Available online at www.sciencedirect.com
ScienceDirect
Phytochemistry 69 (2008) 1631–1640
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

The *O*-Hyp glycosylation code in tobacco and Arabidopsis and a proposed role of Hyp-glycans in secretion

Jianfeng Xu [a], Li Tan [a], Derek T.A. Lamport [b], Allan M. Showalter [c], Marcia J. Kieliszewski [a,*]

[a] *Department of Chemistry and Biochemistry, Ohio University, Athens, OH 45701, United States*
[b] *Department of Environmental and Plant Biology, Ohio University, Athens, OH 45701, United States*
[c] *School of Life Sciences, John Maynard Smith Building, University of Sussex, Falmer, Brighton BN1 9QG, UK*

Received 11 November 2007; received in revised form 30 January 2008
Available online 25 March 2008

Abstract

Most aspects of plant growth involve cell surface hydroxyproline (Hyp)-rich glycoproteins (HRGPs) whose properties depend on arabinogalactan polysaccharides and arabinosides that define the molecular surface. Potential glycosylation sites are defined by an *O*-Hyp glycosylation code: contiguous Hyp directs arabinosylation. Clustered non-contiguous Hyp directs arabinogalactosylation. Elucidation of this code involved a single species, tobacco (*Nicotiana tabacum*) BY-2 cells. However, recent work suggests species variation, perhaps tissue specific Hyp glycosylation. Thus, the extent to which the Hyp glycosylation code is 'global' needs testing. We compared the ability of distantly related Arabidopsis cell cultures to process putative HRGP glycosylation motifs encoded by synthetic genes. The genes included: repetitive *Ser-Pro*, *Ser-Pro*$_2$, *Ser-Pro*$_4$ and an analog of the tomato arabinogalactan-protein, *LeAGP-1ΔGPI*. All were expressed as enhanced green fluorescent protein (EGFP) fusion glycoproteins, designated: AtSO-EGFP (O = Hyp), AtSO$_2$-EGFP, AtSO$_4$-EGFP and AtEGFP-LeAGP-1ΔGPI, respectively. The Arabidopsis glycosylation patterns were essentially similar to those observed in *Nicotiana*: non-contiguous Hyp residues in AtSO-EGFP were glycosylated exclusively with arabinogalactan polysaccharides while contiguous Hyp in AtSO$_2$-EGFP and AtSO$_4$-EGFP was exclusively arabinosylated. Mixed contiguous and non-contiguous Hyp residues in AtEGFP-LeAGP-1ΔGPI were also arabinosylated and arabinogalactosylated consistent with the code. However, slightly more arabinogalactosylated Hyp and less non-glycosylated Hyp in AtEGFP-LeAGP-1ΔGPI than tobacco NtEGFP-LeAGP-1ΔGPI suggested Arabidopsis prolyl hydroxylases have a slightly broader specificity. Arabidopsis Hyp-arabinogalactans differed from tobacco in decreased glucuronic acid content and lack of rhamnose. Yields of the EGFP fusion glycoproteins were dramatically higher than targeted EGFP lacking Hyp-glycomodules. This validates earlier suggestions that the glycosylation of proteins facilitates their secretion.

Keywords: *Arabidopsis thaliana*; *Cruciferae*; *O*-Glycosylation; Arabinogalactan-protein; Hydroxyproline-rich glycoprotein; (β-D-Galactosyl)$_3$-Yariv reagent

1. Introduction

Hydroxyproline-rich glycoproteins (HRGPs) expressed at the plant cell surface are of considerable current interest as they are involved in virtually all aspects of plant growth and development, from fertilization and cytokinesis (Hall and Cannon, 2002) to apoptosis and senescence (Nothnagel, 1997; Showalter, 2001; Motose et al., 2004; Lamport et al., 2006). In the broadest sense, HRGPs comprise numerous cell surface proteins that contain glycosylated hydroxyproline. This includes extensins, arabinogalactan-proteins (AGPs), proline-rich proteins, and numerous other protein chimeras that contain HRGP glycomodules. HRGPs are dominated by *O*-Hyp linked arabinosides (Lamport, 1967) and arabinogalactan polysaccharides (Lamport, 1977; Pope, 1977) which define most of the interactive molecular surface. The 'surface codes' are therefore of intrinsic interest.

Recent work with tobacco BY-2 cells supports the Hyp contiguity hypothesis (Fong et al., 1992; Kieliszewski et al., 1992; Kieliszewski and Lamport, 1994) where a simple

* Corresponding author. Tel.: +1 740 593 9466; fax: +1 740 597 1772.
E-mail address: kielisze@helios.phy.ohiou.edu (M.J. Kieliszewski).

0031-9422/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2008.02.006

code based on peptide sequence directs *O*-Hyp glycosylation: contiguous Hyp blocks, typically the Ser-Hyp_4 motif of extensins, are preferred sites of arabinoside addition, while non-contiguous, clustered Hyp residues, typically Ser-Hyp and Ala-Hyp repeats of the AGPs, are preferred sites of arabinogalactan heteropolysaccharide addition. Earlier we tested this hypothesis by expressing synthetic genes encoding repetitive clustered non-contiguous Pro as Ser-Pro and Ser-Pro_n repeats, or analogs of a naturally occurring tomato (*Solanum lycopersicon*) AGP, LeAGP-1. The products were hydroxylated and glycosylated as predicted (Shpak et al., 1999, 2001; Kieliszewski, 2001; Kieliszewski and Shpak, 2001; Zhao et al., 2002; Tan et al., 2003; Held et al., 2004).

Recently, Estévez et al., 2006 expressed the same synthetic gene constructs and an LeAGP-1 analog in Arabidopsis plants. However, the extent of glycosylation seemed to differ from tissue to tissue. For example, Pro residues in extensin Ser-Pro_4 motifs were hydroxylated and exclusively arabinosylated in BY-2 cells (Shpak et al., 2001). Interestingly when expressed in Arabidopsis leaves the same motif was only 50–60% hydroxylated with addition of arabinogalactan polysaccharide to some of the Hyp residues. Therefore, to ascertain possible differences in glycosylation codes between distantly related species like tobacco and Arabidopsis, we compared Hyp glycosylation in cultured cells of both species. This direct comparison between simple systems eliminated interpretation of HRGP glycosylation profiles arising from a mixture of cell and tissue types as occurs with analysis of whole plant organs as it eliminates the confounding effect of variable prolyl 4-hydroxylase (P4H) expression. Cell cultures tend to express high activities of P4H and they also provide abundant material for Hyp glycosylation profiling. Thus, here we compared the Hyp glycosylation profiles in Arabidopsis cell suspension cultures with those of similar tobacco cultures. Macromolecules expressed included both naturally occurring AGPs and the products of synthetic genes encoding LeAGP-1ΔGPI (Zhao et al., 2002), $(Ser\text{-}Pro)_{32}$ (designated SP), $(Ser\text{-}Pro_2)_{24}$ (designated SP_2), and $(Ser\text{-}Pro_4)_{18}$ (designated SP_4) all expressed as their EGFP fusions.

Finally, we note that addition of an AG glycomodule to EGFP significantly enhances its secretion and thus corroborates the original suggestion that one of the many roles of glycosylation facilitates secretion (Eylar, 1966; Varki, 1993; Borner et al., 2002).

2. Results

Transformation of the Arabidopsis cells with the genes *SP-EGFP*, *SP_2-EGFP*, *SP_4-EGFP* and *EGFP-LeAGP-1ΔGPI* (a gene encoding AtEGFP-LeAGP-1ΔGPI, which lacked the sequence directing addition of a glycosylphosphatidylinositol (GPI) anchor to LeAGP-1) resulted in several cell lines showing green fluorescence in the culture medium. Lines exhibiting the most fluorescence were chosen for propagation in liquid culture and transgene product characterization. Isolation by hydrophobic interaction chromatography (HIC) and reversed-phase chromatography gave yields of AtSO-EGFP, $AtSO_2$-EGFP, $AtSO_4$-EGFP and AtEGFP-LeAGP-1ΔGPI (O = Hyp) that ranged from 25 to 40 mg/l medium. Cell lines expressing ER (endoplasmic reticulum) targeted, but non-glycosylated EGFP secreted <0.05 mg/l EGFP into the medium.

Amino acid composition analyses indicated that all Pro residues were hydroxylated in the isolated $(SO)_{32}$, $(SO_2)_{24}$ and $(SO_4)_{18}$ glycomodules of AtSO-EGFP, $AtSO_2$-EGFP and $AtSO_4$-EGFP (hereafter designated AtSO, $AtSO_2$ and $AtSO_4$) (data not shown), and except for Ser and Hyp, there were no other amino acids detected, confirming the identity and purity of these isolated glycomodules. The composition of AtLeAGP-1ΔGPI (after removal of EGFP) closely resembled that of EGFP-LeAGP-1ΔGPI expressed in tobacco (designated NtEGFP-LeAGP-1ΔGPI) after removal of EGFP (Table 1). N-terminal protein sequencing of AtLeAGP-1ΔGPI yielded TGQTOAAAOVGAKAGTTOOA…. This confirmed the identity and purity of the isolated AtLeAGP-1ΔGPI further corroborated by SDS–PAGE that showed no contaminants (data not shown).

Hyp assays showed that cell walls isolated from leaves and stems of mature Arabidopsis plants contained 0.05% and 0.13% Hyp (dry weight) and the walls of Arabidopsis and BY-2 suspension cultured cells contained 0.3% and 0.5% Hyp, respectively.

AtSO-EGFP was rich in galactose (Gal) and arabinose (Ara) with lesser amounts of glucuronic acid (GlcUA) than

Table 1
Amino acid compositions of AtLeAGP-1ΔGPI, NtLeAGP-1ΔGPI, and the deglycosylated LeAGP-1 polypeptide from tomato compared to those predicted from the cDNA of *LeAGP-1ΔGPI*

Amino acid	Composition (mol%)[a]			
	AtLeAGP-1ΔGPI	NtLeAGP-1ΔGPIa	Tomato LeAGP-1ΔGPI[a]	LeAGP-1ΔGPI cDNA
Hyp	26.3 ± 0.1	25	29	–
Pro	1.9 ± 0.1	3	1	27.5
Asx	4.2 ± 0.2	4	2	3.5
Thr	9.6 ± 0.2	10	10	8.8
Ser	10.5 ± 0.3	10	12	11.7
Gly	3.7 ± 0.2	3	5	3.5
Glx	4.2 ± 0.2	5	3	3.5
Ala	23.3 ± 0.2	22	21	23.5
Val	8.8 ± 0.2	9	9	8.2
Cys	0	0	0	0.0
Met	1.1 ± 0.1	1	0	1.7
Ile	0	0	1	0.0
Leu	1.8 ± 0.2	1	1	1.2
Tyr	0	0	0	0.0
Phe	0	0	1	0.0
His	0	0	0	0.6
Lys	4.6 ± 0.1	4	5	6.4
Arg	0	1	0	0.0
Trp	nd[b]	nd	nd	0.0

Standard deviations represent the measurements from three cell lines.
[a] From Zhao et al. (2002).
[b] Not determined.