  
 TTTTCTGAGCCCGACCACCGCAAGCCTGGCACCCTGAATACCGATCCGTATGTTACCGAAGAGGATAAACGTGCCATTTATAAC  
 ATGGAAAACCTGAGCCAAGAAGAAGCATTGCACTGTTAATCGTCAGTGGGAACCGATGCTGCGTCTGACCCGTTTACCCAG  
 ATTGCAAAATGACAGGTCTGCCTGCAATTAGCATTCCGACATATCTGAGCGAATCTGGTCTGCCTATTGGCACCATTGCTGATGG  
 CAGGGCCAAATTATGATATGGTGTGATCAAATTTGCGACCTTTTGA AAAACACCAGCGCTCAATGTTAAATGGCAGCGCAT  
 CATTGATAAAGAAGTTAAACCGAGCACCGGTCTGATTcAGCCGACCAATAGCCTGTTTAAAGCACATAGCAGCCTGGTTAATCT  
 GGAAGAAAATTCACAGGTTACCCAGGTGAGCATCAGTAAAAAATGGATGAAATCCAGCGTGAAAAACAAACCGAGCGTTATG  
 GCCTATCAGAAAGCACTGCCGAAAACCAGCAGTCATCATCACCATCATTAACTCGAG

XhoI restriction site

Figure 4.6 Recombinant SAG1474 construct DNA sequence

A: Plasmid map of the gene construct, produced in SnapGene Viewer. (GLS Biotech)

B: DNA sequence encoding for SAG1474. NcoI and XhoI restriction enzymes were used to insert the construct DNA into *pET28a*, restriction sites shown in green. Codons were optimised for production in *E. coli* and two bases were altered to remove additional NcoI restriction sites (red).

A – Construct protein sequence

MGNSTETSASVVPTTNTIVQTNDSNPTAKFVSESGQSVIGQVKPDNSAALTTVDTPHHISAPDALKTTQSSPVVESTSTKLTEETYKQK  
DGQDLANMVRSGQVTSEELVNMAYDIIAKENPSLNAVITRRQEAIIEEARLKDNTNQPFLGVPLLKGLGHSIKGETNNGLIYADGKI  
STFDSSYVKKYKDLGFIILGQTNFPEYGWRNITDSKLYGLTHNPWDLAHNAGGSSGSAASGMTPIASGSDAGGSIRIPSSWTGL  
VGLKPTRGLVSNKPDYSYTAHFPLTKSSRDAETLLTYLKKSDQTLVSVNDLKSPLIAYTLKSPMGTEVSQDAKNAIMDNVTFLRKQG  
FKVTEIDLPIGRALMRDYSTLAIGMGGAFSTIEKDLKKHGFTKEDVDPITWAVHVIYQNSDKAELKKSIMEAQKHMDDYRKAMEKL  
HKQFIFLSPTTASLAPLNTDPYVTEEDKRAIYNMENLSQEERIALFNRQWEPMLRRTPTQIANMTGLPAISIPTYLSEGLPIGTMLM  
AGANYDMVLIKFAFFEKHHGFNVKWQRIIDKEVKPSTGLIQPTNSLFAHSSLVNLEENSQVTQVSISKKWWMKSSVKNKPSVMAYQ  
KALPKTSSHHHHH Stop

B – Recombinant protein schematic

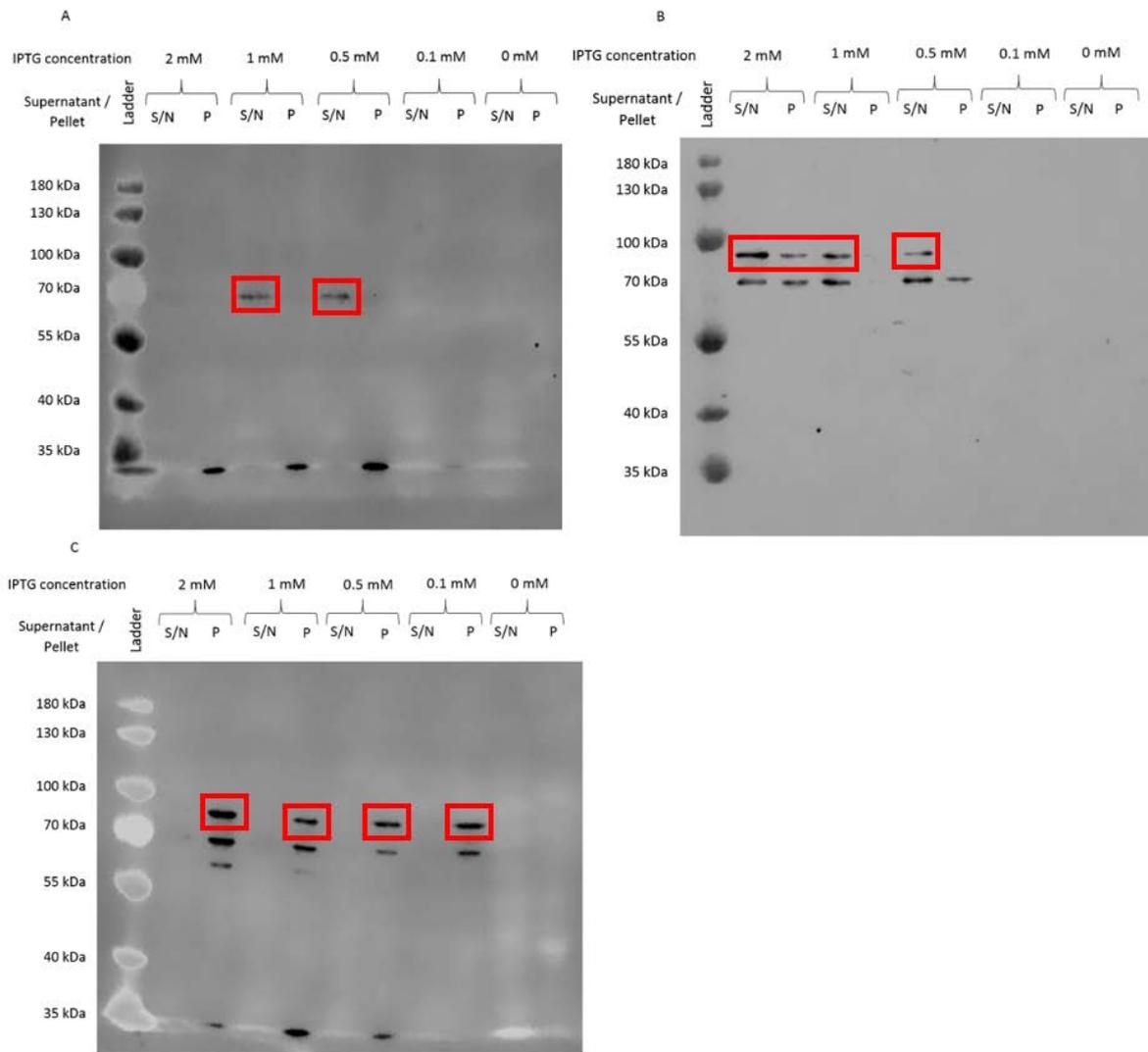


Figure 4.7 Recombinant SAG1474 protein

A: Protein sequence encoding for SAG1474. The N terminal signal sequence and the region to the C terminus of the LPXTG motif have been removed and a histidine tag (orange) has been designed into the sequence. B: Representation of the SAG1474 protein construct.

#### 4.4.2 Induction of SAG1474

In order to investigate expression conditions for optimal SAG1474 production, small induction tests were performed. Each culture was grown at a constant temperature to an OD<sub>600nm</sub> of 0.5, and induced with IPTG at various concentrations for various time points. Cell extracts were harvested, lysed by Bugbuster, and centrifuged to separate insoluble protein found in the pellet from soluble protein in the supernatant. The soluble and insoluble fractions were analysed by western blotting. An anti-polyhistidine tag antibody was used for SAG1474 detection, see figure 4.8. SAG1474 is predicted to be 69.4 kDa. (Gasteiger *et al.*, 2005)



**Figure 4.8 Induction of SAG1474: western blots**

*E. coli* BL21 cells were transformed with *pET28 $\alpha$ -SAG1474*. Cells were grown to an OD<sub>600 nm</sub> of 0.5, before induction with IPTG. A: 37 °C, induced for one hour. B: 37 °C, induced for three hours. C: 20 °C, induced for 16 hours. Cells were lysed by Bugbuster and centrifuged to separate insoluble pellet from the soluble supernatant. Western blots were performed using an anti-his tag antibody for SAG1474 detection. Insufficient protein expression is achieved whilst growing at 37 °C. Growth at 20 °C, caused protein to be found solely in the insoluble fraction. Protein bands representing SAG1474 are indicated by a red box.

Histidine-tagged protein was detected from cells induced with IPTG. There was no histidine-tagged product detected in the absence of IPTG, indicating the plasmid does not exhibit leaky expression in this system. A band corresponding to 85 kDa was detected in experiments B and C; the predicted molecular weight of SAG1474 is 69.4 kDa. Aberrant protein mobility can arise from negatively charged amino acid

content, (Graceffa *et al.*, 1992; Alves *et al.*, 2004) proline residue content (Condemine and Shevchik, 2000) or result from the protein's hydrophobicity. (Shirai *et al.*, 2008)

Western blots showed little protein expression when cells were cultured at 37 °C. There was an increase in SAG1474 detection at three hours in comparison to one hour, suggesting that induction over a longer time period may lead to further enhancement of protein production. However growth curves show cell growth is restricted by IPTG addition, indicating that extension of the induction time may not be viable, see figure 4.9. Protein induction performed at 20 °C for 16 hours, caused SAG1474 to be found in the insoluble fraction. The insoluble fraction contains unfolded proteins, which have aggregated. It may not be possible to re-fold protein into the native conformation. As there is no solved structure for SAG1474 for comparison, it would be difficult to assess whether re-folding into the native formation was successful. Assuming the highest molecular weight band represents SAG1474, western blots show evidence of protein degradation with multiple protein bands of lower molecular weight observed.

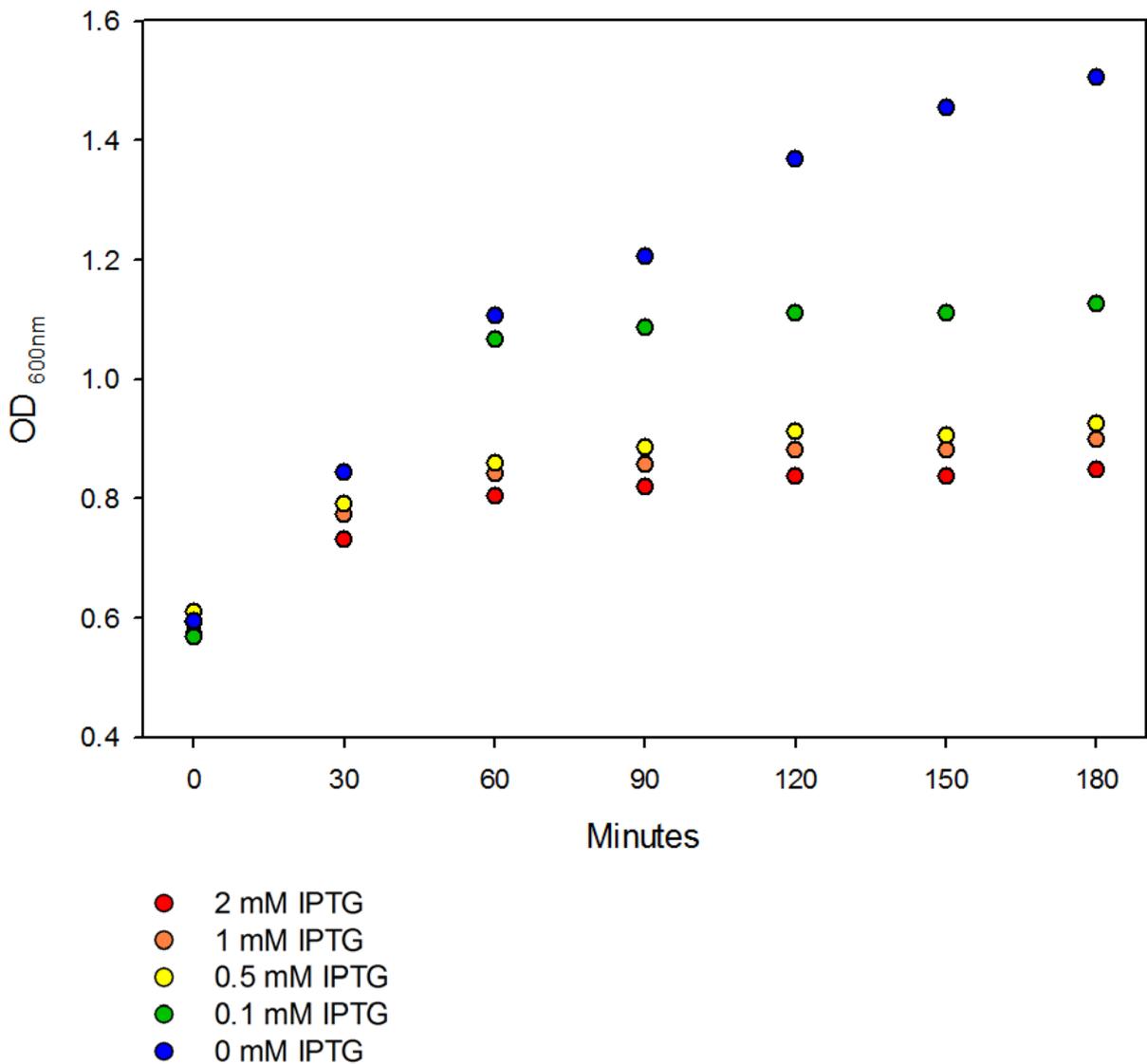


Figure 4.9 Growth curve following induction of SAG1474 with IPTG

The optical density of cell cultures at 600 nm was followed at intervals of 30 minutes.

There is a significant reduction in cell growth in cells induced with IPTG in comparison to non-induced cells.

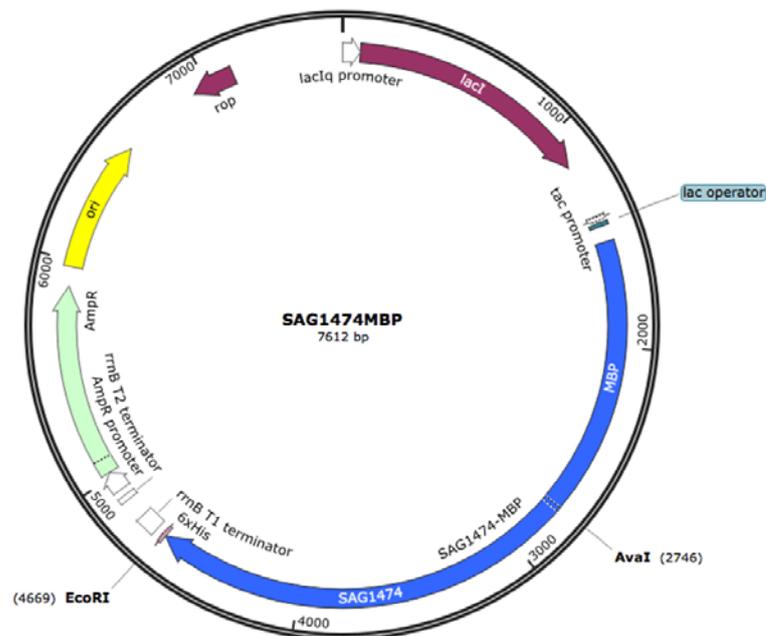
Growth limitation upon SAG1474 expression suggests that SAG1474 is toxic. This is promising information as it may be functional in its putative role as an amidase, and therefore in its folded state. Amidases catalyse amide bond hydrolysis. In its native state as a cell wall associated protein, SAG1474 may function in breaking down peptidoglycan. This plays a role in cell wall growth and restructuring, cell division, and recycling cell wall components. (Scheffers and Pinho, 2005) As SAG1474 is largely unstudied, its precise target is unknown. It may lyse amide bonds in *E. coli* when produced recombinantly, thus causing toxicity.

#### 4.4.3 Design of a SAG1474 fusion protein

Induction of SAG1474 at high temperatures over a short period of time was unable to induce sufficient protein expression, and induction at low temperatures resulted in insoluble unfolded protein; therefore expression of this protein construct was deemed to be an unfeasible route to production. A protein tag was added to the SAG1474 sequence to aid in protein folding, increase expression and reduce protein degradation.

Maltose binding protein (MBP) is a 43 kDa solubility tag (Costa *et al.*, 2014) which has been added to many proteins successfully. (Pryor and Leiting, 1997; Hang *et al.*, 2016; Nguyen *et al.*, 2016) MBP acts as a chaperone protein, preventing self-aggregation of the fusion partner by interaction with the hydrophobic cleft within MBP. (Kapust and Waugh, 1999; Costa *et al.*, 2014) A commercially available plasmid, *pMAL*, includes the gene encoding for maltose binding protein, *malE*, adjacent to a multiple cloning site, allowing fusion partner insertion via the use of a selection of restriction sites. The construct is under the control of a strong promoter,  $P_{Tac}$ , therefore a fusion product will be strongly expressed. There are two variants of the *pMAL* plasmid, one of which directs proteins to the periplasmic space via a signal sequence. (New England BioLabs, 2017b) The periplasm contains fewer proteases, (Baneyx and Mujacic, 2004) and therefore proteins within the periplasm are less susceptible to proteolytic degradation. Periplasmic expression may also aid purification. The periplasmic space contains fewer proteins, (Baneyx and Mujacic, 2004) therefore if the recombinant protein can be isolated from the periplasm without lysing the cell membrane, the protein preparation will contain fewer contaminants. The plasmid variant which directs fusion proteins to the periplasm was selected. A linker region between the gene encoding for MBP and SAG1474 was designed to include a TEV protease site. The protease site allowed for removal of MBP following expression. The plasmid map and construct DNA sequence can be found in figure 4.10. The protein sequence alongside a schematic of the protein construct can be seen in figure 4.11. Details on protein parameters can be found in table 2.4, section 2.4.1.

A – Plasmid map



B – Construct DNA sequence

AvaI restriction site

CTCGGGGAAAACCTGATTTTCAGGGCAATAGCACCGAAACCAGCGCAAGCGTTGTTCCGACCACCAATACCATTGTTTCAGACCAATGATAGCAATCCGACC GAAAATTTGTTAGCGAAAGCGGTGAGAGCGTTATTGGTCAGGTTAAACCGGATAATAGCGCAGCACTGACCACCGTTGATACACCGCATCATATTAGCGCACCGGATGCACTGAAAACACACAGAGCAGTCCGGTTGTTGAAAGCACCAGCACCAAAACCTGACCGAAGAACTATAAACAGAAAGATGGTCAGGATCTGGCAAATATGGTTCGTAGCGGTGAGGTGACCAGCGAAGAAGTGGTGAATATGGCATATGATATTATCGCCAAAGAAAATCCGAGCCTGAATGCAGTTATTACACACGTCGTCAAGAAGCAATTGAAGAAGCACGTAACCTGAAAAGATACCAATCAGCCGTTTCTGGGTGTTCCGCTGCTGGTTAAAGGTCTGGTTCATAGCATTAAAGGTGGCGAAAACCAATAATGGTCTGATTTATGCCGATGGCAAATTTCCACCTTTGATAGCAGCTACGTGAAGAAATACAAAGATCTGGGCTTTATTATCTGGGCCAGACCAATTTCCGGAATATGGTTGGCGCAATATTACCGATAGCAAATGATGGTCTGACCATAATCCGTGGGATTTAGCACATAATGCCGGTGGTAGCAGCGGTGATGTCAGCAGCAATTGCAAGCGGTATGACCCCGATTGCCAGCGGTAGTGATGCAGGCGGTAGCATTTCGTATCCGAGCAGCTGGACCGGTCTGGTGGTCTGAAACCGACACGTGGTCTGGTTAGCAATGAAAAACCGGATTCATATAGCACCGCAGTTCATTTCCGCTGACCAAAAGCAGCCGTGATGCAGAAACCTGCTGACCTATCTGAAAAAAGCGATCAGACCCTGGTTAGCGTTAATGATCTGAAAAGCCTGCCGATTGCATATACCCTGAAAAGTCCGATGGGCACCGAAGTTAGCCAGGATGCAAAAAATGCAATTATGGATAACGTGACCTTTCTGCGCAAACAGGGTTTTAAAGTTACCGAAATGATCTGCCGATCGATGGTCTGCACTGATGCGTGATTATAGCACCCCTGGCAATGGTATGGGTGGTGCATTTAGCACCAATTGAGAAAGACCTGAAAAAACACGGCTTTACCAAAGAAGATGTTGATCCGATTACCTGGCAGTGCATGTTATTTATCAGAAATAGCATAAAGCCGAAGTGAAGAAAAGCATTATGGAAGCCAGAAAACATGGATGATTATCGTAAAGCCATGAAAAACTGCACAAACAGTTTCCGATTTTCTGAGCCGACCACC GCAAGCTGGCACCGCTGAATACCGATCCGTATGTTACCGAAGAGGATAAACGTGCCATTTATAACATGGAAAACTGAGCCAAGAAGAACGCATTGCACTGTTAATCGTCAGTGGGAACCGATGCTGCGTCTGACCCGTTTACACAGATTGCAAATATGACAGGTCTGCCTGCAATTAGCATTCCGACATATCTGAGCGAATCTGGTCTGCCATTGGCACCATGCTGATGGCAGGCGCAAATTATGATATGGTCTGATCAAATTTGCGACCTTCTTTGAAAAACACCATGGCTTCAATGTAAATGGCAGCGCATCTTGATAAAGAAAGTTAAACCAGCAGCCGGTCTGATTCAGCCGACCAATAGCCTGTTAAAGCACATAGCAGCCGGTTAATCTGGAAGAAAATTCACAGGTTACCCAGGTGAGCATCAGTAAAAAATGGATGAAATCCAGCGTGAAAAACAAACCGAGCGTTATGGCCTATCAGAAAGCACTGCCGAAAACAGCAGTCATCATCACCATCATCATTAAGAATTC

EcoRII restriction site

Figure 4.10 Recombinant SAG1474-MBP construct DNA sequence

A: Plasmid map of the gene construct, produced in SnapGene Viewer. (GLS Biotech)

B: DNA sequence encoding for SAG1474. AvaI and EcoRII restriction enzymes were used to insert the construct DNA into *pMAL p5x*, restriction sites shown in green. Codons were optimised for production in *E. coli*.

A – Construct protein sequence

MKIKTGARILALSALTTMMFSASALAKIEEGKLVWINGDKGYNGLAIEVGGKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRF  
GGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLI  
AADGGYAFKYENGYDIKDVGVNAGAKAGLTFVLDLIKXHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTF  
KGQPSKPFVGVLSAGINAASPNKELAKEFLENLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIQMSAFWYA  
VRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGENLYFQGNSTETSASVVPNTNTIVQTNDNSNPTAKFVSESGQSVIGQVKPD  
NSAALTTVDTPHHISAPDALKTTQSSPVVESTSTKLTETTYKQKDGQDLANMVRSGQVTSEELVNMAYDIIAKENPSLNAVITRRQEAIEEA  
RKLKDTNQPFLLVPLLVKGLGHSIKGGETNNGLIYADGKISTFDSSYVKKYKDLGFILGQTNFPEYGWRNITDSKLYGLTHNPWDLAHNAGG  
SSGGSAAAIASGMTPIASGSDAGGSIRIPSSWTGLVGLKPTRGLVSNKPDYSYSTAVHFPLTKSSRDAETLLTYLKKSDQTLVSVNDLKSPIAY  
TLKSPMGTEVSDAKNAIMDNVTLRKGQFKVTEIDLPIDGRALMRDYSTLAIGMGGAFSTIEKDLKKHGFTKEDVDPITWAVHVIYQNSDK  
AELKKSIMEAQKHMDDYRKAMEKLHKQFPIFLSPTTASLAPLNTDPYVTEEDKRAIYNMENLSQEERIALFNRQWEPMLRRTPFTQIANMT  
GLPAISIPTYLSSEGLPIGTMLMAGANYDMVLIKFATFFFEKHGFFNVKWQRIIDKEVKPSTGLIQPTNSLFAKHSLSLVNLEENSQVTQVSISIKK  
WMKSSVKNKPSVMAYQKALPKTSSHHHHHH Stop

B – Recombinant protein schematic

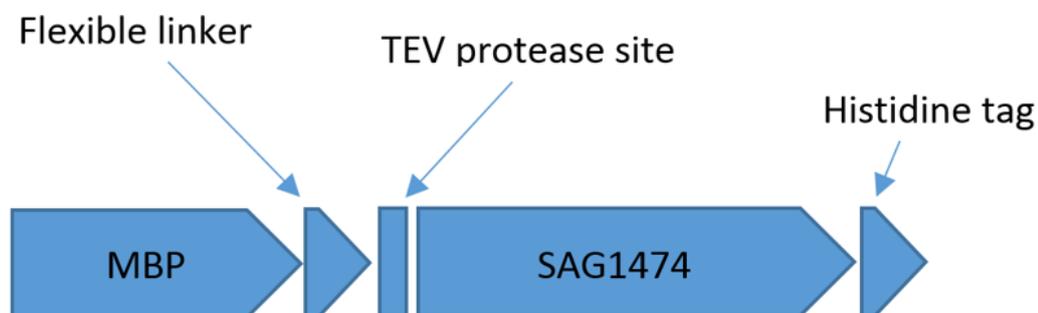
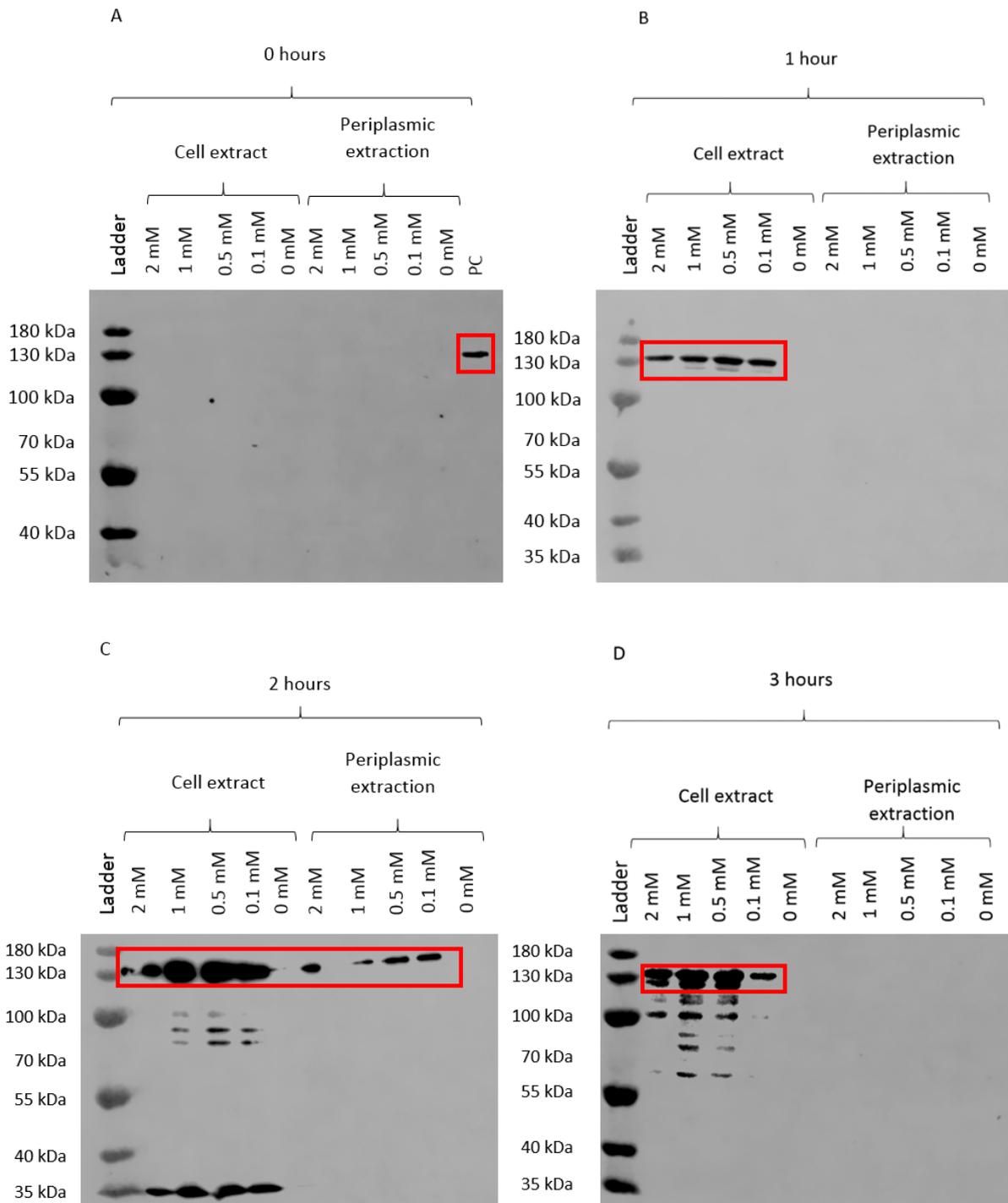


Figure 4.11 Recombinant SAG1474-MBP construct protein

A: Protein sequence inserted into pMAL p5x encoding for production of SAG1474. A TEV protease cleavage site (green) and histidine tag (orange) have been designed into the sequence and the N terminal signal sequence and the region to the C terminus of the LPXTG motif have been removed. B: Representation of the SAG1474-MBP protein construct.

#### 4.4.4 Induction of SAG1474-MBP

A similar process was adopted for the evaluation of optimal conditions for SAG1474-MBP expression as for SAG1474. Induction test cultures were grown at 37 °C and induced for various time points with a range of IPTG concentrations. Periplasmic extraction was performed and these samples, along with complete cell extracts, were analysed by western blotting, see figure 4.12.



**Figure 4.12 Induction of SAG1474-MBP: western blots**

*E. coli* BL21 cells were transformed with *pMAL p5x-SAG1474-MBP*. Cells were grown at 37 °C to an OD<sub>600 nm</sub> of 0.5, before induction with IPTG. A: induced for zero hours. B: induced for one hour. C: induced for two hours. D: induced for three hours. Periplasmic extraction was performed and analysed alongside whole cell extracts. Western blots were performed using an anti-his tag antibody for SAG1474 detection. Positive control (PC) in panel A is a cell extract induced with 2 mM IPTG for 1 hour at 37 °C, added to determine that western blotting procedure was successful. Protein bands representing SAG1474-MBP are indicated by a red box.

Similarly to SAG1474, the plasmid encoding SAG1474-MBP showed no leaky expression in the absence of IPTG. Protein bands were observed at 130 kDa, however prediction software estimates the molecular weight of SAG1474-MBP to be 115 kDa. (Gasteiger *et al.*, 2005) Mass spectrometry confirmed protein identity as SAG1474-MBP, data not shown. Possible reasons for the deviation between observed and expected molecular weights are discussed in reference to SAG1474, see section 4.4.2.

Through analysis of whole cell extracts, maximal expression of SAG1474-MBP was observed following a two hour induction with 0.5 mM IPTG. At induction for two hours bands of lower molecular weight were detected at around 80-100 kDa, suggesting that a fraction of the total protein degrades at various sites within the protein. As the expected molecular weight of SAG1474 is 69.4 kDa, degradation may not fragment the SAG1474 protein. The low molecular weight band at 35 kDa can be removed via protein filtration with an appropriate molecular weight cut off, or via size exclusion chromatography. Degradation was increased when cells were induced for three hours.

Periplasmic extraction was performed on all samples, however detection of histidine-tagged protein was only achieved on one occasion. When repeated under identical conditions it was not possible to successfully extract protein from the periplasm. The SecYEG translocon is responsible for transportation of unfolded maltose binding protein to the periplasm. (Weiss *et al.*, 1988) If the recombinant protein construct folds in the cytoplasm, transportation may not be achievable. Alternatively it is possible that the high expression rate exceeds the rate at which protein can be exported to the periplasm, thus creating a backlog of protein trapped in the cytoplasm. Although periplasmic extraction was unsuccessful, translocation to the periplasm may still have been beneficial due to the reduced protease concentration in comparison to the cytoplasm. Whole cell extracts were used for protein purification. SAG1474-MBP extracted from the whole cell is likely to exist as a mixed population of protein, the full length protein from the cytoplasm, and a signal sequence deficient variant extracted from the periplasm. (Beena *et al.*, 2004)

As seen with induction of SAG1474, SAG1474-MBP expression causes a reduction in cellular growth, see figure 4.13. Toxicity is a sign that the putative amidase is active, and folded. A two hour induction with 0.5 mM IPTG was selected for large scale protein production.

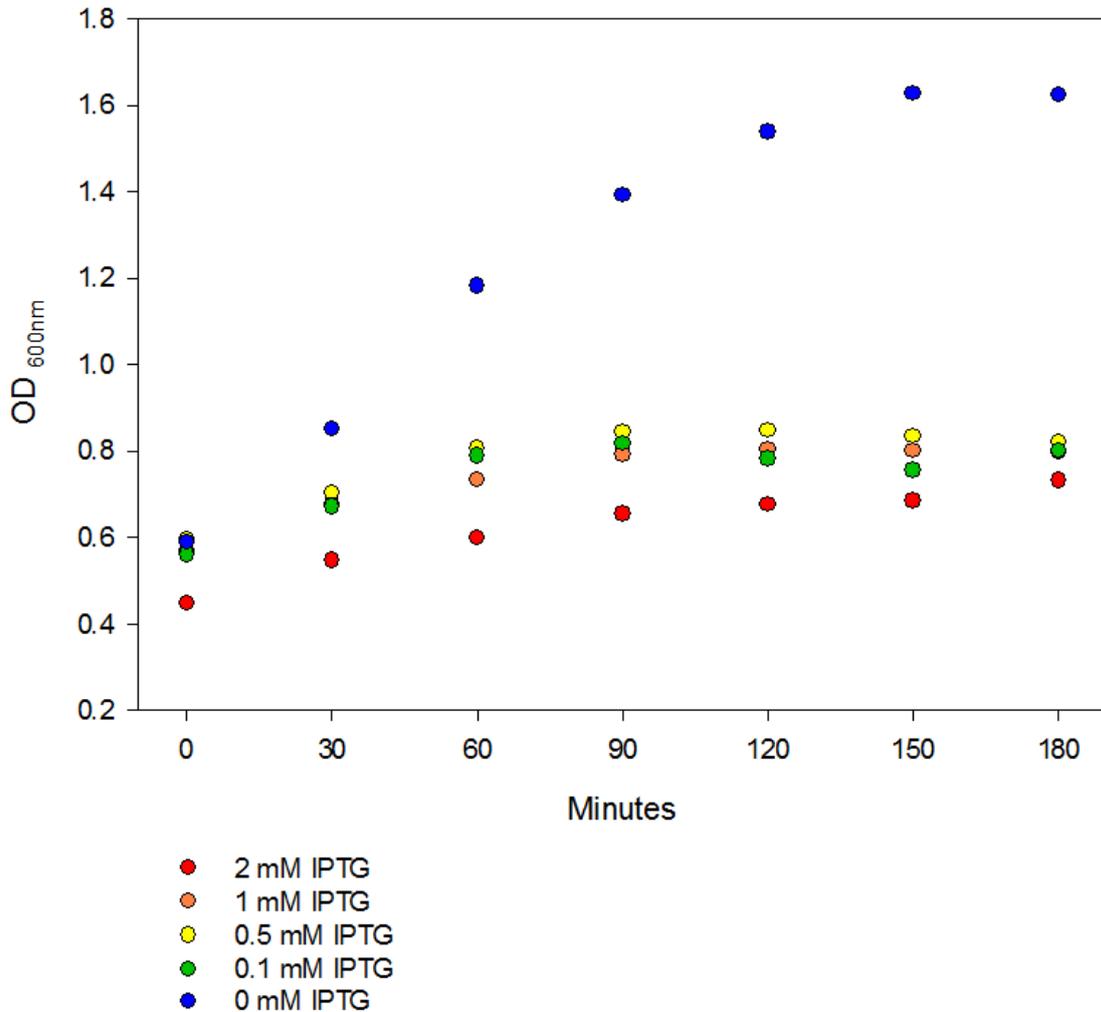
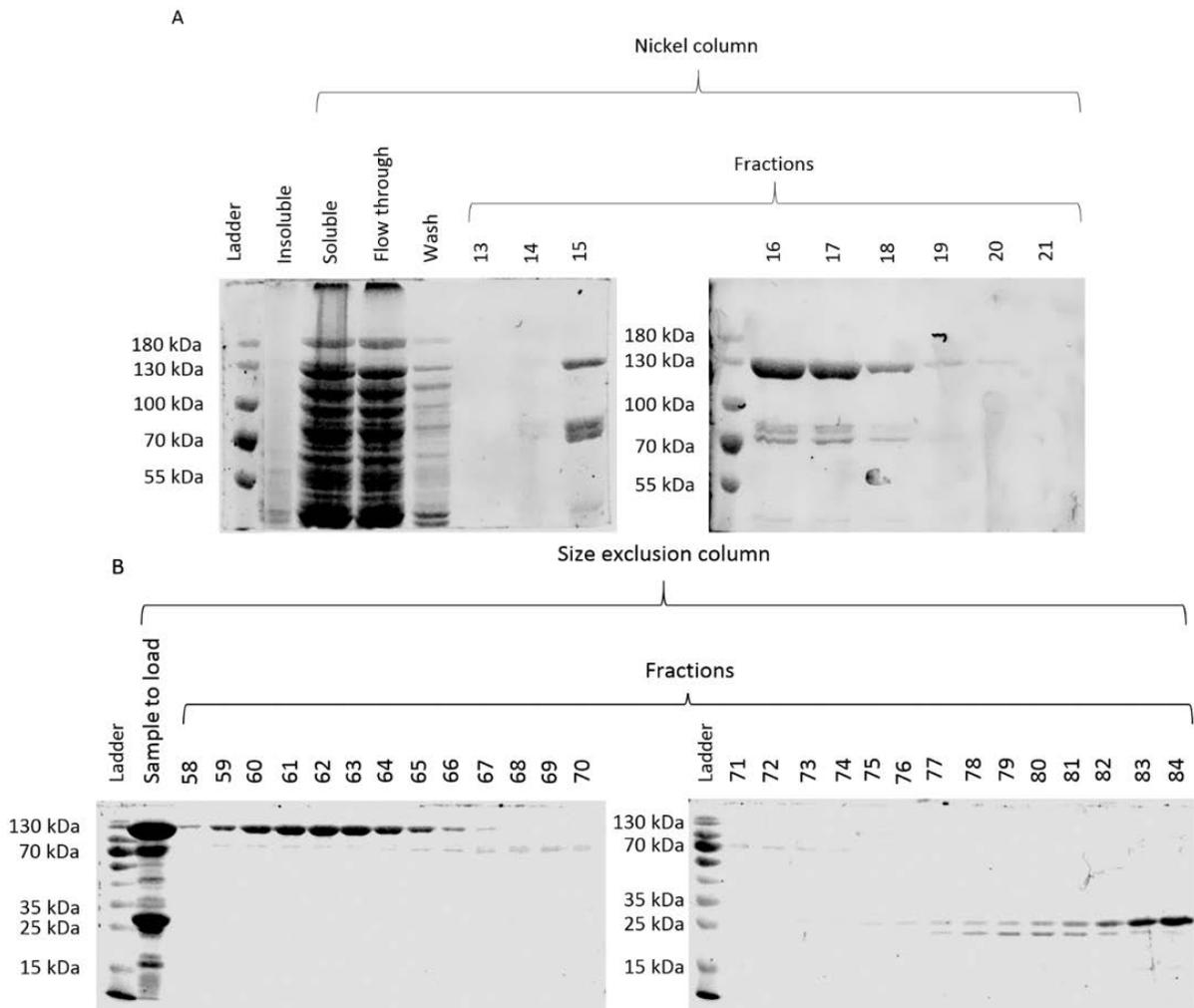


Figure 4.13 Growth curve following induction of SAG1474-MBP with IPTG  
The optical density of cell cultures at 600 nm was followed at intervals of 30 minutes.  
There is a significant reduction in cell growth in cells induced with IPTG in comparison to non-induced cells.

#### 4.4.5 SAG1474-MBP purification

A large scale growth of *E. coli* BL21 *pMALp5xSAG1474-MBP* was conducted by Newcastle University Protein and Proteome Analysis, following the optimised protocol devised by induction tests. Cells were harvested, lysed by Bugbuster and centrifuged

to isolate the soluble protein fraction. Protein samples were purified by nickel affinity chromatography, utilising the nickel-binding properties of the polyhistidine tag. Eluates containing SAG1474-MBP were pooled and further purified by size exclusion chromatography see figure 4.14. A protein migrating at 130 kDa, the same value observed in induction tests was purified. A small sample injection volume and slow running speed ( $1 \text{ ml min}^{-1}$ ) was used to enhance protein separation. This enabled the separation of SAG1474-MBP from the small amount of SAG1474, which was essential for protein characterisation via circular dichroism. A final yield ranging from 0.55-1.1 mg SAG1474-MBP was purified per 1 g cell pellet.



**Figure 4.14 Purification of recombinant SAG1474-MBP**

SDS-PAGE gels show the purification process of recombinant SAG1474-MBP. *E. coli* BL21 cells transformed with *pMALp5xSAG1474* were grown at 37 °C to an OD<sub>600 nm</sub> of 0.5 and induced with 0.5 mM IPTG for two hours. A: Cells were lysed by Bugbuster and centrifuged to isolate soluble protein which was purified by nickel chromatography. B: The sample was further purified using size exclusion chromatography.

Alternative methods of purification were considered and trialed. A TEV protease site was designed into the protein sequence to allow removal of MBP from SAG1474. Following nickel purification, samples were cleaved using TEV protease. The percentage of cleaved protein from total SAG1474 protein increased from around 30% to 80% following incubation with TEV protease, as determined by band densitometry, see figure 4.15. A GST affinity column was used to remove the GST-tagged TEV protease after proteolysis. This step caused an unexpected large loss of SAG1474 protein, with little protein recovered during flow through or column cleaning (data not

shown). The cleaved sample, following GST-purification of TEV protease, exhibited degradation over a two week period, however the un-cleaved protein sample did not (data not shown). This may be due to low levels of protease remaining in the solution and acting non-specifically on the protein sample. Due to the instability, incomplete proteolysis and protein loss during protease removal, TEV protease cleavage was not performed in further protein purifications. Subsequent purification of the mixed protein population was attempted by ion exchange chromatography. A large amount of protein was lost whilst utilising a cationic exchange column (data not shown). For this reason size exclusion chromatography was selected in preference to ion exchange chromatography.

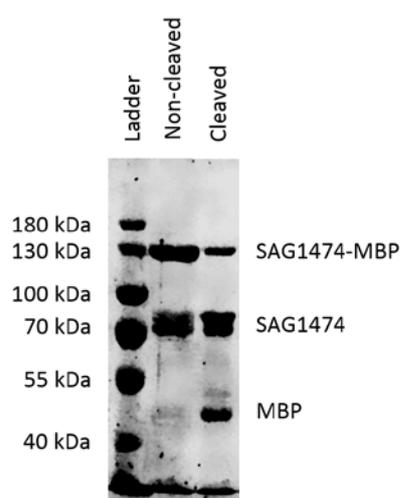


Figure 4.15 TEV protease cleavage of SAG1474-MBP

SDS-PAGE gel showing the SAG1474-MBP protein sample before and after treatment with TEV protease.

The linker region of SAG1474-MBP was designed to be sufficiently long enough for MBP and SAG1474 to fold independently, in their native forms. A previous study has compared the X-ray crystallography structures obtained from free proteins and those bound to MBP. The structures were found to be very similar, with the average difference between atomic positions in the majority of proteins being around 1 Å. (Waugh, 2016) Previous studies have found proteins to be in their active state whilst existing as a MBP fusion product. (Smith and Kotin, 1998; Gholizadeh *et al.*, 2010; Gholizadeh and Kohnehrouz, 2011) Therefore it is valid to structurally characterise SAG1474 in its SAG1474-MBP form.

#### 4.4.6 Structural characterisation of SAG1474-MBP

To assess antibody affinity to protein, the recombinant protein must be folded in its native state. Circular dichroism (CD) was utilised to assess the secondary and tertiary structure of SAG1474-MBP. The CD spectra of MBP was determined and subtracted from that of SAG1474-MBP to discover the structural characteristics of the SAG1474 protein.

Polarised light waves oscillate in single plane. When two polarised light waves, oscillating at an angle difference of  $90^\circ$  with respect to each other, are combined out of phase, the waves are circularly polarised. Circularly polarised light can be right or left handed. Chiral molecules absorb left-handed circularly polarised light to a different extent to right-handed circularly polarised light. All amino acids, except glycine, are chiral, with four distinct groups in an asymmetrical formation around a central carbon atom.

##### 4.4.6.1 Far UV circular dichroism

Amino acids absorb light in the far UV region differently depending on their secondary structure. The CD spectrum of an  $\alpha$ -helical protein shows minima at 222 nm and 208 nm and a maxima at 193 nm. The CD spectrum of  $\beta$ -sheets shows a minima at 218 nm and a maxima at 195 nm. The CD spectrum of an unstructured protein shows a minima at 195 nm, and little activity at over 210 nm. (Greenfield, 2006) The secondary structure of maltose binding protein consists of twice as much  $\alpha$ -helical content as  $\beta$ -sheet; (Spurlino *et al.*, 1991) the spectra observed reflected a species dominated by  $\alpha$ -helices (maxima: 196 nm, minima: 209 nm and 222 nm), see figure 4.16. A spectra indicative of an  $\alpha$ -helical secondary structure was obtained for SAG1474 (maxima: 195 nm, minima: 209 nm and 222 nm). The structure of SAG1474 is unknown, however the circular dichroism spectra can be compared to the structure predicted by I-TASSER and psipred. The predicted structure is dominated by  $\alpha$ -helices and unstructured loops, with a small proportion of  $\beta$ -sheet structure, see table 3.7, section 3.5.5; these results concur. A variety of programmes were utilised to

deconvolute the spectra in order to determine proportions of  $\alpha$ -helix and  $\beta$ -sheet content, however no predictions with high confidence levels were produced. This may be due to a lack of similarity to reference data. (Whitmore and Wallace, 2008)

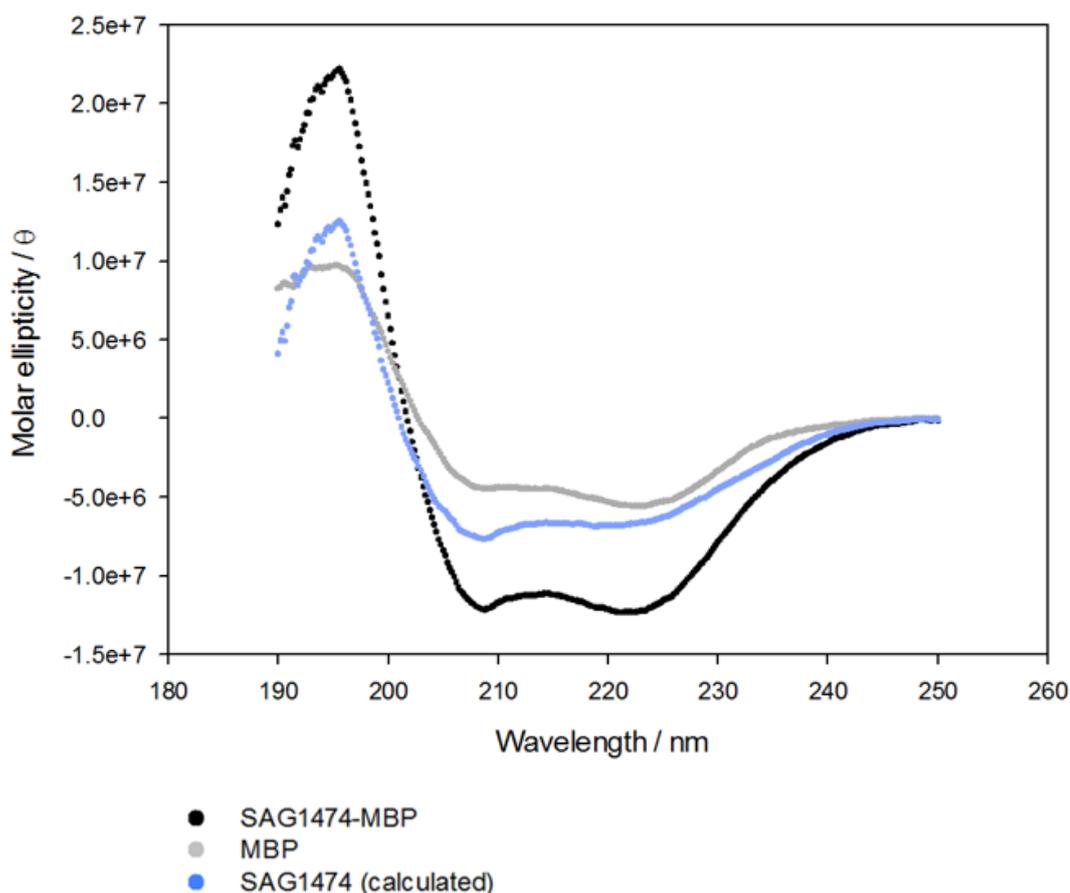


Figure 4.16 Circular dichroism study of SAG1474: Far UV

A: Far UV spectra for SAG1474-MBP, MBP and SAG1474. Each wavelength scan was determined from an average of nine scans. The background CD spectrum from buffer (obtained by averaging the results from three scans) has been taken away. A far UV scan for SAG1474 has been calculated by subtracting the signal obtained for MBP from that of SAG1474-MBP. HT values did not exceed 800 V (data not shown), therefore protein concentration was acceptable. (Fung, n.d.)

#### 4.4.6.2 Near UV circular dichroism

Tertiary structure can be probed with circular dichroism using wavelengths in the near UV region, with signals produced by aromatic residues. Phenylalanine causes a signal at 255-270 nm, tyrosine at 275-282 nm, tryptophan at 290-305 nm and disulphide bonds give rise to indistinct signals. When in a folded state, these residues are held in

an asymmetric (or chiral) formation. Unfolded proteins show a lack of uniformity across the population, therefore circularly polarised light is absorbed to different extents and a signal is not observed. (Kelly and Price, 2000) The near UV spectra for SAG1474-MBP indicates that the protein has three dimensional structure. The spectra for SAG1474, calculated in the same way as for the far UV spectra, shows a peak at 253 nm and 292 nm, and a trough at 280 nm, arising from phenylalanine, tryptophan and tyrosine residues, see figure 4.17.

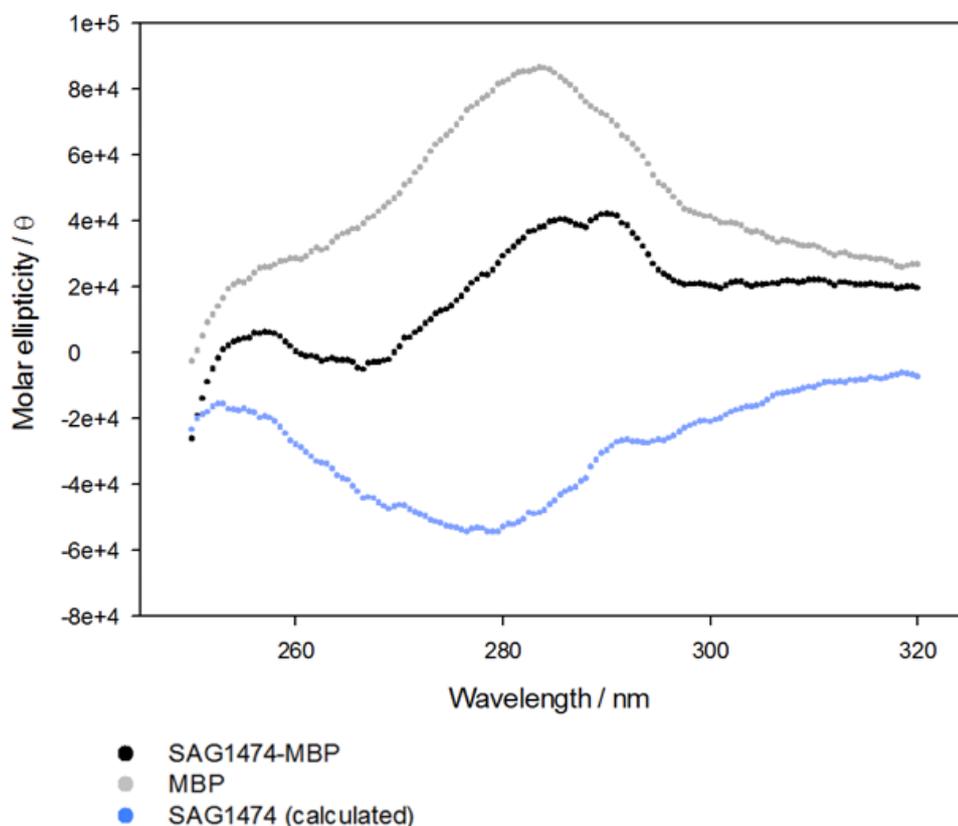


Figure 4.17 Circular dichroism study of SAG1474: Near UV

Near UV spectra for SAG1474-MBP, MBP and SAG1474. Each wavelength scan was determined from an average of nine scans. The background CD spectrum from buffer (obtained by averaging the results from three scans) has been taken away. A near UV scan for SAG1474 has been calculated by subtracting the signal obtained for MBP from that of SAG1474-MBP. HT values did not exceed 800 V (data not shown), therefore protein concentration was acceptable. (Fung, n.d.)

#### 4.4.6.3 Melting temperature circular dichroism

When proteins are heated, their absorption of circularly polarised light is altered due to the loss of folded structure. Unfolded proteins would show no change in CD absorption with increasing temperature. As SAG1474-MBP has a distinctly different CD

absorbance profile at 222 nm over a range of temperatures, the protein must be folded at low temperatures. Melting temperature, is defined as the temperature at which half of the structure exists in its unfolded form. The melting temperatures for SAG1474-MBP and MBP were calculated to be 52 °C and 63 °C respectively. Circular dichroism graphs are found in figure 4.18.

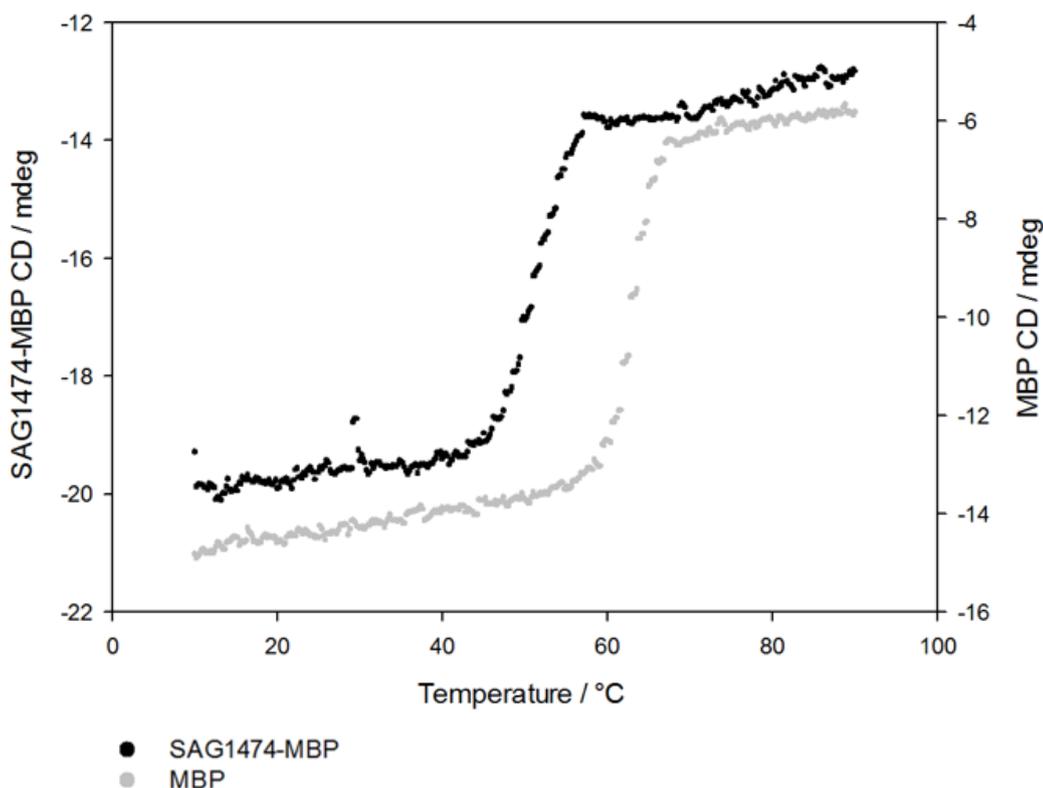


Figure 4.18 Circular dichroism study of SAG1474: Melting temperature

A: Thermal melt spectra for SAG1474-MBP and MBP, measured at 222 nm. Temperature was increased at a rate of 2 °C min<sup>-1</sup>. HT values did not exceed 800 V (data not shown), therefore protein concentration was acceptable. (Fung, n.d.)

#### 4.5 Antibody characterisation against folded protein

The nine most promising antibodies were tested for binding to folded SAG1474-MBP. To confirm antibody affinity to SAG1474-MBP was dependent on binding in the SAG1474 region, MBP binding was used as a negative control. The limit of detection for each antibody was calculated by adding the mean of the MBP negative control to three times its standard deviation. The results can be seen in figure 4.19. ELISA results

show peptide B antibodies exhibited the strongest binding to SAG1474-MBP, these antibodies showed reduced binding to MBP. Peptide A antibodies showed moderate peptide affinities, but limited protein binding, with antibody A1-69B-9-1-18 exhibiting no binding above the limit of detection. This could be due to peptides being not being surface exposed in the protein form or holding a different conformation.

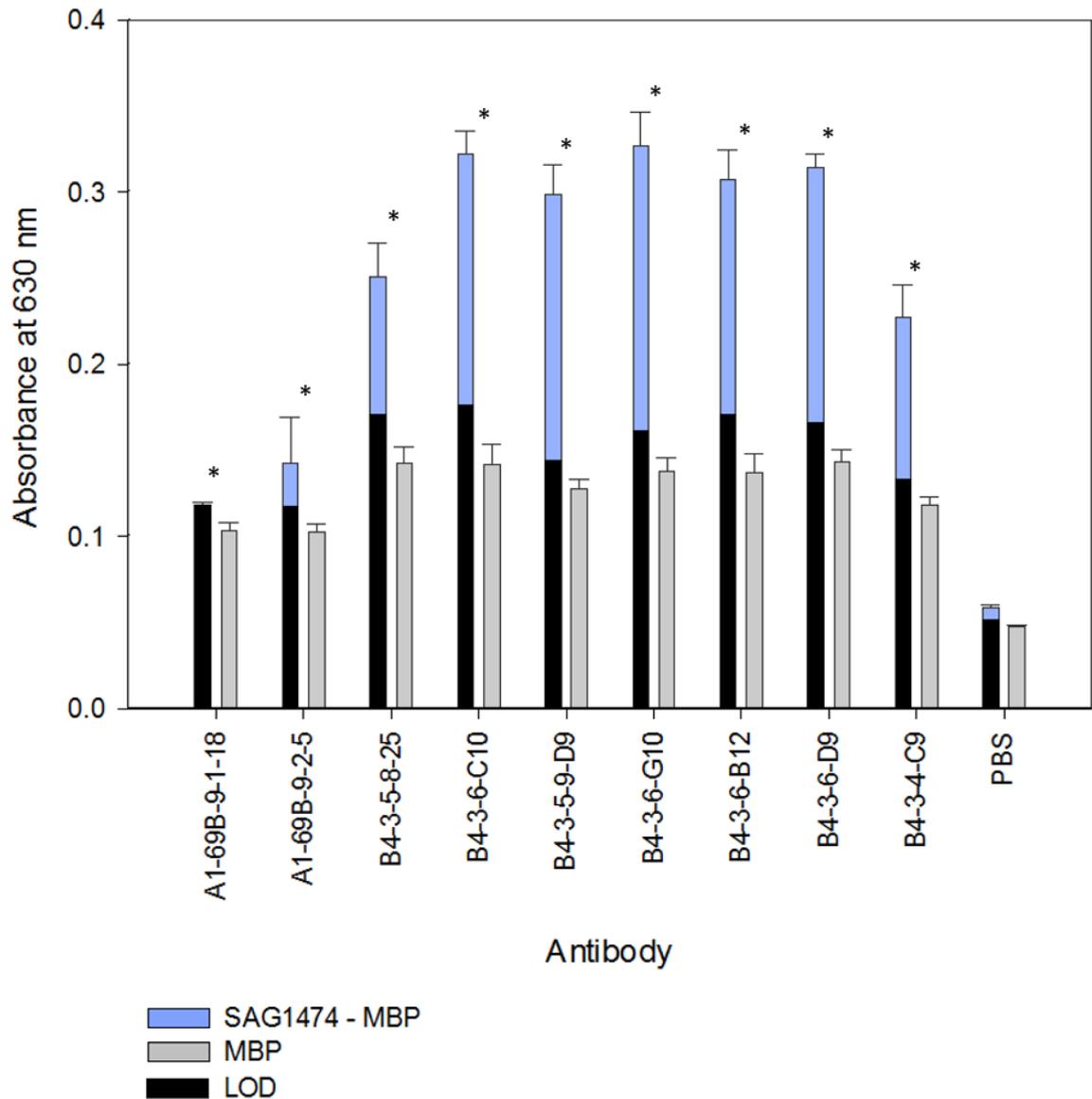


Figure 4.19 Antibody binding to recombinant SAG1474-MBP protein  
 ELISAs were performed to assess antibody binding to SAG1474-MBP protein (blue) and MBP (grey). The limit of detection (mean of the signal obtained from the negative control (MBP) + three times its standard deviation) is shown in black. Plates were coated with protein and incubated with the panel of antibodies at  $2 \mu\text{g ml}^{-1}$ , followed by incubation with HRP-linked anti mouse IgG antibody. Colorimetric change of TMB was measured at 630 nm after 30 minutes. Error bars show standard deviation, n=3.

\* indicates a p value < 0.1, \*\* indicates a p value < 0.05, as determined by a Mann Whitney U test.

#### 4.6 Antibody characterisation against bacterial cells

Antibodies with affinity for peptide B show the most promise and were investigated for binding to bacteria. A critical feature of an antibody for use in a diagnostic test is a lack of cross-reactivity. To assess this, antibody binding to a bacteria belonging to the same genus, *Streptococcus pyogenes*, was examined. SAG1474 has a homologue within *S. pyogenes*, AmiC, which functions as an amidase and has an amino acid identity and similarity to SAG1474 of 36% and 54% respectively. (Glaser *et al.*, 2002) Therefore *S. pyogenes* is a suitable choice as a negative control. Ideally a diagnostic test would require as little processing as possible. An antibody that binds to non-lysed, or whole, cells would be beneficial to reduce the number of sample preparation steps. Antibody binding to lysed and whole cell *S. agalactiae* and *S. pyogenes* samples was investigated, see figure 4.20. *S. agalactiae* serotype III was utilised, as this is the one of the most prevalent serotypes in infectious disease. (Berg *et al.*, 2000; Ippolito *et al.*, 2010; Soares *et al.*, 2013; Vinnemeier *et al.*, 2015) In this example *S. pyogenes* strain M5 Manfredo58 was used.

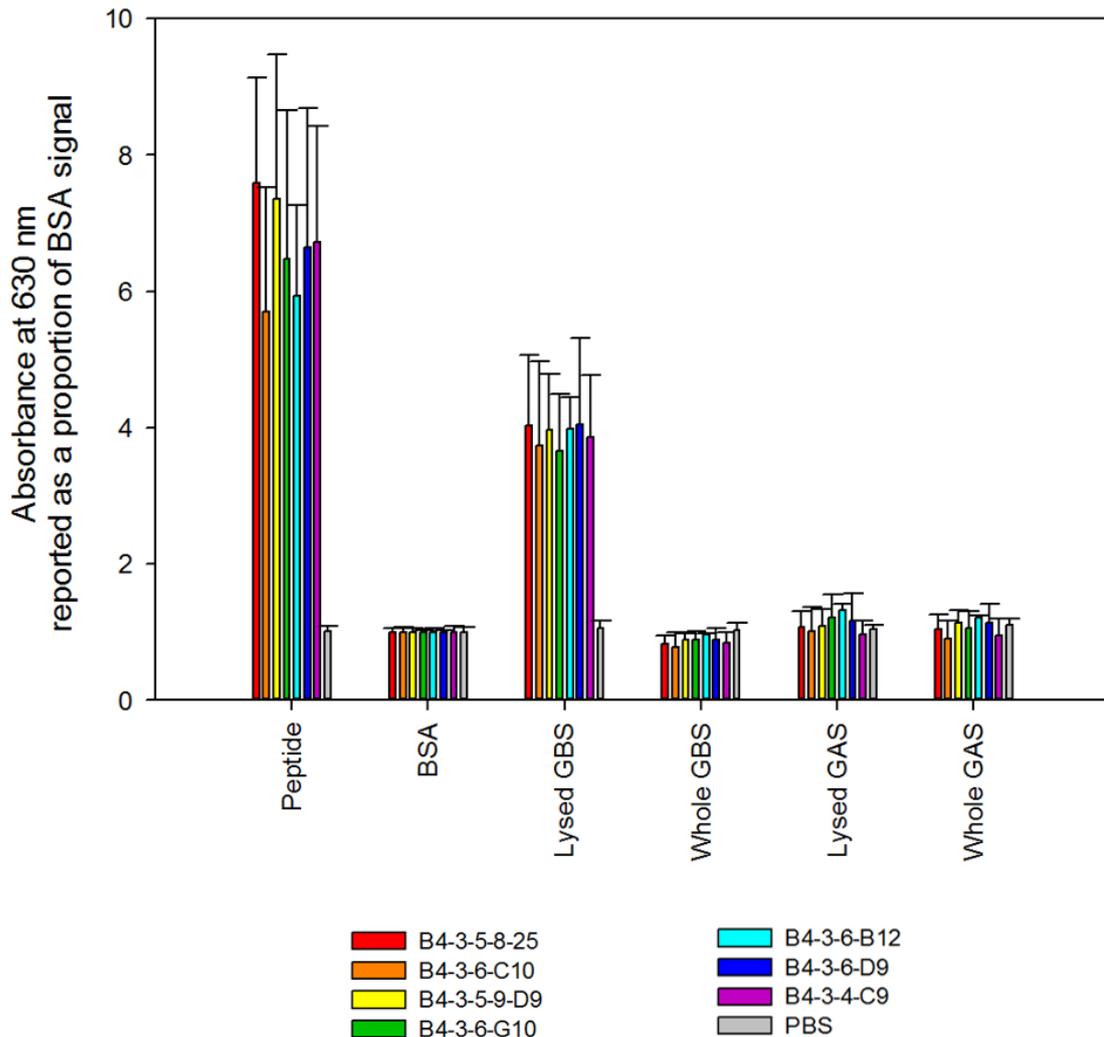


Figure 4.20 Antibody binding to lysed and whole cell *S. agalactiae* and *S. pyogenes* ELISA plates were coated with peptide and lysed and whole-cell group B streptococcal cells. Negative controls of lysed and whole-cell group A streptococcal cells and BSA were also coated onto ELISA wells. Wells were incubated with the panel of antibodies at  $2 \mu\text{g ml}^{-1}$ , followed by incubation with HRP-linked anti mouse IgG antibody. Colorimetric change of TMB was measured at 630 nm after 30 minutes. Error bars show standard deviation. ELISA is able to distinguish between lysed group A and group B streptococcal cells. B4-3-5-8-25, B4-3-6-C10, B4-3-5-9-D9, B4-3-6-G10, B4-3-6-D9, B4-3-4-C9, n=8; B4-3-6-B12, n=4; PBS, n=16.

The signal obtained for antibody binding to lysed *S. agalactiae* cells is lower than to the peptide. When the proportion of target found in the coating material is considered, the result is expected. Wells which are coated in peptide-BSA contain a large number of target sequences, whereas the epitope makes up only a small proportion of the

SAG1474 protein which in turn makes up a small proportion within the *S. agalactiae* lysate. In a sandwich ELISA format SAG1474 would be captured on the plate, eliminating contaminating proteins, therefore enabling production of a higher signal. In respect to differentiating *S. agalactiae* and *S. pyogenes*, the antibodies perform well. However, the antibody was unable to bind whole cell *S. agalactiae*, producing a signal equivalent to that obtained with the BSA negative control. This experiment required a lysis step. Although a desirable feature, whole cell binding is not essential for diagnostics.

#### 4.6.1 Comparison of antibodies

The seven antibodies were analysed to assess which is the most promising for the diagnosis of group B streptococcal disease. Expressing the signal obtained for *S. agalactiae* as a proportion to that obtained for *S. pyogenes* allows the ability of each antibody to discriminate between the target of interest and contaminating bacteria to be compared. The values calculated for each antibody are similar, ranging from 3.0-4.0, see table 4.2.

Antibody	<i>S. agalactiae</i> signal as a proportion of <i>S. pyogenes</i> signal
<b>B4-3-4-C9</b>	4.0
<b>B4-3-6-C10</b>	3.7
<b>B4-3-5-8-25</b>	3.7
<b>B4-3-5-9-D9</b>	3.7
<b>B4-3-6-D9</b>	3.5
<b>B4-3-6-G10</b>	3.1
<b>B4-3-6-B12</b>	3.0
<b>PBS</b>	1.0

**Table 4.2 Comparison of antibody binding to *S. agalactiae* and *S. pyogenes***

The table shows ELISA signals derived from antibody binding to lysed *S. agalactiae* cells as a proportion to those obtained for binding to *S. pyogenes*.

B4-3-6-B12 has a decreased ability to detect lysed *S. agalactiae* cells over lysed *S. pyogenes* cells, which approaches statistical significance ( $p < 0.1$ ) in comparison to B4-3-5-9-D9 and B4-3-6-D9 antibodies, and is statistically significant ( $p < 0.05$ ) with respect to B4-3-5-8-25, B4-3-6-C10 and B4-3-4-C9. There are no statistical differences

between the remaining antibodies, determined by a Kruskal-Wallis test. As there is no outstanding antibody, all antibodies were considered for future work.

#### 4.6.2 Antibody characterisation against *S. agalactiae* serotypes

The purpose of using IDRIS was to aid peptide selection by determining peptides found in all *S. agalactiae* strains and in no other bacteria. A panel of 10 strains representing distinct *S. agalactiae* serotypes was assessed for antibody binding, as well as three *S. pyogenes* strains, see figure 4.21. Antibodies show increased binding to all strains of *S. agalactiae* tested in comparison to those of *S. pyogenes* (p values <0.05, determined by a Mann-Whitney U test, GBS: n=10, GAS: n=3). Thus demonstrating the suitability of these antibodies to be used in a diagnostic test. This test also validates the use of IDRIS to select peptide tokens.

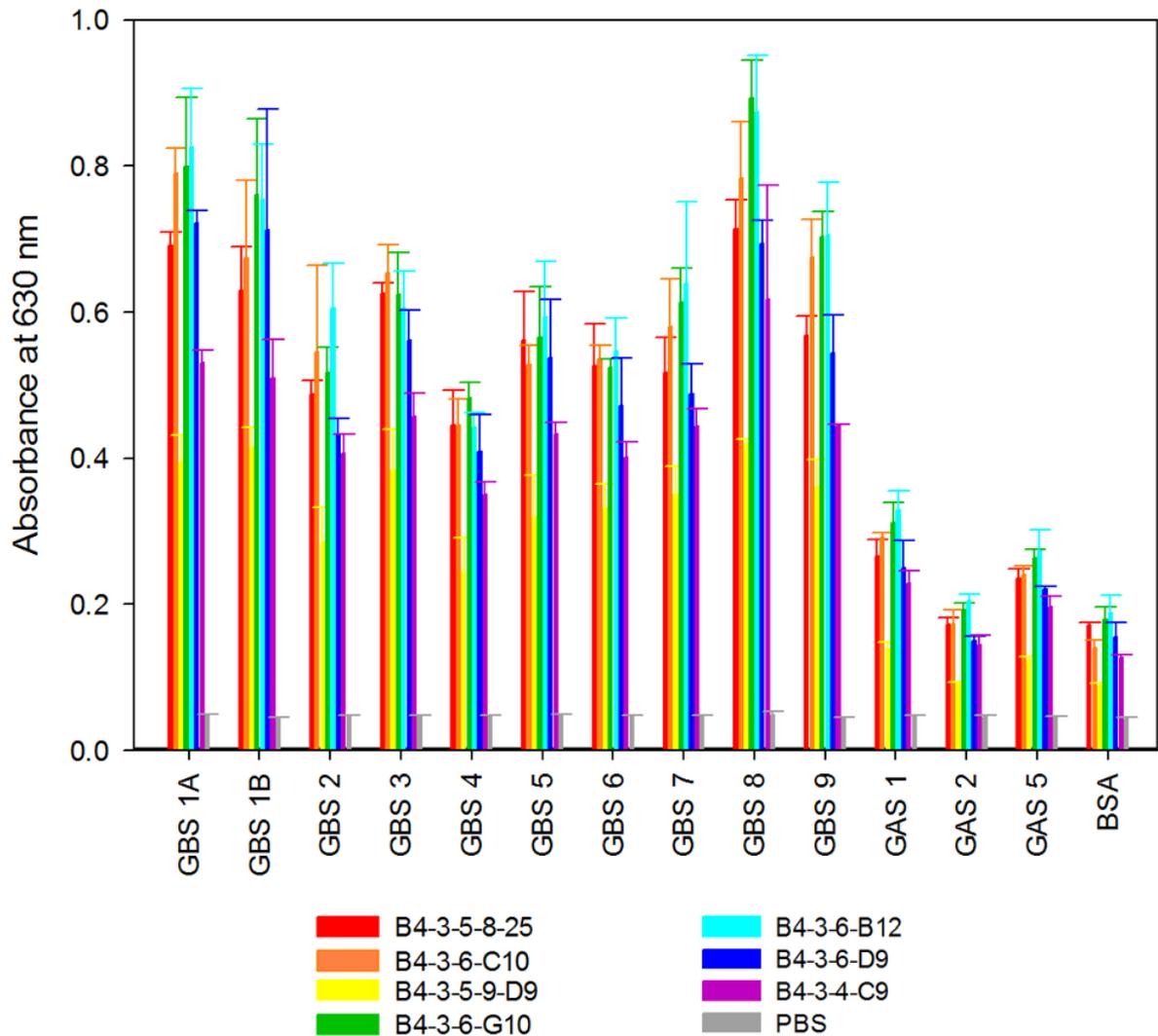


Figure 4.21 Assessment of antibody sensitivity and specificity

10 strains of *S. agalactiae* and three strains of *S. pyogenes* were lysed and coated onto ELISA plates for analysis. Negative control wells were coated in BSA. Wells were incubated with the panel of antibodies at  $2 \mu\text{g ml}^{-1}$ , followed by incubation with HRP-linked anti mouse IgG antibody. Colorimetric change of TMB was measured at 630 nm after 20 minutes. Error bars show standard deviation,  $n=3$ . Results show antibodies exhibit specificity in respect to binding *S. agalactiae* over *S. pyogenes*, and sensitivity to all strains of *S. agalactiae* tested.

#### 4.7 Conclusions and future work

Promising antibodies exhibiting target binding to peptide B (VNLEENSQV) were created. These antibodies are incapable of binding to the exterior of the cell, therefore whole cell diagnostics is not possible. A pair of antibodies, binding distinct regions of the same protein, is required to create an immuno-sandwich detection system. Antibodies generated against peptides A (VQTNDSNPT) and C (NMENLSQEERI) showed little to no binding. Therefore further antibodies need to be created in order to construct a functional diagnostic test. As specificity is ensured by the original antibody, the second antibody may be in the form of a monoclonal antibody or polyclonal antibody raised against the recombinant protein.

Possible reasons for the lack of binding to whole cell *S. agalactiae* include epitope coverage by the cell wall or polysaccharide capsule. It is suggested that the capsule does not entirely cover the bacterial cell at all times, as the Sip protein evokes an immunoprotective response to GBS infection when injected in mice. (Brodeur *et al.*, 2000) It is possible that cells grown under laboratory conditions have a different morphology to those found *in vivo*, cell wall associated proteins may be exposed in patient samples. Alternatively the protein could be partially hidden, therefore producing antibodies against spatially distant sites of the protein may enable whole cell binding.

The antibody signal generated by affinity to lysed *S. agalactiae* is higher than to recombinant SAG1474-MBP. As the protein constitutes a fraction of the cell lysate, and plates were coated at a concentration in excess of the maximal binding efficiency ( $2 \mu\text{g ml}^{-1}$ ), it was expected for the protein derived signal to be higher. It is possible that the target epitope is hidden through incorrect folding of the recombinant protein or concealment by the MBP moiety. Decreased binding may arise from partial or transient blocking of the binding site. As the recombinant protein contains a periplasmic signalling sequence, there may be two protein populations, one in the cytoplasm and one in the periplasm, which may be in two differentially folded states. As periplasmic extraction did not yield sufficient protein, whole cell protein extraction was performed,

and both populations harvested. Creation of a signal sequence deficient SAG1474-MBP construct, which remains within the cytoplasm, should enable a single population of protein to be extracted, which may be correctly folded. Future work may include the purification of native SAG1474 from *S. agalactiae* cells through affinity with the antibodies developed in this work. This would eliminate any effect that recombinant protein production and fusion with MBP causes, however the purification process will be limited by the levels found in the native cell.

Additional work is needed to fully characterise antibody binding affinities to determine which of the seven antibodies raised against peptide B is best suited for group B streptococcal diagnostics. Surface plasmon resonance could be performed to analyse dissociation rates. Antibodies could be isotyped and sequenced to gain more information. Antibody specificity needs to be investigated further. A panel of bacteria which would be present in vaginal or rectal swabs should be investigated. Organisms which contain peptide sequences closely matching the immunogenic peptide should also be evaluated for antibody binding. To validate a diagnostic test, performance in patient samples must be studied. Initially, bacteria-spiked patient samples should be tested, followed by evaluating real patient samples. Ideally this would be performed blinded.

---

## Chapter 5 RPA DNA amplification

---

### 5.1 Introduction

#### 5.1.1 DNA amplification in diagnostic tests

Molecular diagnostics is considered the gold standard for the detection of many infectious diseases. These techniques can exhibit high sensitivity alongside high specificity, and vastly reduce diagnosis times in comparison to traditional microbial cell culture. (Emmadi *et al.*, 2011) The platform is adaptable, and tests for new disease targets can be developed rapidly as there is no reliance on the time-consuming production of affinity molecules, such as antibodies. (Lipman *et al.*, 2005) This means molecular diagnostics can be readily mobilised on the emergence of new pathogens or genetic shifts.

The most common molecular technique is PCR. PCR amplifies specific DNA using a thermostable polymerase, isolated from *Thermus aquaticus*. Template DNA is denatured, target-specific primers bind to the single stranded DNA and strand extension is performed by the polymerase. Each of the three stages, has a distinct optimum temperature, therefore the process is co-ordinated by temperature cycling. This process is typically repeated 25-35 times with DNA concentration doubling with each round of PCR. (Saiki *et al.*, 1988) Simple PCR requires a separate readout step such as gel electrophoresis to visualise the results. Quantitative real time PCR incorporates the readout stage into the amplification process. Fluorescent dyes are used to monitor amplification, either binding non-specifically to the increasing amount of double stranded DNA or binding specifically to amplicons via targeted probes. (Saunders, 2009) PCR's success is due to its high sensitivity. Quantitative PCR enables the detection of single molecules of DNA, with variations of the technique, such as digital PCR, enabling quantification of minority sequences in a sample. (Vogelstein and Kinzler, 1999) PCR is able to specifically identify related bacterial species, for instance 20 intestinal bacterial species were shown to cause no cross-reactivity. (Ott *et al.*, 2004)

Despite PCR's dominance within DNA amplification, it does have considerable drawbacks relevant to the field of diagnostics, one of which is speed. Conventional PCR operates for a specified number of cycles, typically taking over one hour, with an additional 30 minutes per kb of DNA. (New England BioLabs, 2017a) Real time PCR reports the number of cycles required for the fluorescent signal to reach a level equivalent to a specified DNA concentration. Therefore a sample with a high concentration of template DNA will take a short amount of time to report a positive result, however a negative result can only be confirmed in a longer time period. (Saunders, 2009) The large expense of a thermocycler means PCR tends to be performed in centralised healthcare laboratories. (St John and Price, 2014) This increases the time from sample collection to result considerably. Its high price also means it is economically unviable for use in screening of low risk diseases or in developing countries.

#### 5.1.2 Isothermal DNA amplification

Isothermal DNA amplification techniques are alternatives to PCR. As amplification is performed at a single temperature, expensive thermocyclers are not required and equipment costs are reduced. This makes DNA amplification accessible for wider healthcare applications and settings.

There are many variations of isothermal DNA amplification, which can be classified into three major groups. The first group amplifies DNA exponentially, these methods are associated with high sensitivity. Techniques within the second group amplify DNA linearly, reducing the sensitivity, but maximising specificity. The final group amplify DNA linearly and feed the products into a cascade event to increase the signal, thus overcoming the limitations of linear techniques. Exponential DNA amplification techniques are the most studied methods. (Zhao *et al.*, 2015) The properties of commonly used isothermal techniques are detailed in table 5.1. A rapid point-of-care test would ideally show large amplification of DNA at a low temperature over a short period of time. Recombinase polymerase amplification is able to amplify DNA over an extremely short period of time. This time frame is essential for point-of-care testing.

Name	Type	Amplification Efficiency	Amplification Time / mins	Operating Temperature / °C	Reference
RPA	Exponential	10 <sup>12</sup> fold (10 <sup>4</sup> fold every 10 minutes)	5-20	25-42	(Daher <i>et al.</i> , 2016; TwistDx, 2016)
LAMP	Exponential	10 <sup>9</sup> -10 <sup>10</sup> fold	60	60-65	(Sahoo <i>et al.</i> , 2016)
HDA	Exponential	10 <sup>7</sup> fold	120	37-65	(Vincent <i>et al.</i> , 2004; An <i>et al.</i> , 2005)
NASBA	Exponential	10 <sup>9</sup> fold	120	37	(Compton, 1991)
HRCA	Exponential	10 <sup>9</sup> fold	90	65	(Lizardi <i>et al.</i> , 1998; Hamidi <i>et al.</i> , 2015)
PGRCA	Exponential	10 fold every 20 minutes	30-150	60	(Murakami <i>et al.</i> , 2009)
ESDA	Exponential	10 <sup>7</sup> fold	120	37	(Walker <i>et al.</i> , 1992)
LRCA	Linear	10 <sup>3</sup> fold	30-60	37	(Nallur <i>et al.</i> , 2001)
LSDA	Linear	20 fold	40	37	(Joneja and Huang, 2011)

**Table 5.1 Properties of isothermal DNA amplification techniques**

Common isothermal DNA amplification techniques were analysed for their suitability for use as point-of-care diagnostic tests. Amplification efficiency, amplification time and optimal temperature were considered. The techniques evaluated were RPA (recombinase polymerase amplification); LAMP (loop mediated isothermal amplification); HDA (helicase dependent amplification); NASBA (nucleic acid sequence based amplification); HRCA (hyper-branched rolling circle amplification); PGRCA (primer generation rolling circle amplification); ESDA (exponential strand displacement amplification); LRCA (linear rolling circle amplification); and LSDA (linear strand displacement amplification).

### 5.1.3 Recombinase polymerase amplification

RPA employs recombinases and polymerases to amplify short sequences (typically 80-200 bp) of DNA in 5-20 minutes. It operates between 25 °C and 42 °C, and is capable of amplifying single copy numbers of DNA 10<sup>12</sup> fold. It is the only isothermal technique which has been shown to operate using body heat, (Crannell *et al.*, 2014; Daher *et al.*, 2016) making it applicable to resource-limited settings.

Recombinases are utilised in homologous recombination, playing a key role in DNA repair mechanisms. RecA is a well-studied example of a recombinase. Recombinase molecules polymerise along the two primers, creating a structure with a second DNA binding pocket into which the double stranded template binds. (Renkawitz *et al.*, 2014)

The RecA-primer nucleoprotein then searches for regions of homology within the double stranded DNA, it is proposed that this occurs by sliding in either direction along the filament. Once regions of homology are found, base pairing occurs and one strand of the template DNA is displaced, forming the D-loop. The displaced DNA is stabilised by single stranded DNA binding protein (SSB). ATP is generated by creatine kinase, and hydrolysed by RecA, causing RecA to dissociate from the DNA strand. A strand-displacing polymerase can then enter and extend the DNA sequence. Parental strands separate, and polymerisation continues to the end of the DNA strand. DNA sequences are fed back into the system so amplicons are exponentially amplified. (Daher *et al.*, 2016) A schematic of the process is shown in figure 5.1.

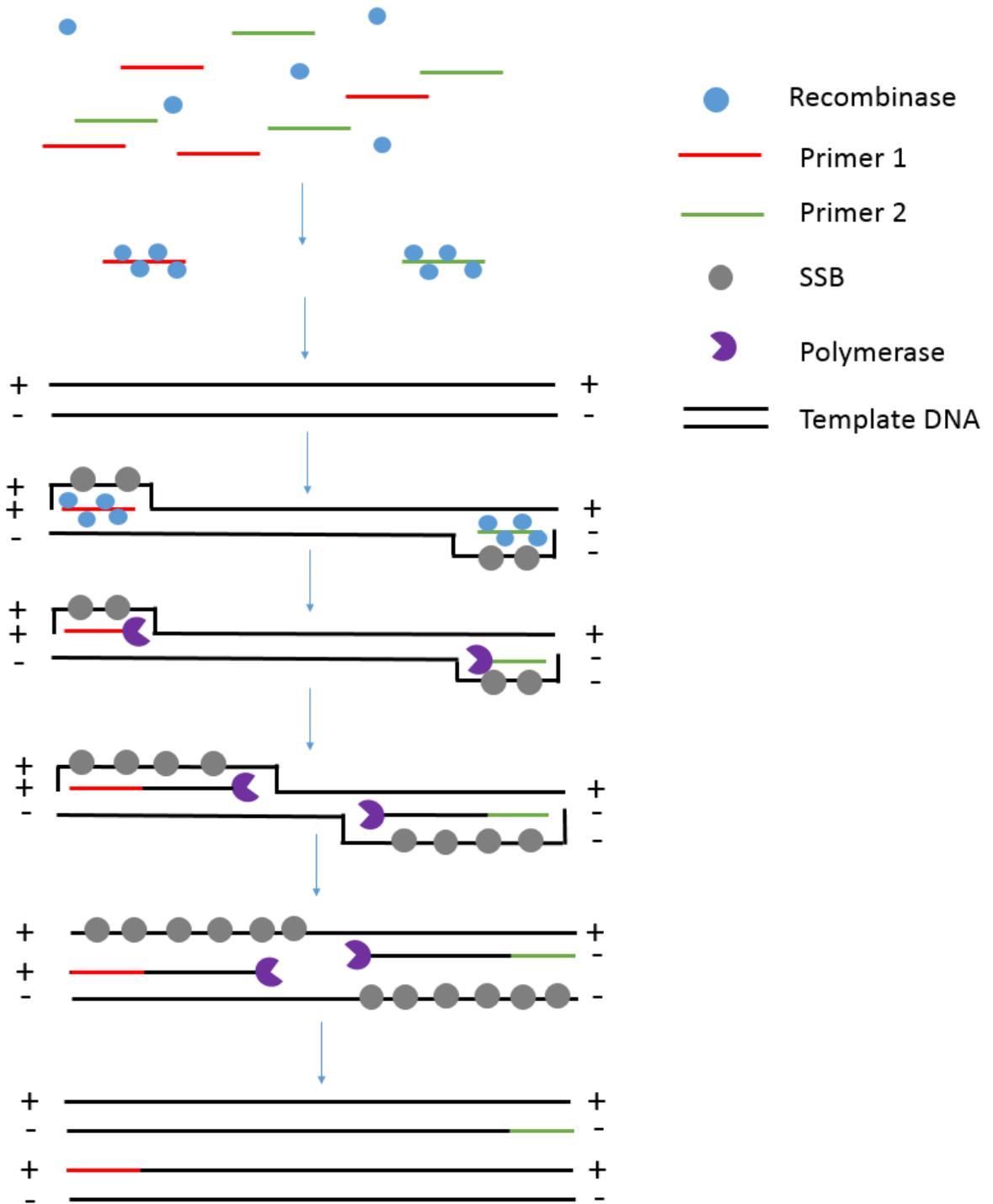


Figure 5.1 Diagrammatic representation of recombinase polymerase amplification  
 A protein mix of recombinase, polymerase and single stranded binding protein (SSB) facilitate the amplification of short regions of template DNA using two single stranded primers. Adapted from (Tröger *et al.*, 2015).

TwistDx offer a number of commercially available variations of the RPA kit which allow detection through various means: agarose electrophoresis (basic kit), fluorescence (exo and fgp kits), and lateral flow assays (nfo kit). Kits are also available which include reverse transcriptase enzymes, to enable detection of RNA. (TwistDx, 2016)

exo probes contain a fluorophore and quencher separated by a THF residue. Once DNA amplification is performed, probes can bind to the double stranded amplicon, in this form the exo enzyme is able to cleave the THF residue. This decouples the fluorophore and quencher. fgp probes contain a quencher and a fluorophore attached to the probe via a C-O-C linker. In its double stranded form, the linker can be attacked by the fgp enzyme, releasing the fluorophore. In both systems fluorescence is observed when released from the quencher and this can be measured in real time.

The TwistDx nfo kit is compatible with lateral flow detection. In this system RPA is performed as usual, however one of the primers is labelled with a 5' biotin. nfo probes contain a 5' FAM protein and a 3' modification that blocks DNA polymerisation, separated by a THF residue. When bound to the double stranded amplicon, nfo cleaves the THF residue, removing the polymerisation block and enabling RPA to extend this strand. The product from this second RPA reaction is an amplicon with a FAM protein attached at one end and a biotin label at the opposing end. Biotin is captured on a lateral flow strip, and a gold labelled anti-FAM antibody is used to detect the RPA product. This strategy has been employed to detect parasite infections, (Cordray and Richards-Kortum, 2015; Nair *et al.*, 2015; Sun *et al.*, 2016) zoonotic viruses (Wu *et al.*, 2016; Yang *et al.*, 2017) and bacterial infections. (Krolov *et al.*, 2014) The technique can be combined with reverse transcription to detect RNA on lateral flow strips. (Yang *et al.*, 2017)

#### 5.1.4 Hybridisation-based detection of RPA amplicon

Fluorometric measurement of RPA reactions increases the cost of the technique, limiting the applications RPA can be utilised in. The nfo probe system incorporates RPA into a lateral flow assay, suitable for cheap point-of-care diagnostics. To decrease

the cost further, an antibody-free system which detects RPA amplicons via nucleic acid hybridisation on lateral flow assays has been devised. (Jauset-Rubio *et al.*, 2016b)

Target-specific primers are designed with 5' single stranded overhang sequences. To ensure the full sequence is not amplified, a three-carbon spacer (C3) modification, which blocks polymerisation, is inserted between the primer and overhang regions. RPA creates an amplicon with a single stranded tail at either end. The complementary DNA sequence to one of the tails acts as the capture probe, which is coated in a line on the lateral flow strip. The complementary sequence to the second tail is covalently bonded to a gold nanoparticle and acts as the reporter probe. When the RPA product flows up the strip it is bound by hybridisation to the capture probe and is linked to signal detection by hybridisation to the reporter probe. This is shown diagrammatically in figure 5.2.

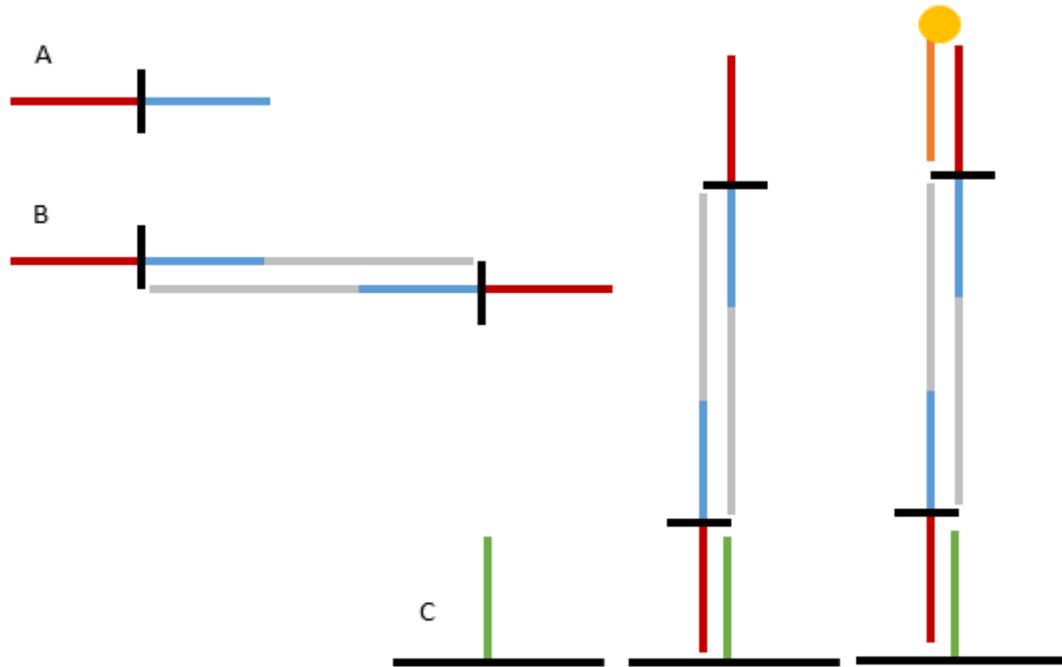


Figure 5.2 Visualisation of tailed RPA detection

A: Tailed-primer. The single stranded overhang (red) is separated from the primer sequence (blue) by a modification which blocks DNA polymerisation (black). B: Amplicon product. RPA extends the DNA (grey) from the primer sequence, creating a double stranded product, with 2 single stranded overhang sequences on opposing strands. C: RPA product binding to the lateral flow strip. Capture probe (green) is attached to the lateral flow strip. This hybridises to its complementary sequence, found in the forward primer overhang in the RPA amplicon. A gold nanoparticle (yellow) is conjugated to the reporter probe (orange) and this hybridises to its complementary sequence in the reverse primer overhang. Thus producing a readable signal.

## 5.2 Recombinase polymerase amplification for the detection of *S. agalactiae*

In previous studies RPA has been employed for the amplification and detection of *S. agalactiae* from vaginal, anal and blood samples. Fluorescent measurements of probes were used to observe amplification reactions. (Daher *et al.*, 2014; Clarke *et al.*, 2016) There is currently no lateral flow-based RPA test for GBS. The aim of this project is to create a *S. agalactiae* detection system to be read on a lateral flow strip, based on the hybridisation of RPA products.

### 5.2.1 Analysis of primers

For a DNA amplification based diagnostic test, a primer pair which is found in all *S. agalactiae* strains, and in no other bacteria is required. Sequences of DNA that correspond to IDRIS tokens were considered. Although the IDRIS system is based on identification of distinctive protein targets, there is an increased chance that the corresponding DNA is also distinctive to *S. agalactiae* strains, in comparison to randomly selected sequences.

Prior to the creation of tailed primers, a panel of non-tailed primers was designed, see table 5.2. It was previously advised that RPA primers should be 30-35 residues in length for optimum efficiency, (TwistDx, 2016) however recently it has been discovered that PCR primers, down to 18 nucleotides in length, are also suitable. (TwistDx, 2017) Therefore sequences of both lengths were selected for analysis. Primer sequences were computationally assessed to predict melting temperatures of folded primer structures and the change in Gibbs' free energy ( $\Delta G$ ) associated with primer self-dimers and hetero-dimers, and to determine amplicon length. The sequences of all currently sequenced *S. agalactiae* strains were aligned and conservation of primer sequences was assessed. The number of bases at which sequence variation occurred between strains was reported as DNA mismatches.

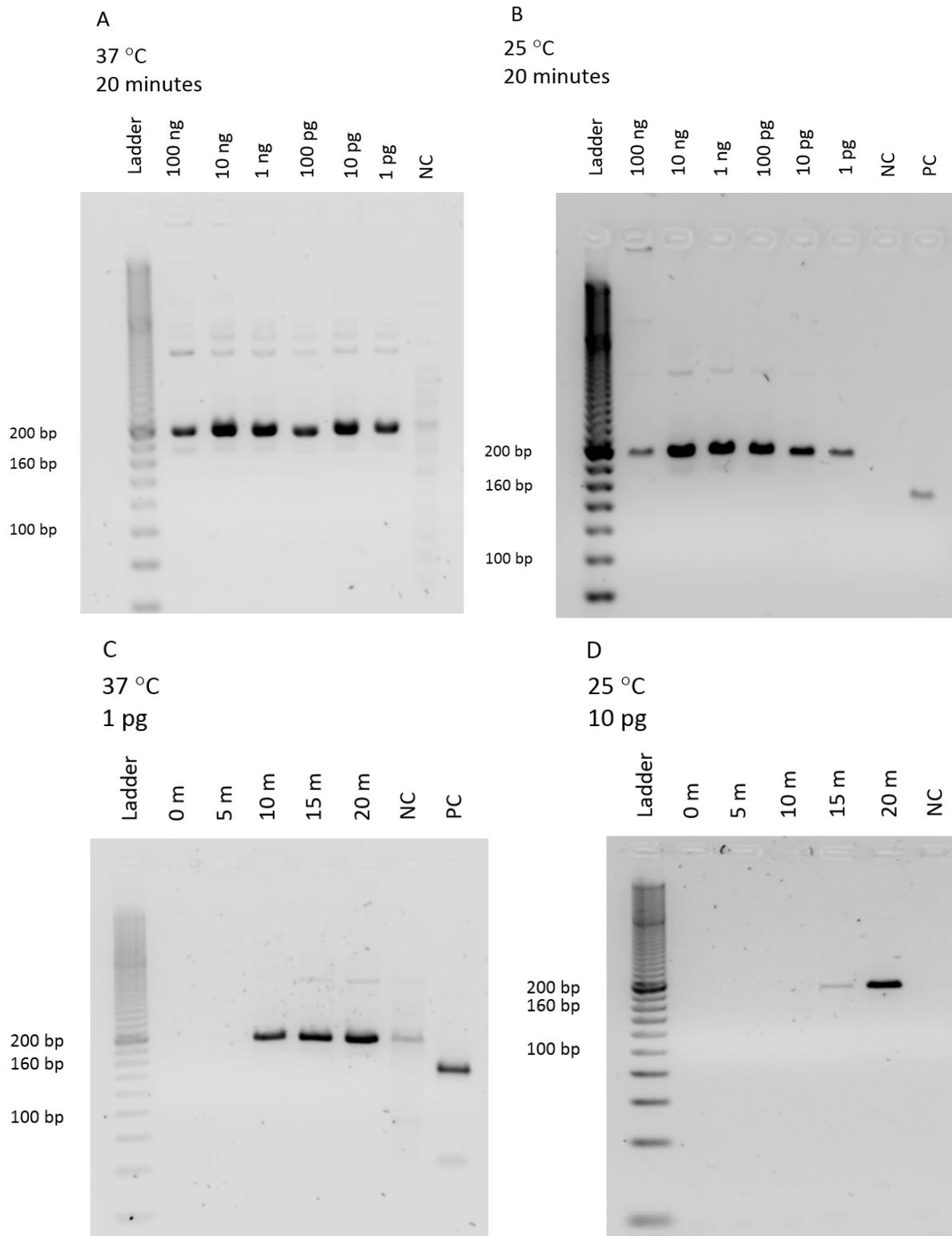
Name	Token	DNA mismatches	Amplicon length	Maximum Hairpin Tm /°C	Maximum Self-dimer $\Delta G$ /kcal mol <sup>-1</sup>	Maximum Hetero-dimer $\Delta G$ /kcal mol <sup>-1</sup>
A Fwd	Yes	0	142	31.2	-5.02	-6.6
A Rev	Yes	0	-	41.6	-7.31	-
B Fwd	Yes	1	191	37.5	-5.02	-6.14
B Rev	No	1	-	30.9	-8.74	-
C Fwd	Yes	1	191	37.3	-4.52	-6.14
C Rev	No	1	-	30.9	-3.14	-
D Fwd	Yes	1	267	32.3	-3.61	-5.47
D Rev	Yes	0	-	28.9	-6.84	-
E Fwd*	Yes	1	185	22.4	-6.34	-6.5
E Rev*	No	0	-	20.2	-3.61	-
F Fwd*	Yes	1	116	32	-6.34	-4.67
F Rev*	Yes	1	-	33.4	-4.85	-
G Fwd	No	1	131	47.6	-6.84	-6.94
G Rev	No	1	-	38.4	-3.55	-
H Fwd	No	0	176	37.3	-3.61	-5.19
H Rev	Yes	1	-	24.6	-5.02	-

**Table 5.2 Analysis of primer properties for amplification of GBS DNA**

The table compares the properties of several primer pairs which may affect RPA. Primer pair E was selected for its favourable thermal properties, with low hairpin melting temperatures and low  $\Delta G$  values associated with formation of self- and hetero-dimers. Primer pair F was selected due to its short amplicon length predicted to enable amplification in a short period of time. Both primer pairs are well conserved with either 0 or 1 mismatches between all GBS strains currently sequenced. Both primer pairs also have at least 1 primer contained within an IDRIS token region, indicating its uniqueness to other bacteria and high specificity. Thermal properties were calculated with the following conditions: 0.5  $\mu$ M oligo, 0 mM Na<sup>+</sup>, 14 mM Mg<sup>2+</sup>, 0.24 mM dNTPs at 25 °C using IDT software.

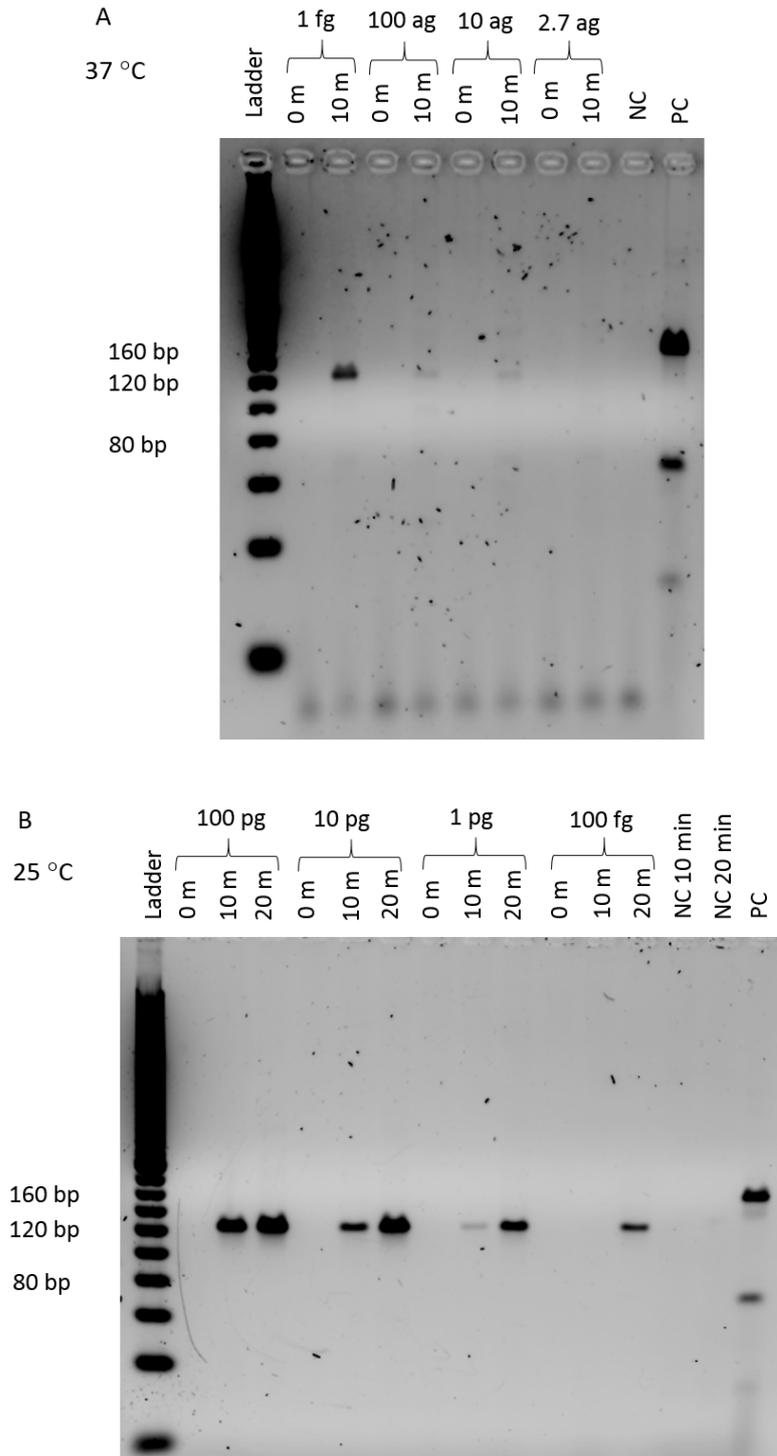
The two most promising primer pairs were pair E and F. Primer pair E had the lowest hairpin melting temperatures, and also showed favourable predicted  $\Delta G$  values of primer dimerisation. It was proposed that this primer pair would function at low temperatures. Primer pair F had the shortest amplicon length and therefore had the potential to be amplified in the shortest time. Untailed versions of primer pair E and F were empirically evaluated in order to select one primer pair to develop into the tailed-primer detection system.

DNA was extracted from *Streptococcus agalactiae* and a 2583 bp region was amplified by PCR. The PCR product was used as the RPA template. Unless specified, DNA from *S. agalactiae* serotype III was utilised, as this is the one of the most prevalent serotypes in infectious disease. (Berg *et al.*, 2000; Ippolito *et al.*, 2010; Soares *et al.*, 2013; Vinnemeier *et al.*, 2015) RPA was performed using primer pair E, see figure 5.3, and primer pair F, see figure 5.4, at various concentrations and for various time points. To analyse amplification, agarose electrophoresis was performed and gels were visualised through a 600 nm filter.



**Figure 5.3 RPA of GBS DNA using primer pair E**

A: RPA: 37 °C, 20 minutes, 100 ng – 1pg. B: RPA: 25 °C, 20 minutes, 100 ng – 1pg. C: RPA: 0-20 minutes, 1 pg. D: 25 °C, 0-20 minutes, 10 pg. No template DNA negative controls (NC) and TwistDx positive controls (PC) were used. RPA was performed and samples ran on 3% agarose gels. Gels were imaged with a Licor Odyssey Fc. RPA using primer pair E is able to amplify 1 pg DNA to a detectable level at 37 °C in 10 minutes, and 10 pg DNA at 25 °C in 15 minutes.



**Figure 5.4 RPA of GBS DNA using primer pair F**

A: RPA: 37 °C, 1 fg–2.7 ag, 0, 10 minutes. No template DNA negative control (NC) 10 minutes, TwistDx positive control (PC) 10 minutes. B: RPA: 25 °C, 100 pg–100 fg, 0, 10, 20 minutes. No template DNA negative control (NC) 10, 20 minutes, TwistDx positive control (PC) 20 minutes. RPA was performed and samples ran on 3% agarose gels. Gels were imaged on a Licor Odyssey Fc. RPA using primer pair F is able to amplify 10 ag DNA to a detectable level at 37 °C in 10 minutes, and 100 fg DNA in 20 minutes at 25 °C.

Using primer pair E, RPA was able to amplify 1 pg DNA to a detectable level at 37 °C in 10 minutes, and 10 pg of DNA at 25 °C in 15 minutes. When RPA was performed with primer pair E at 37 °C for 20 minutes, a faint band was observed in the no template control sample, indicating an amplification product was formed. This is likely to result from contamination from neighbouring reactions. False positive results were not observed in subsequent agarose gels. Using primer pair F, RPA was able to amplify 10 ag DNA to a detectable level at 37 °C in 10 minutes, and 100 fg DNA in 20 minutes at 25 °C. 10 ag of DNA correlates to 3.8 copies of DNA, therefore the amplification is extremely sensitive. The prediction that primer pair E would be best suited to amplifying DNA at low temperatures was incorrect. It appears that despite the prediction that F primers would exist in hairpin structures at 25 °C, DNA amplification is still possible at this temperature. The amplicon size is the dominant feature to determine DNA amplification efficiency, even at low temperatures. It is suggested that DNA clean-up is performed prior to analysis on agarose gel, as RPA proteins may cause bands to appear smeared. (TwistDx, 2016) However as smearing was not observed, DNA clean-up was not performed. A further consideration is that a rapid point of care test procedure would not allow for DNA purification prior to analysis. The presence of proteins may account for the slight difference in observed mobility and the predicted mobility determined by the expected molecular weights, 185 bp using primer pair E and 116 bp using primer pair F.

Amplification of DNA by primer pair F in 10 serotypes (1A, 1B, 2-9) was analysed at 37 °C (shown in figure 5.5) and 25 °C (shown in figure 5.6). Consistent DNA amplification was achieved in all 10 strains at 37 °C. DNA was amplified in all serotypes at 25 °C, however there was a reduction in amplification of DNA extracted from serotypes 1A and 5.

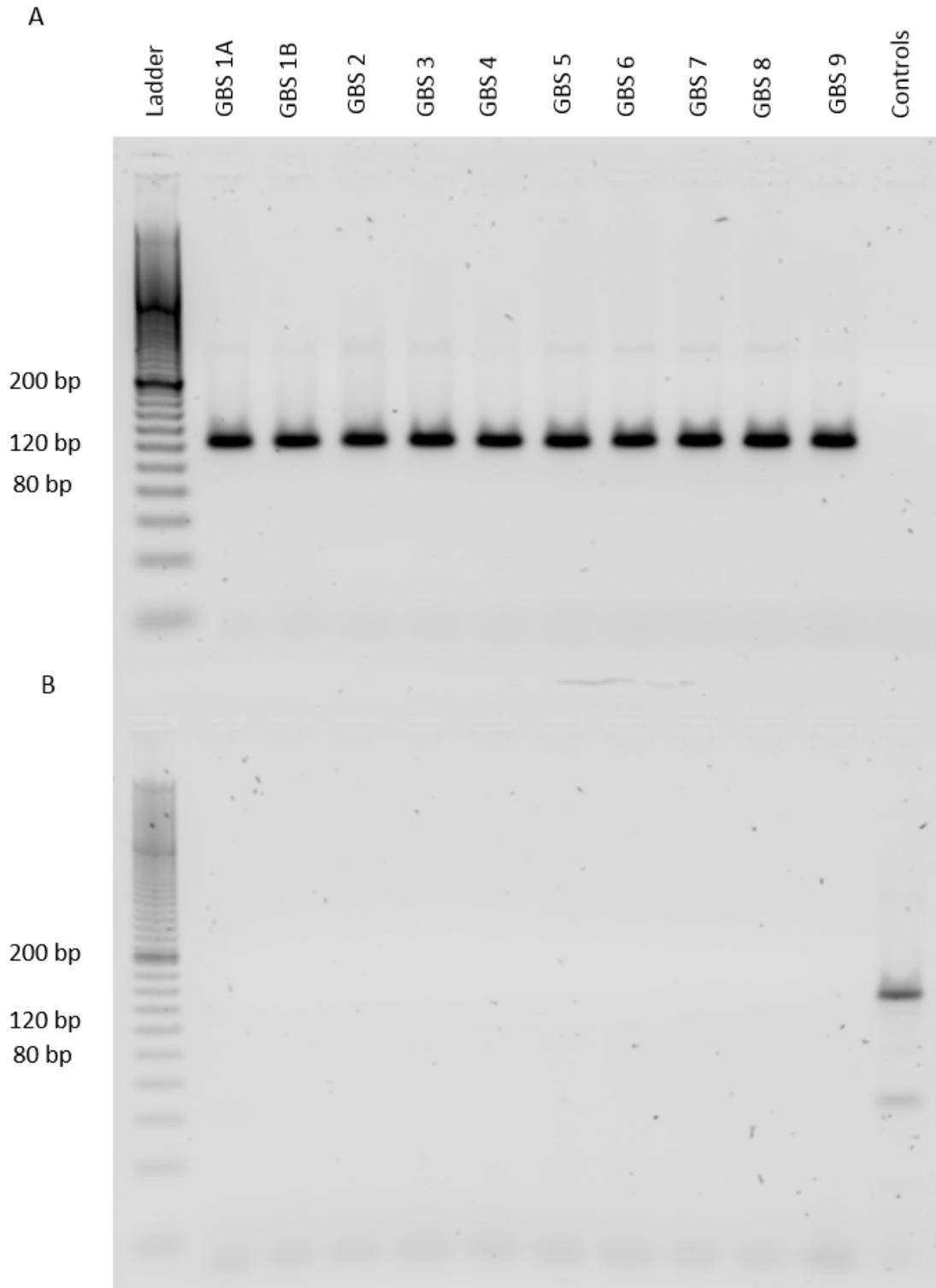


Figure 5.5 RPA of GBS DNA using primer pair F in 10 serotypes at 37 °C

A: RPA: 37 °C, 10 minutes, 10 fg template, no template DNA control. B: RPA: 37 °C, 0 minutes, 10 fg template, TwistDx positive control amplified for 10 minutes. RPA was performed, and samples ran on 3% agarose gels. Gels were imaged on a Licor Odyssey Fc. RPA using primer pair F is able to amplify 10 fg DNA to a detectable level at 37 °C in 10 minutes in all serotypes of GBS.

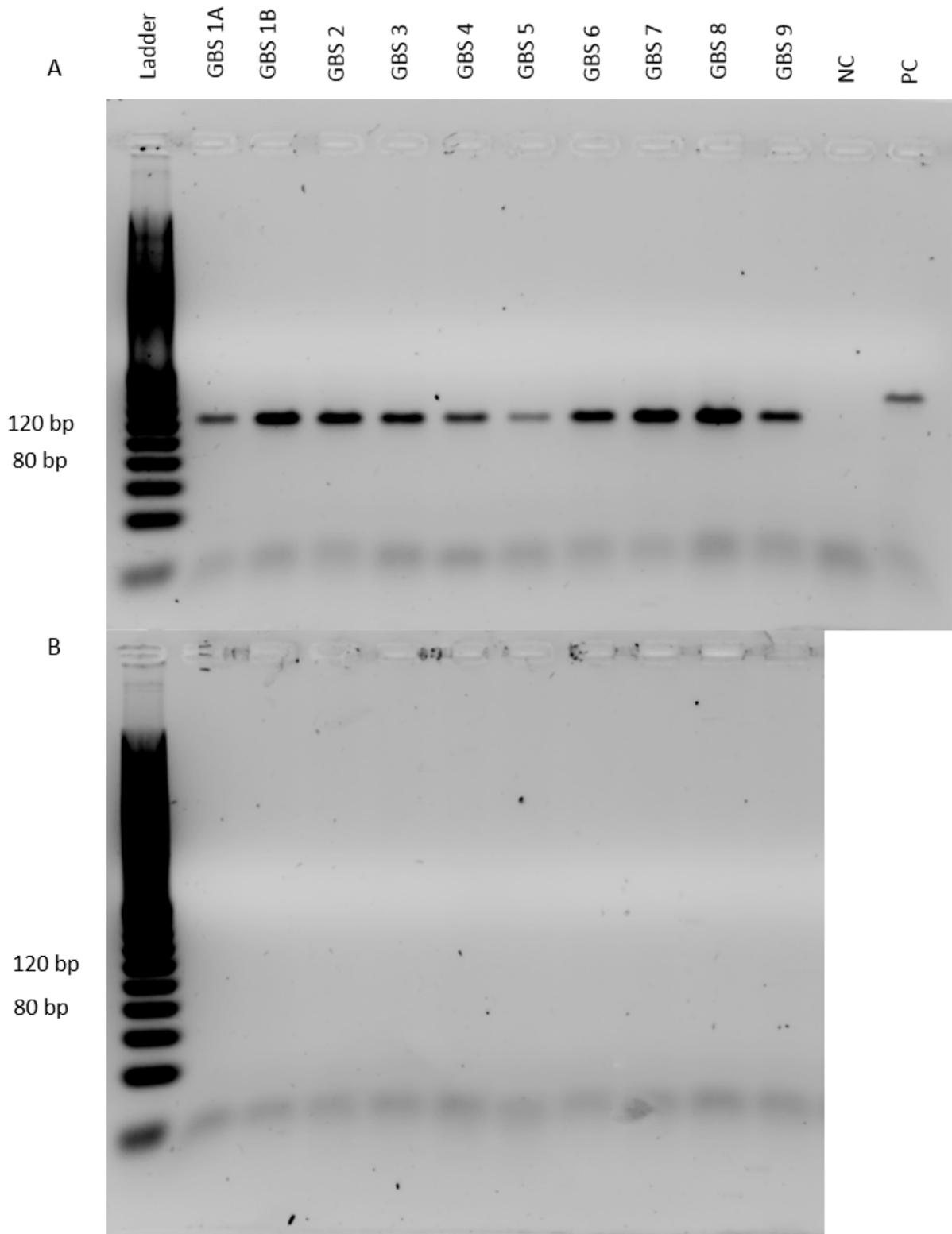


Figure 5.6 RPA of GBS DNA using primer pair F in all 10 serotypes at 25 °C  
 A: RPA: 25 °C, 20 minutes, 1 pg template, no template DNA control, TwistDx positive control. B: RPA: 37 °C, 0 minutes, 1 pg template. RPA was performed, and samples ran on 3% agarose gels. Gels were imaged on a Licor Odyssey Fc. RPA using primer pair F is able to amplify 1 pg DNA to a detectable level at 25 °C in 20 minutes in all serotypes of GBS.

### 5.2.2 Tailed-primer DNA amplification

Primer pair F was selected for modification into tailed primers. Tail sequences were based on previously published sequences. (Jauset-Rubio *et al.*, 2016a) These were added to primer pair F sequences and submitted to be computationally analysed for hairpin melting temperatures and primer dimer  $\Delta G$  values. Any interaction which crossed the overhang-primer junction was discounted, due to likely disruption from the C3 spacer. Overhang sequences were edited to produce more favourable predictions. The final primer dimer  $\Delta G$  values exceed  $-7$  kcal mol<sup>-1</sup>, and hairpin melting temperatures are  $\leq 31$  °C. These results are shown in table 5.3.

Name	Sequence	Maximum Hairpin Tm /°C	Self-dimer $\Delta G$ /kcal mol <sup>-1</sup>	Hetero-dimer $\Delta G$ /kcal mol <sup>-1</sup>
Fwd (original)	GTTTTCCCAGTCACGAC-C3- ACTACTAGACGCCAAGAAGCTATTGAAGAGG	37	-6.34	-7.18
Rev (original)	TGTA AACGACGGCCAGT-C3- TCACCACCTTTAATACTGTGCCCTAACC	45.5	-9.28	-
Fwd (final)	GTAATCACAGACATGCC-C3- ACTACTAGACGCCAAGAAGCTATTGAAGAGG	31	-6.34	-6.85
Rev (final)	TGTA AACGACGACCACT-C3- TCACCACCTTTAATACTGTGCCCTAACC	30.3	-4.85	-

**Table 5.3 Analysis of tailed primer properties.**

IDT software was used to predict the properties of tailed primers which may affect RPA. Sequences from primer pair F were combined with tails from previously published data. These were analysed for temperatures at which hairpins and self- and hetero-dimers form. Tail sequences were edited for more favourable predictions. Structures which bridge the C3 spacers were discounted as the conformation would disrupt inter/intra-molecular binding. The properties were calculated with the following conditions: 0.5  $\mu$ M oligo, 0 mM Na<sup>+</sup>, 14 mM Mg<sup>2+</sup>, 0.24 mM dNTPs at 25 °C.

RPA performance using tailed primers was analysed through observation of TMB turnover by the HRP-conjugated reporter probe. Maleimide coated plates bound capture probe via a 3' thiol group. RPA product was incubated with the plate and hybridised with the capture probe via an overhang sequence. HRP-conjugated reporter probe was introduced, and this hybridised to the RPA product via the second overhang. TMB was added, and HRP caused a colorimetric change to occur in TMB which was

detected on a plate reader at 630 nm. The ELONA readout of the RPA was conducted to validate the detection system prior to moving to lateral flow format.

RPA was executed at 37 °C for 10 and 20 minutes. Using a cut off limit of detection calculated by the mean of the negative (no DNA template) control plus three times its standard deviation, RPA was able to detect 1 pg DNA, corresponding to 380,000 copies, after 10 minutes of amplification, see figure 5.7. After 20 minutes, the amplification efficiency increased 10 fold, with the system able to detect 100 fg (38,000 copies), see figure 5.8.

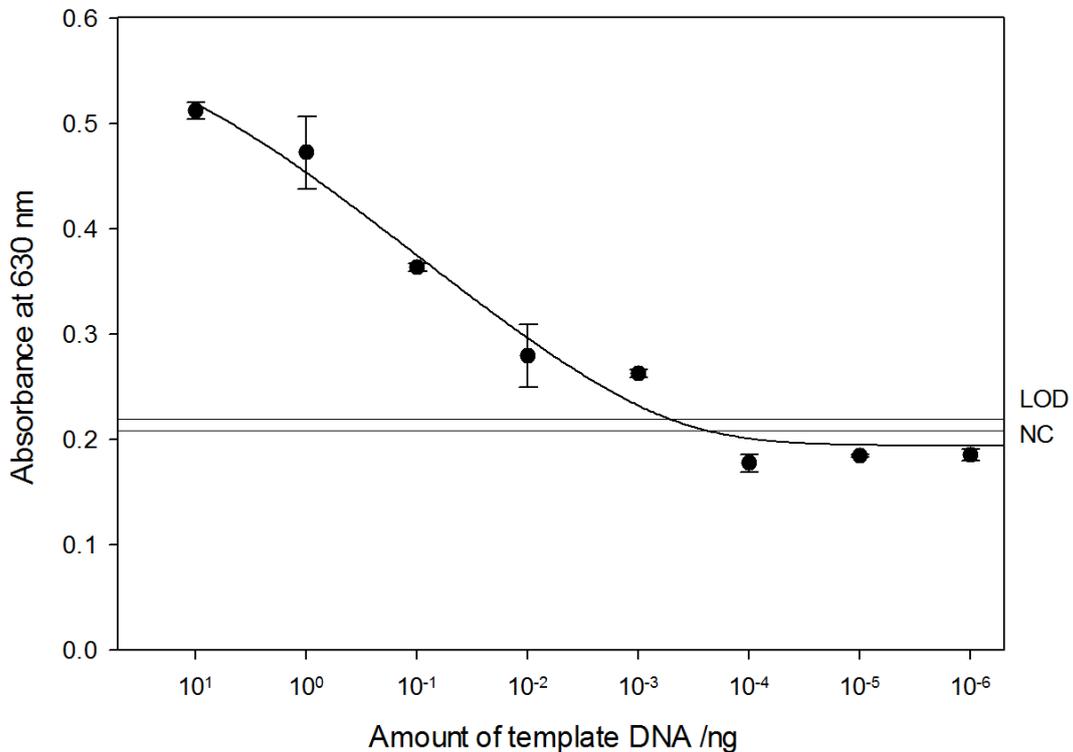
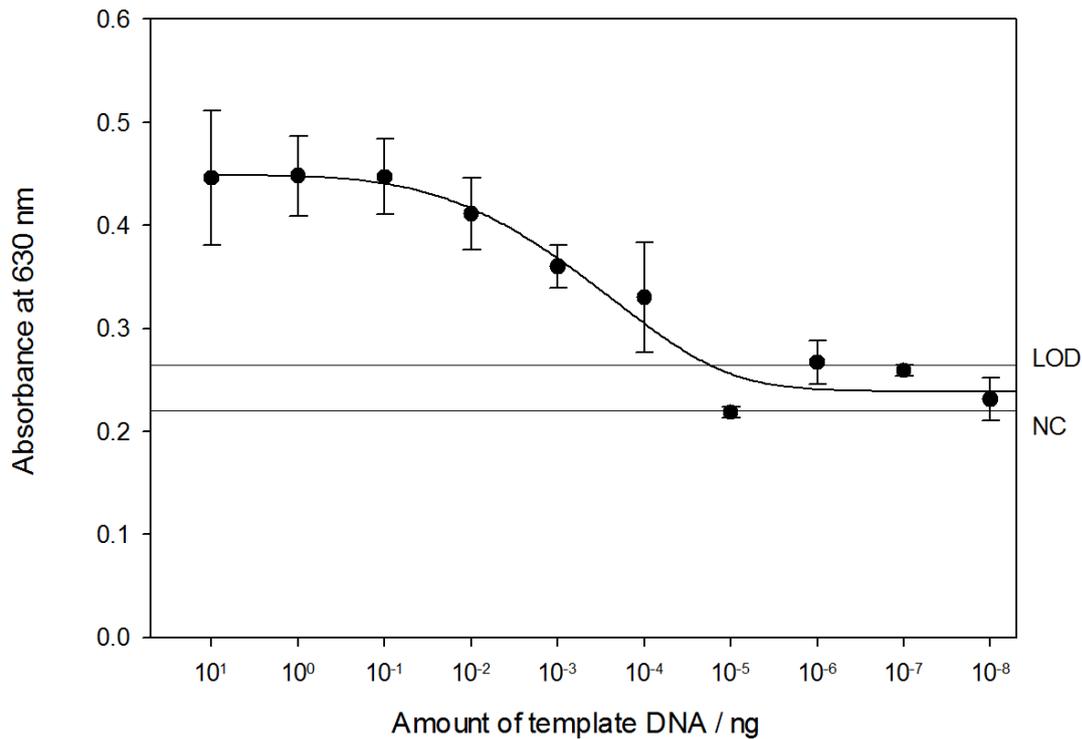


Figure 5.7 Concentration curve of RPA performed at 37 °C for 10 minutes

RPA was performed and samples analysed by TMB turnover, determined using a Tecan Infinite 200. RPA hybridisation was carried out and the colorimetric change of TMB was measured after 5 minutes of incubation. No template DNA negative controls (NC) were used. 1 pg of template DNA, or 380,000 copies, generated a readable signal above the limit of detection. The line is fitted with a sigmoidal, Weibull 4 parameter equation in SigmaPlot,  $R^2 = 0.9588$ .



**Figure 5.8 Concentration curve of RPA performed at 37 °C for 20 minutes**

RPA was performed and samples analysed by TMB turnover, determined using a Tecan Infinite 200. RPA hybridisation was carried out and the colorimetric change of TMB was measured after 5 minutes of incubation. No template DNA negative controls (NC) were used. 100 fg of template DNA, or 38,000 copies, generated a readable signal above the limit of detection. The line is fitted with a sigmoidal, Weibull 4 parameter equation in SigmaPlot,  $R^2 = 0.9598$ .

When RPA was performed on tailed primers at 25 °C for 20 minutes no relationship between template concentration and absorbance was observed, see figure 5.9. Low temperatures appear to be unsuitable for amplification of DNA using tailed primers in this instance. This may be due to interactions between primers and template DNA being sterically hindered by the overhang sequences.

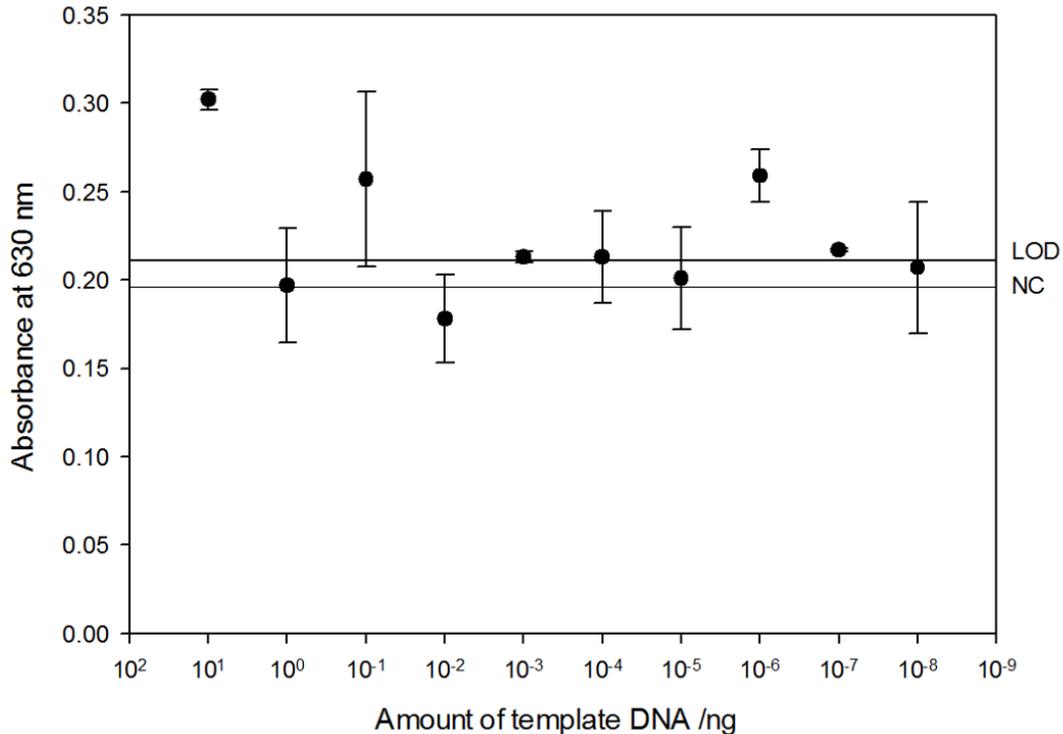


Figure 5.9 Concentration curve of RPA performed at 25 °C for 20 minutes

RPA was performed and samples analysed by TMB turnover, determined using a Tecan Infinite 200. RPA hybridisation was carried out and the colorimetric change of TMB was measured after 5 minutes of incubation. No template DNA negative controls (NC) were used. There is no clear correlation between template DNA concentration and observable signal.

### 5.2.3 Strain sensitivity and specificity

Amplification of DNA using tailed primers was performed on DNA extracted from *S. agalactiae* serotype III in the previously described experiments. To assess whether these tailed primers are suitable for a universal group B streptococcal test, DNA extracted from representative strains of all 10 serotypes of *S. agalactiae* was used as template material. DNA was also extracted from two strains of *S. pyogenes*, a related bacteria, to be used as negative controls. In previous experiments a short fragment of DNA was amplified by PCR for use as template material. In this experiment the extracted DNA was purified, but not amplified, and the total DNA content was used for recombinase polymerase amplification. This was essential as using short fragments of *S. pyogenes* would not qualify as a sufficient test of specificity as RPA could operate on a different region of *S. pyogenes* DNA. Additionally using the full DNA content introduces contaminating DNA, which may hinder the amplification of *S. agalactiae*

DNA. This is a step closer towards the sample which would be used for a diagnostic test. Figure 5.10 shows the discriminative amplification of DNA from *S. agalactiae* over *S. pyogenes* DNA. As there is no signal observed for *S. pyogenes* in excess of the no template control, it can be inferred that no amplification has taken place. The signals obtained from GBS strains (n=10) were statistically different from those obtained for GAS strains (n=2) (p=0.0413, as determined by a Mann Whitney U test).

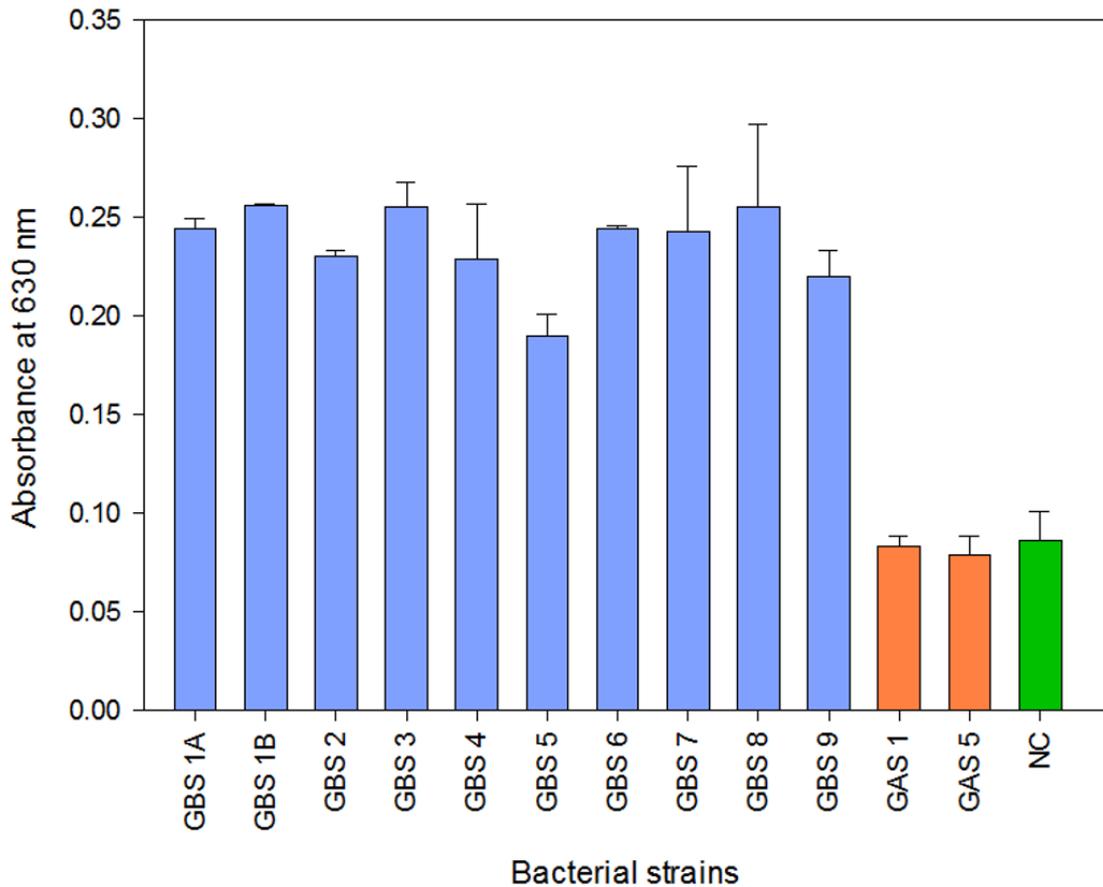


Figure 5.10 RPA of DNA extracted from bacterial cell cultures

DNA samples were extracted and purified from bacterial strains, including 10 serotypes of *S. agalactiae* and two strains of *S. pyogenes*.  $3.65 \times 10^5$  copies of DNA was used as the template material, equivalent to the copy number found in 1 pg of DNA used in the earlier experiments. RPA was performed at 37 °C for 20 minutes and samples analysed by TMB turnover, determined using a Tecan Infinite 200. RPA hybridisation was carried out and the colorimetric change of TMB was measured after 5 minutes of incubation. No template DNA negative controls (NC) were used. The signals obtained from DNA extracted from *S. pyogenes* was equivalent to that of the no DNA template control. The signal from all strains of *S. agalactiae* was around twice the values obtained for *S. pyogenes*.

The previous experiment was performed on DNA extracted from bacterial samples using a DNA purification kit. To determine whether this is necessary RPA was also performed on lysed cell samples. Under these conditions, RPA discriminately amplified DNA from *S. agalactiae* strains and not from *S. pyogenes* strains, see figure 5.11. This indicates that time consuming DNA purification steps are not required, thus increasing the suitability to rapid testing.

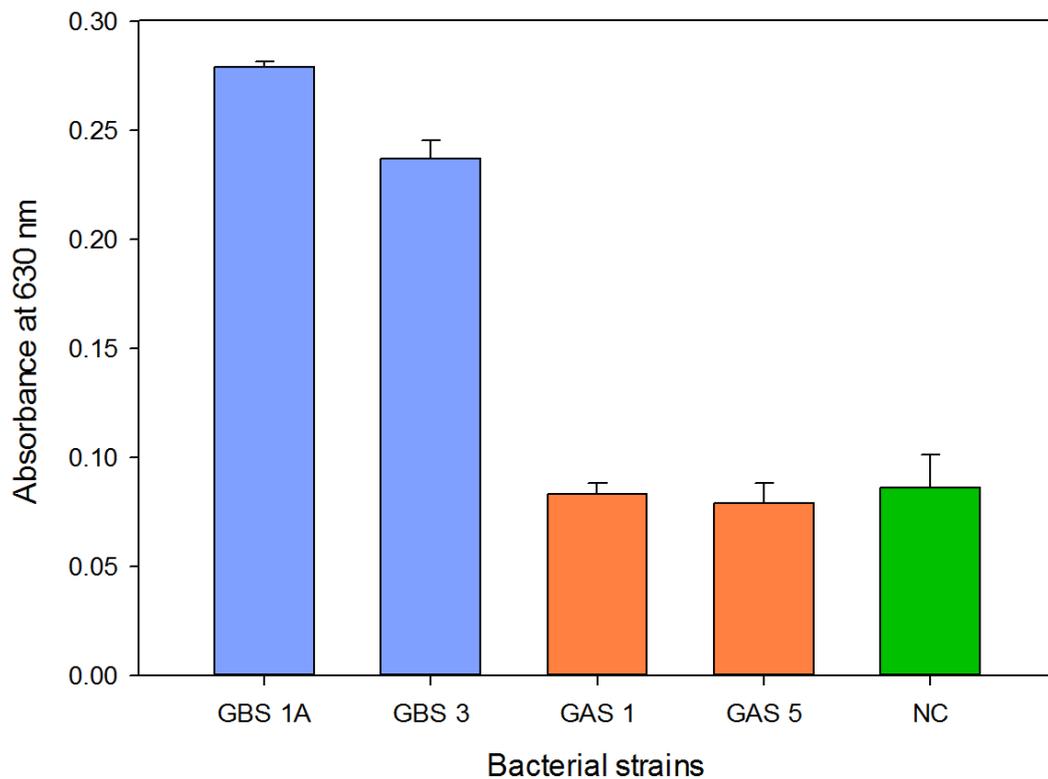


Figure 5.11 RPA directly from lysed cell cultures

The lysates of two serotypes of *S. agalactiae* and two strains of *S. pyogenes* were used as RPA template material. RPA was performed and samples analysed by TMB turnover, determined using a Tecan Infinite 200. RPA hybridisation was carried out at 37 °C for 20 minutes and the colorimetric change of TMB was measured after 5 minutes of incubation. No template DNA negative controls (NC) were used. The signals obtained from DNA extracted from *S. pyogenes* were equivalent to that of the no DNA template control. The signals from both strains of *S. agalactiae* were substantially higher than the values obtained for *S. pyogenes*.

### 5.3 Conclusions and future work

At 37 °C RPA can amplify 38,000 copies of DNA to produce a positive reading on ELONA format in 20 minutes. This rapid amplification is essential for application to point of care testing at the point of labour for prevention of maternal to infant transmission. Further to this, RPA is able to amplify sequences within an example strain from all serotypes of *S. agalactiae*, which is promising for the detection of all clinical strains. In initial specificity tests, RPA shows no cross-reactivity in *S. pyogenes*. Specificity testing will be continued to include assessment of human DNA and DNA from bacteria found in rectal and vaginal patient samples.

Future work will focus on the development of the on-plate format into lateral flow assays. By attaching a biotin label to the capture probe, the capture line can be formed on the lateral flow strip through interaction with streptavidin molecules coated onto the membrane. The HRP moiety on the reporter probe will be replaced by a gold particle which will be visible when multiple gold particles accumulate on the line. To confirm the lateral flow strip is functional, a control line consisting of the DNA sequence found in the reporter-side overhang will be added. This will bind to excess reporter probe and display a signal on the lateral flow strip. Conversion of the plate-based format into lateral flow detection has been achieved in a previous study. (Jauset-Rubio *et al.*, 2016b) The detection limit of the new system will need to be determined. To lower the limit of detection, excess forward primer-overhang may need to be mopped up prior to reaching the test line, as it may compete with amplicon binding. This could be achieved by hybridisation to a sequence complementary to the primer sequence.

Additional work focusing on reducing sample preparation needs to be carried out. Current experiments utilise an hour-long lysis protocol prior to RPA detection, attempts to reduce this have not yet been conducted. Future work will also include the testing of patient samples spiked with *S. agalactiae* bacteria, followed by infected patient samples.

---

## Chapter 6 Peptide aptamer

---

### 6.1 Introduction

In the previous work, antibodies were developed for antigen detection, however there are a number of other protein binding molecules which could be utilised, including peptide aptamers. This chapter investigates the use of a novel peptide aptamer scaffold, to be applied to the detection of *S. agalactiae*.

### 6.2 Aptamers

Peptide and oligonucleotide aptamers mimic antibody recognition; the short sequences, able to hold three-dimensional structures, show affinity to target molecules. Both, peptide and nucleic acid aptamers, exhibit some features which are favourable to antibodies and may find a role to complement antibodies. As there is no reliance on an *in vivo* immune response, aptamers can be generated against toxic or non-immunogenic targets. (Menger *et al.*, 2016; Ospina-Villa *et al.*, 2016) Highly stable aptamers (Schmidt *et al.*, 2004; Song *et al.*, 2011; Song *et al.*, 2012) and those which target molecules under non-physiological conditions (Radom *et al.*, 2013), can be selected for.

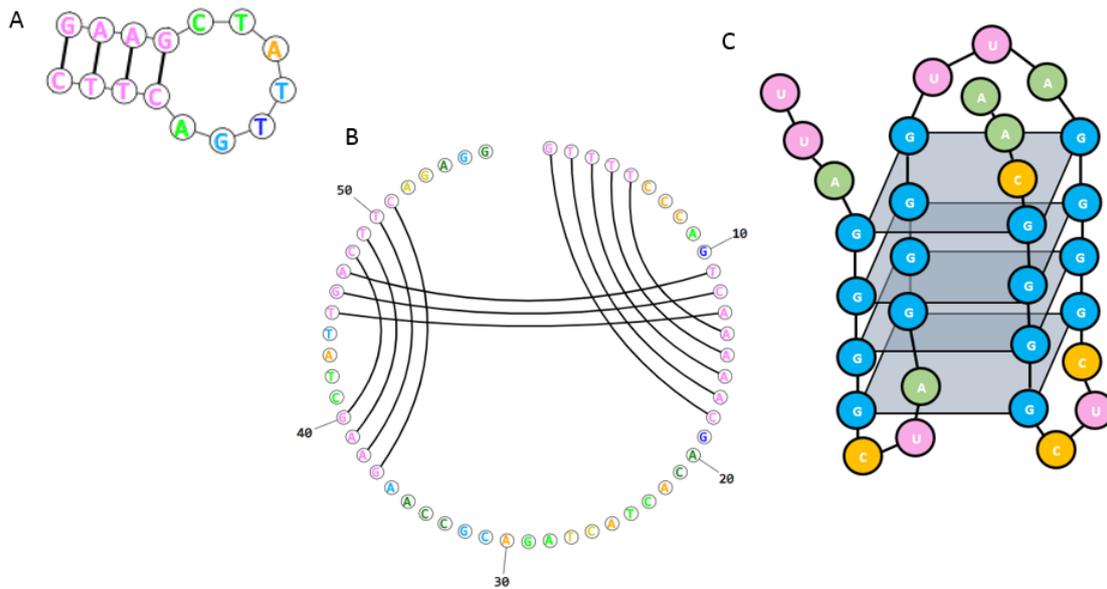
#### 6.2.1 Peptide aptamers

Peptide aptamers may be generated as free entities or as constructs incorporated into a protein scaffold. Peptide aptamer scaffolds, also known as non-immunoglobulin scaffolds or antibody mimetics, act to conformationally restrain aptamers, analogous to antibody constant regions providing a structure for the affinity-carrying CDR regions to be constrained within. The reported advantages are the presentation of structures that would not exist in the free aptamer, increased affinity for the target molecule, decreased entropy of binding (Ladner, 1995) and decreased susceptibility to cleavage.

(Stadler *et al.*, 2014) Aptamers produced for the inhibition of cyclin dependent kinase 2 activity are 1000 times more active constrained within *E. coli* thioredoxin A as opposed to as a free peptide, as indicated by half maximal inhibitory concentration data. (Cohen *et al.*, 1998) Peptide aptamers offer increased target specificity in comparison to oligonucleotides; (McKeague and DeRosa, 2012; Mercier *et al.*, 2017) there is the opportunity for any one of 20 amino acids to occupy each position in the sequence, as opposed to the four variants oligonucleotides possess. When considering that a binding site consists of several contributing monomers, there is scope for a large range of specificity.

### 6.2.2 Nucleic acid aptamers

Peptide aptamers should not be confused with nucleic acid aptamers. Single stranded oligonucleotide aptamers, DNA or RNA-based, possess regions of base pair self-hybridisation which stabilise the aptamer into a unique conformation; common structures include pseudoknots, hairpins, and G-quadruplexes, (Radom *et al.*, 2013) see figure 6.1. Nucleic acid aptamers have shown promise in the field of therapeutics, (Bunka *et al.*, 2010; Mascini *et al.*, 2012; Radom *et al.*, 2013) being particularly suited to *in vivo* roles as they are unaffected by proteases found in the body. Degradation though nuclease activity is a significant problem, however nuclease resistivity can be established through the addition of antisense molecules or nucleotide modification. (Kusser, 2000) Locked nucleic acids have shown to have an increased half-life in the murine blood system. (Schmidt *et al.*, 2004) Oligonucleotide aptamers exhibit high thermostability and are able to refold into their original structure following exposure to high temperatures; a significant advantage over proteins that often undergo irreversible denaturation. (Song *et al.*, 2012) This would be beneficial in ensuring stability in diagnostic applications.



**Figure 6.1 Nucleic acid structures**

A: Hairpin structure. Visualised by RNAstructure, using results from Fold. (Reuter and Mathews, 2010) B: Pseudoknot structure. Visualised by RNAstructure, using results from ProbKnot. (Reuter and Mathews, 2010) C: G-quadruplex structure. Sequence predicted to hold a G-quadruplex structure by GRS mapper. (Kikin *et al.*, 2006)

### 6.2.3 Designing a scaffold

#### 6.2.3.1 Peptide aptamer constraint

When designing a scaffold, the site of aptamer insertion must be selected. Doubly constrained peptides, bound at both termini, are termed peptide aptamers, differing from singly constrained peptides, (Mascini *et al.*, 2012) which are thought of as affinity tags. Aptamers are often inserted into scaffolds at sites utilised for binding activity. (Hoppe-Seyler *et al.*, 2004; Woodman *et al.*, 2005) In the native protein, binding sites are accessible to the solvent to allow ligand binding. Insertion of peptides into these regions ensures accessibility and the opportunity for potential ligands to bind.

Aptamers may be inserted into a single loop, a region of secondary structure, or into multiple sites. Thioredoxin A is a scaffold protein to which aptamers are inserted in a single loop. This loop is highly constrained by disulphide bonds, enabling a high tolerance of peptide insertions, as they impose minimal impact on global protein structure. (Nygren and Skerra, 2004) Affibodies are derived from Staphylococcal protein A, which binds to the Fc portion of antibodies. Residues found to be involved

in this interaction, located across 2 helices, were selected for randomisation and affibodies showing affinity for desired targets were selected. Circular dichroism studies showed that protein structure was not disrupted, demonstrating the high stability of affibodies. (Nord *et al.*, 1995) Anticalins are scaffolds derived from lipocalins. Four loops comprise the entrance to a  $\beta$ -barrel structure, aptamer sequences have been inserted into all loops causing significant changes in conformation in the flexible loop regions, but not disturbing the rigid secondary structure. (Nygren and Skerra, 2004) Affibodies and anticalin- and thioredoxin A- based scaffolds each follow a different strategy in selection of aptamer insertion sites, peptides are inserted into regions of different secondary structure and as continuous peptides or fragmented across multiple locations. In aptamer design an individual approach is likely to be required, based on the scaffold sequence and structure.

#### 6.2.3.2 Functional or non-functional scaffolds

If scaffold structural and functional data is available, the insertion of aptamers may be purposefully designed to either retain or disrupt protein function. Typically scaffolds are selected for their non-functionality, and are designed to perform inhibitory roles in therapeutics (Nagel-Wolfrum *et al.*, 2004) or solely as binding agents in diagnostic applications. (Evans *et al.*, 2008; Davis *et al.*, 2009; Ravalli *et al.*, 2015; Ferrigno, 2016)

In previous studies scaffold domains were fused with Fc antibody domains, effectively adding an effector function to the neutral scaffold. (Ronmark *et al.*, 2002) This product has implications for therapeutics in host-mediated responses, and also indicates the potential for a wide range of functions to be incorporated into recognition molecules. (Liu *et al.*, 2017) Functional aptamers can be created through tagging with nuclear localisation sequences. (Colas *et al.*, 2000) By transporting mislocated proteins to the correct compartment, proteins can become reactivated. Additionally this strategy could be utilised to translocate proteins to foreign localisations rendering them functionally inert. (Hoppe-Seyler *et al.*, 2004) Ubiquitin ligases have been conjugated to peptide aptamers, thus the aptamer's target becomes ubiquitinated. (Colas *et al.*, 2000) This may be used to initiate protein degradation. (Hoppe-Seyler *et al.*, 2004)

There are some examples of scaffolds that retain their native activity and utilise this for the visualisation of signals from molecule detection. GFP (green fluorescent protein) and  $\beta$ -lactamase, often used as reporter molecules for gene expression, have been modified into peptide aptamer scaffolds. (Abedi *et al.*, 1998; Legendre *et al.*, 2002; Rojas *et al.*, 2004; Wang *et al.*, 2014a)

#### 6.2.4 Aptamer scaffolds as diagnostic tools

##### 6.2.4.1 Surface adsorption

Protein adsorption to a surface is often associated with conformational changes. (Nakanishi *et al.*, 2001) Hydrophobic residues re-orientate themselves to maximise contact with a hydrophobic surface, reducing exposure to the hydrophilic solvent. (Hladilkova *et al.*, 2016) Hydrophilic surface adhesion also causes conformational changes; a previous study has found  $\alpha$ -helical content to reduce upon adsorption of  $\alpha$ -chymotrypsin to a hydrophilic surface. (Zoungrana *et al.*, 1997) Changes in protein conformation may inhibit protein function. Stefin A is a cysteine protease inhibitor that exhibits high stability upon surface immobilisation. It was modified into an aptamer scaffold able to retain affinity for CDK2 (cyclin dependent kinase) upon adsorption to glass surfaces. Stefin A shows promise for use in microarrays to assess multiple biomarkers simultaneously. (Song *et al.*, 2011)

##### 6.2.4.2 Specificity

Aptamers have the capability to exhibit high binding partner specificity, essential for correct disease diagnosis. An anticalin molecule modified to show affinity for digoxigenin was found to possess no reactivity against ouabain which is functionally related to digoxigenin. (Schlehuber and Skerra, 2002) Ras is a key molecule in cell signalling pathways. There is a two amino acid difference in the sequence of two allelic variants, RasV<sub>12</sub> and RasA<sub>15</sub>. Peptide aptamers have been developed which are able to differentially bind each variant. (Xu and Luo, 2002)

#### 6.2.4.3 Compatible with label-free detection techniques

For ease of implementation, an ideal aptamer would be compatible with the range of detection techniques currently used in diagnostic laboratories. Observation of antibody binding is often achieved through labelling, typically with fluorescence, radioactivity or enzymatic activity. (Hempen and Karst, 2006) Labelling imposes the risk of altering the conformation of the antibody, potentially reducing or eliminating its affinity for the desired epitope. An example of label-free detection systems is the use of reporter proteins as scaffolds. GFP has been investigated as an aptamer scaffold. Complementarity determining regions from antibodies recognising several toxins were inserted into the scaffold, each retained its affinity for its specific epitope. (Wang *et al.*, 2014a) Similarly, the use of  $\beta$ -lactamase as an aptamer scaffold has been investigated. A signal is obtained through studying the rate of breakdown of penicillin G and nitrocefin spectroscopically. Upon epitope binding, hydrolytic activity of the scaffold increases in respect to penicillin G, and decreases with respect to nitrocefin. (Legendre *et al.*, 2002)

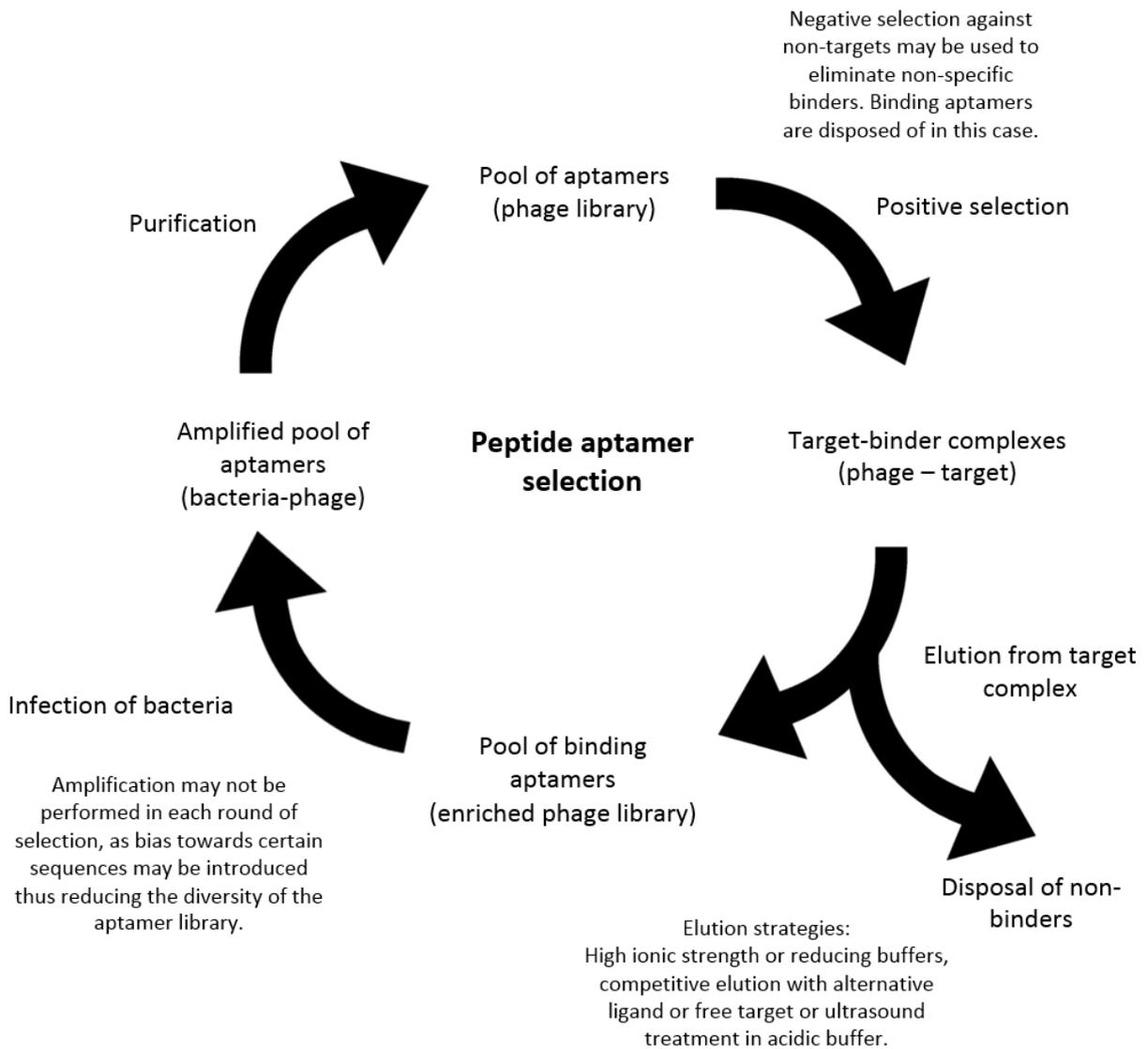
#### 6.2.4.4 The need for additional antigen detection molecules in diagnostics

It is predicted that with the rise of personalised medicine the demand for aptamers will increase. The detection of biomarkers to aid prediction of a patient's response to available treatments in addition to detection of disease-specific biomarkers will require the development of many novel detection molecules. Further to the previously discussed advantages aptamers offer over antibodies, such as suitability for use on surfaces and in label-free detection systems, low production costs are critical; aptamers provide this through high efficiency production and lack of animal involvement. (Song *et al.*, 2011) In many diagnostic techniques two recognition molecules are required, one as a capture molecule and the other as a means of detection. Antibodies or aptamers could be used in either, or both of these roles.

#### 6.2.5 Production and screening of a peptide aptamer library

There are several methods of achieving binding affinity. Aptamers can be created by the direct insertion of an amino acid sequence from one binding molecule to another. Complementarity determining regions of antibodies introduced to a superfolded GFP

variant provided the scaffold with a similar binding affinity as the native antibody. (Wang *et al.*, 2014a) As there is no requirement for the generation of an aptamer library, molecular recognition tools are produced rapidly. This method of generation, however, does not enable novel epitopes to be targeted. Structures computationally predicted to bind the desired epitope may be designed. A research group produced proteins that exhibit 130 nM affinity for their binding partners, using directed evolution the affinity was increased to 180 pM. (Karanicolas *et al.*, 2011) Site-directed mutagenesis can be used to rationally mutate binding sites; prior knowledge of the interaction is required. Alternatively, representative residues of several amino acid types (e.g. polar, negatively charged, and positively charged) may be inserted to each position of a sequence for an efficient method to scan for binding affinity. (Banta *et al.*, 2013) Often a broad spectrum of aptamers is created with no sequence bias. Libraries of peptide aptamer scaffolds only contain variations in a small number of residues, but can comprise of  $10^{15}$  molecules. (Banta *et al.*, 2013) An aptamer should be large enough to include all amino acids that participate in binding the epitope, so the optimal arrangement can be selected. Aptamer length is limited by the likelihood that the insertion causes global protein misfolding of the scaffold and the feasibility of generating a library with sufficient coverage. (Nygren and Skerra, 2004) Affinity maturation is used to select for, refine and optimise binding affinity. (Banta *et al.*, 2013) This process is similar to SELEX, (systematic evolution of ligands by exponential enrichment) which is employed for nucleic acid aptamers. (Ellington and Szostak, 1990; Tuerk and Gold, 1990) Positive discrimination of aptamers exhibiting target binding is used to select candidate aptamers. Multiple rounds of the process are carried out under differing conditions, enabling the production of a high affinity aptamer able to operate in a wide range of environments and functions, (Radom *et al.*, 2013) see figure 6.2. Peptide aptamer library selection requires a technique that links the phenotype to its genetic material, phage display is often used. (Nygren and Skerra, 2004; Banta *et al.*, 2013)



**Figure 6.2 Selection of peptide aptamers exhibiting target affinity**

Multiple rounds of peptide selection are performed under differing conditions to ensure aptamers are selected that show robust specificity for the desired target. Negative selection reduces non-specificity. Information from (Aghebati-Maleki *et al.*, 2016)

In the context of designing an aptamer for use in GBS diagnostics, the complementary determining region of antibodies produced in the previous work could be inserted into a protein scaffold, or a novel peptide aptamer sequence could be selected for.

## 6.3 Design of a reporter-based peptide aptamer

### 6.3.1 Selection of reporter scaffold

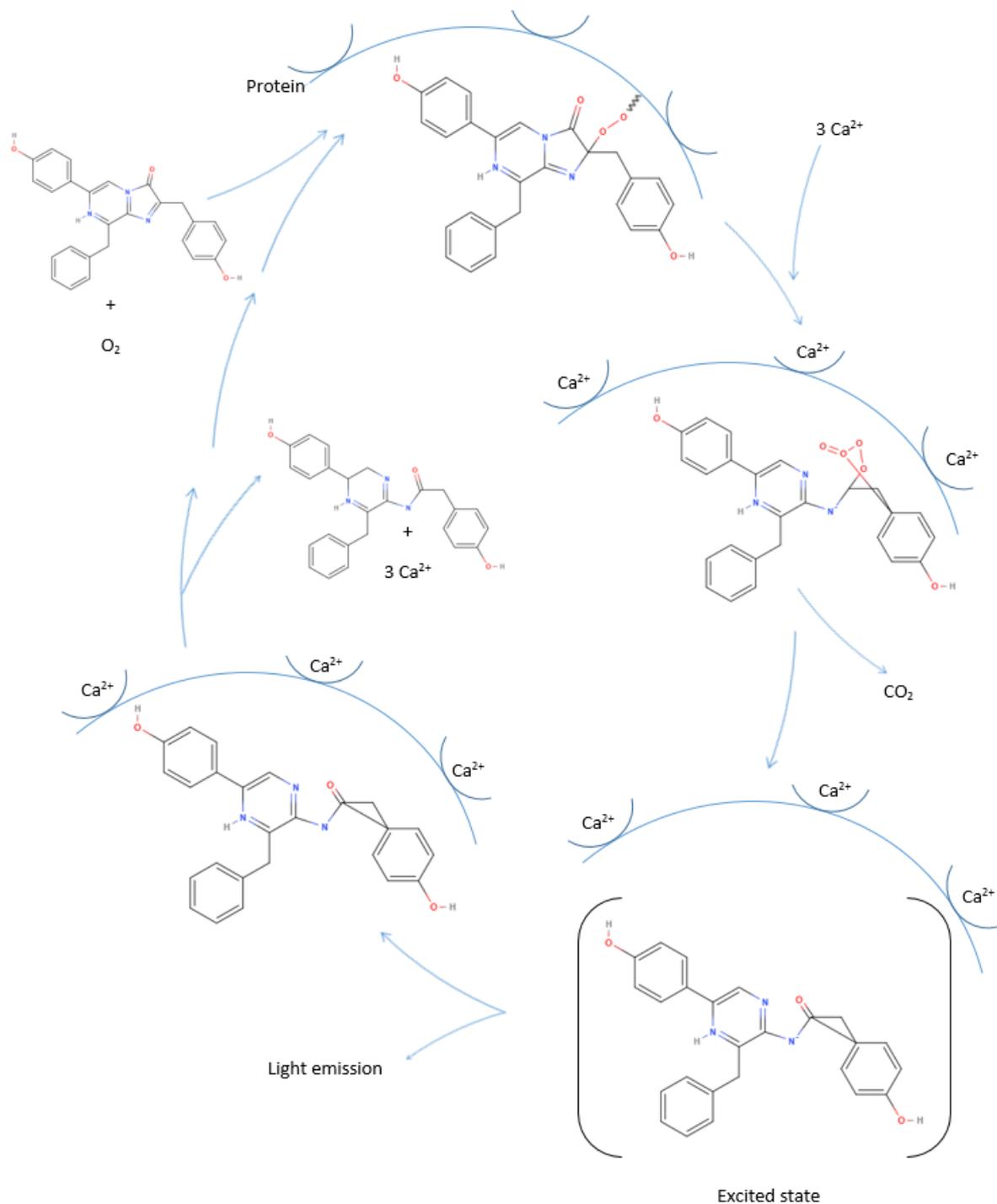
Several proteins were considered for the creation of a novel reporter scaffold including  $\beta$ -galactosidase proteins, HRP, alkaline phosphatase and luciferase proteins.  $\beta$ -galactosidases are enzymes able to hydrolyse the colorimetric substrate X-Gal resulting in a reporter signal. *Penicillium*  $\beta$ -galactosidase is a large 120 kDa protein consisting of five domains, which, as a whole molecule, is too large for aptamer creation due to stability concerns. It is possible that the use of one domain as a scaffold may be achievable, however the active site is formed from regions within two domains (Rojas *et al.*, 2004) and therefore a single domain would be non-functional. Alkaline phosphatase and HRP are two commonly utilised reporter proteins. However due to their large size, possession of disulphide bonds, (Coleman, 1992) and HRP being glycosylated, (Veitch, 2004) these molecules were not pursued as scaffolds, as simple bacterial production could not be exploited. Bioluminescent proteins were also considered. As background signal levels are zero, bioluminescence presents an advantage over fluorescent peptide aptamer scaffolds, such as GFP. Luciferases are bioluminescent proteins containing no post-translational modifications, however they do contain a high ratio of cysteine residues and therefore typically require low efficiency periplasmic expression to be produced. (Loening *et al.*, 2006)

### 6.3.2 Aequorin

Aequorin, another bioluminescent protein, is relatively small compared to other reporter enzymes. According to previous literature, aequorin can be produced in the cytoplasm of *E. coli* (Courchesne and Ozturk, 2003) and therefore it was selected to be assessed for aptamer creation. One limitation may be that it has a slow regeneration process so does not produce signal amplification like most enzyme systems. As aequorin has rapid activity, with signal production lasting for less than 10 seconds (Lewis *et al.*, 2000), molecular detection and therefore diagnoses may be achieved within a very short time scale. Due to extremely low background signals, concentrations of 100 pM of aequorin have been utilised in *in vitro* assay development to produce high signal to noise ratios. (Deo *et al.*, 2001)

Aequorin is a 21.7 kDa calcium-activated photoprotein that, in association with coelenterazine and oxygen, undergoes a conformational change causing light emission at a wavelength of 470 nm. Transfection into bacterial and mammalian cells has assisted the study of calcium cell signalling in living cells. (Jones *et al.*, 2002; Bonora *et al.*, 2013) Aequorin exhibits reasonable heat stability when in the presence of  $\text{Ca}^{2+}$ , following incubation at 45 °C no structural changes are observed. (Inouye, 2004)

Aequorin is formed from four EF-hand domains, each consisting of a helix-loop-helix structure. Three of these domains contain  $\text{Ca}^{2+}$  binding sites; two binding events are required for full activity. Upon calcium binding EF-hand helices re-orientate themselves with respect to each other. (Head *et al.*, 2000) Coelenterazine is oxidised and becomes unstable; carbon dioxide is released and an excited coelenteramide anion is formed. Light is released when this species returns to its ground state, (Jones *et al.*, 1999) for a schematic of the reaction see figure 6.3.



**Figure 6.3 Mechanism of aequorin light emission and regeneration**

$\text{Ca}^{2+}$  binding triggers a conformational change within aequorin which causes coelenterazine to become oxidised. Carbon dioxide is released and the unstable coelenteramide ion is formed. Light is released when the cofactor relaxes to its ground state. The aequorin molecule is regenerated to its original form to enable the process to repeat. The regeneration process is much slower, taking several hours (Shimomura *et al.*, 1993) in comparison to calcium-induced light emission which takes around 10 seconds. (Lewis *et al.*, 2000) Details on mechanism from (Jones *et al.*, 1999), produced using (MolView, 2015).

## 6.4 Sequence design of peptide aptamer scaffold

### 6.4.1 Site of aptamer insertion

A short amino acid sequence, a FLAG tag (DYKDDDK), was inserted into the aequorin structure. The FLAG tag was designed in the 1980s to aid in protein identification and purification through antibody affinity. (Hopp *et al.*, 1988) Insertion of the FLAG tag mimics a peptide aptamer, to assess whether molecular recognition could occur whilst the molecule retained functional activity. Binding of the protein to an anti-FLAG tag antibody would determine that the tag is exposed, essential to enable aptamer-epitope binding events to occur in future variants of the scaffold.

Regions were analysed using structural and functional information to decide which would be most suitable for aptamer insertion, see table 6.1. To minimise the likelihood of structure and function disruption, the sequence was inserted into a loop rather than a sequence of secondary structure. Previous studies have mutated glycine residues involved in calcium binding to determine which sites are important for light emission. Loops within EF-hands 1 and 3 were found to be of significance, glycine mutation within EF-hand 4 did not affect bioluminescence. (Tsuji *et al.*, 1986) Coelenterazine binding is co-ordinated by three tyrosine-histidine-tryptophan triads, which line the cofactor binding cavity. When several of these residues are mutated, protein activity is greatly reduced. (Head *et al.*, 2000) As retaining bioluminescent function is essential for the intended reporting application, regions involved in cofactor and metal binding were avoided.

Region	Structure type	Domain	Considerations for peptide insertion
A	Loop	N-terminus	Aptamer would not be doubly constrained.
B	Helix	EF-hand 1	His16 is likely to interact with coelenterazine; light emission is dependent on the helix moving relative to the C-terminal loop, through disruption of the hydrogen bonding network; domain is involved in Ca <sup>2+</sup> binding.
C	Loop	EF-hand 1	Mutation of central glycine, which prevents Ca <sup>2+</sup> binding, eliminates bioluminescence.*
D	Helix	EF-hand 1	Domain is involved in Ca <sup>2+</sup> binding.
E	Loop	Loop	Previous studies have inserted sequences here.**
F	Helix	EF-hand 2	Helix forms coelenterazine binding cavity; His58 is involved in coelenterazine binding.
G	Loop	EF-hand 2	None.
H	Helix	EF-hand 2	Helix forms coelenterazine binding cavity; Tyr82 and Trp86 are involved in coelenterazine binding.
I	Loop	Loop	Loop brings regions H and J together, forming the coelenterazine binding cavity.
J	Helix	EF-hand 3	Trp108 mutants result in reduced bioluminescence, likely due to loss of interaction with coelenterazine; domain involved in Ca <sup>2+</sup> binding.
K	Loop	EF-hand 3	Mutation of central glycine, which prevents Ca <sup>2+</sup> binding, reduces bioluminescence.*
L	Helix	EF-hand 3	Tyr132 is predicted to interact with coelenterazine and is involved in hydrogen bonding network; domain involved in Ca <sup>2+</sup> binding.
M	Loop	Loop	Close proximity to C-terminus, conformational change of the C-terminus is essential for light emission.
N	Helix	EF-hand 4	Light emission is dependent on the helix moving relative to the C-terminal loop, through disruption of the hydrogen bonding network; domain involved in non-essential Ca <sup>2+</sup> binding.
O	Loop	EF-hand 4	Contains calcium binding glycine residue, calcium binding at this site is not essential for bioluminescence.*
P	Helix	EF-hand 4	Light emission is dependent on the helix moving relative to the C-terminal loop through disruption of the hydrogen bonding network; proposed His169:coelenterazine hydrogen bond; domain involved in non-essential Ca <sup>2+</sup> binding.
Q	Loop	C-terminus	Aptamer would not be doubly constrained; proposed Tyr184:coelenterazine hydrogen bond; C-terminal proline is reported to be essential for function***

**Table 6.1 Analysis of sites in aequorin in which to insert a peptide aptamer.**

The peptide sequence of aequorin has been divided into regions based on the structure determined by X-ray crystallography produced by (Head *et al.*, 2000), PDB code 1EJ3. The likelihood of disruption to protein structure and function following peptide insertion into each of these regions was analysed. Sites which are predicted to be involved with, or in close proximity to, essential binding activity were avoided. Information from: \*(Tsuji *et al.*, 1986); \*\* (Hamorsky *et al.*, 2008); \*\*\* (Ellington and Szostak, 1990; Nomura *et al.*, 1991). Unless specified, data was obtained from (Head *et al.*, 2000).

Regions E and G present no structural evidence to indicate that they would be unsuitable for the insertion of an amino acid sequence. BLAST searches were performed to assess sequence conservation of these regions between species (Altschul *et al.*, 1990); if activity is retained whilst sequences are not conserved, this may indicate that sequences are not essential and may be suitable for aptamer insertion. However, both regions are highly conserved within aequorin molecules, and region E remains highly conserved within similar proteins obelin, clytin and mitrocomin.

Previous studies have shown region E tolerates the insertion of peptides and maintains function; therefore this region was selected, see figure 6.4 for a model of the aequorin structure and site of sequence insertion. Glucose binding protein, a 32 kDa protein, has been inserted into aequorin at region E, splitting aequorin into two fragments to create a glucose sensor. The binding of glucose initiates a conformational change reforming the active site of aequorin, and enabling the bioluminescent reaction to occur. (Hamorsky *et al.*, 2008) It was predicted that if this site can withstand this degree of perturbation, a wide variety of peptides could be inserted with bioluminescent function preserved.

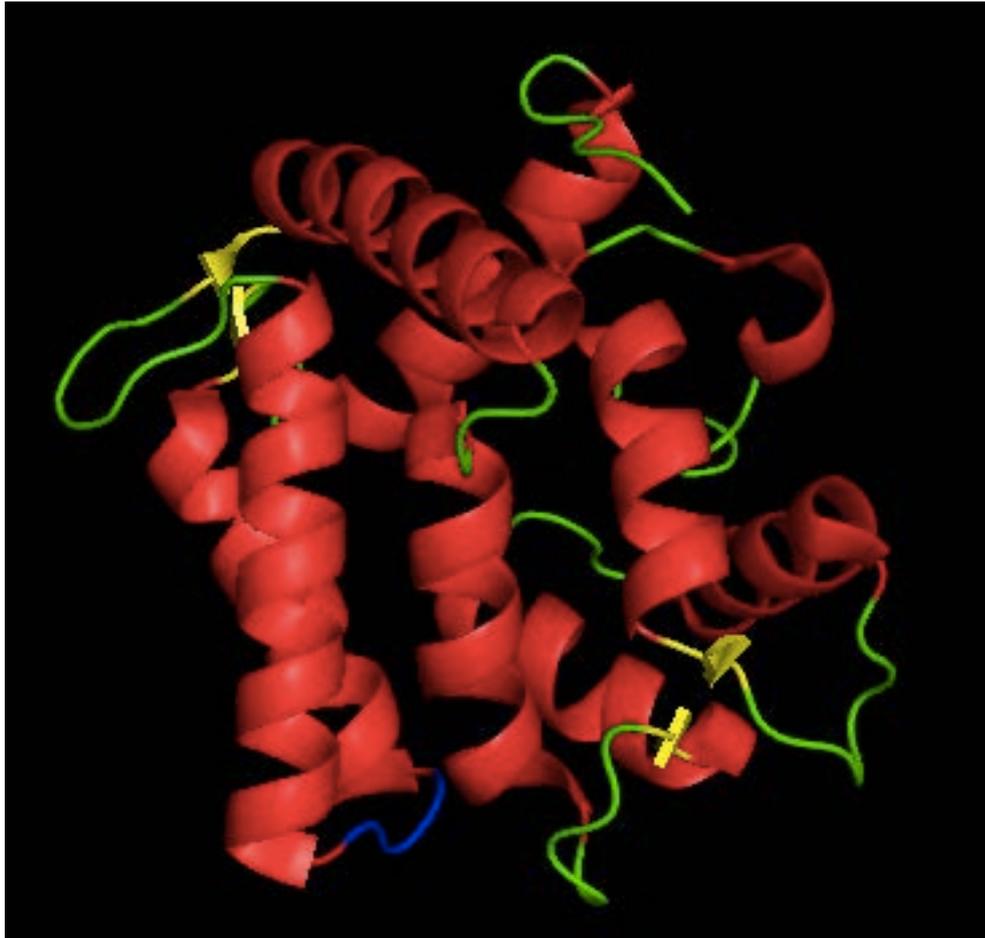


Figure 6.4 Structure of aequorin

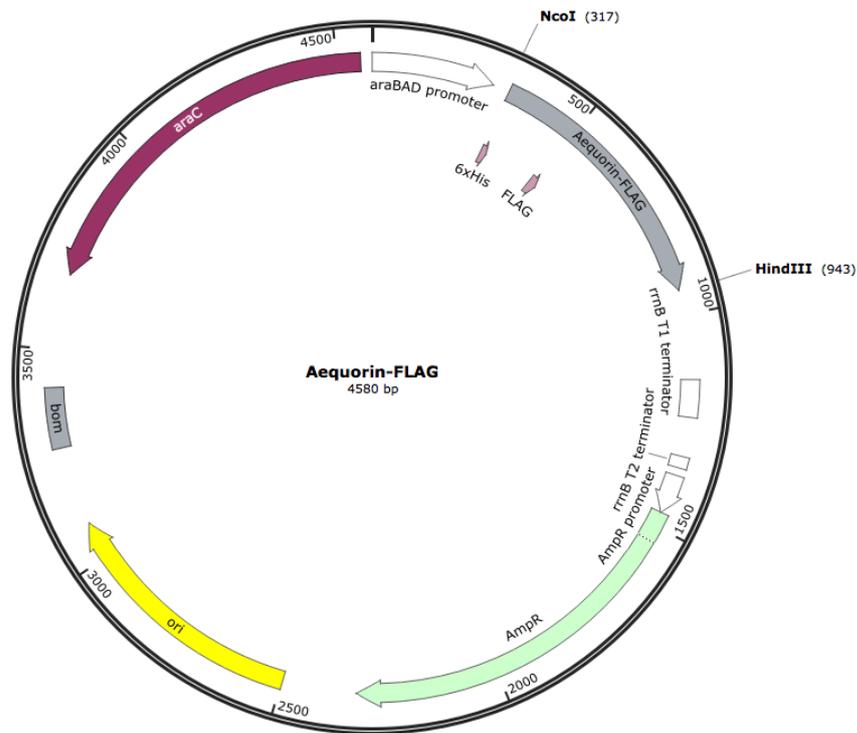
The three-dimensional structure of aequorin was obtained by X-ray crystallography, the structure is coloured by secondary structure (alpha helices red,  $\beta$ -sheets yellow, and unstructured loops green. The loop into which the FLAG tag was inserted into is shown in blue. PDB code 1EJ3 (Head *et al.*, 2000) rendered in PyMOL. (Schrodinger, 2015)

#### 6.4.2 Sequence optimisation

All three cysteine residues were mutated to serine as this variant has shown increased bioluminescent activity in previous studies. ((Kurose *et al.*, 1989; Lewis *et al.*, 2000) Cysteine removal has also been found to improve protein stability as dimerisation is prevented. (Shrestha *et al.*, 2002) Aequorin requires regeneration by replacing coelenteramide with coelenterazine, to enable future calcium-induced light emission. Cysteine bonds are predicted to be involved with this process as the regeneration time of cysteine mutants is much longer than that of the wild-type protein. (Kurose *et al.*, 1989) As regeneration of aequorin is a slow process, with 50% of aequorin re-

established in two hours, (Shimomura *et al.*, 1993) it is not consequential in rapid diagnostics. The lack of disulphide bonds enable production in the cytoplasm, therefore no signal sequence was designed into the construct. A polyhistidine tag was added to the protein to enable purification via nickel affinity. It has been reported that a C-terminal proline residue is essential for bioluminescence. (Nomura *et al.*, 1991) A more recent study has found that sequence insertion to the C-terminus does not disrupt activity, (Deo *et al.*, 2001) however due to conflicting data the C-terminus was avoided and an N-terminal histidine tag was inserted.

Two DNA sequences were designed, one pseudo wild-type variant (Aequorin WT\*) and one variant of aequorin containing the FLAG-tag sequence insertion (Aequorin FLAG). Codon usage was optimised for synthesis in *E. coli* using GeneOptimizer® software (GeneArt, Life Technologies), to maximise chances of high efficiency production. Figure 6.5 shows the construct plasmid map, and DNA and protein sequences.



B

NcoI restriction site

```

CCATGGGCGATCATCATCACCATCATAGCAGCGTTAAACTGACCAGCGATTTTGATAATCC
GCGTTGGATTGGTCGCATAAACACATGTTAACTTTCTGGATGTGAACCACAACGGCAAA
ATTAGCCTGGATGAAATGGTTTATAAAGCCAGCGATATCGTGATCAACAACGATTATAAA
GACGACGATGATAAACTGGGTGCAACACCGGAACAGGCCAAAACGTCATAAAGATGCAGT
TGAAGCATTTTTTGGTGGTGCCGGTATGAAATATGGTGTGAAACCGATTGGCCTGCCTAT
ATTGAAGTTGGAAAAAACTGGCAACCGACGAGCTGAAAAAATGCAAAAAATGAACC
GACCCTGATTCGCATTTGGGGTGATGCACTGTTTGATATCGTTGATAAAGATCAGAAATGGT
GCCATTACACTGGATGAATGGAAAGCATATACCAAAGCAGCAGGTATTATTCAGAGCAGC
GAAGATAGCGAAGAAACCTTTCTGTGTTAGCGATATTGATGAAAGCGGTCAGCTGGATGTT
GATGAAATGACCCGTCAGCATCTGGGTTTTTGGTATACAATGGATCCGGCAAGCGAAAAA
CTGTATGGTGGTGCAGTTCGTAAGCTT
  
```

C

HindIII restriction site

```

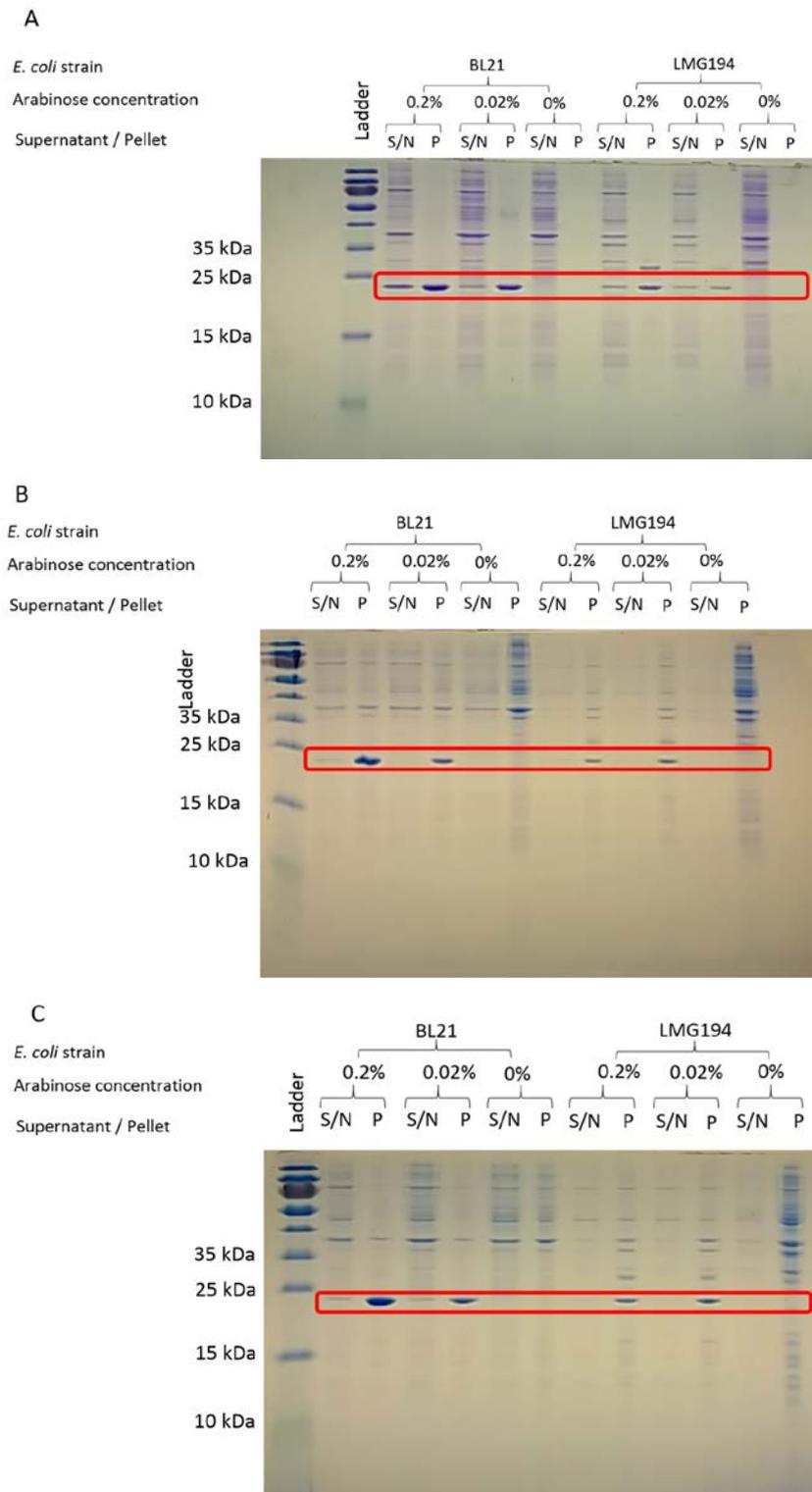
M G H H H H H S S V K L T S D F D N P R W I G R H K H M F N F L D V N H N G K I S L
D E M V Y K A S D I V I N N D Y K D D D K L G A T P E Q A K R H K D A V E A F F G G
A G M K Y G V E T D W P A Y I E G W K K L A T D E L E K Y A K N E P T L I R I W G D A L
F D I V D K D Q N G A I T L D E W K A Y T K A A G I I Q S S E D S E E T F R V S D I D E S
G Q L D V D E M T R Q H L G F W Y T M D P A S E K L Y G G A V P Stop
  
```

Figure 6.5 Recombinant aequorin (WT\* and FLAG) construct DNA and protein sequence. A: Plasmid map of the gene construct, produced in SnapGene Viewer. (GLS Biotech) NcoI and HindIII restriction enzymes were used to insert the construct DNA into *pBAD HisA*. B: Construct DNA sequence encoding for the two variants of recombinant aequorin. The pseudowildtype construct does not include the nucleic acids encoding for the FLAG tag (red). Restriction sites are shown in green. Codons were optimised for production in *E. coli*. C: Protein sequence for the two variants of recombinant aequorin. An N-terminal polyhistidine tag has been added (orange), and three cysteine residues have been mutated to serine residues (blue). The pseudowildtype variant does not include the FLAG tag (red).

## 6.5 Induction of aequorin WT\*

The plasmid constructs were inserted into the vector *pBADHisA* through use of restriction enzymes NcoI and HindIII. *pBADHisA*, containing an arabinose inducible operon, was selected as it has previously been used in successful aequorin production. (Rowe *et al.*, 2010) *pBADHisA-Aeq-WT\** and *pBADHisA-Aeq-FLAG* were transformed into *E. coli* BL21 *pLysS DE3* and *E. coli* LMG194 cells. BL21 is an expression strain often used for induction of the T7 promoter, it is not typically selected for arabinose-based induction as it does not contain an *araBAD* deletion. The expression strain LMG194 lacks the *araBAD* operon, and is therefore unable to breakdown arabinose. This plasmid enables tight control of protein expression via arabinose. Protein expression in both strains was analysed.

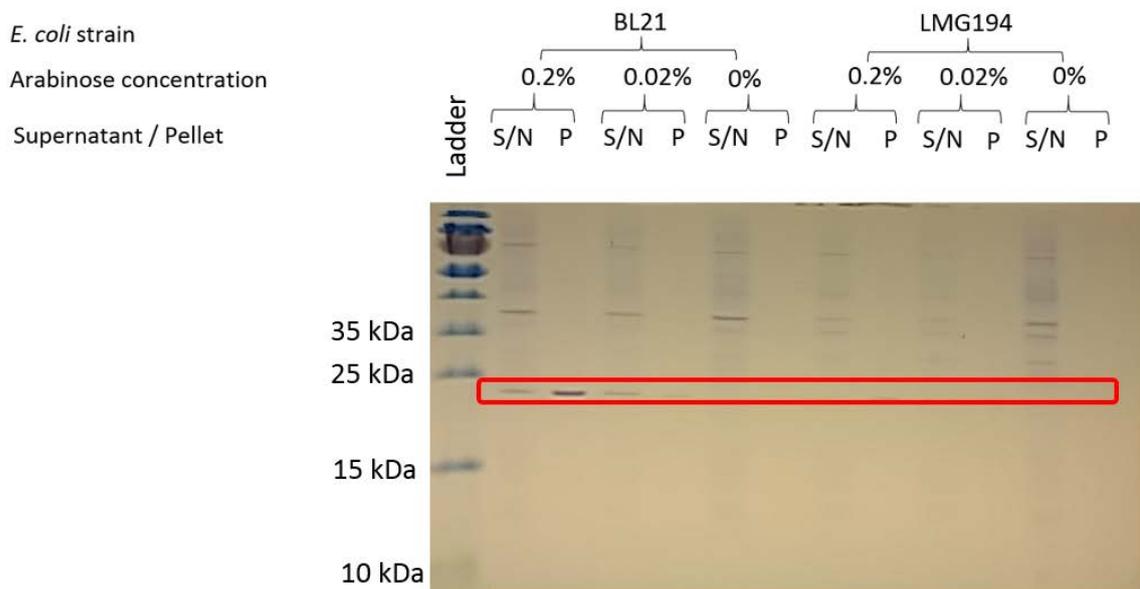
5 ml growth tests were cultured under various induction conditions to assess the optimal parameters (cell type, growth temperature, induction concentration, induction time, and lysis protocol) for aequorin expression. Cultures were grown at a single temperature between 37 °C and 20 °C to an OD<sub>600 nm</sub> of 0.5, where aequorin expression was induced by the addition of L-arabinose at 0.2% and 0.02%. Cells were induced for between one and 20 hours, and harvested. Three lysis protocols were investigated: lysis by Bugbuster, sonication in solution containing 500 mM NaCl and sonication in solution containing 150 mM NaCl, see figure 6.6.



**Figure 6.6 Induction and solubility test for Aequorin WT\*:** comparison of lysis protocols  
 A: lysed by Bugbuster. B: lysed by sonication in 20 mM sodium phosphate 500 mM NaCl.  
 C: lysed by sonication in 20 mM sodium phosphate 150 mM NaCl. *E. coli* cells were grown to an OD<sub>600 nm</sub> of 0.5 and induced with arabinose for 3 hours at 37 °C. Cells were lysed and centrifuged to separate soluble proteins (supernatant) from insoluble proteins (pellet). These were analysed by SDS-PAGE and visualised by Coomassie staining. The red band highlights putative aequorin protein.

Samples were centrifuged to separate soluble and insoluble proteins. Soluble proteins are likely to be in a folded state, whereas unfolded proteins are likely to aggregate, forming inclusion bodies, and therefore will be non-functional. (Palmer and Wingfield, 2004) The process of re-folding proteins adds complexity, and therefore does not lend itself to high efficiency production on a large scale. Additionally, proteins may not re-fold into the correct conformation. Through studying protein production from induction at 37 °C for three hours, the optimal lysis protocol was determined to be chemical lysis by Bugbuster. This resulted in the most protein found in the soluble fraction.

The previous induction at 37 °C for three hours was compared to that performed for one hour. A lower amount of total and soluble aequorin protein was expressed at one hour, see figure 6.7.



**Figure 6.7 Induction and solubility test for Aequorin WT\*:** comparison of time points *E. coli* cells were grown to an OD<sub>600 nm</sub> of 0.5 and induced with arabinose at 37 °C for one hour, followed by lysis by Bugbuster. Cells were lysed and centrifuged to separate soluble proteins (supernatant) from insoluble proteins (pellet). These were analysed by SDS-PAGE and visualised by Coomassie staining. The red band highlights putative aequorin protein.

In order to evaluate the effect of induction temperature, the previous results were compared to induction at 20 °C for 20 hours. Very little aequorin protein was expressed when incubated at the lower temperature, see figure 6.8.

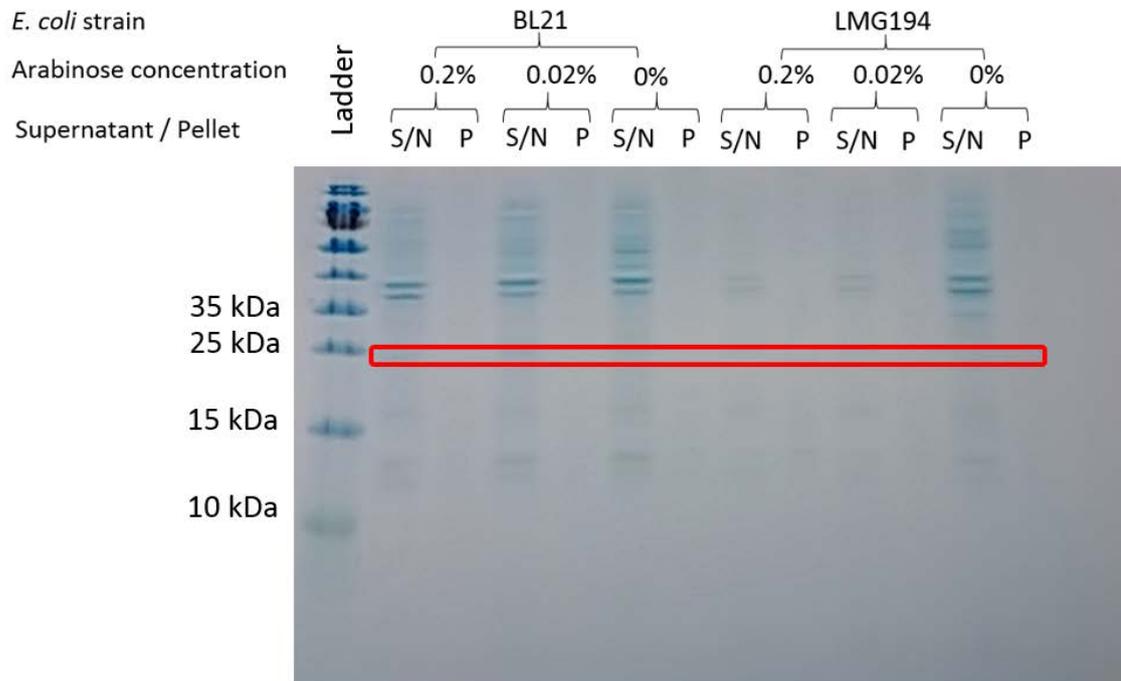


Figure 6.8 Induction and solubility test for Aequorin WT\*: comparison of induction temperatures

*E. coli* cells were grown to an OD<sub>600 nm</sub> of 0.5 and induced with arabinose at 20 °C for 20 hours, followed by lysis by Bugbuster. Cells were lysed and centrifuged to separate soluble proteins (supernatant) from insoluble proteins (pellet). These were analysed by SDS-PAGE and visualised by Coomassie staining. The red band highlights putative aequorin protein.

The presence of 22.6 kDa bands reflected expression of aequorin WT\*. Aequorin WT\* produced by BL21 cells cultured at 37 °C, induced for three hours at a concentration of 0.2% arabinose and lysed by Bugbuster, exhibited highest band density determined by SDS-PAGE analysis. Under the previously described conditions, around 40% of aequorin WT\* was soluble. Although not optimal, it was preferable to purify the smaller amount of soluble aequorin WT\*, than attempt to recover the larger proportion of insoluble protein for use in purification steps.

To confirm that the induced protein observed is recombinant aequorin, western blotting was performed using an anti-histidine tag antibody see figure 6.9. A polyhistidine-tagged protein was detected, which showed inducible behaviour under arabinose addition and corresponded to the correct predicted molecular weight. Thus supporting the conclusion that recombinant aequorin was expressed.

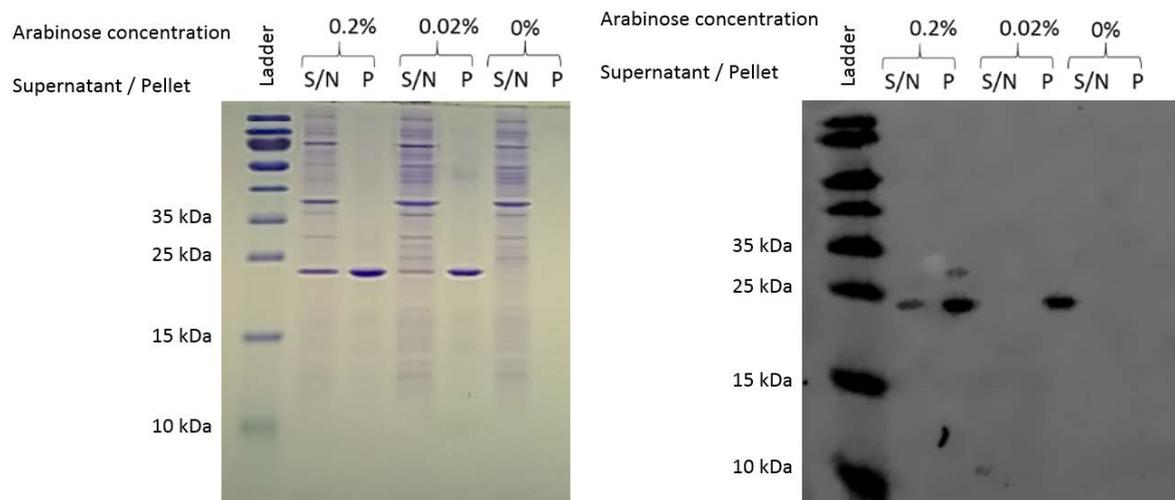


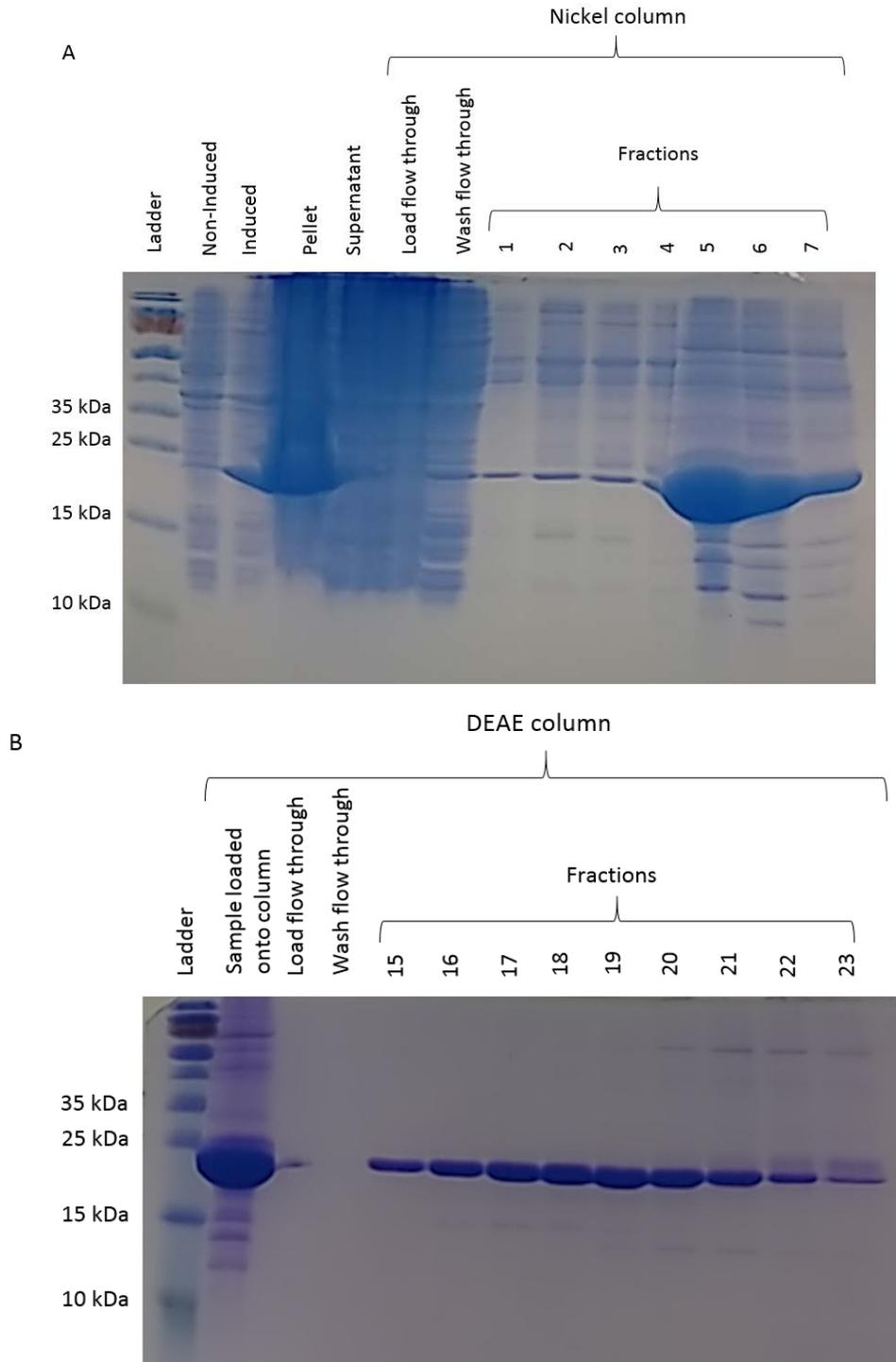
Figure 6.9 Confirmation induced protein is aequorin.

To confirm the induced protein observed is aequorin WT\*, SDS-PAGE gels were analysed by western blotting. An anti-histidine tag antibody was utilised for protein detection. BL21 cells were transformed to express aequorin WT\* were cultured at 37 °C, and induced for three hours. Cells were lysed by Bugbuster and centrifuged to separate soluble (supernatant) and insoluble (pelleted) protein. An arabinose inducible histidine tag-containing protein was detected at the predicted molecular weight.

## 6.6 Purification of aequorin WT\*

Proteins were expressed at the previously described conditions, cell debris and insoluble protein were removed by centrifugation following cell lysis and protein was purified by nickel chromatography. Soluble protein was added to a nickel column, which captured the protein via its polyhistidine tag. Nickel chromatography removed a number of contaminating proteins, however the eluate was an impure sample. The protein sample was further purified by ion exchange chromatography using a positively charged DEAE column. Aequorin WT\* has an isoelectric point of 5.20, as predicted by EXPASy software. (Gasteiger *et al.*, 2005) In a buffer with a pH higher than 5.20 the

protein will have a negative charge, and be able to bind to the positively charged DEAE matrix. Nickel affinity chromatography followed by ion exchange chromatography increased sample purity to between 95-100%, as shown in figure 6.10.

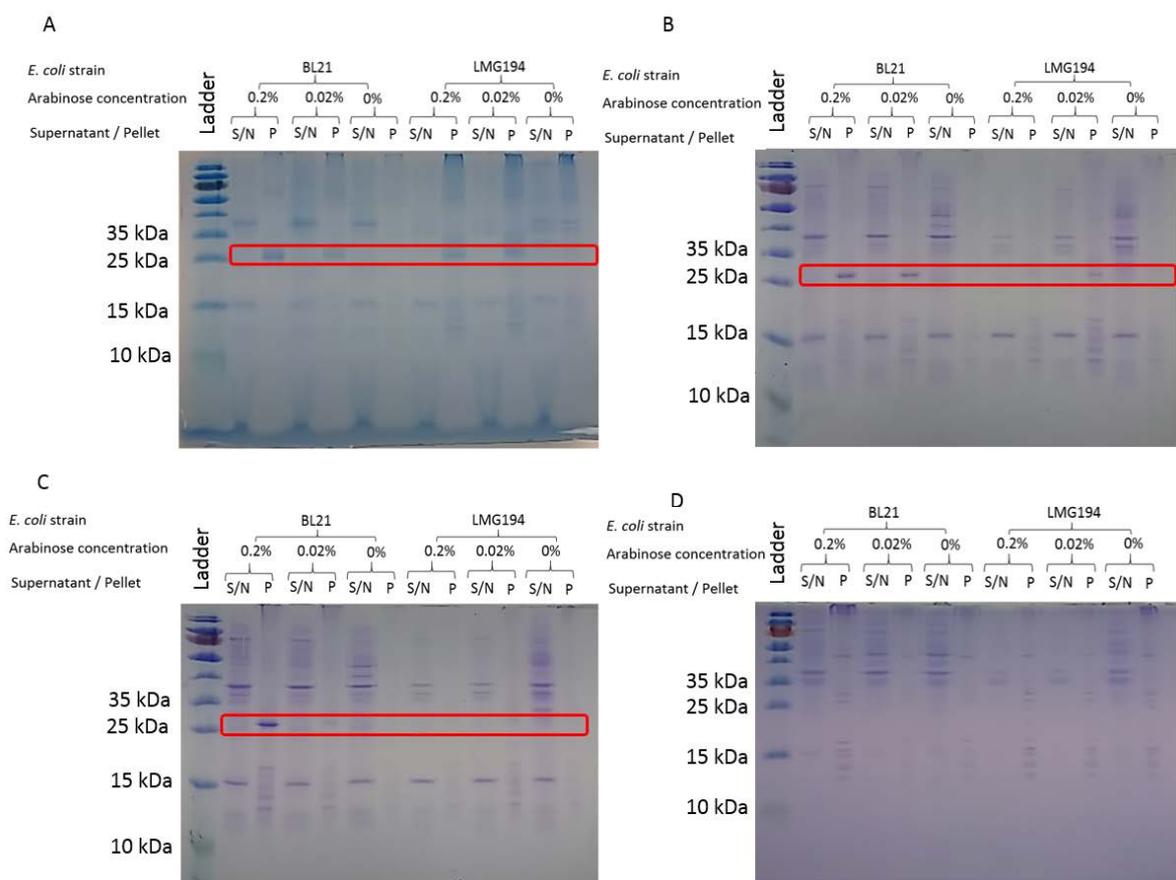


**Figure 6.10 Purification of Aequorin WT\***

A: BL21 aequorin WT\* cells were grown at 37 °C to an OD<sub>600 nm</sub> of 0.5 and induced with 0.2% arabinose for three hours. Cells were lysed with Bugbuster and purified on a nickel column. B: Protein was then loaded onto a DEAE column and purified by ion exchange. Figures are Coomassie stained SDS-PAGE gels.

## 6.7 Induction of aequorin FLAG

Induction tests were performed for aequorin FLAG, varying the same parameters as for aequorin WT\*, see figure 6.11. Aequorin FLAG expression is shown by a band at 23.6 kDa. There is little aequorin protein expressed in the insoluble fraction under any condition examined and no detectable soluble protein produced. Purification was not attempted due to lack of protein expression.



**Figure 6.11 Induction and solubility test for Aequorin FLAG**

A: 37 °C, 3 hours. B: 30 °C 3 hours. C: 30 °C, 4 hours. D: 25 °C, 20 hours. *E. coli* cells were grown to an OD<sub>600 nm</sub> of 0.5 and induced with arabinose under the stated conditions. Cells were lysed and centrifuged to separate soluble proteins (supernatant) from insoluble proteins (pellet). These were analysed by SDS-PAGE and visualised by Coomassie staining. No conditions studied produced soluble aequorin FLAG in significant amounts.

Although aequorin WT\* is able to be produced and purified, there is a large decrease in the amount of protein expressed when the FLAG tag insert is added. As culture growth was not impeded by protein induction, as shown by growth curves, see figure 6.12, it is unlikely that aequorin FLAG caused cell toxicity. The reduction in protein expression and solubility when the FLAG tag was inserted may be due to instability of the global protein structure, and possible degradation within the cell.

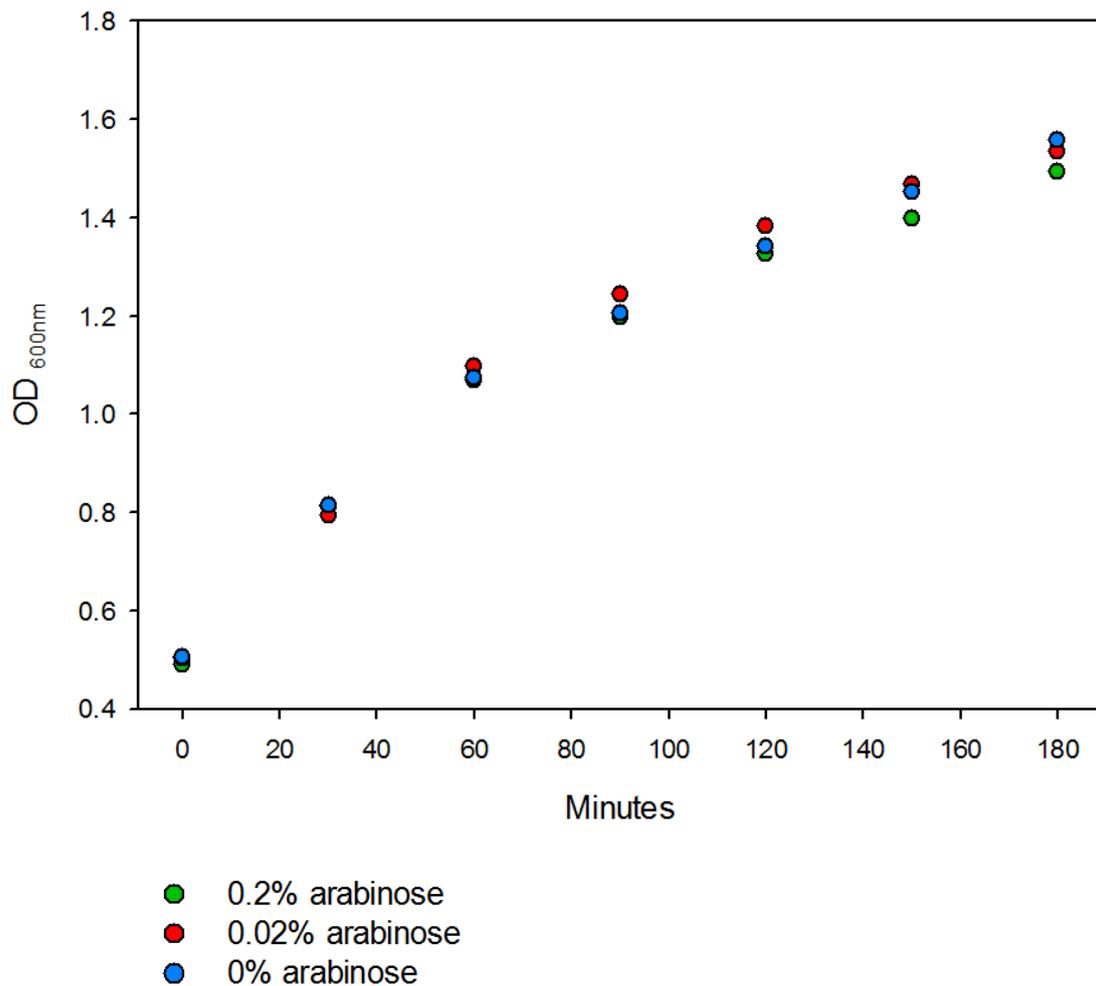


Figure 6.12 Growth curve of cells expressing aequorin FLAG

The growth of cells is unaffected by aequorin FLAG induction, as the growth curve is equivalent in cells with a high concentration of inducer (0.2% arabinose) and non-induced cells. This indicates that the protein is non-toxic.

## 6.8 Conclusion

It may be possible to select an insertion site which does not affect expression or solubility levels to as great an extent. However, as only 40% of the pseudo-native state protein is in its folded form, the protein may not be stable enough to withstand perturbations without significant restructuring. This does not lend itself to the creation of a platform library, which would require a large number of variants to be produced, and large scale production once a successful aptamer has been selected. For this reason research into developing aequorin into an aptamer scaffold was discontinued.

---

## Chapter 7 Discussion

---

### 7.1 Recent research in GBS diagnostics

The gold-standard method for *S. agalactiae* detection, microbial culture, takes several days to provide results, whilst PCR can be prohibitively expensive. Strategies to improve diagnosis of group B streptococcus colonisation include optimising PCR, developing tests based on isothermal DNA amplification, improving sample preparation steps and collecting antibiotic susceptibility data. This section focuses on research published since 2014.

#### 7.1.1 Optimisation of PCR

With the aim of optimising PCR sensitivity and specificity, studies have evaluated the effect of targeting alternative DNA sequences within *S. agalactiae*. PCR targeting the *scpB* gene was found to be more clinically sensitive than amplification of *cfb*. (Fouad *et al.*, 2016) Whilst in another study sensitivity was found to be highest when amplifying 16S rRNA in comparison with targets *cfb*, *atr* and *scpB*. (Mousavi *et al.*, 2016) Recently a CAMP factor negative clinical isolate unable to be detected by GeneXpert was reported. (Savini *et al.*, 2016) This highlights the importance of finding a universal target and shows the potential of IDRIS.

#### 7.1.2 Other methods of detection

As previously discussed, isothermal DNA amplification offers advantages over PCR, such as reduced costs and often quicker completion times. A LAMP-based diagnostic test was found to be 10-100 times more sensitive than PCR and able to detect samples containing 10 cfu per ml, with 100% specificity. This was achieved after 40 minutes. (Yang *et al.*, 2016) A further LAMP-based diagnostic test produced results after 60 minutes of amplification. (Seok *et al.*, 2017)) These times are consistent with intrapartum testing.

An RPA-based test which amplifies the gene encoding for CAMP factor was able to elicit a positive result from as few as 12 genome copies in 10-15 minutes. (Clarke *et al.*, 2016) Another RPA test, directed against the same target, had a sensitivity of 96% and specificity of 100% with respect to RT-PCR. Results were complete within 20 minutes, with positive results available after eight minutes (Daher *et al.*, 2014) and therefore both are also suited to intrapartum testing. The tests utilise the TwistAmp exo system, and therefore require a fluorescent reader to interpret results.

Another technique being evaluated for use in GBS diagnostics is mass spectrometry. Samples were cultured for 18-24 hours in selective broth before analysis, producing a quicker result in comparison to microbial culture but not enabling point-of-care testing. (Abrok *et al.*, 2015)

#### 7.1.3 Alternative detection strategies

LAMP-based amplification of *S. agalactiae* DNA has been performed on paper. Sample solutions travelled along wax printed fluidic channels to the reaction pad, which contained dry reagents. Using dry reagents ensures high stability in a wide range of conditions. During LAMP, pyrophosphate ions are produced, which bind magnesium and cause hydroxynaphthol blue fluorescence to decrease. This reduction in fluorescence was the observable signal. (Seok *et al.*, 2017) Paper-based amplification and detection decreases the complexity of the test, however the system requires a fluorescent reader. A LAMP-based *S. agalactiae* test which is able to be read visually was designed. In positive tests the solution changes colour from orange to green. This eliminates the need for fluorescence readers. (Yang *et al.*, 2016)

#### 7.1.4 Sample preparation

Efficiently obtaining DNA from bacterial samples can be used to lower the limit of detection of subsequent NAAT-based diagnostic tests without requiring a time consuming enrichment step. Including a mechanical bacterial cell lysis process prior to DNA extraction increased DNA yield 50-fold. (Burke *et al.*, 2016)

#### 7.1.5 Antibiotic susceptibility

Whilst there have been no cases of penicillin resistance in the UK, resistance to clindamycin and erythromycin is increasing. (Lamagni *et al.*, 2013) An ELISA was developed to simultaneously detect *S. agalactiae* presence and antibiotic susceptibility. Samples were incubated with antibiotics for up to nine hours prior to evaluation on an ELISA format. (Faro *et al.*, 2016) This may be useful in cases of maternal penicillin allergy.

#### 7.1.6 Identification of multiple pathogens

There are several pathogens which may cause an infant to present with suspected meningitis. In 2015 FDA approval was granted for the PCR-based Biofire Filmarray Meningitis/Encephalitis Panel, which analyses cerebrospinal fluid for the presence of 14 pathogens, including *S. agalactiae*. (Arora *et al.*, 2017) Incorporating GBS detection into a screening panel allows efficient identification of the causative agent to best inform treatment.

#### 7.1.7 Comparison to current work

The two diagnostic tests in development in this project aim to provide rapid feedback to inform clinical decisions during labour. RPA- and LAMP-based tests developed in recent literature also exhibit short time frames (Daher *et al.*, 2014; Clarke *et al.*, 2016; Yang *et al.*, 2016; Seok *et al.*, 2017) consistent with this objective. The proposed antibody- and hybridisation-based lateral flow assays are visualised optically, presenting an advantage over fluorescence-based detection which requires a fluorescence reader for interpretation. Optical detection in solution has been achieved in LAMP-based *S. agalactiae* detection. (Yang *et al.*, 2016) A process of optimising sample preparation may prove beneficial to improving test sensitivity. Alternative methods of improving GBS diagnostics which aim to provide extra information such as antibiotic susceptibility or presence of additional pathogens were not pursued within this project.

## 7.2 IDRIS as a useful tool

IDRIS provides a significant advancement in the selection of bacterial group-distinctive peptides. The programme allows peptide selection, or shortlisting, with no prior knowledge of the protein. In this work the peptides utilised for antibody production, reside in a largely unstudied protein. Bioinformatic analysis performed by IDRIS was evaluated empirically. RT-PCR experiments showed the universal transcription of the selected protein across all serotypes and ELISA results showed expression of a selected peptide (peptide B) in all serotypes of the species of interest and no expression in a closely related species. In summary, data collected so far supports the use of this peptide as a target for GBS diagnostics. Without IDRIS to highlight the ability of the peptide to act as a biomarker for all *S. agalactiae* strains, it would not have been selected.

IDRIS may also be useful in the search for GBS vaccine targets. Protein-based vaccines do not share the same serotype-specific limitations that capsular polysaccharide vaccines have, and may be preferential for vaccine development. (Nuccitelli *et al.*, 2015) IDRIS is able to identify peptides found in all sequenced bacteria in a group of interest and identify those likely to be exposed on the cell surface, two key features for vaccine targets.

## 7.3 ASSURED Diagnostics

The term ASSURED, coined by the World Health Organisation, describes characteristics diagnostic tests should meet. Tests should be affordable, sensitive, specific, user friendly, robust and rapid, equipment-free and deliverable to those who need them. (Mabey *et al.*, 2004) Levels of acceptable characteristics will be dictated by the intended use. The detection tools produced in this study are evaluated using these criteria.

### 7.3.1 Antibody for application to GBS diagnostics

Future work will involve developing the ELISA assay into a lateral flow format following the production of a successful antibody against a second target on the protein. Lateral flow assays provide equipment-free and user friendly readouts, and are able to be deployed at the point of care. These diagnostic tests are low cost to manufacture and often do not require refrigerated storage, and are therefore relatively robust. Lateral flow assays provide rapid results within 5-30 minutes, (Koczula and Gallotta, 2016) which would enable point of care testing at the onset of labour. Intrapartum testing could be used in a universal screening programme or for testing individuals with appropriate clinical presentations. Public Health England has stated that rapid testing would be of particular use in cases of preterm births and premature membrane rupture. (Public Health England, 2015a)

Initial results assessing bacterial sensitivity and specificity are promising, suggesting that the antibodies produced detect all *S. agalactiae* strains and no *S. pyogenes* strains. Additional work will include evaluating a panel of bacteria commonly found in rectal and vaginal swabs for cross reactivity. Testing patient samples will determine clinical sensitivity and specificity values. In lateral flow assays, gold conjugated detection molecules are typically used as reporter signals. As optical immunoassays previously developed for the detection of *S. agalactiae* have struggled with clinical sample sensitivity, (Baker, 1996; Daniels *et al.*, 2009) ultrasensitive state of the art nanoparticles, could be used in replacement of gold particles. (Bruno, 2014; Howes *et al.*, 2014; Hu *et al.*, 2017)

### 7.3.2 RPA for application to GBS diagnostics

RPA applied to a lateral flow format has many of the advantages previously described for antigen detection in lateral flow assays, such as ease of use and applicability to a point of care setting. In comparison to Cepheid's GeneXpert GBS, the equipment demands are reduced, as amplification can be carried out at a constant temperature rather than requiring an expensive thermocycler. Individual Xpert GBS tests cost £38.80, (Jones *et al.*, 2015) whereas RPA coupled to read-out on a lateral flow assay via hybridisation costs €1.15 per assay. This is also cheaper than antibody-based

detection of DNA (Jauset-Rubio *et al.*, 2016b) and traditional screening, which is estimated to cost \$4.76. (El Helali *et al.*, 2012)

In this work DNA amplification was performed for 20 minutes, and it is reported that colour visualisation, resulting from DNA hybridisation onto lateral flow strips, occurs within 30 seconds of sample addition. (Jauset-Rubio *et al.*, 2016b) This time scale is much lower than that of Xpert GBS (Jones *et al.*, 2015) and is consistent with point of care testing. Further work is required to investigate optimal sample preparation steps.

The test in development is able to amplify DNA from all serotypes of *S. agalactiae* and produce no signal (above background) from strains of *S. pyogenes*. Initial specificity testing results are promising, however full analysis using an extended panel of bacteria and clinical samples is required before specificity can be assured. The test in its current ELONA format can detect  $3.8 \times 10^4$  copies of DNA. Further sensitivity tests will be carried out in lateral flow format, initially with DNA and bacterial dilution series, followed by clinical samples.

#### 7.3.3 Aequorin as a peptide aptamer scaffold for application to GBS diagnostics

Peptide insertion into aequorin caused decreased recombinant protein production and solubility; therefore research into the development of aequorin into a peptide aptamer scaffold was discontinued. Selection of an alternate scaffold with reporter function may enable the coupling of antigen detection directly to signal production.

#### 7.4 Conclusion

The work on antibody- and RPA-based GBS diagnostics is promising, with both tests showing sensitivity to all serotypes of *S. agalactiae* and a lack of cross-reactivity to the related species *S. pyogenes*. Sensitivity and specificity assessment of the tests using clinical samples is required before full comparison can be made to currently available tests. Development of both systems to lateral flow formats is intended, with the

objective to produce rapid results in an inexpensive, simple process requiring little equipment which can be performed at the onset of labour. It is proposed that rapid, sensitive and specific testing would aid treatment of colonised mothers, resulting in fewer cases of early onset GBS disease, whilst reducing inappropriate antibiotic prescription.

---

## Chapter 8 References

---

Abedi, M.R., Caponigro, G. and Kamb, A. (1998) 'Green fluorescent protein as a scaffold for intracellular presentation of peptides', *Nucleic Acids Research*, 26(2), pp. 623-630.

Abrok, M., Arcson, A., Lazar, A., Urban, E. and Deak, J. (2015) 'Combination of selective enrichment and MALDI-TOF MS for rapid detection of *Streptococcus agalactiae* colonisation of pregnant women', *Journal of Microbiological Methods*, 114, pp. 23-25.

Aghebati-Maleki, L., Bakhshinejad, B., Baradaran, B., Motallebnezhad, M., Aghebati-Maleki, A., Nickho, H., Yousefi, M. and Majidi, J. (2016) 'Phage display as a promising approach for vaccine development', *Journal of Biomedical Science*, 23.

Al-Maani, A., Streitenberger, L., Clarke, M., Yau, Y.C.W., Kovach, D., Wray, R. and Matlow, A. 29 (2014) 'Nosocomial transmission of group B streptococci proven by positive environmental culture'. *Oman Medical Journal* 376-379.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*, 25(17), pp. 3389-3402.

Alves, V.S., Pimenta, D.C., Sattlegger, E. and Castilho, B.A. (2004) 'Biophysical characterization of Gir2, a highly acidic protein of *Saccharomyces cerevisiae* with anomalous electrophoretic behavior', *Biochemical and Biophysical Research Communications*, 314(1), pp. 229-234.

An, L.X., Tang, W., Ranalli, T.A., Kim, H.J., Wytiaz, J. and Kong, H.M. (2005) 'Characterization of a thermostable UvrD helicase and its participation in helicase-dependent amplification', *Journal of Biological Chemistry*, 280(32), pp. 28952-28958.

Areschoug, T., Stalhammar-Carlemalm, M., Larsson, C. and Lindahl, G. (1999) 'Group B streptococcal surface proteins as targets for protective antibodies: Identification of two novel proteins in strains of serotype V', *Infection and Immunity*, 67(12), pp. 6350-6357.

Arora, H.S., Asmar, B.I., Salimnia, H., Agarwal, P., Chawla, S. and Abdel-Haq, N. (2017) 'Enhanced identification of group B streptococcus and *Escherichia coli* in young infants with meningitis using the Biofire Filmarray Meningitis/Encephalitis panel', *Pediatric Infectious Disease Journal*, 36(7), pp. 685-687.

Ascher, D.P., Wilson, S. and Fischer, G.W. (1991) 'Comparison of commercially available group-B streptococcal latex agglutination assays', *Journal of Clinical Microbiology*, 29(12), pp. 2895-2896.

- Aziz, N., Baron, E.J., D'Souza, H., Nourbakhsh, M., Druzin, M.L. and Benitz, W.E. (2005) 'Comparison of rapid intrapartum screening methods for group B streptococcal vaginal colonization', *Journal of Maternal-Fetal & Neonatal Medicine*, 18(4), pp. 225-229.
- Baker, C.J. (1996) 'Inadequacy of rapid immunoassays for intrapartum detection of group B streptococcal carriers', *Obstetrics and Gynecology*, 88(1), pp. 51-55.
- Baker, C.J., Paoletti, L.C., Rench, M.A., Guttormsen, H.K., Carey, V.J., Hickman, M.E. and Kasper, D.L. (2000) 'Use of capsular polysaccharide-tetanus toxoid conjugate vaccine for type II group B Streptococcus in healthy women', *Journal of Infectious Diseases*, 182(4), pp. 1129-1138.
- Baker, C.J., Paoletti, L.C., Rench, M.A., Guttormsen, H.K., Edwards, M.S. and Kasper, D.L. (2004) 'Immune response of healthy women to 2 different group B streptococcal type V capsular polysaccharide-protein conjugate vaccines', *Journal of Infectious Diseases*, 189(6), pp. 1103-1112.
- Baker, C.J., Paoletti, L.C., Wessels, M.R., Guttormsen, H.K., Rench, M.A., Hickman, M.E. and Kasper, D.L. (1999) 'Safety and immunogenicity of capsular polysaccharide-tetanus toxoid conjugate vaccines for group B streptococcal types Ia and Ib', *Journal of Infectious Diseases*, 179(1), pp. 142-150.
- Baker, C.J., Rench, M.A., Paoletti, L.C. and Edwards, M.S. (2007) 'Dose-response to type V group B streptococcal polysaccharide-tetanus toxoid conjugate vaccine in healthy adults', *Vaccine*, 25(1), pp. 55-63.
- Baldi, P., Pollastri, G., CA, A. and Brunak, S. 8 (2000) 'Matching protein beta-sheet partners by feedforward and recurrent neural networks'. Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology, pp. 25-36.
- Baneyx, F. and Mujacic, M. (2004) 'Recombinant protein folding and misfolding in Escherichia coli', *Nature Biotechnology*, 22(11), pp. 1399-1408.
- Banta, S., Dooley, K. and Shur, O. (2013) 'Replacing antibodies: Engineering new binding proteins', *Annual Review of Biomedical Engineering*, Vol 15, 15, pp. 93-113.
- Beena, K., Udgaonkar, J.B. and Varadarajan, R. (2004) 'Effect of signal peptide on the stability and folding kinetics of maltose binding protein', *Biochemistry*, 43(12), pp. 3608-3619.
- Benitz, W.E., Gould, J.B. and Druzin, M.L. (1999) 'Risk factors for early-onset group B streptococcal sepsis: Estimation of odds ratios by critical literature review', *Pediatrics*, 103(6).
- Benson, J.A., Flores, A.E., Baker, C.J., Hillier, S.L. and Ferrieri, P. (2002) 'Improved methods for typing nontypeable isolates of group B streptococci', *International Journal of Medical Microbiology*, 292(1), pp. 37-42.
- Berardi, A., Rossi, C., Lugli, L., Creti, R., Reggiani, M.L.B., Lanari, M., Memo, L., Pedna, M.F., Venturelli, C., Perrone, E., Ciccia, M., Tridapalli, E., Piepoli, M.,

- Contiero, R., Ferrari, F. and Grp, G.B.S.P.W. (2013) 'Group B streptococcus late-onset disease: 2003-2010', *Pediatrics*, 131(2), pp. E361-E368.
- Berg, S., Trollfors, B., Lagergard, T., Zackrisson, G. and Claesson, B.A. (2000) 'Serotypes and clinical manifestations of group B streptococcal infections in western Sweden', *Clinical Microbiology and Infection*, 6(1), pp. 9-13.
- Bessen, D., Jones, K.F. and Fischetti, V.A. 169 (1989) 'Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever'. *Journal of Experimental Medicine*, pp. 269-283.
- Bishop, E.J., Shilton, C., Benedict, S., Kong, F., Gilbert, G.L., Gal, D., Godoy, D., Spratt, B.G. and Currie, B.J. (2007) 'Necrotizing fasciitis in captive juvenile *Crocodylus porosus* caused by *Streptococcus agalactiae*: an outbreak and review of the animal and human literature', *Epidemiology and Infection*, 135(8), pp. 1248-1255.
- Block, T., Munson, E., Culver, A., Vaughan, K. and Hryciuk, J.E. (2008) 'Comparison of carrot broth- and selective Todd-Hewitt broth-enhanced PCR protocols for real-time detection of *Streptococcus agalactiae* in prenatal vaginal/anorectal specimens', *Journal of Clinical Microbiology*, 46(11), pp. 3615-3620.
- Bloomsbury Centre for Bioinformatics (n.d.) *Overview of prediction methods predict secondary structure (PSIPRED)*. Available at: <http://bioinf.cs.ucl.ac.uk/index.php?id=779> (Accessed: 3rd October 2017).
- Boes, M. (2000) 'Role of natural and immune IgM antibodies in immune responses', *Molecular Immunology*, 37(18), pp. 1141-1149.
- Bogaert, D., de Groot, R. and Hermans, P.W.M. (2004) 'Streptococcus pneumoniae colonisation: the key to pneumococcal disease', *Lancet Infectious Diseases*, 4(3), pp. 144-154.
- Bonora, M., Giorgi, C., Bononi, A., Marchi, S., Patergnani, S., Rimessi, A., Rizzuto, R. and Pinton, P. (2013) 'Subcellular calcium measurements in mammalian cells using jellyfish photoprotein aequorin-based probes', *Nature Protocols*, 8(11), pp. 2105-2118.
- Boris, S. and Barbes, C. (2000) 'Role played by lactobacilli in controlling the population of vaginal pathogens', *Microbes and Infection*, 2(5), pp. 543-546.
- Bornhorst, J.A. and Falke, J.J. (2000) 'Purification of proteins using polyhistidine affinity tags', *Applications of Chimeric Genes and Hybrid Proteins, Pt A*, 326, pp. 245-254.
- Brodeur, B.R., Boyer, M., Charlebois, I., Hamel, J., Couture, F., Rioux, C.R. and Martin, D. (2000) 'Identification of group B streptococcal Sip protein, which elicits cross-protective immunity', *Infection and Immunity*, 68(10), pp. 5610-5618.
- Brooks, S., Apostol, M., Nadle, J., Wymore, K., Haubert, N., Burnite, S., Daniels, A., Hadler, J.L., Farley, M.M., Martell-Cleary, P., Harrison, L.H., Sanza, L.T., Morin, C., Lynfield, R., Albanese, B., Baretta, J., Anderson, B., Cieslak, P., Stefonek, K., Barnes, B., Craig, A.S., Schrag, S.J., Zell, E. and Phares, C.R. (2006) 'Early-onset and late-

onset neonatal group B streptococcal disease - United States, 1996-2004 (Reprinted from MMWR, vol 54, pg 1205, 2005)', *Jama-Journal of the American Medical Association*, 295(12), pp. 1371-1372.

Bruno, J.G. 3 (2014) 'Application of DNA aptamers and quantum dots to lateral flow test strips for detection of foodborne pathogens with improved sensitivity versus colloidal gold'. *Pathogens*, pp. 341-355.

Bunka, D.H.J., Platonova, O. and Stockley, P.G. (2010) 'Development of aptamer therapeutics', *Current Opinion in Pharmacology*, 10(5), pp. 557-562.

Burke, R.M., McKenna, J.P., Cox, C., Coyle, P.V., Shields, M.D. and Fairley, D.J. (2016) 'A comparison of different pre-lysis methods and extraction kits for recovery of *Streptococcus agalacticae* (Lancefield group B *Streptococcus*) DNA from whole blood', *Journal of Microbiological Methods*, 129, pp. 103-108.

Burns, G. and Plumb, J. (2013) 'GBS public awareness, advocacy, and prevention-What's working, what's not and why we need a maternal GBS vaccine', *Vaccine*, 31, pp. D58-D65.

Cammack, R., Atwood, T., Campbell, P., Parish, H., Smith, A., Vella, F. and Stirling, J. (2006) 'Oxford dictionary of biochemistry and molecular biology (2 ed.)'. Oxford University Press.

Casqueiro, J., Casqueiro, J. and Alves, C. 16 (2012) 'Infections in patients with diabetes mellitus: A review of pathogenesis'. *Indian Journal of Endocrinology and Metabolism*, pp. S27-S36.

Centers for Disease Control and Prevention 48 (1999) 'Laboratory practices for prenatal group B streptococcal screening and reporting - Connecticut, Georgia, and Minnesota, 1997-1998'. *Morbidity and Mortality Weekly Report*, pp. 426-428 20.

Cepheid (2018) *Xpert GBS*. Available at: <http://cepheid.com/en/cepheid-solutions/clinical-ivd-tests/sexual-health/xpert-gbs> (Accessed: 22nd February 2018).

Cham, C.M., Xu, H., O'Keefe, J.P., Rivas, F.V., Zagouras, P. and Gajewski, T.F. (2003) 'Gene array and protein expression profiles suggest post-transcriptional regulation during CD8(+) T cell differentiation', *Journal of Biological Chemistry*, 278(19), pp. 17044-17052.

Chaplin, D.D. (2010) 'Overview of the immune response', *Journal of Allergy and Clinical Immunology*, 125(2), pp. S3-S23.

Cheng, Q., Debol, S., Lam, H., Eby, R., Edwards, L., Matsuka, Y., Olmsted, S.B. and Cleary, P.P. (2002a) 'Immunization with C5a peptidase or peptidase-type III polysaccharide conjugate vaccines enhances clearance of group B streptococci from lungs of infected mice', *Infection and Immunity*, 70(11), pp. 6409-6415.

Cheng, Q., Stafslie, D., Purushothaman, S.S. and Cleary, P. (2002b) 'The group B streptococcal C5a peptidase is both a specific protease and an invasin', *Infection and Immunity*, 70(5), pp. 2408-2413.

Cho, C.-Y., Tang, Y.-H., Chen, Y.-H., Wang, S.-T., Yang, Y.-H., Wang, T.-H., Yeh, C.-C., Wu, K.-G. and Jeng, M.-J. (2017) 'Group B Streptococcal infection in neonates and colonization in pregnant women: An epidemiological retrospective analysis'. *Journal of Microbiology, Immunology and Infection*. Available at: <http://dx.doi.org/10.1016/j.jmii.2017.08.004>.

Christie, R., Atkins, N., Munch-Petersen, E. and 22:197-200., A.J.E.B.M.S. 22 (1944) 'A note on a lytic phenomenon shown by group B streptococci'. *The Australian Journal of Experimental Biology and Medical Science*, pp. 197-200.

Chu, Y.W., Tse, C., Tsang, G.K.L., So, D.K.S., Fung, J.T.L. and Lo, J.Y.C. (2007) 'Invasive group B Streptococcus isolates showing reduced susceptibility to penicillin in Hong Kong', *Journal of Antimicrobial Chemotherapy*, 60(6), pp. 1407-1409.

Cieslewicz, M.J., Chaffin, D., Glusman, G., Kasper, D., Madan, A., Rodrigues, S., Fahey, J., Wessels, M.R. and Rubens, C.E. (2005) 'Structural and genetic diversity of group B Streptococcus capsular polysaccharides', *Infection and Immunity*, 73(5), pp. 3096-3103.

Clarke, C., O'Connor, L., Carre-Skinner, H., Piepenburg, O. and Smith, T.J. (2016) 'Development and performance evaluation of a recombinase polymerase amplification assay for the rapid detection of group B streptococcus', *Bmc Microbiology*, 16.

Cohen, B.A., Colas, P. and Brent, R. (1998) 'An artificial cell-cycle inhibitor isolated from a combinatorial Library', *Proceedings of the National Academy of Sciences of the United States of America*, 95(24), pp. 14272-14277.

Colas, P., Cohen, B., Ferrigno, P.K., Silver, P.A. and Brent, R. (2000) 'Targeted modification and transportation of cellular proteins', *Proceedings of the National Academy of Sciences of the United States of America*, 97(25), pp. 13720-13725.

Colbourn, T.E., Asseburg, C., Bojke, L., Philips, Z., Welton, N.J., Claxton, K., Ades, A.E. and Gilbert, R.E. (2007) 'Preventive strategies for group B streptococcal and other bacterial infections in early infancy: cost effectiveness and value of information analyses', *British Medical Journal*, 335(7621), pp. 655-658A.

Coleman, J.E. (1992) 'Structure and mechanism of alkaline-phosphatase', *Annual Review of Biophysics and Biomolecular Structure*, 21, pp. 441-483.

Compton, J. (1991) 'Nucleic-acid sequence-based amplification', *Nature*, 350(6313), pp. 91-92.

Condemine, G. and Shevchik, V.E. (2000) 'Overproduction of the secretin OutD suppresses the secretion defect of an Erwinia chrysanthemii outB mutant', *Microbiology-UK*, 146, pp. 639-647.

Cordray, M.S. and Richards-Kortum, R.R. (2015) 'A paper and plastic device for the combined isothermal amplification and lateral flow detection of Plasmodium DNA', *Malaria Journal*, 14.

Costa, S., Almeida, A., Castro, A. and Domingues, L. (2014) 'Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: the novel Fh8 system', *Frontiers in Microbiology*, 5.

Courchesne, W.E. and Ozturk, S. (2003) 'Amiodarone induces a caffeine-inhibited, MID1-dependent rise in free cytoplasmic calcium in *Saccharomyces cerevisiae*', *Molecular Microbiology*, 47(1), pp. 223-234.

Crannell, Z.A., Rohrman, B. and Richards-Kortum, R. (2014) 'Equipment-free incubation of recombinase polymerase amplification reactions using body heat', *Plos One*, 9(11).

Creti, R., Fabretti, F., Orefici, G. and von Hunolstein, C. (2004) 'Multiplex PCR assay for direct identification of group B streptococcal alpha-protein-like protein genes', *Journal of Clinical Microbiology*, 42(3), pp. 1326-1329.

Cunningham, M.W. (2008) 'Pathogenesis of group A streptococcal infections and their sequelae', *Hot Topics in Infection and Immunity in Children Iv*, 609, pp. 29-42.

Daher, R.K., Stewart, G., Boissinot, M. and Bergeron, M.G. (2014) 'Isothermal recombinase polymerase amplification assay applied to the detection of group B streptococci in vaginal/anal Samples', *Clinical Chemistry*, 60(4), pp. 660-666.

Daher, R.K., Stewart, G., Boissinot, M. and Bergeron, M.G. (2016) 'Recombinase polymerase amplification for diagnostic applications', *Clinical Chemistry*, 62(7), pp. 947-958.

Dahesh, S., Hensler, M.E., Van Sorge, N.M., Gertz, R.E., Schrag, S., Nizet, V. and Beall, B.W. (2008) 'Point mutation in the group B streptococcal *pbp2x* gene conferring decreased susceptibility to beta-lactam antibiotics', *Antimicrobial Agents and Chemotherapy*, 52(8), pp. 2915-2918.

Daniels, J., Gray, J., Pattison, H., Roberts, T., Edwards, E., Milner, P., Spicer, L., King, E., Hills, R.K., Gray, R., Buckley, L., Magill, L., Elliman, N., Kaambwa, B., Bryan, S., Howard, R., Thompson, P. and Khan, K.S. (2009) 'Rapid testing for group B streptococcus during labour: a test accuracy study with evaluation of acceptability and cost-effectiveness', *Health Technology Assessment*, 13(42), pp. 1-+.

Davies, S., Gear, J.E., Mason, C.M. and McIntyre, S.M. (2003) 'Streptococcus grouping latex kits: evaluation of five commercially available examples', *British Journal of Biomedical Science*, 60(3), pp. 136-140.

Davis, J.J., Tkac, J., Humphreys, R., Buxton, A.T., Lee, T.A. and Ferrigno, P.K. (2009) 'Peptide aptamers in label-free protein detection: 2. Chemical optimization and detection of distinct protein isoforms', *Analytical Chemistry*, 81(9), pp. 3314-3320.

De Schutter, K. and Van Damme, E.J.M. (2015) 'Protein-carbohydrate interactions, and beyond', *Molecules*, 20(8), pp. 15202-15205.

de Tejada, B.M., Pfister, R.E., Renzi, G., Francois, P., Irion, O., Boulvain, M. and Schrenzel, J. (2011) 'Intrapartum Group B streptococcus detection by rapid

polymerase chain reaction assay for the prevention of neonatal sepsis', *Clinical Microbiology and Infection*, 17(12), pp. 1786-1791.

de-Paris, F., Machado, A., Gheno, T.C., Ascoli, B.M., de Oliveira, K.R.P. and Barth, A.L. (2011) 'Group B Streptococcus detection: comparison of PCR assay and culture as a screening method for pregnant women', *Brazilian Journal of Infectious Diseases*, 15(4), pp. 323-327.

Delannoy, C.M.J., Crumlish, M., Fontaine, M.C., Pollock, J., Foster, G., Dagleish, M.P., Turnbull, J.F. and Zadoks, R.N. (2013) 'Human Streptococcus agalactiae strains in aquatic mammals and fish', *Bmc Microbiology*, 13.

Deo, S.K., Lewis, J.C. and Daunert, S. (2001) 'C-terminal and N-terminal fusions of aequorin with small peptides in immunoassay development', *Bioconjugate Chemistry*, 12(3), pp. 378-384.

Dillon, H.C., Khare, S. and Gray, B.M. (1987) 'Group-B Streptococcal carriage and disease - A 6-year prospective study', *Journal of Pediatrics*, 110(1), pp. 31-36.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N. and Knight, R. (2010) 'Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns', *Proceedings of the National Academy of Sciences of the United States of America*, 107(26), pp. 11971-11975.

Doro, F., Liberatori, S., Rodriguez-Ortega, M.J., Rinaudo, C.D., Rosini, R., Mora, M., Scarselli, M., Altindis, E., D'Aurizio, R., Stella, M., Margarit, I., Maione, D., Telford, J.L., Norais, N. and Grandi, G. (2009) 'Surfome analysis as a fast track to vaccine discovery', *Molecular & Cellular Proteomics*, 8(7), pp. 1728-1737.

Edgar, R., Domrachev, M. and Lash, A.E. (2002) 'Gene Expression Omnibus: NCBI gene expression and hybridization array data repository', *Nucleic Acids Research*, 30(1), pp. 207-210.

Edmond, K.M., Kortsalioudaki, C., Scott, S., Schrag, S.J., Zaidi, A.K.M., Cousens, S. and Heath, P.T. (2012) 'Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis', *Lancet*, 379(9815), pp. 547-556.

Edwards, M.S. and Baker, C.J. (2005) 'Group B streptococcal infections in elderly adults', *Clinical Infectious Diseases*, 41(6), pp. 839-847.

Edwards, M.S., Kasper, D.L., Jennings, H.J., Baker, C.J. and Nicholsonweller, A. (1982) 'Capsular sialic-acid prevents activation of the alternative complement pathway by type-III, group-B streptococci', *Journal of Immunology*, 128(3), pp. 1278-1283.

Edwards, M.S., Rench, M.A., Rinaudo, C.D., Fabbrini, M., Tuscano, G., Buffi, G., Bartolini, E., Bonacci, S., Baker, C.J. and Margarit, I. (2016) 'Immune responses to invasive group B streptococcal disease in adults', *Emerging Infectious Diseases*, 22(11), pp. 1877-1883.

- Egger, P., Siegrist, C., G, S., D, B. and R, A. 149 (1990) 'Evaluation of two ELISA tests for the rapid detection of group A streptococci'. *European Journal of Paediatrics*, pp. 256-258
- El Helali, N., Giovangrandi, Y., Guyot, K., Chevet, K., Gutmann, L. and Durand-Zaleski, I. (2012) 'Cost and effectiveness of intrapartum group B streptococcus polymerase chain reaction screening for term deliveries', *Obstetrics and Gynecology*, 119(4), pp. 822-829.
- Ellem, J.A., Kovacevic, D., Olma, T. and Chen, C.S.-A. In Press (2017) 'Rapid detection of Group B streptococcus directly from vaginal–rectal specimens using liquid swabs and the BD Max GBS assay'. *Clinical Microbiology and Infection*.
- Ellington, A. and Szostak, J. 346 (1990) 'In vitro selection of RNA molecules that bind specific ligands'. *Nature*, pp. 818-822.
- Emmadi, R., Boonyaratanakomkit, J.B., Selvarangan, R., Shyamala, V., Zimmer, B.L., Williams, L., Bryant, B., Schutzbank, T., Schoonmaker, M.M., Wilson, J.A.A., Hall, L., Pancholi, P. and Bernard, K. (2011) 'Molecular methods and platforms for infectious diseases testing a review of FDA-approved and cleared assays', *Journal of Molecular Diagnostics*, 13(6), pp. 583-604.
- Epalza, C., Goetghebuer, T., Hainaut, M., Prayez, F., Barlow, P., Dediste, A., Marchant, A. and Levy, J. (2010) 'High incidence of invasive group B streptococcal infections in HIV-exposed uninfected infants', *Pediatrics*, 126(3), pp. E631-E638.
- Evans, D., Johnson, S., Laurenson, S., Davies, A.G., Ferrigno, P.K. and Wälti, C. 7 (2008) 'Electrical protein detection in cell lysates using high-density peptide-aptamer microarrays'. *Journal of Biology* 3.
- Farley, M.M. (2001) 'Group B streptococcal disease in nonpregnant adults', *Clinical Infectious Diseases*, 33(4), pp. 556-561.
- Faro, J., Mitchell, M., Chen, Y.-J., Kamal, S., Riddle, G. and Faro, S. 2016 (2016) 'Development of a novel test for simultaneous bacterial identification and antibiotic susceptibility'. *Infectious Diseases in Obstetrics and Gynecology*, pp. 1-7.
- Fernandez-Alonso, M.D., Diaz, D., Berbis, M.A., Marcelo, F., Canada, J. and Jimenez-Barbero, J. (2012) 'Protein-carbohydrate interactions studied by NMR: From molecular recognition to drug design', *Current Protein & Peptide Science*, 13(8), pp. 816-830.
- Ferrigno, P.K. (2016) 'Non-antibody protein-based biosensors', *Biosensor Technologies for Detection of Biomolecules*, 60(1), pp. 19-25.
- Fischetti, V. (2006) 'Surface proteins on gram-positive bacteria, gram-positive pathogens, 2nd edition'. American Society for Microbiology Press, Washington, D.C., pp. 12-25.
- Flanagan, K., Cockell, S., Harwood, C., Hallinan, J., Nakjang, S., Lawry, B. and Wipat, A. (2014) 'A distributed computational search strategy for the identification of

diagnostics targets: application to finding aptamer targets for methicillin-resistant staphylococci', *Journal of integrative bioinformatics*, 11(2), pp. 242-242.

Fouad, M., Zakaria, S., Metwally, L., Aboul-Atta, H. and Kamel, M. 6 (2016) 'Detection of maternal colonization of group B streptococcus by PCR targeting Cfb and ScpB genes'. *Journal of Microbiology, Biotechnology and Food Sciences*, pp. 713-716 1.

Fung, A. (n.d.) 'Jasco J-810 spectropolarimeter basic operating manual'. Jasco, pp. 1-22.

Gao, X.Y., Zhi, X.Y., Li, H.W., Klenk, H.P. and Li, W.J. (2014) 'Comparative genomics of the bacterial genus streptococcus illuminates evolutionary implications of species groups', *Plos One*, 9(6).

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) 'Protein identification and analysis tools on the ExPASy server'. *The Proteomics Protocols Handbook*, pp. 571-607.

GenePOC (2018) *Clinical tests*. Available at: <http://www.genepoc-diagnostics.com/clinical-test/> (Accessed: 22nd February 2018).

Ghaddar, N., Alfouzan, W., Anastasiadis, E., Al Jiser, T., Itani, S.E., Dernaika, R., Eid, T., Ghaddar, A., Charafeddine, A., Dhar, R. and Hajj, H. (2014) 'Evaluation of chromogenic medium and direct latex agglutination test for detection of group B streptococcus in vaginal specimens from pregnant women in Lebanon and Kuwait', *Journal of Medical Microbiology*, 63, pp. 1395-1399.

Gholizadeh, A., Faizi, M.H. and Kohnehrouz, B.B. (2010) 'Induced expression of EcoRI endonuclease as an active maltose-binding fusion protein in Escherichia coli', *Microbiology*, 79(2), pp. 167-172.

Gholizadeh, A. and Kohnehrouz, B.B. (2011) 'Functional fusion expression of sunflower multicystatin in E. coli and its comparison with a single domain cystatin', *Indian Journal of Biochemistry & Biophysics*, 48(6), pp. 375-379.

Glaser, P., Rusniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Msadek, T., Zouine, M., Couve, E., Lalioui, L., Poyart, C., Trieu-Cuot, P. and Kunst, F. (2002) 'Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease', *Molecular Microbiology*, 45(6), pp. 1499-1513.

GLS Biotech *SnapGene Viewer*, *SnapGene software*. Available at: [snapgene.com](http://snapgene.com).

Goodrich, J.S. and Miller, M.B. (2007) 'Comparison of culture and 2 real-time polymerase chain reaction assays to detect group B Streptococcus during antepartum screening', *Diagnostic Microbiology and Infectious Disease*, 59(1), pp. 17-22.

Graceffa, P., Jancso, A. and Mabuchi, K. (1992) 'Modification of acidic residues normalizes sodium dodecyl-sulfate polyacrylamide-gel electrophoresis of caldesmon and other proteins that migrate anomalously', *Archives of Biochemistry and Biophysics*, 297(1), pp. 46-51.

Greenberg, D.N., Ascher, D.P., Yoder, B.A., Hensley, D.M., Heiman, H.S. and Keith, J.F. (1995) 'Sensitivity and specificity of rapid diagnostic tests for detection of group B streptococcal antigen in bacteremic neonates', *Journal of Clinical Microbiology*, 33(1), pp. 193-198.

Greenfield, E.A. (2014) *Generating monoclonal antibodies*. Cold Spring Harbor Laboratory Press.

Greenfield, N.J. (2006) 'Using circular dichroism spectra to estimate protein secondary structure', *Nature Protocols*, 1(6), pp. 2876-2890.

Grodzki, A.C., Berenstein, E., Oliver, C.e. and Jamur, M.e. 588 (2010) 'Antibody purification: Ammonium sulfate fractionation or gel filtration, immunocytochemical. Methods and Protocols (3rd edition). Methods in Molecular Biology'. Humana Press, a part of Springer Science+Business Media, LLC, pp. 15-26.

Guilbert, J., Levy, C., Cohen, R., Delacourt, C., Renolleau, S., Flamant, C. and Bacterial Meningitis, G. (2010) 'Late and ultra late onset Streptococcus B meningitis: clinical and bacteriological data over 6 years in France', *Acta Paediatrica*, 99(1), pp. 47-51.

Hakansson, S., Kallen, K., Bullarbo, M., Holmgren, P.A., Bremme, K., Larsson, A., Norman, M., Noren, H., Ortmark-Wrede, C., Pettersson, K., Saltvedt, S., Sondell, B., Tokarska, M., von Vultee, A. and Jacobsson, B. (2014) 'Real-time PCR-assay in the delivery suite for determination of group B streptococcal colonization in a setting with risk-based antibiotic prophylaxis', *Journal of Maternal-Fetal & Neonatal Medicine*, 27(4), pp. 328-332.

Hamidi, S.V., Ghourchian, H. and Tavoosidana, G. (2015) 'Real-time detection of H5N1 influenza virus through hyperbranched rolling circle amplification', *Analyst*, 140(5), pp. 1502-1509.

Hamorsky, K.T., Ensor, C.M., Wei, Y. and Daunert, S. (2008) 'A bioluminescent molecular switch for glucose', *Angewandte Chemie-International Edition*, 47(20), pp. 3718-3721.

Hancock, D. and O'Reilly, N. 295 (2004) 'Synthetic peptides as antigens for antibody production'. Methods in Molecular Biology, Immunochemical Protocols, Third Edition pp. 13-25.

Hang, B.J., Pan, J.F., Ni, D.S., Zheng, Q., Zhang, X., Cai, J., Huang, L., Wei, P.L. and Xu, Z.N. (2016) 'High-level production of aquaporin Z in Escherichia coli using maltose-binding protein/polyhistidine dual-affinity tag fusion system', *Process Biochemistry*, 51(5), pp. 599-606.

Hauge, M., Jespersgaard, C., Poulsen, K. and Kilian, M. (1996) 'Population structure of Streptococcus agalactiae reveals an association between specific evolutionary lineages and putative virulence factors but not disease', *Infection and Immunity*, 64(3), pp. 919-925.

- Head, J.F., Inouye, S., Teranishi, K. and Shimomura, O. (2000) 'The crystal structure of the photoprotein aequorin at 2.3 angstrom resolution', *Nature*, 405(6784), pp. 372-376.
- Heath, P.G.T., Balfour, G., Weisner, A.M., Efstratiou, A., Lamagni, T.L., Tighe, H., O'Connell, L.A.F., Cafferkey, M., Verlander, N.Q., Nicoll, A., McCartney, A.C. and Grp, P.G.W. (2004) 'Group B streptococcal disease in UK and Irish infants younger than 90 days', *Lancet*, 363(9405), pp. 292-294.
- Hempen, C. and Karst, U. (2006) 'Labeling strategies for bioassays', *Analytical and Bioanalytical Chemistry*, 384(3), pp. 572-583.
- Hildebrand, P.W., Preissner, R. and Frommel, C. (2004) 'Structural features of transmembrane helices', *Febs Letters*, 559(1-3), pp. 145-151.
- Hladilkova, J., Callisen, T.H. and Lund, M. (2016) 'Lateral protein-protein interactions at hydrophobic and charged surfaces as a function of pH and salt concentration', *Journal of Physical Chemistry B*, 120(13), pp. 3303-3310.
- Homer, C.S.E., Scarf, V., Catling, C. and Davis, D. (2014) 'Culture-based versus risk-based screening for the prevention of group B streptococcal disease in newborns: A review of national guidelines', *Women and Birth*, 27(1), pp. 46-51.
- Hopp, T.P., Prickett, K.S., Price, V.L., Libby, R.T., March, C.J., Cerretti, D.P., Urdal, D.L. and Conlon, P.J. (1988) 'A short polypeptide marker sequence useful for recombinant protein identification and purification', *Bio-Technology*, 6(10), pp. 1204-1210.
- Hoppe-Seyler, F., Crnkovic-Mertens, I., Tomai, E. and Butz, K. (2004) 'Peptide aptamers: Specific inhibitors of protein function', *Current Molecular Medicine*, 4(5), pp. 529-538.
- Howes, P.D., Chandrawati, R. and Stevens, M.M. (2014) 'Colloidal nanoparticles as advanced biological sensors', *Science*, 346(6205), pp. 53-+.
- Hsueh, P.R., Teng, L.J., Lee, L.N., Ho, S.W., Yang, P.C. and Luh, K.T. (2001) 'High incidence of erythromycin resistance among clinical isolates of *Streptococcus agalactiae* in Taiwan', *Antimicrobial Agents and Chemotherapy*, 45(11), pp. 3205-3208.
- Hu, L.M., Luo, K., Xia, J., Xu, G.M., Wu, C.H., Han, J.J., Zhang, G.G., Liu, M. and Lai, W.H. (2017) 'Advantages of time-resolved fluorescent nanobeads compared with fluorescent submicrospheres, quantum dots, and colloidal gold as label in lateral flow assays for detection of ractopamine', *Biosensors & Bioelectronics*, 91, pp. 95-103.
- Hughes, R.G., Brocklehurst, P., Steer, P.J., Heath, P., Stenson, B.M., of, o.b.o.t.R.C. and Gynaecologists, O.a. (2017) 'Prevention of early-onset neonatal group B streptococcal disease. Green-top guideline No. 36'. *BJOG: An International Journal of Obstetrics & Gynaecology*.

Inouye, S. (2004) 'Blue fluorescent protein from the calcium-sensitive photoprotein aequorin is a heat resistant enzyme, catalyzing the oxidation of coelenterazine', *Febs Letters*, 577(1-2), pp. 105-110.

Ippolito, D.L., James, W.A., Tinnemore, D., Huang, R.R., Dehart, M.J., Williams, J., Wingerd, M.A. and Demons, S.T. (2010) 'Group B Streptococcus serotype prevalence in reproductive-age women at a tertiary care military medical center relative to global serotype distribution', *Bmc Infectious Diseases*, 10.

Jain, R.G., Rusch, S.L. and Kendall, D.A. (1994) 'Signal peptide cleavage regions - functional limits on length and topological implications', *Journal of Biological Chemistry*, 269(23), pp. 16305-16310.

Jantzen, J., Tzanova, I., Witton, P.K. and Klein, A.M. (1989) 'Rectal pH in children', *Canadian Journal of Anaesthesia-Journal Canadien D Anesthesie*, 36(6), pp. 665-667.

Jauregui, F., Carton, M., Panel, P., Foucaud, P., Butel, M.J. and Doucet-Populaire, F. (2004) 'Effects of intrapartum penicillin prophylaxis on intestinal bacterial colonization in infants', *Journal of Clinical Microbiology*, 42(11), pp. 5184-5188.

Jauset-Rubio, M., Svobodova, M., Mairal, T., McNeil, C., Keegan, N., El-Shahawi, M.S., Bashammakh, A.S., Alyoubi, A.O. and O'Sullivan, C.K. (2016a) 'Aptamer lateral flow assays for ultrasensitive detection of beta-conglutinin combining recombinase polymerase amplification and tailed primers', *Analytical Chemistry*, 88(21), pp. 10701-10709.

Jauset-Rubio, M., Svobodova, M., Mairal, T., McNeil, C., Keegan, N., Saeed, A., Abbas, M.N., El-Shahawi, M.S., Bashammakh, A.S., Alyoubi, A.O. and O'Sullivan, C.K. (2016b) 'Ultrasensitive, rapid and inexpensive detection of DNA using paper based lateral flow assay', *Scientific Reports*, 6.

Ji, Y.D., McLandsborough, L., Kondagunta, A. and Cleary, P.P. (1996) 'C5a peptidase alters clearance and trafficking of group A streptococci by infected mice', *Infection and Immunity*, 64(2), pp. 503-510.

Johnson, D.R. and Ferrieri, P. (1984) 'Group-B streptococcal-IBC protein antigen - Distribution of 2 determinants in wild-type strains of common serotypes', *Journal of Clinical Microbiology*, 19(4), pp. 506-510.

Johri, A.K., Margarit, I., Broenstrup, M., Brettoni, C., Hua, L., Gygi, S.P., Telford, J.L., Grandi, G. and Paoletti, L.C. (2007) 'Transcriptional and proteomic profiles of group B Streptococcus type V reveal potential adherence proteins associated with high-level invasion', *Infection and Immunity*, 75(3), pp. 1473-1483.

Johri, A.K., Paoletti, L.C., Glaser, P., Dua, M., Sharma, P.K., Grandi, G. and Rappuoli, R. (2006) 'Group B Streptococcus: global incidence and vaccine development', *Nature Reviews Microbiology*, 4(12), pp. 932-942.

Joneja, A. and Huang, X.H. (2011) 'Linear nicking endonuclease-mediated strand-displacement DNA amplification', *Analytical Biochemistry*, 414(1), pp. 58-69.

- Jones, D.T. (1999) 'Protein secondary structure prediction based on position-specific scoring matrices', *Journal of Molecular Biology*, 292(2), pp. 195-202.
- Jones, H.E., Holland, I.B. and Campbell, A.K. (2002) 'Direct measurement of free Ca<sup>2+</sup> shows different regulation of Ca<sup>2+</sup> between the periplasm and the cytosol of *Escherichia coli*', *Cell Calcium*, 32(4), pp. 183-192.
- Jones, K., Hibbert, F. and Keenan, M. (1999) 'Glowing jellyfish, luminescence and a molecule called coelenterazine', *Trends in Biotechnology*, 17(12), pp. 477-481.
- Jones, R., Bousfield, D., Willitis, I., Cole, H. and Arber, M. (2015) 'Xpert GBS test for the intrapartum detection of group B streptococcus. Medtech innovation briefing [MIB28]'. Medtech innovation briefing. National Institute for Health and Care Excellence.
- Jorgensen, C.S., Uldum, S.A., Sorensen, J.F., Skovsted, I.C., Otte, S. and Elverdal, P.L. (2015) 'Evaluation of a new lateral flow test for detection of *Streptococcus pneumoniae* and *Legionella pneumophila* urinary antigen', *Journal of Microbiological Methods*, 116, pp. 33-36.
- Juncker, A.S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. and Krogh, A. (2003) 'Prediction of lipoprotein signal peptides in Gram-negative bacteria', *Protein Science*, 12(8), pp. 1652-1662.
- Kaambwa, B., Bryan, S., Gray, J., Milner, P., Daniels, J., Khan, K.S. and Roberts, T.E. (2010) 'Cost-effectiveness of rapid tests and other existing strategies for screening and management of early-onset group B streptococcus during labour', *Bjog-an International Journal of Obstetrics and Gynaecology*, 117(13), pp. 1616-1627.
- 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(8), pp. 1668-1674.
- Karanicolas, J., Com, J.E., Chen, I., Joachimiak, L.A., Dym, O., Peck, S.H., Albeck, S., Unger, T., Hu, W., Liu, G., Delbecq, S., Montelione, G.T., Spiegel, C.P., Liu, D.R. and Baker, D. (2011) 'A de novo protein binding pair by computational design and directed evolution', *Molecular Cell*, 42(2), pp. 250-260.
- Karplus, P.A. and Schulz, G.E. (1985) 'Prediction of chain flexibility in proteins - A tool for the selection of peptide aptamers', *Naturwissenschaften*, 72(4), pp. 212-213.
- Keefe, G.P. (1997) 'Streptococcus agalactiae mastitis: A review', *Canadian Veterinary Journal-Revue Veterinaire Canadienne*, 38(7), pp. 429-437.
- Kelly, S.M. and Price, N.C. (2000) 'The use of circular dichroism in the investigation of protein structure and function', *Current Protein & Peptide Science*, 1(4), pp. 349-384.
- Kenyon, S., Pike, K., Jones, D.R., Brocklehurst, P., Marlow, N., Salt, A. and Taylor, D.J. (2008) 'Childhood outcomes after prescription of antibiotics to pregnant women

with spontaneous preterm labour: 7-year follow-up of the ORACLE II trial', *Lancet*, 372(9646), pp. 1319-1327.

Kikin, O., D'Antonio, L. and Bagga, P.S. (2006) 'QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences', *Nucleic Acids Research*, 34, pp. W676-W682.

Kimura, K., Suzuki, S., Wachino, J., Kurokawa, H., Yamane, K., Shibata, N., Nagano, N., Kato, H., Shibayama, K. and Arakawa, Y. (2008) 'First molecular characterization of group B streptococci with reduced penicillin susceptibility', *Antimicrobial Agents and Chemotherapy*, 52(8), pp. 2890-2897.

Klinzing, D., Ishmael, N., Dunning Hotopp, J., Lin, M., Tetellin, H. and Puopolo, K. (2009) 'Transcriptional responses to elevated temperature in *Streptococcus agalactiae*' *Gene Expression Omnibus*.

Koczula, K.M. and Gallotta, A. (2016) 'Lateral flow assays', *Biosensor Technologies for Detection of Biomolecules*, 60(1), pp. 111-120.

Kohler, G., Howe, S.C. and Milstein, C. (1976) 'Fusion between immunoglobulin-secreting nonsecreting myeloma cell lines', *European Journal of Immunology*, 6(4), pp. 292-295.

Kohler, G. and Milstein, C. (1975) 'Continuous cultures of fused cells secreting antibody of predefined specificity', *Nature*, 256(5517), pp. 495-497.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001) 'Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes', *Journal of Molecular Biology*, 305(3), pp. 567-580.

Krolov, K., Frolova, J., Tudoran, O., Suhorutsenko, J., Lehto, T., Sibul, H., Mager, I., Laanpere, M., Tulp, I. and Langell, O. (2014) 'Sensitive and rapid detection of *Chlamydia trachomatis* by recombinase polymerase amplification directly from urine samples', *Journal of Molecular Diagnostics*, 16(1), pp. 127-135.

Kurose, K., Inouye, S., Sakaki, Y. and Tsuji, F.I. (1989) 'Bioluminescence of the Ca<sup>2+</sup>-binding photoprotein aequorin after cysteine modification', *Proceedings of the National Academy of Sciences of the United States of America*, 86(1), pp. 80-84.

Kusser, W. (2000) 'Chemically modified nucleic acid aptamers for in vitro selections: evolving evolution', *Journal of biotechnology*, 74(1), pp. 27-38.

Ladner, R.C. (1995) 'Constrained peptides as binding entities', *Trends in Biotechnology*, 13(10), pp. 426-430.

Laemmli, U.K. (1970) 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4'. *Nature*, pp. 680-685.

Lamagni, T.L., Keshishian, C., Efstratiou, A., Guy, R., Henderson, K.L., Broughton, K. and Sheridan, E. (2013) 'Emerging trends in the epidemiology of invasive group B streptococcal disease in England and Wales, 1991-2010', *Clinical Infectious Diseases*, 57(5), pp. 682-688.

- Lammler, C., Abdulmawjood, A. and Weiss, R. (1998) 'Properties of serological group B streptococci of dog, cat and monkey origin', *Journal of Veterinary Medicine Series B-Infectious Diseases and Veterinary Public Health*, 45(9), pp. 561-566.
- Lancefield, R.C., McCarty, M. and Everly, W.N. (1975) 'Multiple mouse-protective antibodies directed against group B streptococci - Special reference to antibodies effective against protein antigens', *Journal of Experimental Medicine*, 142(1), pp. 165-179.
- Landwehr-Kenzel, S. and Henneke, P. (2014) 'Interaction of *Streptococcus agalactiae* and cellular innate immunity in colonization and disease', *Frontiers in Immunology*, 5.
- Lang, S.H. and Palmer, M. (2003) 'Characterization of *Streptococcus agalactiae* CAMP factor as a pore-forming toxin', *Journal of Biological Chemistry*, 278(40), pp. 38167-38173.
- Lang, S.H., Xue, J., Guo, Z.W. and Palmer, M. (2007) '*Streptococcus agalactiae* CAMP factor binds to GPI-anchored proteins', *Medical Microbiology and Immunology*, 196(1), pp. 1-10.
- Lateef, S.S., Gupta, S., Jayathilka, L.P., Krishnanchettiar, S., Huang, J.-S. and Lee, B.-S. (2007) 'An improved protocol for coupling synthetic peptides to carrier proteins for antibody production using DMF to solubilize peptides'. *Journal of Biomolecular Techniques*, pp. 173-176 3.
- Legendre, D., Vucic, B., Hougardy, V., Girboux, A.L., Henrioul, C., Van Haute, J., Soumillon, P. and Fastrez, J. (2002) 'TEM-1 beta-lactamase as a scaffold for protein recognition and assay', *Protein Science*, 11(6), pp. 1506-1518.
- Lewis, J.C., Lopez-Moya, J.J. and Daunert, S. (2000) 'Bioluminescence and secondary structure properties of aequorin mutants produced for site-specific conjugation and immobilization', *Bioconjugate Chemistry*, 11(1), pp. 65-70.
- Li, S.L. and Jedrzejewski, M.J. (2001) 'Hyaluronan binding and degradation by *Streptococcus agalactiae* hyaluronate lyase', *Journal of Biological Chemistry*, 276(44), pp. 41407-41416.
- Lin, F.Y.C., Brenner, R.A., Johnson, Y.R., Azimi, P.H., Philips, J.B., Regan, J.A., Clark, P., Weisman, L.E., Rhoads, G.G., Kong, F.H. and Clemens, J.D. (2001) 'The effectiveness of risk-based intrapartum chemoprophylaxis for the prevention of early-onset neonatal group B streptococcal disease', *American Journal of Obstetrics and Gynecology*, 184(6), pp. 1204-1210.
- Lindahl, G., Stalhammar-Carlemalm, M. and Areschoug, T. (2005) 'Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens', *Clinical Microbiology Reviews*, 18(1), pp. 102-+.
- Lipman, N.S., Jackson, L.R., Trudel, L.J. and Weis-Garcia, F. (2005) 'Monoclonal versus polyclonal antibodies: Distinguishing characteristics, applications, and information resources', *Ilar Journal*, 46(3), pp. 258-268.

Lisher, J.P. and Giedroc, D.P. (2013) 'Manganese acquisition and homeostasis at the host-pathogen interface', *Frontiers in Cellular and Infection Microbiology*, 3.

Liu, H.Y., Saxena, A., Sidhu, S.S. and Wu, D.H. (2017) 'Fc engineering for developing therapeutic bispecific antibodies and novel scaffolds', *Frontiers in Immunology*, 8.

Lizardi, P.M., Huang, X.H., Zhu, Z.R., Bray-Ward, P., Thomas, D.C. and Ward, D.C. (1998) 'Mutation detection and single-molecule counting using isothermal rolling-circle amplification', *Nature Genetics*, 19(3), pp. 225-232.

Loening, A.M., Fenn, T.D., Wu, A.M. and Gambhir, S.S. (2006) 'Consensus guided mutagenesis of Renilla luciferase yields enhanced stability and light output', *Protein Engineering Design & Selection*, 19(9), pp. 391-400.

Loesche, W.J. (1986) 'Role of Streptococcus mutans in human dental decay', *Microbiological Reviews*, 50(4), pp. 353-380.

Lukacs, S.L., Schoendorf, K.C. and Schuchat, A. (2004) 'Trends in sepsis-related neonatal mortality in the United States, 1985-1998', *Pediatric Infectious Disease Journal*, 23(7), pp. 599-603.

Mabey, D., Peeling, R.W., Ustianowski, A. and Perkins, M.D. (2004) 'Diagnostics for the developing world', *Nature Reviews Microbiology*, 2(3), pp. 231-240.

Madhi, S.A., Cutland, C.L., Jose, L., Koen, A., Govender, N., Wittke, F., Olugbosi, M., Sobanjo-ter Meulen, A., Baker, S., Dull, P.M., Narasimhan, V. and Slobod, K. (2016) 'Safety and immunogenicity of an investigational maternal trivalent group B streptococcus vaccine in healthy women and their infants: a randomised phase 1b/2 trial', *Lancet Infectious Diseases*, 16(8), pp. 923-934.

mapchart.net (2017) *Simple world map*. Available at: [mapchart.net/world.html](http://mapchart.net/world.html) (Accessed: 6th August 2017).

Markaryan, A., Zaborina, O., Punj, V. and Chakrabarty, A.M. (2001) 'Adenylate kinase as a virulence factor of Pseudomonas aeruginosa', *Journal of Bacteriology*, 183(11), pp. 3345-3352.

Mascini, M., Palchetti, I. and Tombelli, S. (2012) 'Nucleic acid and peptide aptamers: Fundamentals and bioanalytical aspects', *Angewandte Chemie-International Edition*, 51(6), pp. 1316-1332.

McCartney, A.C., Heath, P., Hughes, R., Misfud, A. and Price, E. 109 (2002) 'The ORACLE I and II randomised trial results do not remove the need for prophylaxis against early onset Group B streptococcal disease'. *BJOG*, pp. 1419-1420 12.

McKeague, M. and DeRosa, M.C. Article ID 748913 (2012) 'Challenges and opportunities for small molecule aptamer development'. *Journal of Nucleic Acids*, pp. 1-20.

Melin, P. and Efstratiou, A. (2013) 'Group B streptococcal epidemiology and vaccine needs in developed countries', *Vaccine*, 31, pp. D31-D42.

- Menger, M., Yarman, A., Erdosy, J., Yildiz, H.B., Gyurcsanyi, R.E. and Scheller, F.W. (2016) 'MIPs and aptamers for recognition of proteins in biomimetic sensing', *Biosensors-Basel*, 6(3).
- Mercier, M.C., Dontenwill, M. and Choulier, L. (2017) 'Selection of nucleic acid aptamers targeting tumor cell-surface protein biomarkers', *Cancers*, 9(6).
- Meridian Bioscience (2018) *Illumigene group B streptococcus by Meridian Bioscience* (Accessed: 22nd February 2018).
- Mhalu, F.S. (1976) 'Infection with *Streptococcus agalactiae* in a London hospital', *Journal of Clinical Pathology*, 29(4), pp. 309-312.
- Minitab Inc (2014) 'Minitab statistical software, release 17 for Windows'. State College, Pennsylvania. Minitab ® is a registered trademark of Minitab, Inc.
- Mofatt, B.A. and Ashiharab, H. (2002) 'Purine and pyrimidine nucleotide synthesis and metabolism'. *The Arabidopsis Book*, pp. 1-20.
- MolView (2015) *MolView v2.4*. Available at: molview.org (Accessed: 13th September 2017).
- Montague, N.S., Cleary, T.J., Martinez, O.V. and Procop, G.W. (2008) 'Detection of group B streptococci in Lim broth by use of group B *Streptococcus* peptide nucleic acid fluorescent in situ hybridization and selective and nonselective agars', *Journal of Clinical Microbiology*, 46(10), pp. 3470-3472.
- Mousavi, S.M., Hosseini, S.M., Mashouf, R.Y. and Arabestani, M.R. 54 (2016) 'Identification of group B streptococci using 16S rRNA, *cfb*, *scpB*, and *atr* genes in pregnant women by PCR'. *Acta Medica Iranica* 12.
- Murakami, T., Sumaoka, J. and Komiyama, M. (2009) 'Sensitive isothermal detection of nucleic-acid sequence by primer generation-rolling circle amplification', *Nucleic Acids Research*, 37(3).
- Murphy, K. (2011) 'Innate immunity: The first lines of defense. Janeway's Immunobiology, 8th edition'. Garland Science, pp. 37-73.
- Nagel-Wolfrum, K., Buerger, C., Wittig, I., Butz, K., Hoppe-Seyler, F. and Groner, B. (2004) 'The interaction of specific peptide aptamers with the DNA binding domain and the dimerization domain of the transcription factor Stat3 inhibits transactivation and induces apoptosis in tumor cells', *Molecular Cancer Research*, 2(3), pp. 170-182.
- Nair, G., Rebolledo, M., White, A.C., Crannell, Z., Richards-Kortum, R.R., Pinilla, A.E., Ramirez, J.D., Lopez, M.C. and Castellanos-Gonzalez, A. (2015) 'Detection of *Entamoeba histolytica* by recombinase polymerase amplification', *American Journal of Tropical Medicine and Hygiene*, 93(3), pp. 591-595.
- Nakanishi, K., Sakiyama, T. and Imamura, K. (2001) 'On the adsorption of proteins on solid surfaces, a common but very complicated phenomenon', *Journal of Bioscience and Bioengineering*, 91(3), pp. 233-244.

Nallur, G., Luo, C.H., Fang, L.H., Cooley, S., Dave, V., Lambert, J., Kukanskis, K., Kingsmore, S., Lasken, R. and Schweitzer, B. (2001) 'Signal amplification by rolling circle amplification on DNA microarrays', *Nucleic Acids Research*, 29(23), pp. art. no.-e118.

Navarre, W.W. and Schneewind, O. (1994) 'Proteolytic cleavage and cell-wall anchoring at the LPXTG motif of surface proteins in gram positive bacteria', *Molecular Microbiology*, 14(1), pp. 115-121.

New England BioLabs (2017a) *PCR protocol for Taq DNA polymerase with standard Taq buffer (M0273)*. Available at: <https://www.neb.com/protocols/1/01/01/taq-dna-polymerase-with-standard-taq-buffer-m0273> (Accessed: 1st September 2017).

New England BioLabs (2017b) 'pMAL™ protein fusion & purification system. Instruction manual'. Protein Expression and Analysis, pp. 1-36.

Nguyen, M.T., Krupa, M., Koo, B.K., Song, J.A., Vu, T.T.T., Do, B.H., Nguyen, A.N., Seo, T., Yoo, J., Jeong, B., Jin, J., Lee, K.J., Oh, H.B. and Choe, H. (2016) 'Prokaryotic soluble overexpression and purification of human VEGF165 by fusion to a maltose binding protein tag', *Plos One*, 11(5).

Nizet, V., Gibson, R.L., Chi, E.Y., Framson, P.E., Hulse, M. and Rubens, C.E. (1996) 'Group B streptococcal beta-hemolysin expression is associated with injury of lung epithelial cells', *Infection and Immunity*, 64(9), pp. 3818-3826.

Nobbs, A.H., Rosini, R., Rinaudo, C.D., Maione, D., Grandi, G. and Telford, J.L. (2008) 'Sortase a utilizes an ancillary protein anchor for efficient cell wall anchoring of pili in *Streptococcus agalactiae*', *Infection and Immunity*, 76(8), pp. 3550-3560.

Nogacka, A., Salazar, N., Suarez, M., Milani, C., Arboleya, S., Solis, G., Fernandez, N., Alaez, L., Hernandez-Barranco, A.M., de los Reyes-Gavilan, C.G., Ventura, M. and Gueimonde, M. (2017) 'Impact of intrapartum antimicrobial prophylaxis upon the intestinal microbiota and the prevalence of antibiotic resistance genes in vaginally delivered full-term neonates', *Microbiome*, 5.

Nomura, M., Inouye, S., Ohmiya, Y. and Tsuji, F.I. (1991) 'A C-terminal proline is required for bioluminescence of the Ca<sup>2+</sup> binding photoprotein, aequorin', *Febs Letters*, 295(1-3), pp. 63-66.

Nord, K., Nilsson, J., Nilsson, B., Uhlen, M. and Nygren, P.A. (1995) 'A combinational library of an alpha-helical bacterial receptor domain', *Protein Engineering*, 8(6), pp. 601-608.

Nuccitelli, A., Rinaudo, C.D. and Maione, D. 3 (2015) 'Group B *Streptococcus* vaccine: state of the art'. *Therapeutic Advances in Vaccines*, pp. 76-90 3.

Nygren, P.A. and Skerra, A. (2004) 'Binding proteins from alternative scaffolds', *Journal of Immunological Methods*, 290(1-2), pp. 3-28.

- Ospina-Villa, J.D., Zamorano-Carrillo, A., Castanon-Sanchez, C.A., Ramirez-Moreno, E. and Marchat, L.A. (2016) 'Aptamers as a promising approach for the control of parasitic diseases', *Brazilian Journal of Infectious Diseases*, 20(6), pp. 610-618.
- Otaguiri, E.S., Morguette, A.E.B., Tavares, E.R., dos Santos, P.M.C., Morey, A.T., Cardoso, J.D., Perugini, M.R.E., Yamauchi, L.M. and Yamada-Ogatta, S.F. (2013) 'Commensal *Streptococcus agalactiae* isolated from patients seen at University Hospital of Londrina, Parana, Brazil: capsular types, genotyping, antimicrobial susceptibility and virulence determinants', *Bmc Microbiology*, 13.
- Ott, S.J., Musfeldt, M., Ullmann, U., Hampe, J. and Schreiber, S. (2004) 'Quantification of intestinal bacterial populations by real-time PCR with a universal primer set and minor groove binder probes: a global approach to the enteric flora', *Journal of Clinical Microbiology*, 42(6), pp. 2566-2572.
- Pacha, A., Cian, R.L., Bonofiglio, L., Solari, M., Strada, V., Suarez, M., Vigliarolo, L., Tersigni, C., Mollerach, M. and Lopardo, H. (2017) 'Group B streptococcal necrotizing pneumonia in a diabetic adult patient', *Revista Argentina De Microbiologia*, 49(2), pp. 139-141.
- Palmer, I. and Wingfield, P. Unit 6.3 (2004) 'Preparation and extraction of insoluble (inclusion-body) proteins from *Escherichia coli*'. *Current Protocols in Protein Science*.
- Pancholi, V. and Fischetti, V.A. (1998) 'alpha-enolase, a novel strong plasmin(ogen) binding protein on the surface of pathogenic streptococci', *Journal of Biological Chemistry*, 273(23), pp. 14503-14515.
- Panikkath, D., Tantrachoti, P., Panikkath, R. and Nugent, K. 29 (2016) 'Streptococcus agalactiae pyomyositis in diabetes mellitus'. *Baylor University Medical Center Proceedings*, pp. 290-291 3.
- Paoletti, L.C. and Kasper, D.L. (2002) 'Conjugate vaccines against group B *Streptococcus* types IV and VII', *Journal of Infectious Diseases*, 186(1), pp. 123-126.
- Paoletti, L.C., Wessels, M.R., Michon, F., Difabio, J., Jennings, H.J. and Kasper, D.L. (1992) 'Group-B streptococcus type-II polysaccharide-tetanus toxoid conjugate vaccine', *Infection and Immunity*, 60(10), pp. 4009-4014.
- Paoletti, L.C., Wessels, M.R., Rodewald, A.K., Shroff, A.A., Jennings, H.J. and Kasper, D.L. 62 (1994) 'Neonatal mouse protection against infection with multiple group B streptococcal (GBS) serotypes by maternal immunization with a tetravalent GBS polysaccharide-tetanus toxoid conjugate vaccine'. *Infection and Immunity*, American Society for Microbiology, pp. 3236-3243 8.
- Parker, J.M.R., Guo, D. and Hodges, R.S. (1986) 'New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data - Correlation of predicted surface residues with antigenicity and X-ray derived accessible sites', *Biochemistry*, 25(19), pp. 5425-5432.
- Pellequer, J.L., Westhof, E. and Vanregemortel, M.H.V. (1993) 'Correlation between the location of antigenic sites and the prediction of turns in proteins', *Immunology Letters*, 36(1), pp. 83-100.

- Peltroche-Llacsahuanga, H., Fiandaca, M.J., von Oy, S., Lutticken, R. and Haase, G. (2010) 'Rapid detection of *Streptococcus agalactiae* from swabs by peptide nucleic acid fluorescence in situ hybridization', *Journal of Medical Microbiology*, 59(2), pp. 179-184.
- Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) 'SignalP 4.0: discriminating signal peptides from transmembrane regions', *Nature Methods*, 8(10), pp. 785-786.
- Petri, W. (2011) 'Penicillins, cephalosporins and other B-lactam antibiotics. Goodman & Gilman's The pharmacological basis of therapeutics. 12th edition' Brunton, L., Chabner, B. and Knollman, B. New York.
- Platt, M.W., McLaughlin, J.C., Gilson, G.J., Wellhoner, M.F. and Nims, L.J. (1995) 'Increased recovery of group-B streptococcus by the inclusion of rectal culturing and enrichment', *Diagnostic Microbiology and Infectious Disease*, 21(2), pp. 65-68.
- Pritzlaff, C.A., Chang, J.C.W., Kuo, S.P., Tamura, G.S., Rubens, C.E. and Nizet, V. (2001) 'Genetic basis for the beta-haemolytic/cytolytic activity of group B *Streptococcus*', *Molecular Microbiology*, 39(2), pp. 236-247.
- Pryor, K.D. and Leiting, B. (1997) 'High-level expression of soluble protein in *Escherichia coli* using a His(6)-tag and maltose-binding-protein double-affinity fusion system', *Protein Expression and Purification*, 10(3), pp. 309-319.
- Public Health England (2015a) 'Enriched culture medium test for group B streptococcus infection. Position paper'.
- Public Health England 3 (2015b) 'UK standards for microbiology investigations - Detection of carriage of group B streptococci'. Standards Unit, Bacteriology, Microbiology Services, Public Health England.
- Public Health England (2017) 'Antibiotic awareness key messages 2017'. PHE Publications, pp. 1-15.
- Puopolo, K.M., Madoff, L.C. and Eichenwald, E.C. (2005) 'Early-onset group B streptococcal disease in the era of maternal screening', *Pediatrics*, 115(5), pp. 1240-1246.
- Quidel (2018) *AmpliVue GBS assay*. Available at: <http://www.quidel.com/molecular-diagnostics/testing-strep/amplivue-gbs-assay> (Accessed: 22nd February 2018).
- Radom, F., Jurek, P.M., Mazurek, M.P., Otlewski, J. and Jelen, F. (2013) 'Aptamers: Molecules of great potential', *Biotechnology Advances*, 31(8), pp. 1260-1274.
- Rahman, O., Cummings, S.P., Harrington, D.J. and Sutcliffe, I.C. (2008) 'Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria', *World Journal of Microbiology & Biotechnology*, 24(11), pp. 2377-2382.
- Rao, G.G., Nartey, G., McAree, T., O'Reilly, A., Hiles, S., Lee, T., Wallace, S., Batura, R., Khanna, P., Abbas, H., Tilsed, C., Nicholl, R., Lamagni, T. and Bassett, P.

(2017) 'Outcome of a screening programme for the prevention of neonatal invasive early-onset group B Streptococcus infection in a UK maternity unit: an observational study', *Bmj Open*, 7(4).

Ravalli, A., da Rocha, C.G., Yamanaka, H. and Marrazza, G. (2015) 'A label-free electrochemical affisensor for cancer marker detection: The case of HER2', *Bioelectrochemistry*, 106, pp. 268-275.

Raymond, J., Armengaud, J.B., Lambe, C., Tchetchoua, A., Moulin, F., Lebon, P., Poyart, C. and Gendrel, D. (2007) 'Late-onset neonatal infections caused by group B Streptococcus associated with viral infection', *Pediatric Infectious Disease Journal*, 26(10), pp. 963-965.

Renkawitz, J., Lademann, C.A. and Jentsch, S. (2014) 'DNA damage mechanisms and principles of homology search during recombination', *Nature Reviews Molecular Cell Biology*, 15(6), pp. 369-383.

Reuter, J.S. and Mathews, D.H. (2010) 'RNAstructure: software for RNA secondary structure prediction and analysis', *Bmc Bioinformatics*, 11.

Riedewald, S., Kreutzmann, I.M., Heinze, T. and Saling, E. (1990) 'Vaginal and cervical pH in normal-pregnancy and pregnancy complicated by preterm labor', *Journal of Perinatal Medicine*, 18(3), pp. 181-186.

Ring, A., Braun, J.S., Nizet, V., Stremmel, W. and Shenep, J.L. (2000) 'Group B streptococcal beta-hemolysin induces nitric oxide production in murine macrophages', *Journal of Infectious Diseases*, 182(1), pp. 150-157.

Ring, A., Braun, J.S., Pohl, J., Nizet, V., Stremmel, W. and Shenep, J.L. (2002) 'Group B streptococcal beta-hemolysin induces mortality and liver injury in experimental sepsis', *Journal of Infectious Diseases*, 185(12), pp. 1745-1753.

Rodriguez-Granger, J., Spellerberg, B., Asam, D. and Rosa-Fraile, M. (2015) 'Non-haemolytic and non-pigmented group b streptococcus, an infrequent cause of early onset neonatal sepsis', *Pathogens and Disease*, 73(9).

Rodriguez-Iturbe, B. and Batsford, S. (2007) 'Pathogenesis of poststreptococcal glomerulonephritis a century after Clemens von Pirquet', *Kidney International*, 71(11), pp. 1094-1104.

Rojas, A.L., Nagem, R.A.P., Neustroev, K.N., Arand, M., Adamska, M., Eneyskaya, E.V., Kulminskaya, A.A., Garratt, R.C., Golubev, A.M. and Polikarpov, I. (2004) 'Crystal structures of beta-galactosidase from *Penicillium* sp. and its complex with galactose', *Journal of Molecular Biology*, 343(5), pp. 1281-1292.

Ronmark, J., Hansson, M., Nguyen, T., Uhlen, M., Robert, A., Stahl, S. and Nygren, P.A. (2002) 'Construction and characterization of affibody-Fc chimeras produced in *Escherichia coli*', *Journal of Immunological Methods*, 261(1-2), pp. 199-211.

Rose, N.R. (2015) 'Molecular mimicry and clonal deletion: A fresh look', *Journal of Theoretical Biology*, 375, pp. 71-76.

- Rowe, L., Ensor, M., Mehl, R. and Daunert, S. (2010) 'Modulating the bioluminescence emission of photoproteins by in vivo site-directed incorporation of non-natural amino acids', *Acs Chemical Biology*, 5(5), pp. 455-460.
- Roy, A., Kucukural, A. and Zhang, Y. (2010) 'I-TASSER: a unified platform for automated protein structure and function prediction', *Nature Protocols*, 5(4), pp. 725-738.
- Saar-Dover, R., Bitler, A., Nezer, R., Shmuel-Galia, L., Firon, A., Shimoni, E., Trieu-Cuot, P. and Shai, Y. (2012) 'D-alanylation of lipoteichoic acids confers resistance to cationic peptides in group B streptococcus by increasing the cell wall density', *Plos Pathogens*, 8(9).
- Saha, S. and Raghava, G.P.S. (2004) 'BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties', *Artificial Immune Systems, Proceedings*, 3239, pp. 197-204.
- Sahoo, P.R., Sethy, K., Mohapatra, S. and Panda, D. (2016) 'Loop mediated isothermal amplification: An innovative gene amplification technique for animal diseases', *Veterinary World*, 9(5), pp. 465-469.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) 'Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase', *Science*, 239(4839), pp. 487-491.
- Santi, I., Grifantini, R., Jiang, S.-M., Brettoni, C., Grandi, G., Wessels, M.R. and Soriani, M. (2009) 'CsrRS regulates group B streptococcus virulence gene expression in response to environmental pH: a new perspective on vaccine development', *Journal of Bacteriology*, 191(17), pp. 5387-5397.
- Saunders, N. (2009) *An introduction to real-time PCR*. Caister Academic Press.
- Savini, V., Marrollo, R., De Filippis, R., D'Incecco, E., Imperi, M., Pataracchia, M., Alfaroni, G., Fusilli, P., Coclite, E., D'Incecco, C., Fazii, P. and Creti, R. (2016) 'CAMP-negative group B Streptococcus went unrecognized with Cepheid GeneXpert but was detected by Liofilchem (R) Chromatic StrepB', *International Journal of Clinical and Experimental Pathology*, 9(11), pp. 12017-12020.
- Scheffers, D.J. and Pinho, M.G. (2005) 'Bacterial cell wall synthesis: New insights from localization studies', *Microbiology and Molecular Biology Reviews*, 69(4), pp. 585-+.
- Schlehuber, S. and Skerra, A. (2002) 'Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant - a so-called 'anticalin' - using a molecular random approach', *Biophysical Chemistry*, 96(2-3), pp. 213-228.
- Schmidt, K.S., Borkowski, S., Kurreck, J., Stephens, A.W., Bald, R., Hecht, M., Friebe, M., Dinkelborg, L. and Erdmann, V.A. (2004) 'Application of locked nucleic acids to improve aptamer in vivo stability and targeting function', *Nucleic Acids Research*, 32(19), pp. 5757-5765.

Schrag, S.J., Zywicki, S., Farley, M.M., Reingold, A.L., Harrison, L.H., Lefkowitz, L.B., Hadler, J.L., Danila, R., Cieslak, P.R. and Schuchat, A. (2000) 'Group B streptococcal disease in the era of intrapartum antibiotic prophylaxis', *New England Journal of Medicine*, 342(1), pp. 15-20.

Schrodinger, L. (2015) 'The AxPyMOL molecular graphics plugin for microsoft powerPoint, version 1.8'. PyMOL.

Selkirk, C. 244 (2004) 'Ion exchange chromatography'. Protein Purification Protocols, 2nd edition, pp. 125-131.

Sendi, P., Johansson, L. and Norrby-Teglund, A. (2008) 'Invasive group B streptococcal disease in non-pregnant adults - A review with emphasis on skin and soft-tissue infections', *Infection*, 36(2), pp. 100-111.

Seok, Y., Joung, H.A., Byun, J.Y., Jeon, H.S., Shin, S.J., Kim, S., Shin, Y.B., Han, H.S. and Kim, M.G. (2017) 'A paper-based device for performing loop-mediated isothermal amplification with real-time simultaneous detection of multiple DNA targets', *Theranostics*, 7(8), pp. 2220-2230.

Shabayek, S., Abdalla, S. and Abouzeid, A.M. (2014) 'Serotype and surface protein gene distribution of colonizing group B streptococcus in women in Egypt', *Epidemiology and Infection*, 142(1), pp. 208-210.

Shabayek, S., Bauer, R., Mauerer, S., Mizaikoff, B. and Spellerberg, B. (2016) 'A streptococcal NRAMP homologue is crucial for the survival of *Streptococcus agalactiae* under low pH conditions', *Molecular Microbiology*, 100(4), pp. 589-606.

Shay, J.W. (1987) 'Mechanisms of cell fusion and selection in the generation of hybridomas'. *Methods of Hybridoma Formation* pp. 63-75.

Shimomura, O., Kishi, Y. and Inouye, S. (1993) 'The relative rate of aequorin regeneration from apoaequorin and coelenterazine analogs', *Biochemical Journal*, 296, pp. 549-551.

Shirai, A., Matsuyama, A., Yashiroda, Y., Hashimoto, A., Kawamura, Y., Arai, R., Komatsu, Y., Horinouchi, S. and Yoshida, M. (2008) 'Global analysis of gel mobility of proteins and its use in target identification', *Journal of Biological Chemistry*, 283(16), pp. 10745-10752.

Shockman, G.D. and Barrett, J.F. (1983) 'Structure, function, and assembly of cell-walls of gram-positive bacteria', *Annual Review of Microbiology*, 37, pp. 501-527.

Shrestha, S., Paeng, I.R., Deo, S.K. and Daunert, S. (2002) 'Cysteine-free mutant of aequorin as a photolabel in immunoassay development', *Bioconjugate Chemistry*, 13(2), pp. 269-275.

Sivanandan, S., Soraisham, A.S. and Swarnam, K. 2011 (2011) 'Choice and duration of antimicrobial therapy for neonatal sepsis and meningitis'. *International Journal of Pediatrics*, pp. 1-9 712150.

- Skoff, T.H., Farley, M.M., Petit, S., Craig, A.S., Schaffner, W., Gershman, K., Harrison, L.H., Lynfield, R., Mohle-Boetani, J., Zansky, S., Albanese, B.A., Stefonek, K., Zell, E.R., Jackson, D., Thompson, T. and Schrag, S.J. (2009) 'Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990-2007', *Clinical Infectious Diseases*, 49(1), pp. 85-92.
- Slotved, H.C., Kong, F., Lambertsen, L., Sauer, S. and Gilbert, G.L. (2007) 'Serotype IX, a proposed new *Streptococcus agalactiae* serotype', *Journal of Clinical Microbiology*, 45(9), pp. 2929-2936.
- Smith, R.H. and Kotin, R.M. (1998) 'The Rep52 gene product of adeno-associated virus is a DNA helicase with 3'-to-5' polarity', *Journal of Virology*, 72(6), pp. 4874-4881.
- Soares, G.C.T., Alviano, D.S., Santos, G.D., Alviano, C.S., Mattos-Guaraldi, A.L. and Nagao, P.E. (2013) 'Prevalence of group B streptococcus serotypes III and V in pregnant women of Rio de Janeiro, Brazil', *Brazilian Journal of Microbiology*, 44(3), pp. 869-872.
- Song, J.N., Tan, H., Mahmood, K., Law, R.H.P., Buckle, A.M., Webb, G.I., Akutsu, T. and Whisstock, J.C. (2009) 'Prodepth: Predict residue depth by support vector regression approach from protein sequences only', *Plos One*, 4(9).
- Song, K.-M., Lee, S. and Ban, C. (2012) 'Aptamers and their biological applications', *Sensors*, 12(1), pp. 612-631.
- Song, Q., Stadler, L.K.J., Peng, J. and Ferrigno, P.K. (2011) 'Peptide aptamer microarrays: Bridging the bio-detector interface', *Faraday Discussions*, 149, pp. 79-92.
- Spellerberg, B. and Brandt, C. (2011) *Streptococcus. Manual of clinical microbiology, 10th edition*. American Society for Microbiology.
- Spurlino, J.C., Lu, G.Y. and Quioco, F.A. (1991) 'The 2.3-A resolution structure of the maltose-binding or maltodextrin-binding protein, a primary receptor of bacterial active-transport and chemotaxis', *Journal of Biological Chemistry*, 266(8), pp. 5202-5219.
- St John, A. and Price, C.P. 35 (2014) 'Existing and emerging technologies for point-of-care testing'. *The Clinical Biochemist Reviews*, pp. 155-167 3.
- Staats, J.J., Feder, I., Okwumabua, O. and Chengappa, M.M. (1997) 'Streptococcus suis: Past and present', *Veterinary Research Communications*, 21(6), pp. 381-407.
- Stadler, L.K.J., Tomlinson, D.C., Lee, T., Knowles, M.A. and Ferrigno, P.K. (2014) 'The use of a neutral peptide aptamer scaffold to anchor BH3 peptides constitutes a viable approach to studying their function', *Cell Death & Disease*, 5.
- Stills, H.F. (2005) 'Adjuvants and antibody production: Dispelling the myths associated with Freund's complete and other adjuvants', *Ilar Journal*, 46(3), pp. 280-293.

Stoll, B.J., Hansen, N.I., Sanchez, P.J., Faix, R.G., Poindexter, B.B., Van Meurs, K.P., Bizzarro, M.J., Goldberg, R.N., Frantz, I.D., Hale, E.C., Shankaran, S., Kennedy, K., Carlo, W.A., Watterberg, K.L., Bell, E.F., Walsh, M.C., Schibler, K., Lupton, A.R., Shane, A.L., Schrag, S.J., Das, A., Higgins, R.D. and Eunice Kennedy Shriver Natl Inst, C. (2011) 'Early onset neonatal sepsis: The burden of group B streptococcal and E. coli disease continues', *Pediatrics*, 127(5), pp. 817-826.

Sun, K., Xing, W.W., Yu, X.L., Fu, W.L., Wang, Y.Y., Zou, M.J., Luo, Z.H. and Xu, D.G. (2016) 'Recombinase polymerase amplification combined with a lateral flow dipstick for rapid and visual detection of *Schistosoma japonicum*', *Parasites & Vectors*, 9.

Tazi, A., Reglier-Poupet, H., Dautezac, F., Raymond, J. and Poyart, C. (2008) 'Comparative evaluation of Strepto BID (R) chromogenic medium and Granada media for the detection of Group B streptococcus from vaginal samples of pregnant women', *Journal of Microbiological Methods*, 73(3), pp. 263-265.

Terrone, D.A., Rinehart, B.K., Einstein, M.H., Britt, L.B., Martin, J.N. and Perry, K.G. (1999) 'Neonatal sepsis and death caused by resistant *Escherichia coli*: Possible consequences of extended maternal ampicillin administration', *American Journal of Obstetrics and Gynecology*, 180(6), pp. 1345-1347.

Tettelin, H., Maignani, V., Cieslewicz, M.J., Eisen, J.A., Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I., Read, T.D., Madoff, L.C., Wolf, A.M., Beanan, M.J., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Lewis, M.R., Radune, D., Fedorova, N.B., Scanlan, D., Khouri, H., Mulligan, S., Carty, H.A., Cline, R.T., Van Aken, S.E., Gill, J., Scarselli, M., Mora, M., Iacobini, E.T., Brettoni, C., Galli, G., Mariani, M., Vegni, F., Maione, D., Rinaudo, D., Rappuoli, R., Telford, J.L., Kasper, D.L., Grandi, G. and Fraser, C.M. (2002) 'Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*', *Proceedings of the National Academy of Sciences of the United States of America*, 99(19), pp. 12391-12396.

Thach, T.T., Luong, T.T., Lee, S. and Rhee, D.K. (2014) 'Adenylate kinase from *Streptococcus pneumoniae* is essential for growth through its catalytic activity', *Febs Open Bio*, 4, pp. 672-682.

The European Parliament and the Council of the European Union L 331 (1998) 'Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices'. Official Journal of the European Communities, pp. 1-37.

Thinkhamrop, J., Limpongsanurak, S., Festin, M.R., Daly, S., Schuchat, A., Lumbiganon, P., Zell, E., Chipato, T., Win, A.A., Perilla, M.J., Tolosa, J.E. and Whitney, C.G. (2003) 'Infections in international pregnancy study: Performance of the optical immunoassay test for detection of group B streptococcus', *Journal of Clinical Microbiology*, 41(11), pp. 5288-5290.

Tröger, V., Niemann, K., Gärtig, C. and Kuhlmeier, D. 6 (2015) 'Isothermal amplification and quantification of nucleic acids and its use in microsystems'. *Journal of Nanomedicine & Nanotechnology*, p. 282.

Tsuji, F.I., Inouye, S., Goto, T. and Sakaki, Y. (1986) 'Site-specific mutagenesis of the calcium-binding photoprotein aequorin', *Proceedings of the National Academy of Sciences of the United States of America*, 83(21), pp. 8107-8111.

Tuerk, C. and Gold, L. (1990) 'Systematic evolution of ligands by exponential enrichment - RNA ligands to bacteriophage T4 DNA polymerase', *Science*, 249(4968), pp. 505-510.

TwistDx (2016) *TwistAmp® DNA amplification kits, Combined instruction manual*. Available at: [www.twistdx.co.uk](http://www.twistdx.co.uk) (Accessed: 30th August 2017).

TwistDx (2017) *Using PCR primers with standard recombinase polymerase amplification reagents*. Available at: <https://www.twistdx.co.uk/en/rpa/using-pcr-primers> (Accessed: 1st September 2017).

UK National Screening Committee (2012) 'Antenatal screening for group B streptococcal carriage policy position statement'.

UK National Screening Committee (2017) 'UK NSC group B streptococcus (GBS) recommendation'.

Ulett, K.B., Shuemaker, J.H., Benjamin, W.H., Tan, C.K. and Ulett, G. 6 (2012) 'Group B streptococcus cystitis presenting in a diabetic patient with a massive abdominopelvic abscess: a case report'. *Journal of Medical Case Reports* 237.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. and Leunissen, J.A.M. (2007) 'Primer3Plus, an enhanced web interface to Primer3', *Nucleic Acids Research*, 35, pp. W71-W74.

US Food and Drug Administration (2017) 'Devices@FDA: "group B streptococcus" Administration, U.F.a.D. Available at: <https://www.accessdata.fda.gov/scripts/cdrh/devicesatfda/index.cfm> (Accessed: 29th September 2017).

Van Assche, E., Van Puyvelde, S., Vanderleyden, J. and Steenackers, H.P. (2015) 'RNA-binding proteins involved in post-transcriptional regulation in bacteria', *Frontiers in Microbiology*, 6.

van den Berg, J.P., Westerbeek, E.A.M., van der Klis, F.R.M., Berbers, G.A.M. and van Elburg, R.M. (2011) 'Transplacental transport of IgG antibodies to preterm infants: A review of the literature', *Early Human Development*, 87(2), pp. 67-72.

Veitch, N.C. (2004) 'Horseradish peroxidase: a modern view of a classic enzyme', *Phytochemistry*, 65(3), pp. 249-259.

Verani, J.R., McGee, L. and Schrag, S. 59 (2010) 'Prevention of perinatal group B streptococcal disease revised guidelines from CDC, 2010'. *Morbidity and Mortality Weekly Report*. Department of Health and Human Services Centers for Disease Control and Prevention RR-10.

- Vergnano, S., Embleton, N., Collinson, A., Menson, E., Russell, A.B. and Heath, P. (2010) 'Missed opportunities for preventing group B streptococcus infection', *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 95(1), pp. F72-F73.
- Vincent, M., Xu, Y. and Kong, H.M. (2004) 'Helicase-dependent isothermal DNA amplification', *Embo Reports*, 5(8), pp. 795-800.
- Vinnemeier, C.D., Brust, P., Owusu-Dabo, E., Sarpong, N., Sarfo, E.Y., Bio, Y., Rolling, T., Dekker, D., Adu-Sarkodie, Y., Eberhardt, K.A., May, J. and Cramer, J.P. (2015) 'Group B Streptococci serotype distribution in pregnant women in Ghana: assessment of potential coverage through future vaccines', *Tropical Medicine & International Health*, 20(11), pp. 1516-1524.
- Vlieghe, P., Lisowski, V., Martinez, J. and Khrestchatisky, M. (2010) 'Synthetic therapeutic peptides: science and market', *Drug Discovery Today*, 15(1-2), pp. 40-56.
- Vogelstein, B. and Kinzler, K.W. (1999) 'Digital PCR', *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), pp. 9236-9241.
- Vornhagen, J., Quach, P., Boldenow, E., Merillat, S., Whidbey, C., Ngo, L.Y., Waldorf, K.M.A. and Rajagopal, L. (2016) 'Bacterial hyaluronidase promotes ascending GBS infection and preterm birth', *Mbio*, 7(3).
- Votava, M., Tejkalova, M., Drabkova, M., Unzeitig, V. and Braveny, I. (2001) 'Use of GBS media for rapid detection of group B streptococci in vaginal and rectal swabs from women in labor', *European Journal of Clinical Microbiology & Infectious Diseases*, 20(2), pp. 120-122.
- Walker, G.T., Fraiser, M.S., Schram, J.L., Little, M.C., Nadeau, J.G. and Malinowski, D.P. (1992) 'Strand displacement amplification - An isothermal, in vitro DNA amplification technique', *Nucleic Acids Research*, 20(7), pp. 1691-1696.
- Wang, J.P., Leong, M.C., Leong, E.Z.W., Sen Kuan, W. and Leong, D.T. (2017) 'Clinically relevant detection of Streptococcus pneumoniae with DNA-antibody nanostructures', *Analytical Chemistry*, 89(12), pp. 6900-6906.
- Wang, R., Xiang, S., Zhang, Y., Chen, Q., Zhong, Y. and Wang, S. (2014a) 'Development of a functional antibody by using a green fluorescent protein frame as the template', *Applied and Environmental Microbiology*, 80(14), pp. 4126-4137.
- Wang, Z.F., Guo, C.M., Xu, Y.N., Liu, G.J., Lu, C.P. and Liu, Y.J. (2014b) 'Two novel functions of hyaluronidase from Streptococcus agalactiae are enhanced intracellular survival and inhibition of proinflammatory cytokine expression', *Infection and Immunity*, 82(6), pp. 2615-2625.
- Wannamaker, L. (1965) 'A method for culturing beta hemolytic streptococci from the throat. Committee report, prepared at the request of the Executive Committee on Rheumatic Fever and Congenital Heart Disease.'. New York: American Heart Association.

- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) 'Jalview Version 2-a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189-1191.
- Waugh, D.S. (2016) 'Crystal structures of MBP fusion proteins', *Protein Science*, 25(3), pp. 559-571.
- Weisner, A.M., Johnson, A.P., Lamagni, T.L., Arnold, E., Warner, M., Heath, P.T. and Efstratiou, A. (2004) 'Characterization of group B streptococci recovered from infants with invasive disease in England and Wales', *Clinical Infectious Diseases*, 38(9), pp. 1203-1208.
- Weiss, J., Ray, P. and Bassford, P. 85 (1988) 'Purified SecB protein of Escherichia coli retards folding and promotes membrane translocation of the maltose-binding protein in vitro'. Proceedings of the National Academy of Sciences, pp. 8978-8982.
- Wertheim, H.F.L., Nghia, H.D.T., Taylor, W. and Schultz, C. (2009) 'Streptococcus suis: An emerging human pathogen', *Clinical Infectious Diseases*, 48(5), pp. 617-625.
- Wessels, M.R., Paoletti, L.C., Kasper, D.L., Difabio, J.L., Michon, F., Holme, K. and Jennings, H.J. (1990) 'Immunogenicity in animals of a polysaccharide-protein conjugate vaccine against type-III group-B streptococcus', *Journal of Clinical Investigation*, 86(5), pp. 1428-1433.
- Whitmore, L. and Wallace, B.A. (2008) 'Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases', *Biopolymers*, 89(5), pp. 392-400.
- Wilkinson, H.W. (1977) 'CAMP-disk test for presumptive identification of group-B streptococci', *Journal of Clinical Microbiology*, 6(1), pp. 42-45.
- Williams-Bouyer, N., Reisner, B.S. and Woods, G.L. (2000) 'Comparison of Gen-Probe AccuProbe Group B streptococcus culture identification test with conventional culture for the detection of Group B streptococci in broth cultures of vaginal-anorectal specimens from pregnant women', *Diagnostic Microbiology and Infectious Disease*, 36(3), pp. 159-162.
- Wilson, D.A., Hall, G.S. and Procop, G.W. (2010) 'Detection of group B streptococcus bacteria in LIM enrichment broth by peptide nucleic acid fluorescent in situ hybridization (PNA FISH) and rapid cycle PCR', *Journal of Clinical Microbiology*, 48(5), pp. 1947-1948.
- Wong, H.B. and Lim, G.H. 20 (2011) 'Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV'. Proceedings of Singapore Healthcare, pp. 316-318 4.
- Woodman, R., Yeh, J.T.H., Laurenson, S. and Ferrigno, P.K. (2005) 'Design and validation of a neutral protein scaffold for the presentation of peptide aptamers', *Journal of Molecular Biology*, 352(5), pp. 1118-1133.
- World Health Organization (2006) *State of the art of vaccine research and development. Immunization, vaccines and biologicals*. Available at:

[http://apps.who.int/iris/bitstream/10665/69348/1/WHO\\_IVB\\_06.01\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/69348/1/WHO_IVB_06.01_eng.pdf)  
(Accessed: 30th September 2017).

Wu, Y.D., Zhou, D.H., Zhang, L.X., Zheng, W.B., Ma, J.G., Wang, M., Zhu, X.Q. and Xu, M.J. (2016) 'Recombinase polymerase amplification (RPA) combined with lateral flow (LF) strip for equipment-free detection of *Cryptosporidium* spp. oocysts in dairy cattle feces', *Parasitology Research*, 115(9), pp. 3551-3555.

Xie, Y., Yang, J.L., Zhao, P., Jia, H. and Wang, Q. (2016) 'Occurrence and detection method evaluation of group B streptococcus from prenatal vaginal specimen in Northwest China', *Diagnostic Pathology*, 11.

Xu, C.W. and Luo, Z.J. (2002) 'Inactivation of Ras function by allele-specific peptide aptamers', *Oncogene*, 21(37), pp. 5753-5757.

Yanai, H., Hamasaki, H., Tsuda, N., Adachi, H., Yoshikawa, R., Moriyama, S., Masui, Y. and Mishima, S. 4 (2012) 'Group B streptococcus infection and diabetes: A review'. *Journal of Microbiology and Antimicrobials*, pp. 1-5 1.

Yang, Q. (2011) *Proteomic investigation of the group B streptococcus*. Northumbria University.

Yang, W.H., Cao, C.M., Wang, W.C. and Wang, Y.S. (2016) 'Rapid detection of *Streptococcus agalactiae* infection using a loop-mediated isothermal amplification method', *Jundishapur Journal of Microbiology*, 9(12).

Yang, Y., Qin, X.D., Song, Y.M., Zhang, W., Hu, G.W., Dou, Y.X., Li, Y.M. and Zhang, Z.D. (2017) 'Development of real-time and lateral flow strip reverse transcription recombinase polymerase amplification assays for rapid detection of peste des petits ruminants virus', *Virology Journal*, 14.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. and Brinkman, F.S.L. (2010) 'PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes', *Bioinformatics*, 26(13), pp. 1608-1615.

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. (2000) 'A greedy algorithm for aligning DNA sequences', *Journal of Computational Biology*, 7(1-2), pp. 203-214.

Zhao, Y.X., Chen, F., Li, Q., Wang, L.H. and Fan, C.H. (2015) 'Isothermal Amplification of Nucleic Acids', *Chemical Reviews*, 115(22), pp. 12491-12545.

Zougrana, T., Findenegg, G.H. and Norde, W. (1997) 'Structure, stability, and activity of adsorbed enzymes', *Journal of Colloid and Interface Science*, 190(2), pp. 437-448.