CMLS, Cell. Mol. Life Sci. 58 (2001) 1386–1398
1420-682X/01/101386-13 $ 1.50 + 0.20/0
© Birkhäuser Verlag, Basel, 2001

CMLS Cellular and Molecular Life Sciences

Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins

M. J. Kieliszewski* and E. Shpak

Department of Chemistry and Biochemistry, Ohio University, Athens, Ohio 45701 (USA)
e-mail: kielisze@helios.phy.ohiou.edu

Abstract. Hydroxyproline-rich glycoproteins (HRGPs) are ubiquitous architectural components of the growing plant cell wall, accounting for as much as 10–20% of the dry weight. HRGPs are implicated in all aspects of plant growth and development, including responses to stress. The HRGP superfamily contains three major groups which represent a continuum of peptide periodicity and hydroxyproline-O-glycosylation. These groups range from the highly periodic and lightly arabinosylated repetitive proline-rich proteins (PRPs), through the cross-linked extensins which are periodic and highly arabinosylated, to the arabinogalactan-proteins (AGPs) which are the most highly glycosylated and least periodic. The repetitive units are small, often only four- to six-residue-glycosylated modules viewed hypothetically as functional motifs, or glycomodules. The Hyp contiguity hypothesis predicts that Hyp arabinosylation increases with Hyp contiguity and that clustered noncontiguous Hyp residues are sites of arabinogalactan polysaccharide addition in the AGPs and gums. Recent results involving glycosylation site mapping of endogenous HRGPs and HRGP design using synthetic genes have corroborated the hypothesis. The uses of synthetic genes in HRGP glycosylation site mapping and structural/functional analysis are also discussed.

Key words. Cell wall; hydroxyproline-rich glycoproteins; extensin; arabinogalactan protein; proline-rich protein; O-glycosylation.

Introduction

The plant primary cell wall is a supramolecular, self-assembling, interactive network of structural polysaccharides [1] and HRGPs [2–5]. Three major families comprise the HRGP superfamily: the PRPs, extensins, and AGPs, and although repetitive signature peptide sequences distinguish the three families, they present a continuum of peptide sequence and glycosylation rather than inflexible categories. HRGPs play many different roles in plant growth and development; however, those roles are largely undefined at a molecular level. Significantly, plant HRGPs have no animal homologs for comparative structural and functional analysis. As such, the rich and uniquely 'planty' hydroxyproline O-glycosylation [6] of the HRGP polypeptide backbone must signify functions peculiar to the plant kingdom.

HRGP repetitive modules generally consist of small, simple, four- to six-residue motifs often extensively O-glycosylated by arabinooligosaccharides or arabinogalactan polysaccharides attached to the hydroxyproline residues, and sometimes to serine mainly as single galactosyl residues [7]. The conservation and widespread occurrence of only a few HRGP glycomodules throughout the evolutionary progression implies that they are small functional units [2] whose roles depend on the identity and precise arrangement of the pendent saccharide chains. Indeed, for most HRGPs, carbohydrate essentially encapsulates the peptide backbone, defines the interactive molecular surface, and therefore is the major determinant of function. Thus, HRGP functional analysis necessarily devolves into the rules and roles of glycosylation. We have summarized these rules in the Hyp contiguity hypothesis, which essentially predicts that protein sequence directs

* Corresponding author.

Hyp O-glycosylation [2] as it does for N-glycosylation in other glycoproteins [8].

This review discusses the elucidation of Hyp glycosylation codes in the three major groups of HRGPs, with special emphasis on those codes that direct polysaccharide addition in AGPs and the closely related AGP gums. What are the consequences of decrypting the HRGP glycosylation code? First of all, it will it help us describe and predict HRGP three-dimensional structure. Beyond that, deciphering the code should eventually enable us to relate site-specific HRGP glycosylation to essential molecular functions such as crosslinking, self-association, cell surface binding and wall self-assembly. Here we discuss how the design of synthetic HRGP genes is enabling us to crack HRGP O-glycosylation codes, define the precise three-dimensional structures and properties of HRGP glycomodules, and ultimately relate those structures to specific functions in the plant extracellular matrix.

HRGPs and the Hyp contiguity hypothesis

HRGP carbohydrate content covers a wide range – from ~1% to >95% by weight, involving either an initial arabinofuranosyl β-1,4-Hyp or galactopyranosyl β-1,4-Hyp linkage. Each type of Hyp O-glycosylation, arabinosylation and arabinogalactan-polysaccharide addition has conserved unique architectural features presumably of functional significance. Arabinosylation involves the attachment of linear homooligosaccharides ranging from one to four L-arabinofuranosyl residues [9] (fig. 1). The much larger arabinogalactan heteropolysaccharide substituents possess a 'core' β-1,3 D-galactan backbone [10, 11] periodically decorated with 1,6 additions – small side chains of D-galactose, L-arabinose and often including other sugars such as L-rhamnose and sugar acids such as D-glucuronic acid and its 4-*O*-methyl derivative [12].

A correlation between protein sequence and Hyp arabinosylation (table 1) led us to propose that Hyp glycosylation was not random but dependent on the arrangment of Hyp residues as summarized by the Hyp contiguity hypothesis [2, 13, 14]. This working hypothesis predicts the Hyp residues most likely to be glycosylated and the type of glycosylation based on the following experimental observations:

1) Hyp arabinosylation increases with increasing Hyp contiguity. Thus Hyp tetraarabinosylation occurs predominantly on tetraHyp blocks, typically the Ser-Hyp_4 motif that characterizes extensins (fig. 2 and table 1).

2) Noncontiguous Hyp residues tend not to be arabinosylated, or only occasionally arabinosylated, for example the Lys-Pro-Hyp-Val-Hyp-Val and Ile-Pro-Pro-Hyp-Val repeats of the Douglas fir PHRGP (proline and hydroxyproline-rich glycoprotein) [14] (fig. 3).

3) Arabinogalactan polysaccharide attachment sites involve noncontiguous Hyp residues that are clustered to a greater (e.g. X-Hyp-X-Hyp repeats, where commonly X = Ser, Thr or Ala) or lesser extent (e.g. [X-Hyp-X_n] repeats, where n ~ 1–3) [2] (table 2).

HRGP families reflect these major differences in peptide sequence and glycosylation as follows:

The repetitive proline-rich proteins (PRPs and some nodulins) are the simplest HRGPs in terms of their signa-

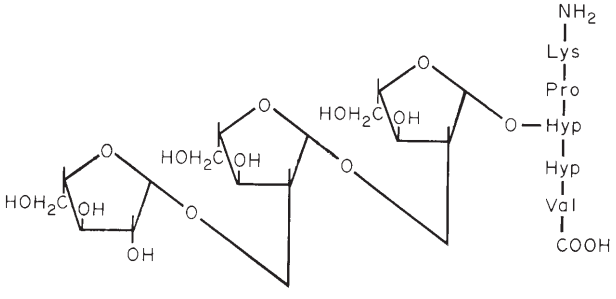

Figure 1. Triarabinooligosaccharide O-linked to hydroxyproline in a glycopeptide isolated from Douglas fir PHRGP. Hydroxyproline arabinosides are attached to Hyp by a β-*O*-glycosidic bond and L-arabinose residues 2 and 3 are β-1,2 linked [9].

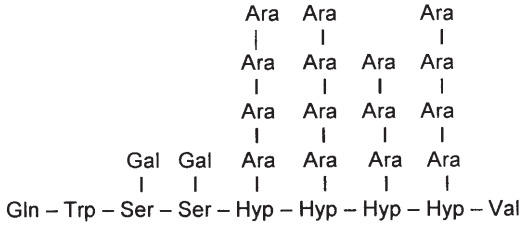

Figure 2. Proposed glycosylated structure of the Ser-Hyp-Hyp-Hyp-Hyp motif from tomato P1 extensin. We degraded tomato P1 extensin with pronase, then isolated glycopeptides. Analysis by Edman degradation and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry indicated this glycopeptide contained 14–16 arabinose residues and 2 galactose residues; 15 arabinose residues attached to Hyp are represented here; however the exact placement of the triarabinoside along the tetraHyp block is unknown.

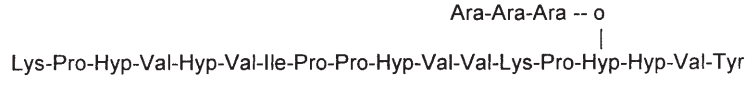

Figure 3. The major repetitive glycopeptide repeat of Douglas fir PHRGP. Glycosylation site mapping of PHRGP showed that the Hyp residue at position 15 in this 18-residue repeat was always arabinosylated, mainly with a triarabinoside. The Hyp residue at position 5 is monoarabinosylated 20% of the time [14].

Table 1. A correlation between Hyp contiguity and Hyp arabinosylation.

HRGP	% NGH[a]	% Hyp-Ara_4	Contiguous Hyp block			Reference
			Hyp_2	Hyp_3	Hyp_4	
Soy PRP1	~100	0	~1	0	0	2, 131
Fir PHRGP	73	~0	++[b]	0	~0	13
Maize THRGP	48	6	+++	0	1	132, 133
Maize HHRGP	34	15	++	+++	+	134
Beet P1	32	4	+++	+	0	135
GAGP	12	8	+++	+++	+	79, 112
Tomato P1	12	38	0	0	+++	129
Potato lectin	8	47	+	0	+++	136
Tomato P2	7	54	0	0	+++	129

[a] NGH, nonglycosylated Hyp.
[b] + indicates the presence of the sequence. For example, one (+) means the sequence occurs, but infrequently, whereas +++ denotes a major repetitive motif.

Table 2. Clustered noncontiguous Hyp repeats in AGPs compared to noncontiguous Hyp in extensins and PRPs.

AGP	Sequence	Reference
Pc1[a]	I**APAP**GA**APTP**SPPTAKSPTATPPTATPPSA**VPVPSP**SK**TP**TA**SPTPSP**VT**APTP**SA	122
Pc2[a]	AEAE**APTP**ALP[b]VVAE**AP**EL**VPTPVPTP**S	123
Na2[a]	I IP[b]A**SPTPAPTP**INEISFPPFSSL**TPTPSPTPAP**AT**APTP**FF	123
GAGP	SOOOT/OL**SOSOTO**TOOO/LGPH	112
TTS[a]	**SPAPSP**TKPPTYSPSKPPVKPPVKPPTKPPTYSPSKPPAKAPVKPP**TPSP**YP[b]**APAP**I	80, 124
PtX3H6[a]	Q**SP**SA**SP**TK**SP**T**TPAP**TTTA**AP**TTT	125
LeAGP1[a]	APTKP[b]K**TPTP**AT**AP**ASAPPTA**VP**V**AP**VT**AP**VT**APT**T**TP**VVA**AP**VS**AP**ASS	126
Na1[a]	Q**AP**GA**SP**AA**SP**KA**SP**V**AP**VAVASPPTAVV**TP**VS**AP**SQ**SP**STAA**SP**SE**SP**LA	127
PO2[a]	AS**AP**A**TP**A**TP**A**TP**A**TP**SKPATPAAVS	128

Extensin	Sequence	Reference
P1	SOOOO**TO**VYKSOOOOVKPYHP**TO**VYKSOOOO**TO**VYKSOOOOVKPYHP**TO**VYK	129
		Shpak and Kieliszewski, unpublished
P3	SOOOO**SO**SOOOOYYYKSOOOO**SO**SOOOOYYYK	68, 130

PRP	Sequence	Reference
PHRGP	KP[b]O**VO**[c]VIPPOVVKPOOVYKPO**VO**VIPPOVVKPOOVYKIPOOVIKP	13
SbPRP2	KPOVYKPOVEKPOVYKPOVEK	32, 131

[a] Sequences deduced from clones do not identify which Pro residues are modified to Hyp.
[b] Pro residues following L, I, Y and K are seldom or never hydroxylated.
[c] The putative code for AGP-polysaccharide addition is clustered XO repeats where X is probably not Val.

ture peptide motifs and glycosylation. They are characterized by short repetitive pentameric or hexameric blocks [15], variations of a Pro-Hyp-(Hyp)-X-Y Lys motif that generally contains no contiguous Hyp [2, 13, 16] or only dipeptidyl Hyp [2, 13]. The PRPs, with relatively few Hyp residues arabinosylated and no arabinogalactan polysaccharide [2, 13, 14], exemplify the low end of the glycosylation range, and carbohydrate plays a lesser role in their function, if at all. The extreme examples are soybeans PRP1 and PRP2, which are essentially composed of $[Pro\text{-}Hyp\text{-}Val\text{-}Tyr\text{-}Lys]_n$ and $[Pro\text{-}Hyp\text{-}Val\text{-}Glu\text{-}Lys\text{-}Pro\text{-}Hyp\text{-}Val\text{-}Tyr\text{-}Lys]_n$ repeats [17, 18] and contain only 1–2% arabinose on a dry weight basis [2]. However, PRPs containing dipeptidyl Hyp are more highly arabinosylated, for example Douglas fir PHRGP [13] (table 1). On this basis, we predict arabinosylation for putative PRPs ENOD2 [19] and ENOD10 [20] and soybean nodulin-75 [21] in keeping with the likelihood that they contain contiguous Hyp residues.

Douglas fir PHRGP was chosen as a first test of the glycosylation patterns predicted from the Hyp contiguity hypothesis because as a PRP it was simple, only lightly arabinosylated, and contained both contiguous and noncontiguous Hyp residues. Glycosylation site mapping of

PHRGP demonstrated that noncontiguous Hyp in the sequence Ile-Pro-Pro-Hyp was never arabinosylated. Likewise, the Hyp residues in the sequence Lys-Pro-Hyp-Val-Hyp were not glycosylated or only rarely was Hyp residue 5 monoarabinosylated. In contrast, the peptide isomer Lys-Pro-Hyp-Hyp-Val was consistently arabinosylated, mainly with triarabinosides (fig. 3). Thus, a change from noncontiguous Hyp to contiguous Hyp in the two peptide structural isomers, Lys-Pro-Hyp-Val-Hyp and Lys-Pro-Hyp-Hyp-Val, dramatically influenced Hyp arabinosylation [14].

PHRGP does contain clustered, noncontiguous Hyp (underlined) in the repetitive module Lys-Pro-Hyp-Val-Hyp (also see fig. 3); however, it is not a substrate for polysaccharide addition. This suggests: (i) that the X in X-Hyp-X-Hyp repeats influences glycosylation, or (ii) that polypeptide rigidity or a polyproline II conformation in addition to sequence may also play a role in directing Hyp arabinosylation. The rigid polyproline II conformation typical of the highly arabinosylated extensins in particular is nucleated by contiguous Hyp/Pro residues (i.e. secondary amino acids which reduce the flexibility of the polypeptide backbone), whereas an HRGP containing only Ser-Hyp-Ser-Hyp repeats and arabinogalactan polysaccharide attached to Hyp has a random coil conformation [22]. However, the series of synthetic genes encoding repetitive (Ser-Pro), ($Ser\text{-}Pro_2$), ($Ser\text{-}Pro_3$) and ($Ser\text{-}Pro_4$) produced glycoproteins in which Hyp glycosylation was clearly dependent on protein sequence rather than polypeptide conformation [22].

The disappearance of soluble PRPs in response to fungal elicitors and wounding [22] has led to the general assumption that PRPs in general are also oxidatively crosslinked in nodule cell walls [16, 24–26], forming covalently crosslinked networks during normal development or in response to pathogens [27–29]. Two additional lines of evidence favor the idea of crosslinked PRPs: First, PRPs such as soybean PRP1 and soybean PRP2 contain the Val-Tyr-Lys motif [17, 30] previously identified as a putative crosslink site in extensin [31]. Second, peptide maps of the anhydrous hydrogen fluoride (HF)-insoluble soybean culture cell wall residue yielded the major PRP2 peptide Lys-Pro-Hyp-Val-Glu-Lys-Pro-Hyp-Val-Tyr-Lys, which implies it is part of a covalently crosslinked network involving protein-protein or protein-phenolic crosslinks [J. Frueauf and M. Kieliszewski, unpublished data]. However, recent examination of the HF-insoluble nodule wall residue of *Medicago truncatula*, known to be rich in various PRPs, showed bound extensin and other non-HRGP protein, but the only evidence of wall-bound, insoluble, crosslinked PRP was indirect. Hyp-glycoside profiles of *Medicago* nodule walls contained a significant amount of nonglycosylated Hyp, suggesting the presence of wall-bound PRPs. In contrast to the *Medicago* nodule, enzymic degradation of the HF-insoluble root wall residue yielded the Val-Tyr-Lys-containing PRP peptide, Lys-Ala-Hyp-Val-Glu-Lys-Pro-Hyp-Val-Tyr [32]. Whereas this result argues for Val-Tyr-Lys involvement in PRP wall insolubilization, the role of PRP glycosylation, or lack of it, in biological function can only be conjectured. However, we have hints from the extensins that Hyp arabinosylation may play a role in wall self-assembly and crosslinking.

The extensins are generally recognized by the $Ser\text{-}Hyp_4$ motif, which nucleates extended polyproline II helices to form extended flexuous rods [32, 33–36]. They contain about 50% carbohydrate and therefore occupy an intermediate position in the glycosylation continuum. Extensins contain monogalactosyl-serine and Hyp-arabinooligosaccharides, but no Hyp-arabinogalactan polysaccharide except for those which also share sequence characteristics of AGPs [36, 37].

Extensins are the only HRGPs for which a specific function in cell wall architecture is defined: they participate in crosslinked scaffolds analogous to collagen; however, their molecular partners in the scaffolds have yet to be completely defined**.** Lamport [38, 39] and Fry [40, 41] have both presented evidence for extensin-extensin crosslinked scaffolds; Mort [42] has provided evidence for extensin-pectin crosslinked scaffolds; and the amino acid composition of the tomato HF-insoluble wall residue [43, 44], which is only about half protein by weight, argues for extensin crosslinks to a nonprotein, noncarbohydrate component. Probably all three types of crosslinked scaffolds exist.

While cell suspension cultures express extensins constitutively, most extensins are expressed in a temporal and tissue-specific manner in planta. Specific extensins are correlated with various aspects of wall architecture such as increased tensile strength of the mechanically stressed wall [45–47], cessation of cell extension, cell wall rigidification [48–51], regeneration of protoplast cell walls [52], lateral root initiation [53, 54], root hair growth [55, 56], phyllotaxy [57], and responses to symbionts and pathogens [16, 24–29, 59]. Not surprisingly, the expression of both PRPs [60–67] and extensins are under quite complex control; for example the extA gene of *Brassica napus* is regulated by four sets of positive and negatively acting cis regions which control wound inducibility, activation in response to tensile stress, and quantitative expression [47].

The Hyp residues in extensins (~40 mole%) occur mainly as tetra-Hyp blocks in the highly conserved [2] repetitive $Ser\text{-}Hyp_4$ peptide motif [7, 9, 68] (table 1). Judging from Lamport's early work with cell wall glycopeptides [68, 69], Mort's work with extensins from cotton [42] and Hyp-glycoside profiles of various extensins, the repetitive $Ser\text{-}Hyp_4$ motifs are extensively arabinosylated with tri- and tetraarabinosides to form glycomodules. We recently confirmed this by mass spectrometry of isolated tomato P1 glycopeptides (fig. 2) [M.

Kieliszewski and R. Orlando, unpublished data]. One likely function of the arabinooligosaccharides is to stabilize via H-bonding the polypeptide backbone [34, 39, 70] in the extended three-residue per turn polyproline-II helix. Several lines of evidence [4, 7, 31, 39, 71–73] led us to propose a unique role for Ser-Hyp_4 glycomodules in extensin-extensin self-recognition [2, 31, 74] resulting in molecular alignment for subsequent crosslinking.

1) Sigmoidal crosslinking kinetics of extensin in vitro suggest cooperative assembly of extensins as a prerequisite for subsequent crosslinking by extensin peroxidase.
2) Crosslinking rates of extensin pronase fragments decrease with fragment size [D. Lamport and M. Kieliszewski, unpublished data].
3) Dearabinosylated extensin monomers do not crosslink in vitro; however, arabinose per se is not involved in the crosslink.
4) HRGP monomers that lack Ser-Hyp_4 modules do not crosslink in vitro.
5) Extensin oligomers are always present in preparations of Ser-Hyp_4-containing crosslinking extensins.
6) Reversible noncovalent association of extensin monomers occurs in vitro.
7) Potato lectin, which contains Ser-Hyp_4 modules but apparently no Val-Tyr-Lys motif, exists as a noncovalently associated dimer. We suggest the association is mediated by its highly arabinosylated Ser-Hyp_4-rich HRGP module.
8) Extensin peroxidase does not crosslink the essentially nonarabinosylated Val-Tyr-Lys containing soybean PRP1 and PRP2 in vitro. Thus, PRP crosslinking involves a mechanism independent of Ser-Hyp_4 alignment.

We suggest that the Ser-Hyp_4 glycomodule helps align extensins for subsequent crosslinking. The resulting juxtaposition of the putative peptide crosslinking modules Thr-Hyp-Val-Tyr-Lys and Tyr-Tyr-Tyr-Lys may enable extensin peroxidase to zip up the insoluble extensin network scaffolds that help build Hyp-rich walls.
AGPs contain up to 95% carbohydrate and exemplify the high end of the glycosylation range. AGPs in the strict sense comprise a homologous group of HRGPs termed classical arabinogalactan proteins [75]. These have β1,3-linked arabinogalactan polysaccharides attached O-glycosidically to certain Hyp residues [11, 76] and arabinooligosaccharides often attached to others [2, 76]. In the broad sense AGPs include glycoproteins that are Hyp poor [77, 78] and HRGPs that also possess extensin peptide motifs, e.g. the histidine-rich HRGP of *Maize* and the gum arabic glycoprotein of *Acacia senegal* [36, 79]. As a further complication, many AGPs exist as chimeras of AGP and non-AGP polypeptides [80, 81], including AGP-extensin chimeras [60][1]. AGPs are highly soluble, yet for many years there has been clear evidence that the plasma membrane binds some AGPs [85–89] confirmed by the recent demonstration of AGP-GPI-anchor involvement [90–92]. Thus AGPs undoubtedly play a role at the plasma membrane-wall interface [85, 87], perhaps in guiding wall self-assembly and organization [93, 94] or as linker proteins between the cell wall and plasma membrane [95–98]. However, unlike the extensins, AGPs are not crosslinked into the wall [11, 97, 100], although limited crosslinking to the dimer level may occur [101]. AGPs also play a role in pollination, pollen tube growth [83, 102, 103] and plant defense responses such as gummosis.
We approach the functional significance of AGP glycosylation by considering the AGP polypeptide backbone as a template that directs side-chain assembly of oligo- and polysaccharides to select regions along the backbone. This raises the question, What determines the relative placement of Hyp-polysaccharide substituents and smaller arabinooligosaccharides on the polypeptide template?
The Hyp contiguity hypothesis predicts that Hyp glycosylation involves a sequence-dependent code [104–106]. Here we describe two experimental approaches to cracking the code. The first identifies glycosylation sites in known HRGPs. The second involves the de novo design of new AGPs and their expression in transgenic plant tissue. This second approach allows a direct test of putative glycosylation codes as well as testing for the phenotypic effects that might be conferred by a 'rogue' AGP.

HRGP glycosylation site characterization

Glycosylation site mapping of known HRGPs via chemical/enzymic degradation

This involves enzymic or chemical degradation of intact glycoproteins to yield diagnostic glycopeptides of only a few amino acid residues where the Hyp residues are either contiguous or noncontiguous, but preferably not both in the same isolated peptide. This approach determines the structure of putative arabinooligosaccharide/polysaccharide attachment sequons in complex naturally occurring glycoproteins. For example, tomato extensin P1 contains the Ser-Hyp_4 glycomodule, which alternates with putative peptide crosslinking modules (underlined):

Ser-Hyp-Hyp-Hyp-Hyp-Thr-Hyp-Val-Tyr-Lys and
Ser-Hyp-Hyp-Hyp-Hyp-Val-Lys-Pro-Tyr-His-Pro-Thr-Hyp-Val-Tyr-Lys

[1] The umbrella term proteoglycan has sometimes been used to describe plant AGPs. However, IUPAC (International Union of Pure and Applied Chemistry) defines proteoglycans [82] as possessing glycosaminoglycans, i.e. oligosaccharides based on repetitive disaccharides of amino sugars and uronic acids. Furthermore, IUPAC proteoglycans are usually sulfated [83], consisting of about 30 members with a homologous protein core [84] that lacks hydroxyproline.

However, difficulties lie in generating and then separating modules containing contiguous Hyp from those that contain only noncontigous Hyp. Both chemical and enzymic degradation approaches are greedy procedures; (glyco)peptide yields diminish rapidly with each subsequent fractionation and therefore make this approach labor intensive.

Glycosylation site mapping of tomato extensins and Douglas fir PHRGP corroborated predictions of the Hyp contiguity hypothesis regarding arabinosylation but said little about codes which direct polysaccharide addition to AGPs and the related gums, except that contiguous Hyp residues are unlikely to be arabinogalactan polysaccharide addition sites. Therefore, we chose gum arabic glycoprotein (GAGP) as the AGP most readily available in quantity and the best characterized for further glycosylation site mapping of polysaccharide addition sites.

GAGP is an extended (~150 nm) 'ropelike' or 'wormlike' macromolecular component [79] of gum arabic secreted by *Acacia senegal* during gummosis. This wound response results in a plastic sealant that plugs cracks in the bark. GAGP has remarkable emulsifying and flavor stabilizing properties [107] which account for the wide industrial use of gum arabic, especially by the food and soft drink industries [108–111]. These properties in part depend on a structure that hypothetically includes extensin and AGP glycomodules in a 19-residue consensus sequence [105, 112]:

Ser-Hyp-Hyp-Hyp-Hyp/Thr/Ser-Leu-**Ser-Hyp-Ser-Hyp-Thr-Hyp-**Thr-Hyp-Hyp-Hyp/Leu-Gly-Pro-His.

Noncontiguous Hyp repeats (bolded and underlined) are characteristic of the classical AGPs, whereas the Ser-Hyp_4 motif is characteristic of extensins. Indeed, the Hyp-glycoside profile of GAGP shows a ratio of Hyp-arabinosides to Hyp-polysaccharide consistent with a highly arabinosylated Ser-Hyp_n (extensin) glycomodule and an (AGP) polysaccharide glycomodule (bolded and underlined). Thus, GAGP presents an opportunity to confirm the relative positions of arabinooligosaccharides and arabinogalactan polysaccharides in the same molecule; the consensus sequence is therefore an acid test of the Hyp contiguity hypothesis predicting arabinose on contiguous Hyp and polysaccharide on noncontigous Hyp residues.

The consensus sequence based on the smallest repetitive subunit of 19 residues detailed here, is twice the size of the earlier 10-residue model that contained one attached polysaccharide [79]. Our current model contains three potential polysaccharide addition sites in each 19-residue repetitive subunit. This is consistent with the published Hyp-glycoside profiles that indicate two Hyp-linked polysaccharides for each 19-residue subunit. Our goal of identifying the precise polysaccharide attachment sites involved degradation of GAGP to yield large glycopeptides composed of short peptides (ideally ~3–5 amino acid residues) that behaved as large glycopeptides on gel filtration due to the attached polysaccharide. The short peptide sequences obtained from large glycopeptides would therefore be diagnostic for polysaccharide attachment sites in the consensus sequence.

The experimental approach involved initial pH 1 treatment of intact GAGP to remove arabinosides attached to Hyp [39]. This left the bulk of the Hyp-polysaccharide intact, ensuring that any large glycopeptides having only 5–10 amino acid residues must contain polysaccharide substituents. Thus, pH 1-treated GAGP followed by mild partial alkaline hydrolysis and size fractionation on Superdex yielded a major retarded peak containing large glycopeptides with a Hyp-polysaccharide to nonglycosylated Hyp ratio of 2:3. Cleavage of the polysaccharide (via anhydrous HF) and subsequent reverse-phase chromatography yielded the following small peptides from the large glycopeptide fraction: a major peptide Hyp-Thr-Hyp-Thr-Hyp-Hyp-Hyp-Gly and minor peptides: Thr-Hyp-Thr-Hyp-Gly and Ser-Hyp-Thr-Hyp-Thr-Hyp-Hyp-Hyp-Gly. Significantly, another Superdex peak containing much larger glycopeptides yielded the sequence Hyp-Ser-Hyp-Thr-Hyp-Thr-Hyp-Hyp-Hyp-Gly, indicating that the simple addition of Hyp-Ser to the sequence Hyp-Thr-Hyp-Thr-Hyp-Hyp-Hyp-Gly probably resulted in another polysaccharide addition site [L. Goodrum and M. Kieliszewski, unpublished data]. Overall, these results suggested that noncontiguous Hyp residues are sites of polysaccharide attachment in GAGP; however, the presence of contiguous Hyp with the noncontiguous Hyp imparted ambiguity. We have resolved this by a new approach to glycosylation site mapping using synthetic genes.

Glycosylation site mapping via synthetic genes

The results described above encouraged us to assess the predictive value of the Hyp contiguity hypothesis and refine it further by designing simple genes to generate endogenous test substrates for the glycosyl transferases involved in HRGP glycosylation [22, 106]. We designed a set of substrates, repetitive $(Ser\text{-}Pro)_n$, $(Ser\text{-}Pro_2)_n$, $(Ser\text{-}Pro_3)_n$, and $(Ser\text{-}Pro_4)_n$ to test the prediction that when targeted through the ER (endoplasmic reticulum)/Golgi, the gene products would be posttranslationally hydroxylated and glycosylated with arabinogalactan polysaccharide or arabinooligosaccharides, depending on the arrangement of Hyp residues. We also designed another repetitive substrate, $(GAGP_{consensus})_n$, based on the GAGP consensus sequence. This contained both noncontiguous and contiguous Hyp repeats and tested the predicted addition of both polysaccharide and arabinooligosaccharide. The experiments involved the following stages: (i) synthetic gene and plasmid construction and culture transformation, (ii) purification of secreted HRGP fusion proteins, (iii) structural analysis of new HRGPs.

Synthetic gene and plasmid construction

This approach applies methods used to express in *Escherichia coli*, repetitive proteins normally made by animal cells, such as mussel byssus threads [113], spider web silk [114], elastin [115] and collagen [116]. Here we applied these methods to the design and production of new HRGP glycoproteins in plant cells.

HRGPs are made up of small repetitive modules and therefore are well suited for genetic engineering using synthetic genes. Briefly, we designed overlapping synthetic oligonucleotide pairs encoding a predicted HRGP glycomodule [22, 106]. For example, the simple repeating dipeptide $(SO)_n$ (where S = Ser and O = Hyp) is predicted by the Hyp contiguity hypothesis to direct attachment of arabinogalactan polysaccharide to Hyp. It was encoded by the internal repeat oligonucleotide pair:

```
TCA CCC TCA CCA TCT  CCT TCG  CCA TCA CCC
          GGT AGA GGA AGC GGT AGT GGG
          AGT GGG AGT
S    P    S    P    S    P    S    P    S    P
```

Head-to-tail polymerization of these internal repeat pairs via their complementary sticky ends resulted in a gene ladder (fig. 4) with genes ranging in size from 1 repeat to greater than 30. Each end of the gene was capped with synthetic oligonucleotide sets, designated 5′ or 3′ linkers, having a 'sticky' end that enabled the 5′ linker set to anneal to the 5′ end of the gene and the 3′ linker set to anneal to the 3′ end of the gene. Both sets of linkers contained unique restriction sites for insertion into plasmid vectors [106].

Our plant transformation vector was pBI121 based, but encoded a tobacco extensin synthetic signal sequence [106] and the enhanced green fluorescent protein (EGFP; Clontech) in place of the GUS reporter gene. The synthetic genes were sandwiched between the signal sequence and EGFP gene (figs. 5 and 6). After transformation, BY2 tobacco lines were selected for kanamycin resistance and green fluorescence [22, 106, 117] (fig. 7).

Purification of the gene products

Secretion of the synthetic gene products into the culture medium allowed rapid collection of the crude products free of intracellular contaminants. The genes were designed without polyHis modules or other such 'purification' tags to avoid influencing the structure or function of the gene product by introducing metal binding or potential crosslinking sites. Immunoaffinity purification via anti-EGFP antibodies was an alternative, but is expensive and inefficient considering the product yields required for extensive biochemical and biophysical characterization. A straightforward chromatographic approach using gel filtration, hydrophobic interaction, or ion exchange and reverse-phase matrices combined with sensitive in-line fluorometric detection allowed large-scale

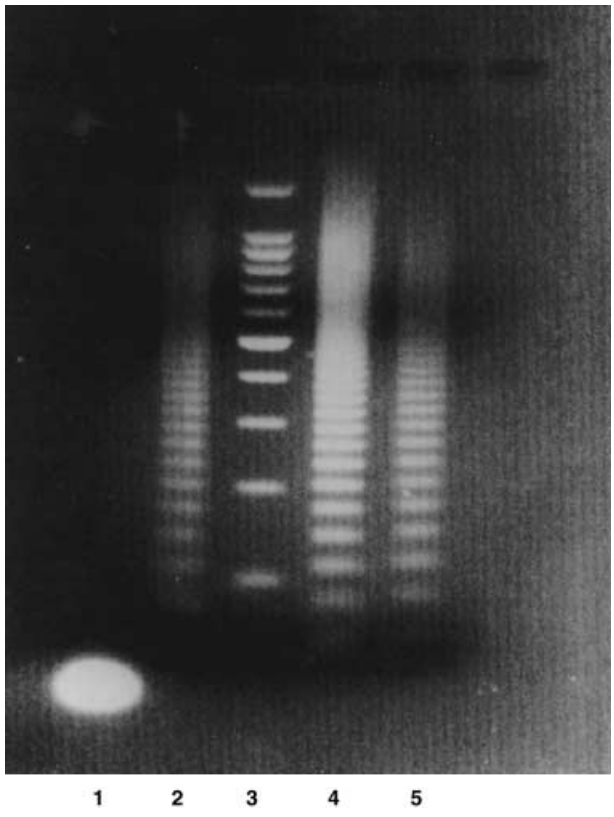

Figure 4. Synthetic gene ladder encoding Ser-Pro repeats. Complementary, nonpalindromic overlapping ends of the internal repeat oligonucleotides allowed an orderly head-to-tail polymerization to yield genes ranging in size from 1 internal repeat to >30 (Lane 4, labeled SP). Lane 1 contains the 5′-linker set and lanes 2 and 5 contain the internal repeat cassettes capped with either the 5′ linkers (lane 2) or the 3′ linkers (lane 5). Initial gene cassette design involved internal repeat oligonucleotides. Lane 3 contains 100–1500-bp markers.

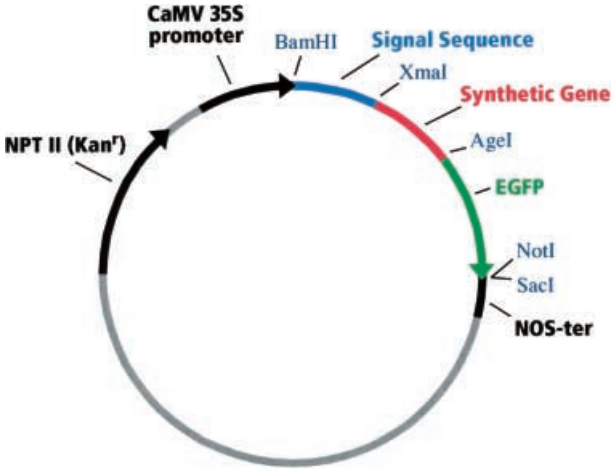

Figure 5. Plasmid map of pBI121-Sig-EGFP, the pBI121-derived vector encoding an extensin signal sequence and EGFP. The plasmid retained its 35S cauliflower mosaic virus promoter, but the pBI121 GUS reporter gene was replaced with synthetic DNA encoding an extensin signal sequence [13] and the gene for EGFP (Clontech). Synthetic gene cassettes were sandwiched between the signal sequence and EGFP as XmaI/AgeI fragments.

A

```
                                                                    ↓
GGATCCGCAATGGGAAAA ATGGCTTCTCTATTTGCCACATTTTTAGTGGTTTTAGTGTCACTTAGCTTAGCACAAACAACCCGGGCC  87
          M  G  K  M  A  S  L  F  A  T  F  L  V  V  L  V  S  L  S  L  A  Q  T  T  R  A    26

TCACCCTCACCATCTCCTTCGCCATCACCCTCACCCTCACCATCTCCTTCGCCATCACCCTCACCCTCACCATCTCCTTCGCCATCA  174
S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S    55

CCCTCACCCTCACCATCTCCTTCGCCATCACCCTCACCCTCACCATCTCCTTCGCCATCACCCTCACCCTCACCATCTCCTTCGCCA  261
P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P  S  P     84

TCACCCTCACCCTCACCGGTCGCCACC– egfp  288
S  P  S  P  S  P  V  A  T           93
```

B

```
                                                                   ↓
GGATCCGCAATGGGAAAAATGGCTTCTCTATTTGCCACATTTTTAGTGGTTTTAGTGTCACTTAGCTTAGCACAAACAACCCGGGAC  87
          M  G  K  M  A  S  L  F  A  T  F  L  V  V  L  V  S  L  S  L  A  Q  T  T  R  D    26

TCACCCTCACCAACTCCTACCGCACCACCTGGTCCACACTCTCCACCACCAACATTGTCACCCTCACCA ACTCCTACCGCACCACCT  174
S  P  S  P  T  P  T  A  P  P  G  P  H  S  P  P  P  T  L  S  P  S  P  T  P  T  A  P  P    55

GGTCCACACTCACCACCACCAACATTGTCACCCTCACCAACTCCTACCGCACCACCTGGTCCACACTCACCACCACCATCATTGTCA  261
G  P  H  S  P  P  P  T  L  S  P  S  P  T  P  T  A  P  P  G  P  H  S  P  P  P  S  L  S    84

CCCTCACCGGTCGCCACC- egfp  279
P  S  P  V  A  T          90
```

Figure 6. Synthetic genes encoding Sig-$(Ser\text{-}Pro)_{32}$-EGFP and Sig-$(GAGP)_3$-EGFP (*A* and *B*, respectively). The signal sequences (designated Sig above) are underlined, and the predicted signal peptide cleavage sites are marked with arrows.

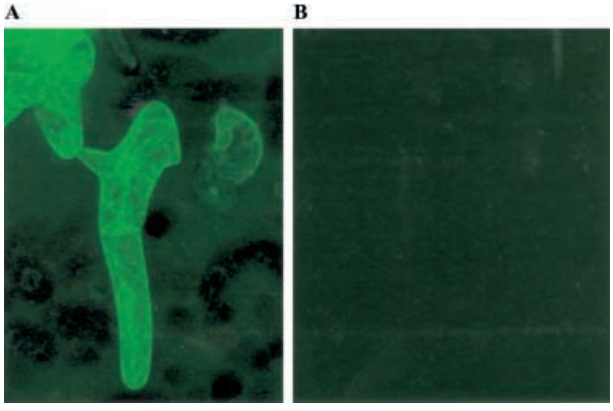

Figure 7. Fluorescence micrograph of tobacco callus cells transformed with Sig-$(GAGP)_3$-EGFP. (*A*) The signal sequence directed the gene products through the posttranslational machinery for Pro hydroxylation and Hyp glycosylation, then out to the extracellular matrix. The medium is alight with EGFP, outlining nontransformed cells, which appear as black ghosts. Panel *B* features control nontransformed tobacco cells which show no autofluoresence. The microscope was a Molecular Dynamics Sarastro 2000 confocal laser scanning microscope using a 488-nm laser wave length filter, 510-nm primary beam splitter and a 510-nm barrier filter.

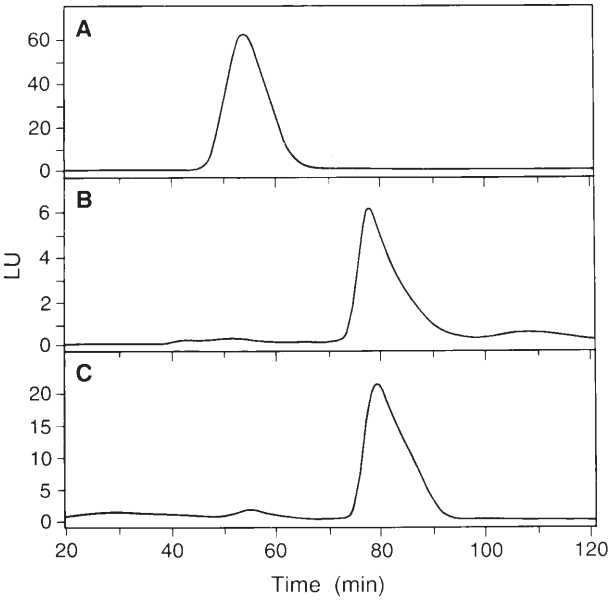

Figure 8. Gel permeation chromatography of $(Ser\text{-}Hyp)_{32}$-EGFP in the unconcentrated culture medium of transformed tobacco cells (*A*) compared with EGFP controls. Panel (*B*) is the profile of medium from cells expressing EGFP (sans glycomodule) targeted to the extracellular matrix ; panel (*C*) is 10 µg of EGFP from Clontech. We collected spent medium and fractionated it on a Superose-12 gel permeation column (Pharmacia), monitoring for EGFP fluorescence (excitation = 488 nm, emission = 520 nm). Posttranslational hydroxylation of Pro residues and glycosylation of Hyp caused a dramatic increase in the molecular mass of $(Ser\text{-}Hyp)_{32}$-EGFP, which eluted at ~53 min (*A*); the void volume eluted at ~45 min. EGFP targeted to the extracellular matrix (*B*) did not contain any Hyp.

isolation of the expression products (fig. 8). Importantly for future structural work, especially experiments involving nuclear magnetic resonance (NMR) and glycoprotein crystallization, the purified product yields were high [> 27 mg/l medium for $(SO)_{32}$-EGFP and ~8 mg/l medium for $(GAGP)_3$-EGFP].

Structural analysis of new AGPs

Analysis of the gene products of $(Ser\text{-}Pro_2)_n$ and $(Ser\text{-}Pro_4)_n$ indicated that virtually all Pro residues in the HRGP glycomodules were hydoxylated, and the resulting Hyp residues were glycosylated exclusively with arabinooligosaccharides. This was consistent with the Hyp contiguity hypothesis, which predicted arabinosylation of contiguous Hyp blocks.

In contrast, the products of $(Ser\text{-}Pro)_n$-EGFP, $(Ser\text{-}Pro_3)_n$-EGFP and GAGP-EGFP contained arabinogalactan polysaccharide, identifying them as arabinogalactan proteins [75, 97, 118]. The hydroxyproline glycoside profile of the of $(Ser\text{-}Pro)_n$-EGFP glycoprotein indicated the exclusive presence of Hyp-polysaccharide; Hyp-arabinooligosaccharides and nonglycosylated Hyp were absent. On the other hand, sequence analysis of $(GAGP)_3$-EGFP confirmed the predicted presence of both contiguous and noncontiguous Hyp, while hydroxyproline glycoside profiles of $(GAGP)_3$-EGFP showed Hyp polysaccharide, Hyp arabinosides and non-glycosylated Hyp consistent with the predictions of the Hyp contiguity hypothesis and with the known Hyp glycoside profile of native GAGP [79, 106].

The presence of arabinogalactan polysaccharide and arabinooligosaccharides in the $(Ser\text{-}Pro_3)_n$-EGFP glycoprotein was a surprise, as we expected all the Pro residues to be hydroxylated to form contiguous Hyp blocks. Contiguous Hyp blocks are predicted arabinosylation sites. However, amino acid composition and sequence analysis indicated that $(Ser\text{-}Pro_3)_n$ is incompletely hydroxylated yielding mixed contiguous/noncontiguous Hyp and a corresponding mixture of Hyp-arabinosides and Hyp-polysaccharides.

Significantly, the secondary structure of the glycosylated and deglycosylated modules as determined by circular dichroism confirmed that Hyp glycosylation is indeed sequence driven rather than a result of protein conformation [22].

General conclusions

Synthetic genes for HRGPs have both theoretical and practical ramifications. As analogs of existing periodic glycoproteins, 'designer HRGPs' are already helping decipher glycosylation codes. The exclusive presence of arabinogalactan-polysaccharide in $(Ser\text{-}Hyp)_n$-EGFP demonstrated that X-Hyp repeats are sites of polysaccharide attachment, at least when X = Ser. The designer HRGPs containing only contiguous Hyp residues, $(Ser\text{-}Hyp_2)_n$-EGFP and $(Ser\text{-}Hyp_4)_n$-EGFP, demonstrated that contiguous Hyp blocks do indeed code for Hyp arabinosylation. $(GAGP)_3$-EGFP demonstrated that addition of contiguous Hyp to an X-Hyp-X-Hyp-X-Hyp module resulted in Hyp-arabinosylation as well as addition of polysaccharide to Hyp. Other genes encoding Ala-Hyp or Thr-Hyp repeats will test the possibility that Ala-Hyp and Thr-Hyp repeats direct addition of polysaccharide to Hyp, whether or not those polysaccharides differ structurally from those placed on Ser-Hyp repeats, and help us assess the influence of amino acid side chains on Hyp glycosylation. Genes encoding $(X\text{-}Hyp\text{-}X)_n$, where n is not Hyp) repeats will test the maximum spacing allowed between Hyp residues before the enzymes which add galactose to Hyp no longer recognize it as substrate (i.e. the enzymes involved in Hyp-polysaccharide addition).

To what extent is the Hyp glycosylation code universal? Does it extend throughout the plant kingdom? The results here showing that tobacco adds both arabinogalactan and oligoarabinosides to the GAGP consensus sequence suggest that *Leguminosae* and *Solanaceae* share the same Hyp glycosylation code. Length of the galactan backbone and the details of its side-chain decoration might be sequence directed (consider Ser-Hyp or Thr-Hyp vs. Ala-Hyp repeats) or on the other hand, species specific and a function of the glycosyltransferase repertoire. Indeed, these synthetic gene products, after HF-deglycosylation, also provide simple substrates for defining the precise in vitro specificity of putative glycoprotein glycosyltransferases deduced from genomic sequences.

Designer HRGP might also help relate site-specific glycosylation (and hydroxylation) to essential molecular functions that include control of molecular conformation [i.e. as detected by circular dichroic and nuclear magnetic resonance (NMR) spectra], intermolecular crosslinking in vitro and in vivo, self-association in vitro, and cell wall self-assembly [22, 106]. We view HRGPs as block copolymers made up of a mixture of small functional modules involved in various aspects of wall self-assembly such as polysaccharide-binding modules, protein-binding modules, crosslinking modules and so on [2]. The multiplicity of HRGPs sharing common functional modules presents a special problem in defining general HRGP function via gene knockout or antisense technology because of functional redundancy where one HRGP may fill in for another. On the other hand, single repetitive glycomodules may act as competitive inhibitors of module binding and allow us to analyze HRGP function one module at a time. For example, if the extensin Ser-Hyp_4 module is required in recognition or self-recognition, a concurrently expressed repetitive Ser-Hyp_4 designer module should compete for the same binding site and inhibit the crosslinking or other function of all Ser-Hyp_4-containing extensins. Furthermore, the isolation of these novel gene products should permit us to identify their macromolecular interactions and reactions in vitro, and hence help describe how specific HRGP modules function in the extracellular matrix.

Finally, when was the Hyp glycosylation code first elaborated? It may even predate the plant-animal divergence

judging from the report of short Hyp-heterooligosaccharides containing arabinose and galactose in the primitive protist Chlamydomonas [119, 120].

At the utilitarian level, synthetic genes for HRGPs and plant gums provide a unique avenue to the de novo design of novel glycoprotein biopolymers of controlled structure and properties for use as emulsifiers, stabilizers, adjuvants, expanders and excipients in the food and pharmaceutical industries. The two novel structurally correct HRGPs discussed here, an analog of gum arabic glycoprotein and the novel $(SO)_{32}$ AGP, were produced in high yields and demonstrate the feasibility of this approach. Given the intense interest in rational plant improvement and the recent complaint that such improvement is currently limited by a shortage of cloned genes for important traits [116], why not design your own?

1 Albersheim P., Darvill A. G., O'Neill M. A., Schols H. A. and Voragen A. G. J. (1996) An hypothesis: the same six polysaccharides are components of the primary cell walls of all higher plants. Prog. Biotechnol. **14:** 47–55

2 Kieliszewski M. J. and Lamport D. T. A. (1994) Extensin: repetitive motifs, functional sites, posttranslational codes and phylogeny. Plant J. **5:** 157–172

3 Lamport D. T. A. and Catt J. W. (1981) The Encyclopedia of Plant Physiology, vol. 13 B, Glycoproteins and enzymes of the cell wall. In: Plant Carbohydrates II: Extracellular Carbohydrates, Tanner W. and Loewus F. A. (eds), pp. 133–165, Springer, New York

4 Showalter A. M. (1993) Structure and function of plant cell wall proteins. Plant Cell **5:** 9–23

5 Varner J. E. and Lin L. S. (1989) Plant cell wall architecture. Cell **56:** 231–239

6 Lamport D. T. A. (1967) Hydroxyproline-O-glycosidic linkage of the plant cell wall glycoprotein extensin. Nature **216:** 1322–1324

7 Lamport D. T. A., Katona L. and Roerig S. (1973) Galactosyl serine in extensin. Biochem. J. **133:** 125–131

8 Shakin-Eshleman S. H., Kasturi L., Spitalnik S. L. and Mellquist J. (1998) The Amino Acid Following an Asn-X-Ser/Thr Sequon Is an Important Determinant of N-Linked Core Glycosylation Efficiency. Biochemistry **37:** 6833–6837

9 Akiyama Y., Mori M. and Kato K. (1980) ^{13}C-NMR analysis of hydroxyproline arabinosides from *Nicotiana tabacum.* Agric. Biol. Chem. **44:** 2487–2489

10 Aspinall G. O. (1970) Plant gums. In: The Carbohydrates: Chemistry and Biochemistry, vol. **2B,** pp. 522–536, Pigman W. and Hordant D. (eds), Academic Press, New York

11 Fincher G. B., Stone B. A. and Clarke A. E. (1983) Arabinogalactan-Proteins: structure, biosynthesis and function. Annu. Rev. Plant Physiol. **34:** 47–70

12 Defaye J. and Wong E. (1986) Structural studies of gum arabic, the exudate polysaccharide from *Acacia senegal*. *Carbohydr.* Res. **150:** 221–231

13 Kieliszewski M., de Zacks R., Leykam J. F. and Lamport D. T. A. (1992) A repetitive proline-rich protein from the gymnosperm Douglas Fir is a hydroxyproline-rich glycoprotein. Plant Physiol. **98:** 919–926

14 Kieliszewski M. J., O'Neill M., Leykam J. and Orlando R. (1995) Tandem mass spectrometry and structural elucidation of glycopeptides from a hydroxyproline-rich plant cell wall glycoprotein indicate that contiguous hydroxyproline residues are the major sites of hydroxyproline-O-arabinosylation. J. Biol. Chem. **270:** 2541–2549

15 Chen J. and Varner J. E. (1985) Isolation and characterization of cDNA clones for carrot extensin and a proline-rich 33-kDa protein. Proc. Natl. Acad. Sci. **82:** 4399–4403

16 Wilson R. C., Long F., Maruoka E. M. and Cooper J. B. (1994) A new proline-rich early nodulin from *Medicago truncatula* is highly expressed in nodule meristematic cells. Plant Cell **6:** 1265–1275

17 Hong J. C., Nagao R. T. and Key J. L. (1987) Characterization and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean. J. Biol. Chem. **262:** 8367–8376

18 Lindstrom J. T. and Vodkin L. O. (1991) A soybean cell wall protein is affected by seed color genotype. Plant Cell **3:** 561–571

19 Dehio C. and De Bruijn F. J. (1992) The early nodulin gene SrEnod2 from *Sesbania rostrata* is inducible by cytokinin. Plant J. **2:** 117–128

20 Lobler M. and Hirsch A. M. (1993) A gene that encodes a proline-rich nodulin with limited homology to PsENOD12 is expressed in the invasion zone of *Rhizobium meliloti*-induced alfalfa root nodules. Plant Physiol. **103:** 21–30

21 Franssen H. J., Nap J.-P., Gloudemans T., Stiekema W., van Dam H., Govers F. et al. (1987) Characterization of cDNA for nodulin-75 of soybean: a gene product involved in early stages of root nodule development. Proc. Natl. Acad. Sci. **84:** 4495–4499

22 Shpak E., Barbar E., Leykam J. F. and Kieliszewski M. J. (2001) Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum.* J. Biol. Chem. 11272–11278

23 Bradley D. J., Kjellbom P. and Lamb C. J. (1992) Elicitor- and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: a novel, rapid defense response. Cell **70:** 21–30

24 Pichon M., Journet E.-P., Dedieu A., de Billy F., Truchet G. and Barker D. G. (1992) *Rhizobium meliloti* elicits transient expression of the early nodulin gene ENOD12 in the differentiating root epidermis of transgenic alfalfa. Plant Cell **4:** 1199–1211

25 Scheres B., van de Wiel C., Zalensky A., Horvath B., Spaink H., Van Eck H. et al. (1990) The ENOD12 gene product is involved in the infection process during the pea-rhizobium interaction. Cell **60:** 281–294

26 van de Wiel C., Scheres B., Franssen H., van Lierop M.-J., van Lammeren A. et al. (1990) The early nodulin transcript ENOD2 is located in the nodule parenchyma (inner cortex) of pea and soybean root nodules. EMBO J. **9:** 1–7

27 Kawalleck P., Schmelzer E., Hahlbrock K. and Somssich E. E. (1995) Two pathogen-responsive genes in parsley encode a tyrosine-rich glycoprotein (hrgp) and an anionic peroxidase. Mol. Gen. Genet. **247:** 444–452

28 Stark-Urnau M. and Mendgen K. (1995) Sequential addition of plant glycoproteins and polysaccharides at the host-parasite interface of *Uromyces vignae* and *Vigna sinensis*: evidence for endocytosis and secretion. Protoplasma **186:** 1–11

29 Wycoff K. L., Powell P. A., Gonzales R. A., Corbin D. R., Lamb C. and Dixon R. A. (1995) Stress activation of a bean hydroxyproline-rich glycoprotein promoter is superimposed on a pattern of tissue-specific developmental expression. Plant Physiol. **109:** 41–52

30 Datta K. and Marcus A. (1990) Nucleotide sequence of a gene encoding soybean repetitive proline-rich protein 3. Plant Mol. Biol. **14:** 285–286

31 Schnabelrauch L. S., Kieliszewski M. J., Upham B. L., Alizedeh H. and Lamport D. T. A. (1996) Isolation of pI 4.6 extensin peroxidase from tomato cell suspension cultures and identification of Val-Tyr-Lys as putative intermolecular cross-link site. Plant J. **9:** 477–489

32 Frueauf J., Dolata M., Leykam J. F., Lloyd E. A., Gonzales M., Van den Bosch K. et al. (2000) Peptides isolated from cell

walls of *Medicago truncatula* nodules and uninfected roots. Phytochemistry, **55**: 429–438
33 Lamport D. T. A. (1977) Structure, biosynthesis and significance of cell wall glycoproteins. In: Recent Advances in Phytochemistry, pp. 79–115, Loewus F. A. and Runeckles V. C. (eds), Plenum New York.
34 van Holst G.-J. and Varner J. E. (1984) Reinforced polyproline II conformation in a hydroxyproline-rich glycoprotein from carrot root. Plant Physiol. **74:** 247–251
35 Heckman J. W., Terhune B. T. and Lamport D. T. A. (1988) Characterization of native and modified extensin monomers and oligomers by electron microscopy and gel filtration. Plant Physiol. **86:** 848–856
36 Kieliszewski M. J., Kamyab A., Leykam J. F. and Lamport D. T. A. (1992) A histidine-rich extensin from *Zea mays* is an arabinogalactan protein. Plant Physiol. **99:** 538–547
37 Lind J. L., Bacic A., Clarke A. E. and Anderson M. A. (1994) A style-specific hydroxyproline-rich glycoprotein with properties of both extensins and AGPs. Plant J. **6:** 491–502
38 Lamport D. T. A. (1965) The protein component of primary cell walls. Adv. Bot. Res. **2:** 151–218
39 Lamport D. T. A. (1980) Structure and function of plant glycoproteins. In: The Biochemistry of Plants, Preiss J. (eds), pp. 501–541, Academic Press, New York
40 Fry S. C. (1982) Isodityrosine, a new cross-linking amino acid from plant cell-wall glycoprotein. Biochem. J. **204:** 449–455
41 Brady J. D. and Fry S. C. (1997) Formation of di-isodityrosine and loss of isodityrosine in the cell walls of tomato cell-suspension cultures treated with fungal elicitors or H_2O_2. Plant Physiol. **115:** 87–92
42 Qi X., Behrens B. X., West P. R. and Mort A. J. (1995) Solubilization and partial characterization of extensin fragments from cell walls of cotton suspension cultures. Evidence for a covalent cross-link between extensin and pectin. Plant Physiol. **108:** 1691–1701
43 Mort A. J. (1978) Partial characterization of extensin by selective degradation of cell walls. Ph.D. dissertation, Michigan State University, East Lansing, Michigan
44 Mort A. J. and Lamport D. T. A. (1977) Anhydrous hydrogen fluoride deglycosylates glycoproteins. Analyt. Biochem. **82:** 289–309
45 Tire C., Rycke R. D., De Loose M., Inze D., Van Montagu M. and Engler G. (1994) Extensin gene expression is induced by mechanical stimuli leading to local cell wall strengthening in *Nicotiana plumbaginifolia.* Planta **195:** 175–181
46 Shirsat A. H., Bell A., Spence J. and Harris J. N. (1996) The *Brassica napus* extA extensin gene is expressed in regions of the plant subject to tensile stresses. Planta **199:** 618–624
47 Elliott K. A. and Shirsat A. H. (1998) Promoter regions of the extA extensin gene from Brassica napus control activation in response to wounding and tensile stress. Plant Mol. Biol. **37:** 675–687
48 Cleland R. E. and Karlsnes A. (1967) A possible role for hydroxyproline-containing proteins in the cessation of cell elongation. Plant Physiol. **42:** 669–671
49 Klis F. M. and Eeltink H. (1979) Changing arabinosylation patterns in wall-bound hydroxyproline in bean cultures. Planta **144:** 479–484
50 Monro J. A., Penny D. and Bailey R. W. (1976) The organization and growth of primary cell walls of *Lupin* hypocotyl. Phytochem. **15:** 1193–1198
51 Sadava D. and Chrispeels M. J. (1973) Hydroxyproline-rich cell wall protein (extensin): role in the cessation of elongation in excised pea epicotyls. Biochim. Biophys. Acta **227:** 278
52 Cooper J. B., Heuser J. E. and Varner J. E. (1994) 3,4-Dehydroproline inhibits cell wall assembly and cell division in tobacco protoplasts. Plant Physiol. **104:** 747–752
53 Ahn J. H., Choi Y., Kwon Y. M., Kim S.-G., Cho Y. D. and Lee J. S. (1996) A novel extensin gene encoding a hydroxyproline-rich glycoprotein requires sucrose for its wound-inducible expression in transgenic plants. Plant Cell **8:** 1477–1490
54 Keller B. and Lamb C. J. (1989) Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation. Genes Dev. **3:** 1639–1646
55 Arsenijevic-Maksimovic I., Broughton W. J**.** and Krause A. (1997) Rhizobia modulate root-hair-specific expression of extensin genes. Mol. Plant-Microbe Interact. **10:** 95–101
56 Christiansen H., Hansen A. C., Vijn I., Pallisgaard N., Larsen K., Yang W.-C. et al. (1996) A novel type of DNA-binding protein interacts with a conserved sequence in an early nodulin ENOD12 promoter. Plant Mol. Biol. **32:** 809–821
57 Basile D. V. (1990) Morphoregulatory role of hydroxyproline-containing proteins in liverworts. In: Bryophyte Development: Physiology and Biochemistry, pp. 225–243, Chopra R. N. et al., CRC Press, Baton Rouge, LA
58 Hammerschmidt R., Lamport D. T. A. and Muldoon E. P. (1984) Cell wall hydroxyproline enhancement as an early event in the resistance of cucumber to *Cladosporium cucumerinum*. Physiol. Plant Pathol. **24:** 43–47
59 Sheng J., D'Ovidio R. and Mehdy M. C. (1991) Negative and positive regulation of a novel proline-rich protein mRNA by fungal elicitor and wounding. Plant J. **1:** 345–354
60 Suzuki H., Wagner T. and Tierney M. L. (1993) Differential expression of 2 soybean (Glycine max L) proline-rich protein genes after wounding, Plant Physiol. **101:** 1283–1287
61 Ahn J. H., Choi Y., Kim S. G., Kwon Y. M., Choi Y. D. and Lee J. S. (1998) Expression of a soybean hydroxyproline-rich glycoprotein gene is correlated with maturation of roots. Plant Physiol. **116:** 671–679
62 Bucher M., Schroeer B., Willmitzer L. and Riesmeier J. (1997) Two genes encoding extensin-like proteins are predominantly expressed in tomato root hair cells. Plant Mol. Biol. **35:** 497–508
63 Hood E. E., Shen Q. X. and Varner J. E. (1988) A developmentally regulated hydroxyproline-rich glycoprotein in maize pericarp cell walls. Plant Physiol. **87:** 138–142
64 Raggi V. (1998) Hydroxyproline-rich glycoprotein accumulation in TMV-infected tobacco showing systemic acquired resistance to powdery mildew. J. Phytopathol. **146:** 321–325
65 Ruiz-Avila L., Ludevid M. D. and Puigdomenech P. (1991) Differential expression of a hydroxyproline-rich cell-wall protein gene in embryonic tissues of *Zea Mays* L. Planta **184:** 130–136
66 Sheng J., Jeong J. and Mehdy M. C. (1993) Developmental regulation and phytochrome-mediated induction of mRNAs encoding a proline-rich protein, glycine-rich proteins, and hydroxyproline-rich glycoproteins in *Phaseolus vulgaris* L. Proc. Natl. Acad. Sci. USA **90:** 828–832
67 Tierney M. L., Wiechert J. and Pluymers D. (1988) Analysis of the expression of extensin and p33-related cell wall proteins in carrot and soybean. Mol. Gen. Genet. **211:** 393–399
68 Lamport D. T. A. (1973) Is the primary cell wall a protein-glycan network? First International Protoplast Colloquium, Versailles. INRA., Collogues Internationaux CNRS, no. 212, 27–31
69 Lamport D. T. A. (1969) The isolation and partial characterization of hydroxyproline-rich glycopeptides obtained by enzymic degradation of primary cell walls. Biochemistry **8:** 1155–1163
70 Stafstrom J. P. and Staehelin L. A. (1986) The role of carbohydrate in maintaining extensin in an extended conformation. Plant Physiol. **81:** 242–246
71 Allen A. K., Desai N. N., Neuberger A. and Creeth J. M. (1978) Properties of potato lectin and the nature of its glycoprotein linkages. Biochem. J. **171:** 665–674

72 Ashford D., Manon R., Allen A. K. and Neuberger A. (1982) Structural studies of the carbohydrate moieties of lectins from potato (*Solanum tuberosum*) tubers and thornapple (*Datura stramonium*) seeds. Biochem. J. **201:** 199–208

73 Brownleader M. D., Byron O., Rowe A., Trevan M., Welham K. and Dey P. M. (1996) Investigations into the molecular size and shape of tomato extensin. Biochem. J. **320:** 577–583

74 Leffler H. (1997) Introduction to galactins. *Trends Glycosci. Glycotechnol.* **9:** 9–19

75 Sommer-Knudsen J., Bacic A. and Clarke A. E. (1997) Hydroxyproline-rich plant glycoproteins. Phytochem. **47:** 483–497

76 Pope D. G. (1977) Relationships between hydroxyproline-containing proteins secreted into the cell wall and medium by suspension-cultured *Acer pseudoplatanus* cells. Plant Physiol. **59:** 894–900

77 Baldwin T. C., McCann M. C. and Roberts K. (1993) A novel hydroxyproline-deficient arabinogalactan protein secreted by suspension-cultured cells of *Daucus carota*. Plant Physiol. **103:** 115–123

78 Mollard A. and Joseleau J. P. (1994) *Acacia senegal* cells cultured in suspension secrete a hydroxyproline-deficient arabinogalactan-protein. Plant Physiol. Biochem. **32:** 703–709

79 Qi W., Fong C. and Lamport D. T. A. (1991) Gum arabic glycoprotein is a twisted hairy rope: a new model based on O-galactosylhydroxyproline as the polysaccharide attachment site. Plant Physiol. **96:** 848–855

80 Cheung A. Y., May B., Kawata E. E., Gu Q. and Wu H. (1993) Characterization of cDNAs for stylar transmitting tissue-specific proline-rich proteins in tobacco. Plant J. **3:** 151–160

81 Sommer-Knudsen J., Clarke A. E. and Bacic A. (1996) A galactose-rich, cell-wall glycoprotein from styles of *Nicotiana alata*. Plant J. **9:** 71–83

82 Sharon N. (1986) IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature of glycoproteins, glycopeptides and peptidoglycans. Glyconjugate J. **3:** 123–134

83 Pye D. A. (1997) Glycosaminoglycans: sulfated polysaccharides of the cell surface and extracellular matrix. Bioelectrochem. Princ. Pract. **5:** 360–384

84 Iozzo R. V. (1998) Matrix proteoglycans: from molecular design to cellular function. Annu. Rev. Biochem. **67**: 609–652

85 Knox R. B. and Clarke A. E. (1977) Localization of AGPs in plant tissues. AGP News **1:** 42–45

86 Komalavilas P., Zhu J. and Nothnagel E. A. (1991) Arabinogalactan-proteins from the suspension culture medium and plasma membrane of rose cells. J. Biol. Chem. **266:** 15956–15965

87 Larkin P. (1977) Arabinogalactan proteins in protoplast agglutination studies. AGP News **1:** 36–38

88 Norman P. M., Kjellbom P., Bradley D. J., Hahn M. G. and Lamb C. J. (1990) Immunoaffinity purification and biochemical characterization of plasma membrane arabino-galactan-rich glycoproteins of *Nicotiana glutinosa.* Planta **181:** 365–373

89 Zhu J.-K., Bressan R. A. and Hasegawa P. M. (1993) Loss of arabinogalactan-proteins from the plasma membrane of NaCl-adapted tobacco cells. Planta **190:** 221–226

90 Youl J. J., Bacic A. and Oxley D. (1998) Arabinogalactan-proteins from *Nicotiana alata* and Pyrus communis contain glycosylphosphatidylinositol membrane anchors. Proc. Natl. Acad. Sci. USA **95:** 7921–7926

91 Svetek J., Yadav M. P. and Nothnagel E. A. (1999) Presence of a glucosylphosphatidylinositol lipid anchor on rose arabinogalactan proteins. J. Biol. Chem. **274**: 14724–14733

92 Sherrier J. D., Prime T. A. and Dupree P. (1999) Glucophosphatidylinositol-anchored cell-surface proteins from *Arabidopsis.* Electrophoresis **20**: 2027–2035

93 Roy S., Jauh G. Y., Hepler P. K. and Lord E. M. (1998) Effects of Yariv phenylglycoside on cell wall assembly in the lily pollen tube. Planta **204:** 450–458

94 Taylor L. P. and Hepler P. K. (1997) Pollen germination and tube growth. Annu. Rev. Plant Physiol. Plant Mol. Biol. **48:** 461–491

95 Kwan J. S. and Morvan H. (1995) Characterization of extracellular beta(1,4)-xylan backbone O-substituted by arabinogalactans type II in a plant cell suspension. Carbohydr. Polym. **26:** 99–107

96 Knox J. P. (1995) Developmentally regulated proteoglycans and glycoproteins of the plant cell surface. FASEB J. **9:** 1004–1012

97 Nothnagel E. A. (1997) Proteoglycans and related components in plant cells. Int. Rev. Cytol. **174:** 195–291

98 Pont-Lezica R. F., McNally J. G. and Pickard B. G. (1993) Wall-to-membrane linkers in onion epidermis: some hypotheses. Plant Cell Environ. **16:** 111–123

99 Bacic A., Du H., Stone B. A. and Clarke A. E. (1996) Arabinogalactan proteins: a family of cell-surface and extracellular matrix plant proteoglycans. Essays Biochem. **31:** 91–101

100 Gane A. M., Craik D., Munro S. L. A., Howlett G. J., Clarke A. E. and Bacic A. (1995) Structural analysis of the carbohydrate moiety of arabinogalactan-proteins from stigmas and styles of *Nicotiana alata*. Carbohydr. Res. **277:** 67–85

101 Kjellbom P., Snogerup L., Stohr C., Reuzeau C., McCabe P. F. and Pennell R. I. (1997) Oxidative crosslinking of plasma membrane arabinogalactan proteins. Plant J. **12:** 1189–1196

102 Cheung A. Y., Zhan X., Wang H. and Wu H. (1996) Organ-specific and agamous-regulated expression and glycosylation of a pollen tube growth-promoting protein. Proc. Natl. Acad. Sci. **93:** 3853–3858

103 Rubinstein A. L. (1995) Investigation of the role of extensin-like proteins in pollination and pollen tube growth (hydroxyproline, glycoproteins, maize). Diss. Abstr. Int. B **56:** 3606–3606

104 Hansen J. E., Lund O., Nielsen J. O., Hansen J.-E. S. and Brunak S. (1996) O-GLYCBASE: a revised database of O-glycosylated proteins. Nucleic Acids Res. **24:** 248–252

105 Kieliszewski M. J. (1999) Synthetic genes for plant gums containing repeat motifs. Int. Patent Appl. PIXXD2 and Publ. WO 9903978

106 Shpak E., Leykam J. and Kieliszewski M. J. (1999) Synthetic genes for the elucidation of glycosylation codes in arabinogalactan- and other hydroxyproline-rich glycoproteins of plants. Proc. Natl. Acad. Sci. USA **96:** 14736–14741

107 Chikamai B. N., Banks W. B., Anderson D. M. W. and Weiping W. (1996) Processing of gum arabic and some new opportunities. Food Hydrocolloids **10:** 309–316

108 Phillips G. O., Takigami S. and Takigami M. (1996) Hydration characteristics of the gum exudate from *Acacia senegal*. Food Hydrocoll. **10:** 11–19

109 Prakash A. and Mangino M. E. (1990) The effects of added proteins on the functionality of gum arabic in soft drink emulsion systems. Food Hydrocolloids **4:** 177–184

110 Randall R. C., Phillips G. O**.** and Williams P. A. (1988) The role of the proteinaceous component on the emulsifying properties of gum arabic. Food Hydrocolloids **2:** 131–140

111 Ray A. K., Bird P. B., Iacobucci G. A. and Clark B. C. Jr (1995) Functionality of gum arabic. Fractionation, characterization and evaluation of gum fractions in citrus oil emulsions and model beverages. Food Hydrocolloids **9:** 123–131

112 Goodrum L. J., Patel A., Leykam J. F. and Kieliszewski M. J. (2000) Gum arabic glycoproteins contains glycomodules of both extensin and arabinogalactan-glycoproteins. Phytochemistry **54:** 99–106

113 Salerno A. J. and Goldberg I. (1993) Cloning, expression, and characterization of a synthetic analog to the bioadhesive precursor protein of the sea mussel *Mytilus edulis*. Appl. Microbiol. Biotechnol. **39:** 221–226

114 Lewis R. V., Hinman M., Kothakota S. and Fournier M. (1996) Expression and purification of a spider silk protein: a new

strategy for producing repetitive proteins. Protein Expression Purif. **7:** 400–406
115 Zhang X., Urry D. W. and Daniell H. (1996) Expression of an environmentally friendly synthetic protein- based polymer gene in transgenic tobacco plants. Plant Cell Reports **16:** 174–179
116 Heslot H. (1998) Artificial fibrous proteins: a review. Biochimie **80:** 19–31
117 McCormick S., Niedermeyer J., Fry J., Barnason A., Horsch R. and Fraley R. (1986) Leaf disc transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumefaciens*. Plant Cell Reports **5:** 81–84
118 Jermyn M. A. and Yeow Y. M. (1975) A class of lectins present in the tissues of seed plants. Aust. J. Plant Physiol. **2:** 501–531
119 Miller D. H., Lamport D. T. A. and Miller M. (1972) Hydroxyproline heterooligosaccharides in *Chlamydomonas*. Science **176:** 918–920
120 Woessner J. P. and Goodenough U. W. (1992) Zygote and vegetative cell wall proteins in *Chlamydomonas reinhardtii* share a common epitope, $(SerPro)_x$. Plant Sci. **83:** 65–76
121 Mehta A., Natarajan K., Raina A. and Datta A. (1994) A step towards developing transgenic plants with high nutritional quality. Proc. Indian. Natl. Sci. Acad., Part B. **60:** 375–380
122 Chen C.-G., Pu Z.-Y., Moritz R. L., Simpson R. J., Bacic A., Clarke A. E. et al. (1994) Molecular cloning of a gene encoding an arabinogalactan-protein from pear (*Pyrus communis*) cell suspension culture. Proc. Natl. Acad. Sci. USA **91:** 10305–10309
123 Mau S.-L., Chen C.-G., Pu Z.-Y., Moritz R. L., Simpson R. J., Bacic A. et al. (1995) Molecular cloning of cDNAs encoding the protein backbones of arabinogalactan-proteins from the filtrate of suspension-cultured cells of *Pyrus communis* and *Nicotiana alata*. Plant J. **8:** 269–281
124 Chen C.-G., Mau S.-L. and Clarke A. E. (1993) Nucleotide sequence and style-specific expression of a novel proline-rich protein gene from *Nicotiana alata*. Plant Mol. Biol. **21:** 391–395
125 Loopstra C. A. and Sederoff R. R. (1995) Xylem-specific gene expression in loblolly pine. Plant Mol. Biol. **27:** 277–291
126 Li S. and Showalter A. M. (1996) Cloning and developmental/stress regulated expression of a gene encoding a tomato arabinogalactan protein. Plant Mol. Biol. **32:** 641–652
127 Du H., Simpson R. J., Moritz R. L., Clarke A. E. and Bacic A. (1994) Isolation of the protein backbone of an arabinogalactan-protein from the styles of *Nicotiana alata* and characterization of a corresponding cDNA. Plant Cell **6:** 1643–1653
128 Qiu X., Wu Y., Du S. and Erickson L. (1997) A new arabinogalactan protein-like gene expressed in the pollen of alfalfa. Plant Sci. **124:** 41–47
129 Smith J. J., Muldoon E. P., Willard J. J. and Lamport D. T. A. (1986) Tomato extensin precursors P1 and P2 are highly periodic structures. Phytochem. **25:** 1021–1030
130 Zhou J., Rumeau D. and Showalter A. M. (1992) Isolation and characterization of two wound-regulated tomato extensin genes. Plant Mol. Biol. **20:** 5–17
131 Hong J. C., Nagao R. T. and Key J. L. (1989) Developmentally regulated expression of soybean proline-rich cell wall protein genes. Plant Cell **1:** 937–943
132 Kieliszewski M. and Lamport D. T. A. (1987) Purification and partial characterization of a hydroxyproline-rich glycoprotein in a graminaceous monocot, *Zea mays*. Plant Physiol. **85**: 823–827
133 Kieliszewski M. J., Leykam J. F. and Lamport D. T. A. (1990) Structure of the threonine-rich extensin from *Zea mays*. Plant Physiol. **92**: 316–326
134 Kieliszewski M. J., Kamyab A., Leykam J. F. and Lamport D. T. A. (1992) A histidine-rich extensin from *Zea mays* is an arabinogalactan protein. Plant Physiol. **99**: 538–547
135 Li X., Kieliszewski M. and Lamport D. T. A. (1990) A Chenopod extensin lacks repetitive tetrahydroxyproline blocks. Plant Physiol. **92**: 327–333
136 Kieliszewski M. J., Showalter A. M. and Leykam J. F. (1994) Potato lectin: a modular protein sharing sequence similarities with the extensin family, the hevein lectin family, and snake venom disintegrins (platelet aggregation inhibitors). Plant J. **5**: 849–861


To access this journal online:
http://www.birkhauser.ch