Investigation of sequon engineering for improved O-glycosylation by the human polypeptide N-acetylgalactosaminyl transferase T2 isozyme and two orthologues

Nicole K. Thompson, Leif T. N. LeClaire, Samantha Rodriguez Perez and Warren W. Wakarchuk*

Dept. of Biological Sciences, University of Alberta, Edmonton AB, T6G2E9

* Corresponding author: wakarchuk@ualberta.ca

Abstract

We have been developing bacterial expression systems for human mucin-type O-glycosylation on therapeutic proteins, which is initiated by the addition of α-linked GalNAc to serine or threonine residues by enzymes in the GT-27 family of glycosyltransferases. Substrate preference across different isoforms of this enzyme is influenced by isoform-specific amino acid sequences at the site of glycosylation, which we have exploited to engineer production of Core 1 glycan structures in bacteria on human therapeutic proteins. Using RP-HPLC with a novel phenyl bonded phase to resolve intact protein glycoforms, the effect of sequon mutation on O-glycosylation initiation was examined through in vitro modification of the naturally O-glycosylated human interferon α-2b, and a sequon engineered human growth hormone. As part of the development of our glycan engineering in the bacterial expression system we are surveying various orthologues of critical enzymes to ensure complete glycosylation. Here we present an in vitro enzyme kinetic profile of three related GT-27 orthologues on natural and engineered sequons in recombinant human interferon α2b and human growth hormone where we show a significant change in kinetic properties with the amino acid changes. It was found that optimizing the protein substrate amino acid sequence using Isoform Specific O-Glycosylation Prediction (ISOGlyP, http://isoglyp.ute.edu/index.php) resulted in a measurable increase in $k_{cat}/K_m$, thus improving glycosylation efficiency. We showed that the Drosophila orthologue showed superior activity with our human growth hormone designed sequons compared to the human enzyme.
Introduction

Biological functions of protein O-linked glycans are essential and extensively varied in mammals. Many proteins with therapeutic applications rely on proper glycosylation for increased serum half-life through decreased proteolysis and a reduction in clearance by hepatic lectins. In contrast to serum-derived therapeutic proteins which are often heavily glycosylated with N-linked glycans (1) proteins in the cytokine or hormone families have simple O-glycan structures which are often missing in the recombinant versions, as they are produced in bacterial expression systems leading to the need for chemical modification with polyethylene glycol polymers to sustain serum half-life (2). A recent paper on human peptide hormone O-glycans shows that these glycans can protect peptide hormones from proteolysis leading to increased serum half-life and better bio-availability (3). Recently we developed an expression platform to produce cytokines and hormones in a bacterial expression system with the goal of providing the benefit of both ease of production and adding authentic human O-glycan structures (4). Our test proteins are the natively glycosylated interferon α-2b (IFNα2b) and the human growth hormone (hGH) which carries an (inactivating) O-glycan on ~50% of the 24kDa isoform produced in humans (5). These proteins share a common protein fold of a tetrameric helical bundle (Figure 1) and offer an excellent platform to explore sequon engineering for the optimized addition of O-glycans (IFNα2b), or the introduction of a new glycosylation site (hGH) as potential means of increasing circulating half-life of these therapeutic proteins. We are currently optimizing this expression platform so that we can use synthetic biology to produce complex O-glycans on a variety of therapeutic proteins and peptides.

Figure 1. Location of unstructured loop containing native site of glycosylation for IFNα2b (left from a NMR model, PDB 2LMS to visualize the loop which is missing in crystallographic data), and the equivalent loop on hGH (right, from the PDB file 3HHR where hGH is bound to its receptor) being targeted for sequon engineering.

Those O-glycans which begin an α-linked N-acetylgalactosamine residue are often referred to as mucin-type O-glycans, owing to the heavy presence of these glycans on all mucins.
The O-GalNAc glycans come in 8 core types and can be a single monosaccharide or large, branched structures (6). Humans have a family of 20 polypeptide α-N-acetylgalactosaminyltransferases (GalNAc-Ts) which add the first residue of these O-glycan core structures. These have been extensively reviewed (7-9), and it has been observed that these enzymes are both redundant yet specific to human glycobiology (10).

The GalNAc-T enzymes belong to the CAZy GT-27 glycosyltransferase family (11) and they catalyze the covalent addition of an α-GalNAc moiety from a UDP-α-D-GalNAc donor onto a serine or threonine residue generating the structure known as the Tn antigen, which is then further elongated by multiple glycosyltransferases in the Golgi. GalNAc-T’s have distinct catalytic and lectin domains which are important in the buildup of complex mucin-type glycosylation where multiple residues are glycosylated (7). Within that diverse GT-27 glycosyltransferase family, the GalNAc-T2 isozyme is conserved across diverse eukaryotic organisms, indicating how important the requirement of these O-glycans is. This glycosyltransferase has multiple substrate specificities and has been shown to be promiscuous (12). Unlike their N-linked counterparts, O-GalNAc glycans do not have a precise consensus amino acid sequence (sequon) where the transfer takes place. Some general considerations have emerged based on natural sites of glycosylation as well as extensive synthetic peptide modification work (8). The consensus observations are a strong bias for proline 3 residues C-terminal to the site of glycosylation and a strong preference of threonine over serine for glycosylation (13). This was also validated in vivo in tissue cultured cells (14). Like other GalNAc-Ts, the crystal structure of human GalNAc-T2 shows a proline pocket and other features which are consistent with the observed sequon preference (15).

There has been much published about the sequence specificity used by the GalNAc-T family (7,16). This led to the development of the sequon utilization prediction software: Isoform Specific O-Glycosylation Prediction (ISOGlyP, found at https://isoglyp.utep.edu/). This was developed by Gerken and co-workers (8) and can be used to tune the sequon to improve site occupancy in an in vivo situation (4). This is a boon for the synthetic biology approach because it means you can tailor the protein glycosylation site rather than having to include many GalNAc-T enzymes.

For our synthetic biology approach to the bacterial addition of Core 1 glycans on therapeutic cytokines and hormones we need to use enzymes from a variety of organisms, as we
found eukaryotic Core 1 synthases do not work in the *E. coli* cell (unpublished observations). Previous research suggests that GalNAc-T1/T2 are the isozymes likely to initiate O-glycan synthesis (17). As we were focused on creating a single site of glycosylation, we then concentrated on the GalNAc-T2 isozyme which had been successfully expressed in bacteria. The expression of functional human GalNAc-T2 in *E. coli* was reported (18,19) but a side-by-side comparison of the expression of GalNAc-T2 enzymes from diverse sources in *E. coli* has not been reported. Because the GalNAc-T2 isozyme is conserved in other species, and active eukaryotic glycosyltransferase expression is not uniform in *E. coli*, we sought to examine if orthologues might offer an advantage in the first step of the synthesis of Core 1 glycans on human therapeutic proteins with altered sites of O-glycosylation. In this work we evaluate the *in vitro* performance of three orthologues of GalNAc-T2, produced in *E. coli*, on sequon-engineered versions of recombinant IFNα2b and hGH to determine the effect of sequon manipulation through a kinetic analysis, using reverse phase HPLC to quantitate intact protein modification.
Experimental

Expression and purification of GalNAc-T2 catalytic domain fusions with the *E. coli* maltose binding protein

The glycosyltransferase expression plasmids have been previously described (4). MBP fusions of hGalNAc-T2 (UniProtKB: Q10471, aa 51-447), DmGalNAc-T2 (UniProtKB: Q6WV19, aa 89-493), and CeGalNAc-T4 (UniProtKB: Q81136, aa 65-461) were co-expressed with hPDI in *E. coli* Shuffle® T7 Express (New England Biolabs) in 2x YT medium (16 g yeast extract, 10g tryptone, 5 g NaCl per litre) with 25 μg/mL chloramphenicol, 150 μg/mL ampicillin, and 0.2% (w/v) glucose. The strains were grown at 30°C until an optical density at 600 nm (OD<sub>600</sub>) of 0.5, then induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20°C. After overnight incubation, the cell pellet was resuspended in amylose column binding buffer (20 mM Tris, 200 mM NaCl, 1 mM EDTA, pH 7.4), including Dnase I (Bio Basic) and protease inhibitor (cOmplete™ EDTA-free protease inhibitor cocktail tablet, Roche Life Science). The suspension was lysed by bead mill homogenization with 0.1 mm ceramic beads at 4 m/sec for 60 sec. Cell debris was removed by centrifugation at 18,000 RCF for 30 mins, and the supernatant was filtered through a 0.45 μm polyethersulfone membrane. The clarified supernatant was then applied to a MBPTrap resin cartridge (Cytiva Life Sciences) and the MBP-tagged transferases were eluted with a gradient of 0-10 mM maltose in binding buffer. Typically yields were ~10mg/L of culture.

Expression and purification of target proteins

The target proteins IFNα2b-WT, IFNα2b-MUT, and hGH-MUT1 have been previously described (4) and are based on the pET21b-6xHis-GB1 expression plasmids. The two additional hGH sequon mutants reported in this work were obtained as synthetic genes (Integrated DNA Technologies) and cloned in the same fashion. IFNα2b-WT/MUT (amino acids 24-188) and hGH-MUT1/MUT2/MUT3 (amino acids 27-201) were expressed in *E. coli* Origami™ 2(DE3) (Novagen) in 2x YT medium with 150 μg/mL ampicillin. The strains were grown at 37°C until an OD<sub>600</sub> of 0.4, then induced with 0.5 mM IPTG at 20°C. After overnight incubation, the cell pellet was lysed and clarified as described for the transferases but using IMAC column binding buffer (50 mM HEPES, 300 mM NaCl, pH 8.0). COmplete™ His-tag purification resin (Roche Life Science) was used to purify the polyhistidine-tagged GB1-fusion proteins with a gradient
elution from 10-500 mM imidazole in binding buffer. To remove the imidazole, the proteins were further dialyzed at 4 °C into 50 mM HEPES pH 7.0 using 10 kDa MWCO SnakeSkin™ dialysis tubing (Thermo Scientific).

**Transferase assays**

Activities of the GalNAc-T enzymes were assessed in 20 μL reactions containing 50 mM HEPES pH 7.0, 5 mM MnCl₂, 1 mM UDP-GalNAc, 25 μg/mL GalNAc-T, and 0.1-1.6 mg/mL protein substrate (sample data shown in Figure S1). Reactions were incubated at 30°C and stopped at repeating time intervals by 1:1 addition of 10 mM EDTA in 50% acetonitrile. Transferase assays were repeated with two biological replicates of each enzyme, and kinetic parameters were generated by GraphPad Prism ver. 8.0. The specific activity of hGalNAc-T2 expressed with and without the chaperone hPDI was also assessed in 20 μL reactions containing 50 mM HEPES pH 7.4, 10 mM MnCl₂, 1 mM UDP-GalNAc, 10-50 μg/mL enzyme and 0.6 mg/mL IFNα2b-MUT. Reactions were incubated and stopped in the same manner. The level of target conversion was less than <=20% so these were still initial rates. Units are micromoles/min/mg of enzyme.

**One-pot transferase assays with hGH sequon mutants using BODIPY-sialic acid detection**

A one-pot synthesis reaction was performed with each of the three hGH sequon mutants as summarized in Figure 2. This one pot synthesis reaction results in the addition of sialyl-T antigen on the purified protein, where the sialic acid is tagged with a BODIPY fluorophore (20). The assays contained 50 mM HEPES pH 7.0, 10 mM MnCl₂, 0.1 mM UDP-GalNAc, 0.2 mM UDP-Gal, 0.2 mM BODIPY-9-Neu5Ac-CMP, 1 mg/mL hGH, 25 μg/mL GalNAc-T, 100 μg/mL T synthase (C1GalTA), and 100 μg/mL pST3Gal1. Again, assays were repeated with two biological replicates of each GalNAc-T, and reactions were incubated and stopped in the same manner. The reactions were analyzed again by intrinsic fluorescence instead of the BODIPY label so that both starting material and product could be measured.

**Figure 2.** Schematic for one-pot synthesis of BODIPY-conjugated sialyl-T antigen.
HPLC analysis of transferase assays

The mobile phase was a mixture of 0.1% trifluoroacetic acid (TFA) in water (solvent A) and 0.1% TFA in acetonitrile (solvent B). All samples were diluted in solvent A, then 300 ng of target protein was injected onto a BioResolve RP mAb polyphenyl column (450Å, 2.7 μm, 3 mm x 100 mm; Waters Corporation). 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. The elution was performed as a 10-minute gradient at a flow rate of 1 mL/min and a temperature of 60°C. For IFNα2b-WT and IFNα2b-MUT, the gradient was from 40-45% solvent B, while for hGH-MUT1/MUT2/MUT3 the gradient was from 49-50% solvent B.
Results

Sequence comparison of GalNAc-Ts from human, Drosophila, and Caenorhabditis

A structure-based sequence alignment (Figure 3) using hGalNAc-T2 as a guide (PDB: 2FFU) indicated highly conserved regions across the three GalNAc-Ts we are investigating. We want to find the one best suited for expression in E. coli and modification of therapeutic proteins. These were the human GalNAc-T2 (hGalNAc-T2), the Drosophila melanogaster GalNAc-T2, (DmGalNAc-T2, 66% identity to human), and the Caenorhabditis elegans GalNAc-T4 (CeGalNAc-T4, 83% identity to DmGalNAc-T2 and 53% identity to hGalNAc-T2). CeGalNAc-T4 was chosen for its similarity to the human T2 isozyme, and was shown to be able to glycosylate peptides derived from human proteins (21). DmGalNAc-T2 has been examined for specificity compared to the human orthologue and was shown to have a similar activity on peptide substrates derived from human proteins (16).

Figure 3. Structure based sequence alignment of GT-27 orthologues.

Secondary structure elements were extracted from the PDB file 2FFU for hGalNAc-T2 and used as guide to align the DmGalNAc-T2 and CeGalNAc-T4 orthologues. The structure-based alignment was generated with Expresso (22). The figure was prepared with ESPript 3.0 (23). The residues in red are identical for all three orthologues, those in yellow are conservative changes for all three orthologues.

Expression of the catalytic domains of GT-27 orthologues in E. coli

Our expression system makes use of the E. coli maltose binding protein as a fusion partner (24). We used synthetic genes which were designed to remove the sequences for type II membrane anchor and part of the stem region, as well as the lectin domain. The purity of the proteins is shown in Figure 4. Some proteolytic fragments can be seen in the profiles.

Figure 4. SDS-PAGE analysis of purified GT-27 orthologues, as MBP-fusion proteins.

Two independent batches of purified protein are shown. Lane 1, hGalNAc-T2; Lane 2, DmGalNAc-T2; Lane 3, CeGalNAc-T4; Lane 4, Standards. Lanes 5-7 are a second growth and
purification of the same enzymes shown in Lanes 1-3. The gel is stained with Coomassie brilliant blue.

The GT-27 enzymes have multiple disulfide bonds, and thus do not fold correctly in the cytoplasm of *E. coli* due to its overall reducing redox state. Commercial strains like Shuffle® and Origami® have mutations rendering the cytoplasm oxidising rather than reducing (25). Shuffle strains additionally carry the dsbC gene with a deletion which produces this normally periplasmic disulfide bond isomerase in the cytoplasm. For the GT-27 enzymes, the presence of cytoplasmic DsbC is not sufficient to permit expression of highly active enzyme. Activity of the GT-27 enzymes was further enhanced ~ 4-fold by co-expression of the folding chaperone human protein disulfide isomerase (hPDI) on a compatible plasmid derived from pACYC184 (4). An illustration of the enzyme activity enhancement is shown in Figure 5. With the IFNa2b-MUT substrate, the specific activity of hGalNAc-T2 was 10.4 mU/mg when expressed with hPDI and 2.3 mU/mg without hPDI.

**Figure 5.** HPLC analysis of hGalNAc-T2 activity expressed without and with hPDI. The left panel shows an intact protein reaction (IFNa2b-MUT) with purified hGalNAc-T2 expressed without hPDI, the right panel is the same reaction but with purified enzyme expressed with hPDI.

**Using ISOGlyP to engineer the sequons**

The target proteins IFNa2b and hGH have similar protein folds in that they both have a core of four alpha-helices, with an unstructured loop on the side of the protein facing away from the receptor (IFNa2b PDB 3S9D) (26) (hGH PDB 3HHR)(27). IFNa2b is naturally glycosylated on Thr106 (28) which is within a flexible loop which we analyzed with ISOGlyP (29) to see what changes would be required to increase the efficiency of hGalNAc-T2. As hGH is not normally O-glycosylated on the equivalent loop, we had to engineer a site for hGalNAc-T2 to use it as a substrate. As we previously published (4) a two amino acid substitution in IFNa2b resulted in vastly improved bacterial glycosylation (>85% modification), and a four amino acid rearrangement in hGH (Table S1) gave >60% protein modification for bacterial glycosylation.
Still, the mechanism by which better overall glycosylation was achieved remained unclear, so we undertook to examine enzyme kinetics on the intact proteins to assess changes in enzyme parameters. Additionally, the initial engineered sequon in our hGH-MUT1 protein was sub-optimally glycosylated compared to that of IFNα2b-MUT. What we encountered with hGH-MUT1 was that the first Thr with the higher score was not the preferred site of addition despite the higher score (4). We sought to improve that by creating two more sequon variants based on much higher ISOGlyP scores (Table 1).

Table 1. ISOGlyP scores for designed hGH variants

| Protein       | Loop Sequence | GalNAc-T2 score |
|---------------|---------------|----------------|
| hGH - native  | DGSPR_TGQIFK  | 0.28           |
| hGH-MUT1      | DGPTPTGQIFK   | 17.59 minor*   |
| hGH-MUT1      | DGPTPTGQIFK   | 4.35 major*    |
| hGH-MUT2      | DGQPQPTPFFK   | 52.56          |
| hGH-MUT2      | DGQPQPTPFFK   | 6.64           |
| hGH-MUT3      | DGQPQTAGFFK   | 40.83          |

The bold T is the threonine the GalNAc-T2 score refers to.

* reference 4

Enzyme kinetic assays on intact proteins

In our previous work with bacterial glycosylation, we showed that reversed phase HPLC was a valuable quantitative tool for the analysis of protein modification (4). We compared three target proteins and three GT-27 orthologues, and the data are summarized in Figure 6 and Table 2. The error bars on the graphs are quite large, which we believe reflects handling of different batches of enzyme (i.e. from independent expression cultures) for the replicate values. With the wild type sequon in IFNα2b-WT we see the best activity from the human enzyme. Comparing the wild type sequon, GVGVTTEPL, to the designed sequon, GPQPTEPL, we see a substantial increase in both enzyme rate and catalytic efficiency for all three orthologues. The hGalNAc-T2 enzyme rate increases 19-fold, the DmGalNAc-T2 enzyme increases 200-fold, and the CeGalNAc-T4 enzyme rate increase is 42-fold. This rate enhancement is accompanied by some
moderate changes in the $K_m$ values, the largest of which is in CeGalNAc-T4 where the $K_m$ decreases nearly 4-fold. Overall, the $k_{cat}/K_m$ shows all the enzymes have similar specificity for the improved sequon.

**Figure 6.** Michaelis-Menten curves of the GT-27 enzymes on IFNα2b-WT, IFNα2b-MUT and hGH-MUT1. Note that the scales for the Y-axis are not the same for this comparison. Reaction rates are substantially higher for all three enzymes on the sequon mutant, and the sequon engineered version of hGH is used by all three orthologues. The sequon being modified is shown below each graph. Biological and technical replicates were used, and the error is shown as standard deviation on the graphs.

**Table 2.** Kinetic parameters of GT-27 orthologues on IFNα2b-WT and the enhanced O-glycosylation sequon mutants.

| Target Protein | $K_m$ (μM) | $V_{max}$ (pmol/min) | $k_{cat}$ (min$^{-1}$) | $k_{cat}/K_m$ (min$^{-1}$ nM$^{-1}$) |
|----------------|------------|----------------------|------------------------|-------------------------------------|
| **IFNα2b-WT** |            |                      |                        |                                     |
| hGalNAc-T2     | 18.9 ± 5.79| 2.21 ± 0.28          | 0.39 ± 0.05            | 20.5 ± 6.81                         |
| DmGalNAc-T2    | 15.7 ± 4.72| 0.33 ± 0.04          | 0.06 ± 0.01            | 3.78 ± 1.58                         |
| CeGalNAc-T4    | 9.44 ± 5.30| 0.31 ± 0.06          | 0.06 ± 0.01            | 5.94 ± 4.42                         |
| **IFNα2b-MUT** |            |                      |                        |                                     |
| hGalNAc-T2     | 14.1 ± 4.51| 39.1 ± 5.9           | 6.86 ± 0.06            | 487 ± 156                           |
| DmGalNAc-T2    | 23.2 ± 5.61| 61.6 ± 8.3           | 11.0 ± 1.48            | 474 ± 178                           |
| CeGalNAc-T4    | 4.67 ± 2.00| 12.8 ± 1.8           | 2.29 ± 0.31            | 489 ± 277                           |
| **hGH-MUT1**   |            |                      |                        |                                     |
| hGalNAc-T2     | 43.4 ± 22.2| 16.8 ± 4.89          | 2.95 ± 0.86            | 68.0 ± 40.0                         |
| DmGalNAc-T2    | 91.4 ± 49.1| 6.94 ± 2.63          | 1.24 ± 0.47            | 13.6 ± 12.4                         |
| CeGalNAc-T4    | 37.0 ± 21.9| 5.71 ± 1.83          | 1.02 ± 0.33            | 27.6 ± 25.1                         |

We could not compare the native hGH to the engineered sequon version because the native version is not a substrate for GalNAc-T2. The sequon engineered hGH-MUT1 shows a kinetic profile unique from IFNα2b-MUT, as hGalNAc-T2 has a much better reaction rate than the other two enzymes. The $K_m$ for hGH-MUT1 is approximately 10-fold higher than the engineered IFNα2b-MUT but has a similar $k_{cat}$. Again, all the enzymes show much better
reactivity than what was observed with the IFNα2b-WT sequon, but not as big a change as we observed for the modified IFNα2b.

**Sequon engineering of hGH as assessed by one-pot reactions**

We wanted to examine all three sequon mutations of hGH with *in vitro* modification reactions to examine the kinetic parameters but found that GalNAcT-2 reaction products from hGH-MUT2/MUT3 do not separate well enough under our RP-HPLC conditions to permit direct detection of GalNAc transfer by the GalNAc-T orthologues. To improve resolution, we instead developed a one-pot reaction sequence similar to a published method (30) in which the protein substrate is labelled with a BODIPY-tagged sialyl-T antigen (BDP-SiaT). Under these reaction conditions we could then monitor the rate of BDP-SiaT synthesis to compare relative changes in specific activity between the three orthologues on all three hGH mutants (Figure 7 and Figure S2). Transfer efficiency is improved for all orthologues on the redesigned sequons, with DmGalNAc-T2 showing the greatest magnitude of improvement in specific activity for hGH-MUT2/MUT3 compared to the original hGH-MUT1 substrate.

**Figure 7.** Specific activity of GT-27 orthologues on three sequon mutants of hGH

Using a one-pot reaction with limiting amounts of the GT-27 orthologues, the three hGH mutants were evaluated side-by-side. Using RP-HPLC analysis to detect BODIPY-tagged glycan structures, a relative specific activity could be calculated for these reactions.
Discussion

The examination of the enzymes which initiate mucin-type glycosylation has been extensive and the enzymology of many enzyme family members (GT-27) has been examined largely with synthetic peptides (8,16,31,32). These GT-27 GalNAc-T’s are found in many eukaryotes as mucin-type glycosylation is highly conserved. The recombinant expression of many of these has been reported, although not in an active form in E. coli. In the development of specialized expression systems to produce active eukaryotic glycosyltransferases with multiple disulfide bonds, we have shown that co-expressed folding chaperones can be used to produce multi-milligram amounts of active and relatively pure GT-27 orthologues, which can have up to five disulfide bonds when including the lectin domain. There is some proteolysis of the fusion proteins prior to purification, as is particularly evident with the DmGalNAc-T2 construct presenting two major bands by SDS-PAGE, but this does not appear to compromise enzyme activity.

From the extensive peptide work performed by others (9,16) the isozyme hGalNAcT-2 was shown to have a wider substrate specificity compared to many other human GT-27 family members. Some similarities in activity were also noted with the DmGalNAc-T2, and to a lesser extent CeGalNAc-T4 (33). Each of these enzymes has a native set of protein substrates, however it is not known if the DmGalNAc-T2 or CeGalNAc-T4 ever glycosylate proteins with structures like the cytokines we used here. So, it is not entirely surprising that the wild type sequon in IFNα2b-WT is not used very well. This is also true for the human enzyme; the IFNα2b-WT sequon is predicted by ISOGlyP to be preferred by the GalNAc-T1 isoform, which is consistent with what we have previously observed (4).

What was very striking with the IFNα2b-MUT sequon was the exceptionally large improvement in reaction rate and turn-over for DmGalNAc-T2 compared to the other enzymes. The human enzyme showed a marked improvement with the engineered IFNα2b sequon as well, but not to the same degree as the Drosophila enzyme. This suggests that substrate recognition by GT-27 orthologues with a folded protein substrate can be effectively manipulated through the ISOGlyP algorithm. With the large change in activity for the DmGalNAc-T2 enzyme on the modified IFNα2b-MUT, it would appear to be the enzyme of choice for the in vitro modification of a single sequon on that protein. In contrast to IFNα2b-MUT, the first de novo sequon design in hGH-MUT1 only showed moderate rate of glycosylation by all three enzymes. This fits with
what we observed in bacterial cells that the lower ISOGlyP scoring Thr (4.53 vs 17.59). This suggests the orientation of the first Thr residue must be such that it is poorly accessible. The subsequent sequon designs tested in hGH-MUT2 and hGH-MUT3 further demonstrate some correlation between glycosylation efficiency and ISOGlyP score across all orthologues, and DmGalNAc-T2 greatly surpasses both other enzymes in specific activity. Although ISOGlyP scores only reflect propensity for glycosylation by human GalNAc-T isoforms, the enhancement values for DmGalNAc-T2 reported in literature show an even stronger preference for proline at the -3, -1, +1, and +3 positions compared to hGalNAc-T2 (16). This offers some explanation towards the disproportionate improvement of DmGalNAc-T2 specific activity compared to hGalNAc-T2 or CeGalNAc-T2 despite hGH-MUT3 having a lower ISOGlyP score than hGH-MUT2. However, it must also be noted that the structural complexity of an intact protein substrate introduces additional factors influencing glycosylation efficiency which cannot entirely be accounted for by a peptide-based prediction tool. This was further reflected by changes in glycoform resolution via RP-HPLC, where mutating only a few amino acids introduced significant variability to the retention shift generated by a single GalNAc addition, ultimately requiring us to change our approach for measuring the rate of glycosylation for the hGH mutants. This suggests some rigidity with respect to orientation of the loop region containing the O-glycosylation sequon, thus influencing the overall surface hydrophobicity of the protein as well as accessibility of the modified threonine.

Of course, we have shown only two target protein examples here, and we are designing sequons in other cytokine-like proteins to probe their glycosylation efficacy both in vitro and in bacterial cells with the goal of establishing a versatile approach to producing human glycans on a variety of protein substrates. It will be essential for therapeutics with these sequon mutations to also maintain biological activity. We previously showed that the IFNα2b-MUT appears as active as the wild type (4), and are currently following up on the activity of the hGH variants described here. We are also looking into structural strategies for introducing multiple sequons to investigate serum half-life benefits from having multiple glycans per protein.

We have not yet tested many members of the GT-27 family, but we were successful with bovine GalNAc-T1 and hGalNAc-T2 (4), and now the *D. melanogaster* and *C. elegans* orthologues for hGalNAc-T2. It appears that the general folding of GT-27 enzymes can be efficient in *E. coli* with assistance from protein disulfide isomerases. We can generate enough
enzyme for detailed kinetic studies, and perhaps even structural studies if the MBP tag were cleaved off. We are currently planning to investigate the expression of other orthologues, GT-27 family members and sequon combinations in folded proteins. This will expand our access to these important protein modifying glycosyltransferases, and in combination with other glycosyltransferases permit the further engineering of bioactive O-glycans.

**Data availability**

All data will be made available upon request.

**Acknowledgments**

We would like to acknowledge funding from the Canadian Glycomics Network [http://dx.doi.org/10.13039/501100009056](http://dx.doi.org/10.13039/501100009056) project TP1, and the University of Alberta. We thank Dr. Lyann Sim, University of British Columbia for preparation of the GT-27 sequence alignment figure. Figure 2 was created with [BioRender.com](http://www.biorender.com).

**Competing interests**

No competing interests.

**References**

1. Walsh, G., and Jefferis, R. (2006) Post-translational modifications in the context of therapeutic proteins. *Nature Biotechnology* **24**, 1241-1252
2. Costa, M. B., Picon, P. D., Sander, G. B., Cuni, H. N., Silva, C. V., Meireles, R. P., Góes, A., Batoreu, N. M., Maia, M. L. S., Albuquerque, E. M., Matos, D. C. S., and Saura, P. L. (2018) Pharmacokinetics comparison of two pegylated interferon alfa formulations in healthy volunteers. *BMC pharmacology & toxicology* **19**, 1
3. Madsen, T. D., Hansen, L. H., Hintze, J., Ye, Z., Jelbari, S., Andersen, D. B., Joshi, H. J., Ju, T., Goetze, J. P., Martin, C., Rosenkilde, M. M., Holst, J. J., Kuhre, R. E., Goth, C. K., Vakhreshev, S. Y., and Schjoldager, K. T. (2020) An atlas of O-linked glycosylation on peptide hormones reveals diverse biological roles. *Nature Communications* **11**, 4033
4. Du, T., Buenbrazo, N., Kell, L., Rahmani, S., Sim, L., Withers, S. G., DeFrees, S., and Wakarchuk, W. (2019) A Bacterial Expression Platform for Production of Therapeutic Proteins Containing Human-like O-Linked Glycans. *Cell chemical biology* **26**, 203-212.e205
5. 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
6. Brockhausen, I., and Stanley, P. (2015) O-GalNAc Glycans. in Essentials of Glycobiology (Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. eds.), Cold Spring Harbor Laboratory Press

Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved., Cold Spring Harbor (NY). pp 113-123

7. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology 22, 736-756

8. 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

9. Ten Hagen, K. G., Tran, D. T., Gerken, T. A., Stein, D. S., and Zhang, Z. (2003) Functional characterization and expression analysis of members of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family from Drosophila melanogaster. J Biol Chem 278, 35039-35048

10. de Las Rivas, M., Lira-Navarrete, E., Gerken, T. A., and Hurtado-Guerrero, R. (2019) Polypeptide GalNAc-Ts: from redundancy to specificity. Current opinion in structural biology 56, 87-96

11. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42, D490-495

12. 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

13. Daniel, E. J. P., Las Rivas, M., Lira-Navarrete, E., García-García, A., Hurtado-Guerrero, R., Clausen, H., and Gerken, T. A. (2020) Ser and Thr acceptor preferences of the GalNAc-Ts vary among isoenzymes to modulate mucin-type O-glycosylation. Glycobiology 30, 910-922

14. Schjoldager, K. T., Joshi, H. J., Kong, Y., Goth, C. K., King, S. L., Wandall, H. H., Bennett, E. P., Vakhrushev, S. Y., and Clausen, H. (2015) Deconstruction of O-glycosylation--GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. EMBO Rep 16, 1713-1722

15. Fritz, T. A., Raman, J., and Tabak, L. A. (2006) Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-2. The Journal of biological chemistry 281, 8613-8619

16. 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

17. Pratt, M. R., Hang, H. C., Ten Hagen, K. G., Rarick, J., Gerken, T. A., Tabak, L. A., and Bertozzi, C. R. (2004) Deconvoluting the functions of polypeptide N-alpha-acetylgalactosaminyltransferase family members by glycopeptide substrate profiling. Chem Biol 11, 1009-1016

18. 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

19. 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

20. Abukar, T., Rahmani, S., Thompson, N. K., Antonescu, C. N., and Wakarchuk, W. W. (2021) Development of BODIPY labelled sialic acids as sialyltransferase substrates for direct detection of terminal galactose on N- and O-linked glycans. Carbohydr Res 500, 108249
21. Hagen, F. K., and Nehrke, K. (1998) cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase sequence homologs from Caenorhabditis elegans. J Biol Chem 273, 8268-8277

22. Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., Taly, J. F., and Notredame, C. (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res 39, W13-17

23. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42, W320-324

24. 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

25. Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggis, 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

26. Thomas, C., Moraga, I., Levin, D., Krutzik, P. O., Podoplelova, Y., Trejo, A., Lee, C., Yarden, G., Vleck, S. E., Glenn, J. S., Nolan, G. P., Pehler, J., Schreiber, G., and Garcia, K. C. (2011) Structural linkage between ligand discrimination and receptor activation by type I interferons. Cell 146, 621-632

27. 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

28. 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

29. Mohl, J. E., Gerken, T. A., and Leung, M.-Y. (2020) ISOGlyP: de novo prediction of isoform-specific mucin-type O-glycosylation. Glycobiology

30. Malekan, H., Fung, G., Thon, V., Khedri, Z., Yu, H., Qu, J., Li, Y., Ding, L., Lam, K. S., and Chen, X. (2013) One-pot multi-enzyme (OPME) chemoenzymatic synthesis of sialyl-Tn-MUC1 and sialyl-T-MUC1 glycopeptides containing natural or non-natural sialic acid. Bioorganic & medicinal chemistry 21, 4778-4785

31. Wandall, H. H., Hassan, H., Mirgorodskaya, E., Kristensen, a. K., Roepstorff, P., Bennett, E. P., Nielsen, P. a., Hollingsworth, M. a., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) Substrate specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. The Journal of biological chemistry 272, 23503-23514

32. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2011) Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. Glycobiology 22, 736-756

33. Hagen, F. K., and Nehrke, K. (1998) cDNA Cloning and Expression of a Family of UDP-N-acetyl-D-galactosamine:PolypeptideN-Acetylgalactosaminyltransferase Sequence Homologs from Caenorhabditis elegans. Journal of Biological Chemistry 273, 8268-8277
Rate Limiting Step

hGH \rightarrow \text{GalNAc-T} \rightarrow T \text{ synthase} \rightarrow \text{ST3Gal1}
Specific Activity (mU/mg)

- hGalNAC-T2
- DmGalNAC-T2
- CeGalNAC-T4

hGH-MUT1
hGH-MUT2
hGH-MUT3