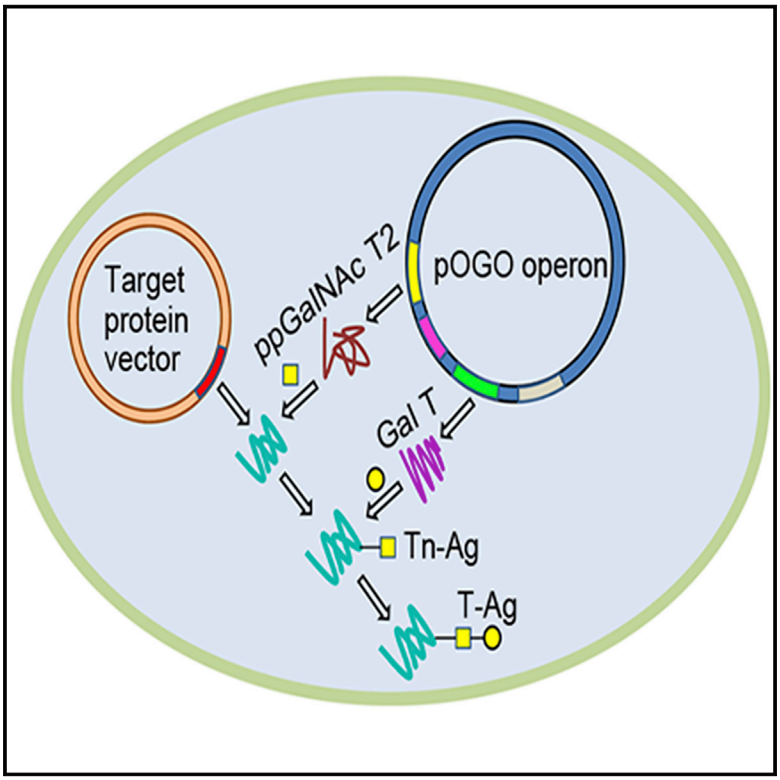
Article

Cell Chemical Biology

A Bacterial Expression Platform for Production of Therapeutic Proteins Containing Human-like O-Linked Glycans

Graphical Abstract



Authors

Ting Du, Nakita Buenbrazo, Laura Kell, ..., Stephen G. Withers, Shawn DeFrees, Warren Wakarchuk

Correspondence

wwakarchuk@ryerson.ca

In Brief

Du et al. describe a synthetic biology approach to producing mammalian-like O-glycans in bacteria. They demonstrate that this glycan can be produced on two examples of human therapeutic proteins. This demonstration of a platform technology for glycosylation has the potential to improve serum half-life of therapeutic proteins made in bacteria.

Highlights

- Designed and expressed a synthetic core-1 O-glycan synthesis operon in *E. coli*
- Improved the O-glycosylation sequon through sequence manipulation
- Produced glycosylated, biologically active interferon-α2b
- Demonstrated a platform expression system for core-1 O-glycan addition

Du et al., 2019, Cell Chemical Biology *26*, 203–212
February 21, 2019
https://doi.org/10.1016/j.chembiol.2018.10.017

CellPress

Cell Chemical Biology
Article

A Bacterial Expression Platform for Production of Therapeutic Proteins Containing Human-like O-Linked Glycans

Ting Du,[1,4] Nakita Buenbrazo,[1] Laura Kell,[1] Sadia Rahmani,[1] Lyann Sim,[2] Stephen G. Withers,[2] Shawn DeFrees,[3] and Warren Wakarchuk[1,5,6,*]
[1]Department of Chemistry and Biology, Ryerson University, Toronto, ON M5B 2K3, Canada
[2]Department of Chemistry, University of British Columbia, Vancouver, BC V6T 1Z1, Canada
[3]La Jolla Pharmaceutical Company, San Diego, CA 92121, USA
[4]Present Address: Dana-Farber Cancer Institute, Harvard Medical School Boston, Boston, MA, USA
[5]Lead Contact
[6]Twitter: twitter@wakarchuk
*Correspondence: wwakarchuk@ryerson.ca
https://doi.org/10.1016/j.chembiol.2018.10.017

SUMMARY

We have developed an *Escherichia coli* strain for the *in vivo* production of O-glycosylated proteins. This was achieved using a dual plasmid approach: one encoding a therapeutic protein target, and a second encoding the enzymatic machinery required for O-glycosylation. The latter plasmid encodes human polypeptide N-acetylgalactosaminyl transferase as well as a β1,3-galactosyl transferase and UDP-Glc(NAc)-4-epimerase, both from *Campylobacter jejuni*, and a disulfide bond isomerase of bacterial or human origin. The effectiveness of this two-plasmid synthetic operon system has been tested on three proteins with therapeutic potential: the native and an engineered version of the naturally O-glycosylated human interferon α-2b, as well as human growth hormone with one engineered site of glycosylation. Having established proof of principle for the addition of the core-1 glycan onto proteins, we are now developing this system as a platform for producing and modifying human protein therapeutics with more complex O-glycan structures in *E. coli*.

INTRODUCTION

Recombinant therapeutic proteins represent a large and growing percentage of approved drugs. However, the efficacy of these drugs can be compromised by short *in vivo* half-lives caused by rapid clearance and proteolysis. In such cases achieving the desired therapeutic outcome requires frequent administration of large doses, bringing with it a higher risk of toxicity, and considerable patient inconvenience and discomfort, not to mention high cost. Many approaches have been taken to improve the *in vivo* activity and circulating half-life of such proteins (Kontermann, 2011), including PEGylation (Bailon and Won, 2009), or fusion with other protein elements such as the immunoglobulin G Fc fragment (Kuai et al., 2000; Levin et al., 2015).

Indeed, the industry standard approach of PEGylation of therapeutic proteins is not without its problems. This popular method to improve serum half-life is achieved through the chemical addition of polyethylene glycol (PEG; reviewed in Harris and Chess, 2003). Recent data, however, suggest that immunogenicity of PEG and the production of anti-PEG antibodies (Verhoef et al., 2014), as well as the accumulation of non-metabolized polyethylene glycol in tissues, are serious concerns (Caliceti and Veronese, 2003). Alternative protein-modification strategies are therefore needed, and glycosylation is an obvious choice.

Most therapeutic proteins are naturally glycosylated with either N- or O-linked glycans, these being the most common and structurally complex natural post-translational modifications (PTMs) (Walsh et al., 2005). In eukaryotic cells both N- and O-glycosylation start in the ER with further decoration taking place in the Golgi. These glycans play important roles in protein folding, solubility, stability, serum half-life, and biological function (Sinclair and Elliott, 2005; Solá and Griebenow, 2010). Consequently, glycoengineering has been suggested as a general strategy to improve the efficacy of therapeutic proteins (Solá and Griebenow, 2010). For example, an engineered, hyperglycosylated form of erythropoietin (EPO), bearing two additional N-glycosylation sites, maintained a 3-fold longer serum half-life in humans compared with its parent protein (Kiss et al., 2010).

However, different expression systems can yield glycoproteins with quite different glycoform profiles. For example, yeast expression systems will add yeast-specific high mannose structures to the protein (Gemmill and Trimble, 1999), while insect cell lines produce their own unique glycans, which are distinct from those of mammals (Jarvis et al., 1998). The presence of incorrect glycan structures or incorrect attachment sites can adversely affect the pharmacokinetic behavior of the therapeutics, and in some cases can lead to immunogenic responses (Kuriakose et al., 2016). Hence, to produce biologically relevant, human-compatible glycan structures, the expression systems to produce therapeutic glycoproteins must be carefully chosen.

Large-scale production of recombinant proteins can be achieved in many expression systems, including bacteria,

insect, yeast, plants, and mammalian cell lines. Nonetheless, eukaryotic production systems such as mammalian cell lines (e.g., Chinese hamster ovary cells or human embryonic kidney cells) remain the preferred approach for glycosylated human therapeutic proteins since they tend to produce correctly folded products with natural glycosylation (Walsh and Jefferis, 2006). However, these eukaryotic production processes are expensive, time-consuming, and still of relatively low productivity.

Bacterial expression systems offer the advantages of fast growth, well-characterized genetics, confirmed safety background, and high protein yield, making them one of the most important industrial expression hosts. However, their reducing environment makes disulfide bond formation challenging, and the fact that they do not naturally produce N- or O-glycosylated products limits their utility. Recent cellular engineering approaches have improved the expression of human proteins in bacteria. Commercial strains, such as *Escherichia coli* SHuffle from New England Biolabs, or Origami from Novogen, offer an oxidizing cytoplasmic environment (Bessette et al., 1999; Lobstein et al., 2012), and SHuffle strains also contain constitutively expressed, cytoplasmic, disulfide bond isomerase (DsbC) to aid in the formation of disulfide bonds and promote proper protein folding. Furthermore, co-expression with other molecular chaperones, such as sulfhydryl oxidase from *Saccharomyces cerevisiae* and human protein disulfide isomerase (hPDI) can improve protein folding and yield (Nguyen et al., 2011). There has also been a report of cytoplasmic expression of disulfide-containing proteins in normal *E. coli* strains that express a mixture of sulfhydryl oxidase and hPDI (Gąciarz et al., 2017).

Since *E. coli* does not naturally produce N- or O-glycosylated proteins with mammalian-style glycans, attempts have been made to address this deficiency and gain proof of principle for both N- and O-linked glycosylation in *E. coli* (Merritt et al., 2013; Pandhal and Wright, 2010). For example, a tri-mannosyl chitobiose N-glycan was successfully synthesized and transferred to the engineered target eukaryotic protein by expression of several glycosyltransferases in *E. coli* (Valderrama-Rincon et al., 2012). To our knowledge, in only one case has nascent O-glycosylation in *E. coli* been published in peer-reviewed literature and that involved simply the initial addition of the GalNAc to a series of mucin-derived peptides and an antibody Fab fragment, although yields of protein were not reported. Subsequent chemical modifications to add PEG to that single sugar were achieved *in vitro* (Henderson et al., 2011). Additionally there are as yet unpublished data (S.D.) from a United States patent (US20090311744A1), which show that about 30% of a test protein can be modified with the disaccharide Galβ-1,3-GalNAcα (core-1 carbohydrate structure) in *E. coli*. The expression of O-glycan-containing recombinant proteins has also been demonstrated in engineered plants (Castilho et al., 2012; Yang et al., 2012).

Protein O-glycosylation occurs naturally in some bacterial species using the PilO or PglL systems (Faridmoayer et al., 2007). This system has been exploited for making conjugate vaccines to bacterial polysaccharide antigens (Pan et al., 2016), but has significant limitations in its ability to make authentic mammalian O-glycans. Since the site of addition on a mammalian therapeutic protein is part of the recognition sequon by the ppGalNAcT enzymes, it is not clear how the PglL oligosaccharyltransferase would be able to discern a mammalian O-linked sequon. In addition, these enzymes require undecaprenyl-linked oligosaccharides and transport of the target protein to the periplasmic space. Transport of the target mammalian proteins to the periplasm is possible but may limit the overall yields compared with cytoplasmic expression (Alanen et al., 2015). Fortunately, success with mammalian glycosyltransferase expression in *E. coli* suggests that the authentic mammalian O-glycans can be made in *E. coli* (Lauber et al., 2015; Rao et al., 2009).

Many human therapeutic proteins could be considered for such production: to start out we have considered two possibilities. The cytokine human interferon-α2b (hIFNα2b) is well known for its antiviral, antiproliferative, and immunomodulatory function. This protein is naturally O-glycosylated with a NeuAcα2,3-Galβ1,3 [NeuAcα2,6]-GalNAcα carbohydrate structure on Thr106 (Adolf et al., 1991; Goodbourn et al., 2000). Currently recombinant hIFNα2b is produced as inclusion bodies in a bacterial system and, as such, lacks the O-glycan found on the mammalian cell-derived protein. In this case, the refolding of the inclusion bodies results in a low recovery and loss of activity (Srivastava et al., 2005). A second issue is the relatively rapid clearance of this non-glycosylated form, necessitating the addition of PEG to improve the serum half-life (Glue et al., 2000). This protein can be produced as a fully glycosylated version in the HEK293 cell line, but this process requires 2 weeks of cell culture per batch of protein (Loignon et al., 2008). Many years ago, the mammalian expression of an engineered variant with four N-linked glycosylation sites was reported, but this does not appear to have been adopted by industry (Reddy et al., 2001), suggesting that a more native glycosylation would be more desirable.

Human growth hormone (hGH) has been used therapeutically for almost four decades (Cai et al., 2014). It is used to treat growth hormone deficiency and has other therapeutic applications in the treatment of Turner syndrome, Prader-Willi syndrome, chronic renal insufficiency, idiopathic short stature in children, and AIDS-related wasting. The natural form of pituitary hGH has been shown to be a complex mixture of isoforms, only one of which is O-glycosylated at Thr60 with the NeuAcα-2,3-Galβ1,3-GalNAcα trisaccharide, which is predicted to interfere with receptor binding (Bustamante et al., 2009), but the clinically used recombinant form is not glycosylated. Recombinant hGH therefore has a very short serum half-life that necessitates PEGylation or fusion protein strategies to stabilize it in the circulation for therapeutic application (Cai et al., 2014).

O-GalNAc or mucin-type O-glycosylation is initiated by the polypeptide GalNAc transferase (ppGalNAcT), which belongs to a large glycosyltransferase family including up to 20 members in humans (Ten Hagen et al., 2003). This enzyme adds a single GalNAc residue, which is also known as the Tn-antigen (Tn-Ag). We have used human ppGalNAcT2, as it appears to be the most promiscuous isoform of this enzyme family (Gerken et al., 2008; Lauber et al., 2015), along with the bacterial β1,3-galactosyltransferase CgtB from *Campylobacter jejuni* (Bernatchez et al., 2007) to synthesize the Gal-β1,3-GalNAcα core structure (T antigen [T-Ag] or core-1 O-glycan). To permit glycosylated protein production we had to introduce the UDP-Glc(NAc)-4 epimerase from *C. jejuni* (Bernatchez et al., 2005) to provide the necessary levels of UDP-GalNAc, along with a

Table 1. Sequon Engineering Using IsoGlyP

Protein	Peptide	ppGalNAcT1 Score	ppGalNAcT2 Score
IFNα2b	GVGV**T**^{106}ETPL	11.45	6.39
IFNα2b	GVGVTE**T**^{108}PL	0.40	0.72
IFNα2b*	GPQP**T**^{106}ETPL	7.96	23.05
hGH	GSPR**T**^{137}GQIF	0.20	0.29
hGH*	GP**T**^{135}PTGQIF	4.52	19.47
hGH*	GPTP**T**^{137}GQIF	0.54	4.53

The asterisks designate the protein as an engineered version.
Residue expected to be glycosylated is shown in boldface, and the score refers to the boldface residue. Sequence numbering refers to the mature form of the protein.

cytoplasmic version of *E. coli* disulfide bond isomerase DsbC (Bessette et al., 1999), or hPDI to aid in folding (Gąciarz et al., 2017). Here, we demonstrate that the authentic core-1 O-glycan structure can be produced on the target recombinant proteins hIFNα2b and hGH in a good yield in an *E. coli* system by co-expressing the target protein with a constructed "O-glycosylation operon," thereby demonstrating the feasibility of this approach in *E. coli*.

RESULTS AND DISCUSSION

Engineering Challenges for Glycoprotein Production in *E. coli*

The addition of the first monosaccharide to the protein is a crucial step in building the core-1 O-glycans. Our approach was to use the well-studied human ppGalNAcT2 isozyme, which has been demonstrated to exhibit a broad recognition of O-glycan sequons through studies with synthetic peptides (Gerken et al., 2011). The enzyme has been expressed in *E. coli* and shown to partially modify peptide sequences near the C termini of fusion proteins (Henderson et al., 2011). We sought to first optimize this initiation step both by increasing expression levels and by exploration of several *E. coli* strains bearing different disulfide bond isomerases that are needed to attain optimal enzyme activity. Furthermore, we wanted to explore whether we could also increase overall target protein modification through simple engineering of their glycosylation sequons. Such sequon engineering makes use of an algorithm developed by Gerken et al. (2011) based upon extensive analysis of ppGalNAcT site specificity. The glycosylation propensity of each sequon is expressed as an IsoGlyP score.

Expression and *In Vivo* Glycosylation of Recombinant Target Proteins in *E. coli*

Our first target protein, the human hIFNα2b, is naturally O-glycosylated on Thr106, which is found on a flexible loop in the structure that does not contact the receptor (PDB: 3SD9) (Thomas et al., 2011). Additionally, the mature protein contains two intramolecular disulfide bonds, Cys1-Cys98 and Cys29-Cys138, the latter being vital for biological activity (Klein et al., 1988; Morehead et al., 1984). Application of the sequon preference algorithm to the hIFNα2b protein confirmed our experimental observation that although both ppGalNAcT1 and -2 can modify the protein *in vitro*, the native sequon is a better match to the preference of ppGalNAcT1 (Figure S1). Since we wish to use a single GalNAc transferase, ppGalNAcT2, in our glycosylation operon, we modified the sequon to PQPT as shown in Table 1, since this gave a significantly higher IsoGlyP score for ppGalNAcT2 without extensive changes to the loop sequence.

Our second target protein is hGH, which has been reported to be O-glycosylated at Thr60, a site that interferes with receptor binding (Bustamante et al., 2009) but is currently marketed as a non-glycosylated, PEGylated material to increase serum half-life (Cai et al., 2014). hGH contains a four-helix bundle structure linked by two disulfide bridges, Cys53-Cys165 and Cys182-Cys189 (Junnila et al., 2013), both of which are required for biological function. Our goal was to explore the addition of core-1 O-glycosylation at a non-native site that would not interfere with receptor binding as an alternative way of improving serum half-life. Using the crystal structure of hGH bound to its receptor from PDB: 3HHR (de Vos et al., 1992) we identified a loop at residues 133–144 remote from the binding site which could be used as a potential glycosylation site (Figure S1).

The data in Table 1 show the original and designed sequons and the IsoGlyP scores for ppGalNAcT1 and -2 and their engineered versions hGH* and hIFNα2b*, designated as such by the asterisk. The second T in the designed hGH sequon (T137) has a score of 4.53 for ppGalNAcT2, suggesting that it would be a poorer substrate compared with T135. The IsoGlyP score for the second T in the designed hIFNα2b sequon is very low, suggesting it would not be used by these ppGalNAcT enzymes. The corresponding mutants were constructed and tested *in vitro* and then with the synthetic operons. High-performance liquid chromatography (HPLC) assays reveal that the sequon-engineered hGH was not a substrate *in vitro* for the ppGalNAcT1 but is a substrate for ppGalNAcT2 (Figure S2).

To express these proteins in *E. coli*, we synthesized codon-optimized genes encoding the processed mature protein and cloned them into a fusion protein vector, GB1-pET-21b (a kind gift from Dr. Steven Smith, Queens University). This GB1 peptide has been shown to enhance solubility and improve expression for several small proteins in *E. coli* (Zhou and Wagner, 2010). With these proteins in hand we then confirmed that they could be O-glycosylated by ppGalNAcT2 and CgtB on the natural or designed sites *in vitro* by specific enzymatic labeling of the T-Ag with a fluorescent sialic acid derivative, intact protein mass spectral analysis, and HPLC analysis (data not shown). This gave us confidence that the proteins were likely to be glycosylated *in vivo* by the enzymes we had chosen for inclusion in the synthetic pOGO operon.

Construction of the O-Glycan Synthesis Operon pOGO

To produce the O-glycan *in vivo*, we constructed a synthetic operon encoding the required enzymes in a plasmid we named pOGO. All the genes were inserted and assembled into the pACYC184 backbone that had been modified to contain the promoter region from pCWori^{+} (Wakarchuk et al., 1994) (Figures 1A and S3). The genes for the maltose binding protein (MBP) fusion to human ppGalNAcT2 and the UDP-Glc(NAc)-4 epimerase (GNE) gene from *C. jejuni* (Bernatchez et al., 2005) were driven by a strong tandem Ptac promoter arrangement from pCWori^{+}. A second Ptac promoter was inserted, followed by the synthetic gene for β1,3-galactosyl transferase (CgtB) from *C. jejuni* fused

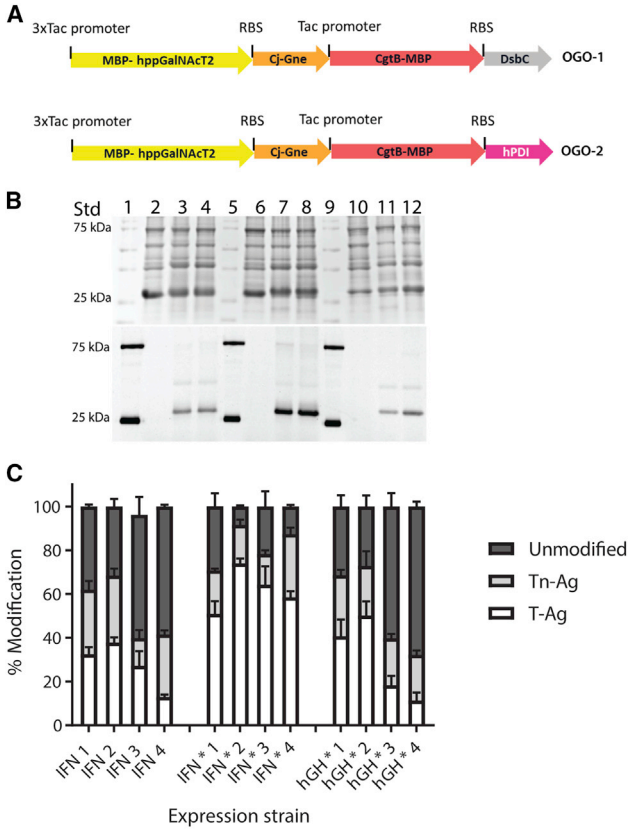


Figure 1. Expression of Target Proteins with the Synthetic Operons OGO-1 and OGO-2

(A) Scheme of the synthetic operons pOGO1 and pOGO2. The elements are an N-terminal MBP fusion of human ppGalNAcT2; a ribosome binding site; the UDP-Hex/NAc C4 epimerase; a Ptac promoter; β1,3-galactosyltransferase (core-1 synthase); a ribosome binding site (RBS); disulfide bond isomerase DsbC (OGO-1); and disulfide bond isomerase hPDI (OGO-2). The RBS was from the intergenic region of the *E. coli* Gal operon between *galE* and *galT*.

(B) Target co-expression with pOGO and *in vivo* O-glycosylation verification from the Origami 2 (DE3) strain. The top panel is Coomassie blue staining of target protein GB1-hIFNα2b, GB1-hIFNα2b*, and GB1-hGH* expressed in *E. coli* with OGO-1 and OGO-2. For lanes 1, 5, and 9 molecular weight standards, two of the marker proteins at 75 kDa and 25 kDa are fluorescently tagged by the manufacturer. Lanes 2–4 are expression of GB1-hIFNα2b alone, expression with OGO-1, and expression with OGO-2. Lanes 6–8 are expression of GB1-hIFNα2b* alone, expression with OGO-1, and expression with OGO-2. Lanes 10–12 are expression of GB1-hGH* alone, expression with OGO-1, and expression with OGO-2. Bottom panel shows detection of T-Ag using the ST3Gal1 assay with 9-BDP-Neu5Ac-CMP as visualized on a blue-light transilluminator.

(C) Glycoform distribution in co-expression strains. Biological replicates of the proteins were generated, and purified protein was analyzed by reverse-phase HPLC. The average of glycoform distribution from three independent growth experiments was placed into a bar-graph format where each section represents a different glycoform. The unmodified protein is shown in dark gray; light gray is the Tn-Ag, and white is the T-Ag modified. IFN refers to GB1-hIFNα2b and IFN* the sequon-engineered GB1-hIFNα2b*. hGH* refers to the GB1-hGH*. Strain 1 is *E. coli* Origami 2 (DE3) with OGO-1, strain 2 is *E. coli* Origami 2 (DE3) with OGO-2, strain 3 is SHuffle Express T7 with OGO-1, and strain 4 is SHuffle Express T7 with OGO-2.

to a C-terminal MBP (Bernatchez et al., 2007) and leaderless DsbC gene or hPDI. We found that the second promoter was necessary to ensure expression of those genes at high enough levels to promote protein modification (data not shown).

To confirm whether the enzymes were functional, we transformed the plasmid containing the operon into several *E. coli* hosts and assayed the individual glycosyltransferases as previously described (Bernatchez et al., 2007). We found that all three host strains, SHuffle T7, SHuffle T7 Express, and Origami 2 (DE3) expressed the active enzymes required to glycosylate the target protein (data not shown). While the SHuffle strains do provide cytoplasmic DsbC, we found it necessary to include a protein disulfide isomerase to obtain higher levels of enzyme activity from the ppGalNAcT2 expressed in Origami 2 (DE3).

Target Co-expression with the Operon pOGO in *E. coli*

Experiments to test the *in vivo* modification of our target proteins were performed using the three protein variants, two OGO operons, and two *E. coli* strains. This was achieved using a dual plasmid expression system, wherein the corresponding target protein constructs were co-expressed with the OGO operons. The expression levels of the co-expressed GB1-hIFNα2b, GB1-hIFNα2b*, and GB1-hGH* were not greatly affected in the dual plasmid strain compared with being expressed alone (Figure 1B).

After purification we initially used lectin blotting with the Peanut agglutinin to confirm the presence of T-Ag modification on the purified proteins, but switched to using an *in vitro* sialyltransferase reaction to more rapidly detect the T-Ag on the target proteins in crude lysates or purified target proteins. This involved a labeling reaction using recombinant porcine α-2,3-sialyltransferase (ST3Gal1) (Rao et al., 2009) to specifically transfer a fluorescently labeled neuraminic acid from cytidine-5′-monophospho-9-BODIPY-N-acetylneuraminic acid (BDP-9-Neu5Ac-CMP) onto the terminal galactose monosaccharide of the T-Ag. This modification can be then visualized on a blue-light or UV transilluminator after separation of the reaction mixture on an SDS-PAGE gel (Figure 1B). We believe that the weakly labeled bands are incompletely denatured aggregated forms of the target proteins, which persist because we do not boil the samples prior to running these gels to preserve the labeled sialic acid residue. In the absence of a target protein, this assay fails to label any proteins (Figure S4).

Confirmation and Quantification of the O-Glycans on Target Proteins Produced *In Vivo*

The absolute amount of modified target protein was accurately determined using an HPLC-based assay (Li et al., 2014) (Figure S5). Identities of the species on the HPLC chromatograms were confirmed using *in vitro* modified protein standards (data not shown). The HPLC assay provides glycoform quantitation that matches well with the data from the intact mass analysis. The HPLC data from three biological replicates were converted to a bar graph (Figure 1C), which shows glycoforms present in various co-expression samples. Under these conditions we observed that a total of 91% of the GB1-hIFNα2b* was modified with 74% of that as T-Ag. The overall modification for the GB1-hGH was lower, at 72%, with 55% as the T-Ag modified form.

Table 2. Intact Mass Data for GB1-Fusion Proteins Co-expressed with OGO Operons

Protein	Theoretical Mass	Observed Mass	Difference
GB1-hIFNα2b	28,079.6	27,945	−134.65 −131.2 Met −3.45 = 2 disulfides
GB1-hIFNα2b + HexNAc	28,147.6	28,147	<1
GB1-hIFNα2b + HexNAc-Hex	28,309.6	28,310	<1
GB1-hIFNα2b*	28,146.7	28,012	−134.7 −131.2 Met −3.5 = 2 disulfides
GB1-hIFNα2b* + HexNAc	28,216.7	28,215	~1.7
GB1-hIFNα2b* + HexNAc-Hex	28,378.7	28,377	~1.7
GB1-hGH*	30,894.5	30,760	−134.54 −131.2 Met −3.3 = 2 disulfides
GB1-hGH* + HexNAc	30,962.4	30,963	<1
GB1-hGH* + HexNAc-Hex	31,124.4	31,125	<1

The masses of the glycosylated forms are calculated assuming that the initiator Met is processed, and that both disulfides have formed.

Co-expression with the OGO-1/2 operons resulted in target proteins with masses consistent with the addition of GalNAcα and Gal-β1,3-GalNAcα, respectively. These results corroborate the presence of Tn-Ag and T-Ag on all the co-expressed target proteins. Thus, these engineered sequon mutants GB1-IFNα2b* and GB1-hGH* have folded correctly with the two expected disulfides, the N-terminal Met has been processed off, and glycosylation has occurred predominately at a single site of modification (Table 2; Figures 2 and 3).

We analyzed the three purified proteins further by examining the peptide fragments to identify the sites of glycosylation (Tables 3 and S1). After the proteins had been digested with both trypsin and chymotrypsin, liquid chromatography-tandem mass spectrometry analysis revealed the sites of modification. Greater than 95% sequence coverage was obtained for both interferon proteins in the analysis, while GB1-hGH* had 89% sequence coverage. In all cases the glycosylation was confirmed at the expected amino residue with a mixture of HexNAc only and Hex-HexNAc, consistent with the intact mass analysis. We also unexpectedly observed some T-Ag from the native IFN sequon at the second Thr residue of the VTETP sequon. The more surprising finding was that the modification of the Thr residues in the engineered hGH sequon showed that the second Thr of the PTPT sequence was better represented in the peptides that were analyzed. In no case did we observe a peptide with two Hex-HexNAc modifications, or two mono-HexNAc modifications, although there were peptides with a HexNAc at Thr214 and Hex-HexNAc at Thr216. This may reflect the preference of the GalT rather than the ppGalNAcT since its normal substrate is not a protein. There was also evidence of a T-Ag-containing disaccharide seven residues downstream of the expected hGH sequon. In addition to the major modifications, some traces of HexNAc-only modification were seen in the GB1-linker portion of all three proteins as is shown in Table S1. Overall the peptide data provided more insights into trace modifications than did the intact mass data. The peptide mass analysis thus revealed more heterogeneity than did the intact mass analysis, but since the data were not quantitative we cannot say with certainty how significant this heterogeneity is for each protein. It does, however, suggest that the IsoGlyP prediction, useful though it is, does not work perfectly in the context of a protein, and that empirical data still need to be collected on a protein-by-protein basis.

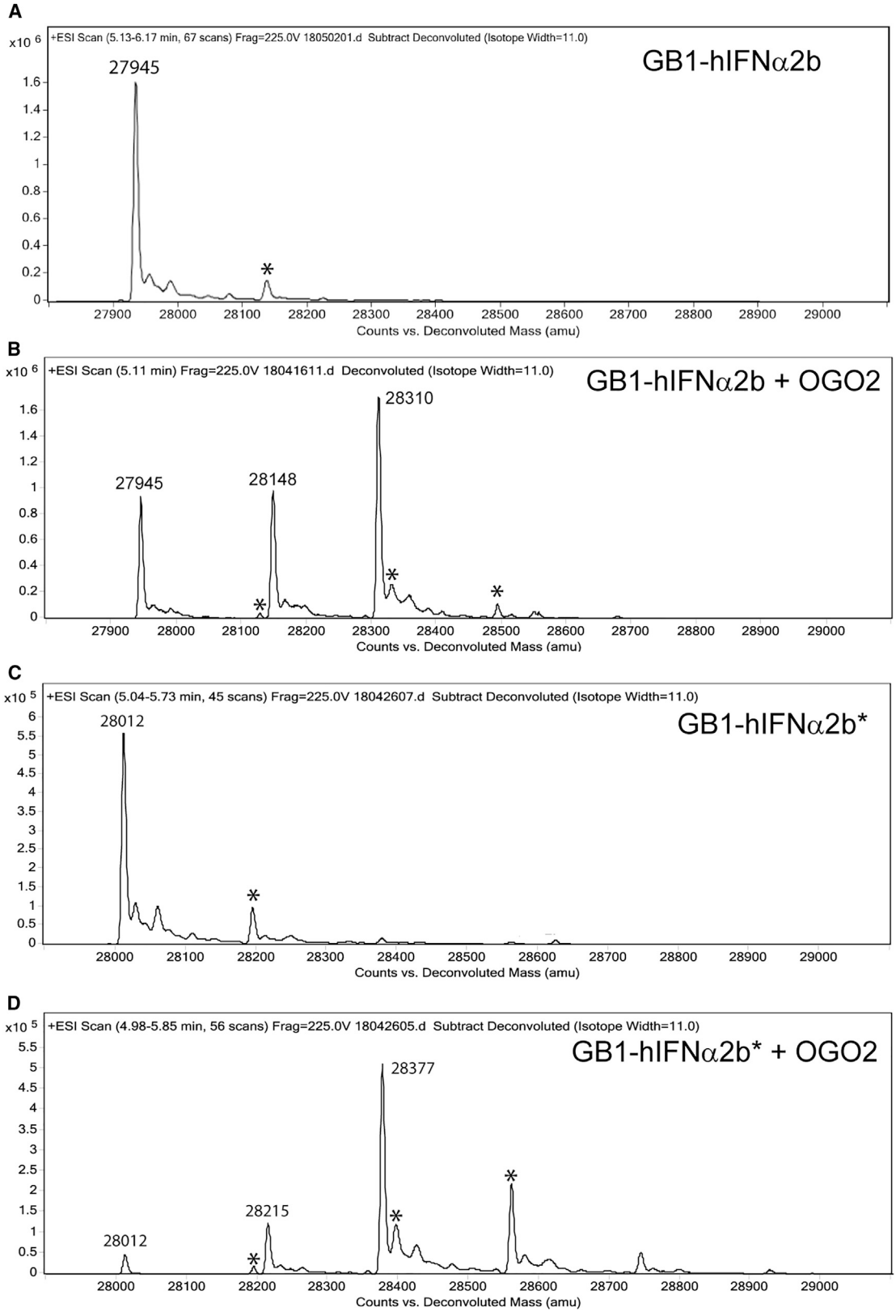
Yields of the glycosylated target proteins were obtained from isolation of the protein from 1 L of culture medium. SHuffle T7 Express produces a higher yield of target protein, whereas Origami 2 (DE3) strain results in lower yield but in some cases higher glycosylation. On average a 4.0-g pellet of cells is produced from 1 L of culture for both *E. coli* strains. Isolated yields from the SHuffle T7 Express strain were up to 80 mg/L for GB1-hIFNα2b and GB1-hIFNα2b*, and up to 67 mg/L for GB1-hGH*. The isolated yield from Origami 2 (DE3) strains yields for GB1-hIFNα2b and GB1-hIFNα2b* were up to 70 mg/L, and for GB1-hGH* up to 60 mg/L.

These assays established that the Origami 2 (DE3) strain produced higher proportions of the T-Ag modified protein in combination with OGO-2 and when the operon contains hPDI. This fits with our finding of around 2-fold higher activity of ppGalNAcT2 when co-expressed with hPDI in Origami 2 (DE3), compared with having DsbC only. This effect is less pronounced with the GB1-hIFNα2b* sequon mutant because the ppGalNAcT2 enzyme more efficiently glycosylates the engineered sequon. The better overall glycosylation in Origami 2 (DE3) might stem from the slightly lower overall expression level of the target proteins in Origami 2 (DE3) where the level of sugar nucleotides available would then be less of a limitation. The efficiency of galactosylation needs to increase to raise the overall conversion to T-Ag, where again there may be a deficiency in the quantity of UDP-Gal required for the reaction. Both the Origami 2 (DE3) and SHuffle express T7 strains are deficient in UDP-Gal synthesis as they are *galE galK* mutants (NEB and Novogen strain genotypes), so a restoration of the Gal operon may augment protein glycosylation efficiency.

The lower efficiency of the GB1-hGH* modification, even though the IsoGlyP score was 19.47, suggests a need to evaluate other variations on the sequon to obtain one that is more efficiently modified. Nevertheless, we have convincingly demonstrated with two examples that we can use the IsoGlyP score to engineer the O-glycosylation sequon for improved glycosylation from ppGalNAcT2. This gives us confidence that we will be able to further optimize our current targets and take on other targets for glycosylation.

Recombinant IFNα2b Biological Activity *In Vitro*

To assess the potency of the recombinant GB1-hIFNα2b relative to a reference hIFNα2b, we examined the activation of the signal transducer and activator of transcription (STAT1) protein. In this cell-based assay, IFNα2b stimulation causes dimerization of the receptor subunits and activation of signals through a well-studied Janus-activated kinase (JAK)-STAT pathway (Rawlings et al., 2004). Within this pathway, STAT1 phosphorylation occurs



(legend on next page)

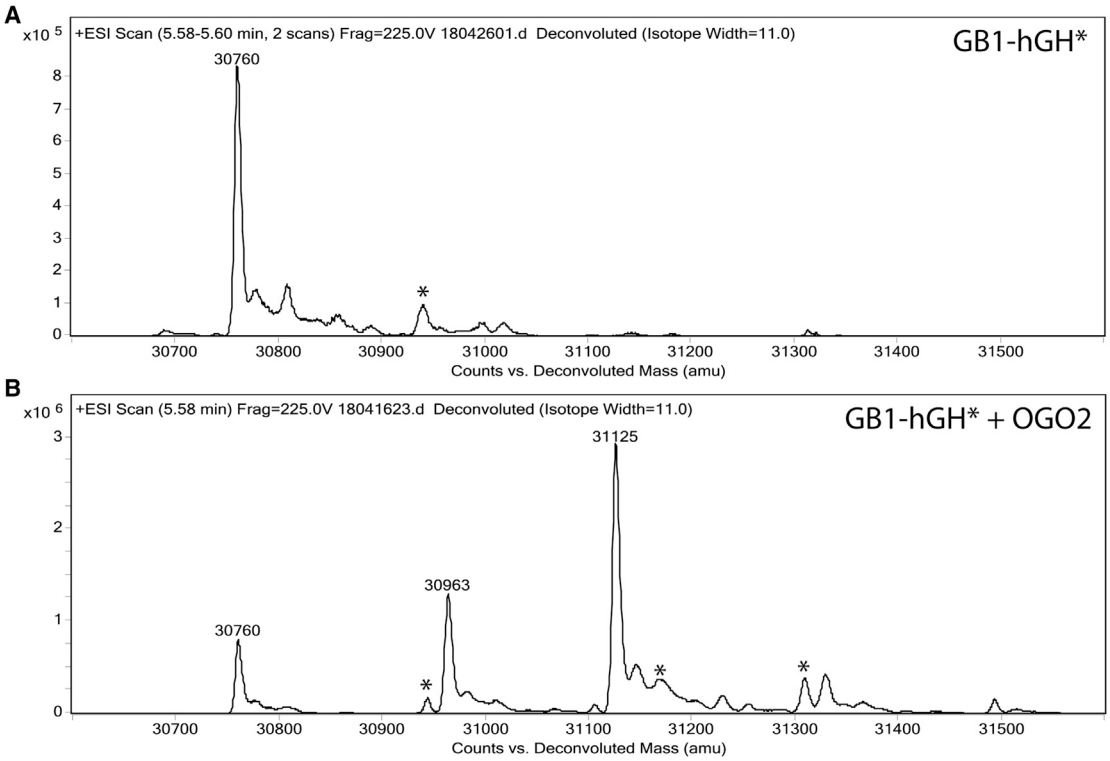


Figure 3. Intact Mass Analysis of GB1-hGH*

(A) GB1-hGH* unmodified and (B) GB1-hGH* with OGO-2 co-expression in the Origami 2 (DE3) background. The parent protein mass was 30,760 Da; the 30,963 peak is from the HexNAc modified, and the 31,125 peak from the Hex-HexNAc modified. The peaks labeled with an asterisk are AEBSF adducts (+183 Da) generated from the protease inhibitor we used.

downstream of IFNα2b binding to IFNαR, eliciting STAT dimerization and subsequent translocation into the nucleus to initiate IFN-stimulated gene transcription (Takaoka and Yanai, 2006). We performed the bioassay with retinal pigment epithelial (ARPE-19, henceforth RPE) cells, previously shown to respond to IFNα (Fasler-Kan et al., 2013).

The range of functional responsiveness of commercial hIFNα2b was measured in different concentrations (2.5 ng/mL, 5.0 ng/mL, and 10 ng/mL) stimulated for 10 min to identify an optimal concentration of IFNα2b to achieve submaximal activation of STAT1 within this system (Figure 4A). From this, 5.0 ng/mL IFNα2b was selected as the optimal treatment condition for comparison of commercial hIFNα2b and GB1-hIFNα2b, given that this elicits significant but submaximal STAT1 phosphorylation. Importantly, treatment of RPE cells with 5.0 ng/mL GB1-hIFNα2b or GB1-hIFNα2b* elicited a similar time-dependent phosphorylation of STAT1 to that observed upon treatment with 5.0 ng/mL hIFNα2b (Figures 4B and 4C), demonstrating the comparable potency of the hIFNα2b, GB1-hIFNα2b, and GB1-hIFNα2b* in binding IFNαR and eliciting downstream signals. As a negative control, we also treated cells with the GB1-hGH fusion protein, produced in a protocol identical to that of GB1-hIFNα2b, which did not elicit STAT1 phosphorylation (Figure 4). This control demonstrates that the activation of STAT1 by GB1-hIFNα2b was (1) not due to the possible action of the GB1 portion of the protein on the IFN receptor, and (2) not due to any non-specific STAT1 activation via endotoxin contamination from the protein purified from *E. coli*.

We also evaluated the glycosylated forms of these proteins (Figure S6) and showed that the activity of these IFNα2b derivatives was independent of their glycosylation, as has been demonstrated previously (Adolf et al., 1991). We have not performed similar assays on our hGH* target protein, but a recent publication (Nguyen et al., 2011) showed that hGH can be produced from expression in *E. coli* as a soluble protein with appropriate disulfide bonds and biological activity. As we continue to engineer this protein, we will need to incorporate a biological assay for hGH to validate that the sequon-engineered versions are still biologically active. We also measured the circular dichroism spectra of both forms of each protein and did not see any significant changes between unmodified and modified proteins (data not shown).

Figure 2. Intact Mass Analysis of GB1-hIFNα2b and GB1-hIFNα2b*

(A and B) GB1-hIFNα2b unmodified (A) and GB1-hIFNα2b with OGO-2 (B) co-expression in the Origami 2 (DE3) strain. The unmodified mass of the protein is 27,945 Da (without the initiator Met, and with two disulfide bonds), the addition of the Tn-Ag (+204 Da) and T-Ag (+366 Da) is seen upon co-expression of OGO-1. (C and D) The same experimental set up with the sequon-engineered GB1-hIFNα2b* (C), where the unmodified protein has a mass of 28,012 Da, and GB1-hIFNα2b* with OGO-2 (D). The peaks marked with an asterisk are attributed to AEBSF adducts from the protease inhibitor mixture we used (+183 Da).

Table 3. Glycopeptide Analysis Data of the Glycosylated Proteins

Protein	Tn-Ag Modification	Spectral Counts[a]	T-Ag Modification	Spectral Counts
GB1-IFN	VIQGVGV**T**ETPLMK	17/63	VIQGVGV**T**ETPLMK	17/63
			VIQGVGVTE**T**PLMK	4/66
GB1-IFN[b]	VIQGPQP**T**ETPLMK	3/11	VIQGPQP**T**ETPLMK	6/11
GB1-hGH[b]	LEDGP**T**PTGQIFK	15/214	LEDGP**T**PTGQIFK	11/214
	LEDGPTP**T**GQIFK	36/214	LEDGPTP**T**GQIFK	56/214
			LEDGPTPTGQIFKQ**T**YSK	2/11
	TSLCFSESIP**T**PSNR	10/211		
GB1	**T**LKGE**TT**TEAVDAA**T**AEK	traces[b]		

[a]Spectral counts are the total number of spectra obtained from peptides with the indicated modification compared with the total number of spectra for the peptide.
[b]All the spectral count data for the glyco-modifications are given in Table S1.

Conclusion

The results we have presented provide the solid base of a platform technology for the O-glycosylation of therapeutic proteins in *E. coli* strains. In the development of this platform we considered the need for multiple ppGalNAcT enzymes to initiate the process, but decided to start with only one enzyme and to investigate the manipulation of the sequon to work better with this base enzyme. However, we remain interested in alternative ppGalNAcT enzymes that may function as better catalysts for use in *E. coli*. A possible limitation of the amino acid sequence changes in the target proteins is that they may alter the biological activity of the protein of interest. Reinvestigation of the biological consequences of these changes is therefore going to be important. The potential also exists for enhanced expression/activity with these modified proteins, which could provide new therapeutic variants to explore. While this is a potential problem, the glycosylation sites we introduce are not predicted to interact with the receptors and thus should be less disruptive than a non-targeted modification.

The next iteration of this glyco-modification will be to introduce the sialic acids that are needed to improve the serum half-life of these proteins. We have previously approached this *in vitro* through polysialylation (Lindhout et al., 2011). The T-Ag intermediate can be modified *in vitro* to carry a single sialic acid, as we show in our detection system, but we plan to engineer strains that would introduce multiple sialic acid residues and would envision having the proteins polysialylated *in vivo* as our long-term goal. Development of this next phase will require genome-level engineering, as the total number of extra genes needed for multiple sialic acid additions to the core-1 O-glycan would make the two-plasmid system unwieldy.

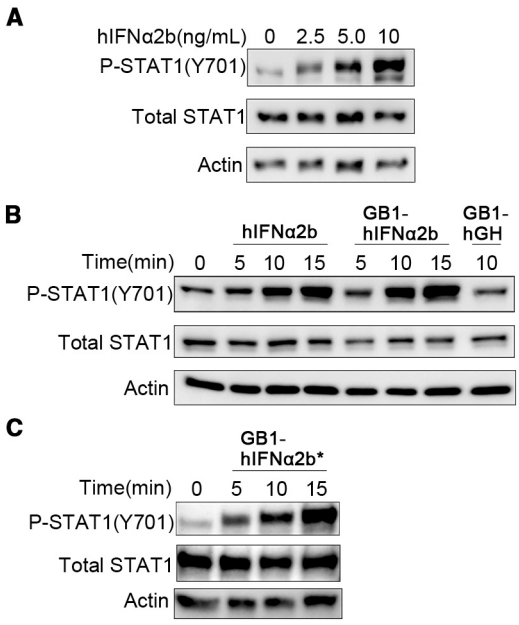


Figure 4. IFNα2b Biological Activity *In Vitro* in a Cell-Based Assay
Retinal pigment epithelial cells (ARPE-19) were stimulated with (A) reference hIFNα2b (2.5 ng/mL, 5.0 ng/mL, and 10 ng/mL) for 10 min (n = 3), (B) 5.0 ng/mL hIFNα2b or 5.0 ng/mL GB1-hIFNα2b, or 5.0 ng/mL GB1-hGH (n = 3), and (C) 5.0 ng/mL GB1-hIFN-α2b* (n = 3) for the indicated times. Whole-cell lysates were prepared, resolved by immunoblotting, and probed for anti-P-STAT (Y701), anti-STAT1, and anti-actin (loading control). Shown are representative immunoblots.

SIGNIFICANCE

Therapeutic glycoproteins are an important area of drug development. Unfortunately, their efficient production remains a significant challenge due to cost and lack of control of the glycan structures produced. The application of synthetic biology to reprogram *Escherichia coli* to make human-style glycans is an attractive option to overcome these barriers of having to perform fermentation with eukaryotic cells to produce these proteins. Some protein drugs are already made in *E. coli*, but they lack the correct post-translational glycan modification found on the native protein. The ability to develop specific strains that would be tailored to produce designed glycans is now feasible, with recent successes in producing human therapeutic proteins and biosynthetic enzymes in *E. coli* strains. We have shown here that we have established a bacterial expression platform for production of therapeutics bearing human-like O-linked core-1 glycans. Using a synthetic operon on one plasmid and a target protein on a compatible second plasmid, we produced the T-antigen modification on three proteins. We have also demonstrated that we can manipulate the O-glycosylation sequon and that such sequence alterations substantially

change the overall glycosylation levels on those proteins, such that we have up to 90% of the target protein modified. This insight into manipulation of the O-glycan sequon, and production of the core-1 O-glycan will aid us to further refine the constructs to optimize and expand the complexity of the glycan with other types of glycosyltransferases, and target proteins to find combinations that provide soluble and fully glycosylated biologically active material.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial Strains
- METHOD DETAILS
  - Bacterial Strains and Growth Conditions
  - Plasmids, DNA Constructs, and Primers
  - Protein Expression and Purification
  - Recombinant Therapeutics Glycosylation *In Vitro*
  - Co-expression with Operon pOGO and *In Vivo* Glycosylation Test with ST3Gal1 Enzyme
  - Glycoform Analysis by HPLC
  - Western Blot
  - Intact Mass Analysis
  - Glycopeptide Analysis
  - Database Searching
  - CD Spectra
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, two tables, and four data files and can be found with this article online at https://doi.org/10.1016/j.chembiol.2018.10.017.

ACKNOWLEDGMENTS

W.W. and S.W. were funded through a grant from the Canadian Glycomics Network https://doi.org/10.13039/501100009056. We thank the Analytical Core at the Canadian Glycomics Network for the intact mass analysis of the target proteins. We thank Sussex Research Laboratories for the kind gift of 9-azido sialic acid, and synthetic peptides for testing enzyme activity. We thank Denis Brochu, Cathy Lorbetski, and Dr. Stephane Bernatchez, National Research Council Canada, for preliminary work on the OGO operon. We thank Nicole Thompson for providing Figure S4.

AUTHOR CONTRIBUTIONS

T.D.: Investigation, Visualization, Writing – Original Draft. N.B. and L.K.: Investigation, Validation, Visualization, Writing – Review and Editing. S.R.: Methodology, Validation, Visualization. L.S.: Investigation, Methodology, Writing – Review and Editing. S.D.: Resources, Conceptualization. S.G.W.: Funding Acquisition, Resources, Supervision, Conceptualization, Writing – Review and Editing. W.W.: Funding Acquisition, Conceptualization, Supervision, Visualization, Writing – Review and Editing, Project Administration.

DECLARATION OF INTERESTS

S.D. is a co-applicant on US patent US20090311744A1, which is related to this work.

Received: July 13, 2018
Revised: September 7, 2018
Accepted: October 19, 2018
Published: November 29, 2018

REFERENCES

Adolf, G.R., Kalsner, I., Ahorn, H., Maurer-Fogy, I., and Cantell, K. (1991). Natural human interferon-alpha 2 is O-glycosylated. Biochem. J. *276* (Pt 2), 511–518.

Alanen, H.I., Walker, K.L., Lourdes Velez Suberbie, M., Matos, C.F.R.O., Bönisch, S., Freedman, R.B., Keshavarz-Moore, E., Ruddock, L.W., and Robinson, C. (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. Biochim. Biophys. Acta *1853*, 756–763.

Bailon, P., and Won, C.Y. (2009). PEG-modified biopharmaceuticals. Expert Opin. Drug Deliv. *6*, 1–16.

Bauer, M.C., Xue, W.F., and Linse, S. (2009). Protein GB1 folding and assembly from structural elements. Int. J. Mol. Sci. *10*, 1552–1566.

Bernatchez, S., Gilbert, M., Blanchard, M.C., Karwaski, M.F., Li, J., Defrees, S., and Wakarchuk, W.W. (2007). Variants of the beta 1,3-galactosyltransferase CgtB from the bacterium *Campylobacter jejuni* have distinct acceptor specificities. Glycobiology *17*, 1333–1343.

Bernatchez, S., Szymanski, C.M., Ishiyama, N., Li, J., Jarrell, H.C., Lau, P.C., Berghuis, A.M., Young, N.M., and Wakarchuk, W.W. (2005). A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in *Campylobacter jejuni*. J. Biol. Chem. *280*, 4792–4802.

Bessette, P.H., Aslund, F., Beckwith, J., and Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. Proc. Natl. Acad. Sci. U S A *96*, 13703–13708.

Bustamante, J.J., Gonzalez, L., Carroll, C.A., Weintraub, S.T., Aguilar, R.M., Muñoz, J., Martinez, A.O., and Haro, L.S. (2009). O-Glycosylated 24-kDa human growth hormone (hGH) has a mucin-like biantennary disialylated tetrasaccharide attached at Thr-60. Proteomics *9*, 3474–3488.

Cai, Y., Xu, M., Yuan, M., Liu, Z., and Yuan, W. (2014). Developments in human growth hormone preparations: sustained-release, prolonged half-life, novel injection devices, and alternative delivery routes. Int. J. Nanomedicine *9*, 3527–3538.

Caliceti, P., and Veronese, F.M. (2003). Pharmacokinetic and biodistribution properties of poly(ethylene glycol)-protein conjugates. Adv. Drug Deliv. Rev. *55*, 1261–1277.

Castilho, A., Neumann, L., Daskalova, S., Mason, H.S., Steinkellner, H., Altmann, F., and Strasser, R. (2012). Engineering of sialylated mucin-type O-glycosylation in plants. J. Biol. Chem. *287*, 36518–36526.

de Vos, A.M., Ultsch, M., and Kossiakoff, A.A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. Science *255*, 306–312.

Faridmoayer, A., Fentabil, M.A., Mills, D.C., Klassen, J.S., and Feldman, M.F. (2007). Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. J. Bacteriol. *189*, 8088–8098.

Fasler-Kan, E., Barteneva, N.S., Ketterer, S., Wunderlich, K., Reschner, A., Nurzhanova, A., Flammer, J., Huwyler, J., and Meyer, P. (2013). Human cytokines activate JAK-STAT signaling pathway in porcine ocular tissue. Xenotransplantation *20*, 469–480.

Gąciarz, A., Khatri, N.K., Velez-Suberbie, M.L., Saaranen, M.J., Uchida, Y., Keshavarz-Moore, E., and Ruddock, L.W. (2017). Efficient soluble expression of disulfide bonded proteins in the cytoplasm of *Escherichia coli* in fed-batch fermentations on chemically defined minimal media. Microb. Cell Fact. *16*, 108.

Gemmill, T.R., and Trimble, R.B. (1999). Overview of N- and O-linked oligosaccharide structures found in various yeast species. Biochim. Biophys. Acta *1426*, 227–237.

Gerken, T.A., Jamison, O., Perrine, C.L., Collette, J.C., Moinova, H., Ravi, L., Markowitz, S.D., Shen, W., Patel, H., and Tabak, L.A. (2011). Emerging

paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. J. Biol. Chem. *286*, 14493–14507.

Gerken, T.A., Ten Hagen, K.G., and Jamison, O. (2008). Conservation of peptide acceptor preferences between *Drosophila* and mammalian polypeptide-GalNAc transferase ortholog pairs. Glycobiology *18*, 861–870.

Glue, P., Fang, J.W., Rouzier-Panis, R., Raffanel, C., Sabo, R., Gupta, S.K., Salfi, M., and Jacobs, S. (2000). Pegylated interferon-alpha2b: pharmacokinetics, pharmacodynamics, safety, and preliminary efficacy data. Hepatitis C Intervention Therapy Group. Clin. Pharmacol. Ther. *68*, 556–567.

Goodbourn, S., Didcock, L., and Randall, R.E. (2000). Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures. J. Gen. Virol. *81*, 2341–2364.

Harris, J.M., and Chess, R.B. (2003). Effect of pegylation on pharmaceuticals. Nat. Rev. Drug Discov. *2*, 214–221.

Henderson, G.E., Isett, K.D., and Gerngross, T.U. (2011). Site-specific modification of recombinant proteins: a novel platform for modifying glycoproteins expressed in *E. coli*. Bioconjug. Chem. *22*, 903–912.

Jarvis, D.L., Kawar, Z.S., and Hollister, J.R. (1998). Engineering N-glycosylation pathways in the baculovirus-insect cell system. Curr. Opin. Biotechnol. *9*, 528–533.

Junnila, R.K., Wu, Z., and Strasburger, C.J. (2013). The role of human growth hormone's C-terminal disulfide bridge. Growth Horm. IGF Res. *23*, 62–67.

Karwaski, M.F., Wakarchuk, W.W., and Gilbert, M. (2002). High-level expression of recombinant *Neisseria* CMP-sialic acid synthetase in *Escherichia coli*. Protein Expr. Purif. *25*, 237–240.

Kiss, Z., Elliott, S., Jedynasty, K., Tesar, V., and Szegedi, J. (2010). Discovery and basic pharmacology of erythropoiesis-stimulating agents (ESAs), including the hyperglycosylated ESA, darbepoetin alfa: an update of the rationale and clinical impact. Eur. J. Clin. Pharmacol. *66*, 331–340.

Klein, M.L., Bartley, T.D., Lai, P.H., and Lu, H.S. (1988). Structural characterization of recombinant consensus interferon-alpha. J. Chromatogr. A *454*, 205–215.

Kontermann, R.E. (2011). Strategies for extended serum half-life of protein therapeutics. Curr. Opin. Biotechnol. *22*, 868–876.

Kuai, L., Wu, C., Qiu, Q., Zhang, J., Zhou, A., Wang, S., Zhang, H., Song, Q., Liao, S., Han, Y., et al. (2000). Plasminogen activator inhibitor-1 fused with erythropoietin (EPO) mimetic peptide (EMP) enhances the EPO activity of EMP. J. Pept. Res. *56*, 59–62.

Kuriakose, A., Chirmule, N., and Nair, P. (2016). Immunogenicity of biotherapeutics: causes and association with posttranslational modifications. J. Immunol. Res. *2016*, 1298473.

Lauber, J., Handrick, R., Leptihn, S., Durre, P., and Gaisser, S. (2015). Expression of the functional recombinant human glycosyltransferase GalNAcT2 in *Escherichia coli*. Microb. Cell Fact. *14*, 3.

Levin, D., Golding, B., Strome, S.E., and Sauna, Z.E. (2015). Fc fusion as a platform technology: potential for modulating immunogenicity. Trends Biotechnol. *33*, 27–34.

Li, Y., Rao, C., Tao, L., Wang, J., Lorbetskie, B., and Girard, M. (2014). Improved detection of variants in recombinant human interferon alpha-2a products by reverse-phase high-performance liquid chromatography on a core-shell stationary phase. J. Pharm. Biomed. Anal. *88*, 123–129.

Lindhout, T., Iqbal, U., Willis, L.M., Reid, A.N., Li, J., Liu, X., Moreno, M., and Wakarchuk, W.W. (2011). Site-specific enzymatic polysialylation of therapeutic proteins using bacterial enzymes. Proc. Natl. Acad. Sci. U S A *108*, 7397–7402.

Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M. (2012). SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb. Cell Fact. *11*, 56.

Loignon, M., Perret, S., Kelly, J., Boulais, D., Cass, B., Bisson, L., Afkhamizarreh, F., and Durocher, Y. (2008). Stable high volumetric production of glycosylated human recombinant IFNalpha2b in HEK293 cells. BMC Biotechnol. *8*, 65.

Merritt, J.H., Ollis, A.A., Fisher, A.C., and DeLisa, M.P. (2013). Glycans-by-design: engineering bacteria for the biosynthesis of complex glycans and glycoconjugates. Biotechnol. Bioeng. *110*, 1550–1564.

Morehead, H., Johnston, P.D., and Wetzel, R. (1984). Roles of the 29-138 disulfide bond of subtype A of human alpha interferon in its antiviral activity and conformational stability. Biochemistry *23*, 2500–2507.

Nguyen, V.D., Hatahet, F., Salo, K.E., Enlund, E., Zhang, C., and Ruddock, L.W. (2011). Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of *E. coli*. Microb. Cell Fact. *10*, 1.

Pan, C., Sun, P., Liu, B., Liang, H., Peng, Z., Dong, Y., Wang, D., Liu, X., Wang, B., Zeng, M., et al. (2016). Biosynthesis of conjugate vaccines using an O-linked glycosylation system. MBio *7*, e00443–00416.

Pandhal, J., and Wright, P.C. (2010). N-Linked glycoengineering for human therapeutic proteins in bacteria. Biotechnol. Lett. *32*, 1189–1198.

Rao, F.V., Rich, J.R., Rakic, B., Buddai, S., Schwartz, M.F., Johnson, K., Bowe, C., Wakarchuk, W.W., Defrees, S., Withers, S.G., et al. (2009). Structural insight into mammalian sialyltransferases. Nat. Struct. Mol. Biol. *16*, 1186–1188.

Rawlings, J.S., Rosler, K.M., and Harrison, D.A. (2004). The JAK/STAT signaling pathway. J. Cell Sci. *117*, 1281–1283.

Reddy, K.R., Wright, T.L., Pockros, P.J., Shiffman, M., Everson, G., Reindollar, R., Fried, M.W., Purdum, P.P., 3rd, Jensen, D., Smith, C., et al. (2001). Efficacy and safety of pegylated (40-kd) interferon alpha-2a compared with interferon alpha-2a in noncirrhotic patients with chronic hepatitis C. Hepatology *33*, 433–438.

Sinclair, A.M., and Elliott, S. (2005). Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J. Pharm. Sci. *94*, 1626–1635.

Solá, R.J., and Griebenow, K. (2010). Glycosylation of therapeutic proteins: an effective strategy to optimize efficacy. BioDrugs *24*, 9–21.

Srivastava, P., Bhattacharaya, P., Pandey, G., and Mukherjee, K.J. (2005). Overexpression and purification of recombinant human interferon alpha2b in *Escherichia coli*. Protein Expr. Purif. *41*, 313–322.

Takaoka, A., and Yanai, H. (2006). Interferon signalling network in innate defence. Cell Microbiol. *8*, 907–922.

Ten Hagen, K.G., Fritz, T.A., and Tabak, L.A. (2003). All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. Glycobiology *13*, 1R–16R.

Thomas, C., Moraga, I., Levin, D., Krutzik, P.O., Podoplelova, Y., Trejo, A., Lee, C., Yarden, G., Vleck, S.E., Glenn, J.S., et al. (2011). Structural linkage between ligand discrimination and receptor activation by type I interferons. Cell *146*, 621–632.

Valderrama-Rincon, J.D., Fisher, A.C., Merritt, J.H., Fan, Y.Y., Reading, C.A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M., and DeLisa, M.P. (2012). An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. Nat. Chem. Biol. *8*, 434–436.

Verhoef, J.J., Carpenter, J.F., Anchordoquy, T.J., and Schellekens, H. (2014). Potential induction of anti-PEG antibodies and complement activation toward PEGylated therapeutics. Drug Discov. Today *19*, 1945–1952.

Wakarchuk, W.W., Campbell, R.L., Sung, W.L., Davoodi, J., and Yaguchi, M. (1994). Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. Protein Sci. *3*, 467–475.

Walsh, C.T., Garneau-Tsodikova, S., and Gatto, G.J., Jr. (2005). Protein post-translational modifications: the chemistry of proteome diversifications. Angew. Chem. Int. Ed. *44*, 7342–7372.

Walsh, G., and Jefferis, R. (2006). Post-translational modifications in the context of therapeutic proteins. Nat. Biotechnol. *24*, 1241–1252.

Yang, Z., Drew, D.P., Jørgensen, B., Mandel, U., Bach, S.S., Ulvskov, P., Levery, S.B., Bennett, E.P., Clausen, H., and Petersen, B.L. (2012). Engineering mammalian mucin-type O-glycosylation in plants. J. Biol. Chem. *287*, 11911–11923.

Zhou, P., and Wagner, G. (2010). Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies. J. Biomol. NMR *46*, 23–31.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-P-STAT1 (Y701)	Cell Signaling Technologies	Cat#: 9167S; RRID: AB_561284
Rabbit monoclonal anti-STAT1	Cell Signaling Technologies	Cat#: 9172S; RRID: AB_10693929
Rabbit monoclonal anti-actin	Cell Signaling Technologies	Cat#: 8456S; RRID: AB_10998774
Recombinant protein		
Human interferon alpha2b recombinant (*E.coli*)	Pestka Biomedical Laboratories (PBL) Assay Science	Cat#: 11,105-1
Porcine ST3Gal1 as MBP fusion	Rao et al., 2009	n/a
Deposited data used		
IFNα2b receptor complex	Thomas et al., 2011	PDB 3SD9
hGH receptor complex	de Vos et al., 1992	PDB 3HHR
Model System Cell Line		
Human: Retinal Pigment Epithelial (ARPE-19)	American Type Cell Culture (ATCC)	RRID: CVCL_0145
Bacterial strains		
SHuffle^{Tm} T7 Express	New England Biolabs	Cat#: C3029J
DH10B	Invitrogen (Thermofisher)	Cat#: 18297010
Origami2^{Tm} (DE3)	Novogen (EMDMillipore)	Cat#: 71408
Recombinant DNA, plasmids and genes		
pET21B with GB1 fusion protein	Dr. Stephen Smith, Queens University, Canada	n/a
pCWMalET	Bernatchez et al., 2007	n/a
pACYC184	New England Biolabs	Cat: # E4152 Discontinued 2018
IFNα2b (Δ23 aa)	BioBasic	UniProtKB:P01563
hGH (Δ26 aa)	BioBasic	UniProtKB: P01241
DsbC (Δ22 aa)	BioBasic	UniProtKB:P0AEG6
hPDI	BioBasic	UniProtKB:P07237
GNE	Bernatchez et al., 2005	UniProtKB:Q0P9C3
CgtB-MPB	Bernatchez et al., 2007	
ppGalNAcT2 (Δ51aa)	Biobasic	UniProtKB: Q10471
ppGalNAcT1 (Δ42aa)	Biobasic	UniProtKb:Q07537
Oligonucleotides		
See Table S2	N/A	N/A
Software and Algorithms		
SnapGene 4.24	GSL Biotech LLC	RRID:SCR_015052
Prism 7	Graphpad Software	https://www.graphpad.com/scientific-software/prism/
Scaffold PTM	Proteomesoftware.com	RRID:SCR_014345
Pymol 2.0	Pymol.org	RRID:SCR_000305

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact: Dr. Warren Wakarchuk, wwakarchuk@ryerson.ca.

A material transfer agreement may be required by the University for strains and plasmids developed from this work.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial Strains

Culture Conditions

All strains were routinely grown in Luria–Bertani (LB) liquid medium and agar plates (1.5% w/v), DH10B at 37°C and SHuffle^{Tm} T7 Express at 30°C respectively. For protein expression the strains were grown in 2YT (Invitrogen, USA) medium. For protein expression all cultures were grown in shake flasks with no more than 10% of the volume as culture media. This was in an orbital shaker at 180-200 rpm at a temperature of 20–30°C depending on the experiment. When needed, antibiotics were added to the media at the following concentrations: chloramphenicol (Cm) at 25μg/mL, ampicillin (Amp) at 150 μg/mL. All strains were maintained as glycerol stocks at −80°C. All strains were from obtained from commercial sources as indicated below but were not further validated in our laboratory.

DH10B, Invitrogen (Thermofisher) Catalogue 18297010. Genotype: F– mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara-leu)7697 galU galK λ– rpsL(StrR) nupG.

SHuffle^{Tm} Express T7, New England Biolabs, Catalogue C3029J Strain. Genotype: fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB sulA11 R(mcr-73::miniTn10–TetS)2 [dcm] R(zgb-210::Tn10 –TetS) endA1 Δgor Δ(mcrC-mrr)114::IS10

Origami^{Tm}(DE3), Novogen (EMDMillipore) Catalogue 71408. Genotype: Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL F′[lac+ lacIq pro] (DE3) gor522::Tn10 trxB (StrR, TetR)

Eukaryotic Cell Culture

Adult Retinal Pigment Epithelial Cell Line-19; NTC-200. Adult Retinal Pigment Epithelial cell line-19; NTC-200, were obtained from the ATCC, but were not authenticated in our laboratory. Sex of cell = Male.

RRID:CVCL_0145.

ARPE-19; henceforth RPE cells (Human non-immortalized Retinal Pigment Epithelial, male) (ATCC) were maintained in Dulbecco's Modified Eagle's Media (DMEM/F12) (Life Technologies) supplemented with 10% fetal bovine serum, 100 U/mL penicillin, and 100 μg/mL streptomycin (Life Technologies) antibiotics in a humidified incubator at 37°C supplied with 95% O_2 and 5% CO_2. For passaging, cells were washed with phosphate buffered saline and detached with trypsin/EDTA-solution followed by resuspension in fresh DMEM before cells were transferred to new culture flasks.

METHOD DETAILS

Bacterial Strains and Growth Conditions

Three *E. coli* strains were used in this work. For gene cloning we used DH10B and for protein expression both SHuffle^{Tm} T7 Express stain (C3029J, from New England Biolabs), and Origami^{Tm} 2 (DE3), from Novagen, were employed. For plasmid construction and strain selection strains were grown in Luria–Bertani (LB) liquid medium and agar plates (1.5% w/v), DH10B at 37°C and SHuffle^{Tm} T7 Express at 30°C respectively. For protein expression the strains were grown in 2YT (Invitrogen, USA) medium. When needed, antibiotics were added to the media at the following concentrations: chloramphenicol (Cm) at 25μg/mL, ampicillin (Amp) at 150 μg/mL.

Plasmids, DNA Constructs, and Primers

All the synthesized DNA fragments were codon-optimized for *E. coli* expression and purchased from Bio Basic (Canada) as plasmid constructs cloned into the pUC57 vector. Oligonucleotides were ordered from IDT via The Centre for Applied Genomics, Hospital for Sick Children, Toronto. Plasmids were prepared using the GeneJET Plasmid Miniprep kit (Thermo Fisher, USA) and DNA fragments were isolated from agarose gels employing the GeneJET Gel Extraction Kit (Thermo fisher, USA). For operon construction, the backbone plasmid pACYC184 was obtained from NEB (USA). The MBP-ST3Gal1-expression vector was the same as previously published (Rao et al., 2009).

For target protein expression we used a fusion protein vector which expressed the 56 aa immunoglobulin G-binding domain 1 of *Streptococcal* Protein G, GB1 (Bauer et al., 2009) along with an N-terminal poly-histidine tag for facile purification. The GB1-pET-21b plasmid was a kind gift of Dr. Steven Smith, Queen's University, Canada. This GB1 peptide has been shown to enhance solubility and improve expression for several small proteins in *E. coli*. The *E. coli* SHuffle^{Tm} T7 Express strain (NEB), which constitutively expresses a chromosomal copy of DsbC, was used as an expression host for the proteins, along with the Origami^{Tm} 2 (DE3) strain (Novogen), which does not express a cytoplasmic protein disulfide isomerase.

The human therapeutic protein IFNα2b gene fragment (amino acids 24-188) (Gene ID:3440, UniProtKB:P01563) was amplified from the synthetic gene vector with primers IFNfor (5'-CGGGATCCTGTGACCTGCCGCAAACC-3') and IFNre (5'-ACGCGTCGACTTAT TCTTTGCTACGCAGAGATTCT -3') by Phusion High-Fidelity DNA polymerase (NEB). The IFNα2b fragment and GB1-pET-21b vector were digested with *Bam*HI and *Sal*I, gel extracted, ligated, and subsequently transformed into *E. coli* DH10B. The construct GB1-pET-21b::IFNα2b was screened by colony PCR with designed T7 primers and verified by restriction digests and confirmed by DNA sequencing. This protein is referred to as GB1- hIFNα2b. The sequon change was performed using base substitution mutagenesis by overlap extension PCR. Two fragments which contained the engineered sequon were amplified using with two pairs of primers; IFN-MutP1 (5'- CCGGCGTAGAGGATCGAGATCTCGATCCC -3') and IFN-MutP2 (5'- cgttggttgtggaccTTGGATAACGCACG -3'), IFN-MutP3 (5'- ggtccacaaccaacgGAGACGCCACTG -3') and IFN-MutP4 (5'- CGATGGCCCACTACGTGAACCATCACCCTAATC -3')

(the lowercase section represents the overlapping sequence). After final amplification of the IFNα2b fragment using IFN-MutP1 and IFN-MutP4, the insert was digested with *Nde*I and *Sal*I and cloned as described above into the GB1-pET21b. This protein is called GB1-IFNα2b*.

The hGH gene was designed to express amino acids 27-201 and engineered with potential O-glycosylation sites based on the crystal structure of the hGH bound to its receptor (PDB file 3HHR) (de Vos et al., 1992) protein sequence (UniProtKB: P01241, Figure S2). The hGH gene fragment was amplified with hGHfor (5'- CGGGATCCTTCCCCACTATTCCTCTCTC -3') and hGHre (5'- ACGCGTCGACTTAGAAACCGCAGGAGC -3') by Phusion High-Fidelity DNA polymerase (NEB, USA) cloned as was described for hIFNα2b above.

To re-engineer the glycosylation site, overlap extension PCR was applied to construct the sequon. Two fragments which contain the engineered sequon were amplified with two pairs of primers, hGHfor and hGH-MutP2(5'- AATttggccggtcggcgttggACCATCCTCG AGGCGACCCATA -3'), hGH-MutP3(5'- GGTccaacgccgaccggccaaATTTTTAAACAAACTTATTCGAAAT -3') and hGH-MutP4(5'- GC ACTACGTGAACCATCACCCTAATC -3') (lower case represent the overlapping part). After the final amplification, the hGH fragment and GB1-pET-21b vector were digested with *Bam*HI and *Dra*III and then cloned as described above. This protein is referred to as GB1-hGH*.

The operon plasmid was constructed in multiple steps. The complete sequence is available upon request, the operon sequence can be found in the Supplemental Information section. The plasmid pACYC184 was first amplified with primers designed to introduce PciI and ZraI sites. This amplicon contained the Cm^R gene as well as the p15A origin of replication. The promoter and ppGalNAcT2 gene (UniProtKB: from pCW:MalE-ppGalNAcT2 pCW was PCR amplified with a new PciI site to allow cloning and was added as an PciI-XbaI fragment to this intermediate vector. The *Campylobacter* GNE gene (UniProtKB:Q0P9C3) was cloned from CPG-13 (Bernatchez et al., 2005) and inserted into pACYC184::MBP- ppGalNac T2 by *Xba*I and *Sph*I. A synthetic construct was used to introduce the Tac promoter, the C-terminally truncated MBP fusion CJL-136 β1,3-galactose-transferase CgtB-MalE gene (Bernatchez et al., 2007)) and a *dsbC* gene fragment which produces an N-terminally truncated version of the chaperone DsbC which is cytoplasmic (UniProtKB:P0AEG6). Between these two genes we used the *E. coli* Gal operon intergenic region between *galE* and *galT* as it contained a ribosome binding site (RBS) with optimal spacing for such an operon. This sequence XXX**TAA**GGAACGACC**ATGXXX** had the stop codon TAA overlap with the RBS, AGGA, with 6 bases until the start codon of the next gene. This first construct was known as OGO-1.

For replacement of DsbC with human protein disulfide isomerase, hPDI (UniProtKB:P07237) was produced as a synthetic gene (Bio Basic) codon optimized for expression in *E. coli*. PCR primers were designed to add the restriction sites SalI and XhoI for the replacement construction. This operon is designated as OGO-2.

5'AAGCTTGTCGACTAGGAGGTCACatgGATGCACCGGAAGAAGAGGACC3' hPDI-F

5'gggatcctctagactcgagTTACAGTTCGTCTTTCACTGCTTTC3' hPDI-R.

Protein Expression and Purification

The expression plasmids including the GB1-pET-21b::IFNα2b, GB1-pET-21b:: hGH, pCWMalE::ppGalNAc T2 and pCW-CgtB-MalE were each transformed separately into the SHuffle™ Express T7, and Origami™ 2 (DE3) strains. Selected transformants were grown at 30°C in 2YT medium until an optical density at 600 nm (OD600) of 0.5 was reached. Temperature was subsequently decreased to 20°C and 0.25 mM isopropyl-thio β-D-galactoside (IPTG) was added for overnight induction. After overnight incubation, protein expression levels were examined by boiling cells directly in SDS-gel-loading buffer and analyzed on an SDS-PAGE gel.

For recombinant target protein purification, the cell pellet derived from the culture (200 ml up to 1L) was resuspended in IMAC column binding buffer (50 mM Tris, 500 mM NaCl, pH 8, at 10 mL/g of cell pellet) including recommended amount of DNase (Bio Basic) and protease inhibitor (complete protease inhibitor cocktail tablet, Roche life science). The suspension was lysed using an Emulsiflex (Avestin, Canada) homogenizer and cell debris removed by centrifuging at 4°C, 10,000 rpm for 30 min. Supernatant was further clarified by ultra-centrifugation (45,000 rpm for 1 h) and filtering through a 0.45 μm filter. COmplete™ His-tag purification column (Roche life science) was used to purify the polyhistidine-tagged GB1-fusion protein with a gradient elution (50 mM Tris, 500 mM NaCl, from 10 mM to 500 mM imidazole, pH 8). The proteins were further purified by Superdex 75 gel filtration (GE Healthcare Life Sciences) using a 50 mM Tris-HCl, 100 mM NaCl, pH 8 buffer. All the protein quantitation was done using the Pierce BCA protein assay kit (Thermo fisher).

Recombinant Therapeutics Glycosylation *In Vitro*

To test whether the recombinant therapeutic fusion-proteins can be glycosylated by the enzymes we selected and expressed in this work, we performed the *in vitro* enzymatic modification with the purified proteins. The ppGalNAcT enzymes were cloned as MBP fusions using a synthetic gene construct which encoded a protein with a N-terminal truncation cloned into the pCW::MalET vector (Bernatchez et al., 2007) . The Tn-antigen synthesis by MalE-ppGalNAcT1 (ppGalNAcT1 Δ42aa, from UniProtKb:Q07537) or MalE-ppGalNAcT2 (ppGalNAcT2 Δ51aa) enzymes was performed in 20 mM HEPES pH7.2, 10 mM $MnCl_2$, 1 mM UDP-GalNAc as the donor substrate and 5 mg/ml target protein and, with 0.1 mg/ml ppGalNAc enzyme. After 1 hr at 30°C, the reactions were stopped by addition of SDS-PAGE loading buffer, or 10 mM EDTA when the samples were to be analyzed further. CgtB protein modification reactions were performed as previously described (Bernatchez et al., 2007).

Co-expression with Operon pOGO and *In Vivo* Glycosylation Test with ST3Gal1 Enzyme

After assembling the operon into pOGO-1 and 2, the plasmids were transformed into the SHuffle^{Tm} Express T7, and Origami^{Tm} 2 (DE3) strains and protein expression was induced as described above for the individual plasmids. Enzyme activities were confirmed with synthetic peptide substrates by thin layer chromatography or C18-RP HPLC as previously described (Bernatchez et al., 2007). Once it was verified that the operon produced functional enzymes, the strains containing pOGO were constructed with the GB1-pET-21b::IFNα2b, GB1-pET-21b::IFNα2b* and GB1-pET-21b::hGH* plasmids introduced via transformation. The dual plasmid strains were selected from LB plates which contained both chloramphenicol and ampicillin. Protein expression was induced using the same conditions as those used for the single plasmid expression. The crude proteins were extracted from the cells lysed with B-PER Bacterial Protein Extraction reagent (Thermo fisher) or homogenized with an EmulsiFlex-C3 (Avestin). The recombinant therapeutics co-expressed with the pOGO operons were purified as described for the unmodified proteins.

The enzyme MBP-ST3Gal1 (α-2,3-sialyltransferase1 from *Sus scrofa*) was used to test the presence of T-Ag in crude extracts or purified protein of the co-expressed therapeutics. Reactions were performed in 50 mM HEPES pH7.2, 0.05 mM 9-BDP-Neu5Ac-CMP, 0.1 mg/ml ST3Gal1 enzyme and 5 μl crude extracts or 5 mg/mL purified protein at RT for 1 h. Reactions were stopped by adding SDS-PAGE loading buffer and samples were loaded into 15% SDS-PAGE gel and imaged on a blue light tray (Gel Doc™ EZ Imager Bio-Rad) prior to being stained with Coomassie blue. The mol. wt. standard we used was the Bio-Rad two colour standard which has two fluorescent proteins at 25 kDa and 75 kDa. The 9-BDP-Neu5Ac-CMP was synthesized in 2 steps from 9-azido sialic acid (a gift of Sussex Research Chemicals). First the 9-azido sialic acid was labelled with BODIPY FL alkyne (Lumiprobe, USA) using copper catalyzed reactions conditions as suggested by the manufacturer, and then enzymatically coupled to CMP as has been previously described for normal sialic acid (Karwaski et al., 2002). The product was purified by chromatography on Superdex peptide (10 x300 mm) in 50 mM NH_4HCO_3 in 12% acetonitrile. The absorbance at 504 nm was used to quantitate the product (Extinction at 504 = 80,000 $L \cdot mol^{-1} \cdot cm^{-1}$).

Glycoform Analysis by HPLC

The HPLC method performed was based on a previously reported technique (Li et al., 2014). In this work the gradient profile was modified and is unique for each target protein. All other aspects of the method are identical to those of the referenced work. The column was an BIOshell^{Tm} A400 Protein C18, (3.4μm, 15cmx3.0mm; 400Å pore size, Supelco, USA). The mobile phase was a mixture of 0.2% TFA in water (solvent A) and 0.2% TFA in acetonitrile (solvent B). Gradient profiles for the three target proteins are as follows: for GB1-hIFNα2b, 0-5 min, B: 30%; 5-25 min, B: 45-55%; 25-35 min, B: 30%; for GB1-hIFNα2b*, 0-5 min, B: 30%; 5-25 min, B: 40-60%; 25-35 min, B: 30%, for GB1-hGH*, 0-5 min, B: 50%; 5-25 min, B: 50-65%; 25-35 min, B: 50%. Prior to being run, samples were diluted in solvent A, then 500 ng of target protein was injected onto the column. A Shimadzu Prominence Series HPLC was used with fluorometric detection (Shimadzu RF-20A) with an excitation wavelength of 280 nm and emission wavelength of 335 nm.

Western Blot

RPE cells (RRID:CVCL_0145) were seeded onto 12-well plates and incubated at 37°C supplied with 95% O2 and 5% CO2 overnight. Confluent RPE cells were incubated in low serum DMEM for 1 hour, after which they were stimulated with reference hIFN-α2b (2.5 ng/mL, 5.0 ng/mL and 10.0 ng/mL) (obtained from Pestka Biomedical Laboratories, USA) or GB1-hIFNα2b (5.0 ng/mL) for 10 minutes. For direct comparison, RPE cells were stimulated with hIFN-α2b (5ng/mL) or GB1-hIFN-α2b (5ng/mL) or GB1-hIFN-α2b*T (5ng/mL) for 5, 10, and 15 minutes.

For *in vitro* bioassay of hIFNα2b, whole-cell lysates were prepared in Laemmli Sample Buffer (0.5 M Tris pH 6.8, Glycerol, 10% SDS, 10% β-mercaptoethanol, and 5% bromophenol blue) supplemented with 1 mM sodium orthovanadate, 10 nM okadaic acid, and 20 nM Protease inhibitor, all from BioShop, Burlington, ON. Lysates were then passed through a 27.5-gauge syringe and heated for 15 minutes at 65°C. Lysates from whole-cell extracts were probed by western blotting as described here except: whole-cell lysates were loaded onto 10% polyacrylamide gels and subjected to glycine-Tris SDS-PAGE. The transferred membrane was washed with buffer containing 1X TBS (0.1% Tween-20) and blocked with 3% BSA for 1 hour. Membranes were then incubated with primary antibodies in 1.0% BSA as suggested by the manufacturer. Antibodies used for western blotting were as follows: anti-phospho-STAT1 (Y701) (9167S), anti-STAT1 (9172S) and anti-actin (8456S) were all from Cell Signaling Technology. The rabbit IgG HRP conjugate was detected using Luminata Crescendo Western HRP substrate (Millipore, USA) and imaged on the ChemiDoc™ MP Imaging System (Bio-Rad).

Intact Mass Analysis

After the purification described above the modified and unmodified target proteins were buffer exchanged and desalted by Size Exclusion Chromatography on Superdex 75 (GE Healthcare Life Sciences, Canada) in 50 mM ammonium bicarbonate (NH_4HCO_3). After lyophilization, samples were subjected to intact-MASS analysis at the GlycoNet Mass Spectrometry Core Facility at the University of Alberta. For protein molecular weight determination reverse phase high performance liquid chromatography followed by detection using mass spectrometry (RP-HPLC-MS) was performed using an Agilent 1200 SL HPLC System with a Phenomenex Aeris 3.6μm, WIDEPORE XB-C8, 200Å, 2.1 x 50 mm with guard column. The following gradient was used: the column was washed after loading of the sample using a 0.5 mL/min flow rate and 2% mobile phase B, for 1 minute to effectively remove salts. Elution of the proteins was done by using a linear gradient from 2% to 45% mobile phase B over a period of 6 minutes, 45% to 98% mobile phase B over a period of 1 minute, kept at 98% mobile phase B over a period of 3 minutes and back to 2% mobile phase B over a period of

1 minute. Column temperature was 40°C. Mass spectra were acquired in positive ionization mode using an Agilent 6220 Accurate-Mass TOF HPLC/MS system (Santa Clara, CA, USA) equipped with a dual sprayer electrospray ionization source with the second sprayer providing a reference mass solution. Mass correction was performed for every individual spectrum using peaks at m/z 121.0509 and 922.0098 from the reference solution. Mass spectrometric conditions were drying gas 10 L/min at 325°C, nebulizer 20 psi, mass range 100-3000 Da, acquisition rate of ~1.03 spectra/sec, fragmentor 225 V, skimmer 65 V, capillary 4000 V, instrument state 4 GHz High Resolution. Data analysis was performed using the Agilent MassHunter Qualitative Analysis software package version B.03.01 SP3.

Glycopeptide Analysis

Proteins were dissolved in 8 M urea, 50 mM Tris-HCl (pH 8), 4 mM DTT in a reaction volume of up to 50 ul. DTT was added to final concentration of 10 mM, and the reaction heated to 60°C for 30 minutes. Iodoacetamide is added to a final concentration of 10 mM for 15 minutes, after which the reaction is quenched with 40 mM DTT. Samples were diluted with 50 mM NH_4HCO_3 until concentration was <1 M urea. Modified trypsin (MS grade) was added to a final protease:protein ratio of 1:50-1:100 (w/w), and the sample incubated overnight at 37°C. Samples were analyzed on a Orbitrap analyzer (Q-Exactive, ThermoFisher, San Jose, CA) outfitted with a nanospray source and EASY-nLC nano-LC system (ThermoFisher, San Jose, CA). Lyophilized peptide mixtures were dissolved in 11 μL of 0.1% formic acid and 5 μL were loaded onto a 75 μm x 15 cm PepMax RSLC EASY-Spray column filled with 2 μM C18 beads (ThermoFisher San, Jose CA) at a pressure of 800 Bar. Peptides were eluted over 60 min at a rate of 250 nl/min using a 0 to 35% acetonitrile gradient in 0.1% formic acid. Peptides were introduced by nano-electrospray into the Q-Exactive mass spectrometer (Thermo-Fisher). The instrument method consisted of one MS full scan (400–1600 m/z) in the Orbitrap mass analyzer with an automatic gain control (AGC) target of 1e6, maximum ion injection time of 120 ms and a resolution of 70,000 followed by 10 data-dependent MS/MS scans with a resolution of 17,500, an AGC target of 5e5, maximum ion time of 100 ms, and one microscan. The intensity threshold to trigger a MS/MS scan was set to 3.3e4. Fragmentation occurred in the HCD trap with normalized collision energy set to 27. The dynamic exclusion was applied using a setting of 8 seconds. Single charged species, as well as species with charges greater than 8 were excluded.

Database Searching

Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by XCalibur version 3. All MS/MS samples were analyzed using PEAKS Studio (Bioinformatics Solutions, Waterloo, ON Canada; version 8.0 (2016-09-08)). PEAKS Studio was set up to search the human uniprot database modified to include specific recombinant proteins, 7000 entries) assuming the digestion enzyme Trypsin+Chymotrypsin. PEAKS Studio was searched with a fragment ion mass tolerance of 0.020 Da and a parent ion tolerance of 10.0 PPM. Carbamidomethylation of cysteine was specified as a fixed modification. Deamidation of asparagine and glutamine, oxidation of methionine, hexNAc of serine and threonine, and hex1HexNAc1 of serine and threonine were specified in PEAKS Studio as variable modifications.

CD Spectra

The target proteins were purified as described above. After dialysis into 50 mM NH_4HCO_3, samples were diluted to final concentration of 0.2mg/mL in a 1 mm path rectangular cell and were analyzed on a Jasco 810 spectropolarimeter (Jasco International) at 25°C. The instrument and thermal control unit were controlled with Spectra Manager Software (JASCO International). Spectra were measured from 190-250 nm with a step size of 0.1 nm and a response time of 1 sec. The average signal represents five scans.

QUANTIFICATION AND STATISTICAL ANALYSIS

The HPLC analysis of glycoforms is reported with error bars that were calculated in Microsoft Excel as standard deviation from peak integration from the three sets of data we collected for each protein.

For the Western blot analysis, ''n'' refers to the number of biological repeats that were performed for each experiment.