Toward Stable Genetic Engineering of Human O-Glycosylation in Plants[^C][^W][^OA]

Zhang Yang, Eric P. Bennett, Bodil Jørgensen, Damian P. Drew[^2], Emma Arigi, Ulla Mandel, Peter Ulvskov, Steven B. Levy, Henrik Clausen, and Bent L. Petersen[^*]

Department of Molecular Biology and Genetics, Faculty of Science and Technology, Aarhus University, Flakkebjerg, 4200 Slagelse, Denmark (Z.Y.); Department of Plant Biology and Biotechnology, Faculty of Science, University of Copenhagen, 1871 Frederiksberg C, Denmark (D.P.D., P.U., B.L.P.); Department of Agriculture and Ecology, Faculty of Science, University of Copenhagen, 1871 Frederiksberg C, Denmark (B.J.); and Center for Glycomics, Departments of Cellular and Molecular Medicine and School of Dentistry, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen N, Denmark (Z.Y., E.P.B., E.A., U.M., S.B.L., H.C.)

Glycosylation is the most abundant and complex posttranslational modification to be considered for recombinant production of therapeutic proteins. Mucin-type (N-acetylgalactosamine [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 the synthesis of the donor substrate UDP-GalNAc and one or more polypeptide GalNAc-transferases for incorporating GalNAc residues into proteins. Here, we have stably engineered O-glycosylation capacity in two plant cell systems, soil-grown Arabidopsis (Arabidopsis thaliana) and tobacco (Nicotiana tabacum) Bright Yellow-2 suspension culture cells. Efficient GalNAc O-glycosylation of two stably coexpressed 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 Bright Yellow-2 cells. In summary, stably engineered mammalian type O-glycosylation was established in transgenic plants, demonstrating that plants may serve as host cells for the production of recombinant O-glycoproteins. However, the present stable implementation further strengthens the notion that elimination of endogenous posttranslational modifications may be needed for the production of protein therapeutics.

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 the design and optimization of host cells aimed for recombinant expression of biologics. Significant success has been achieved with designed N-glycosylation in 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., 2012). O-Glycosylation is initiated by a family of up to 20 UDP-GalNAc polypeptide N-acetylgalactosaminytransferases (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 the intent, for example, to target proteins to antigen-presenting cells and augment immunity in vaccine strategies. Therefore, it may be desirable to build glycosylation capacity in host cells de novo in order to attain complete design freedom without an extensive

[^C]: This work was supported by the Danish Council for Strategic Research (grant no. 09–063090), the Danish Council for Independent Research (grant no. 09–066624/274–09–0314), a program of excellence from the University of Copenhagen, and the Danish National Research Foundation.

[^W]: Present address: Wine Science and Business, School of Agriculture, Food, and Wine, University of Adelaide, Urrbrae, South Australia 5064, Australia.

[^OA]: Open Access articles can be viewed online without a subscription.

[^1]: The online version of this article contains Web-only data.

[^2]: The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Bent L. Petersen (blp@life.ku.dk).

[^C]: Some figures in this article are displayed in color online but in black and white in the print edition.

[^OA]: Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.112.198200
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 (Cricetus griseus) ovary (CHO) cells (Durocher and Butler, 2009). CHO cells produce monosialylated 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 (Homo sapiens) O-glycosylation from scratch (Amano et al., 2008).

However, since endoplasmic reticulum-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 coculturing in rhodamine-3-acetic acid, but complete elimination of O-mannosylation is presumably not possible, as knockout 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 the 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 glycan attachment and glycan structures (Sasaki et al., 1987; 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 the custom design and buildup of O-glycosylation from scratch with respect to both sites of O-glycan attachment and O-glycan structures.

Daskalova et al. (2010) initially started engineering plants for mammalian O-glycosylation and suggested that GalNAc O-glycosylation in Nicotiana benthamiana plants required a Glc(Nac) C4-epimerase, a UDP-Gal (Nac) transporter, and a human GalNAc-T. We subsequently tested a variety of design strategies for the 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 the 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 the introduction of a UDP-GalNAc transporter is not fully clarified, but the study by Daskalova et al. (2010) was solely based on lectin labeling, and the authors subsequently demonstrated that the lectins used detect endogenous plant proteins (Reaves et al., 2011), causing serious concerns with regard to the original findings. Our study included the 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 subcellular substrate pools are increasingly difficult to control as the engineering effort increases in complexity. Stable engineering of especially the glycosylation machineries is thus generally needed at an early point, as seen in the engineering efforts to humanize plants with respect to N-linked protein glycosylation (Strasser et al., 2008; Castilho et al., 2010; Kaulfürst-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, soil-grown Arabidopsis (Arabidopsis thaliana) and tobacco (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 the introduction of a UDP-Gal(NAc) transporter in Arabidopsis and tobacco BY-2 suspension cells. We also found that Pro hydroxylation and Hyp-linked arabinosylation produced by the endogenous plant PTM machinery were considerably more pronounced in stable engineered cells compared with our transiently engineered cells. Interestingly, for suspension cell cultures, the presence of a P4H 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 *N. benthamiana*, we identified optimal two-component GalNAc O-glycosylation machinery designs (Fig. 1), which conferred efficient O-glycosylation of coexpressed 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 polypeptide 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 the following: (1) the approximately 11-kD 3.5-tandem repeat (TR) sequence of the cancer-associated mucin MUC1 (MUC1-3.5TR) N terminally fused to a His6 and T7 tag (MUC1); (2) MUC1-3.5TR C terminally fused to yellow fluorescent protein (YFP; MUC1-YFP);...
(3) MUC1-2.5TR embedded into GFP [GF(MUC1)P]; (4) the full-length secreted IFNa2B cytokine tagged with a T7 tag and an AGP type SP_{10} glycomodule (IFN-SP_{10}; Fig. 1B); and (5) 1.2TR of MUC16 tagged with T7 and His tags (MUC16; Supplemental Fig. S1B). All reporter constructs included an N-terminal signal peptide sequence for direction into the secretory pathway.

We first tested the expression and stability of reporter substrates. Expression of the small approximately 11-kD MUC1 3.5TR (MUC1) reporter in Arabidopsis and water-grown Lemma minor conferred only weak accumulation (Fig. 2A, lanes 1 and 2; 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 an approximately 20-fold increased accumulation in Arabidopsis (Fig. 2A, lanes 4 and 5) as compared with expression of MUC1 alone (Fig. 2A, lanes 1 and 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 coexpression of the O-glycosylation machinery T2-2A-CytoEpi, as evidenced by western-blot analysis using a GalNAc-MUC1 glycoform (Tn-MUC1)-specific monoclonal antibody (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 three potential N-glycosylation sites and up to 30 O-glycosylation sites, and peptide N-glycosidase F treatment of MUC16-containing medium fractions resulted in mobility shifts to an apparent slightly lower molecular weight 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 IFNa2B tagged with a SP_{10} glycomodule (IFN-SP_{10}) was also tested as an O-glycosylation substrate in Arabidopsis. Coexpression with the O-glycosylation machinery CytoEpi-T2 resulted in a mobility shift and reactivity with the Vicia villosa (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 the 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 TR from approximately 5% to approximately 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 in the BY-2 medium fractions prompted us to look further into the fate of MUC1-YFP. A BY-2 transgenic cell line stably expressing MUC1-YFP and the O-glycosylation

---

**Figure 2.** Implementation of O-glycosylation in Arabidopsis and tobacco BY-2 suspension cells. A, SDS-PAGE western-blot analysis of four Arabidopsis lines expressing either MUC1 (lanes 1 and 2) or MUC1-YFP (lanes 4 and 5) alone. The contrast of the image was adjusted to visualize very weak bands for MUC1 (lanes 1 and 2). Lane 3 shows results for the wild type ( WT). 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). The absence (−) or presence (+) of O-glycosylation machinery is indicated above the lanes. C, An Arabidopsis line transgenic for T7-tagged full-coding secreted IFNa2B expressed alone (−) or coexpressed with the O-glycosylation machinery CytoEpi-T2 (+). The glycosylation of His tag-purified IFNa2B was detected by VVA lectin-blot analysis. Total protein extracts from transgenic Arabidopsis leaves or BY-2 cell 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 nonglycosylated MUC1. Approximately 30 μg of total protein was loaded in each lane.
machinery T2-2A-CytoEpi failed to produce detectable levels of the MUC1-YFP protein (37 kD) 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-kD protein on SDS-PAGE gels and reacted with a GFP polyclonal antibody (Fig. 3, B and C) but not with a MUC1-specific MAb (Fig. 3D). The yield was estimated to be approximately 25 mg L⁻¹, as judged from Coomassie blue-stained SDS-PAGE gels (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 SP10 glycomodule (Xu et al., 2010) to the C terminus of MUC1 also failed to confer the 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 His₈ tag and N terminally with a Myc tag was embedded in the loop between Asp-196 and Gly-197 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 nickel-affinity chromatography was now feasible, and the isolated product migrated as a single band (approximately 40 kD) as judged by Coomassie blue-stained SDS-PAGE gels (Fig. 4B, right panel), suggesting that MUC1 was stabilized by embedding it into GFP. Endoproteinase 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 three GalNAc residues (Fig. 5A). However, evidence was also found for the modification of Pro to Hyp and of Hyp with three (approximately 45%) and less frequently six (approximately 5%) Ara residues attached (Fig. 5A). The MUC1 repeat contains five Pro residues, of which up to four 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-d-old BY-2 wild-type 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 nonboiled medium after 24 h of incubation (Fig. 4C, lanes 41–45), as evidenced by western-blot analysis with the MUC1-specific MAb 5E10. For the corresponding Tn-MUC1 glycoform-specific western blot, see Supplemental Figure 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 Pro hydroxylase activity were selected (ethyl-3,4-dihydroxybenzoate, 3,4-dehydroproline, 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 Pro hydroxylation of MUC1 TRs and the accompanying attachment of pentose sugars (Fig. 5B; 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 Pro
hydroxylase inhibitors conferred a minor reduction in the Pro hydroxylation level, from approximately 20% to approximately 10%, without an accompanying reduced level of expressed MUC1-YFP (Supplemental Fig. S7).

MUC1-YFP Is Glycosylated and Hyp Is 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 Tr-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 blue staining of purified MUC1-YFP from Arabidopsis leaves revealed that the

Figure 5. Inhibition of Pro hydroxylation of MUC1 in BY-2 cells. A, MALDI-TOF-MS analysis of endoproteinase Asp-released MUC1 tandem repeats (DTRPAGSTAPPAHGVTSAP; indicated by arrows) from Tr-GF(MUC1)P expressed in BY-2 cells coexpressing CytoEpi-T2. The analysis revealed the presence of 3 mol of attached GalNAc (i.e. complete GalNAc-T2-mediated GalNAc occupancy of sites), with varying Pro hydroxylations of up to 4 mol per tandem repeat. Hydroxylated MUC1 peptides were further modified by the attachment of mainly three, and less frequently of six, Ara residues. B, MALDI-TOF-MS analysis of MUC1 tandem repeats (indicated by arrows) of GF(MUC1)P coexpressed with CytoEpi-T2 in BY-2 cells grown in the presence of a P4H inhibitor. Growth medium was exchanged at day 7 with fresh medium including 200 μM 2,2-DP, followed by incubation in similar growth conditions. [See online article for color version of this figure.]
major product was YFP without the MUC1 entity (Fig. 6A, 2). MUC1-YFP of the roots was almost completely degraded, as judged by Coomassie blue-stained SDS-PAGE gels (Fig. 6A, 3, right panel), with the MUC1-YFP being only detectable in HIC-enriched elution fractions (Fig. 6A, 3, left panel). Nickel-affinity purification of embedded GF(MUC1)P coexpressed with CytoEpi-T2 revealed very little degradation (Fig. 6B).

Both MUC1-YFP (Fig. 6C) and embedded GF(MUC1)P (Fig. 6D) coexpressed with T2-2A-CytoEpi and CytoEpi-T2, respectively, were subjected to Asp-N digestion for release of the MUC1 TR 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 nonglycosylated MUC1-1TR with two Pro hydroxylations (mass-to-charge ratio of 1,918; Fig. 6, C and D). Peaks corresponding to Tn-MUC1-1TR with one to three GalNAc attachments and one to four Pro hydroxylations were observed (Fig. 6, C and D). Minor additional products, corresponding to the addition of three Ara residues, were also identified (Fig. 6C).

DISCUSSION

We previously established mammalian GalNAc O-glycosylation capacity transiently in plants. GalNAc O-glycosylation capacity required the 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: (1) tobacco BY-2 suspension cells, which offers potential for the secretion of products (Hellwig et al., 2004) and has low reproductive constraints with regard to the propagation and manipulation of transgenic lines; and (2) soil-grown Arabidopsis, which has a comprehensive transferred DNA knockout repertoire (http://signal.salk.edu/) enabling, for example, the manipulation of endogenous-derived PTMs. A previous report proposed that a three-component machinery (additionally including a UDP-GalNAc-T) was required for the 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-T is not required.

We did find unexpected differences in efficiency of the glycosylation machinery designs when introduced stably compared with 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 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 to 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 polypeptide constructs with respect to the order of proteins may be important.

We previously observed that the MUC1 reporter substrate was somewhat unstable when expressed transiently in plants (Yang et al., 2012). Here, we observed even more severe problems with the stability of the small MUC1 reporter substrate, whether expressed alone (MUC1) or in fusion with the carrier YFP (MUC1-YFP; Figs. 1 and 3). We also showed that incubation of nonglycosylated 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. Therefore, we tested an alternative strategy where we embedded 2.5 TRs of MUC1 with tags into GFP (Fig. 4A), as proposed previously (Kobayashi et al., 2008). Those 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 without substantial loss of fluorescence capacity. Coexpression of the embedded construct GF(MUC1)P with the O-glycosylation machinery CytoEpi-T2 resulted in the 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 three GalNAc residues, which corresponds to the substrate specificity of GalNAc-T2 (Bennett et al., 1998). Approximately 50% of the GF(MUC1)P in the BY-2 medium further contained up to four Hyp residues as well as three and six Ara residues (Fig. 5A). Introduction of a single construct for the expression of GF(MUC1)P and the O-glycosylation machinery CytoEpi-2A-T2 resulted in moderate accumulation of GF(MUC1)P (approximately 5 μg g⁻¹ fresh weight) in Arabidopsis (Fig. 6B), which was modified with zero to three GalNAc residues, zero to four Pro 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 IFNα2B in Arabidopsis resulted in expression of the intact product, which was glycosylated when coexpressed with the O-glycosylation machinery CytoEpi-T2 (Fig. 2C). Likewise, 1.2 TRs of MUC16, which also realizes a full mature secreted form of the cytokine IFNα2B, resulted in moderate accumulation of GF(MUC1)P (approximately 5 μg g⁻¹ fresh weight) in Arabidopsis (Fig. 6B), which was modified with zero to three GalNAc residues, zero to four Pro hydroxylations, and minor arabinosides (Fig. 6D).
Figure 6. MUC1-YFP and GF(MUC1)P expressed with or without O-glycosylation machinery in Arabidopsis. A, SDS-PAGE analysis of a transgenic Arabidopsis line expressing a combined construct (MUC1-YFP-T2-2A-CytoEpi) comprising both the reporter MUC1-YFP and the O-glycosylation machinery T2-2A-CytoEpi. Shown are MUC1-specific and Tn-MUC1-specific western-blot analysis of total leaf extracts (1) and MUC1-YFP purified from leaves (2) and roots (3) by HIC as visualized by Coomassie blue-stained SDS-PAGE gels and MUC1-specific MAb 5E10 western blots of the eluates. B, His tag purification of embedded GF(MUC1)P coexpressed with the O-glycosylation machinery CytoEpi-T2 as visualized by Coomassie blue-stained SDS-PAGE gels and MUC1-specific MAb 5E10 western blot. C and D, MALDI-TOF-MS analysis of MUC1 tandem repeats (DTRPAGSTAPPAGHVTAS) indicated by arrows) from His tag-purified Asp-N-digested MUC1-YFP (C) and GF(MUC1)P (D) coexpressed with the O-glycosylation machineries T2-2A-CytoEpi and CytoEpi-T2, respectively. [See online article for color version of this figure.]
Plant Cell Factories with Mucin-Type O-Glycosylation

small MUC1-3.5TR may include a lack of proper folding and a lack of protective O-glycosylation, where native mammalian mucins are large (greater than 200 kD) heavily O-glycosylated glycoproteins abundant in mucous layers, where they hydrate, lubricate, and protect cells from proteases as well as from pathogens. In plants, hundreds of genes encode proteins involved in proteolytic processes (Rawlings et al., 2008), with an estimated 1,900 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 require, for the production of some pharmaceutical proteins, tailored knockout 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 the recombinant production of human therapeutics in plants. Pro hydroxylation is initiated by a family of P4Hs, and the resulting Hyp residues may subsequently be glycosylated with arabinogalactan or shorter arabinosides. P4Hs are membrane-bound type II proteins found in the endoplasmic reticulum and shorter arabinosides. P4Hs are membrane-bound type II proteins found in the endoplasmic reticulum and extending into the Golgi apparatus (Yuasa et al., 2005).

The high protease activity in the BY-2 medium (Robertson et al., 1997; Okushima et al., 2000) may require, for the production of some pharmaceutical proteins, tailored knockout 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 the recombinant production of human therapeutics in plants. Pro hydroxylation is initiated by a family of P4Hs, and the resulting Hyp residues may subsequently be glycosylated with arabinogalactan or shorter arabinosides. P4Hs are membrane-bound type II proteins found in the endoplasmic reticulum and extending into the Golgi apparatus (Yuasa et al., 2005).

The high protease activity in the BY-2 medium (Robertson et al., 1997; Okushima et al., 2000) may require, for the production of some pharmaceutical proteins, tailored knockout 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 the recombinant production of human therapeutics in plants.

Our interest in designing plant cell factories with engineered human-style O-glycosylation stems from a need for the 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 Tr-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 the 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 and 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 nonfeasible 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 PTMs 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 the expression of glycoprotein pharmaceuticals in plants.

MATERIALS AND METHODS

Transformation and Plant Growth Conditions

Tobacco (Nicotiana tabacum) BY-2 suspension cells were cultivated by shaking (25°C, 130 rpm) in the dark. Transformation and fermentation of BY-2 cells are described by Mayo et al. (2006). Agrobacterium tumefaciens strain
DNA Constructs for Plant Transformation

Open-source vectors used for A. tumefaciens-mediated expression and transformation are pBI121 (GenBank accession no. AY781296), pCAMBIA 2300 (GenBank accession no. AF234315), and pCAMBIA 1302 (GenBank accession no. AF234298). For the legacy of open-source pCAMBIA binary transformation are pBI121 (GenBank accession no. AY781296), pCAMBIA 2300, and transformation of transgenic Arabidopsis plants were as described (Egelund et al., 2007). Transformation and maintenance of Lemma minor were done as described (Rival et al., 2008). A. tumefaciens-mediated expression was done essentially as described in our previous study (Yang et al., 2012).

Nickel Chromatography

Cleared supernatants (50 mL, 20,000 g for 30 min) were incubated with 0.5 mL of nickel-nitrioltriacetic acid agarose beads (Qiagen) for 2 h (4°C) under gentle rolling. Beads were washed with 10 mL of 20 mM Tris-HCl (pH 8.0) and eluted with 5 × 1 mL of elution buffer (50 mM NaPO4, 250 mM NaCl, and 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 d was collected by centrifugation at 10,000 × g for 5 min (wild-type BY-2 medium fraction). Approximately 5 μg of purified MUC1-YFP (transiently produced in leaves of Nicotiana benthamiana plants with/without O-glycosylation machinery) and embedded GF(MUC1)P (purified from intracellular or extracellular BY-2 cells with/without O-glycosylation machinery) was added to 7-d-old 1 mL of unboiled or boiled (for 15 min) wild-type BY-2 medium fractions, which were further incubated at 25°C under shaking (130 rpm) for an additional 24 or 48 h under the same conditions and ultimately subjected to MUC1 and Tr-MUC1-specific (MAbs SE10 and SE5, respectively) western blot analysis.

Endoprotease Asp Digestion of MUC1-TR and Sample Purification

Purified MUC1-YFP (approximately 25 μg) was incubated with 1 μg of endoprotease Asp-N from Pseudomonas fragi (Roche) in 300 μL of 100 mM Tris-HCl (pH 8.0) for 16 h 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, 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−1) and trypsin (10 ng μL−1) were added into the digestion buffer (25 mM NH4HCO3 and 10% acetonitrile). The digest was cleaned on a C18 Zip-Tip (Millipore). Briefly, the digestion mixture was dissolved in 20 μL of 0.1% trifluoroacetic acid and drawn through the column, desalted with 0.5% formic acid, and eluted with 100% acetonitrile.

MALDI-TOF-MS

Lyophilized peptides were dissolved in 20 μL of water. The MALDI matrix 2,5-dihydroxybenzoic acid (Sigma-Aldrich) was dissolved (25 g L−1) 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.
(Perceptive Biosystem) equipped with delayed extraction; external calibration was used.

**Supplemental Data**

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Expression of the mucin type O-glycoproteins MUC1 and MUC16 in L. minor and BY-2 cells, respectively.

**Supplemental Figure S2.** SDS-PAGE western-blot analysis of 2A-linked O-glycosylation machinery.

**Supplemental Figure S3.** Time course of Tn-MUC1 and (Tn)-GF(MUC1)P in BY-2 cell medium.

**Supplemental Figure S4.** Prior introduction of O-glycosylation machinery enhanced the glycosylation level of MUC1-YFP in a transient expression system.

**Supplemental Figure S5.** Chemical inhibitors of P4Hs reduce extracellular accumulation of GF(MUC1)P and inhibit growth of BY-2 cells.

**Supplemental Figure S6.** Chemical inhibition of proline hydroxylation of MUC1 expressed in BY-2 cells.

**Supplemental Figure S7.** Transient expression of MUC1-YFP with coexpression of different C4 P4Hs in *N. benthamiana*.

**ACKNOWLEDGMENTS**

pET23-WbpP® was a generous gift from J.S. Lam (University of Guelph). Takashi Murayama (Juntendo University School of Medicine) is acknowledged for sharing the construct modified GFP. Dr. Jose Estevez (Universidad de Buenos Aires) is thanked for helpful suggestions with relation to Pro hydroxylation inhibition experiments.

Received April 22, 2012; accepted July 11, 2012; published July 12, 2012.

**LITERATURE CITED**


Yang et al.


Copyright © 2012 American Society of Plant Biologists. All rights reserved.