Running head: Plant Cell Factories with Engineered Mucin Type O-Glycosylation

Corresponding author: Bent L. Petersen
Address: Department of Plant Biology and Biotechnology, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark
Tel: 45-35332670
Fax: 45-35333300
Email: blp@life.ku.dk

Research area: Biochemical Processes and Macromolecular Structures
Towards Stable Genetic Engineering of Human O-glycosylation in Plants

Zhang Yang¹, Eric P. Bennett⁴, Bodil Jørgensen³, Damian P. Drew²†, Emma Arigi⁴, Ulla Mandel⁴, Peter Ulvskov², Steven B. Levery⁴, Henrik Clausen⁴, Bent L. Petersen²*

¹ Department of Molecular Biology and Genetics, Faculty of Science and Technology, Aarhus University, Flakkebjerg, 4200 Slagelse, Denmark

² Department of Plant Biology and Biotechnology, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

³ Department of Agriculture and Ecology, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

⁴ Center for Glycomics, Departments of Cellular and Molecular Medicine and School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen N, Denmark
Financial source:
This work was supported by the Danish Council for Strategic Research (DSF: 09-063090), Danish Council for Independent Research (DFF: 09-066624/274-09-0314), a program of excellence from the University of Copenhagen, and the Danish National Research Foundation.

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

*Corresponding author: Bent L. Petersen; e-mail: blp@life.ku.dk
Abstract
Glycosylation is the most abundant and complex posttranslational modification to be considered for recombinant production of therapeutic proteins. Mucin-type (GalNAc-type) O-glycosylation is found in eumetazoan cells, but absent in plants and yeast, making these cell types an obvious choice for de novo engineering of this O-glycosylation pathway. We previously showed that transient implementation of O-glycosylation capacity in plants requires introduction of synthesis of the donor substrate UDP-GalNAc and one or more polypeptide GalNAc-transferases for incorporating GalNAc residues into proteins (Yang et al., 2012). Here, we have stably engineered O-glycosylation capacity in two plant cell systems, i.e. soil grown Arabidopsis thaliana and Nicotiana tabacum Bright Yellow 2 (BY-2) suspension culture cells. Efficient GalNAc O-glycosylation of two stably co-expressed substrate O-glycoproteins was obtained, but a high degree of proline hydroxylation and hydroxyproline linked arabinosides, on a mucin (MUC1) derived substrate was also observed. Addition of the Prolyl 4-hydroxylase inhibitor 2,2-dipyridyl, however, effectively suppressed proline hydroxylation and arabinosylation of MUC1 in BY-2 cells. In summary, stably engineered mammalian type O-glycosylation was established in transgenic plants, demonstrating that plants may serve as host cells for production of recombinant O-glycoproteins. However, the present stable implementation further strengthens the notion that elimination of endogenous posttranslational modifications may be needed for production of protein therapeutics.
Introduction

Posttranslational modifications (PTMs) of recombinant protein therapeutics are important with glycosylation being by far the most important PTM to address. *N*- and *O*-linked glycosylation are often essential for biological activity and pharmacokinetic properties (Li and d’Anjou, 2009). Recombinant production of glycoprotein therapeutics has so far generally aimed to mimic the glycosylation found on naturally produced human glycoproteins primarily to avoid adverse immunological reactivity and improve circulatory half-life. Considerable efforts have thus been devoted to design and optimization of host cells aimed for recombinant expression of biologics. Significant success has been achieved with designed *N*-glycosylation in both higher eukaryotes such as plants (Strasser et al., 2008; Gomord et al., 2010), lower eukaryotes such as yeast (Gerngross, 2004), and even extended to some prokaryotes (Hug et al., 2011). In contrast, limited progress has been made with respect to *O*-glycosylation. *O*-glycosylation is one of the most complex regulated PTMs of proteins and also one of the least understood (Bennett et al., 2011). *O*-glycosylation is initiated by a family of up to 20 UDP-GalNAc: polypeptide *N*-acetyl-galactosaminytransferases (GalNAc-Ts), which allows for huge variability of *O*-glycan attachments to proteins in different cells, a feature not found for most other types of protein glycosylation. Engineering glycosylation in host cells may also aim to produce aberrant truncated glycoforms that mimic those found in disease with intend to *e.g.* target proteins to antigen presenting cells and augment immunity in vaccine strategies. It may therefore be desirable to build glycosylation capacity in host cells *de novo* in order to attain complete design freedom without extensive requirement for gene elimination. This is particularly desirable for mammalian *O*-glycosylation where cell-specific repertoires of the many GalNAc-Ts direct the glycosylation capacity.

Most glycoprotein therapeutics are produced in Chinese hamster ovary cells (CHO) (Durocher and Butler, 2009). CHO cells produce mono- and disialylated core 1 (Galβ1-3GalNAcα1-O-Ser/Thr) *O*-glycan structures (Sasaki et al., 1987). Recently the repertoire of GalNAc-Ts expressed in CHO cells was shown to include only four isoforms (GalNAc-T2, T7, T11, and T19) (Xu et al., 2011), thus demonstrating that CHO cells have limited capacity for *O*-glycosylation. Insect cells produce immature *O*-glycans generally without sialic acids and have an as yet unknown GalNAc-T repertoire (Thomsen et al., 1990). Yeast does not produce GalNAc-type *O*-glycosylation, but *O*-mannosylates similar Ser/Thr residues as GalNAc *O*-
glycosylation in higher eukaryotes. The yeast *Pichia pastoris* was previously used as a model system to engineer the first step of human *O*-glycosylation from scratch (Amano et al., 2008). However, since ER localized *O*-mannosylation precedes GalNAc *O*-glycosylation in the Golgi apparatus, considerable *O*-mannosylation was found on glycoproteins expressed in the engineered cells. This competing *O*-mannosylation was partly blocked by co-culturing in rhodanine-3-acetic acid, but complete elimination of *O*-mannosylation is presumably not possible, as knock-out of three or more of the seven *O*-mannosyltransferases is lethal (Strahl-Bolsinger et al., 1999).

Plants are emerging as an attractive alternative platform for expression of glycoproteins (Cox et al., 2006; Sainsbury and Lomonossoff, 2008; Pogue et al., 2010). Several studies have demonstrated that plants can efficiently produce complex mammalian proteins, such as monoclonal antibodies with humanized *N*-glycan structure (Cox et al., 2006; Strasser et al., 2008; Castilho et al., 2010; Pogue et al., 2010). Whereas plants and mammalian cells have identical core *N*-linked glycosylation, plants do not perform GalNAc *O*-glycosylation (Gomord and Faye, 2004; Saint-Jore-Dupas et al., 2007; Bennett et al., 2011). Plants, however, produce another type of *O*-glycosylation, which is primarily found in the hydroxylproline (Hyp) rich glycoprotein (HRGP) superfamily, where e.g. arabinogalactan (AG) and arabinosides are attached to hydroxyprolines (Hyp’s) of arabinogalactan proteins (AGPs) and extensins, respectively. This type of *O*-glycosylation requires the initial action of prolyl 4-hydroxylases (P4Hs), and although a consensus motif for this type of modification has not been defined, it appears to occur mainly in regions with high density of Pro residues (Shimizu et al., 2005; Estevez et al., 2006; Jamet et al., 2008; Lamport et al., 2011). The Hyp contiguity hypothesis predicts arabinosylation of contiguous Hyps (≥2) with extensin type β-linked arabinofuranosides (β-Araf) of up to four residues in length and galactoarabinosylation of clustered noncontiguous Hyp residues in AGPs (Shpak et al., 1999; Zhao et al., 2002; Tan et al., 2010). Several small peptide hormones and plasma membrane kinases involved in cell wall sensing have also been shown to be arabinosylated (for recent reviews see (Ringli, 2010; Matsubayashi, 2011)). While small peptide hormones and plasma membrane kinases invariably carry the Hyp-(β-Araf) side-chain, extensin β-Araf side-chains may be shorter or additionally capped with a α-linked Araf. The amino acid sequence ‘SerProProPro(Pro)’ directs arabinosylation in extensins (Kieliszewski et al. 2011),
but a glycosylation motif has not yet been identified in the plasma membrane kinases or the small peptide hormones. Plants also produce O-linked galactosylation of Ser residues, which is found on some vacuolar proteins such as sweet potato sporamin of the Solanaceae family (Solanaceous lectin-type O-galactosylation) and in the plant specific extensin SerProProPro(Pro) motif repeat of cell wall extensins (Lamport et al., 1973; Yong-Pill and Chrispeels, 1976; Matsuoka et al., 1995; Kieliszewski and Shpak, 2001). The restricted occurrence or sequence specificity of these plant specific types of O-glycosylation, should make plants an ideal candidate cell system for custom design and build-up of O-glycosylation from scratch with respect to both sites of O-glycan attachment and O-glycan structures.

Daskalova et al. (Daskalova et al., 2010) initially started engineering plants for mammalian O-glycosylation and suggested that GalNAc O-glycosylation in Nicotiana benthamiana L plants required a Glc(NAc) C4-epimerase, a UDP-Gal(NAc) transporter, a human GalNAc-T. We subsequently tested a variety of design strategies for introduction of GalNAc O-glycosylation in plants using transient expression of the glycogenes in N. benthamiana and found that O-glycosylation capacity was achieved by introduction of only the Pseudomonas aeruginosa Glc(NAc) C4-epimerase (WbpP) (Creuzenet et al., 2000) and a human GalNAc-T. The apparent discrepancy with regard to the requirement for introduction of a UDP-GalNAc transporter is not fully clarified, but the study by Daskalova et al. was solely based on lectin labeling and the authors subsequently demonstrated that the lectins used detect endogenous plant proteins (Reaves et al., 2011) shedding serious concerns with regard to the original findings. Our study included production of recombinant glycoproteins in quantities allowing for a detailed characterization of the introduced glycosylation by mass spectrometry (Yang et al., 2012). However, stoichiometry and subcellular-localization of the introduced glycoenzymes and sub-cellular substrate pools are increasingly difficult to control as the engineering effort increases in complexity. Stable engineering of especially the glycosylation machineries are thus generally at an early point needed as for example seen in the engineering efforts to humanize plants with respect to N-linked protein glycosylation (Strasser et al., 2008; Castilho et al., 2010; Kaulfurst-Soboll et al., 2011; Nagels et al., 2011; Schoberer and Strasser, 2011; Yin et al., 2011)

Here, we have established stable O-glycosylation capacity in two plant cell
systems, i.e. soil grown Arabidopsis thaliana (Arabidopsis) and Nicotiana tabacum Bright Yellow 2 (BY-2) suspension culture cells. In agreement with our previous studies using transient expression (Yang et al., 2012), we found that efficient O-glycosylation capacity was obtained without introduction of a UDP-Gal(NAc) transporter in Arabidopsis and tobacco BY-2 suspension cells. We also found that Pro hydroxylation and hydroxyproline linked arabinosylation produced by the endogenous plant PTM machinery were considerably more pronounced in stable engineered cells compared to our transiently engineered cells. Interestingly, for suspension cell cultures the presence of a prolyl 4-hydroxylase inhibitor resulted in a marked reduction of the endogenous PTMs. In summary, plants were stably engineered with efficient O-glycosylation machineries, but further host cell engineering aimed at eliminating Pro hydroxylation in O-glycoproteins prone for this endogenous modification appear to be needed.

RESULTS
Engineering GalNAc O-Glycosylation in Plants

Previously, using transient expression in Nicotiana benthamiana we identified optimal two-component GalNAc O-glycosylation machinery designs (Fig. 1), which conferred efficient O-glycosylation of co-expressed reporter substrates (Yang et al., 2012). Of the earlier transiently tested O-glycosylation machinery constructs, T2-2A-CytoEpi and CytoEpi-T2 (Fig. 1A), expressing Golgi targeted GalNAc-T2 and cytoplasmic epimerase as a single polyprotein interspaced by the 2A self cleavage sequence or as separate proteins, respectively, conferred efficient O-glycosylation in the stable implementations.

Stably introduced O-glycosylation reporter constructs included: i) the ca. 11 kDa 3.5 tandem repeat sequence of the cancer-associated mucin MUC1 (MUC1-3.5TR), N-terminally fused to a His6 and T7 tag (MUC1), ii) MUC1-3.5TR C-terminally fused to YFP (MUC1-YFP), iii) MUC1-2.5TR embedded into GFP (GF(MUC1)P), iv) the full length secreted IFNα2B cytokine tagged with a T7 tag and an AGP type (SP)10 glycomodule (IFN-SP10) (Fig. 1B), and v) 1.2TR of MUC16 tagged with T7 and His6 tags (MUC16) (Supplemental Fig. S1B). All reporter constructs included an N-terminal signal peptide sequence for direction into the secretory pathway.
We first tested expression and stability of reporter substrates. Expression of the small ca. 11 kDa MUC1 3.5TR (MUC1) reporter in Arabidopsis and water grown *Lemna minor* conferred only weak accumulation (Fig. 2A, lanes 1-2 and Supplemental Fig. S1A), but was not detectable in BY-2 cells (data not shown) as evidenced by SDS-PAGE Western blot analyses. Expression of the fusion reporter substrate MUC1-YFP (Fig. 1B), however, resulted in agreement with the transient implementations (Yang et al. 2012) in a ~20 fold increased accumulation in Arabidopsis (Fig. 2A, lanes 4-5) as compared to expression of MUC1 alone (Fig. 2A, lane 1-2). Expression of MUC1-YFP in BY-2 cell callus also resulted in detectable intracellular accumulation (Fig. 2B, lane 1). Intracellular *O*-glycosylation of MUC1-YFP in BY-2 cells was achieved by co-expression of the *O*-glycosylation machinery T2-2A-CytoEpi as evidenced by Western blot analysis using a GalNAc-MUC1 glycoform (Tn-MUC1) specific MAb 5E5 (Fig. 2B, lane 2). However, neither MUC1-YFP nor Tn-MUC1-YFP was detectable in the BY-2 cell medium (data not shown). Expression of another mucin-type substrate, MUC16, also conferred accumulation in the BY-2 medium. The MUC16 reporter construct design contains 3 potential *N*-glycosylation and up to 30 *O*-glycosylation sites, and PNGase treatment of MUC16 containing media fractions resulted in mobility shifts to an apparent slightly lower MW on SDS-PAGE Western blots in agreement with the presence of *N*-glycosylation sites that are utilized by the endogenous plant *N*-glycosylation machinery (Supplemental Fig. S1B).

The secreted form of IFNα2B tagged with a SP_{10} glycomodule (IFN-SP_{10}), was also tested as an *O*-glycosylation substrate in Arabidopsis. Co-expression with the *O*-glycosylation machinery CytoEpi-T2 resulted in a mobility shift and reactivity with the VVA lectin (Fig. 2C). IFNα2B only has a single *O*-glycosylation site (GVGVT^{132}ETPLM^{137}), which is known to be *O*-glycosylated specifically by GalNAc-T2 (DeFrees et al., 2006). The observed mobility shift and VVA labeling clearly indicate that the protein was *O*-glycosylated (Fig. 2C), although mass spectrometry analysis is needed for confirmation of glycosylation site.

Finally, using transient expression of the glycogenes, prior introduction of the glycosylation machinery appeared to enhance the full occupancy of the three GalNAc-T2 specific *O*-glycosylation sites in the MUC1 tandem repeat from ~ 5% to ~ 40% (Supplemental Fig. S4) perhaps pointing to the significance of providing ample amounts of donor substrate and transferase enzyme.
MUC1 of MUC1-YFP is degraded in BY-2 cell medium

Repeated failed attempts to detect MUC1 expressed alone or in fusion with the carrier YFP protein in the BY-2 media fractions prompted us to look further into the fate of in particular MUC1-YFP. A BY-2 transgenic cell line stably expressing MUC1-YFP and the \( O \)-glycosylation machinery T2-2A-CytoEpi failed to produce detectable levels of the MUC1-YFP protein (37 kDa) in BY-2 medium as evidenced by MUC1 specific Western blot analysis (data not shown). His-tag affinity purification of MUC1-YFP from spent medium of the BY-2 cell line was also not possible (data not shown). Concentration by hydrophobic interaction chromatography (HIC), however, revealed the presence of substantial amounts of fluorescent protein (Fig. 3A), which migrated as a 25 kDa protein on SDS-PAGE gels and reacted with a GFP polyclonal Ab (Fig. 3B+C), but not with a MUC1 specific MAb (Fig. 3D). The yield was estimated to be approximately 25 mg/L as judged from Coomassie stained SDS-PAGE (Fig. 3B). Based on this, we concluded that MUC1-YFP was partially degraded in the BY-2 medium resulting in release of the C-terminal His-tag and the N-terminal MUC1 sequence. A construct expressing a fusion of an AGP type (SP)\textsuperscript{10} glyco-module (Xu et al., 2010) to the C-terminal of MUC1 also failed to confer accumulation of MUC1 in the BY-2 medium (data not shown).

Embedding and glycosylation stabilize secreted MUC1 in BY-2 medium

Previous studies have shown that embedding of small otherwise labile proteins in GFP may confer increased stabilization (Kobayashi et al., 2008). Accordingly, we prepared an embedded MUC1 construct, designated GF(MUC1)P, where a MUC1-2.5TR sequence flanked C-terminally with a (His)\textsubscript{8} and N-terminally with a Myc tag, was embedded in the loop between Asp196 and Gly197 of GFP (Fig. 4A). A transgenic BY-2 cell line expressing the embedded GF(MUC1)P substrate together with the \( O \)-glycosylation machinery CytoEpi-T2 appeared to accumulate fluorescence in callus cells (Fig. 4B, left panel). Isolation of GF(MUC1)P from the medium by Ni-affinity chromatography was now feasible and the isolated product migrated as a single band (~ 40 kDa) as judged by Coomassie stained SDS-PAGE (Fig. 4B, right panel), suggesting that MUC1 was stabilized by embedding it into GFP. Endo-Asp mediated release of the MUC1 TR 20-mer peptide from the purified product with subsequent matrix-assisted laser desorption ionization time-of-flight mass
spectrometry (MALDI-TOF-MS) confirmed a uniform incorporation of 3 GalNAc residues (Fig. 5A). However, evidence was also found for modification of Pro to Hyp, and of Hyp with 3 (~45%) and less frequently 6 (~5%) arabinose residues attached (Fig. 5A). The MUC1 repeat contains 5 proline residues of which up to 4 were hydroxylated.

**O-glycosylation further stabilizes MUC1**

An *in vitro* degradation model was used to evaluate the effect of *O-*glycosylation on MUC1 stability in BY-2 medium. Purified GF(MUC1)P of BY-2 cell lysates, purified glycosylated Tn-GF(MUC1)P of BY-2 medium, and purified MUC1-YFP and Tn-MUC1-YFP, transiently produced in *N. benthamiana* plants, were added to 7 day-old BY-2 wt medium with and without prior boiling (Fig. 4). While all MUC1 substrates were stable in boiled medium (Fig. 4C), only the Tn-GF(MUC1)P was stable in non-boiled medium after 24 h incubation (Fig. 4C, lane 41-45), as evidenced by Western blot analysis with MUC1 specific mAb 5E10. For the corresponding Tn-MUC1 glycoform specific Western blot, see Supplemental Fig. S3.

**2,2-Dipyridyl inhibits Pro Hydroxylation and Arabinosylation in BY-2 cells**

In order to block or reduce formation of the observed endogenous PTMs on the GF(MUC1)P reporter substrate, three inhibitors of proline hydroxylase activity were selected (ethyl-3,4-dihydroxybenzoate (EDHB), 3,4-dehydroproline (DHP) and 2,2-dipyridyl (2,2 DP)) (Schmidt et al., 1991; Velasquez et al., 2011) and added to the BY-2 medium at various concentrations. Addition of all three inhibitors reduced the accumulated level of GF(MUC1)P in the medium as well as cell growth rates (Supplemental Fig. S5). Only 2,2 DP was able to efficiently suppress proline hydroxylation of MUC1 tandem repeats and the accompanying attachment of pentose sugars (Fig. 5B and Supplemental Fig. S6). The larger arabinogalactans and the short arabinofuranosides are the only known Hyp-linked glycans in plants (Lamport et al. 2011). In the transient expression system, addition of the proline hydroxylase inhibitors conferred a minor reduction on the proline hydroxylation level from ~20% to ~10% without an accompanying reduced level of expressed MUC1-YFP (Supplemental Fig. S7).
MUC1-YFP is glycosylated and Hyp arabinosylated in glycoengineered Arabidopsis

Expression of the MUC1 constructs and their use as glycosylation targets were also tested in soil grown Arabidopsis. SDS-PAGE western blot analysis of Arabidopsis lines expressing a single construct encoding combined expression of MUC1-YFP and the O-glycosylation machinery T2-2A-CytoEpi (Fig. 1C) revealed the presence of two closely migrating bands when probed with MAb 5E10 specific to MUC1 (Fig. 6A-1). Only the upper of these bands reacted with MAb 5E5, which is specific to Tn-MUC1 (Fig. 6A-1). The MUC1-YFP was purified from leaves and roots by HIC and O-glycosylation was confirmed by MALDI-TOF-MS analysis. Coomassie staining of purified MUC1-YFP from Arabidopsis leaves revealed that the major product was YFP without the MUC1 entity (Fig. 6A-2). MUC1-YFP of the roots was almost completely degraded as judged by Coomassie stained SDS-PAGE (Fig. 6A-3, right panel), with the MUC1-YFP being only detectable in HIC enriched elution fractions (Fig. 6A-3, left panel). Ni-affinity purification of embedded GF(MUC1)P co-expressed with CytoEpi-T2 revealed very little degradation (Fig. 6B).

Both MUC1-YFP (Fig. 6C) and embedded GF(MUC1)P (Fig. 6D) co-expressed with T2-2A-CytoEpi and CytoEpi-T2, respectively, were subjected to Asp-N digestion for release of the MUC1 tandem repeat for subsequent MALDI-TOF-MS analysis. Approximately 30% and 70% of MUC1-YFP (Fig. 6C) and embedded GF(MUC1)P (Fig. 6D), respectively, were O-glycosylated. The dominant peak was found to be non-glycosylated MUC1-1TR with 2 proline hydroxylations (m/z 1918) (Fig. 6C+D). Peaks corresponding to Tn-MUC1-1TR with 1-3 GalNAc attachments and 1-4 proline hydroxylations were observed (Fig. 6C+D). Minor additional products, corresponding to addition of three arabinose residues, were also identified (Fig. 6C).

DISCUSSION

We previously established mammalian GalNAc O-glycosylation capacity transiently in plants. GalNAc O-glycosylation capacity required introduction of a two-
component O-glycosylation machinery basically consisting of a C4-epimerase and a GalNAc-T (Yang et al., 2012). Here we have established O-glycosylation capacity stably in two different plant cell systems: i) tobacco BY-2 suspension cells, which offers potential for secretion of products (Hellwig et al., 2004) and has low reproductive constraints with regards to propagation and manipulation of transgenic lines; and ii) soil grown A. thaliana, which has a comprehensive T-DNA knock-out repertoire (http://www.signal.salk.edu/) enabling, e.g., manipulation of endogenous derived PTMs. A previous report proposed that a three-component machinery (additionally including a UDP-GalNAc-transporter) was required for introduction of O-glycosylation in plants (Daskalova et al., 2010), but our results using transient (Yang et al., 2012) and now also stable introduction of O-glycosylation machineries clearly demonstrate that the UDP-GalNAc-transporter is not required.

We did find unexpected differences in efficiency of the glycosylation machinery designs when introduced stably compared to transiently. The two most efficient designs in the transient implementations (HA-T2-2A-CytoEpi and CytoEpi-2A-T2) failed to express the enzyme proteins and glycosylate reporter substrates in stable implementations (Supplemental Fig. S2A). Surprisingly, removal of the N-terminal Hemagglutinin (HA) tag in HA-T2-2A-CytoEpi markedly improved production of the two functional enzymes in stable implementations (Fig. 2B). Degradation of all 2A-derived polyproteins was observed as evidenced by Western blot analysis (Supplemental Fig. S2B). Also, the one-gene-one-promoter O-glycosylation machinery design (CytoEpi-T2), which in transient implementations proved markedly less efficient than the 2A linked designs, conferred a similar O-glycosylation efficiency as T2-2A-CytoEpi in a stable implementation in BY-2 cells (Fig. 4C, lanes 41-46). There are to our knowledge few reports of stably integrated 2A-linked implementations where the two proteins are destined for different subcellular localizations. Our results suggest that the design of 2A linked polyprotein constructs with respect to the order of proteins may be important.

We previously observed that the MUC1 reporter substrate was somewhat instable when expressed transiently in plants (Yang et al., 2012). Here, we observed even more severe problems with stability of the small MUC1 reporter substrate whether expressed alone (MUC1) or in fusion with the carrier YFP (MUC1-YFP) (Fig. 1 and 3). We also showed that incubation of non-glycosylated and glycosylated MUC1-YFP in BY-2 medium conferred a complete protease mediated breakdown of
MUC1, resulting in the trimmed resistant stable barrel YFP structure. We therefore tested an alternative strategy where we embedded 2.5 tandem repeats of MUC1 with tags into GFP (Fig. 4A) as previously proposed by Kobayashi and co-workers (Kobayashi et al., 2008). These authors demonstrated that 67 amino acid residues could be inserted into a loop between the β-strands located on the opposite side of the N- and C-termini of the GFP protein without substantial loss of fluorescence capacity. Co-expression of the embedded construct GF(MUC1)P with the O-glycosylation machinery CytoEpi-T2 resulted in secretion of GalNAc-glycosylated GF(MUC1)P in the BY-2 medium (Fig. 5A). MALDI-TOF-MS analysis revealed that GF(MUC1)P was uniformly glycosylated with 3 GalNAc residues, which corresponds to the substrate specificity of GalNAc-T2 (Bennett et al., 1998). Approx. 50% of the GF(MUC1)P in the BY-2 medium further contained up to 4 Hyps as well as 3 and 6 arabinose residues (Fig. 5A). Introduction of a single construct for expression of GF(MUC1)P and the O-glycosylation machinery CytoEpi-2A-T2, resulted in moderate accumulation of GF(MUC1)P (~5 µg/g fresh weight) in Arabidopsis (Fig. 6B), which was modified with 0-3 GalNAc residues, 0-4 proline hydroxylations and minor arabinosides (Fig. 6D).

In agreement with our previous transient implementation (Yang et al., 2012), introduction of the mature secreted form of the cytokine IFNa2B in Arabidopsis resulted in expression of the intact product, which was glycosylated when co-expressed with the O-glycosylation machinery CytoEpi-T2 (Fig. 2C). Likewise, a 1.2 TR of MUC16, which also realizes a full structural unit with 3 potential N-glycosylation and up to 30 O-glycosylation sites, was expressed and shown to be N-glycosylated by the endogenous N-glycosylation machinery in BY-2 cells (Supplemental Fig. S1). Explanations for the apparent instability of the small MUC1-3.5TR may include lack of proper folding and of protective O-glycosylation, where native mammalian mucins are large (>200 KDa) heavily O-glycosylated glycoproteins abundant in mucus layers, where they hydrate, lubricate and protect cells from proteases as well as from pathogens. In plants, hundreds of genes encode for proteins involved in proteolytic processes (Rawlings et al., 2008), with an estimated 1900 genes in Arabidopsis being implicated directly or indirectly in peptide bond hydrolysis (Schaller, 2004; Smalle and Vierstra, 2004; Benchabane et al., 2008). The high protease activity in the BY-2 medium (Robertson et al., 1997; Okushima et al., 2000) may for production of some pharmaceutical proteins require tailored knock-out
of specific proteases. It is well established that $N$- and $O$-linked-glycosylation may enhance protein stability, probably by acting as a physical barrier to protease mediated degradation (Li and d'Anjou, 2009). The finding that $O$-linked GalNAc glycosylation protected the MUC1 protein offers an alternative approach to combating protease-mediated degradation.

The plant-specific Hyp $O$-glycosylation is an obvious concern for recombinant production of human therapeutics in plants. Proline hydroxylation 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 (Yuasa et al., 2005). Thirteen putative P4Hs have been identified in Arabidopsis (Showalter et al., 2010), and our understanding of the functions of these in terms of peptide/protein substrate specificities and sequence contexts (which proteins and where) is still limited. This is especially true for transgene expression of proteins. Hyps for example were identified on human erythropoietin (Weise et al., 2007) and IgA1 (Karnoup et al., 2005) recombinantly expressed in Physcomitrella and maize, respectively, and Hyp-linked $O$-glycosylation were additionally identified on IgA1, specifically in the Pro/Ser/Thr rich hinge region where GalNAc $O$-glycosylation occur in man (Karnoup et al., 2005). More recently, Pinkhasov et al. (2011) found Hyp formation ($P^{11}$ or $P^{12}$) and Hyp substituted with three Ara/f residues in a MUC1 tandem repeat transiently expressed in plants. This finding was confirmed in our studies with an additional interesting observation, namely that expression of MUC1 from stably integrated constructs markedly increased the level of proline hydroxylation and subsequent Hyp linked arabinosides. Lamport and co-workers recently summarized plant Hyp-$O$-glycosylation, with non-contiguous Hyp residues being glycosylated exclusively with arabinogalactan polysaccharides, and contiguous Hyp residues being glycosylated exclusively with arabinosides (Kieliszewski et al. 2011). We demonstrated that substantial reduction in Hyp modifications could be achieved by addition of the 2,2 DP inhibitor but, clearly, targeted knock-out of one of more P4Hs acting on the expressed proteins of interest would be more efficient. Several reports suggest involvement of the three P4Hs, AtP4H2, AtP4H5, and AtP4H13, in hydroxylation of contiguous Pro motifs in particular (Hieta and Myllyharju, 2002; Tiainen et al., 2005; Keskiaho et al., 2007; Velasquez et al., 2011), perhaps pointing to these enzymes as first targets for engineering plants with knock-
out of plant-specific proline hydroxylation and Hyp-linked arabinosylation. The clinical significance of proline hydroxylations and Hyp-linked arabinosylations is unknown or at best speculative (Petersen et al., 1998; Leonard et al., 2005; Altmann, 2007; Manduzio et al., 2012, Pinkhasov et al. 2011). In relation to the present glycoengineering, these endogenous derived PTMs did not appear to affect the GalNAc T2 mediated O-glycosylation of MUC1.

Our interest in designing plant cell factories with engineered human style O-glycosylation stems from a need for production of MUC1 and other vaccines with aberrant O-glycosylation. We have previously identified cancer-associated immunodominant GalNAc-glycopeptide epitopes in MUC1 mucin that are not covered by immunological tolerance (Sorensen et al., 2006; Sabbatini et al., 2007; Tarp et al., 2007). We have also shown that chemoenzymatically produced Tn-MUC1 glycopeptides induce IgG antibodies in cancer patients (Sabbatini et al., 2007), and spontaneous IgG antibodies to the same epitope are found in some cancer patients at the time of diagnosis (Wandall et al., 2010; Blixt et al., 2011). The absence of GalNAc-type O-glycosylation makes plants an ideal cell system for engineering O-glycosylation de novo and production of aberrant O-glycosylation glycoforms as shown in this study.

The absolute requirement of a carrier protein for stabilization of the small MUC1-2.5TR or MUC1-3.5TR proteins, which in the BY-2 cell system needed to be embedded into the carrier GFP and glycosylated, plus the add-on presence of endogenous PTMs, render production of this particular small vaccine candidate in plants non-feasible or at best cumbersome. In summary, stably engineered plants with the capacity to perform mammalian O-glycosylation on recombinantly expressed secreted target proteins were established. Problems with degradation and plant specific posttranslational modifications were identified, highlighting the need for additional engineering for plants to become versatile production platforms of therapeutics. Although in its infancy, engineering O-glycosylation capacities in plants may eventually be combined with engineered human N-glycosylation (Strasser et al., 2008; Kaulfurst-Soboll et al., 2011; Nagels et al., 2011; Schoberer and Strasser, 2011), to generate more general platforms for expression of glycoprotein pharmaceuticals in plants.
MATERIALS AND METHODS

Transformation and Plant Growth Condition

Bright Yellow 2 (BY-2) suspension cells were cultivated by shaking (25°C, 130 rpm) in the dark. Transformation and fermentation of BY-2 cell are described in (Mayo et al., 2006). *Agrobacterium tumefaciens* strain C58C1 pGV3850 was used for both stable transformation and transient agrobacterium mediated expression. Transformation of *Arabidopsis thaliana* was performed as described (Horsch et al., 1986). Growth conditions and transformation of transgenic *Arabidopsis thaliana* plants were performed as described (Egelund et al., 2007). Transformation and maintenance of *Lemna minor* were done as described (Rival et al., 2008). Agrobacterium mediated expression was done essentially as described in our previous study (Yang et al., 2012).

DNA Constructs for Plant Transformation

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 is an intermediate *E. coli* only vector (Odell et al., 1985), 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 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.

Designs of expression constructs used are depicted in Fig. 1. Secreted reporter constructs for O-glycosylation were designed with either of the following signal peptide sequences: i) *Nt*SP derived from *Nicotiana tabacum* proline-rich protein 3 (UniProt accession number Q40502: Sequence: MGKMASLFAASLLVVLVSLSLA); or ii) *Pp*SP derived from *Physcomitrella patens* aspartic protease (EMBL accession number AJ586914: Sequence: MGASRSVRLAFFLVLVVLVLAALAEA), and these were generated as described previously (Yang et al., 2012). A novel construct with 2.5 tandem repeat (TR) sequence of MUC1 embedded in GFP (GF(MUC1)P) was prepared as follows (Fig. 1). The MUC1-2.5TR sequence was inserted into GFP.
protein after Asp196 into the NcoI site of modified GFP (MfGFP) (Kobayashi et al., 2008), followed by PCR mediated insertion of the signal peptide for secretion. The primer sets were SSGFPfor (5’-gagctccatggtaagactaatctttttctttctatctgtcctttcctctctatattgctggcggacgaagcaggagggggaggt-3’) and GFPrev (5’-catatgttggctcagctctctactcattg-3’) and the resulting PCR fragment were cloned into SacI site of pPS48 under control of the Ubi promoter and the Ubi terminator. Finally the whole expression cassette was inserted into HindIII site of pCAMBIA2300, yielding pC2300D-UbiPro-NtSP-GF(MUC1-2.5TR)P-UbiTerm (GF(MUC1)P).

Design of the O-glycosylation machinery was essentially as described previously (Yang et al., 2012) (Fig. 1A and C), and consisted of a cytosolic Pseudomonas auriginosa Glc(NAc) C4-epimerase (CytoEpi) WbpP (Creuzenet et al., 2000) (GenBank accession number AAF23998.1) and human polypeptide N-acetylgalactosaminytransferase 2 (GalNAc-T2) (UniProt accession number X85019). The O-glycosylation machinery was implemented from a single construct either with each gene driven by a single promoter or a polycistronic transcript encoding two genes interspaced by nucleotides encoding the FMDV 2A auto cleavable sequence (El Amrani et al., 2004; Szymczak et al., 2004). Construction was carried out as follows: Constructs encoding the epimerase and the GalNAc-T2 were assembled by inserting the XbaI-35SPro-CytoEpi-35STerm-XbaI of construct CytoEpi into the XbaI site of construct T2, resulting in pC1302D-35SPro-T2-35STerm;35SPro-CytoEpi-35STerm (CytoEpi-T2). This construct was also made as a 2A-linked polycistronic construct. A modified version of pC1302-35SPro-HA-T2-2A-CytoEpi-35STerm (HA-T2-2A-CytoEpi) (Yang et al., 2012) without a HA-tag at N-terminus of GalNAc-T2 was also prepared as follows: T2-2A-CytoEpi was cloned by PCR using T2For (5’-ccatgcctgctgcggcggcgtcggcgggaggt-3’) and PFwbpRev (5’-gagctctcaactgtcgtcgctcgctcgctctctctttaaaagcgttgaacag-3’) primer set, followed by insertion into SacI site of pPS48 vector. The entire fragment containing 35SPro-T2-2A-CytoEpi-35STerm was inserted into the HindIII site of pCAMBIA1302, yielding pC1302-35SPro-T2-2A-CytoEpi-35STerm (T2-2A-CytoEpi). The 2A linker region of T2-2A-CytoEpi, encodes GSGQTLNFDLLKLADVESNPG↓PM, where italics designate the 2A sequence and the arrow indicates the auto-cleavage site (El Amrani et al., 2004; Szymczak et al., 2004). The XbaI digested fragment of 35SPro-PpSP-MUC1-3.5TR-YFP(His)↓-35STerm was inserted into T2-2A-CytoEpi, yielding pC1302-35SPro-
Preparation of Total Protein Extracts

Approximately one gram of freshly harvested leaves or roots 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 Imidazol, 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 (Petersen et al., 2009). Monoclonal antibodies (MAbs) to T7-tag and MUC16 (M11) were obtained from Invitrogen and Dako, respectively. MAbs to MUC1 and GalNAc-MUC1 (Tn-MUC1), 5E10 and 5E5, respectively, were described previously (Tarp et al., 2007). HRP-conjugated Vicia villosa lectin (VVA; EY Laboratories) was also used for detection of Tn glycoforms.

Hydrophobic Interaction Chromatography

MUC1-YFP was purified by hydrophobic interaction chromatography (HIC) using ÄKTA FPLC System with a HiTrap HIC column (5 ml, GE), pre-packed with Phenyl Sepharose 6 Fast Flow. BY-2 cell medium fraction (100 ml), additionally adjusted to include 3 M NaCl and 20 mM Tris-buffer (pH 8.0), was centrifuged 25,000 x g for 20 min. The supernatant was loaded (2ml/min) to the pre-equilibrated HIC column and then washed with 20 ml wash buffer, i.e. 2 M NaCl in 20 mM Tris-buffer (pH 8.0). MUC1-YFP was eluted by water in 10 fractions of 2 ml.

Ni-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 (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).

**Degradation Study of MUC1-YFP and embedded GF(MUC1)P in BY-2 medium**

Culture medium (50 ml) of wild type BY-2 cells grown for 7 days was collected by centrifugation 10,000 g for 5 min (wt BY-2 medium fraction). Ca. 5 µg purified MUC1-YFP (transiently produced in leaves of *N. benthamiana* plants with/without O-glycosylation machinery) and embedded GF(MUC1)P (purified from intracellular or extracellular of BY-2 cells with/without O-glycosylation machinery), was added to 7 day old 1 ml of un-boiled or boiled (for 15 min) wt BY-2 medium fractions, which was further incubated at 25°C under shaking (130 rpm) for an additional 24 h or 48h under the same conditions and ultimately subjected to MUC1 and Tn-MUC1 specific (MAb 5E10 and 5E5, respectively) Western blot analysis.

**Endo-Asp Digestion of MUC1-TR and Sample Purification**

Purified MUC1-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 under shaking (600 rpm). In the case of embedded GF(MUC1)P, in-gel digestion (Shevchenko et al., 2006) was used to release MUC1 tandem repeats from the GFP protein, as in-gel protein digestion of embedded GF(MUC1)P was found to give a better recovery compared to the ‘in-solution’ method. Both Asp-N (10 ng/µl) and trypsin (10 ng/µl) were added into the digestion buffer (25 mM NH₄HCO₃, 10% acetonitrile (ACN)). The digest was cleaned on a C18 Zip-Tip (Millipore). Briefly, the digestion mixture was dissolved in 20 µL 0.1% trifluoracetic acid (TFA) and drawn through the column, desalted with 0.5% formic acid, and eluted with 100% ACN.

**Matrix-Assisted Laser-Desorption Time of Flight Mass Spectrometry (MALDI-TOF-MS)**

Lyophilized peptides were dissolved in 20 µl water. The MALDI matrix 2,5-dihydroxybenzoic acid (Sigma-Aldrich) was dissolved (25 g/L) in a 1:1 mixture of water and methanol. Samples were prepared for analysis by placing 0.5 µl of sample solution on a MALDI target, followed by 0.5 µl of matrix solution which was then air dried. All mass spectra were acquired in the linear mode on a Voyager-Elite MALDI
time of flight mass spectrometer (Perseptive Biosystem Inc., Framingham, MA) equipped with delayed extraction; external calibration was used.

Acknowledgements

pET23-WbpP was a generous gift from JS. Lam. Takashi Murayama, Juntendo University School of Medicine, is acknowledged for sharing the construct MfGFP. Dr. Jose Estevez, Universidad de Buenos Aires, Argentina, is thanked for helpful suggestions with relation to proline hydroxylation inhibition experiments.
LITERATURE CITED


Keskihao K, Hieta R, Sormunen R, Myllyharju J (2007) Chlamydomonas reinhardtii has multiple prolyl 4-hydroxylases, one of which is essential for proper cell wall assembly. The Plant cell 19: 256-269


fluorescent protein tag for a wide variety of protein research. PLoS One 3: e3822


Schoberer J, Strasser R (2011) Sub-compartmental organization of Golgi-resident N-glycan processing enzymes in plants. Molecular plant 4: 220-228


Figure Legends:

**Figure 1.** Construct designs for engineering O-glycosylation. (A) Depiction of O-glycosylation machinery construct designs. Designs included Flag-tagged cytoplasmic targeted *Pseudomonas aeruginosa* C4-epimerase WbpP (CytoEpi) and Golgi targeted polypeptide N-acetyl-galactosaminyltransferase 2 (T2) expressed as a single polycistronic-protein interspaced by the 2A self-cleaving sequence (cleavage site indicated by an arrow) (T2-2A-CytoEpi), or as separate transcriptional units (CytoEpi-T2). (B) Depiction of O-glycosylation reporter constructs used. MUC1-3.5TR (MUC1) with and without a C-terminal fusion to YFP (MUC1-YFP), MUC1 was embedded into GFP protein (GF(MUC1)P). Full coding IFNα2B cytokine with a C-terminal T7 and (SP)10 glycomodule IFN-SP10. All reporter constructs included N- or C-terminal (His)6 tags and N-terminal signal peptide sequences (SS). (C) Depiction of a single combined construct with reporter MUC1-YFP and the 2A linked O-glycosylation machinery under control of two promoters (MUC1-YFP-T2-2A-CytoEpi).

**Figure 2.** Implementation of O-glycosylation in *Arabidopsis thaliana* and Tobacco Bright Yellow 2 (BY-2) suspension cells. (A) SDS-PAGE Western blot analysis of four *A. thaliana* lines expressing either MUC1 (lanes 1-2) or MUC1-YFP (lanes 4-5) alone. Contrast of the picture was adjusted to visualize very weak bands for MUC1 (lanes 1-2). (B) SDS-PAGE Western blot analysis of two tobacco BY-2 suspension cell lines expressing MUC1-YFP alone (lane 1) and MUC1-YFP together with the 2A linked O-glycosylation machinery T2-2A-CytoEpi (lane 2). Absence (-) or presence (+) of O-glycosylation machinery is indicated above the lanes. (C) An *A. thaliana* line transgenic for T7-tagged full coding secreted IFNα2B expressed alone (-) or co-expressed with the O-glycosylation machinery CytoEpi-T2 (+). Glycosylation of His-tag purified IFNα2B was detected by VVA lectin blot analysis. Total protein extracts from transgenic *A. thaliana* leaves or BY-2 cells callus were loaded and blots reacted with MUC1 specific MAbs 5E10 and 5E5, where 5E5 is specific for GalNAc-glycosylated MUC1 (Tn-MUC1) and does not react with non-glycosylated MUC1. Approximately 30 μg total protein was loaded in each lane.
Figure 3. MUC1-YFP is expressed and glycosylated in BY-2 cells, but MUC1 is degraded in the medium. (A) UV light picture of fractions (E1 through E8) from Hydrophobic interaction chromatography (HIC) of MUC1-YFP secreted from BY-2 cells co-expressing T2-2A-CytoEpi, (B) SDS-PAGE Coomassie stained analysis of fractions, (C) Western blot with anti-GFP Ab, and (D) with anti-MUC1 (5E10). 10 μl of 2 ml eluate fractions were applied in each lane.

Figure 4. O-glycosylation and embedding in GFP stabilize MUC1 in BY-2 cell culture. (A) Construct design for embedding MUC1-2.5TR into GFP (GF(MUC1)P) and corresponding barrel structure showing the loop into which MUC1-2.5 TR flanked with (His)$_8$ and C-Myc tags were inserted after Asp196 in the loop between the β-strands (blue) located opposite to the N- and C-termini. The figure was adapted from Kobayashi et al. 2008. (B) Fluorescent microscopy of a stable BY-2 transgenic line co-expressing GF(MUC1)P with O-glycosylation machinery CytoEpi-T2 (left panel), and SDS-PAGE Coomassie staining of Ni-chromatography purified secreted GalNAc glycosylated GF(MUC1)P (Tn-GF(MUC1)P) of this line (right panel). (C) Analysis of degradation of MUC1-YFP and GF(MUC1)P in BY-2 cell culture medium by SDS-PAGE Western blotting; 1) MUC1-YFP (lanes 1-10) and GalNAc glycosylated MUC1-YFP (Tn-MUC1-YFP) (lanes 11-20) transiently produced in leaves of N. benthamiana plants; 2) intracellular embedded GF(MUC1)P (lanes 21-30) purified from a stably transformed BY-2 cell line expressing GF(MUC1)P alone; 3) intracellular (lane 31-40) and extracellular (lane 41-50) embedded (Tn)-GF(MUC1)P purified from a transgenic BY-2 cell line additionally co-expressing CytoEpi-T2. The isolated proteins were added to 7 day-old un-boiled (-) or boiled (+) wt BY-2 medium fractions, which were further incubated for up to 24 h under the same conditions. Approximately 5 μg purified proteins were added to 1 ml BY-2 medium fractions, respectively, of which ca. 15 μl were loaded and blots probed with anti-MUC1 MAb 5E10. The corresponding Tn-MUC1 specific Western blot using MAb 5E5 is presented in Supplemental Fig. S3.

Figure 5. Inhibition of proline hydroxylation of MUC1 in BY-2 cells. (A) MALDI-TOF analysis of endo-Asp released MUC1 tandem repeat (DTRPAPGSTAPPAHGVTSA, indicated by arrow) from Tn-GF(MUC1)P expressed in BY-2 cells co-expressing CytoEpi-T2. The analysis revealed the
presence of 3 moles of attached GalNAc, i.e. complete GalNAc-T2 mediated GalNAc occupancy of sites, with varying proline hydroxylations of up to 4 moles per tandem repeat. Hydroxylated MUC1 peptides were further modified by attachment of mainly by 3-, and less frequently by 6 arabinose residues. (B) MALDI-TOF-MS analysis of MUC1 tandem repeat (indicated by arrow) of GF(MUC1)P co-expressed with CytoEpi-T2 in BY-2 cells grown in the presence of prolyl 4-hydroxylase (P4H) inhibitor. Growth medium was exchanged at day 7 with fresh medium including 200 µM of 2,2-Dipyridyl (2,2 DP) followed by incubation at similar growth conditions.

**Figure 6.** MUC1-YFP and GF(MUC1)P expressed with or without O-glycosylation machinery in *Arabidopsis thaliana*. (A) SDS-PAGE analysis of a transgenic *A. thaliana* line expressing a combined construct (MUC1-YFP-T2-2A-CytoEpi), comprising both the reporter MUC1-YFP and the O-glycosylation machinery T2-2A-CytoEpi. MUC1 specific and Tn-MUC1 specific Western blot analysis of total leaf extracts (1), MUC1-YFP purified from leaves (2) and roots (3) by Hydrophobic interaction chromatography (HIC) as visualized by coomassie stained SDS-PAGE and MUC1 specific MAb 5E10 Western blots of the eluates. (B) His-tag purification of embedded GF(MUC1)P co-expressed with O-glycosylation machinery CytoEpi-T2 as visualized by coomassie stained SDS-PAGE and MUC1 specific MAb 5E10 Western blot analysis (C) MALDI-TOF-MS analysis of MUC1 tandem repeat (DTRPAPGSTAPPAHGVTSAP, indicated by arrow) from His-tag purified Asp-N digested MUC1-YFP, and (D) GF(MUC1)P co-expressed with O-glycosylation machineries T2-2A-CytoEpi and CytoEpi-T2, respectively.
Figure 1. Construct designs for engineering O-glycosylation. (A) Depiction of O-glycosylation machinery construct designs. Designs included Flag-tagged cytoplasmic targeted Pseudomonas aeruginosa C4-epimerase WbpP (CytoEpi) and Golgi targeted polypeptide N-acetyl-galactosaminytransferase 2 (T2) expressed as a single polycistronic-protein interspaced by the 2A self-cleaving sequence (cleavage site indicated by an arrow) (T2-2A-CytoEpi), or as separate transcriptional units (CytoEpi-T2). (B) Depiction of O-glycosylation reporter constructs used. MUC1-3.5TR (MUC1) with and without a C-terminal fusion to YFP (MUC1-YFP), MUC1 was embedded into GFP protein (GF(MUC1)p). Full coding IFNα2B cytokine with a C-terminal T7 and (SP)10 glycomodule IFN-SP10. All reporter constructs included N- or C-terminal (His)6 tags and N-terminal signal peptide sequences (SS). (C) Depiction of a single combined construct with reporter MUC1-YFP and the 2A linked O-glycosylation machinery under control of two promoters (MUC1-YFP-T2-2A-CytoEpi).
Figure 2. Implementation of O-glycosylation in Arabidopsis thaliana and Tobacco Bright Yellow 2 (BY-2) suspension cells. (A) SDS-PAGE Western blot analysis of four A. thaliana lines expressing either MUC1 (lanes 1-2) or MUC1-YFP (lanes 4-5) alone. Contrast of the picture was adjusted to visualize very weak bands for MUC1 (lanes 1-2). (B) SDS-PAGE Western blot analysis of two tobacco BY-2 suspension cell lines expressing MUC1-YFP alone (lane 1) and MUC1-YFP together with the 2A linked O-glycosylation machinery T2-2A-CytoEpi (lane 2). Absence (-) or presence (+) of O-glycosylation machinery is indicated above the lanes. (C) An A. thaliana line transgenic for T7-tagged full coding secreted IFNα2B expressed alone (-) or co-expressed with the O-glycosylation machinery CytoEpi-T2 (+). Glycosylation of His-tag purified IFNα2B was detected by VVA lectin blot analysis. Total protein extracts from transgenic Arabidopsis leaves or BY-2 cells callus were loaded and blots reacted with MUC1 specific MAb 5E10 and 5E5, where 5E5 is specific for GalNAc-glycosylated MUC1 (Tn-MUC1) and does not react with non-glycosylated MUC1. Approximately 30 µg total protein was loaded in each lane.
Figure 3. MUC1-YFP is expressed and glycosylated in BY-2 cells, but MUC1 is degraded in the medium. (A) UV light picture of fractions (E1 through E8) from Hydrophobic interaction chromatography (HIC) of MUC1-YFP secreted from BY-2 cells co-expressing T2-2A-CytoEpi, (B) SDS-PAGE Coomassie stained analysis of fractions, (C) Western blot with anti-GFP Ab, and (D) with anti-MUC1 (5E10). 10 µl of 2 ml eluate fractions were applied in each lane.
Figure 4. O-glycosylation and embedding in GFP stabilize MUC1 in BY-2 cell culture. (A) Construct design for embedding MUC1-2.5TR into GFP (GF(MUC1)P) and corresponding barrel structure showing the loop into which MUC1-2.5 TR flanked with (His)$_8$ and C-Myc tags were inserted after Asp196 in the loop between the $\beta$-strands (blue) located opposite to the N- and C-termini. The figure was adapted from Kobayashi et al. 2008. (B) Fluorescent microscopy of a stable BY-2 transgenic line co-expressing GF(MUC1)P with O-glycosylation machinery CytoEpi-T2 (left panel), and SDS-PAGE Coomassie staining of Ni-chromatography purified secreted GalNAc glycosylated GF(MUC1)P (Tn-GF(MUC1)P) of this line (right panel). (C) Analysis of degradation of MUC1-YFP and GF(MUC1)P in BY-2 cell culture medium by SDS-PAGE Western blotting; 1) MUC1-YFP (lanes 1-10) and GalNAc glycosylated MUC1-YFP (Tn-MUC1-YFP) (lanes 11-20) transiently produced in leaves of N. benthamiana plants; 2) intracellular embedded GF(MUC1)P (lanes 21-30) purified from a stably transformed BY-2 cell line expressing GF(MUC1)P alone; 3) intracellular (lane 31-40) and extracellular (lane 41-50) embedded (Tn)-GF(MUC1)P purified from a transgenic BY-2 cell line additionally co-expressing CytoEpi-T2. The isolated proteins were added to 7 day-old un-boiled (-) or boiled (+) wt BY-2 medium fractions, which were further incubated for up to 24 h under the same conditions. Approximately 5 µg purified proteins were added to 1 ml BY-2 medium fractions, respectively, of which ca. 15 µl were loaded and blots probed with anti-MUC1 MAb 5E10. The corresponding Tn-MUC1 specific Western blot using MAb 5E5 is presented in Supplemental Fig. S3.
Figure 5. Inhibition of proline hydroxylation of MUC1 in BY-2 cells. (A) MALDI-TOF analysis of endo-Asp released MUC1 tandem repeat (DTRPAPGSTAPPAGVTSAP, indicated by arrow) from Tn-GF(MUC1)P expressed in BY-2 cells co-expressing CytoEpi-T2. The analysis revealed the presence of 3 moles of attached GalNAc, i.e. complete GalNAc-T2 mediated GalNAc occupancy of sites, with varying proline hydroxylations of up to 4 moles per tandem repeat. Hydroxylated MUC1 peptides were further modified by attachment of mainly by 3-, and less frequently by 6 arabinose residues. (B) MALDI-TOF-MS analysis of GF(MUC1)P co-expressed with CytoEpi-T2 in BY-2 cells grown in the presence of prolyl 4-hydroxylase (P4H) inhibitor. Growth medium was exchanged at day 7 with fresh medium including 200 µM of 2,2-Dipyridyl (2,2 DP) followed by incubation at similar growth conditions.
Figure 6. MUC1-YFP and GF(MUC1)P expressed with or without O-glycosylation machinery in Arabidopsis thaliana. (A) SDS-PAGE analysis of a transgenic A. thaliana line expressing a combined construct (MUC1-YFP-T2-2A-CytoEpi), comprising both the reporter MUC1-YFP and the O-glycosylation machinery T2-2A-CytoEpi. MUC1 specific and Tn-MUC1 specific Western blot analysis of total leaf extracts (1), MUC1-YFP purified from leaves (2) and roots (3) by Hydrophobic interaction chromatography (HIC) as visualized by coomassie stained SDS-PAGE and MUC1 specific MAb 5E10 Western blots of the eluates. (B) His-tag purification of embedded GF(MUC1)P co-expressed with O-glycosylation machinery CytoEpi-T2 as visualized by coomassie stained SDS-PAGE and MUC1 specific MAb 5E10 Western blots of the eluates. (C) MALDI-TOF-MS analysis of MUC1 tandem repeat (DTRPAPGSTAPPAGVTSAP, indicated by arrow) from His-tag purified Asp-N digested MUC1-YFP, and (D) GF(MUC1)P co-expressed with O-glycosylation machineries T2-2A-CytoEpi and CytoEpi-T2, respectively.