<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Glycosylation and glycoengineering of recombinant antibody fragments in Escherichia coli</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>O’ Connor, Iain</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2016-11-14</td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/6158">http://hdl.handle.net/10379/6158</a></td>
</tr>
</tbody>
</table>

Some rights reserved. For more information, please see the item record link above.
Glycosylation and glycoengineering of recombinant antibody fragments in *Escherichia coli*

A thesis submitted to the National University of Ireland, Galway for the degree of Doctor of Philosophy by:

Iain O' Connor, B.Sc.

Recombinant Protein Production Laboratory,

Discipline of Microbiology, School of Natural Sciences

College of Science, National University of Ireland, Galway

Research Supervisor: Dr. J. Gerard Wall

Head of Discipline: Dr. J. Gerard Wall

November 2016
# Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of contents</td>
<td>i</td>
</tr>
<tr>
<td>List of abbreviations</td>
<td>vi</td>
</tr>
<tr>
<td>List of figures</td>
<td>ix</td>
</tr>
<tr>
<td>List of tables</td>
<td>xii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xiii</td>
</tr>
<tr>
<td>Declaration</td>
<td>xiv</td>
</tr>
<tr>
<td>Abstract</td>
<td>xv</td>
</tr>
<tr>
<td>Chapter 1: General Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Recombinant protein production</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Glycosylation in recombinant hosts</td>
<td>4</td>
</tr>
<tr>
<td>1.3 <em>E. coli</em> as an expression host</td>
<td>6</td>
</tr>
<tr>
<td>1.4 Glycosylation in <em>E. coli</em></td>
<td>10</td>
</tr>
<tr>
<td>1.5 Antibody sources</td>
<td>13</td>
</tr>
<tr>
<td>1.6 Antibody structure</td>
<td>15</td>
</tr>
<tr>
<td>1.7 Antibody fragments</td>
<td>17</td>
</tr>
<tr>
<td>1.8 Antibody fragment isolation and expression in <em>E. coli</em></td>
<td>19</td>
</tr>
<tr>
<td>1.9 Immobilisation approaches</td>
<td>21</td>
</tr>
<tr>
<td>1.10 Immobilised antibody fragments</td>
<td>24</td>
</tr>
<tr>
<td>1.11 Scope of this thesis</td>
<td>25</td>
</tr>
<tr>
<td>Chapter 2: Materials and Methods</td>
<td>27</td>
</tr>
<tr>
<td>2.1 Materials</td>
<td>28</td>
</tr>
<tr>
<td>2.1.1 Suppliers and reagents</td>
<td>28</td>
</tr>
<tr>
<td>2.1.2 Bacterial strains, plasmids and primers</td>
<td>29</td>
</tr>
<tr>
<td>2.1.3 Antibiotics</td>
<td>33</td>
</tr>
<tr>
<td>2.1.4 Antibodies, lectins and detection reagents</td>
<td>33</td>
</tr>
<tr>
<td>2.1.5 Molecular weight markers</td>
<td>33</td>
</tr>
<tr>
<td>2.1.6 General multi-component buffers</td>
<td>33</td>
</tr>
<tr>
<td>2.1.7 Protein purification buffers</td>
<td>34</td>
</tr>
</tbody>
</table>
2.1.8 Culture media

2.1.8.1 Lysogeny broth (LB)

2.1.8.2 ZYP-5052

2.1.8.3 Pichia pastoris media

2.2. Methods

2.2.1 Bacterial growth and protein expression

2.2.2 Electrocompetent cell preparation

2.2.3 Transformation by electroporation

2.2.4 Transformation by heat-shock

2.2.5 Protein expression by induction via IPTG

2.2.6 Auto-induction

2.2.7 P. pastoris growth and protein expression

2.2.8 Protein extraction

2.2.8.1 Periplasmic protein extraction from E. coli

2.2.8.2 Protein extraction from P. pastoris

2.2.9 Protein purification and cleanup

2.2.9.1 Affinity chromatography using immobilised nickel

2.2.9.2 Affinity chromatography using cobalt resin

2.2.10 Protein processing and analysis

2.2.10.1 SDS-PAGE

2.2.10.2 Protein staining

2.2.10.3 Western blotting

2.2.10.4 Glycan detection/staining

2.2.10.5 Lectin blotting

2.2.10.6 Protein quantification

2.2.10.7 Glycoprotein oxidation

2.2.11 Enzyme-Linked Immunosorbent Assay (ELISA)

2.2.11.1 Indirect ELISA

2.2.11.2 Inhibition ELISA
2.2.12 Protein immobilisation 43

2.2.12.1 Adsorption to microplates 43
2.2.12.2 Covalent attachment to microplates 44
2.2.12.3 Immobilisation on TiO$_2$-coated 316L stainless steel 44
2.2.12.4 Disk washing, antigen incubation, and imaging 44
2.2.12.5 Densitometry analysis of immobilisations 45

2.2.13 Oligosaccharide extension 45

2.2.14 Lectin array 45

2.2.15 Molecular techniques – DNA 46

2.2.15.1 Plasmid isolation and purification 46
2.2.15.2 Restriction enzyme digestion 46
2.2.15.3 Dephosphorylation and ligation 46
2.2.15.4 Polymerase chain reaction 47
2.2.15.5 Agarose gel electrophoresis 47
2.2.15.6 DNA purification and concentration 47
2.2.15.7 Mutagenesis 47
2.2.15.8 Sequencing 48

2.2.16 Molecular techniques – RNA 48

2.2.16.1 RNA isolation from Pichia pastoris 48
2.2.16.2 cDNA synthesis 48

2.2.17 Computational analyses 49

2.2.17.1 Codon usage analysis 49
2.2.17.2 Sequence alignment and secondary structure prediction 49

Chapter 3: Glycoprotein expression in E. coli and engineering of the glycan chain 50

3.1 Introduction 51

3.2 Production of glycosylated scFvs in E. coli 56

3.2.1 Selection of scFv for glycoengineering 56
3.2.2 Optimisation of glycoprotein production and recovery

3.3 Generation of alternative pgl constructs

3.3.1 Effect on glycan length of site-directed mutagenesis of pglJ

3.3.2 Investigation of pgl locus redundancy by truncation mutagenesis

3.3.3 Western blot-based characterisation of glycoprofiles

3.3.3.1 DIG-specific analysis of glycoproteins

3.4 Functionality of glycoproteins

3.4.1 Antigen binding

3.4.2 Immobilisation of scFvs using adsorption

3.4.2.1 Adsorption - antigen-binding

3.4.3 Covalent immobilisation

3.4.3.1 Time course of covalent immobilisation of scFvs

3.4.3.2 Antigen-binding by covalently immobilised scFvs

3.4.4 Analysis of antigen binding by oxidised scFvs

3.5 Discussion

Chapter 4: Investigation into extension of the established Campylobacter jejuni N-linked glycan in Escherichia coli and additional characterisations of glycosylated scFvs

4.1 Introduction

4.2 Investigating protein sialylation in E. coli

4.3 CgtB-mediated extension and expression in E. coli

4.4 cgtB expression in P. pastoris

4.5 CgtB purification and oligosaccharide extension

4.6 Alternate characterisation of glycoproteins

4.6.1 Lectin array analysis

4.6.2 Lectin blot analysis of glycoproteins

4.7 Glycan-mediated immobilisation of scFvs

4.8 Discussion
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(NH$_4$)$_2$SO$_4$</td>
<td>Ammonium sulfate</td>
</tr>
<tr>
<td>*</td>
<td>Stop codon</td>
</tr>
<tr>
<td>2H12</td>
<td>Anti-domoic acid scFv</td>
</tr>
<tr>
<td>4M5.3</td>
<td>Anti-fluorescein scFv</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic force microscopy</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>AOX1</td>
<td>Alcohol oxidase 1 promoter</td>
</tr>
<tr>
<td>APTES</td>
<td>3-(aminopropyl) triethoxysilane</td>
</tr>
<tr>
<td>Asn</td>
<td>Asparagine</td>
</tr>
<tr>
<td>AURIF</td>
<td>Auristatin F</td>
</tr>
<tr>
<td>Bac</td>
<td>Bacillosamine</td>
</tr>
<tr>
<td>BMGY</td>
<td>Buffered glycerol-complex medium</td>
</tr>
<tr>
<td>BMMY</td>
<td>Buffered methanol-complex medium</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>Cam</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td>CDR</td>
<td>Complementarity-determining region</td>
</tr>
<tr>
<td>C$_H$</td>
<td>Heavy chain constant domain</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary cell</td>
</tr>
<tr>
<td>C$_L$</td>
<td>Light chain constant domain</td>
</tr>
<tr>
<td>C-terminal</td>
<td>Carboxyl terminal</td>
</tr>
<tr>
<td>Cy5</td>
<td>Cyanine5 NHS ester</td>
</tr>
<tr>
<td>dH$_2$O</td>
<td>Deionised water</td>
</tr>
<tr>
<td>DHB</td>
<td>2,5-dihydroxybenzoic acid</td>
</tr>
<tr>
<td>DIG</td>
<td>Digoxigenin</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DsbA/C</td>
<td>Disulfide bond isomerase A/C</td>
</tr>
<tr>
<td>dsFv</td>
<td>Disulfide-stabilised Fv antibody fragment</td>
</tr>
<tr>
<td>EDA</td>
<td>Ethylenediamine</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediamine tetraacetic acid</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>Endo H/F</td>
<td>Endoglycosidase H/F</td>
</tr>
<tr>
<td>Fab</td>
<td>Antigen-binding fragment</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>GalNAc</td>
<td>N-Acetylgalactosamine</td>
</tr>
<tr>
<td>Glc</td>
<td>Glucose</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-Acetylglicosamine</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>Water</td>
</tr>
<tr>
<td>H$_2$SO$_4$</td>
<td>Sulphuric acid</td>
</tr>
<tr>
<td>HAMA</td>
<td>Human anti-mouse antibody</td>
</tr>
<tr>
<td>HER2</td>
<td>Human epidermal growth factor receptor 2</td>
</tr>
<tr>
<td>His tag</td>
<td>Hexahistidine tag</td>
</tr>
<tr>
<td>Hrp</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>Hsp</td>
<td>Heat shock protein</td>
</tr>
<tr>
<td>IC$_{50}$</td>
<td>Half maximal inhibitory concentration</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilised metal ion affinity chromatography</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>KH2PO4</td>
<td>Monopotassium phosphate</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani broth</td>
</tr>
<tr>
<td>LOS</td>
<td>Lipooligosaccharide</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>mAb</td>
<td>Monoclonal antibody</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix assisted laser desorption/ionisation time-of-flight</td>
</tr>
<tr>
<td>Man</td>
<td>Mannose</td>
</tr>
<tr>
<td>MBP</td>
<td>Maltose-binding protein</td>
</tr>
<tr>
<td>MgCl2</td>
<td>Magnesium chloride</td>
</tr>
<tr>
<td>MgSO4</td>
<td>Magnesium Sulphate</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MWCO</td>
<td>Molecular weight cut off</td>
</tr>
<tr>
<td>Na2HPO4.2H2O</td>
<td>Di-sodium hydrogen phosphate dihydrate</td>
</tr>
<tr>
<td>NeuAc</td>
<td>Neuraminic acid (Sialic acid)</td>
</tr>
<tr>
<td>NH2</td>
<td>Primary amine</td>
</tr>
<tr>
<td>N-linked/glycosylated</td>
<td>Asparagine-linked/glycosylated</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic spectroscopy</td>
</tr>
<tr>
<td>N-terminal</td>
<td>Amino terminal</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>O-linked/glycosylated</td>
<td>Serine/threonine-linked/glycosylated</td>
</tr>
<tr>
<td>Omp</td>
<td>Outer membrane protein</td>
</tr>
<tr>
<td>OST</td>
<td>Oligosaccharyltransferase</td>
</tr>
<tr>
<td>pAb</td>
<td>Polyclonal antibody</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PeI B</td>
<td>Pectate lyase B</td>
</tr>
<tr>
<td>pg/l</td>
<td>C. jejuni protein glycosylation gene family</td>
</tr>
<tr>
<td>PNGase F</td>
<td>Peptide-N-Glycosidase F</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>QCM-D</td>
<td>Quartz crystal microbalance with dissipation monitoring</td>
</tr>
<tr>
<td>R²</td>
<td>Coefficient of determination</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RPP</td>
<td>Recombinant protein production</td>
</tr>
<tr>
<td>SBA</td>
<td>Soybean lectin from <em>Glycine Max</em></td>
</tr>
<tr>
<td>scFv</td>
<td>Single chain variable fragment</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Strep tag</td>
<td>Streptavidin tag</td>
</tr>
<tr>
<td>SUMO</td>
<td>Small ubiquitin-like modifier protein</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris buffered saline</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco etch virus</td>
</tr>
<tr>
<td>VH</td>
<td>Heavy chain variable domain</td>
</tr>
<tr>
<td>VH H</td>
<td>Single domain antibodies</td>
</tr>
<tr>
<td>VL</td>
<td>Light chain variable domain</td>
</tr>
<tr>
<td>YPD S</td>
<td>Yeast extract peptone dextrose medium</td>
</tr>
</tbody>
</table>
## Units of measurement

<table>
<thead>
<tr>
<th>Units</th>
<th>Symbols</th>
<th>Prefixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seconds</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>Minutes</td>
<td>min</td>
<td></td>
</tr>
<tr>
<td>Hours</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>Litres</td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>Gravities</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>Base pairs</td>
<td>bp</td>
<td></td>
</tr>
<tr>
<td>Absorbance</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Ampere</td>
<td>Amp</td>
<td></td>
</tr>
<tr>
<td>Molarity</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>Grams</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>Metres</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>Daltons</td>
<td>Da</td>
<td></td>
</tr>
<tr>
<td>Voltage</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>Degrees Celsius</td>
<td>°C</td>
<td></td>
</tr>
<tr>
<td>Moles</td>
<td>mol</td>
<td></td>
</tr>
</tbody>
</table>

### Unit prefixes

- **centi**: $10^{-2}$
- **milli**: $10^{-3}$
- **micro**: $10^{-6}$
- **nano**: $10^{-9}$
- **pico**: $10^{-12}$
- **femto**: $10^{-15}$
List of figures

Fig 1.1. A. *C. jejuni*-directed heptasaccharide glycan chain and associated *pgl* locus-encoded glycosyltransferase activities. B. *N*-linked core oligosaccharide structure attached to eukaryotic glycoproteins. 11

Fig 1.2. Structure of a whole IgG molecule and derived antigen-binding antibody fragments. 16

Fig 3.1. i Schematic representation of the *C. jejuni* locus that produces *N*-linked glycoproteins in *E. coli*. ii *C. jejuni*-directed glycan chain and associated *pgl* locus glycosyltransferase activities. 53

Fig. 3.2 Alignment of predicted amino acid sequences of *pglJ* and *pimA*. 55

Fig. 3.3 Comparison of glycosylated 4M5.3 scFv expression by IPTG induction and auto-induction. 57

Fig. 3.4 Analysis by whole protein staining and western blot analysis of purification of glycosylated 4M5.3 scFv produced by auto-induction conditions in the presence of 1X glucose or 5X glucose, concentrations. 58

Fig. 3.5 Western blot analysis of soluble and resoluble extracts of glycosylated 4M5.3 scFv soluble and resoluble extracts produced by auto-induction conditions with 1X and 5X glucose. 59

Fig. 3.6 Anti-polymhistidine antibody-based detection of BSA-FITC binding by extracted soluble and resoluble scFv adsorbed on 96-well microplates. 60

Fig 3.7 Comparison of glycoforms of 4M5.3 scFv produced in *E. coli* cells harbouring plasmids containing a variety of pACYC*pgl* constructs. 61

Fig 3.8 DIG-mediated detection of glycoforms of 4M5.3 scFv produced in the presence of different pACYC*pgl* constructs. 66

Fig 3.9 Comparative inhibition ELISAs of 4M5.3 scFvs. 67

Fig. 3.10 Quantification of scFvs remaining in solution after immobilisation on ELISA plates. 69

Fig. 3.11 Amount of scFv remaining on ELISA plate wells after adsorption of 6 µg ml⁻¹ (11.25 pmol) of each scFv under varying conditions. 70

Fig. 3.12 Amount of scFv remaining on ELISA plates after 2 h adsorption. 71

Fig 3.13 Comparison of antigen binding of surface-adsorbed scFv glycovariants. 73

Fig 3.14 Schematic of scFv immobilisation on functionalised surfaces. 74

Fig 3.15 Analysis of immobilisation of oxidised/unoxidised full-length glycoprotein and oxidised shortened glycoprotein on amine-functionalised plates. 75

Fig 3.16 Ratio of glycosylated scFv : unglycosylated scFv immobilised on aminated plates with increasing incubation times. 76

Fig 3.17 Quantification of immobilisation of scFv molecules on aminated
Fig 3.18 Comparison of amount of glycosylated and unglycosylated scFvs immobilised on aminated surfaces at varying concentrations.

Fig 3.19 Determination of antigen binding of covalently immobilised scFvs.

Fig 3.20 Comparison of antigen binding abilities of scFv variants immobilised on aminated surfaces.

Fig. 3.21 Inhibition ELISAs of oxidised 4M5.3 scFvs.

Fig 4.1 Hypothetical sialylated decasaccharide structure produced by the activities of a combination of C. jejuni protein glycosylation and lipooligosaccharide sialylation machineries.

Fig 4.2 pIG6-2H12-NeuAc scFv expression vector containing 2H12 scFv and additional C. jejuni genes involved in LOS sialylation.

Fig 4.3 pIG6-2H12-cgtB expression vector containing 2H12 scFv and downstream cgtB with a C-terminal streptavidin tag.

Fig 4.4 cgtB codon usage analysis for expression in P. pastoris.

Fig 4.5 Analysis of intracellular proteins prepared from P. pastoris cells after cgtB expression.

Fig 4.6 Analysis of IMAC-based purification of CgtB from the intracellular fraction of P. pastoris cells after expression.

Fig 4.7 Lectin array analysis of all glycoprotein variants with reported GalNAc- and α-linked galactose-binding lectins.

Fig 4.8 Clustering analysis of glycoprotein variants with GalNAc and α-linked galactose binding lectins.

Fig 4.9 Lectin array analysis subset of full-length glycosylated and unglycosylated scFv variants.

Fig 4.10 Lectin array analysis subset of full-length glycosylated, truncated pglH, and unglycosylated scFv variants.

Fig 4.11 Lectin binding analysis of subset of (A) full-length glycosylated, truncated pglJ, and unglycosylated scFv variants; and (B) pglJ modifications demonstrating the inter-sample relationships: single site mutagenesis; double single site mutagenesis; truncated.

Fig. 4.12 Analysis of binding of SBA to glycoprotein variants in lectin array.

Fig 4.13 SBA lectin-based detection of glycoforms of 4M5.3 expressed in E. coli cells harbouring different pACYCpgl constructs.

Fig 4.14 Comparison of normalised response of SBA-based detection of glycoprotein variants in lectin array and blot experimental setups.

Fig 4.15 Imaging of scFvs adsorbed or covalently immobilised on stainless steel aminated disk surfaces before and after washes.

Fig 4.16 Fluorescence microscopy images of glycosylated 4M5.3 scFv
immobilised on stent surfaces.

**Fig 4.17** AFM topography image of oxidised 4M5.3 scFv covalently immobilised on aminated stent material.

**Fig 4.18** 3D topographical image of oxidised 4M5.3 scFv, covalently immobilised on aminated stent material, generated from AFM data.

**Fig i** MALDI-MS/MS sequencing of 20 µg full-length heptasaccharide containing scFv generated *E. coli* containing the native *C. jejuni pgl* locus.
# List of tables

<table>
<thead>
<tr>
<th>Table 2.1</th>
<th>Bacterial strains used in this study.</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.2</td>
<td>Plasmids used in this study.</td>
<td>30</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Primers used in this study.</td>
<td>31</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>GalNAc-, α-Gal-, and Gal-binding lectins analysed in the lectin array study.</td>
<td>106</td>
</tr>
<tr>
<td>Table i</td>
<td>Lectins used in lectin array analysis.</td>
<td>175</td>
</tr>
</tbody>
</table>
Acknowledgements

First and foremost I would like to thank my supervisor Dr. Gerard Wall for his extensive support and guidance throughout my project, from making contingency plans when experiments didn't pan out to celebrating (in the grand scheme of things) marginal results and progress. Particularly I would like to thank Dr. Wall for his rapid and incredibly thorough corrections in constructing this thesis.

I would like to thank my colleagues and friends in the Recombinant protein production group, particularly Claire, Gosia, Maura, Chus, and Akshay not only for their support and guidance but for the many laughs and good days we had working and relaxing outside of the lab. I would like to additionally thank Dr. Soshee and Dr. Kumar for the wealth of knowledge and ideas they contributed to my project and for their engaging scientific discussion.

I consider myself very lucky to have worked in a department wherein postgraduates and staff are happy to lend a helping hand by providing materials or knowledge at a moment's notice and are amenable to relaxing on a Friday evening with a pint. I would like to thank Kate, Maria, Beth, Kerrie, and Aileen for your help and support throughout. I would like to give special mention to Dr. Paul Mc Cay who was one of the friendliest, most knowledgeable individuals I have had the pleasure to engage with and befriend, he will be sadly missed and forever remembered.

Amongst the staff of Microbiology there are many who go above and beyond the call of duty in their work and support. I would like to thank Dr. O’ Byrne who provided valuable guidance throughout the years in GRC meetings. Additionally Caroline, Mike, Ann, Katrina, and especially Maurice, all of whom keep the department going quietly but without whom it would quickly grind to a halt. Additionally I would like to thank Dr. Kilcoyne and Dr. Gerlach for their aid with lectin analyses.

To my mother Peggy and sister Chloe, thank you for your love and support always. Mathan, thank you for being you.

Finally I would like to thank my wife and best friend Meaghan for everything, particularly your love and patience in seeing me through to completion of this project, I look forward to seeing the next chapter in our lives.
Declaration

The work contained in this thesis is all my work. I have not obtained a degree in this University, or elsewhere based on this work.

_______________________

Iain O’ Connor
Abstract

Recombinant protein production (RPP) in *Escherichia coli* benefits from a long history of research which has facilitated the development of a plethora of strains and vectors suited to RPP applications. The identification and functional transfer to *E. coli* of an N-linked protein glycosylation system, *pgl*, from *Campylobacter jejuni* has recently enabled investigation of novel glycoprotein-based applications previously not possible in the organism. To date, a number of glyco-conjugate vaccines have been successfully produced in *E. coli*, and work towards “humanisation” of the characteristic *C. jejuni* glycan chain is continuing in order to produce glycoproteins suited to *in vivo* applications. Meanwhile, the *C. jejuni* glycan chain has been successfully attached to single-chain antibody fragments (scFvs) to mediate their covalent attachment to functionalised surfaces for improved immunosensor development.

In this study, the *pgl* system was further engineered by mutagenic analysis of the glycosyltransferase PglJ – which is responsible for addition of the third sugar of the heptasaccharide N-glycan chain. Mutagenesis of E276 and E284 of the functionally important EX$_7$E motif resulted in glycoprofiles indicating that glycan chain extension was not completely eliminated in either glycovariant, in contrast to published results with homologous glycosyltransferase enzymes. Inactivation of PglJ by termination of translation or by creation of a E276A/E284A double mutant completely eliminated glycosyltransferase activity of PglJ. Termination of translation of PglH, which further extends the glycan chain produced by PglJ, was also carried out to confirm the partial activity of the PglJ single-site mutants.

Disaccharide-linked scFvs produced in *E. coli pglJ* cells were compared to fully (heptasaccharide-containing)-glycosylated and unglycosylated scFvs in surface attachment and antigen-binding studies. Adsorption of all scFvs onto polystyrene occurred at similar rates whereas both glycosylated protein variants displayed a specific and more rapid attachment (up to five-fold greater coating) to amine-functionalised surfaces than their unglycosylated counterpart. Upon adsorption, the three scFv variants exhibited similar antigen binding while the covalently attached, glycosylated scFvs displayed up to eight-fold higher antigen binding than the
unglycosylated scFvs. This demonstrated the potential application of the truncated glycan in protein immobilisation, while minimising its immunogenicity and potential susceptibility of the sugar chain to degradation by glycosidase enzymes in vivo. Comparison of adsorption and covalent attachment of scFvs to functionalised cardiovascular stent materials confirmed the efficacy of immobilising scFvs for such applications via the shortened glycan tag. Meanwhile, the use of atomic force microscopy provided insight into the spatial positioning and distribution of covalently-immobilised scFvs on the stent material.

A lectin array analysis was carried out to query glycovariants produced by E. coli cells harbouring the engineered C. jejuni pgl constructs against a broad panel of carbohydrate-binding lectins. This rapid and multivariate typing investigation identified SBA as a candidate lectin that was used to distinguish between the glycovariant sequences in western blots.

Efforts to extend the C. jejuni-pgl-encoded glycan chain through the action of the CgtB enzyme derived from the LOS locus of C. jejuni, identified a number of bottlenecks in cgtB expression in E. coli and P. pastoris. Potential solutions to express the relevant glycosyltransferases and mediate terminal capping of E. coli-produced glycoproteins with sialic acids were established for future investigations.

These results illustrate the potential of expression of recombinant glycoproteins and their glycan engineering in E. coli in fundamental studies of glycosyltransferase enzymes and in biotechnological applications. Bioconjugates that exhibit improved immobilisation and antigen binding over their absorbed, non-glycosylated counterparts can be readily produced via the incorporation of a minimal recombinant tag. The study also establishes characterisation approaches to detect and analyse subtle glycan changes in recombinant glycoproteins. The C. jejuni pgl glycosylation system utilised in this work will likely see increased adoption and continued engineering in RPP in the near future and this study contributes to its ongoing adaptation for applications in fields such as diagnostics and therapeutics.
Chapter 1:
General Introduction
1.1 Recombinant protein production

Obtaining proteins of interest for biomedical applications traditionally required their isolation and purification from their natural environments, which could be a laborious, expensive, source-limited, and time-consuming process. In the case of therapeutic proteins sourced from animal hosts, such as insulin (Banting et al., 1922), this can result in prohibitive costs for the end-user, potentially denying life-saving or -altering medication to affected patients (Dyck et al., 2003; Holden et al., 2011). The development of recombinant expression technologies has supplanted the need for use of animal sources of many proteins as it increases availability of target proteins, provides improved safety compared to protein isolation from natural hosts or the environment, and increases the potential for engineering of the proteins to introduce beneficial modifications (Andersen and Krummen, 2002).

Recombinant protein production (RPP) is typically carried out in a number of common bacterial, eukaryotic, or cell-free systems. Cell-free systems include those based on cell extracts of *Escherichia coli* (Kigawa et al., 1999), wheat germ (Sawasaki et al., 2002; Vinarov et al., 2004), and rabbit reticulocytes (Zhang et al., 1996). These systems are typically employed to produce proteins that are difficult to express in natural hosts, such as cytotoxic proteins (Avenaud et al., 2004), membrane proteins (Katzen et al., 2005), or for isotopic labelling of proteins for NMR studies (Matsuda et al., 2007; Sobhanifar et al., 2009). Cell-free systems have not become widely adopted due to their high cost and difficulties with scale-up. In recent years, however, production costs have decreased significantly, which may lead to an increase in their adoption in RPP settings in the future (Casteleijn et al., 2013).

Eukaryotic systems are favoured for production of complex, often mammalian, proteins which rely on the host's ability to carry out high-throughput synthesis and polypeptide folding, and to introduce additional post-translational modifications (PTMs) which may be beyond the capability of bacterial systems. The most commonly used eukaryotic hosts include Chinese Hamster Ovary (CHO) cells and the yeasts *Saccharomyces cerevisiae* and *Pichia pastoris*. Insect cell lines transfected with baculovirus expression vectors (Summers, 2006) have also been used to successfully express a wide range of proteins, particularly therapeutic proteins with *N*-linked glycosylation. The main drawback of insect-based systems is
the need to infect new cells for each round of protein synthesis, unlike the continuous batch capabilities of yeast, prokaryotic, and mammalian cells (Yin et al., 2007).

Yeast cells are an attractive expression platform due to the low cost of production and inexpensive culture conditions, the ease of their genetic manipulation, and their potential to achieve high product yields by maintaining high cell densities (Cereghino and Cregg, 2000). Of particular interest for industrial production, expressed proteins can be secreted into the extracellular environment by native processes, thereby allowing simplified recovery of target proteins. *P. pastoris* in particular secretes few native proteins, thereby also simplifying subsequent downstream purification steps (Macauley-Patrick et al., 2005). Yeast systems further benefit from shuttle vectors which allow genetic manipulations to be carried out in *E. coli*, thereby exploiting its faster replication times, followed by protein expression, using the same vector, in the yeast host (Sikorski and Hieter, 1989).

Mammalian cells, such as CHO cells, are typically the favoured platform for the production of therapeutic proteins, with 33% of all biopharmaceutical products approved from 2010 to 2014 produced in CHO cells (Walsh, 2014). Mammalian cells are preferred for their capability to perform PTMs, the ability to establish a stable cell line for the constitutive expression of a target protein, and their adaptability to growth conditions such as serum-free or suspension growth conditions, which are suited to large-scale bioreactors (Kim et al., 2012). The main drawback of mammalian cells as a production host, particularly in considering the associated costs, is their susceptibility to viral contamination. CHO cells are more resistant to many viruses than other mammalian hosts, however, and further efforts are being made to understand the mechanisms behind this phenomenon and to establish strains with increased resistance (Xu et al., 2011; Mascarenhas, 2016).

Therapeutic proteins are those produced for the treatment of a specific condition *in vivo*. These include diverse proteins such as cytokines, antibodies, hormones, vaccine components, protein-drug conjugates, and growth factors for the treatment of conditions as varied as haemophilia, diabetes, infertility, anaemia, hepatitis, and cancers (Walsh, 2014; Banga, 2015). When considering the production of biopharmaceutical products and particularly therapeutic proteins, the market is
dominated by mammalian cells, *E. coli*, and yeast cells in that order, commanding a cumulative 94.5% of approved products in the 2010-2014 period (Walsh, 2014). Meanwhile, the number of biopharmaceuticals approved for clinical use has steadily risen, from 118 in 2003 (Spada and Walsh, 2004) to 246 by 2014 (Walsh, 2014). *E. coli* has seen a steady decline in its proportional contribution over this time but it remains the most commonly used non-mammalian system for the production of therapeutics, accounting typically for almost 30% of approved recombinant products (Graumann and Premstaller, 2006; Huang *et al*., 2012). If a protein does not require numerous PTMs or complex assembly, *E. coli* is often the go-to expression host due to its low cost, fast growth, and well-established regulatory processes (see Section 1.3 for more detailed information).

Many proteins undergo covalent modification of one form or another after they emerge from the ribosome. These include sulphation, phosphorylation, disulfide bond synthesis, glycosylation, acylation, acetylation, gamma-carboxylation, and proteolytic processing (Walsh and Jefferis, 2006) – all processes which can affect protein structure, folding, or functionality to varying degrees (Walsh, 2009). Therefore, it is clear that post-translational modifications of a protein must be carefully considered when choosing a suitable host for its recombinant production: large, complex proteins with high numbers of PTMs, often of human origin, are better suited to production in eukaryotic systems such as mammalian cells (Jayapol *et al*., 2007) while yeast cells, both unmodified and engineered for improved recombinant protein production (Gerngross, 2004), are finding increasing application in the production of therapeutic proteins (Kim *et al*., 2015).

### 1.2 Glycosylation in recombinant hosts

One of the most important PTMs, particularly in relation to proteins of therapeutic interest, is glycosylation as it is critical to the function and pharmacokinetic properties of many protein therapeutics. The main types of glycosylation are *N*-linked and *O*-linked, in which the associated glycan is attached to proteins via the amide nitrogen of asparagine or the hydroxyl group of serine/threonine amino acid residues, respectively (Lodish *et al*., 2000). The structures and assembly of *N*- and *O*-linked oligosaccharides also vary greatly: *O*-linked oligosaccharides typically contain only one to four monosaccharides whereas...
$N$-linked oligosaccharides can be much more complex, with polysaccharides composed of multiple repeating units and branched chains being typical.

With regard to the recombinant production of glycoproteins in particular, each commonly used expression host has advantages and drawbacks in glycan chain assembly and no single system is suitable for production of all glycoproteins. CHO cells can exhibit a high degree of heterogeneity in their glycosylation patterns so the resultant glycoprotein can be highly immunogenic (Yang et al., 2015). To circumvent this, stable cell lines that allow for increased accuracy and consistency in their glycosylation patterns have been generated, at great expense, through multiple rounds of selection and adaptation (Walsh and Jefferis, 2006).

Insect cell lines harbouring a baculovirus expression vector system are also capable of carrying out $N$-linked glycosylation and are used for the production of some therapeutic glycoproteins. They tend to produce much simpler glycans than their mammalian counterparts, with terminal mannose or $N$-Acetylgalactosamine (GalNAc) residues as opposed to the terminal $N$-Acetylglucosamine (GlcNAc) or sialic acid residues that are characteristic of many mammalian glycoproteins and critical for their half-life in vivo (Harrison and Jarvis, 2006). Attempts to date to produce glycans that are structurally identical to the human molecules have utilised transgenic insect cell lines containing the necessary glycosyltransferase profile, which has met with some success in producing identical biantennary $N$-glycans structurally to those produced by mammalian cells (Hollister et al., 2002) and other glycan structures identical to those currently produced in CHO cells (Palmberger et al., 2012).

Protein glycosylation in yeast is complicated by the tendency of yeast strains to hyperglycosylate their glycoproteins: after forming a core glycan structure similar to a precursor of mammalian $N$-glycans, up to 200 mannose residues can be added to the glycan chain, thereby increasing its size several-fold and creating a heterogenous, non-mammalian-like product (Conde et al., 2004) that is likely to confer unwanted immunogenicity in vivo and to be cleared rapidly from the human bloodstream due to recognition by human mannose receptors (Lehle et al., 1995; Jacobs and Callewaert, 2009). Desired structural features such as epitopes for immune recognition of conjugate vaccines may also be masked by the extended
glycosylation, thereby lowering the efficacy of the recombinant molecules (Kniskern et al., 1994). Similarly to optimisation strategies with other recombinant hosts, recent yeast strain engineering efforts have resulted in limiting of hyperglycosylation and, importantly, further humanising of the yeast-produced oligosaccharides by reducing mannosylation and incorporating heterologous glycosyltransferases to assemble a more human-like N-glycan signature (Wildt and Gerngross, 2005; Meehl and Stadheim, 2014; Hamilton and Zha, 2015).

1.3 **E. coli as an expression host**

*E. coli* is the prokaryotic host of choice when carrying out recombinant protein expression in a microbial system. The main other prokaryotic host, *Bacillus subtilis*, produces high titres of proteases, thereby requiring the use of less production-efficient protease-deficient strains (He et al., 1991; Demain and Vaishnav, 2009). *E. coli* has long been utilised as a research tool and as such its use as a recombinant protein expression host builds on this prior research. Its well-characterised genetics has enabled the production of a diverse variety of strains and vectors suited to particular RPP applications (Ramos et al., 2004; Blount, 2015), while its fast growth, simple fermentation pathways, and ability to reach high cell densities (Riesenberg et al., 1991) allows for rapid establishment of stable cultures for high product titres to be achieved. In recent years, *E. coli* has been used to produce a broad variety of approved therapeutic proteins such as the recombinant human growth hormone Accretropin, an inhalable recombinant insulin Afrezza, or any number of antibody fragment-based products such as the anti-inflammatory Cimzia, with more approved every year (Food and Administration, 2008; Huang et al., 2012; Walsh, 2014).

The long history of study of *E. coli* has led to a vast molecular toolbox being available for use with the organism for *de novo* protein expression or for optimisation of RPP experiments. High-copy-number plasmids with selectable markers and strong, tightly-controlled promoters enable regulated, high-level gene expression (Rosano and Ceccarelli, 2014), while features such as affinity tags (e.g. hexahistidine, streptavidin) can be inserted into target protein N- or C- termini or internal to the protein to facilitate their detection and purification. These recombinant tags can be retained for use in downstream applications such as targeted
immobilisation or they can be removed upon specific cleavage by proteases such as tobacco etch virus (TEV) protease (Riechmann et al., 1992), prior to applications such as in vivo use in which they could elicit undesirable responses (Arnau et al., 2006). Proteins of interest may also be expressed with larger fusion partners, such as maltose-binding protein (MBP) (di Guana et al., 1988; Kapust and Waugh, 1999), glutathione S-transferase (GST) (Völkel et al., 1998), protein A (Uhlén et al., 1983), or small ubiquitin-like modifier protein (SUMO) (Marblestone et al., 2006), to improve their production or enhance their solubility.

Normal protein expression in E. coli is directed to the cytoplasm, one of the membranes, the inter-membrane periplasmic space, or the extracellular environment. The extracellular secretion mechanisms of E. coli are generally less efficient than cytoplasmic or periplasmic expression and are not often preferred in RPP settings. Targeting proteins to the cytoplasmic space is preferable for ease of expression and high yields but the compartment is unsuitable for proper formation of many proteins due to its reducing environment meaning that disulfide bonds are not stably formed and maintained (Derman and Beckwith, 1991). Several strains and engineering approaches have been developed to overcome this obstacle, such as the introduction of a sulphydryl oxidase and disulfide isomerase to the cytoplasm to carry out disulfide bond formation, and strategies to increase solubility of the proteins expressed in the cytoplasm (Stewart et al., 1998; Bessette et al., 1999; Nguyen et al., 2011).

High recombinant protein expression can frequently lead to polypeptide aggregation and inclusion body formation, however. While this is unfavoured in many research settings as it yields poorly accessible, insoluble protein molecules of interest, it is frequently exploited in industrial production processes as proteins contained in inclusion bodies are less susceptible to degradation by proteases and are suitable for long term storage prior to further extraction (Graumann and Premstaller, 2006). Recovery of the target protein from inclusion bodies then typically requires harsh denaturation and refolding into their correct (active) conformation. Due to the complexity of refolding, low yields of active protein are often attained and physical analytical methodologies are required to follow the molecular refolding process. As a result, many research groups – including our own – prefer to optimise "functional
production” of proteins of interest, albeit usually at the expense of overall protein yields. Reduction of recombinant protein aggregation and inclusion body formation can be achieved in many cases by management of growth media, growth rates, and production rates. *E. coli* strains and expression plasmids have been developed to alleviate inclusion body formation by co-expression of molecular chaperones for improved folding or increased proteolysis of misfolded proteins (Nishihara et al., 2000; Deuerling et al., 2003). The use of a fusion partner such as thioredoxin can also increase solubility of the target protein, thereby alleviating inclusion body formation (LaVallie et al., 1993), while the use of additional solubility-enhancing strategies such as reducing inducer concentration, lowering expression temperatures, altering media composition, and changing the affinity tag employed for detection and purification can also improve protein production and/or folding (Gopal and Kumar, 2013; Kim et al., 2013; Zhu et al., 2013). Unfortunately, it is not currently possible to identify in silico the bottleneck limiting the folding of a “difficult to express” protein and so this optimisation of expression remains a process of trial and error.

Another approach to reduce formation of inclusion bodies is to secrete the target protein to the periplasm, as recombinant proteins in the periplasm have a much lower propensity towards inclusion body formation than those expressed in the cytoplasm (Miot and Betton, 2004). Periplasmic expression of recombinant proteins is also favourable due to the natural capacity of the cellular compartment for disulfide bond formation due to its oxidative environment, while the lower number of proteases and native proteins than in the cytoplasmic space can lead to enhanced solubility, stability, and recovery of the target protein (Choi et al., 2006). Yields of recombinant proteins expressed in the periplasmic space are typically lower than of those recovered from the cytoplasm however, due to the added complexity of subcellular transport and membrane translocation of the target polypeptide – as well as the limited capacity of the membrane translocation machinery (de Marco, 2009; Schlegel et al., 2013). The primary pathways used for protein transport to the *E. coli* periplasm are the general secretory (Sec) and twin-arginine translocation (Tat) pathways. The Sec pathway utilises an N-terminal signal peptide, such as *pelB*, *ompA*, or *dsbA/C*, attached to a protein substrate to direct transport of the unfolded polypeptide through a membrane-bound translocase, whereupon the signal peptide is cleaved (Kudva et al., 2013). The Tat pathway, meanwhile, acts on already folded
substrates containing a conserved N-terminal twin-arginine motif and has therefore been shown to be capable of transporting heterologous proteins in active forms (Santini et al., 2001; Thomas et al., 2001; Kim et al., 2005).

Further benefits of the traditional research into the genetics of *E. coli* are that a variety of strains and plasmid combinations are now available to allow expression of heterologous proteins containing features that traditionally present a difficulty for *E. coli*. While the lac promoter has been widely used for RPP, it relies on negative control and can be subject to basal levels of leak-through protein expression without induction. Therefore, the analogous T7 promoter employed by the pET vector system (Studier et al., 1990), that relies on positive control and thereby significantly reduces leaky expression (Rosano and Ceccarelli, 2014), has seen increased use in conjunction with the specialised *E. coli* host strain BL21(DE3). *E. coli* BL21(DE3) provides the T7 RNA polymerase necessary to drive transcription under the control of the lacUV5 promoter. This polymerase is highly stringent and highly active upon induction so that up to 50% of all cellular proteins can be pET-encoded under optimal conditions (Novagen, 2006). The pET system is an excellent example of approaches to improve cloning and gene expression in *E. coli*, with up to 44 different plasmids in the family, containing different combinations of antibiotic resistance markers, affinity tags, signal sequences, protease cleavage sites, and fusion partners to improve the solubility, folding, and even immunofluorescence of target proteins (Novagen, 2006).

Further improvements in RPP have seen the development of strains such as *E. coli* Rosetta which alleviates differences in codon bias between gene donor species and host *E. coli* cells, providing rare tRNAs to facilitate efficient translation, thereby avoiding premature translational termination and enabling expression of otherwise intransigent or truncated proteins in *E. coli* (Novy et al., 2001). As previously discussed, the co-expression of molecular chaperones can also aid in solubilisation and folding of recombinant proteins and is a commonly used troubleshooting approach to overcome expression problems (Georgiou and Valax, 1996; de Marco et al., 2007). Over-expression of disulfide bond isomerases DsbA or DsbC, members of the thioredoxin family (Zapun et al., 1995), can also be used to enhance disulfide bond formation in target proteins as well as, in the case of DsbC,
catalysing shuffling of already formed disulfides and even mediating protein folding. In regards to more specialised post-translational modifications, work has been carried out to successfully acetylate recombinant proteins in *E. coli* (Charbaut *et al.*, 2002), a PTM which appears to be more common in prokaryotes than previously thought (Hu *et al.*, 2010). Importantly for recombinant production of the many mammalian proteins with attached glycans, a major breakthrough in the field of RPP in *E. coli* has seen the introduction of a system for N-linked glycosylation in *E. coli*, enabling glycoprotein synthesis for their myriad of biotechnological and biomedical applications (Wacker *et al.*, 2002).

1.4 Glycosylation in *E. coli*

Previously, glycosylation was thought to be a process unique to eukaryotes but it is now clear that it occurs in archaea and bacteria also, and is even a key pathogenicity mechanism for many pathogenic species (Szymanski and Wren, 2005). The first bacterial N-linked protein glycosylation pathway was discovered in *Campylobacter jejuni*, leading to characterisation of the resultant protein glycosylation locus, *pgl*, pathway. Interruption of *pglE* and *pglB* individually, which are involved in bacillosamine synthesis and oligosaccharide transfer respectively, resulted in a loss of immunogenicity, colonisation, adherence, and invasion (Szymanski *et al.*, 2002) thereby demonstrating the important role of this glycan synthesis pathway in *C. jejuni* (Larsen *et al.*, 2004). Investigation of the structure and assembly of the *C. jejuni*-directed, Asn-linked oligosaccharide revealed it to be a sequentially assembled heptasaccharide, attached to the target protein at an N-terminally extended version of the corresponding eukaryotic consensus sequence: Asp/Glu-Y-Asn-X-Ser-Thr (D/E-Y-N-X-S/T) where Y and X can be any amino acid except proline (Kowarik *et al.*, 2006), whereas N-X-S/T is the eukaryotic recognition sequence (Schwarz and Aebi, 2011). The *C. jejuni* *pgl* cluster was successfully transferred to *E. coli* on a low-copy pACYC vector, enabling the host cells to carry out N-linked protein glycosylation and the associated mechanisms to be elucidated (Wacker *et al.*, 2002). In brief, the *pgl* cluster encodes glycosyltransferases PglC, PglA, PglJ, PglH, and PglI for cytoplasmic assembly of the oligosaccharide on a lipid-linked precursor (Und-P), followed by the action of the PglK flippase to translocate the lipid-linked oligosaccharide across the cytoplasmic membrane into
the periplasm and attachment of the glycan chain to the target protein at the specified recognition site by the oligosaccharyltransferase (OST) PglB (Linton et al., 2005).

This system, which allows for efficient, controlled production of homogenously glycosylated proteins in *E. coli*, has enormous potential application by allowing consideration of the use of the bacterium to produce many of the 70% of candidate therapeutic proteins thought to be glycosylated (Sethuraman and Stadheim, 2006). Further development of the system in *E. coli*, coupled with modification of the native *C. jejuni* machinery, may also facilitate glycoengineering of protein products in *E. coli* for improved stability and pharmacokinetics. While the extant *C. jejuni* system produces the heptasaccharide GalNAcα1-4GalNAcα1-4(Glcβ1-3)GalNAcα1-4GalNAcα1-4GalNAcα1-3Bacβ1-NAsn (Figure 1.1 A) which differs greatly from eukaryotic glycans (Figure 1.1 B) in size and composition, its attachment of the glycan chain to a similar recognition sequence in proteins in particular, opens the possibility of engineering of the bacterial mechanisms and enzymes to tailor the produced oligosaccharide to a composition more suited to specific, e.g. *in vivo*, applications.

(A)

(B)

**Fig 1.1** A. *C. jejuni*-directed heptasaccharide glycan chain and associated *pgl* locus-encoded glycosyltransferase activities. B. *N*-linked core oligosaccharide structure attached to eukaryotic glycoproteins. Image is reproduced with permission from: Schwarz and Aebi 2011.
Regarding engineering of the extant system, the critical OST PglB has been identified to have a relaxed substrate specificity, enabling it to attach alternate oligosaccharide sequences to target proteins (Wacker et al., 2006; Chen et al., 2007). This has been exploited to transfer a Staphylococcus aureus capsular polysaccharide to a detoxified S. aureus α toxin protein carrier in E. coli, resulting in a viable, multicomponent vaccine which protected rabbits and mice against S. aureus bacteremia and pneumonia (Wacker et al., 2014). Similar approaches to exploit the broad oligosaccharide specificity of PglB have yielded bioconjugate vaccines against Burkholderia pseudomallei (Garcia-Quintanilla et al., 2014), Shigella dysenteriae type 1 (Ihssen et al., 2010; Ravenscroft et al., 2016), and extra-intestinal pathogenic E. coli (van den Dobbelsteen et al., 2016).

The relaxed specificity of PglB has also been exploited for the production of a diagnostic tool for human brucellosis by transfer of the Yersinia enterocolitica O:9 polysaccharide to a carrier protein coupled to magnetic beads (Ciocchini et al., 2013), as well as a conjugate vaccine against Brucella abortus (Iwashkiw et al., 2012), and a diagnostic for shiga-toxin-producing E. coli (Melli et al., 2015). Recent saturation mutagenesis-based engineering of PglB has led to the identification of enzyme variants with significantly increased efficiency of transfer of non-native polysaccharide chains, which could enable development of additional bioconjugate vaccines that were previously not identifiable due to a prohibitively low OST efficiency of the native enzyme (Ihssen et al., 2015). Another interesting study developed PglB variants with reduced specificity in their acceptor site amino acid sequences, thereby allowing transfer of glycans onto a more diverse range of proteins, including those containing the shorter acceptor sequences typical of eukaryotic glycosylation machineries. These engineered PglB variants could allow for the glycosylation of many more proteins in E. coli expression platforms also (Ollis et al., 2014), thus broadening the potential use of the bacterium in glycoprotein production.

As discussed above, the heptasaccharide produced by the C. jejuni pgl system is unsuitable for glycosylation of therapeutic proteins for in vivo applications other than as bioconjugate vaccines, due to its differences from mammalian glycans (Figure 1.1 B). A number of approaches to “humanise” the bacterial glycan sequence
have emerged over the past decade. In one such case, the glycosylation machinery was modified to yield a lipid-linked GlcNAc rather than lipid-linked bacillosamine, followed by standard, sequential assembly of the characteristic *C. jejuni* glycan chain. Following attachment to the target protein and purification from the periplasmic space, the glycoprotein was trimmed by an endoglycosidase to yield a GlcNAc-tagged protein upon which a mammalian-like glycan was then built by transglycosylation of a pre-assembled N-glycan species *in vitro* (Schwarz *et al.*, 2010). Production of a eukaryotic-like glycan structure has also been achieved by combining the use of eukaryotic glycosyltransferases and *C. jejuni* PglB to create Man$_3$GlcNAc$_2$ glycans, the core structure of eukaryotic glycans (Valderrama-Rincon *et al.*, 2012) which has previously been found to be sufficient for the activity and pharmacokinetics of glycoprotein therapeutics (Van Patten *et al.*, 2007). Finally, an antigenic glycoprotein was synthesised through a combination of *in vitro* and *in vivo* techniques, employing glycosyltransferases from four different species, and recombinant expression and glycan assembly in *E. coli* (Hug *et al.*, 2011). Meanwhile, the *pgl* system has also been shown to be capable of glycosylating outer membrane proteins, which could have implications for cell surface display of glycan motifs (Fisher *et al.*, 2011), with potential applications in diverse fields such as vaccination and drug delivery. These reports of increasingly complex and diverse mechanisms designed to exploit non-native protein glycosylation machineries in *E. coli* continue to broaden its potential capacity as a platform for synthesis of both prokaryotic and eukaryotic glycans.

### 1.5 Antibody sources

Antibodies are widely used in biotechnological applications such as therapeutics, immunoassays, and immunosensors. Whole antibodies have traditionally been obtained from polyclonal or monoclonal sources. Polyclonal antibodies are produced and purified from immunised animal hosts, and contain not only heterogenous antibodies against the desired antigen, but also a collection of unrelated, and potentially cross-reacting antibodies can be obtained (Kindt *et al.*, 2007). Polyclonal antibodies typically require significant purification to remove these unrelated antibodies but, while the purification strategy reduces the yield of the desired antibody, it usually will not be sufficient to generate a homogenous antibody population. In addition, significant batch-to-batch variation is observed between
polyclonal antibody preparations, amounts of antibody produced are limited and finite, and there are considerable ethical concerns (and high costs) related to the continuing use of animal hosts in antibody production in this manner (Lipman et al., 2005).

Monoclonal antibodies (mAbs) are antibody populations with a single antigenic specificity that are produced from a single immortalised B-lymphocye, using technology developed by Köhler and Milstein in 1975 (Köhler and Milstein, 1975). An immunised mouse B cell is fused with a cancerous myeloma cell to form continuously replicating hybridoma cells, which are screened to identify those secreting the associated mAb, thereby providing an immortal source of homogenous antibody molecules that bind a desired target (for overview, see (Marx et al., 1997; Ward, 1999)). Monoclonal antibody production offers advantages over polyclonal production processes outlined above in that it does not suffer from inter-batch variability or the presence of heterogenous antibodies, and the produced antibodies are typically highly specific and targeted against single epitopes rather than multiple epitopes (Köhler and Milstein, 1975). Monoclonal antibodies generated in this manner were initially broadly adopted as in vitro diagnostic reagents but soon found widespread application in protein purification due to their often exquisite specificity for specific antigens in crude extracts, even when immobilised on immunoabsorbent columns (Dasch et al., 1989; Kindt et al., 2007).

The therapeutic applications of monoclonal antibodies were initially limited by their rapid clearance from the host, mediated by a human anti-mouse antibody (HAMA) response to the murine-derived antibodies in patients (De Nardo et al., 1995). This led to a well-documented pathway of generating “human-like” antibodies, firstly through chimeric man-mouse monoclonal antibody hybrids (Morrison et al., 1984), followed by chimeric, humanised antibodies retaining only murine hypervariable loops in an otherwise human protein (Kettleborough et al., 1991), and finally, the emergence of technologies for recombinant isolation and production of entirely human antibodies in vitro (Sharp et al., 1994). The success of chimeric mAbs can be seen in the fact that a number are still utilised in cancer therapies, such as Rituximab for non-Hodgkin’s lymphoma (Maloney et al., 1997), Cetuximab for colorectal cancer (Cunningham et al., 2004), and Trastuzumab for treatment of HER2-overexpressing breast cancer (Slamon et al., 2001), as well as a
constant stream of new chimeric mAb variants that continue to be developed for the treatment of diverse cancer targets (Lin et al., 2015; Challita-Eid et al., 2016). Meanwhile, entirely human mAbs which avoid the use of potentially problematic animal sources, are a rapidly growing area for targeted therapeutic products (Nelson et al., 2010) and have been employed in the treatment of rheumatoid arthritis with Adalimumab (Weinblatt et al., 2003), and of osteoporoses with the human monoclonal antibody, AMG 162 (Bekker et al., 2004). With the number and range of human monoclonal antibodies progressing through and continuing to enter clinical trials, this field seems certain to continue to grow over the coming decades (Nelson et al., 2010; Reichert, 2015).

Immunoglobulin-based applications such as therapeutics, diagnostics, and protein purification initially utilised intact, monoclonal antibodies of the IgG format. Recent decades have seen smaller antibody-based fragments become the dominant players in these fields, however (Scott et al., 2012), particularly for ligand-binding applications in which the effector functions of the Fc antibody stem are unnecessary (e.g. in vitro diagnostics) or even undesirable (e.g. to avoid removal by, and damage to, FcγR-bearing phagocytes during drug delivery). This fragmentation of mAbs – enzymatically (Porter, 1959) or, more commonly, using recombinant techniques (Cabilly et al., 1989) – has enabled the expression of smaller, often human-derived fragments in hosts such as E. coli. This has facilitated in turn the development of engineered, added-function reagents such as antibody-drug conjugates and antibody-radioactive isotope conjugates for targeted drug delivery (Hamblett et al., 2004; Senter, 2009) or in vivo imaging or monitoring of therapeutic activity (Goldenberg et al., 1978; Fung et al., 2016), respectively.

1.6 Antibody structure

Immunoglobulins or antibodies are antigen-binding proteins secreted by B cells as part of the immune response. They are capable of immense antigen-binding diversity through their highly variable regions, while maintaining the capacity to elicit and mediate a comparatively limited number of immune effector functions through their conserved regions. The genetic basis of this duality of complexity and function is now well understood, and described in detail elsewhere (Kim et al., 1981; Tonegawa, 1983). The typical antibody molecule is approximately 150 kDa, Y-shaped, and
representative of the IgG class (Figure 1.2 (A)). IgGs are heterodimers, composed of two identical heavy chains of approximately 50 kDa each and two light chains of approximately 25 kDa each; chains are covalently linked by interchain disulfide bonds. Each heavy and light chain contains an amino-terminal variable (V_H and V_L) region while the light chains also have a single constant domain (C_L) and the heavy chains each consist of three constant domains (C_H1; C_H2; C_H3). Variable and constant domains each contain approximately 100 amino acid residues and fold into a characteristic "immunoglobulin-like fold" that is stabilised by a single intradomain disulfide bond (Edelman, 1973).

The fragment crystallisable region, or Fc, of immunoglobulins consists of the C_H2 and C_H3 constant domains of the heavy chain, in the stem of the molecule. This region is responsible for immune responses such as activation of the complement cascade (Michl et al., 1979) or interaction with Fc receptor-bearing immune cells for activation (Chang, 1985; Nimmerjahn and Ravetch, 2008). The Fc stem can also increase the serum half-life of antibodies through the actions of the neonatal Fc receptor, effectively recycling IgGs and evading their proteolysis or degradation (Woof and Burton, 2004; Ward et al., 2005). Meanwhile, the "arms" of the IgG molecule, which mediate antigen binding, are the Fab or antigen-binding fragments.
and are composed of the light chains and the heavy chain $V_H$ and $C_H1$ domains, with the heavy and light chains joined by an interchain disulfide bond. The functions of the $C_H1$ and $C_L$ domains are thought to be to extend the Fab arms to allow their interaction with antigens and to increase rotation of the Fab arms via their flexible hinge regions (Wrigley et al., 1983).

The variable domains of the heavy and light chains each contain three hypervariable loops which determine the antigen binding specificity of the antibody molecule; as its structure is "complementary" to the bound antigen or epitope, this region is termed the complementarity determining region (CDR). The CDRs are interspersed in the variable domains between relatively conserved regions, known as framework regions, from which the hypervariable loops extend to form an antigen-binding pocket at the end of each arm of the IgG (Collis et al., 2003). In an antigen-binding event, at least four of the six hypervariable loops in each antigen-binding pocket typically make contact with the antigen epitope. Antibody diversity is achieved through the enormous variability between CDRs in different antibodies (Kabat, 1988; MacCallum et al., 1996), while throughout an immune response the affinity of antibodies produced by a B cell clone can also increase several fold through affinity maturation, which incorporates stochastic somatic hypermutation, followed by B-cell selection based on affinity and proliferation (Tas et al., 2016). Though extremely well suited to its role in immunosurveillance and immunoprotection, the modular structure of the antibody molecule also leads to clear possibility for its engineering using recombinant technologies to remove unnecessary elements, enhance desired properties, and tailor make antibody or antibody fragment variants to suit an application's needs.

### 1.7 Antibody fragments

Studies into reducing the size and composition of mAbs were originally driven by the need to obtain high affinity antigen-binding molecules without the immune effector components, as well as the inability of expression hosts such as *E. coli* to express recombinant full-length antibody molecules. Fragmentation of whole antibody molecules initially involved enzymatic digestion by pepsin, which generated a bivalent F(ab')$_2$ and a released Fc stem, or digestion by papain, which generated two homologous Fab units, each of which contained a single antigen-
binding pocket (Figure 1.2 (B)) (Porter, 1959). Recombinant DNA technology has enabled individual domains or combinations of domains to be reassembled in a variety of formats of varying size, affinity, and specificity, while typically retaining (or improving upon) the antigen-binding properties of the parent antibody (Skerra and Plückthun, 1988). This approach also allows for modifications to be made to molecules such as improved stability via engineering of disulfide bonds (Schumacher et al., 2013), or the introduction of affinity tags for purification or analysis (Nilsson et al., 1997). Meanwhile the considerably smaller size of antibody fragments favours their recombinant production and facilitates their application in fields such as enhanced tumour penetration (Yokota et al., 1992), drug delivery (Danhier et al., 2010), immobilisation and immunodetection (Bonroy et al., 2006; Hu et al., 2007).

Early studies with murine-derived V_H domains (Figure 1.2 (E)) (Ward et al., 1989) yielded functional single-domain fragments, but these were difficult to obtain in a soluble format and displayed a much reduced affinity compared to Fabs or their parent antibody (Holliger and Hudson, 2005). The isolation of single variable-like domains from camelids and cartilaginous fish have since yielded improved single domain antibodies (V_HH) with higher affinities (De Genst et al., 2004), but alternative, multi-domain fragments remain the sector leader as they typically perform better in antigen-binding and stability studies, and are more readily produced in recombinant forms.

Monovalent multi-domain antibody fragments were among the first recombinant antibody fragments developed, including Fabs, scFvs, dsFvs, and Fvs. Fv fragments are heterodimers consisting of non-covalently linked V_H and V_L domains (Figure 1.2 (D)) (Skerra and Plückthun, 1988). These are typically unstable due to the lack of an interchain disulfide bond between the heavy and light chain components and are prone to aggregation (Glockshuber et al., 1990). DsFvs use the same domain format but incorporate an interchain disulfide bond, which significantly increases stability of the molecule (Brinkmann et al., 1993). ScFvs, or single chain variable fragments, are heterodimeric fragments composed of a V_H and a V_L domain that are covalently joined by a flexible polypeptide linker, all of which is translated as part of a single peptide chain (Figure 1.2 (C)). ScFvs can be created
with V_H-V_L or V_L-V_H domain ordering and with a variety of linker lengths, but the typical linker length of 15-20 amino acids allows scFv monomer formation and avoids dimerisation (Desplancq et al., 1994; Wörn and Plückthun, 2001). ScFvs are an attractive and increasingly preferred antibody format as they are relatively robust and typically retain the affinity and specificity of their parent antibody, while their much-reduced size makes them suited to in vivo applications such as drug delivery, imaging, or immunotargeting, in which their size facilitates tissue penetration (Wu and Senter, 2005).

Multivalent antibody fragments are produced largely for specialised applications that require improved antigen binding efficacy (by avoiding loss of valency), combined with reduced size or the removal of unnecessary domains. Diabodies, triabodies, and tetrabodies have all been created to investigate the increased avidity afforded by their additional antigen-binding sites (Plückthun and Pack, 1997; Nuñez-Prado et al., 2015). They are typically engineered by shortening the inter-domain polypeptide linker, which encourages inter- rather than intra-molecular association of monomeric units into stable, multimeric, antigen-binding molecules (Cuesta et al., 2010), as well as exploiting peptide motifs to drive multimerisation, such as leucine zippers or Fos-Jun pairs (Kostelny et al., 1992; Plückthun and Pack, 1997). Using a similar approach, multiply-specific antibody fragments which possess binding sites for multiple antigens can be synthesized. Bispecific antibody fragments have been created in this manner to enhance avidity by binding more than one epitope on a target antigen (Neri et al., 1995) and in targeting the HIV-1 envelope (Asokan et al., 2015), while bispecific scFvs that bind epitopes from two different molecules have been used to recruit natural killer cells (potent cytotoxic lymphocytes that can kill tumour cells), to combat tumour growth, resulting in a potent antitumour effect (Schmohl et al., 2016). Whilst multivalent antibody fragments appear to present a powerful and unique opportunity for immuno-intervention in specific applications, their use can be complicated by poorer folding or increased aggregation during their recombinant production (Schmiedl et al., 2000), thereby lowering yields and increasing production costs.

1.8 Antibody fragment isolation and expression in E. coli

As previously discussed, E. coli is a simple, low-cost, and reproducible
expression platform for expression of small, low-complexity recombinant proteins. The variety of strains, plasmids, and expression methods developed with the host allows for successful expression of a wide range of recombinant proteins, including antibody fragments. Initial identification of antibody fragments that bind a target ligand can be carried out by a variety of methods in which antibody fragment genes are incorporated into a combinatorial gene library, followed by display of encoded proteins on bacteriophages, ribosomes, or cells (Smith, 1985; Clackson et al., 1991; Hanes et al., 1998; Wronska et al., 2016), and selection of those binding the target. Affinity maturation of antibody fragments by iterative rounds of selection in this manner can greatly increase the number of high-affinity fragments obtained in many of these approaches (Fodey et al., 2011). The process can also incorporate mutagenesis steps that increase the diversity of the fragments and allow isolation of binding fragments with increased affinity, as demonstrated by affinity improvement of the anti-fluorescein 4-4-20 scFv into the anti-fluorescein scFv 4M5.3, which was used in this work, with an increase of up to 1800 fold achieved using this approach (Boder et al., 2000; Midelfort et al., 2004; Midelfort and Wittrup, 2006). In vitro mutagenic techniques used in this manner can be random (Chowdhury and Pastan, 1999) or may target specific areas such as the CDRs in antibody fragments, to mimic the natural in vivo procedure of affinity maturation (Ho et al., 2005).

The use of E. coli as an expression host allows a range of well developed protein engineering techniques to be exploited to further enhance properties of the target scFv. The compact size and simple composition of scFvs allows for ease of integration of elements such as removable affinity tags for detection and purification with potentially minimal impact on downstream solubility and activity (Arnau et al., 2006; Hu et al., 2012). The oxidising environment and presence of disulfide bond isomerases in the periplasm leave it well suited to expression and folding of disulfide bond-containing antibody fragments (Zapun et al., 1995; Kolaj et al., 2009); scFvs are typically directed to the periplasm with a native E. coli N-terminal signal peptide such as ompA or pelB, as previously discussed (Section 1.3). The periplasmic space also aids scFv purification due to the lower number of bacterial proteins present while secretion to the extracellular medium is also possible, which may suit industrial-setting production of scFvs (Mukherjee et al., 2004). E. coli has been shown to be a reliable production host for scFvs with achievable yields of up to 450
mg/l, albeit in high-density cultures (Harrison and Keshavaraz-Moore, 1996), while yields in the range of 1-20 are more commonly achieved in shake-flask approaches (Kipriyanov et al., 1997; Studier, 2005; Kim et al., 2007), though these yields vary widely with the properties of the scFvs (Ge et al., 1995; Knappik and Plückthun, 1995).

Engineered scFvs are currently employed in a wide array of in vivo techniques. A humanised scFv-cytotoxic fusion protein, Gb-H22, was produced using an scFv specific for CD64, located on acute myeloid leukemia cells, and granzyme B, an active protease released by cytotoxic T cells and natural killer cells. This targeted fusion protein was found to be highly specific and as active as free granzyme B, thereby demonstrating the efficacy of functionalised scFvs in tumour targeting (Stahnke et al., 2008). Similar tumour-targeting scFv fusions such as 425(scFv)- and αHER2(scFv)-SNAP-AURIFs targeting breast cancer tumour markers have also shown their potency (Woitok et al., 2016). ScFvs benefit from increased tissue penetration and faster clearance rates due to their small size, which are exploited in the use of radiolabelled scFvs for imaging and delivery of radionuclides. Radiolabelled scFvs have been used to image carcinoembryonic antigen-producing cancers successfully (Begent et al., 1996) and to deliver γ-emitter radionuclides conjugated to tumour targeting scFvs (Kenanova and Wu, 2006).

ScFvs are increasingly employed in in vitro applications also, including immunosensors, immunoaffinity chromatography, and enzyme immunoassays such as enzyme-linked immunosorbent assays (ELISAs). Not only does their reduced size allow for greater surface densities of the immunoactive molecule compared to mAbs to be achieved but their ease of engineering allows for further enhancement of their application-relevant properties. Additional tags or labels which can allow for multicomponent detection through fluorescent (Ueda and Dong, 2014) or radiodetection can be incorporated (Waibel et al., 1999) as well as engineered attachment sites (Yoon et al., 2016) and recombinant tags (Hortiguera et al., 2015) to improve the orientation and stability, and reduce the leaching of immobilised scFvs from the immunosupport (Hu et al., 2007; Hu et al., 2013).

1.9 Immobilisation approaches

Immobilised antibodies and antibody fragments have been employed in a
range of in vitro and in vivo applications, such as target detection and quantification, ligand purification, delivery of an attached drug payload, and biofunctionalisation of surfaces (reviewed in Wronska et al., 2016). At its most basic level, coating a surface with antibodies or fragments can enable the capture of a target analyte for downstream processes; accordingly the size, specificity, stability, orientation, and additional functionalities of the immobilised antibody moiety all effect the efficacy of the assay or process. As previously discussed (Section 1.7 & 1.8), antibody fragments are attractive target-binding molecules due to their size, specificity, and capacity for modification compared to mAbs. The choice of immobilisation method plays a critical role in sensor design and their downstream performance in their surface-tethered state.

Adsorption of proteins onto surfaces relies on hydrophobic, hydrophilic and electrostatic forces, and is the most commonly employed method for antibody immobilisation in standard assays, such as immunoblotting and ELISAs (Jesionowski et al., 2014). This method can be attractive for some applications as it does not require specialised conditions, relies on passive processes, is cheap to perform, and does not require engineering of the associated antibody molecule (Jung et al., 2008). Adsorption is an unattractive option for assays requiring high sensitivity, specificity, and stability, however. Many proteins are subject to denaturation upon adsorption, with as little as only 5-10% of immobilised antibodies active after passive adsorption in some cases (Butler et al., 1993). In addition, antibodies adsorb to surfaces in many different orientations and therefore the antigen-binding pocket of individual molecules is frequently inaccessible, thereby potentially lowering the sensitivity and reproducibility of an immunoassay (Jung et al., 2008). Adsorbed antibodies can also form multiple protein layers on the surface, which reduces sensitivity and cost-effectiveness of the associated assay. Significant time and research are required to optimise adsorption steps to favour monolayer formation (Nojiri et al., 1993). Adsorbed antibodies can also desorb from supports over time, leading to their displacement in in vivo applications by molecules with a higher affinity for the employed surface (Avenaud et al., 2004) and a lower long-term stability of assays based on adsorbed molecules.

Antibodies and antibody fragments can be covalently attached to supports using approaches that optimise their stability and orientation on the surface (Kalia et
The inherent benefits of this tactic has made covalent immobilisation the preferred antibody attachment method in the development of immunoassays that require stringency and prolonged stability, such as in reusable and regeneratable immunosensors. Examples of these include highly specific immunosensors for the detection of tumour markers (Limbut et al., 2006; Zhang et al., 2008) or of staphylococcal enterotoxin A, the most commonly encountered toxin in *S. aureus* food poisoning outbreaks (Salmain et al., 2012).

Covalent immobilisation is carried out by reacting a functional chemical group on a surface with a functional chemical group on the protein molecule to form a covalent bond; covalent linking between a surface and a protein molecule is frequently referred to as bioconjugation (Srivastava et al., 2014). Covalent conjugation strategies can be random, in which case largely uncontrolled modification of protein sites results in heterogeneously activated molecules for interaction with the surface. When used in chemical or enzymatic cross-linking of proteins to surfaces, this type of approach can result also in intramolecular cross-linking within proteins or protein-protein interactions, thereby negatively affecting the conformation and activity of the immobilised protein (Danczyk et al., 2003).

Site-specific covalent modification approaches depend on greater chemoselectivity and regioselectivity to target sites, which are often unique within a protein molecule, followed by reaction with a specific functional group on the relevant surface (reviewed in (Srivastava et al., 2014)). Whilst many covalent bioconjugation techniques exploit naturally occurring cysteine groups, alternative naturally-occurring moieties such as amine, carboxyl, carbonyl, hydroxyl, and sulfhydryl groups are also frequently utilised as these are common in proteins and their relevant chemistries are well understood (Dugas et al., 2010). Alternatively, unique functional sites can be introduced into proteins by recombinant methods (Wu et al., 2009). Native chemical ligation is a selective bioconjugation technique which involves forming a thioester bond between an unpaired N-terminal cysteine residue in the protein molecule and in the support material, thereby providing a well-oriented stable bioconjugate for use in downstream analysis such as biosensor (Helms et al., 2007) or fluorescent probe development (Yuan et al., 2012). It is still a relatively new process for the development of bioconjugates and its potential application scope has yet to be realised in the analytical field (Gori and Longhi, 2016). Additional site-
specific conjugation avenues include incorporating unnatural amino acids into proteins via recombinant technologies thereby allowing greater control over the conjugation site and stoichiometry of the resultant bioconjugate than in the case of naturally occurring moieties (Deiters et al., 2004; Axup et al., 2012). The potential of this linking approach is demonstrated by its use to create a potent antibody-drug conjugate by introducing \( p \)-acetylphenylalanine to an antibody and antibody fragment for conjugation to an auristatin derivative via a stable oxime linkage (Axup et al., 2012; Currier et al., 2016).

### 1.10 Immobilised antibody fragments

As discussed in Section 1.7, antibody fragments exhibit properties of stability, size, binding affinity and engineering potential that make them better suited than their whole antibody counterparts to use in a number of applications. ScFvs are well suited to biosensor assays as their reduced size and absence of non-ligand-binding domains reduces the occurrence of non-specific interactions and background noise in assays (Backmann et al., 2005). They have been employed in label free biosensors, such as extremely sensitive piezoimmunosensors in which scFvs were engineered to contain a cysteine within the inter-domain linker for interaction with the gold surface of a quartz crystal microbalance, thereby forming a self-assembled monolayer with high specificity for the target antigen. Comparison to immobilised Fab and mAb molecules revealed the scFv to exhibit much reduced nonspecific binding while maintaining a high sensitivity to the target analyte in complex samples (Shen et al., 2005; Shen et al., 2007).

A range of scFvs with polymer-binding tags have also been engineered for the development of biosensors, such as scFvs with polystyrene-binding and poly(methylmethacrylate)-binding tags (Kumada et al., 2011; Kumada et al., 2014). Using this approach, scFvs are adsorbed to the appropriate surface but in the correct orientation by locating the polymer-binding tag distal to the antigen-binding site. Nanoparticle-scFv conjugates have been developed for improved targeting and delivery of the associated nanoparticle payload by linking tumour targeting scFvs to surface-enhanced Raman spectroscopy (SERS) nanoparticles for delivery of a chemotherapeutic agent or \textit{in vivo} spectroscopic detection (Qian et al., 2008), or for the delivery of small interfering RNA and microRNA simultaneously to
downregulate target genes involved in tumour growth and metastasis (Chen et al., 2010). Similarly, a tumour-targeting scFv was covalently conjugated via its terminal amino group to carboxyl functional groups on mesoporous silica nanoparticles to deliver the humanised antitumour monoclonal antibody Bevacizumab (Zhang et al., 2015). Delivery of Bevacizumab via this nanoparticle approach significantly improved its cellular uptake due to the specific targeting afforded by the scFv, which is of considerable benefit due to the severe systemic side effects the therapeutic can exhibit upon administration (Zhang et al., 2015).

The development of the C. jejuni pgl system in E. coli has enabled the addition of another easily incorporated tag into recombinantly produced proteins. The use of the short amino acid glycan acceptor sequence lowers the associated risk for structural perturbation of the scFvs compared to larger and more complex additions, such as protease-cleavable affinity tags or MBP. The attachment of the glycan to the scFv is highly site specific and homogenous, and optimal design and localisation of the attachment site in the scFv molecule can result in well-oriented, stably immobilised scFvs. Oxidation of the glycoprotein generates an aldehyde group which, upon reaction with an amine-coated surface, forms a stable, covalent imine linkage. This method has been employed in our group previously and presents a relatively simple method for the generation of stable, glycoscFv-based immunosensors (Hu et al., 2013; Hortigüela et al., 2015). This approach has also been used to coat stent surfaces with cell-specific scFvs capable of capturing epithelial progenitor cells via cell-specific scFvs (Foerster et al., 2016). These studies demonstrate the efficacy of the described glycan tag for directing scFv immobilisation and it is likely that this technique will become more widely employed due to the increasing ease with which glycosylated proteins can be generated in recombinant expression systems.

1.11 Scope of this thesis

The main objective of this study was to investigate engineering of the C. jejuni protein glycosylation system pgl in E. coli for avenues to modulate the size of the glycan chain attached to the target scFv, and to characterise the effects of the resultant changes on protein activity when covalently immobilised via the attached glycan. The first objective was to investigate by mutagenesis the activity and
molecular mechanism of action of the glycosyltransferase PglJ, as described in Chapter 3. The second objective was to study the effect of the truncated glycan compared to the native *C. jejuni* heptasaccharide on covalent immobilisation of the scFv, as outlined in Chapter 3. The third objective was to combine lectin-based analyses with available microscopy techniques to characterise the glycan changes introduced in Chapter 3 and to determine the effectiveness of immobilising scFvs covalently via the glycan chain (Chapter 4). The fourth objective was to investigate extension of the native *pgl*-produced heptasaccharide by recombinant expression of the *C. jejuni* galactosyltransferase, CgtB, as described in Chapter 4. Finally, the principle conclusions of this research and potential future research directions are discussed in Chapter 5.
Chapter 2:
Materials and Methods
2.1 Materials

2.1.1 Suppliers and reagents

All materials were provided by the following suppliers unless otherwise stated:

Abcam, 330 Cambridge Science Park, Cambridge, CB4 0FL, U.K.
BDH Merck Ltd, Glasgow, U.K.
Beckman Coulter Inc., Brea, California, U.S.A.
Bioline Reagents Limited, Unit 16 The Edge Business Centre, Humber Road, London, NW2 6EW, U.K.
Bio-Rad Laboratories, Hemel Hempstead, Hertfordshire, U.K.
Corning Life Sciences, 836 North Street, Building 300 Suite 3401, Tewksbury, MA, 01876, U.S.A.
Cruinn Diagnostic Ltd., 5b/6b Hume Centre, Park West Industrial Estate, Nangor Road, Dublin 12, Ireland.
Expedeon Ltd., Unit 12, Buckingway Business Park, Anderson Road, Swavesey, Cambridgeshire, CB24 4AE, U.K.
GE Healthcare, Little Chalfont, Buckinghamshire HP7 9NA, U.K.
Greiner Bio-One GmbH, Bad Haller Str. 32, 4550 Kremsmünster, Austria.
IBA GmbH, Rudolf-Wissel-Str. 28, 37079 Goettingen, Germany.
Invitrogen Life Technologies Ltd., Inchinnan Business Park, Paisley, U.K.
Jena Bioscience GmbH, Loebstedter Strasse 71, D-07749 Jena, Germany.
Medical Supply Company Ltd., Damastown, Mulhuddart, Dublin 15, Ireland.
Melford Laboratories, Bildeston Road, Chelsworth, Ipswich, Suffolk, IP7 7LE, U.K.
Merck Millipore Ltd, Watford, Hertfordshire, U.K.
MWG Biotech AG.,Anzinger Str. 7a, 85560, Ebersberg, Germany.
MyBio Ltd., Kilkenny Research & Innovation Centre, St. Kieran’s College, Kilkenny, Ireland.
New England Biolabs Ltd. (NEB), Hitchin, Hertfordshire, SG4 0TY, U.K.
Nippon Genetics Europe GMBH, Binsfelder Strasse 77, 52351 Dueren, Germany.
Novagen, 441 Charmany Drive, Madison, Wisconsin 53719, U.S.A.
Qiagen Ltd, Fleming Way, Crawley, West Sussex , RH10 9NQ, U.K.
R&D System, Inc., 614 McKinley Place NE, Minneapolis, MN 55413, U.S.A.
Roche Diagnostics Ltd., East Essex, U.K.
2.1.2 Bacterial strains, plasmids and primers

The bacterial strains used in this study are presented in Table 2.1. Glycerol stocks of strains were prepared in 2 ml cryovials by addition of 1.125 ml of overnight culture to 0.375 ml of 60 % glycerol (v/v) to provide a final concentration of 15 % glycerol. Glycerol stocks were then snap frozen in liquid nitrogen and stored at -80 °C. Recovery of strains from glycerol stocks was achieved by streaking on LB agar with appropriate selective agents and grown overnight at 37 °C. Single colonies were subsequently used to inoculate liquid overnight cultures for downstream use. Plasmids were sourced as described in Table 2.2. All primers used in this study were produced by Eurofins Genomics.

<table>
<thead>
<tr>
<th>E. coli strain</th>
<th>Genotype</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLM37</td>
<td>W3110 (F λ rph-1 INV[rrnD, rrnE]) ΔwecA</td>
<td>Strain used for glycoprotein expression. wecA deletion interrupts native GlcNAc transferase</td>
<td>Professor Markus Aebi, ETH Zurich, Switzerland</td>
</tr>
<tr>
<td>TOP10</td>
<td>F mcrA D (mrr-hsdRMS-mcrBC) f80lacZDM15 DlacX74 deoR recA1 araD139 D (ara-leu) 7697 galU galK rpsL (StrR) endA1 nupG</td>
<td>Designed for cloning and plasmid propagation</td>
<td>Invitrogen</td>
</tr>
</tbody>
</table>
Table 2.2  Plasmids used in this study.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pIG6-4M5.3-gly</td>
<td>pIG6 expression vector utilising ompA leader sequence for periplasmic direction; lac promoter; F1-based origin of replication; Amp&lt;sup&gt;R&lt;/sup&gt;; insert – 4M5.3-gly: anti-fluorescein scFv with an N-terminal his-tag and a C-terminal N-linked glycosylation tag</td>
<td>Ge et al. 1995, Hu et al. 2013</td>
</tr>
<tr>
<td>pIG6-2H12-gly</td>
<td>As above but with anti-domoic acid scFv 2H12 in place of 4M5.3-gly scFv</td>
<td>Hu et al. 2005</td>
</tr>
<tr>
<td>pIG6-2H12-NeuAc</td>
<td>pIG6-2H12-gly including additional genes from Campylobacter jejuni OH4384 LOS locus to investigate sialylation</td>
<td>Internal lab vector</td>
</tr>
</tbody>
</table>
Chapter 2

<table>
<thead>
<tr>
<th>Oligo name</th>
<th>Sequence</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYCpgl</td>
<td>Vector containing <em>C. jejuni</em> glycosylation <em>pgl</em> locus for synthesis, assembly, and attachment of heptasaccharide to target protein</td>
<td>Wacker et al. 2002</td>
</tr>
<tr>
<td>pET22 (+)</td>
<td>Expression vector with T7 promoter and terminator flanking MCS; <em>pelB</em> leader sequence for subcellular targeting; Amp(^R);</td>
<td>Novagen</td>
</tr>
<tr>
<td>pPICZαA</td>
<td>Vector designed for expression, secretion, and purification of recombinant proteins in <em>P. pastoris</em> and for cloning in <em>E. coli</em>; 6xHis tag; <em>c-myc</em> epitope; Zeo(^R).</td>
<td>Invitrogen</td>
</tr>
</tbody>
</table>

### Table 2.3

Primer used in this study.

<table>
<thead>
<tr>
<th>Oligo name</th>
<th>Sequence</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>plG4M5.3preEcoRV</td>
<td>ACCGTAGCGCAGGGCC</td>
<td>Forward primer for <em>EcoRI</em>-HindIII exchange in cloned genes</td>
</tr>
<tr>
<td>plG4M5.3EcoRIr</td>
<td>CAGGTCGAATTTCTTATTATTAT</td>
<td>Reverse primer for <em>EcoRI</em>-HindIII exchange in cloned genes</td>
</tr>
<tr>
<td>plG4M5.3EcoRIrl</td>
<td>TTCACAGGTCGAATTCTTATTATAG</td>
<td>Redesigned reverse primer for HindIII replacement</td>
</tr>
<tr>
<td>pIGSialClustAmpF</td>
<td>ACCCGCCCGCGCTTTAAGGC</td>
<td>Sialylation genes amplification primer - forward</td>
</tr>
<tr>
<td>pIGSialClustAmpR</td>
<td>AATGGGCGCACATTGTGCC</td>
<td>Sialylation genes amplification primer - reverse</td>
</tr>
<tr>
<td>cgtBFor</td>
<td>CTAAAGAATTCTCTGACCTACCATCACCACCACCACCACCACAACATGGAACG</td>
<td>Amplification of <em>cgtB</em> from plG6-2H12-NeuAc4, adding BamHI/EcoRV for cloning and a C-terminal strep tag - forward</td>
</tr>
<tr>
<td>cgtBRev</td>
<td>TATGGGATCTGTGGCAACATGTTGGCAACATGTTGGGAAATTCTGGC</td>
<td>Amplification of <em>cgtB</em> from plG6-2H12-NeuAc4, adding BamHI/EcoRV for cloning and a C-terminal strep tag - reverse</td>
</tr>
<tr>
<td>OmpCgtForOver</td>
<td>TCCGGCGATGTGTTATTGGTTATGGTTAAAAATTTTCAA</td>
<td>Primer overlap of <em>ompA</em> to <em>cgtB</em> - forward</td>
</tr>
<tr>
<td>OmpCgtRevOver</td>
<td>TGAAAAATTAAACATACGACATCGACGGA</td>
<td>Primer overlap of <em>ompA</em> to <em>cgtB</em> - reverse</td>
</tr>
<tr>
<td>OmpAmpFor</td>
<td>TACGAAATTTCTAGATACCCGACCATCGCGA</td>
<td><em>ompA</em> amplification primer - forward</td>
</tr>
<tr>
<td>cgtBHindRev</td>
<td>CCCTTTATGAAGCTTTGTTTCGTCGCAAC</td>
<td>HindIII addition to <em>cgtB</em>-</td>
</tr>
<tr>
<td>Primer Name</td>
<td>Sequence Information</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>cgtB-pET22For</td>
<td>CTCCGGCGGGATCCGTTAAAAATTTCGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cloning of cgtB to pET22 (+) - forward</td>
<td></td>
</tr>
<tr>
<td>cgtB-pET22Rev</td>
<td>ATTTTTTCAGCGGCCGCTTTTCAAAACTGCG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cloning of cgtB to pET22 (+) - reverse</td>
<td></td>
</tr>
<tr>
<td>a827c</td>
<td>GAATTTTTTCGTTTTGCTTTTTCAAACTGCG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>First single site mutagenesis for pglJ primer - forward</td>
<td></td>
</tr>
<tr>
<td>a827c_anti</td>
<td>CTCTTCAATTTAAAACATTAGGTGTGCGAAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>First single site mutagenesis for pglJ primer - reverse</td>
<td></td>
</tr>
<tr>
<td>pACYCseq</td>
<td>AGAGTTAAAAATTAGAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sequencing primer for pglJ mutagenesis</td>
<td></td>
</tr>
<tr>
<td>pACYCseq2.0</td>
<td>GGTGAAGGTTGCTTAAAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Improved sequencing primer for pglJ mutagenesis</td>
<td></td>
</tr>
<tr>
<td>pPICZAcgtBFor</td>
<td>CTCCGGCGATGCTGATTTGCTTCAAATCATTTATTTTTCAATCATCTTAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cloning of cgtB to pPICZα - forward</td>
<td></td>
</tr>
<tr>
<td>pPICZAcgtBRev</td>
<td>ATTTTTTTTCAAAACTGCGGGATGTCTAGAAACAATTCTTTAAATTTTCAATCATCTTAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cloning of cgtB to pPICZα - reverse</td>
<td></td>
</tr>
<tr>
<td>pglJmuta2.0For</td>
<td>GTTATTGCAAGCTTAGGGTTGCTTGAAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second single site mutagenesis primer for pglJ mutagenesis - forward</td>
<td></td>
</tr>
<tr>
<td>pglJmuta2.0Rev</td>
<td>CAAGCTAAGGCTTGCAATTAAAACATTTGGAAGGCC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second single site mutagenesis primer for pglJ mutagenesis - reverse</td>
<td></td>
</tr>
<tr>
<td>pglJStopFor</td>
<td>GAGGTGCAATAACATTATTATAGGTGCGAATTACACTAGGAAAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Translation termination mutagenesis of pglJ - forward</td>
<td></td>
</tr>
<tr>
<td>pglJStopRev</td>
<td>CTAGGTCTTGGCAATTACATTTGGAAGGCC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Translation termination mutagenesis of pglJ - reverse</td>
<td></td>
</tr>
<tr>
<td>pglJknockseq</td>
<td>GGTGGTGCTGAAAGAGTTC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primer for sequencing of translation termination site of pglJ</td>
<td></td>
</tr>
<tr>
<td>cgtBRev13</td>
<td>CATTGAAAAACACTAATCAAACACTCTC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transcription termination check on cgtB in <em>P. pastoris</em>; (short) - reverse</td>
<td></td>
</tr>
<tr>
<td>cgtBRev32</td>
<td>AAAATTTTTTTTATAATACATTTGGAAGGCC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primer for transcription termination check on cgtB in <em>P. pastoris</em>; (middle) - reverse</td>
<td></td>
</tr>
<tr>
<td>B-actinFor</td>
<td>GGTATTTGCTGAGCGTATGAAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Housekeeping gene sequencing in <em>P. pastoris</em> - forward</td>
<td></td>
</tr>
<tr>
<td>B-actinRev</td>
<td>CCACCGATCCATACG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Housekeeping gene</td>
<td></td>
</tr>
</tbody>
</table>
### Chapter 2

<table>
<thead>
<tr>
<th>Primer/Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAGTACT</td>
<td>Sequencing in <em>P. pastoris</em> - reverse</td>
</tr>
<tr>
<td>pPICZAcgtBFor</td>
<td>Alternative primer for cgtB transcription confirmation - forward</td>
</tr>
<tr>
<td>pgIHmutFor</td>
<td>Translation termination mutagenesis of pgI - forward</td>
</tr>
<tr>
<td>pgIHmutRev</td>
<td>Translation termination mutagenesis of pgI - reverse</td>
</tr>
<tr>
<td>pgIHseq</td>
<td>Sequencing for pgI mutagenesis</td>
</tr>
</tbody>
</table>

#### 2.1.3 Antibiotics

Antibiotics were used for strain and plasmid selection as follows unless otherwise stated: ampicillin (100 µg ml\(^{-1}\)) in Milli-Q H\(_2\)O; chloramphenicol (34 µg ml\(^{-1}\)) in 100 % EtOH; Zeocin™ (25-200 µg ml\(^{-1}\)) in Milli-Q H\(_2\)O.

#### 2.1.4 Antibodies, lectins and detection reagents

Monoclonal anti-polyhistidine peroxidase-conjugated antibody, produced in mouse (Sigma-Aldrich), was used in western blotting and ELISA analysis at 1:1000 and 1:3000, respectively. *Strep*MAB-Classic conjugated to horseradish peroxidase was used in western blotting at 1:320000 (IBA Life-Sciences). Streptavidin-HRP was used in lectin blotting at 1:5000-1:10000 (R&D Systems). Biotinylated lectins were kindly provided by Dr. Kilcoyne of the Alimentary Glycoscience Research Cluster, NUI Galway and were typically employed at 1-2 µg ml\(^{-1}\).

#### 2.1.5 Molecular weight markers

Hyperladders I & IV from Bioline were used in agarose gel electrophoresis. PageRuler Plus Prestained Protein Ladder from Thermo Scientific was used in SDS-PAGE analysis.

#### 2.1.6 General multi-component buffers

All solutions were made using Milli-Q or distilled water unless otherwise stated.

**PBS:** 137 mM NaCl; 2.7 mM KCl; 10 mM phosphate buffer
\( \text{Na}_2\text{HPO}_4: \text{KH}_2\text{PO}_4 \); pH 7.4

**TBS**: 20 mM Tris; 100 mM NaCl; pH 7.4

**TBST**: TBS containing 2 % Tween-20

**Lectin buffer**: 20 mM Tris; 100 mM NaCl; 1 mM CaCl\(_2\); 1 mM MgCl\(_2\); 0.05 % Tween-20; pH 7.4

### 2.1.7 Protein purification buffers

All buffers used in nickel-Immobilised Metal ion Affinity Chromatography (IMAC) were adjusted to pH 7.4 and filtered with a 0.45 µm cellulose acetate filter. The following buffers were based on the IMAC buffer (20 mM sodium phosphate buffer, 0.5 M NaCl) with additions as listed: equilibration buffer (10 mM imidazole; 2 % Tween-20); wash buffers (20-100 mM imidazole; ± 2 % Tween-20); elution buffers (200-500 mM imidazole).

### 2.1.8 Culture media

All culture media was autoclaved at 121 °C for 15 min in a Labo autoclave (Sanyo) unless otherwise stated. Filter-sterilisation was performed using a 0.22 µm syringe filter (Sarstedt Filtropur).

#### 2.1.8.1 Lysogeny broth (LB)

LB broth was prepared by the addition of 20 g of LB broth powder (Sigma) per 1000 ml dH\(_2\)O. LB agar was prepared by the addition of 15 g of Agar powder (Sigma) per 1000 ml LB broth.

#### 2.1.8.2 ZYP-5052

ZYP-5052 is a defined medium described by Studier *et al.* (2005) and was prepared from the following stocks:

**ZY**

ZY was prepared by the addition of 10 g N-Z Amine® A powder (Sigma) and 5 g yeast extract powder (Sigma) per 1000 ml.

**NPS (20X)**
The complex phosphate buffer was prepared by the addition of 66 g \((\text{NH}_4)_2\text{SO}_4\) (ammonium sulfate) (Sigma), 136 g \(\text{KH}_2\text{PO}_4\) (monopotassium phosphate) (Sigma), 177.99 g \(\text{Na}_2\text{HPO}_4.2\text{H}_2\text{O}\) (di-sodium hydrogen phosphate dihydrate) (Sigma) per 1000 ml dH\(_2\)O.

5052 (50X)

The carbon and energy source solution, 5052, was prepared by the addition, in order, of 250 g glycerol (VWR), 730 ml dH\(_2\)O, 25 g glucose (Sigma), 100 g \(\alpha\)-lactose monohydrate (Sigma). The solution was filter sterilised and stored at 4 °C.

To make a 1 litre working solution of ZYP-5052, 928 ml of ZY, 1 ml of MgSO\(_4\) (1 M), 20 ml of 5052 (50X), 50 ml of NPS (20X), and appropriate amounts of required antibiotics for selection of bacterial strains and plasmids, were mixed and made up to 1 litre with ZY.

2.1.8.3 **Pichia pastoris media**

The following defined media were prepared as outlined in the *Pichia* Expression Kit user manual (Invitrogen, 2014): low salt LB medium; yeast extract peptone dextrose medium (YPDS); buffered glycerol complex medium (BMGY); buffered methanol-complex medium (BMMY).

2.2. **Methods**

2.2.1 **Bacterial growth and protein expression**

All overnight cultures were prepared with 5 ml of LB with the appropriate antibiotic(s) and inoculated with a single colony from LB agar plates unless otherwise stated. Shaking of bacterial cultures was carried out at 225 rpm unless otherwise stated.

2.2.2. **Electrocompetent cell preparation**

Cells were streaked from glycerol stocks on LB agar plates containing the required antibiotics, and incubated overnight at 37 °C. A single colony was used to inoculate 10 ml LB containing the required antibiotics and grown overnight at 37 °C. The overnight culture was diluted 40-fold into 400 ml LB in 2 l Erlenmyer flasks and grown with shaking at 37 °C. Upon reaching an OD\(_{600}\) of 0.4, cells were placed on
ice for 20 min in pre-chilled centrifuge tubes. Cells were harvested by centrifugation at 4000 g, 4 °C for 20 min using a Beckman Avanti J-20 XP and JLA-16.250 rotor. Cells were resuspended in ice-cold 400 ml 1 mM HEPES, pH 7.0, followed by two centrifugation and wash steps with ice-cold 10 % glycerol. The final cell pellet was resuspended in 500-1000 µl 10 % glycerol, and stored for future use by snap freezing in liquid nitrogen in aliquots of 55 µl and transferring to -80 °C.

2.2.3  Transformation by electroporation

Electrocompetent cells as prepared above (Section 2.2.2) and the desired plasmid DNA for transformation into cells were maintained on ice. Electroporation cuvettes from Cell Projects Ltd. with a 1 mm gap were chilled on ice for 20 min. A 150-400 ng amount of DNA was mixed gently with 50 µl thawed cells and stored on ice for 1 min. The mixture was pipetted into the cuvette and pulsed at 1800 V in Eppendorf Electroporator 2510. Cells were immediately recovered by addition of 450 µl pre-warmed LB, and transferred to a sterile 1.5 ml centrifuge tube and incubated at 37 °C with 250 rpm shaking for 1 h. After recovery, transformed cells were concentrated by centrifugation at 2500 g, resuspended in a 100-300 µl volume of fresh LB depending on the required number of spread plates, and spread on LB agar plates with appropriate antibiotics, followed by incubation overnight at 37 °C.

2.2.4  Transformation by heat-shock

Chemically competent cells obtained from listed suppliers were thawed on ice. A 200-400 ng amount of DNA was added to 50 µl thawed cells and incubated on ice for 30 min. The transformation mixture was incubated at 42 °C in a water bath for 45 sec followed by storage on ice for 2 min. Pre-warmed LB was then added to a final volume of 500 µl LB and incubated at 37 °C with 250 rpm shaking for 1 h. After recovery, transformed cells were plated as in Section 2.2.3.

2.2.5  Protein expression by induction via IPTG

The OD$_{600}$ of overnight cultures was measured and used to inoculate 50-500 ml LB containing appropriate antibiotics in baffled Erlenmyer flasks to a starting OD$_{600}$ of 0.05. Cultures were incubated at 37 °C with 225 rpm shaking, with periodic monitoring of OD$_{600}$. Upon reaching an OD$_{600}$ of 0.5, protein expression was induced by the addition of 0.1-1.0 mM isopropyl β-D-1-thiogalactopyranoside.
(IPTG). Protein expression was allowed to progress for periods of 3 h, 5 h, or 16 h. Cultures were subsequently harvested by centrifugation at 8000 g and pellets were stored at -20 °C for future analysis.

2.2.6 Auto-induction

The OD_{600} of overnight cultures was measured and used to inoculate 50-500 ml ZYP-5052 containing appropriate antibiotics in baffled Erlenmyer flasks to a starting OD_{600} of 0.05. Cultures were incubated at 37 °C with 225 rpm shaking, with periodic monitoring of OD_{600}. Upon reaching an OD_{600} of 0.9-1.0, cultures were incubated at 25 °C and protein expression was allowed to continue for 16-48 h. Cultures were subsequently harvested by centrifugation at 8000 g and pellets were stored at -20 °C for future analysis.

2.2.7 P. pastoris growth and protein expression

P. pastoris GS115 cells were grown from glycerol stocks and prepared for electroporation according to the pPICZα user manual. All expression protocols and media compositions were derived from the Pichia Expression Kit User Guide (Invitrogen, 2014). Following electroporation, successful transformants were identified by colony PCR and glycerol stocks were prepared from these and stored at -70 °C. Cell growth and induction was carried out using BMGY and BMMY, respectively, and was supplemented with 0.002 % histidine. Initial expression analyses were carried out in 5 ml cultures, followed by scaled up expression as required. Culture supernatants and cells were screened for the presence of expressed proteins by SDS-PAGE and western blotting.

2.2.8 Protein extraction

2.2.8.1 Periplasmic protein extraction from E. coli

All reagents were pre-chilled on ice or at 4 °C and the volumes used were based on the original culture volume. Cell pellets were recovered from -20 °C and allowed to thaw on ice. Cell pellets were resuspended in 1 l PBS, pH 7.4 / 1 of expression followed by centrifugation at 8000 g, 4 °C, for 20 min. Cell pellets were then resuspended in 30 ml l^{-1} 0.75 M sucrose, 100 mM Tris-HCl, pH 7.5 followed by the addition of 60 ml l^{-1} 1 mM EDTA and incubation at room temperature for 10 min.
with rocking. A 60 U l⁻¹ aliquot of rLysozyme was added, followed by incubation for 30 min at room temperature and 1 h on ice with rocking. Subsequently 8.6 ml l⁻¹ 0.25 M MgCl₂ was added followed by 5 mg l⁻¹ DNase I and incubation for 15 min on ice. The entire solution was centrifuged at 8000 g, 4 °C for 10 min and the resulting supernatant was stored as the “soluble” periplasmic extract. The pellet was resuspended in 200 ml PBS per litre of expression and centrifuged at 8000 g, 4 °C for 10 min and the resulting supernatant was stored as the “resoluble” periplasmic extract. The remaining pellet was resuspended in 200 ml PBS per litre of expression and samples were taken as the “insoluble” sample. Soluble and resoluble samples were subsequently filtered through a 0.45 µm cellulose acetate filter and stored at 4 °C for further analysis.

2.2.8.2 Protein extraction from P. pastoris

All reagents were prepared according to the Pichia Expression Kit User Guide (Invitrogen, 2014) for the isolation of intracellular and secreted protein. Samples were taken at timepoints throughout expression experiments and subsequently centrifuged at 13000 g for 3 min. Supernatants from centrifuged samples were stored as the secreted fraction and remaining cell pellets were stored at -80 °C. Intracellular expression was analysed from whole cell samples by thawing pellets, resuspending in 100 µl ml⁻¹ breaking buffer, and addition of an equal volume of 0.5 mm acid-washed glass beads (Sigma). Samples were subjected to 8 cycles of vortexing for 30 seconds followed by incubation on ice for 30 seconds, following which samples were centrifuged at 13000 g for 10 min and the resultant supernatant was stored for analysis as the intracellular protein fraction.

2.2.9 Protein purification and cleanup

Dialysis was used to effect buffer exchange of protein-containing samples and utilised high retention dialysis tubing, MWCO 12400 (Sigma) or Slide-A-Lyzer™ dialysis cassettes, MWCO 3500 (Thermo Scientific) for volumes greater or less than or 3 ml, respectively. Dialysis was carried out from 4-16 h at 4 °C with stirring. Concentration of proteins was carried out using Vivaspin 2 concentrators, 10 kDa MWCO, according to the manufacturer’s instructions. Protein purifications were typically carried out at 4 °C.
2.2.9.1 Affinity chromatography using immobilised nickel

Protein solutions were dialysed against 5 l IMAC buffer prior to purification by IMAC. Columns used were 1 ml HisTrap™ HP (GE Healthcare), charged with Ni\(^{2+}\) ions. Columns were stripped and recharged between runs as per the recommended protocol (GE Healthcare, 2005). IMAC was carried out either using disposable syringes or a peristaltic pump (IDEX REGLO Digital MS-4/8). All buffers were stored on ice prior to and during IMAC; pump tubing and HisTrap columns were immersed in ice where possible. Post-dialysis, protein solutions were equilibrated in 10 mM imidazole containing 2 % Tween-20. Columns were washed with 15 ml equilibration buffer prior to loading of the protein solution. Protein loading and elution were carried out at 1 ml min\(^{-1}\) and wash steps were carried out at a flow rate of 1-2 ml min\(^{-1}\).

After protein loading, 1 ml of wash buffers 1 and 2 were applied in tween to the column per ml of protein solution, followed by 1 ml wash buffer 3 per 5 ml protein solution and up to 3 ml wash buffer 4. Elution of his-tagged proteins was carried out by application of elution buffers 1 and 2 and eluate was collected in 1 ml fractions. Eluate was stored at 4 °C and protein-containing eluates were dialysed against PBS prior to immediate use or frozen in 50-1000 µl aliquots for storage.

2.2.9.2 Affinity chromatography using cobalt resin

Protein solutions were dialysed against PBS prior to purification using cobalt resin. Post-dialysis, protein solutions were equilibrated in 5 mM imidazole containing 2 % Tween-20. Columns used were Thermo Scientific Pierce Disposable Plastic Column with 0.5 ml resin-bed volumes. All buffers were stored on ice prior to and during affinity purification. Columns were prepared by the addition of 0.5-1.0 ml 6 % Cobalt-IDA Agarose (Jena Bioscience), followed by elution of any residual ethanol post-settling. All buffer and protein solutions were allowed to settle in the column prior to elution.

Columns were prepared for protein addition by washing with 10 column volumes (CV) of PBS. Protein solutions were loaded and passed through the column up to three times to allow for maximum protein binding. The resin was washed by the addition of 15 CV of PBS containing 15 mM imidazole, followed by elution
using 5 CV of PBS containing 300 mM imidazole in 0.5 ml steps. Elutates were stored at 4 °C until analysed and protein containing eluates were dialysed against PBS for immediate use or were snap frozen in 50-1000 µl aliquots for storage.

2.2.10 Protein processing and analysis

2.2.10.1 SDS-PAGE

SDS-PAGE was carried out using a dual mini slab electrophoresis system (ATTO Corporation). Protein separation was carried out on 12 % Bis-Tris gels following a standardised protocol (Laemmli, 1970). Samples for SDS-PAGE were prepared with either 2X or 5X sample reducing buffer prepared according to Roche Lab FAQs manual (Roche Applied Science, 2011).

2.2.10.2 Protein staining

Following SDS-PAGE, gels were rinsed with dH₂O and 5 ml InstantBlue (Expedeon) was added. Staining was allowed to proceed for 30 min to 4 h at room temperature with rocking. Removal of excess InstantBlue™ solution was achieved by multiple washes in dH₂O, followed by imaging using an Epson Perfection scanner.

2.2.10.3 Western blotting

Following SDS-PAGE, gels for blotting were transferred onto Amersham Protran™ 0.2 µm nitrocellulose blotting membrane (GE Healthcare) in a semi-dry transfer apparatus at 60 mA for 45 min. Western blotting was carried out by standard protocol (Burnette, 1981). Blocking was carried out with a 5 % skim milk powder solution, while protein detection was achieved using an enzyme-conjugated anti-tag (his/strep) antibody. Colour was developed using 3,3’,5,5’-tetramethylbenzidine (TMB, Sigma) as substrate. Membranes were imaged using an Epson Perfection scanner.

2.2.10.4 Glycan detection/staining

Glycan detection/staining was carried out utilising the DIG Glycan Detection Kit (Roche) and based on method A in the manufacturer's instructions, using 0.1-10 µg protein. For each analysis the same concentration of protein was used for all
samples. Glycoprotein was dialysed or diluted 1:1 in a total volume of 20 µl 0.1 M sodium acetate buffer prior to oxidation. Oxidation was carried out by the addition of 2 µl sodium meta-periodate solution to samples, followed by mixing and incubation in the dark for 20 min at room temperature. Excess periodate was destroyed by the addition of 10 µl sodium disulfite solution followed by mixing and incubation for 5 min at room temperature. Hydrazide binding was achieved by adding 5 µl DIG-3-O-succinyl-ε-aminocaproic acid hydrazide and incubating for 1 h at room temperature. Samples were prepared for analysis by the addition of 15 µl 4X SDS sample buffer and incubation at 98 °C for 8 min.

Denatured samples were separated by SDS-PAGE and transferred to nitrocellulose membranes as detailed in Section 2.2.10.3. Membranes were blocked with 5 % skim milk powder solution in TBS for 1 h and washed 3 times with 50 ml TBS for 10 min each. Membranes were incubated with alkaline phosphatase-conjugated anti-digoxigenin antibody (10 µl conjugate in 10 ml TBS) at room temperature for 1 h and washed 3 times with 50 ml TBS for 10 min each. Staining was achieved by incubation with 10 ml staining solution: 200 µl NBT/X-phosphate solution in 10 ml TBS, pH 9.5. Detection was allowed to occur from a few minutes to overnight depending on development of the signal, followed by several washes with distilled water to stop the reaction. Membranes were imaged using an Epson Perfection scanner.

2.2.10.5 Lectin blotting

For analysis, 1-2 µg purified proteins were denatured and separated by SDS-PAGE and transferred onto Amersham™ Protran™ 0.2 µm nitrocellulose blotting membrane (GE Healthcare) as described above. Membranes were subsequently blocked for 1-16 h with shaking (4 °C for incubations above 1 h, room temperature for incubations of up to 1 h) in 1 % periodate-treated Bovine Serum Albumin (pBSA) in lectin buffer. After blocking, the lectin of interest was immediately added to the membrane at a concentration of 1-2 µg ml⁻¹ and incubated for up to 2 h at room temperature. After lectin incubation, the membrane was washed three times in TBST and then incubated with streptavidin-HRP at 1:5000-1:10000 for 1 h at room temperature. Membranes were then washed twice in TBST and once in TBS followed by development with TMB. Membranes were imaged using an Epson
2.2.10.6 Protein quantification

Concentrations of purified proteins were determined using a NanoDrop 2000c spectrophotometer (Thermo Scientific) using 2 µl of samples and based on the calculated molecular extinction coefficients and molecular weights of the relevant proteins. Alternatively, a Bio-Rad DC™ Protein Assay Kit was utilised, with protein concentrations calculated from a linear fit trend line of BSA.

2.2.10.7 Glycoprotein oxidation

Samples to be oxidised were equilibrated in 0.1 M sodium acetate buffer, pH 5.5 by dialysis. A stock solution of sodium *meta*-periodate (Thermo Scientific) was prepared in 0.1 M sodium acetate buffer and added to the protein solution to a final concentration of 10 mM. Oxidation was carried out in the dark for 30 min at room temperature with rocking. Protein solutions were immediately dialysed to the appropriate buffer.

2.2.11 Enzyme-Linked Immunosorbent Assay (ELISA)

2.2.11.1 Indirect ELISA

All analyses were performed in triplicate. Wash steps were carried out with 200 µl 0.45 µm-filtered PBS; for three and five wash-steps, the first or first two washes, respectively, contained 0.02 % Tween-20. BSA and BSA-Fluorescein isothiocyanate (FITC) conjugate were immobilised in individual wells of a Nunc MaxiSorp® flat-bottom 96 well plate (Thermo Scientific) at 10 µg ml⁻¹ overnight at 4 ºC. After washing three times, wells were blocked with 250 µl 2 % BSA containing 0.02 % Tween-20 in PBS for 60 min at 37 ºC, followed by three washes. A 50 µl volume of purified scFv, at concentrations from 10-0.001 µg ml⁻¹, was incubated on plates at 25 ºC for 1 hr, followed by three washes. Bound scFvs were detected using peroxidase-conjugated anti-polyhistidine antibody (Sigma), diluted 1:3000 in 1 % BSA and 0.02 % Tween-20 in PBS, for 60 min at 37 ºC. Plates were washed five times and reactions developed with 50 µl TMB for up to 10 min at room temperature. Colour development was stopped with the addition of 50 µl of 1 N H₂SO₄. The optical density of each well was then measured at 450 nm using a Tecan
2.2.11.2 Inhibition ELISA

Inhibition ELISAs were carried out as described for indirect ELISAs up to the blocking step. Blocking was carried out using 250 µl of 2 % BSA containing 0.02 % Tween-20 in PBS for 60 min at 37 °C. Simultaneous with the blocking step, 4M5.3 scFvs were incubated with fluorescein, with constant shaking, at a range of concentrations from 30 nM-3 mM for 60 min at 25 °C. Wells were washed as described previously (Section 2.2.11.1) and scFvs pre-incubated with fluorescein were added to wells at 50 µl per well and incubated for 60 min at 37 °C. Wells were subsequently washed with 200 µl PBS five times. Bound scFvs were detected using peroxidase-conjugated anti-polyhistidine antibody (Sigma), diluted 1:3000 in PBS containing 1 % BSA and 0.02 % Tween-20, with incubation for 60 min at 37 °C. Plates were washed five times and 50 µl TMB added per well. After incubation for up to 10 min at room temperature, colour development was stopped by the addition of 50 µl of 1 N H₂SO₄. The optical density of each well was measured at 450 nm using a Tecan GENios microplate reader.

2.2.12 Protein immobilisation

2.2.12.1 Adsorption to microplates

Adsorption-based immobilisations were carried out on Nunc MaxiSorp® flat-bottom 96 well plates. ScFvs were prepared for immobilisation by dialysing into PBS, pH 7.4 and diluted to concentrations from 0.12-6.0 µg ml⁻¹. All analyses were performed in triplicate. Wash steps were carried out with 240 µl of 0.45 µm-filtered PBS, with the first of three wash steps or first two of five wash steps also containing 0.02 % Tween-20. Typically, 60 µl of scFv was incubated in wells from 2.5-120 min at 25 °C. Post-immobilisation, the solution was pooled from triplicate wells and stored for analysis. Plates were washed three times, followed by blocking of unbound sites with 300 µl of 2 % BSA and 0.02 % Tween-20 in PBS for 60 min at 37 °C. Post-blocking, plates were washed three times.

Fluorescein was prepared at 0.1 µg ml⁻¹ in PBS and 60 µl was incubated in each well. In parallel, 60 µl monoclonal anti-polyHistidine-peroxidase antibody (Sigma), diluted 1:3000 in PBS containing 1 % BSA and 0.02 % Tween-20, was
added to replicate wells and plates were incubated for 60 min at 37 °C. After incubation, the solution was removed from wells and transferred to a black, flat-bottom, 96 well microplate (Greiner) for excitation at 485 nm and measurement of emission at 520 nm using a Tecan GENios microplate reader. Fluorescein concentrations were determined by comparison to a standard curve and amounts of bound fluorescein were determined by subtracting calculated fluorescence concentrations in these solutions from the initial concentrations used. Anti-polyhistidine measurements were carried out as in ELISAs (Section 2.2.1) with 60 µl TMB and H₂SO₄ used per well.

2.2.12.2 Covalent attachment to microplates

Covalent immobilisation of scFvs was carried out on amine-coated 96 well plates (Corning® PureCoat™). ScFvs were oxidised and dialysed into PBS, pH 7.4, according to Section 2.2.10.7 prior to immobilisation, unless otherwise stated. Initial incubations of scFv on plates were carried out in the concentration range 0.12-6.0 µg ml⁻¹ for 2.5 min unless otherwise stated. The remaining procedure was carried out according to the adsorption immobilisation protocol outlined in Section 2.2.12.1.

2.2.12.3 Immobilisation on TiO₂-coated 316L stainless steel

Samples to be immobilised were equilibrated in 150 mM sodium phosphate buffer containing 300 mM NaCl, pH7.0. Disks were kindly provided by Balton Sp. z o.o. as TiO₂-coated 316L stainless steel disks (10 mm x 10 mm x 1 mm) with or without added NH₂ surface groups. ScFv immobilisation was carried out by incubation of 25 µl of 50 µg ml⁻¹ scFvs on disks for 60 min at 37 °C. Disks were rinsed four times with phosphate buffer and briefly allowed to dry at 37 °C prior to incubation with 50 µl anti-6X His tag® antibody, DyLight® 488/650 (Abcam), 1:60 dilution in PBS, for up to 16 h at 4°C and removal of unbound DyLight by four PBS washes.

2.2.12.4 Disk washing, antigen incubation, and imaging

Disks that had been incubated with scFvs were washed with high-salt-based PBS (300 mM NaCl) and/or low-pH-based PBS (pH 4.5). Washes were carried out up to five times and excess solution was removed by shaking prior to imaging. Imaging was carried out with an Olympus BX51 inverted fluorescence microscope.
after each wash. Antigen binding was investigated by incubating 25-50 µl fluorescein at 10-1000 ng ml\(^{-1}\) for 60 min at 37 °C with disks, with gentle agitation at 60 rpm. Unbound antigen was removed by washing with PBS prior to re-imaging. Images were captured with a typical exposure of 50 ms and Cy5 and FITC filters were used for imaging as appropriate. Disks were imaged in the same area throughout each experiment. Image processing was carried out using a Fiji image processing package (Schindelin \textit{et al.}, 2012; Schneider \textit{et al.}, 2012).

2.2.12.5 Densitometry analysis of immobilisations

After protein immobilisation, remaining scFv-containing solutions were denatured and analysed by SDS-PAGE and western blotting, alongside freshly prepared standards of the scFv for comparison. Membranes were scanned using an Epson Perfection scanner and image processing was carried out with a Fiji image processing package (Schindelin \textit{et al.}, 2012; Schneider \textit{et al.}, 2012), based on a guideline on quantifying protein on membranes by densitometry measurements (Miller, 2010). Trend lines were generated from the analysed standards and used to calculate the amount of scFv present in supernatants and, by subtraction, the amount of scFv attached to wells.

2.2.13 Oligosaccharide extension

The protocol was carried out as described by Blixt (Blixt \textit{et al.}, 2005). Purified glycosylated 4M5.3 scFv and purified CgtB were each dialysed into 20 mM MES buffer, pH 6.0. All reaction components were also prepared in 20 mM MES buffer, pH 6.0. Reactions were carried out in 350 µl reactions at 37 °C, with 120 rpm shaking, over a 16 h period and aliquots were taken at predetermined timepoints. Reactions were carried out in the presence of 5.25 µg CgtB (15 µg ml\(^{-1}\)) and either 35 µg or 17.5 µg scFv (100 and 50 µg ml\(^{-1}\) respectively), in a solution containing 1 % BSA, 50 nmol UDP-α-Galactose disodium salt (Merck millipore) and 100 mM MnCl\(_2\). Samples removed from reactions were immediately denatured in 2X SDS sample buffer and stored at -20 °C for analysis by SDS-PAGE and western blotting.

2.2.14 Lectin array

Lectin array analysis was carried out by Dr. Michelle Kilcoyne of the Alimentary Glycoscience Research Cluster, NUI Galway. After expression and
purification of the relevant scFvs, these were prepared at concentrations of up to 50 µg ml\(^{-1}\) for the analysis. ScFvs were labelled with DyLight® 650 by Dr. Kilcoyne. The lectin array utilised FITC-labelled lectins and was carried out as previously described (Gerlach et al., 2014). Clustering of results was carried out using normalised fluorescence data and utilised Hierarchical Clustering Explorer v3.5 (Human-Computer Interaction Lab, 2015).

2.2.15 Molecular techniques – DNA

All molecular techniques involving DNA were performed according to Sambrook and Russell (2001) unless otherwise stated.

2.2.15.1 Plasmid isolation and purification

Single colonies of \(E.\ coli\) cells containing the desired plasmid were used to inoculate 5 ml LB cultures with the required selective agents, followed by incubation at 37 °C overnight with shaking. Plasmid DNA was isolated from cells using the Isolate II Plasmid Mini Kit by Bioline as per the provided instructions (Bioline, 2016).

2.2.15.2 Restriction enzyme digestion

All restriction enzymes used were obtained from Fermentas. Conventional enzyme-based digests were performed at 37 °C for 1-16 hr. FastDigest enzyme-based digests were performed at 37 °C for 20 min to 1 hr.

2.2.15.3 Dephosphorylation and ligation

FastAP™ Thermosensitive Alkaline Phosphatase (Thermo Scientific) was used for vector dephosphorylation prior to ligations to prevent self-ligation of linearised plasmid DNA. Dephosphorylation reactions were carried out according to the “protocol for nucleic acid dephosphorylation” provided with the product (Thermo Scientific, 2012).

Ligation reactions were carried out using T4 DNA ligase (Thermo Scientific). Ligations were typically carried out at concentration ratios ranging from equimolar to a 5:1 molar ratio of insert DNA to plasmid DNA. For cloning of PCR products, the Strataclone PCR Cloning Kit by Agilent Technologies was employed according to the recommended protocol (Agilent Technologies, 2015).
2.2.15.4 Polymerase chain reaction

Polymerase chain reaction (PCR) cycles were carried out using a Techne Flexigene thermal cycler. Annealing temperatures for primers were calculated using Oligo Property Scan, provided by Eurofins Genomics (available at: https://ecom.mwgdna.com/services/webgist/mops.tcl). Amplification reactions were typically composed of: 200 $\mu$M each dNTP; 0.5 $\mu$M forward and reverse primers; 0.02 U $\mu$l$^{-1}$ Phusion High-Fidelity DNA Polymerase (Thermo-Scientific), with the necessary buffer and water for each reaction.

2.2.15.5 Agarose gel electrophoresis

Agarose gel electrophoresis was utilised for the separation and analysis of DNA and RNA in a variety of applications. Nucleic acids were typically separated on 1-2 % agarose gels at 90 V for up to 70 min. Visualisation of nucleic acids was enabled by Midori Green Advance, produced by Nippon Genetics Europe GmbH, using a fluorescence G:Box from SynGene using GeneSnap (SynGene).

2.2.15.6 DNA purification and concentration

After agarose gel electrophoresis, bands of interest were excised from gels and purified using either High Pure PCR Product Purification Kit (Roche Applied Science, 2010) or Agarose Gel DNA Extraction Kit (Roche Applied Science, 2008). Concentration of DNA was carried out by ethanol precipitation using sodium acetate buffer to equalise ion concentrations.

2.2.15.7 Mutagenesis

Mutageneses were carried out using either the QuikChange Lightning Site-Directed Mutagenesis Kit by Agilent Technologies or the In-Fusion® HD Cloning Kit by Clontech® Laboratories, following the recommended protocols (Clontech® Laboratories, 2014; Agilent Technologies, 2015). Modifications for performing site-directed mutagenesis (Raman and Martin, 2014) are described below.

Agilent kit-based mutageneses were carried out using 20-100 ng template DNA per reaction, with the intra-cycle extension times increased to 10 min per cycle and the final extension time increased to 10 min. Alternative competent cells were used as the provided cells contained the same antibiotic resistance (Cam$^R$) as the
vector template. Clontech kit-based mutageneses were carried out in 10 µl reactions with 50 ng template per reaction and the cloning-enhancer was replaced by DpnI treatment (Fermentas).

Active site mutageneses of pgIJ were performed by single-site modifications in which the two glutamic acids of the enzyme’s EX7E motifs were changed to alanines. Mutagenic primers were designed using the Agilent primer design tool (www.agilent.com/genomics/qcpd) or by following design guidelines outlined in the In-Fusion® user manual, depending on the kit used. Translational termination mutageneses were carried out by insertion of multiple stop codons upstream of identified active sites.

2.2.15.8 Sequencing

DNA sequencing was performed by Eurofins Genomics or Source Bioscience.

2.2.16 Molecular techniques – RNA

All molecular techniques involving RNA were performed on ice or at 4 °C where possible.

2.2.16.1 RNA isolation from P. pastoris

Total RNA was isolated from P. pastoris cells alone and cells containing the expression vector pPICZα with and without the insert of interest. Isolation of total RNA was performed after 48 h of recombinant protein expression, using ZR Fungal/Bacterial RNA MiniPrep by Zymo Research as per the recommended protocol (Zymo Research, 2014) and a Mini-BeadBeater 24 (BioSpec Products, Inc.). Isolated RNA was DNase-treated using a DNA-free™ DNA removal kit by Ambion® as per the recommended protocol (Ambion®, 2012) and analysed by agarose gel electrophoresis.

2.2.16.2 cDNA synthesis

cDNA was synthesised using Thermo Scientific Maxima H Minus First Strand cDNA Synthesis Kit as per the user manual (Thermo Scientific, 2013). Reactions were performed with 1 µg RNA and a combination of a random hexamer
primer and a gene-specific primer. Synthesised cDNA was used immediately for transcript detection or aliquoted and stored at -70 °C for later use.

2.2.17 Computational analyses

2.2.17.1 Codon usage analysis

Codon usage analysis was carried out using the GenScript Rare Codon Analysis Tool by GenScript (GenScript, 2014) unless otherwise stated. Supporting analyses were also performed with the *E. coli* Codon Usage Analyzer 2.1 provided by The Maduro Lab, University of California, Riverside (Maduro, 1996). Additional analysis was carried out visually with codon usage tables.

2.2.17.2 Sequence alignment and secondary structure prediction

*pglJ* (NCBI GenBank CP000814.1), *pglH* (NCBI GenBank Y11648.1) and *pimA* (NCBI YP_887254.1) amino acid sequences were aligned using ClustalW2 multiple sequence alignment tool (Larkin *et al.*, 2007; Goujon *et al.*, 2010). Porter Protein Secondary Structure Prediction software (Pollastri and McLysaght, 2005) was used to determine the optimal site for single site mutagenesis of *pglJ*. 
Chapter 3:
Glycoprotein expression in *E. coli* and engineering of the glycan chain
3.1 Introduction

The first established bacterial $N$-linked protein glycosylation pathway, encoded by the $pgl$ locus, was characterised in the human enteropathogen $C$ jejuni 81116 and shown to assemble a homogenous Asn-linked heptasaccharide on the protein backbone (Szymanski et al., 1999; Young et al., 2002). The 16-kilobase $pgl$ locus of $C$. jejuni consists of multiple glycosyltransferases and sugar synthesis enzymes within a single operon which function together to provide the necessary subunits and machinery for assembly of glycan chains (Figure 3.1 (i)) (Wacker et al., 2002). Interruption of integral $pgl$ genes, resulted in a loss of immunogenicity, colonisation, adherence and invasion, thereby indicating the essential role of this pathway in pathogenicity (Szymanski et al., 2002). Functional transfer of the $pgl$ operon to $E$. coli enabled targeted $N$-linked glycosylation (Alaimo et al., 2006; Kelleher and Gilmore, 2006) which provided an ideal platform to investigate the function of each $pgl$ gene product – and to produce recombinant glycoproteins in the heterologous host (Wacker et al., 2002).

Expression of pACYC$pgl$ in $E$. coli results in the glycan chain being assembled sequentially in the cytoplasm on a lipid-linked precursor, undecaprenyl phosphate (Und-P) (Linton et al., 2005) and subsequent block transfer of the chain from a nucleotide-activated donor to the protein backbone. Initial attachment of a $\beta 1$-linked bacillosamine is mediated by the glycosyltransferase PglC (Linton et al., 2005), followed by transfer of an $\alpha 1,3$-linked $N$-acetylgalactosamine (GalNAc) by PglA (Szymanski et al., 2002; Glover et al., 2005; Linton et al., 2005; Weerapana et al., 2005) and an $\alpha 1,4$-linked GalNAc by PglJ (Glover et al., 2005; Linton et al., 2005). PglH is then responsible for addition of three $\alpha 1,4$-linked GalNAcs, a mechanism which is self-limiting by way of product inhibition (Glover et al., 2005; Kelly et al., 2006; Troutman and Imperiali, 2009). The final oligosaccharide synthesis step is carried out by PglI which is responsible for the addition of a $\beta 1,3$-linked branched glucose (Kelly et al., 2006).

Upon assembly of the heptasaccharide the flippase PglK transports the oligosaccharide to the periplasm (Alaimo et al., 2006; Reid et al., 2009; Verchere and Menon, 2015) and the $N$-oligosaccharyltransferase PglB mediates attachment of the oligosaccharide to a protein with the conserved acceptor sequence D/E-Y-N-X-
S/T, where Y and X can be any residue except proline (Kowarik et al., 2006). This system is well suited to co-expression of recombinant proteins, such as scFvs, which are directed to the periplasmic space for disulfide bridge formation and folding (Glockshuber et al., 1992). Additionally, the necessary acceptor sequence can be easily incorporated into the protein of interest at a location that does not impact on folding or activity.

The downstream usefulness and the yields of glycoproteins produced in *E. coli* using this *pgl* glycosylation system are potentially limited by the size and composition of the resultant glycan chains (Szymanski et al., 2002; Jaffé et al., 2014). Efforts have been made to modify the glycan sequence recombinantly to produce glycoproteins with eukaryotic-like compositions due to their presumed reduced likelihood of immunogenicity *in vivo* (Schwarz et al., 2010; Valderrama-Rincon et al., 2012). This engineering is facilitated by the relaxed glycan chain specificity of the oligosaccharyltransferase responsible for transfer of the oligosaccharide to the protein, PglB, which can be exploited to transfer alternate glycans to the acceptor protein and therefore produce more diverse glycoproteins (Wacker et al., 2002; Linton et al., 2005; Chen et al., 2007; Cuccui and Wren, 2015). This approach has been utilised in the recent production of glycoconjugates, including bioconjugate vaccine candidates against *Burkholderia pseudomallei* (Garcia-Quintanilla et al., 2014), *Shigella dysenteriae* type 1 (Ihssen et al., 2010), and extra-intestinal pathogenic *E. coli* (van den Dobbelsteen et al., 2016). While fundamental studies into manipulating the composition and functionality of the glycan chain and glycosyltransferases, respectively, are beneficial in themselves, the approach in this work was to investigate further manipulation of the existing glycosylation machinery in order to produce alternate glycoproteins of reduced glycan chain length and composition that may prove more immunogenically discrete.

A goal of this work, therefore, was to reduce the extant glycan chain length attached to scFvs produced in the recombinant *E. coli* system. It was hypothesised that a shorter glycan chain would allow for less flexibility of the scFv upon surface attachment, thereby potentially improving the orientation and antigen-binding of the immobilised scFv, as well as potentially reducing immunogenicity and susceptibility to degradation of the molecules by glycosidase enzymes *in vivo*. This bacterial expression platform would also establish a system of modifying the glycan chain
without the necessity for enzymatic trimming \textit{in vitro} (Schwarz \textit{et al.}, 2010). The ability to selectively modify glycans in this manner would further develop the potential to construct glycoconjugates to order for vaccine development and custom glycan synthesis applications.

(i)

(ii)

\textbf{Fig 3.1}. (i) Schematic representation of the \textit{C. jejuni} locus that produces \textit{N}-linked glycoproteins in \textit{E. coli}. Image is reproduced with permission from: Linton \textit{et al.}, 2005. (ii) \textit{C. jejuni}-directed glycan chain and associated pg\textit{l} locus glycosyltransferase activities. (a) Full-length heptasaccharide generated by native \textit{C. jejuni} pg\textit{l} locus; (b) Trisaccharide expected to be produced in the absence of PglH and PglI activity; (c) Disaccharide product expected to be generated in the absence of PglH, PglI, and PglJ activity. Coloured boxes on right indicate enzymes responsible for monosaccharide addition to glycan chains.

The \textit{C. jejuni} glycosylation system assembles the glycan sequentially through the action of several glycosyltransferases situated on the pg\textit{l} locus (Figure 3.1). Due to the specificity of the glycosyltransferases, the ability to truncate the GalNAcα1-4GalNAcα1-4(Glcβ1-3)GalNAcα1-4GalNAcα1-4GalNAcα1-3Bacβ1-NAsn
heptasaccharide chain (Figure 3.1 (ii) (a)) by inactivation of a single glycosyltransferase was investigated. PglJ, which is responsible for addition of the third sugar of the N-glycan, and PglH, responsible for addition of the subsequent three GalNAcs of the glycan chain, were selected for inactivation. As PglH requires an α1-4 linked GalNAc as an acceptor substrate, it was hypothesised that inactivation of PglJ would preclude PglH-mediated extension of the remaining terminal α1-3 linked GalNAc. As the glucosyltransferase PglI, which requires as template the α1,4-linked GalNAcs produced by PglH, would similarly be unable to attach the branched β1,3-Glc moiety (Figure 3.1 (ii)) (Kelly et al., 2006), PglJ inactivation was hypothesised to yield the N-linked disaccharide, GalNAcα1-3Bacβ1-NAsn.

While crude gene excision or elimination of regulatory signals could be used to eliminate enzyme activity, directed mutagenesis of the PglJ enzyme active site provides a more informative approach to understanding its activity. Previously, Troutman & Imperiali demonstrated the essential role of conserved regions amongst glycosyltransferases (Troutman and Imperiali, 2009) through comparison of PglH and the crystallised mannosyltransferase PimA. This led to identification of a conserved EX7E motif which was essential for transferase activity in PimA. Mutation of the corresponding motif in PglH in their work saw glycosyltransferase activity greatly reduced. Based on their study, two similar EX7E motifs were identified in PglJ in this work. Mutagenic investigation of terminal residues in the EX7E motif in PglJ which aligned with the PimA and PglH sequences was carried out (Figure. 3.2) (PglH alignment not shown). This was complemented by inactivation of PglH and deletion mutagenesis of PglJ to aid in further understanding of the pgl pathway and identification of resultant glycoprotein products.
Fig. 3.2 Alignment of predicted amino acid sequences of pglJ and pimA. Predicted secondary structures are identified above and below the relevant sequences. α-helices are highlighted in green, β-sheets are highlighted in blue. Dark shading denotes conserved sequences and grey shading denotes homology between sequences. EX:E motifs are in red text and indicated above the relevant locations.

Specialised methods used to assess altered glycosyltransferase activity include mass-spectrometry techniques such as MALDI-TOF (Wacker et al., 2002; Wuhrer et al., 2005; Morelle et al., 2009) and tracer/isotope labelling to detect changes in oligosaccharide length and composition (Troutman and Imperiali, 2009), and monosaccharide analysis by analytical chromatography (Hardy et al., 1988; Cox et al., 2006; Royle et al., 2008). Previous studies have also demonstrated the value of detecting glycoprotein products with lectins specific for particular glycans (Wacker et al., 2002; Linton et al., 2005). In this work, we combined sensitive tag-based detection of products to identify changes in chain length by as little as a single saccharide with lectin arrays, previously used to characterise complex cell-surface glycans in prokaryotic and mammalian samples (Kilcoyne et al., 2014; Collin et al., 2016; Ross et al., 2016), to mine for alterations in lectin binding of
glycoengineered products, and single lectin-based detection of glycoproteins in a western blot format to rapidly screen for compositional changes in glycoproducts. This analysis was further supported by the use of glycan-specific detection of proteins on membranes using a DIG glycan detection kit based on aldehyde-hydrazide reactions that detected all glycans.

As previously demonstrated in our group, scFvs provide an excellent platform for development of immunoassays on solid supports (Hu et al., 2006; Hu et al., 2007). Covalent tethering of scFvs to surfaces can further improve antigen-binding due to better orientation of scFvs, better access to binding pockets, increased stability of the immobilised scFvs, and the use of more stringent washing to reduce non-specifically bound proteins (Hu et al., 2013; Hortigüela et al., 2015). Previous work in the group had also established a mechanism of covalently attaching glycosylated scFvs to appropriately activated surfaces via the full-length C. jejuni-directed glycan chain (Hortigüela and Wall, 2013; Hu et al., 2013; Hortigüela et al., 2015). In this work glyco-scFv variants were immobilised via adsorption and covalent binding, followed by investigation of attachment and antigen-binding of the scFvs. Glycoprotein variants with fewer sugar residues might be expected to lead to reduction of unwanted immunogenic responses elicited by the prokaryotic glycan signature (Szymanski et al., 1999; Zilbauer et al., 2008), reduced flexibility of the glycan chain for improved ligand binding of immobilised scFvs (Liu et al., 2012), and lower susceptibility to glycosidase activity in in vivo applications, such as functionalising stent surfaces with cell-capturing scFvs (Foerster et al., 2016). Based on these characteristics, it was hypothesised that the ability to express in E. coli glyco-scFvs with shorter oligosaccharides which could be immobilised efficiently while retaining ligand binding properties would constitute a preferred platform for future production of glycosylated scFvs for surface immobilisation applications.

3.2 Production of glycosylated scFvs in E. coli

3.2.1 Selection of scFv for glycoengineering

Initial experiments compared expression and purification of glycosylated versions of the anti-domoic acid scFv 2H12 (Hu et al., 2005; Hu et al., 2006; Hu et al., 2007) and the anti-fluorescein scFv 4M5.3 (Hu et al., 2013) (Protein Data Bank: 1X9Q) that have previously been studied in our research group. 2H12 scFv has been
used in environmental immunoassay development but its study is constrained by the limited availability and high cost of its domoic acid antigen, whereas 4M5.3 scFv, an enhanced affinity descendant of 4-4-20 scFv (Boder et al., 2000; Midelfort et al., 2004), binds cheap and readily available fluorescein. Expression stocks for production of glycosylated scFv were established by co-transformation of the pIG6 scFv expression vector (Ge et al., 1995) and the pACYCpgl vector containing the C. jejuni protein glycosylation locus (Wacker et al., 2002) into E. coli CLM37. The CLM37 strain contains a deletion in the wecA gene, the protein product of which can introduce an undesired saccharide, GlcNAc, into the glycan chain and so introduce heterogeneity to the glycan sequence (Rick et al., 1994; Linton et al., 2005). The use of E. coli CLM37 is therefore necessary to ensure homogeneity of the glycan chains produced. 4M5.3 scFv was found to typically express at higher titres than 2H12 scFv in initial E. coli expression analyses with yields typically in the range of 1.2–5.0 mg per l of bacterial culture under a variety of expression conditions, compared with 0.6–1.2 mg per l of culture in the case of 2H12 scFv. Preliminary comparison of IPTG-based induction and auto-induction conditions also found that the latter yielded higher expression yields and recovery of a higher amount of soluble protein (Figure 3.3). Due to the combination of higher yields and lower cost of antigen-binding assays, 4M5.3 scFv was chosen for investigation in glycoengineering studies.

![Fig. 3.3. Comparison of glycosylated 4M5.3 scFv expression by IPTG induction and auto-induction. (1) Molecular weight marker; (2) IPTG 16 h; (3) IPTG 5 h; (4) IPTG 3 h; (5) auto-induction 48 h; (6) auto-induction 24 h; (7) auto-induction 6 h. Arrow indicates product of expected size, full-length glycosylated scFv was approximately 34 kDa; unglycosylated scFv was 32 kDa; lower bands are predicted degradation products.](image)

### 3.2.2 Optimisation of glycoprotein production and recovery

In order to enhance overall production and increase the yield of soluble, active 4M5.3 glyco-scFv protein reported elsewhere (Hu et al., 2013; Hortigüela et
al., 2015), a variety of modifications to the basic expression protocol were investigated. Amongst these, auto-induction (Studier, 2005) was typically carried out for 48 h to maximise biomass generation and protein production. Meanwhile, increasing substantially the concentration of glucose in the 5052 auto-induction medium component from 2.5 % to 12.5 % led to significantly higher protein expression in cultures, most likely due to the increase in cellular biomass present before induction of expression. Coomassie staining and western blot analysis of histagged glyco-scFvs produced under these conditions indicated increased yields of purified, full-length glyco-scFv of up to 2.3-fold (Figure 3.4).

![Fig. 3.4](image)

**Fig. 3.4** Analysis by whole protein staining ((A) and (C)) and western blot analysis (B) and (D) of purification of glycosylated 4M5.3 scFv produced by auto-induction conditions in the presence of 1X glucose (A) & (B) or 5X glucose, concentrations (C & D). (1) Molecular weight marker; (2) soluble protein fraction pre-purification; (3) column flow through; (4) wash 1; (5) wash 2; (6) eluate 1; (7) eluate 2; (8) eluate 3. Arrow indicates product of expected size, full-length glycosylated scFv was approximately 34 kDa; unglycosylated scFv was 32 kDa; lower bands are predicted degradation products.

Approaches to increase scFv recovery from the *E. coli* periplasm were also investigated as insoluble cellular fractions were frequently observed to contain similar levels of his-tagged product to soluble fractions after cell fractionation and protein extraction. Additional washes of cell pellets after extraction with TSE and PBS buffers identified that a single added PBS wash could recover an additional 30-60 % of "resoluble" protein compared to that recovered in the initial soluble fraction (Figure 3.5). Soluble and resoluble fractions were purified in parallel by IMAC and quantified by the Bio-Rad Bradford-based protein assay, prior to comparison of the activities of the two fractions.
Fig. 3.5 Western blot analysis of soluble and resoluble extracts of glycosylated 4M5.3 scFv soluble and resoluble extracts produced by auto-induction conditions with 1X and 5X glucose. (1) Molecular weight marker; (2), (3) soluble and resoluble proteins, respectively, produced with 1X glucose; (4), (5) soluble and resoluble proteins, respectively, produced with 5X glucose. All samples were electrophoresed and blotted on the same gel and membrane; the image has been separated to omit lanes not relevant to the present analysis. Arrow indicates band of expected size of full-length glycosylated scFv, full-length glycosylated scFv was approximately 34 kDa; unglycosylated scFv was 32 kDa; lower bands are predicted degradation products.

Indirect ELISAs were carried out with soluble and resoluble fractions to confirm the latter contained soluble, active protein and not inactive misfolded molecules. These revealed higher fluorescein binding levels by the resolubilised protein fraction (Figure 3.6), with an increase in binding ranging from 2.5 to 7.3-fold. This verification of the functionality of the "resolubilised" fraction allowed soluble and resoluble fractions to be combined into a single sample in subsequent analyses. Finally, while previous work in the group had reported a significantly higher ratio of glycosylated:unglycosylated scFv upon increasing the antibiotic concentration used for selection of E. coli cells containing the medium copy number (15-20) pIG6 expression vector (ampicillin: 100 μg ml⁻¹ to 200 μg ml⁻¹) and the low copy number (10-12) pACYCpgl plasmid (chloramphenicol: 34 μg ml⁻¹ to 68 μg ml⁻¹) during glyco-scFv expression (Hu et al., 2013), it did not prove possible to replicate these results in our investigation and so “standard” antibiotic concentrations of 100 μg ml⁻¹ ampicillin and 34 μg ml⁻¹ chloramphenicol were used for plasmid maintenance and in recombinant protein expression throughout the work.
3.3 Generation of alternative pgl constructs

As outlined in Section 2.2.17.2, the predicted amino acid sequences of pglJ and pglH were aligned with each other and with that of PimA which contains an essential EX₇E motif also present in PglH (Troutman and Imperiali, 2009). This analysis was combined with secondary structure predictions to identify the EX₇E motif most likely to be responsible for glycosyltransferase activity in PglJ (Figure 3.2). Finally, these analyses were used to design primers for mutagenesis using either the QuickChange Lightning Site-Directed Mutagenesis Kit by Agilent or the In-Fusion® HD Cloning Kit by Clontech® Laboratories. Resultant changes in glycoprotein profiles due to pgl gene modification were identified and characterised in western blots.

3.3.1 Effect on glycan length of site-directed mutagenesis of pglJ

The first glutamic acid of the ²⁷⁶EX₇E²⁸⁴ motif in pglJ, E276, was substituted for an alanine using the primer set a278c/a827c_anti to create the pACYCpglJ²⁷⁶ single site mutant (Figure 3.2). Following sequencing of four clones using the primer pacycseq2.0 which annealed 122 bp upstream of the mutagenesis site two with the desired modified sequence were identified. These were then assessed by co-expression with scFvs harbouring the glycan acceptor sequence.
Expression of the 4M5.3 scFv in *E. coli* cells harbouring the mutated pACYCpglJ \(^{276}\) plasmid yielded unglycosylated and glycosylated bands characteristic of expression in the presence of the unmutated vector, but also an additional protein product. While of apparently intermediate molecular weight, the band was closer in size to the unglycosylated than glycosylated scFv and was thought likely to be a glycoprotein containing the disaccharide side chain predicted from PglJ inactivation (Figure 3.7, lane 7). Protein of the size of full-length glyco-scFv was also unexpectedly present. This was thought to result from residual activity remaining in the mutated PglJ enzyme after the single site mutation, or from a relaxed substrate specificity of another glycosyltransferase in the pathway, allowing it to utilise the alternative acceptor saccharide remaining in the absence of PglJ activity. The full-length glycoprotein was typically produced at much lower levels in expressions carried out in the presence of the mutated rather than wild type *pglJ*, indicating a lower efficiency of the PglJ-associated glycosyltransferase activity during assembly of the full-length glycan chain.

![Fig 3.7](image)

**Fig 3.7** Comparison of glycoforms of 4M5.3 scFv produced in *E. coli* cells harbouring plasmids containing a variety of pACYCpgl constructs. (1) Molecular weight marker; scFvs produced in cells containing: (2,9) pACYCpgl; (3, 10) no *pgl* genes; (4) truncated *pglJ* (pACYCpglJ); (5) double site mutagenesis (pACYCpglJ\(^{276:284}\)); (6) truncated *pglH* (pACYCpglH); (7) single site mutagenesis (pACYCpglJ\(^{276}\)); (8) combination single site mutagenesis and truncated *pglH* (pACYCpglJ\(^{276}\)pglH). Arrow indicates expected size of unglycosylated scFv, expected size of: full-length glycosylated scFv was approximately 34 kDa; trisaccharide glycosylated scFv was approximately 32.6 kDa; disaccharide scFv was approximately 32.4 kDa; unglycosylated scFv was 32 kDa; lower bands are predicted degradation products.

To further inhibit glycosyltransferase activity in the mutated PglJ protein, the terminal glutamic acid residue of the \(^{276}\)E\(^{284}\) motif, E284, was also changed to alanine by site-directed mutagenesis. Secondary structure analysis indicated that this residue may be more exposed in the protein structure than E276 (Figure 3.2).
Mutagenesis of the terminal residue E284, was carried out using the primer set pglJmuta2.0For/Rev to produce pACYCpglJ284, followed by screening and sequence confirmation. Expression of 4M5.3 scFv in cells harbouring this pACYCpglJ284 vector yielded an identical glycoprotein profile to cells harbouring pACYCpglJ276 (not shown). This result confirmed that substitution of either glutamic acid residue in the putative EX7E active site motif of PglJ significantly altered the resulting glycoprofile - but provided no insight into whether the remaining, apparently full-length glycoprotein resulted from residual PglJ activity or continued activity in its absence by another glycosyltransferase. In order to gain further insight into this, it was decided to mutate both glutamic residues in the same enzyme molecule and to determine the subsequent impact on resultant glyco profiles.

Mutagenesis of E284 was carried out using pACYCpglJ276 as template and primers pglJmuta2.0For/Rev. After sequence confirmation of the plasmid now termed pACYCpglJ276,284, 4M5.3 scFv co-expression was again carried out and analysed as before. No detectable full-length glycoprotein was obtained in this case (Figure 3.7, lane 5). Instead, unglycosylated and potentially disaccharide-containing scFvs were evident in western blot analysis, as well as lower molecular weight proteins thought to be degradation products.

3.3.2 Investigation of pgl locus redundancy by truncation mutagenesis

Due to results of analysis of single and double site mutageneses of the EX7E motif on glycosyltransferase activity (Troutman and Imperiali, 2009), the most likely explanation for the continued presence of full-length glycoprotein in scFv expression in the presence of mutated PglJ was thought to be residual activity remaining in modified PglJ enzymes, thereby providing sufficient glycan chain template for subsequently-acting enzymes in the pathway to generate full-length glycan chains. To investigate this, mutagenesis to completely inactivate PglJ was carried out by inserting a translation termination cassette upstream of the predicted pglJ active sites. This was designed to replace an existing I-H-F-L-E amino acid motif in the pglJ sequence 151 bp 3’ to the primary ATG with a *-H-*L-* sequence where * represents a stop codon, with minimum modification of extant nucleotides and use of TAA as the preferred stop codon in E. coli. Unmodified pACYCpgl was used as a template for this mutagenesis with the primer pair pglJStopFor/Rev to generate
pACYC<sub>pglJ</sub> and clones were confirmed by sequencing using the primer pglJknockseq. Expression of 4M5.3 scFv in the presence of the pACYC<sub>pglJ</sub> vector yielded the same glycoprofile pattern (Figure 3.7, lane 4) as seen in the case of expression in the presence of pACYC<sub>pglJ</sub><sup>276,284</sup>; these protein products were of the size expected of unglycosylated scFv and scFv with an attached disaccharide resulting from the activities of PglC and PglA enzymes (Figure 3.1 (ii) (c)). No full-length glycoprotein-sized product was detectable indicating complete elimination of PglJ activity.

To complement PglJ mutagenic analysis of the <i>pgl</i> pathway and in the absence of an <i>N</i>-glycan mass-spectrometry analysis capability to support this work, mutagenesis experiments were designed to inactivate PglH, as had been carried out for PglJ. A similar translation termination cassette was inserted using the primer pair pglHmutF/pglHmutR, to replace an existing V-S-I-I-K amino acid motif at 94 bp 3’ of the primary ATG with *-S-*I-* in the <i>pglH</i> sequence, as per the parameters considered for <i>pglJ</i>, to generate pACYC<sup>pglH</sup><sub>J</sub>. The construct was confirmed by sequencing using the primer pglHseq. Expression of 4M5.3 scFv in cells harbouring this pACYC<sup>pglH</sup><sub>J</sub> vector yielded a glycoprofile similar to that observed in cells harbouring pACYC<sup>pglJ</sup><sup>276,284</sup> or pACYC<sup>pglJ</sup>, with the single difference that a higher molecular weight protein product was clearly visible in the case of the <i>pglH</i> expression (Figure 3.7, lane 6) in the place of the "di-saccharide-containing scFv" observed in the case of the mutated <i>pglJ</i> clones. This difference was consistently observed in repeated electrophoretic analyses and was more clearly visualised upon optimising the separation time as presented in Figure 3.7. This result was consistent with the expected outcome of mutagenesis of PglH, in which the glycan chain would be assembled up to trisaccharide length by the activities of glycosyltransferases, following which extension would stop in the absence of PglH to add the subsequent α-1,4-linked GalNAc residue (Figure 3.1 (ii) (b)). Following confirmation of this profile, work proceeded to investigate activities of the mutated PglJ enzymes in the absence of PglH activity, and analysis of the resultant glycoprofiles.

The <i>pglH</i> gene was inactivated in the pACYC<sup>pglJ</sup><sup>276</sup> plasmid by utilising the pglHmutF/pglHmutR primer pair, resulting in pACYC<sup>pglH</sup><sup>276</sup><sub>J</sub>. Expression of 4M5.3 scFv in cells harbouring this plasmid yielded the glycoprofile shown in Figure 3.7, lane 8. The three main bands in this profile were considered to be, lower
to upper, unglycosylated scFv, glyco-scFv containing the disaccharide resulting from an absence of PglJ activity, and a higher molecular weight protein, smaller than full-length glycoprotein, which corresponded in size with the “trisaccharide-linked scFv” previously observed in the pACYCpglH expression analysis. By comparing the uppermost band arising from this construct to those of pACYCpglJ276 and pACYCpglH clones, it can be inferred that the upper band in pACYCpglJ276 is generated through PglH activity as the pglJ276, truncated PglH combination effectively reduced the largest molecular weight band to the size of that generated in the presence of pACYCpglH construct.

3.3.3 Western blot-based characterisation of glycoprofiles

Comparison of unglycosylated scFv (Figure 3.7, lanes 3 and 10) and scFv expressed in the presence of pglJ inactivations (truncation and double site mutagenesis (lanes 4 and 5 respectively)), and the middle band in single site mutations (Figure 3.7, lanes 7 and 8), reflected a size difference expected to correspond to the disaccharide GalNAcα1,3-Bac which should have a combined predicted mass of 383.4 Da. Furthermore the predicted trisaccharide produced in the presence of pglH modifications was clearly different in its size, (predicted as 605.02 Da) to that of the disaccharide (Figure 3.7, lane 6). The combined banding pattern of the di- and trisaccharide glycoproteins was observed in lane 8 and further supported interpretation of the individual banding patterns. The full-length glycoprotein was distinguishable from its unglycosylated and shorter glycosylated counterparts due to the additional mass of 1448.4 Da conferred by the heptasaccharide (Figure 3.7, lanes 2 and 9). This was further supported by western blot analyses, as outlined below.

3.3.3.1 DIG-specific analysis of glycoproteins

In order to complement size-based analysis of glyco-scFvs, samples were analysed using a "DIG Glycan Detection Kit" which detects glycans on nitrocellulose membranes by oxidation of adjacent hydroxyl groups in the oligosaccharides to aldehydes and labeling of the aldehydes with hydrazide-linked digoxigenin (DIG) for anti-DIG antibody-based detection. This revealed a similar banding pattern in terms of protein sizes (Figure 3.8) as detected by the peroxidase-conjugated anti-polyhistidine antibody (Figure 3.7), with the expected exception that the unglycosylated protein was not detected. Bands were noticeably less sharp than
in the case of hexahistidine-based detection of the same proteins, which has been repeatedly observed to result from oxidation of the purified scFvs (Hu et al., 2013).

Nevertheless, the relative sizes of the glyco-scFv products were apparent, with the predicted disaccharide-linked scFv present in lanes 2 and 3 (corresponding to pglJ truncation and pglJ double mutagenesis, respectively) obviously smaller in size that the predicted trisaccharide-linked scFv in lane 4 (pglH'), and in turn clearly smaller than the full-length, heptasaccharide-linked scFv present in lane 7 (intact pgl genes). Lane 5 (pglJ276 mutation) displayed a less distinct pattern but with bands of apparently the same molecular weights as lanes 2 and 3, and lane 7 visible, which is further displayed in the contrast-enhanced image of Figure 3.8 (lane 9), thereby matching the banding pattern observed upon anti-polyhistidine detection (Figure 3.7, lane 7). Lane 6 is similarly less intense but a band of the same apparent molecular weight as in lane 4 is visible, indicating the presence of a trisaccharide-linked scFv which is further demonstrated in the contrast-enhanced image of Figure 3.8 (lane 10). Due to the lower resolution of this detection method, it was not possible to distinguish between di- and tri-saccharide linked scFvs in lanes 6/10 using this approach. The absence of any signal from the apparently unglycosylated scFv (lane 8) confirms the specificity of the method – and the presence of the unglycosylated scFv products in all of lanes 2-10 of the hexahistidine-mediated detection in Figure 3.7. This complementary approach thereby validates the data and analysis outlined in Figure 3.7. It is a useful addition to analysing modification of the scFv-attached glycan due to its specificity for glycosylated products and its ability to distinguish between glycoproteins differing by as little as a single GalNAc in their glycan compositions; these results are further supported by lectin blot data presented in Chapter 4.
3.4 Functionality of glycoproteins

Beyond simple detection of the engineered glycoproteins, it was important to compare their antigen-binding activities to their standard full-length glycoprotein and unglycosylated counterparts if they were to be suitable for use in either in vitro assays or in vivo applications. The shortest glycoprotein produced, the disaccharide glyco-scFv produced in the presence of pACYCpgl\textsuperscript{J}\textsubscript{276,284} plasmid, was selected for this detailed investigation of binding and immobilisation, and comparison with the unglycosylated and fully glycosylated scFvs.

3.4.1 Antigen binding

Competitive ELISAs had been performed previously with unglycosylated and full-length 4M5.3 glyco-scFvs (Hu et al., 2013). In these, it was demonstrated that the unglycosylated and glycosylated scFvs exhibited binding affinities of 2.89 ± 0.77 x 10\textsuperscript{-12} and 3.4 ± 0.5 x 10\textsuperscript{-12} M, for fluorescein, respectively.

In the present work, inhibition ELISAs were carried out with the three 4M5.3 scFvs of interest by pre-incubating the scFvs with a range of fluorescein concentrations prior to binding surface-immobilised FITC and detection by peroxidase-conjugated anti-polyhistidine antibody. The expected decrease in signal
with increasing inhibition was observed for all scFvs (Figure 3.9). The data was normalised and log-transformed to calculate the IC$_{50}$ of each with GraphPad Prism using Nonlinear Regression analysis which employed an inhibition "log (inhibitor) vs. normalised response" equation. The IC$_{50}$ of the unglycosylated, full glycan and short glycan 4M5.3 scFvs were determined as $3.43 \pm 1.44 \times 10^{-7}$, $9.83 \pm 1.11 \times 10^{-7}$ and $1.89 \pm 1.16 \times 10^{-6}$ M, respectively, indicating retention of scFv binding activity in the glycosylated molecules.

![Fig 3.9 Comparative inhibition ELISAs of 4M5.3 scFvs: unglycosylated (no pACYPgl genes); fully glycosylated (+pACYPgl); and with shortened glycan chain (+pACYPglJ$_{276,284}$).](image)

### 3.4.2 Immobilisation of scFvs using adsorption

Comparison of immobilisation of the three scFv variants was initiated by using an adsorption approach to generate a baseline of surface binding and subsequent activity of the scFvs, prior to commencing covalent immobilisations. Relatively non-specific protein adsorption relies on hydrophilic, hydrophobic, and electrostatic interactions between protein molecules and the target surface (Jung et al., 2008). Consequently, adsorbed scFvs may be randomly oriented and generally exhibit a poor ratio of antigen bound:immobilised scFv (Hortigüela and Wall, 2013). In consideration of immobilisation methods, therefore, it is often beneficial to direct protein immobilisation on surfaces via covalent interactions if possible. This may result in more stably attached, better- and more uniformly-oriented protein molecules.
across a surface compared to adsorption-based approaches (Tominaga et al., 2005). Covalent immobilisation of proteins can also lead to better consistency and reproducibility between experiments, and reduction in the batch-to-batch variability often associated with adsorbed proteins (Howell et al., 1998; Torrance et al., 2006; Jung et al., 2008). Comparison of outcomes of scFv adsorption analyses with subsequent covalent immobilisations in this study would therefore be instructive in determining if improved activity could be observed in scFvs immobilised in a directed, targeted manner.

Adsorption-based immobilisation of scFvs was carried out in 96-well polystyrene plates. Based on previous antigen-binding ELISA analysis, 1-3 µg ml\(^{-1}\) (1.9-5.6 pmol) scFv was immobilised as this had been determined to be the optimal range to avoid saturation and facilitate detection of antigen binding. As nanodrop readings and BCA assays of protein concentrations were found to vary greatly at scFv concentrations of less than 20 µg ml\(^{-1}\) (37.5 pmol), an alternative method was required for determining the amount of scFv was bound to plates. Western blot detection and comparison to standard curves was determined to consistently detect 2 µg ml\(^{-1}\) (3.8 pmol) of his-tagged scFv (120 ng protein), while 1 µg ml\(^{-1}\) was frequently detected also. By immobilising 6 µg ml\(^{-1}\) (11.25 pmol) scFv on plates for different time intervals and estimating the amounts of unbound protein by western blot, scFv concentrations in the region of 2 µg ml\(^{-1}\) could be adsorbed onto plates for antigen-binding analyses. Quantification of protein on western blots was achieved using comparative densitometry analysis (Figure 3.10), which was found to be accurate, frequently returning an \(R^2\) value > 0.95. Western blot detection was also queried to quantify scFv removed in wash steps subsequent to immobilisation but was below the detectable limits of this approach thereby supporting the stability of the adsorbed scFvs and the analyses of post-immobilisation supernatants.
Chapter 3

Fig. 3.10 Quantification of scFvs remaining in solution after immobilisation on ELISA plates. (A) Western blot used for densitometry analysis. Lane 1: Molecular weight marker; lanes 2–4: triplicate samples of protein-containing solution after scFv immobilisation; lane 5: 6 µg ml⁻¹ scFv standard; lane 6: 5 µg ml⁻¹ scFv standard; lane 7: 4 µg ml⁻¹ scFv standard; lane 8: 3 µg ml⁻¹ scFv standard; lane 9: 2 µg ml⁻¹ scFv standard. (B) Standard curve generated from densitometry readings of standards on western blot (R² = 0.9868).

Initial analyses showed that all three scFvs continued to adsorb onto plates over a 2 h time course (Figure 3.11 (A)) but a corresponding increase in signal was not observed upon detection by an anti-polyhistidine-peroxidase antibody in ELISA. This was thought to be due to layering of protein molecules on the ELISA well surface upon continued immobilisation, thereby masking the initially adsorbed protein and leading to a disproportional increase in signal with increasing amounts of immobilised protein. It would be of interest to reduce the amount of scFv adsorbed to less than a theoretical monolayer to determine if increasing the amount of immobilised protein up to monolayer levels would lead to a proportional increase in immobilisation and antigen binding.
Fig. 3.11 Amount of scFv remaining on ELISA plate wells after adsorption of 6 µg ml$^{-1}$ (11.25 pmol) of each scFv under varying conditions over (A) 10-120 min; (B) 2.5-10 min.

An additional time course of scFv adsorption from 2.5 to 10 min identified that 5 min immobilisations displayed the most consistent results within the desired concentration range for immobilised scFv across replicates (Figure 3.11 (B)). These parameters led to adsorption of 2.5 (± 1.0) µg ml$^{-1}$ (4.7 pmol ± 1.9 pmol) of each of the three scFvs being investigated. The amounts of non-adsorbed protein were measured by western blot analysis, as outlined above, and were consistent with measurements of surface-adsorbed scFvs generated by the indirect ELISA approach.
3.4.2.1 Adsorption - antigen-binding

Initial antigen binding experiments with scFvs immobilised for 5 min yielded inconsistent results between experiments and the immobilised scFvs exhibited low antigen:scFv binding ratios. This could be due to removal of immobilised scFv in wash steps so that insufficient scFv remained to carry out antigen binding, or because this concentration of adsorbed scFv yields too low a concentration of correctly-oriented scFv to provide a detectable binding response. These results were in contrast to earlier ELISA-based experiments in which 1-3 µg ml⁻¹ scFv concentrations demonstrated clear antigen binding but these prior experiments reported scFv molecules by the use of highly sensitive anti-polyhistidine peroxidase conjugated antibody, whereas the present work relied on the more variable reporting mechanism of fluorescent detection, which is also susceptible to bleaching and high background readings. In order to detect antigen binding of adsorbed scFvs, immobilisation was therefore carried out for 2 h with an initial concentration of 6 µg ml⁻¹ which enabled adsorption of 4-5 µg ml⁻¹ (7.5-9.4 pmol) of each scFv (Figure 3.12). This was followed by determination of the ability of the adsorbed scFvs to bind 0.1 µg ml⁻¹ fluorescein (18 pmol) (Figure 3.13) following a number of range finding experiments to determine the lowest consistently detectable concentration of fluorescein (not shown).

![Graph showing scFv remaining on plate](image)

**Fig. 3.12** Amount of scFv remaining on ELISA plates after 2 h adsorption. Initial scFv concentrations were 6 µg ml⁻¹.

In order to determine the effect on antigen binding of low immobilised scFv
concentrations, scFv adsorption was carried out for 2 h as above with scFv variants at 6, 3, 1.2 and 0.12 µg ml⁻¹ (11.3, 5.6, 2.3, and 0.2 pmol) scFvs, followed by investigation of the ability of the adsorbed molecules to bind 0.1 µg ml⁻¹ (18 pmol) fluorescein. As densitometric analysis is not sufficiently sensitive to detect non-adsorbed scFv remaining in solution at the lower incubated concentrations, the antigen:scFv ratios plotted in Figure 3.13 below were calculated using the amounts of scFv incubated in wells rather than the amounts measured on the plates. The analysis indicated that unglycosylated and both glycosylated scFvs bound fluorescein at a similar level (ranging from 1.50 ± 0.11 pmol fluorescein at 0.225 pmol scFv immobilised to 4.86 ± 0.92 pmol fluorescein at 11.25 pmol scFv immobilised) (Figure 3.13 (A)). The expected reduction in the ratio of antigen:antibody at higher immobilised scFvs levels was observed for all three scFvs (Figure 3.13(B)).
Fig 3.13 Comparison of antigen binding of surface-adsorbed scFv glycovariants. (A) Amount of fluorescein bound at different concentrations of immobilised scFvs. (B) Estimated ratio of antigen molecules bound per scFv molecule at increasing concentrations of incubated scFv.

3.4.3 Covalent immobilisation

To properly assess the potential use of glycosylated scFvs in surface immobilisation applications, a method for protein attachment was required which would yield well oriented, stably attached scFvs via a directed method. As previously demonstrated in our group (Hu et al., 2013), upon oxidation by sodium meta-periodate, cis-glycol groups contained in the glycan chain of the full-length glycoprotein (+pACYCpgl), which are also present in the shortened glycan
(\(+\text{pACYCpglJ}^{276,284}\)), generate aldehydes which can readily form covalent bonds with amine (NH\(_2\)) groups present in an aminated surface (Figure 3.14). As such glycan-containing scFvs attach much more readily to aminated surfaces than unglycosylated molecules and do so in a well-oriented manner, providing a much better platform for protein immobilisation than via physiosorption.

Fig 3.14 Schematic of scFv immobilisation on functionalised surfaces. The 4M5.3 scFv is shown as a ribbon structure, with \(\alpha\) helices in red, \(\beta\) sheets in yellow and flexible loops in green. The glycan chain, attached to the C-terminus of the scFv, is shown as seven sugar monomers, including the scFv-proximal bacillosamine in red, five N-acetylglucosamine units in yellow and the chemical structure of the \(\beta1,3\)-branched glucose that undergoes oxidation shown for clarity of the immobilisation chemistry. The scFv binding pocket is arrowed. Image is reproduced with permission from: Hu et al., 2013.

Oxidised and non-oxidised full-length glycoproteins were immobilised on Corning PureCoat polystyrene 96-well microplates, which have a uniform, functionalised amine coating. Detection of immobilised scFvs was carried out via their hexahistidine tag. Immobilisation of scFvs for 2-16 h revealed maximal levels of immobilised scFv were already reached after 2 h, which was therefore used in subsequent experiments. ScFv immobilisations were carried out with 2.75 and 27.5 \(\mu\text{g ml}^{-1}\) (5.16 and 51.56 pmol) oxidised shortened glycoprotein (\(+\text{pACYCpglJ}^{276,284}\)), oxidised full-length, and unoxidised full-length glycoproteins (\(+\text{pACYCpgl}\)). The unoxidised glycoprotein exhibited a much reduced attachment to the surface (Figure 3.15), as expected, due to the lack of glycan moieties for oxidation and reaction with
the surface amines. At 2.75 µg ml\(^{-1}\) initial scFv concentration, the shortened glycoprotein displayed a significantly reduced signal compared to the full-length scFv (68 %) though repeated analyses revealed somewhat inconsistent results in which the shortened and full-length glycan-modified scFvs could exhibit from this level up to similar efficiencies of immobilisation. Hexahistidine tag-based detection method was carried out in parallel with all covalent immobilisation and antigen-binding experiments to control for differences in levels of immobilised scFv between samples.

![Graph](image)

**Fig 3.15** Analysis of immobilisation of oxidised/unoxidised full-length glycoprotein (+pACYCpgl) and oxidised shortened glycoprotein (+pACYCpgl\(^{276,284}\)) on amine-functionalised plates.

### 3.4.3.1 Time course of covalent immobilisation of scFvs

Investigation of scFv immobilisation over a 5-60 min period indicated a reduction in preferential specificity of the glycosylated scFv with increasing incubation lengths (Figure 3.16). After just 5 min, three times more immobilised glycoprotein than unmodified scFv was successfully immobilised, falling to only approximately 1.7-fold after 60 min.
Therefore, it was decided to analyse immobilisation over a shorter timeframe, with scFvs immobilised for 1, 2.5, 5, 7.5, and 10 min. It was clear in this analysis that immobilisation of the glycosylated scFv occurred much more rapidly than that of the unglycosylated scFv (Figure 3.17). Notably, after 2.5 min, the amount of glycosylated scFv immobilised was over 5 times higher than that of unglycosylated scFv (Figure 3.17B), following which the glycosylated:unglycosylated ratio dropped to a consistent 1.86 (± 0.04)-fold difference, very similar to the 1.7-2.0 ratio observed at incubations of 10 min and longer in the previous analysis (Figure 3.16). On this basis, a 2.5-min incubation was selected for immobilisation of glycosylated scFvs in order to maximise attachment of the glyco-scFvs while minimising surface tethering of the unglycosylated scFv molecules via reactions not mediated by the oxidised N-glycan chain.

**Fig 3.16** Ratio of glycosylated scFv : unglycosylated scFv immobilised on aminated plates with increasing incubation times. Specificity for the glycosylated molecule dropped over time.
Figure 3.17 Quantification of immobilisation of scFv molecules on aminated surface over time. (A) Comparison of amounts of unglycosylated and full-length glycosylated scFv immobilised after 1 to 10 min incubations. (B) Ratio of glycosylated:unglycosylated scFv immobilised at the same time points.

Figure 3.3 reveals that in expression and purification of the glycosylated scFv, approximately 33% of purified protein, as determined by western blotting and densitometry, was typically unglycosylated. In order to assay the same amounts of scFv glycovariants of interest in immobilisation analyses, glycosylated scFv preparations were used at 3 µg ml\(^{-1}\), containing an estimated 2 µg ml\(^{-1}\) full-length glycosylated scFv, and 1 µg ml\(^{-1}\) unglycosylated scFv for comparison. To eliminate
the possibility that the remaining 1 µg ml\(^{-1}\) unglycosylated scFv was responsible for signal differences observed between glycosylated and unglycosylated scFv samples in immobilisation experiments, this amount of unglycosylated scFv was also immobilised for 2.5 min alone and in parallel with 2 µg ml\(^{-1}\) unglycosylated and glycosylated scFvs (Figure 3.18). As the less concentrated, unglycosylated sample displayed only 6.53 (± 2.3) % of the signal observed with the immobilised glycosylated protein, it could be concluded that the unglycosylated scFv carried over in glycosylated scFv samples made only a relatively minor contribution to the up to 5-fold increase in glyco-scFv immobilised over unglycosylated scFv found to be attached to the surface under optimal conditions (Figure 3.17).

![Graph showing Absorbance 450nm](image)

**Fig 3.18** Comparison of amount of glycosylated and unglycosylated scFvs immobilised on aminated surfaces at varying concentrations.

### 3.4.3.2 Antigen-binding by covalently immobilised scFvs

scFv adsorption experiments were carried out with an initial concentration of 6 µg ml\(^{-1}\) (11.25 pmol) scFvs as this was necessary to allow for densitometric detection of scFv molecules remaining in solution after immobilisation (see Figure 3.10). Densitometry experiments established that the same concentration of scFvs was not detectable after oxidation, however. In this case 18 µg ml\(^{-1}\) (33.75 pmol) scFv proved the limit of detection and quantification. As this meant it was no longer possible to measure amounts of scFvs remaining in solution after immobilisation experiments, therefore, it was also not possible to determine the amounts of scFvs
successfully immobilised on the surface. As a consequence, antigen binding measurements determined with oxidised scFvs were expressed relative to the initial concentration of scFv incubated with surfaces, rather than the immobilised protein.

A comparison of the fluorescein binding of scFvs immobilised for 2.5 min or for 2 h (Figure 3.19) revealed that up to 34.9 % of the available 0.1 µg ml⁻¹ (18 pmol) fluorescein was bound by both glycosylated scFv formats after only 2.5 min (Figure 3.19 (A)). Meanwhile, the unglycosylated scFv reached maximal fluorescein-binding signals of only 6.5-7.9 and 3.6-3.8 times lower than the two glycosylated molecules after 2.5 min and 2 h, respectively (Figure 3.19). Importantly, the full and partially glycosylated scFvs exhibited fluorescein binding of 1.1- and 1.5-fold higher, respectively than their adsorbed counterpart upon immobilisation for 2 h. Unglycosylated scFvs bound fluorescein with a 3.7-fold decrease over their adsorbed counterpart after 2 h incubations. The net overall impact of glycan-mediated covalent immobilisation was calculated to be a 1.13-fold increase in antigen-binding ability of the fully-glycosylated scFv over the oxidised, unglycosylated fragment, therefore, while the improvement for the short glycan-linked scFv was 1.51-fold.
Fig 3.19 Determination of antigen binding of covalently immobilised scFvs. Fluorescein bound by oxidised scFvs immobilised on aminated surface for (A) 2.5 min, and (B) 2 h was measured. ScFvs were immobilised at 6 µg ml\(^{-1}\) (11.25 pmol) and fluorescein was added for antigen binding at an initial concentration of 0.1 µg ml\(^{-1}\) (18 pmol). Significance was determined by multiple comparison analysis via a one-way ANOVA test.

The ability of lower concentrations of immobilised scFv to bind antigen was also measured, for which the three scFv variants were immobilised for 2.5 min at initial concentrations of 6, 3, 1.2, and 0.12 µg ml\(^{-1}\) scFv, indicating a clearly higher fluorescein-binding ability of the glycosylated molecules (Figure 3.20 (A)). As described previously (Section 3.4.2.1), antigen:antibody molar ratios were then inferred from the amounts of scFv incubated with surfaces rather than those actually
immobilised (Figure 3.20 (B)). While unglycosylated scFvs attached to the wells exhibited negligible binding of fluorescein at all scFv concentrations investigated (Figure 3.20 (A)), both glyco-scFv variants exhibited much higher antigen-binding, close to the 1:1 ratio expected, with a similar reduction in antigen:scFv ratio to that described previously (Figure 3.13) at higher scFv concentrations (Figure 3.20 (B)).

Fig 3.20 Comparison of antigen binding abilities of scFv variants immobilised on aminated surfaces. (A) Fluorescein binding by scFv incubated with the support at varying concentrations; (B) Estimated specific binding ratio of each scFv: molecules fluorescein/molecules scFv incubated with the surface.

3.4.4 Analysis of antigen binding by oxidised scFvs

Based on previous studies in the field (Hu et al., 2013) antigen-binding of scFvs was expected to be higher with oriented immobilised glycosylated scFv compared with adsorbed molecules: in this study, 6 µg ml\(^{-1}\) (11.25 pmol) full-length
glycosylated and short glycan 4M5.3 scFvs bound 5.7 ± 1.2 pmol and 5.2 ± 0.7 pmol fluorescein respectively when adsorbed onto ELISA plates (Figure 3.13 (A)) compared to 5.4 ± 1.6 and 7.4 ± 2.0 pmol fluorescein, (Figure 3.20 (A)), when the same molecules were covalently immobilised, Hu et al., meanwhile, observed as much as a 2.7-fold increase in antigen-binding from covalently immobilised scFvs compared to their adsorbed counterparts which is to be expected based on anticipated gains in orientation and stability. Due to the lower than expected antigen-binding overall, inhibition ELISAs were carried out with all three oxidised scFvs to determine any possible effects of oxidation on their antigen binding (Figure 3.21). IC_{50}s of the unglycosylated, full-length glycosylated and shortened glycosylated 4M5.3 scFvs were determined to be 4.97 ± 1.5 x 10^{-6}, 7.37 ± 1.5 x 10^{-6}, and 2.16 ± 1.2 x 10^{-5} M, respectively, after oxidation, which represented reductions in IC_{50} of approximately 14.5-fold in the case of the unglycosylated scFv, 7.5-fold for the full-length glycosylated scFv, and an 11.4-fold for the scFv with the shortened glycan chain. This indicates that oxidation of scFvs decreased the antigen-binding capability of scFvs significantly, presumably due to changes occurring throughout the rest of the protein molecule upon oxidation, whereas covalent immobilisation of the scFvs (Figure 3.20) revealed a marginal increase in antigen-binding due to immobilisation.

These results demonstrate the potential efficacy of directing immobilisation via the glycan tag but progress is currently limited by the effects of oxidation on scFv activity. The approach nevertheless provides a selective method to mediate specific and rapid surface attachment of glycosylated molecules. Given potential future elucidation of milder oxidation conditions could lessen the impact on the protein conformation and on antigen-binding capability, the approach may have the potential to provide a platform for selective, covalent, and oriented immobilisation of glycosylated proteins, while minimising adsorption of competing or undesired molecules.
Fig. 3.21 Inhibition ELISAs of oxidised 4M5.3 scFvs: unglycosylated (-pACYCpgl); fully glycosylated (+pACYCpgl); containing shortened glycoprotein (+pACYCpglJ\textsuperscript{276,284}).

3.5 Discussion

*Escherichia coli* is widely used in research and industrial settings as a host for recombinant protein production (RPP) due to factors such as its low cost of nutrients and quick doubling time. As *E. coli* has been in use for more than three decades, it has been extensively studied and benefits from well characterised genetics and a wide range of recombinant tools, specialised hosts, and vectors for production of recombinant proteins (Ramos *et al.*, 2004; Blount, 2015). While almost 30 % of all approved therapeutics are produced in *E. coli* (Graumann and Premstaller, 2006; Huang *et al.*, 2012), the host is limited by its natural inability to glycosylate proteins. The introduction of a protein glycosylation system from *Campylobacter jejuni* into *E. coli* (Wacker *et al.*, 2002) represented a milestone for the production of products such as glycoconjugates, vaccines and other glycan-functionalised proteins in *E. coli* (Fisher *et al.*, 2011; Ciocchini *et al.*, 2013; Ravenscroft *et al.*, 2016). This system has been utilised within our group to generate glycosylated scFvs for controlled attachment to aminated surfaces, resulting in stable, well oriented antibody fragments for improved detection of antigens (Hu *et al.*, 2013; Hortigüela *et al.*, 2015).

The main goal of this aspect of the present work was to modify the glycosyltransferase activity within the *C. jejuni pgl* locus in order to produce alternatively glycosylated scFvs in *E. coli*. In so doing, fundamental insights into the activity of the PglJ glycosyltransferase were also elucidated. Considerable
optimisation of previous glycoprotein expression in *E. coli* and recovery of soluble protein was also carried out, as well as comparison of antigen binding by different glyco-scFv variants. Finally, glycan-mediated immobilisation of scFvs was carried out with covalent attachment of scFvs on appropriately prepared surfaces.

Initial comparison and optimisation of expression of an anti-fluorescein scFv 4M5.3 and an anti-domoic acid scFv 2H12 previously studied in our group was carried out to select an appropriate scFv for glycosylation investigation and downstream analysis. Increasing the glucose concentration in auto-induction expression procedures was utilised to enhance total cellular biomass prior to induction, thereby increasing the yield of protein. Exploitation of the well-characterised lac operon allowed investigation into increasing the concentration of glucose by as much as 5-fold to significantly increase scFv expression yields due to the increased culture densities reached. Introduction of an additional PBS wash step during subsequent recovery of the soluble periplasmic protein extract after cell fractionation was also found to recover an additional 30-60% of active scFv, a step which was therefore employed in all future extraction procedures. These combined modifications provided a suitable and robust platform to produce sufficiently high levels of glycosylated scFvs in *E. coli* for downstream analysis, with the 4M5.3 fluorescein-binding molecule selected for investigation based on its higher yields.

Glycan chain assembly in the *C. jejuni pgl* system is carried out sequentially by a series of glycosyltransferases (Figure 3.1). In order to shorten the scFv-attached glycan, PglJ was identified as a candidate enzyme for mutagenesis and inactivation, leading to an expected N-linked disaccharide composed of GalNAcα1-3Bacβ1-NAsn (Figure 3.1, (c)). Based on the work of Troutman and Imperiali, in which PglH active sites were identified by comparison with the crystallised mannosyltransferase PimA in which the EX₇E motif has been shown to be integral to sugar nucleotide binding (Alexander *et al.*, 2004; Guerin *et al.*, 2007; Troutman and Imperiali, 2009), two similar EX₇E active site motifs were identified in PglJ for investigation. Single-site mutageneses of the two motifs and co-expression with the 4M5.3 scFv yielded the expected combination of unglycosylated protein and a shortened glyco-scFv product but also, unexpectedly, the persistence of a higher molecular weight band of a size similar to that of full-length glycosylated scFv. The presence of this apparently full-length glycoprotein was thought to result from either residual activity of PglJ or a
relaxed substrate specificity of the downstream glycosyltransferase PglH, allowing it to elongate a glycan chain terminating with an α1,3-linked GalNAc rather than its recognised acceptor substrate of a terminal α1,4-linked GalNAc. A similar relaxed acceptor sequence specificity has been observed in the *C. jejuni* oligosaccharyltransferase PglB, leading to it being exploited for glycoprotein engineering in *E. coli* (Feldman *et al.*, 2005). It is possible, therefore, that PglH might also have a non-unique acceptor substrate, leading to assembly and attachment to the protein of a hexasaccharide instead of the native *C. jejuni* heptasaccharide attached in the presence of a functional PglJ. Further chain elongation analysis and, particularly, mass spectrometry sequencing of the resultant products would help to identify this unexpected extended chain product.

Troutmann and Imperiali, however, reported that point mutations in the EX\_7\_E motif of PglH significantly reduced that enzyme’s glycosyltransferase activity. Given the homology between PglH and PglJ (Linton *et al.*, 2005), mutagenesis of the chosen EX\_7\_E motif in PglJ was expected to greatly reduce or eliminate its activity. Analysis of *pgl* constructs in which both active sites of the targeted EX\_7\_E motif (E276, E284) were therefore mutagenised revealed only shortened glycoprotein being expressed, leading to the conclusion that targeting of only one terminus of the motif had been insufficient to completely inactivate PglJ. This work confirmed the importance of the EX\_7\_E motif in the glycosyltransferase activity of the enzyme but indicated a different structure or molecular mechanism in PglJ compared with PglH (Troutman and Imperiali, 2009).

Further interruption of PglJ production by insertion of translation terminating cassettes upstream of potential active sites yielded glycoprotein profiles similar to those obtained in the double mutagenesis experiments. This provided further evidence that the full-length glycoprotein observed in singly-mutated PglJ expression experiments resulted from incomplete inactivation of PglJ, thereby providing sufficient template for PglH and subsequent PglI activity to assemble the full-length, "normal" glycoprotein. This also indicated that the flippase, PglK, which is responsible for transfer of the glycan chain from the cytoplasm, where it is assembled, to the periplasm, can transfer a diverse range of completed and partial glycoproteins, indicating a relaxed specificity for substrate on its part. While the mechanism for this is unclear, other transmembrane protein flippases, such as *E. coli*...
Wzx that is involved in LPS synthesis, have been demonstrated to have a requirement for GlcNAc at the reducing end of the glycan chain, while Alaimo et al. posited a requirement for an Und-P-linked 2-acetamido group for lipid-linked oligosaccharide flippase activities of enzymes including PglK and Wzx in *E. coli* (Alaimo et al., 2006). In the case of PglK in this work, it appears that assembly of the encoded glycan continues on the undecaprenyl-linked precursor and the resultant truncated glycan chain is transferred by PglK in the absence of a GlcNAc residue at its reducing end, thereby indicating an as of yet uncharacterised substrate capability of the *C. jejuni* flippase enzyme.

While multiple shortened glycoproteins were produced by mutagenesis of *pglJ* and *pglH*, a disaccharide-linked scFv was the shortest glycoprotein produced by glycosyltransferase manipulation in the work. As well as the importance of orientation, stability, and availability of the antigen-binding pocket in *in vitro* applications of immobilised scFvs, there are additional considerations for *in vivo* applications such as in the capture of endothelial progenitor cells to promote stent endothelialisation (Foerster et al., 2016). In this case, utilisation of a shorter glycan chain attached to the protein rather than the heptasaccharide resulting from the native *C. jejuni* machinery to mediate covalent, oriented attachment might also reduce the potential for an unwanted immunogenic response to the bacterial glyco motif, as well as potentially lessening susceptibility to glycosidase enzymes in the bloodstream by removing enzyme target sites or limiting accessibility.

This glyco-scFv was therefore selected based on its short glycan size for subsequent analysis in a variety of immobilisation-based assays, with comparisons carried out to the unglycosylated 4M5.3 scFv and the fully glycosylated variant. Antigen-binding of the three scFvs was determined and IC<sub>50</sub>s found to differ between the three by less than 10-fold, with these differences possibly resulting at least in part from the presence of other (glyco-)scFv variants in the fully- and partially-glycosylated scFv preparations due to incomplete scFv glycosylation in the *E. coli* host cells. scFvs were initially immobilised on polystyrene surfaces by adsorption, which is largely mediated by electrostatic forces, hydrophobic and hydrophilic interactions, and conformational changes within the protein (Nakanishi et al., 2001; Kusnezow and Hoheisel, 2003) -features which make this method much less specific than controlled, oriented methods and can lead to protein denaturation. Employing
adsorption results in protein molecules attaching to the surface in a variety of
orientations while the non-covalent nature of the protein-surface interactions leave
the potential for desorption depending on the downstream conditions, thereby further
reducing both efficiency and the potential longer term applications of this
immobilisation approach. Incorrect orientation of antibody or antibody fragment
molecules upon adsorption has been reported to lead to the binding pocket being
inaccessible to antigen in as many as 95 % of immobilised molecules (Davies et al.,
1994). Adsorption studies with the three scFvs in this work demonstrated that these
readily adsorbed within a 2 h period and typically bound approximately 0.33-0.5
picomole fluorescein per picomole adsorbed scFv (Figure 3.13 (B)), though this
molar ratio varies with scFv and antigen concentrations, and was significantly lower
with increasing concentration of scFv in particular, as expected (Figure 3.13 (B)).

Covalent immobilisation of the glycosylated scFvs was achieved by
oxidation of the $cis$-glycol groups in glycan chains using sodium $meta$-periodate,
thereby generating aldehyde groups for covalent bond formation with primary amino
groups on the surface of functionalised PureCoat plates. Time course experiments
revealed rapid immobilisation of the two glycosylated scFvs, while the
unglycosylated scFvs accumulated much more slowly on the same support. After
immobilisation for just 2.5 min, the glyco-scFvs also exhibited much higher
antigen:scFv ratios than the unglycosylated scFv, most likely indicating the correct
glycan-mediated orientation of the former on the surface. Even when immobilisation
was allowed to proceed for 2 h to accumulate unglycosylated scFv on the surface
also, the glycosylated scFvs still displayed much higher antigen-binding efficiency
than unglycosylated scFv (Figure 3.19). No significant difference was observed in
antigen-binding ability between the two glycoprotein formats. This work was a
highly promising indication of the ability to immobilise the scFv via a much
shortened glycan chain that provided significant improvements over its adsorbed
scFv counterpart while retaining the immobilisation capability and antigen binding
of the full-length glyco-scFv.

Investigation of the effects of oxidation on antigen binding by scFvs
demonstrated a much reduced level of antigen-binding of oxidised scFvs than in the
case of unoxidised molecules. This may be due to conformational changes induced
by the oxidation process and it would be instructive to identify less harsh or more
glycan-specific oxidation approaches that avoid damage to the protein molecule in order to avoid or reduce this effect in the future.

These experiments demonstrate the capability to intervene in the heterologous protein glycosylation pathway to alter the length of the glycan chain attached by the pgl system in E. coli to order. As the described modification did not significantly modify either antigen binding or immobilisation capability of the glycosylated scFvs compared to the full-length glycoprotein, the work raises the possibility of creating glycoengineered proteins that retain the advantages of the full-length glycan (covalent, oriented immobilisation) while avoiding some of their potential limitations (immunogenicity, susceptibility to hydrolysis). The work also provides greater insight into the functionality of the identified conserved glycosyltransferase motif EX7E in C. jejuni PglJ. These experiments can provide a platform for further modification of the pgl locus to engineer attached glycans of differing lengths or sequences, or for further extension of the glycan chain by providing a truncated foundation upon which more complex glycans can be attached (Schwarz et al., 2010). The disaccharide produced in this work could also be used to coat target proteins with multiple glycans for alternative attachment or construction of increased-potency conjugate vaccines through the use of additional glycan acceptor sequences which could be preferable to larger heptasaccharide glycan chains.
Chapter 4:
Investigation into extension of the established 
Campylobacter jejuni N-linked glycan in Escherichia coli 
and additional characterisations of glycosylated scFvs
4.1 Introduction

Recombinant protein production in *E. coli* is limited in its capacity to carry out post translational modifications (PTMs) and accordingly many therapeutic products have traditionally been produced in eukaryotic hosts. Post translational modifications are important for many essential properties and functions of proteins, such as their folding, stability, activity, and immunocompatibility. PTMs include sulphation, folding, glycosylation, phosphorylation, and disulfide bond synthesis. *N*-linked glycosylation is the most abundant post translation modification in human proteins, including many important biopharmaceuticals (Elliott *et al.*, 2003; Jefferis, 2005). In addition to conferring specific glycan-mediated functionalities in the case of molecules such as immunoglobulins (Varki, 1993; Arnold *et al.*, 2007; Hutzler *et al.*, 2008), *N*-linked glycans are also critical in protecting many proteins against proteolysis, assisting their folding (Imperiali and O’Connor, 1999; Shental-Bechor and Levy, 2008), or increasing their half-life *in vivo* (Solá and Griebenow, 2010).

When choosing an appropriate platform for recombinant production of human proteins, it is of note that approximately half of all human proteins (Wong, 2005) and 70% of candidate therapeutic proteins (Sethuraman and Stadheim, 2006) are glycosylated. In spite of this, one third of currently approved recombinant therapeutics are produced in *E. coli*-based expression platforms (Graumann and Premstaller, 2006; Walsh, 2014) even though natural glyco-products produced in this manner are typically aglycosylated (Huang *et al.*, 2012). This lack of their native glycosylation in *E. coli*-produced proteins can lead to significantly reduced activity when compared to the same product produced in Chinese Hamster Ovary (CHO) cells, as demonstrated in the case of interferon-β (Runkel *et al.*, 1998). This discrepancy between the large proportion of human proteins that have attached glycans and the inability of *E. coli* - the otherwise dominant protein expression platform - to accurately produce these products, has led to extensive efforts to “functionalise” traditional *E. coli* expression platforms with a glycosylation capability.

The introduction of an *N*-linked protein glycosylation system from *C. jejuni* into *E. coli* (Wacker *et al.*, 2002) appeared to open up the use of *E. coli* for the production of a wide variety of glycoproteins of research and industrial interest. In
its original format, this recombinant C. jejuni glycosylation system was not entirely fit-for-purpose in terms of producing human-like glyco-products, however, as the resultant glycoproteins bear a distinctive prokaryotic glycan signature which elicits an immunogenic response in humans in vivo (Weerapana and Imperiali, 2006). A number of efforts to alter the sequence length, composition and protein-glycan linkage of the native C. jejuni protein glycosylation machinery have followed, therefore, in an attempt to create an E. coli host cell capable of more accurately recreating humanised glycoprotein products.

Amongst these, the base glycan synthesis pathway was modified to produce a lipid- as opposed to bacillosamine-linked GlcNAc, followed by sequential assembly of the remaining glycan chain and attachment to the target protein in the normal manner. The glycan can then be trimmed from the purified protein by an endoglycosidase in vitro to yield a GlcNAc-tagged protein, to which a mammalian-like glycan can be attached by transglycosylation of a pre-assembled N-glycan species (Schwarz et al., 2010). Other studies in which the pACYCpgl-mediated glycosylation system has been engineered include modification of the glycan chain length to investigate glycosyltransferase activity (Glover et al., 2005; Linton et al., 2005; Troutman and Imperiali, 2009) and to exploit the oligosaccharyltransferase PglB to transfer modified glycans in applications such as assembly of conjugate vaccines, as discussed in Chapter 3. Desired modifications to glycan length include both its shortening, to reduce potential immunogenicity or glycosidase susceptibility in vivo, and lengthening, to introduce human-like motifs such as terminal capping of chains with sialic acid for improved pharmacokinetics (Lindhout et al., 2011). Sialylation of N-glycans is typically catalysed by α2,3- and α2,6-sialyltransferases (Kornfeld and Kornfeld, 1985; Wen et al., 1992), resulting in an α-linked sialic acid (Marth and Grewal, 2008). Sialylation is common in human proteins and is associated with a range of biological processes such as signalling and intercellular adhesion, functions which are dependent on the terminal position of the sialic acid(s) on glycan chains (Varki, 2007). Due to its key role in host immune regulation and terminal location on mammalian N-glycans, an array of pathogens have developed pathways with which to display sialic acid, either synthesised or scavenged from the environment, on their surfaces, resulting in enhanced pathogenicity (Vimr et al., 2004) or evasion of host cell recognition (Vimr and Lichtensteiger, 2002). Therefore,
while assembly of sialylated, eukaryotic-like glycoproteins is not carried out by *C. jejuni*, the species has been shown to be capable of sialylation in its lipooligosaccharide (LOS) metabolism (Louwen *et al.*, 2008; Ellström *et al.*, 2014). While the considerable metabolic burden on host *E. coli* cells of maintaining the ~20-kb pACYCpgl plasmid, and producing the 14 *pgl*-encoded protein glycosylation proteins required for glycosylation (Nothaft and Szymanski, 2010), may limit their ability to co-express additional glycosyltransferases with recombinant glycoproteins, previous studies have utilised yeast/bacterial artificial chromosome-based vectors in *E. coli* to maintain and express much larger constructs (Tao and Zhang, 1998), as such we nevertheless chose to investigate in this work whether enzymes specific for LOS sialylation in *C. jejuni* could be incorporated into *E. coli* cells harbouring the existing *pgl* system, resulting in the ability to sialylate recombinant proteins in the bacterial host. The addition of sialylation to the recombinant protein production arsenal in *E. coli* could greatly enhance the production of existing therapeutic products and enable expansion of the use of the host to incorporate many other proteins currently outside its capability.

In addition to *E. coli*, the yeast *Pichia pastoris* was utilised for recombinant expression of the galactosyltransferase *cgtB* from *C. jejuni* in this work. The use of *P. pastoris* in heterologous protein expression builds on research and molecular tools developed with *Saccharomyces cerevisiae* but it offers advantages over *S. cerevisiae* due to its slower fermentation rates, resulting in lower toxic by-product accumulation, and reduced hyperglycosylation, in particular hypermannosylation, of proteins (Cereghino and Cregg, 2000; Macauley-Patrick *et al.*, 2005). Expression of recombinant proteins in *P. pastoris* also exploits its methylotrophic characteristics in the use of the strong *AOX1* promoter which allows tightly regulated expression of gene products upon methanol induction (Tschopp *et al.*, 1987). Recombinant proteins can also be expressed intracellularly or can be secreted by the use of a number of signal sequence options, with the latter approach often preferred as *P. pastoris* secretes few native proteins, thus allowing relative ease of isolation and purification of recombinant molecules from culture media (Macauley-Patrick *et al.*, 2005).

Established methods for analysing oligosaccharide composition and length typically exploit glycosidase activities (Zhao *et al.*, 1997), such as that of Peptide-N-
Glycosidase F (PNGase F; cleaves the linkage between the asparagine residue in the protein and the innermost saccharide residue) or endoglycosidases (Endo H, F; leave a single Asn-linked GlcNAc residue), to release the glycan from the protein backbone via the non-reducing terminus. The former approach is complicated in the case of oligosaccharide products of the C. jejuni pgl system by the absence of known glycosidases that cleave the characteristic bacillosamine (on the reducing end of the oligosaccharide)-asparagine linkage. Meanwhile, the potential information available from endoglycosidase analysis of the glycan chain is limited by the proliferation of α1,4-linked GalNAcs in the C. jejuni glycan (Figure 3.1). Due to this lag in technical expertise in analysing prokaryotic glycans – and the unavailability of a mass spec-based glycan sequencing facility to support the present work – the outcomes of glycoengineering of the pgl system were initially evaluated using western blot analyses, followed by investigation using a novel, lectin array approach.

Lectins are carbohydrate-binding proteins that have been isolated from animal, bacterial, and, primarily, plant sources and may exhibit specificity for precise carbohydrate structures (Weis and Drickamer, 1996). Hundreds of lectins have been characterised and many are commercially available for glycan analysis, including multiple variants for many specific carbohydrates and linkages. Numerous lectin-based glycan detection assays have been developed which combine ease of access, reproducibility, and the facility to analyse multiple parameters (e.g. glycosylation pattern, specific linkages) simultaneously (Pilobello et al., 2005). In recent years, lectin-based detections have increasingly been used to classify glycans in a diverse range of applications, such as distinguishing bacterial strains and cell-surface changes under various environmental conditions (Kilcoyne et al., 2014), identifying potential roles in pathogenicity of glycoforms of proteins (Hsu et al., 2006), identifying markers for cell maturation in human invertebral disks (Collin et al., 2016), and developing cancer diagnostics (Tao et al., 2008). We utilised a lectin array developed in-house by the Alimentary Glycoscience Research Cluster (AGRC) at NUI Galway. The group has previously successfully applied this array or preceding versions in the characterisation of individual glycoproteins, bacterial species, and human tissue types (Kilcoyne et al., 2014; Collin et al., 2016). As the binding of many lectins to bacterial glycoproteins has been poorly studied to date, this array-based approach presented a valuable opportunity to circumvent these
limitations in typing glycoproteins produced by engineered pgl constructs in this work. In addition, the approach had the potential to generate important, fundamental information on the binding specificities of some of the displayed lectins. The microarray format also enabled a more efficient comparison of interaction of multiple glycoprotein samples with a wide panel of lectins than would be possible with more traditional, blot-based approaches.

Lectin array analyses were preceded, however, by screening of glycoproduct molecular weights in hexahistidine tag-based and DIG-glycan-based western blot analysis (Section 3.3.3), which easily distinguished fully glycosylated from unglycosylated scFv products based on the size differences between their differentiating oligosaccharides. By optimisation of the method, a variety of saccharide chains of intermediate lengths could also be identified and characterised. Lectin array analyses was then used to select appropriate individual lectins for use in lectin blots.

4.2 Investigating protein sialylation in E. coli

Previous work carried in our group investigated supplementing the C. jejuni-derived protein glycosylation machinery in E. coli with relevant “sialylation genes” from the LOS locus from C. jejuni OH4384 (S. Robin, unpublished data). The cluster of genes cloned into E. coli cells harbouring pACYCpgl encoded β1,3-galactosyltransferase CgtB, sialyltransferase CstII, and the three enzymes responsible for synthesis of the common sialic acid donor substrate CMP-NeuAc and which have been shown to be involved in sialic acid: NeuA1, NeuB1, and NeuC1 (Karlyshev et al., 2005). The goal was to extend the existing protein-linked heptasaccharide by three sugar monomers to yield the decasaccharide NeuAca2,8-NeuAca2,3-Galβ1,3-GalNAca1,4-GalNAca1,4-(Glcβ1,3)-GalNAca1,4-GalNAca1,4-GalNAca1,3-Bacβ1-NAsn (Figure 4.1).
Chapter 4

Fig 4.1 Hypothetical sialylated decasaccharide structure produced by the activities of a combination of C. jejuni protein glycosylation and lipooligosaccharide sialylation machineries. Extended section is highlighted and the glycosyltransferase activities associated with addition of each sugar monomer are indicated by colouring.

The glycan-extension cluster was cloned into the scFv-containing pIG6 expression vector to yield the 9510 bp pIG6-2H12NeuAc (Figure 4.2). Analysis of domoic acid-binding 2H12 scFv expressed from this vector, in E. coli cells also harbouring pACYCpgl, did not result in detectable scFvs of increased molecular weight that would be expected from extension of the glycan chain. Initial experiments to detect extended-chain glyco-scFvs produced using this experimental setup in the present work incorporated strategies to increase glycoprotein expression, such as increasing glucose concentrations in auto-induction expression experiments and recovering additional scFv from cell extractions (see Section 3.2.2) but scFvs of the expected increased molecular weight were not detected. Work therefore switched to focus on the anti-fluorescein 4M5.3 scFv due to its higher expression levels and the vector pIG6-4M5.3-NeuAc was produced by restriction digestion substitution of the scFv into the pIG6-2H12-NeuAc expression vector (Figure 4.2).
4.3 CgtB-mediated extension and expression in E. coli

Investigation of standard approaches to increase overall scFv expression, including induction methods, times, increased cell densities, increased recovery steps, and modification of antibiotic controls, did not yield detectable levels of extended 4M5.3 glycoprotein in host E. coli cells. This lack of extension was thought most likely to result from an inability to transfer the β1,3-linked Galactose to the terminal GalNAc of the glycan chain in the first stage of the designed glycan extension, for two reasons: firstly, because no glycoprotein product of increased size corresponding to a Gal-extended glycan could be detected and, secondly, as the natural substrate of the CgtB enzyme in C. jejuni is N-terminal β1,4-linked GalNAc (Gilbert et al., 2000), rather than the N-terminal α1,4-linked GalNAc available in the present experimental setup – whereas CgtB-mediated extension of the chain by the β1,3-linked Gal residue, if successful, would generate the natural glycan substrate for subsequent addition of α2,3-linked, followed by α2,8-linked NeuAc residues by CstII.

Fig 4.3 pIG6-2H12-cgtB expression vector containing 2H12 scFv and downstream cgtB with a C-terminal streptavidin tag.

Therefore, cgtB was subcloned into the standard pIG6 expression vector, downstream of the 2H12 scFv gene for increased expression and with a C-terminal
strep tag (Strep-tag II) (Figure 4.3) for its detection and purification in parallel to that of the hexahistidine-tagged scFv. Co-expression analyses did not detect hexahistidine-tagged glycoproteins of increased size. Meanwhile strep-tag-mediated detection of CgtB by western blotting was complicated by binding of the StrepMAB-Classic-HRP antibody reagent to an E. coli protein of the same molecular weight as CgtB (38.6 kDa) and attempts to purify CgtB from 0.5 l expression experiments using a Strep-Tactin® Superflow® column yielded no distinct bands at the size expected, indicating low-level or no expression of the enzyme in the host E. coli cells but reverse-transcription PCR (RT-PCR) was not carried out to investigate expression levels.

Further attempts to express cgtB at higher levels in E. coli included its expression with an N-terminal ompA leader peptide for secretion to the periplasm for improved folding; its expression in the absence of other recombinant proteins, including the scFv, to minimise metabolic stress on the host cells; its co-expression with E. coli disulfide bond isomerase DsbC (Zapun et al., 1995) as CgtB contains up to four disulfide bonds (DiANNA Disulphide Bond partner prediction (Ferrè and Clote, 2005; Ferrè and Clote, 2005; Ferrè and Clote, 2006)) and disulfide bond formation in the periplasm has been observed to be a rate-limiting step in E. coli (Georgiou and Valax, 1996; Nakamoto and Bardwell, 2004); its expression in E. coli BL21(DE3) cells using the strongly inducible pET-22b(+) expression vector; co-production of a panel of molecular chaperones (Hsp60 family, Hsp70 family, trigger factor (Nishihara et al., 1998; Nishihara et al., 2000)) for improved folding; its expression in E. coli Rosetta (DE3)pLacI cells (Novy et al., 2001) that incorporated additional tRNAs for all 8.6 % of cgtB codons classified as “rare” (GenScript Rare Codon Analysis Tool; E. coli Codon Usage Analyzer 2.1 (Maduro, 1996; GenScript, 2014)) and that might inhibit translation; and a variety of combinations of these approaches. No CgtB protein was detected by western blotting under any of these experimental conditions. Based on this exhaustive investigation and the distinctly different codon usage profile in the cgtB gene compared to E. coli it was decided to move the focus of expression of the enzyme into a new host platform.

4.4 cgtB expression in P. pastoris

Following review of a number of candidate hosts for production of CgtB, the
methylotrophic yeast *P. pastoris* was selected due to: its compatibility with available equipment and infrastructure used for *E. coli* expression; the existence of expertise and know-how in the group with the organism; and codon usage analysis of *cgtB* that demonstrated a much higher compatibility with expression in *P. pastoris* (only 1 % of *cgtB* codons below the 30 % threshold of codon abundance and none below the 10 % threshold (Figure 4.4; (Fuhrmann et al., 2004)), compared with 42 % and 8.6 %, respectively, for its expression in *E. coli*.

*cgtB* was subcloned into the *Pichia* expression vector pPICZα in the "A" reading frame, generating a construct including a C-terminal *c-myc* epitope and a polyhistidine tag for detection and purification, and an N-terminal α-factor secretion signal, expressed from the vector's *AOX1* promoter. Expressions were carried out at a small-scale (50 ml) over a 120 h period followed by upscaling (250 ml).

![Fig 4.4 cgtB codon usage analysis for expression in P. pastoris, carried out using graphical codon usage analyser (Fuhrmann et al., 2004). Grey bars indicate a tRNA abundance of less than 30 % in P. pastoris.](image)

Analysis of proteins in culture supernatants by western blotting and staining did not detect hexahistidine-tagged proteins. Proteins were detected by anti-hexahistidine antibody at low levels in cell lysates, however, though with a distinct double band pattern of higher molecular weights (approximately 45 kDa and 39.8 kDa) than expected (38.6 kDa) for mature CgtB (Figure 4.5, lanes 6, 7). As the CgtB protein had been produced as a fusion protein with the α-factor signal peptide but was not successfully secreted from the *Pichia* cells, it was hypothesised that the
signal peptide was not cleaved by either of the two cleavage mechanisms (Kev2, Ste13) which would result in a 47.9 kDa protein, potentially corresponding to the larger, approximately 45 kDa product in this analysis.

The Ste13 cleavage site is located two amino acids downstream of the Kex2 site and immediately prior to the CgtB peptide. Therefore, this larger product is likely to result from complete non-cleavage of the dual-signal peptide. Although many recombinant proteins are successfully secreted in *P. pastoris* with the α-factor signal sequence (Cereghino *et al*., 2002), several cases have also been observed in which the signal processing and secretion efficiency vary widely due in part to the structure of the protein of interest. In some cases the use of an alternative signal sequence overcame this bottleneck and would be the most likely avenue for further attempts at expression of CgtB in *P. pastoris* (Cereghino *et al*., 2002).

The lower band observed in anti-hexahistidine blots closely matched the expected size of processed CgtB but was found exclusively in intracellular protein preparations. As signal peptide cleavage should operate during secretion, it is unclear why the lower molecular weight product is observed intracellularly. While it is clear that some extracellularly secreted protein could remain in cells during harvesting, and intracellular protein fractions were 10X concentrated relative to culture supernatants upon electrophoresis, it seems unlikely that a high enough proportion of cleaved and secreted protein could remain attached to the expressing cells to explain the observed results. The shorter protein product may therefore be more likely to result from degradation during CgtB production, which might be investigated by the use of protease-deficient *P. pastoris* strains for expression which are widely adopted for industrial-scale expression of commercial proteins (Higgins and Cregg, 1998; Cereghino and Cregg, 2000; Macauley-Patrick *et al*., 2005).
Analysis of intracellular proteins prepared from *P. pastoris* cells after cgtB expression. (1) Molecular weight marker; Intracellular protein prepared from cells after (2) 0 h; (3) 2.5 h; (4) 16 h; (5) 24 h; (6) 48 h; (7) 72 h of expression. Expected size of CgtB was 38.6 kDa.

### 4.5 CgtB purification and oligosaccharide extension

Multiple expressions and purifications of CgtB from *P. pastoris* intracellular fractions under a variety of conditions yielded a maximum of 3 ml purified protein at 250 µg from a 150 ml expression (Figure 4.6). With an approximate size of 38 kDa, the purified protein was likely the lower band observed in intracellular fractions previously (Figure 4.4) and indicates that the unrecovered, larger product was likely largely insoluble protein. While at a very low concentration (84 µg ml⁻¹), the ability of the purified protein product to add the desired β1,3-linked Gal to the extant heptasaccharide to form an octasaccharide-linked glycoprotein was nevertheless investigated. A similar approach using CgtB contained within crude cell extract from fermentor scale *E. coli* expressions was previously reported to be successful (Blixt *et al.*, 2005) but no increase in the size of the acceptor glyco-scFv was detected in the present analysis. The bottleneck in determining the ability of CgtB to extend the current glycan in the desired manner therefore remained the availability of sufficient recombinant CgtB enzyme to carry out a more detailed analysis. Future work to overcome this, while outside the scope of the present project, might include synthesis of the cgtB sequence in a codon optimised manner for expression in *E. coli*, which would also allow incorporation of tools such as the co-expression of molecular chaperones for improved folding or for disulfide bond formation and isomerisation.
Fig 4.6 Analysis of IMAC-based purification of CgtB from the intracellular fraction of P. pastoris cells after expression. (A) Coomassie stained gel image. (B) Contrast-enhanced image of gel shown in (A). (1) Molecular weight marker; (2) soluble protein extract; (3) soluble protein extract after dialysis; (4) column flow-through; (5) column wash; (6 - 11) column elutions. Expected size of CgtB was 38.6 kDa.

In the case of continued expression attempts in P. pastoris, investigating alternative signal sequences could be beneficial to overcome potential related incompatibilities of the α-factor signal sequence with CgtB. Several well established and novel (Huang et al., 2011) signal sequences are available for exploitation with recombinant proteins in P. pastoris, while mutations in the α-factor signal sequence have also been demonstrated to lead to up to 40 % increases in yield of secreted target proteins (Lin-Cereghino et al., 2013). A parallel approach to enable increasing CgtB expression in P. pastoris might be adoption of a large scale continuous culture model though expression levels of the order of only five to six fold are often associated with such approaches (Evans and Ratledge, 1983; Goodrick et al., 2001). Nevertheless, combining these optimisation approaches with modifications in media composition and induction conditions might yield sufficient CgtB protein to carry out N-glycan extension as planned. These improvements in expression could be further supplemented with accessing more sensitive glycan analysis methods such as mass spectrometry to investigate whether any addition of the N-terminal Gal was occurring, at levels below the detection limits of our current analyses.

4.6 Alternate characterisation of glycoproteins

4.6.1 Lectin array analysis
Lectin-binding analysis of unglycosylated and glycosylated scFvs, expressed and purified as part of this project, was carried out by Dr. Michelle Kilcoyne of the Alimentary Glycoscience Research Cluster, NUI Galway. 4M5.3 scFvs produced by the following eight E. coli clones harbouring differently modified pgl genes (see Section 3.3) were subjected to this microarray analysis at 2.5 μg ml⁻¹: unglycosylated (-pACYCpglJ); full-length glycosylated (+pACYCpglJ); pglJ single site mutagenesis of E276 (+pACYCpglJ²⁷⁶); pglJ single site mutagenesis of E284 (+pACYCpglJ²⁸⁴); pglJ double site mutagenesis (+pACYCpglJ²⁷⁶,²⁸⁴); truncated pglJ (+pACYCpglJ); truncated pglH (+pACYCpglH); single site pglJ mutagenesis, truncated pglH (+pACYCpglJ²⁷⁶,pglH). All samples were also analysed for lectin binding after co-incubation with 100 mM Galactose to ensure specificity of any observed binding. Results of binding of the eight scFv variants to the 50 lectins on the array (Appendix i) were examined and data for GalNAc-specific lectins (Table 4.1) was analysed first due to their expected binding to the C. jejuni glycan chain. These results were shown in Figure 4.7, with normalised data for fluorescence intensity. Although the full-length glycoprotein containing the standard C. jejuni heptasaccharide displayed a clearly different binding pattern to the other scFvs, high similarity between pglJ-modified constructs was observed but lectin binding profiles were not visibly related to the expected differences in glycan composition between the proteins, other than the presence or absence of GalNAc.

Hierarchical clustering analysis was carried out with complete linkage and Euclidean distance to determine similarities between the lectin-binding profiles of the various glycoforms (Figure 4.8). This analysis provided further support for the similarity of the pglJ-modified constructs and indicated that these likely retained GalNAc in their chain sequences, as predicted from their genetic compositions (Figure 4.8). It was also observed that the full-length, heptasaccharide glycoprotein shared only 21 % similarity lectin-binding results with its closest neighbour, indicating a marked difference between the heptasaccharide and all other samples. These discrepancies with predicted binding profiles might result from a number of factors: lectin binding specificities have largely been determined with plant and animal glycans rather than bacterial proteins, which may lead to differences in recognition of characteristic prokaryotic sequence motifs. Furthermore, while the AGRC group have previously used lectin arrays to determine general glycan-binding
profiles from complex samples (Kilcoyne et al., 2014), the patterns observed from these studies did not translate well to distinguishing between highly similar glycoproteins and the lack of homogeneity of some of our glycan patterns (Figure 3.7) may therefore also have an impact on lectin binding profiles.
Fig 4.7 Lectin array analysis of all glycoprotein variants with reported GalNAc- and α-linked galactose-binding lectins. Legend for the lectin abbreviations is provided in Table 4.1.
Table 4.1. GalNAc-, α-Gal-, and Gal-binding lectins analysed in the lectin array study.

<table>
<thead>
<tr>
<th>Lectin</th>
<th>Organism</th>
<th>Common Name</th>
<th>Reported specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIA, Jacalin</td>
<td><em>Artocarpus integrifolia</em></td>
<td>Jack fruit lectin</td>
<td>Gal (sialylation tolerant)</td>
</tr>
<tr>
<td>RPBAI</td>
<td><em>Robinia pseudoacacia</em></td>
<td>Black locust lectin</td>
<td>Gal, GalNAc</td>
</tr>
<tr>
<td>SNA-II</td>
<td><em>Sambucus nigra</em></td>
<td>Sambucus lectin-II</td>
<td>Gal, GalNAc</td>
</tr>
<tr>
<td>DBA</td>
<td><em>Dolichos biflorus</em></td>
<td>Horse gram lectin</td>
<td>GalNAc</td>
</tr>
<tr>
<td>GHA</td>
<td><em>Glechoma hederacea</em></td>
<td>Ground ivy lectin</td>
<td>GalNAc</td>
</tr>
<tr>
<td>SBA</td>
<td><em>Glycine max</em></td>
<td>Soy bean lectin</td>
<td>GalNAc</td>
</tr>
<tr>
<td>VVA</td>
<td><em>Vicia villosa</em></td>
<td>Hairy vetch lectin</td>
<td>GalNAc</td>
</tr>
<tr>
<td>BPA</td>
<td><em>Bauhinia purpurea</em></td>
<td>Camel’s foot tree lectin</td>
<td>GalNAc/Gal</td>
</tr>
<tr>
<td>WFA</td>
<td><em>Wisteria floribunda</em></td>
<td>Japanese wisteria lectin</td>
<td>GalNAc/sulfated GalNAc</td>
</tr>
<tr>
<td>HPA</td>
<td><em>Helix pomatia</em></td>
<td>Garden snail lectin</td>
<td>α-GalNAc</td>
</tr>
<tr>
<td>GSL-I-A4</td>
<td><em>Griffonia simplicifolia</em></td>
<td>Griffonia lectin-I A4</td>
<td>GalNAc</td>
</tr>
<tr>
<td>PA-I</td>
<td><em>Pseudomonas aeruginosa</em></td>
<td>Pseudomonas lectin</td>
<td>Gal, Gal derivatives</td>
</tr>
<tr>
<td>EEA</td>
<td><em>Euonymous europaeus</em></td>
<td>Spindle tree lectin</td>
<td>α-Gal</td>
</tr>
<tr>
<td>GSL-I-B4</td>
<td><em>Griffonia simplicifolia</em></td>
<td>Griffonia/Bandeiraea lectin-I</td>
<td>α-Gal</td>
</tr>
<tr>
<td>MPA</td>
<td><em>Maclura pomifera</em></td>
<td>Osage orange lectin</td>
<td>α-Gal</td>
</tr>
<tr>
<td>VRA</td>
<td><em>Vigna radiata</em></td>
<td>Mung Bean Lectin</td>
<td>α-Gal</td>
</tr>
<tr>
<td>MOA</td>
<td><em>Marasmius oreades</em></td>
<td>Fairy ring mushroom lectin</td>
<td>α-Gal</td>
</tr>
</tbody>
</table>

Results of the microarray analysis were analysed to identify candidate lectin(s) that could be used to easily distinguish scFv glycoforms in lectin-based western blots. Due to the generally low binding in the majority of non-GalNAc binding lectins, GalNAc- and galactose-binding (as these should behave similarly to GalNAc-binding lectins (AGRC, personal communication, May 17, 2016)) lectins were focussed on (Figure 4.7 and Table 4.1). A major focus for selection of lectins for further analysis was a distinct difference between unglycosylated and glycosylated samples, represented in Figure 4.9, followed by a clear ability to distinguish between intermediate glycans (Figure 4.10-4.11). Differences between unglycosylated, glycosylated, and individual mutageneses of *pglJ* (Figure 4.11) and *pglH* (Figure 4.10) were analysed, with the disparity between lectin-binding profiles of the *pglJ*-modified constructs highlighted in Figure 4.11 (B). A primary candidate identified through this analysis was Soybean agglutinin lectin (SBA) from *Glycine max*, which is reported to selectively bind α- and β-N-acetylgalactosamine residues.
(Weis and Drickamer, 1996). As can be seen from the results (Figure 4.12), SBA clearly distinguished between fully-glycosylated and unglycosylated scFv (Figure 4.9), while it yielded an intermediate level of binding to scFv produced in a truncated pglH background (Figure 4.10), and to two of the three scFvs produced in modified pglJ E. coli cells (Figure 4.11 (B)). This is consistent with its selective detection of the pgl-produced heptasaccharide which contains five α-linked GalNAcs Fig 3.1 (a).

**Fig 4.8** Clustering analysis of glycoprotein variants with GalNAc and α-linked galactose binding lectins (complete linkage) (pgl284 was omitted to avoid repetition of pgl276 glycoprofile). Higher intensity binding is displayed in red, lower intensity binding is displayed in green. Clustering analysis grouped glycovariants based on similarity of lectin binding profiles for multiple lectins with associated grouping demonstrated with dendrograms.

**Fig 4.9** Lectin array analysis subset of full-length glycosylated (gly) and unglycosylated (ungly) scFv variants; normalised response is taken from full panel of lectins shown in Figure 4.7.
Fig 4.10 Lectin array analysis subset of full-length glycosylated (gly), truncated \textit{pglH} (H), and unglycosylated (ungly) scFv variants; normalised response is taken from full panel of lectins shown in Figure 4.7.

(A)
Figure 4.11 Lectin binding analysis of subset of (A) full-length glycosylated (gly), truncated pgLI, and unglycosylated (ungly) scFv variants; and (B) pgLI modifications demonstrating the inter-sample relationships: single site mutagenesis (J^276); double single site mutagenesis (J^276,284); truncated. Normalised responses are taken from full panel of lectins represented in Figure 4.7.

As the SBA array data revealed a relatively strong signal amongst the GalNAc-binding lectins and a distinctive difference in binding of the heptasaccharide (five GalNAc species) and the unglycosylated sample, albeit with the variable binding to the intermediate-length glycopeptides also observed amongst all GalNAc-binding lectins (Figure 4.12), this was selected for analysis in individual-lectin blot format. ScFvs expressed in the pgI cells (predicted to contain two GalNAcs) exhibited a reduced signal compared to the heptasaccharide. Meanwhile, scFvs produced in E. coli pgI mutant cells are expected to yield potentially heterogenous mixtures of scFvs (see Section 3.3) containing one GalNAc (completely inactivated PglJ), one and two GalNAcs (incompletely inactivated PglJ and completely inactivated PglH), or one-five GalNAcs (incompletely inactivated PglJ). Of these glycan species, SBA bound scFvs expressed in pgI and pgI^276 pgI^2 pgI^276,284 E. coli cells at a reduced level compared with unglycosylated scFv, and scFvs expressed in pgI^276 E. coli cells at a similar level to the unglycosylated protein (Figure 4.12). Binding to scFv from E. coli pgI^276,284 cells was particularly inconsistent with expected results. Broader analysis of binding results identified that this glycoprotein exhibited the highest signal with 15 of the 17 focussed lectins –
higher even than the fully glycosylated scFv. This was despite results previously outlined in Figure 3.7 that indicated its likely glycan composition as containing just a single GalNAc.

**Fig. 4.12** Analysis of binding of SBA to glycoprotein variants in lectin array. Numbers in brackets indicate the number of GalNAcs expected to occur in the relevant glycan chain; numbers separated by '/' indicate the occurrence of a mix of more than one glycan chain types.

The discrepancy between results observed with the SBA lectin array and with anti-polyhistidine peroxidase-based western blots, wherein an analysis of intermediate glycan chain length does not correlate between the two approaches is potentially explained by examining the suitability of SBA and indeed all lectins as candidates for distinguishing glycan chain differences on short oligosaccharides. SBA is a large 110 kDa tetramer with four binding sites per molecule (Dessen et al., 1995). Due to its large size it is unlikely that more than one SBA molecule can bind to a pgl-produced glycoprotein, whether it contains one or five GalNAcs. Additionally most lectins best recognise sugars in terminal positions in glycan chains (Weis and Drickamer, 1996), therefore lectins may be better suited to determining the number of oligosaccharide chains as opposed to the length of bound chains as multiple binding events are unlikely to occur within a single glycan chain associated with this project. Nevertheless, the microarray investigation yielded sufficient data to justify further evaluation of the potential of SBA to distinguish between the scFv glycoforms of interest in this work in individual lectin blots.
4.6.2 Lectin blot analysis of glycoproteins

Based on the lectin array analysis described above and previous reports of the use of SBA in analysing *C. jejuni pgl*-directed glycans in western blots (Linton *et al.*, 2002; Wacker *et al.*, 2002; Linton *et al.*, 2005), scFvs produced in this work from the *pgl*-engineered constructs were investigated on membranes using biotinylated SBA and peroxidase-conjugated streptavidin for detection (Figure 4.13).

**Fig 4.13** SBA lectin-based detection of glycoforms of 4M5.3 expressed in *E. coli* cells harbouring different pACYC*pgl* constructs. (1) Molecular weight marker; scFvs expressed by cells containing (2) pACYC*pglJ*; (3) pACYC*pglJ276,284*; (4) pACYC*pglH*; (5) pACYC*pglJ276*; (6) pACYC*pglJ276*pglH*; (7) pACYC*pglH*; (8) No *pgl* genes. Expected size of full-length glycosylated scFv was approximately 34 kDa; trisaccharide glycosylated scFv was approximately 32.6 kDa; disaccharide scFv was approximately 32.4 kDa; unglycosylated scFv was 32 kDa.

Lectins, despite their high specificity for particular carbohydrate structures, often exhibit weaker, non-specific binding to peptides. In this investigation, weak binding to unglycosylated scFv can be observed (Figure 4.13; lane 8), both in the case of the scFv expressed in the absence of *pgl* genes and throughout most other samples due to the presence of unglycosylated scFv in these also, as can be seen in Chapter 3 (Figure 3.7). The results clearly demonstrate products of the expected sizes in lanes 2 and 3 (*pglJ* and *pglJ* double mutant, respectively), in which scFvs with an attached disaccharide are expected to be produced, and in lane 7 (unmodified *pgl* genes), where the fully glycosylated scFv is apparent as the upper, dominant band. The upper band in lane 4 (*pglH*) is characteristic of the trisaccharide observed in previous *pglH* knockout analyses but the basis for the much stronger signal observed with the lower, unglycosylated scFv is unclear in this case. Meanwhile, lane 5 (*pglJ* single mutant) contains scFv containing the disaccharide associated with
Alterations (also present in lanes 2, 3) and a higher molecular weight, apparently full-length glyco-scFv (also visible in lane 7) that typically persists in *pglJ* single site mutageneses. This latter product (in lanes 5, 7) appears to stain more strongly with SBA-based detection than previously observed in anti-histidine based detections (Figure 3.7, lanes 7 and 2/9 respectively), most likely due to the specificity of SBA for GalNAc residues. Finally, lane 6 (*pglJ*<sup>276</sup>*pglH*) also appears to exhibit scFv containing the disaccharide associated with *pglJ* alterations (lanes 2, 3, 5). It is noteworthy that a number of bands detectable in hexahistidine-based probing of the same samples are not evident with this SBA binding, such as the putative trisaccharide-linked scFv in the case of the *pglJ*<sup>276</sup>*pglH<sup>-</sup> construct (Figure 3.7, lane 8). The reasons for the apparent lack of detection of these additional bands is unclear but may be due to differing sensitivity of the lectin for the glyco-scFv products, compared to the anti-hexahistidine antibody. Glyco-scFv's with less intense signals in this SBA blot were also observed to exhibit lower signals in general in the multi-lectin array.

A comparison of densitometry-measured detection of glyco-scFvs in the SBA blot and of fluorescence intensity readings obtained from SBA in the lectin array are shown in Figure 4.14, with both sets of data normalised against the highest signals obtained in the respective assays. As expected, patterns of SBA binding of the scFvs were broadly similar between the two approaches. While somewhat higher signals are apparent for *pglJ* in the SBA blot and for the unglycosylated scFv in the array setup, experiments would need to be repeated a number of times to check these apparent differences. Nevertheless, the comparison indicated the potential to use lectin arrays as a first-screen to identify lectins that could be used to discriminate between glycoprotein variants in a blot detection experiment, with the added use of recombinant tag (e.g. 6His) tag detection to provide additional information for the analysis. This body of work also demonstrated the unsuitability of the queried lectins for distinguishing between different short glycan chains, particularly when the associated differences are based on changes of the number of repeat units of a single monosaccharide; while lectin analysis is clearly applicable in validation of the presence of glycans and has been shown to be useful in differentiation of longer polysaccharide chains (Neu *et al.*, 2001), it does not appear to be well suited to this short prokaryotic glycan chain. As has previously been indicated, the reported glycan
specificity of lectins may not always be accurate with potential cross-reactivities and unreported secondary binding specificities common, while lectin specificities have also been shown to differ between mono-, oligo-, and polysaccharides (Strathmann et al., 2002). Furthermore, the binding specificity of most lectins is much less well characterised for bacterial than eukaryotic samples, as communicated by the AGRC. With the potential for different binding specificities of lectins than those reported, it would be of interest to repeat the lectin array analyses and query a broader pool of lectins which were not originally considered, thereby potentially discovering lectins more suited to validation of alternate glycans produced by bacteria harbouring the modified pgl constructs.

![Fig 4.14](image_url) Comparison of normalised response of SBA-based detection of glycoprotein variants in lectin array (purple) and blot (orange) experimental setups.

### 4.7 Glycan-mediated immobilisation of scFvs

Previous work in our group established methods to immobilise glyco-scFvs via their glycan chains on ethylenediamine (EDA)-agarose beads, leading to higher ligand-binding activities and increased stabilities over non-covalently adsorbed scFvs (Hu et al., 2013). As the combination of covalent attachment and directed orientation of the binding pocket were attractive features of this methodology, we investigated in the present study whether the approach could be extended to differently functionalised surfaces and/or to glycoengineered scFvs prepared in this work.
Amine-functionalised titanium-coated stainless steel 316L disks were provided by a collaborator (synthesised and coated by Mr Łukasz Wasyluk, Balton Sp zoo, Warsaw, Poland; amine-functionalised by Prof. Halina Podbielska, Wrocław Technological University (WrUT), Poland) for scFv immobilisation assays. This material is of interest in an ongoing cardiovascular stent functionalisation project. Amine-functionalised disks were composed of 316L stainless steel with a proprietary titania-based coating which increased hydroxyl group density for silanisation with 3-(aminopropyl) triethoxysilane (APTES), the terminal group of which provided a uniform layer of primary amines (Foerster et al., 2016). Full-length glycosylated fluorescein-binding 4M5.3 scFvs were oxidised as outlined previously (Section 2.2.10.7) and their covalent attachment to the aminated surfaces was carried out via the resultant aldehyde groups generated in the glycan chain. Detection of scFvs was then carried out using a fluorescently labelled anti-hexahistidine antibody which allowed direct visualisation of immobilised scFv molecules via fluorescence microscopy.

Initial optimisation experiments involved identification of a suitable scFv labelling approach: a Cyanine5 NHS ester (Cy5) was initially used but found to react with the aminated disks while an anti-histidine DyLight® 488 operated within the excitation and emission spectra of fluorescein, leading to the adoption of DyLight® 650, which operated at a similar excitation and emission spectra to Cy5, for scFv labelling.

Non-oxidised and oxidised full-length glycosylated 4M5.3 scFvs were immobilised at concentrations from 50-250 µg ml⁻¹ while low- (5 mM NaCl, 5 mM sodium phosphate, pH 6.0) and medium- (120 mM NaCl, 150 mM sodium phosphate, pH 6.0) salt buffers were investigated to reduce non-specific binding. High-salt, low pH washes (100 mM NaCl, 300 mM sodium phosphate, pH 4.5) were also utilised to reduce non-specific binding of protein, as well as a range of blocking regimens.

This led to determination of optimal conditions for immobilisation of oxidised and non-oxidised scFvs in 120 mM sodium phosphate buffer, 150 mM sodium chloride, pH 7.4 and in 5 mM sodium phosphate buffer, 5 mM sodium chloride, pH 7.4, respectively, with the higher salt conditions used for oxidised scFvs.
to reduce non-specific immobilisation of protein molecules. Following immobilisation of 1.25 µg of oxidised or non-oxidised scFv in this manner, disks were washed as outlined previously (Section 2.2.12.4) to remove non-specifically bound protein and surfaces were imaged (Figure 4.15 (A;C)). Further washes with high salt and low pH buffers were used to further remove non-specifically bound protein before re-imaging (Figure 4.15 (B;D)) and measurement of fluorescein binding by the surface-immobilised scFvs. Imaging analysis revealed a higher level of oxidised, glycosylated scFv to be bound to the disks than to non-oxidised scFv. As immobilisation of oxidised scFv was expected to be via covalent reaction, the observation of higher scFv signals after disk washing was also expected, with reduction in signal likely to be attributable to removal of non-specifically (non-covalently) attached scFv and photobleaching of the fluorophore before imaging (Vogelsang et al., 2008). Overall, the work established conditions for largely uniform coating of the disk surfaces with glycan-oxidised scFvs, while non-oxidised scFvs were observed in occasional larger, likely aggregated occurrences on the disk surfaces.

Analysis of fluorescein binding of the disk-immobilised scFvs was not possible as incubated fluorescein molecules appeared to crystallise on the disk surface, leading to no detectable antigen-binding signal under any scFv immobilisation conditions (not shown). The relatively large numbers of disks necessary to further investigate and optimise antigen binding were not available due to an increasing focus of the collaborating manufacturer on the use of application-relevant scFvs and so no further investigation was carried out with the 4M5.3 scFv-functionalised surfaces. Similarly immobilised scFvs against endothelial cells prepared in our group, however, demonstrated cell capture and mediated surface coating of disks with cells when immobilised using the procedure developed and described here (Foerster et al., 2016), indicating the potential of scFvs to be covalently attached in this manner while maintaining antigen binding properties.
Further to the results of scFv immobilisation outlined above, oxidised and non-oxidised variants of the full-length glyco-scFv were expressed, purified and fluorescently tagged with anti-hexahistidine DyLight® 488 for immobilisation by partners on cardiovascular stents, using the parameters described for flat disk analysis above (Figure 4.16). Amine-functionalised titanium-coated stainless steel 316L stents were manufactured at Balton and functionalised at WrUT as described above, followed by scFv immobilisation and stent washing by the WrUT collaborator. Imaging by fluorescence microscopy clearly demonstrated the efficacy of utilising covalent immobilisation of oxidised scFvs in this manner, with a clear coating of the stent surface evident (Figure 4.16 (D)). While further investigation of the stability of immobilisation and of antigen binding of the scFvs was not carried out, these results, combined with the ligand-binding evaluation carried out with disk-immobilised scFvs described above, indicate the potential application of the surface-
attachment approach in functionalising a wide variety of surfaces for in vivo applications.

![Image](image1.png)

**Fig 4.16** Fluorescence microscopy images of glycosylated 4M5.3 scFv immobilised on stent surfaces. (A) non-oxidised scFv on aminated stent; (B) oxidised scFv on non-aminated stent; (C) oxidised scFv on aminated stent; (D) oxidised scFv on aminated stent (note lower magnification) ((D) has had brightness increased by 10% for the legibility).

In order to better understand the basis of surface-immobilisation of the glycosylated anti-fluorescein scFv, the scFv was prepared and purified, in unglycosylated and glycosylated formats, and provided to collaborators at WrUT for immobilisation and atomic force microscopy (AFM) analysis. ScFvs were immobilised on aminated stent material at WrUT as previously described and AFM used to provide topographical and phase contrast information which can be interpreted to indicate scFv position (Figure 4.17) and height (Figure 4.18) on the surface as well as surface coating composition (Figure not shown). AFM analysis demonstrated the distribution of scFvs on the surface and their height. The data indicated a non-uniform distribution of the scFvs across the surface, with peak heights that exceeded that of the calculated molecular size, suggesting possible aggregation of the scFvs. By comparison, cysteine-functionalised scFvs subjected to
the same analysis on thiol-displaying surfaces exhibited more uniform distribution and expected peak heights (M. Kopaczyńska, personal communication, 21 January, 2016), though *E. coli* expression of the scFvs was severely compromised by cysteine-mediated aggregation of the modified antibody fragments. Overall, therefore, the combined results from disk coating and stent functionalisation above, together with this AFM study, indicate sufficient potential in the coating approach to warrant more detailed characterisation and evaluation for future application in diverse material functionalisation applications.

**Fig 4.1** AFM topography image of oxidised 4M5.3 scFv covalently immobilised on aminated stent material indicating height and distribution of immobilised scFvs. Intensity of bright spots correlates to the height of immobilised scFvs as related to the legend on the right of the image.

**Fig 4.18** 3D topographical image of oxidised 4M5.3 scFv, covalently immobilised on aminated stent material, generated from AFM data. Peaks represent surface distribution and molecular heights of immobilised scFvs.
4.8 Discussion

There exists an obvious requirement for additional options for production and engineering of glycoproteins in *E. coli*. The current *C. jejuni* *pgl*-mediated system is capable of producing a defined heptasaccharide glycan signature or, by exploiting truncated glycan assembly pathways, shortened glycoproteins consisting of subsets of the heptasaccharide. Components of the *pgl* pathway have also been exploited to carry out modified functions, as in the case of the oligosaccharyltransferase PglB and its adoption to mediate attachment of alternate glycan structures to proteins with the consensus acceptor sequence for production of conjugate vaccines (Ihssen *et al.*, 2010; Garcia-Quintanilla *et al.*, 2014; Ihssen *et al.*, 2015; van den Dobbelsteen *et al.*, 2016). The protein glycosylation locus represents a potent system for glycoprotein production in *E. coli*, therefore with its full range of potential applications yet to be realised or developed. The work presented in this chapter describes efforts to further develop the potential of this system by extending the extant oligosaccharide, thereby producing an oligosaccharide of increased length and incorporating alternate monosaccharides to provide a basis for future sialylation, as well as exploitation of the attached glycan moiety to utilise glycoproteins in biotechnological applications.

Initial work to extend the glycan chain to a sialylated glycoform identified no detectable extension. From this, expression of the galactosyltransferase CgtB in *E. coli* in a variety of formats was investigated, incorporating alternate recombinant tags, molecular chaperone co-expression, and expression in a variety of strains. Although expression in a codon-bias alleviating strain did not lead to expression of detectable CgtB, this was compounded by slower growth rates and lower cell densities achieved with the strain compared to typical expression strains, likely due to the increased metabolic burden of additional plasmids (Diaz Ricci and Hernández, 2000; Ow *et al.*, 2006). Nevertheless, *in silico* analysis identified codon bias as a potential bottleneck for expression in *E. coli* as up to 42 % of the *C. jejuni* cgtB codons were present below 30 % abundance in *E. coli*. For comparison, the coding sequence of the well-expressed 4M5.3 scFv has no codons below this abundance level, further illustrating the necessity of adapting codon usage for heterologous gene expression. This difference originates from the *C. jejuni* source of cgtB (Muto and Osawa, 1987) with this species having a GC content of 30.6-36.4 %, compared to 50.79 % for *E. coli* (Muto and Osawa, 1987; Poly *et al.*, 2004; Mohan and
Stevenson, 2013; Lowe, 2016). For future cgtB expression studies in *E. coli*, the coding sequence should clearly be synthesised *de novo* to optimise its codon usage profile for the host organism (Burgess-Brown *et al.*, 2008), after which the multiple optimisation steps investigated in this study could be re-studied to further improve production of the CgtB protein.

*CgtB* expression in *P. pastoris* was investigated utilising a well established system for high-level expression and secretion. Yeasts are typically employed for the expression of heterologous proteins such as those sourced from other eukaryotes rather than *E. coli* as a result of their capability to carry out post-translational modifications including glycosylation, as previously discussed. Yeast strains generally express at much higher levels than their prokaryotic counterparts and are readily employed in industrial settings for their highly efficient secretion pathways (Daly and Hearn, 2005) and high titre expression levels (Demain and Vaishnav, 2009). In addition to alleviating codon usage chokepoints identified in *E. coli*, *P. pastoris* is highly efficient at disulfide bond formation and as such would be of interest for *cgtB* which exhibits four disulfide bonds (White *et al.*, 1994). Considering its GC content of 41% (De Schutter *et al.*, 2009), *P. pastoris* is also much closer to that of *C. jejuni* (30.6-36.4%) thus potentially favouring successful translation of the *cgtB* gene in this host system.

As described in Section 4.4, the yields of purified CgtB obtained from expression in *P. pastoris* were much lower than expected. Possible problems with secretion could be relatively easily addressed, as discussed, but low expression levels would likely require in-depth optimisation to achieve more efficient protein expression in this host. An alternative approach would be to express the protein intracellularly but it would be advantageous to retain the use of this broadly efficient secretion system in *P. pastoris* if possible. Increased expression levels could also be achieved by switching to a continuous culture model but as previously discussed this would likely not be singularly sufficient to achieve the desired levels of expression in the absence of additional optimisation of growth and expression, as discussed below.

Further investigation into the following aspects of expression have been used previously to increase expression of heterologous proteins in *P. pastoris*: altering
media composition; optimisation of GC content; utilisation of alternate strains; and adoption of alternative integrations sites or models (Sreekrishna et al., 1997). The most readily queryable of these factors is media composition and conditions, including modification of the concentration of the induction component - typically methanol - to identify the optimal induction concentration; switching from a complex medium to a defined medium with a set pH to inhibit certain proteases activities which could impact on the low expression levels observed (Macauley-Patrick et al., 2005); aeration; and alteration of the lengths of biomass generation and induction stages to increase cell density and reduce lag phases for glycerol depletion (Cos et al., 2006). It is clear that multiple avenues could be investigated to improve CgtB expression in P. pastoris in the future, with the main drawback being the considerably longer doubling time of P. pastoris compared to E. coli, which would result in a significant time investment to investigate multiple different parameters. Additionally in both E. coli and P. pastoris-based expression analyses it would be of interest to assess cgtB expression via RT-PCR analysis to further illuminate the suitable avenue which should be investigated to obtain CgtB.

Oligosaccharide extension analysis was carried out to test functionality of any produced galactosyltransferase CgtB. In this work, purified heptasaccharide-glycosylated scFv, galactose in an appropriate donor substrate format, the relevant enzyme cofactors, and the appropriate conditions were provided to catalyse addition of a β1,3-linked galactose to the extant oligosaccharide. No detectable size increase was observed however. The natural substrate of the CgtB enzyme in its native host is N-terminal β1,4-linked GalNAc (Gilbert et al., 2000), it is therefore possible that catalysis of a β1,3-linked GalNAc will not be possible with CgtB, despite this, relaxed acceptor substrate specificity has been a noted feature of several bacterial glycosyltransferases (Hoffmeister et al., 2003; Minami and Eguchi, 2007) and cannot be ruled out until sufficient CgtB is obtained to test this hypothesis. As the work was based on a protocol for oligosaccharide extension incorporating crude E. coli cell extracts from large scale fermentations expressing CgtB (Blixt et al., 2005), co-factors or other components available in the crude cell extract, not present in this work, may have made oligosaccharide extension viable. Perhaps more likely, the lack of extension may have been due to insufficient levels of CgtB present in this work: with the range of bottlenecks potentially limiting progress in this experiment,
the basis of this failure will be unclear until higher levels of CgtB can be obtained for further investigation.

Alternate characterisation of the glycosylation pattern and effects of \textit{pgl} alterations was necessary in the absence of mass spectrometry based analysis. As the exact oligosaccharide composition had previously been determined by us (Appendix ii) and by others (Wacker et al., 2002), select lectins were utilised for analysis of the glycoproteins produced in this project, with specific focus on those previously reported to have specificity for N-Acetylgalactosamine or galactose. Although previous analyses carried out by the AGRC had distinguished between multiple strains or markers for tissue maturity (Collin et al., 2013; Kilcoyne et al., 2014; Collin et al., 2016), the work carried out in this project was concerned with differentiating between purified glycoproteins with differences in the number of units of GalNAc and the presence/absence of glucose, for which lectins were not incorporated due to spatial limitations.

As discussed below specifically in relation to SBA, lectin analysis is useful in characterising the presence or absence of different monosaccharides but is limited in its capacity to differentiate shorter glycans, due both to simple physical characteristics and their propensity for binding terminal monosaccharides. In this regard the sensitivity of lectin arrays is in its infancy compared to much better-established, rapid, and sensitive nucleic acid microarrays which have been readily employed in clinical diagnostics, environmental monitoring, expression profiling, comparative genomics, and other applications over the past two decades (O'Connor and Glynn, 2010). Due to the uncharacterised interaction of many studied lectins with bacterial glycans, it would be of interest to either analyse a broad range of lectins in the current array format to identify alternative primary or secondary binding specificities to \textit{pgl}-generated glycans, or to immobilise the \textit{C. jejuni} glycan signature in glycan or glycoprotein format on an array and query a panel of labelled lectins to determine differential binders, similar to approaches used to identify glycan-binding proteins (Blixt et al., 2004). Lectin arrays have previously demonstrated their suitability for broad glycan analysis with complex samples, providing distinctive glycan-binding profiles of multiple lectins (Angeloni et al., 2005; Zheng et al., 2005; Kilcoyne et al., 2014): more detailed evaluation of bacterial glycan-specific lectin responses may allow for identification of lectins more
suited to distinguishing between modified *pgl*-generated glycans, thereby facilitating similar broader characterisation of specific bacterial glycoprofiles.

SBA had previously been used in western blotting for detection of *pgl*-produced glycoproteins (Linton *et al.*, 2002; Wacker *et al.*, 2002; Linton *et al.*, 2005). In our array format, the differences in SBA-binding between unglycosylated and full-length glycoprotein were clear but intermediate glycans did not yield signals corresponding to the expected signal differences (Figure 4.12). The binding affinity of lectins can vary greatly depending on the structure of the considered glycan (Cummings and Etzler, 2009), however, and the interaction of SBA with this bacterial-produced, truncated oligosaccharide structure was previously uncharacterised. SBA is 110 kDa and contains four carbohydrate binding sites per molecule, as previously discussed (Section 4.6.1) it is likely that the differing oligosaccharide composition of glycoproteins investigated in this work would not permit differential binding by SBA, i.e. the observed SBA binding signal is not proportional to chain length or GalNAc composition of the bound glycoprotein, beyond the presence of a single GalNAc monomer. Binding proportional to glycan chain length has, however, been reported with polysaccharides (Dam and Brewer, 2010). Furthermore, lectins with the same binding affinity typically share fundamental structural features and binding mechanisms, it seems unlikely that a much higher potential for differentiation could be expected from additional lectins with substrate specificities similar to SBA (Weis and Drickamer, 1996).

Upon consideration of the full range of galactose and N-acetylgalactosamine binding lectins in array format (Figure 4.7), followed by hierarchical clustering (Figure 4.8), it was determined that specific GalNAc-mediated binding was occurring but that it lacked the resolution or specificity to consistently identify differences between oligosaccharides containing one, two, or five GalNAcs. The wide range of lectins analysed did allow selection of appropriate lectins for lectin blot analysis, however.

Lectin blot-based analyses were found to broadly correlate with array analysis and to support the predicted changes to glycoprotein composition conferred by *pgl* alterations. SBA lectin blot analysis detected low levels of unglycosylated scFv but were largely specific for glycosylated scFvs and as such provided
supplementary evidence to support interpretations of earlier analysis by anti-polyhistidine antibody and DIG-glycan-based western blots. The individual SBA lectin blot (Figure 4.13) was largely in agreement with the SBA lectin array data (Figure 4.14), apart from the signal observed for scFv produced in the presence of pACYCpglJ (J).

Additional methods for demonstrating covalent immobilisation of glycosylated scFv and antigen binding were also developed. Covalent immobilisation of oxidised, glycosylated scFvs on amine-coated PureCoat 96-well microplates demonstrated the efficacy of pursuing covalent immobilisation of the scFv via the glycan and its subsequent ability to carry out antigen-binding compared to an adsorbed counterpart. The preliminary immobilisation work carried out on amine-functionalised titanium-coated stainless steel 316L disks demonstrated a viable future platform for carrying out antigen-binding comparisons between adsorbed and covalently bound scFvs. While immobilisation conditions were developed to favour covalent attachment of scFvs, experimental conditions would need to be relaxed to allow increased adsorption to occur if comparisons of antigen-binding of covalently attached and adsorbed scFvs were to be made. Stability of covalently bound scFvs and bound antigen could be likewise further investigated in a range of conditions such as temperature, over time to determine stability, and for resistance to enzymatic degradation, similarly to that performed in assays employing fluorescent immunohistochemistry (Kocbek et al., 2007; Hortigüela et al., 2015).

The supporting evidence provided by AFM analysis established the spatial positioning, orientation and distribution of immobilised scFvs and could be incorporated in situ to carry out additional analyses such as measuring antigen binding efficacy of immobilised scFvs and the binding forces of the antibody-antigen interaction as other groups have demonstrated (Hinterdorfer et al., 1996; Allen et al., 1997; Ros et al., 1998). This would supplement antigen-binding analyses carried out in the microplate format. Covalent immobilisation on amine-functionalised titanium-coated stainless steel 316L stents demonstrated the efficacy of oxidised, glycosylated scFv immobilisation on an aminated surface via a covalent linkage and the importance of the presence and proper activation of each component in this process. The results presented for aminated microplates and the aminated stent disk material, coupled with characterisation via AFM analysis, demonstrate the viability of
covalently immobilising glycosylated scFvs on appropriate surfaces, and the molecules' subsequent antigen-binding and stability. These results exhibit the potential of this system for developing a designed functionalised surface. Coordinated binding of a glycosylated molecule in this manner could also be envisaged to be broadly applicable to develop bioconjugates for biomolecule detection or binding. By glycosylating alternative targets, meanwhile, this covalent immobilisation method could be employed to direct immobilisation of any glyco-molecule for the development of biosensors, in therapeutic delivery, and in creating bioactive surfaces for in vivo and in vitro approaches.

Microscopic-based characterisations of immobilised glycosylated scFv were carried out with the heptasaccharide, C. jejuni-directed glycan variant. It would be of interest to perform the same analysis with the disaccharide variant to investigate any detectable changes. Although improved antigen-binding was not observed with the shortened glycoprotein, neither was binding diminished. It would therefore be of interest to investigate the long term stability of immobilised glyco-scFvs in vitro when exposed to glycosidase enzymes - assuming potentially reduced accessibility of substrate sites in the shortened format. Any increased stability would be of significant interest in in vivo applications such as the development of functionalised stent surfaces with immobilised human EPC capturing scFvs (Foerster et al., 2016). The shortened glycan chain could also benefit the construction of bioconjugates for therapeutics, such as nanoparticle-based models for drug delivery and radiolabelling/radionuclide delivery, due both to reduced glycosidase sensitivity, and potentially reduced immunogenicity of the attached glycan due to its reduced size. This model could also benefit from further humanisation of the attached glycan to further reduce the risk of an immunogenic response as discussed below (Chapter 5).

The experiments presented herein demonstrated a detailed investigation of the expression of C. jejuni cgtB in E. coli and P. pastoris, and attempts to carry out oligosaccharide extension with the expressed material. Improved results may be obtained in the future upon codon optimisation of the cgtB gene. The glycan and glycoprotein characterisation described demonstrated additional confirmation of oligosaccharide changes mediated by the pgl pathway engineering and established methods for studying and interpreting changes conferred to glycoproteins expressed in the presence of these constructs. Additionally, microscopy-based approaches
displayed the capability of the glycosylated scFv to be immobilised covalently on an aminated surface and the potential for visualising its activity *in situ* on functionalised surfaces.
Chapter 5: Discussion
5.1 Discussion

The goal of this project was to investigate the extant *C. jejuni* protein glycosylation system expressed in *E. coli* for ways to shorten and lengthen the glycan chain attached to target proteins, and evaluate associated changes in immobilisation and/or activity of variant glycoproteins upon their covalent immobilisation to supports via the glycan. The specific objectives towards achieving this goal were: enhancing recombinant production of scFv antibody fragments in *E. coli*; engineering the glycosylation pathway by site-directed mutagenesis to produce a shortened glycan chain; extending the glycan chain by expression of additional glycosyltransferases; characterising the glycan composition and binding of produced glycoproteins; and comparing the activity of glycoproteins upon directed immobilisation. The body of work also highlights the potential for producing engineered glycoproteins with a view to adding to the current toolkit for custom glycoprotein synthesis in *E. coli*.

The antibody fragment with which expressions and glycan modifications were characterised was selected based on its higher expression levels than other scFvs used in our group, and the ease and low cost of working with its antigen fluorescein, thereby facilitating downstream activity-related assays. The anti-fluorescein scFv 4M5.3 was previously generated by affinity mutagenesis from its progenitor 4-4-20 (Kranz and Voss, 1981; Boder *et al.*, 2000) and was recognised as the highest known affinity in an engineered antibody fragment. The underlying mechanisms for this affinity improvement were determined to be multiple minor structural alterations that work in concert to enhance affinity including residue changes which were not directly related to antigen-binding (Midelfort *et al.*, 2004; Midelfort and Wittrup, 2006). The 4M5.3 scFv has been used previously in our group for a range of expression and biochemical studies and is well suited for investigating expression conditions and downstream analyses (Hu *et al.*, 2013; Hortigüela *et al.*, 2015).

The characterisation and subsequent functional transfer of the first established bacterial *N*-linked protein glycosylation pathway to *E. coli* in 2002 (Wacker *et al.*, 2002) paved the way for a plethora of new research to be carried out in recombinant glycoprotein production. It has also enabled consideration of a range
of additional applications for *E. coli*, from the production of proteins with mammalian-like glycosylation patterns (Schwarz *et al.*, 2010) to the synthesis of highly effective conjugate vaccines (Price *et al.*, 2016; Ravenscroft *et al.*, 2016). The continued development of the technology will provide a platform for the production of additional therapeutic proteins in *E. coli* (Pandhal and Wright, 2010; Baeshen *et al.*, 2015). It is likely that the platform will continue to build on existing publications and patents in multiple directions, including - importantly - further humanisation of the current oligosaccharide, as evidenced in recent humanised sialylation of *O*-linked glycoproteins in *E. coli* (Skretas *et al.*, 2009).

Previous work in our group has utilised the glycan tag attached to a *C. jejuni* pgl-directed glycosylated scFv to mediate covalent attachment of the glycoprotein to aminated surfaces (Hu *et al.*, 2013; Hortigüela *et al.*, 2015). The glycan tag is situated on the carboxyl terminus of the scFv; by covalently immobilising via this distally-located tag the fluorescein binding ability of the antigen-binding pocket should be unaffected. It was of interest in this work to modify the glycan chain length in order to investigate potential effects on the capacity of glycosylated scFvs to be covalently immobilised or the biochemical properties of the immobilised protein. By employing a targeted mutagenesis approach to generate shortened glycan chains which could then be extended by the actions of tailored glycosyltransferases or by the addition of pre-assembled glycan chains, we hoped to circumvent limitations of current platforms established for synthetic synthesis of custom glycans for biomedical applications (Hurevich and Seeberger, 2014).

Shortened glycans would also present an attractive option with regard to reducing potential immunogenicity of the standard *C. jejuni* heptasaccharide; for an *in vivo* application such as functionalising stents (Foerster *et al.*, 2016), the inherent immunogenicity of the *C. jejuni* glycan could pose a serious risk factor in biomaterial or device rejection (Szymanski *et al.*, 1999). By reducing the glycan chain length, the amount of exposed glycan available for immune recognition should be significantly reduced while continuing to allow oriented surface tethering of the attached protein. In combination with methods to glycosylate proteins at multiple locations within a target region (Fisher *et al.*, 2011), this could allow for increased stability of the bound glycoprotein by incorporating multiple shortened glycans in the place of a single full-length glycan.
The approach employed in this project to extend the glycan chain is designed to ultimately lead to sialylation of the produced glycoproteins. Sialylation is an important determinant of human immunotolerance of glycoproteins. With sialic acids commonly found in the terminal position of eukaryotic glycoproteins and glycolipids, this process also plays a key role in the pathogenicity mechanisms of many bacteria (Severi et al., 2007). Importantly also for therapeutic applications, sialylation has been used to confer increased serum half-life on engineered proteins (Lindhout et al., 2011). As a number of mammalian sialyltransferases have been demonstrated to be poorly expressed in recombinant form in bacterial hosts (Tsukamoto et al., 2007), bacterial sources of suitable sialyltransferases have been investigated. Many such enzymes have been found to display more relaxed acceptor specificity than their mammalian counterparts, which could aid in the transfer of diverse heterogenous sialic acids to proteins in RPP systems (Gilbert et al., 1997). Although many bacterial sialyltransferases purified to date have also exhibited low yields and/or activities, efforts to improve these enzyme characteristics have met with considerable success, such as the use of directed-site-saturation mutagenesis to increase by up to five-fold the activity of an α2,6-sialyltransferase enzyme, with minimal structural perturbations detected in the improved enzyme (Choi et al., 2014).

Cloning of the genes encoding glycosyltransferase enzymes for synthesis of the appropriate sialic acid donors and the precursor glycans necessary for sialylation was carried out previously in our group (S. Robin, unpublished data). Expression of these genes and evaluation in an E. coli production platform was unsuccessful, however, in achieving detectable sialylation of the target protein. Extensive analysis was carried out in the present study to extend this previous work: the galactosyltransferase CgtB was identified as the primary target to initiate extension of the glycan chain as it transfers the first additional monosaccharide at the non-reducing end of the chain (Gilbert et al., 2000). Multiple engineering and expression attempts, in both E. coli and P. pastoris, yielded only barely detectable levels of purified CgtB enzyme which did not successfully extend the glycan chain, either due to insufficient amounts of CgtB available or inactivity of the enzyme. This body of work indicated that glycan extension in this manner would likely require a significant further effort for successful realisation, particularly if further
sialyltransferases were to be incorporated for subsequent addition of sugar monomers to the chain. Codon optimisation of relevant genes for the expression host might be critical for successful expression of heterologous glycosyltransferases, while the identification of alternative bacterial sources of glycosyltransferases might also be a fruitful approach: a functional α2,3-sialyltransferase, WbwA, recently isolated from a pathogenic *E. coli* strain (Czuchry *et al.*, 2015) was shown to be active with a Galβ1,3-GalNAc acceptor glycan in synthesising an O antigen precursor (Wang *et al.*, 2014). Prokaryotic sialyltransferases of varied regioselectivity have been successfully expressed in *E. coli* and it is clear therefore, that there is potential to identify the necessary synthesis and transferase activities from prokaryotic sources to recreate desired glycan chain sequences in host systems such as *E. coli* (Drouillard *et al.*, 2010; Richard *et al.*, 2016). Recent directed engineering of human sialyltransferases, including site-directed mutagenesis of exposed hydrophobic regions and expression with solubility-enhancing fusion partners, have also allowed for their greatly improved expression in *E. coli* and enhancement of their activities (Ortiz-Soto and Seibel, 2016). These studies provide potential insight into possible approaches that might be used to improve expression of enzymes such as CgtB in future work.

Use of the pACYCpgl vector to enable glycosylation of acceptor-sequence-tagged proteins necessitates expression in *E. coli* ΔwecA CLM37 strain to avoid native transfer of a GlcNAc, rather than the pgl mediated GalNAc, to the lipid-linked precursor of the glycan chain thereby ensuring homogeneity of the resultant glycoproteins. Enhancement of scFv expression and folding by our in-house system of molecular chaperone co-expression (Hu *et al.*, 2007; Hortigüela *et al.*, 2015) was not possible due to the incompatibility of the chloramphenicol-resistant Takara plasmid system (Nishihara *et al.*, 1998; Nishihara *et al.*, 2000) with the similarly Cam<sup>R</sup> pACYCpgl vector. In any case, the metabolic burden associated with plasmid maintenance and replication in *E. coli* is well established and the introduction of a third, 5-11 kb chaperone-encoding plasmid to cells already harbouring a scFv-expression vector (~ 4.6 kb), and a glycosylation vector (~ 20 kb) may be overly detrimental to protein expression (Glick, 1995; Carneiro *et al.*, 2013). In this work, expression optimisation instead employed approaches such as modifications to induction and periplasmic protein extraction which would be broadly applicable
across many *E. coli* strains, including CLM37. A number of the improvements established in this work have now been incorporated into standard group protocols to enhance expression and recovery of all soluble proteins produced in *E. coli* in the lab.

A consistent feature of glycoprotein expression in the present *C. jejuni* pgl-based experimental setup was the production of a mixture of unglycosylated and glycosylated scFv molecules with all pgl constructs. The persistence of unglycosylated scFv in bacterial extracts suggests a rate-limiting step in the synthesis of glycan chain precursors, assembly of the glycan chain, transport across the cytoplasmic membrane, or transfer of the assembled glycan to the scFv. While covalent immobilisation of oxidised, glycosylated scFvs is mediated by the glycan chain and should lead to elimination of the unglycosylated scFv, and blot-based analyses were developed to distinguish amongst differently-sized glycoforms, it would clearly be preferable to produce entirely homogenous glycoproteins for use in downstream analyses and for ease of interpretation of results. In the case of inefficient or inadequate pgl gene expression causing this bottleneck, a simple approach such as lowering the expression temperature to assist scFv folding would likely still result in unbalanced over-production of the scFv due to its significantly higher-copy-number plasmid than in the case of pACYCpgl (Rosano and Ceccarelli, 2014). Therefore, re-cloning of the scFv gene into a low-copy-number plasmid might favour production of homogenously glycosylated glycoproteins by producing the scFv at a similar rate to the requisite glycans; furthermore the use of a dual low-copy plasmid setup would be expected to lead to better folding of the expressed protein and a higher quality yield of glycosylated scFv overall (Glick, 1995; Jones *et al.*, 2000). In terms of glycan transfer, the capability of the oligosaccharyltransferase PglB, albeit recombinantly produced in the host *E. coli* cells in this work, should be largely unaffected as it is attaching *C. jejuni*-defined glycans to its native acceptor sequence. A recent study has shown that the position and sequence of the acceptor sequence within an scFv can greatly influence glycosylation efficiency, however, while larger scale cultures can also benefit from change of the pgl vector resistance cassette from chloramphenicol to kanamycin (Lizak *et al.*, 2011). Therefore it may be beneficial in future work to investigate alternative sites for placement of the glycan attachment sequence in the scFv (logically either at the N-terminus or in the
inter-domain linker) in order to check if more complete glycosylation could be achieved.

Previously, the PglJ glycosyltransferase had been determined using, knockout mutagenesis, to catalyse the transfer of an α1,4-linked GalNAc to the non-reducing end of the N-linked GalNAcα1,3-Bac disaccharide (Linton et al., 2005). By modelling our site-directed mutageneses on similar work carried out with PglH glycosyltransferase that acts next in the glycan chain (Troutman and Imperiali, 2009), the criticality of their previously flagged amino acids to activity of the glycosyltransferase was investigated. Unlike in the case of PglH, wherein substitution of one of the two conserved glutamic acids in the $^{276}$EX$^2_{7}$E$^{284}$ motif was sufficient to curtail glycosyltransferase activity, it was noted with PglJ that complete inhibition of the enzymes activity in glycan extension was only observed upon substitution of both glutamic acids; single residue changes only reduced the glycosyltransferase efficiency, resulting in a mix of disaccharides (truncated) and heptasaccharides (full-length) being transferred to target scFvs. Previous studies with EX$^7$E-containing glycosyltransferases have assigned an essential role to the first Glu residue in the motif in the transferase activity of the enzymes (O'Reilly et al., 2006) which was therefore first targeted in the present study whereas mutagenesis of the second Glu residue resulted in only partial inactivation of the enzymes (Abdian et al., 2000). Given our results, the role of this motif in glycosyltransferase activity has clearly not yet been fully elucidated: while it is clear that it can play an essential role as a whole, the importance of the individual glutamic acid residues may have been previously overstated, or they may assume different importance in different enzyme homologues due to their positions in the structure of the glycosyltransferase or other differences between enzymes. Nevertheless, this method of truncating the $C. jejuni$ glycan chain can be readily employed for synthesising a protein-linked oligosaccharide of desired length in $E. coli$. This approach can aid in building custom glycans and could be amalgamated into current systems to produce humanised $N$-glycans in $E. coli$, such as enzymatic trimming of a constructed glycan in vitro to generate the appropriate acceptor sequence for the humanised glycan (Schwarz et al., 2010). Site-directed mutagenesis of this type could simplify such procedures by reducing the need for enzymatic trimming of glycoproteins post-purification, thereby also increasing product yields by removing additional
A glycoengineering approach for producing shortened glycans in mammalian cells by truncating the synthesis pathway has recently been developed (Meuris et al., 2014). This system yields trisaccharide-tagged glycoproteins with attached Neu5Ac-α2,3-Gal-β1,4-GlcNAc N-glycans, in addition to monosaccharide- and disaccharide-tagged glycoproteins at levels far lower than the heterogeneity observed in glycoproteins produced in unmodified mammalian cells. Recombinant IgGs produced in this setup retained their native molecules' binding rates of neonatal Fc receptors but exhibited marked reductions in binding to Fc-γ receptors, thereby increasing their retention in vivo and enabling them to remain at the required load for therapeutic effects for up to 12 days longer than the unmodified IgGs. This clearly demonstrates the potential value in vivo of this type of "GlycoDelete" approach. Our tailoring of the C. jejuni pgl glycan machinery, similarly, in combination with future expected work to sialylate glycoproteins in E. coli, has potential to significantly add to the recombinant protein production toolbox by broadening the scope of applications for which "human-like" recombinant glycoproteins can be produced in the bacterial host.

Lectin microarrays were investigated in this study to distinguish between recombinantly produced glycoproteins with different glycan sequences. Analyses focussed largely on reported α1,4-GalNAc binding lectins, as the main expected difference between the recombinant glycoproteins resulting from pgl pathway engineering was the number of repeated α1,4-linked GalNAc units contained within the attached glycan chains. Although the array approach revealed a clear difference between the lectin binding profiles of fully glycosylated and unglycosylated scFvs, delineation of signatures between di-, tri-, and hepta-saccharide-linked proteins was less clear. Ultimately, the lectin microarray analysis successfully identified potential lectins that could be used to characterise the different glycoforms in western blot-based analyses. While the lectin microarray approach has enormous potential in glycan analysis and will find increasing usage in the detection of simple and complex oligosaccharides in applications such as glycan profiling of cancers and tissue characterisation (Kuno et al., 2005; Uchiyama et al., 2008; Nishijima et al., 2012; Collin et al., 2016), the approach may not be sufficiently sensitive to discriminate between glycoprotein variants, as in this work, and particularly when...
those variants differ only by the numbers of repeating units of a particular monomer within the attached oligosaccharide. The 110 kDa size of the SBA lectin used in this analysis almost certainly precludes multiple binding events occurring along a single heptasaccharide chain. Although it has been shown to display increased binding affinity with an increasing number of saccharide units (Dam and Brewer, 2010), it is likely that this differentiation between different glycan chain lengths is limited to, or at least most efficient in the case of, polysaccharides in which there is the space for multiple binding events to occur, unlike the short oligosaccharides of interest in this project.

As previously discussed, extensive western blot analysis was employed to characterise the glycoforms expressed from alternative pgl constructs. As the sequence of the full-length heptasaccharide resulting from full pgl pathway expression had previously been confirmed by MALDI-MS by our group (Appendix ii) and others (Wacker et al., 2002; Lizak et al., 2011), it was possible to interpret differences between glycoprofiles on western blots that arise from modification of glycosyltransferase enzymes. While anti-polyhistidine-based detection provides a crude analysis of differences between glycoforms, the addition of glycan-specific DIG- and lectin-based western blot analyses complemented these results and provided further evidence to support their interpretation. Lectin blotting is an efficient preliminary method for analysis of complex and defined glycoprotein samples (Cao et al., 2013) and continues to find broad usage in the detection and characterisation of glycoproteins (Ross et al., 2016). Lectins specific for the pgl-encoded glycan chain (containing multiple α-linked GalNACs) were also previously used for detection and differentiation of glycan patterns from unmodified pgl constructs (Linton et al., 2002; Wacker et al., 2002; Young et al., 2002). DIG-glycan-specific detection can be employed in a range of glycan detection applications (Zafred et al., 2013; Flood-Nichols et al., 2014; Çiftci et al., 2016); although this approach is generally employed in parallel with lectin-based differentiation between molecules, in this project the DIG-glycan detection kit could differentiate between glycoforms produced using different pgl constructs, thereby supporting results observed in western blot analyses. Although the combined western blot and glycan analyses approach were not nearly as comprehensive as glycan mass-spectrometry analysis, in the absence of such a technical capability for this work, their use in
combination with data from the previous characterisation of the heptasaccharide nevertheless provided a sufficiently detailed package of data to interpret results of \( pgl \) pathway engineering.

Protein adsorption is a low-cost, low-specificity, and low-complexity method of immobilisation which typically yields molecules with sufficient functionality and stability for use in methods that require low stringency. This has led to its widespread application in coating of inorganic materials, such as single-walled carbon nanotubes for use as paraoxon detectors (Liu \textit{et al.}, 2007), or the adsorption of enzymes to silicates for biosensor development (Küchler \textit{et al.}, 2015). ScFvs are readily employed as bioconjugates in biosensors or functionalised surfaces, and as tools for targeted therapeutic delivery. Adsorbed scFvs have been widely utilised for, e.g., the development of a sensitive immuno-QCM-based sensor by modified and unmodified scFvs to demonstrate the efficacy of a directed adsorption approach (Kumada \textit{et al.}, 2011), and in our group for the development of a sensitive environmental immunosensor (Hu \textit{et al.}, 2006; Hu \textit{et al.}, 2007).

The method by which scFvs are immobilised is critical to the functionality, stability, and feasibility of the downstream application, however, and as such is an important consideration in design of a bioconjugate. Due in part to the emergence of improved surface functionalisation methodologies and the ease of engineering of proteins by recombinant systems, covalent attachment approaches are increasingly utilised for the development of bioconjugates, providing high specificity, sensitivity, and versatility (Hermanson, 2013; Smith \textit{et al.}, 2013; Liébana and Drago, 2016). Covalent immobilisations are typically used for applications in which non-specific adsorption of analytes is undesirable, and a high activity of the immobilised protein is paramount, as adsorbed proteins can suffer from extensive denaturation (Gray, 2004). Covalently immobilised protein bioconjugates offer a potentially improved interface for most applications in which their adsorbed counterparts currently find widespread use: biosensor development, such as in the development of a glucose biosensor via covalently immobilised glucose oxidase (Fu \textit{et al.}, 2011); the development of a novel paper-based immunoassay platform as a rapid, easy-to-use biosensor (Cao \textit{et al.}, 2015); the development of a novel biodegradable bioconjugate for the delivery of covalently immobilised vascular endothelial growth factor, leading to improved and more rapid vascularisation in treated rat models (Miyagi \textit{et al.}...
al., 2011); and the use of epidermal growth factor-binding mAbs, immobilised on single-walled carbon nanotubes, to deliver anticancer radionuclides (Spinato et al., 2016).

ScFvs have been covalently immobilised on supports via a range of recombinant tags, innate or engineered amino acids, and chemical modifications, as discussed in Chapter 1. The recent development of the pgl-based protein glycosylation system in E. coli has allowed for use of the attached oligosaccharide to covalently immobilise proteins onto surfaces with primary amines, via aldehyde groups generated by mild oxidation of the glycan. This method has previously been used in our group to generate immunosensors for the detection of fluorescein (Hu et al., 2013), functionalised polymers with improved antigen-binding of three glycosylated scFvs compared to their adsorbed counterparts (Hortigüela et al., 2015), and functionalised stent surfaces that exhibit improved cell coating and re-endothelialisation, mediated via tethered scFvs (Foerster et al., 2016). Immobilised glycans have also been readily employed in microarray technologies to detect and analyse glycan binding proteins (Rillahan and Paulson, 2011), leading to elucidation of cell-adhesion mechanisms (Nimrichter et al., 2004) and development of better understanding of host immunity mechanisms (Arthur et al., 2014). Other research on immobilisation of the pgl-encoded heptasaccharide involved validation of a glycophage system used to detect glycan-binding proteins via the full-length heptasaccharide or alternative glycan substrate products of the C. jejuni OST PglB (Çelik et al., 2015). This technology could also be readily adapted to transfer custom glycans to target proteins and, in combination with the present method for synthesising shortened glycans, could be a valuable platform for investigating the potential immunogenicity of presented glycans in vitro by analysing their glycan-binding proteins. As discussed above also, the shortened glycan tags developed in this project may represent an attractive option via which proteins can be immobilised for in vivo applications as this may combine increased long term stability of the immobilised proteins over adsorbed molecules with lower susceptibility to glycosidases than in the full-length glycans.

Antigen-binding experiments in this study revealed no difference in affinity for antigen of disaccharide-tagged scFvs compared to heptasaccharide-tagged scFvs. Due to this, it may be beneficial to produce the disaccharide-linked molecule for
immobilisation applications in order to reduce somewhat the metabolic burden on the cells associated with synthesising the extended glycan chain. It would be of interest to measure comparative growth rates, protein expression levels, and glycosylation efficiency in such clones in the future. Consideration of potential optimisation strategies for the covalent immobilisation procedure would also be worth investigating: milder oxidation could be beneficial to reduce immobilisation of unglycosylated scFv molecules and loss of binding activity in oxidised scFvs. Meanwhile, adsorption of a recently-developed approach to transfer multiple glycans to different sites in a single protein could be used to enhance immobilisation or long-term stability of covalently-attached scFvs, mediated by multiple glycan-amine interactions (Fisher et al., 2011). As the amounts of covalently-bound scFv remaining on aminated plates after immobilisations were not quantifiable due to their low concentrations, it would also be of interest in future studies, to incorporate more sensitive detection methods, such as QCM-D (Chen et al., 2010) or sensitive protein quantification kits, e.g., FluoroProfile protein quantification kit by Sigma or the NanoOrange protein quantitation kit by Thermo Scientific to provide this information. Alternatively, the concentration of initial protein and/or dimensions of the support surface could be significantly increased to allow for western blot-based quantification of scFvs remaining in the supernatant post-immobilisations.

5.2 Conclusion

The goals of this project were to investigate modifying the glycan chain encoded by the C. jejuni pgl pathway for attachment to scFvs in E. coli and to analyse the effects of this varying chain length on the scFv molecules. Current studies incorporating the C. jejuni glycan machinery either utilise the full-length, unadulterated heptasaccharide chain or focus largely on exploitation of the broad substrate specificity of the oligosaccharyltransferase PglB to transfer non-native glycans to proteins of therapeutic interest for potentially lucrative applications such as conjugate vaccine development. In this work, progress was made on shortening the glycan chain through an easily adoptable approach which will add to future glycoengineering studies; additionally, the shortened glycan was demonstrated to still mediate covalent immobilisation of the glycosylated scFvs on appropriate surfaces. Progress was also made towards synthesising an extended glycan in E. coli,
leading ultimately to a sialylated recombinant protein.

The method by which the glycan chain was shortened was based on studies with the glycosyltransferase PglH (Troutman and Imperiali, 2009). It provides insight into the role of the conserved EX\textsubscript{7}E motif in activity of the PglJ enzyme compared to in previously characterised mannosyltransferases. This analysis also demonstrates that it is possible to identify critical features for protein function without available structural data, but that the exact binding mechanism for the sugar nucleotide donor may vary between dissimilar glycosyltransferases.

Covalent immobilisation and antigen-binding activity of the disaccharide-linked scFv were the same as in the heptasaccharide-linked molecule. Ongoing investigation of the stability and glycosidase susceptibility of the immobilised scFvs will determine whether the shorter glycan confers improved, long-term stability and viability of the immobilised scFv in applications of interest in our group, such as mediating endothelial cell attachment to cardiovascular stents (Foerster et al., 2016) or therapeutic delivery of scFv-bearing particles for regeneration of degenerated tissues (Cunningham et al., 2013).

Overall, this work has modified the heterologous protein glycosylation machinery in \textit{E. coli} to produce a disaccharide-glycosylated scFv, and extended understanding of the molecular mechanisms of the \textit{pgl} protein glycosylation machinery of \textit{C. jejuni}. In the process, the study produced a potentially beneficial alternate glycan chain for attachment to target proteins in \textit{E. coli}, thereby expanding the recombinant molecular toolbox available for protein expression in \textit{E. coli}. 


References


Casteleijn MG, Urtti A, Sarkhel S. Expression without boundaries: Cell-free protein
References


Desplancq D, King DJ, Lawson ADG, Mountain A. Multimerization behaviour of single chain Fv variants for the tumour-binding antibody B72.3. Protein Eng 7 (1994) 1027-1033.


Dugas V, Elaissari A, Chevalier Y. Surface sensitization techniques and recognition receptors immobilization on biosensors and microarrays, Recognition Receptors in Biosensors: Springer. 2010, pp. 47-134.


References


Helms B, van Baal I, Merkx M, Meijer EW. Site-Specific Protein and Peptide


Jones KL, Kim S-W, Keasling JD. Low-Copy Plasmids can Perform as Well as or Better Than High-Copy Plasmids for Metabolic Engineering of Bacteria. Metab Eng 2 (2000) 328-338.


Kapust RB, Waugh DS. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8 (1999) 1668-1674.


References


Mohan V, Stevenson M, Campylobacter jejuni genomes exhibit notable GC variation within housekeeping genes 2013.


Reichert JM. Antibodies to watch in 2015. mAbs. 2015; pp. 1-8


Roche Applied Science. Agarose Gel DNA Extraction Kit. 2008. Last accessed on
References

07-09-16. https://pim-eservices.roche.com/LifeScience/Document/2ba0e6ed-d2ed-e311-98a1-00215a9b0ba8


Shen Z, Yan H, Parl FF, Mernaugh RL, Zeng X. Recombinant Antibody


References


References


Uhlén M, Nilsson B, Guss B, Lindberg M, Gatenbeck S, Philipson L. Gene fusion


Varki A. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 3 (1993) 97-130.


Appendices
Appendix i:

Table i Lectins used in lectin array analysis.

<table>
<thead>
<tr>
<th>Lectin</th>
<th>Organism</th>
<th>Common name</th>
<th>Reported specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIA, Jacalin</td>
<td>Artocarpus integrifolia</td>
<td>Jack fruit lectin</td>
<td>Gal (sialylation tolerant)</td>
</tr>
<tr>
<td>RPbAI</td>
<td>Robinia pseudoacacia</td>
<td>Black locust lectin</td>
<td>Gal, GalNAc</td>
</tr>
<tr>
<td>SNA-II</td>
<td>Sambucus nigra</td>
<td>Sambucus lectin-II</td>
<td>Gal/GalNAc</td>
</tr>
<tr>
<td>SJA</td>
<td>Sophora japonica</td>
<td>Pagoda tree lectin</td>
<td>β-GalNAc</td>
</tr>
<tr>
<td>DBA</td>
<td>Dolichos biflorus</td>
<td>Horse gram lectin</td>
<td>GalNAc</td>
</tr>
<tr>
<td>GHA</td>
<td>Glechoma hederacea</td>
<td>Ground ivy lectin</td>
<td>GalNAc</td>
</tr>
<tr>
<td>SBA</td>
<td>Glycine max</td>
<td>Soy bean lectin</td>
<td>GalNAc</td>
</tr>
<tr>
<td>VVA</td>
<td>Vicia villosa</td>
<td>Hairy vetch lectin</td>
<td>GalNAc</td>
</tr>
<tr>
<td>BPA</td>
<td>Bauhinia purpurea</td>
<td>Camel's foot tree lectin</td>
<td>GalNAc/Gal</td>
</tr>
<tr>
<td>WFA</td>
<td>Wisteria floribunda</td>
<td>Japanese wisteria lectin</td>
<td>GalNAc/sulfated GalNAc</td>
</tr>
<tr>
<td>HPA</td>
<td>Helix pomatia</td>
<td>Garden snail lectin</td>
<td>α-GalNAc</td>
</tr>
<tr>
<td>GSL-I-A4</td>
<td>Griffonia simplicifolia</td>
<td>Griffonia lectin-I A4</td>
<td>GalNAc</td>
</tr>
<tr>
<td>ACA</td>
<td>Amaranthus caudatus</td>
<td>Amaranth</td>
<td>Sialylated/Gal-β-(1,3)-GalNAc</td>
</tr>
<tr>
<td>ABL</td>
<td>Agaricus bisporus</td>
<td>Edible mushroom lectin</td>
<td>Gal-β-(1,3)-GalNAc, GlcNAc</td>
</tr>
<tr>
<td>PNA</td>
<td>Arachis hypogaea</td>
<td>Peanut lectin</td>
<td>Gal-β-(1,3)-GalNAc</td>
</tr>
<tr>
<td>GSL-II</td>
<td>Griffonia simplicifolia</td>
<td>Griffonia/Bandeiraea lectin-II</td>
<td>GlcNAc</td>
</tr>
<tr>
<td>sWGA</td>
<td>Triticum vulgaris</td>
<td>Succinyl WGA</td>
<td>GlcNAc</td>
</tr>
<tr>
<td>DSA</td>
<td>Datura stramonium</td>
<td>Jimson weed lectin</td>
<td>GlcNAc</td>
</tr>
<tr>
<td>STA</td>
<td>Solanum tuberosum</td>
<td>Potato lectin</td>
<td>GlcNAc oligomers</td>
</tr>
<tr>
<td>LEL</td>
<td>Lycopersicum esculentum</td>
<td>Tomato lectin</td>
<td>GlcNAc-β-(1,4)-GlcNAc</td>
</tr>
<tr>
<td>Calsepa</td>
<td>Calystegia sepium</td>
<td>Bindweed lectin</td>
<td>Man/Maltose</td>
</tr>
<tr>
<td>NPA</td>
<td>Narcissus pseudonarcissus</td>
<td>Daffodil lectin</td>
<td>α-(1,6)-Man</td>
</tr>
<tr>
<td>GNA</td>
<td>Galanthus nivalis</td>
<td>Snowdrop lectin</td>
<td>Man-α(1,3)-</td>
</tr>
<tr>
<td>HHA</td>
<td>Hippeastrum hybrid</td>
<td>Amaryllis agglutinin</td>
<td>Man-α(1,3)-Man-α(1,6)-</td>
</tr>
<tr>
<td>ConA</td>
<td>Canavalia ensiformis</td>
<td>Jack bean lectin</td>
<td>Man, Glc, GlcNAc</td>
</tr>
<tr>
<td>Lch-B</td>
<td>Lens culinaris</td>
<td>Lentil isolecitin B</td>
<td>Man, fucose dependent</td>
</tr>
<tr>
<td>Lch-A</td>
<td>Lens culinaris</td>
<td>Lentil isolecitin A</td>
<td>Man, fucose</td>
</tr>
<tr>
<td>Code</td>
<td>Species</td>
<td>Name</td>
<td>Lectin</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>PSA</td>
<td>Pisum sativum</td>
<td>Pea lectin</td>
<td>Man, fucose</td>
</tr>
<tr>
<td>TJA-I</td>
<td>Trichosanthes japonica</td>
<td>TJA One</td>
<td>Sialic acid-α-(2,6)-Gal(NAc)</td>
</tr>
<tr>
<td>WGA</td>
<td>Triticum vulgäris</td>
<td>Wheat germ agglutinin</td>
<td>NeuAc/GlcNAc</td>
</tr>
<tr>
<td>MAA</td>
<td>Maackia amurensis</td>
<td>Maackia agglutinin</td>
<td>Sialic acid-α-(2,3)-Gal(NAc)</td>
</tr>
<tr>
<td>SNA-I</td>
<td>Sambucus nigra</td>
<td>Sambucus lectin-I</td>
<td>Sialic acid-α-(2,6)-Gal(NAc)</td>
</tr>
<tr>
<td>CCA</td>
<td>Cancer antennarius</td>
<td>California crab</td>
<td>9-O-acetyl-sialic acid</td>
</tr>
<tr>
<td>PHA-L</td>
<td>Phaseolus vulgaris</td>
<td>Kidney bean leukoagglutinin</td>
<td>tri-/tetra-antennary β-Gal/Gal-β(1,4)-GlcNAc</td>
</tr>
<tr>
<td>PHA-E</td>
<td>Phaseolus vulgaris</td>
<td>Kidney bean erythroagglutinin</td>
<td>biantennary, bisecting GlcNAc,β-Gal/Gal-β(1,4-)GlcNAc</td>
</tr>
<tr>
<td>RCA-I/120</td>
<td>Ricinus communis</td>
<td>Castor bean lectin I</td>
<td>Gal-β-(1,4)-GlcNAc</td>
</tr>
<tr>
<td>AMA</td>
<td>Arum maculatum</td>
<td>Lords and Ladies agglutinin</td>
<td>Gal-β-(1,4)-GlcNAc</td>
</tr>
<tr>
<td>CPA</td>
<td>Cicer arietinum</td>
<td>Chickpea lectin</td>
<td>Complex glycopeptides</td>
</tr>
<tr>
<td>CAA</td>
<td>Caragana arborescens</td>
<td>Pea tree lectin</td>
<td>Gal-β-(1,4)-GlcNAc</td>
</tr>
<tr>
<td>ECA</td>
<td>Erythrina crista-galli</td>
<td>Cocks comb/coral tree lectin</td>
<td>Gal-β-(1,4)-GlcNAc oligomers</td>
</tr>
<tr>
<td>TJA-II</td>
<td>Trichosanthes japonica</td>
<td>TJA Two</td>
<td>Fuc-α(1,2)Gal(NAc)-β(1,4)</td>
</tr>
<tr>
<td>AAL</td>
<td>Aleuria aurantia</td>
<td>Orange peel fungus lectin</td>
<td>α-(1,6)-linked Fuc</td>
</tr>
<tr>
<td>LTA</td>
<td>Lotus tetragonolobus</td>
<td>Lotus lectin</td>
<td>α-(1,3)-linked Fuc</td>
</tr>
<tr>
<td>UEA-I</td>
<td>Ulex europaeus</td>
<td>Gorse lectin-I</td>
<td>α-(1,2)-linked Fuc</td>
</tr>
<tr>
<td>PA-I</td>
<td>Pseudomonas aeruginosa</td>
<td>Pseudomonas lectin</td>
<td>Gal, Gal derivatives</td>
</tr>
<tr>
<td>EEA</td>
<td>Euonymous europaeus</td>
<td>Spindle tree lectin</td>
<td>α-Gal</td>
</tr>
<tr>
<td>GSL-I-B4</td>
<td>Griffonia simplicifolia</td>
<td>Griffonia/Bandeiraea lectin-I</td>
<td>α-Gal</td>
</tr>
<tr>
<td>MPA</td>
<td>Maclura pomifera</td>
<td>Osage orange lectin</td>
<td>α-Gal</td>
</tr>
<tr>
<td>VRA</td>
<td>Vigna radiata</td>
<td>Mung Bean Lectin</td>
<td>α-Gal</td>
</tr>
<tr>
<td>MOA</td>
<td>Marasmium oreades</td>
<td>Fairy ring mushroom lectin</td>
<td>α-Gal</td>
</tr>
</tbody>
</table>
Appendix ii

Fig i MALDI-MS/MS sequencing of 20 µg full-length heptasaccharide containing scFv generated E. coli containing the native C. jejuni pgI locus. Samples were digested by proteinase K, purified by carbon column, and analysed by a MALDI 4800 TOF in positive reflectron mode using 2,5-dihydroxybenzoic acid (DHB) as a matrix. The identified glycan chain is inset in the upper image with peaks identified with their corresponding labels. The analysis was kindly carried out by the Aebi Microbial Glycobiology Laboratory ETH Zurich.
Appendix iii

Co-authored publication list:


