A Comparative Study of GT27 Family *N*-acetylgalactosaminyltransferases

by

Biprajit Sanyal

B. Sc., University of Toronto, 2015

A thesis presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Science

in the program of

Molecular Science

Toronto, Ontario, Canada, 2019

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for scholarly research. I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for scholarly research.

I understand that my thesis may be made electronically available to the public.

A Comparative Study of GT27 Family *N*-acetylgalactosaminyltransferases

Biprajit Sanyal, 2019

Master of Science, Molecular Science

Ryerson University

Abstract

Therapeutic proteins face short half lives *in vivo*. Their high costs and associated toxicity effects of increasing dosage warrant exploration of methods to increase serum half-life. These proteins can be produced with native or engineered glycosylation sites, which has been shown to be an effective means of prolonging serum half-life. Engineered *E. coli* represents an economical route of production. I have been able to produce, purify and test the activity of three *N*-acetylgalactosaminyltransferase isoform 2 in *Escherichia coli* and show glycosylation on peptides derived from Interleukin 29. I followed the activity of these enzymes on three candidate therapeutic proteins via lectin blotting. Data suggest the *Homo sapiens* orthologue of GalNAcT2 is the most efficient enzyme in the *in vitro* assays with all candidate therapeutic protein substrates displaying the Tn-antigen. Future research should investigate continuous assays for precise results as well as assaying native peptides as opposed to non-native ones.

Acknowledgements

I would like to thank my parents and sister for their unconditional love and support through the years, and the years to come. They have somehow put up with me the longest out of anybody and I aspire for that level of strength and endurance.

I would like to thank my supervisor Dr. Warren Wakarchuk for all the years of guidance, advice, support, and most of all, his patience. Through the good and the bad, he has believed in me, forgiven me, and given me every chance to succeed. The opportunities and diversity of techniques I learned from working with him helped me become a way better scientist.

I would like to thank my committee members Dr. Sabatinos and Dr. Nitz for their help and advice throughout the course of my studies. I would also like to thank Dr. Steve Withers' group for their gifts of BODIPY-NHS throughout the entire course of my degree.

Special thanks to Dr. Ting Du for her guidance, her plasmids, her competent cells, and her lovely gifts through the years. I would also like to thank Dr. Bettina Janesch for her guidance, her ability to push oneself to do better, and the chocolates. Thank you to Hirak Saxena for the folding chaperone plasmid, always being there to bounce ideas off, and his raw and unforgiving honesty. I would certainly not have been able to do this without the training, advice, guidance, the raw honesty and emotional support from Nakita Buenbrazo and Laura Kell. Thank you for training me on pretty much all the equipment. Thank you for always lending a helping hand inside and outside the lab.

Furthermore, thanks to my friends for being the greatest distractions I could have ever asked for; even though I didn't ask.

Table of Contents

List of Tables

List of Figures

List of Appendices

Introduction

Current Challenges in Protein Therapeutics

Protein therapeutics represent a class of drugs which has high potential for effective treatment of many diseases and to improve the quality of lives for ailing patients worldwide. Examples of highly used therapeutic proteins include insulin for diabetes and clotting factors (Factor VII, VIII, and IX) for blood clotting disorders in addition to antibodies which are becoming widely used (Carter 2006). Many diseases could be cured by introducing recombinant proteins which have become useless as a result of mutation in the genome. Unfortunately, many therapeutic proteins are rapidly cleared as seen by Wang and Prueksaritanont (2010) and they present a model of how to predict systemic clearance. Current approaches for therapeutic protein production depend on protein engineering, and not all the proteins that could be therapeutic have been able to be produced as recombinant proteins. Once design and purification are complete and tested *in vitro*, therapeutic proteins face more challenges upon their introduction and testing *in vivo*. Therapeutics are often produced in non-human animals as seen in the example of insulin where the proinsulin gene is inserted into another organism such as *Saccharomyces cerevisiae* and then isolated and purified (Donner 2015). It is important to study a wide variety of recombinant therapeutic proteins which is made difficult due to the human system's unfortunately excellent immune system.

A significant hurdle in recombinant therapeutic protein application is their often short serum half-lives which can be seen for Fc-fusion proteins by Suzuki *et al.* (2010). Externally introduced therapeutic proteins are rapidly cleared *in vivo* thereby positing significant challenges in the treatment of many diseases as Gillies *et al.* (1993) noticed cytokine fusion proteins, although biologically active, were cleared rapidly. The low serum half-life of therapeutic

proteins makes oral administration of pharmaceuticals impossible especially considering the degradation of proteins in the stomach. This opens the route for more invasive drug delivery routes which are often not favourable in their own rights. Effective therapeutic protein production requires research from many sub-disciplines of biochemistry to see what strategies can be taken to improve their serum longevity. Even with therapeutic antibodies, there are still examples of recombinant antibodies having immunogenic responses as LoBuglio *et al.* (1989) observed one of their patients were reactive to the murine variable region of a chimeric antibody. Dobson *et al.* (2016) showed a method of removing self interaction by a monoclonal antibody by mutating three hydrophobic residues which also improved serum persistence. Therapeutic protein engineering represents a research topic where the protein itself is modified to make it more effective *in vivo*. The mutagenic strategy is only effective where the modified sites have no effect on the activity of the therapeutic protein. This may not be the case for all potential therapeutic proteins as alterations to amino acid residues can interfere with folding and stability. One of the solutions to rapid clearance of therapeutic proteins is to increase their dosage. Due to the nature of many drugs, cellular and organ toxicities can be a side-effect of increased dosage of drugs (O'Brien *et al.* 2006). Much research has been performed in trying to improve the protein structure overall and reduce non-specific tissue binding as reviewed by Vugmeyster *et al.* (2012). A suitable method of increasing the longevity of proteins is by glycosylating them (Saxena *et al.* 1997).

A common improvement of therapeutic protein half-life is the chemical addition of polyethylene glycol (PEG) polymers to the protein in a process known as PEG-ylation (Brygier *et al.* 1993). The benefits here include increasing the overall size of a protein to prevent its rapid clearance via renal filtration. PEG also increases the hydrophilicity of more hydrophobic

substances allowing them to be present and stored in solution (Wu *et al.* 2014). Furthermore, PEG has been shown to increase entry into cells in the cosmetic industry (Jang *et al.* 2015). Abuchowski *et al.* (1977) showed the addition of these polymers onto bovine catalase showing very low immunoreactivity, resistance to proteolytic degradation and increased circulating half life. Unfortunately, soon after this method was discovered Fisher (1978) reported instances of allergic reactions which suggest that there is a potential for immunogenicity in the event of repeated dosage of PEGylated proteins. Because many therapeutic proteins are PEGylated, it may be possible that there can be future immunogenic responses that may increase in severity over time. There is also potential that the very large PEG polymers can sterically hinder receptor activity of many therapeutic proteins depending on the amino acid location of PEG polymers (Grace *et al.* 2005). Furthermore, Zhang *et al.* (2011) reported in their study that there were toxicity effects upon the liver after analysis of PEG coated gold nanoparticles in mice. This accumulation in the liver caused increases in transaminases indicative of liver damage. Verhoef *et al.* (2014) also noticed the formation of anti-PEG antibodies *in vivo* further highlighting the potential of harmful immunogenic responses to externally produced therapeutic proteins. As effective and helpful PEGylation has been in healthcare there is still the need to explore alternative methods of increasing serum half-life of therapeutic proteins as it is important to have a toolbox to rapidly adapt to the production of therapeutic proteins for a variety of diseases.

New developments in therapeutic protein production represents a large growth opportunity for industry. This would hopefully increase treatments for both human and veterinary applications, and potentially lower the cost of dosage for these protein-based drugs. This is a huge industry and significant improvements can be made to improve treatment and the

cost of treatment, which can have a broad application to improvement of quality of life for people who need protein therapeutics.

Benefits of Glycosylation for Protein Targets

Glycosylation is a ubiquitous process involved in countless cellular process in all biological systems. Lack of successful glycosylation leads to embryonic lethality in mammals; an example of which can be seen in Chan *et al.* (2016) where models of disrupted phosphomannomutase 2 were created in mice, where many of the mice with mutated genotypes dying prenatally. Glycosylation disorders can be heritable as seen in GMPPA-congenital disorder of glycosylation (Koehler *et al.* 2013). Loss of function mutations in many of these enzymes, as seen in Freeze *et al.* (2012) review, can lead to disruptions in protein folding, growth factor signalling and pathogen binding. The importance of glycosylation in biology cannot be understated. Protein glycosylation refers to the process of covalently linking carbohydrate moieties to amino acid residues on the protein backbone. Protein glycosylation has been shown to occur mostly on the amino acid asparagine for *N*-linked glycans, or onto serine/threonine for *O*-linked glycans (**Figure 1**).

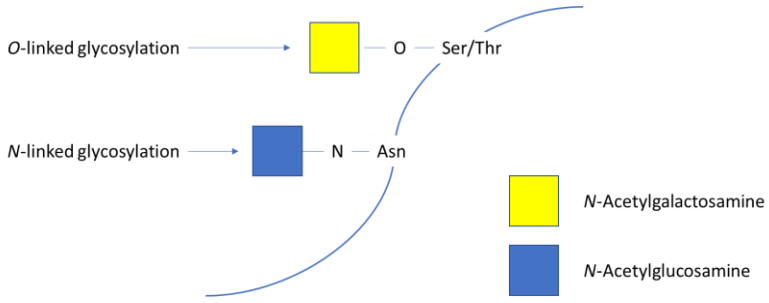


Figure 1. Types of Glycosylation: A representative diagram showing the first sugar additions on a sample protein backbone indicating the first steps in *N*-linked and *O*-linked glycosylation.

Glycosylation of proteins has a critical function in cellular recognition, signaling, viability, and quality control of the proteins in the endoplasmic reticulum. The presence of extracellular glycans also facilitates motility and adhesion (Arora *et al.* 2005). Diversity in glycans can lead to many implications in the function of the glycosylated protein due to potential changes in adhesion as seen by Asada *et al.* (1997) where increased branched glycans were correlated with malignant cancer phenotypes. The exploration of glycosyltransferases is therefore highly beneficial to elucidate functional relationships between proteins and glycans. This is important as there is evidence in literature highlighting species specific glycosylation patterns for proteins (Raju *et al.* 2000).

Current research being performed on the pharmacological properties of therapeutic proteins shows the presence of glycan chains to have positive effects on protein stability and serum half-lives which has been a trend observed in many reviews of the field detailed in Solá and Grebenow (2010) in *BioDrugs*. A therapeutic protein once introduced to a human system instantly faces the danger of non-specific proteases and bulk clearance. In Solà's 2009 review, it was proposed that the reactive side chains of proteins themselves can potentially make them more sensitive to broad scale non-specific degradation. The protection of enzymes from degradation via glycosylation was observed early as 1975 (Vegarud and Christensen 1975). The presence of glycans on the proteins alone can reduce proteolysis due to sterically hindering proteases.

Ngantung *et al.* (2006) presented a summary of glycoproteins and compared their sialylated and desialylated forms (sialic acid being the primary terminal glycan) and found the presence of the terminal sialic acid can positively influence the circulating half life of

glycoproteins by significant amounts of time where glycoproteins lacking terminal sialylation being removed in the liver via asialoglycoprotein receptors (Roggenbuck *et al.* 2012).

Glycosylation of recombinant therapeutic proteins is primarily performed in eukaryotic cell systems with many therapeutic proteins themselves being isolated from human plasma fractionation or expressed in Chinese Hamster Ovary cells (Solà and Griebenow 2009). This is highly time consuming and expensive with many considerations needing to occur to ensure proper protein production while simultaneously creating a positive glycosylation environment for therapeutic proteins. Although highly effective and replicatetd throughout literature, eukaryotic protein production systems are expensive, and maintenance of eukaryotic cell lines may be an economic bottleneck for many companies; with Lagasse *et al.* (2017) describing produced therapeutic proteins as some of the most expensive on the market. It is therefore necessary to generate a system which can produce therapeutic glycoproteins quickly and economically.

Production of Therapeutic Proteins and Glycosylation Enzymes in Lower Complexity Organisms

Many therapeutic proteins are currently produced in *Escherichia coli*, and Baeshen *et al.* (2015) estimates that around 30% of therapeutic proteins are produced in *E. coli*. One of the major challenges in clinical efficacy of recombinant *E. coli* therapeutic proteins is the lack of human like glycosylation patterns. The exploration of simultaneous production and glycosylation of therapeutic proteins by recombinant glycosyltransferases produced in *E. coli* represents an excellent potential way to produce a variety of therapeutic proteins.

Enzymes capable of glycosylating proteins are widely known in literature with many studies focusing on specificity (Elhammer *et al.* 1993; Gerken *et al.* 2004). In earlier work from

our lab, Lindhout *et al.* (2011) presented a model showing bacterial systems can produce enzymes which can glycosylate therapeutic protein candidates resulting in improved pharmacokinetic profiles of the therapeutic proteins. In that work, they were able to obtain different sources of glycosyltransferases and produce them in *E. coli.* These recombinant glycosyltransferases could successfully extend existing *N*-glycans on alpha-1-antitrypsin (A1AT) obtained commercially. This work showed that it is possible for recombinant glycosyltransferases to successfully manipulate the glycans on therapeutic proteins *in vitro* and improve the pharmacokinetics of that protein.

The bigger challenge is producing both the enzymes and target therapeutic proteins in *E. coli*. Folding difficulties arise for eukaryotic proteins produced in prokaryotic cells and special considerations are needed when producing complex eukaryotic enzymes in *E. coli* and other prokaryotic model organisms. The folding challenges are due to the cytoplasm of *E. coli* being primarily a reducing environment (Stewart *et al.* 1998). New England Biolabs (NEB) SHuffle series of competent *E. coli* strains contain favourable mutations to produce properly folded proteins. These include the deletion of membrane proteases like OmpT, thioredoxin reductase (*tox*), glutathione reductase (*gor*), and a leaderless DsbC gene (Lobstein *et al.* 2012). DsbC is a prokaryotic disulphide bond isomerase where the wildtype location is at the periplasm (Zhuo *et al.* 2014), while the SHuffle phenotype retains this protein in the cytosol.

Enzymes are able to be produced when their genes are expressed alongside additional proteins such as folding chaperones. A soluble form of human protein disulphide isomerase (hPDI) expressed in *E. coli* in was discussed by Alanen *et al.* (2003). This isomerase has been used with success and has greatly improved production of proteins in *E. coli* cytoplasm (Lauber *et al.* 2015; Nguyen *et al.* 2011). Nguyen *et al.* (2011) in *Microbial Cell Factories* reported the

use of human Protein Disulphide Isomerase (hPDI) to assist folding of proteins and another group was able to do this as well in minimal media (Gaciarz *et al.* 2017).

Lauber *et al.* (2015) successfully produced active UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase 2 (GalNAcT2) in *E. coli* in experiments where they used the hPDI described by Nguyen *et al.* (2011). The GalNAcT enzymes are crucial in initiating the *O*-glycosylation process by the addition of a α-GalNAc moiety on to a serine or threonine residue. This process occurs in the Golgi apparatus and is a common post translational modification before secretion of proteins (Fleischer 1981). Lauber *et al.* (2015) produced and purified this enzyme and could verify its activity on many sample peptides. These findings suggest that *E. coli* systems can produce glycosyltransferase enzymes which have activity on varying substrates. This work also showed the capacity of multiple plasmid expression to help the formation of a glycosyltransferase.

Polypeptide *N*-acetylgalactosaminyltransferases

The ppGalNAcT2 enzymes, belonging to the GT-27 family catalyze the covalent addition of a α-GalNAc moiety from the donor molecule UDP-α-D-GalNAc on to a serine or threonine residue generating the structure know as the Tn antigen. This α-GalNAc residue is at the heart of several O-glycan core structures reviewed extensively by Bennett *et al.* (2012). This initial addition of GalNAc is the first step of *O*-glycosylation which can then be elongated via several other glycosyltransferases and terminally sialylated. The glycosyltransferase mechanism employed by this protein is a retaining mechanism for the anomeric carbon leaving the alpha configuration in tact. The ppGalNAcT2 enzyme is a conserved enzyme across a variety of diverse complexity eukaryotic organisms indicating how important the requirement of the O-glycan is. It is a broad-spectrum glycosyltransferase with multiple substrate specificities. The

ideal acceptor peptide sequence was found to be Pro-Gly-Pro-Thr-Pro-Gly-Pro (Gerken *et al.* 2006). Proline residues proximal to the glycosylation site was most favourable for glycosylation (Nehrke *et al.* 1996). Furthermore, Fritz *et al.* (2006) in their analysis of the ppGalNAcT2 crystal structure showed a non-polar association of a proline residue three residues down the acceptor threonine to Trp282 in the core GalNAcT2 structure. This creates a finger-like grasp of the peptide onto the enzyme. The activity of this enzyme was observed to be hindered when there were large aromatic side chains flanking the serine or threonine to be glycosylated (O'Connell *et al.* 1992). The ppGalNAc-T2 protein has distinct catalytic and lectin domains (**Figure 2**). For ppGalNAc-T2, the lectin domain does not affect catalytic activity on non-glycosylated peptides (Fritz *et al.* 2006 and Wandall *et al.* 2007). Taus *et al.* (2014) also reported no significant influence of glycosylation by the lectin domain in the *Biomphalaria glabrata* ppGalNAcT2 enzyme. This would allow for this domain to be excluded when designing a recombinant construct for production in *E. coli* as this would remove one disulfide bond and an unnecessary domain. It is therefore important to produce different GalNAcT2 proteins in *E. coli* to determine which one will be best for our future strain engineering work.

Three sources of GalNAcT2 enzymes were chosen for comparison in this work the sources being genes encoding for GALNT2 (*Homo sapiens*), PGANT2 (*Drosophila melanogaster*), and Gly4 (*C. elegans*). These enzymes code for proteins similar in size and their sequence similarities suggest structural similarity making them excellent candidates for comparison (**Figure 3**). One of the key features being conserved across the candidate enzymes are the conserved cysteines. Furthermore, the locations of the lectin domains are conserved as well. The *D. melanogaster* isoform has 66% similarity to the human one while the *C. elegans* isoform has 55% similarity to the human one. By assaying activity of these enzymes on common

recombinant therapeutic proteins and synthetic peptide candidates, it can be seen which of these three enzymes are most broadly reactive and suitable for the contribution of the enzyme to a model where therapeutic proteins are produced and glycosylated in *E. coli*. This is the first comparison of various GalNAcT2 enzymes being produced from *E. coli* and their activity being compared on pharmacologically relevant substrates. Previous studies have been performed on ppGalNAcT enzymes on their mucin derived peptides and computationally derived random ones (Gerken *et al.* 2006), but few explore their potential activities on therapeutic protein sites or sites not normally glycosylated in the proteins.

Substrates for GalNAcT2

Many publications concerning ppGalNAcTs focus on their activity on mucin derived peptides (Fritz *et al.* 2006; Taus *et al.* 2014; Lauber *et al.* 2015). Although this would be excellent to gauge the general activity of these transferases, the substrate tolerance of ppGalNAcT2 activity would not be able to be elucidated by this method. It is therefore beneficial to observe the activity of these enzymes on peptides derived from therapeutic protein candidates. In our lab, two test proteins are being examined for *in vivo* glycosylation in *E. coli*. Interferon α-2b and a next generation therapeutic to replace IFNα2b, IL-29 (Uze and Monneron 2007). In support of this project, I have been investigating glycosylation of peptides derived from native IL-29.

IL-29 belongs to a group of molecules known as cytokines which upon binding to their respective receptor, will cause gene expression leading to increased production of immune system effectors in the host. An example of the use of IL-29 is in Hepatitis C (Pagliaccetti *et al.* 2008). IL-29 itself is effective in combatting viral infections with recent publications indicating its role in treatments of cancer (Witte *et al*. 2010). IL-29 does have a single *N*-glycosylation site

however it does not have *O*-glycosylation. As the crystal structure of this cytokine is known (Miknis *et al.* 2010), along with the binding behaviour, it is possible to design glycosylation sites without interfering with the receptor binding activity of the cytokine (**Figure 4**). The longer-term goal of our lab's effort is to produce an engineered version of cytokine proteins, that will contain fully elongated O-glycans to improved serum half-life. The scope of my work will focus on the addition of only the initiating α-GalNAc moiety on peptide candidates. Because IL29's peptide backbone is not commonly found to be *O*-glycosylated, observation of the substrate kinetic parameters of IL-29 derived peptides on the recombinant GalNAcT2 will inform the effectiveness of the recombinant glycosyltransferases I produced are.

Assaying glycosyltransferase activity on peptides are facilitated by labelling with 7-[3-[(2,5-Dioxopyrrolidin-1-yl)oxy]-3-oxopropyl]-5,5-difluoro-1,3-dimethyl-5H-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-4-ium-5-uide (BODIPY-NHS). Peptides are labelled by BODIPY(BDP) through the BDP-NHS ester allowing high quantum yields with high brightness (**Figure 5**). The difference between maximal absorption and emission spectra is incredibly small for BDP making its Stokes shift favourable for analysis with multiple other dyes. Polarity changes associated to the addition of a sugar molecule are used to track reaction progress using quantitative methods.

Along with peptide candidates, therapeutic protein candidates are explored as well. Because glycosyltransferases are normally active on fully folded polypeptides, it needs to be seen if the recombinant glycosyltransferases produced in this work are capable of glycosylating entire polypeptides produced from our *E. coli* expression system. Chosen candidates are a synthetic IL29 protein that has been named IL29-3G due to three engineered glycosylation sites being present, IFNα2B, and hGH. All three candidate proteins have a GB1 fusion at the *N*-terminal of the protein

as a solubility aid for these proteins in *E. coli* (Du *et al.* submitted). These substrates are detailed in **Table 1.** The enzymes themselves contain maltose binding protein (MBP) fusions at the *N*-terminus as a solubility aid as the use of the to improve solubility along with simplifying purification has been seen in literature (Sun *et al.* 2011).

The sugar nucleotide UDP-α-D-GalNAc is very expensive, so we have been exploring coupled reactions with the C4-epimeric sugar nucleotide UDP-α-D-GlcNAc. Our lab developed a C-4 hexose/hexNAc epimerases Bernatchez *et al.* (2005) named CPG-13, which can be used in coupled reactions with the transferase. UDP-α-D-GlcNAc is a highly abundant in normal *E. coli* where one of its functions is to assist in the formation of peptidoglycan. UDP-α-D-GalNAc is not normally present in *E. coli*, and so because we ultimately want to see how the ppGalNAcT enzymes perform *in vivo*, we wanted to make sure they could function in coupled assays with the normal UDP-α-D-GlcNAc donor and the C4 UDP-Hex/HexNAc epimerase.

Objectives and Scope of this Work

In this work, I aim to compare the activity of three recombinant glycosyltransferases from unique source organisms expressed in *E. coli*. After successfully purifying them, I tested their activity on two fluorescently labelled IL-29 derived peptides. The enzymes were expressed and grown in the presence of another plasmid containing a gene for hPDI which is known in literature to improve folding of proteins and theoretically the activity of enzymes. The cooperativity of these enzymes was also explored alongside epimerase enzymes in coupled reactions.

Quantitative substrate kinetics were performed using high performance liquid chromatography (HPLC) methods with C-18 reverse phase column chromatography being used to separate glycosylated and non-glycosylated peptides to track reaction progress. A qualitative

exploration of the activity of the enzymes studied here were performed as well using Soy Bean Agglutinin (SBA) lectin binding blots on proteins subjected to our glycosyltransferases. This will show how effective recombinant glycosyltransferase enzymes would be in glycosylating full size therapeutic protein candidates.

I aim to contribute a glycosyltransferase enzyme to the formation of a prokaryotic glycosylation system which can simultaneously generate a therapeutic protein while also glycosylating them all while every protein is being folded properly due to the presence of hPDI. It is my hope that the glycosyltransferase enzyme can glycosylate proteins in the same system as downstream glycosyltransferases. This will allow the identification of which of the three candidate GalNAcT2 enzymes are most suitable for expression in *E. coli*, glycosylating non-native O-glycosylation sequences and activity on target proteins derived from expression in *E. coli*. Hopefully, these multiple analyses of glycosyltransferases will give new insight on enzyme activity on specific therapeutic protein substrates to further develop methods for human like *O*-glycosylation on pharmacologically relevant proteins.

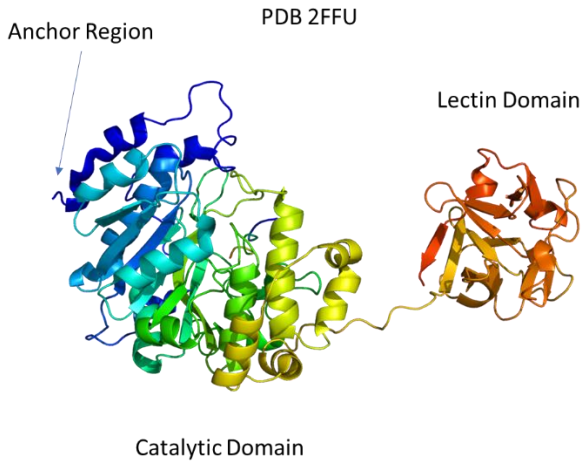


Figure 2. Crystal Structure of ppGalNAcT2: A PyMol rendition of GalNAcT2 based on crystal structure (PDB:2FFU) submitted by Fritz *et al.* (2006). Shown are catalytic and lectin domains of the enzyme sourced from *Homo sapiens*.

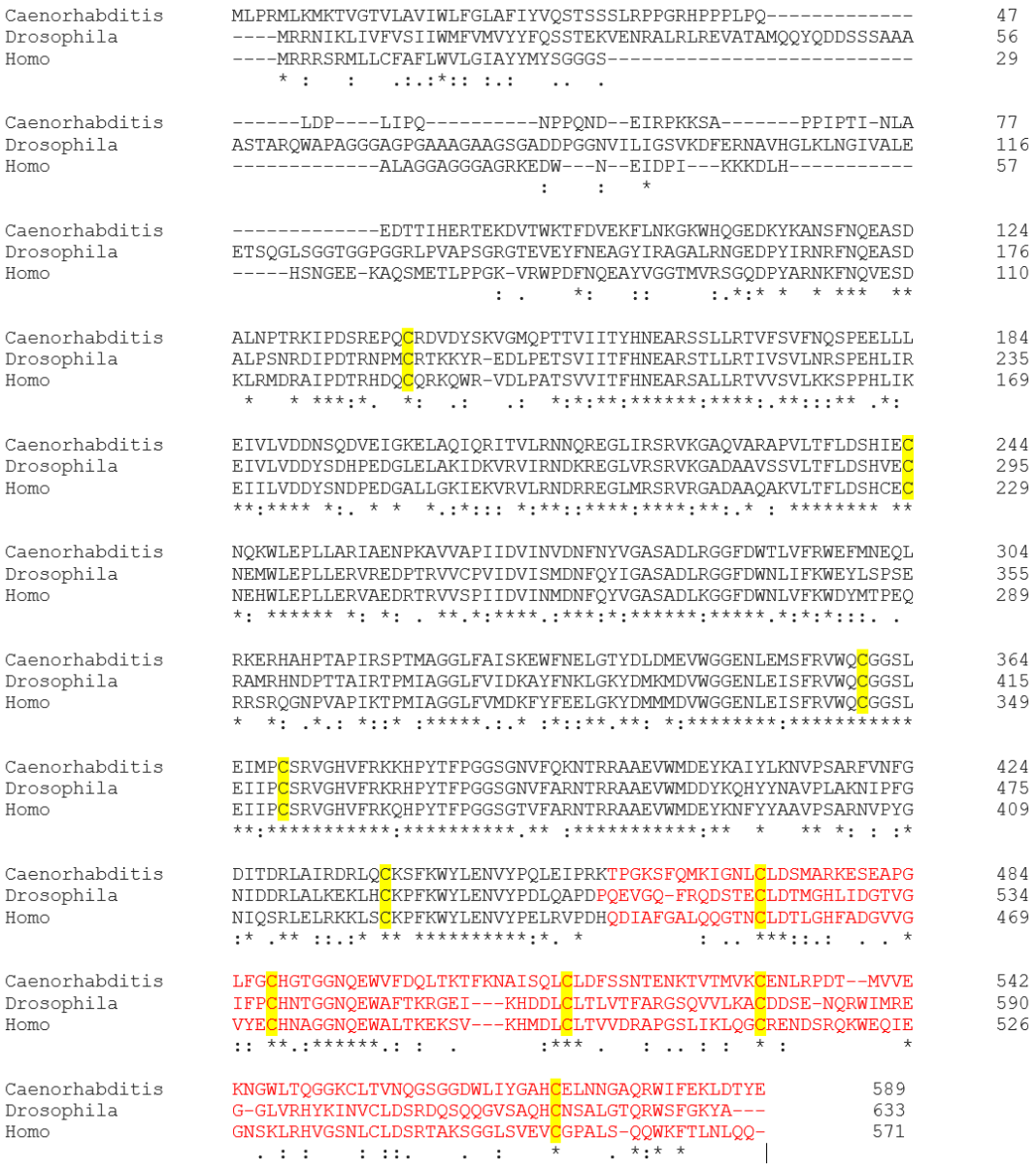


Figure 3. Sequence Alignment of GalNAcT2 Isoforms: Sequence alignment of GalNAcT2 from *Homo sapiens* (UniProt: Q10471), *Drosophila melanogaster* (UniProt: Q6WV19), and *Caenorhabditis elegans* (UniProt: Q8I136) generated using Clustal Omega. Highlighted in yellow are conserved cysteines while red outlines represent the lectin domain.

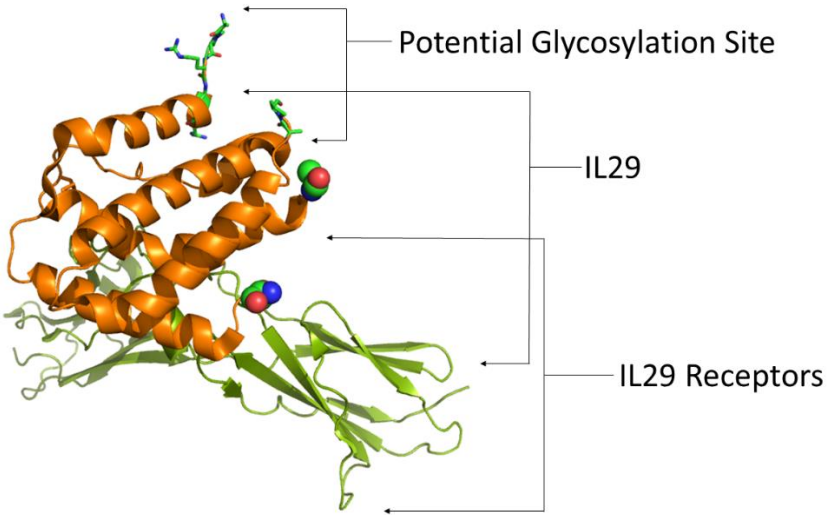


Figure 4. IL29 Crystal Structure: A figure showing potential glycosylation site of IL29 and the bases for the $IL29_1$ and $IL29_2$ peptides. Structure based on work by Miknis *et al.* (2010). Image generated using PyMol based on PDB file 3OG4

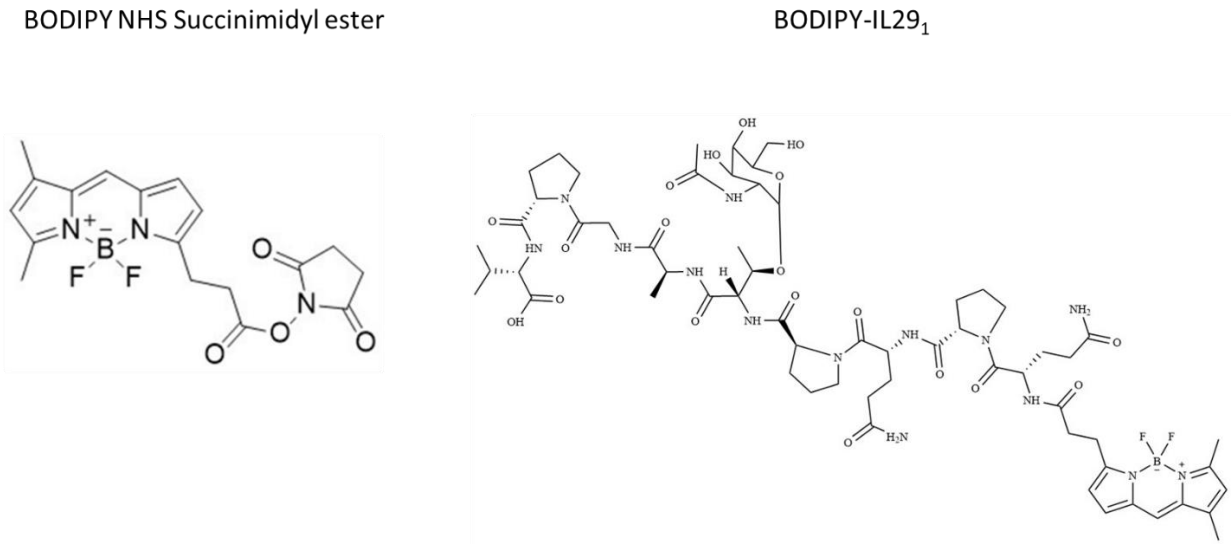


Figure 5. BODIPY Peptide Labelling: Substrates involved in reactions with GalNAcT2. Structure on the left represents the free BODIPY NHS probe in its succinimidyl ester form while the structure on the right represents the probe bound to synthetic $IL29_1$ peptide.

Table 1: List of Substrates and Sequences in this Work with Highlighted Glycosylation Sites and Fusion Partner Indicators (Red)

Name	Sequence
$IL29_1$	QPQP**T**AGPV
$IL29_2$	GPVP**T**SQPT

GB1-IL29-3G	MSGSHHHHHHGMQYKLALNG KTLKGETTTE AVDAATAEKV FKQYANDNGV DGEWTYDDAT KTFTVTEPGG PASENLYFQG SGP**T**P**T**SQPT INTAGCHIGR FKSLSPQELA SFKKARDALE ESLKLKNWSC SSPVFPGNWD LRLLQVRERP VALEAELALT LKVLEAAAGP ALEDVLDQPL HTLHHILSQL QACIQP**T**PTI NTVPVGRLHH WLHRLQEAPK KESAGCLEAS VTFNLFRLLT RDLKYVADGD LCLGVTETPE ST
GB1-IFNα2B	MSGSHHHHHH GMQYKLALNG KTLKGETTTE AVDAATAEKV FKQYANDNGV DGEWTYDDAT KTFTVTEPGG PASENLYFQG SCDLPQTHSL GSRRTLMLLA QMRRISLFSC LKDRHDFGFP QEEFGNQFQK AETIPVLHEM IQQIFNLFST KDSSAAWDET LLDKFYTELY QQLNDLEACV IQGVGV**T**ETP LMKEDSILAV RKYFQRITLY LKEKKYSPCA WEVVRAEIMR SFSLSTNLQE SLRSKE
GB1-hGH	MSGSHHHHHH GMQYKLALNG KTLKGETTTE AVDAATAEKV FKQYANDNGV DGEWTYDDAT KTFTVTEPGG PASENLYFQG SFPTIPLSRL FDNAMLRAHR LHQLAFDTYQ EFEEAYIPKE QKYSFLQNPQ TSLCFSESIP **T**PSNREETQQ KSNLELLRIS LLLIQSWLEP VQFLRSVFAN SLVYGASDSN VYDLLKDLEE GIQTLMGRLE DGSPTINTIF KQTYSKFDTN SHNDDALLKN YGLLYCFRKD MDKVETFLRI VQCRSVEGSC GF
GB1 Fusion Partner	MSGSHHHHHH GMQYKLALNG KTLKGETTTE AVDAATAEKV FKQYANDNGV DGEWTYDDAT KTFTVTEPGG PASENLYFQG

Methods

Selection of *N*-acetylgalactosaminyltransferases isoform 2 (GalNAcT2)

Enzyme sequences available from UniProt were analyzed for *H. sapiens* ppGalNAcT2 (UniProt: Q10471), *D. melanogaster* PGANT2 (UniProt: Q6WV19), and *C. elegans* Gly4 (Q8I136) were aligned using ClustalW. Sequence similarities allowed the removal of lectin domains and signal peptide sequences. The *D. melanogaster* isoform was 66% similar to the *H. sapiens* one while the *C. elegans* isoform was 55% similar to the *H. sapiens* isoform. Synthetic genes of ppGalNAcT2, PGANT2, and Gly4 were ordered from amino acids 51-447, 89-493, and 65-461 respectively. To facilitate gene cloning, synthetic genes were ordered from BIOBasic (Markham, Ontario, Canada) and these genes were obtained in pUC57 vectors with NdeI and SalI as the restriction sites.

Creation of HGT-12 Strain for Baseline Testing

A GalNAcT2 fragment with its lectin domain existing as a Maltose Binding Protein (MBP) fusion was excised from its source OGO-6 (**Appendix 1**) using restriction enzymes BamHI and XbaI obtained from New England Biolabs Canada and inserted into plasmid pCW (**Appendix 2**) to create the HGT-12 plasmid (**Appendix 3**). DNA solutions were obtained using plasmid mini-prep kits from Sigma Aldrich and instructions were used as supplied by the manufacturer. Fragment and vector was ligated using T4-DNA ligase (NEB Canada) using instructions specified by the manufacturer. Fully ligated plasmid was verified by sequencing at The Hospital for Sick Children (Toronto, Ontario, Canada) using primers listed in **Table 3**. Verified plasmid was transformed into electrocompetent SHuffle Express (NEB Canada). A

glycerol stock with SHuffle Express containing the HGT-12 plasmid was created at 10% glycerol and stored at -80 ^{o}C.

Preparation and Purification of Fluorescent Substrate for Enzymatic Assays

Non-native peptide sequences were obtained for the IL-29 cytokine derived from its loop occurring from amino acids 139 to 147 based on its PDB structure 3OG6. Peptides QPQPTAGPV (894 g/mol) and GPVPTSQPT (883 g/mol) were commercially synthesized from Bio Basic Inc. (Markham, Ontario, Canada). These peptides were labelled IL-29$_1$ and IL29$_2$ respectively. Approximately 2.2 µmol of peptides were dissolved in DMF (Sigma Aldrich).

BODIPY-NHS FL succinimidyl ester (MW 389.16 g/mol) was obtained from the Withers Lab (Vancouver, British Columbia, Canada) and was also dissolved in DMF. 1.8 mg of peptide and 2.0 mg of the succinimidyl ester were combined in a 200 µL reaction with 25 mM of Sodium Borate pH 8.0. The reaction occurred for two hours at room temperature and then overnight at 4 ^{o}C while being protected from light. Reaction progress was monitored by Thin Layer Chromatography (TLC) on silica gel plates (EMD Millipore Corporation, Billerica, MA, USA) and developed in 4:2:1:0.1 solvent containing ethyl acetate: methanol: water: acetic acid respectively.

Purification of the BODIPY labelled peptides occurred via size exclusion chromatography using Superdex peptide© resin (GE Healthcare) packed into a 74 mL column at 0.5 mL/min with absorbance detection at 504 nm. Reactions were diluted to 500 µL using column buffer consisting of 50 mM ammonium bicarbonate pH 6.0 and 25% acetonitrile (ACN) and loaded onto the gel filtration column using an AKTA system (GE Healthcare). Complete separation of labelled peptide and free dye components were verified using qualitative TLC analysis (**Figure 6**). Labelled

fractions were concentrated using the LabConco CentriVap system spinning overnight at room temperature and stored at -20 ^{o}C until used.

Crude Activity Assays of OGO-6

OGO-6 is a Wakarchuk lab plasmid containing sequence coding for a maltose binding protein fusion of ppGalNAcT2, a C-4 hexose/hexNAc epimerase previously characterized in Bernatchez *et al.* (2005), and a β1,3-galactosyltransferase Core-1 synthase enzyme (CgtB) also characterized by the Wakarchuk group (Bernatchez *et al.* 2007). This plasmid also contains a coded leaderless DsbC sequence. Plates of SHuffle Express *E. coli* expressing OGO-6 were made using selection via chloramphenicol, then were subjected to plasmid extraction using a mini-prep kit, upon which, the plasmid obtained was transformed into electrocompetent Shuffle Express. Overnight cultures in 125 mL flasks with 30 mL of 2YT media were created at 30 ^{o}C and then 5 mL was transferred into a 0.5L 2YT culture supplemented with chloramphenicol (30 µg/mL) and 0.2 % glucose grown at 30 ^{o}C. The growth of the cells was followed by observing the optical density at 600 nm every hour until 0.4 OD was reached upon which 0.5 mM IPTG was added and growing temperature was reduced to 20 ^{o}C. The cells were harvested by centrifugation at 7500 x *g* in a Fiberlite F14-6x250y rotor (Thermo Fisher Scientific). Obtained pellets were mechanically lysed by grinding in a cooled mortar and pestle using Celite (diatomaceous earth Sigma-Aldrich). Lysate was resuspended in 20 mM Tris pH 7.5, 200 mM NaCl, and 1 mM EDTA (Buffer A);EDTA-free protease inhibitor (Sigma-Aldrich) was also included in the mixture along with DNAase I. Solution was then transferred to 50 mL Oakridge tubes and centrifuged at 24, 000 x *g* in a Fiberlite F18-12x50 rotor with temperature being maintained at 4 ^{o}C. The supernatant was then filtered through a 0.22 µm membrane. A reaction was set up with the following components: 5 µL of crude lysate in a 10 µL reaction with 20 mM Tris pH 7.5, 10 mM $MnCl_2$, 0.1 mM BDP-

peptide, and 1.5 mM UDP-GlcNAc. Reactions were allowed to proceed at hour intervals whereupon 2 µL of the mixture were stopped with equal amounts of 50% ACN and 10 mM EDTA (stopping solution). Stopped reaction mixtures were spotted on TLC plates and developed in 4:2:1:0.1 solvent containing ethyl acetate: methanol: water: acetic acid respectively. CgtB was tested in a 10µL reaction with the following conditions: 50 mM HEPES pH 7.0, 10 mM $MnCl_2$, 1 mM UDP-Gal, 0.2% Triton X100, and 0.5 mM BDP-GalNAc. Reaction was started upon the addition of 4µL of crude extract and stopped using stopping solution. Controls were generated by exclusion of crude cell extract.

Gel Electrophoresis

15% polyacrylamide gels of 1 mm thickness were made manually using templates and equipment from BioRAD. The lower gel was created in ten millilitres providing enough solution for two gels. The mixture consisted of 3.5 mL of deionized water, 2.7 mL of 4X 0.5% SDS containing Tris-HCl buffered at pH 8.8, 3.75 mL of 30% Bis:Tris Acrylamide (BioShop, Burlington, ON, Canada), 80 µL of 10% APS (BioShop, Burlington, ON, Canada), and 5 µL of 1, 2-Bis(dimethylamino)ethane (TEMED; BioBasic, Markham, ON, Canada). After polymerization the stacking layer was created with 1.6 mL deionized water, 665 µL of 4X 0.5% SDS containing Tris-HCl buffered at pH 6.8, 325 µL of Bis:Tris Acrylamide, 40 µL of 10% APS, and 5 µL of TEMED. 1 mm 15 well combs were placed in the solution and the gel was allowed to polymerize.

Protein samples were diluted 1:10 and denatured in SDS loading dye after heating at 90^{o}C for ten minutes. SDS loading dye solution was created as a 4X stock with 200 mM Tris-Cl pH 6.8, 400 mM DTT, 0.4% bromophenol blue, and 40% glycerol. 10 µL samples were loaded with 2 µL of BioRAD's Precision Plus dual colour standards. Samples were run using BioRAD's gel

equipment systems and power sources. Samples were run for 10 minutes at 120V and subsequently at 160V for 50 minutes. Upon completion Coomassie solution was poured over the finished gels and allowed to stain for an overnight period. Gels were destained using a solution of 10% methanol and 10% acetic acid and imaged using the BioRAD GelDoc system and ImageLab software.

Growth of HGT-13 and HGT-14

The human GalNAcT2 isoform was inserted into two expression plasmids. One plasmid pCWMalET (**Appendix 4**) was developed previously by the Wakarchuk lab and the other is commercially available pMalC5X (**Appendix 5**) from New England Biolabs (NEB). HGT-13 (**Appendix 6**) refers to the protein produced from the pCWMalET plasmid with a 17-amino acid linker region cleavable by thrombin while HGT-14 (**Appendix 7**) refers to the protein produced from the pMalC5X containing a 23-amino acid linker region cleavable by Factor Xa (vector and construct information available in **Table 4** and **Table 5**). Glycerol stocks of these strains in SHuffle Express (NEB) were streaked on ampicillin resistant plates and grown overnight at 30 ^{o}C as per recommended growing conditions for the SHuffle Express strain. 35 mL overnight cultures were grown at 30 ^{o}C overnight containing 150 µM of ampicillin. 5 mL of the culture was transferred to 500 mL 2YT broth supplemented with 0.2% glucose and 150 µM of ampicillin and grown until OD 0.4 whereupon induction with 0.2 mM IPTG was performed and incubation temperature reduced to 20 ^{o}C. The twenty-degree temperature incubation occurred overnight, and cultures were centrifuged at 7500 x *g* for 15 minutes at 4 degrees and pellets were stored in the -20 ^{o}C freezer until usage. 1 mL samples were taken from cultures before and after induction to assess successful induction of protein. These samples were centrifuged at 24, 000 x *g* for 1 minute and pellets were resuspended in 50 µL of MilliQ water for the pre-induction pellets and 500 µL of MilliQ water for post induction. 20 µL of these solutions were added to 10

µL of loading dye and 10 µL of water and then boiled at 95 ^{o}C for ten minutes. These samples were loaded on to SDS gels and subject to 120V for ten minutes to ensure sample uniformly passed the stacking gel. Gels were then subjected to 160V of current for 50 minutes. Gels were run using a Bio Rad PowerPac^{TM} Basic Power Supply system.

Purification and Activity Assays of HGT-13 and HGT-14

Cells were harvested and lysed as described in the OGO-6 protocol and supernatant was loaded on pre-equilibrated GE Healthcare 5mL MBP-Trap columns. MBP-Trap columns were preequilibrated with 5 column volumes of Buffer A as described above. Purification was facilitated using the AKTA system at 4 ^{o}C at a flow rate of 5 mL/min. Column was washed with 10 column volumes of Buffer A and proteins were eluted with Buffer A with the addition of 10 mM Maltose (Sigma Aldrich) (Buffer B). Sizes and purity of protein preparations were verified using SDS gel electrophoresis of fractions just after introduction of maltose.

Enzymes were assayed in 10 µL with the following conditions: 20 mM Tris pH 7.5, 10 mM $MnCl_2$, 1 mM UDP-GalNAc, and 0.1 mM BDP-IL29$_1$. Reactions were started upon the addition of 2 µL of the enzyme mixture resulting in a 0.1 µg/µL enzyme concentration. Control reactions were made with the same conditions while excluding UDP-GalNAc. Reactions were stopped at one and two-hour time intervals with 1:1 ratio of reaction to 50% ACN and 10 mM EDTA solution (Stopping Solution). Reactions were spotted on TLC plates and resolved using 4:2:1:0.1 solvent. Reactions were then visualized through the blue tray on the BioRad GelDoc. Quantification of proteins were performed after initial activity testing using Thermo Fisher's BCA kit and instructions were followed as supplied by the manufacturer using Bovine Serum Albumin (Bio-Shop) as the standard.

Growth Condition Optimization and Testing of Recombinant *Drosophila melanogaster* and *Caenorhabditis elegans* GalNAcT2

Glycerol stocks for the enzymes from *D. melanogaster* and *C. elegans* expressed in pCWMalET with SHuffle Express as the host strain were streaked on Ampicillin agar plates (150 µg/µL) and grown at 30 ^{o}C and were then transferred into overnight cultures. 1 mL of the overnight cultures were transferred to 100 mL 2YT cultures supplemented with 0.2% glucose and 150 µg/µL ampicillin. Independent cultures were created with one set being dedicated to being uninduced with the other being induced. Cultures were grown at 30 ^{o}C until OD ~0.4 and then induced with 0.2 mM IPTG and one batch was incubated at 16 ^{o}C while another batch was incubated at 30 ^{o}C. Control cultures were also moved over to the altered expression temperatures as well. Induction was verified via SDS-PAGE. Further growth and inductions were also done with 20 ^{o}C being the expression temperature. These generated pellets were then crudely assayed using the same procedure as the OGO-6 crude lysate analysis with the reaction conditions from the HGT-13 and HGT-14 analysis as above. Further DGT-100 and CGT-100 pellets were obtained, purified and tested under the same conditions as HGT-13.

Verification of MBP Fusion Protein Presence using Western Blotting

Purified protein samples were loaded onto an SDS gel and membranes were transferred onto PVDF activated in methanol. Transfer occurred over 90 minutes in a Bio Rad system at 100 V while maintaining the system at 4 ^{o}C. Membranes were blocked with 5% skim milk overnight and antibody binding was performed using 0.02 µg/µL anti-MBP antibody (Sigma-Aldrich) conjugated to horse radish peroxidase. Blot was imaged using Luminata Crescendo (EMD Millipore) and shown in **Figure 6**.

Activity of GalNAcT2 Enzymes in Concert with C-4 Hexose Epimerases

Purified GalNAcT2 enzymes were assayed alongside purified C-4 Hexose Epimerase enzymes named CPG-13 (**Appendix 10**) (Bernatchez *et al.* 2005) and *Ecgne2* (Guo *et al.* 2006) the latter of which was created into ECE-01 by the Wakarchuk Lab by insertion of the *Ecgne2* gene into plasmid VEK-06 (**Appendix 11**; ECE-01 in VEK-06 detailed in **Appendix 12**). 20 µL reactions were set up with 20 mM Tris pH 7.5, 10 mM $MnCl_2$, 1.5 mM UDP-GlcNAc and 0.1 mM BDP-$IL29_1$. CPG-13 reactions contained 0.05 µg/µL of epimerase enzyme while Ecgne2 reactions contained 0.18 µg/ µL of enzyme. All reactions contained 0.5 µg/ µL of GalNAcT2 and were carried out at 30 ^oC with time points being taken by stopping 2 µL aliquots with stopping solution. Reactions were visualized using TLC. Control reactions were also maintained which excluded epimerase and UDP-GlcNAc.

Optimization of pH Conditions for GalNAcT2 Reactions

Ranges of pH tested for all three GalNAcT2 proteins. Candidate concentrations were 25 mM sodium acetate pH 5.0, 25 mM sodium acetate pH 5.5, 25 mM MES pH 6.0, 25 mM MES pH 6.5, 50 mM HEPES pH 7.0, 50 mM HEPES pH 7.5, and 50 mM HEPES pH 8.0. Other reaction components were maintained at 10 mM $MnCl_2$, 1 mM UDP-GalNAc and 0.05 mM BDP-$IL29_1$. Reactions proceeded for 30 minutes at 30^oC and then stopped and two further technical replicates were performed. Stopped reactions were then diluted to 0.3 µM of BDP-$IL29_1$ with a solution of 10 mM ammonium acetate, 2.5% ACN and 0.1% TFA and analyzed using reverse phase C-18 column chromatography using a 1.5 mL Restek column on a Shimadzu HPLC system at a flow rate of 0.5 mL/min with fluorescence absorption at 504 nm and emission at 514 nm. Percent product formation values were obtained and converted to specific activity values at the given pH. Samples were injected onto the column after one MilliQ sample and three

blank samples to ensure column equilibration. A negative control with excluded UDP-GalNAc was also included at the beginning and end of samples.

Analysis of $MnCl_2$ Concentrations and their Impact on GalNAcT2 Activity

The effect of different $MnCl_2$ concentrations were briefly explored by maintaining a constant pH state in 10µL reactions in pH 6.0 buffer while varying manganese concentrations at 5 mM, 10 mM, and 15 mM. Reactions occurred at 30^{o}C for a total of thirty minutes and then stopped with stopping solution. Other conditions were maintained as above.

Exploration of Buffer Effects on Enzyme Activity

HGT-13 reactions were carried out at conditions of 25 mM sodium acetate pH 5.5, 25 mM MES pH 5.5, 25 mM citric acid pH 5.5, 25 mM MES pH 6.0 and 25 mM Citric acid pH 6.0. Other components were maintained as in the pH analysis with enzyme concentration of HGT-13 being maintained at 0.024 µg/µL. Reaction was allowed to proceed as detailed above. Reactions were stopped after 30 minutes and repeated twice more. Reactions were analysed on HPLC as detailed above.

Creation, Growth, Purification and Testing of Folding Chaperone Assisted GalNAcT2s

Initial creation of the multiple plasmid strain for GalNAcT2 was performed by making the chloramphenicol resistant AP-01 plasmid (**Appendix 13**) in expressed in SHuffle T7 Express (NEB), (which was created by Hirak Saxena coding for human protein disulfide isomerase first reported by Nguyen *et al.* (2011)) competent using Thermo Scientific's Transformaid kit. The plasmid for HGT-13 was then transformed into the competent SHuffle T7 Express strain using instructions from the Transformaid kit. Transformants were grown on LB agar plates containing 150 µg/µL ampicillin and 25 µg/µL chloramphenicol. Colonies had 10% glycerol stocks made

for further usage. Further stocks of competent Shufftle T7 Express strains already containing the AP-01 plasmid were obtained from the Wakarchuk lab and the plasmids for DGT-100 and CGT-100 were transformed and glycerol stocks were generated as well. Folding chaperone assisted enzymes were grown, purified and tested as was done on the single plasmid expressed GalNAcT2 enzymes being mindful to maintain additional antibiotic presence at 25 µg/µL of chloramphenicol. Lysis, however, was performed with the aid of an Emulsiflex C-5 system (Avestin) at a pressure of 15000 psi allowing several passes through the system to ensure complete lysis. Tubing on the Emulsiflex was kept cool using ice. The lysate from the system was centrifuged for 30 minutes at 12000 rpm and the supernatant was syringe filtered through a 0.2 µm membrane. Lysed material was processed and purified as stated above using 5mL MBP-Trap columns thoroughly washed with 0.5 M NaOH between each isoform purification.

GalNAcT2 Reactions on Protein Substrates and Detection via Lectin Blotting

Reactions were set up with substrates GB1-IL29-3G, GB1-IFNα2b, and GB1-hGH. Aside from IL29-3G where the substrate concentration was 0.5 µg/µL, all other substrate concentrations were maintained at 1 µg/µL. Enzyme concentrations were maintained at 1:20 of the respective protein concentrations. Reactions occurred in 50 µL with 25 mM sodium acetate pH 5.5, 10 mM $MnCl_2$ and 1 mM UDP-GalNAc. Control reactions were set up excluding UDP-GalNAc. Reactions took place overnight at 30 ^oC and were stopped using 4X SDS-Loading dye followed by boiling at 95 degrees.

15% SDS-PAGE gels were prepared and membrane transfers were performed on PVDF overnight at 4 ^oC in a cold room at 30V. Blots were blocked with 5% BSA in PBST (PBS 7.4 with 0.2% Tween-20) for one hour followed by overnight incubation at 4 ^oC in SBA binding buffer containing PBST, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 1 mM $MgCl_2$ and 0.4 µg/µL SBA lectin

conjugated to Horse Radish Peroxidase (HRP) (Sigma Aldrich from a stock of 2 mg/mL). The PBST and salt solution was first syringe filtered through a 0.2 µm membrane before addition of lectin solution. Six 1X PBS rinses were performed followed by imaging using 1mL Luminata Crescendo with a 30 second incubation. Images were taken at 5 seconds post exposure.

Preparation of Mass Spectrometry Samples of IFNα2b

100 µg of GB1-IFNα2b was glycosylated in a 200 µL reaction with 25 mM sodium acetate pH 5.5, 10 mM $MnCl_2$ and 1 mM UDP-GalNAc. Reactions occurred over the course of 48 hours to ensure completion and buffer exchanged to 10 mM Ammonium Bicarbonate pH 6.0 using Superdex G75 filtration (GE Healthcare) using the AKTA at room temperature. A control reaction was also prepared by excluding UDP-GalNAc and buffer exchanged using Superdex 75 filtration. Samples were then sent to the SPARC BioCentre at SickKids for intact mass analysis.

K_m Determination of GalNAcT2 Enzymes

Sufficient enzyme dilutions were prepared for linear activity detection and 10 µL 0.05 mM, 0.1 mM, 0.2 mM, 0.4 mM, and 0.8mM reactions of BODIPY labelled $IL29_1$ and $IL29_2$ were prepared. Other relevant conditions were maintained at 10 mM $MnCl_2$ and 1mM UDP-GalNAc. Reactions were allowed to proceed at 30 ^oC created from a master mix allowing for time points at 10, 15, 20, 25, 30 minutes for each substrate concentration. Reactions were stopped using stopping solution at necessary times. A negative control reaction was prepared at 0.4 mM substrate concentration and had UDP-GalNAc excluded. Reactions were analyzed on HPLC as described above. Percent product formation data was converted to molar quantities and slopes were analyzed for kinetic parameters with three different enzyme preparations using PRISM Pad software.

Table 2: List of Strains Utilized in this Study

Strain	Genotype
Origami 2	StrR, TetR. Δ(ara-leu) 7697 ΔlacX74 ΔphoA pvuII phoR araD139 ahpC galE galK rpsL F'[lac+ lacIq pro] gor522::Tn10 trxB.
SHuffle Express	fhuA2 [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
SHuffle Express T7	fhuA2 lacZ::T7 gene1 [lon] ompT ahpCgalλatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 ΔgorΔ(mcrC-mrr)114::IS10

Table 3: List of Primers Involved in this Work

Primer Name	Sequence (5' to 3')	Base Count and Direction	Tm
rCW-70	AGG CCC TTT CGT CTT CAA GCA GAT C	25 Reverse	56^{o}C
MB1041	GGT GAT CAA CGC CGC CAG CGG TCG	24 Forward	65^{o}C

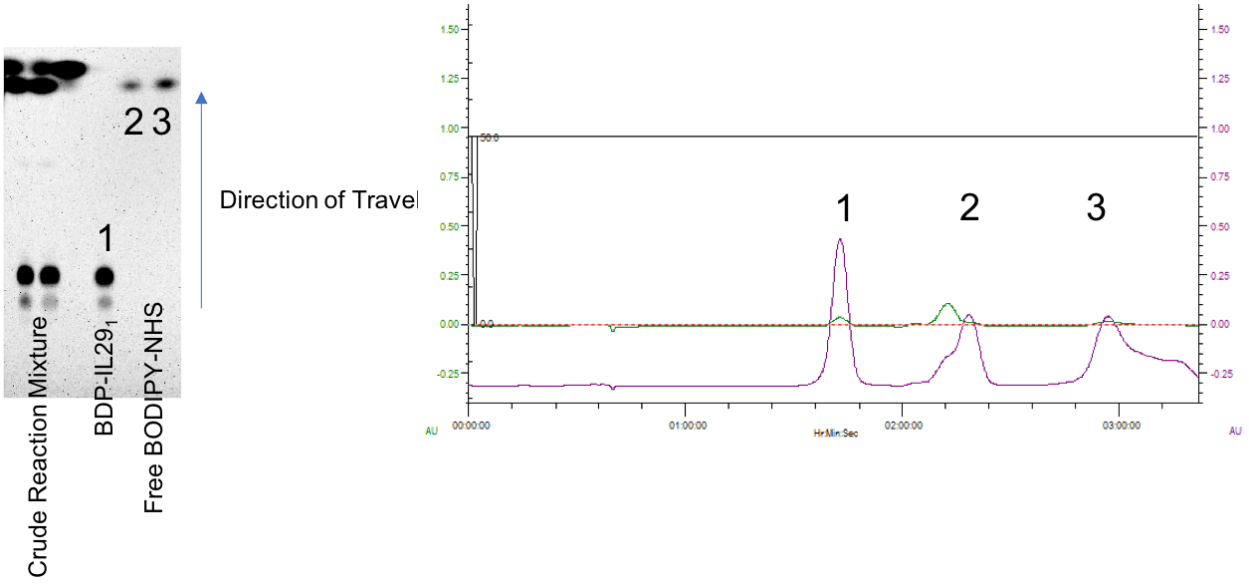


Figure 6. BDP-IL291 Purification: TLC of the crude reaction mixture of BDP-NHS and IL29$_1$ peptide showing the effectiveness of gel filtration. Two left most spots represent the crude reaction mixture while the two right most spots are later filtration fractions indicative of free BODIPY-NHS. BDP-IL29$_1$ has been successfully separated from free BDP-NHS as seen in the BDP-IL29$_1$ lane showing no higher R_f spots higher. The right pane shows the corresponding peaks that were spotted on the TLC. Purple line represents absorbance at 504 nm while the green line represents absorbance at 280 nm.

Table 4: Vectors Used

Vector Name	Resistance	Features
pCW	Ampicilin	-
pCWMalET	Ampicilin	N-terminal MBP tag cleavable by thrombin
pMalC5X	Ampicilin	N-terminal MBP tag cleavable by Factor Xa
VEK-06	Ampicilin	N-terminal polyhistidine tag
VEK-08	Chloramphenicol	

Table 5: Constructs Used

Construct Name	Gene	Expressed Amino Acids	Predicted Mass (kDa)	Tag	Vector	Source
HGT-12	*GALNT2*	52-571	102.29	N-terminal MBP tag	pCW	*H. sapiens*
HGT-13	*GALNT2*	51-447	87.78	N-terminal MBP tag	pCWMalET	*H. sapiens*
HGT-14	*GALNT2*	51-447	88.57	N-terminal MBP tag	pMalC5X	*H. sapiens*
DGT-100	*PGANT2*	89-493	88.11	N-terminal MBP tag	pCWMalET	*D. melanogaster*
CGT-100	*Gly4*	65-461	87.66	N-terminal MBP tag	pCWMalET	*C. elegans*
CPG-13	*gne*		78.77	N-terminal MBP tag	pCWMalET	*C. jejuni* NCTC 11168
ECE-01	*gne2*		40.19	N-terminal His tag	VEK-06	*E. coli* O:86
AP-01	*hPDI*	18-508		-	-	*H. sapiens*

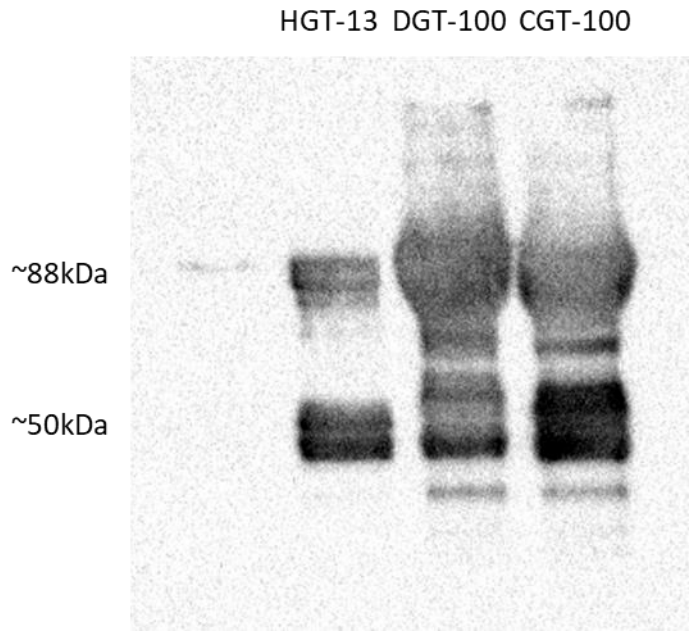


Figure 7. MBP Product Verification: Western blot of all three purified GalNAcT2 enzymes with 0.02 µg/µL anti-MBP antibody conjugated to HRP on a PVDF membrane exposed for 10 seconds. Full length enzyme associated signal is observed at ~88kDa while anything beneath represents degradation product. Signal disappears after ~50 kDa indicating the mass of the maltose binding protein.

Results

Verification of *N*-acetylgalactosaminyltransferase, C-4 hexose epimerase, and Core-1 Galactosyltransferase from previous recombinant plasmid design

Crude assay of whole cell crude lysate of OGO-6 including negative control excluding UDP-GlcNAc is shown in **Figure 8** and **Figure 9**. **Figure 8** shows activity of the GalNAc transferase from the OGO-6 operon through the presence of fluorescent signal present in the lanes denoted (+) suggesting the formation of BDP-$IL29_1$-GalNAc. Reactions, when resolved on TLC, plates show degradation products of free BODIPY and other BODIPY labelled degradation peptide products. **Figure 9** shows the activity of the core-1 β-1,3-galactosyltransferase (CgtB) also expressed from the OGO-6 construct. The presence of fluorescence signal on BODIPY labelled GalNAc, in addition to the starting material, suggests the formation of BDP-GalNAc-Gal. The right lane in **Figure 9** shows the presence of di-galactosylated product as well which is represented by the lowest band. Negative control lanes (denoted (-)) excluding crude extract show no signal of decreased R_f. **Figure 10** shows the restriction fragments associated with successful excision of the ppGalNAcT2 gene from OGO-6 and into plasmid pCW. **Figure 11** shows relatively low expression of HGT-12 enzyme in comparison to the expression from OGO-6 as seen by the bands appearing at 102 kDa. Peak analysis from HPLC chromatograms performed on GalNAcT2 reactions on $IL29_1$ showed 75% reaction completion after one hour (**Figure 12**) resulting in an enzyme activity value for HGT-12 on $IL29_1$ of 6.25 mU/mL.

Growth, Purification, and Initial Assays of HGT-13 and HGT-14

Expression of HGT-13 and HGT-14 constructs were not optimal at IPTG concentrations of 0.5 mM and reduction of IPTG concentration to 0.2 mM in culture showed improved protein production. Crude assays from higher IPTG concentrations did not show activity while those

from lower concentrations did (data not shown). Furthermore, protein production was improved at 20°C induction temperature in comparison to protocols with 25°C. **Figure 13** shows that after one hour of reaction time on substrate $IL29_1$, product bands of 49% and 53% are obtained for HGT-13 and HGT-14 respectively using Image Lab Software. This resulted in enzyme activity values of 4.1 mU/mL and 4.4 mU/mL for HGT-13 and HGT-14 respectively. The enzyme solutions before and after inductions shows presence of fusion protein at ~88 kDa. There are lower weight degradation products with HGT-14 with higher band intensity potentially indicating increased degradation product (**Figure 14**).

Production of Recombinant GalNAcT2 from *Drosophila melanogaster* and *Caenorhabditis elegans* and Testing of Crude Lysates

D. melanogaster and *C. elegans* isoforms of GalNAcT2 were only successfully expressed in pCWMalET. **Figure 15** shows the lack of obvious distinct full length fusion product at 89 kDa in comparison to the IPTG exclusion controls denoted as (-). Despite the lack of obvious presence of full length fusion protein, the pCWMalET constructs showed a higher presence of degradation product at ~45 kDa. Testing of crude extracts of the cell pellets from the 16°C expression (shown in **Figure 16**) suggest that despite high degradation of the fluorescent substrate $IL29_1$, there is still activity indicating presence of enzyme though not at full length. Whole cell lysates expressed in **Figure 15** via pCWMalET were tested on substrate $IL29_1$ and highlighted in **Figure 16**. Further purification of these pellets with MBP affinity chromatography did not yield active protein. Active fusion protein for both DGT-100 and CGT-100 were found only after incubation for 20 hours at 20°C post-induction. All three GalNAcT2 enzymes still presented varying amounts of degradation product after MalE affinity chromatography (**Figure 17**). Initial enzyme activity values were as follows; for CGT-100 on $IL29_1$ and $IL29_2$ were 27

mU/mL and 26 mU/mL respectively; for DGT-100 on $IL29_1$ and $IL29_2$ were 6.2 mU/mL and 5.8 mU/mL respectively; and for HGT-13 on $IL29_1$ and $IL29_2$ the activity values were 5 mU/mL and 7.2 mU/mL respectively.

Co-expression of Folding Chaperone Alongside GT-27 Enzymes

Co-expression of GalNAcT2 expression plasmids and AP-01 was verified using NdeI digests with the representative figure of HGT-13 and AP-01 being shown in **Figure 18**. To compare preparations with and without hPDI, the designation MPS-45 was given to the folding chaperone containing preparation while the single plasmid expression enzyme remained at HGT-13. *H. sapiens* enzymes assayed under both forms (MPS-45 and HGT-13) maintaining protein concentrations of 0.2 µg/µL showed specific activities of 63 mU/mg for the single plasmid expressed protein while the multiple plasmid preparation gave a specific activity value of 125 mU/mg. **Figure 19** highlights the doubling of activity due to presence of the accessory human protein disulfide isomerase. Protein expression patterns were not different with degradation products being present in preparations with and without folding chaperone. After these findings HGT-13, DGT-100 and CGT-100 were all co-expressed with hPDI.

Ability of Recombinant Glycosyltransferases to Cooperate with Recombinant C-4 Hexose Epimerases *in vitro*

Epimerases co-assayed with CGT-100 showed coupled activity of HexNAc epimerases with the GalNAcT2 preparations (**Figure 20;** CPG-13 right pane and ECE-01 left pane). The left pane begins with the reaction lane containing both epimerase and UDP-GlcNAc, the epimerase exclusion lane, and the UDP-GlcNAc exclusion lane. The right pane shows the no epimerase lane, the reaction lane, and the no UDP-GlcNAc lane from left to right. CGT-100 concentrations

were maintained at 0.07 μg/μL and epimerase concentrations were 0.1 μg/μL. Working epimerase is indicated due to the lack of signal present in controls without epimerase and without UDP-GlcNAc. Despite both showing different activities there is still evidence of formation of GalNAc-modified BDP-IL29$_1$ for both epimerases in concert with the GT-27 enzymes. All enzymes showed activity in concert with these epimerase enzymes replicating results seen from OGO-6 assays performed earlier.

Optimization of Reaction Conditions

Optimal reaction conditions were obtained by testing the activity of the enzymes on the BDP-IL29$_1$ substrate in different buffer conditions. No differences in activity were observed for different concentrations of $MnCl_2$ (data not shown). **Figure 21** demonstrates the highest specific activity for the *H. sapiens* isoform of GalNAcT2 occurring at pH 5.5 in 25 mM sodium acetate with a specific activity of 674 mU/mg. The *D. melanogaster* isoform exhibits the highest specific activity beginning at pH 5.5 without reduction in specific activity onwards for the pH levels tested shown in **Figure 22**. The *C. elegans* isoform shows similar results as the *H. sapiens* isoform where the maximum specific activity observed occurs at pH 5.5 shown in **Figure 23**.

The representative enzyme used to test for buffer effects was HGT-13 at 0.025 μg/μL in 10 μL reactions performed three times. **Figure 24** shows that citric acid has decreased specific activity at both pH 5.5 and pH 6.0. The reactions in sodium acetate and MES showed higher specific activities with sodium acetate at pH 5.5 showing the highest activity for the pH values tested. All reactions were allowed to proceed to 30% conversion before stopped.

Attempt at Saturating Enzymes with BODIPY Labelled IL29$_1$ and IL29$_2$ Peptide Loop Derivatives

Saturation of enzymes using BDP-IL29$_1$ were not successful as reactions showed severe substrate inhibition at higher concentrations. **Figure 25** shows initial increase in activity upon increasing substrate concentration, but the rate severely decreases at 0.2 mM of BDP-IL29$_1$ on HGT-13. This figure does not conform to the regular Michaelis-Menten model for easy determination of enzyme kinetics based off analysis on Prism (GraphPad). The highest rate of activity was observed at 0.1 mM, and Km can only be estimated to be ~0.05 mM. The estimated Vmax based on a non-substrate inhibition model would occur ~5.8 pmol of product per minute (SD:0.68 n=3).

D. melanogaster (DGT-100) isoform reactions performed similarly as HGT-13 reactions are highlighted in **Figure 26.** At the same concentration of enzyme, the peak activity was observed before the reaction rate starts to decline at higher concentrations. The peak reaction rate observed here under a non-substrate inhibition model would occur at ~1.2 pmol product/min (SD:0.11, n=3). **Figure 27** shows the reaction for CGT-100 with no pattern at all. The estimated Vmax based on a non-substrate inhibition model would occur ~1.1 pmol of product per minute (SD:0.4, n=3). Between all three enzymes tested, at the same three concentrations of enzyme, the most effective isoform is the *Homo sapiens* isoform for this substrate solely based on the highest rate achieved due to lack of IL29 peptide data on GalNAcT2 in literature. **Figure 28** shows the lowest HGT-13 rate is still higher than the highest rates from CGT-100 and DGT-100. Estimated K_m for these enzymes assuming half of estimated V_{max} for all enzymes would be around 0.05 mM. This is not the result seen in CGT-100 as there is no pattern at all even with occurrence of glycosylation suggesting this substrate is not compatible with CGT-100.

Enzymes assayed on BDP-IL29$_2$ showed results which conform to the Michaelis-Menten model more than BDP-IL29$_1$. Enzyme concentrations were increased to 0.025 μg/μL in hopes of increasing active site in the reaction system after observing the 0.05 mM substrate concentration did not reach completion after 25 minutes. Results for assays involving BDP-IL29$_2$ show less reduction in rate at higher concentrations than the assays involving BDP-IL29$_1$. **Figure 29** shows the curve obtained from a substrate saturation reaction involving BDP-IL29$_2$ being glycosylated by HGT-13. The evidence of substrate inhibition is not as obvious although there is a decrease in rate at 0.4 mM of substrate. Michaelis-Menten fit values obtained from Prism gives kinetic parameters of 31.72 pmol/min for V_{max} and 0.085 mM for K_m. **Figure 30** highlights a similar trend observed in the DGT-100 reaction on the same substrate at the same enzyme concentration. Kinetic parameters obtained from PRISM software give a V_{max} value of 33.55 pmol/min and a K_m value of 0.084 mM. Both instances show a decrease in activity for the highest concentration of substrate at 0.4 mM. **Figure 31** and **Figure 32** show the chromatograms and reaction time curves for HGT-13 on IL29$_2$ at 0.05 and 0.2 mM of BODIPY substrate respectively. Both concentrations show an increase over time and the 0.2 mM concentration curve has a higher rate of activity based on slope.

Verification of *O*-linked Glycosylation on Recombinant Protein Substrates

Qualitative methods were utilized to observe effectiveness of glycosylation on recombinant test proteins GB1-IL29-3G, GB1-IFNα2B, and GB1-hGH and shown with their predicted glycosylation sites, excluding their GB1 fusion tag in **Figure 4**, **Figure 33**, and **Figure 34** respectively. Lectin blotting analysis showed all enzymes active on protein substrates with some enzymes having differential effects. Luminescence observed shows the ability of all three isoforms of enzymes to glycosylate the test protein GB1-IL29-3G which has a mass of 29 kDa

(**Figure 35**). In the SDS-PAGE portion of the figure slight gel shifts are observed in the lanes denoted (+) for the respective enzymes. The gel shift is not as pronounced in the CGT-100 lane. The ability of CGT-100 to glycosylate this particular test protein can only be concluded via the lectin blot. The signal strength of the band corresponding to glycosylated GB1-IL29-3G by CGT-100 is also not as strong as the products from the other isoform reactions. This trend was also observed in reactions with the test protein substrate GB1-IFNα2B (**Figure 36**) with substrate mass occurring at 28 kDa. As for GB1-IL29-3G, recombinant GB1-IFNα2B seems broadly reactive as well with luminescence intensities for HGT-13 and DGT-100 matching each other while CGT-100 does not appear to proceed with the same rate as the other reactions even after ensuring ample time for reaction completion. This trend for both discussed test protein substrates were observed over multiple trials across different batches of enzymes (n=3). The relatively low activity from CGT-100 was also observed in the recombinant test protein GB1-hGH reactions. There is a clear decrease in substrate glycosylation from HGT-13 to DGT-100 to CGT-100 indicated by the luminescent signal observed at 28 kDa (**Figure 37**). Based on results from SBA lectin blots on the test proteins, HGT-13 is the most broadly reactive GalNAc transferase isoform tested, while CGT-100 represents the least effective. The ability to notice differences in migration of proteins on SDS-PAGE was compared for high stock substrates hGH and GB1-IFNα2B. **Figure 38** highlights the results from the lectin blots showing slight gel shifts in the HGT-13 and DGT-100 lanes while gel shifts are not as obvious in the CGT-100 lanes. The only CGT-100 lane showing a band shift is the GB1-IFNα2B lane.

As GB1-IFNα2B was observed to be a great substrate in blotting and mass spectrometry results from other colleagues (Du *et al.*, submitted), 100 μg of it was glycosylated by HGT-13 and analyzed for intact mass for glycosylated by HGT-13. **Figure 39** shows panes A and B

corresponding to non-glycosylated and glycosylated forms with masses 27943 Da and 28147 Da respectively. The increase in mass is an indicator that mass increase had occurred on the protein. The difference of 204 Da in mass does correspond to the addition of an *N*-acetylhexosamine (Kolarich *et al.* 2012).

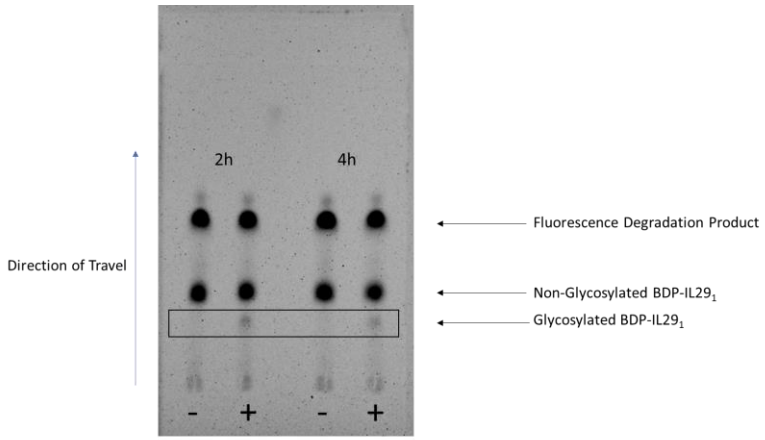


Figure 8. Crude whole cell extract activity assay of GalNAcT2 and HexNAc epimerases of OGO-6: Tested reaction is ppGalNAcT2 on 0.1 mM IL29$_1$. Lanes denoted (-) indicate the exclusion of UDP-GlcNAc from the reaction mixture. Reaction resolved on silica plates using 4:2:1:0.1 solvent containing ethyl acetate: methanol: water: acetic acid respectively.

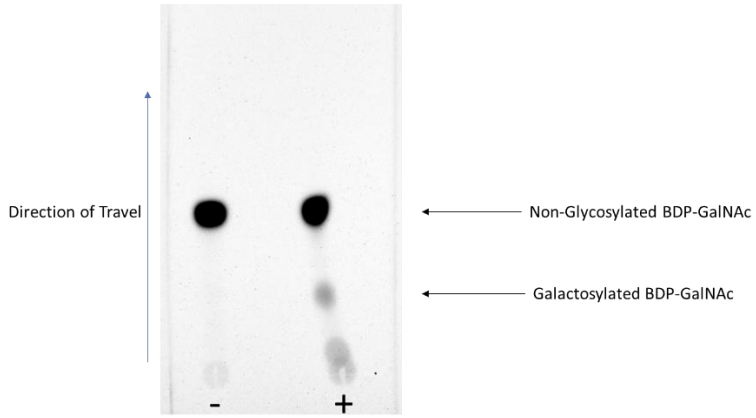


Figure 9. Crude whole cell extract activity assay of OGO-6: Tested reaction is CgtB (β-1,3-galactosyltransferase) on 0.5 mM BDP-GalNAc. Lanes denoted (-) indicate the exclusion of crude cell extract from the reaction mixture. Reaction resolved on silica plates using 4:2:1:0.1 solvent containing ethyl acetate: methanol: water: acetic acid respectively.

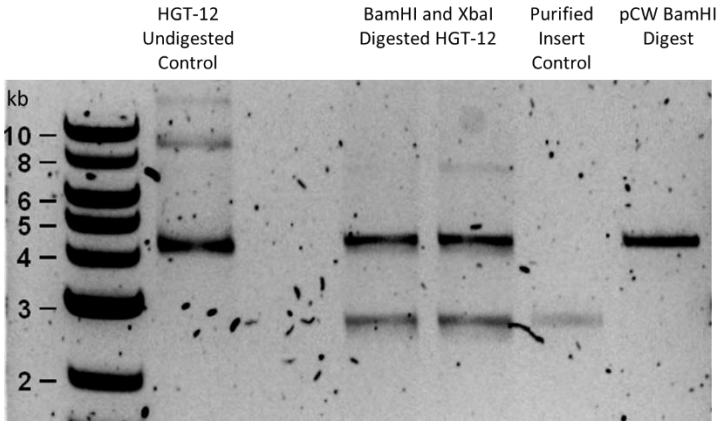


Figure 10. Verification of Insert Fragments in HGT-12: A 0.8% agarose gel imaged using the assistance of SYBR Safe dye to visualize bands. A look at the genetic components required to create the HGT-12 expression plasmid.

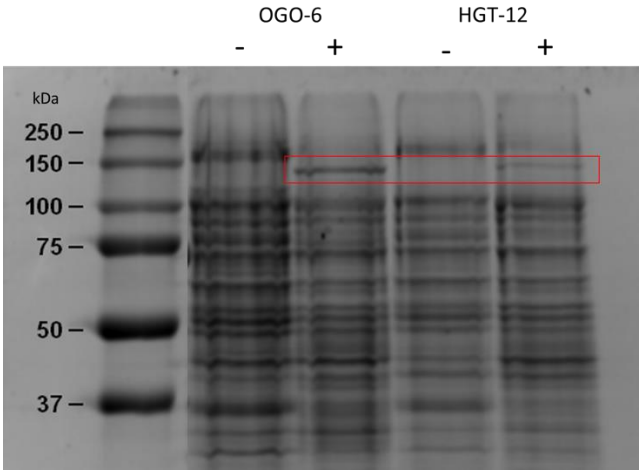


Figure 11. GalNAcT2 Induction Verification of HGT-12: A 12% gel highlighting the expression of HGT-12 expected at 102 kDa. Lanes denoted (-) indicate exclusion of IPTG. Expression of HGT-12 protein is higher from the OGO-6 plasmid than the HGT-12 plasmid.

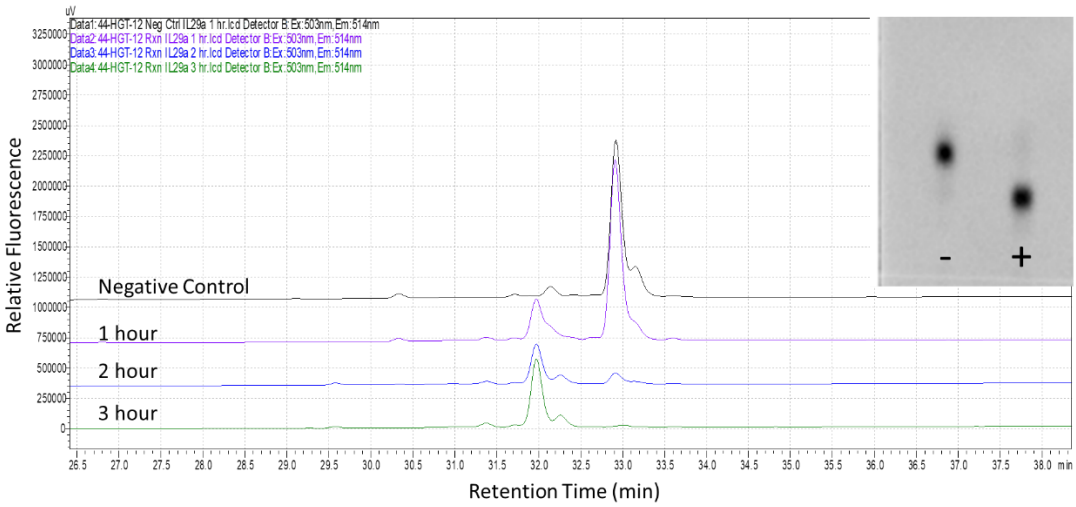


Figure 12: HPLC chromatogram showing relative fluorescence as a function of retention time including the sample TLC in the top right. Peaks on the left are product peaks indicating glycosylated BDP-IL29$_1$ peptide while peaks on the right are reactant peaks indicating non-glycosylated material. The inset pane shows the TLC of the reaction after three hours with the (-) lane indicating the exclusion of UDP-GalNAc. Reaction conditions for the GalNAcT2 assay are 20 mM Tris pH 7.5, 10 mM $MnCl_2$, 1 mM UDP-GalNAc, and 0.1 mM BDP-IL29$_1$.

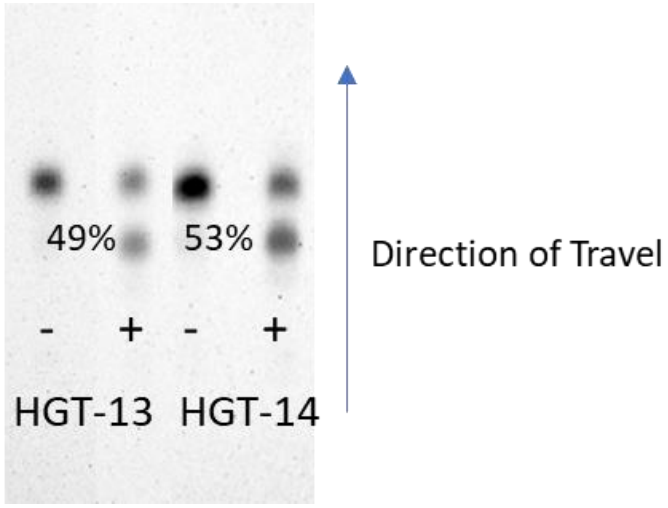


Figure 13. TLC of enzymatic preparations of HGT-13 and HGT-14 reactions after one hour: Lanes denoted (-) indicate exclusion of UDP-GalNAc. Signals of lower R_f are representative of glycosylated product. The substrate in this reaction is IL29$_1$. Enzymes preparations consisted of some maltose binding protein degradation product. Reaction conditions for the GalNAcT2 assay are 20 mM Tris pH 7.5, 10 mM $MnCl_2$, 1 mM UDP-GalNAc, and 0.1 mM BDP-IL29$_1$.

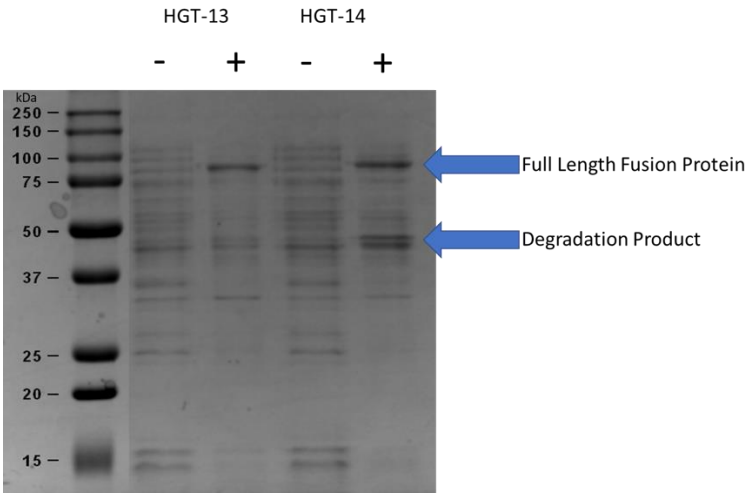


Figure 14. GalNAcT2 Expression from HGT-13 and HGT-14: A 12% SDS gel showing the results of whole cell lysates from non-induced (-) and induced (+) samples of HGT-13 and HGT-14. Full length fusion protein is present at ~88 kDa along with degradation products at lower mass from 45-50 kDa. Increased degradation product is noted in HGT-14.

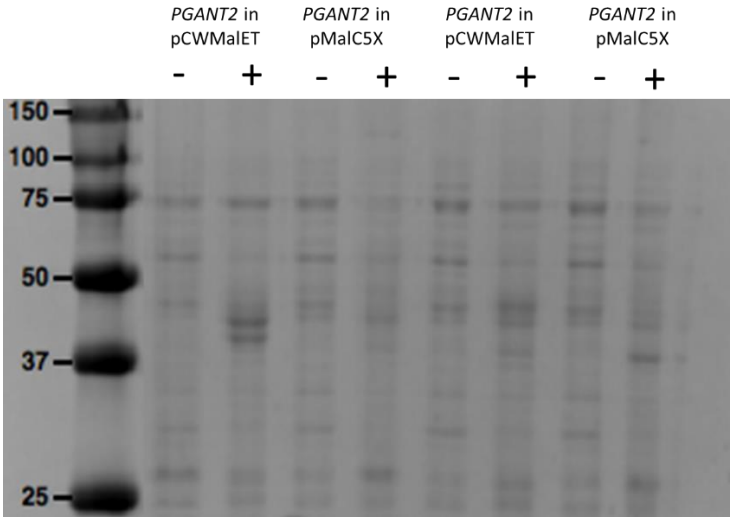


Figure 15. PGANT2 Production in Two Different Vectors: A 12% SDS gel showing the results of whole cell lysates from non-induced (-) and induced (+) samples of PGANT2 enzymes expressed in pCWMalET and pMalC5X. Four lanes from the right of the figure are induction temperatures of 30ºC while the other lanes are induction temperatures of 16ºC.

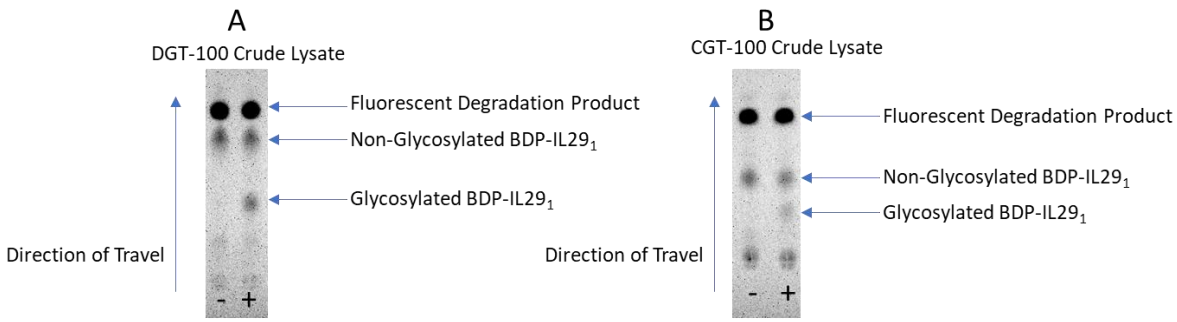


Figure 16. Crude Lysate Activity from DGT-100 and CGT-100: Crude extracts of DGT-100 and CGT-100 induced at 16ºC. Pane A shows the reaction from DGT-100 while Pane B shows the reaction from CGT-100. Lanes denoted (-) have UDP-GalNAc excluded.

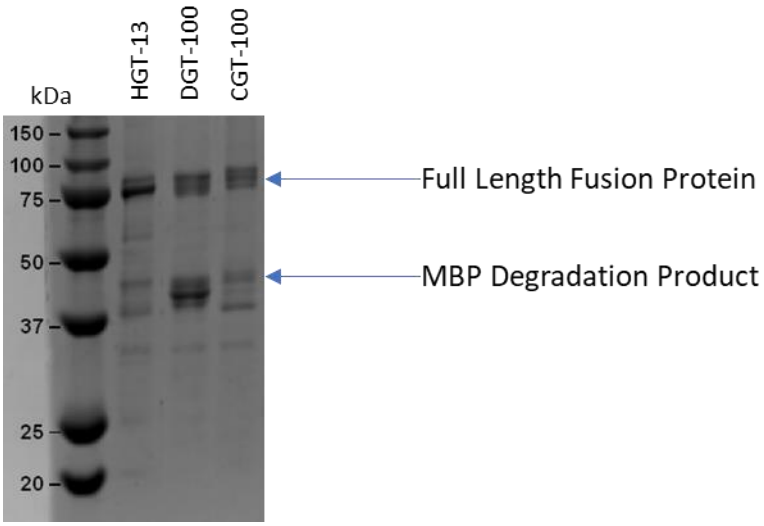


Figure 17. Purification of Three GalNAcT2 Isoforms: A 12% SDS gel showing purified GT-27 enzymes at the same concentration. Full length MBP fusions are observed at 88 kDa for all three enzymes with MBP degradation product also being observed higher than 37 kDa.

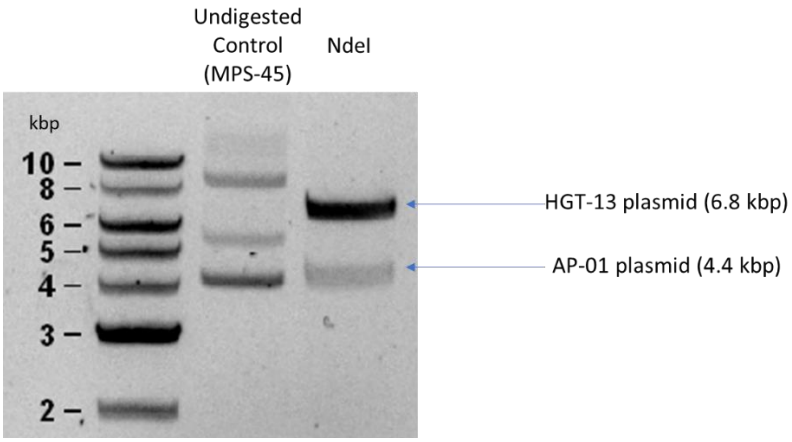


Figure 18. Verification of Multiple Plasmid Transformation: A 0.8% agarose gel with linearized plasmids in MPS-45 (HGT-13 plasmid and AP-01 plasmid) showing presence of two linearized plasmids in comparison to non-digested plasmid preparations.

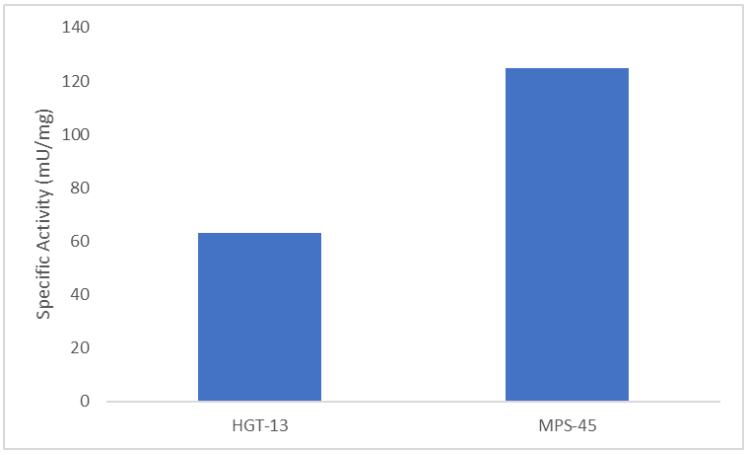


Figure 19. Increase in GalNAcT2 Activity by Coexpression with hPDI: Specific activity comparison of single plasmid HGT-13 and folding chaperone assisted HGT-13 showing almost doubling of activity. Specific activity of HGT-13 was 63 mU/mg while MPS-45 was 125 mU/mg.

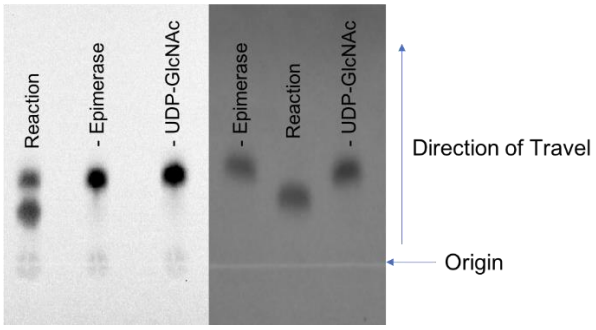


Figure 20. Coupled GalNAcT2 Assays with HexNAc Epimerases: TLC of epimerases CPG-13 (right) and ECE-01 (left) in concert with CGT-100. Reactions maintained same epimerase concentrations of 0.1 μg/μL for one hour.

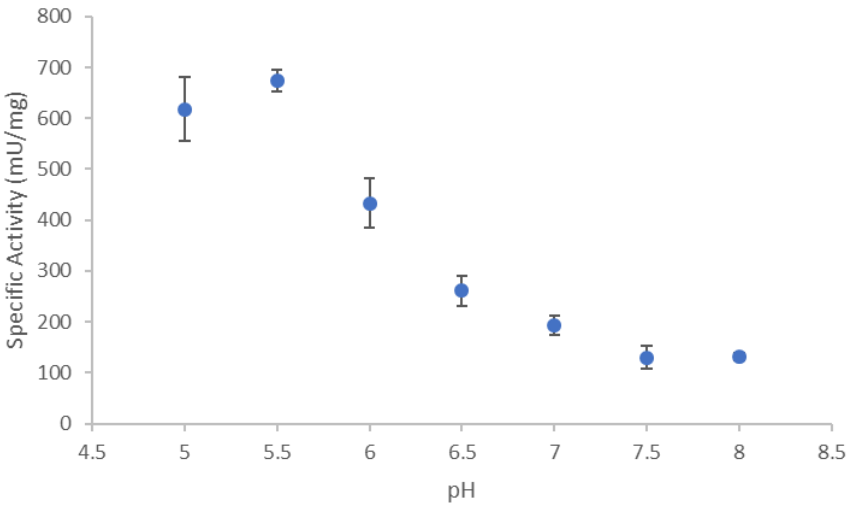


Figure 21. Activity of HGT-13 Over Varying pH: Optimal specific activity of HGT-13 occurring at pH 5.5. Reactions stopped after thirty minutes were analyzed and found to have the highest specific activity towards 0.05 mM BDP-IL29$_1$ at 674 mU/mg.

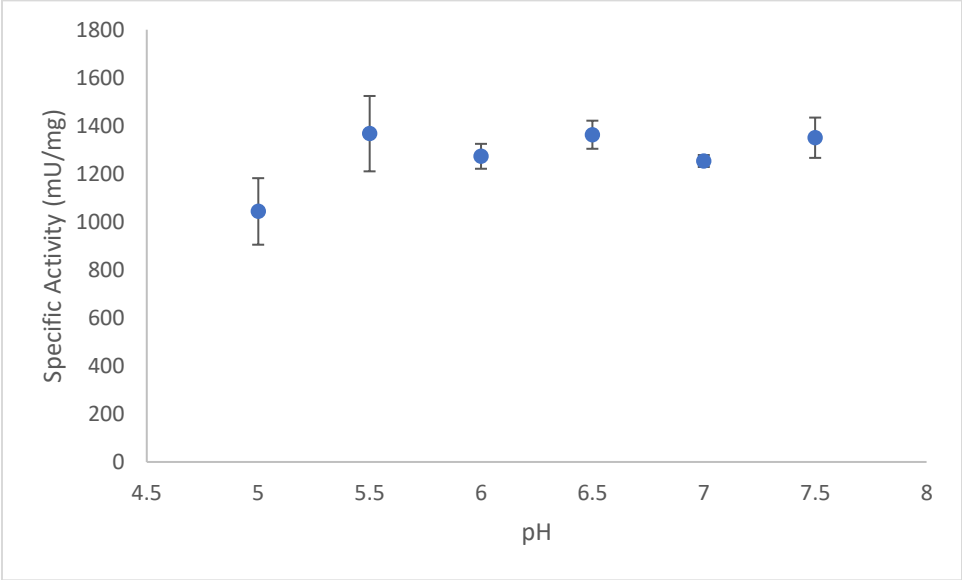


Figure 22. Activity of DGT-100 Over Varying pH: Optimal specific activity of DGT-100 does not have an effect after pH 5.5. Reactions stopped after thirty minutes were analyzed and found to have the highest specific activity towards 0.05 mM BDP-IL29$_1$ at 1367 mU/mg.

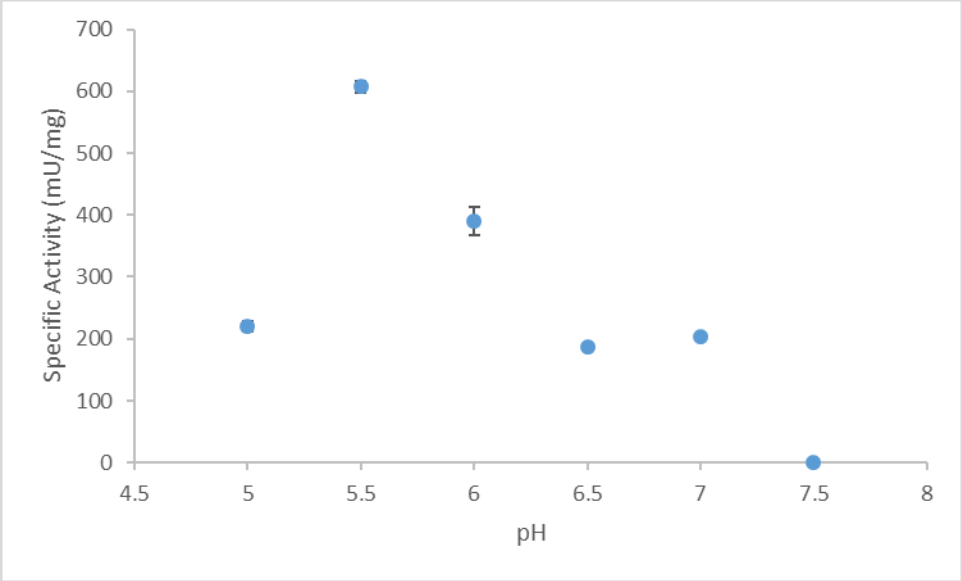


Figure 23. Activity of CGT-100 Over Varying pH: Optimal specific activity of CGT-100 occurs at pH 5.5. Reactions stopped after thirty minutes were analyzed and found to have the highest specific activity towards 0.05 mM BDP-IL29$_1$ at 608 mU/mg. Each point has an error bar but they are very minimal.

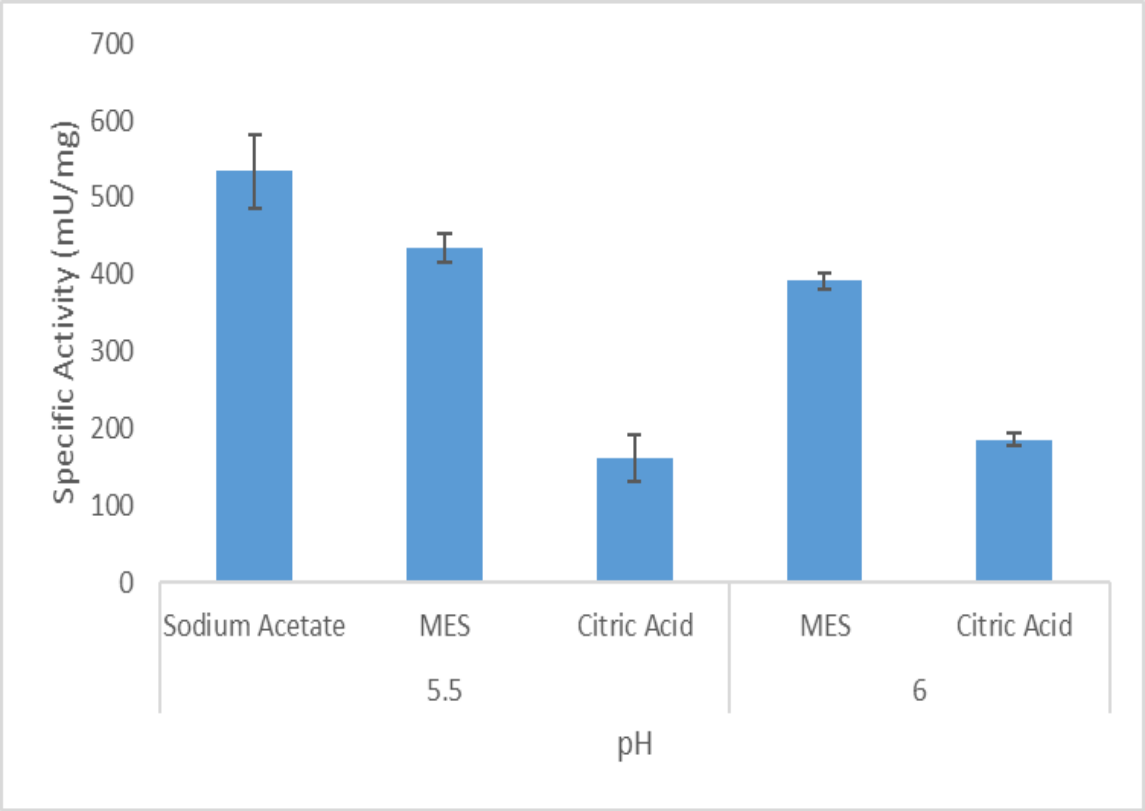


Figure 24. Buffer Effects on GalNAcT2 Activity: Effects of different buffers on activity of HGT-13 at pH 5.5 and 6.0. Citric acid has an extremely negative effect on GalNAcT2 activity while MES does not have as pronounced negative effect.

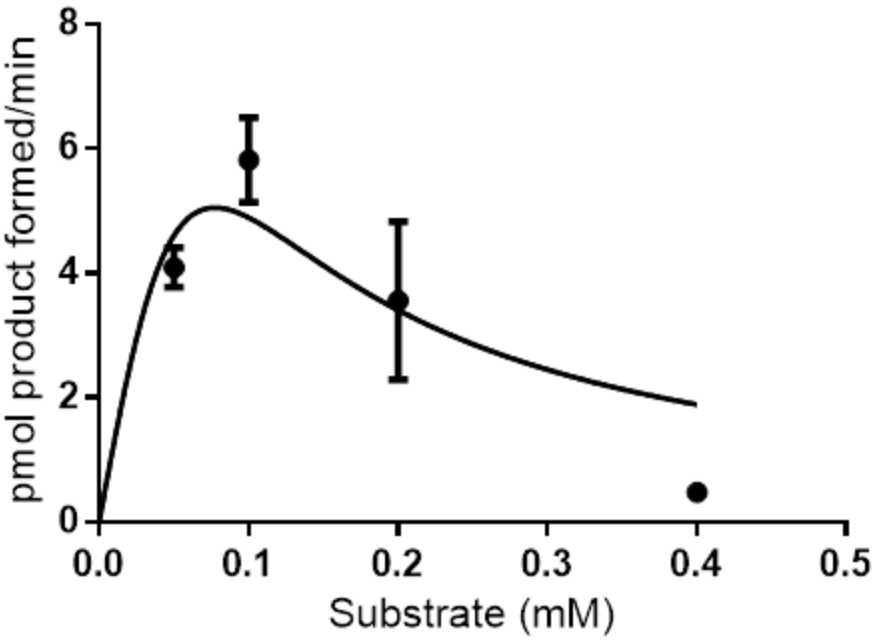


Figure 25. Substrate saturation curve for HGT-13 on IL29$_1$: Maximum rate of reaction occurs at 0.1 mM with maximum V_{max} of 5.8 pmol product formed per minute. Estimated K_m occurs at approximately 0.05 mM. Enzyme is inhibited by increasing concentration of substrate.

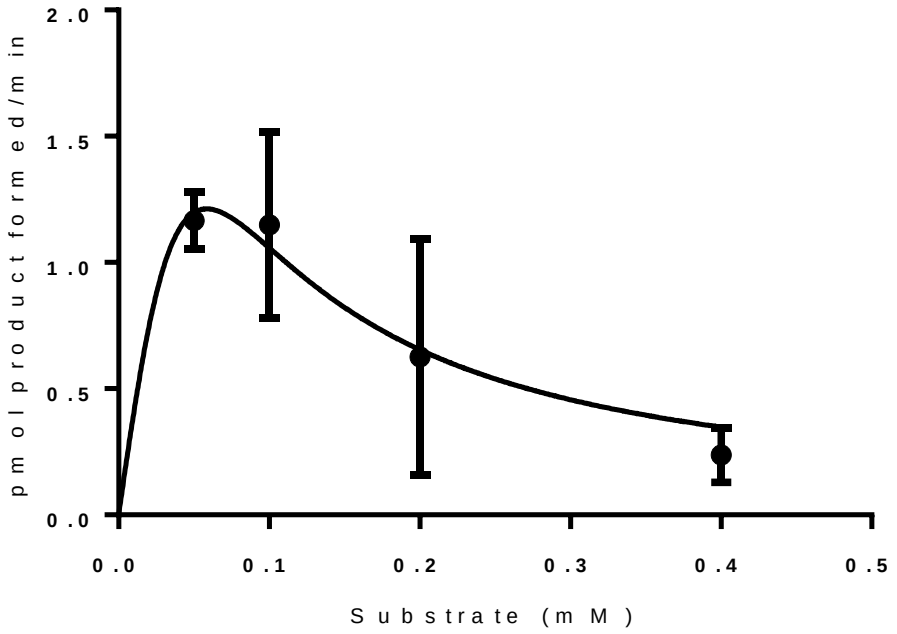


Figure 26. **Substrate saturation curve for DGT-100 on IL29**$_1$: Maximum rate of reaction occurs at 0.1 mM with maximum V_{max} of 1.1 pmol product formed per minute. Estimated K_m occurs at approximately 0.05 mM. Enzyme is inhibited by increasing concentration of substrate. Reaction performed over three biological replicates.

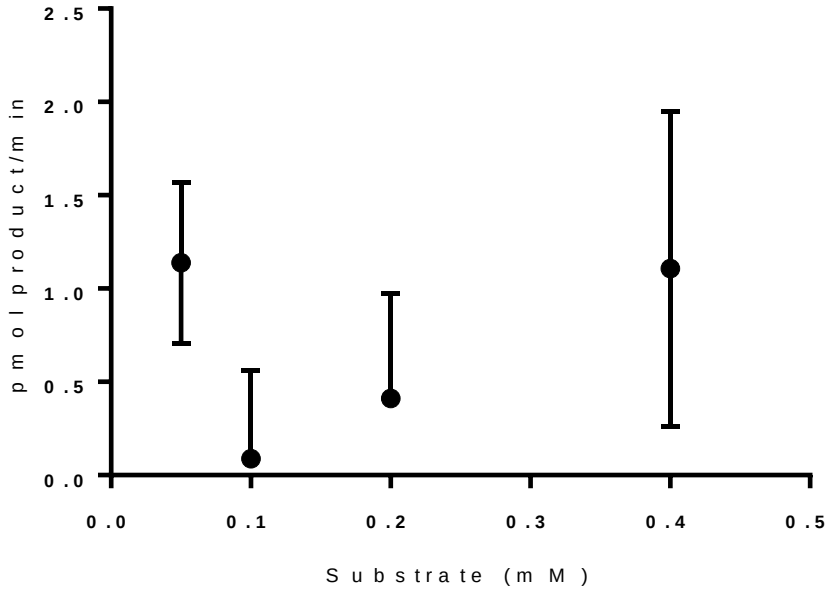


Figure 27. **Substrate saturation curve for CGT-100 on IL29$_1$**: Maximum rate of reaction occurs at 0.1 mM with maximum V_{max} of 1.1 pmol product formed per minute. Enzyme is inhibited by increasing concentration of substrate. Reaction occurred over three biological replicates.

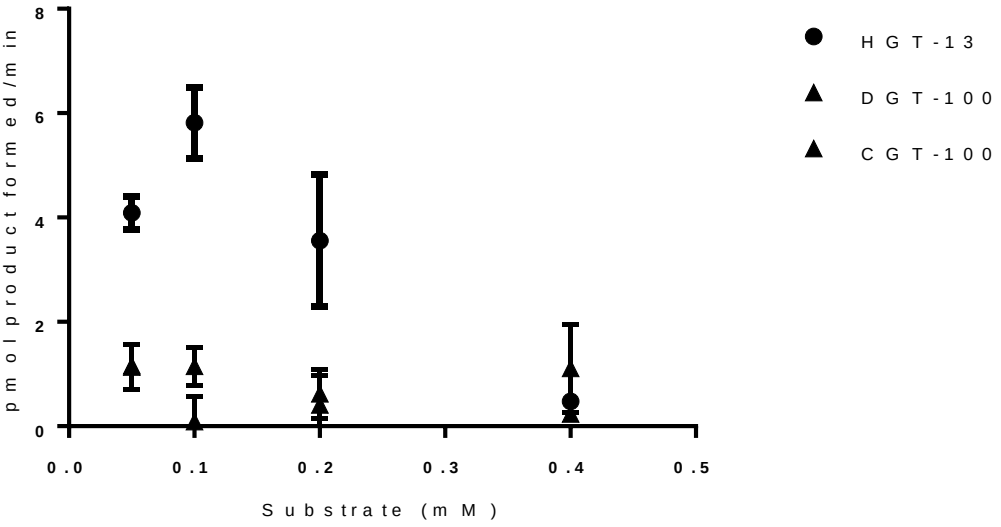


Figure 28. **Compilation of Substrate Saturation curves of GT-27 enzymes on BDP-IL29$_1$** : Overall rates of reaction are highest with HGT-13 on IL29$_1$.

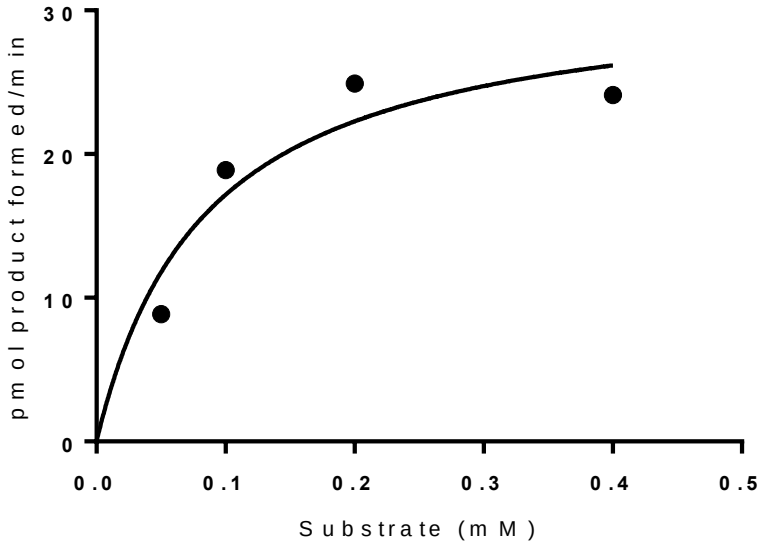


Figure 29. Substrate saturation curve for HGT-13 on IL29$_2$: Enzyme is slightly inhibited by increasing concentration of substrate. V_{max} occurs at 31.72 pmol product formed per minute with a K_m of 0.085 mM.

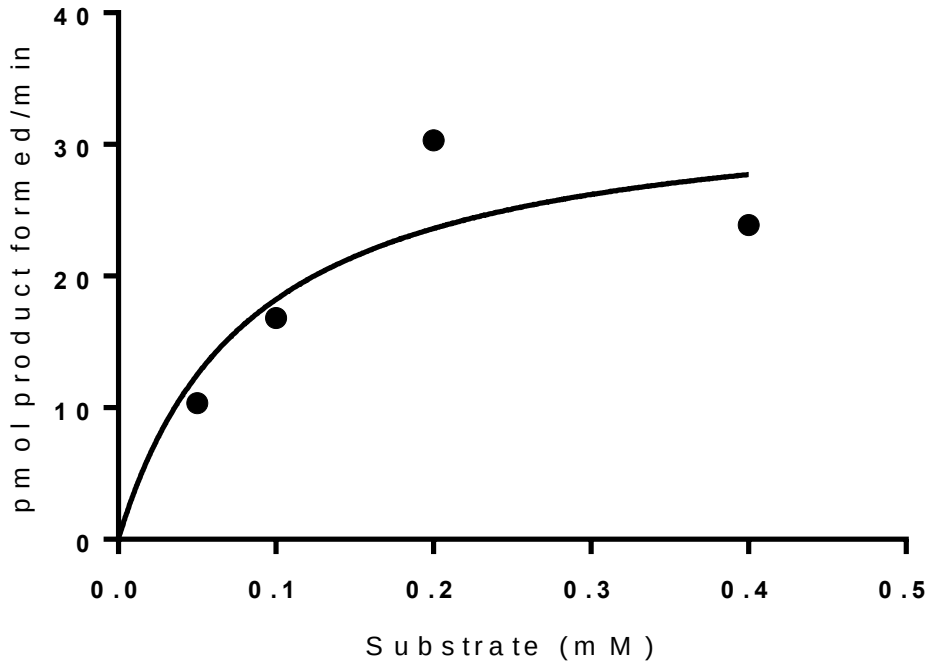
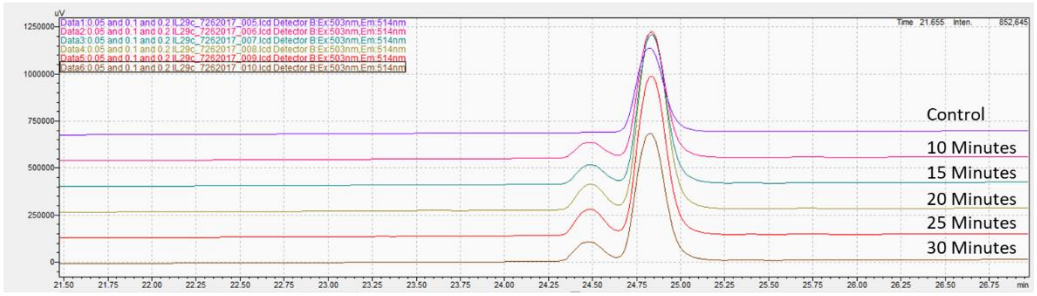


Figure 30. Substrate saturation curve for DGT-100 on IL29$_2$: Curve suggests inhibition at higher concentrations of substrate. V_{max} occurs at 33.55 pmol product formed per minute with a K_m of 0.084 mM.

A



B

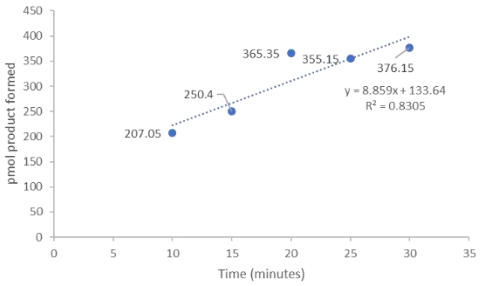


Figure 31. Reaction Progress Curve for HGT-13 on 0.05 mM IL29$_2$: (A) Chromatogram traces of HGT-13 on 0.05 mM IL29$_2$ with product peaks represented by the peaks to the left and reactant to the right and (B) their corresponding product formation quantities. Area underneath product curves were calculated as percentages and then used to calculate product formed based on available substrate quantities. Reactions were stopped using 50% Acetonitrile and 10 mM EDTA and diluted to 0.3 µM BODIPY-IL29$_2$.

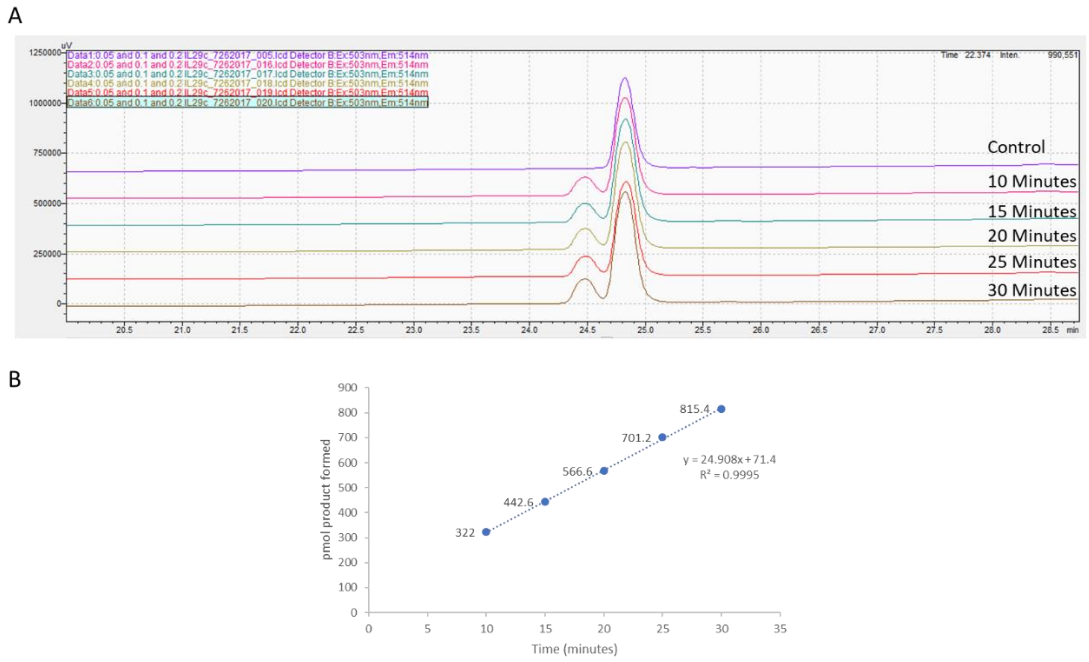


Figure 32. Reaction Progress Curve for HGT-13 on 0.2 mM IL29$_2$: (A) Chromatogram traces of HGT-13 on 0.2 mM IL29$_2$ with product peaks represented by the peaks to the left and reactant to the right and (B) their corresponding product formation quantities. Area underneath product curves were calculated as percentages and then used to calculate product formed based on available substrate quantities. Reactions were stopped using 50% Acetonitrile and 10 mM EDTA and diluted to 0.3 µM BODIPY-IL29$_2$.

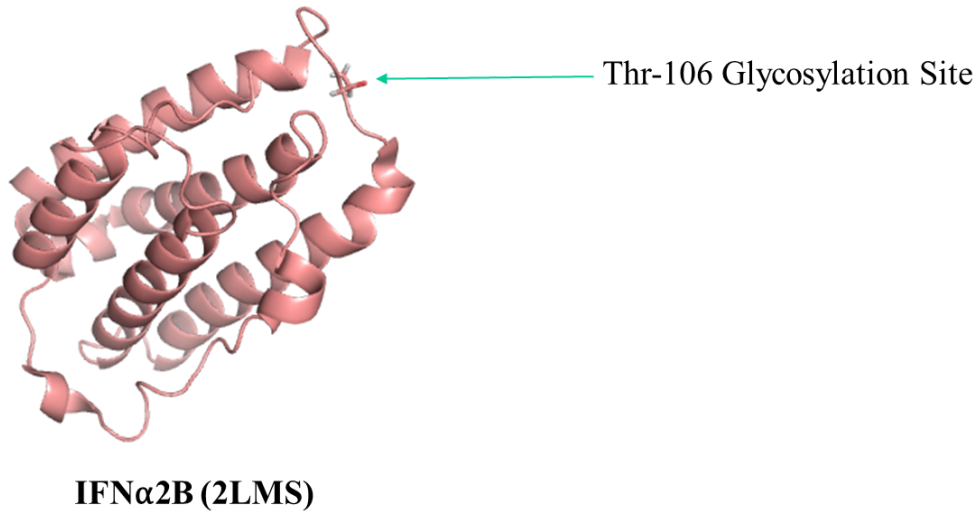


Figure 33. Crystal Structure of IFNα2B: Crystal Structure of IFNα2B with expected glycosylation site in its wild type form. Expected glycosylation site with the GB1 solubility tag occurs at Thr-186.

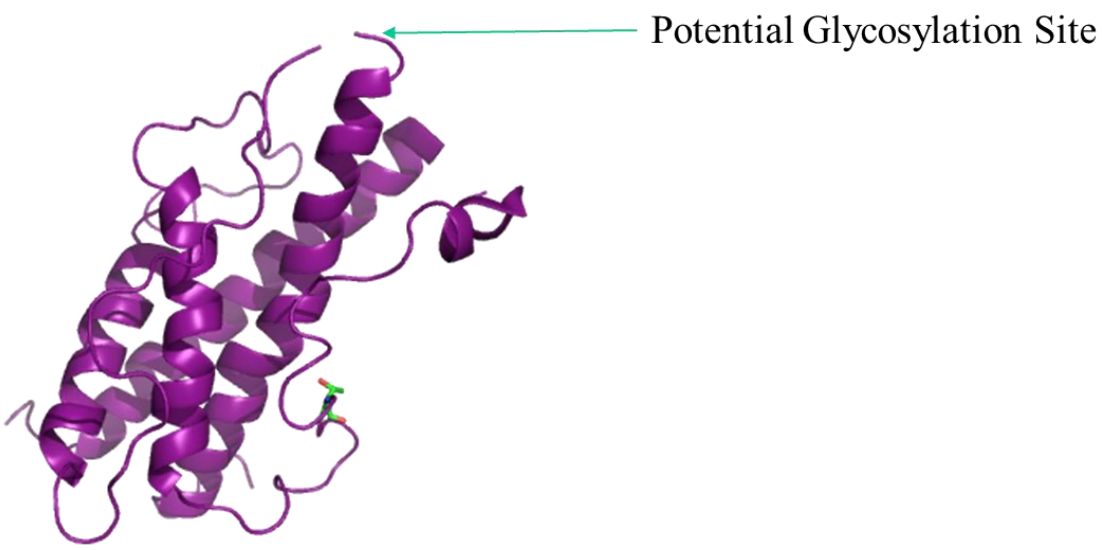


Figure 34. Crystal Structure of hGH: Crystal Structure of hGH with expected glycosylation site in its wild type form. Expected glycosylation site in the GB1 fusion occurs at T-140.

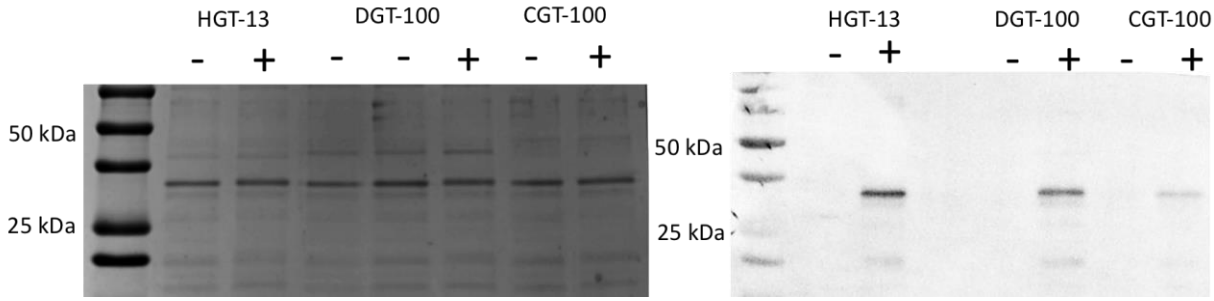


Figure 35. SBA Lectin Blot analysis using 0.4 µg/µL of lectin on IL29-3G from three GT-27 family enzymes: Left pane shows 15% SDS gels with assayed enzymes showing slight gel shifts associated with increasing glycan presence. Right pane shows 10 second exposure lectin blots with superimposed colorimetric ladder. Lanes denoted (-) excluded UDP-GalNAc from the reaction. Mass of IL29-3G is 29 kDa.

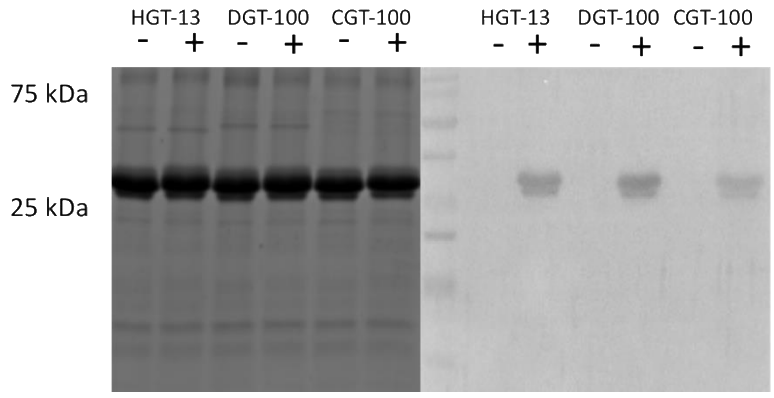


Figure 36. **SBA Lectin Blot analysis using 0.4 µg/µL of lectin on IFNα2B from three GT-27 family enzymes**: Left pane shows 15% SDS gels with assayed enzymes showing slight gel shifts associated with increasing glycan presence. Right pane shows 10 second exposure lectin blots with superimposed colorimetric ladder. Lanes denoted (-) excluded UDP-GalNAc from the reaction. Mass of IFNα2B is 28 kDa.

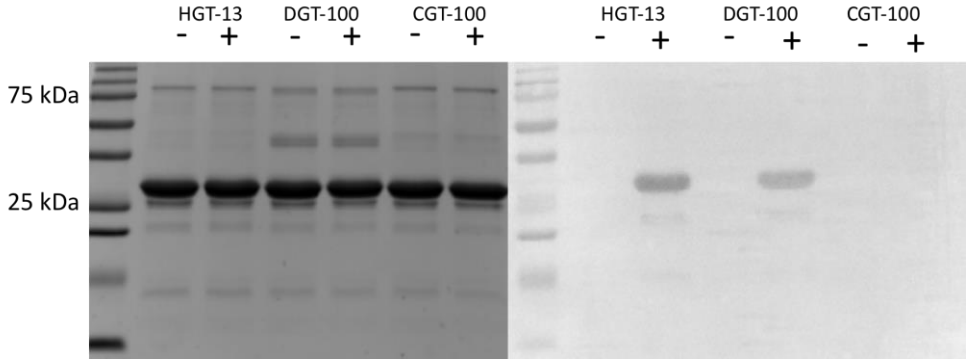


Figure 37. **SBA Lectin Blot analysis using 0.4 µg/µL of lectin on hGH from three GT-27 family enzymes**: Left pane shows 15% SDS gels with assayed enzymes showing slight gel shifts associated with increasing glycan presence. Right pane shows 10 second exposure lectin blots with superimposed colorimetric ladder. Lanes denoted (-) excluded UDP-GalNAc from the reaction. Mass of hGH is 28 kDa.

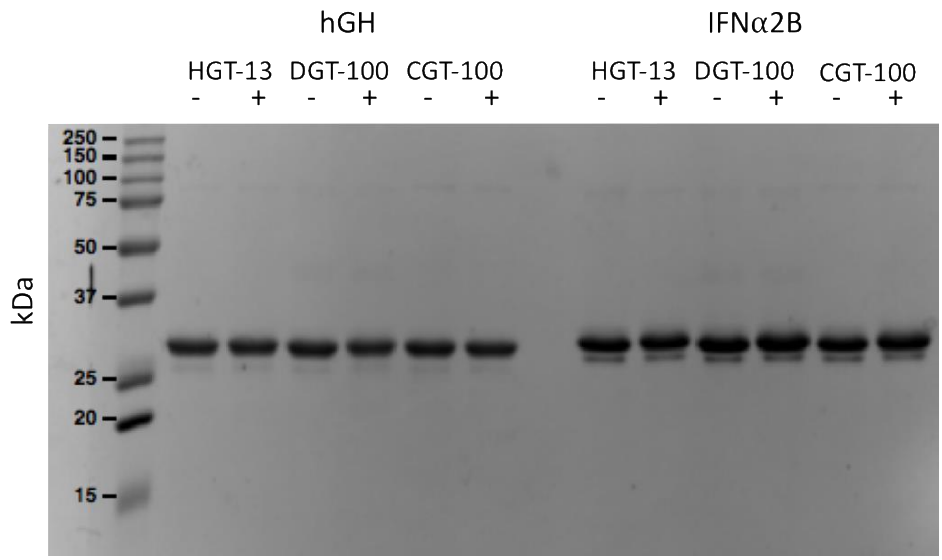


Figure 38. SDS-PAGE of Lectin Blotting Reactions on GB1-hGH and GB1-IFNα2B: SDS-PAGE of GalNAcT2 reactions on GB1-hGH and GB1-IFNα2B showing band shifts in response to glycosylation. Substrate protein content is 3.7 µg.

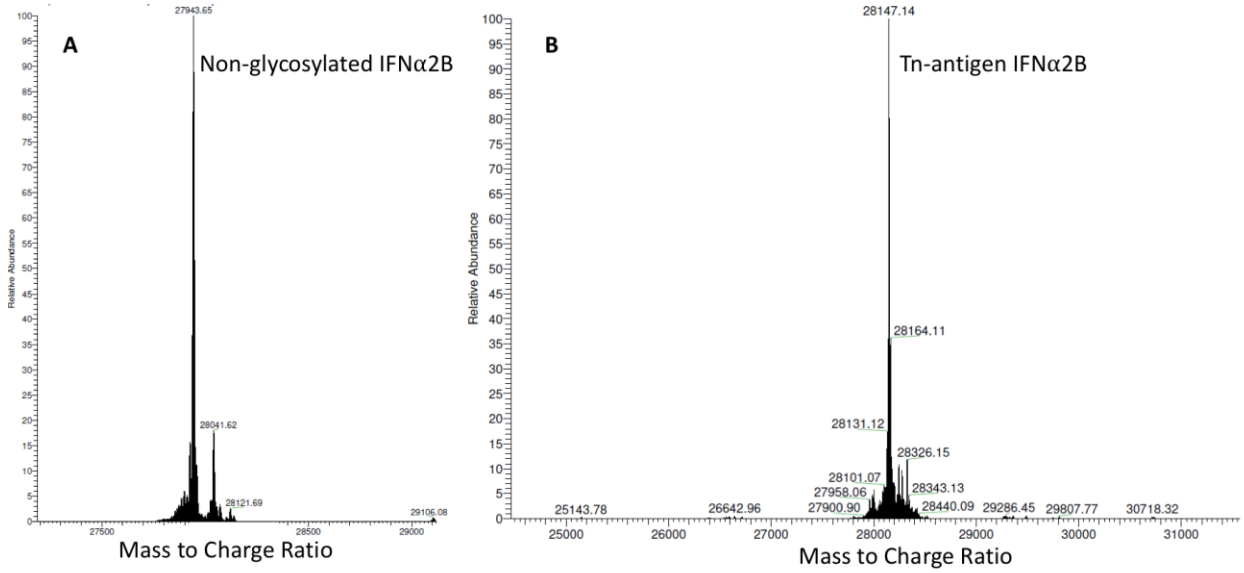


Figure 39. Mass Spectrometry Analysis of HGT-13 on IFNα2B: Mass Spectrometry analysis showing non-glycosylated (pane A) and glycosylated (pane B) GB1-IFNα2B. Masses for non-glycosylated and glycosylated forms are 27943 Da and 28147 Da respectively.

Discussion

Work performed by Lauber *et al.* (2015) showed the possibility of expressing active GalNAcT2 in *E. coli* from *Homo sapiens.* I have been able to express and purify active recombinant GalNAcT2 from *D. melanogaster* and *C. elegans.* Both peptide and protein substrates were tested and to our knowledge this work shows the first effective usage of *D. melanogaster* and *C. elegans* glycosyltransferases sourced from *E. coli*.

Work from this thesis shows that the GT-27 enzymes produced from *E. coli* can work in concert with other enzymes. The *in vitro* capabilities have been shown here particularly with the GalNAcT2 and C-4 hexNAc epimerases. Work submitted by Du *et al.* has shown the HGT-13 enzyme has been producing glycosylated substrates *in vivo*. These findings suggest a GalNAcT2 isoform excluding its lectin domain can be incorporated into a prokaryotic glycosylation system without loss of function. Previous exclusion of the lectin domain was also not seen to have significant activity differences for non-glycosylated substrates but only showed negative effects on activity of glycopeptides (Fritz *et al.* 2006; Lira-Navarrete *et al.* 2015). As this work focused mainly on initiating the *O*-glycan chain, the effects between glycopeptide and non-glycosylated peptide substrates have not been explored. This finding in literature however can back the finding that the *H. sapiens* isoform without is lectin domain is a suitable candidate to initiate glycosylation on human therapeutic proteins.

Initial experiments were focused on which plasmid expression system to use for the purification of the human GalNAcT2, for this I explored pCWMalET and pMalC5X as potential expression vectors. The difference between the two was that the NEB plasmid had a longer linker region of 23 amino acids compared to 17 and is cleavable by Factor Xa in comparison to thrombin. Yang *et al.* (2015) showed the increase in activity of onconase fusion proteins in a

Pichia Pastoris expression system suggesting longer linkers have resulting in higher activities and this made the usage of pMalC5X a better candidate before the experiments. Unfortunately, apart from the *H. sapiens* isoform, I was unable to produce *D. melanogaster* and *C. elegans* GalNAcT2s. It is likely this longer linker region made it difficult to obtain full-length protein due to the assumed constant degradation of it in the cytoplasm through proteolytic activity around the linker region. Full length fusion protein was not observed for the *D. melanogaster* and *C. elegans* proteins. As the pCWMalET could produce full length fusion for all three isoforms of GalNAc-T2, this was the obvious expression plasmid to proceed with.

Nguyen *et al.* (2011) reported the use of hPDI in order to facilitate expression of complex proteins in *E. coli.* I expressed the pCWMalET plasmids in strains that had hPDI plasmids already expressed and noticed a doubling in activity of the GalNAcT2 enzymes. These enzymes still produced MBP degradation products regardless of the improvement of activity. Lower induction concentrations of IPTG were observed to produce more enzyme. This is unsurprising as Malakar *et al.* (2012) noticed that higher IPTG concentrations resulted in a higher metabolic cost on *E. coli*. Therefore, with minimal IPTG it is obvious that there is a higher yield of active enzyme. These experiments resulted in a total average enzyme content of 10 mg per 250 mL of culture. Further research as far as prokaryotic hosts producing recombinant enzymes can explore the inclusion of the folding chaperone (hPDI) in the genome of the host *E. coli* strain. This protocol did produce a higher protein content than seen in Horynova *et al.* (2012). They do have a lower mass of enzyme as they only have secretory signal Igκ (65.8 kDa total fusion product) while I had a much larger MalE fusion product (~88kDa). Another potential exploration route is the use of a temperature sensitive promoter to ensure enzymes are properly folded. San-Miguel *et al.* (2013) discovered induction at OD_{600}=1 after 48-72 h 4^{o}C would increase soluble protein

yield by three-fold. The inclusion of hPDI can potentially increase total active protein content both for glycosylation systems and therapeutic protein production systems in *E. coli*. A comparison would be interesting between temperature based induction or IPTG induction systems to see which would be more effective to produce complex enzymes.

Coupled activity with different enzymes were tested *in vitro* using two different published C-4 hexose epimerases (both published epimerases can epimerize galactose and GalNAc). One characterized from *E. coli* O86:B7 by Guo *et al.* (2006) and another characterized from a *Campylobacter jejuni* source by Bernatchez *et al.* (2005). I noted successful reactions, where function controls excluding epimerase and sugar showed no activity. Without the presence of UDP-GlcNAc, the C-4 hexose epimerase enzyme is unable to generate UDP-GalNAc to be used as a substrate for the GalNAc Transferase reaction and therefore glycosylation does not occur. Lack of activity in the epimerase exclusion lane signifies the produced enzyme does not have GlcNAc transferase activity. The GalNAc transferase drives the epimerization of UDP-GlcNAc to GalNAc as GalNAc becomes added to serine and threonine residues pushing the system's equilibrium towards the formation of more GalNAc. Comparison between the two epimerases were not performed in this work however literature states the *E. coli* sourced epimerase has a K_m value of 0.37 mM for UDP-GalNAc and 0.32 mM for UDP-GlcNAc (Guo *et al.* 2006), while the *C. jejuni* sourced epimerase shows K_m values of 0.109 mM for UDP-GlcNAc and 0.107 mM for UDP-GalNAc (Bernatchez *et al.* 2005). Based on reported values alone the protein named CPG-13 would represent the ideal epimerase candidate for future reactions. From experience purifying both proteins, CPG-13 was easier to purify and equal volume cultures at least ten times more total enzyme content than ECE-01. However, the *E. coli* epimerase shows a slight preference towards the formation of UDP-GalNAc. This comparison

should be further explored as the CPG-13 protein was purified with the aid of a MalE fusion while ECE-01 was done so with a polyhistidine tag. Literature examples described earlier do show increasingly stable and soluble proteins with the MalE fusion.

Using *isoglyp* software (University of Texas at El Paso, El Paso, Tx; http://isoglyp.utep.edu/index.php) the substrate peptides were predicted to be great substrates showing scores of 41 and 19 for $IL29_1$ and $IL29_2$ respectively. The scores indicate the predicted rate of glycosylation where 1 would indicate a neutral rate of glycosylation. Scores larger than 1 indicate a higher likelihood of glycosylation. Furthermore, literature examples and analysis of ppGalNAcT2 showed that a proline residue +3 from the glycosylation site (S/T) increases the chance of glycosylation at that serine or threonine residue (Gerken *et al.* 2006) when they tested alongside random peptide substrates. Fritz *et al.* (2006) suggested this was due to the hydrophobic association with the proline and W282 on the enzyme catalytic domain. The results described in this assay may be a bit skewed as O'Connell *et al.* (1993) suggested that the glycosylation of threonine residues is more effective *in vitro* than *in vivo*.

Kinetic assays failed to saturate the enzymes with the substrate tested. $IL29_1$ which is the peptide derived from the original IL29 loop sequence did not saturate any of the enzymes as I noticed substrate inhibition after 0.1 mM concentrations of BDP-IL29a. For all the enzymes studied HGT-13 did show the highest V_{max} over three replicates. I am fortunate to have used 0.05 – 0.1 mM of BDP-$IL29_1$ for all the activity assays or I may not have noticed activity or been able to make the conclusions that I can make. Because a noticeable peak occurs in the curves at 0.1 mM, I can assume this is near the V_{max} even though not much clarity can be obtained from these results and any conclusions made on K_m are broad estimates. This was seen by Wandall *et al.* (1997) where their erythropoietin peptide inhibiting the enzyme activity in their enzyme

characterization. This is even though Kotenko *et al.* (2003) noted six possible *O*-glycan sites in erythropoietin. Based on what I have observed a potential glycosylation site is no guarantee of the substrate being able to saturate enzyme.

I noticed Michaelis-Menten like kinetics in the few trials I was able to perform on $IL29_2$ peptide using HGT-13 and DGT-100. The PRISM software was able to fit kinetic values on the enzymes for substrate $IL29_2$ better than $IL29_1$. Similar values were obtained for both HGT-13 and DGT-100. The reasoning behind this could be that the nonpolar BODIPY group, compounded with the more hydrophobic $IL29_1$ peptide, prevents the reaction from occurring by potentially interfering with the coordination between the UDP group and the catalytic domain of the enzyme at higher concentrations. Viewing the peptide sequence shows less hydrogen bonding capabilities in $IL29_1$. For a more accurate comparison for these enzymes, among those produced in literature, I should have performed preliminary assays for the produced enzymes on the MUC1 peptide as this is one of the first peptides used to test activity in earlier analyses of GalNAc transferases.

Using SBA lectin binding we could see all three of our compared enzymes showing activity on protein substrates IL29-3G, hGH, and IFNα2B. IL293G and IFNα2B were able to be glycosylated by all three enzymes whereas hGH was unable to be glycosylated by CGT-100. This is likely due to the physiological and immunological adaptations that glycosylation enzymes from more complex organisms such as *H. sapiens* and *D. melanogaster* have undertaken as each organism requires different defense and non-self recognition systems Furthermore, although the enzymes could glycosylate IL29-3G, this was only due to the engineered glycosylation sites input by Du *et al.* (unpublished observations), and likely the wild type would not have as strong indicators of glycosylation. The wild-type IL29 enzyme was created but was highly difficult to

purify across many attempts, even after co-expressing the wildtype IL29 fusion plasmid with hPDI. Furthermore, IL29-3G was also difficult to produce and make soluble for reactions. IFNα2B was the most successful glycosylation and purification candidate also seen from other researchers in the Wakarchuk Lab. The simple mass spectrometry analysis performed in this work showed incorporation of HexNAc. This interferon is an excellent candidate for the glycosylation system of the Wakarchuk Lab.

Lectin blotting is another successful way to qualitatively identify success of glycosylation. The verification of presence of Tn-antigen is difficult in western blotting as Loureiro *et al.* (2015) discussed the difficulty of antibody development in their specificities and low affinities. The steric hindrance with the glycan so close to the protein backbone does make it difficult for high specificity. I used SBA as the lectin, but the lectin domain from GalNAc transferases can also recognize the Tn antigen and there are other lectins that may recognize GalNAc as well such as *Vicia villosa* lectin. Yoshimura *et al.* (2012) showed the lectin domain of GalNAcT3 being able to recognize sugars on unnatural glycopeptides. This is useful as many therapeutic proteins have unnatural glycosylation sites. Un-natural glycosylation sites are sites which have been glycosylated in sites which would not be glycosylated in nature, but the introduction of the glycan improves its serum half-life.

Conclusion

I set out to produce three GalNAcT2 enzymes from *Homo sapiens*, *Drosophila melanogaster* and *Caenorhabditis elegans* and compare their activities and kinetics on synthetically produced peptides $IL29_1$ and $IL29_2$ whose sequences or their parent proteins are not normally *O*-glycosylated. I was unable to saturate the enzymes with peptide $IL29_1$ but results for $IL29_2$ look more promising and research on this may produce more desirable results on the ability of prokaryotic recombinant GalNAcT enzymes to glycosylate non-native substrates at least for isoform 2. The produced enzymes were successful in glycosylating protein substrates as well, with CGT-100 having the lowest glycosylation of tested substrates. Mass spectrometry analysis further signified the ability of *H. sapiens* isoform of GalNAcT2 being able to glycosylate IFNα2B based on increase in substrate mass. The produced enzymes are also cooperative with epimerase enzymes which can introduce the Tn-antigen on substrates from the much cheaper donor UDP-GlcNAc.

This work shows the first instance of production, to my knowledge, of the *D. melanogaster* and *C. elegans* isoforms of GalNAcT2 being produced in *E. coli.* This is the first work showing any of these enzymes on BODIPY derived peptide substrates and showing activity on full length folded proteins. However, due to the substrates being non-traditional it does make it difficult to compare activities among other works in literature. The nature of the protein production here is also different from literature; as MalE fusions were used due to them increasing solubility of proteins and the ease of their purification. The inclusion of the MalE fusion tag and the coexpression of the hPDI protein can potentially improve the production and activity of many enzymes which have yet to be produced in *E. coli*.

I should have begun glycosyltransferase assays using the MUC1 derived peptide that is commonly used. Although this substrate is not universal, it is still the tested peptide in many GalNAcT2 related literature. Assays on MUC1 may not be able to give insight on glycosylation of therapeutic substrates but they would allow comparisons across research. This convention would be able to truly explain which produced enzyme is more active in the context of this work compared to those performed earlier. The enzymes, when assayed on the IL29 derived peptides, were assumed to be active on their native substrates.

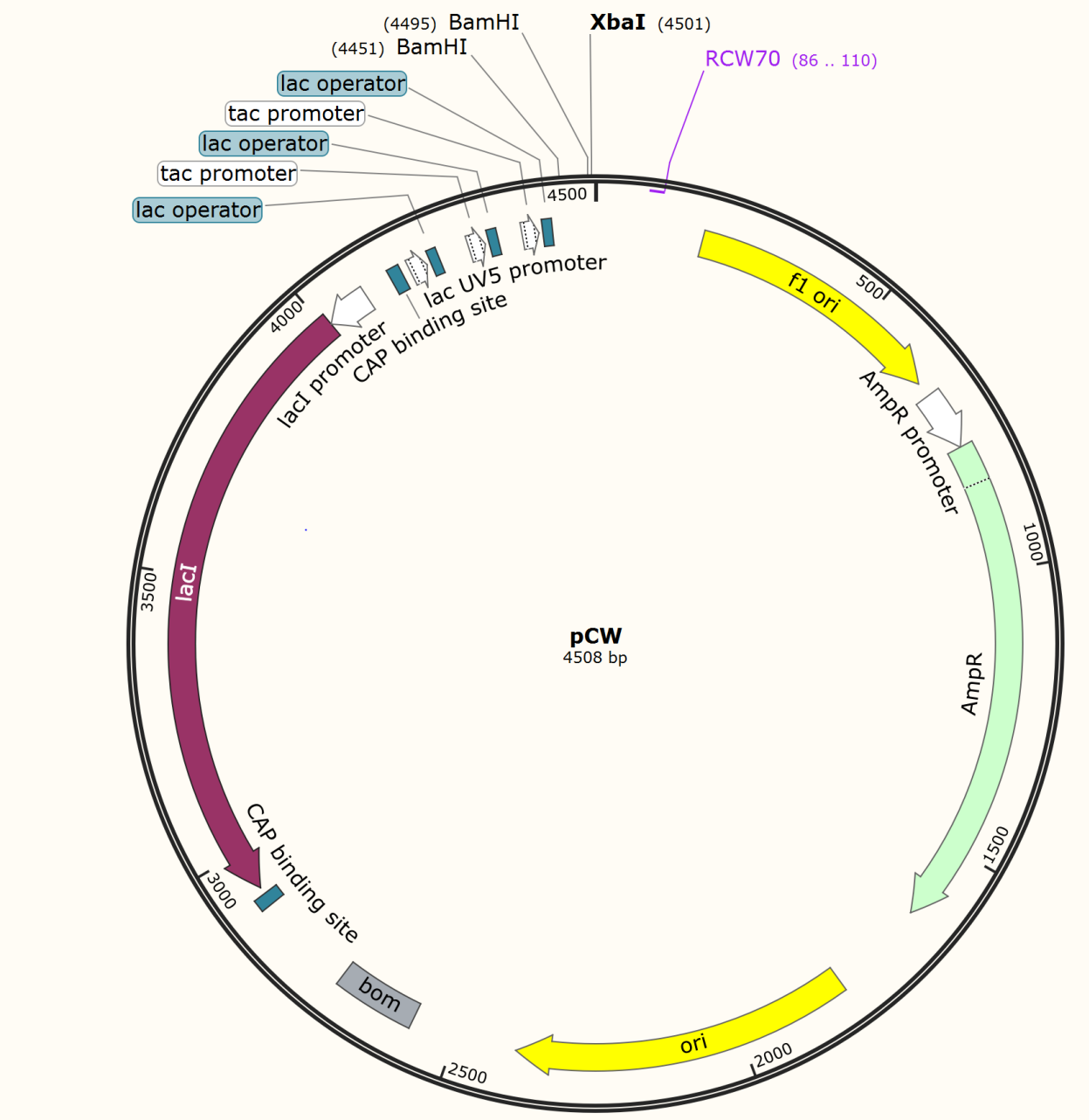
Reactions were individually stopped at different time points. This method had difficulty in producing perfect linear product formation curves. In this design as well, the bottleneck becomes production of substrate. This was the case for both peptide and protein portions of the enzyme comparison. The substrate saturation reactions used large quantities of BODIPY labelled peptide and much time was spent in producing and purifying the peptides. The proteins as well, particularly in the case of GB1-IL29-3G was highly difficult to produce. The GB1-IL29-3G protein was not soluble in higher than 0.5 mg/mL concentrations and therefore this substrate could not be used in as high concentrations as the GB1-hGH and GB1-IFNα2B were. Should further *in vitro* reactions be necessary with IL29, they should be done with commercially acquired cytokines. For future kinetics work, the enzymes should be studied in a continuous assay potentially using UDP-Glo (Zegzouti *et al.* 2013). This would make studying kinetics highly efficient as the fluorescence from the formation of free UDP would allow the reactions to be studied in real time as opposed to using stopped reactions. Furthermore, in the event of comparisons to enzymes discussed in literature, I should do initial assays involving native peptides. Native peptides may be different when comparing substrates across different sources of enzymes, particularly between human and non-human isoforms. Although *isoglyp* indicated that

the substrates would be glycosylated, it gives no guarantee of the substrate saturating the enzyme.

This work contributes to the formation of the Tn antigen in order to generate the *O*-glycan on therapeutic proteins. The enzyme without the lectin domain is capable of glycosylating free peptides as seen in literature and therefore it can glycosylate proteins as they are being produced concurrently by *E. coli*. By including the *H. sapiens* isoform without its lectin domain in the glycosylation operon, the metabolic burden on *E. coli* would be reduced. This is due to less amino acids and nutrients being dedicated to the production of a domain which for these purposes are not needed. This exclusion of the lectin domain would ensure better folding due to the decrease in available cysteines in the enzyme forming improper disulphide bonds. This would allow for the quick production of Tn-antigen containing therapeutic substrates as they are being produced by *E. coli* as well. Research from this work attempts to be the first step in a long journey of trial and error which would lead to the formation of glycosylated therapeutic proteins with complete *O*-glycan profiles capable of having therapeutic effects on patients.

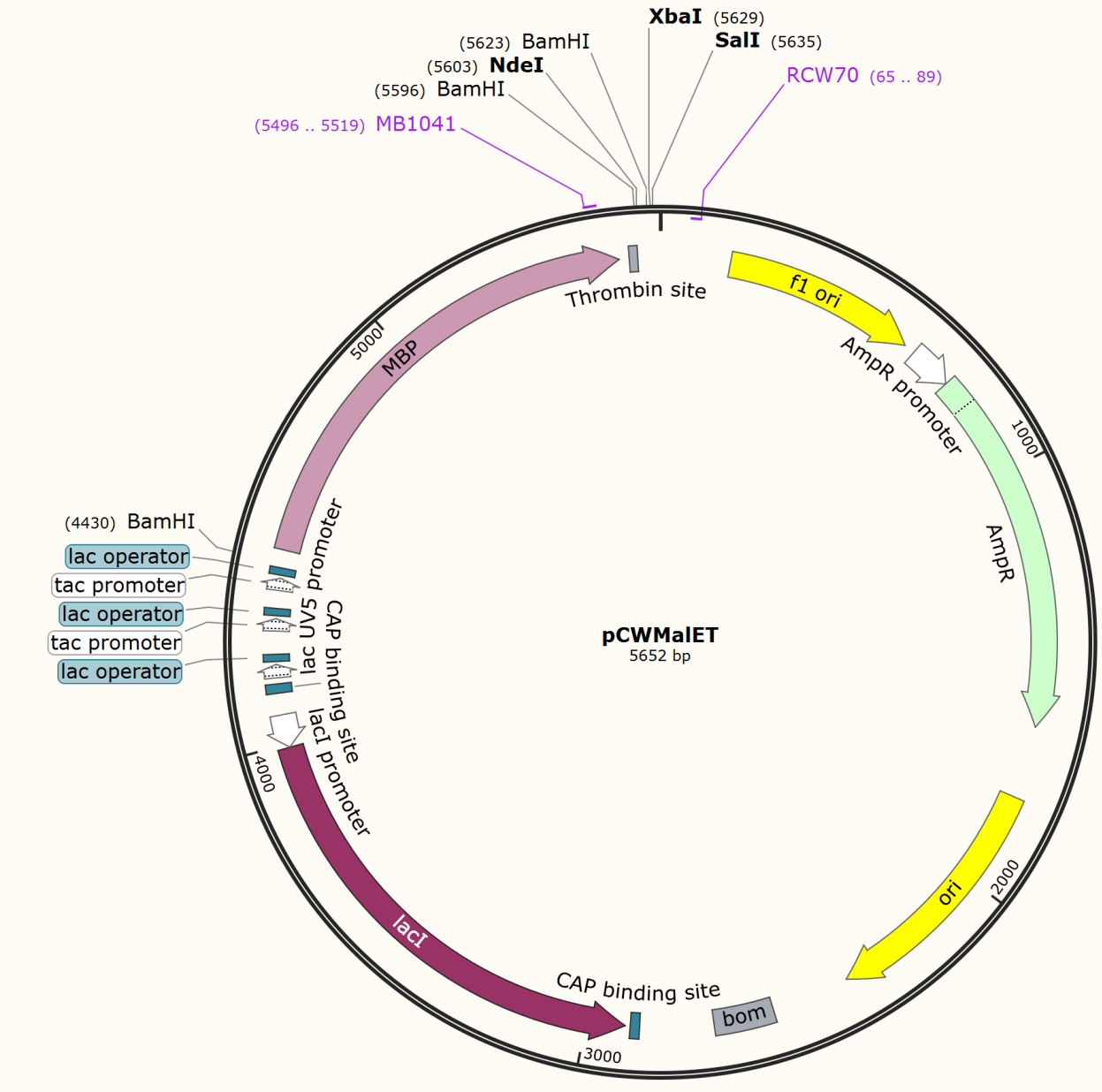
Appendices

Appendix 1: Plasmid OGO-6 operon coding for *H. sapiens* ppGalNAcT2 with its lectin domain as an MBP fusion. Features include chloramphenicol resistance (CmR), and LacI promoter for IPTG based induction.

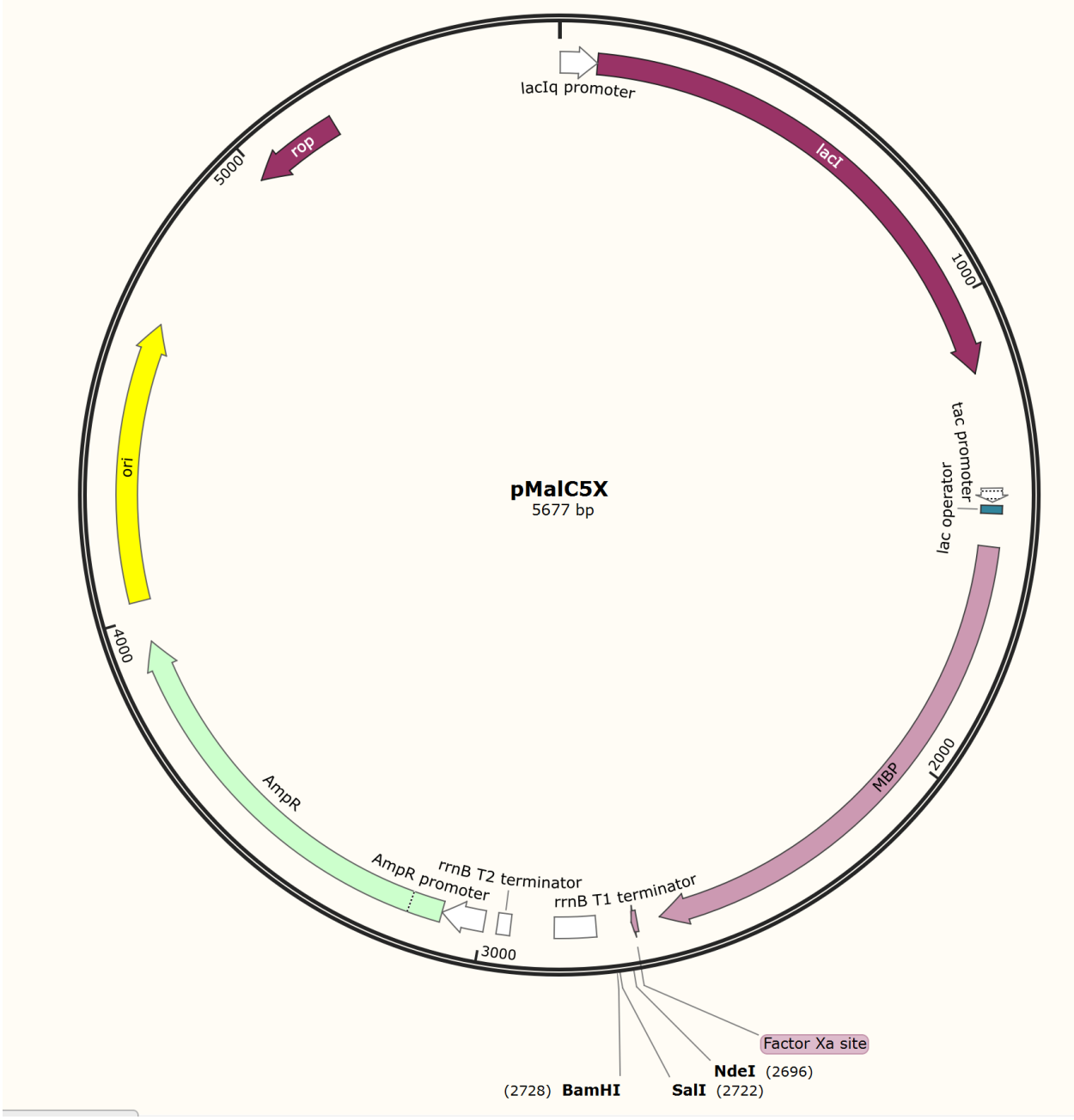


Appendix 2: Plasmid map showing vector pCW. Features include ampicillin resistance (AmpR), and LacI promoter for IPTG based induction.

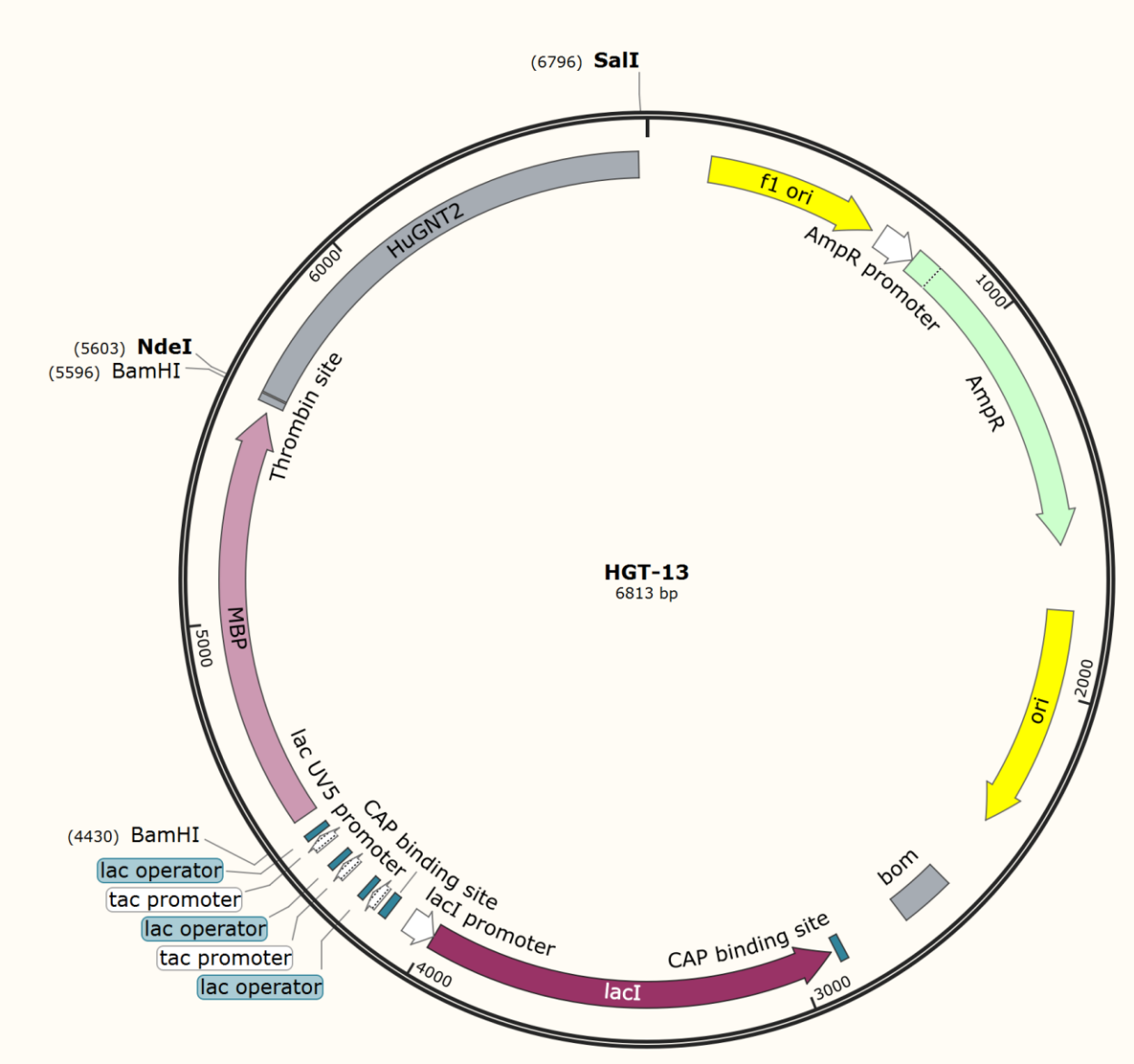
Appendix 3: Plasmid map showing the excision fragment of MBP-ppGalNAcT2 from OGO-6 from BamHI to XbaI. Features are the same as pCW vector with the inclusion of the ppGalNAcT2 enzyme with a Factor Xa linker between the enzyme and the Maltose Binding Protein(MBP).



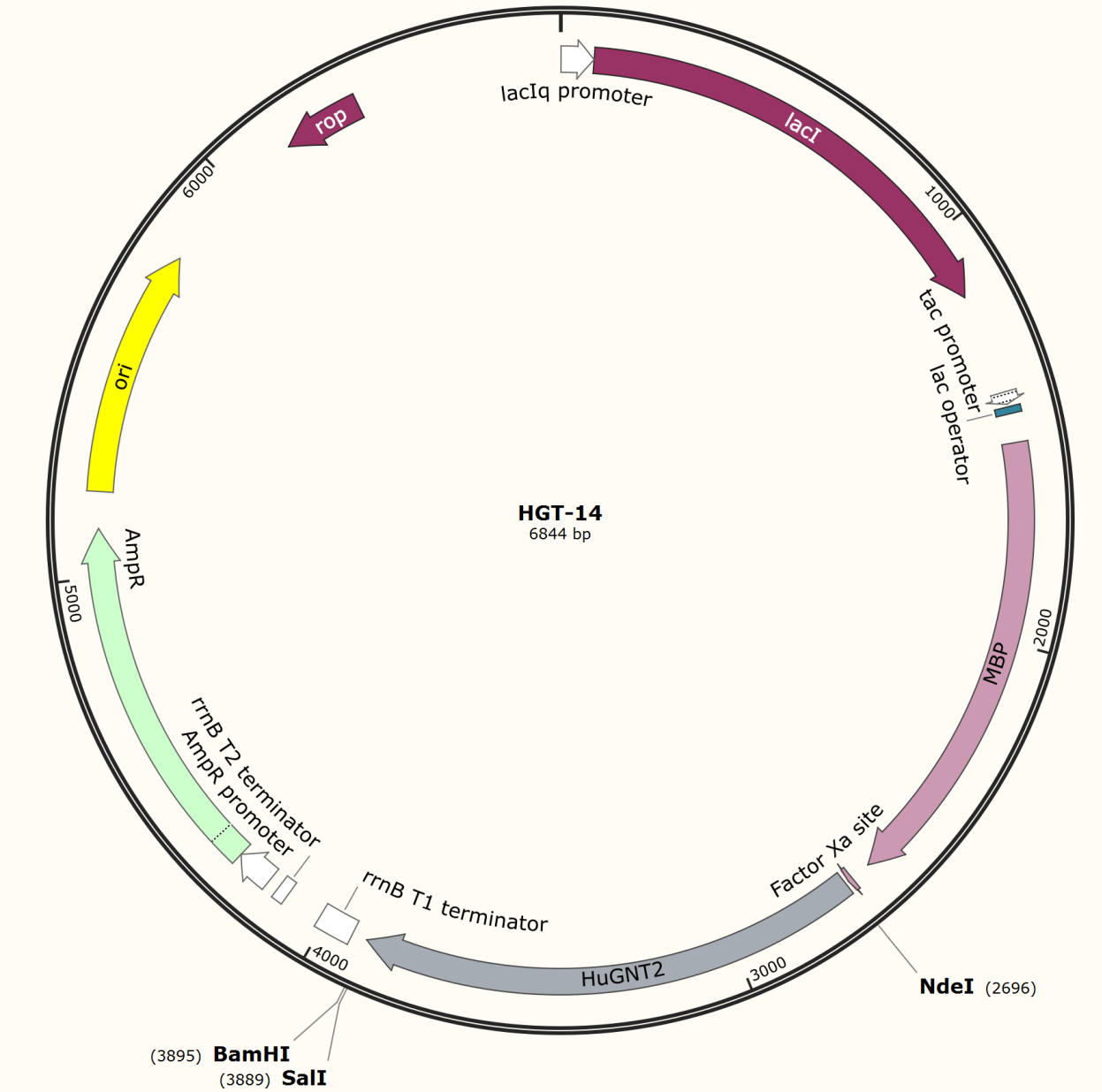
Appendix 4: Plasmid map showing vector pCWMalET. Features include ampicillin resistance (AmpR), and LacI promoter for IPTG based induction. A MBP linker is included as well with a thrombin cleavage site.



Appendix 5: Plasmid map showing vector pMalC5X. Features include ampicillin resistance (AmpR), and LacI promoter for IPTG based induction. A MBP linker is included as well with a Factor Xa cleavage site.



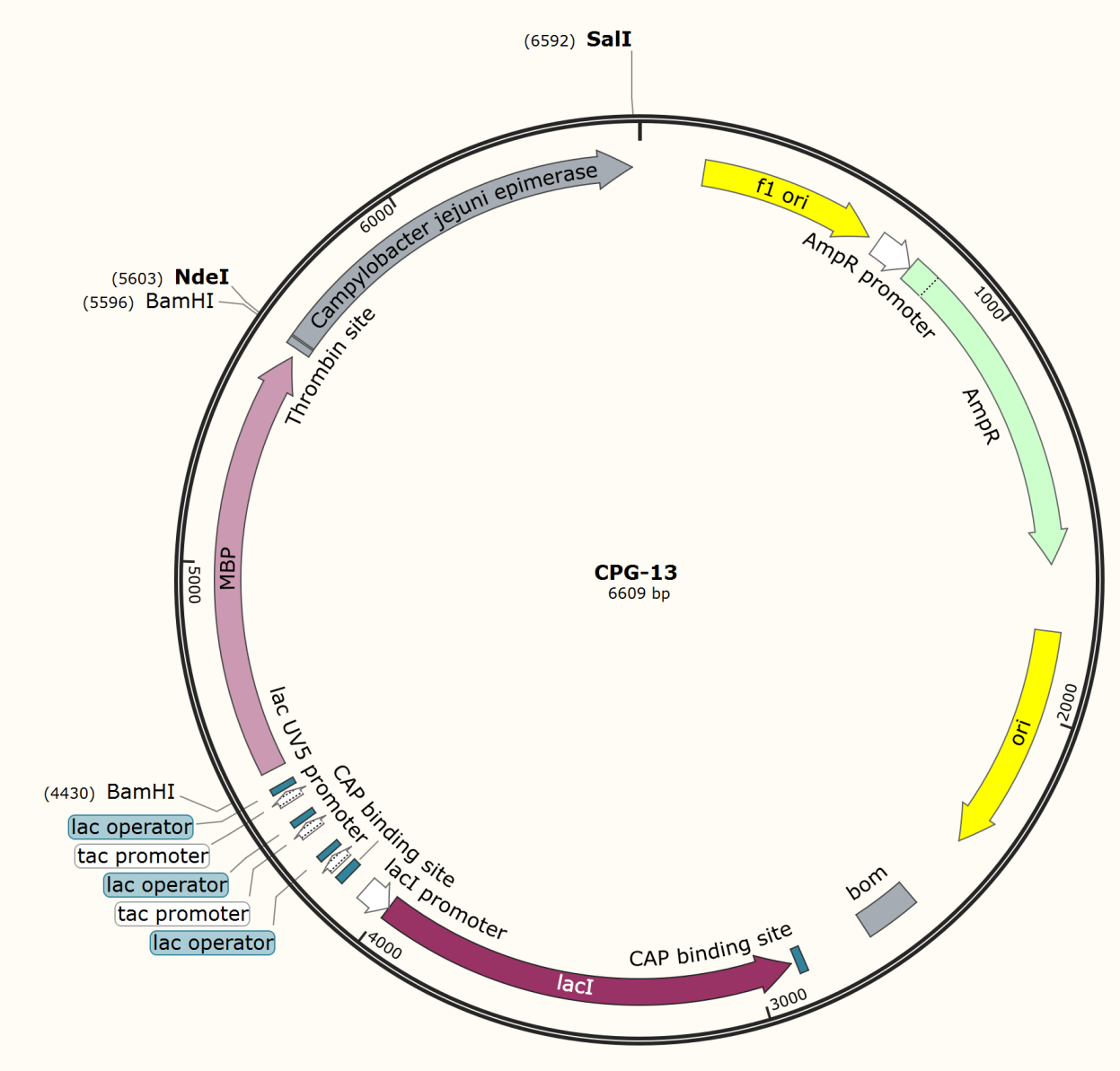
Appendix 6: Plasmid map of ppGalNAcT2 from *H. sapiens* excluding the lectin domain in plasmid pCWMalET. Features are the same as pCW vector with the inclusion of the ppGalNAcT2 enzyme with a thrombin cleavable 17 amino acid linker between the enzyme and the Maltose Binding Protein(MBP).



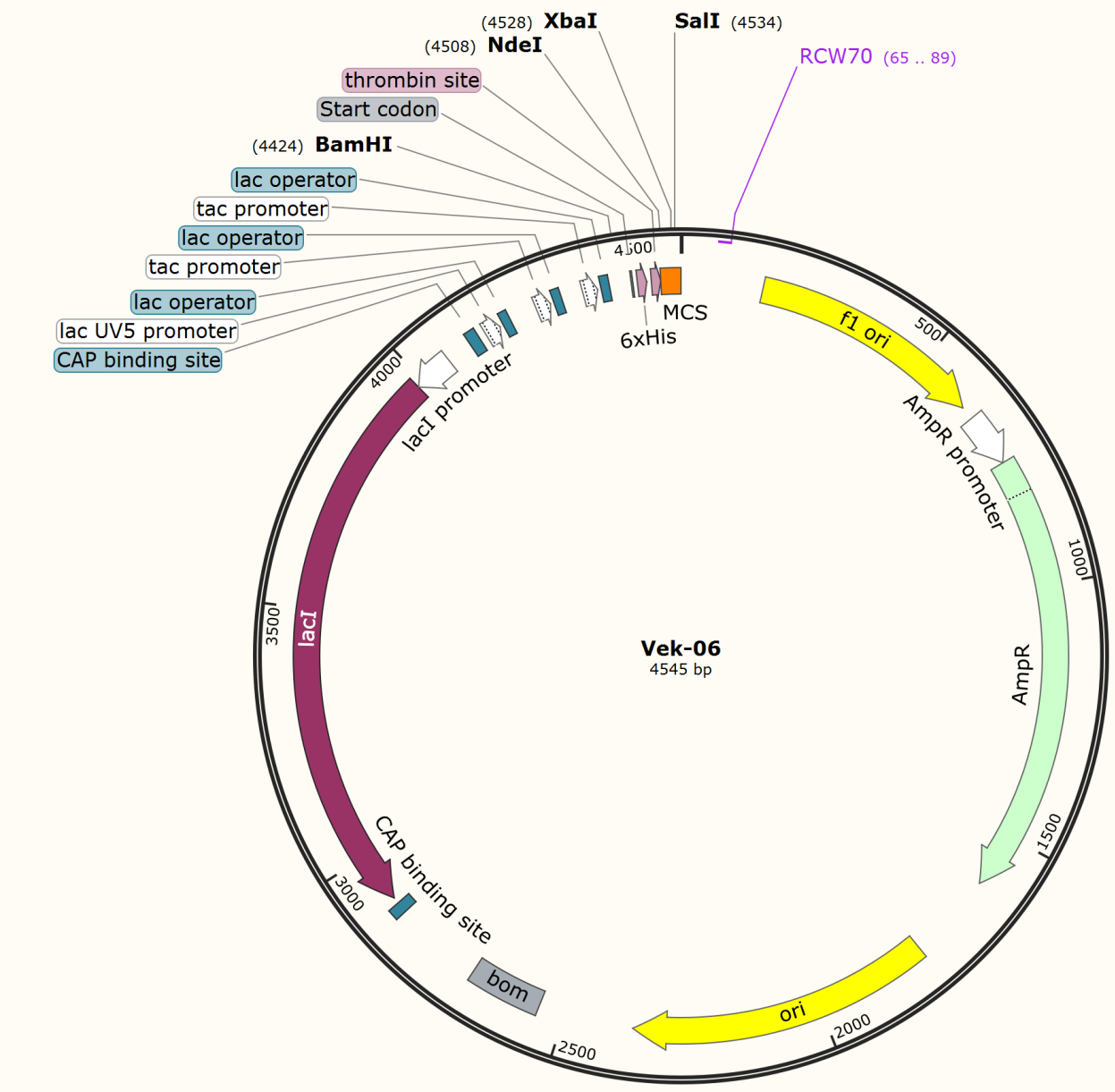
Appendix 7: Plasmid map of ppGalNAcT2 from *H. sapiens* excluding the lectin domain in plasmid pMalC5x. ppGalNAcT2 enzyme has a Factor Xa cleavable amino acid linker between the enzyme and the Maltose Binding Protein(MBP).

Appendix 8: Plasmid map of *PGANT2* from *D. melanogaster* excluding the lectin domain in plasmid pCWMalET. Features are the same as pCW vector with the inclusion of the GalNAcT2 enzyme with a thrombin cleavable 17 amino acid linker between the enzyme and the Maltose Binding Protein(MBP).

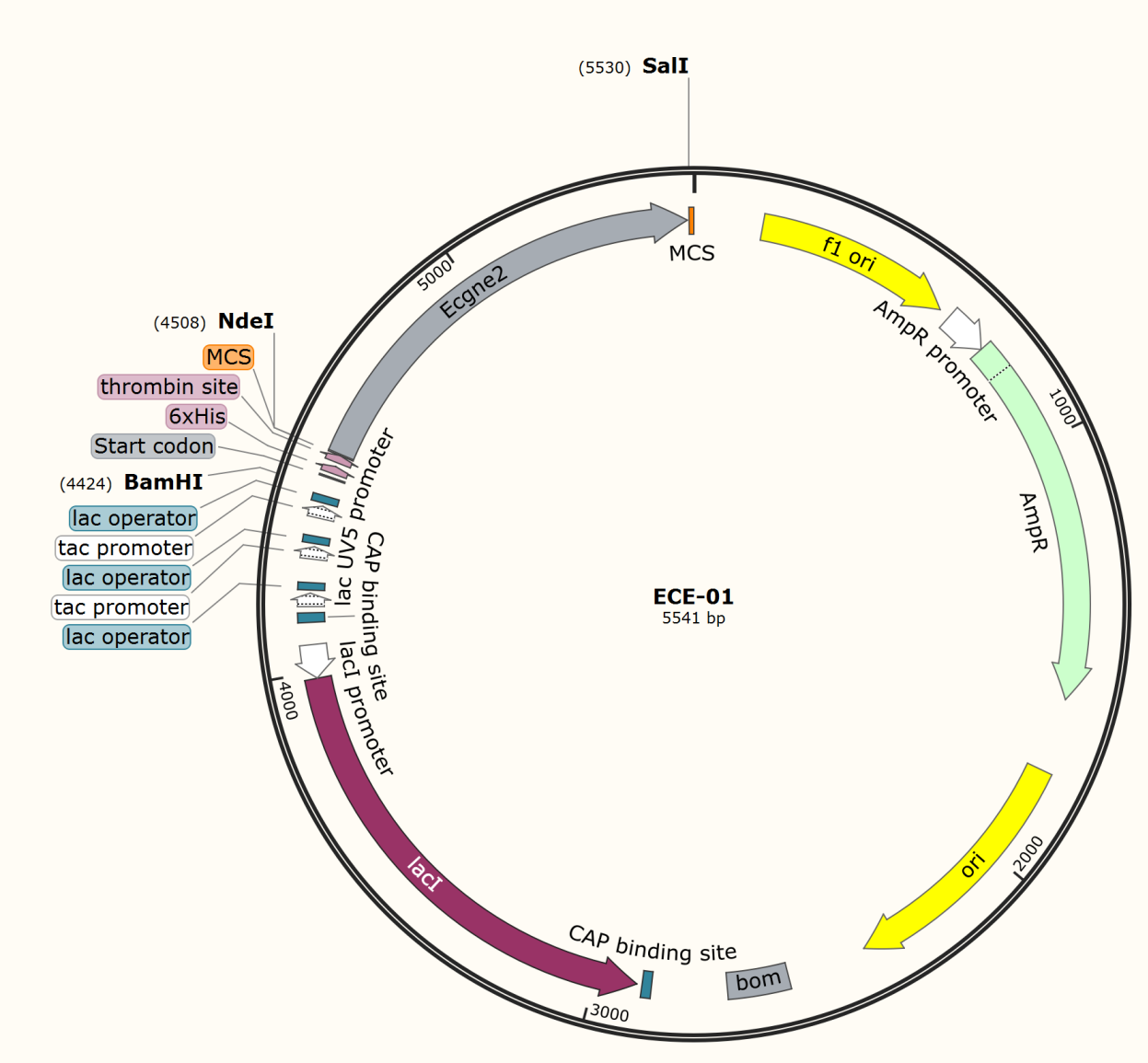
Appendix 9: Plasmid map of *Gly4* from *C.elegans* excluding the lectin domain in plasmid pCWMalET. Features are the same as pCW vector with the inclusion of the GalNAcT2 enzyme with a thrombin cleavable 17 amino acid linker between the enzyme and the Maltose Binding Protein (MBP).



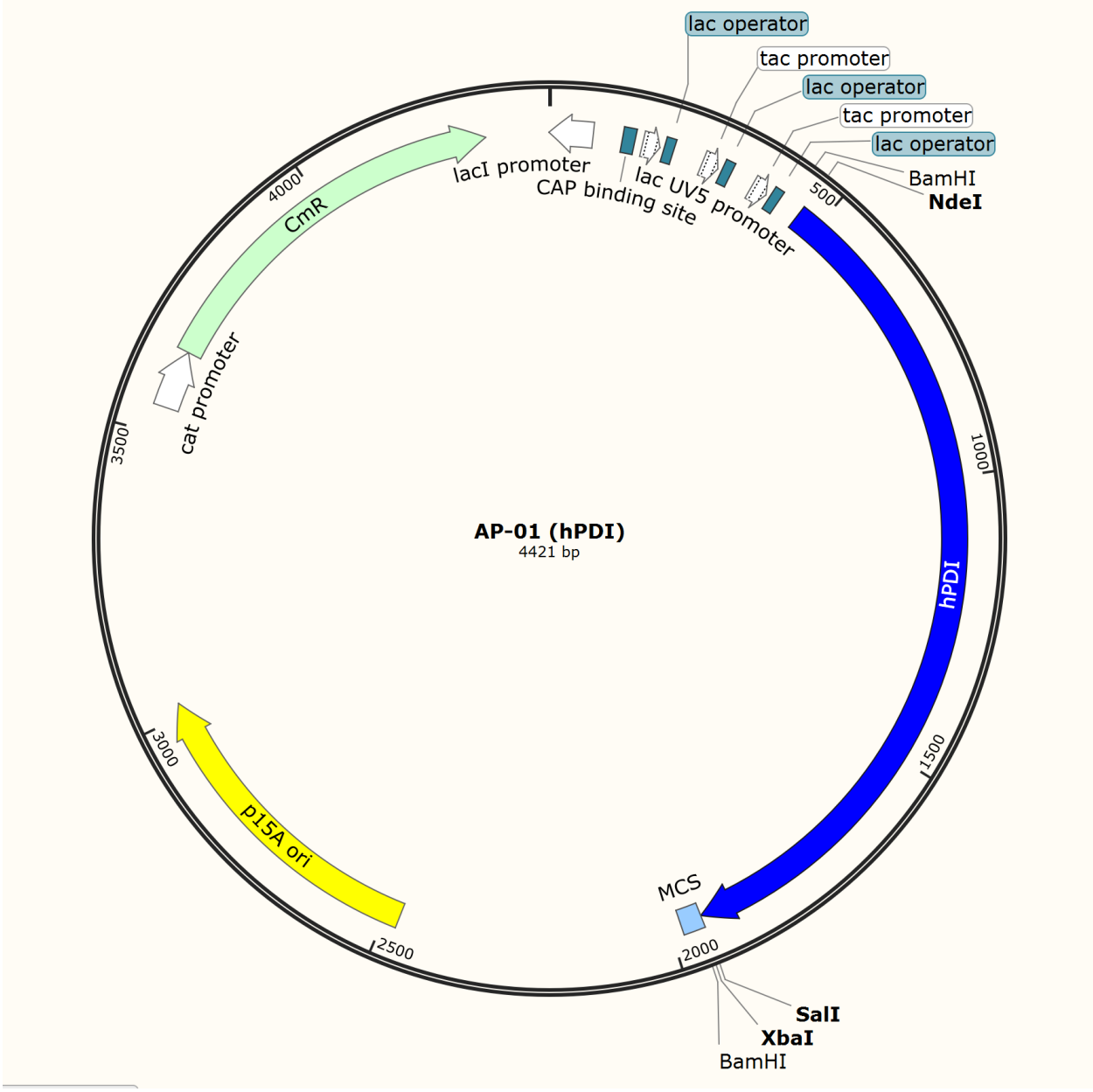
Appendix 10: Plasmid map of CPG-13 in pCWMalET.



Appendix 11: Plasmid map of VEK-06. Features include a lac promoter, N-terminal polyhistidine tag, ampicillin resistance and a thrombin cleavage site.



Appendix 12: Plasmid map of *Ecgne2* inserted into VEK-06.



Appendix 13: Plasmid AP-01 expressing human Protein Disulphide Isomerase (hPDI). Features include chloramphenicol resistance and lacUV5 promoter to drive continuous expression.

References

1. Abuchowski, A., J. R. McCoy, N. C. Palczuk, T. van Es, and F. F. Davis. 1977. 'Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase', *J Biol Chem*, 252: 3582-6.
2. Alanen, H. I., R. A. Williamson, M. J. Howard, A. K. Lappi, H. P. Jantti, S. M. Rautio, S. Kellokumpu, and L. W. Ruddock. 2003. 'Functional characterization of ERp18, a new endoplasmic reticulum-located thioredoxin superfamily member', *J Biol Chem*, 278: 28912-20.
3. Arora, Shiwani K., Alice N. Neely, Barbara Blair, Stephen Lory, and Reuben Ramphal. 2005. 'Role of Motility and Flagellin Glycosylation in the Pathogenesis of Pseudomonas aeruginosa Burn Wound Infections', *Infection and Immunity*, 73: 4395-98.
4. Asada, M., Furukawa, K., Segawa, K., Endo, T., and Kobata, A. 1997. 'Increased Expression of Highly Branched *N*-Glycans at Cell Surface is Correlated with the Malignant Phenotypes of Mouse Tumor Cells', *Cancer Research*, 57(6): 1073-1080.
5. Baeshen, M. N., A. M. Al-Hejin, R. S. Bora, M. M. Ahmed, H. A. Ramadan, K. S. Saini, N. A. Baeshen, and E. M. Redwan. 2015. 'Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives', *J Microbiol Biotechnol*, 25: 953-62.
6. 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(6): 736-756.
7. Bernatchez, S., M. Gilbert, M. C. Blanchard, M. F. Karwaski, J. Li, S. Defrees, and W. W. Wakarchuk. 2007. 'Variants of the beta 1,3-galactosyltransferase CgtB from

the bacterium Campylobacter jejuni have distinct acceptor specificities', *Glycobiology*, 17: 1333-43.

8. Bernatchez, S., C. M. Szymanski, N. Ishiyama, J. Li, H. C. Jarrell, P. C. Lau, A. M. Berghuis, N. M. Young, and W. W. Wakarchuk. 2005. 'A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in Campylobacter jejuni', *J Biol Chem*, 280: 4792-802.
9. Brygier, J., Gelbcke, M., Guermant, C., Nijs, M., Baeyens-Volant, D., and Looze, Y. 1993. 'Covalent attachment of poly(ethyleneglycol) to peptides and proteins', *Applied Biochemistry and Biotechnology*, 42: 127-135.
10. Carter, P. J. 2006. 'Potent antibody therapeutics by design', *Nat Rev Immunol*, 6(5): 343-57.
11. Chan, Barden, Michelle Clasquin, Gromoslaw A. Smolen, Gavin Histen, Josh Powe, Yue Chen, Zhizhong Lin, Chenming Lu, Yan Liu, Yong Cang, Zhonghua Yan, Yuanfeng Xia, Ryan Thompson, Chris Singleton, Marion Dorsch, Lee Silverman, Shin-San Michael Su, Hudson H. Freeze, and Shengfang Jin. 2016. 'A mouse model of a human congenital disorder of glycosylation caused by loss of PMM2', *Human Molecular Genetics*, 25: 2182-93.
12. Dobson, Claire L., Paul W. A. Devine, Jonathan J. Phillips, Daniel R. Higazi, Christopher Lloyd, Bojana Popovic, Joanne Arnold, Andrew Buchanan, Arthur Lewis, Joanne Goodman, Christopher F. van der Walle, Peter Thornton, Lisa Vinall, David Lowne, Anna Aagaard, Lise-Lotte Olsson, Anna Ridderstad Wollberg, Fraser Welsh, Theodoros K. Karamanos, Clare L. Pashley, Matthew G. Iadanza, Neil A. Ranson, Alison E. Ashcroft, Alistair D. Kippen, Tristan J. Vaughan, Sheena E.

Radford, and David C. Lowe. 2016. 'Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo', *Scientific Reports*, 6: 38644.

13. Donner, T. 2015. 'Insulin – Pharmacology, Therapeutic Regimens And Principles Of Intensive Insulin Therapy', *Endotext*, South Dartmouth (MA): MDText.com, Inc. Available from https://www.ncbi.nlm.nih.gov/books/NBK278938.
14. Elhammer, A. P., Poorman, R. A., Brown, E., Maggiora, L.L., Hoogerheide J. G., and Kézdy, F. J. 1993. 'The specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase as inferred from a database of in vivo substrates and from the in vitro glycosylation of proteins and peptides', *Journal of Biological Chemistry*, 268, 10029-10038.
15. Fisher, A. A. 1978. 'Immediate and delayed allergic contact reactions to polyethylene glycol', *Contact Dermatitis*, 4: 135-8.
16. Fleischer, B. 1981. 'Orientation of glycoprotein galactosyltransferase and sialyltransferase enzymes in vesicles derived from rat liver Golgi apparatus', *J Cell Biol*, 89(2): 246-55.
17. Freeze, H. H., E. A. Eklund, B. G. Ng, and M. C. Patterson. 2012. 'Neurology of inherited glycosylation disorders', *Lancet Neurol*, 11: 453-66.
18. Fritz, T. A., J. Raman, and L. A. Tabak. 2006. 'Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-2', *J Biol Chem*, 281: 8613-9.
19. Gąciarz, Anna, Narendar Kumar Khatri, M. Lourdes Velez-Suberbie, Mirva J. Saaranen, Yuko Uchida, Eli Keshavarz-Moore, and Lloyd W. Ruddock. 2017.

'Efficient soluble expression of disulfide bonded proteins in the cytoplasm of Escherichia coli in fed-batch fermentations on chemically defined minimal media', *Microbial Cell Factories*, 16: 108.

20. Gerken, T.A., Tep, C., and Rarick, J. 2004. 'Role of peptide sequence and neighboring residue glycosylation on the substrate specificity of the uridine 5'-diphosphate-alpha-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyl transferases T1 and T2: kinetic modeling of the porcine and canine submaxillary gland mucin tandem repeats', *Biochemistry*, 43(30): 9888-900.
21. Gerken, T. A., J. Raman, T. A. Fritz, and O. Jamison. 2006. 'Identification of common and unique peptide substrate preferences for the UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates', *J Biol Chem*, 281: 32403-16.
22. Gillies, S.D., Young, D., Lo, K. M., and Roberts S. 1993. 'Biological activity and in vivo clearance of antitumor antibody/cytokine fusion proteins', *Bioconjug Chem*, 4(3): 230-5.
23. Grace, M.J., Lee, S., Bradshaw, S., Chapman, J., Spond, J., Cox, S., DeLorenzo, M., Brassard, D., Wylie, D., Cannon-Carlson, S., Cullen, C., Indelicato, S., Voloch, M., and Bordens, R. 2005. 'Site of Pegylation and Polyethylene Glycol Molecule Size Attenuate Interferon-α Antiviral and Antiproliferative Activities through the JAK/STAT Signaling Pathway', *Journal of Biological Chemistry*, 280 (8): 6327-6336.
24. Guo, H., L. Li, and P. G. Wang. 2006. 'Biochemical characterization of UDP-GlcNAc/Glc 4-epimerase from Escherichia coli O86:B7', *Biochemistry*, 45: 13760-8.

25. Horynova, M., K. Takahashi, S. Hall, M. B. Renfrow, J. Novak, and M. Raska. 2012. 'Production of N-acetylgalactosaminyl-transferase 2 (GalNAc-T2) fused with secretory signal Igkappa in insect cells', *Protein Expr Purif*, 81: 175-80.
26. Jang, H. J., C. Y. Shin, and K. B. Kim. 2015. 'Safety Evaluation of Polyethylene Glycol (PEG) Compounds for Cosmetic Use', *Toxicol Res*, 31: 105-36.
27. Koehler, K., Malik, M., Mahmood, S., Gießelmann, S., Beetz, C., Hennings, J. C., Huebner, A. K., Grahn A., Reunert, J., Nurnberg, G., Thiele, H., Altmuller, J., Nurnberg, P., Mumtaz, R., Babovic-Vuksanovic, D., Basel-Vanagaite, L., Borck, G., Bramswig, J., Muhlenberg, R., Sarda, P., Sikiric, A., Anyane-Yeboa, K., Zeharia, A., Ahmad, A., Coubes, C., Wada, Y., Thorsten, M., Vanderschaeghe, D., Schaftingen, E. V., Kurth, I., Huebner, A., Hubner, C. A. 2013. 'Mutations in GMPPA Cause a Glycosylation Disorder Characterized by Intellectual Disability and Autonomic Dysfunction', *Am J Hum Genet*, 93(4): 727-734.
28. Kolarich, Daniel, Pia H. Jensen, Friedrich Altmann, and Nicolle H. Packer. 2012. 'Determination of site-specific glycan heterogeneity on glycoproteins', *Nature Protocols*, 7: 1285.
29. Kotenko, S. V., G. Gallagher, V. V. Baurin, A. Lewis-Antes, M. Shen, N. K. Shah, J. A. Langer, F. Sheikh, H. Dickensheets, and R. P. Donnelly. 2003. 'IFN-lambdas mediate antiviral protection through a distinct class II cytokine receptor complex', *Nat Immunol*, 4: 69-77.
30. Lagassé, H. A. D., Alexaki, A., Simhadri, V. L., Katagiri, N. H., Jankowski, W., Sauna, Z. E., and Kimchi-Sarfaty, C. 2017. Recent advances in (therapeutic protein) drug development. *F1000Res*, 6: 113.

31. Lauber, Jennifer, René Handrick, Sebastian Leptihn, Peter Dürre, and Sabine Gaisser. 2015. 'Expression of the functional recombinant human glycosyltransferase GalNAcT2 in Escherichia coli', *Microbial Cell Factories*, 14: 3.
32. Lindhout, Theresa, Umar Iqbal, Lisa M. Willis, Anne N. Reid, Jianjun Li, Xin Liu, Maria Moreno, and Warren W. Wakarchuk. 2011. 'Site-specific enzymatic polysialylation of therapeutic proteins using bacterial enzymes', *Proceedings of the National Academy of Sciences of the United States of America*, 108: 7397-402.
33. Lira-Navarrete, E., M. de Las Rivas, I. Companon, M. C. Pallares, Y. Kong, J. Iglesias-Fernandez, G. J. Bernardes, J. M. Peregrina, C. Rovira, P. Bernado, P. Bruscolini, H. Clausen, A. Lostao, F. Corzana, and R. Hurtado-Guerrero. 2015. 'Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates protein O-glycosylation', *Nat Commun*, 6: 6937.
34. 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', 11: 56.
35. LoBuglio, A. F., Wheeler, R. H., Trang, J., Haynes, A., Rogers, K., Harvey, E. B., Sun, L., Ghrayeb, J., and Khazaeli M. B. 1989. 'Mouse/human chimeric monoclonal antibody in man: kinetics and immune response', *Proc Natl Acad Sci U S A*, 86(11): 4220-4224.
36. Loureiro, L. R., M. A. Carrascal, A. Barbas, J. S. Ramalho, C. Novo, P. Delannoy, and P. A. Videira. 2015. 'Challenges in Antibody Development against Tn and Sialyl-Tn Antigens', *Biomolecules*, 5: 1783-809.
37. Malakar, P., and K. V. Venkatesh. 2012. 'Effect of substrate and IPTG concentrations

on the burden to growth of Escherichia coli on glycerol due to the expression of Lac proteins', *Appl Microbiol Biotechnol*, 93: 2543-9.

38. Miknis, Zachary, Eugenia Magracheva, Wei Li, Alexander Zdanov, Sergei V. Kotenko, and Alexander Wlodawer. 2010. 'Crystal structure of the complex of human interferon-λ1 with its high affinity receptor interferon-λR1', *Journal of molecular biology*, 404: 650-64.
39. Nehrke, K., F. K. Hagen, and L. A. Tabak. 1996. 'Charge distribution of flanking amino acids influences O-glycan acquisition in vivo', *J Biol Chem*, 271: 7061-5.
40. Ngantung, F. A., P. G. Miller, F. R. Brushett, G. L. Tang, and D. I. Wang. 2006. 'RNA interference of sialidase improves glycoprotein sialic acid content consistency', *Biotechnol Bioeng*, 95: 106-19.
41. Nguyen, Van Dat, Feras Hatahet, Kirsi E. H. Salo, Eveliina Enlund, Chi Zhang, and Lloyd W. Ruddock. 2011. 'Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli', *Microbial Cell Factories*, 10: 1.
42. O'Brien, P. J., Irwin, W., Diaz, D., Howard-Cofield, E., Krejsa, C. M., Slaughter, C. M., Gao, B., Kaludercic, N., Angeline, A., Bernardi, P., Brain, P., and Hougham, C. 2006. 'High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening', *Arch Toxicol*, 80 (9): 580-604.
43. O'Connell, B. C., F. K. Hagen, and L. A. Tabak. 1992. 'The influence of flanking sequence on the O-glycosylation of threonine in vitro', *J Biol Chem*, 267: 25010-8.
44. O'Connell, B. C., and L. A. Tabak. 1993. 'A comparison of serine and threonine O-

glycosylation by UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase', *J Dent Res,* 72: 1554-8.

45. Pagliaccetti, Nicole E., Roger Eduardo, Steven H. Kleinstein, Xinmeng Jasmine Mu, Prasanthi Bandi, and Michael D. Robek. 2008. 'Interleukin-29 Functions Cooperatively with Interferon to Induce Antiviral Gene Expression and Inhibit Hepatitis C Virus Replication', *The Journal of Biological Chemistry*, 283: 30079-89.
46. Raju, T. S., J. B. Briggs, S. M. Borge, and A. J. Jones. 2000. 'Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics', *Glycobiology,* 10: 477-86.
47. Roggenbuck, D., Mytilinaiou, M.G., Lapin, S.V., Reinhold, D., and Conrad, K. 2012. 'Asialoglycoprotein receptor (ASGPR): a peculiar target of liver-specific autoimmunity', *Auto-Immunity Highlights,* 3 (3): 119-125.
48. San-Miguel, Teresa, Pedro Pérez-Bermúdez, and Isabel Gavidia. 2013. 'Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature', *SpringerPlus*, 2: 89.
49. Saxena, A., Raveh, L., Ashani, Y., and Doctor B. P. 1997. 'Structure of glycan moieties responsible for the extended circulatory life time of fetal bovine serum acetylcholinesterase and equine serum butyrylcholinesterase', *Biochemistry*, 36(24): 7481-9.
50. Sola, R. J., and K. Griebenow. 2009. 'Effects of glycosylation on the stability of protein pharmaceuticals', *J Pharm Sci*, 98: 1223-45.
51. Sola, R. J., and K. Griebenow. 2010. 'Glycosylation of therapeutic proteins: an

effective strategy to optimize efficacy', *BioDrugs*, 24: 9-21.

52. Stewart, E. J., F. Aslund, and J. Beckwith. 1998. 'Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins', *Embo j*, 17: 5543-50.
53. Sun, Qiao-Yang, Ling-Wen Ding, George P Lomonossoff, Yong-Bing Sun, Ming Luo, Chaoqiong Li, Liwen Jiang, and Zeng-Fu Xu. 2011. Improved expression and purification of recombinant human serum albumin from transgenic tobacco suspension culture, *J Biotechnol*, 155(2): 164-72.
54. Suzuki, T. Ishii-Watabe, A., Tada, M. Kobayashi, T. Kanayasu-Toyoda, T. Kawanishi, T., and Yamaguchi, T. 2010. 'Importance of neonatal FcR in regulating the serum half-life of therapeutic proteins containing the Fc domain of human IgG1: a comparative study of the affinity of monoclonal antibodies and Fc-fusion proteins to human neonatal FcR', *J Immunol*, 184 (4): 1968-76.
55. Taus, Christopher, Markus Windwarder, Friedrich Altmann, Reingard Grabherr, and Erika Staudacher. 2014. 'UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyl-transferase from the snail Biomphalaria glabrata – substrate specificity and preference of glycosylation sites', *Glycoconjugate Journal*, 31: 661-70.
56. Uze, G., and D. Monneron. 2007. 'IL-28 and IL-29: newcomers to the interferon family', *Biochimie*, 89: 729-34.
57. Vegarud, Gerd, and Terje B. Christensen. 1975. 'Glycosylation of proteins: A new method of enzyme stabilization', *Biotechnology and Bioengineering*, 17: 1391-97.
58. Verhoef, J. J., J. F. Carpenter, T. J. Anchordoquy, and H. Schellekens. 2014.

'Potential induction of anti-PEG antibodies and complement activation toward PEGylated therapeutics', *Drug Discov Today,* 19: 1945-52.

59. Vugmeyster, Y., Xu, X., Theil, F. P., Khawli, L. A., and Leach, M. W. 2012. 'Pharmacokinetics and toxicology of therapeutic proteins: Advances and challenges', *World J Biol Chem*, 3(4): 73-92.
60. Wandall, H. H., H. Hassan, E. Mirgorodskaya, A. K. Kristensen, P. Roepstorff, E. P. Bennett, P. A. Nielsen, M. A. Hollingsworth, J. Burchell, J. Taylor-Papadimitriou, and H. Clausen. 1997. 'Substrate specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3', *J Biol Chem*, 272: 23503-14.
61. Wandall, H. H., F. Irazoqui, M. A. Tarp, E. P. Bennett, U. Mandel, H. Takeuchi, K. Kato, T. Irimura, G. Suryanarayanan, M. A. Hollingsworth, and H. Clausen. 2007. 'The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc: lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation', *Glycobiology*, 17: 374-87.
62. Wang, W., and Prueksaritanont, T. 2010. 'Prediction of human clearance of therapeutic proteins: simple allometric scaling method revisited', *Biopharm Drug Dispos*, 31(4): 253-63.
63. Witte, K., E. Witte, R. Sabat, and K. Wolk. 2010. 'IL-28A, IL-28B, and IL-29: promising cytokines with type I interferon-like properties', *Cytokine Growth Factor Rev*, 21: 237-51.
64. Wu, J., Zhao, C., Lin, W., Hu, R., Wang, Q., Chen, H., Li, L., Chen, S., and Zheng, J. 2014. 'Binding characteristics between polyethyleneglycol (PEG) and proteins in

aqueous solution', *Journals of Materials Chemistry B*, 2: 2983-2992.

65. Yang, G. G., X. Y. Xu, Y. Ding, Q. Q. Cui, Z. Wang, Q. Y. Zhang, S. H. Shi, Z. Y. Lv, X. Y. Wang, J. H. Zhang, R. G. Zhang, and C. S. Xu. 2015. 'Linker length affects expression and bioactivity of the onconase fusion protein in Pichia pastoris', *Genet Mol Res*, 14: 19360-70.
66. Yoshimura, Y., A. S. Nudelman, S. B. Levery, H. H. Wandall, E. P. Bennett, O. Hindsgaul, H. Clausen, and S. Nishimura. 2012. 'Elucidation of the sugar recognition ability of the lectin domain of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase 3 by using unnatural glycopeptide substrates', *Glycobiology*, 22: 429-38.
67. Zegzouti H., Engel L.,Vidugiris G. and Goueli S. 2013. Detection of glycosyltransferase activities with homogenous bioluminescent UDP detection assay. *Glycobiology*, 23: 1340–1341.
68. Zhang, X. D., D. Wu, X. Shen, P. X. Liu, N. Yang, B. Zhao, H. Zhang, Y. M. Sun, L. A. Zhang, and F. Y. Fan. 2011. 'Size-dependent in vivo toxicity of PEG-coated gold nanoparticles', *Int J Nanomedicine*, 6: 2071-81.
69. Zhuo, X. F., Y. Y. Zhang, Y. X. Guan, and S. J. Yao. 2014. 'Co-expression of disulfide oxidoreductases DsbA/DsbC markedly enhanced soluble and functional expression of reteplase in Escherichia coli', *J Biotechnol*, 192 Pt A: 197-203.