Secondary Structure and Post-Translational Modifications of the Leucine-Rich Repeat Protein PGIP (Polygalacturonase-Inhibiting Protein) from *Phaseolus vulgaris*†

Benedetta Mattei,‡ Maria Scala Bernalda,‡ Luca Federici,§ Peter Roepstorff,# Felice Cervone,‡ and Alberto Boffi*,§

Dipartimento di Biologia Vegetale, Universitá di Roma "La Sapienza", Piazzale Aldo Moro 5, 00185 Roma, Italy, CNR Centro di Biologia Molecolare and Dipartimento di Scienze Biochimiche, Universitá di Roma "La Sapienza", Piazzale Aldo Moro 5, 00185 Roma, Italy, and Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense University, Campusvej 55, DK-5230 Odense, Denmark

Received July 27, 2000; Revised Manuscript Received October 18, 2000

ABSTRACT: A detailed analysis of the secondary structure has been carried out on the polygalacturonase-inhibiting protein (PGIP) from *Phaseolus vulgaris*, a leucine-rich repeat (LRR) protein present in the cell wall of many plants. Far-UV CD and infrared spectroscopies coupled to constrained secondary structure prediction methods indicated the presence of 12 α- and 12 β-segments, thus allowing a schematic representation of three domains of the protein, namely, the central LRR region and the two cysteine-rich flanking domains. Peptides from endoproteinase-degraded PGIP were analyzed by mass spectrometry, and four disulfide bonds were identified. Mass spectrometric analysis in combination with glycosidase treatments revealed two N-linked oligosaccharides located on Asn 64 and Asn 141. The main structure resembled the typical complex plant N-glycan consisting of a core pentasaccharide β1,2-xylosylated, carrying an α1,3-fucose linked to the innermost *N*-acetylglucosamine and one outer arm *N*-acetylglucosamine residue. The schematic representation of PGIP structural domains is discussed in the framework of the structure and function of LRR proteins.

Polygalacturonase-inhibiting proteins (PGIPs),[1] present in the cell wall of many plants, belong to a super-family of leucine-rich repeat (LRR) proteins specialized for recognition of nonself molecules and rejection of pathogens (*1*). PGIPs are structurally related to several products of resistance genes recently cloned in plants (*2*) and to several receptor-like proteins involved in plant development (*3*).

Some PGIPs are strongly ionically bound to the plant cell wall, while others are readily extracted with dilute buffers. They inhibit the activity of fungal endopolygalacturonases and in vitro favor the accumulation of elicitor-active oligogalacturonides (*4*). In PGIP from *Phaseolus vulgaris*, the domain from amino acid 69 to 326 can be divided into a set of 10.5 tandemly repeating units, each derived from modifications of a 24-amino acid peptide; the repeat element matches the extracytoplasmic consensus LxxLxx**LxxLxLxx**-NxLxGxIPxx (*5*). A comparison of the LRR motifs of PGIP and other proteins indicate that although consensus sequences display high similarity, the number of repeats may vary significantly in different species (*5*−*14*) (Figure 1). The amino acids of PGIP that determine specificity and affinity for fungal PGs occur at positions internal to the conserved **xxLxLxx** motif, which is predicted to form a β-sheet/β-turn structure (*15*).

PGIPs present intriguing similarities with a class of animal extracytoplasmic LRR proteoglycans named SLRPs (small leucine-rich proteoglycans), which include decorin, biglycan, fibromodulin, lumican, and epiphycan (*16*). PGIPs and SLRPs share several properties: they are soluble proteins of the extracellular matrix, and exhibit similar molecular sizes, a similar number of LRRs in the central domain, as well as the presence of cysteine-rich clusters flanking the central domain. SLRPs are involved in matrix assembly, orienting and ordering of collagen fibrils during ontogeny and in pathological processes such as wound healing and tissue repair. Decorin and perhaps other SLRPs regulate at several levels the production and assembly of the extracellular matrix and hence the remodeling of connective tissue (*16*). Posttranslational modifications may account for different functionalities in these molecules: for example, the existence of different SLRPs with one or several glycosaminoglycan chains and diverse patterns of N-glycosylation has been reported, and this diversity is thought to have a functional role (*16*). By analogy with the function of animal SLRPs, involvement of PGIPs in plant matrix assembly and tissue morphogenesis would not be surprising. The wall surrounding plant cells, which can be considered an equiva-

† This research was supported by the European Community Grant QLK3-CT99-089, the Institute Pasteur-Fondazione Cenci Bolognetti and the Giovanni Armenise-Harvard Foundation, the CNR Target Project "Biotechnology" to F.C. and the Danish Biotechnology programme.

* To whom correspondence should be sent. E-mail: Boffi@axrma.uniroma1.it; fax: 39 06 44 40 062.

‡ Dipartimento di Vegetale, Universitá di Roma "La Sapienza".

§ CNR Centro di Biologia Molecolare and Dipartimento di Scienze Biochimiche, Universitá di Roma "La Sapienza".

University of Southern Denmark.

[1] Abbreviations: PGIP, polygalacturonase-inhibiting protein; LRR, leucine-rich repeat; SATR, solution attenuated total reflectance; GlcNAc, *N*-acetylglucosamine; Man, mannose; Xyl, xylose; Fuc, fucose.

10.1021/bi0017632 CCC: $20.00
Published on Web 12/13/2000

Protein	Consensus	LRR repeats	Ref
PGIP	LxxLKxLxxLdLSxNxLxG----xIPxx	10	(5)
RLK5	LxxLxxLxxLxLxxNxLSG----xIPxx	21	(6)
Cf-2.1	LGNLxxLxxLxLxxNxLxG----SIPxx	36	(7)
Cf-9	LxxLxxLxxLDLSSNNLxG----xIPSx	28	(8)
Xa21	LxxLxxLxxLDLSSNNLxG----xIPxx	23	(9)
BRI1	axxxxxLxxLxLSxNxaSG----xIPxx	25	(10)
CLAVATA1	axxaxxLxxLxLxxNxLxG----xIPxx	23	(11)
ERECTA	axxaxxLxxLxLxxNxLxG----xIPxx	20	(12)
sLRP	axxlxxLxxLxLxxNxIxx----Ixxxx	7-12	(13)
pRI	xxxLxxLxLxx$^{C}/_{N}$xLxxxgaxxLxxaLxx	15	(14)

FIGURE 1: Comparison of the LRR motif of PGIP and those of other LRR proteins. x stands for any amino acid; a indicates an aliphatic amino acid residue (L, I, F, V, M); where isoleucine is more represented, I is indicated. S stands for S or T, d stands for an acidic residue (D or E). Cf- and Xa21 are products of resistance genes in tomato and in rice, respectively. RLK5, BR11, CLAVATA1, and ERECTA are products of development-related genes isolated in Arabidopsis sLRP (small leucine-rich proteoglycans) are a class of animal proteins comprising decorin, biglycan, fibromodulin, PRELP, keratocan, osteoadherin, epiphycan, and osteoglycin. pRI is the porcine ribonuclease inhibitor.

lent of the extracellular matrix existing in animal tissues, is a source of signals that plants use to detect pathogens and signals from within the cell determine how the wall loosens, yields, and is remodeled during growth and differentiation. In this connection, it is worth mentioning the recent finding that PGIP in vitro interacts with methylated pectins and that tomato plants overexpressing a PGIP from *Phaseolus vulgaris* display an altered morphology (Bellincampi et al., personal communication).

There is currently no information available on the three-dimensional structure of plant LRR proteins; the difficulty in applying generalizations from the structure of one member of the LRR superfamily to other distantly related members prompted us to gain more detailed information on PGIP. A prediction has been made on the structure of the LRR domain of PGIP using a model of the plant-specific LRR motif derived by Kajava (*17*) as a template for the single 24 amino acid LRR. However, PGIP, as well as decorin and byglican, displays cysteine-rich N-terminal and C-terminal regions, in addition to the central LRR domain. In the present work, we report the first detailed study of the secondary structure and posttranslational modifications of PGIP.

MATERIALS AND METHODS

Protein Purification. PGIP-2 was purified from *Nicotiana benthamiana* plants infected with PVX as previously described (*15*). Leaves were homogenized in 1 M NaCl (2 mL/g), incubated under gentle shaking for 1 h at 4 °C, and centrifuged for 20 min at 8000*g*. The supernatant was filtered, and proteins were precipitated in 75% $(NH_4)_2SO_4$ o/n at 4 °C. The dialyzed proteins were mixed with a suspension of DEAE cellulose (DE52, Whatman, Kent, U.K) preequilibrated with 20 mM sodium acetate, pH 5. The nonabsorbed proteins were chromatographed on a Mono-S column (Pharmacia, Milan, Italy), eluted with a linear gradient of 0−0.5 M NaCl in 20 mM sodium acetate, dialyzed, and purified to homogeneity by an affinity-based procedure on a Sepharose-*A. niger* PG column (*18*).

Circular Dichroism Spectroscopy. CD spectra were measured with a Jasco J-715 spectropolarimeter (Jasco Ltd., Japan). Protein solutions of about 1 mg/mL in 5 mM phosphate buffer, pH 7.0, were placed in a strain-free 1-mm quartz cuvette, and spectra were measured between 260 and 180 nm with the following set up: bandwidth 2 nm, time constant 2 s, scan rate 20 nm/min, N_2 purging rate 25 L/min. Eight spectra corrected to background (phosphate buffer) were averaged and successively smoothed with an 8 points Savisky Golay smoothing procedure. The CD spectrum was normalized to $\Delta\epsilon$ units using the molar extinction coefficient at 280 nm $\epsilon_{280} = 25\,200\ M^{-1}\ cm^{-1}$.

Infrared Spectroscopy. Solution attenuated total reflectance (SATR) FTIR spectra were measured on a Nicolet Magna 760 spectrometer equipped with an MCT detector. A thermostatable ZnSe micro CIRCLE cell (Spectra-Tech, Madison, WI) was used. Protein solutions (150 μL) were placed in the cell at 20 °C and 512 interferograms at 4 cm^{-1} resolution, Mertz apodization, and two levels of zero filling were collected and averaged. The SATR spectra of PGIP as well as all the spectra in the database were collected under identical conditions according to the method of Oberg and Fink (*19*). The following proteins were obtained in the laboratory: azurin (1AZU), cytochrome *c*551 (351C), hemoglobin (4HHB), scapharca hemoglobin (1HbI), nucleoside diphosphate kinase (1NPK), and eglin (1EGL). Alpha-amylase (1BAG), alpha-lactalbumin (1HFZ), beta-lactoglobulin (1BEB), lysozime (1LYZ), catalase (8CAT), insulin (9INS), superoxide dismutase (2SOD), carbonic anhydrase (2CAB), lectin (7WGA), myoglobin (1PMB), pepsin (5PEP), trypsin (1TPO), trypsin inhibitor (6PTI), mellitin (2MLT), papain (1PAP), and trypsinogen (1TGN) were obtained from Sigma Aldrich Co. All proteins (2−5 mg/mL) were dialyzed against 5 mM phosphate buffer at pH 7.0 for 32 h (four changes), and the same dialysis buffer was used to measure the "flush" spectrum (2 mL of buffer were run in the cell before measuring the flush spectrum). The amide I region of the spectrum, from 1800 to 1600 cm^{-1}, was used for calculations. The criteria for inclusion in the database were the following: (i) reproducibility of the spectrum over three distinct data collection, (ii) flatness of the spectrum, after "flush" subtraction, in the region 1780−1720 cm^{-1} (the

difference between the minimum and maximum absorbance value in this region must not exceed 2% of the maximum intensity of the amide I signal), (iii) scaling of the "flush" spectrum to be subtracted must not exceed 0.99 ± 0.1. The spectra were not smoothed nor baseline corrected and were normalized to the area of the amide I region (1800−1600 cm^{-1}). All spectra are available upon request to Boffi@axrma.uniroma1.it.

Secondary Structure Analysis. Analysis of CD data for secondary structure determination was carried out with Matlab-based programs (the Math Works Inc., South Natick, MA) written to perform the same type of analysis on both CD and SATR data. The first program is based on the random variable selection method (*20*, *21*) and the second on the locally linearized method (*22*, *23*). The reader is addressed to the references cited for the details. In both programs, the CD and secondary structure databases of 29 proteins were taken from Johnson's spectra collection, and the secondary structure database was constructed according to Sreerama et al. (*23*). Six eigenvalues were considered in our versions of the locally linearized and random variable selection methods.

Constrained secondary structure predictions were carried out by iterating the GOR4 (*24*) sequence based prediction program within the limits imposed by a set of four decision constants, one for each secondary structure element (i.e., α, β, turn and coil). The decision constants were calculated according to the method of Rahmelow (*25*). As a last step, the elements of secondary structures obtained with the GOR-constrained predictions were compared to the number and average length of the secondary structure elements obtained from CD and IR spectroscopy by applying the criteria established by Sreerama et al. (*23*). In this method, the overall α or β structure content is divided in two substructures, a regular (α_r, β_r) and a distorted (α_d, β_d) one. The number of secondary structure elements is obtained assuming that four residues (two at the beginning and two at the end of the segment) in each α-helix and two residues in each β-strand belong to the distorted configuration.

Generation and Separation of Peptides. Prior to endoproteinase digestion, PGIP was reduced and alkylated: 5 μL of 45 mM dithiothreitol (DTT) was added to 60 μg of PGIP, and the sample was incubated at 50 °C for 15 min. Then 5 μL of 100 mM iodoacetamide were added, and the alkylation was allowed to proceed for 15 min at room temperature. After reduction and alkylation, PGIP was digested either with trypsin (Promega, Madison, WI) or with endoproteinase AspN (Boehringer Mannheim GmbH, Frankfurt, Germany) using an enzyme/substrate ratio of 1:50 (w/w) and 1:200 (w/w), respectively, and incubating for 18 h at 37 °C. Separation of endoproteinase-generated peptides was performed by high-performance liquid chromatography (HPLC) on a Nucleosil C_{18} column (4 by 250 mm; 10-μm particle size; 30-nm pore size). The peptides were eluted by using a linear gradient with 0.1% trifluoroacetic acid (TFA) as buffer A and 90% acetonotrile−0.08% TFA as buffer B. The eluate was monitored at 214 nm. All fractions were collected manually and freeze-dried.

Digestion of Glycopeptides with Exoglycosidases. β-*N*-acetylhexosaminidase (*D. pneumoniae*) and α-mannosidase, (*Jack bean*) were purchased from Boehringer Mannheim GmbH, Frankfurt, Germany. Aliquots containing 50 pmol of the purified glycopeptides were lyophilized and redissolved in 5 μL of 50 mM ammonium acetate, pH 5.0. 1 milliunit of β-*N*-acetylhexosaminidase and 50 milliunits of α-mannosidase were added sequentially. Incubation with each exoglycosydase was carried out for 20 h at 37 °C. Aliquots were removed for mass spectrometry analysis before addition of the new enzyme.

Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. Solvents and chemicals: HPLC-grade trifluoroacetic acid (TFA), acetonitrile (CH3CN) (all purchased from Rathburn Ltd., Scotland), 1,4-dithiothreitol (DTT) (Sigma, St. Louis, MO). Matrixes used were 2,5-dihydroxybenzoic acid (DHB) (Hewlett-Packard, Palo Alto, CA) 100 μL of 100 mM solution lyophilized and redissolved in an equal volume of 30% aqueous acetonitrile, 0.1% TFA; 4-hydroxy-α-cyanocinnaminic acid (HCCA) (Sigma, St. Louis, MO) 10 μg/μL dissolved in 70% aqueous acetonitrile, 0.1% TFA; and sinapinic acid (SA) (3,5-dimethoxy-4-hydroxycinnamic acid) (Sigma, St. Louis, MO) 20 μg/μL dissolved in 70% aqueous acetonitrile, 0.1% TFA.

MALDI mass spectra were acquired on a Voyager-Elite MALDI-TOF (PerSeptive Biosystems Inc., MA) mass spectrometer, operated either in positive ion linear mode or in positive ion reflector mode. HPLC-purified samples were prepared for MALDI analysis by mixing 0.8 μL of an HPLC fraction with 0.5 μL of matrix solution directly on the target. The instrument was equipped with delayed ion extraction, the delay time used was 200 ns in the linear mode and 260 ns in the reflector mode. Mass spectra were recorded as single-shot spectra by using a UV laser operated at 337 nm and an acceleration voltage of 20 kV. Typically, 100 to 256 single-shot spectra were averaged to give the final spectrum. The spectra were externally calibrated by using HP peptide standard (oxytocin, arginine-8-vasopressin, angiotensin I, somatostatin, chicken atrial natriuretic peptide (ANP), human r insulin, r hirudin) (Hewlett-Packard, Palo Alto, CA).

Edman Degradation (Amino Acid Sequencing). Edman degradation of peptides (100 pmol) was performed on an HP-G1000A protein sequencer, connected to an HP-1090LC HPLC system (Hewlett-Packard) for identification of phenylthiohydantoin (PTH) derivatives. Sequencing was carried out by the standard protocol provided by the manufacturer.

RESULTS

Secondary Structure of PGIP. The overall line shape of the CD spectrum of PGIP, reported in Figure 2, does not show any significant resemblance with the spectra reported in the database. The low $\Delta\epsilon$ and the position of both the minimum (216 nm) and the maximum (188 nm), however, are within the range of all-β proteins (*26*).

The second derivative and Fourier self-deconvolved SATR spectra in the amide I region display a rich pattern that allows the identification of several secondary structure-related peaks (Figure 3). The most prominent line in the spectrum is at 1636 cm^{-1}, in a frequency region dominated by the contribution of β-strand structures (*27*, *28*). Within the same spectral region, the shoulder at 1630 cm^{-1} is likewise assigned to a β-strand component. The β-strand assignment is completed by the identification of the high-frequency component of the β type amide I normal mode at 1682 cm^{-1} (*28*). The band at 1652 cm^{-1} is readily identified as an α helical structure

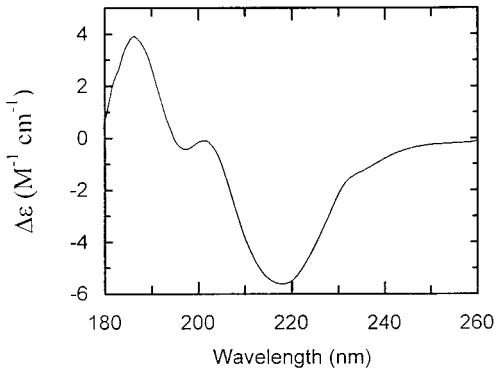

FIGURE 2: Far UV CD spectrum of PGIP. The spectrum was measured at 20 °C on a 1 mg/mL protein solution in 5 mM phosphate buffer at pH 7.0. The spectrum has been normalized in $\Delta\epsilon$ units by standard procedures.

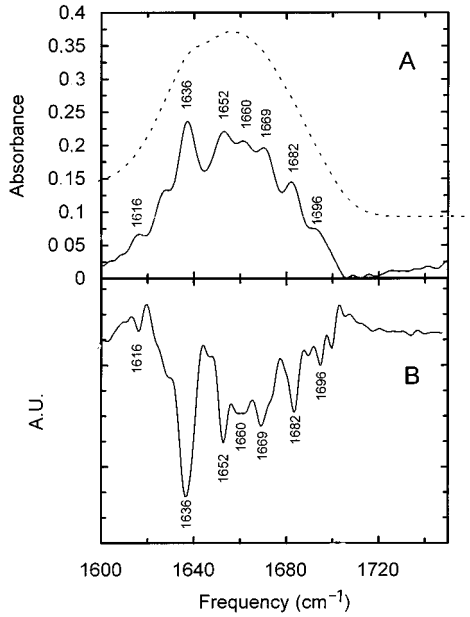

FIGURE 3: Fourier self-deconvolved (A) and second derivative (B) SATR spectra of PGIP. The spectrum was measured at 20 °C on a 2 mg/mL protein solution in 5 mM phosphate buffer at pH 7.0. A ZnSe circle cell was used. The original spectrum is shown in panel A as a dashed line. Deconvolution parameters for the amide I band were set with a half-bandwidth of 18 cm^{-1} and *K* value at 2.6 (panel A, continuous line). The second derivative was calculated after an 11 points Savitzky Golay smoothing procedure.

contribution while the smaller features at 1660 and 1696 cm^{-1} are assigned as turns and/or unordered structures.

Both CD and SATR amide I spectra were analyzed for secondary structure determination with the same algorithms based on the random variable selection method (*20*, *21*) and on the locally linearized method (*22*, *23*). The analysis shows that there is a reasonably good agreement in the secondary structure determinations obtained with different techniques and by applying different algorithms (Figure 4) considering that (i) the CD spectrum of PGIP does not bear any clear resemblance with reported CD spectra and (ii) the CD and SATR reference data sets have only eight proteins in common.

A consistent quantification of β-strand ($\beta_r + \beta_d = 34.6 \pm 1.4\%$), α-helix ($\alpha_r + \alpha_d = 16.6 \pm 2.8\%$), and turns ($21.5 \pm 1.5\%$) is obtained from the average of the CD and IR determinations that allows a prediction of the number and

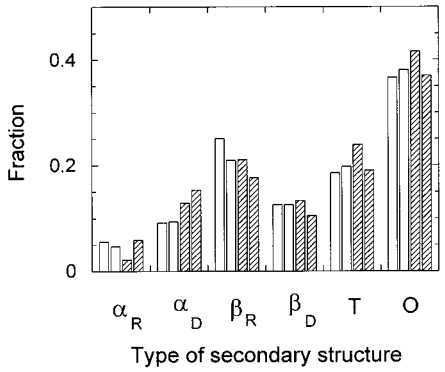

FIGURE 4: Bar graph of secondary structures obtained from CD and SATR multivariate data analysis. The secondary structure types are α_R, alpha regular; α_D, alpha distorted; β_R, beta regular; β_D, beta distorted; T, turns; and O, other structures. White bars are relative to the analysis of CD data performed according to the random variable selection method (right) and to the locally linearized model (left). Striped bars refer to the same analysis applied to the infrared amide I band.

of the average length of α-helical and β-strand segment (see Figure 4). On the basis of the criteria established by Sreerama et al. (*23*), the major α-helical component belongs to a distorted α-helical type, indicating that 52 ± 4 residues are arranged in a number of short helical segments (10 segments with an average length of 5 ± 1 residues). β-structure elements are also arranged in a series of 12 segments of 8 ± 1 residues average length. These findings were used in connection with the GOR4-constrained secondary structure prediction method to assign specific amino acid sequences to the secondary structure segments (*24*). Convergence was obtained after a few iterations, using the overall amounts of α, β, turn, and coil determined by combined CD and IR analysis. The number of helical segments obtained by the constrained prediction method is 12 instead of the 10 predicted by spectroscopic methods alone, and all helical segments are four residues long with the exception of the first, N-terminal segment, which is 6−8 residues long (see Figure 7). In turn, the 12 β-segments determined spectroscopically have been confirmed by the constrained prediction. Last, it should be mentioned that unconstrained secondary structure prediction methods slightly underestimated the length of the β-segments in the LRR domain.

Disulfide Bonds in PGIP. The primary sequence of PGIP deduced from the cDNA sequence was verified by mass spectrometric analysis of reduced, alkylated, and endoproteinase-digested PGIP. The tryptic peptides covered 93.6% of the PGIP sequence, including the N- and C-terminal parts of the protein with eight cysteine residues. Only the smaller peptides were missing, and combined with the analysis of peptides obtained by digestion with endoproteinase AspN the sequence coverage was 100%.

On the basis of sequential MALDI-MS analysis of nonreduced, reduced, and reduced and alkylated peptides from a tryptic digest of the native protein, four intramolecular disulfide bonds were identified. Endoproteinase digestion of nonreduced PGIP was used in combination with HPLC and MALDI-TOF-MS to determine the presence and position of the disulfide bridges (Figure 5). PGIP was digested with trypsin in nonreducing conditions, and peptides were separated by HPLC and analyzed by MALDI-TOF before and after chemical reduction with DTT. Two disulfide bridges

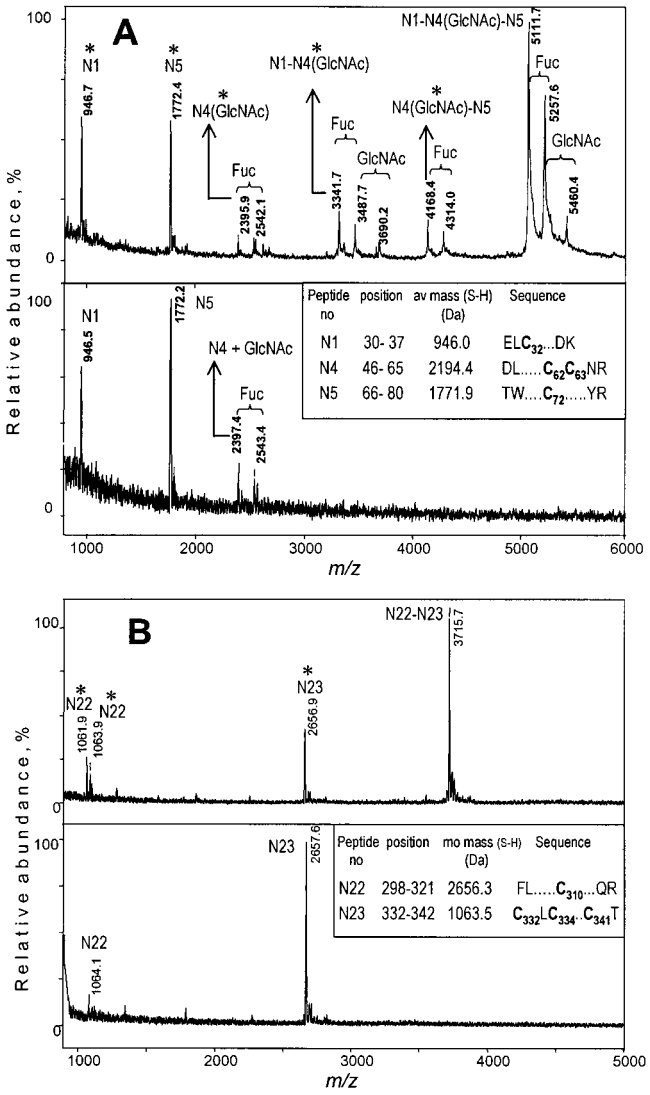

FIGURE 5: (A) MALDI spectra of fraction 39 obtained by RP HPLC analysis of the tryptic digest of PGIP-2 before reduction (top) and after reduction with DTT (bottom). Spectra were obtained in positive ion linear mode using DHB matrix. Asterisks mark peaks corresponding to reduced peptides arising from fragmentation of the disulfide-linked peptides. The table shows the calculated average molecular masses of peptides N1, N4, and N5 derived from tryptic digestion. (B) MALDI spectra of fraction 41 obtained by RP HPLC analysis of the tryptic digest of PGIP-2 before reduction (top) and after reduction with DTT (bottom). Spectra were obtained in positive ion reflector mode using DHB matrix. The table shows the calculated monoisotopic molecular masses of peptides N22 and N23 derived from tryptic digestion.

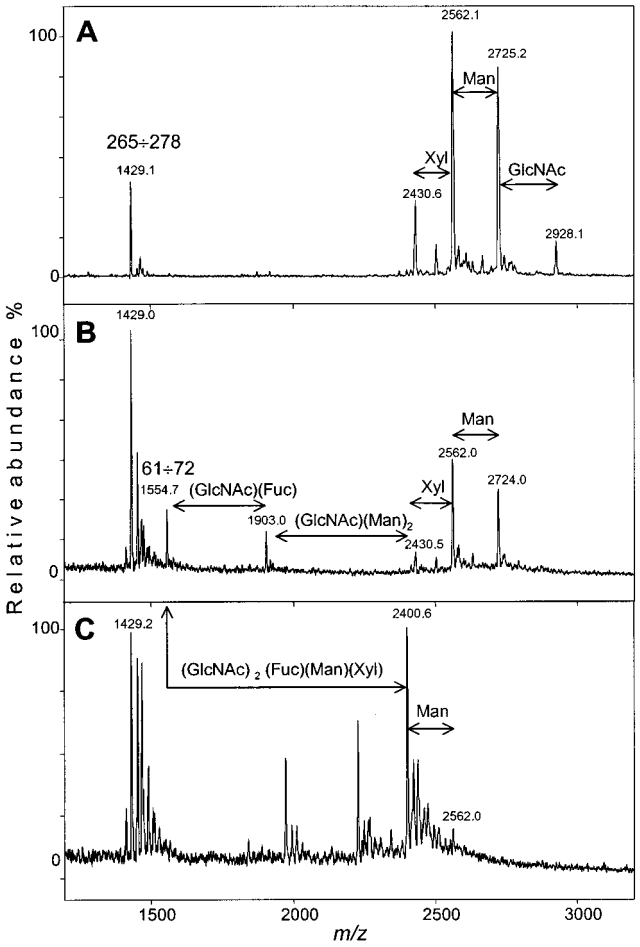

FIGURE 6: (A) MALDI spectrum of fraction 20 obtained by RP HPLC separation of the AspN digest of PGIP. (B) MALDI spectrum of the same fraction after treatment with *β-N*-acetylhexosaminidase. The peak at *m/z* 1554.7 corresponds to the nonglycosylated form of the peptide. The observed mass shift of the main component corresponds to one GlcNAc residue. (C) MALDI spectrum of the same fraction after sequential digestion with first *β-N*-acetylhexosaminidase followed by α-mannosidase. The observed mass shift of the main component corresponds to two Man residues. Spectra were obtained in positive ion linear mode using DHB matrix. Two possible structures are proposed for the N-linked carbohydrate: GlcNAc*β*1,2Manα1,6(Manα1,3)(Xyl*β*1,2)Man*β*1,4GlcNAc*β*1,4-(Fucα1,3)GlcNAc Manα1,6(GlcNAc*β*1,2Manα1,3)(Xyl*β*1,2)Man*β*1,-4GlcNAc*β*1,4(Fucα1,3)GlcNAc.

were identified at the N-terminal end between the peptide N4 containing Cys 62 and Cys 63 and the two peptides N1 and N5 containing Cys 32 and Cys 72, respectively (Figure 5, panel A). The major peak in the spectrum of the nonreduced fraction (*m/z* 5111.7) corresponds to N1−N4−N5 peptide with one GlcNAc residue attached at Asn 64. Ions corresponding to the individual peptides N1, N4, N5, as well as combinations of N1−N4 and N4−N5 were also observed in the spectra (asterisked in the figure). Since this sample has not been reduced, these peptides must arise due to a reduction in the mass spectrometric procedure or due to prompt fragmentation of the S−S bond, as previously reported by Patterson and Katta (*29*). After on-target reduction with DTT, the mass spectrum shows three peaks at *m/z* 946.5, 1772.2, 2397.4, corresponding to the individual reduced peptides N1, N4, and N5. We were not able to discriminate which of the peptides N1 and N5 was attached to which of the two adjacent cysteine residues 62 and 63 in N4. At the C-terminal end, one disulfide bridge was identified between Cys 310 and one of the Cys residues of the C-terminal peptide N23 (Figure 5, panel B). The major peak at *m/z* 3715.7 corresponds to peptide N22−N23. The mass spectrum shows also in this case the ions corresponding to fragmentation or reduction of the disulfide bonds at *m/z* 2656.9 and 1063.9. The observed ion at *m/z* 1061.9 suggests the presence of a second disulfide bridge internal to peptide N23. The on-target reduction with DTT causes a shift of ca. 2.0 Da in the mass of this peptide (observed *m/z* 1064.1), due to the reduction of an internal disulfide bridge (Figure 5, panel B, bottom). To determine which cysteine residue is involved in interpeptide disulfide bridge with Cys 310, the nonreduced peptides were submitted to Edman degradation.

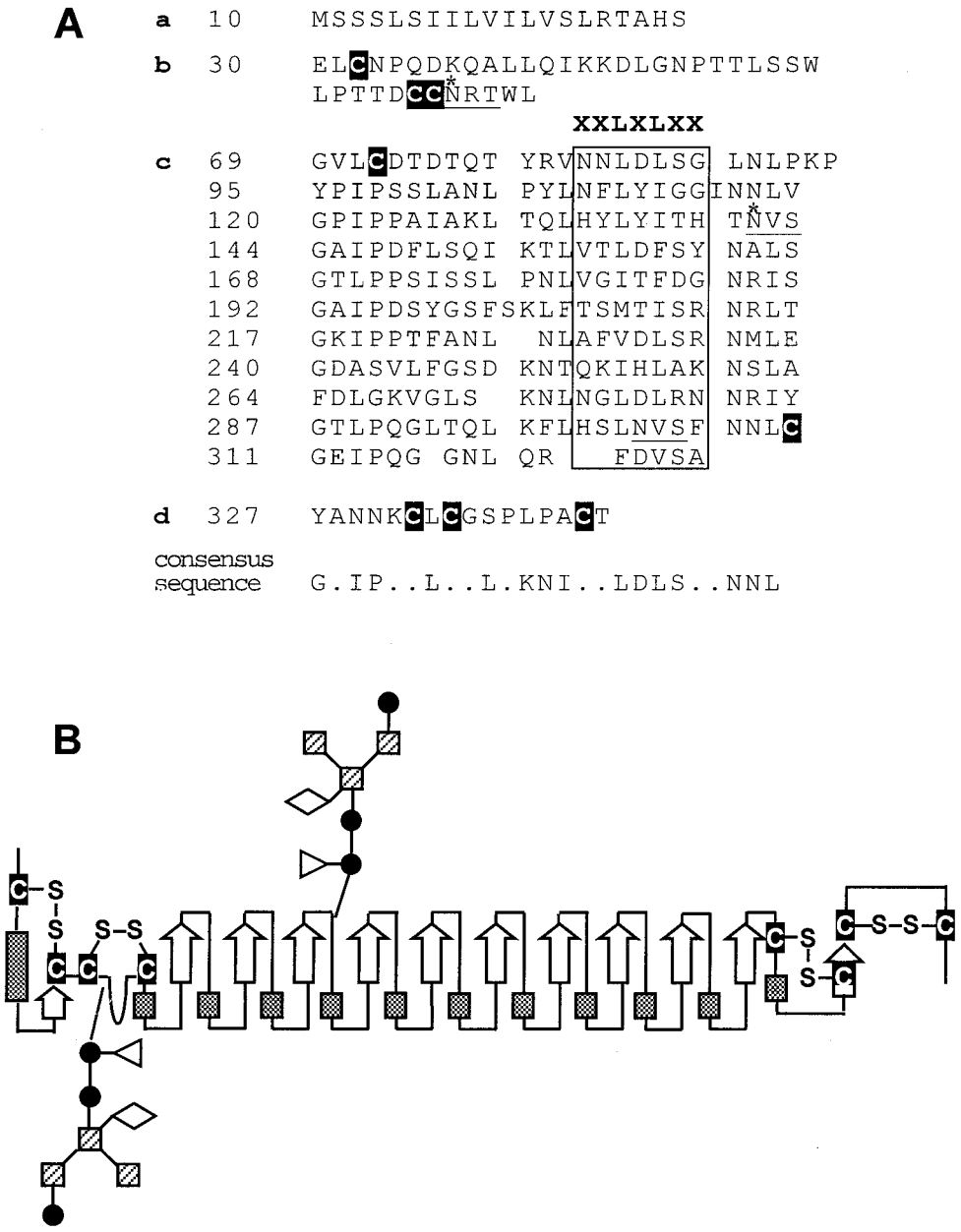

FIGURE 7: (A) PGIP-2 LRR structure. a = signal peptide, b = NH_2-terminus of the mature protein, c = 10.5 LRRs and d = COOH terminus. Cys residues are highlighted. Putative glycosylation sites are underlined, while the attachment sites of the two N-glycans identified are indicated by an asterisk. The box indicates the area of the protein predicted to form the β-sheet/β-turn structural motifs. Amino acids are numbered according to ref *6*. (B) Schematic drawing of PGIP secondary structure elements. Arrows indicate β-strands and boxes indicate α-helices. Glycan structure has been sketched as follows: (●) GlcNAc, (□) Man, (▽) Fuc, (◇) Xyl.

After one sequencing cycle, the peptides were manually eluted from the sequencer column and subjected to mass spectral analysis: an observed peak at *m*/*z* 2750.8 corresponds to peptide 299−321 linked to PTH−Cys 332, and a peak at *m*/*z* 960.7 corresponds to the C-terminal peptide 333−342 (data not shown). After on-target reduction with DTT, the observed peaks were at *m*/*z* 2513.6 corresponding to the reduced peptide 299−321, and at *m*/*z* 962.6 corresponding to the reduced C-terminal peptide 333−342. We therefore conclude that Cys 310 and Cys 332 share a disulfide bond and that another disulfide bond is formed between Cys 334 and Cys 341.

Glycosylation of PGIP. PGIP was reduced, alkylated, and digested either with endoproteinase AspN or with trypsin, and the proteolytic fragments were separated by RP HPLC and analyzed by MALDI-TOF-MS (Figure 6). Glycopeptides were recognized in the mass spectra by their carbohydrate microheterogeneity that gave rise to multiple peaks with mass differences corresponding to the carbohydrate residues (*30*). In plants, N-linked oligosaccharides have typical structures that differ from those found in mammalian glycoproteins: a specific characteristic of plant N-glycans is the absence of sialic acid and the presence of β(1,2)-linked xylose residues and α(1,3)-fucosylation of the innermost *N*-acetyl glucosamine (*31*). N-glycans were identified in PGIP attached to Asn 64 and Asn 141, positioned on the predicted Asn−Xxx−Thr/Ser consensus site.

One peptide obtained by endoproteinase AspN digestion of PGIP produced the mass spectrum shown in Figure 6, panel A. Several peaks were identified around *m*/*z* 2000−3000, separated by approximately 203 (*N*-acetylhexosamine), 162 (hexose), and 132 (pentose) and therefore likely corre-

sponding to different glycoforms. The oligosaccharide was sensitive to digestion by β-*N*-acetylhexosaminidase, resulting in a mass decrease of 203 Da corresponding to one GlcNac residue. The mass difference from the main component to the nonglycosylated form of the peptide (*m*/*z* value of 1554.74 corresponding to peptide Asp61−Cys72 containing the Asn_{64}−Xxx−Thr consensus site for glycosylation) corresponds to the mass of the core structure ($GlcNac_2Man_3$) containing a fucose and a xylose. Sequential digestion with α-mannosidase decreased the mass of the main component by 324 Da, corresponding to removal of two Man residues. The glycan structure proposed on the basis of the observed mass differences is consistent with one of the typical plant complex-type N-glycans (*32*), consisting of a core-pentasaccharide containing α(1,3)-fucose and β(1,2)-xylose residues, respectively, linked to the innermost *N*-acetylglucosamine and to the β-mannose residue of the core, and one outer-arm β(1,2)-*N*-acetylglucosamine. The latter residue can be linked either to the Manα1,6 or to the Manα1,3 of the trimannosyl core, leading to two potential structures:

(i) GlcNAcβ1,2Manα1,6(Manα1,3)(Xylβ1,2)Manβ1,4GlcNAcβ1,4(Fucα1,3)GlcNAc or (ii) Manα1,6(GlcNAcβ1,2Manα1,3)(Xylβ1,2)Manβ1,4GlcNAcβ1,4(Fucα1,3)GlcNAc.

The same structure could be deduced for the glycan attached to Asn 141 (data not shown). A third consensus site for glycosylation is present at Asn 303 in the tenth LRR of PGIP but mass spectrometric characterization of PGIP digested with different endoproteinases could detect only the nonglycosylated peptide.

DISCUSSION

In all described LRR proteins, consensus sequences for the repeating units are surprisingly similar considering the range of organisms and the widely divergent functions of the different proteins. The great similarity among these proteins indicates a strong selection pressure for conservation of this structure. However, modeling of the LRR superfamily performed by Kajava predicts that the different subfamilies can exhibit differences in their tertiary structure (*17*). It has also been shown that even members of the same subfamily such as decorin and biglycan may differ significantly in secondary structure and that the glycosaminoglycan chains influence the conformational stability of these proteins (*33*). It has been proposed that the LRR domain of PGIP may have an extensive arch-shaped ligand-binding surface facilitated by a protein fold resembling that of the α/β coiled fold, similar to that observed in the crystal structure of pRI (*34*). The body of experimental data obtained from CD and FTIR spectroscopy and the subsequent constrained secondary structure prediction indicate the presence of 11 small, distorted α-helical segments and 10 typical LRR β-motifs. Thus, a total of 10.5 LRR elements have been proposed. The short α-helical portion of each repeat in PGIP, with respect to pRI, suggests that the curvature of the arch-shaped surface may be less pronounced than in pRI. Moreover, a first guess for the secondary structure of the cysteine-rich C- and N-terminal flanking domains has been obtained that assign a 6−8 residue α-helical region followed by a short β_segment to the N-terminal region and a further β segment to the C-terminal part.

The final picture (see Figure 7) indicates that the total number of helical segments, obtained from the constrained secondary structure prediction method, is 12 instead of the 10 predicted by spectroscopic methods alone and all helical segments are four residues long with the exception of the first, N-terminal segment, which is 6−8 residues long. The secondary structure assignments, combined with the identification of four disulfide bridges and two glycosylation sites allow for a schematic representation of the whole PGIP structure thus expanding the previous structural assignment based upon homology modeling of the LRR domain with the RI tridimensional structure (*15*).

The four cysteine residues at the N-terminal end form two disulfide bridges, and at the C-terminal end four similarly spaced cysteine residues in a 27-amino acid stretch are also bridged by disulfide bonds, indicating that these linkages are important to maintain structures of local importance and to provide additional stabilization of the secondary and tertiary structure. The alignment of PGIP sequences (citrus, kumquat, *P. trifoliata*, pear, apple, apricot, kiwi, tomato, bean, and soybean) shows that the positions of cysteine residues are highly conserved in all the PGIPs. This could reflect the presence of conserved structural features of general importance for the biological function of these LRR proteins.

The position and number of consensus sites for N-linked glycosylation are not highly conserved among the PGIPs from different plants, indicating that the glycans probably play more individual roles in the structure of the PGIPs than the disulfide bonds. It has also been suggested that glycosylation of asparagine residues within the conserved β-strand/β-turn motif might be expected to interfere with ligand binding (*35*). Consistent with this suggestion, just one consensus site for glycosylation is present within this motif at Asn 303 in the tenth LRR of PGIP and mass spectrometric characterization of endoproteinase digested peptides of PGIP could detect only the nonglycosylated peptide.

REFERENCES

1. Cervone, F., Castoria, R., Leckie, F., De Lorenzo, G. (1997) in *Signal Transduction in Plants* (Aducci, P., Ed.) pp 153−177, Birkäuser Verlag, Basel/Switzerland.
2. Jones, D. A., Jones, J. D. G. (1997) *Adv. Bot. Res. 24*, 89−166.
3. Becraft, P. W. (1998) *Trends in Plant Science 3*, 384−388.
4. De Lorenzo, G., Cervone, F. (1997) in *Plant-Microbe Interactions* (Stacey, G., and Keen, N. T., Ed.) Vol. 3, pp 76−93, Chapman and Hall, New York.
5. De Lorenzo, G., Cervone, F., Bellincampi, D., Caprari, C., Clark, A. J., Desiderio, A., Devoto, A., Forrest, R., Leckie, F., Nuss, L. (1994) *Biochem. Soc. Trans. 22*, 396−399.
6. Walker, J. C. (1993) *Plant J. 3*, 451−456.
7. Dixon, M. S., Jones, D. A., Keddie, J. S., Thomas, C. M., Harrison, K. and Jones, J. D. (1996) *Cell 84*, 451−459.
8. Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J., and Jones, J. D. G. (1994) *Science 266*, 789−459.
9. Song, W., Wang, G., Chen, L., Kim, H., Pi, L., Holsten, T., Gardner, J., Wang, B., Zhai, W., Zhu, L., Fauquet, C., and Ronald, P. (1995) *Science 270*, 1084−1086.
10. Li, J., and Chory, J. (1997) *Cell 90*, 919−938.
11. Clark, S. E., Williams, R. W., and Meyerowitz, E. M. (1997) *Cell 89*, 575−585.
12. Torii, K. U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R. F., and Komeda, Y. (1996) *Plant Cell 8*, 735−746.

13. Krishnan, P., Hocking, A. M., Sholtz, J. M., Pace, C. N., Holik, K. K., and McQuillan, D. J. (1999) *J. Biol. Chem. 274*, 10945−10950.
14. Kobe, B., and Deisenhofer, J. (1993) *Nature 366*, 751−756.
15. Leckie, F., Mattei, B., Capodicasa, C., Hemmings, A., Nuss, L., Aracri, B., De Lorenzo, G., and Cervone, F. (1999) *EMBO J. 18*, 2352−2363.
16. Iozzo, R. V. (1997) *Crit. Rev. Biochem. Mol. Biol. 32*, 141−174.
17. Kajava, A. V. (1998) *J. Mol. Biol. 277*, 519−527.
18. Cervone, F., Hahn, M. G., De Lorenzo, G., Darvill, A., and Albersheim, P. (1989) *Plant Physiol. 90*, 542−548.
19. Oberg, K. A., and Fink, A. L. (1998) *Anal. Biochem. 256*, 92−106.
20. Manavalan, P., and Johnson, W. (1987) *Anal. Biochem. 167*, 76−85.
21. Johnson, W. (1999) *Protein-Struct. Funct. Genet. 35*, 307−312.
22. Van Stokkum, I. H. M., Spoelder, H. J. W., Bloemendal, M., van Grondelle, R., and Groen, F. C. A. (1990) *Anal. Biochem. 191*, 110−118.
23. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (1999) *Protein Sci. 8*, 370−380.
24. Garnier, J., Gibrat, J. F., and Robson, B. (1996) *Methods Enzymol. 266*, 540−553.
25. Schubach, B. J., and Rahmelow, K. (1997) *Biochim. Biophys. Acta-Protein Struct. Mol. Enzymol. 1340*, 72−80.
26. Venyaminov, S. Y., and Vassilenko, K. S. (1994) *Anal. Biochem. 222*, 176−184.
27. Susi, H., and Byler, D. M. (1983) *Biochem. Biophys. Res. Commun. 115*, 391−397.
28. Olinger, J. M., Hill, D. M., Jakobsen, R. J., and Brody, R. S. (1986) *Biochim. Biophys. Acta 869*, 89−98.
29. Patterson, S. D., and Katta, V. (1994) *Anal. Chem. 66*, 3727−3732.
30. Mirgorodskaya, K., Krogh, T. N., and Roepstorff, P. (2000) in *Methods in Molecular Biology. Proteins and Peptide Analysis - New Mass Spectrometric Applications* (Chapman, J. R., Ed.) Humana Press, Totowa, New Jersey.
31. Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Laine, A. C., Gomord, V., and Faye, L. (1998) *Plant Mol. Biol. 38*, 31−48.
32. Zeleny R., Altmann F., and Praznik, W. (1999) *Phytochem. 51*, 199−210.
33. Krishnan, P., Hocking, A. M., Scholtz, J. M., Pace, C. N., Holik, K. K., and McQuillan, D. J. (1999) *J. Biol. Chem. 274*, 10945−10950.
34. Thomas, C. M., Jones, D. A., Parniske, M., Harrison, K., Balint-Kurti, P. J., Hatzixanthis, K., and Jones, J. D. G. (1997) *Plant Cell 9*, 2209−2224.

BI0017632