Plant Physiol. (1992) 98, 919–926
0032-0889/92/98/0919/08/$01.00/0

Received for publication May 28, 1991
Accepted August 23, 1991

A Repetitive Proline-Rich Protein from the Gymnosperm Douglas Fir Is a Hydroxyproline-Rich Glycoprotein[1]

Marcia Kieliszewski*[2], Renate de Zacks, Joseph F. Leykam, and Derek T. A. Lamport

Biochemistry Department (M.K., J.F.L.) and Michigan State University-Department of Energy Plant Research Laboratory (R.d.Z., D.T.A.L.), Michigan State University, East Lansing, Michigan 48824

ABSTRACT

Intact cell elution of suspension cultures derived from Douglas fir, *Pseudotsuga menziesii* (Mirbel) Franco, yielded two extensin monomers, the first hydroxyproline-rich glycoproteins (HRGPs) to be isolated from a gymnosperm. These HRGPs resolved on Superose-6 gel filtration. The smaller monomer was compositionally similar to angiosperm extensins like tomato P1. The larger monomer had a simple composition reminiscent of repetitive proline-rich proteins (RPRPs) from soybean cell walls and contained proline, hydroxyproline, and sugar; hence designated a proline-hydroxyproline-rich glycoprotein (PHRGP). The simple composition of the PHRGP implied a periodic structure which was confirmed by the simple chymotryptic map and 45-residue partial sequence of the major proline-hydroxyproline-rich glycoprotein chymotryptide 5: Lys-Pro-Hyp-Val-Hyp-Val-Ile-Pro-Pro-Hyp-Val-Val-Lys-Pro-Hyp-Hyp-Val-Tyr-Lys-Pro-Hyp-Val-Hyp-Val-Ile-Pro-Pro-Hyp-Val-Val-Lys-Pro-Hyp-Hyp-Val-Tyr-Lys-Ile-Pro-Pro(Hyp)-Val-Ile-Lys-Pro. Proline-hydroxyproline-rich glycoprotein chymotryptide 5 contained an 18-residue tandem repeat devoid of tetra(hydroxy)-proline or serine; it also contained two instances of the five-residue motif Hyp-Hyp-Val-Tyr-Lys and five of the general Pro-Pro-X-X-Lys motif, thereby establishing its homology with typical angiosperm RPRPs and extensins from tomato, petunia, carrot, tobacco, sugar beet, and *Phaseolus*. Unlike the nonglycosylated soybean RPRP, the highly purified Douglas fir PHRGP was lightly glycosylated, confirmed by a quantitative hydroxyproline glycoside profile, indicating that extensins can range from highly glycosylated hydroxyproline to little or no glycosylated hydroxyproline. Comparison of extensin sequence data strongly indicates that a major determinant of hydroxyproline glycosylation specificity is hydroxyproline contiguity: extensins with tetrahydroxyproline blocks are very highly arabinosylated (>90% hydroxyproline glycosylated), tri- and dihydroxyproline are less so, and single hydroxyproline residues perhaps not at all. Despite high yields of extensins eluted from intact cells, the Douglas fir cell wall itself was hydroxyproline poor yet remarkably rich in protein (>20%), again emphasizing the existence of other structural cell wall proteins that are neither HRGPs nor glycine-rich proteins.

Recently, we (21) suggested that the new classes of HRGPs[3], RPRPs, and proline-rich nodulins are demonstrably homologous with extensin and are, therefore members of the same protein family. This connection was initially obscured by highly variable sequences and by the fact that the first extensin sequences contained the "diagnostic" motif Ser-Hyp-Hyp-Hyp-Hyp which was absent from RPRPs and nodulins (2, 5, 6, 12, 13, 19). In contrast, the repetitive motif Pro-Pro-Val-Tyr-Lys (and its relatives) is common to many extensins (6, 21, 26, 28), RPRPs (2, 5, 7, 8, 13, 22, 23), and nodulins (12, 25, 32).

Although a mixture of both conserved and variable domains characterizes the extensin HRGP family, polysaccharide wall components are more highly conserved (29, 30). Thus, it seems of profound interest to ask why one structural component of the primary wall should be intrinsically more variable than the others yet still retain highly conserved motifs.

Therefore, we are attempting to define functional domains of the extensins, first by identifying the most highly conserved motifs, followed by correlation of those domains and their variants with biological function; hence, our current attempts to isolate extensin monomers from plants more "primitive" than the advanced angiosperms whose extensins have been studied almost exclusively.

So far, we have isolated extensin monomers from cell cultures of several gymnosperms: the primitive *Gingko* and *Cycas*, the advanced *Gnetum*, and the typical gymnosperm Douglas fir (*Pseudotsuga*) described here. Our culture line derived recently (June 1989) from the crown of 4-week-old Douglas fir seedlings yielded a finely divided, somewhat clumpy, culture with no evidence of secondary thickening. Intact cell elution of rapidly growing cells yielded two extensin-like monomers resolved by Superose-6 gel filtration. The smaller monomer was compositionally similar to typical angiosperm tomato glycosylated extensin type 1 (26–28) and remains to be characterized further. As the larger monomer was notably deficient in serine with a simple composition reminiscent of RPRPs (8, 23), we characterized it by peptide

[1] Supported by U.S. Department of Agriculture grant No. 88-37261-3682 and U.S. Department of Energy grant No. DE-FG02-91ER2-0021.

[2] Present address: Complex Carbohydrate Research Center, The University of Georgia, Athens, GA 30602.

[3] Abbreviations: HRGP, hydroxyproline-rich glycoprotein; ngh, nonglycosylated hydroxyproline; PC5, PHRGP chymotryptide 5; PHRGP, proline-hydroxyproline-rich glycoprotein; RPRP, repetitive proline-rich protein; THRGP, threonine-hydroxyproline-rich glycoprotein; dH_2O, distilled water; AU, absorbance units; mAU, milliabsorbance units.

mapping, sugar analysis, and a quantitative hydroxyproline glycosylation profile.

The simple chymotryptic peptide map implied a periodic structure, confirmed by a 45-residue partial sequence of the major chymotryptide which contained an 18-residue tandem repeat devoid of tetra(hydroxy)proline or serine. The major peptide also contained two instances of the five-residue motif Hyp-Hyp-Val-Tyr-Lys and five instances of the general Pro-Pro-X-X-Lys motif, thereby establishing its homology with typical angiosperm RPRPs (2, 5, 7, 8, 12, 13, 22, 25, 31, 32) and extensins from tomato (27, 28), petunia (26), carrot (6, 31), tobacco (M. Kieliszewski, unpublished sequence), and sugar beet (21). Sugar analysis showed small, but significant, amounts of arabinose, suggesting that Douglas fir extensin was lightly glycosylated, which was confirmed by alkaline hydrolysis and positive identification of hydroxyproline arabinosides. This is the first isolation of a gymnosperm extensin and the first indication that some RPRPs are glycosylated.

MATERIALS AND METHODS

Douglas Fir Suspension Cultures

Seeds of Douglas fir, *Pseudotsuga menziesii* (Mirbel) Franco (Lincoln National Forest) were obtained from F.W. Schumacher Co., Inc., Sandwich, MA, and sown on a vermiculite/perlite mixture (1:1). After 4 weeks, hypocotyl and crown segments were excised and surface sterilized by dipping in 95% ethanol, followed by 5 min in 25% Clorox bleach, and then rinsing with sterile dH_2O. Callus grew on solid medium (pH 5.6) containing one-half concentration Murashige-Skoog salts, 25 g/L sucrose, 5×10^{-6} M 2,4-D, 2.5×10^{-6} M benzyladenine, 10^{-6} M L-glutamine, and 2.5 g/L Gelrite (Kelco). The suspension-cultured cells were maintained on medium (pH 5.6) containing one-half strength Murashige-Skoog salts, 10^{-5} M 2,4-D, 20 g/L glucose, 10^{-6} M L-glutamine, shaken on a rotary shaker at 22°C, and subcultured every 7 to 9 d.

Cell Wall Preparation

Douglas fir cell walls were prepared by suspending the cells in dH_2O, cooling in an ice bath, and sonicating the cells 10 min, followed by washing with 100 mM $AlCl_3$, then alternating dH_2O washes, and pelleting the walls by bench-top centrifugation until the walls were free of cell debris, judged by microscopic examination.

Preparation of Crude HRGP by Intact Cell Elution

Batches of crude Douglas fir HRGP were prepared from 6- or 7-d cultures harvested on a 1-L coarsely sintered funnel, followed by a water wash and then gentle agitation of the cell pad in 500 mL of 50 mM $AlCl_3$ for 3 min before final suction. The eluate was reduced in volume to 100 mL by lyophilization before adding TCA (final concentration, 10%, w/v, 18 h, 4°C). Centrifugation (13,000*g*, 45 min, SS-34 rotor) yielded a hydroxyproline-rich supernatant which was dialyzed 36 h at 4°C and then freeze dried. We designated this TCA-soluble fraction "crude PHRGP" and purified it as described below (Fig. 1).

Gel Filtration Chromatography on Superose-6

Crude PHRGP (10–30 mg) was dissolved in 2 mL 200 mM sodium phosphate buffer (pH 7) and 0.02% sodium azide and injected onto a Pharmacia preparative Superose-6 fast protein liquid chromatography gel filtration column, which eluted at a flow rate of 60 mL/h. The eluant was then collected and dialyzed against dH_2O, and peaks P1, P2, and P3 were freeze dried (Fig. 2).

Polysulfoethyl Aspartamide SCX Ion Exchange Chromatography

Superose P1 (1–7 mg) was dissolved in Poly A start buffer (10 mM NaH_2PO_4/phosphoric acid, 10% MeCN, pH 3), applied to a semipreparative Polysulfoethyl Aspartamide column (9.4 × 200 mm) (PolyLC Inc., Columbia, MD) equilibrated with Poly A start buffer, and eluted with a buffered 0 to 1 M NaCl gradient at a flow rate of 1.5 mL/min, monitoring at A_{220} and A_{280}.

SDS Gel Electrophoresis

PHRGP (7 μg) was loaded in 20 μL sample buffer (0.01 M Trizma base, 1% SDS, 0.001 M EDTA, 5% β-mercaptoethanol) and separated onto a 12% polyacrylamide gel. The gels were run in Tris-Gly buffer (0.025 M Trizma base, 0.192 M glycine, 0.1% SDS). Protein was stained with 0.2% Coomassie Brilliant Blue R-250 in water:ethanol:acetic acid (25:25:10, v/v).

Sugar Analysis

Sugars were analyzed as their alditol acetates by GC using a 6-foot × 2-mm i.d. PEG succinate 224 column programmed

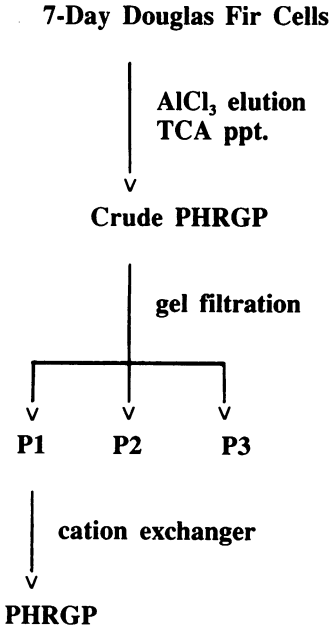

Figure 1. PHRGP purification flow chart. ppt., Precipitate.

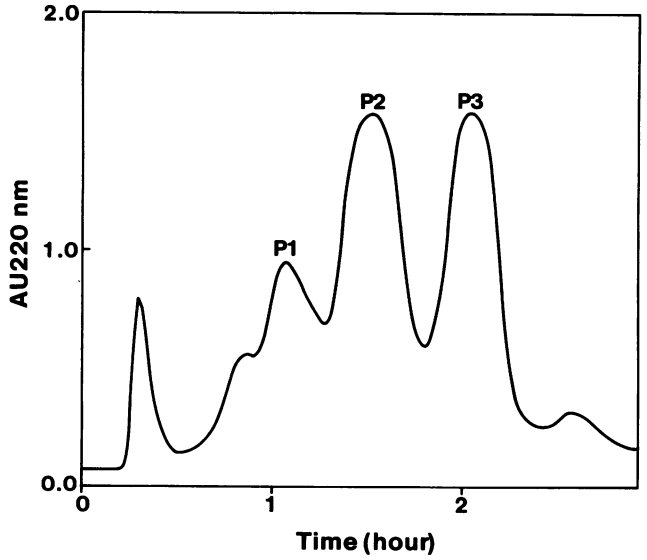

Figure 2. Gel filtration on Superose-6 of the Douglas fir crude PHRGP. Peak (P1) contained the PHRGP, peak 2 (P2) contained an HRGP whose amino acid composition resembled tomato glycosylated type 1 extensins from the advanced herbaceous dicots (27), and peak 3 was hydroxyproline poor.

from 130 to 180° at 4°C/min for neutral sugars. Data capture was by a Compaq 386 with Nelson Turbochrom II software.

Hydroxyproline Arabinoside Profile

Hydroxyproline arabinosides were determined after alkaline hydrolysis (0.44 N $Ba(OH)_2$, 18 h, 105°C) of PHRGP or Douglas fir cell walls and careful neutralization with concentrated H_2SO_4. The supernatant fraction was then centrifuged and freeze dried. The lyophilate was redissolved in dH_2O, and 0.5 mL containing 100 to 200 μg hydroxyproline was applied to a 75- × 0.6-cm column (H^+ form) of Technicon Chromobeads C washed with water and eluted with a 0 to 0.5 N HCl gradient. The postcolumn hydroxyproline assay reaction was monitored at A_{560}.

Hydroxyproline Assay

The hydroxyproline content was determined after hydrolysis (6 N HCl, 110°C, 18 h) of cell wall and PHRGP preparations as described earlier, involving alkaline hypobromite oxidation and subsequent coupling with acidic Ehrlich's reagent and monitoring at A_{560}.

Amino Acid Analysis

Amino acid compositions were determined with a Pickering cation exchange column (3-mm i.d. × 150 mm) eluted by Pickering buffers A, B, and C. Postcolumn fluorometric detection involved NaOCl oxidation and orthophthalaldehyde coupling which allowed hydroxyproline and proline detection. Data capture was by a Compaq 386 with Nelson Turbochrom II software.

Digestion of Glycosylated PHRGP with Chymotrypsin

Glycosylated PHRGP (7 mg) was digested overnight with chymotrypsin (Worthington) (2% sodium bicarbonate [pH 8], 10 mM $CaCl_2$, substrate:enzyme ratio 100:1) and then freeze dried.

Dearabinosylation of PHRGP Chymotryptides with 0.1 N HCL

The chymotryptic peptides were dearabinosylated by heating in 0.1 N HCl (pH 1) at 100°C for 1 h, the peptides were freeze dried, and peptide mapping was done as described below (Fig. 3).

HPLC Peptide Mapping

We obtained peptide maps of the deglycosylated chymotryptic peptides via reverse phase HPLC on a Hamilton PRP-1 column (4.1-mm i.d. × 150 mm) (Fig. 3) using programmed gradient elution (0.5 mL/min) with the mobile phase solvents A (0.1% TFA) and B (0.1% TFA in 80% [v/v] aqueous acetonitrile). For resolution of peptides, the gradient began at 100% A and increased (0.5%/min) from 0 to 50% B in 100 min, as described previously (28). Absorbance was monitored at 220 and 280 nm on a Hewlett-Packard photodiode array spectrophotometer. After initial peptide fractionation, the major peptide was rerun through PRP-1 to ensure purity before peptide sequencing.

Automated Edman Degradation

Peptides were sequenced by the Michigan State University Biochemistry Department Macromolecular Facility on a 477A Applied Biosystems, Inc., gas phase sequencer.

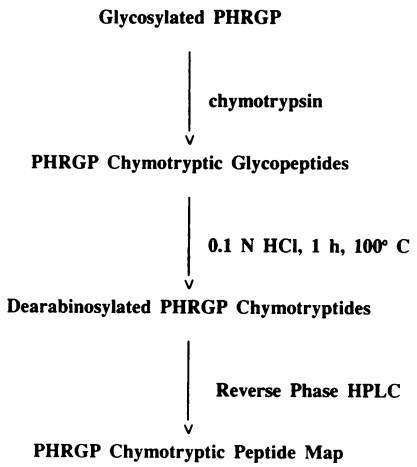

Figure 3. Flow chart for the preparation of PHRGP chymotryptides. Glycosylated PHRGP was digested with chymotrypsin, and the peptides were dearabinosylated in HCl and then fractionated on a Hamilton PRP-1 reverse phase column.

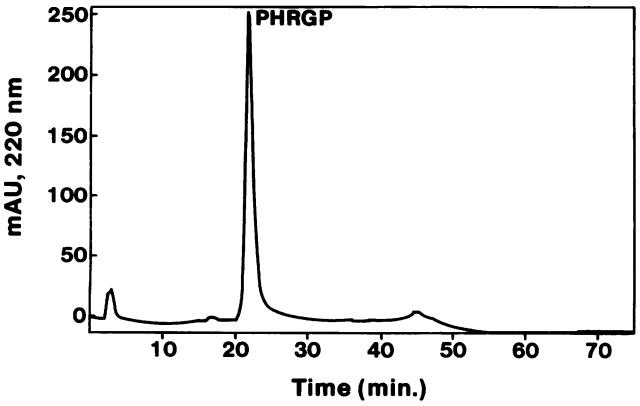

Figure 4. Cation exchange chromatography of PHRGP. Superose P1 fractionated on a Polysulfoethyl Aspartamide column to yield a pure PHRGP.

RESULTS

PHRGP Yield and Hydroxyproline Content

The cells yielded 430 to 588 μg (TCA soluble) crude PHRGP/g cells fresh weight. The crude PHRGP contained 80 to 100 μg hydroxyproline/mg dry weight compared with only 3.4 μg hydroxyproline/mg dry weight in the cell wall fraction.

Fractionation of the Crude PHRGP

Gel filtration yielded three major fractions, P1, P2, and P3, with a hydroxyproline content of 20, 12.8, and 0.2% (w/w),

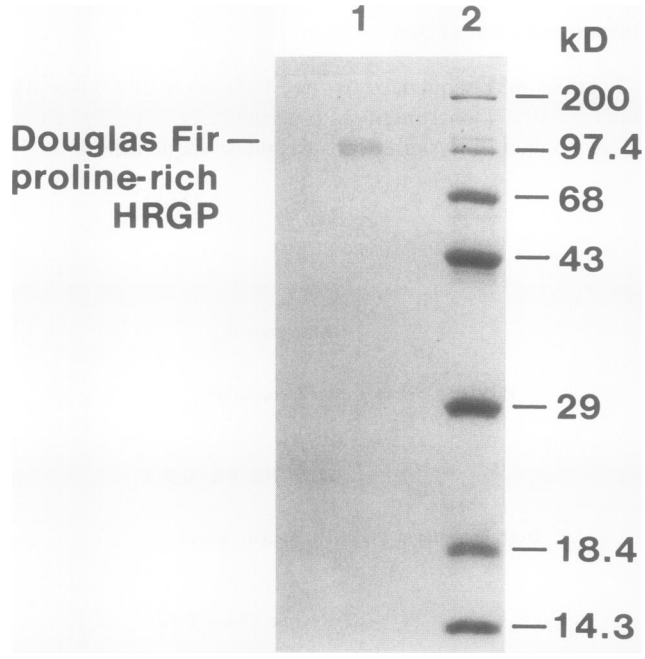

Figure 5. SDS-PAGE of the PHRGP. Lane 1, 7 μg PHRGP; lane 2, 3 μg molecular mass standards. Standards: myosin (H chain), 200 kD; phosphorylase b, 97.4 kD; BSA, 68 kD; ovalbumin, 43 kD; carbonic anhydrase, 29 kD; β-lactoglobulin, 18.4 kD; lysozyme, 14.3 kD.

Table I. *Amino Acid Compositions of Douglas Fir PHRGP, Douglas Fir P2, a Soybean RPRP, and Tomato Extensin P1*

Amino Acid	Douglas Fir PHRGP	Douglas Fir P2	Soybean 33-kD RPRP[a]	Tomato Extensin P1[b]
		mol %		
Hyp	28.1	29.2	20.0	33.5
Asx	0.0	0.0	1.2	1.8
Thr	1.5	4.5	0.4	7.2
Ser	2.5	6.8	0.0	9.5
Glx	0.0	2.9	8.9	1.9
Pro	21.3	14.2	21.5	8.3
Gly	0.6	2.0	0.0	1.6
Ala	0.0	4.1	0.0	3.2
Val	17.3	6.5	17.3	5.0
Ile	8.9	0.9	1.1	1.8
Leu	0.0	1.8	0.0	0.8
Tyr	4.9	4.6	11.5	6.3
Phe	0.0	0.7	0.0	1.3
Lys	11.0	14.9	17.3	10.5
His	1.2	6.0	0.0	7.1
Arg	1.9	0.6	0.0	1.2

[a] From Datta *et al.* (8). [b] From Smith *et al.* (27).

respectively (Fig. 2). Cation exchange chromatography of P1 on Polysulfoethyl Aspartamide yielded PHRGP as the single major peak (Fig. 4) which migrated as a single band on SDS-PAGE with an apparent molecular mass of 97.4 kD (Fig. 5).

Composition of Purified PHRGP and Superose-P2

Six amino acids Hyp, Pro, Val, Ile, Tyr, and Lys accounted for 91 mol% of the PHRGP polypeptide. Although Superose peak 2 was also rich in hydroxyproline and proline, it was notably richer in serine and histidine but contained less valine and isoleucine (Table I). Neutral sugar analysis accounted for 18% (w/w) of the PHRGP as arabinose with 3% (w/w) galactose. The arabinose to hydroxyproline ratio was 1.3:1, indicating a relatively lightly glycosylated polypeptide backbone, which the hydroxyproline arabinoside profile confirmed: 73% of the hydroxyproline was nonglycosylated, with Hyp-Ara_1 and Hyp-Ara_3 as the major glycosylated species, whereas Hyp-Ara_4 was virtually absent (Table II).

PHRGP Chymotryptic Peptide Map

HPLC fractionation of the chymotryptic digest after de-arabinosylation of the digest in 0.1 N HCl gave a peptide map consisting of five major peptides (Fig. 6). We collected peptide PC5 for amino acid analysis and sequencing.

Amino Acid Composition and Edman Degradation of PC5

The amino acid composition of PC5 corroborated the partial sequence and also indicated that the complete peptide contained 49 residues (Table III). Automated Edman degradation of PC5 gave a 45-residue partial sequence in which an 18-residue identical repeat made up the first 36 amino acids, and the pentameric motif (Pro/Hyp)-Hyp-Val-X-Lys occurred five times (Fig. 7).

Table II. *Hydroxyproline Arabinoside Profiles of Extensins from Douglas Fir, Sugar Beet[a], Maize[b], and Tomato[c], and the Douglas Fir Cell Wall*

Hyp-Ara$_n$	Douglas Fir		Sugar Beet P1	Maize THRGP	Tomato P1
	PHRGP	Wall			
			% of total Hyp		
Hyp	73	10	32	48	12
Hyp-Ara	11	6	31	15	9
Hyp-Ara$_2$	5	6	17	6	8
Hyp-Ara$_3$	11	45	16	25	33
Hyp-Ara$_4$	0	20	4	6	38

[a] From Li *et al.* (21). [b] From Kieliszewski and Lamport (17). [c] From Smith *et al.* (25).

Amino Acid Composition and Hydroxyproline Arabinoside Profile of the Douglas Fir Cell Wall

Amino acid analyses showed the Douglas fir wall was >20% (w/w) protein. Furthermore, the wall was hydroxyproline poor but rich in Asx, Glx, Ala, Gly, and Leu (Table IV). In direct contrast to the PHRGP, the fir cell wall hydroxyproline residues were mostly glycosylated, with appreciable amounts of Hyp-Ara$_4$ as the predominant species (Table II).

DISCUSSION

Salt eluates of suspension-cultured Douglas fir contained two putative HRGP extensin monomers judging from their chromatographic properties and chemical composition, including significant hydroxyproline arabinosylation of at least one of the monomers. The smaller monomer resembled the composition of a typical type 1 angiosperm extensin monomer (*e.g.* tomato P1) (Table I) and remains for future study. Because the larger monomer (97 kD via SDS-PAGE) (Figs. 2 and 5) was rich in both proline and hydroxyproline (Table I) and was also glycosylated (Table II), we designated it a PHRGP. The PHRGP was notably poor in serine, yet rich in valine and lysine, suggesting an affinity with the repetitive proline-rich cell wall proteins isolated from cultured soybean, which lack the Ser-Hyp$_4$ block characteristic of the first studied dicot extensins (2, 8, 12–14, 22).

A PHRGP chymotryptic peptide map obtained from an apparently incomplete digest of the glycosylated PHRGP (presumably due to steric hindrance by the arabinosides) gave PC5 which was the most easily purified of the major peptides and whose 45-residue (partial) sequence included an 18-residue tandem repeat: (Lys-Pro-Hyp-Val-Hyp-Val-Ile-Pro-Pro-Hyp-Val-Val-Lys-Pro-Hyp-Hyp-Val-Tyr-Lys)$_2$ (Figs. 6 and 7). This 45-residue sequence also contained two instances of the five-residue motif Hyp-Hyp-Val-Tyr-Lys and five instances of the general repetitive pentameric motifs: X-Pro-Val-Tyr-Lys, Pro-Pro-Val-X-Lys, Pro-Pro-X-Tyr-Lys (where proline is often hydroxylated). The Douglas fir PHRGP is, therefore, homologous with typical angiosperm RPRPs (2, 7, 8, 13) and extensins from tomato (28), carrot (6), petunia (26), sugar beet (21), and tobacco (26) (M. Kieliszewski, unpublished sequence). Thus, extensins, which we define by sequence homology (9), encompass proteins that contain proline/hydroxyproline residues ranging from highly contiguous (*i.e.* dicot extensins that contain the Ser-(Pro)Hyp$_{4-5}$ motif; refs. 6, 11, 15, and 28) to highly dispersed as in:

1. Nodulins and Douglas fir PHRGP: *e.g.* a (Pro)-Pro-Pro-His-Glu-Lys motif as in Nodulin-75 and ENOD2 (12, 32); Pro-Hyp-Hyp-Val-Tyr-Lys and Pro-Pro-Hyp-Val-Val-Lys in the PHRGP (Fig. 7);
2. RPRPs, typically a Pro-Hyp(Pro)-Val-Tyr/Glu-Lys motif (2, 13, 22);
3. "Split-block" extensin from sugar beet, consisting of two motifs, the split block itself, Ser-Hyp-Hyp-[Val-His-Glu-Tyr]-Hyp-Hyp, and its associated Thr-Hyp-Val-Tyr-Lys motif (note a single-base change converts this motif to the RPRP motif) (21); and
4. Graminaceous monocot THRGP extensin: Thr-Hyp-Ser-

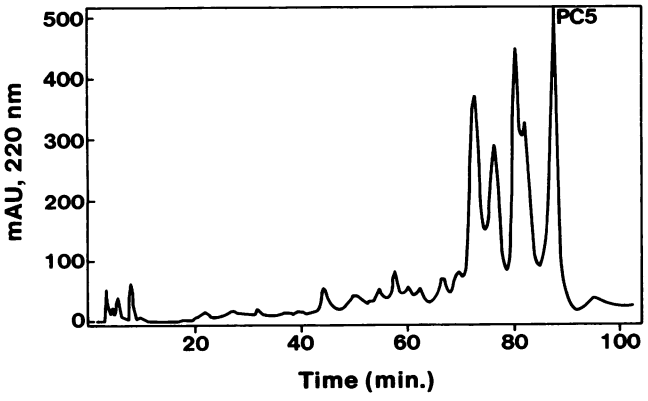

Figure 6. HPLC peptide map of the Douglas fir PHRGP. PC5 was collected for sequencing.

Table III. *Amino Acid Molar Ratios of the Major Chymotryptic Peptide Sequenced from Douglas Fir PHRGP*

	Composition by Amino Acid Analysis	Composition of Partial Sequence
Amino acid		
Hyp	17.1	12
Pro	12.6	10
Val	6.6	11
Lys	5.7	6
Ile	3.6	4
Tyr	3.0	2
Residues	48.6	45

Lys-Pro-Hyp-Val-Hyp-Val-Ile-Pro-Pro-Hyp-Val-Val-Lys-Pro-Hyp-Hyp-Val-Tyr-

Lys-Pro-Hyp-Val-Hyp-Val-Ile-Pro-Pro-Hyp-Val-Val-Lys-Pro-Hyp-Hyp-Val-Tyr-

Lys-Ile-Pro-Hyp-Hyp-Val-Ile-Lys-Pro

Figure 7. A partial amino acid sequence of PHRGP PC5.

Hyp-Lys-Pro-Hyp-Thr-Pro-Lys-Pro-Thr-Hyp-Hyp-Thr-Tyr (17, 18).

Significantly, Lys-Pro is never hydroxylated in any known extensin peptide, including the PHRGP described here. Therefore, we predict that nodulins and RPRPs also follow this rule; hence, the recent suggestion of Lys-Hyp in a nodulin (25) seems unlikely.

The PHRGP was only lightly arabinosylated (73% ngh), mainly as Hyp-Ara_3, whereas Hyp-Ara_4, whose abundance is characteristic of some dicot extensins (10, 20, 24, 27, 33), was virtually absent. Significant glycosylation of the PHRGP raises the possibility that both the RPRPs and soybean nodulins may also be glycosylated except that the very low arabinose content (1.6% dry weight) of soybean RPRPs (8) (>98% ngh; M. Kieliszewski, unpublished data) virtually excludes significant hydroxyproline glycosylation. This exceptionally high ngh in RPRPs points to a possible correlation between hydroxyproline arabinosylation and polypeptide sequence: highly contiguous hydroxyproline residues are highly glycosylated in extensins from tomato (9% ngh) (27), carrot (3% ngh) (33), and inferred from melon (6% ngh in the wall itself[4]) (24). However, interspersed hydroxyproline residues seem to be glycosylated less and, in the most dispersed (RPRPs), virtually not at all. For example, the lightly glycosylated maize THRGP (approximately 50% ngh) contains frequent dipeptidyl hydroxyproline (17, 18), whereas the split-block sugar beet extensin, with di- and tripeptidyl hydroxyproline (21), is more highly glycosylated (32% ngh) than the THRGP. On the other hand, the nonglycosylated soybean RPRPs (20 mol% hydroxyproline) probably lack contiguous hydroxyproline[5] (22) and are not glycosylated (8); yet, the very similar Douglas fir PHRGP (28 mol% hydroxyproline) is lightly glycosylated (73% ngh). This apparent contradiction is actually quite consistent with the peptide sequence data which show some dipeptidyl hydroxyproline in the PHRGP

[4] Inferred from the sequence of the major peptide isolated from melon cell walls (11) and hydroxyproline arabinoside profiles of both melon wall (10, 24) and melon HRGPs (24).

[5] Thus far, only one peptide sequence (an N-terminus) has been determined for an RPRP: XXX-Tyr-Glu-Lys-Pro-Hyp-Ile-Tyr-Lys-Pro-Hyp-Val-Tyr-Thr- (22), which corresponds to the N-terminus of a soybean RPRP clone first isolated by Hong *et al.* (13).

Table IV. *Amino Acid Compositions of Cell Walls From Tomato[a], Sugar Beet[b], Maize[c], Rice, Asparagus[c], and Douglas Fir*

Amino Acid	Tomato Culture	Sugar Beet Culture	Maize Culture	Maize Coleoptile	Rice Meristem	Asparagus Culture	Douglas Fir Culture
Hyp	28.5	0.4	1.1	2.6	0.6	2.1	0.8
Asx	4.0	12.0	10.4	8.9	11.9	8.7	9.4
Thr	4.6	5.1	5.1	5.2	4.8	4.9	5.3
Ser	14.2	7.3	6.9	7.4	6.8	7.0	7.5
Glx	2.8	12.0	9.3	10.2	11.3	10.1	11.8
Pro	3.9	5.1	3.7	5.0	7.6	5.8	5.4
Gly	3.3	9.5	10.7	10.5	11.5	13.2	9.5
Ala	3.2	8.7	10.6	11.6	14.2	10.4	9.0
Val	7.0	6.5	6.4	6.8	5.6	5.0	6.7
Cys	n.d.[d]	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Met	0.8	1.6	1.7	0.4	1.0	0.8	0.7
Ile	1.8	4.8	4.2	4.4	3.0	4.1	5.0
Leu	2.5	9.7	10.3	10.2	8.4	7.6	10.6
Tyr	6.3	1.3	1.9	1.0	0.4	1.7	1.9
Phe	1.3	3.8	4.0	3.2	3.3	3.2	3.5
Lys	10.5	7.1	6.2	6.1	6.8	7.0	6.6
His	2.7	2.2	2.1	2.3	1.6	2.9	1.6
Arg	1.2	3.3	4.7	4.3	1.3	4.9	4.6

[a] From Smith *et al.* (27). [b] From Li *et al.* (21). [c] From Benbow (3). [d] n.d., Not determined.

(Fig. 7). On this basis, we predict that at least some of the nodulins, such as ENOD2 and nodulin-75 (12, 32) which are likely to contain contiguous hydroxyproline (cDNA clones show frequent blocks of three contiguous proline residues) will also be lightly glycosylated. Thus, a correlation between hydroxyproline contiguity and glycosylation may help predict the predominant extensin type in a given tissue or species. For example, a survey of hydroxyproline arabinosides in the plant kingdom (20) now implies that the Ser-Hyp-Hyp-Hyp-Hyp (contiguous) motif is generally characteristic of the most abundant dicot extensins (3–13% ngh), whereas the more interspersed extensins may be more abundant in monocots and lower plants. The intriguing exception is the primitive gymnosperm *Gingko*, in which 13% ngh in the wall implies a predominance of highly contiguous hydroxyproline (20).

Interestingly, the PHRGP hydroxyproline-arabinoside profile was quite different from that of the Douglas fir wall (approximately 10% ngh), so the PHRGP itself (and RPRPs and hydroxyproline/proline-rich nodulins?) can only account for a small proportion, if any, of the covalently bound HRGPs in the wall and may perhaps remain as a non-cross-linked monomer, contrary to the prevailing notion that they are covalently built into the wall (8).

Despite the yield of Douglas fir elutable extensins, the wall itself is surprisingly hydroxyproline poor and, as previously noted by Burke *et al.* (4), is unusually rich in non-HRGP protein. This protein (or proteins?) has hitherto remained unrecognized as a major wall component; based purely on compositional analyses, it is widespread, occurring in other species ranging from herbaceous dicots, such as sugar beet (21), to the nongraminaceous monocot asparagus (3) and the graminaceous monocots rice and maize (Table III).

It is tempting to speculate about the origin or interconversion of extensin types. For example, as discussed earlier (21) it is easy to see how the split-block extensin of sugar beet either arose from (or gave rise to!) the canonical dicot type I motif, Ser-Hyp-Hyp-Hyp-Hyp-Thr-Hyp-Val-Tyr-Lys (28), simply by shifting the insertion/deletion sequence from the end of a tetrahydroxyproline block to its middle. Whether one should also view other extensins such as nodulins and RPRPs as split-block types is currently a moot point. One might argue that their insertion/deletion sequences are represented by His-Glu-Lys-Pro (nodulin) (32) and Val-Tyr/Glu-Lys-Pro (RPRP) (13) sandwiched between two Hyp-Hyp or Pro-Hyp dipeptides. That view may be correct if extensin evolution has really involved the apparent rampant shuffling of structural motifs by insertion, deletion, and transposition. On the other hand, the notable lack of serine in RPRPs and nodulins would appear to preclude the argument, were it not for the presence of a few serine residues in the PHRGP (see Table I) and in the recently described nodulins from pea (25, 32).

Finally, we must consider the possible origin of extensin by homoplasy, *i.e.* through convergent evolution of generally proline-rich proteins. However, this seems quite unlikely because glycosylated hydroxyproline-rich extensin-like molecules occur as major wall components in primitive protists such as *Chlamydomonas* (34) which predate the metaphytes. On the other hand, the major metazoan matrix glycoprotein, collagen, which can be regarded as the animal analog of extensin, is absent from protists, appearing first only in the lower metazoa, sponges (1). Interestingly, although collagens contain hydroxyproline (invariably nonglycosylated), they are essentially glycine-rich (33 mol%) proteins. Thus, it is not entirely inconceivable that collagen evolved from translation products of the G-rich noncoding strand of extensin as suggested by Keller and Lamb (16).

LITERATURE CITED

1. **Adams E** (1978) Invertebrate collagens. Science **202:** 591–598
2. **Averyhart-Fullard V, Datta K, Marcus A** (1988) A hydroxyproline-rich protein in the soybean cell wall. Proc Natl Acad Sci USA **85:** 1082–1085
3. **Benbow L** (1991) Isolation and characterization of extensins from the non-graminaceous monocot, *Asparagus*. Master's thesis. Michigan State University, East Lansing, MI
4. **Burke D, Kaufman P, McNeil M, Albersheim P** (1974) The structure of plant cell walls. VI. A survey of the walls of suspension-cultured monocots. Plant Physiol **54:** 109–115
5. **Chen J, Varner JE** (1985) Isolation and characterization of cDNA clones for carrot extensin and a proline-rich 33-kDa protein. Proc Natl Acad Sci USA **82:** 4399–4403
6. **Chen J, Varner JE** (1985) An extracellular matrix protein in plants: characterization of a genomic clone for carrot extensin. EMBO J **4:** 2145–2151
7. **Datta K, Marcus A** (1990) Nucleotide sequence of a gene encoding soybean repetitive proline-rich protein 3. Plant Mol Biol **14:** 285–286
8. **Datta K, Schmidt A, Marcus A** (1989) Characterization of two soybean repetitive proline-rich proteins and a cognate cDNA from germinated axes. Plant Cell **1:** 945–952
9. **Doolittle RF** (1981) Similar amino acid sequences: chance or common ancestry? Science **214:** 149–159
10. **Esquerre-Tugaye MT, Lafitte C, Mazau D, Toppan A, Touze A** (1979) Cell surfaces in plant-microorganism interactions. II. Evidence for the accumulation of hydroxyproline-rich glycoproteins in the cell wall of diseased plants as a defense mechanism. Plant Physiol **64:** 320–326
11. **Esquerre-Tugaye MT, Lamport DTA** (1979) Cell surfaces in plant-microorganism interaction. I. A structural investigation of cell wall hydroxyproline-rich glycoproteins which accumulate in fungus infected plants. Plant Physiol **64:** 314–319
12. **Franssen HJ, Nap J-P, Gloudemans T, Stiekema W, van Dam H, Govers F, Louwerse J, van Kammen A, Bisseling T** (1987) Characterization of cDNA for nodulin-75 of soybean: a gene product involved in early stages of root nodule development. Proc Natl Acad Sci USA **84:** 4495–4499
13. **Hong JC, Nagao RT, Key JL** (1987) Characterization and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean. J Biol Chem **262:** 8367–8376
14. **Hong JC, Nagao RT, Key JL** (1989) Developmentally regulated expression of soybean proline-rich cell wall protein genes. Plant Cell **1:** 937–943
15. **Keller B, Lamb CJ** (1989) Specific expression of a novel cell wall hydroxyproline-rich glycoprotein in lateral root initiation. Genes Dev **3:** 1639–1646
16. **Keller B, Sauer N, Lamb CJ** (1988) Glycine-rich cell wall protein in bean. Gene structure and association of the protein with the vascular system. EMBO J **7:** 3625–3633
17. **Kieliszewski M, Lamport DTA** (1987) Purification and partial characterization of a hydroxyproline-rich glycoprotein in a graminaceous monocot, *Zea mays*. Plant Physiol **85:** 823–827
18. **Kieliszewski M, Leykam JF, Lamport DTA** (1990) Structure of the threonine-rich extensin from *Zea mays*. Plant Physiol **92:** 316–326
19. **Lamport DTA** (1969) The isolation and partial characterization of hydroxyproline-rich glycopeptides obtained by enzymic degradation of primary cell walls. Biochemistry **8:** 1155–1163
20. **Lamport DTA, Miller DH** (1971) Hydroxyproline arabinosides in the plant kingdom. Plant Physiol **48:** 454–456

21. **Li X-B, Kieliszewski M, Lamport DTA** (1990) A chenopod extensin lacks repetitive tetrahydroxyproline blocks. Plant Physiol **92:** 327–333
22. **Lindstrom JT, Vodkin LO** (1991) A soybean cell wall protein is affected by seed color genotype. Plant Cell **3:** 561–571
23. **Marcus A, Greenberg J, Averyhart-Fullard V** (1991) Repetitive proline-rich proteins in the extracellular matrix of the plant cell. Physiol Plant **81:** 273–279
24. **Mazau D, Rumeau D, Esquerre-Tugaye MT** (1988) Two different families of hydroxyproline-rich glycoproteins in melon callus: biochemical and immunochemical studies. Plant Physiol **86:** 540–546
25. **Scheres B, Van de Wiel C, Zalensky A, Horvath B, Spaink H, Van Eck H, Zwartkruis F, Wolters A-M, Gloudemans T, Van Kammen A, Bisseling T** (1990) The ENOD12 gene product is involved in the infection process during the pea-rhizobium interaction. Cell **60:** 281–294
26. **Showalter AM, Rumeau D** (1990) Molecular biology of plant cell wall hydroxyproline-rich glycoproteins. *In* WS Adair, RP Mecham, eds, Organization and Assembly of the Animal and Plant Extracellular Matrix. Academic Press, New York, pp 247–281
27. **Smith JJ, Muldoon EP, Lamport DTA** (1984) Isolation of extensin precursors by direct elution of intact tomato cell suspension cultures. Phytochemistry **23:** 1233–1239
28. **Smith JJ, Muldoon EP, Willard JJ, Lamport DTA** (1986) Tomato extensin precursors P1 and P2 are highly periodic structures. Phytochemistry **25:** 1021–1030
29. **Thomas JR, Darvill AG, Albersheim P** (1989) Isolation and structural characterization of the pectic polysaccharide rhamnogalacturonan II from walls of suspension-cultured rice cells. Carbohydr Res **185:** 261–277
30. **Thomas JR, Darvill AG, Albersheim P** (1989) Rhamnogalacturonan. I. A pectic polysaccharide that is a component of monocot cell walls. Carbohydr Res **185:** 279–305
31. **Tierney ML, Wiechart J, 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
32. **van de Wiel C, Scheres B, Franssen H, van Lierop M-J, van Lammeren A, van Kammen A, Bisseling T** (1990) The early nodulin transcipt ENOD2 is located in the nodule parenchyma (inner cortex) of pea and soybean root nodules. EMBO J **9:** 1–19
33. **van Holst G-J, Varner JE** (1984) Reinforced polyproline II conformation in a hydroxyproline-rich cell wall glycoprotein from carrot root. Plant Physiol **74:** 247–251
34. **Woessner JP, Goodenough UW** (1989) Molecular characterization of a zygote wall protein: an extensin-like molecule in *Chlamydomonas reinhardtii*. Plant Cell **1:** 901–911