Engineering Mammalian Mucin-Type O-Glycosylation in Plants

Zhang Yang1, Damian P. Drew2,*, Bodil Jørgensen3, Ulla Mandel1, Søren S. Bach2, Peter Ulyskov2, Steven B. Levery4, Eric P. Bennett4, Henrik Clausen4, Bent L. Petersen2*

1 Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, Flakkebjerg, 4200 Slagelse, Denmark
2 Department of Plant Biology and Biotechnology, Faculty of Life Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark
3 Department of Agriculture and Ecology, Faculty of Life Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark
4 Center for Glycomics, Department of Cellular and Molecular Medicine and School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen N, Denmark

Running title: Mucin-Type O-Glycosylation in Plants

* To whom correspondence should be addressed: Department of Plant Biology and Biotechnology, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark. Tel.: 45-35333379; Fax: 45-35333300. E-mail: blp@life.ku.dk

1Current address: Wine Science and Business, School of Agriculture, Food and Wine, University of Adelaide, Wine Innovation Central Bld, Urrbrae, SA 5064 Australia

Background: Plants lack mammalian GalNAc-type protein O-glycosylation. Results: Transient expression of a Glc(NAc) C4-epimerase and a polypeptide GalNAc-transferase in Nicotiana benthamiana resulted in O-glycosylation. Conclusion: Mammalian O-glycosylation can be established in plants. Significance: Plants may serve as host cells for recombinant production of custom designed O-glycoproteins.

Mucin-type O-glycosylation is an important posttranslational modification (PTM) that confers a variety of biological properties and functions to proteins. This PTM has a particularly complex and differentially regulated biosynthesis rendering prediction and control of where O-glycans are attached to proteins, and which structures are formed, difficult. Since plants are devoid of GalNAc type O-glycosylation, we have assessed requirements for establishing human GalNAc O-glycosylation de novo in plants with the aim of developing cell systems with custom designed O-glycosylation capacity. Transient expression of a Pseudomonas aeruginosa Glc(NAc) C4-epimerase and a human polypeptide GalNAc-transferase in leaves of Nicotiana benthamiana resulted in GalNAc O-glycosylation of co-expressed human O-glycoprotein substrates. A chimeric YFP construct containing 3.5 tandem repeat sequence of MUC1, was glycosylated with up to 3 and 5 GalNAc residues when co-expressed with GalNAc-T2 and a combination of GalNAc-T2 and GalNAc-T4, respectively, as determined by mass spectrometry. O-glycosylation was furthermore demonstrated on a tandem repeat of MUC16 and interferon α 2b. In plants, prolines in certain classes of proteins are hydroxylated and further substituted with plant specific O-glycosylation; unsubstituted hydroxyprolines were identified in our MUC1 construct. In summary, this study demonstrates that mammalian type O-glycosylation can be established in plants and that plants may serve as a host cell for production of recombinant O-glycoproteins with custom designed O-glycosylation. The observed hydroxyproline modifications, however, call for additional future engineering efforts.

Most recombinant produced biological therapeutics are glycoproteins, and selection of host cells for their production is critical for therapeutic effects and safety because the glycan moieties can interact with lectin scavenger receptors and immune cells. Traditionally, selection of host cells has aimed at producing glycoproteins with normal human
glycosylation, i.e. mature glycans similar to those found on the natural glycoproteins, typically complex type N-glycans and core 1 O-glycans, all with sialic acid capping (1). However, aberrant glycosylation may be desirable, e.g. with respect to design of vaccines aimed at eliciting immunity to specific glycoforms found on virus particles and virus infected cells or on cancer cells (2). We have previously shown that the cancer-associated mucin MUC1 contains immunodominant aberrant O-glycopeptide epitopes to which cancer-specific IgG antibodies can be elicited in man (3), and we have developed a chemoenzymatic strategy for synthesis of such vaccine glycopeptides (3). However, sustainable recombinant expression systems for production of vaccines targeting cancer cells with immature O-glycosylation or envelope virus glycoproteins are still in demand.

O-glycosylation is controlled by a family of up to 20 UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases (GalNAc-Ts); the repertoire of GalNAc-Ts present in a given cell defines the pattern and density with which proteins are decorated with O-glycans. Normal mammalian cells typically produce elongated and/or branched O-glycans capped with sialic acids or different blood group related structures. The most frequently used host cell for recombinant expression of therapeutic proteins is CHO, which generally produces mono- and disialylated core 1 (Galβ1-3GalNAca1-O-Ser/Thr) O-glycan structures. Recently, it was shown that CHO-K1 expresses four GalNAc-Ts (GalNAc-T2, -T7, -T11, and -T19) (4). Because all eumetazoan cells express multiple GalNAc-Ts and generally extend and cap O-glycans, it is desirable to develop a cell system for recombinant production in which the entire O-glycosylation pathway can be custom-built both with respect to the GalNAc-T repertoire for custom design of sites of O-glycosylation, as well as the repertoire of glycosyltransferases involved in extending and capping O-glycans.

Yeast and plants offer cell systems suitable for recombinant expression without the capacity for mammalian GalNAc O-glycosylation (5,6). Neither species contain genes homologous to the GalNAc-T gene family that initiates O-glycosylation (7). While GalNAc and the donor substrate UDP-GalNAc are not found in yeast, the presence of these compounds in plants is still disputed (5,8-10). Most of the enzymes known to elongate O-glycans in mammals have no close homologs in plants (11-13). These features should make it possible to custom design and build a capacity for O-glycosylation from the bottom-up in plants. In yeast the endogenous protein O-mannosylation initiated in the ER is likely to compete for substrate sites with the Golgi-localized GalNAc O-glycosylation. In spite of this Narimatsu and colleagues (14), introduced O-GalNAc glycosylation into yeast and obtained recombinant production of reporter constructs with GalNAc and core1 O-glycosylation, but also a degree of O-mannosylation. With the use of a rhodanine-3-acetic acid derivative added to the culture medium O-mannosylation was reduced, but not completely eliminated.

Plants on the other hand do not have competing O-glycosylation pathways for Ser/Thr residues. Plants produce another type of protein O-glycosylation, whereby prolines are hydroxylated to yield hydroxyproline (Hyp), which may be substituted with various Hyp O-glycans (reviewed in Lamport et al. (2011)) (15). Hyp conversion occurs in the ER and extends into the Golgi apparatus (16,17) and this modification is preferentially found at proline (Pro)-rich repetitive sequence motifs (15). Whereas the short Pro rich hinge region of IgA1 in humans is substituted with mucin-type O-linked glycans, recombinant expression in plants results in several Hyp modifications and further substitution with short arabinosides (18). More recently, expression of a tandem repeat of the human mucin MUC1, which in humans is decorated with O-glycans, was also found to contain Hyp O-glycosylation (19). The key step in plant specific Hyp O-glycosylation is regulated by thirteen C4-hydroxylases (P4Hs), which individually, according to initial studies (17,20-22), are not required for viability and growth, and in principle one or more of these could be mutated to eliminate undesirable modifications on recombinant glycoproteins. It is therefore conceivable that plants can be engineered to eliminate specific unwanted Hyp modifications that are required by proteins of interest during recombinant expression. Plants should therefore offer a suitable and safe system for recombinant production of glycoprotein pharmaceuticals.

In a recent study Daskalova et al. (23) studied transient expression of a GalNAc-T in combination with a Glc(NAc) C4-epimerase and a UDP-Gal(NAc) transporter in Nicotiana benthamiana L. plants. They used a chimeric reporter construct with a short sequence from the human mucin, MUC1, as acceptor for O-glycosylation, and used questionable VVA lectin blotting results to conclude that GalNAc O-glycosylation could be established in
plants with co-expression of GalNAc-T2 alone. They also concluded and that co-expression of both an epimerase and a transporter were required for enhancing lectin reactivity with the reporter, while neither the epimerase nor the transporter alone improved lectin reactivity. The results were not corroborated by structural analysis, and it is known that lectin reactivity poses particular problems in plants (5,24). Thus, the original proposal of presence of sialic acids in plant glycoproteins (25,26), was later shown likely to be due to contaminants (27). Furthermore, Daskalova and colleagues more recently reported that the VVA lectin recognize non-carbohydrate epitopes by Western blotting in plants (28) casting serious doubts on the conclusions drawn in their previous publication (23).

In this report, we have carefully tested different design strategies for the introduction of GalNAc mucin-type O-glycosylation into N. benthamiana plants. The results are essentially in complete disagreement with those reported by Daskalova et al (23), which highlights the need for structural verification of products produced with glycoengineering. We used transient expression of enzymes and reporter substrates to identify requirements for establishing efficient glycosylation of mammalian O-glycoproteins (See Fig. 1). We found no evidence of GalNAc O-glycosylation when a reporter substrate was expressed in plants alone or with an appropriate GalNAc-T, GalNAc-T2. Basic O-glycosylation capacity was, however, achieved by introduction of a Glc(NAc) C4-epimerase and a GalNAc-T, which was validated through co-expression of three O-glycosylation target proteins. We did observe Hyp modifications in a MUC1 based substrate, which indicates that use of plants for production of O-glycoprotein therapeutics will require additional strategies to eliminate the corresponding endogenous proline hydroxylases.

**EXPERIMENTAL PROCEDURES**

Inoculation and Growth Conditions of N. benthamiana - Growth conditions were as described in Egelund et al. (29). Agrobacterium tumefaciens strain C58C1 pGV3850 was used for agrobacterium mediated expression in N. benthamiana which was performed essentially as described by Sainsbury and Lomonossoff (30). In brief, agrobacteria were transformed with constructs by electroporation and selected with the appropriate antibiotics. Agrobacteria cultures were grown overnight in Luria-Bertani (LB) medium, harvested by centrifugation and resuspended in a buffer containing 10 mM MES, 10 mM MgCl2, and 100 µM acetylsyringone (OD600 = 0.5) and left at 20°C for 2 h. Leaves of 3-4 week old N. benthamiana plants were infiltrated with the bacterial cell suspensions using 1-mL syringes and leaf material was collected for analysis after 5-6 days. To increase expression of transgenes, all experiments included co-infiltration of agrobacteria carrying a p19 gene-construct, a viral protein specifically inhibiting posttranscriptional gene silencing (31).

**DNA Constructs for Plant Transformation and Transient Expression** - Vector constructs used are depicted in Fig. 2A. Open source vectors used for Agrobacterium mediated expression and transformation are: pBI121 (GenBank accession number AY781296); pCAMBIA 2300 (GenBank accession number AF234315); pCAMBIA 1302 (GenBank accession number AF234298). For legacy of open source pCAMBIA binary vectors see http://www.cambia.org. pPS48 (32) is an intermediate E. coli only vector, which contains the cauliflower mosaic virus 35S promoter followed by the 35S terminator interspaced by a multiple cloning site (MCS), into which genes of interest (goi) were cloned. Entire transcriptional units (35S-Pro-goi-35S-term) were then excised using XbaI or HindIII and ligated into the MCS of the pCAMBIA-derived plant expression plasmids. The Nicotiana tabacum Ubiquitin promoter and terminator were synthesized by GenScript, USA. O-glycosylation target constructs were directed to the secretory pathway by N-terminal fusion of nucleotides encoding one of the following signal peptide sequences: 1) NtSP derived from Nicotiana tabacum proline-rich protein 3 (UniProt accession number Q40502), encoding the sequence: MGKMASLFASLLVVLVVSLA. 2) PpSP derived from Physcomitrella patens aspartic protease (EMBL accession number AJ586914) encoding the sequence: MGASRSVLAFFLVLVLAALAEA.

The following constructs expressing mucins and other O-glycosylation target peptides were assembled or synthesized synthetically. The MUC1-3.5 tandem repeat (TR) (33) codon optimized for expression in Arabidopsis thaliana, was synthesized with an N-terminal PpSP sequence (EUROFINS, MWG, Germany) and then utilized as template for PCR amplification using the primers SPMUC1F or and MUC1Rev (sequence of all primers are listed in Table S1). The resulting PCR product was then N-
O-glycosylation machinery consisted of cytosolic Pseudomonas aeruginosa Glc(NAc) C4-epimerase (Epi) WbpP (35) (GenBank accession number AAF23998.1), human GalNAc-T2 (UniProt accession number X85019) and GalNAc-T4 (GenBank accession number NP_003765.2). O-glycosylation machineries were implemented as single constructs either with each gene driven by their own promoter or as a polycistronic transcript interspaced by the 2A auto cleavable sequence (FMDV) (36,37) under control of a single promoter driving expression of the resulting 2A linked fusion protein. Construction of O-glycosylation machineries encoding C4-epimerase and GalNAc-Ts, with each gene brought under control of a separate promoter, was carried out as follows: 1) Cloning of the N-terminal Flag-tagged (MDYKDDDD) epimerase into pCAMBIA2300 vector (pC2300) involved PCR-amplification using the primer set PF1-2 and Primer Rev, which was sub-cloned into the SacI site of construct T2-35SPro-MUC1-3.5TR-YFP-His, yielding pC2300D-35SPro-MUC1-3.5TR-YFP-His, resulting in the 2A linker region of T2-2A-CytoEpi, The 2A linker region of T2-2A-CytoEpi encodes QGSQTLNFDLKLADVESNPG↓PMD, where italics designate the 2A sequence and arrow ↓ being the auto-cleavage site (36,37). 2) The reverse order, encoding (from N- to C-terminus) cytosolic epimerase followed by the 2A sequence in pC130035Su in accordance to the USER cloning method delineated by Nour-Eldin et al. and Geu-Flores et al. (34,39), yielding pC1302-35SPro-T2-2A-CytoEpi-35STerm (T2-2A-CytoEpi), Two additional target peptides, under control of the ubiquitin promoter and terminator system were created: 1) pC2300D-UbiPro-NISP-INFα2B-GM-UbiTerm (INFα2B) encoding interferon α2B (GenBank accession number AE255838.1), INFα2B, with an N-terminal signal peptide NtSP and C-terminal glyco-module ((SP)10), T7 (MASMTGGQQMGG) and His6, tag. 2) pC2300D-UbiPro-NISP-MUC16-1.2TR-UbiTerm (MUC16) encoding MUC16-1.2TR (33) (GenBank accession number AF414442.2) with an N-terminal signal peptide NtSP and C-terminal T7 and His6, tag, was codon optimized for expression in Arabidopsis and synthesized by GenScript, USA. O-glycosylation machinery construct encoding C4-epimerase and GalNAc-T2 interspaced by the 2A auto-cleavage site (36,37). Three additional O-glycosylation machineries were made as 2A-linked polycistronic constructs: 1) GalNAc-T2 was PCR amplified with primer-introduced N-terminal hemagglutinin tag (HA-tag) using the primers HAT2For and HAT2Rev. N-terminal Flag-tagged epimerase was PCR-amplified using the primer set PFwbppFor and PFwbppRev. The two PCR fragments were then ligated and interspaced with a sequence encoding the 2A sequence in pC130035Su in accordance to the USER cloning method delineated by Nour-Eldin et al. and Geu-Flores et al. (34,39), yielding pC1302-35SPro-T2-2A-CytoEpi-35STerm (T2-2A-CytoEpi). Three additional O-glycosylation machineries were made as 2A-linked polycistronic constructs: 1) GalNAc-T2 was PCR amplified with primer-introduced N-terminal hemagglutinin tag (HA-tag) using the primers HAT2For and HAT2Rev. N-terminal Flag-tagged epimerase was PCR-amplified using the primer set PFwbppFor and PFwbppRev. The two PCR fragments were then ligated and interspaced with a sequence encoding the 2A sequence in pC130035Su in accordance to the USER cloning method delineated by Nour-Eldin et al. and Geu-Flores et al. (34,39), yielding pC1302-35SPro-T2-2A-CytoEpi-35STerm (T2-2A-CytoEpi).
2A sequence was PCR amplified using pBI121-35SPro-MUC1-3.5TR-YFP-2A-GolgiEpi-2A-T2-NosTerm as template and the primers PFlagFor and PT2Rev, yielding the fragment GolgiEpi-2A-T2, which was inserted back into pBI121 vector using the XbaI and ScaI sites, yielding pBI121-35SPro-GolgiEpi-2A-T2-NosTerm (GolgiEpi-2A-T2).

Preparation of Leaf Total Protein Extracts - Approximately one gram of freshly harvested leaves were frozen in liquid N₂ and comminuted using a pestle and mortar with 2 ml extraction buffer A (50 mM NaPO₄, 250 mM NaCl, 5 mM Imidazole, pH 8.0) containing Complete Proteinase Inhibitor (Roche) and 1 mM phenylmethanesulfonylfluoride. The sample was incubated for 10 min on ice and insoluble material was pelleted by centrifugation (20,000 x g) for 10 min, the supernatant was recovered and stored at -20°C.

SDS-PAGE Western Blotting - SDS-PAGE Western blot analysis was performed as previously described (40). Monoclonal antibodies (MAbs) to the T7 tag, and M11 to MUC16 were obtained from Invitrogen and Dako, respectively. MAbs to MUC1 and GalNAc-MUC1, SE10 and SE5, respectively, have been described previously (41). Lectin blot analysis was performed essentially as Western blot analysis, except the antibody was HRP-conjugated Vicia villosa lectin (VVA; EY Laboratories).

Ni-affinity Chromatography - Cleared supernatants (50 ml, 20,000 x g for 30 min) were incubated with 0.5 ml of Ni-NTA agarose beads (Qiagen) for 2 h at 4°C under gentle rolling. Beads were washed for 10 min with 20 ml of wash-buffer (50 mM NaPO₄, 250 mM NaCl, 20 mM Imidazole, pH 8.0), and eluted with 5 x 1 ml of elution-buffer (50 mM NaPO₄, 250 mM NaCl, 250 mM Imidazole, pH 8.0).

Endo-Asp Digestion of MUC-3.5TR and Sample Purification - Purified MUC1-3.5TR-YFP (~25 µg) was incubated with 1 µg endoproteinase Asp-N from Pseudomonas fragi (Roche) in 300 µL 100 mM Tris-HCl (pH 8.0) for 16 hours at 37°C with shaking. The digest was cleaned on a C18 Zip-Tip (Millipore). Briefly, the digestion mixture was dissolved in 20 µL 0.1% TFA and drawn through the column, desalted with 0.5% formic acid, and eluted with acetonitrile. In some instances the released 20-mer MUC1 tandem repeat peptide was further isolated by HPLC using a Dionex system. Trifluoroacetic acid (TFA) was added to samples (0.05% v/v) before application to a C12 column (150 x 4.6 mm Jupiter Proteo with 90Å pore size, 4 micron particle size, Phenomenex). Chromatographic separation was obtained in a two eluent system where eluent A was 0.05% TFA in water and eluent B was 0.05% TFA in acetonitrile, and the pump speed was a constant 0.5 mL min⁻¹. From 0-5 min, the eluent was 5% B, from 5-35 min, eluent B increased in a linear gradient to 40%, and from 35-45 min eluent B increased to 100%.

Matrix-assisted Laser Desorption Ionization Time-of-flight (MALDI-TOF) - Lyophilized peptides were dissolved in 20 µL water. The MALDI matrix was 25 g/L 2,5-dihydroxybenzoic acid (Sigma-Aldrich) dissolved in a 1:1 mixture of water and methanol. Samples were prepared for analysis by placing 0.5 µL of sample solution on a probe tip followed by 0.5 µL of matrix. All spectra were obtained in the linear mode and calibrated using external calibration. All mass spectra were acquired on a Voyager-Elite MALDI time of flight mass spectrometer (Perseptive Biosystem Inc., Framingham, MA), equipped with delayed extraction.

Characterization of O-glycosylation Sites by ETD-MS² - Products of O-glycosylation were characterized by electrospray ionization-linear ion trap-Fourier transform mass spectrometry (ESI-LIT-FT-MS) in an LTQ-Orbitrap XL hybrid spectrometer (Thermo-Scientific) equipped for electron transfer dissociation (ETD) for peptide sequence analysis by MS/MS (MS²) with retention of glycan site-specific fragments. Samples were dissolved in methanol-water (1:1) containing 1% formic acid and introduced by direct infusion via a TriVersa NanoMate ESI-Chip interface (Advion BioSystems) at a flow rate of ~100 nl/min and 1.4 kV spray voltage. Mass spectra were acquired in positive ion FT mode using parameters similar to previous studies (42), except at a nominal resolving power of either 30,000 or 60,000. MS¹ spectra, in which multiple charge states were observed, were deconvoluted for clarity of presentation using the Xtract function in the Xcalibur data analysis software package (Thermo Fisher). ETD-MS² spectra were analyzed by comparison with theoretical c- and z- fragment m/z values calculated for all positional combinations of HexNAc residues distributed on the all potential S and T glycosylation sites in the sequence. Where potential hydroxylation of Pro was observed, positional combinations including one or more Hyp residues were also calculated. Calculations were performed using the web-based Protein Prospector.
RESULTS

Design of GalNAc O-Glycosylation Engineering in Plants - Depicted in Fig. 1 are the principle constituents needed for GalNAc O-glycosylation. Introduction of UDP-GalNAc synthesis requires a Glc/GlcNAc C4-epimerase. High concentrations of nucleotide sugars in Golgi are maintained by transporters, and it has been suggested that plants lack a specific transporter (43,44) for UDP-GalNAc transporters, and it has been suggested that plants lack a specific transporter (43,44) for UDP-GalNAc transporters (43,45-47). We therefore first investigated whether introduction of an epimerase, Pseudomonas aeruginosa cytoplasmic epimerase (WbpP) (35), and a human GalNAc-T, GalNAc-T2 (48), would be sufficient for O-glycosylation. A series of construct designs were prepared and tested to address the optimal order of coding regions, efficiency of introducing a 2A self-cleaving sequence, and effect of targeting the epimerase to ER, Golgi or the cytoplasm (Fig. 2A). For expression and localization of O-glycosylation machinery components see Supplemental Fig. S1.

Expression of a MUC1-YFP Reporter Substrate in Plants - We initially assessed O-glycosylation capacity by expressing a known acceptor substrate for GalNAc-T2, which was based on the tandem repeat sequence of MUC1 and linked to YFP (MUC1-YFP). When expressed alone, the reporter MUC1-YFP substrate migrated as a ~45 kDa protein and reacted with MAb 5E10 (reactive with unglycosylated and all glycoforms of MUC1), but not with MAb 5E5 (reactive with GalNAc-glycosylated MUC1), indicating that the MUC1 reporter substrate was not substituted with GalNAc in wildtype plants (Fig. 2B, lane 1). This was subsequently confirmed by mass spectrometric analysis of the MUC1 peptide enzymatically digested from the expressed MUC1-YFP protein (Fig. 3D), where a product corresponding to unmodified MUC1 (m/z 1887) was present, but no products corresponding to glycosylated MUC1 were observed. This analysis did uncover evidence of Hyp modification of MUC1. The presence of a ~16 Dalton increment (m/z 1903) in the MS spectrum of MUC1-YFP corresponds to hydroxylation of a single proline. ESI-FT-ETD-MS analysis further demonstrated that proline P11 of the MUC1 tandem repeat (DTPAPGSTAPP11PAHGVTSAP) was hydroxylated (Yang Z. and Petersen B.L., unpublished results). Based on peak sizes this Hyp modification was a minor constituent and evidence of other modifications including Hyp O-glycosylation was not observed (Fig. 3D).

Construct Designs of O-Glycosylation Machinery - Co-expression of either C4-epimerase WbpP or GalNAc-T2 alone with the MUC1-YFP reporter substrate did not result in detectable O-glycosylation (Fig. 2B, lanes 2-3). In contrast co-expression of MUC1-YFP with a combination of C4-epimerase and GalNAc-T2 resulted in detectable O-glycosylation as evidenced by reactivity with MAb 5E5 and a slight mobility shift (1-2 kDa) on SDS-PAGE derived Western blotting (Fig. 2B, lanes 4-8). The 2A auto-cleaving motif is a 20 amino acid sequence derived from the foot-and-mouth virus, which through an intra-ribosomal skipping event during translation results in synthesis of two protein products (49). Insertion of the 2A sequence between two coding regions in a polyprotein construct allows for expression of two independent proteins under control of a single promoter, and it has been widely tested in many eukaryotic expression systems (49,50). The 2A construct design is an attractive approach for achieving coordinated expression of multiple proteins. Tests of different construct designs for expression of the epimerase and GalNAc-T2 demonstrated that only the 2A linked bicistronic constructs (T2-2A-CytoEpi and Cyto-Epi-2A-T2) produced efficient GalNAc O-glycosylation (Fig. 2B, lanes 4-5). The attempts to direct the epimerase to the Golgi (GolgiEpi-2A-T2) or to the ER (EREpi-T2) did not result in substantial glycosylation (Fig. 2B, lanes 6-7). Furthermore, expression of the cytoplasmic epimerase and GalNAc-T2 from separate promoters resulted in detectable, but substantially lower efficiency in glycosylation (Fig. 2B, lane 8) indicating that the 2A sequence is essential for the vector design.

Structural Analysis of MUC1-YFP Expressed in Glycoengineered N. benthamiana - MUC1-YFP co-expressed with or without T2-2A-CytoEpi was purified by metal-affinity chromatography and analysed via SDS-PAGE and VVA lectin blotting (Fig. 3A). The increase in molecular mass exhibited by MUC1-YFP when co-expressed with T2-2A-CytoEpi was readily apparent on a highly separated SDS-PAGE gel (Fig. 3A, left panel). HRP-conjugated VVA lectin, specific for GalNAc moieties, reacted with MUC1-YFP co-expressed with T2-2A-Epi, but not with MUC1-YFP expressed alone (Fig. 3A, right panel). Purified
MUC1-YFP was enzymatically digested with endo-Asp-N to release individual MUC1 tandem repeats (Fig. 3B). Analysis of the products of this digest by HPLC enabled the separation of differentially modified MUC1 tandem repeats (Fig. 3C). The released MUC1 expressed in plants without O-glycosylation machinery eluted at 22.9 min and MS analysis of the isolated product confirmed this (Yang Z. and Petersen B.L., unpublished results). MS analysis of the total MUC1-YFP digest revealed the presence of unmodified MUC1, but no products corresponding to GalNAc modifications (Fig. 3D). The HPLC chromatogram of digested MUC1-YFP from plants co-expressing the O-glycosylation machinery, T2-2A-CytoEpi, revealed additional products eluting between 20.5 and 22.3 min (Fig. 3C, lower panel). MS analysis of these products identified MUC1 tandem repeats with 1-3 GalNAc moieties (Fig. 3E). This HPLC and MS analysis revealed that a substantial amount of the tandem repeat sequence was modified by 1-3 moles GalNAc, which is in agreement with the fact that GalNAc-T2 can attach GalNAc up to three of the five potential sites in the MUC1 tandem repeat sequence (Ser5, Thr3, Thr15) (38). Two minor peaks at 21.1 and 22.3 min were found when MUC1-YFP was expressed alone (Fig. 3C, upper panel). MALDI-TOF analysis of these fractions showed the presence of MUC1-1TR with 1 Hyp (only in the 22.3 min fraction), YFP peptide fragments, and apparent contaminants (Yang Z. and Petersen B. L., unpublished results). Analysis of MUC1-YFP co-expressed with the CytoEpi-2A-T2 or GolgiEpi-2A-T2 O-glycosylation machinery constructs confirmed incorporation of GalNAc, albeit at lower levels than observed with the T2-2A-CytoEpi construct (Fig. 3F, G). Approximately 65% of total MUC1-YFP was glycosylated by the most efficient O-glycosylation machinery, T2-2A-CytoEpi, and MUC1-YFP was estimated to accumulate to approximately 12 µg/g (fresh weight) equivalent to ca. 1% of the total soluble protein.

Co-expression of GalNAc-T2 and -T4 Completes GalNAc O-Glycosylation of MUC1 - Complete O-glycosylation of all five potential sites of the MUC1 tandem repeat requires the coordinate action of GalNAc-T4 and GalNAc-T2 (38). We therefore co-expressed GalNAc-T4 (T4) with T2-2A-CytoEpi and MUC1-YFP, which resulted in a further minor shift in SDS-PAGE mobility of MUC1-YFP as evidenced by 5E5 MAb mediated Western blot analysis (Fig. 4A). MS analysis of purified MUC1-YFP digested with endo-Asp-N confirmed incorporation of up to 5 moles GalNAc per tandem repeat of MUC1 (m/z 2902) when co-expressed with GalNAc-T2 and -T4 (Fig. 4B). Again a minor amount of hydroxylation of P11 was observed (m/z 1903).

O-glycosylation of Human MUC16 and INFα2B - Two other human O-glycoproteins were tested as substrates. A construct expressing 1.2 tandem repeat of MUC1 (MUC1-T7), i.e. encoding 223 amino acids with ~30 putative O-glycosylation sites and 3 N-glycosylation sites, was co-expressed with the T2-2A-CytoEpi glycosylation machinery. A clear shift in SDS-PAGE mobility (~45kd to ~50kd) of the T7 tagged protein combined with reactivity with VVA was observed with co-expression of T2-2A-CytoEpi (Fig. 5A). Co-expression of a construct encoding the full secreted INFα2B cytokine tagged with T7 and an Arabinogalactan Protein (AGP) type (SP)10 glycomodule (INFα2B-T7-AGP) with the T2-2A-CytoEpi construct resulted in a mobility shift of a minor fraction of the protein (~30kd to ~31kd) and reactivity with VVA (Fig. 5B). INFα2B has a single O-glycosylation site (GVGVTETPLM), which is known to be O-glycosylated by GalNAc-T2 (51). The observed mobility shifts and VVA labelling are indicative of O-glycosylation, but further MS analysis is needed to confirm glycosylation sites.

DISCUSSION

Plants represent one eukaryotic cell system in which animal O-glycosylation can be built from scratch and custom designed for a particular purpose. Because the capacity for GalNAc-type O-glycosylation is highly regulated and dynamic in mammalian cells it is desirable to develop cell systems with defined capacity for O-glycosylation for production of recombinant therapeutics and vaccines. In this study, we established GalNAc O-glycosylation capacity in N. benthamiana plant cells by expression of a UDP-GlcNAc C4-epimerase and GalNAc-Ts, and showed that three different co-expressed secreted substrates were O-glycosylated. In the present study, we expressed GalNAc-T2 and epimerase transiently in plants using both monocistronic constructs and bicistronic 2A linked constructs to screen different permutations of optimal vector design for the glycosylation machinery. In the present transient implementations the bicistronic 2A construct designs proved most efficient (Fig. 2B).

The GolgiEpi-2A-T2 and EREpi-T2 constructs were designed to test if ectopic expression of the
C4-epimerase in the ER-Golgi lumen would convert ER-Golgi lumen UDP-GlcNAc to UDP-GalNAc \textit{in situ} and allow GalNAc O-glycosylation. However, targeting epimerase to cytoplasm was most effective as evidenced by glycosylation of the co-expressed acceptor substrate MUC1 (Fig. 2B). Factors including local substrate (UDP-GlcNAc) and resulting product (UDP-GalNAc) concentrations, the presence of chaperones and pH differences in various compartments may influence the glycosylation efficiency. Other factors may be inefficient 2A sequence cleavage and aberrant localization of the individual proteins. For example, de Felice and Ryan reported that a two times 2A construct encoding a Golgi targeted cyano fluorescence protein (CFP), Golgi targeted YFP and a cytosolic puromycin resistance protein (PAC) (i.e. GT-CFP-2A-GT-YFP-2A-PAC), resulted in Golgi localized CFP and YFP aberrantly localized in the mitochondria (52). Our attempts to exploit the ER/Golgi GlcNAc pools by targeting the epimerase to the ER or the Golgi proved less efficient than targeting it to the cytosol. This demonstrates that plant cells efficiently transports UDP-GalNAc into the secretory pathway and either do not produce GlcNAc and UDP-GalNAc or if so, only in minute amounts. The results provide a basis for further exploration of plant cells as glycoengineered hosts for recombinant expression of O-glycoproteins.

Considerable efforts have been devoted to glycoengineering host cells to accommodate human type glycosylation on recombinant expressed therapeutics, which have led to remarkable improvements especially for protein N-glycosylation. Human type N-glycosylation has been engineered in yeast with unprecedented control of homogeneity of N-glycan structures added to proteins (53). In contrast, engineering of the abundant GalNAc-type O-glycosylation has lagged behind. Yeast cells were originally engineered to enable formation of GalNAcea and Gal\(\beta\)1–3GalNAcea O-glycosylation (14). The engineered glycosylation machinery comprised a UDP-GlcNAc C4-epimerase, a UDP-Gal(NAc) transporter, and the glycosyltransferases GalNAc-T2 and C1GalT1. While yeast is a desirable expression system, it suffers from competing O-mannosylation initiated in the ER. O-mannosylation is essential for yeast (54), but O-mannosyl glycans are immunogenic in man (55,56). Whereas it is not possible to knock out all O-mannosyltransferases, some reduction of O-mannosylation was achieved by adding a rhodanine-3-acetic acid derivative to the culture medium (14).

While the present study was being completed, Daskalova et al. reported on similar glycoengineering attempts in \textit{N. benthamiana} (23). Based on VVA lectin labelling experiments the authors concluded that co-expression of a GalNAc-T with a reporter substrate was sufficient to produce low levels of glycosylation, whereas co-expression of both a UDP-Gal(NAc) transporter and a C4-epimerase were required for enhanced O-glycosylation. These results are in complete disagreement with the results presented here, where efficient glycosylation was achieved by co-expression of a GalNAc-T with an epimerase. We can only reconcile the different findings with erroneous interpretation of VVA lectin reactivity as was in fact reported by Daskalova and colleagues only recently (28). In this paper, Daskalova and co-workers presented evidence of non-carbohydrate specific binding of VVA lectin to tobacco protein extracts, which could not be inhibited by hapten sugar, such as D-GalNAc. Furthermore, the claim that a specific UDP-Gal(NAc) transporter is required for GalNAc O-glycosylation in plants (23) is controversial. Although UDP-GalNAc transport capacity has not been directly demonstrated in plants, many transporters are known to have UDP-Glc/Gal and UDP-GlcNAc/GalNAc transport capacities (57). Here we clearly demonstrate efficient O-glycosylation without exogenous introduction of a dedicated UDP-GalNAc transporter. This was also found with the additional functional expression of the GalNAc-T4 isoform, which is unique among the GalNAc-T isoforms by being required for complete GalNAc O-glycosylation of all five potential sites in the MUC1 tandem repeat sequence as well as having a very high Km for UDP-GalNAc (38).

Plants have a unique hydroxyproline linked O-glycosylation, which has been found on recombinant expressed human IgA1 (18) and most recently MUC1 (19). This modification is initiated by a family of prolyl 4-hydroxylases (P4Hs) and the resulting Hyps may subsequently be glycosylated with arabinogalactan or shorter arabinosides. P4Hs are membrane bound type II proteins found in the ER and extending into the Golgi apparatus (16). Thirteen putative P4Hs have been identified in Arabidopsis (58). In the present study a minor amount of hydroxylation of P\textsuperscript{11} of the MUC1 tandem repeat was observed. Other PTMs, including hydroxyproline linked O-glycosylation, were not encountered. The recent findings of Pinkhasov et al (2011), where either P\textsuperscript{11} or P\textsuperscript{12} of
the MUC1 tandem repeat was further substituted with 3 L-arabinofuranoses. This hydroxyproline linked O-glycosylation may be rationalized by the use of different promoter systems driving the transient expression of the MUC1 reporter, where Pinkhasov et al. (2011) employed a potent viral based expression system, perhaps causing increased contact with the endogenous plant PTM machinery.

Our preliminary studies suggest that the number of Pro residues in the MUC1 tandem repeat undergoing hydroxylation, and the degree of hydroxylation appear to increase in Arabidopsis plants and tobacco suspension BY-2 cells stably transformed with the same MUC1 reporter construct used in this study (Yang Z. and Petersen B.L., unpublished results). In plants, protein O-glycosylation is primarily found in Hyp-rich glycoproteins (HRGPs) of the cell wall, where Ser and in particular Hyp residues may be O-glycosylated. Dipeptidyl proline sequences in angiosperm HRGPs prone to hydroxylation were recently summarized by Kieliszewski et al. (2011) (59). Consistently hydroxylated sequences included the motifs AP, SP, PP, and the latter extensin-type motif is found in the APPA motif of the MUC1 tandem repeat. Several reports suggest involvement of the three P4Hs, AtP4H2, AtP4H5, and AtP4H13, in hydroxylation of consecutive Pro motifs in particular (17,20-22), perhaps pointing to these enzymes as first targets for engineering.

Glycoengineering of host cells generally seeks to achieve human glycosylation produced by normal cells, which involves complex-type N-glycans and/or core1 O-glycans capped with sialic acids to ensure circulation of injectable therapeutics for therapeutic effect. However, another important purpose of engineering O-glycans is to modulate immunogenic glycoforms produced by cancer cells and virus infected cells. Changes in O-glycosylation are a hallmark of cancer cells, and immunogenic short immature aberrant O-glycans are termed pan-carcinoma antigens (60). Our interest in engineering O-glycosylation from scratch stems from our previous identification of immunodominant aberrant O-glycopeptide epitopes in the cancer-associated MUC1 mucin, which are not covered by immunological tolerance (41,61,62). Vaccination with MUC1 glycopeptides with the truncated Tn O-glycoform produced by GalNAc-T2 and -T4 in combination produce IgG antibodies with cancer-specific reactivity (61), and spontaneous IgG antibodies to the same epitope are found in many cancer patients at time of diagnosis (3,63). Recombinant production of such vaccines will require a host cell that produces similar truncated O-glycans and glycoengineered plants as reported here may provide such a system. Interestingly, the acquisition of Hyp in the MUC1 repeat may not adversely affect immunogenicity, but rather stimulate immunity as recently reported for unglycosylated MUC1 vaccine produced in plants (19).

In the present study we have elucidated requirements for engineering plants with the capacity to perform initiation of human type O-glycosylation. O-glycosylation capacities may be engineered in e.g. tobacco with humanized N-glycosylation (64-69), which may serve as a more general platform for transient (70) as well as stable expression of glycoprotein pharmaceuticals. The results highlight that additional engineering is needed for plants to become versatile production platforms of therapeutics.
FIGURE 1. Depiction of the initiation process of mucin-type glycosylation in plants. A, The donor substrate UDP-GalNAc is synthesized from UDP-GlcNAc by Pseudomonas aeruginosa C4-epimerase (WbpP) protein. UDP-GalNAc is transported into the Golgi lumen by endogenous plant sugar nucleotide transporter(s) where Golgi localized GalNAc-T2 and -T4 catalyze the transfer of GalNAc onto the target peptide(s) destined for secretion in the secretory pathway.

FIGURE 2. O-glycosylation of the MUC1 reporter construct. A, The O-glycosylation target protein MUC1-YFP was N-terminally fused to a signal peptide for direction into the secretory pathway, followed by a (His)₆ tag. Gene constructs for introduction of O-glycosylation capacity into plants, comprising Flag tagged cytoplasmic, ER or Golgi targeted epimerase and native Golgi targeted GalNAc-T2, either expressed as a single polycistronic-protein, interspaced by the 2A self-splicing sequence (cleavage site indicated by an arrow), i.e. T2-2A-CytoEpi, CytoEpi-2A-T2 and GolgiEpi-2A-T2, or from separate promoters, i.e. EREpi-T2 and CytoEpi-T2. B, SDS-PAGE Western blot analysis of MUC1-YFP expressed alone (lane 1), together with cytoplasmic epimerase (lane 2) or GalNAc-T2 (lane 3). MUC1-YFP co-expressed with either of the 2A linked O-glycosylation machinery constructs (T2-2A-CytoEpi or CytoEpi-2A-T2) (lanes 4-5), 2A linked Golgi targeted epimerase and GalNAc-T2 (lane 6), Golgi targeted GalNAc-T2 with ER targeted epimerase (EREpi) (lane 7), and cytoplasmic epimerase (lane 8) expressed from unlinked glycosylation machinery. Absence or presence of O-glycosylation machinery is indicated above the lanes with (-) or (+), respectively. Total protein extracts from transiently transformed N. benthamiana leaves were loaded and blots reacted with MUC1 specific MAbs 5E10 and 5E5, where 5E5 is specific for GalNAc-glycosylated MUC1 and does not react with unglycosylated MUC1. Approximately 30 µg total protein was loaded in each lane.

FIGURE 3. Structural analysis of MUC1-YFP expressed in N. benthamiana. A, Purification of His-tagged MUC1-YFP. Left panel: SDS-PAGE Coomassie staining of MUC1-YFP expressed (-) alone and (+) together with 2A linked glycosylation machinery (T2-2A-CytoEpi). Right panel: Staining with Vicia villosa agglutinin (VVA) lectin for detection of GalNAc glycosylation. B, Schematic illustration of endo-Asp-N digestion of O-glycosylated MUC1. C, Analytical HPLC purification of endo-Asp-N digested MUC1-YFP expressed alone (upper panel) or together with O-glycosylation machinery (lower panel) separating released 20mer MUC1 repeat sequences with and without GalNAc-glycosylation. D-G, mass spec analysis of purified endo-Asp-N MUC1-YFP products. D, MUC1-YFP expressed alone. E, MUC1-YFP co-expressed with the O-glycosylation machinery (T2-2A-
CytoEpi). F, MUC1-YFP co-expressed with the inverse ordered glycosylation machinery (CytoEpi-2A-T2). G, MUC1-YFP co-expressed with 2A linked Golgi targeted epimerase and GalNAc-T2 (GolgiEpi-2A-T2). Each spectrum is a deconvoluted average centroid m/z representation. Peaks consistent with the MUC1-YFP sequence are labeled with monoisotopic m/z. These correspond to DTRPAPGSTAPPAHGVTSAP (31-50, 51-70, or 71-90, calculated monoisotopic m/z 1886.9355; +1 x Oxidation [i.e., Pro → Hyp], calculated monoisotopic m/z 1902.9304); DTLVNRIELKGIDFKE (220-235, calculated monoisotopic m/z 1890.0331); DGPVLLPDNYHLSYQSALS (293-312, calculated monoisotopic m/z 2217.1186); DFFKSAMPEGYVQERTIFFK (185-204, +1 x Oxidation [of Met or Pro], calculated monoisotopic m/z 2456.1955); DGNILGHKLEYNYNSHNVYITA (236-257, calculated monoisotopic m/z 2535.2263). Asterisks (*) denote peaks consistent with mono-, di- and triglycosylated DTRPAPGSTAPPAHGVTSAP peptide (31-50, 51-70, or 71-90, calculated monoisotopic m/z 2090.0149, 2293.0943, and 2496.1736, respectively).

**FIGURE 4. Co-expression of GalNAc-T2 and -T4 completes GalNAc O-glycosylation of MUC1.** A, Western blot analysis of target construct MUC1-YFP (lane 1) alone (-) or co-inoculated (+) with either of the 2A linked glycosylation machinery constructs T2-2A-CytoEpi (lane 2) or T2-2A-CytoEpi and ectopically with GalNAc-T4 (lane 3). Co-expression of GalNAc-T4 with the 2A linked glycosylation machinery (T2-2A-CytoEpi) conferred glycosylation of all 5 potential GalNAc O-glycosylation sites in MUC1-1TR. B, Mass spec analysis of end-Asp-N digested MUC1-YFP co-expressed with 2A linked glycosylation machinery (T2-2A-CytoEpi) and GalNAc-T4, demonstrating glycosylation of the three GalNAc-T2 specific sites and the two additional GalNAc that are added by GalNAc-T4 (DT\textsubscript{T}\textsubscript{2}RPAPGS\textsubscript{T}\textsubscript{2}T\textsubscript{2}APPAHGVT\textsubscript{T}\textsubscript{2}S\textsubscript{T}\textsubscript{4}AP), *i.e.* complete GalNAc occupancy of sites.

**FIGURE 5. O-glycosylation of human Mucin 16 (MUC16) reporter and Interferon α2B.** A, Western blot and lectin staining analysis of Mucin 16-1,2TR (MUC16), and B, Interferon α2B (INF-α2B), alone (-) or co-inoculated (+) with the 2A linked glycosylation machinery construct T2-2A-CytoEpi. MAb M11 is a specific antibody to MUC16, and MAb T7 is a T7-tag antibody. Glycosylation of MUC16 and INF-α2B were shown by staining with *Vicia villosa* agglutinin (VVA). 30 ug of total protein was loaded in each lane. The two constructs expressing MUC16 and INF-α2B are shown with the potential glycosylation sites (S or T) denoted by underlining.
1. Elliott, S., Lorenzini, T., Asher, S., Aoki, K., Brankow, D., Buck, L., Busse, L., Chang, D., Fuller, J., Grant, J., Hernday, N., Hokum, M., Hu, S., Knudten, A., Levin, N., Komorowski, R., Martin, F., Navarro, R., Osslund, T., Rogers, G., Rogers, N., Trail, G., and Egrie, J. (2003) *Nature biotechnology* **21**, 414-421

2. Tarp, M. A., and Clausen, H. (2008) *Biochim Biophys Acta* **1780**, 546-563

3. Wandall, H. H., Blixt, O., Tarp, M. A., Pedersen, J. W., Bennett, E. P., Mandel, U., Ragupathi, G., Livingston, P. O., Hollingsworth, M. A., Taylor-Papadimitriou, J., Burchell, J., and Clausen, H. (2010) *Cancer Res* **70**, 1306-1313

4. Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M. R., Neff, N., Passarelli, B., Koh, W., Fan, H. C., Wang, J., Gui, Y., Lee, K. H., Betenbaugh, M. J., Quake, S. R., Famili, I., Palsson, B. O., and Wang, J. (2011) *Nature biotechnology* **29**, 735-741

5. Gerlach, J. Q., Kilcoyne, M., Eaton, S., Bhavanandan, V., and Joshi, L. (2011) *Adv Exp Med Biol* **705**, 257-269

6. Gomord, V., and Faye, L. (2004) *Curr Opin Plant Biol* **7**, 171-181

7. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2011) *Glycobiology*

8. Kishimoto, T., Watanabe, M., Mitsui, T., and Hori, H. (1999) *Arch Biochem Biophys* **370**, 271-277

9. Kilcoyne, M., Shah, M., Gerlach, J. Q., Bhavanandan, V., Nagaraj, V., Smith, A. D., Fujiyama, K., Sommer, U., Costello, C. E., Olszewski, N., and Joshi, L. (2009) *J Plant Physiol* **166**, 219-232

10. Alonso, A. P., Piasecki, R. J., Wang, Y., LaClair, R. W., and Shachar-Hill, Y. (2010) *Plant physiology* **153**, 915-924

11. Hashimoto, K., Tokimatsu, T., Kawano, S., Yoshizawa, A. C., Okuda, S., Goto, S., and Kaniehisa, M. (2009) *Carbohydr Res* **344**, 881-887

12. Henrissat, B., Surolia, A., and Stanley, P. (2009)

13. Seveno, M., Bardor, M., Paccalet, T., Gomord, V., Lerouge, P., and Faye, L. (2004) *Nature biotechnology* **22**, 1351-1352; author reply 1352-1353

14. Amano, K., Chiba, Y., Kasahara, Y., Kato, Y., Kaneko, M. K., Kuno, A., Ito, H., Kobayashi, K., Hirabayashi, J., Jigami, Y., and Narimatsu, H. (2008) *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3232-3237

15. Lamport, D. T. A., Kieliszewski, M. J., Chen, Y., and Cannon, M. C. (2011) *Plant physiology* **156**, 11-19

16. Yuasa, K., Toyooka, K., Fukuda, H., and Matsuoka, K. (2005) *Plant J* **41**, 81-94

17. Velasquez, S. M., Ricardi, M. M., Dorosz, J. G., Fernandez, P. V., Nadra, A. D., Pol-Fachin, L., Egelund, J., Gille, S., Harholt, J., Ciancia, M., Verli, H., Pauly, M., Bacic, A., Olsen, C. E., Ulvskov, P., Petersen, B. L., Somerville, C., Iusem, N. D., and Estevez, J. M. (2011) *Science (New York, N Y)* **332**, 1401-1403

18. Karnoup, A. S., Turkelson, V., and Anderson, W. H. (2005) *Glycobiology* **15**, 965-981
19. Pinkhasov, J., Alvarez, M. L., Rigano, M. M., Piensook, K., Larios, D., Pabst, M., Grass, J., Mukherjee, P., Gendler, S. J., Walmsley, A. M., and Mason, H. S. (2011) *Plant biotechnology journal*

20. Hietanen, P., Myllyharju, J. (2005) *The Journal of biological chemistry* **280**, 1142-1148

21. Keskiaho, K., Hietanen, R., Sormunen, R., and Myllyharju, J. (2007) *The Plant cell* **19**, 256-269

22. Reaves, M. L., Lopez, L. C., and Daskalova, S. M. (2011) *Research in Plant Biology* **1(4)**: 49-54

23. Egelund, J., Obel, N., Ulvskov, P., Geshi, N., Pauly, M., Bacic, A., and Petersen, B. L. (2007) *Plant molecular biology* **64**, 439-451

24. Sainsbury, F., and Lomonossoff, G. P. (2008) *Plant physiology* **148**, 1212-1218

25. Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T., and Halkier, B. A. (2007) *Nucleic acids research* **35**, e55

26. Petersen, B. L., Egelund, J., Damager, I., Faber, K., Jensen, J. K., Yang, Z., Bennett, E. P., Scheller, H. V., and Ulvskov, P. (2009) *Glycoconjugate journal* **26**, 1235-1246

27. Tarp, M. A., Sorensen, A. L., Mandel, U., Paulsen, H., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2007) *Glycobiology* **17**, 197-209
42. Schjoldager, K. T., Vester-Christensen, M. B., Bennett, E. P., Levery, S. B., Schwientek, T., Yin, W., Blixt, O., and Clausen, H. (2010) The Journal of biological chemistry 285, 36293-36303
43. Berninsone, P., and Hirschberg, C. B. (1998) Ann NY Acad Sci 842, 91-99
44. Hirschberg, C. B., Robbins, P. W., and Abeijon, C. (1998) Annu Rev Biochem 67, 49-69
45. Berninsone, P. M., and Hirschberg, C. B. (2000) Curr Opin Struct Biol 10, 542-547
46. Segawa, H., Kawakita, M., and Ishida, N. (2002) Eur J Biochem 269, 128-138
47. Handford, M., Rodriguez-Furlan, C., and Orellana, A. (2006) Braz J Med Biol Res 39, 1149-1158
48. White, T., Bennett, E. P., Takio, K., Sorensen, T., Bonding, N., and Clausen, H. (1995) The Journal of biological chemistry 270, 24156-24165
49. Halpin, C. (2005) Plant biotechnology journal 3, 141-155
50. Donnelly, M. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D., and Ryan, M. D. (2001) The Journal of general virology 82, 1013-1025
51. DeFrees, S., Wang, Z. G., Xing, R., Scott, A. E., Wang, J., Zopf, D., Gouty, D. L., Sjoberg, E. R., Panneerselvam, K., Brinkman-Van der Linden, E. C., Bayer, R. J., Tarp, M. A., and Clausen, H. (2006) Glycobiology 16, 833-843
52. de Felipe, P., and Ryan, M. D. (2004) Traffic (Copenhagen, Denmark) 5, 616-626
53. Gerngross, T. U. (2004) Nature biotechnology 22, 1409-1414
54. Willer, T., Valero, M. C., Tanner, W., Cruces, J., and Strahl, S. (2003) Curr Opin Struct Biol 13, 621-630
55. Yip, C. L., Welch, S. K., Klebl, F., Gilbert, T., Seidel, P., Grant, F. J., O'Hara, P. J., and MacKay, V. L. (1994) Proceedings of the National Academy of Sciences of the United States of America 91, 2723-2727
56. Kuroda, K., Kobayashi, K., Kitagawa, Y., Nakagawa, T., Tsumura, H., Komeda, T., Shinmi, D., Mori, E., Motoki, K., Fuji, K., Sakai, T., Nonaka, K., Suzuki, T., Ichikawa, K., Chiba, Y., and Jigami, Y. (2008) Appl Environ Microbiol 74, 446-453
57. Liu, L., Xu, Y. X., and Hirschberg, C. B. (2010) Semin Cell Dev Biol 21, 600-608
58. Showalter, A. M., Liang, Y., Faik, A., Kieliszewski, M., Tan, L., and Xu, W. L. (2010) Plant physiology 154, 632-642
59. Kieliszewski, M. J., Lamport, D. T. A., Tan, L., and Cannon, M. C. (2010) Hydroxyproline-Rich Glycoproteins: Form and Function. in Annual Plant Reviews, Wiley-Blackwell. pp 321-342
60. Springer, G. F. (1984) Science 224, 1198-1206
61. Sabbatini, P. J., Ragupathi, G., Hood, C., Aghajanian, C. A., Juretzka, M., Iasonos, A., Hensley, M. L., Spassova, M. K., Ouerfelli, O., Spriggs, D. R., Tew, W. P., Konner, J., Clausen, H., Abu Rustum, N., Dansihefsky, S. J., and Livingston, P. O. (2007) Clin Cancer Res 13, 4170-4177
62. Sorensen, A. L., Reis, C. A., Tarp, M. A., Mandel, U., Ramachandran, K., Sankaranarayanan, V., Schwientek, T., Graham, R., Taylor-Papadimitriou, J., Hollingsworth, M. A., Burchell, J., and Clausen, H. (2006) Glycobiology 16, 96-107
63. Blixt, O., Bueti, D., Burford, B., Allen, D., Julien, S., Hollingsworth, M., Gammerman, A., Fentiman, I., Taylor-Papadimitriou, J., and Burchell, J. M. (2011) *Breast Cancer Res* **13**, R25

64. Strasser, R., Stadlmann, J., Schahs, M., Stiegler, G., Quendler, H., Mach, L., Glossl, J., Weterings, K., Pabst, M., and Steinkellner, H. (2008) *Plant biotechnology journal* **6**, 392-402

65. Castilho, A., Strasser, R., Stadlmann, J., Grass, J., Jez, J., Gattinger, P., Kunert, R., Quendler, H., Pabst, M., Leonard, R., Altmann, F., and Steinkellner, H. (2010) *The Journal of biological chemistry* **285**, 15923-15930

66. Nagels, B., Van Damme, E. J. M., Pabst, M., Callewaert, N., and Weterings, K. (2011) *Plant physiology* **155**, 1103-1112

67. Schoberer, J., and Strasser, R. (2011) *Molecular plant* **4**, 220-228

68. Kaulfurst-Soboll, H., Rips, S., Koiwa, H., Kajiura, H., Fujiyama, K., and von Schaewen, A. (2011) *The Journal of biological chemistry* **286**, 22955-22964

69. Yin, B.-J., Gao, T., Zheng, N.-Y., Li, Y., Tang, S.-Y., Liang, L.-M., and Xie, Q. (2011) *Protein & cell* **2**, 41-47

70. Pogue, G. P., Vojdani, F., Palmer, K. E., Hiatt, E., Hume, S., Phelps, J., Long, L., Bohorova, N., Kim, D., Pauly, M., Velasco, J., Whaley, K., Zeitlin, L., Garger, S. J., White, E., Bai, Y., Haydon, H., and Bratcher, B. (2010) *Plant biotechnology journal* **8**, 638-654
GalNAc-T2
UDP-GlcNAc/Glc 4-epimerase
(WbpP)
UDP-GlcNAc/Glc
UDP-GalNAc
Golgi
plant UDP-Gal/GalNAc
transporter

UDP-(UDP-GalNAc)
protein
GalNAc-T2
GalNAc-T4
O-glycosylated protein

Yang et al. Fig. 1
Yang et al. Fig. 2

A

MUC1-YFP

| S  | SS  | MUC1-3.5 TR | YFP | His tag |
|----|-----|-------------|-----|---------|
| 1  | 28  | 103         | 341 | 347     |

T2-2A-CytoEpi

| Flag tag | Epimerase | 2A | GalNAc-T2 |
|----------|-----------|----|-----------|
| 1        | 571       | 591| 600       |

CytoEpi-2A-T2

| Flag tag | Epimerase | 2A | GalNAc-T2 |
|----------|-----------|----|-----------|
| 1        | 370       | 390| 615       |

GolgiEpi-2A-T2

| T2 TMD | Epimerase | 2A | GalNAc-T2 |
|--------|-----------|----|-----------|
| 1      | 138       | 486| 595       |

EREpi-T2

| Flag tag | Epimerase | KDEL | GalNAc-T2 |
|----------|-----------|------|-----------|
| 1        | 21        | 30   | 389       |

CytoEpi-T2

| Flag tag | Epimerase | GalNAc-T2 |
|----------|-----------|-----------|
| 1        | 360       | 571      |

B

GalNAc-MUC1 \{ MUC1 \}

| Lane | CytoEpi | T2 | T2-2A-CytoEpi | CytoEpi-2A-T2 | GolgiEpi-2A-T2 | EREpi-2A-T2 | CytoEpi-T2 |
|------|---------|----|---------------|---------------|----------------|-------------|------------|
| 1    | -       | -  | +             | +             | +              | +           | -          |
| 2    | -       | -  | +             | +             | +              | +           | -          |
| 3    | -       | +  | +             | +             | +              | +           | -          |
| 4    | -       | +  | +             | +             | +              | +           | -          |
| 5    | -       | +  | +             | +             | +              | +           | -          |
| 6    | -       | +  | +             | +             | +              | +           | -          |
| 7    | -       | +  | +             | +             | +              | +           | -          |
| 8    | -       | +  | +             | +             | +              | +           | -          |

GalNAc-MUC1 \{ MUC1 \}

| Lane | MUC1 (MAb 5E10) | GalNAc-MUC1 (MAb 5E5) |
|------|-----------------|-----------------------|
| 1    |                 |                       |
| 2    |                 |                       |
| 3    |                 |                       |
| 4    |                 |                       |
| 5    |                 |                       |
| 6    |                 |                       |
| 7    |                 |                       |
| 8    |                 |                       |
Yang et al. Fig. 3

A

T2-2A-CytoEpi

GalNAc-MUC1
MUC1

Coomassie Stain
VVA

B

Endo-Asp
Endo-Asp
Endo-Asp
Endo-Asp

Sig.pep.-VTSAP-
DTRPAPGSTAPPANGTSAP

n=3; MW=1887

C

Endo-Asp digest of MUC1-YFP

Ab=210 nm

Endo-Asp digest of MUC1-YFP + T2-2A-CytoEpi

Ab=210 nm

D

MUC1-YFP

1886.948
1890.044
1902.941
1949.983
2217.128
2217.120
2456.201
2456.197
2535.231
2535.226

E

MUC1-YFP + T2-2A-CytoEpi

1890.034
1902.932
1949.973
2154.178
2217.120
2293.096
2496.176
2496.195
2535.226

F

MUC1-YFP + CytoEpi-2A-T2

1886.948
1886.947
1902.941
1902.929
2090.027
2154.200
2217.120
2293.293
2496.195

G

MUC1-YFP + GolgiEpi-2A-T2
Yang et al. Fig. 4

A

- 49
+ + - + T2-2A-CytoEpi + T4
- 38
kDa

GalNAc-MUC1
MUC1

| lane | 1 | 2 | 3 | 1 | 2 | 3 |
|------|---|---|---|---|---|---|
| MUC1 |   |   |   | MUC1 |   |   |
| (MAb 5E10) |   |   |   | (MAb 5E5) |   |   |

B

MUC1-YFP + T2-2A-CytoEpi + T4

1 Pro-Hyp
1902.927
1924.879
GalNAc-MUC1-1TR
2090.012
2154.173
GalNAc-MUC1-1TR
2293.092
2496.173
GalNAc-MUC1-1TR
2699.252
2902.252

1886.935
Yang et al. Fig. 5

A

SS | MUC16-1.2-TR | T7 | His tag

- 38 kDa
- 49

T2-2A-CytoEpi

GalNAc-MUC16
MUC16

MAb: M11 VVA

B

SS | INF-α2B | T7 | SP₁₀ | His tag

- 38 kDa
- 49
- 38
- 28

T2-2A-CytoEpi

GalNAc-INF
INF

MAb: T7 VVA
Engineering mammalian mucin-type O-glycosylation in plants
Zhang Yang, Damian P. Drew, Bodil Joergensen, Ulla Mandel, Soeren S. Bach, Peter Ulvskov, Steven B. Levery, Eric P. Bennett, Henrik Clausen and Bent L. Petersen

J. Biol. Chem. published online February 14, 2012

Access the most updated version of this article at doi: 10.1074/jbc.M111.312918

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2012/02/14/M111.312918.DC1