



-Linked glycoengineering for human therapeutic proteins in bacteria

Jagroop Pandhal, Phillip C. Wright

► To cite this version:

Jagroop Pandhal, Phillip C. Wright. -Linked glycoengineering for human therapeutic proteins in bacteria. *Biotechnology Letters*, 2010, 32 (9), pp.1189-1198. 10.1007/s10529-010-0289-6 . hal-00591183

HAL Id: hal-00591183

<https://hal.science/hal-00591183>

Submitted on 7 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N-Linked glycoengineering for human therapeutic proteins in bacteria

Jagroop Pandhal, Phillip C. Wright*

ChELSI Institute, Biological and Environmental Systems Group, Department of Chemical and Process Engineering, University of Sheffield, Sheffield, UK

* Corresponding author: Department of Chemical and Process Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK; Tel No. +44 (114) 2227577; fax +44(0)114 2227501; email: p.c.wright@sheffield.ac.uk

Abstract

Approx. 70% of human therapeutic proteins are *N*-linked glycoproteins, and therefore host cells for production must contain the relevant protein modification machinery. The discovery and characterisation of the *N*-linked glycosylation pathway in the pathogenic bacterium *Campylobacter jejuni*, and subsequently its functional transfer to *Escherichia coli*, presents the opportunity of using prokaryotes as cell factories for therapeutic protein production. Not only could bacteria reduce costs and increase yields, but the improved feasibility to genetically control microorganisms means new and improved pharmacokinetics of therapeutics is an exciting possibility. This is a relatively new concept, and progress in bacterial *N*-glycosylation characterisation is reviewed and metabolic engineering targets revealed.

Key words

N-Linked glycosylation, therapeutic proteins, *Campylobacter jejuni*, *E.coli*, metabolic engineering

Introduction

Therapeutic proteins for human use have been traditionally prepared from human or even animal sources with the well-known example, insulin, being derived from the pancreas. With the rapid increase in demand for therapeutics combined with the development in genetic engineering tools, recombinant expression systems are

by far the most favourable method of production. Bacteria, yeast, plant, insect and mammalian cell lines are the most common hosts, and human insulin produced in the well-studied Gram-negative bacterium, *Escherichia coli*, was brought into the market in 1982 (Johnson 1983). However, the structure of insulin consists of only two polypeptide chains linked by two inter-chains and one intra-chain disulphide bridge. Essentially, post-translational modifications are absent, and this contrasts significantly with the majority of therapeutic proteins. The most common, complex and energy demanding of these modifications is glycosylation, and approx. 70% of therapeutic proteins in the three phases of drug development are glycosylated (Sethuraman and Stadheim 2006). Because these modifications are essential for protein functionality, host cells must be chosen on their ability to perform them. For this reason, only eukaryotic cell lines can be used to produce N-type glycosylated therapeutic proteins, as they have the correct molecular machinery capable of including this structural variation, or so we thought.

Glycosylation is the covalent addition of oligosaccharides (sugars) to a polypeptide backbone. This carbohydrate moiety, also referred to as a glycan, can make up on average 20% of the total protein molecular weight, but can be upwards of 90%. The specific addition of these glycans affects protein structure, biological activity, plasma half-life and tissue targeting (Varki 1993). Several elaborate glycosylation types have been identified that result in glycoproteins being secreted out of cells, incorporated into membranes or the cytoplasm and nucleus. There are five different classes of glycosylation (see below) but this does not mean one type is restricted to one protein, for example, some immunoglobulins contain two types (Fanger and Smyth 1972).

- i) *N*-linked: Glycans attach to nitrogen of asparagine or arginine side-chain
- ii) *O*-linked: Glycans attach to hydroxyl group of serine, threonine, tyrosine, hydroxylysine, or hydroxyproline side-chains
- iii) *P*-linked: phospho-glycans linked through the phosphate of a phosphoserine
- iv) *C*-linked: Glycans attach to a carbon on a tryptophan side chain
- v) *G*-linked: Components of glycosphosphatidylinositol anchor.

For this review, we concentrate on the *N*-Linked glycosylation, as this is the most widely distributed glycan-peptide bond found in therapeutic proteins. *N*-Linked glycosylation occurs when the membrane protein complex called the oligosaccharyltransferase, recognises the Asn-Xaa-Ser/Thr (A-X-S/T) sequon (Lehle and Tanner 1978), although only approximately 66% of sequons are glycosylated due to further structural requirements

(Apweiler *et al.* 1999). Most significantly, from a therapeutic point of view, the exact structure, type and location of the glycan groups (called macro and microheterogeneity), affect both the function and efficacy of the proteins. The incorrect presence or composition of glycans can affect the pharmacokinetic properties and lead to immunogenic responses. Using host systems which cannot correctly glycosylate proteins would ultimately lead to rapid clearance of the drug. Furthermore, it has been reported that as much as 80% of erythropoietin produced in Chinese hamster ovary (CHO) cells is wasted due to incomplete glycosylation (Jacobs and Callewaert 2009).

The requirement for glycoproteins to be of ‘human-type’ leads to the question- what exactly do human glycoproteins look like? It helps to answer this question by looking at the process in a step-wise manner (Varki *et al.* 1999):

- i) On the cytoplasmic side of the endoplasmic reticulum (ER), *N*-acetylglucosamine phosphate transferase adds a GlcNAc group to lipid carrier dolichol-PP.
- ii) This lipid-linked oligosaccharide (LLO) is transferred to the luminal side of the ER through a membrane transporter (RFT1).
- iii) The core oligosaccharide (Glc₃Man₉GlcNAc₂) is assembled on the dolichol-PP
- iv) The oligosaccharyltransferase transfers Glc₃Man₉GlcNAc₂ onto a specific asparagine site on the nascent polypeptide chain.
- v) Glycoside hydrolases, called glucosidase I and II, then trim this glycan.
- vi) α -1,2-Mannosidase removes a specific terminal α -1,2-mannose residue producing a Man₈GlcNAc₂ structure.
- vii) Man₈GlcNAc₂-containing glycoproteins are then transferred to the Golgi where several α -1,2-mannosidases remove mannose to yield Man₅GlcNAc₂.
- viii) Glycosyltransferases assist in the addition of a diverse variety of monosaccharide units, for example, GalNAc, fucose, sialic and galactose, producing complex oligosaccharides.

The structure of the human core oligosaccharide Glc₃Man₉GlcNAc₂ is given in **Figure 1**, and further examples of complex oligosaccharide structures are illustrated in the literature (Sethuraman and Stadheim 2006). The processes that occur in the ER are highly conserved in lower and higher eukaryotes.

The intention of this review is to present an overview of how the concept of glycoengineering in bacteria for production of human therapeutics has evolved since the discovery of *N*-glycosylation in bacteria. Targets for making this system successful are revealed through specific studies which have uncovered crucial cellular mechanisms, and further targets are suggested for improving overall production rates and hence making the bio-process, a viable one.

Host cells

CHO cells are presently the preferred host for therapeutic protein production (Sheeley *et al.* 1997). The main advantage of CHO cells is the wealth of information available regarding their implementation as protein production cell factories (although interestingly there is no public CHO genome sequence available). Their growth characteristics are well-defined from a scaling-up perspective, and technical advances mean they are relatively simple to manipulate genetically through transfection. Most significantly, they offer a post-translationally modified expression product (Sheeley *et al.* 1997). However, there are major disadvantages in using these cell lines. The costs associated with cell maintenance and growth can be high, and a relatively long time is required for the process of growing cells, expressing relevant genes, and finally harvesting the protein product. There is also the issue of transmitting viruses and prions. Perhaps the most important disadvantage of CHO cells from a biotechnology point of view is the inevitable production of different types of glycoprotein in addition to the glycoprotein desired. This heterogeneous mixture can also be the consequence of differing culture conditions, further complicating the matter (Yuen *et al.* 2005). Therefore, methods of purification are required, increasing overall production costs. Metabolic engineering of CHO cells has seen some advances. For example, the introduction of three proteins, alkaline phosphatase, p21 and CCAAT/enhancer-binding protein α , led to arrests in cell proliferation of growing cell cultures, and thereby increased recombinant protein production up to fifteen times (Fussenegger *et al.* 1998). More recently, the use of antisense DNA and gene targeting have improved glycosylated protein yield (Warner 1999).

Other host cell types include insect cell lines and insect larvae. Although naturally unable to produce complex oligosaccharides, some success has been achieved with sialylated forms using Mimic insect cells expressing five mammalian genes encoding glycosyltransferases (Legardinier *et al.* 2005). Transgenic plants have also been investigated for human glycoprotein production, mainly to overcome the lack of human sialic acid and galactose glycans and remove immunogenic xylose and α 1,3-fucose glycans (Gomord and Faye 2004; Strasser *et al.*

2004). Transgenic animals have been used as host cells (Houdebine 2002), although issues regarding serum half-life have arisen due to natively occurring high-mannose type glycans (van Berkel *et al.* 2002). Yeast expression systems have provided the most exciting advances of late in regard to using alternative host cells (Hamilton *et al.* 2003; Hamilton and Gerngross 2007; Li *et al.* 2006; Potgieter *et al.* 2009). Compared to previously discussed systems, the relatively shorter growth time means potential productivity rates are higher and scaling up fermentation is a well-understood procedure. Yeasts do not naturally perform the trimming associated with human-like N-glycosylations (discussed in the introduction), but share the same initial biosynthesis pathway (Figure 1). The yeast *Pichia pastoris* has been engineered to eliminate adverse glycans and produce human glycosylations, and very importantly to a high degree of homogeneity (Hamilton *et al.* 2003). This ability to control the glycoform of recombinant proteins has far-reaching advantages over rival host cell platforms, as screening different products led to the discovery of therapeutic proteins of enhanced efficacy (Li *et al.* 2006).

Glycosylation in bacteria

Originally believed to be a modification exclusive to Eukarya, glycosylation is now known to occur in Bacteria and Archaea (Abu-Qarn *et al.* 2008; Messner 2004; Yurist-Doutsch *et al.* 2008). The S-layer glycoprotein of *Halobacterium halobium* (*salinarum*) was the first published example of a prokaryotic glycoprotein (Mescher *et al.* 1974). This led to a plethora of reports about S-layer glycoproteins in non-eukaryotes (Schaffer and Messner 2004) and recently, Mobili *et al.* (2009) used periodic acid-Schiff staining of polyacrylamide gel fractionated proteins to reveal that all extracted *Lactobacillus kefir* S-layer proteins were glycosylated (Mobili *et al.* 2009). In addition to S-layer glycoproteins in bacteria, the pili and flagelli glycoproteins have also received a lot of attention. For example, Peng *et al.* (2008) have identified a gene product, Gap3, which is involved in glycosylation and secretion of a newly discovered family of high molecular weight serine-rich pili glycoproteins, which exist in gram-positive bacteria *Streptococci* and *Staphylococci* (Peng *et al.* 2008). Interestingly, the exponential increase in research in this field led to the discovery that glycans in bacteria seem to more structurally and compositionally diverse than their eukaryotic counterparts (Guerry *et al.* 2006; Young *et al.* 2002). Most of the studies on bacterial glycosylation are motivated by the number of pathogenic species which perform this modification (Inga and Schmidt 2002), and in particular, Gram-negative bacteria, where

most glycoproteins are associated with virulence factors. Unique *O*-linked glycan structures were discovered in the flagellin of the opportunistic pathogen *Pseudomonas aeruginosa*, together with site-specific information using β -elimination with ammonium hydroxide. Eleven heterogeneous glycans are attached to the polypeptide backbone through a rhamnose residue and two genes (*orfA* and *orfN*) are involved in their attachment. A concise review of glycan site and structure determination specifically in bacteria is given elsewhere (Hitchen and Dell 2006). *Campylobacter* spp. are examples of Gram-negative human-gut mucosal pathogens, and they are responsible for the majority of cases of gastroenteritis worldwide. In 2002, the first structure of a bacillosamine (2,4-diacetamido-2,4,6-trideoxyglucose) *N*-linked glycan (Figure 2c) (using MALDI-TOF mass spectrometry and nano-NMR techniques) was presented (Young *et al.* 2002). By mutating an essential gene, *pglB*, evidence that the encoded enzyme was involved in the glycosylation process was presented by the presence of unmodified proteins. Lectin chromatography using GalNAc specific binding, and soybean agglutinin assisted in these studies (Linton *et al.* 2002). Subsequently, the entire *N*-linked glycosylation pathway was described in detail using mutant studies and western blotting (Linton *et al.* 2002; Szymanski *et al.* 2002; Young *et al.* 2002). The widespread cases of *Campylobacter* spp. related disease meant this organism was fully sequenced (Parkhill *et al.* 2000), and has been now been characterised in detail at a genetic and post-translational modification level. The *pgl* gene cluster responsible is shown in Figure 2a. It consists of five putative glycosyltransferases required to assemble the heptasaccharide on a lipid carrier, an oligosaccharyltransferase (PglB) required to transfer this onto the protein, a sugar nucleotide modifying enzyme and enzymes involved in sugar biosynthesis.

The significant breakthrough arrived when Wacker *et al.* (2002) transferred the Pgl pathway into *E. coli*, the workhorse of molecular biology, and successfully produced glycoproteins, although in tiny quantities. This opened the door to using this well-understood bacterium for human glycoprotein production. It is useful here to examine the advantages of using a bacterium such as *E. coli* to produce human glycoproteins. From a molecular standpoint, *E. coli* has a published genome sequence, and is facile to genetically modify. It is the most studied of all organisms; thus its metabolic pathways are well elucidated and there is a wealth of literature allowing more accurate predictions of its behaviour in various conditions or in responses to genetic changes (Baneyx 1999). Generally, bacteria can produce higher titres of product (volumetric productivity $\text{mg l}^{-1} \text{h}^{-1}$) even compared to yeast (Müller *et al.* 2006), and are certainly cheaper to culture with shorter fermentation times. Reduced viral contamination issues often associated with mammalian cell culture is also a vital advantage. More specifically,

bacteria are less sensitive to glycosylation changes, and essentially cell survival is not dependent upon the glycosylation process, and cell death is therefore unlikely. This means glycosylation control is more amenable in *E. coli*, leading to less heterogeneous glycoprotein products. Further, controlling glycan type and site can lead to increased efficacy as detailed earlier (Li *et al.* 2006). The potential to modify protein structure, half-life, immunogenicity, cellular uptake and target recognition is extremely exciting, and bacteria would be an ideal host compared to eukaryotes where hundreds of enzymes are involved, complicating process control.

More recent work has increased understanding of this *N*-glycosylation system. In bacteria, the *N*-glycosylation process occurs independently from protein translocation machinery, and therefore fully folded proteins can be glycosylated (Kowarik *et al.* 2006). This makes the process a true post-translational modification, as opposed to eukaryotes where it is more of a co-translational one. The putative oligosaccharyltransferase in *C. jejuni*, PglB, recognises an extended sequon compared to oligosaccharyltransferases in eukaryotes (Figure 2), in addition, recognition requires a negatively charged amino acid side chain at position -2 (Kowarik *et al.* 2006). By performing elegant mutational and structural experiments, the relaxed specificity of PglB has been demonstrated (Linton *et al.* 2005). This means that variant glycan structures can be added to the polypeptide backbones. It is now known that proteins native to *E. coli* interact with the *pgl* locus (Linton *et al.* 2005). One example is Wzx, an ABC transporter which can replace the transport activity of PglK, and lead to the provision of incomplete glycans (Kelly *et al.* 2006). Another example is the WecA protein, an undecaprenylphosphate (UDP) GlcNAc-1-phosphate transferase that has functions in lipid-linked glycan biosynthesis. This latter example showed that a lipid linked intermediate is involved in the process, and confirmed the earlier hypothesis by Wacker *et al.* (2002) that this leads to a heptasaccharide glycan starting with the HexNAc sugar residue, as opposed to the bacillosamine. This pioneering work has contributed significantly for the prospects of glycoengineering in *E. coli*.

Targets and progress for humanizing bacteria

There are several major differences between the *N*-glycosylation process in humans and *Campylobacter* species (Figure 1), including consensus sequence recognition, lipid carrier type, method of glycan transfer and the actual glycan groups themselves. All will require further research in order to humanise glycoproteins produced in bacteria.

As specified earlier, the process of *N*-glycosylation in eukaryotes and bacteria is co-translational and post-translational, respectively.. Therefore, bacteria can glycosylate completely folded proteins most likely because the consensus sites are presented in displayed regions of the protein (Kowarik *et al.* 2006; Rangarajan *et al.* 2007). Eukaryotic proteins are in a flexible form prior to glycosylation, and are subsequently correctly folded. To avoid missing essential *N*-glycosylation sites in therapeutic glycoproteins, the glycosylation sites would need to be in flexible positions at the time of the modification (Kowarik *et al.* 2006).

As shown in Figure 1, the *C. jejuni* system uses a specific sugar donor and attaches bacillosamine to asparagine, using the enzyme glycosyl-1-phosphate transferase (encoded by PglC (Glover *et al.* 2006)). Unfortunately PglC cannot use UDP-GlcNAc or UDP-GalNAc as donors, the primary glycans in human proteins. Therefore, the bacterial system requires an enzyme that can use GlcNAc as a donor substrate, and attach this sugar to the acceptor (UDP-PP). Bacterial systems do contain proteins that can transfer HexNAc sugars to the membrane bound polyprenol phosphate acceptor as a function of peptidoglycan biosynthesis. Examples include WecA, MraY, and WbpL which transfer GlcNAc, MurNAc-pp and FucNAc/QuiNAc from sugar donors, respectively (Price and Momany 2005). Therefore, WecA is a candidate protein here for GlcNAc transfer, and this strategy is discussed later. Moreover, studies on the catalytic mechanisms and substrate specificities in bacterial glycosyltransferases have revealed carbohydrate recognition domains which could be manipulated in the future (Price and Momany 2005).

As well as the actual transfer of the oligosaccharide to the lipid anchor, another essential process is the assembly of the LLO acceptor. The availability of the lipid carrier portion dolichol phosphate (DOL-P) in eukaryotes and undecaprenol (UND-P) in prokaryotes, is a key factor for LLO generation. This potentially impacts on the *N*-glycosylation capability of a cell, and therefore it is useful to understand how it is generated. In addition, the recycling pathway for UND-P synthesis from preformed UND-PP, which is ultimately released into the periplasm, may also contribute to the total UND-P pool available for the initiation of lipid-linked glycan biosynthesis (Tatar *et al.* 2007). LLO are made up of hydrophobic polymers called polyisoprenoid alcohols. In *E. coli* the precursors to polyisoprenoid alcohols are isopentenyl pyrophosphate (IPP), and dimethylallyl pyrophosphate (DMAPP). Both these precursors are generated in *E. coli* by the DOXP pathway, a possible target for metabolic engineering (Figure 3). In eukaryotes, the dolichol pathway occurs in the ER membrane, and is the process for generating the characteristic human tetradecasaccharide (Glc₃Man₉GlcNAc₂) on the DOL-

PP carrier (Burda and Aebi 1999). This is assisted by a series of glycosyltransferases, which have been elucidated via mutagenesis studies, and have been designated the *ALG* genes. Work on the *ALG* gene products has progressed significantly in recent years, and this is crucial for humanising bacterial protein products (Bickel *et al.* 2005; Chantret *et al.* 2005; Gao *et al.* 2005; Weerapana and Imperiali 2006). The introduction of *ALG* genes into bacteria is essential, as they lack the glycan trimming and elaboration steps that are present in eukaryotes, and crucial for complex oligosaccharide synthesis (Figure 1). Furthermore, these genes require engineering to recognise the UND-PP-GlcNAc, as opposed to the DOL-PP-GlcNAc motif they currently identify.

Complex human glycans are characterised by terminal sialic acid residues that have a strong effect on the half-life of glycoproteins in the serum. Not only do incorrectly sialylated proteins reduced life expectancy but additional sialic acids increase the half-life of recombinant human erythropoietin (Egrie and Browne 2001; Elliott *et al.* 2003). Sialylation is performed by sialtransferases and, unfortunately, the expression of such mammalian glycosyltransferases in bacteria leads to very low production of functional product. However, Skretas *et al.* (2009) recently expressed human sialyltransferase ST6GalNAcI, an enzyme that sialylates *O*-linked glycoproteins, in *E. coli*. This is the first example of the functional expression of human glycosyltransferases in bacteria. Impressively, by generating thioredoxin and glutaredoxin/glutathione mutant cells, oxidising conditions were created in the cytoplasm, and hence functionality of soluble ST6GalNAcI was maintained (Skretas *et al.* 2009). Co-expression of specific cellular chaperones increased this yield further. Importantly, this study has implications for the functional expression of human glycosyltransferases (*N*-linked and *O*-linked) in *E. coli*.

The oligosaccharyltransferase in eukaryotes is an enzyme comprised of eight membrane bound subunits including the STT3 unit in yeast, that has been directly implicated in the process of catalytically adding sugars to asparagine (Kelleher and Gilmore 2006). *Campylobacter* protein PglB is homologous to STT3 (Dempski and Imperiali 2002), and both contain the highly conserved WWDYG amino acid sequence. The flexibility of PglB, presented previously, adds encouragement to the concept of controlling its function in bacterial expression systems. In *E. coli*, *O*-antigen LPS biosynthesis involves the addition of glycan groups to a lipid A core. Replacing the enzyme *O*-antigen ligase with PglB illustrated that various *O*-antigen glycans could be transferred to acceptor proteins (Feldman *et al.* 2005). These findings combined with further engineering work means PglB

could potentially be used to transfer complex oligosaccharides in N-linked glycosylation. However, there are significant hurdles at present for using PglB. Firstly, the difference in consensus sequences means the existing N-X-S/T in eukaryotes is not recognised by PglB, although this can be overcome by engineering the bacterial consensus D/E-Z-N-X-S/T into eukaryotic proteins. Moreover, work uncovering structural and functional characteristics of the protein could contribute to altering the enzyme's substrate specificity, and lead to eukaryotic consensus recognition.

In 2004, a new strategy for generating glycoproteins in *E. coli* was reported (Zhang *et al.* 2004). They expanded on previous breakthroughs regarding the co-translational process of incorporating unnatural amino acids into proteins using evolved tRNA synthetase/tRNA pairs (Wang *et al.* 2001). They provide evidence for successful incorporation of GlcNAc (importantly the primary human-type glycan) onto serine residues on myoglobin, and also demonstrated that expressing a glycosyltransferase can lead to the addition of the galactose sugar onto this glycan. However, this illustration of a unique strategy for producing homogenous glycoforms in *E. coli* was unwarranted as the authors retracted the paper in late 2009 as a consequence of their inability to replicate the results (Zhang *et al.* 2009). This is a good example of just how complex the process of glycosylation is, and how a variety of factors can affect the final protein product.

More recently, a combined (*in vivo* and *vitro*) method for producing eukaryotic N-glycoproteins has been demonstrated (Schwarz *et al.* 2010). In this study, the Pgl pathway is introduced into *E. coli*, minus the genes coding for sugar biosynthesis (*pglD*, *pglE* and *pglF*) and transfer (*pglC*). This meant that the native WecA protein in *E. coli* (referred to previously) was able to transfer GlcNAc-1-phosphate to UND-P making the primary glycan GlcNAc (human-like) as opposed to bacillosamine. Subsequent purification and exo- α -N-acetylgalactosaminidase digestion meant GlcNAc tagged proteins could be used as a substrate for *in vitro* transglycosylation reactions. When an excess of Man₃GlcNAc was added to these reactions, a branching Man₃GlcNAc₂ structure was obtained on *Campylobacter* protein AcrA as well as eukaryotic proteins, human IgG-Fc (CH2 domain) and the single-chain antibody F8 (Schwarz *et al.* 2010).

When *E. coli* is used for high-level recombinant protein expression, it is important to achieve a desirable balance between anabolic and catabolic reactions. Assembling oligosaccharide groups could represent a bottleneck in recombinant glycoprotein production that would ultimately lead to very small amounts of glycoprotein

production. Low levels of the glycoprotein, AcrA, were produced when the Pgl pathway was transferred to *E.coli* (Wacker *et al.* 2002). Moreover, *E. coli* needs to be able to produce its own precursors for glycosylation, primarily because of the high cost of adding the required chemicals to the media, and the relatively poor uptake efficiency of glycans with free hydroxyl groups (Sarkar *et al.* 1995). *E. coli* generates the correct glycan GalNAc using the pathway shown in Figure 3 and this was illustrated previously in an *in vitro* study using a combination of six purified recombinant enzymes (Shao *et al.* 2002). Processes including central carbon metabolism (glycolysis), nucleotide sugar metabolism and peptidoglycan synthesis are all involved in production.

As mentioned previously, although *E. coli* can glycosylate *C. jejuni* proteins PEB3 and AcrA (Wacker *et al.* 2002), it does so very inefficiently. A study using global proteomics technique- isobaric tags for relative and absolute quantification (iTRAQ), metabolic maps on graphs (MMG) and pseudo-selective reaction monitoring (SRM) led to forward engineering of metabolism in an *E.coli* strain which contained the Pgl pathway and AcrA protein (Pandhal *et al.* submitted). Production of the glycosylated protein was improved using this systems biology based approach by over 300%. Further manipulation of the major and minor pathways would ultimately lead to a fitter strain for glycoprotein production, and engineered eukaryotic cells e.g. CHO cells, can reveal already tried and tested successful strategies, which could feasibly be applied to bacteria (Warner 1999).

Conclusions

Now that it is known that the most popular bacterial host for recombinant protein production can perform *N*-glycosylation, the race (academically and industrially) to achieve the ultimate goal and produce human-type glycoproteins is on. Increased understanding of this modification in *C. jejuni* has revealed interesting insights into shared and contrasting processes with eukaryotic glycosylation, essential for achieving the first '*E.coli*-produced human glycoprotein' ambition. Humanizing the yeast, *P. pastoris*, has been achieved through a multitude of techniques, and there is no reason why similar principles cannot be applied in *E. coli*, particularly whilst hurdles are continuing to be overcome. Metabolic engineering (or synthetic biology) strategies are taking a variety of forms due to the better availability and accessibility of sophisticated genomic and post-genomic tools, as well as the application of systems biology approaches to reveal metabolic bottle-necks in host cells. With a combination of these research outcomes, and from lessons learnt from previous studies on glycoprotein-

producing mammalian cells, glycoengineering in *E. coli* for human therapeutic drug production holds a potentially sweet future.

Acknowledgements

The authors acknowledge funding from the UK's Biotechnology and Biological Sciences Research Council (BBSRC) through the Bioprocess Research Industry Club (BRIC) programme (ref BBF0048421), and also from the Engineering and Physical Sciences Research Council (EPSRC) (ref EP/E036252/1).

References

- Abu-Qarn M, Eichler J, Sharon N (2008) Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr Opin Struct Biol* 18:544-550.
- Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1473:4-8.
- Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. *Curr Opin Biotechnol* 10:411-421.
- Bickel T, Lehle L, Schwarz M, *et al.* (2005) Biosynthesis of lipid-linked oligosaccharides in *Saccharomyces cerevisiae*: Alg13p and Alg14p form a complex required for the formation of GlcNAc(2)-PP-dolichol. *J Biol Chem* 280:34500-34506.
- Burda P, Aepli M (1999) The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* 1426:239-257.
- Chantret I, Dancourt J, Barbat A, *et al.* (2005) Two proteins homologous to the N- and C-terminal domains of the bacterial glycosyltransferase Murg are required for the second step of dolichyl-linked oligosaccharide synthesis in *Saccharomyces cerevisiae*. *J Biol Chem* 280:9236-9242.
- Dempski RE, Jr., Imperiali B (2002) Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Curr Opin Chem Biol* 6:844-850.
- Egrie JC, Browne JK (2001) Development and characterization of novel erythropoiesis stimulating protein (NESP). *Br J Cancer* 84 Suppl 1:3-10.
- Elliott S, Lorenzini T, Asher S, *et al.* (2003) Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat Biotech* 21:414-421.
- Fanger MW, Smyth DG (1972) Oligosaccharide units of rabbit IgG multiple CHO attachment sites. *Biochem J* 127:757-765.
- Feldman MF, Wacker M, Hernandez M, *et al.* (2005) Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc Natl Acad Sci U S A* 102:3016-3021.
- Fussenegger M, Schlatter S, Datwyler D, *et al.* (1998) Controlled proliferation by multigene metabolic engineering enhances the productivity of Chinese hamster ovary cells. *Nat Biotech* 16:468-472.
- Gao XD, Tachikawa H, Sato T, *et al.* (2005) Alg14 recruits Alg13 to the cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-N-acetylglucosamine transferase required for the second step of N-linked glycosylation. *J Biol Chem* 280:36254-36262.
- Glover KJ, Weerapana E, Chen MM, *et al.* (2006) Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry* 45:5343-5350.
- Gomord V, Faye L (2004) Posttranslational modification of therapeutic proteins in plants. *Curr Opin Plant Biol* 7:171-181.
- Guerry P, Ewing C, Schirm M, *et al.* (2006) Changes in flagellin glycosylation affect *Campylobacter* autoagglutination and virulence. *Mol Microbiol* 60:299-311.
- Hamilton SR, Bobrowicz P, Bobrowicz B, *et al.* (2003) Production of complex human glycoproteins in yeast. *Science* 301:1244-1246.
- Hamilton SR, Gerngross TU (2007) Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr Opin Biotechnol* 18:387-392.
- Hitchen PG, Dell A (2006) Bacterial glycoproteomics. *Microbiology* 152:1575-1580.

Houdebine LM (2002) Antibody manufacture in transgenic animals and comparisons with other systems. *Curr Opin Biotechnol* 13:625-629.

Inga B, Schmidt MA (2002) Never say never again: protein glycosylation in pathogenic bacteria. *Mol Microbiol* 45:267-276.

Jacobs PP, Callewaert N (2009) N-glycosylation engineering of biopharmaceutical expression systems. *Curr Mol Med* 9:774-800.

Johnson IS (1983) Human insulin from recombinant DNA technology. *Science* 219:632-637.

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

Kelly J, Jarrell H, Millar L, *et al.* (2006) Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. *J Bacteriol* 188:2427-2434.

Kowarik M, Numao S, Feldman MF, *et al.* (2006) N-Linked Glycosylation of Folded Proteins by the Bacterial Oligosaccharyltransferase. *Science* 314:1148-1150.

Legardinier S, Klett D, Poirier J-C, *et al.* (2005) Mammalian-like nonsialyl complex-type N-glycosylation of equine gonadotropins in MimicTM insect cells. *Glycobiology* 15:776-790.

Lehle L, Tanner W (1978) Glycosyl transfer from dolichyl phosphate sugars to endogenous and exogenous glycoprotein acceptors in yeast. *Eur J Biochem* 83:563 - 570.

Li H, Sethuraman N, Stadheim TA, *et al.* (2006) Optimization of humanized IgGs in glycoengineered *Pichia pastoris*. *Nat Biotechnol* 24:210-215.

Linton D, Allan E, Karlyshev AV, *et al.* (2002) Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. *Mol Microbiol* 43:497-508.

Linton D, Dorrell N, Hitchen PG, *et al.* (2005) Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol Microbiol* 55:1695-1703.

Mescher MF, Strominger JL, Watson SW (1974) Protein and carbohydrate composition of the cell envelope of *Halobacterium salinarum*. *J Bacteriol* 120:945-954.

Messner P (2004) Prokaryotic glycoproteins: unexplored but important. *J Bacteriol* 186:2517-2519.

Mobili P, Serradell Mde L, Trejo S, *et al.* (2009) Heterogeneity of S-layer proteins from aggregating and non-aggregating *Lactobacillus kefir* strains. *Antonie Leeuwenhoek* 95:363-372.

Müller D, Bayer K, Mattanovich D: Potentials and limitations of prokaryotic and eukaryotic expression systems for recombinant protein production – a comparative view. The 4th Recombinant Protein Production Meeting. Barcelona, Spain: Microbial Cell Factories, 2006.

Parkhill J, Wren BW, Mungall K, *et al.* (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* 403:665-668.

Peng Z, Wu H, Ruiz T, *et al.* (2008) Role of gap3 in Fap1 glycosylation, stability, in vitro adhesion, and fimbrial and biofilm formation of *Streptococcus parasanguinis*. *Oral Microbiol Immunol* 23:70-78.

Potgieter TI, Cukan M, Drummond JE, *et al.* (2009) Production of monoclonal antibodies by glycoengineered *Pichia pastoris*. *J Biotechnol* 139:318-325.

Price NP, Momany FA (2005) Modeling bacterial UDP-HexNAc: polyprenol-P HexNAc-1-P transferases. *Glycobiology* 15:29R-42R.

Rangarajan ES, Bhatia S, Watson DC, *et al.* (2007) Structural context for protein N-glycosylation in bacteria: The structure of PEB3, an adhesin from *Campylobacter jejuni*. *Protein Sci* 16:990-995.

Sarkar A, Fritz T, Taylor W, *et al.* (1995) Disaccharide uptake and priming in animal cells: inhibition of sialyl Lewis X by acetylated Gal beta 1->4GlcNAc beta-O-naphthalenemethanol. *Proceedings of the National Academy of Sciences U S A* 92:3323-3327.

Schaffer C, Messner P (2004) Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiology* 14:31R-42R.

Schwarz F, Huang W, Li C, *et al.* (2010) A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat Chem Biol*.

Sethuraman N, Stadheim TA (2006) Challenges in therapeutic glycoprotein production. *Curr Opin Biotechnol* 17:341-346.

Shao J, Zhang J, Kowal P, *et al.* (2002) Donor substrate regeneration for efficient synthesis of globotetraose and isoglobotetraose. *Appl Environ Microbiol* 68:5634-5640.

Sheeley DM, Merrill BM, Taylor LC (1997) Characterization of monoclonal antibody glycosylation: comparison of expression systems and identification of terminal alpha-linked galactose. *Anal Biochem* 247:102-110.

Skretas G, Carroll S, DeFrees S, *et al.* (2009) Expression of active human sialyltransferase ST6GalNAcI in *Escherichia coli*. *Microb Cell Fact* 8:50.

Strasser R, Altmann F, Mach L, *et al.* (2004) Generation of *Arabidopsis thaliana* plants with complex N-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose. *FEBS Lett* 561:132-136.

Szymanski CM, Burr DH, Guerry P (2002) *Campylobacter* protein glycosylation affects host cell interactions. *Infect Immun* 70:2242-2244.

Tatar LD, Marolda CL, Polischuk AN, *et al.* (2007) An *Escherichia coli* undecaprenyl-pyrophosphate phosphatase implicated in undecaprenyl phosphate recycling. *Microbiology* 153:2518-2529.

van Berkel PH, Welling MM, Geerts M, *et al.* (2002) Large scale production of recombinant human lactoferrin in the milk of transgenic cows. *Nature biotechnology* 20:484-487.

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

Varki A, Cummings R, Esko J, *et al.* (1999) *Essentials of Glycobiology*. Cold Spring Harbor, New York Cold Spring Harbor Laboratory Press.

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.

Wang L, Brock A, Herberich B, *et al.* (2001) Expanding the genetic code of *Escherichia coli*. *Science* 292:498-500.

Warner TG (1999) Enhancing therapeutic glycoprotein production in chinese hamster ovary cells by metabolic engineering endogenous gene control with antisense DNA and gene targeting. *Glycobiology* 9:841-850.

Weerapana E, Imperiali B (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 16:91R-101R.

Young NM, Brisson JR, Kelly J, *et al.* (2002) Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J Biol Chem* 277:42530-42539.

Yuen CT, Storing P, Tiplady R, *et al.*: Relationships between the N-glycan structures and biological activities of recombinant human erythropoietins produced using different culture conditions and purification procedures. *Glycobiology and Medicine*, 2005:141-142.

Yurist-Doutsch S, Chaban B, VanDyke DJ, *et al.* (2008) Sweet to the extreme: protein glycosylation in Archaea. *Mol Microbiol* 68:1079-1084.

Zhang Z, Gildersleeve J, Yang YY, *et al.* (2004) A new strategy for the synthesis of glycoproteins. *Science* 303:371-373.

Zhang Z, Gildersleeve J, Yang YY, *et al.* (2009) Retraction. *Science* 326:1187.

Figure legends

Fig. 1 A comparison of *N*-linked glycosylation in eukaryotes (humans) and prokaryotes (*C. jejuni*). DOL-PP: Dolichyl-pyrophosphate linked tetradecasaccharide UND-PP: Undecaprenyl-pyrophosphate linked deptasaccharide Rft1p : ATP-independent, bi-directional,membrane-spanning flippase, X and Z can be any amino acid except for proline. The calnexin/calreticulin cycle is used for maintaining protein quality control. *In bacteria *N*-glycosylation can occur independently of the protein translocation machinery (Kowarik *et al.* 2006).

Fig. 2 *N*-Linked glycosylation in *C. jejuni*. a) Pgl pathway proteins and functions b) Glycan structure. The numbers match the proteins in Figure 2a indicating which proteins are responsible for the addition of certain sugar groups (revealed through mutational and structural studies (Linton *et al.* 2005). PglH, which transfers three terminal GalNAc residues to the carrier polyisoprene, works by product accumulation inhibition (Troutman and Imperiali 2009) c) Glycan structures including a representation of the glycosylamide linkage of DATDH to asparagine in polypeptide chains.

Fig. 3 Pathways relevant to glycosylation in *E.coli* a) DOXP pathway for lipid carrier synthesis b) sugar precursor generation. Enzymes in blue were highlighted for UDP-GalNAc regeneration by Shao *et al.* (2002).