Book Section:
Strutton, B., Jaffé, S.R.P., Pandhal, J. et al. (2017) Generation of Recombinant N-Linked Glycoproteins in E. coli. In: Burgess-Brown, N.A., (ed.) Heterologous Gene Expression in E.coli. Methods in Molecular Biology, 1586 (1586). Springer New York, pp. 233-250. ISSN: 1064-3745.

https://doi.org/10.1007/978-1-4939-6887-9_15

Generation of Recombinant *N*-linked Glycoproteins in *E. coli*

Benjamin Strutton[1], Stephen R.P Jaffé[1], Jagroop Pandhal[1], Phillip C. Wright[2]

[1]Department of Chemical and Biological Engineering, ChELSI Institute, University of Sheffield, Sheffield, UK

[2]Faculty of Science, Agriculture and Engineering, Newcastle University, Newcastle Upon Tyne, UK

E-mail correspondence: j.pandhal@sheffield.ac.uk

Running head: Generation of Recombinant N-linked Glycoproteins in *E. coli*

Abstract

The production of *N*-linked recombinant glycoproteins is possible in a variety of biotechnology host cells, and more recently in the bacterial workhorse, *Escherichia coli*. This methods chapter will outline the components and procedures needed to produce *N*- linked glycoproteins in *E. coli*, utilizing *Campylobacter jejuni* glycosylation machinery, although other related genes can be used with minimal tweaks to this methodology. To ensure a successful outcome, various methods will be highlighted that can confirm glycoprotein production to a high degree of confidence, including the gold standard of mass spectrometry analysis.

Keywords: Glycosylation, *N*-linked glycoproteins, Posttranslational modifications, *E. coli*, Glycoprotein validation

1 Introduction

The process of glycosylation, a posttranslational modification estimated to be present on half of all human proteins [1. 2], can affect a wide variety of protein properties including its function, localization and half-life [3-5]. Due to these properties, 40 % of the drugs approved in the current recombinant therapeutics market are glycosylated, with this figure predicted to increase [6]. This sizeable market share means that glycoproteins contribute significantly to the biopharma industry, which in 2013 was calculated to be worth $US140 billion [7].

Of the approved glycotherapeutics, approximately 70 % are currently being produced in the eukaryotic cell line, Chinese hamster ovary (CHO) cells [8], which, through the utilization of their inherent glycosylation machinery, can mimic human type glycans, yielding therapeutic proteins with the desired glycosylation profile that will not induce an immunogenic response in humans [9].
In the non-glycosylated therapeutic protein market, the main host cell factory is *E. coli,* which generates around 30 % of the 151 recombinant therapeutics approved by either the EMA or the FDA [10]. With recent advancements in the synthetic biology toolkit for *E. coli,* this host now has the ability to express full length monoclonal antibodies [11, 12], and generate recombinant *O*- and *N*-linked glycoproteins [13-17]. These new capabilities present the opportunity to generate recombinant glycoproteins within *E. coli*.

The initial discovery of bacterial *N*-linked glycosylation in *C. jejuni* and its successful transfer into *E. coli* [16, 18] opened up this area of research. As the field has advanced and accumulated knowledge on the bacterial glycosylation system, variations on the initial system containing the defined pgl pathway allowed the first steps in creating a model cell chassis. This has provided the opportunity for researchers to experiment with different genetic components within the system, replacing them with alternative parts from a variety of organisms, including eukaryotes [14, 19]. This led to the creation of a wide range of different glycan structures that could be attached [14, 16, 19], with a commercially successful example of this coming from the generation of glyco vaccines by GlycoVaxyn [20, 21].

With an ever expanding number of glycan structures still being discovered in the archaeal and prokaryotic domains of life, the potential to transfer them to target proteins for currently unexplored uses is a promising prospect for the field [22, 23].

Although the option to engineer this system is apparent, in order to be able to carry out bacterial *N*-linked glycosylation, there are a few core components required. Firstly, the genes required for sugar biosynthesis within the cytoplasm must be functional [24]. Second of all, the glycosyltransferases which transfer these monosaccharides and sequentially build the glycan on the lipid anchor must be present [25]. An enzyme to flip the anchored glycan across to the periplasm needs to be expressed [26, 27], and finally, an oligosaccharyl transferase (OSTase) that transfers the glycans to the target protein must also be functional within the system [28]. The structure and composition of the glycan, how efficiently it is flipped across the membrane and ultimately transferred to the target protein, are dependent on the genes introduced [14, 29, 30]. Here, we will outline the required components for producing glycoproteins in *E. coli* based upon the original *C. jejuni* system.

1.1 Oligosaccharyl Transferase and the Consensus Sequence that it can recognize within the Target Protein

In eukaryotes, *N*-linked glycosylation occurs at the consensus site of N-X-S/T, where X can be any amino acid except proline. The recognition of this site and subsequent transfer of glycan onto the asparagine residue is dependent on a multi subunit (OSTase) complex with a core functional unit known as STT3 [31]. In bacteria, this OST is a large single protein, with the most commonly utilized transferase being a periplasm located, membrane bound protein called pglB [32-34]. The native form of this protein recognizes a stricter glycosylation sequon with the requirement of a negatively charged amino acid at the -2 position, giving the consensus sequence, D/E-X-N-X-S/T, again with X being any amino acid except proline. When thinking about utilizing this system for glycoprotein production, it is vital that this sequence is present, unless the machinery has been modified to include an OSTase that recognizes the eukaryotic glycosylation sequence [30]. Even so, this may not be sufficient for glycosylation to occur. Within eukaryotes, the process takes place in multiple steps as the protein is

folded through two eukaryotic organelles, the ER and golgi [35, 36], making it a cotranslational process [37]. Within these organelles, the glycan structure can be built up and subsequently trimmed down by processing enzymes before final glycans are added to produce the mature glycan structure [38]. In bacteria, glycosylation typically takes place in the periplasm of the cell [22], and is seen as a single step block transfer of the final glycan to the target protein [39], which occurs posttranslationally. Due to this occurring on a fully folded substrate, it requires the consensus sequence to be situated in a flexible region of the protein that is accessible to the OSTase [39].

1.2 Periplasmic Localization

As mentioned in the previous section, attachment of the glycan onto the protein within bacteria occurs in the periplasm. Therefore, the target protein, glycan and OSTase need to be localized to this compartment of the cell. PglB, as highlighted previously, is a membrane bound protein located in the periplasm, so is already present. As for the target protein, there are multiple pathways available that will direct the target protein to this part of the cell, including the TAT export system [40], SRP pathway [41], and sec transport system [42]. The most utilized methodology is the sec transport system, which requires the addition of a 22 amino acid leader sequence at the N-terminus of the polypeptide chain. This can be engineered through molecular cloning of the expression plasmid, but it is recommended that protein expression and translocation rates be attenuated prior to glycoprotein production. Finally, the glycan of interest must be present in the periplasm. Within the bacterial system, the glycans are sequentially built upon an undecaprenyl-pyrophosphate (UND-pp) lipid anchor by various glycosyltransferases. Depending on the glycan structure, certain sugar biosynthesis genes must also be expressed that utilize molecules from the central carbon metabolism, modifying them to generate any unnatural glycan precursors such as bacillosamine, as well as the common monosaccharides like UDP-GlcNAc [43]. Once these have been generated, the glycosyltransferases transfer them to the glycan being built upon the lipid anchor. Once the glycans are attached, the UND-pp linked glycan molecule can be flipped across the inner membrane by a flippase, such as the *C. jejuni* pglK [34], or the native *E. coli* protein wzX [44] so the glycan now faces the periplasm [19]. With wzX and waaL, *E. coli* has a native system that utilizes periplasmic UND-pp linked glycans for

attachment to a lipid A core, for subsequent presentation on the cell surface [15]. It is therefore recommended that for recombinant *N*-linked glycoprotein expression, the competing protein that transfers the glycans to the lipid A core, waaL, is removed as seen in the W3110 mutant, CLM24 [15]. The presence of waaL may be utilized to detect the presentation of glycans on the surface as a means of checking whether they are being expressed (or not) [32, 45].

1.3 Conclusion of Requirements

The target recombinant protein must contain an asparagine residue situated in a consensus sequence, located in a flexible region of the protein, so that the OSTase of choice is able to recognize and bind. Once expressed, it must also be directed to the periplasm for this type of glycosylation to occur, which can be achieved by utilizing a number of export systems. Alongside expression of the target protein, glycosylation machinery must be incorporated into the cell, containing any genes needed for specific sugar biosynthesis, the required glycosyltransferases to build the glycan, a flippase that can recognize the glycan as a substrate, and an OSTase that will also recognize the glycan. This should all preferably take place in a bacterial strain where the waaL pathway is inhibited, unless glycan production is being checked by cell surface display.

1.4 General Analysis

In order to determine if the target protein has been successfully glycosylated, it is necessary to perform a periplasmic extraction, to release proteins that have the potential to be glycosylated. In order to determine whether *N*-glycosylation of the target protein is successful, a number of methodologies can be applied with varying degrees of speed and accuracy. These include western blots, where a mass shift for the addition of a glycan is observed [32], lectin peroxidase screen whereby a lectin that binds to a target glycan is bound [32] or the gold standard is the use of mass spectrometry, which can provide both protein sequence and glycan information [33]. By following the methodology and workflow stated (see **Figure 1**), the user should be able to generate *N*-linked glycoproteins in *E. coli* and validate this production using a variety of techniques. For more details on bacterial glycosylation please refer to this extensive review paper [46].

2 Materials

All solutions should be prepared using either nuclease-free water or HPLC grade water along with analytical grade reagents. All solutions should be made up at room temperature unless otherwise stated. All examples and methodologies listed are for the generation of IFNα2b with an *N*-linked glycan structure of $GalNAc_5GlcNAC$ within *E. coli*, utilizing the plasmids pJExpress*IFNa2b* and pACYC*pgl*2.

2.1 General and Experiment Specific Reagents, Strains and Plasmids

2.1.1 Antibiotic Stock Solutions

1. 50 mg/mL Kanamycin, filter sterilized.
2. 30 mg/mL Chloramphenicol in 100 % ethanol.

2.1.2 Bacterial Strains

1. *E. coli* W3110.
2. *E. coli* CLM24.

2.1.3 Glycosylation Machinery

1. pACYC*pgl*2.

2.1.4 Target Protein

1. pJexpress*IFN*α*2b*.

2.2 Cell Surface Expression of Glycans

1. Luria Broth (10 g/L tryptone, 10 g/L NaCl 5 g/L yeast extract).
2. Petri dishes.
3. Protran Nitrocellulose membranes (Fisher Scientific).
4. PBS containing 2 % (v/v) TWEEN® 20.

5. PBS with 0.05 % (v/v) TWEEN® 20, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 1 mM $MgCl_2$ and 3 µg of soybean agglutinin lectin peroxidase.
6. Immobilon™ chemiluminescent HRP (Fisher Scientific).
7. ImageQuant™ RT ECL (GE Healthcare) fitted with temperature cooled 16-bit CCD Camera.
8. 1-Step TMB Ultra blotting solution (Thermo Fisher Scientific).

2.3 Target Protein Expression and Localization

1. Luria Broth: 10 g/L Tryptone, 10 g/L NaCl 5 g/L Yeast extract.
2. 1 mM IPTG.

2.4 Periplasmic Protein Extraction and Quantification

2.4.1 Extraction

1. Periplasmic lysis buffer composed of 20 % (w/v) sucrose, 30 mM Tris-HCl, pH 8.0, and 1 x HALT (Pierce) in HPLC grade water.
2. Centrifuge capable of spinning 50 mL falcon tubes at 3,000 x *g,* 4°C.

2.4.2 Quantification

1. Bradford reagent (Sigma Aldrich).
2. RC DC protein assay kit II (Bio-Rad).

2.5 SDS-PAGE Gels and Analysis

2.5.1 SDS-PAGE

1. NuPAGE Novex 4-12 % Bis-Tris protein gels, 1.0 mm, 12 well (Life Technologies).
2. 20 x Novex MOPS SDS running buffer (Life Technologies).
3. Novex Sharp Pre-stained protein standard (Life Technologies).
4. Hoefer SE300 miniVE integrated vertical electrophoresis unit (Hoefer Inc).
5. Centrifuge capable of spinning 1.5 mL microcentrifuge tubes at 13, 000 x *g*.

2.5.2 Western Blotting

1. iBlot™ Gel Transfer Device (Thermo Fisher Scientific).
2. iBlot™ Gel Transfer Stacks (Thermo Fisher Scientific).
3. Blocking Buffer consisting of: 5 % (w/v) blocking powder (Bio-Rad) in TBS 0.05 % (v/v) Tween® 20.
4. 6 x Histidine HRP linked antibody (Abcam).
5. 1-Step TMB Ultra blotting solution (Thermo Fisher scientific).

2.5.3 Lectin Screen

1. iBlot™ Gel Transfer Device (Thermo Fisher Scientific).
2. iBlot™ Gel Transfer Stacks (Thermo Fisher Scientific).
3. Protran Nitrocellulose membranes from (Thermo Fisher Scientific).
4. PBS containing 2 % (v/v) TWEEN® 20 (Sigma Aldrich).
5. PBS with 0.05 % (v/v) TWEEN® 20, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 1 mM $MgCl_2$ and 3 µg of soybean agglutinin lectin peroxidase (Sigma Aldrich).
6. Immobilon™ chemiluminescent HRP (Fisher Scientific).
7. ImageQuant™ RT ECL (GE Healthcare) fitted with temperature cooled 16-bit CCD Camera.
8. 1-Step TMB Ultra blotting solution (Thermo Fisher scientific).

2.6 Mass Spectrometry Analysis

2.6.1 In-Solution Digest

1. 100 mM ammonium bicarbonate made in HPLC grade water.
2. Protein Lobind tubes (Eppendorf).
3. 50 mM DTT stock solution.
4. 100 mM Iodoacetamide stock solution.
5. Trypsin solution made using lyophilised trypsin protease, MS grade (Pierce).
6. 94.5 % (v/v) HPLC grade H_2O, 5 % (v/v) acetonitrile, 0.5 % (v/v) TFA.
7. Water bath

8. Benchtop microcentrifuge for 1.5 mL microcentrifuge tubes

2.6.2 C18 Clean Up

1. Pierce™ C18 Spin Columns (Thermo Fisher Scientific).
2. Protein Lobind tubes (Eppendorf).
3. Centrifuge capable of spinning down 1.5 mL microcentrifuge tubes at 13, 000 x *g*.
4. Vacuum concentrator centrifuge (Eppendorf)

2.6.3 LC-MS

1. Sonicating water bath
2. Vortex mixer
3. Centrifuge capable of spinning down 1.5 mL microcentrifuge tubes at 13, 000 x *g*.
4. Automated LC vials.
5. Automated LC vial caps.
6. maXis Q-ToF mass spectrometer (Bruker Daltonics).
7. Ultimate 3000 HPLC system (Dionex).
8. HPLC buffer A consisting of: 96.9% HPLC grade water, 3% Acetonitrile, 0.1% Formic acid
9. HPLC buffer B consisting of: 3% HPLC grade water, 96.9% Acetonitrile, 0.1% Formic acid

3 Methods

All examples and methodologies listed are for the generation of IFNα2b with an *N*-linked glycan structure of $GalNAc_5GlcNAC$ within *E. coli*, utilizing the plasmids pJExpress*IFNa2b* and pACYC*pgl*2.

3.1 Cell Surface Expression of Glycans

If checking glycan production, use bacterial strain W3110. If checking waaL deletion, use your modified strain or in our case CLM24 (see **Figures 2 and 3**).

1. Transform the desired *E. coli* strain with pACYC*pgl*2.

2. Inoculate 1 mL of LB, with the appropriate antibiotics, in a sterile 1.5 mL centrifuge tube with a colony of the bacterial strain containing pACYC*pgl*2.
3. Incubate the tube at 37 °C for 16 hours in a shaker at 180 rpm.
4. Measure the O.D. 600 and normalize the O.D. down to 0.6 using sterile LB.
5. Dilute the cells by a factor of 1 in 75,000 to a final volume of 1 mL.
6. Take 100 µL of the diluted culture and streak out onto an LB agar plate with correct antibiotics and leave to incubate for 16 hours at 37 °C (*see* **Notes 1 and 2**).
7. Soak a piece of nitrocellulose paper, cut to fit inside a petri dish, for 5 mins in the required antibiotics for the maintenance of the plasmid in the strain.
8. Place the nitrocellulose paper in a flow hood for 5 mins before placing the paper over the bacterial colonies using a clean pair of laboratory tweezers, and incubating for 3 hours at 37 °C (*see* **Note 3**).
9. Remove the nitrocellulose paper from the agar plate and block the membrane in PBS containing 2 % (v/v) TWEEN® 20 for 2 mins at 20 °C.
10. Wash the membrane twice (10 mins each) in PBS.
11. Incubate the membrane in PBS with 0.05 % (v/v) TWEEN® 20, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 1 mM $MgCl_2$ and 3 µg of soybean agglutinin lectin peroxidase (specific for GalNAc) for 16 hours at 20 °C (*see* **Note 4**).
12. Wash the membrane for 2 x 10 mins in PBS.
13. Detect colonies using Immobilon™ chemiluminescent HRP substrate with ImageQuant™ RT ECL (GE Healthcare), fitted with a cooled 16-bit CCD camera.
14. If this type of camera is not available, the screen can be developed using approximately 10 mL TMB-Ultra Blotting solution and leaving the membrane to develop for between 5-30 mins depending on the desired intensity and the level of background development.
15. Wash with 2 x 5 mins washes with HPLC grade water to stop development of the membrane (*see* **Note 5**).

3.2 Target Protein Expression and Localization

3.2.1 Bacterial Growth and Protein Expression

Required to check protein of interest is being expressed and translocated to the periplasm (see **Figure 4**).

1. Inoculate 10 mL of LB containing the appropriate antibiotics with a colony of CLM24 pJexpress*IFNa2b* and grow for 16 hours at 37 °C with shaking at 180 rpm.
2. Using the starter culture, inoculate 100 mL of LB containing the appropriate antibiotics and leave to grow at 37 °C with shaking at 180 rpm.
3. When the O.D. 600 of the culture reaches 0.5, induce target protein expression with 1 mM IPTG.
4. Leave the bacteria to express the protein of interest for 4 hours at 30 °C with shaking at 180 rpm.
5. Measure the final O.D. 600 of the culture, collecting 40 O.D.'s worth of the culture by spinning at 3,000 x *g* for 10 mins at 4 °C.
6. Discard supernatant and continue to periplasmic protein extraction (*see* **Note 6**).

3.2.2 Periplasmic Protein Extraction

1. Gently resuspend the bacterial pellet in 1 mL of ice cold periplasmic lysis buffer, keeping the suspension in a 50 mL Falcon tube.
2. Leave the pellet to gently roll on ice for 1.5 – 2 hours.
3. Spin down the cell debris at 3,000 x *g* for 10 mins at 4 °C, collecting the supernatant as the periplasmic extract (*see* **Note 7**).
4. Pellet can be retained and the rest of the soluble fraction extracted to analyze the quantity of the target protein that has not been exported to the periplasm but has still been expressed.

3.2.3 Quantification of the Periplasmic Extract

Many common laboratory methods can be used here including a Bradford assay, RC/DC assay or nanodrop. Due to the frequent nature of these protocols, we advise that the researcher follows the detailed protocols that are widely available for these quantification techniques.

3.2.4 SDS-PAGE

1. Pre-cast protein gels were used to run the gel based analysis.
2. In a single LoBind 1.5 mL centrifuge tube, add the required volume of the periplasmic extract to place 5 µg of protein into the tube. Along with the sample add, 5 µL of 4 x LDS sample buffer, as well as 2 µL of 10 x reducing agent. Using distilled water, make the total volume up to 24 µL.
3. Aliquot 10 µL of pre-stained protein standard into a fresh LoBind tube.
4. Boil the samples and protein ladder for 10 mins at 70 °C.
5. Leave the samples to cool before spinning all tubes down at 13,000 x *g* for 1 min.
6. Load the gel into the gel tank and fill the internal chamber with fresh 1 x MOPS SDS running buffer (*see* **Note 8**). Fill the rest of the tank with 1 x MOPS SDS running buffer. Remove the plastic comb from the top of the gel and discard (*see* **Note 9**).
7. Load protein ladder and samples onto the protein gel (*see* **Note 10**).
8. Place the lid on top of the gel tank and run the gel at a constant 200 V for 50 mins (*see* **Note 11**).
9. Following the running of the gel, remove the gel tank lid, remove the plastic encased gel and carefully open along the seam at the edge of the casing. Remove the gel and place into a clean plastic container (*see* **Note 12**). Discard the used plastic casing.

3.2.5 Western Blotting

1. Western blotting will be conducted using the Invitrogen iBlot system and compatible transfer stacks.
2. Briefly wash the acrylamide gel with 2 x MilliQ grade water to remove any excess SDS
3. Following the manufacturer's protocol place the protein gel onto the membrane and complete the stack setup (*see* **Note 13**).
4. Fasten lid and transfer the proteins onto the nitrocellulose membrane using the P3 transfer program for a run time of 7 mins.

5. Once run, discard all of the excess stack, including the gel, and using tweezers, carefully remove the membrane and submerge it in sufficient blocking buffer and leave rocking in this solution for 1 hour at room temperature (*see* **Note 14**).
6. Wash the membrane with 3 x washes in TBS-T (0.05 % (v/v) Tween 20) for 10 mins each at room temperature, discarding the wash solution after each repeat.
7. During the second wash, run a 1 in 10,000 dilution of the 6 x Histidine residue antibody in 20 mL blocking buffer and place on the rocker during the rest of the wash steps.
8. Remove the final wash solution and cover the membrane in the antibody solution.
9. Leave this to incubate at 4 °C for 16 hours, preferably with gentle shaking.
10. Discard the antibody solution and conduct 5 x 5 mins wash steps with TBS-T (0.05 % (v/v) Tween 20).
11. Post washing, cover the membrane with approximately 10 mL of TMB-Ultra Blotting solution, leaving the membrane to develop for between 5-30 mins, depending on the desired intensity and the level of background development.
12. Wash with 2 x 5 mins washes with HPLC grade water to stop development of the membrane.

3.3 Expression of Glycoprotein

See **section 3.2.1** as the protocol for this is the same needed for glycoprotein expression except the starting bacterial strain must contain the plasmids required for both the target protein expression and the glycosylation machinery. Due to the presence of two plasmids, an extra antibiotic will be required for plasmid maintenance. If a mass spectrometry approach is going to be utilized for glycoprotein validation, it is important that when extracting the periplasmic proteins (**section 3.2.2**) no protease inhibitor cocktail is present in the lysis buffer mentioned earlier and that the sample is kept cold at all times (*see* **Note 15**).

3.4 Validation of Glycoprotein Production using Gel based methods

3.4.1 Western Blot Analysis

Expecting to see multiple bands compared to the control due to the mass shift that occurs with the attached glycan (see **Figures 5 and 6**). Run subsequent quantification steps (**section 3.2.3**), SDS-PAGE (**section 3.2.4**) and Western blot (**section 3.2.5**) analysis as outlined above as well as a lectin screen specific to the sugars incorporated in the glycan.

3.4.2 Lectin Screen

Sugar specific lectin should bind to the glycans if present.

1. Follow the same protocol as the western blotting up to step 5 with dismemberment of the blotting stack but instead of placing the membrane in the western blocking buffer, the membrane must be washed with PBS containing 2 % (v/v) Tween 20 for 2 mins at room temperature.
2. Post blocking, wash the membrane twice with PBS for 10 mins at room temperature.
3. Incubate the membrane for 16 hours at room temperature on a shaker in PBS with 0.05 % (v/v) TWEEN® 20, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 1 mM $MgCl_2$ and 3 µg of soybean agglutinin lectin peroxidase (specific for GalNAc).
4. Repeat the wash steps conducted in **step 2** before covering the membrane with approximately 10 mL of TMB-Ultra Blotting solution, leaving the membrane to develop for between 5-30 mins depending on the desired intensity and the level of background development.
5. Wash with 2 x 5 mins washes with HPLC grade water to stop development of the membrane.

3.5 Validation of Glycoprotein Production using Mass Spectrometry Analysis

Expecting to see the presence of diagnostic ions correlating to the sugars involved in the glycan (see **Figures 5 and 7**).

3.5.1 In-Solution Digest

1. In a single LoBind 1.5 mL centrifuge tube, add the required volume of the periplasmic extract to place 5 µg of protein into the tube.
2. Make up the total volume to 40 µL using 100 mM ammonium bicarbonate.

3. Add a sufficient quantity of 50 mM DTT to achieve a final DTT concentration of 4 mM in the final solution.
4. Incubate for 60 mins in a 56 °C water bath.
5. Gently spin down at 1,000 x *g* for 1 min
6. Add a sufficient quantity of 100 mM Iodoacetamide to achieve a final Iodoacetamide concentration of 8 mM in the final solution.
7. Incubate in the dark at room temperature for 30 mins.
8. Gently spin down at 1,000 x *g* for 1 min
9. Add the protease at a ratio of 1:25 (protease:protein)
10. Leave digest for 18 hours at 37 °C
11. Dry down the samples in a vacuum concentrator centrifuge (*see* **Note 16**).
12. Resuspend samples in 94.5 % (v/v) HPLC grade water, 5 % (v/v) ACN and 0.5 % (v/v) TFA

3.5.2 C18 Clean Up

Prior to mass spectrometry analysis, the sample should undergo a C18 clean up procedure. Due to the frequency of this technique and the wide range of columns available, it is recommended that the researcher follows the protocol that is supplied by the manufacturer. A recommended column is given in **section 2.6.2**.

3.5.3 LC-MS

1. Add 10 - 20 µL of 94.5 % (v/v) HPLC grade water, 5 % (v/v) ACN and 0.5 % (v/v) TFA to the dried peptides post C18 cleanup.
2. Sonicate and vortex the samples for a minute each and then centrifuge the tubes at 13,000 x *g* for 30 secs to pool the liquid at the bottom of the LoBind centrifuge tube.
3. Transfer liquid into a vial compatible with the automated LC system (*see* **note 17**).

Using a maXis Q-TOF mass spectrometer (Bruker), perform high selectivity pseudo-selective reaction monitoring (pSRM) of the targeted glycopeptides.

3.5.4 Pseudo-Selective Reaction Monitoring (pSRM)

1. Select the correct *m/z* values for the desired glycopeptides (*see* **Note 18)** and operate the Q-TOF in MRM mode with an *m/z* window of 2 *m/z*.
2. Run the mass spectrometer in positive mode with an *m/z* window of 3.
3. Implement a rolling collision energy of 100-150 % with an MRM scan mode.
4. Start online HPLC with 5 % of HPLC buffer B and 95 % of HPLC buffer A and run for 5 min.
5. Run a 30 min gradient program, increasing the concentration of HPLC buffer B up to 90 %.
6. Once the gradient is at 90 % of HPLC buffer B, maintain that level for 10 mins.
7. End the program by returning the HPLC buffer B concentration to 5 %.

3.5.5 Data Analysis

1. Select pSRM scans obtained from the analysis in subheading (section **3.5.4**) and analyze using DataAnalysis v.4.1.
2. Identify diagnostic ions present for the sugars generated in the expected glycan structure (204.08 for HexNAc, 366.14 for HexHexNAc) in MRM scans that add up to the expected *m/z* value for the glycopeptide.

4 Notes

1. This dilution factor should be significant enough to achieve single, non-touching colonies on the plate. If this is not the case, alterations may need to be made.
2. If very small colonies are seen, leave them to grow for another few hours to become more established.
3. Make sure tweezers are ethanol wiped before use and try to remove any air bubbles that can interfere with transfer.
4. Solution containing soybean agglutinin lectin peroxidase can be cloudy when all components are mixed together.

5. Try to take an image of the developed gel as soon as possible after development, if left the blot may dry out or continue to develop if not washed properly.
6. Try to remove as much residual LB as possible.
7. Supernatant may need to be centrifuged a second time to ensure no cell debris is present in the periplasmic extract as it can late interfere with quantification.
8. Ensure no leaks are present when filling the internal volume of the SDS-PAGE gel. This is an indication that the gel isn't sealed in properly.
9. Carefully remove the plastic comb to expose the wells in the SDS-PAGE gel. Wells are very delicate and easily damaged.
10. Protein loading tips (Bio-Rad) are recommended for loading SDS-PAGE gels.
11. If the target protein has a large MW protein, the gel can be run for longer to separate out bands of interest from one another more distinctly. Pgl2 glycans ($GalNAc_5GlcNAC$) are roughly 1.2 kDa.
12. Handle SDS-PAGE gel carefully as it is extremely delicate and partial to breaking (ripping). Wetting heavily with distilled water is advised to prevent this occurring.
13. Ensure that when building up the stacks there are no air bubbles between layers as they can interfere with the transfer.
14. Coloured protein ladder should now be visible on the membrane and no longer present in the acrylamide gel.
15. No protease inhibitor cocktail should be used as it may interfere with the protease used for protein digestion prior to mass spectrometry analysis.
16. At this point, the dried peptides can be stored at -20 °C and be analyzed at a later date.
17. Ensure there are no bubbles present in the LC vials once the sample has been transferred into it.
18. Be very careful when calculating MRM values. Make use of an in silico tool to calculate the m/z values for different peptides, taking into account the protease used, chemical modifications and the glycan structure used.

Acknowledgements

The authors acknowledge and are deeply thankful for funding from the UK's Biotechnology and Biological Sciences Research Council (BBSRC) through the Bioprocess Research Industry Club (BRIC) programme (BB/K011200/1 and BB/F004842/1).

References

1. Apweiler R, Hermjakob H , Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochimica et Biophysica Acta (BBA)-General Subjects 1473: 4-8
2. Rich JR , Withers SG (2008) Emerging methods for the production of homogeneous human glycoproteins. Nature Chemical Biology 5: 206-215
3. Elbein AD (1991) The role of *N*-linked oligosaccharides in glycoprotein function. Trends in biotechnology 9: 346-352.
4. Lusch A, Kaup M, Marx U, et al (2013) Development and analysis of alpha 1-antitrypsin neoglycoproteins: the impact of additional *N*-glycosylation sites on serum half-life. Molecular pharmaceutics 10: 2616-2629
5. Moore KL, Stults NL, Diaz S, et al (1992) Identification of a specific glycoprotein ligand for P-selectin (CD62) on myeloid cells. The Journal of cell biology 118: 445-456
6. Walsh G (2010) Biopharmaceutical benchmarks 2010. Nature biotechnology 28: 917-924
7. Walsh G (2014) Biopharmaceutical benchmarks 2014. Nature biotechnology 32: 992-1000
8. Higgins E (2010) Carbohydrate analysis throughout the development of a protein therapeutic. Glycoconjugate journal 27: 211-225

9. Yoo EM, Yu LJ, Wims LA, et al (2010) Differences in *N*-glycan structures found on recombinant IgA1 and IgA2 produced in murine myeloma and CHO cell lines. MAbs 2: 320-334

10. Ferrer-Miralles N, Domingo-Espín J, Corchero JL, et al (2009) Microbial factories for recombinant pharmaceuticals. Microbial cell factories 8: 1-8

11. Mazor Y, Van Blarcom T, Mabry R, et al (2007) Isolation of engineered, full-length antibodies from libraries expressed in *Escherichia coli*. Nature biotechnology 25: 563-565

12. Chan CEZ, Lim APC, Chan AHY, et al (2010) Optimized expression of full-length IgG1 antibody in a common *E. coli* strain. PLoS One 5: e10261 1-8

13. Faridmoayer A, Fentabil MA, Mills DC, et al (2007) Functional characterization of bacterial oligosaccharyltransferases involved in *O*-linked protein glycosylation. Journal of bacteriology 189: 8088-8098

14. Srichaisupakit A, Ohashi T, Misaki R, et al (2015) Production of initial-stage eukaryotic *N*-glycan and its protein glycosylation in *Escherichia coli*. Journal of bioscience and bioengineering 119: 399-405

15. Feldman MF, Wacker M, Hernandez M, et al (2005) Engineering *N*-linked protein glycosylation with diverse *O*-antigen lipopolysaccharide structures in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America 102: 3016-3021

16. Wacker M, Linton D, Hitchen PG, et al (2002) *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science 298: 1790-1793

17. Strutton B, Pandhal J , Wright PC (2013) Expanding the bioprocessing toolbox of *Escherichia coli* through metabolic engineering and synthetic biology: an emerging glycosylation chassis. Pharmaceutical Bioprocessing 1: 221-224

18. Szymanski CM, Yao R, Ewing CP, et al (1999) Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Molecular microbiology 32: 1022-1030

19. Schwarz F, Huang W, Li C, et al (2010) A combined method for producing homogeneous glycoproteins with eukaryotic *N*-glycosylation. Nature Chemical Biology 6: 264-266

20. Wacker M, Wang L, Kowarik M, et al (2013) Prevention of *Staphylococcus aureus* infections by glycoprotein vaccines synthesized in *Escherichia coli*. Journal of Infectious Diseases: 1551-1561

21. Hatz CF, Bally B, Rohrer S, et al (2015) Safety and immunogenicity of a candidate bioconjugate vaccine against *Shigella dysenteriae* type 1 administered to healthy adults: A single blind, partially randomized Phase I study. Vaccine 33: 4594-4601

22. Nothaft H , Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. Nature Reviews Microbiology 8: 765-778

23. Nothaft H , Szymanski CM (2013) Bacterial Protein *N*-glycosylation: New Perspectives and Applications. Journal of Biological Chemistry 288: 6912-6920

24. Guerry P, Ewing CP, Schoenhofen IC, et al (2007) Protein glycosylation in *Campylobacter jejuni*: partial suppression of pglF by mutation of pseC. Journal of bacteriology 189: 6731-6733

25. Glover KJ, Weerapana E , Imperiali B (2005) In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic *N*-linked glycosylation. Proceedings of the National Academy of Sciences of the United States of America 102: 14255-14259

26. Lehrman MA (2015) Flipping a Lipid-Linked Oligosaccharide? You Must Whip It! Trends in biochemical sciences 40: 715-717

27. Alaimo C, Catrein I, Morf L, et al (2006) Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. The EMBO journal 25: 967-976

28. Ihssen J, Haas J, Kowarik M, et al (2015) Increased efficiency of *Campylobacter jejuni N*-oligosaccharyltransferase PglB by structure-guided engineering. Open biology 5: 1-10

29. Ihssen J, Kowarik M, Dilettoso S, et al (2010) Production of glycoprotein vaccines in *Escherichia coli*. Microbial cell factories 9: 1-13

30. Ollis AA, Zhang S, Fisher AC, et al (2014) Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. Nature Chemical Biology 10: 816-822

31. Kelleher DJ , Gilmore R (2006) An evolving view of the eukaryotic oligosaccharyltransferase. Glycobiology 16: 47R-62R

32. Pandhal J, Woodruff L, Jaffe S, et al (2013) Inverse metabolic engineering to improve *Escherichia coli* as an N-glycosylation host. Biotechnology and bioengineering 110: 2482-2493

33. Pandhal J, Ow SY, Noirel J, et al (2011) Improving *N*-glycosylation efficiency in *Escherichia coli* using shotgun proteomics, metabolic network analysis, and selective reaction monitoring. Biotechnology and bioengineering 108: 902-912

34. Pandhal J, Desai P, Walpole C, et al (2012) Systematic metabolic engineering for improvement of glycosylation efficiency in *Escherichia coli*. Biochemical and Biophysical Research Communications 419: 472-476

35. Aebi M (2013) *N*-linked protein glycosylation in the ER. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 1833: 2430-2437

36. D'Angelo G, Uemura T, Chuang C-C, et al (2013) Vesicular and non-vesicular transport feed distinct glycosylation pathways in the Golgi. Nature 501: 116-120

37. Ruiz-Canada C, Kelleher DJ , Gilmore R (2009) Cotranslational and posttranslational *N*-glycosylation of polypeptides by distinct mammalian OST isoforms. Cell 136: 272-283.

38. Bieberich E (2014) Synthesis, processing, and function of *N*-glycans in *N*-glycoproteins. Glycobiology of the Nervous System. Springer 47-70

39. Kowarik M, Numao S, Feldman MF, et al (2006) *N*-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. Science 314: 1148-1150

40. Alanen HI, Walker KL, Suberbie MLV, et al (2015) Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the *Escherichia coli* Tat pathway in the absence of prior disulfide bond formation. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 1853: 756-763

41. Lee YJ , Jeong KJ (2013) Enhanced production of antibody fragment via SRP pathway engineering in *Escherichia coli*. Biotechnology and bioprocess engineering 18: 751-758.

42. Mergulhao F, Summers D , Monteiro G (2005) Recombinant protein secretion in *Escherichia coli*. Biotechnology advances 23: 177-202

43. Samuel G , Reeves P (2003) Biosynthesis of *O*-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and *O*-antigen assembly. Carbohydrate Research 338: 2503-2519

44. Valderrama-Rincon JD, Fisher AC, Merritt JH, et al (2012) An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. Nature Chemical Biology 8: 434-436

45. Jaffé SR, Strutton B, Pandhal J, et al (2015) Inverse Metabolic Engineering for Enhanced Glycoprotein Production in *Escherichia coli*. Glyco-Engineering: Methods and Protocols. 17-35

46. Jaffe SR, Strutton B, Levarski Z, et al (2014) *Escherichia coli* as a glycoprotein production host: recent developments and challenges. Current opinion in biotechnology 30: 205-210

Figure Captions

Figure 1. A workflow showing all the key points required in designing, generating and validating the expression of recombinant *N*-linked glycoproteins in *E. coli*.

Figure 2. Diagrammatic workflow outlining the strains for the glycan production screen.

Figure 3. Cell surface expression of glycans and confirmation of the production of a *waaL* knockout strain. (**A**) Cell surface screening with a functional *waaL* pathway showing that functional glycan machinery is being expressed. (**B**) Cell surface screening to show that glycans produced are not being exported, confirming the inhibition of the *waaL* pathway.

Figure 4. Diagrammatic workflow outlining the process to check the target protein expression and localisation.

Figure 5. Diagrammatic workflow outlining the process of glycoprotein production and validation.

Figure 6. Validation of glycoprotein production using Western Blotting. (**A**) Expression of the target protein without the presence of the glycosylation machinery. (**B**) Expression of the target protein with the glycosylation machinery producing the multiple glycoforms.

Figure 7. Mass spectrometry validation of glycoprotein production. Spectra of a tryptically digested Interferon α2b glycopeptide with the diagnostic oxonium ions of 204 and 366 highlighted.