Phytochemistry 69 (2008) 1973–1982

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Primary structure and glycan moiety characterization of PD-Ss, type 1 ribosome-inactivating proteins from *Phytolacca dioica* L. seeds, by precursor ion discovery on a Q-TOF mass spectrometer

Angela Chambery [1], Antimo Di Maro [1], Augusto Parente *

Dipartimento di Scienze della Vita, Seconda Università di Napoli, Via Vivaldi 43, I-81100 Caserta, Italy

ARTICLE INFO

Article history:
Received 8 January 2008
Received in revised form 1 April 2008
Available online 29 May 2008

Keywords:
N-glycosylation
Precursor ion discovery
Mass spectrometry
Ribosome-inactivating protein
Phytolacca dioica

ABSTRACT

Seeds from *Phytolacca dioica* L. contain at least three *N*-glycosylated PD-Ss, type 1 ribosome-inactivating proteins (RIPs), which were separated and purified to homogeneity by conventional chromatographic techniques. ESI–Q-TOF mass spectrometry provided the accurate M_r of native PD-S1 and PD-S3 (30957.1 and 29785.1, respectively) and the major form PD-S2 (30753.8). As the amino acid sequence of PD-S2 was already known, its disulfide pairing was determined and found to be Cys34–Cys262 and Cys88–Cys110. Further structural characterization of PD-S1 and PD-S3 (N-terminal sequence determination up to residue 30, amino acid analysis and tryptic peptide mapping) showed that the three PD-Ss shared the entire protein sequence. To explain the different chromatographic behaviour, their glycosylation patterns were characterized by a fast and sensitive mass spectrometry-based approach, applying a precursor ion discovery mode on a Q-TOF mass spectrometer. A standard plant paucidomannosidic *N*-glycosylation pattern [Hex_3, $HexNAc_2$, $deoxyhexose_1$, $pentose_1$] was found for PD-S1 and PD-S2 on Asn120. Furthermore, a glycosylation site carrying only a HexNAc residue was identified on Asn112 in PD-S1 and PD-S3. Finally, considering the two disulfide bridges and the glycan moieties, the experimental M_r values were in agreement with the mass values calculated from the primary structure. The complete characterization of PD-Ss shows the high potential of mass spectrometry to rapidly characterize proteins, widespread in eukaryotes, differing only in their glycosylation motifs.

1. Introduction

Ribosome-inactivating proteins (RIPs; rRNA *N*-β-glycosidases; EC 3.2.2.22) are single-chain (type 1) or heterodimeric (types 2 and 3) cytotoxic enzymes endowed with rRNA *N*-glycosidase, and more extensively, with adenine polynucleotide glycosylase (APG) activity on substrates such as RNA, poly(A) and DNA (Stirpe and Battelli, 2006). The most studied and characterized effects of polynucleotide depurination are the resulting irreversibly inactivation of ribosomes and the block of protein synthesis. Nevertheless, some RIPs posses even a DNase activity, being able to cleave supercoiled plasmids [pBR322 dsDNA], thus generating relaxed and linear molecules (Aceto et al., 2005). RIPs are present in a number of higher plants and have been found also in fungi, algae and bacteria (Girbes et al., 2004). They have been studied for long time for their biological activities (*i.e.* broad-spectrum antiviral, antifungal, and antibacterial activities) and bio-medical applications (*i.e.* immunotoxins for anti-cancer and immunosuppressive therapy) (Bolognesi and Polito, 2004; Parikh and Tumer, 2004; Stirpe and Battelli, 2006).

Phytolacca dioica, woody plant introduced in Italy from South America and belonging to the Phytolaccaceae family, contains several RIPs. Some of them have been already isolated and characterized by our research group: four RIP isoforms, named PD-Ls1-4, from summer leaves of adult *P. dioica* (Di Maro et al., 1999) and three isoforms from the seeds, apparently with various degree of glycosylation (Parente et al., 1993). Protein glycosylation has long been recognized as a common post-translational modification, especially for proteins destined to vacuolar and/or extra-cellular environment (Lerouge et al., 1998; Rayon et al., 1998). Carbohydrates are linked to seryl or threonyl residues (*O*-glycosylation) or to asparaginyl residues (*N*-glycosylation). In the latter case the consensus sequences fall into the N-X-S/T/C motifs, X denoting any amino acid except proline (Rayon et al., 1998). The protein *N*-glycosylation in plants has a great impact both on their physico-chemical properties and, in several cases, on their biological function (Ceriotti et al., 1998). Unlike the B chain of type 2 RIPs, type 1 RIP glycosylation and its possible functional implications

* Corresponding author. Tel.: +39 0823 274583/535; fax: +39 0823 274571.
E-mail address: augusto.parente@unina2.it (A. Parente).
[1] These authors contributed equally to this work.

0031-9422/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2008.04.005

have been poorly investigated. Daubenfeld and colleagues report the glycosylation motifs of gelonin, isolated from the seeds of *Gelonium multiflorum* (Daubenfeld et al., 2005). On the basis of its glycan structure, they suggest that gelonin is located in the vacuole of seed cells. We have previously reported the primary structure determination of PD-S2, major RIP form isolated from *P. dioica* seeds and the purification to homogeneity of two minor forms (PD-S1 and PD-S3; Del Vecchio Blanco et al., 1997).

In the present work we report: (i) the setting-up of an optimized purification procedure from *P. dioica* seeds of the three PD-Ss; (ii) the determination of their accurate molecular masses by ESI/Q-TOF mass spectrometry; (iii) the tryptic peptide mapping for amino acid sequence determination of PD-S1 and PD-S3; (iv) the disulfide pairing of PD-S2; and (v) the elucidation of the oligosaccharide structures of all three PD-Ss by applying a quick and selective mass spectrometric approach based on precursor ion discovery (PID) mode on a Q-TOF mass spectrometer.

2. Results

2.1. PD-Ss purification and mass spectrometry analysis

For PD-Ss isolation, total proteins were extracted from *P. dioica* seeds in phosphate-buffered saline and acid precipitated with acetic acid (pH 4.0). Soluble proteins were fractionated by cation exchange chromatography on *S*-Sepharose, gel-filtration on Sephacryl S-100 HR and a further cation exchange chromatography on CM-Sepharose, according to a previous published procedure (Parente et al., 1993). However, minor contaminants that could affect the following mass spectrometry analyses were revealed by SDS–PAGE (data not shown). Therefore, to obtain a more homogeneous preparation, a further purification step by FPLC was added, as described in Section 4. Single peaks at retention times of 33.5, 38.2 and 43.9 min were eluted for PD-S1, PD-S2 and PD-S3, respectively (Fig. 1).

The relative molecular masses of native PD-Ss were then determined by ESI/Q-TOF mass spectrometry, after protein desalting by RP-HPLC. The ESI-MS spectra, which revealed the presence of differently multiple charge species of the same protein, were deconvoluted using maximum entropy techniques (Ferrige et al., 1992), which allowed us to determine the accurate M_r of PD-S1 (30957.1), PD-S2 (30753.8) and PD-S3 (29785.1).

The total sugar composition of the glycan chains of PD-S2 had been previously determined, revealing a 2.22% neutral sugar content (Parente et al., 1993). It was then hypothesized that, also for PD-S1 and PD-S3, the heterogeneity of the intact protein, revealed by mass spectrometry analyses, could derive, at least in part, by variations in one or more of their sugar moieties.

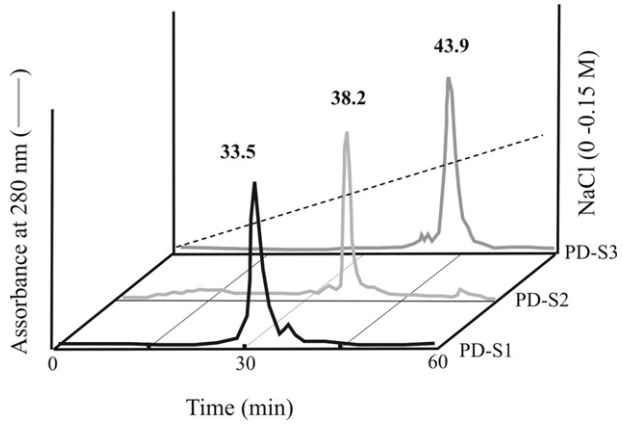

Fig. 1. FPLC elution profiles of PD-S1, PD-S2 and PD-S3 on an AKTA Purifier System from cation exchange chromatography using a Source 15S PE 4.6/100 column. Experimental conditions are described in the text.

Therefore, complete characterization of PD-S1 and PD-S3 was undertaken. For N-terminal sequence determination the two proteins were chromatographed on RP-HPLC and the resulting peak fractions were subjected to automated Edman degradation. This analysis allowed us to obtain the sequence of the first 30 N-terminal amino acid residues (1-VSTITFDVGS ATISKYTTFL ESLRNQAKDP-30), which were identical between the two proteins and, in turn, identical to those of PD-S2 (Del Vecchio Blanco et al., 1997; Parente et al., 1993). Furthermore, amino acid compositions similar to that of PD-S2 were also obtained for PD-S1 and PD-S3 (Table 1).

Hitherto, the three *P. dioca* seed RIPs had: (i) identical N-terminal sequences up to residue 30; (ii) similar amino acid compositions; (iii) different elution times from ion exchange chromatographies; and (iv) different molecular masses as determined by Q-TOF mass spectrometry. Furthermore, it was already known that the three PD-Ss inhibited protein synthesis in a rabbit reticulocyte lysate, with different IC_{50} values of 0.12, 0.06 and 0.08 nM for PD-S1, PD-S2 and PD-S3, respectively (Parente et al., 1993).

2.2. Peptide mapping of PD-S1 and PD-S3

In order to elucidate the structural differences of PD-Ss and considering their identical amino acid compositions, we decided to perform a comparative tryptic peptide mapping of PD-S1 and PD-S3, using the PD-S2 primary structure (Del Vecchio Blanco et al., 1997), as reference peptide profile.

The PD-S1 and PD-S3 tryptic maps, resulting from RP-HPLC and obtained as described in Section 4, are reported in Fig. 2C and B, respectively, and compared to that of PD-S2 (Fig. 2A). A comparison of the three chromatograms revealed a high reproducibility. Indeed, with few exceptions, almost all peptides of PD-S1 and PD-S3 eluted with the same retention times with respect to those of PD-S2. In particular, for peak TP120, a shift to higher retention times (46.3 min) was registered in the PD-L3 map, whereas the same elution times (42.7 min) were found for TP120 peaks in PD-S2 and PD-S1 maps (see Fig. 2). Furthermore, for peak TP103, a slight shift to lower retention times (35.5 min) was detected in PD-S1 and PD-S3 maps with respect to that (36.4 min) of the corresponding peak in PD-S2 map (see Fig. 2).

Peaks TP120 and TP103 were collected from RP-HPLC and subjected to Edman degradation for sequence determination. In both sequenced peptides, two potential consensus sequences for *N*-glycosylation (NIS for peptide TP120 in PD-S3 and NAS for peptide TP103 in PD-S2) were found (Table 2). Furthermore, as previously reported for PD-S2 (Del Vecchio Blanco et al., 1997), the residue corresponding to Asn120 in PD-S3 was not assigned also in PD-S1 (reported as X in Table 2). Furthermore, the amino acid residues at position 112 (Asn in PD-S2) were not assigned in PD-S1 and PD-S3 (reported as X in Table 2). On the contrary, positions 120 and 112 of PD-S3 and PD-S2, respectively, were identified as Asn by Edman degradation, proving that their side chains do not carry glycan moieties.

2.3. Precursor ion discovery for identification and characterization of glycopeptides by LC/ESI-MS/MS

Peptide mapping and Edman degradation experiments reported above strongly suggested the presence of glycan chains linked to all three PD-Ss. Therefore, we decided to characterize them by using a recently developed mass spectrometric approach, successfully used to study protein post-translational modifications such as phosphorylation (Bateman et al., 2002; Lu et al., 2005b), glycosylation (Zhang et al., 2004), *S*-nitrosylation (Lu et al., 2005a) and, more

Table 1
Amino acid compositions of PD-S1, PD-S2 and PD-S3

Amino acid	PD-S2 from the sequence	PD-S2[a] from acid hydrolyzate	PD-S3[a] from acid hydrolyzate	PD-S1[a] from acid hydrolyzate
Asx	30	29.82 (30)	29.74 (30)	29.92 (30)
Thr	17	17.87 (18)	17.90(18)	17.71 (18)
Ser	28	26.80 (27)	26.96 (27)	26.77 (27)
Glx	22	22.57 (23)	22.65 (23)	22.88 (23)
Pro	11	11.47 (11)	11.33 (11)	11.37 (11)
Gly	13	14.20 (14)	14.36 (14)	14.27 (14)
Ala	17	17.80 (18)	17.69 (18)	17.92 (18)
½ Cys	4	3.67 (4)	3.59 (4)	3.78 (4)
Val	17	16.25 (16)	16.33 (16)	16.41 (16)
Met	5	4.59 (5)	4.71 (5)	4.95 (5)
Ile	17	16.27 (16)	16.21 (16)	16.39 (16)
Leu	24	23.30 (23)	23.41 (23)	23.47 (23)
Tyr	12	10.97 (11)	11.17 (11)	11.33 (11)
Phe	9	8.87 (9)	9.09 (9)	8.69 (9)
His	3	2.85 (3)	3.13 (3)	2.93 (3)
Lys	23	21.78 (22)	21.67 (22)	21.93 (22)
Arg	11	11.91 (12)	11.73 (12)	11.67 (12)
Trp	2	nd	nd	nd
Total amino acids	265	262	262	262

The closest integers of the calculated number of residues are given in parentheses.
[a] Residues are expressed as number of residues/mol of protein.

recently, for gonadotropin-releasing hormone (GnRH) detection in complex biological mixtures (Chambery et al., 2008). The method is based on the "precursor ion discovery" (PID) mode, that exploits the capability of the Q-TOF analyzer to quickly record alternate mass spectra at low and high collision energy of both precursor and product ion spectra, following liquid chromatographic separation of complex tryptic digests. The great advantages of the PID mode compared to conventional tandem MS experiments are due to the high selectivity of glycopeptide identifications. Briefly, in the PID experiment the voltage on the collision cell is rapidly switched between low (10 V) and high (30 V) voltage to obtain spectra of intact and fragmented peptides, respectively. For glycopeptide detection, the method has been set up by monitoring specific and diagnostic carbohydrate b ions [such as $(Hex)^{+1}$, *m/z* 163.060; $(HexNAc)^{+1}$, *m/z* 204.084; and $(HexHexNAc)^{+1}$, *m/z* 366.139)], deriving from the preferential fragmentation of glycopeptides at high collision energies, due to the very labile nature of glycosidic bonds. Only if these criteria (*i.e.* when the diagnostic sugar product ions are detected) are met during the high collision energy scan, the instrument automatically switches to record the full MS/MS spectra on target precursor ion (*i.e.* the glycosylated peptide) to acquire structural data before reverting to acquisition of alternate spectra in the two survey modes.

A representative PID chromatogram obtained with CapLC/Q-TOF for PD-S1 tryptic digest is shown in Fig. 3. The base peak intensity (BPI) chromatograms for the lower and higher collision energy (CE) data are reported in Fig. 3A and B, respectively. The MS/MS BPI chromatogram, triggered by the characteristic carbohydrate b ions (Fig. 3C), clearly showed that the only tandem mass spectra acquired during the chromatographic separation were registered for the doubly charged ion at *m/z* 1386.56 (precursor ion, 2770.74), at the retention time of 24.4 min and the triply-charged ion at *m/z* 567.58 (precursor ion, 1699.70), at the retention time of 33.25 min. In contrast, when a classical LC-MS/MS approach in data dependent analysis mode was performed on the same sample, the complicated peptide elution profiles in the BPI (Fig. S1A, Supplementary material), contained very populated MS/MS traces (Fig. S1B, S1C and S1D, Supplementary material), due to the high number of tryptic peptides in the mixture. Under these conditions, the manual review of data to identify glycopeptides, can be very laborious without a strategy for the selection of relevant data. It must be also underlined that tryptic digestions yield about five times more peptides than glycopeptides. The latter are often heterogeneous owing to the variable sugar content and, furthermore, glycopeptides have a much lower sensitivity than peptides in ESI. Nevertheless, the MS and MS/MS data acquired from survey scan experiments were useful to obtain a nearly complete coverage of the three PD-Ss, resulting in the determination of the primary structures of PD-S1 and PD-S3, found to be identical to that of PD-S2 (data not shown). These results were expected and further confirmed the reported peptide mapping and amino acid analyses of the three proteins (see earlier).

Interestingly, as reported above, when PD-S2 and PD-S3 tryptic digests were analyzed using the PID mode, only the doubly charged ion at *m/z* 1386.62 and the triply-charged ion at *m/z* 567.62, respectively, were detected in the MS/MS trace (Fig. S2A and S2B, respectively, see Supplementary material). The glycopeptides identified in the PID experiments were then deconvoluted on the basis of fragment ion charge states, and their spectra were analyzed to gain insights into the glycosylation site and glycan chain structure and composition. The MS/MS spectrum of the doubly charged ion at *m/z* 1386.62 (theoretical molecular mass, 2771.24 Da), identified in PD-S1 PID analysis and potentially corresponding to a glycopeptide, is shown in the *m/z* range 1400–2775 Da, where the larger glycopeptide fragments are observed (Fig. 4). The molecular ion at *m/z* 1600.75, corresponding to the deglycosylated peptide TP120 (theoretical molecular mass, 1600.75 Da), was clearly detected. The characteristic ion (peptide + HexNAc) was also found at *m/z* 1803.80, with a rather high intensity. The clear-cut fragmentation pattern allowed the complete reconstruction of the oligosaccharide chain on peptide TP120 [Hex_3, $HexNAc_2$, $deoxyHex_1$, $Pentose_1$ ($[M+H^+]$, 1171 u)]. Our mass spectrometry analysis confirms previous data reporting, for RIP carbohydrates, glycan structures constituted by a standard plant paucidomannosidic *N*-glycosylation pattern [Man_3, $GlcNAc_2$, Fuc_1, Xyl_1] (Barbieri et al., 1993; Daubenfeld et al., 2005). Also, in the lower mass region of the MS/MS spectrum of the glycopeptide at *m/z* 1386.62, several diagnostic sugar-derived ions were revealed, facilitating the carbohydrate moiety assignments (data not shown). The triply-charged ion at *m/z* 567.62 (theoretical molecular mass, 1699.86), was analyzed in a similar manner. Also in this case, the molecular ion at *m/z* 1497.83, corresponding to the deglycosylated peptide TP103 (theoretical molecular mass, considering the presence of a *S*-pyridylethylcysteine residue, 1497.70 Da), was clearly detected (data

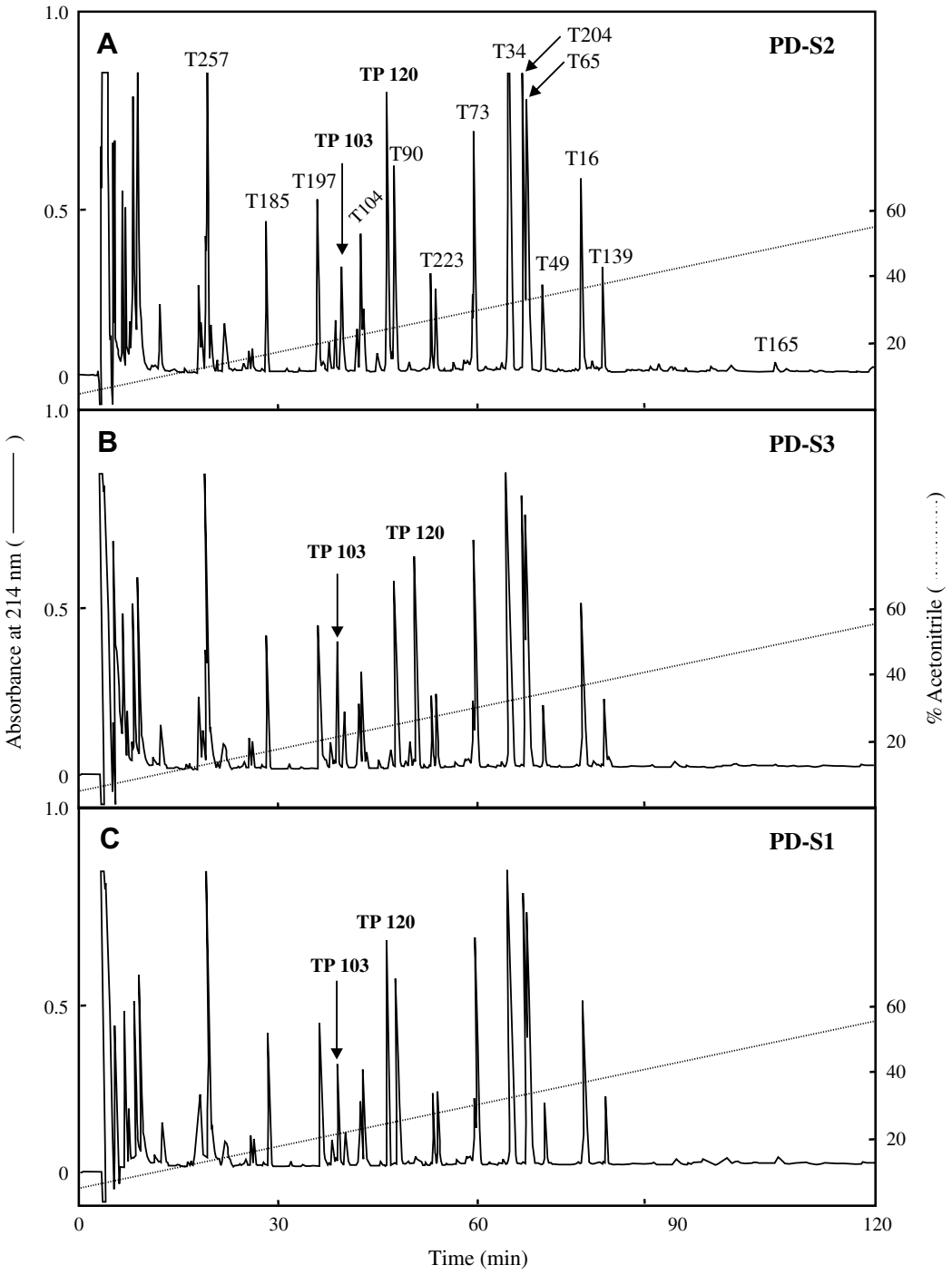

Fig. 2. RP-HPLC mapping of tryptic peptides of PD-S1 (C) and PD-S3 (B) compared to that of PD-S2 (A). Shifted peptides are labelled in bold as TP120 and TP103, while the other peaks, whose elution time does not change between tryptic maps, are indicated as reference only on PD-S2 elution profile. The labelling has been performed according to the sequence position of the first amino acid residue of tryptic peptides.

not shown). A difference of 203 Da was measured between the mass of the deglycosylated peptide and the mass of the precursor ion at *m/z* 1700.88, suggesting the presence on this peptide of a HexNAc residue (data not shown).

Taking into account the molecular masses of the sugar moieties and the presence of two disulfide bridges (see Sections 2.1 and 2.4), the M_r values of all PD-Ss were found in good agreement with those experimentally determined by ESI/Q-TOF mass spectrometry on native proteins. The presented data indicate that PD-S1, PD-S2 and PD-S3 share the same primary structure, but differ for their glycosylation patterns on Asn120 and Asn112. In particular, PD-S2 is glycosylated on Asn120 and PD-S3 on Asn112, while PD-S1 is glycosylated on both asparaginyl residues (Fig. 5). A so-called paucidomannosidic *N*-glycosylation pattern (Rayon et al., 1998) has been found on Asn120, while only a HexNAc residue is linked to Asn112.

2.4. Localization of intra-chain disulfide bridges by MALDI-TOF MS

Primary structure analyses revealed that a total of four cysteinyl residues are present in the PD-S2 sequence: Cys34, Cys88, Cys110 and Cys262 (Del Vecchio Blanco et al., 1997). As previously deter-

Table 2
Amino acid sequences, determined by Edman degradation, and sequence position (referred to PD-S2) of peptides TP120 and TP103 (see Fig. 2)

Peptide ID	Protein	Sequence position	Sequence
TP120	PD-S1	120–133	**X**ISYDSSYPALENK
	PD-S2	120–133	**X**ISYDSSYPALENK
	PD-S3	120–133	NISYDSSYPALENK
TP103	PD-S1	103–116	KDVETTLC**X**ASSR
	PD-S2	103–116	KDVETTLCNASSR
	PD-S3	103–116	KDVETTLC**X**ASSR

The potential glycosylation sites for peptide TP120 (NIS) in PD-S3 and TP103 (NAS) in PD-S2, respectively, are underlined. Residues not assigned by Edman degradation (X) are reported in bold.

mined by the Ellman reagent, PD-S2 contains no free sulfhydryl groups, suggesting the occurrence of two intramolecular disulfide bridges. To experimentally identify the S–S pairing of PD-Ss, we performed a comparative MALDI-TOF mapping using tryptic peptides derived from PD-S2, the most abundant isoform, in absence and presence of β-mercaptoethanol as reducing agent (see Section 4).

In Fig. 6 the spectra obtained for the two samples are compared. Two signals at *m/z* 2572.14 and 1667.69 are clearly identified in the spectrum of the tryptic digest (see arrows in Fig. 6B and spectrum magnifications in Supplementary Fig. 3). It can be calculated that these ions correspond, respectively, to tryptic peptides with sequence positions 34–48/257–265 (theoretical [M+H$^+$]: 2571.17) and 88–89/104–116 (theoretical [M+H$^+$]: 1667.75), linked through disulfide bridges. After β-mercaptoethanol treatment, the two peaks disappeared, whereas two additional ion signals with *m/z* 1658.96 and 1392.80, corresponding to tryptic peptides with sequence positions 34–48 (theoretical [M+H$^+$]: 1658.80) and 104–116 (theoretical [M+H$^+$]: 1392.64), respectively, were clearly detected (see arrows in Fig. 6A and spectrum magnifications in Fig. S4A and S4B of Supplementary material). Although not easily visible in the full range spectrum, also peptide with sequence positions 257–265 (theoretical [M+H$^+$]: 915.39) was detected in the spectrum of the reduced mixture (see spectrum magnification in Fig. S4C, Supplementary material). The dipeptide 88–89 (theoretical [M+H$^+$]: 278.13) could not be detected due to its low molecular weight. Overall, from the reported data, the pattern of disulfide bridges can be assigned with confidence as Cys34 linked to Cys262 and Cys88 to Cys110.

3. Discussion

The amino acid sequence and the glycosylation motifs of three PD-Ss, ribosome-inactivating proteins isolated from *P. dioica* seeds, have been characterized by LC-peptide mapping assisted by Edman degradation and ESI–Q-TOF MS.

PD-Ss consist of three different post-translational modified forms. Mapping analysis of PD-S1 and PD-S3 tryptic peptides is consistent with the amino acid sequence previously published for PD-S2 (Del Vecchio Blanco et al., 1997). Furthermore, by ESI–Q-TOF MS in PID mode, the standard plant paucidomannosidic *N*-glycosylation pattern (Hex_3, $HexNAc_2$, $deoxyhexose_1$, $pentose_1$) was found for PD-S1 and PD-S2 on Asn120. It has been found in other RIPs (Barbieri et al., 1993; Daubenfeld et al., 2005) and observed for the majority of known vacuolar glycoproteins (Lerouge et al., 1998). As for other type 1 RIPs (Daubenfeld et al., 2005), also for PD-S1 and PD-S2 a vacuolar localization in *P. dioica* seed cells could be hypothesized.

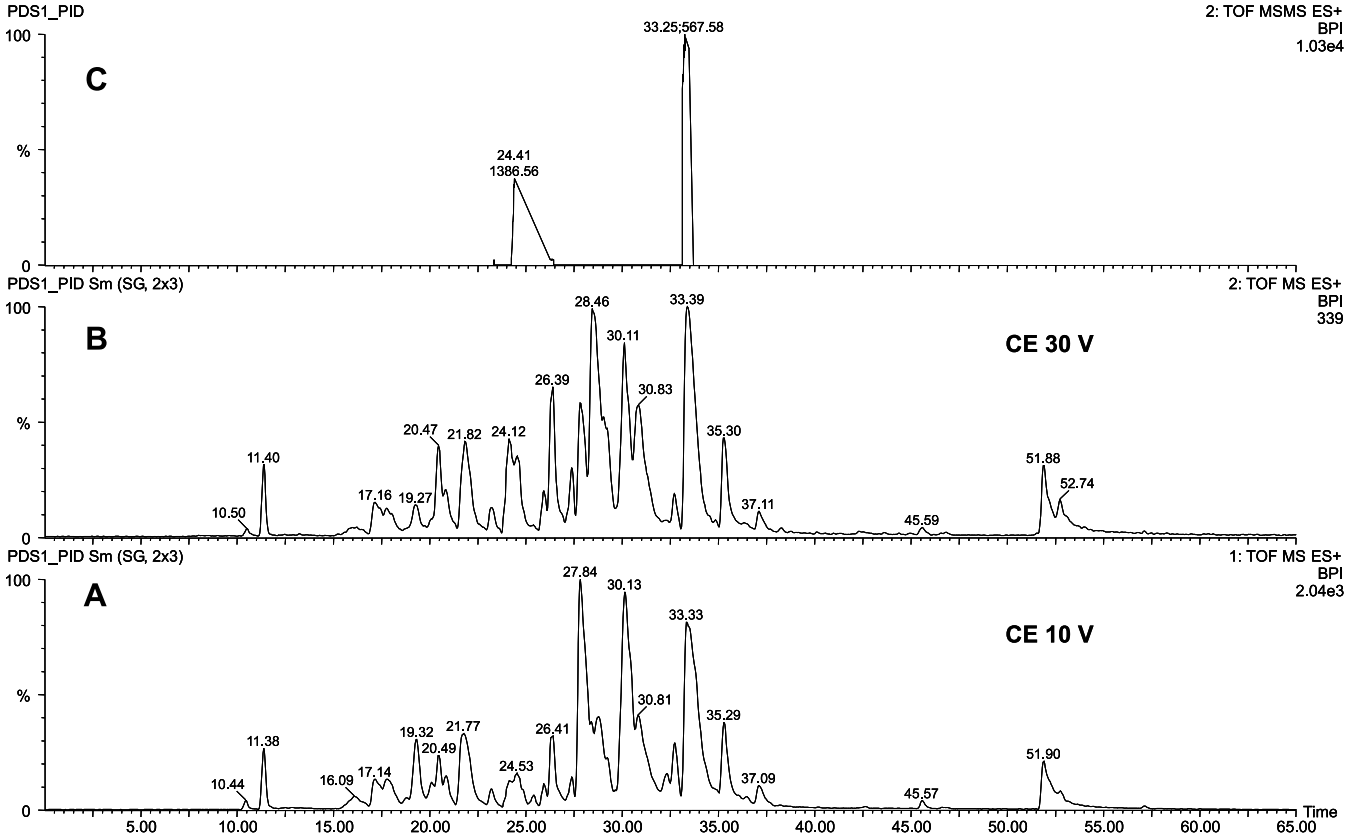

Fig. 3. Base peak intensity (BPI) chromatograms deriving from LC/MS acquisition in PID mode of PD-S1 tryptic digest. Low and high collision energy (CE) chromatograms are reported in (A) and (B), respectively. (C) MS/MS trace triggered by detection of carbohydrate b ions (see Section 4).

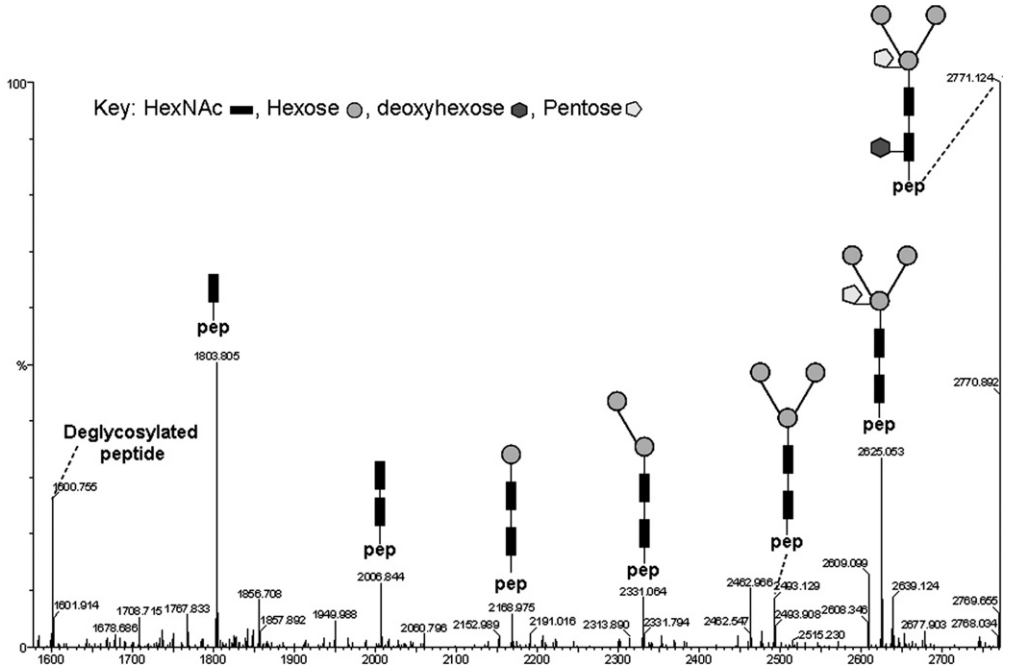

Fig. 4. MS/MS spectrum of the doubly charged ion at *m/z* 1386.62 (theoretical molecular mass, 2771.24 Da), identified in PD-S1 PID analysis. The region of *m/z* 1400–2775 Da, in which the larger glycopeptide fragments have been identified, is reported. The molecular ion at *m/z* 1600.75, corresponding to the deglycosylated peptide α (theoretical molecular mass, 1600.75 Da), is marked. The complete reconstruction of the oligosaccharide chain on peptide TP120 [Hex_3, $HexNAc_2$, $deoxyhexose_1$, $pentose_1$ ([$M+H^+$], 1171 u)] is also reported.

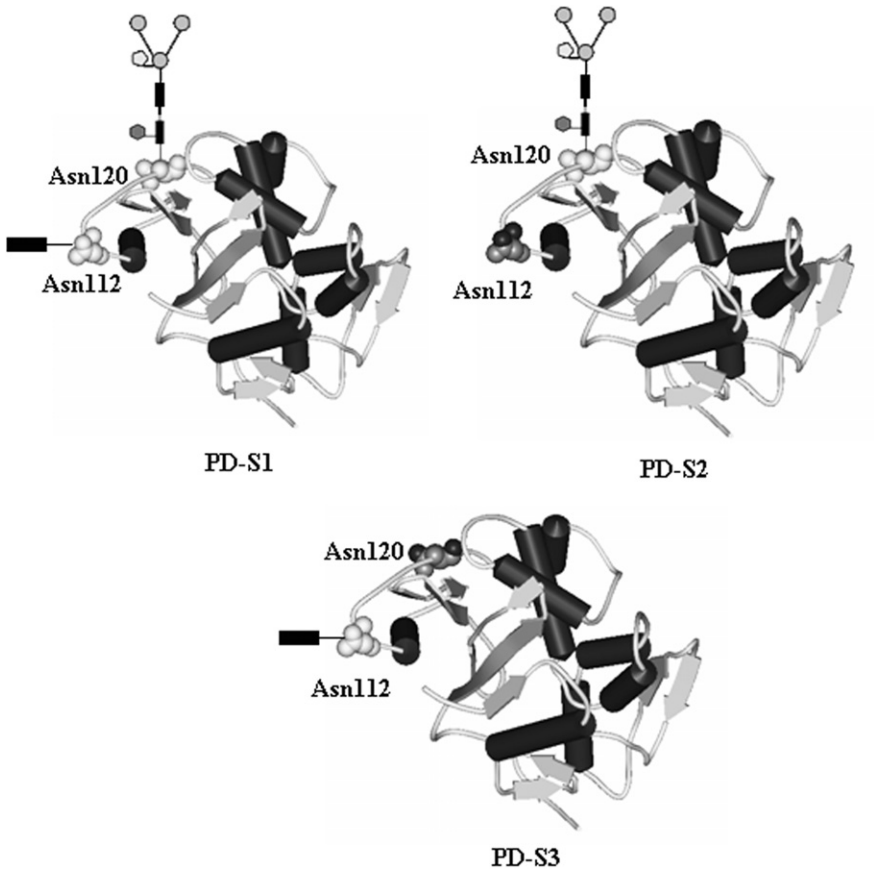

Fig. 5. Schematic representation of the post-translational modifications of the three PD-S forms. The PD-Ss model has been obtained using the comparative protein modelling as described in Section 4. Asn112 and Asn120, when glycosylated (see Fig. 4 for symbol key), have been reported in light grey.

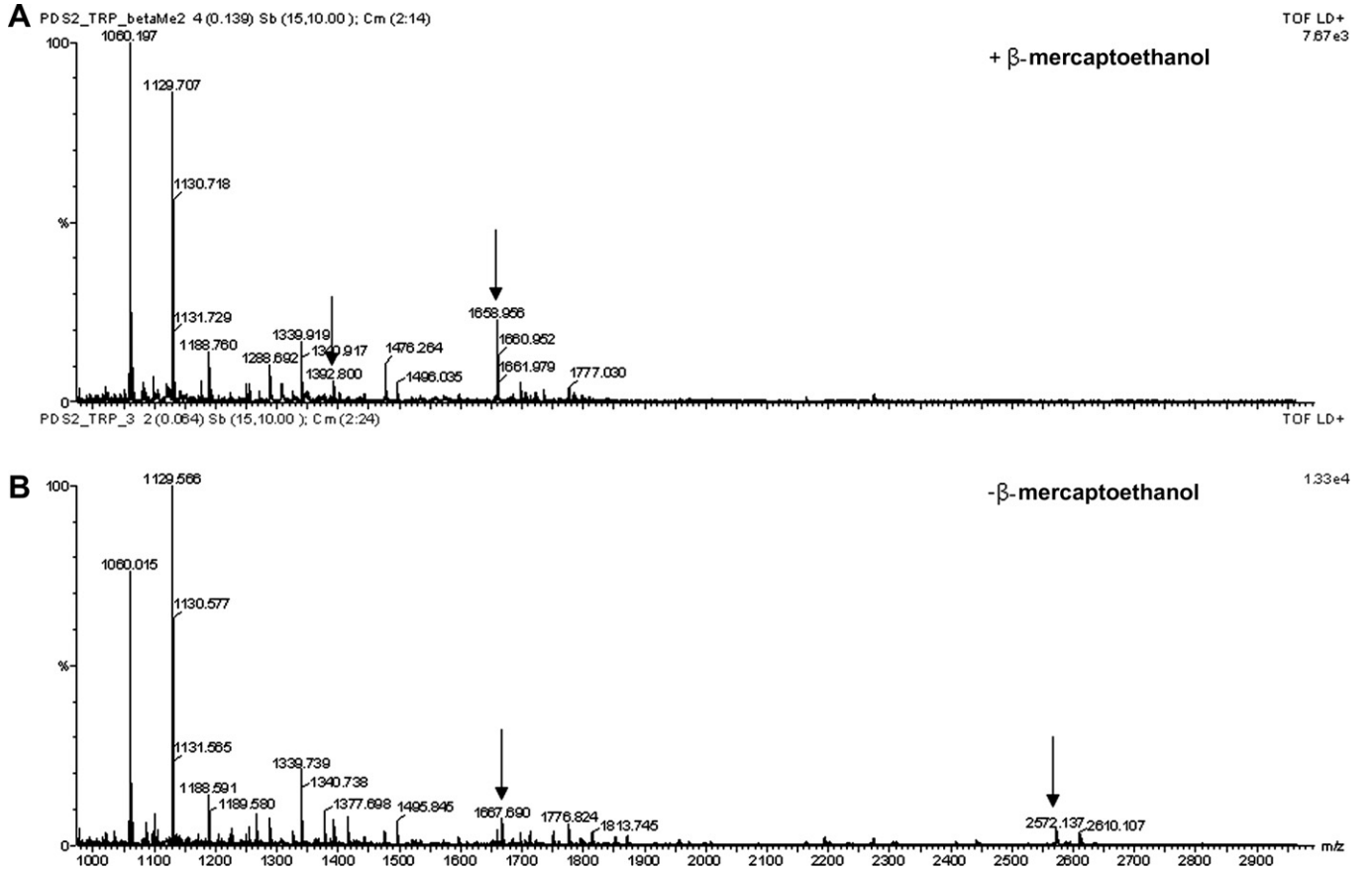

Fig. 6. MALDI-TOF spectra of PD-S2 tryptic peptides both in the presence (A) and absence (B) of β-mercaptoethanol as reducing agent. The arrows indicate the cysteine containing peptides (see text for details).

Worth noting, a glycosylation site carrying only a HexNAc residue has been identified for the first time in type 1 RIPs on Asn112 in PD-S1 and PD-S3.

Most functional proteins in eukaryotes, including plants, are known to be glycosylated. To date, the function of protein glycosylation is still an area of intense investigation in glycobiology. *N*-glycosylation occurs on nascent chains as a post-translational modification of proteins, whereas the subsequent de-*N*-glycosylation takes place later, as a new post-translational re-modification of proteins to gain their function (Ceriotti et al., 1998). In particular, the *N*-glycosylation of plant proteins starts with the transfer to endoplasmic reticulum (ER) of the oligosaccharide precursor $Glc_3Man_9GlcNAc_2$ from a dolichol lipid carrier to specific asparaginyl residues on the nascent polypeptide chain, included in the Asn-Xaa-Ser/Thr/Cys consensus sequence. The precursor is subsequently modified by glycosidases and glycosyl transferases during the transport of glycoproteins to their final location (Rayon et al., 1998).

Several roles have been ascribed to *N*-linked glycans, such as prevention of proteolytic degradation, induction of the correct folding and influence on protein conformation, stability and biological activity, involvement in protein recognition and cell–cell adhesion processes (Elbein, 1991; Lis and Sharon, 1993; Sharon and Lis, 1993; Varki, 1993). Regarding protein folding and stability, a direct contribution of *N*-glycans has been related to the increase of protein solubility, the reduced tendency to aggregate, but also to the presence of additional hydrogen bonds and hydrophobic interactions between the oligosaccharide and the protein (O'Connor and Imperiali, 1996; Wyss and Wagner, 1996). After long debate about the direct, indirect or null effects of the glycan chains in glycoprotein folding and structure, a widely accepted view is that *N*-glycosylation directly affects local and/or global conformational dynamics of the nascent polypeptide by imposing structural constrains and interactions that limit the number of folding pathways and prevent aggregation of folding intermediates. This may not necessarily directly affect the final protein structure, but it can increase both rate and yield of protein maturation processes (Ceriotti et al., 1998).

Furthermore, the influence of *N*-glycans on protein folding could also be related to protein sorting, as a correct folding is, in most cases, a pre-requisite for transport to the Golgi complex (Pedrazzini et al., 1997; Vitale and Ceriotti, 2004). Indeed, the functions of *N*-glycans, and in particular of the paucidomannosidic type, in the targeting of proteins into the vacuole, the plant equivalent of animal lysosome, has been extensively studied (Rayon et al., 1998). Recently, the growing interest in protein plant *N*-glycosylation is mainly related to the biotechnological applications that exploit the transgenic plants as system for the expression of therapeutic recombinant glycoproteins (Lerouge et al., 2000). As a consequence, *N*-glycosylation of recombinant proteins is one of the major problems for the production of functional glycoproteins, supporting the need to have experimental systems for further investigation of plant protein glycosylation. To date, most plant glycoproteins have been selected for *N*-glycosylation studies mainly on the basis of their abundance, as often they accumulate in large amount in plant seeds as storage proteins, such as the bean model glycoprotein phytohaemoagglutinin (Rayon et al., 1998). In addition, glycan structural analyses have rarely been carried out on several glycoproteins from the same plant. As a consequence, the heterogeneity of plant systems used for these studies makes any comparison very difficult.

In this context, the three glycosylated forms of PD-Ss, isolated from *P. dioica* seeds and characterized in the present work, appear to constitute an excellent model system for the study of the plant

protein *N*-glycosylation because they: (i) are expressed in large amounts; (ii) can be easily and quickly purified by classical chromatographic techniques; and (iii) posses two *N*-glycosylation sites carrying discrete oligosaccharidic moieties on the same primary structure bulk.

4. Experimental

4.1. Materials

Seeds of *P. dioica* L., collected in autumn from a single female plant grown in the Botanical Garden of the University of Naples, Federico II, were a kind gift of Drs. P. De Luca and G. Siniscalco Gigliano. Materials for chromatography were described elsewhere (Di Maro et al., 1999). HPLC grade solvents and reagents were obtained from Carlo Erba (Milan, Italy). Trypsin (sequencing grade) was purchased from Boehringer Mannheim (Monza, Italy). All other reagents and chemicals were of analytical grade. Reagents for amino acid analyses and for automated sequence determination were from Biochrom Ltd. (Cambridge, UK) and Applera Italia (Monza, Italy) respectively. The following buffers and solvents have been used: buffer A, 5 mM Na-phosphate, pH 7.2, containing 0.15 M NaCl; buffer B, 10 mM Na-acetate, pH 4.5; buffer C, 5 mM Na-phosphate pH 7.2, containing 1 M NaCl; buffer D: 5 mM Na-phosphate, pH 7.2; buffer E: 5 mM Na-phosphate, pH 7.2, containing 0.35 M NaCl. Solvent A, 0.1% TFA; solvent B, acetonitrile containing 0.1% TFA; solvent C, 5% CH_3CN, containing 0.1% formic acid solvent; solvent D, 2% CH_3CN in MilliQ water; solvent E, 95% CH_3CN in MilliQ water.

4.2. PD-Ss purification

PD-Ss were purified essentially according to the procedure published by Parente et al. (1993) with some modifications. Four sequential chromatographies were used: (i) a Streamline SP column (ϕ 3 × 10 cm, Amersham Pharmacia; Milan, Italy) in place of the *S*-Sepharose ion exchange, equilibrated in buffer B and eluted with buffer C; (ii) a gel-filtration on Sephacryl S-100 column (ϕ 2.6 × 120 cm) as second step, equilibrated and eluted with buffer E; and (iii) a CM-Sepharose column (ϕ 1.4 × 25 cm) equilibrated in buffer D and eluted with a NaCl gradient up to 0.15 M (reservoir A, 500 ml, reservoir B, 500 ml; total volume 1l). Modifications were also introduced in the final step to obtain homogeneous protein preparations suitable for mass spectrometry analyses. Indeed, a further purification step was introduced by FPLC on an AKTA Purifier System (Amersham Pharmacia; Milan, Italy), using a Source 15S PE 4.6/100 column, equilibrated in buffer D and eluted applying a linear gradient from 0 to 100% of buffer A over 60 min.

4.3. Protein electrophoresis

Homogeneity of isolated proteins was determined by SDS–PAGE with a Mini-Protean II mini-gel apparatus (Bio-Rad; Milan, Italy), using 6% (w/v) stacking polyacrylamide gel and 12% (w/v) separation gel (Laemmli and Favre, 1973).

4.4. N-terminal amino acid sequencing by Edman degradation

For the amino acid sequence analysis by automated Edman degradation, native PD-S1 and PD-S3 were purified by RP-HPLC using the experimental conditions described in Section 4.5 for protein desalting. Purified proteins were dried, resuspended in 50% (v/v) acetonitrile, adsorbed on a Biobrene™-treated filter and analyzed by Edman degradation on a Procise Model 491C sequencer (Applied Byosystem Inc., Foster City (USA); Chambery et al., 2006).

4.5. Reduction, S-pyridylethylation and RP-HPLC purification of proteins and peptides

Prior to tryptic digestion, PD-S1 and PD-S3 were reduced and *S*-pyridylethylated with 4-vinylpyridine as previously reported (Chambery et al., 2006). The modified proteins were desalted by RP-HPLC using a Vydac C_4 column (0.46 × 25 cm; 5 μm particle size) and eluted with a linear gradient of solvent A and solvent B over 60 min, at a flow rate of 1 ml/min. Separation of tryptic peptides by RP-HPLC was performed on a Waters Breeze instrument equipped with a Waters Symmetry C_{18} column (0.46 × 15 cm; 5 μm particle size). Peptide elution was obtained using a linear gradient of solvent A and solvent B, from 5% to 55% of solvent B over 150 min.

4.6. Analytical methods

Amino acid analyses were performed as described (Del Vecchio Blanco et al., 1997). Experimental conditions for enzymatic cleavage with trypsin were as described previously (Di Maro et al., 2001). The enzyme was added in three steps with a finale enzyme-to-substrate ratio of 1:50 (w/w) and the reaction was carried out at 37 °C for 4 h.

4.7. Mass spectrometry of native proteins

The relative molecular masses (M_r) of PD-S1, PD-S2 and PD-S3 were determined using a Q-TOF *Micro* mass spectrometer after protein desalting by RP-HPLC as described in Section 4.5 (Waters, Milford, MA, USA) equipped with a CapLC system as previously reported (Chambery et al., 2006). The capillary source voltage and the cone voltage were set at 3000 V and 43 V, respectively. The source temperature was kept at 80 °C and nitrogen was used as a drying gas (flow rate about 50 l/h). RP-HPLC purified proteins at a concentration of about 1 pmol/μl were infused into the system at a flow rate of 5 μl/min. The acquisition and deconvolution of data were performed by Mass Lynx 4.0 software.

4.8. Identification glycopeptides by LC-MS/MS using the precursor ion discovery mode

Tryptic peptides were separated by means of a modular CapLC system connected to the Z-spray source of a Q-TOF *Micro* (Waters Manchester, UK) equipped with the *LockSpray* interface as reported (Chambery et al., 2006).

The sample was loaded onto a Symmetry300 C_{18} pre-column (0.18 × 23.5 mm, 5 μm particle size; Waters Manchester, UK) at a flow rate of 20 μl/min and desalted for 5 min using solvent C as mobile phase at a flow rate of 5 μl/min. The sample was then eluted from the C_{18} pre-column and directed onto a Symmetry-C_{18} analytical column (15 cm × 300 μm ID) for peptide separation. The chromatographic separation was carried out using a linear gradient from 5% solvent D to 55% solvent E, both containing 0.1% formic acid over 30 min at a flow rate of 5 μl/min. Electrospray MS and MS/MS data were acquired on a Q-TOF *Micro* mass spectrometer fitted with a Z-spray electrospray ion source. The mass spectrometer was operated in the positive ion mode with a source temperature of 80 °C and with a potential of 3000 V applied to the capillary probe. The instrument was calibrated with a multi-point calibration using selected fragment ions, resulted from the collision-induced decomposition (CID) of human [Glu_1]-fibrinopeptide B [500 fmol/μl in CH_3CN/H_2O (50:50), 0.1% formic acid] at an infusion rate of 5 μl/min in the TOF MS/MS mode. The instrument resolution in MS/MS mode for the [Glu_1]-fibrinopeptide B fragment ion at *m/z* 684.3469, was found to be above 5000 FWHM (full width at half-maximum). To perform standard LC-MS/MS experi-

ments on PD-S tryptic digests, the automatic function switching (data directed analyses) was employed. In this procedure, the MS/MS experiment is conducted without any prior knowledge of the origin of the precursor ions. Charge state recognition was adopted to select doubly- and triply-charged precursor ions for the MS/MS experiments, which also includes the automated selection of the collision energy based on charge and mass. A maximum of three precursor masses was defined for concurrent MS/MS acquisition from a single MS survey scan. MS/MS fragmentation spectra were collected from *m/z* 50 to *m/z* 1600.

The glycopeptide-containing fractions were identified with the instrument operating in precursor ion discovery (PID) mode (Bateman et al., 2002; Chambery et al., 2008), where the voltage on the collision cell was switched alternately between high (30 V) and low (10 V) collision energy every second. Upon detection of carbohydrate b ions (*m/z* 163.060, 204.084, 292.0954 and 366.139) with a threshold of 10 counts/s, the instrument switched to MS/MS mode and selected the most intense multiply charged ions (from 2^+ to 7^+) for fragmentation. Only when the selected precursor ion yielded the expected carbohydrate b ions the MS/MS was fully acquired. Mass spectrometer parameters were set as follows: cone voltage, 35 V; capillary voltage, 3.0 kV; collision energy for MS/MS data depended on the charge state and *m/z* range; MS data acquisition range, *m/z* 300–2000; MS/MS data acquisition range *m/z* 50–1500. For both MS and MS/MS acquisitions scan time and inter-scan delay were set to 1 s and 0.1 s, respectively. The MS/MS data were processed using MaxEnt3 in MassLynx 4.0 software (Waters Manchester, UK) for de-isotoping and deconvolution. Carbohydrate structure analysis reconstruction was manually performed with the assistance of Carbotools of the Biolynx application in MassLynx 4.0 software.

4.9. Localization of intra-chain disulfide bridges by MALDI-TOF MS

To assess the cysteinyl pairing, a tryptic peptide mapping was performed by MALDI-TOF mass spectrometry on PD-S2 (the most abundant isoform), both in absence and presence of β-mercaptoethanol as reducing agent. PD-S2 (2 μg) was subjected to tryptic digestion in 10 mM Tris·Cl, pH 8.4, 20 mM $CaCl_2$, 10% CH_3CN with two subsequent additions of trypsin (1:200 for 3 h, then 1:100 for 2 h, at 37 °C). At the end of digestion, an aliquot (50 ng) of the tryptic digest was treated with 0.01% β-mercaptoethanol for 10 min at room temperature. Tryptic peptides, before and after disulfide reduction, were then desalted on ZipTip C18 (Millipore, Billerica, USA) and analyzed with a MALDI-TOF LR (Waters, Manchester, UK). Prior to spectra acquisition, 1 μl of each peptide solution was mixed with 1 μl of saturated α-cyano-4-hydroxycinnamic acid matrix solution [10 mg/ml in ethanol:acetonitrile (1:1; v:v), containing 0.1% trifluoroacetic acid] and a droplet of the resulting mixture (1 μl) placed on the mass spectrometer's sample target. The droplet was dried at room temperature. Once the liquid was completely evaporated, the sample was loaded into the mass spectrometer and analyzed. The instrument was externally calibrated using a tryptic alcohol dehydrogenase digest (Waters, Manchester, UK) in reflectron mode. All spectra were processed and analyzed using the MassLynx 4.0 software (Waters, Manchester, UK).

4.10. Molecular modelling of PD-Ss

A model for PD-Ss was constructed using the SWISS-MODEL system available online at http://swissmodel.expasy.org/ for comparative protein modelling. Crystallographically derived coordinates of the pokeweed antiviral protein (PDB entry code 1APA) refined to 2.3 Å resolution, were used as a template structure (Ago et al., 1994).

Acknowledgements

This research was supported by Grants from the Ministero Italiano dell'Istruzione, dell'Università e della Ricerca (MIUR; PRIN 2005/2007), and funds from the Seconda Università di Napoli.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.phytochem.2008.04.005.

References

Aceto, S., Di Maro, A., Conforto, B., Siniscalco, G.G., Parente, A., Delli Bovi, P., Gaudio, L., 2005. Nicking activity on pBR322 DNA of ribosome inactivating proteins from *Phytolacca dioica* L. leaves. Biol. Chem. 386, 307–317.

Ago, H., Kataoka, J., Tsuge, H., Habuka, N., Inagaki, E., Noma, M., Miyano, M., 1994. X-ray structure of a pokeweed antiviral protein, coded by a new genomic clone, at 0.23 nm resolution. A model structure provides a suitable electrostatic field for substrate binding. Eur. J. Biochem. 225, 369–374.

Barbieri, L., Battelli, M.G., Stirpe, F., 1993. Ribosome-inactivating proteins from plants. Biochim. Biophys. Acta 1154, 237–282.

Bateman, R.H., Carruthers, R., Hoyes, J.B., Jones, C., Langridge, J.I., Millar, A., Vissers, J.P., 2002. A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. J. Am. Soc. Mass Spectrom. 13, 792–803.

Bolognesi, A., Polito, L., 2004. Immunotoxins and other conjugates: pre-clinical studies. Mini Rev. Med. Chem. 4, 563–583.

Ceriotti, A., Duranti, M., Bollini, R., 1998. Effects of *N*-glycosylation on the folding and structure of plant proteins. J. Exp. Bot. 49, 1091–1103.

Chambery, A., de Donato, A., Bolognesi, A., Polito, L., Stirpe, F., Parente, A., 2006. Sequence determination of lychnin, a type 1 ribosome-inactivating protein from *Lychnis chalcedonica* seeds. Biol. Chem. 387, 1261–1266.

Chambery, A., Severino, V., D'Aniello, A., Parente, A., 2008. Precursor ion discovery on a hybrid quadrupole-time-of-flight mass spectrometer for gonadotropin-releasing hormone detection in complex biological mixtures. Anal. Biochem. 374, 335–345.

Daubenfeld, T., Hossann, M., Trommer, W.E., Niedner-Schatteburg, G., 2005. On the contentious sequence and glycosylation motif of the ribosome inactivating plant protein gelonin. Biochem. Biophys. Res. Commun. 333, 984–989.

Del Vecchio Blanco, F., Bolognesi, A., Malorni, A., Sande, M.J., Savino, G., Parente, A., 1997. Complete amino-acid sequence of PD-S2, a new ribosome-inactivating protein from seeds of *Phytolacca dioica* L. Biochim. Biophys. Acta 1338, 137–144.

Di Maro, A., Ferranti, P., Mastronicola, M., Polito, L., Bolognesi, A., Stirpe, F., Malorni, A., Parente, A., 2001. Reliable sequence determination of ribosome – inactivating proteins by combining electrospray mass spectrometry and Edman degradation. J. Mass Spectrom. 36, 38–46.

Di Maro, A., Valbonesi, P., Bolognesi, A., Stirpe, F., De Luca, P., Siniscalco Gigliano, G., Gaudio, L., Delli Bovi, P., Ferranti, P., Malorni, A., Parente, A., 1999. Isolation and characterization of four type-1 ribosome-inactivating proteins, with polynucleotide:adenosine glycosidase activity, from leaves of *Phytolacca dioica* L. Planta 208, 125–131.

Elbein, A.D., 1991. The role of N-linked oligosaccharides in glycoprotein function. Trends Biotechnol. 9, 346–352.

Ferrige, A.G., Seddon, M.J., Green, B.N., Jarvis, S.A., Skilling, J., 1992. Disentangling electrospray spectra with maximum entropy. Rapid Commun. Mass Spectrom. 6, 707–711.

Girbes, T., Ferreras, J.M., Arias, F.J., Stirpe, F., 2004. Description, distribution, activity and phylogenetic relationship of ribosome-inactivating proteins in plants, fungi and bacteria. Mini Rev. Med. Chem. 4, 461–476.

Laemmli, U.K., Favre, M., 1973. Maturation of the head of bacteriophage T4. I. DNA packaging events. J. Mol. Biol. 80, 575–599.

Lerouge, P., Bardor, M., Pagny, S., Gomord, V., Faye, L., 2000. *N*-glycosylation of recombinant pharmaceutical glycoproteins produced in transgenic plants: towards an humanisation of plant *N*-glycans. Curr. Pharm. Biotechnol. 1, 347–354.

Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Laine, A.C., Gomord, V., Faye, L., 1998. *N*-glycoprotein biosynthesis in plants: recent developments and future trends. Plant Mol. Biol. 38, 31–48.

Lis, H., Sharon, N., 1993. Protein glycosylation. Structural and functional aspects. Eur. J. Biochem. 218, 1–27.

Lu, X.M., Lu, M., Tompkins, R.G., Fischman, A.J., 2005a. Site-specific detection of *S*-nitrosylated PKB alpha/Akt1 from rat soleus muscle using CapLC-Q-TOF(micro) mass spectrometry. J. Mass Spectrom. 40, 1140–1148.

Lu, X.M., Lu, M.Y., Fischman, A.J., Tompkins, R.G., 2005b. A new approach for sequencing human IRS1 phosphotyrosine-containing peptides using CapLC-Q-TOF(micro). J. Mass Spectrom. 40, 599–607.

O'Connor, S.E., Imperiali, B., 1996. Modulation of protein structure and function by asparagine-linked glycosylation. Chem. Biol. 3, 803–812.

Parente, A., De Luca, P., Bolognesi, A., Barbieri, L., Battelli, M.G., Abbondanza, A., Sande, M.J., Gigliano, G.S., Tazzari, P.L., Stirpe, F., 1993. Purification and partial characterization of single-chain ribosome-inactivating proteins from the seeds of *Phytolacca dioica* L. Biochim. Biophys. Acta 1216, 43–49.

Parikh, B.A., Tumer, N.E., 2004. Antiviral activity of ribosome inactivating proteins in medicine. Mini Rev. Med. Chem. 4, 523–543.

Pedrazzini, E., Giovinazzo, G., Bielli, A., de Virgilio, M., Frigerio, L., Pesca, M., Faoro, F., Bollini, R., Ceriotti, A., Vitale, A., 1997. Protein quality control along the route to the plant vacuole. Plant Cell 9, 1869–1880.

Rayon, C., Lerouge, P., Faye, L., 1998. The protein *N*-glycosylation in plants. J. Exp. Bot. 49, 1463–1472.

Sharon, N., Lis, H., 1993. Carbohydrates in cell recognition. Sci. Am. 268, 82–89.

Stirpe, F., Battelli, M.G., 2006. Ribosome-inactivating proteins: progress and problems. Cell Mol. Life Sci. 63, 1850–1866.

Varki, A., 1993. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 3, 97–130.

Vitale, A., Ceriotti, A., 2004. Protein quality control mechanisms and protein storage in the endoplasmic reticulum. A conflict of interests? Plant Physiol. 136, 3420–3426.

Wyss, D.F., Wagner, G., 1996. The structural role of sugars in glycoproteins. Curr. Opin. Biotechnol. 7, 409–416.

Zhang, C., Doherty-Kirby, A., Huystee Rv, R., Lajoie, G., 2004. Investigation of cationic peanut peroxidase glycans by electrospray ionization mass spectrometry. Phytochemistry 65, 1575–1588.