An in vitro one pot synthetic biology approach to emulating Golgi O-glycosylation divergence in MUC1 and tumor associated MUC1

Abdullateef Nashed[1,2] and Kevin J. Naidoo[1,2*].

[1]Scientific Computing Research Unit Address, University of Cape Town, Rondebosch 7701, [2]Department of Chemistry, University of Cape Town, Rondebosch 7701.

*To whom correspondence should be addressed

KEYWORDS: In vitro synthetic biology, Cancer, MUC1 antigens, systems chemical glycobiology.

ABSTRACT: Peptide O-glycosylation is a non-template driven process dependent on an orchestrated collaboration of glycosyltransferases (GTs) in the ER and golgi apparatus. An *in vitro* modelling of this requires an understanding of GT specificities, kinetics and their spatial distribution along the ER-Golgi axis. This study explores the specificity and kinetics of O-GalNAc glycosylation on a designed tandem repeat of MUC1 (23mer), a protein significantly implicated in cancer progression. Using an *in vitro* one-pot synthetic biology approach, MUC1 glycopeptides were produced, and GT specificity for cancer-associated antigens Tn, T, and their sialylated forms were assessed. The impact of GT re-localization from the ER to Golgi on glycosylation patterns was modelled and analyzed. The model suggests that presenting MUC1 to GALNTs in isolation from other GTs, namely C1GALT1 and ST6GALNAC1 significantly increases the extent of site occupancy by Tn a tumour epitope. This mimics the ER localization of GALNTs associated with cancer. In contrast when MUC1 is presented to GALNTs in combination with C1GALT1 there is a decline in observed occupancy. Our modelling shows this to be the result of a competition between the secondary glycosylation of the GalNac sites by C1GALT1 and the GALNTs' lectins that need to bind to the GalNac sites to further proceed with the primary glycosylation of adjacent free sites. In the case of normal MUC1 C1GALT1 shows greater specificity to the under-occupied (GalNAc3-23mer) form of Tn. In the cancer case ST6GALNAC1 shows greater specificity to the fully occupied form (GalNAc5-23mer). These results suggest a mechanism of regulation of tumour associated MUC1 antigens independent of GTs expression level.

1. INTRODUCTION

A prediction of the concentration of glyconjugates or the extent to which and how they are glycosylated is not possible with current experimental and computational tools. Simply, the post translational event of peptide/protein glycosylation is a non-template driven process that relies on more than glycoenzyme gene expression data or even glycoenzyme expression levels itself.[1] A case in point is while glycosyltransferase genes can be used to classify cancer[2] this knowledge cannot inform the difference between cancer and healthy glycoconjugate expressions. The characteristic high degree of sialylation in tumour tissues[3] and the structural changes in glycans cannot be directly corelated by the genes expressing for those sialyltransferase. This is in part because the complex glycosylation pathways of a cell are intimately connected and intertwined with other critical metabolic and regulatory networks of that cell.[4] Consequently, developing a systems biology model of glycoconjugate metabolic networks requires multiple components. However, a critical first step is to construct a developmental model that will mimic the biosynthesis processes, within the ER and Golgi apparatus. Taking MUC1 as an example, a systems model must explain the preferential construction of a normal glycosylated state over a tumour-associated (TA) MUC1 glycosylated state while assigning key drivers to the MUC1 and TA-MUC1 glycopeptide outcomes.

Chemoenzymatic synthesis is the standard method for producing glycopeptides where the core peptide is chemically synthesized, followed by the enzymatic synthesis of the glycan one sugar moiety at a time. Advances in understanding enzyme specificities and mechanisms have enabled the synthesis of more complex glycopeptides,[5] expanding the dimensions of glycopeptide arrays. For example, Yoshimura et al.[6] produced an array of 20 MUC1 glycopeptides covering the Tn and T antigens and their sialylated forms, STn and ST, respectively. These glycans were synthesized at each of the five possible glycosylation sites of the MUC1 tandem repeat. The initial glycosylation was done by a chemical addition of the GalNAc residue at the selected site. This was followed with enzymatic glycosylation to complete the glycan structures.[6] Good yields for all the single site glycosylation's were achieved in this way. Alternatively, a synthetic biology approach using genetically engineered human embryonic kidney (HEK) cells was developed.[7] This was achieved by rationally modifying the endogenous glycosylation pathways through knock-in or knock-out of specific GTs. What results are cells that can synthesize specific glycan structures and can either be displayed on cell surface or on a probe protein designed for secretion.[7] The advantage of the synthetic biology approach over *in vitro* is that a range of glycoconjugate structures that are biologically possible can be produced from a disease-specific genotype, so inferring the engineered glycosylation pathway. Further, the glycoconjugate binding and interaction can be studied the in context of their cell-displayed (and cellularly generated) form. Shortcomings of this approach are that i) the produced glycoconjugates rely on the endogenous machinery of the cells and cannot be completely controlled and optimized to avoid side reactions and incomplete glycosylation; ii) the pathways modifications are limited to the minimal cell survival needs of the selected glycosylation pathways. These drawbacks result in a heterogeneity of expressed structures that do not make a designed singular glycosylated structure, needed for functional studies, possible. Since the intention is to map and mimic the *in vivo* GTs kinetics, selectivity and their mechanistic action, the *in vitro* methods are preferable as they are able to produce single purifiable and spectroscopically verified structures. Moreover, models must be capable of mimicking the alterations of *in vivo* glycosylation that results from a spatial-temporal rearrangement of the distribution and presentation of GTs to the substrate in the ER-golgi.[8, 9] The localization of GTs, such as GALNTs has been found to be a regulatory mechanism involved in cancer phenotypes via altering O-GalNAc glycan structures and level.[10, 11]

Mucin 1 (MUC1), a highly glycosylated transmembrane protein overexpressed in epithelial cancers, contributes to tumorigenesis, immune evasion and suppression, metastasis, and ultimately poor prognosis. Its extracellular O-glycosylation weakens drug sensitivity by acting as a barrier and modulates signaling pathways, leading to decreased drug permeation and increased cancer cell survival. MUC1 O-glycosylation has been proposed as a target to enhance drug sensitivity and efficacy.[12] Here we show a MUC1 glycosylation model that accounts for the spatial and temporal regulation of glycosyltransferases (GTs). Specifically, an account of i) the kinetic parameters governing GT activities and substrate specificities, ii) GT concentrations and their distribution along the ER-Golgi axis must form part of the modelling. An accurate and consistent measure of kinetics across the entirety of all GTs participating in the biosynthesis is an essential tool. Previously, we reported the UGC assay, as such a universal tool.[13] We show here that this glycosylation model is achieved through an *in vitro* one-pot synthetic biology approach using a peptide fusion protein expression system. Contrived ER-golgi conditions are created to discover the O-GalNAc glycosylation of the MUC1 peptide that produces the cancer associated antigens Tn, T and their respective sialylated forms sTn and sT. The model was used to assess specificity of different GTs, involved in the synthesis of these antigens, to the five unique MUC1 tandem repeat glycosylation sites.

2. RESULTS AND DISCUSSION

2.1 Modelling MUC1 sequential Golgi glycosylation

Glycosylation of proteins takes place in the ER-Golgi system via the action of different glycoenzymes, namely, glycosyltransferase and glycosidase enzymes. These enzymes have been found distributed across specific cisternae.[8, 9, 14] The distribution of GTs is determined by the length of their transmembrane domains in correlation to thickness and lipid content of the cisternae membrane. The localization of GTs along the ER-Golgi axis is dynamic, and these enzymes are constantly trafficked in both directions via a complex but tightly regulated vesicle system involves COP-I and COP-II vesicles[14, 15]. The localization of the enzymes participating in the glycosylation in different cisternae results in assembly of the enzymes in compartments acting on the target protein sequentially (Figure 1, A).

2.1.1 The Sequential One-pot synthesis method:

To capture the spatial-temporal segregation of glycosylation pathways, a sequential one-pot enzymatic synthesis is needed (Figure 1, B). In this synthesis, the products of each assembly of GTs are analysed in terms of structures and reactivity with the subsequent enzymes in the pathway to aid the design of the following assembly (pot) and extract kinetics data assisting model construction.

To facilitate this, the peptide of interest was expressed in a fusion protein form core peptide of interest (MUC1 in this work) and a carrier protein, to be expressed in E. coli. Superfolder green fluorescence protein sfGFP was selected for its folding efficiency, minimized dimerization, and enhanced solubility[16]. To minimize the possibility of interaction with the peptide, sfGFP was separated from the peptide with linker (Linker 2) that contains three rigid units, and three flexible units (from N to C direction). Another rigid linker (Linker 1) was also used for better steric presentation the N-terminus His-tag improving affinity-based purification. A tobacco etch virus (TEV) protease recognition sequence is also added to cleave the peptide in its native sequence at any point of the synthesis (Figure 1, C and S1, B). These features aim to enable in vitro enzymatic glycosylation and facilitate simple one-step purification (Figure 1, E).

2.1.2 Illustrating MUC1 vs TA-MUC1 case study:

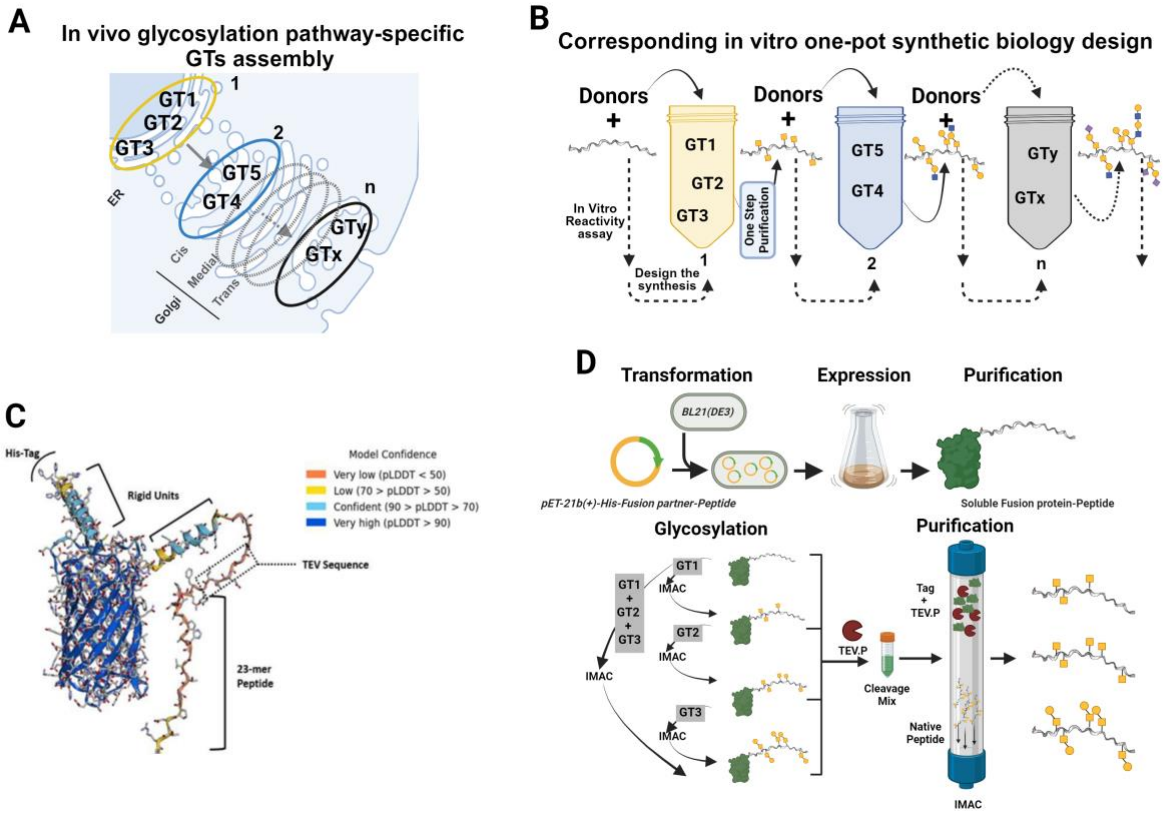

Figure 1. Design of the in vitro synthesis of glycopeptides. A. Distribution of glycosyltransferases across the ER-Golgi axis. B. Design of the in vitro synthetic biology mimicking the compartmentation in the ER-Golgi axis. C. The structure of the fusion protein for the expression of the desired peptide as predicted by AlphaFold. D. The pipeline of expression and enzymatic glycosylation of MUC1 peptide.

Peptide Sequence

Polypeptide GalNAc-transferases (GALNTs) catalyze the addition of GalNAc to serine or threonine residues, initiating the biosynthesis of GalNAc O-linked glycans and forming the GalNAc residue α-linked to Ser/Thr, Thomsen-nouvelle (Tn) antigen. This family of enzymes, comprising 20 human isoforms divided into two families based on substrate preference. Family I GALNTs mainly target unglycosylated peptides, while Family II GALNTs target O-GalNAc glycosylated peptides. Some GALNTs such as GALNT1 and GALNT2 are widely expressed across many tissues, whereas some are tissue specific.[17] Each GALNT has a catalytic domain and a lectin-like binding domain, with the catalytic domain containing sites crucial for substrate specificity and the lectin domain binds pre-glycosylated sites. The structure of the two domains and the linker between them determine the site specificity and directionality of the lectin-assisted glycosylation.[18] The

specificity of the isoforms GALNT1, 2 and 4 was studied intensively[19-22] either using MUC1 peptide contains multiple tandem repeats or a single repeat with extension to the N terminal to the first T residue TAP-24 peptide (Figure 2, B). These designs were helpful to reveal the specific direct and lectin assisted glycosylation for these enzymes (Figure 2, A). However, for the purpose of quantifying enzyme specificities and synthesizing biologically relevant glycopeptides that capture the structural aspects in their native biological context, the optimum sequence must account for all the determinant variables of GALNTs specificities, and each variable is only represented once in the sequence. The following criteria was set for the optimum peptide sequence: (1) includes all the five unique potential glycosylation sites of MUC1 tandem repeat, and each site is only represented once in the sequence, (2) none of the sites is terminal and a sufficient extension of the tandem repeat sequence in both direction of the site is needed to account for the motif specificity of the catalytic domain, (3) the frame of the sequence must be optimized for the position of each site in relation to the rest of the sites to account for the direction specificity of the lectin domain in GALNTs. A design of 27-mer peptide, which was obtained by removing the N terminal T from TAP24 and expanding the C terminal sequence until before the next T8 (Figure S1, A). the 27-mer peptide satisfies the first and second criteria. However, it lacks S19 and/or T20 left to S9 and T13, which are required for GALNT4 glycosylation of S9 and T13.[22] However, the 27mer was included as a negative control for GALNT4. An optimum sequence was obtained by shifting the frame of the sequence to start with S19 and T20 from the N terminal end of the peptide (Figure 2, B). Despite being an unusual frame compared to the designs commonly used in literature, this sequence is expected to address the shortcomings of the 27-mer

The role of distribution and levels of GTs in MUC1 to TA-MUC1 transformation

The GalNAc glycosylated sites are subject to either sialylation via ST6GALNAC1, galactosylation (addition of Gal) via C1GALT1 and its chaperone C1GALT1C1 (Cosmc), or N-acetylglucosaminylation (addition of GlcNAc) via B3GNT6 (Figure 2, C). GALNTs and C1GALT1 are localized in cis Golgi whereas ST6GALNAC1 is distributed across all the cisternae of Golgi.[8, 9, 14] No specific localization of B3GNT6

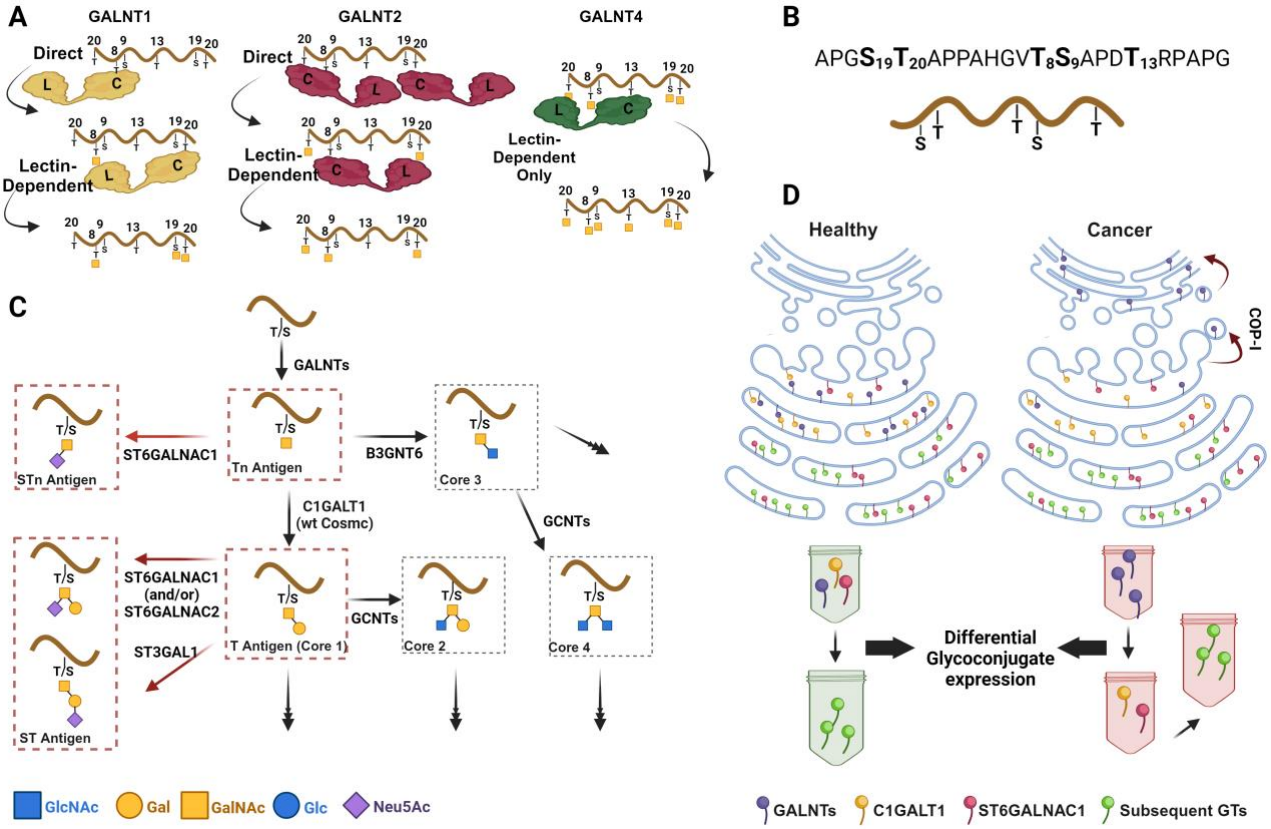

Figure 2. MUC1 case study. A. initiation of O-glycosylation via GALNTs. Glycosylation by GALNTs can be direct via the catalytic domain (C) recognition of the motif surrounding the glycosylation site. Lectin domain (L) assists subsequent sites glycosylation via binding to pre-glycosylated sites. Proffered Directionality of the subsequent glycosylation is also shown. GALNT4 lacks the direct glycosylation mechanism. B. Sequence of the 23-mer peptide showing the potential glycosylation sites in bold, the order of the sites is based on the VNTR repeat unit (Figure S1, A). C. O-GalNAc glycosylation pathway showing the main structures (cores). Terminal reactions are indicated with red arrows. Cancer associated structures are labeled with red frames. D. Differential localization of GALNTs between normal and cancer epithelial attributed to COP-1 mediated retrograde activation in cancer and the corresponding synthesis design.

has been reported. The localization of GALNTs has been found altered in response to EGF stimulation of SRC, the proto-oncogene in cancer[23] via COP-1 mediated retrograde from cis Golgi to the ER. Other GT enzymes were not affected[10]. This relocation results in overexpression of Tn in the ER and that a fraction of this Tn transits to the cell surface without modification, while another fraction transits with modification to T antigen. In patient samples, the study found that the mean expression of Tn in breast cancer tissue samples was 4.5

folds higher than in normal tissues. From the samples with high Tn expression, 70 % of the samples showed ER localization of GALNTs (inferred indirectly from ER localization of Tn), whereas no significant loss of C1GALT1 was detected in these samples.[10, 11]

In addition to relocation of GALNTs, Expression of O-GalNAc glycosylation enzymes has also been found altered in cancer when compared to normal cells. ST6GALNAC1 was found upregulated in almost all cancer types[24] and C1GALT1 downregulated mainly due to Cosmc mutation or epigenetic alteration of both C1GALT1 and Cosmc[25, 26]. The model in this work will focus on the impact of redistribution of GALNTs in altering glycan structures independently of enzymes levels and will be used to illustrate the utility of the one-pot in vitro synthesis (Figure2, D). Therefore, all activities and kinetics experiments is performed on a standardized concentration of the enzymes at 250 nM.

2.2 Initiation of MUC1 glycosylation (GalNAc glycosylation)

Expression of the 23mer and the 27mer was carried out in *E. coli* BL21 (DE3), a single-step purification using IMAC yielded around 60 mg of pure fusion protein from 250 ml cell culture. The tag cleavage using TEV protease was also performed and SDS-PAGE gel confirmed complete cleavage for the naked peptide and the peptide with different glycosylation sites (Figure S2).

2.2.1 GALNTs specificity

The performance of the 23mer peptide was compared to the 27mer by measuring the relative reactivity of GALNT1, GALNT2, and GALNT4, individually and in various combinations, using the UGC assay (Figure 3A for the 23mer and Figure S3 for the 27mer). The primary observation from both experiments is that GALNT4 alone does not react with either peptide confirming the strictly lectin-dependent mechanism. On the other hand, GALNT1 and GALNT2 could individually react with both peptides due to their direct mechanism. The rates of glycosylation via GALNT1 or GALNT2 were overall lower with the 23mer and decreased with time (Figure 3, A (a)), whereas the rates were higher and constant with the 27mer (SI). This can be explained by the depletion of the direct-glycosylation specific sites T13 and T20 for GALNT1 and GALNT2, respectively in the 23mer case, and the absence of secondary lectin-assisted sites in the preferable direction. On the other hand, the 27mer permits the continuation of glycosylation after the first site to the following lectin-assisted sites (Figure 3, A and SI). This can also be confirmed by the synergistic effect (rate for combination is larger than the sum of rates of individual reactions) when GALNT1 and GALNT2 were combined; in the case of 23-mer, the synergistic effect was more significant than the 27-mer indicating that glycosylation of the initial sites in the 23-mer via direct mechanisms of one enzyme provides the required conditions for the lectin-dependent mechanism for the other enzyme. In the 27mer, GALNT4 showed no activity even when combined with GALNT1 or GALNT2. In contrast, GALNT4 exhibited significant activity with the 23mer peptide when combined with GALNT2 indicated by the enhancement of the glycosylation rate when compared with the reaction of GALNT2 individually. GALNT2 is known to have a preferred specificity to T20 via direct catalytic domain recognition. In the 23mer, T20 is at the N-terminal to T13 and S9 (GALNT4 specific sites), and Pre-glycosylation at T20 left to T13 and S9 sites is the prerequisite for GALNT4 lectin-dependent specificity.[22] To rule out the possibility of the participation of the tag in the reaction, the reactivity of a mixture of the three GALNTs with cleavage products (tag and naked peptide) was tested and the results show that the enzymes do not glycosylate the tag (data not shown). The evidence presented thus far confirms the efficiency of the fusion protein design in terms of expression level, solubility, cleavage, and the inertness of the tag to the GALNTs tested in this work. Additionally, 23mer has proved valid to capture the reaction mechanism of GALNT4 and the direct mechanism for both GALNT1 and GALNT2. However, the reported lectin-dependent mechanisms for GALNT1 and GALNT2 can be better studied using the 27mer.

2.2.2 In vitro synthesis of Tn

Next, the in vitro synthesis was carried out to explore the possible structures enabled by extensive glycosylation via different combinations of GALNTs (Figure 3, B(a)). LC-MS results confirmed the structure and purity of the peptide (i). The extensive glycosylation via GALNT1 and GALNT2 results in glycosylation of 2 or 3 sites in case of GALNT1 and 3 sites in case of GALNT2. While glycosylation of T8 and T20 can be explained by the direct mechanism of GALNT1 and GALNT2, respectively, the observed glycosylation of the other two sites cannot be attributed to the lectin-dependent mechanisms previously reported for GALNT1 and GALNT2 activity on MUC1[20, 22] and observed in the short-term recorded activity (Figure 3, A (b)) because the 23mer does not provide the directionality required for this mechanism. However, a randomized peptide sequences study found that GALNT1 can catalyze a long-range N terminal lectin-dependent mechanism while GALNT 2 can participate in both N and C long-range lectin-dependent mechanism,[27] which may explain the results found here. No further glycosylation beyond three sites was observed when GALNT1 and GALNT2 were combined (Figure 3, B, Peak v). When the product of GALNT1 and GALNT2 combination was glycosylated with GALNT4, all five sites were glycosylated. This results indicate that two sites are GALNT4 specific and are glycosylated via a lectin-dependent mechanism. The consensus from all previously reported results is that T8 and T2 are specific sites for GALNT1 and GALNT2, respectively and S9 and T13 are GALNT4 specific. Taking all this into account, a conclusion can be made that GALNT1, GALNT2, or in combination can glycosylate S19 T20 T8 utilizing direct and lectin-dependent mechanisms. However, GALNT1 is less efficient than GALNT2 in completing the glycosylation on either S19 or T20 (Figure 3, B (c)).

It is widely accepted that the GalNAc site occupancy increases with over-expression and ER relocation of GALNTs.[10, 11] The effect of these two factors was confirmed in the extensive glycosylation performed here as a simulation of expanding the action of GALNTs in isolation of other GTs. However, some sites, such as T20, S19, and T8 were more redundant and can be targeted by multiple GALNTs, such as GALNT1 and GALNT2 in this work, and GALNT3 reported previously[20, 22]. These sites are the first sites to be glycosylated. Other sites are strictly lectin-dependent (S9 and T13) and are they are dependent on prior glycosylation of the other sites, which classifies them as late sites. The lectin-dependent mechanism is responsible for high-density GalNAc-O-glycosylation.[21] This differential site occupancy was also found to be associated with cancer transformation. The site saturation was observed in MUC1 expressed in cancer cells when compared with normal MUC1 in breast milk[28]. The GalNAc3-23mer and GalNAc5-23mer synthesized here will be used as models for site occupancy.

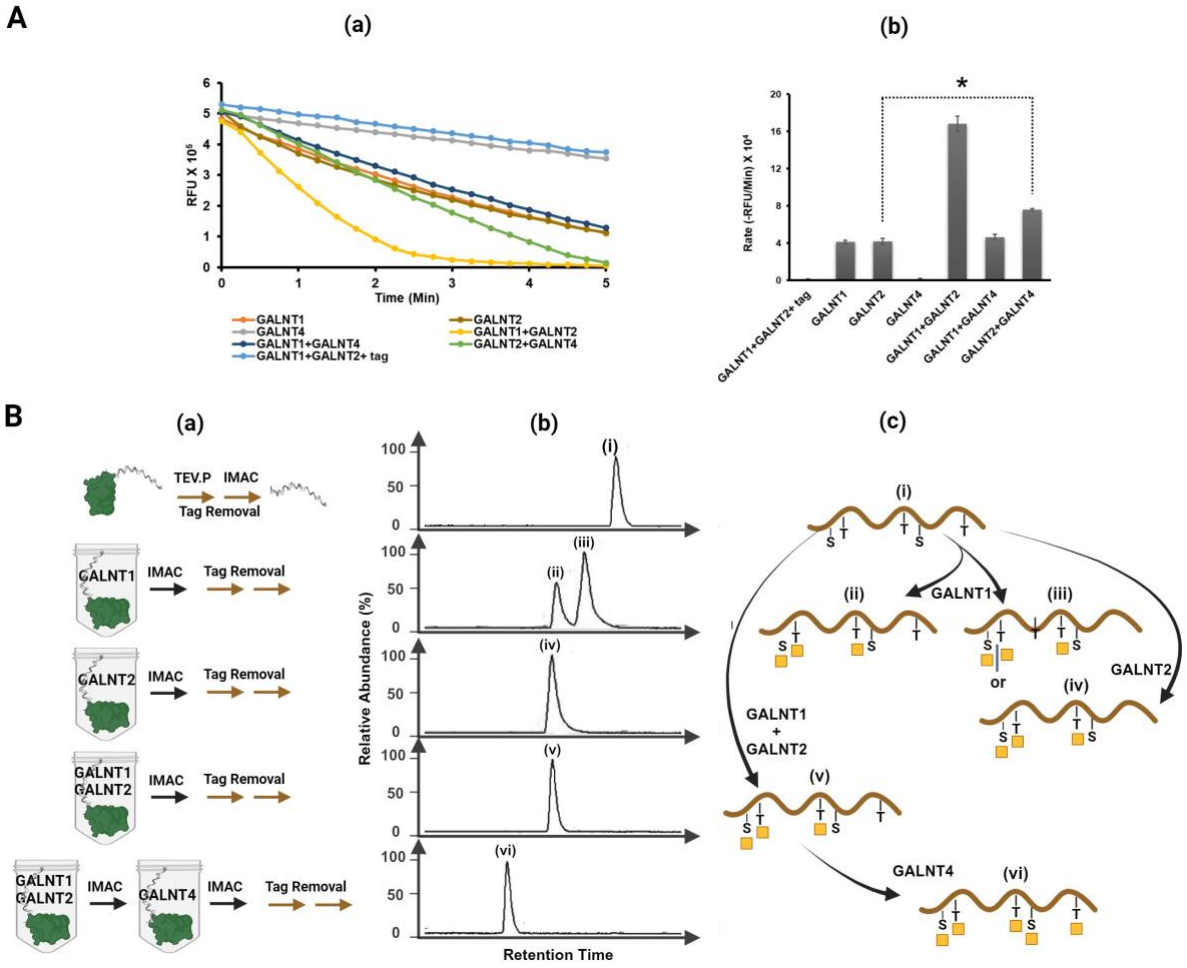

Figure 3. In vitro emulation of the initiation of O-GalNAc glycosylation on the 23mer via GALNTs. A. Reactivity of the 23mer peptide with GALNTs: (a). Progress of glycosylation reactions via different combinations of GALNTs recorded using the UGC assay. Control reaction with GALNT1 and GALNT2 with the tag is included (b). Rates of the reaction recorded in the linear range for each reaction, data presented as (Mean ± SD). pairs marked with asterisk indicates p value < 0.01. B. In vitro synthesis of GalNAc-glycosylated 23mer: (a). Brief description of synthesis designs and steps mirroring the reactivity combinations in panel A. (b) Abundance of the products of each reaction as measured by liquid chromatography (Supportive Information). (c) Structures of the products of each reaction as determined from the LC-MS data (Supportive Information) and the synthesis pathway leads to these structures as concluded from the preceding Data.

2. 3 Subsequent Glycosylation of Tn

The sequential addition of one or two sugars to Tn antigen generates diverse glycan core structures, with eight different cores identified. The predominantly structures cores 1-4 are shown in Figure 2, C), and cores 5-8 are rare.[29] Core 1, also known as T antigen, is formed by adding galactose to Tn antigen via a β1-3 bond, a process catalyzed by T synthase (C1GALT1) with the help of its chaperone C1GALT1C1 (Cosmc). Core 3 is synthesized by adding GlcNAc to Tn antigen through a β1-3 bond, catalyzed by B3GNT6. ST6GALNAc1 sialylates Tn antigen to sialyl-Tn (sTn) that terminates the synthesis preventing the structures to undergo further glycosylation via C1GALT1 or B3GNT3. Core 2 and Core 4 are synthesized by extending Core 1 and core 3 structures via GCNTs enzymes that can be extended to more complex structures.

2.3.1 Competition S6GALNAc1, C1GALT1 and B3GNT6 on the different states of site occupancy by GalNAc

The reactivity of the three enzymes were compared for GalNAc3-23mer or GalNAc5-23mer as representatives of sub-saturation and full saturation (full occupancy) of MUC1 potential glycosylation sites, respectively. Serial dilutions of both substrates were prepared by normalizing the concentrations to the number of GalNAc-glycosylated sites (Figure 4, A). While the overall activity of C1GALT1 was much higher than that of ST6GALNAC1, the results show that ST6GALNAC1 preferably glycosylates full occupancy sites (GALNT4 specific sites) whereas C1GALT1 prefers the sub saturation sites. The selectivity of both C1GALT1 and ST6GALNAC1 is more pronounced as MUC1 concentration increases. The kinetics parameters derived from the dose-response of ST6GALNAC1 to both acceptors indicate that the enzyme has a lower affinity but higher turnover to fully glycosylated peptide (lower K_m and higher V_{max}) when compared to the subsaturated peptide. This observation suggests that ST6GALNAC1 prefers full occupancy form only when MUC1 is over-expressed which is a phenotypic of cancer. On the other hand, no significant difference in K_m values of C1GALT1 was observed indicating that C1GALT1 prefers the direct sites at any concentration of MUC1. Glycosylation of both acceptors with B3GNTs was very slow and undetectable at lower concentrations of the substrates. However, the rates at the highest tested concentration (380 µM) did not show differential specificity.

Finally, the activity of ST6GALNAC2 was also tested with both acceptors, and no significant glycosylation was detected despite the confirmation of expression of ST6GALNAC2 in active form when tested with asialofetuin (Data not shown).

2.3.2 In vitro synthesis of Core 1 (T antigen)

The synthesis of Core 1 via C1GALT1 was carried out in different sequential one-pot designs simulating the different GALNTs distribution: when C1GALT1 competes with GALNTs i.e. the co-localization in cis Golgi, and when GALNTs act on the peptide first in isolation i.e. when they re-localize to the ER. When C1GALT1 was mixed with GALNT1 and GALNT2 (Figure 4, B (a)), the chromatogram showed a major peak of a $(Gal\text{-}GalNAc)_2$-23mer and a minor peak of $(Gal\text{-}GalNAc)_3$-23mer as identified from the mass spectrometer results. The

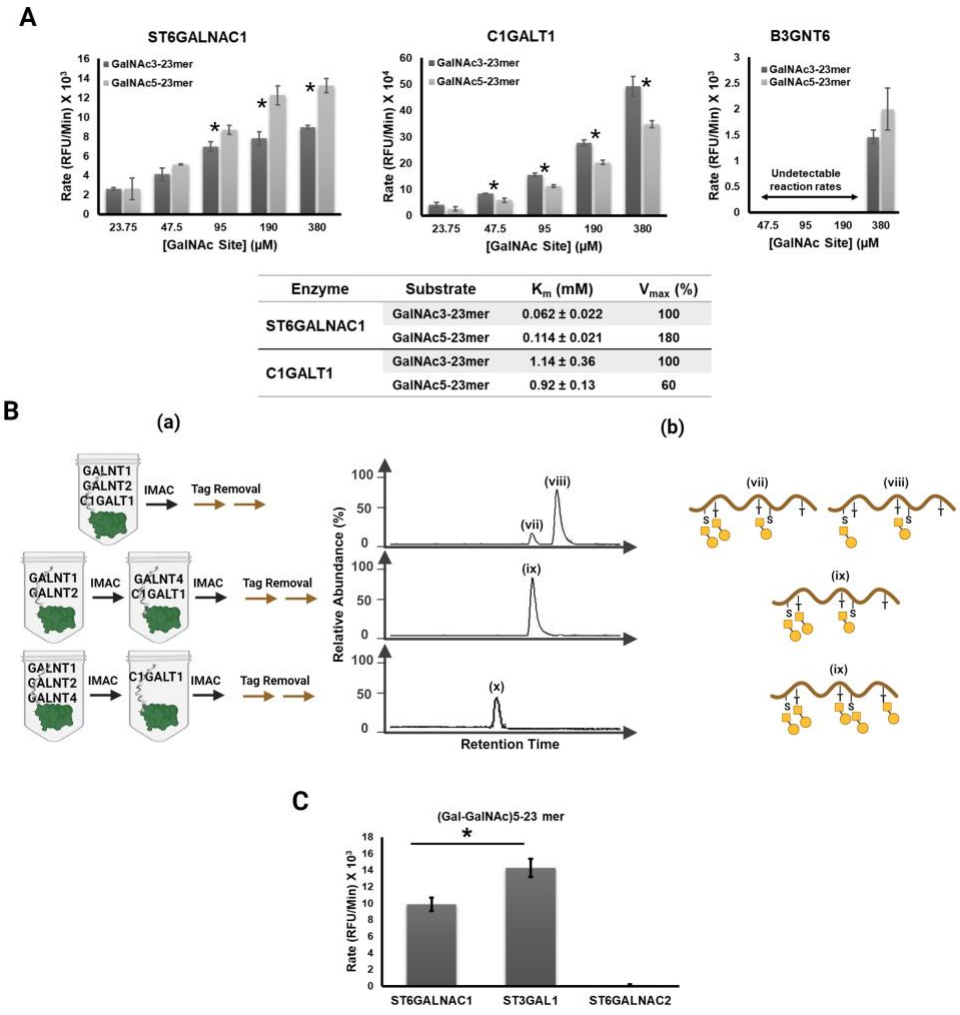

Figure 4. Assessment of first layer of structure extension after initiation (Figure 2, C). A. Effect of site occupancy by GalNAc on the specificity of ST6GALNAC1, C1GALT1 and B3GNT6. Serial dilution of Concentrations of GalNAc3-23mer and GalNAc5-23mer were calculated per GalNAc site occupancy. Data is presented as (mean ± SD). B. Effect of GALNTs C1GALT1 competition on site occupancy. (a) synthesis designs of different combinations of sequential synthesis reflecting C1GALT1 competition with GALNT1 and GALNT2, and with GALNT4, top and middle designs, respectively, and C1GALNT1 in isolation after complete glycosylation by the GALNTs (bottom design). (b) shows the LC peaks and structures deduced from analyzing mass spectrometry results (Supportive Information). C. Sialylation of Core 1 via ST3GAL1, ST6GALNAC1, and ST6GALNAC2. Bas in panels A and C are represented as means ± SD), pairs marked with asterisk indicates p value < 0.01.

results suggest that upon incorporation of GalNAc via the direct mechanism of GALNT1 and GALNT2 on T8 and T20, respectively, C1GALT1 competes with both GALNTs lectin domain on these sites. The result of this competition is the synthesis of Core1 structures and inhibition of the subsequent site GalNAc glycosylation via the lectin-dependent mechanism of GALNTs. This results also suggest that GalNAc glycosylation of S19 via GALNT1 and/or GALNT2 is catalysed via a lectin-dependant mechanism. Similar inhibition of the GALNT4 lectin-dependant mechanism can be concluded when C1GALT1 and GALNT4 reacted with the product of glycosylation by GALNT1 and GALNT2 as only three sites were identified with core 1 structures (Figure 4, B).

Finally, when simulating ER localization of GALNTs, C1GALT1 produced five sites occupied by core 1. The effect of co-localization of GALNTs with C1GALT1 on site occupancy can be extrapolated to their co-localization with ST6GALNAC1. This conclusion is supported by a study confirmed the effect of ST6GALNAC1 overexpression in reducing site occupancy by 25 % in Chinese Hamster Ovary (CHO) cells.[14]

2.3.3 Sialylation of Core 1

Core 1 can be sialylated via ST3GAL1 to form sialyl-3-T antigen or via ST6GALNAC1 or ST6GALNAC2 to form sialyl-6-T antigen. However, despite the reported activities of the three enzymes on T antigen, the specificities of these enzymes were not tested comprehensively using standardized substrates. Additionally, details of ST6GALNAC1 vs ST6GALNAC2 specificities for Tn and T are conflicting across studies and the models (in vitro vs in vivo).[30, 31] Both ST6GALNAC1 and GALNAC2 show no reactivity with the stand alone GalNAc or Gal-GalNAc acceptor, which confirms the peptide core requirement for both enzymatic activities.[30, 31] That was confirmed when both showed significant reactivity with asialofetuin (data not shown). We showed above that ST6GALNAC1 (but not ST6GALNAC2) reacts with Tn antigen. Same observation was true for T antigen (Figure 4, C). When compared to ST6GALNAC1, ST3GAL1 showed significantly higher reactivity with T antigen.

2.4 sTn expression: GALNTs relocation and Sialylation of Tn model

The in vitro reconstruction of the MUC1 GalNAc-O-glycosylation pathway delivered the following details of the pathway: the competitive specificity of each enzyme to each substrate, the possible structures produced by each enzyme, effect of the sequence of each enzymatic reaction and compartmentation (co-localization) on the glycosylation products profile. To illustrate the utility of the approach, the effect of GALNTs relocating to the ER on site occupancy was confirmed. This effect was found to be produced by two mechanisms: the isolation of the GALNT in the ER elongates the reaction time with the substrate allowing complete glycosylation and the saturation of the sites via the lectin-dependant mechanisms, the second mechanism depends on preventing the inhibition of the lectin-dependant mechanism of GALNTs by glycosylating the initial GalNAc moieties by the subsequent enzymes, such as C1GALT1 and ST6GALNAC1.

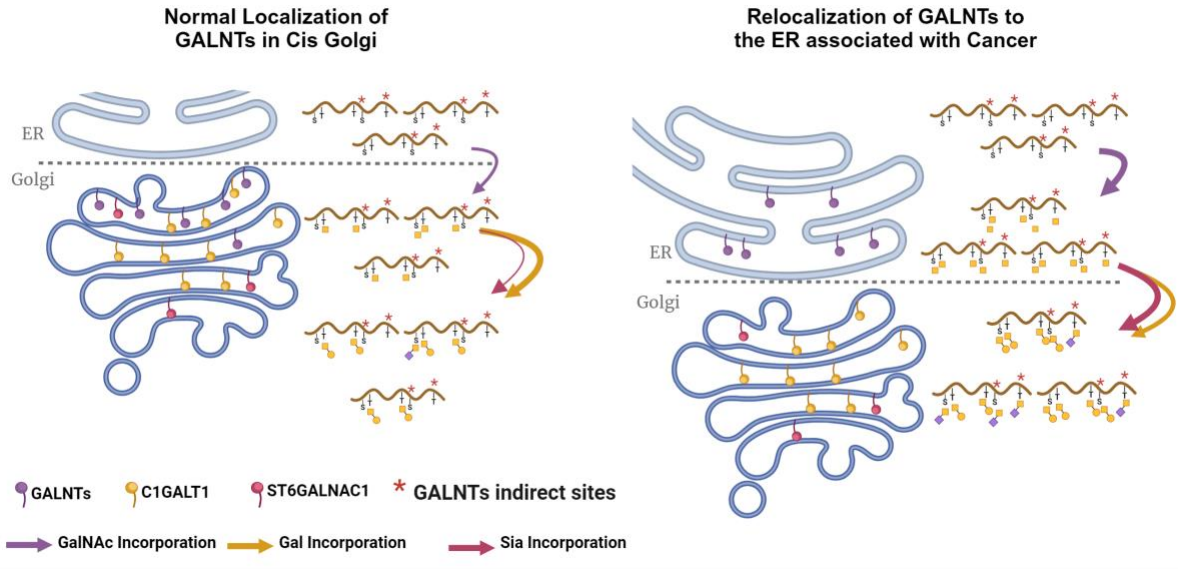

Figure 5. Proposed model of the effect of the localization of GALNTs on site occupancy and sialylation of Tn antigen.

The in vitro enzyme specificity and competition at each step of the synthesis revealed that B3GNT6 specificity to MUC1 Tn sites are negligible compared to those of C1GALT1 and ST6GALNAC1, which suggests that core 3 and core 4 structures are less predominant in MUC1, which supports previous observation when O-GalNAc glycans of MUC1 were compared between normal epithelial breast cell lines and breast cancer cell lines and core 3 and core 4 structures were absent from both type of cells[32]. The role of B3GNT6 and Core 3 structures are found more significant in other mucins such as MUC2, which is the major mucin expressed in colon cancer[33], highlighting the potential peptide specificity of B3GNT6 to be responsible for the weak activity observed in this work. Relative specify of ST6GALNAC1 and ST6GALNAC2 for Tn and T antigens suggest that ST6GALNAC1 and not ST6GALNAC2 is responsible for the α-2 sialylation of these two antigens.

Interestingly, the specific reactivity of C1GALT1 and ST6GALNAC1 to either the fully occupied or under saturated MUC1 site by Tn antigen revealed that C1GALT1 preferred the sub-saturation state (associated with T8, S19, and T20) whereas ST6GALNAC1 prefers the full saturation state (associated with S9 and T13). In fact, when evaluating the chemoenzymatic synthesis approach followed by Yoshimura et al. 2019, ST6GALNAC1 showed significant specificity to the threonine T13 compared to other threonine residues T8 and T20. Additionally, the serin S19 was not glycosylated by the enzyme whereas low reactivity with S9 was detected[34]. The kinetics parameters calculated in this work (The table in Figure 4, A) revealed that the preference of ST6GALNAC1 to the full occupancy form is only presented when MUC1 expression is high. Interestingly, over expression of MUC1, full occupancy, and the over expression of sTn are all established cancer associated phenotypes. What this result show is that the upregulation of sTn can be partly attributed to the occupancy status marked by the occupancy of S9 and T13 that are the slow and delayed sites to be GalNAc glycosylated via GALNTs. This preference is more pronounced as MUC1 levels increase. Since occupancy is enhanced by GALNTs ER to Golgi retrograde relocation. The upregulation of the overall sialylation of Tn antigen can be an indirect effect of this retrograde mechanism and independent of the enzyme's levels (Figure 5). This notion can be echoed in data reported by Swell et al. 2006 that showed that not all sTn positive samples showed significant ST6GALNAC1 expression.[14]

3. Materials and Methods

3.1 Expression of MUC1 peptide fusion protein and tag removal

Gene sequences encoding MUC1 peptide fusion proteins were synthesize and inserted in the pET-21b(+) E. coli expression vector by BIOMATIK (Ontario, Canada). The sequence of the synthesized fusion protein is provided in the supplementary information (Figure S1). Transformation of *E. Coli.* BL21 (Sigma-Aldrich, Cat. no. CMC0014) was carried out using heat shock protocol and grown on Luria Broth (LB) agar plates supplemented with ampicillin to a final concentration of 50 µg/ml at 37 °C. Expression was carried out in Terrific Broth (TB) medium overnight at 16 °C with shaking at 150 rpm. Glucose was added to a concentration of 2 % wt/vol at all stages of expression. Expression was induced by isopropyl β-D-thiogalactoside (IPTG, Sigma-Aldrich, Cat. No. 16758) to a final concentration of 1 mM at OD600 of 0.4-0.6.

The cell pellets from E. coli expression were lysed by incubating at 4 °C for 4 hours in an IMAC binding buffer consisting of 20 mM Tris-HCl, 5 mM Imidazole-HCl, 500 mM NaCl, 0.05% (w/v) sodium azide, and 10% (v/v) glycerol at pH 7.9. This buffer was supplemented with a protease inhibitor tablet (cOmplete, Sigma Aldrich, Cat. no. 11873580001) and 20 mg of lysozyme per 10 ml of buffer. The cell lysate was then clarified through centrifugation at 48,000 RCF for 30 minutes followed by filtration using a 0.45 µM sterile filter. IMAC was conducted on a protein liquid chromatography (FPLC) system ÄKTA Start utilizing 1 ml HiTrap Chelating High-Performance columns (Cytiva, Cat. no. 17-0408-01). Proteins were eluted with a gradient ranging from 5 to 500 mM imidazole-HCl over 15 minutes at a flow rate of 1 ml/min. The eluted fractions were collected, and SDS-PAGE was used to verify protein purity and size. The purified enzymes were quantified with the Bradford protein assay kit (ThermoFisher, A55866) and stored at -80 °C in a freezing buffer comprising 20 mM Tris-HCl, 150 mM NaCl, and 10% (v/v) glycerol at pH 7.6.

TEV protease was expressed from the expression plasmid pRK793 (Addgene plasmid #8827; http://n2t.net/addgene:8827; RRID: Addgene_8827), in *E. coli*. BL21(DE3)-RIL cells as described previously. [13] Briefly, Tag removal was performed overnight at 4 °C by incubating the TEV protease with the fusion proteins at an optimized ratio (Supportive Information, Figure S2) in a buffer of 50 mM Tris-HCl, 0.5 mM EDTA, 1 mM DTT, pH 8.0

3.2 Expression of Glycosyltransferases

HEK293F cells (R79007, Thermo Fisher) were a gift from E. Sturrock (University of Cape Town, South Africa). All expression plasmids for GTs used in this work were purchased from the plasmid repository DNASU. These expression vectors are the pGEn2 vectors constructed by Moremen et al. (2018). GTs expressed from pGEn2 vectors are N-terminally tagged with signalling peptide-8X His-Avi tag-super folder GFP-Tev protease recognition sites. To remove the tag and obtain the enzymes in their native sequence, purified (tagged) proteins were incubated overnight at 4 °C with TEV protease at a ratio of 1:5 (TEV protease: fusion protein) in TEV protease buffer (50 mM Tris-HCl, 0.5 mM EDTA, 1 mM DTT, pH 8.0). The mixture was then processed by an IMAC to collect the untagged enzymes from the flow through fractions. Enzymes were stored at −80 °C in freezing buffer (20 mM Tris-HCl, 150 mM NaCl and 10% (v/v) Glycerol, pH 7.6).

3.3 Enzyme assay

The assay was enzymatic reaction were monitored using the UGC assay developed previously, the components of the assay are: l-Lactic dehydrogenase (LDH, L2500), PK (P1506), cytidine 5′-monophosphate (CMP, C1006), bovine serum albumin (BSA, A3059), *N*-(2-hydroxyethyl) piperazine-*N*′-(2-ethanesulfonic acid), 4-(2-hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES, H3375). The induction of reaction and preparation of reaction mixtures were performed following the standard protocol.[13] Details of concentrations of the substrates and the enzymes and specific conditions are mentioned in the supportive information.

3.4 Preparative glycosylation

The glycosylation of the fusion protein with GALNT1/2 and GALNT4 was optimized to maximize yield and minimize protein aggregation during incubation. The optimization process involved refreshing the reaction by adding enzymes and donor halfway through incubation, using non-denaturing detergents like Triton-X 100, and maintaining an optimum temperature of 37°C. The protocol balanced increasing substrate concentrations above their Km values with cost constraints and the risk of peptide aggregation. A stable stock solution of the fusion protein at 717 µM was used. The final optimized protocol involves a fusion protein concentration greater than 400 µM, with UDP-GalNAc added in two batches: the first at a 2:1 ratio to potential glycosylation sites and the second after 12 hours at a 1:1 ratio. Enzymes are also added in two batches, starting with 1 µM and adding half that amount after 12 hours. The reaction buffer consists of 25 mM Tris-HCl (pH 7.4), 50 mM NaCl, 10% glycerol, 0.25% Triton-X 100, and 10 mM MnCl2. For example, with a final fusion protein concentration of 400 µM and three glycosylation sites, the first batch of UDP-GalNAc would be 2400 µM and the second 1200 µM.

3.5 LC-MS analysis

LC-MS analysis was performed using a Q-Exactive quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) coupled with a Dionex Ultimate 3000 nano-UPLC system, and data were acquired with Xcalibur v4.1.31.9, Chromeleon v6.8 (SR13), Orbitrap MS v2.9 (build 2926), and Thermo Foundations 3.1 (SP4). Peptides were prepared in a solution contains 0.1% (v/v) formic acid (FA) and 2% (v/v) acetonitrile (ACN). Final concentration of the peptides were estimated to be around 10 nM, with a volume equivalent to 50 fmol of peptide injected per sample. Samples were trapped on a PepMap100 C18 column and separated using a ReproSil-Pur 120 C-18-AQ column with a multi-step gradient of Solvent A (0.1% FA in LC water) and Solvent B (0.1% FA in ACN). The mass spectrometer was operated in positive

ion mode at a capillary temperature of 320°C and an electrospray voltage of 1.95 kV. Full scan and data-dependent MS/MS settings were used to determine MS1 and MS2 m/z distributions, detailed in the Supportive Information.

3.6 Statistics

All of the data points were repeated three times. Shapiro–Wilk normality tests were performed to confirm the normal distribution of the data. Normality tests and regressions were performed using GraphPad Prism 8 software. The data are presented as (mean value) ± (standard deviation).

Funding Sources

The National Research Foundation (NRF) CPPR 466624 and the South African Medical Research Council Self-Initiated Grant (SAMRC SIG) 416090.

ACKNOWLEDGMENT

This work is based in part upon research supported by the National Research Foundation (NRF) CPPR 466624 grant (K.J.N.) and the South African Medical Research Council Self-Initiated Grant (SAMRC SIG) 416090 grant. A.N. Thanks to the Scientific Computing Research Unit (SCRU) for graduate fellowship funding. We thank the Centre for High-Performance Computing for providing additional computing resources.

REFERENCES

(1) Pothukuchi, P.; Agliarulo, I.; Russo, D.; Rizzo, R.; Russo, F.; Parashuraman, S. Translation of Genome to Glycome: Role of the Golgi Apparatus. FEBS Lett 2019, 593 (17), 2390-2411. DOI: 10.1002/1873-3468.13541.

(2) Ashkani, J.; Naidoo, K. J. Glycosyltransferase Gene Expression Profiles Classify Cancer Types and Propose Prognostic Subtypes. Scientific Reports 2016, 6, 26451. DOI: 10.1038/srep26451.

(3) Pinho, S. S.; Reis, C. A. Glycosylation in Cancer: Mechanisms and Clinical Implications. Nat Rev Cancer 2015, 15 (9), 540-555, Review. DOI: 10.1038/nrc3982.

(4) Murrell, M. P.; Yarema, K. J.; Levchenko, A. The Systems Biology of Glycosylation. ChemBioChem 2004, 5 (10), 1334-1347. DOI: 10.1002/cbic.200400143.

(5) Goth, C. K.; Mehta, A. Y.; McQuillan, A. M.; Baker, K. J.; Hanes, M. S.; Park, S. S.; Stavenhagen, K.; Hjortø, G. M.; Heimburg-Molinaro, J.; Chaikof, E. L.; et al. Chemokine Binding to Psgl-1 Is Controlled by O-Glycosylation and Tyrosine Sulfation. Cell Chemical Biology 2023, 30 (8), 893-905.e897. DOI: 10.1016/j.chembiol.2023.06.013 (acccessed 2024/05/02).

(6) Yoshimura, Y.; Denda-Nagai, K.; Takahashi, Y.; Nagashima, I.; Shimizu, H.; Kishimoto, T.; Noji, M.; Shichino, S.; Chiba, Y.; Irimura, T. Products of Chemoenzymatic Synthesis Representing Muc1 Tandem Repeat Unit with T-, St- or Stn-Antigen Revealed Distinct Specificities of Anti-Muc1 Antibodies. Scientific Reports 2019, 9 (1), 16641. DOI: 10.1038/s41598-019-53052-1.

(7) Narimatsu, Y.; Joshi, H. J.; Nason, R.; Van Coillie, J.; Karlsson, R.; Sun, L.; Ye, Z.; Chen, Y. H.; Schjoldager, K. T.; Steentoft, C.; et al. An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. Mol Cell 2019, 75 (2), 394-407 e395. DOI: 10.1016/j.molcel.2019.05.017.

(8) Rabouille, C.; Hui, N.; Hunte, F.; Kieckbusch, R.; Berger, E. G.; Warren, G.; Nilsson, T. Mapping the Distribution of Golgi Enzymes Involved in the Construction of Complex Oligosaccharides. J Cell Sci 1995, 108 (Pt 4), 1617-1627. DOI: 10.1242/jcs.108.4.1617.

(9) Bui, S.; Mejia, I.; Diaz, B.; Wang, Y. Adaptation of the Golgi Apparatus in Cancer Cell Invasion and Metastasis. Front Cell Dev Biol 2021, 9, 806482. DOI: 10.3389/fcell.2021.806482.

(10) Gill, D. J.; Chia, J.; Senewiratne, J.; Bard, F. Regulation of O-Glycosylation through Golgi-to-Er Relocation of Initiation Enzymes. J Cell Biol 2010, 189 (5), 843-858. DOI: 10.1083/jcb.201003055.

(11) Gill, D. J.; Tham, K. M.; Chia, J.; Wang, S. C.; Steentoft, C.; Clausen, H.; Bard-Chapeau, E. A.; Bard, F. A. Initiation of Galnac-Type O-Glycosylation in the Endoplasmic Reticulum Promotes Cancer Cell Invasiveness. Proc Natl Acad Sci U S A 2013, 110 (34), E3152-3161. DOI: 10.1073/pnas.1305269110.

(12) Xi, X.; Wang, J.; Qin, Y.; Huang, W.; You, Y.; Zhan, J. Glycosylated Modification of Muc1 Maybe a New Target to Promote Drug Sensitivity and Efficacy for Breast Cancer Chemotherapy. Cell Death Dis 2022, 13 (8), 708. DOI: 10.1038/s41419-022-05110-2.

(13) Nashed, A.; Naidoo, K. J. Universal Glycosyltransferase Continuous Assay for Uniform Kinetics and Inhibition Database Development and Mechanistic Studies Illustrated on St3gal1, C1galt1, and Fut1. ACS Omega 2024, 9 (15), 17518-17532. DOI: 10.1021/acsomega.4c00485.

(14) Sewell, R.; Bäckström, M.; Dalziel, M.; Gschmeissner, S.; Karlsson, H.; Noll, T.; Gätgens, J.; Clausen, H.; Hansson, G. C.; Burchell, J.; et al. The St6galnac-I Sialyltransferase Localizes Throughout the Golgi and Is Responsible for the Synthesis of the Tumor-Associated Sialyl-Tn -Glycan in Human Breast Cancer. J Biol Chem 2006, 281 (6), 3586-3594. DOI: 10.1074/jbc.M511826200.

(15) Pantazopoulou, A.; Glick, B. S. A Kinetic View of Membrane Traffic Pathways Can Transcend the Classical View of Golgi Compartments. Frontiers in Cell and Developmental Biology 2019, 7. DOI: ARTN 153

10.3389/fcell.2019.00153.
(16) Pedelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. Nat Biotechnol 2006, 24 (1), 79-88. DOI: 10.1038/nbt1172.
(17) Pratt, M. R.; Hang, H. C.; Ten Hagen, K. G.; Rarick, J.; Gerken, T. A.; Tabak, L. A.; Bertozzi, C. R. Deconvoluting the Functions of Polypeptide N-Alpha-Acetylgalactosaminyltransferase Family Members by Glycopeptide Substrate Profiling. Chem Biol 2004, 11 (7), 1009-1016. DOI: 10.1016/j.chembiol.2004.05.009.
(18) de Las Rivas, M.; Lira-Navarrete, E.; Gerken, T. A.; Hurtado-Guerrero, R. Polypeptide Galnac-Ts: From Redundancy to Specificity. Curr Opin Struct Biol 2019, 56, 87-96. DOI: 10.1016/j.sbi.2018.12.007.
(19) Hassan, H.; Reis, C. A.; Bennett, E. P.; Mirgorodskaya, E.; Roepstorff, P.; Hollingsworth, M. A.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H. The Lectin Domain of Udp-N-Acetyl-D-Galactosamine: Polypeptide N-Acetylgalactosaminyltransferase-T4 Directs Its Glycopeptide Specificities. J Biol Chem 2000, 275 (49), 38197-38205. DOI: 10.1074/jbc.M005783200.
(20) Wandall, H. H.; Hassan, H.; Mirgorodskaya, E.; Kristensen, A. K.; Roepstorff, P.; Bennett, E. P.; Nielsen, P. A.; Hollingsworth, M. A.; Burchell, J.; Taylor-Papadimitriou, J.; et al. 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 1997, 272 (38), 23503-23514. DOI: 10.1074/jbc.272.38.23503.
(21) Wandall, H. H.; Irazoqui, F.; Tarp, M. A.; Bennett, E. P.; Mandel, U.; Takeuchi, H.; Kato, K.; Irimura, T.; Suryanarayanan, G.; Hollingsworth, M. A.; et al. 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 2007, 17 (4), 374-387. DOI: 10.1093/glycob/cwl082.
(22) Coelho, H.; Rivas, M. L.; Grosso, A. S.; Diniz, A.; Soares, C. O.; Francisco, R. A.; Dias, J. S.; Companon, I.; Sun, L.; Narimatsu, Y.; et al. Atomic and Specificity Details of Mucin 1 O-Glycosylation Process by Multiple Polypeptide Galnac-Transferase Isoforms Unveiled by Nmr and Molecular Modeling. JACS Au 2022, 2 (3), 631-645. DOI: 10.1021/jacsau.1c00529.
(23) Luo, J.; Zou, H.; Guo, Y.; Tong, T.; Ye, L.; Zhu, C.; Deng, L.; Wang, B.; Pan, Y.; Li, P. Src Kinase-Mediated Signaling Pathways and Targeted Therapies in Breast Cancer. Breast Cancer Res 2022, 24 (1), 99. DOI: 10.1186/s13058-022-01596-y.
(24) Garnham, R.; Scott, E.; Livermore, K. E.; Munkley, J. St6gal1: A Key Player in Cancer. Oncol Lett 2019, 18 (2), 983-989. DOI: 10.3892/ol.2019.10458.
(25) Zeng, J.; Mi, R.; Wang, Y.; Li, Y.; Lin, L.; Yao, B.; Song, L.; van Die, I.; Chapman, A. B.; Cummings, R. D.; et al. Promoters of Human Cosmc and T-Synthase Genes Are Similar in Structure, yet Different in Epigenetic Regulation. J Biol Chem 2015, 290 (31), 19018-19033. DOI: 10.1074/jbc.M115.654244.
(26) Ju, T.; Lanneau, G. S.; Gautam, T.; Wang, Y.; Xia, B.; Stowell, S. R.; Willard, M. T.; Wang, W.; Xia, J. Y.; Zuna, R. E.; et al. Human Tumor Antigens Tn and Sialyl Tn Arise from Mutations in Cosmc. Cancer Res 2008, 68 (6), 1636-1646. DOI: 10.1158/0008-5472.CAN-07-2345.
(27) Revoredo, L.; Wang, S.; Bennett, E. P.; Clausen, H.; Moremen, K. W.; Jarvis, D. L.; Ten Hagen, K. G.; Tabak, L. A.; Gerken, T. A. Mucin-Type O-Glycosylation Is Controlled by Short- and Long-Range Glycopeptide Substrate Recognition That Varies among Members of the Polypeptide Galnac Transferase Family. Glycobiology 2016, 26 (4), 360-376. DOI: 10.1093/glycob/cwv108.
(28) Hanisch, F. G.; Muller, S. Muc1: The Polymorphic Appearance of a Human Mucin. Glycobiology 2000, 10 (5), 439-449. DOI: 10.1093/glycob/10.5.439.
(29) Wang, S.; Chen, C.; Gadi, M. R.; Saikam, V.; Liu, D.; Zhu, H.; Bollag, R.; Liu, K.; Chen, X.; Wang, F.; et al. Chemoenzymatic Modular Assembly of O-Galnac Glycans for Functional Glycomics. Nat Commun 2021, 12 (1), 3573. DOI: 10.1038/s41467-021-23428-x.
(30) Kono, M.; Tsuda, T.; Ogata, S.; Takashima, S.; Liu, H.; Hamamoto, T.; Itzkowitz, S. H.; Nishimura, S.; Tsuji, S. Redefined Substrate Specificity of St6galnac Ii: A Second Candidate Sialyl-Tn Synthase. Biochem Biophys Res Commun 2000, 272 (1), 94-97. DOI: 10.1006/bbrc.2000.2745.
(31) Marcos, N. T.; Pinho, S.; Grandela, C.; Cruz, A.; Samyn-Petit, B.; Harduin-Lepers, A.; Almeida, R.; Silva, F.; Morais, V.; Costa, J.; et al. Role of the Human St6galnac-I and St6galnac-Ii in the Synthesis of the Cancer-Associated Sialyl-Tn Antigen. Cancer Res 2004, 64 (19), 7050-7057. DOI: 10.1158/0008-5472.CAN-04-1921.
(32) Lloyd, K. O.; Burchell, J.; Kudryashov, V.; Yin, B. W.; Taylor-Papadimitriou, J. Comparison of O-Linked Carbohydrate Chains in Muc-1 Mucin from Normal Breast Epithelial Cell Lines and Breast Carcinoma Cell Lines. Demonstration of Simpler and Fewer Glycan Chains in Tumor Cells. J Biol Chem 1996, 271 (52), 33325-33334. DOI: 10.1074/jbc.271.52.33325.
(33) An, G.; Wei, B.; Xia, B.; McDaniel, J. M.; Ju, T.; Cummings, R. D.; Braun, J.; Xia, L. Increased Susceptibility to Colitis and Colorectal Tumors in Mice Lacking Core 3-Derived O-Glycans. J Exp Med 2007, 204 (6), 1417-1429. DOI: 10.1084/jem.20061929.
(34) Yoshimura, Y.; Denda-Nagai, K.; Takahashi, Y.; Nagashima, I.; Shimizu, H.; Kishimoto, T.; Noji, M.; Shichino, S.; Chiba, Y.; Irimura, T. Products of Chemoenzymatic Synthesis Representing Muc1 Tandem Repeat Unit with T-, St- or Stn-Antigen Revealed Distinct Specificities of Anti-Muc1 Antibodies. Sci Rep 2019, 9 (1), 16641. DOI: 10.1038/s41598-019-53052-1.