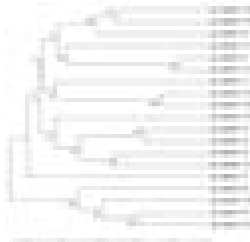
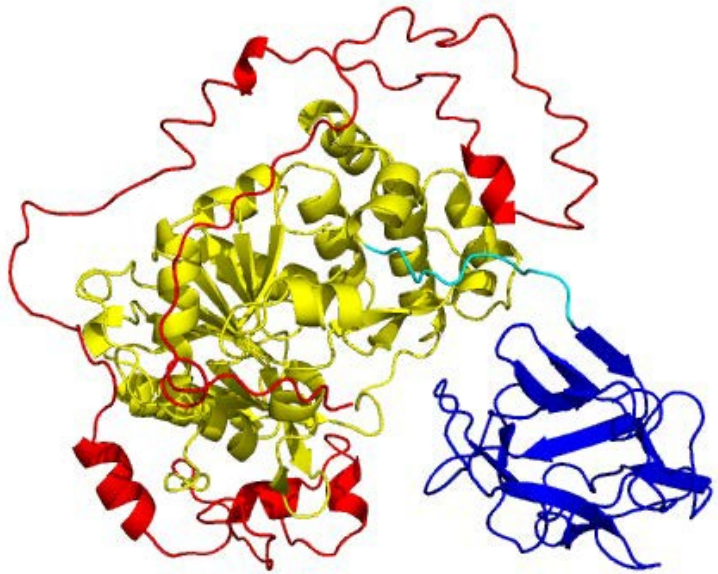
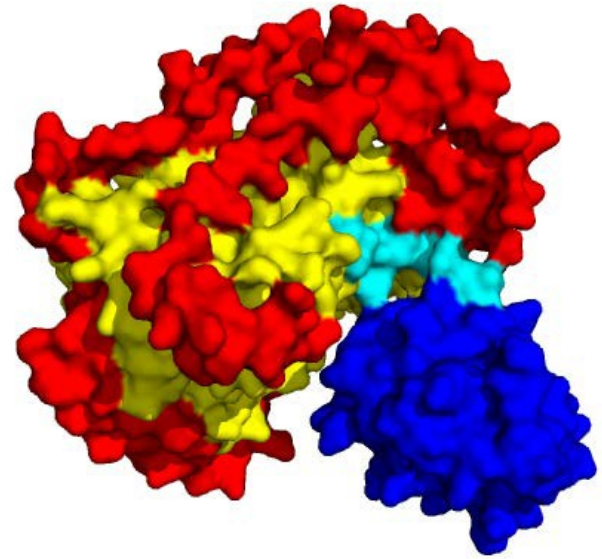






UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase-6 (pp-GalNAc-T6): Role in cancer and prospects as a drug target

Running title: pp-GalNAc-T6 in cancer

ABSTRACT

UDP-*N*-acetyl-D-galactosamine: polypeptide *N*-acetylgalactosaminyl transferase-6 (pp-GalNAc-T6) is a member of the *N*-acetyl-D-galactosamine transferase family. It catalyzes the addition of *N*-acetyl-D-galactosamine to proteins, often the first step in O-glycosylation of proteins. Glycosylated proteins play important roles *in vivo* in the cell membrane. These are often involved in cell-cell adhesion, cytoskeleton regulation and immune recognition. pp-GalNAc-T6 has been shown to be upregulated in a number of types of cancer. Abnormally glycosylated forms of mucin 1 (substrate of the enzyme), are used clinically as a biomarker for breast cancer. There is potential for other products of the pp-GalNAc-T6 catalyzed reaction to be used. It is also possible that pp-GalNAc-T6 itself could be used as a biomarker, since levels of this protein tend to be low in non-malignant tissues. pp-GalNAc-T6 has been implicated in malignant transformation and metastasis of cancer cells. As such, it has considerable potential as a target for chemotherapy. To date, no selective inhibitors of the enzyme have been identified. However, general inhibitors of the enzyme family result in reduced cell surface *O*-linked glycosylation and induce apoptosis in cultured cells. Thus, a selective inhibitor of pp-GalNAc-T6 is likely to target cancer cells and could be developed into a novel anticancer therapy.

Keywords

UDP-GalNAc, UDP-*N*-acetyl-D-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase-6, cancer, metastasis, biomarker, *GALNT6*

INTRODUCTION

UDP-*N*-acetyl-D-galactosamine: polypeptide *N*-acetylgalactosaminyl transferase-6 (pp-GalNAc-T6; EC 2.4.1.41; UniProt: Q8NCL4) is an enzyme first identified in 1999 [1, 2]. Polypeptide *N*-acetylgalactosaminyl transferases are classified in the GT27 family, by the CAZy glycosyltransferase classification [3]. This group of enzymes catalyses the addition of *N*-acetyl-D-galactosamine (GalNAc) to threonine or serine residues in proteins using UDP-*N*-acetyl-D-galactosamine (UDP-GalNAc) as the donor, generating an oxygen-linked GalNAc (*O*-GalNAc) with the release of UDP [4, 5]. The addition of GalNAc to a peptide is often the first step in the process of glycosylation [6]. The second step involves GalNAc: glycosyltransferases adding further monosaccharides to the *O*-GalNAc, in a step-wise process, creating an *O*-glycan chain [7]. *O*-glycans play important roles in the structure and function of membrane proteins, and changes in these proteins are thought to lead to malignant transformation of tissues [8].

Post-translational modification of proteins by glycosylation occurs in the Golgi apparatus and endoplasmic reticulum of cells. Like other family members, pp-GalNAc-T6 resides on the lumen side of the membrane of the Golgi apparatus [1, 9-11]. It is a type II membrane protein, in which its N-terminal tail spans the Golgi membrane [12]. When Src is activated, pp-GalNAc-T family members move into the endoplasmic reticulum [12].

pp-GalNAc-T6 is a protein in a family of 20 GalNAc transferases [13]. The family members differ in their specificity towards target proteins, although some overlap exists. They also differ in the cells in which they are expressed although, again, there is overlap [14]. The proteins recognise the serine or threonine glycosylation site on the target protein as well as the nearby amino acid residues and secondary structure [15]. pp-GalNAc-T1 and pp-GalNAc-T2 have a preference for proteins that have yet to be glycosylated and are inhibited by neighbouring glycosylations [16]. However, pp-GalNAc-T4 and pp-GalNAc-T10 prefer previously glycosylated proteins [17-19]. pp-GalNAc-T6 also appears to be in this latter category [2].

MOLECULAR GENETICS

The human *GALNT6* gene, encodes pp-GalNAc-T6 and is located on chromosome 12 at locus 12q13.13. The open reading frame is 1869 base pairs long. The gene has 12 exons arranged in a pattern very similar to that of *GALNT3* [1]. The similarity in sequence and the layout of the two genes suggest that it may be a product of a late gene duplication event [1, 20]. Phylogenic analysis of the protein sequences show that pp-GalNAc-T3 is the most closely related member of the GalNAc family to pp-GalNAc-T6, with isoforms pp-GalNAc-T8, 9 and 18 being the most distant (Figure 1).

GALNT6 expression is tissue and developmental stage dependent. It has been shown to be upregulated when cells are differentiating, and in malignancy [21-24]. Expression is minimal in normal tissues [21]. Northern blot analysis on 26 organ samples was used to determine the tissue expression pattern and showed that the presence of the *GALNT6* transcript in tissues from fibroblasts, brain, trachea, pancreas and placenta [1]. However, Western blot analyses showed that the pp-GalNAc-T6 protein is barely detectable in normal human tissues [21].

STRUCTURE

pp-GalNAc-T6 has a predicted, native molecular mass of 71159 Da and is 622 amino acids in length. It was demonstrated that it is 64% similar to its family member pp-GalNAc-T3 at the protein level. It differs largely in the first ~140 residues of the N-terminal domain, and is largely conserved in the active site and a manganese ion binding site. Since no structural data is available on this isoform, a homology model was constructed (Figure 2; for details of how the model was constructed see figure legend). Key functional regions that have been shown to be important in pp-GalNAc-T10 (the only family member for which a high resolution structure has been published, PDB: 2D7R; [19]) were used to identify the corresponding parts of pp-GalNAc-T6. These analyses predicted that the protein is composed of multiple domains (Figure 2). The protein is composed of a hydrophobic membrane spanning signal sequence (residues 9-28), an N-terminal, largely random coil cytoplasmic tail, a GT1-type type catalytic subdomain A (residues 176-285), and a catalytic subdomain B (residues 348-410) joined to an R-type lectin domain (residues 496-622) by a flexible unstructured linker. The lectin domain has a typical β-trefoil fold. The cytoplasmic tail has been proposed to be involved with peripheral Golgi membrane protein tethering complexes [25].

The length of the linker differs between family members and has been shown to affect the specificity of the enzyme. pp-GalNAc-T1 and 2 have longer linkers in comparison to the shorter linkers seen in pp-GalNAc-T10 These longer linkers are seen in the isoforms that require a substrate that has been already glycosylated. A smaller distance between the catalytic and lectin domain and a narrower cleft between them increases the effective concentration of the protein substrate aiding the rate of reaction [19]. From the model it must be noted that this isoform of the protein has an extended N-terminal domain that differs from the other isoforms (Figure 2). It appears in the model to be mostly unstructured. This is unlikely to be the case, and it is probably a transmembrane region (such regions are typically modelled less well by homology modelling software). In the crystal structure of pp-GalNAc-T10 this region was not present as membrane spanning regions are often difficult to crystallise [19]. The manganese ion binding site is conserved in all pp-GalNAc-T family members. The ion binding site has six ligands: two histidine residues (predicted to be residues His271 and His407 in pp-GalNAc-T6), one asparate (Asp269), two oxygens from the UDP-*N*-acetylgalactosamine substrate and a water molecule [19].

FUNCTION

The lectin domain recognises glycosylated polypeptide substrates [26]. The domain contains the three repeats of a CCQxW motif, in subdomains α, β and γ. In pp-GalNAc-T10 the β subdomain interacts with the GalNAc and it seems likely that the same is true in pp-GalNAc-T6 [19]. Threonine residues in the acceptor protein are generally more effective galactose acceptors than serines [27]. However, the residues neighboring the serine/threonine are also important in determining whether the protein will be a substrate or not, particularly in positions -3 and +3 of the acceptor residue. Proline residues in positions +1 and +3 also help drive substrate preference, and charged residues close by have a negative effect [28, 29].

MECHANISM

There has been considerable discussion in the literature about the mechanism of action of this family of proteins. Understanding their mechanism would be an important initial step for assessing if pp-GalNAc-T6 is a viable drug target. The conservation of the

catalytic domain makes it likely that the catalytic mechanism is also well-conserved throughout the family. Therefore, studies on other family members are likely to apply to pp-GalNAc-T6.

There has been two mechanisms proposed [19, 30]. The first is a double replacement mechanism, in which the side chain amide NH_2 from Asn-383 (Gln-346 in pp-GalNAc-T10) acts as a nucleophile and attacks the anomeric carbon of the GalNAc ring, and this leads to the formation of an enzyme-glycosyl intermediate [19]. Asparagine does not normally act as a nucleophile; however it has been proposed that the glycosidic bond in UDP-GalNAc is strained and is elongated making it very susceptible to nucleophilic attack when found in the enzyme binding pocket [19]. The intermediate state is proposed to be stabilised by eight hydrogen bonds [19]. The reaction ends when the β-phosphate oxygen from the UDP donates electrons to the hydrogen from the hydroxyl group of the serine residue of the protein substrate. The Mn^{2+} interacts with the diphosphate and stabilises it [31]. The hydroxyl group then nucleophilically attacks the anomeric carbon joining the protein substrate and the GalNAc [19]. Electrons from the Asn-GalNAc bond return to the amide of the asparagine. This returns the enzyme to its initial state and thus makes it ready of the next UDP-GalNAc substrate [19].

The other proposed mechanism is a front-side attack mechanism [30]. This was postulated as an alternative mechanism because of doubt about the ability of asparagine and glutamine to act as nucleophiles as well as the potential nucleophile's distance from the β-phosphate oxygen (approximately 7 Å In pp-GalNAc-T2) [30]. In the front side attack mechanism the hydroxyl group of the protein substrate acts as the nucleophile attacking the anomeric carbon. This creates an oxocarbenium ion-like transition state whereby both the oxygen from the serine residue and the β-phosphate oxygen are trivalent. The β-phosphate oxygen then donates electrons to the serine hydroxyl group which nucleophilically attacks the anomeric carbon as in the double replacement mechanism [32].

The reaction is believed to be made possible because of the conformational stress that causes the anomeric carbon of UDP-GalNAc to be elongated and thus susceptible to attack. The intermediate state is stabilised by eight hydrogen bonds with the protein. In pp-GalNAc-T10 this is mediated by Gln346 [19]. In the homology model of pp-GalNAc-T6 (Figure 2) this corresponds to Asn383. A second nucleophilic attack forms a glycosidic bond between the enzyme, UDP-GalNAc and the substrate protein forming an intermediate.

The isoforms differ in their kinetic parameters depending on the substrate. Of the substrates studied with pp-GalNAc-T6 to date, the highest V_{max} has been measured with fibronectin (2.37 nmol/min/μg) with a corresponding K_m of 3.23 mM. Its V_{max} and K_m are reduced to 1.43 nmol/min/μg and 0.61mM respectively when the substrate is HIV_{HIB}gp120 [1].

ROLE IN CANCER

Expression of pp-GalNAc-T6 is often upregulated in malignant tissue. Dysregulation of the gene coding for pp-GalNAc-T6 has been suggested to be an early event in tumorgenisis [33]. This can result in increased glycosylation of proteins in sites that would normally be unoccupied. This can change the structure and function of these

proteins making them oncogenic [33]. In malignant tissue glycoproteins such as mucin are often truncated, and the *O*-GalNAc is not extended as expected in non-malignant tissue. In adenocarcinomas these are often seen as core antigen carbohydrates (e.g. TF and Tn antigens) [34]. Altered tumour-associated core antigens are seen more in malignancy than oncogenes. They are also are more strongly associated with cancer progression [8]. Correct glycosylation is important in controlling cell-cell interactions, a key part in invasion and metastasis.

Mucins are large glycoproteins that have many *O*-oligosaccharides attached to the protein structure. The saccharide units can comprise more than 50% of the mass of the mucin [35, 36]. MUC1 is a transmembrane mucin that interacts with EGFRs, ERα, and β-catenin [37]. Mucins sterically hinder adhesion molecules enabling the malignant cell to more easily become detached and metastasise [38]. Mucins also aid the tumour in evading the immune system by preventing immune cells from binding as well as stopping antigen presentation [39]. MUC1 has been implicated in mammary cancer [21]. Antigens derived from this protein, CA 27.29 and CA 15-3, are biomarkers used to predict breast cancer recurrence [40, 41]. The *O*-oligosaccharides are added to the protein by pp-GalNAc-T6 [21]. These glycosylations add stability to the protein and play a role in the regulation of the cytoskeleton and proliferation of malignant cells in breast cancer [42]. These truncated glycans play a role in altered cell-cell adhesion, malignant transformation and cancer metastasis. These breast cancer cells are able to evade the immune system due to their different antigenic appearance and the breakdown of cell–cell and cell–matrix contacts allows the malignant cells to migrate and metastasise [37]. These altered glycosylation patterns can be used as a marker in premalignant and malignant cancer [43].

GalNAc-T6 is expressed in 81% of breast carcinomas [34]. A statistically significant correlation was observed between pp-GalNAc-T6 expression and tumour stage in breast cancers. pp-GalNAc-T6 is strongly expressed in myoepithelial cells of the breast [34]. Myoepithelial cells are seen as ‘natural tumour suppressors’ because they have been implicated in angiogenesis, polarity and growth of cancer cells [34]. Knockdown of *GALNT6* by miRNA in breast cancer cells enhanced cell-cell adhesion and cell growth. These cells showed to have increased β-catenin and E-cadherin molecule expression than the control cells [21]. pp-GalNAc-T6 has been found to be over expressed in the bone marrow of patients with breast cancer, and has been associated with cancer recurrence [44]. A statistically significant higher risk of recurrence was found in people without lymph node involvement, in comparison to pp-GalNAc-T6 negative patients. RT-PCR analysis on these cells has been suggested as a way to monitor risk of metastasis and response to treatment [44].

pp-GalNAc-T1, T2, T3 and T6 are expressed in oral cell carcinoma [45-48]. They *O*-glycosylate fibronectin forming glycosylated onco-foetal fibronectin (GOF) in the stroma of malignant cells [49]. An *O*-glycan is added to a threonine in the IIICS domain of the fibronectin. GOF has an altered spiral shape, and it is only present in tissues with pathology such as in oral carcinoma, breast carcinoma and rheumatoid arthritis [50]. The presence of GOF is correlated with survival rates of patients with oral cell carcinoma, and either pp-GalNAc-T3 or T6 must be present to allow GOF to form. Neither isoform could be detected in stromal fibroblasts but are present in malignant cells [49]. In malignant cells, the pp-GalNAc-T6 mediated glycosylation of fibronectin tends to result in a large number of short oligosaccharide chains being added to the

surface of the protein [21, 51-53]. This stabilizes the fibronectin molecule promoting an invasive, cancerous phenotype in the cells [54].

pp-GalNAc-T6 has been associated with venous invasion in gastric carcinomas, in the body of the stomach [9]. It is expressed in 79% of gastric carcinomas tested in comparison to 69% in normal gastric mucosa. The percentage was higher in intestinal and diffuse type carcinomas (81% and 86%). In the precancerous state of intestinal metaplasia the percentage presence dropped to 52% [9]. This may reflect the differentiation traits of the tissue. pp-GalNAc-T6 has been suggested as a marker to predict venous invasion in gastric carcinoma [9].

pp-GalNAc-T6 has also been studied once in pancreatic cancer [23, 55]. pp-GalNAc-T6 expression is rare in normal ductal epithelium of the pancreas. However it is expressed in 51% of patients with pancreatic adenocarcinoma. The expression of pp-GalNAc-T6 was mainly seen in early American Joint Committee on Cancer stage, well to moderately differentiated, small tumours [55]. In contrast to other cancer types studied, pp-GalNAc-T6 positive patients had a longer overall survival time than those without the protein. It has been suggested as a prognostic indicator, as in its absence it may recommend the use of more aggressive treatment. However, one disadvantage is that it is not expressed in undifferentiated cancer [55]. Further study will be required to confirm these results as the sample size was limited in this study. In contrast with breast and oral cancers it appears that the inhibition of pp-GalNAc-T6 would have a negative impact on patients' survival as it may enhance cancer progression, although the mechanism of this is not yet known.

The abnormal glycosylation patterns on malignant cells can be used as biomarkers to identify cancer [56]. As many pp-GalNAC-T6 products are associated with the cell membrane it makes them more accessible to antibodies. Thus, it is realistic to propose that they could be used as therapeutic markers. Some of these highly expressed in cancers include Tn, T and STn antigens. Each of these antigens has varying degrees of *O*-glycan chain length and composition [57, 58].

Bloodstream MUC1 (CA15-3) is currently used as a biomarker of prognosis and response to therapy in people with breast cancer. Elevated levels indicate a poor prognosis or treatment failure. CA15-3 cannot be used for screening purposes as it is not only expressed in breast cancer but also in other cancers, as well as some inflammatory conditions such as ovarian hyperstimulation syndrome [59]. pp-GalNAc-T6 itself has potential as a biomarker since it is up-regulated in many types of cancer. One issue would be the development of antibodies (or other reagents) which could reliably recognize this isoform over the other family members.

POTENTIAL AS A DRUG TARGET

Since pp-GalNAc-T6 is not present in non-cancerous tissues, it is a promising a drug target candidate. Inhibition of it would have limited consequences on normal healthy tissue, helping to reduce the side effects encountered with many current treatments. It has also been suggested that altered *O*-glycans in tumours could be targeted by vaccines [60-62]. However, there are some drawbacks to targeting pp-GalNAc-T6. The sequence, structural and mechanistic similarity to pp-GalNAc-T3 may make it challenging to find a compound that inhibits only pp-GalNAc-T6. However, given that there is some functional redundancy between pp-GalNAc-T3 and pp-GalNAc-T6, there

may also be benefits in inhibiting both isoforms in circumstances where both have been clearly linked to pathology. pp-GalNAc-T family enzymes are generally assayed by the detection of glycosylated peptide products using MALDI-TOF mass spectrometry and radioactive substrates [63-65]. These methods would be difficult to carry out with a large number of compounds (typically >500,000 in a high throughput screen) [1]. An enzyme-linked lectin assay (ELLA) has been developed in which glycosylation of an immobilised peptide substrate is detected with a lectin linked to horse radish peroxidase [66]. This has some potential for development into a highly parallel, high throughput assay format. It has also been difficult to identify efficient GalNAcylation targets and sites. However, "SimpleCell" lines have enabled the creation of cell lines that are deficient in genes involved in glycosylation [67]. This may improve our understanding of members of the GalNAc-T family roles in cells.

To date, only a limited range of inhibitors of pp-GalNAc-T family members have been reported. The majority of these are natural metabolites (or close analogues thereof) and are, therefore, unlikely to have much potential as drugs. For example, a range of nucleotide triphosphates (including ATP, ADP, AMP, CTP, GTP and UTP) inhibit pp-GalNAc-T activity purified from the mammary gland of cows [68]. It is not clear what isoform(s) were being inhibited in this study (which was performed before the diversity of the pp-GalNAc-T family was known). UDP and UTP have been shown to competitively inhibit rat and bovine pp-GalNAc-T family members with respect to UDP-*N*-acetylgalactosamine strongly suggesting that these compounds interact at the active site [69, 70]. *N*-acetylgalactosamine and a range of related compounds (including benzyl, phenyl, *p*-nitrophenyl and *o*-nitrophenyl derivatives) have been shown to antagonise the interaction between the substrate and the lectin domain [71]. UMP-hexadecanesulfonic-anhydride (hexadecane-1-sulfonic (((2R,3S,4R,5R)-5-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl phosphoric) anhydride; CAS: 134282-88-5; Scheme 1) competitively inhibits pp-GalNAc-T [72]. The IC_{50} value was 160 μM, which suggests too low an affinity to be a viable drug-like molecule. Furthermore, it seems likely that this compound will be a generic inhibitor of the pp-GalNAc-T family and may also affect other enzymes which have UDP or UDP-sugars as substrates. However, it is possible that, through modification of this compound, it may be possible to obtain selective inhibitors of human pp-GalNAc-T6. It may also be possible to study other substrate analogues, including compounds which combine some similarity to both the UDP-*N*-acetylgalactosamine and the polypeptide substrates.

In an alternative approach, a number of UDP-sugar analogues have been synthesised in which the carbon-oxygen bond between the sugar and the diphosphate moiety is replaced with a carbon-carbon-oxygen bond system [73]. It was hypothesised that these would act as competitive inhibitors of enzymes which have UDP-sugar substrates, especially those where cleavage of the sugar from UDP is required as part of the mechanism [73]. However, these compounds have not been tested for their inhibitory potential. Again, it seems likely that they would act relatively non-specifically and that considerable medicinal chemistry optimisation would be required before they could be considered viable drug leads.

Screening of a library of over 1300 uridine derivatives using an enzyme-linked lectin assay (ELLA) identified two compounds which inhibit pp-GalNAc-T1 [66]. Both compounds incorporate a trihydroxylated benzene group (Scheme 2). The IC_{50} values

for inhibition of pp-GalNAc-T1 were in the 20-25 μM range for both compounds [66]. They also inhibit other members of the family (with IC_{50} values in the range 5-40 μM); however, pp-GalNAc-T6 was not tested. No inhibition of other sugar transferases was detected [66]. These data suggest that these inhibitors are specific to the pp-GalNAc-T family, but do not differentiate between family members. One of these compounds (1-68A; (E)-2,3,4-trihydroxybenzaldehyde *O*-(((2*R*,3*S*,4*R*,5*R*)-5-(2,4-dioxo-3,4-dihydropyrimidin-1(2*H*)-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl) oxime) reduced cell surface glycosylation and induced apoptosis [74]. Since this compound inhibits several members of the pp-GalNAc-T family with roughly equivalent potency, it is not possible to say which family member(s) were the most important in this outcome. Nevertheless, these data demonstrate that pp-GalNAc-T family members play key roles in cells, with inhibition resulting in apoptosis. Given that pp-GalNAc-T6 is up-regulated in a range of different types of cancer and that this may result in promotion of the cancerous phenotype (see above), targeting pp-GalNAc-T6 may result in the selective induction of apoptosis in cancer cells – a highly desirable outcome in cancer chemotherapy. Although the pp-GalNAc family members are likely to be similar in structure and catalytic mechanism, there are examples of drugs which differentiate between highly similar enzymes. For example, sildenafil (Viagra) selectively inhibits one human phosphodiesterase isoform and Clorsulon inhibits highly conserved glycolytic enzymes yet selects for helminth parasite enzymes over the mammalian host enzymes [75, 76].

CONCLUSION

Thus far relatively little structural and biochemical information is available for pp-GalNAc-T6. Much of this has been predicted from the properties of other members of the pp-GalNAc-T family, and in particular pp-GalNAc-T3 due to its high sequence similarity. However from the clinical and tissue based studies that have been carried out it seems to have significant potential. It is up regulated in cancer, with little expression in other cells. Therefore, it has a variety of possible uses in therapy; it may be used as a biomarker itself, as a drug target or a product of its reaction could be used as a biomarker. Given that it is particularly involved in metastasis it has the potential to be targeted in instances where all other treatment options fail. This protein warrants future study. Furthermore, the study of glycosylation events and patterns, and how these change in the cancerous state, presents an important frontier in molecular oncology.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers who made many useful and constructive suggestions to improve this review.

FIGURE AND SCHEME LEGENDS

Figure 1: An UPGMA tree with bootstrap analysis (2000 replicates) of the protein sequences of the human pp-GalNAc-T family [77, 78]. The distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site [79]. The analysis involved 19 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 448 positions in the final dataset. Sequences were aligned using ClustalW [80] and the tree generated using MEGA 7 [81-83]. The alignment includes all currently known pp-GalNAc-T isoforms and also proteins classified as "UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase like" (pp-GalNAc-TL). pp-GalNAc-T15 is also known as pp-GalNAc-T2L, pp-GalNAc-T16 as pp-GalNAc-T1L, pp-GalNAc-T18 as pp-GalNAc-T4L and pp-GalNAc-T20 as both pp-GalNAc-T17 and pp-GalNAc-T6L.

Figure 2: Homology model of pp-GalNAc-T6. The N-terminal domain is depicted in mid grey, the catalytic domain is in light grey and the lectin domain in dark grey. The predicted structure is shown in a ribbon representation (left) and a space-filling representation (right). The model was constructed using Phyre 2, in the intensive mode [84] using sequence NP_009141, minimized using YASARA [85] and visualized using PyMol (www.pymol.org). The protein structure file is provided as electronic supplementary information.

Scheme 1: The structure of UMP-hexadecanesulfonic-anhydride, an inhibitor of pp-GalNAc-T family enzymes.

Scheme 2: Two uridine derivatives identified from a medium-throughput screen which inhibit members of the pp-GalNAc-T family of enzymes. The names are those used in the original paper reporting their synthesis and inhibitory potential [66].

References

[1] Bennett, E.P.; Hassan, H.; Mandel, U.; Hollingsworth, M.A.; Akisawa, N.; Ikematsu, Y.; Merkx, G.; van Kessel, A.G.; Olofsson, S.; Clausen, H., Cloning and characterization of a close homologue of human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase-T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy. *J Biol Chem,* **1999**, *274*, (36), 25362-25370.

[2] Ten Hagen, K.G.; Tetaert, D.; Hagen, F.K.; Richet, C.; Beres, T.M.; Gagnon, J.; Balys, M.M.; VanWuyckhuyse, B.; Bedi, G.S.; Degand, P.; Tabak, L.A., Characterization of a UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase that displays glycopeptide N-acetylgalactosaminyltransferase activity. *J Biol Chem,* **1999**, *274*, (39), 27867-27874.

[3] Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res,* **2009**, *37*, (Database issue), D233-238.

[4] White, T.; Bennett, E.P.; Takio, K.; Sorensen, T.; Bonding, N.; Clausen, H., Purification and cDNA cloning of a human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. *J Biol Chem,* **1995**, *270*, (41), 24156-24165.

[5] Kato, K.; Takeuchi, H.; Kanoh, A.; Mandel, U.; Hassan, H.; Clausen, H.; Irimura, T., N-acetylgalactosamine incorporation into a peptide containing consecutive threonine residues by UDP-N-acetyl-D-galactosaminide:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology,* **2001**, *11*, (10), 821-829.

[6] Hurtado-Guerrero, R., Recent structural and mechanistic insights into protein O-GalNAc glycosylation. *Biochem Soc Trans,* **2016**, *44*, (1), 61-67.

[7] Van den Steen, P.; Rudd, P.M.; Dwek, R.A.; Opdenakker, G., Concepts and principles of O-linked glycosylation. *Critical reviews in biochemistry and molecular biology,* **1998**, *33*, (3), 151-208.

[8] Hakomori, S., Tumor-associated carbohydrate antigens defining tumor malignancy: basis for development of anti-cancer vaccines. *Adv Exp Med Biol,* **2001**, *491*, 369-402.

[9] Gomes, J.; Marcos, N.T.; Berois, N.; Osinaga, E.; Magalhaes, A.; Pinto-de-Sousa, J.; Almeida, R.; Gartner, F.; Reis, C.A., Expression of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 in gastric mucosa, intestinal metaplasia, and gastric carcinoma. *J Histochem Cytochem,* **2009**, *57*, (1), 79-86.

[10] Beaman, E.M.; Brooks, S.A., The extended ppGalNAc-T family and their functional involvement in the metastatic cascade. *Histol Histopathol,* **2014**, *29*, (3), 293-304.

[11] Rottger, S.; White, J.; Wandall, H.H.; Olivo, J.C.; Stark, A.; Bennett, E.P.; Whitehouse, C.; Berger, E.G.; Clausen, H.; Nilsson, T., Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus. *J Cell Sci,* **1998**, *111 (Pt 1)*, 45-60.

[12] 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.

[13] Ten Hagen, K.G.; Fritz, T.A.; Tabak, L.A., All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology,* **2003**, *13*, (1), 1r-16r.
[14] Gerken, T.A.; Revoredo, L.; Thome, J.J.; Tabak, L.A.; Vester-Christensen, M.B.; Clausen, H.; Gahlay, G.K.; Jarvis, D.L.; Johnson, R.W.; Moniz, H.A.; Moremen, K., The lectin domain of the polypeptide GalNAc transferase family of glycosyltransferases (ppGalNAc Ts) acts as a switch directing glycopeptide substrate glycosylation in an N- or C-terminal direction, further controlling mucin type O-glycosylation. *J Biol Chem,* **2013**, *288*, (27), 19900-19914.
[15] O'Connell, B.C.; Hagen, F.K.; Tabak, L.A., The influence of flanking sequence on the O-glycosylation of threonine in vitro. *J Biol Chem,* **1992**, *267*, (35), 25010-25018.
[16] Gerken, T.A.; Tep, C.; Rarick, J., 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,* **2004**, *43*, (30), 9888-9900.
[17] 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.
[18] Cheng, L.; Tachibana, K.; Zhang, Y.; Guo, J.; Kahori Tachibana, K.; Kameyama, A.; Wang, H.; Hiruma, T.; Iwasaki, H.; Togayachi, A.; Kudo, T.; Narimatsu, H., Characterization of a novel human UDP-GalNAc transferase, pp-GalNAc-T10. *FEBS Lett,* **2002**, *531*, (2), 115-121.
[19] Kubota, T.; Shiba, T.; Sugioka, S.; Furukawa, S.; Sawaki, H.; Kato, R.; Wakatsuki, S.; Narimatsu, H., Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J Mol Biol,* **2006**, *359*, (3), 708-727.
[20] Bennett, E.P.; Weghuis, D.O.; Merkx, G.; van Kessel, A.G.; Eiberg, H.; Clausen, H., Genomic organization and chromosomal localization of three members of the UDP-N-acetylgalactosamine: polypeptide N-acetylgalactosaminyltransferase family. *Glycobiology,* **1998**, *8*, (6), 547-555.
[21] Park, J.H.; Nishidate, T.; Kijima, K.; Ohashi, T.; Takegawa, K.; Fujikane, T.; Hirata, K.; Nakamura, Y.; Katagiri, T., Critical roles of mucin 1 glycosylation by transactivated polypeptide N-acetylgalactosaminyltransferase 6 in mammary carcinogenesis. *Cancer Res,* **2010**, *70*, (7), 2759-2769.
[22] Andergassen, U.; Liesche, F.; Kolbl, A.C.; Ilmer, M.; Hutter, S.; Friese, K.; Jeschke, U., Glycosyltransferases as Markers for Early Tumorigenesis. *Biomed Res Int,* **2015**, *2015*, 792672.
[23] Tarhan, Y.E.; Kato, T.; Jang, M.; Haga, Y.; Ueda, K.; Nakamura, Y.; Park, J.H., Morphological Changes, Cadherin Switching, and Growth Suppression in Pancreatic Cancer by GALNT6 Knockdown. *Neoplasia,* **2016**, *18*, (5), 265-272.
[24] Liesche, F.; Kolbl, A.C.; Ilmer, M.; Hutter, S.; Jeschke, U.; Andergassen, U., Role of N-acetylgalactosaminyltransferase 6 in early tumorigenesis and formation of metastasis. *Molecular medicine reports,* **2016**, *13*, (5), 4309-4314.
[25] Smith, R.D.; Lupashin, V.V., Role of the conserved oligomeric Golgi (COG) complex in protein glycosylation. *Carbohydr Res,* **2008**, *343*, (12), 2024-2031.

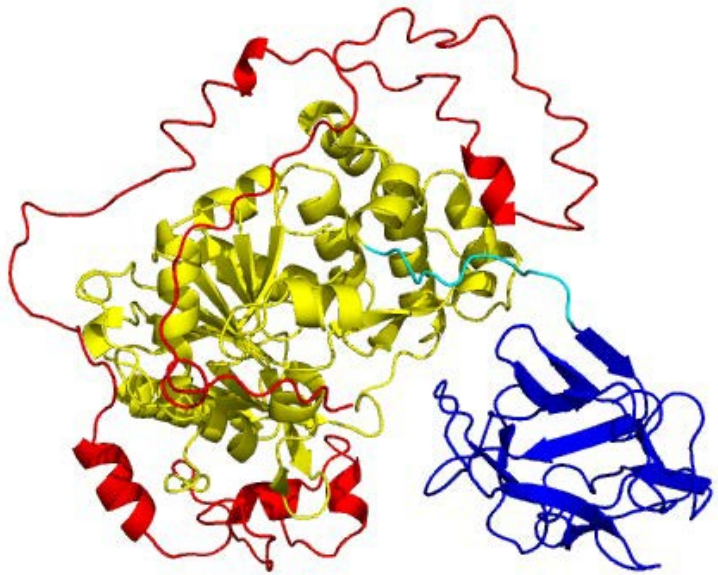
[26] Raman, J.; Fritz, T.A.; Gerken, T.A.; Jamison, O.; Live, D.; Liu, M.; Tabak, L.A., The catalytic and lectin domains of UDP-GalNAc:polypeptide alpha-N-Acetylgalactosaminyltransferase function in concert to direct glycosylation site selection. *J Biol Chem,* **2008**, *283*, (34), 22942-22951.
[27] Nehrke, K.; Ten Hagen, K.G.; Hagen, F.K.; Tabak, L.A., Charge distribution of flanking amino acids inhibits O-glycosylation of several single-site acceptors in vivo. *Glycobiology,* **1997**, *7*, (8), 1053-1060.
[28] Gerken, T.A.; Jamison, O.; Perrine, C.L.; Collette, J.C.; Moinova, H.; Ravi, L.; Markowitz, S.D.; Shen, W.; Patel, H.; Tabak, L.A., Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J Biol Chem,* **2011**, *286*, (16), 14493-14507.
[29] Gerken, T.A.; Ten Hagen, K.G.; Jamison, O., Conservation of peptide acceptor preferences between *Drosophila* and mammalian polypeptide-GalNAc transferase ortholog pairs. *Glycobiology,* **2008**, *18*, (11), 861-870.
[30] Milac, A.L.; Buchete, N.V.; Fritz, T.A.; Hummer, G.; Tabak, L.A., Substrate-induced conformational changes and dynamics of UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferase-2. *J Mol Biol,* **2007**, *373*, (2), 439-451.
[31] Bennett, E.P.; Mandel, U.; Clausen, H.; Gerken, T.A.; Fritz, T.A.; Tabak, L.A., Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology,* **2012**, *22*, (6), 736-756.
[32] Gomez, H.; Rojas, R.; Patel, D.; Tabak, L.A.; Lluch, J.M.; Masgrau, L., A computational and experimental study of O-glycosylation. Catalysis by human UDP-GalNAc polypeptide:GalNAc transferase-T2. *Org Biomol Chem,* **2014**, *12*, (17), 2645-2655.
[33] Brooks, S.A.; Carter, T.M.; Bennett, E.P.; Clausen, H.; Mandel, U., Immunolocalisation of members of the polypeptide N-acetylgalactosaminyl transferase (ppGalNAc-T) family is consistent with biologically relevant altered cell surface glycosylation in breast cancer. *Acta histochemica,* **2007**, *109*, (4), 273-284.
[34] Berois, N.; Mazal, D.; Ubillos, L.; Trajtenberg, F.; Nicolas, A.; Sastre-Garau, X.; Magdelenat, H.; Osinaga, E., UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 as a new immunohistochemical breast cancer marker. *J Histochem Cytochem,* **2006**, *54*, (3), 317-328.
[35] Devine, P.L.; McKenzie, I.F., Mucins: structure, function, and associations with malignancy. *Bioessays,* **1992**, *14*, (9), 619-625.
[36] Strous, G.J.; Dekker, J., Mucin-type glycoproteins. *Critical reviews in biochemistry and molecular biology,* **1992**, *27*, (1-2), 57-92.
[37] Carraway, K.L., 3rd; Funes, M.; Workman, H.C.; Sweeney, C., Contribution of membrane mucins to tumor progression through modulation of cellular growth signaling pathways. *Current topics in developmental biology,* **2007**, *78*, 1-22.
[38] Yuan, Z.; Wong, S.; Borrelli, A.; Chung, M.A., Down-regulation of MUC1 in cancer cells inhibits cell migration by promoting E-cadherin/catenin complex formation. *Biochem Biophys Res Commun,* **2007**, *362*, (3), 740-746.
[39] Suzuki, Y.; Sutoh, M.; Hatakeyama, S.; Mori, K.; Yamamoto, H.; Koie, T.; Saitoh, H.; Yamaya, K.; Funyu, T.; Habuchi, T.; Arai, Y.; Fukuda, M.; Ohyama, C.; Tsuboi, S., MUC1 carrying core 2 O-glycans functions as a molecular shield

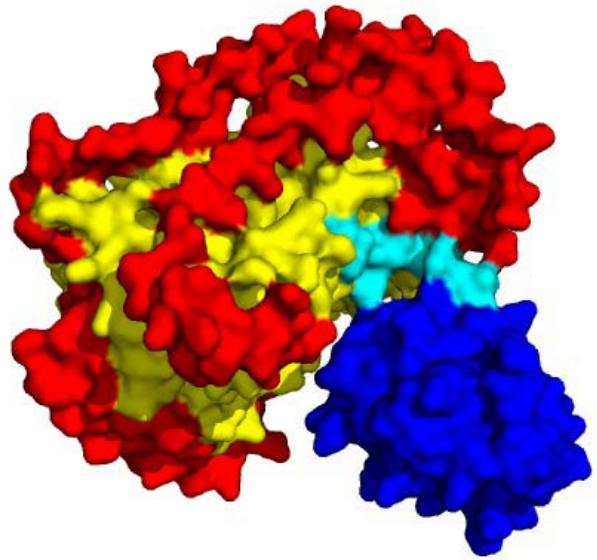
against NK cell attack, promoting bladder tumor metastasis. *Int J Oncol,* **2012**, *40*, (6), 1831-1838.
[40] Lumachi, F.; Basso, S.M., Serum tumor markers in patients with breast cancer. *Expert Rev Anticancer Ther,* **2004**, *4*, (5), 921-931.
[41] Duffy, M.J.; Duggan, C.; Keane, R.; Hill, A.D.; McDermott, E.; Crown, J.; O'Higgins, N., High preoperative CA 15-3 concentrations predict adverse outcome in node-negative and node-positive breast cancer: study of 600 patients with histologically confirmed breast cancer. *Clin Chem,* **2004**, *50*, (3), 559-563.
[42] Tsuboi, S.; Hatakeyama, S.; Ohyama, C.; Fukuda, M., Two opposing roles of O-glycans in tumor metastasis. *Trends Mol Med,* **2012**, *18*, (4), 224-232.
[43] Pinho, S.S.; Osorio, H.; Nita-Lazar, M.; Gomes, J.; Lopes, C.; Gartner, F.; Reis, C.A., Role of E-cadherin N-glycosylation profile in a mammary tumor model. *Biochem Biophys Res Commun,* **2009**, *379*, (4), 1091-1096.
[44] Freire, T.; Berois, N.; Sonora, C.; Varangot, M.; Barrios, E.; Osinaga, E., UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (ppGalNAc-T6) mRNA as a potential new marker for detection of bone marrow-disseminated breast cancer cells. *International journal of cancer,* **2006**, *119*, (6), 1383-1388.
[45] Cheong, S.C.; Chandramouli, G.V.; Saleh, A.; Zain, R.B.; Lau, S.H.; Sivakumaren, S.; Pathmanathan, R.; Prime, S.S.; Teo, S.H.; Patel, V.; Gutkind, J.S., Gene expression in human oral squamous cell carcinoma is influenced by risk factor exposure. *Oral oncology,* **2009**, *45*, (8), 712-719.
[46] Warner, G.C.; Reis, P.P.; Jurisica, I.; Sultan, M.; Arora, S.; Macmillan, C.; Makitie, A.A.; Grenman, R.; Reid, N.; Sukhai, M.; Freeman, J.; Gullane, P.; Irish, J.; Kamel-Reid, S., Molecular classification of oral cancer by cDNA microarrays identifies overexpressed genes correlated with nodal metastasis. *International journal of cancer,* **2004**, *110*, (6), 857-868.
[47] Shen, L.I.; Liu, L.; Yang, Z.; Jiang, N., Identification of genes and signaling pathways associated with squamous cell carcinoma by bioinformatics analysis. *Oncology letters,* **2016**, *11*, (2), 1382-1390.
[48] Mandel, U.; Hassan, H.; Therkildsen, M.H.; Rygaard, J.; Jakobsen, M.H.; Juhl, B.R.; Dabelsteen, E.; Clausen, H., Expression of polypeptide GalNAc-transferases in stratified epithelia and squamous cell carcinomas: immunohistological evaluation using monoclonal antibodies to three members of the GalNAc-transferase family. *Glycobiology,* **1999**, *9*, (1), 43-52.
[49] Wandall, H.H.; Dabelsteen, S.; Sorensen, J.A.; Krogdahl, A.; Mandel, U.; Dabelsteen, E., Molecular basis for the presence of glycosylated onco-foetal fibronectin in oral carcinomas: the production of glycosylated onco-foetal fibronectin by carcinoma cells. *Oral oncology,* **2007**, *43*, (3), 301-309.
[50] Matsuura, H.; Hakomori, S., The oncofetal domain of fibronectin defined by monoclonal antibody FDC-6: its presence in fibronectins from fetal and tumor tissues and its absence in those from normal adult tissues and plasma. *Proc Natl Acad Sci U S A,* **1985**, *82*, (19), 6517-6521.
[51] Burchell, J.M.; Mungul, A.; Taylor-Papadimitriou, J., O-linked glycosylation in the mammary gland: changes that occur during malignancy. *Journal of mammary gland biology and neoplasia,* **2001**, *6*, (3), 355-364.
[52] Brockhausen, I., Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO Rep,* **2006**, *7*, (6), 599-604.

[53] Chia, J.; Goh, G.; Bard, F., Short O-GalNAc glycans: regulation and role in tumor development and clinical perspectives. *Biochim Biophys Acta,* **2016**, *1860*, (8), 1623-1639.
[54] Park, J.H.; Katagiri, T.; Chung, S.; Kijima, K.; Nakamura, Y., Polypeptide N-acetylgalactosaminyltransferase 6 disrupts mammary acinar morphogenesis through O-glycosylation of fibronectin. *Neoplasia,* **2011**, *13*, (4), 320-326.
[55] Li, Z.; Yamada, S.; Inenaga, S.; Imamura, T.; Wu, Y.; Wang, K.Y.; Shimajiri, S.; Nakano, R.; Izumi, H.; Kohno, K.; Sasaguri, Y., Polypeptide N-acetylgalactosaminyltransferase 6 expression in pancreatic cancer is an independent prognostic factor indicating better overall survival. *Br J Cancer,* **2011**, *104*, (12), 1882-1889.
[56] Reis, C.A.; Osorio, H.; Silva, L.; Gomes, C.; David, L., Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol,* **2010**, *63*, (4), 322-329.
[57] Hakomori, S., Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer Res,* **1996**, *56*, (23), 5309-5318.
[58] Springer, G.F., T and Tn, general carcinoma autoantigens. *Science,* **1984**, *224*, (4654), 1198-1206.
[59] Tuccillo, F.M.; de Laurentiis, A.; Palmieri, C.; Fiume, G.; Bonelli, P.; Borrelli, A.; Tassone, P.; Scala, I.; Buonaguro, F.M.; Quinto, I.; Scala, G., Aberrant glycosylation as biomarker for cancer: focus on CD43. *Biomed Res Int,* **2014**, *2014*, 742831.
[60] Keding, S.J.; Danishefsky, S.J., Prospects for total synthesis: a vision for a totally synthetic vaccine targeting epithelial tumors. *Proc Natl Acad Sci U S A,* **2004**, *101*, (33), 11937-11942.
[61] Slovin, S.F.; Keding, S.J.; Ragupathi, G., Carbohydrate vaccines as immunotherapy for cancer. *Immunology and cell biology,* **2005**, *83*, (4), 418-428.
[62] Lo-Man, R.; Vichier-Guerre, S.; Perraut, R.; Deriaud, E.; Huteau, V.; BenMohamed, L.; Diop, O.M.; Livingston, P.O.; Bay, S.; Leclerc, C., A fully synthetic therapeutic vaccine candidate targeting carcinoma-associated Tn carbohydrate antigen induces tumor-specific antibodies in nonhuman primates. *Cancer Res,* **2004**, *64*, (14), 4987-4994.
[63] Iwasaki, H.; Zhang, Y.; Tachibana, K.; Gotoh, M.; Kikuchi, N.; Kwon, Y.D.; Togayachi, A.; Kudo, T.; Kubota, T.; Narimatsu, H., Initiation of O-glycan synthesis in IgA1 hinge region is determined by a single enzyme, UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2. *J Biol Chem,* **2003**, *278*, (8), 5613-5621.
[64] 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.; Clausen, H., 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.
[65] Bennett, E.P.; Hassan, H.; Hollingsworth, M.A.; Clausen, H., A novel human UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase, GalNAc-T7, with specificity for partial GalNAc-glycosylated acceptor substrates. *FEBS Lett,* **1999**, *460*, (2), 226-230.

[66] Hang, H.C.; Yu, C.; Ten Hagen, K.G.; Tian, E.; Winans, K.A.; Tabak, L.A.; Bertozzi, C.R., Small molecule inhibitors of mucin-type O-linked glycosylation from a uridine-based library. *Chem Biol,* **2004**, *11*, (3), 337-345.
[67] Schjoldager, K.T.; Vakhrushev, S.Y.; Kong, Y.; Steentoft, C.; Nudelman, A.S.; Pedersen, N.B.; Wandall, H.H.; Mandel, U.; Bennett, E.P.; Levery, S.B.; Clausen, H., Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc Natl Acad Sci U S A,* **2012**, *109*, (25), 9893-9898.
[68] Takeuchi, M.; Yoshikawa, M.; Sasaki, R.; Chiba, H., Purification and Characterization of UDP-N-Acetylgalactosamine: κ-Casein Polypeptide N-Acetylgalactosaminyltransferase from Mammary Gland of Lactating Cow. *Agricultural and Biological Chemistry,* **1985**, *49*, (4), 1059-1069.
[69] Sugiura, M.; Kawasaki, T.; Yamashina, I., Purification and characterization of UDP-GalNAc:polypeptide N-acetylgalactosamine transferase from an ascites hepatoma, AH 66. *J Biol Chem,* **1982**, *257*, (16), 9501-9507.
[70] Wragg, S.; Hagen, F.K.; Tabak, L.A., Kinetic analysis of a recombinant UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase. *J Biol Chem,* **1995**, *270*, (28), 16947-16954.
[71] Wandall, H.H.; Irazoqui, F.; Tarp, M.A.; Bennett, E.P.; Mandel, U.; Takeuchi, H.; Kato, K.; Irimura, T.; Suryanarayanan, G.; Hollingsworth, M.A.; Clausen, H., 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.
[72] Hatanaka, K.; Slama, J.T.; Elbein, A.D., Synthesis of a new inhibitor of the UDP-GalNAc:polypeptide galactosaminyl transferase. *Biochem Biophys Res Commun,* **1991**, *175*, (2), 668-672.
[73] Schafer, A.; Thiem, J., Synthesis of novel donor mimetics of UDP-Gal, UDP-GlcNAc, and UDP-GalNAc as potential transferase inhibitors. *J Org Chem,* **2000**, *65*, (1), 24-29.
[74] Tian, E.; Ten Hagen, K.G.; Shum, L.; Hang, H.C.; Imbert, Y.; Young, W.W., Jr.; Bertozzi, C.R.; Tabak, L.A., An inhibitor of O-glycosylation induces apoptosis in NIH3T3 cells and developing mouse embryonic mandibular tissues. *J Biol Chem,* **2004**, *279*, (48), 50382-50390.
[75] Timson, D.J., Metabolic Enzymes of Helminth Parasites: Potential as Drug Targets. *Curr Protein Pept Sci,* **2016**, *17*, (3), 280-295.
[76] Wang, H.; Liu, Y.; Huai, Q.; Cai, J.; Zoraghi, R.; Francis, S.H.; Corbin, J.D.; Robinson, H.; Xin, Z.; Lin, G.; Ke, H., Multiple conformations of phosphodiesterase-5: implications for enzyme function and drug development. *J Biol Chem,* **2006**, *281*, (30), 21469-21479.
[77] Sneath, P.H.; Sokal, R.R. *Numerical taxonomy. The principles and practice of numerical classification*, **1973**.
[78] Felsenstein, J., Confidence limits on phylogenies: an approach using the bootstrap. *Evolution,* **1985**, 783-791.
[79] Zuckerkandl, E.; Pauling, L., Evolutionary divergence and convergence in proteins. *Evolving genes and proteins,* **1965**, *97*, 97-166.
[80] Chenna, R.; Sugawara, H.; Koike, T.; Lopez, R.; Gibson, T.J.; Higgins, D.G.; Thompson, J.D., Multiple sequence alignment with the Clustal series of programs. *Nucleic acids research,* **2003**, *31*, (13), 3497-3500.

[81] Kumar, S.; Nei, M.; Dudley, J.; Tamura, K., MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics,* **2008**, *9*, (4), 299-306.
[82] Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S., MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution,* **2011**, *28*, (10), 2731-2739.
[83] Kumar, S.; Stecher, G.; Tamura, K., MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol,* **2016**, *33*, (7), 1870-1874.
[84] Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E., The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols,* **2015**, *10*, (6), 845-858.
[85] Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K., Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins,* **2009**, *77 Suppl 9*, 114-122.





pp-GalNAc-T6 is over-expressed in cancer cells, but under-explored in terms of its biochemistry. Current evidence suggests that selective inhibition of this enzyme may result in an effective, novel anti-cancer and anti-metastasis treatment.