Glycoconjugate Journal (1996) **13:** 575–583

Post-translational proteolytic processing and the isolectins of lentil and other Viciae seed lectins

N. MARTIN YOUNG[1]*, DAVID C. WATSON[1] and PIERRE THIBAULT[2]

[1]*Institute for Biological Sciences, National Research Council of Canada, Ottawa, Canada K1A 0R6*
[2]*Institute for Marine Biosciences, National Research Council of Canada, Halifax, Canada B3H 3Z1*

Received 12 July 1995, revised 27 September 1995

Electrospray mass spectrometry was used to identify precisely the proteolytic cleavage points within, and at the C-termini of, the proprotein forms of four Viciae lectins that give rise to their two-chain forms. The lectins examined were the pea and lentil lectins, favin and the *Lathyrus odoratus* lectin, which represent each of the four genera in this tribe. The molecular mass data showed single β-chain forms for each lectin, with masses consistent with the available sequence and glycopeptide data, indicating that each came from a single proprotein. In contrast, the pea, lentil and *L. odoratus* α-chains occurred in as many as five forms, due to multiple C-terminal cleavage points. Only favin showed a single α-chain form. The α-chain mass data were again consistent with the sequence information available, except for the lentil lectin α-chain which was re-determined by protein sequencing. The two isolectin forms of this protein were shown to arise from α-chain species with and without residue Lys53. The mass spectrum of concanavalin A was also examined and both the single-chain form and the two fragment forms showed no evidence of C-terminal heterogeneity.

Keywords: mass spectrometry, lectins, post-translational processing

Introduction

Proteins often occur in multiple isoforms detectable by methods such as iso-electric focusing. There are two possible origins for such forms, either genetic polymorphism or differences in post-translational processing. It has been laborious to discriminate between these two types of heterogeneity by protein chemical methods. However, the recent development of electrospray mass spectrometry [1] greatly facilitates this type of investigation by providing rapid and precise molecular mass measurement on much smaller amounts of protein. By comparison of the observed molecular mass with that calculated from the gene sequence and possible alterations arising from post-translational processing of the protein product, it is often possible to discriminate between the two mechanisms. Plant lectins frequently occur in isolectin forms, and we have used this approach to show that several single-chain legume lectins are polymorphic due to C-terminal proteolysis [2, 3] and that Moraceae lectins have a high number of isolectin forms due to combined genetic and post-translational differences [4].

Legume seed lectins have frequently undergone post-translational proteolysis from proprotein forms. This processing is best known for concanavalin A (Con A), where not only is a central glycopeptide excised but also a peptide ligation occurs in up to two-thirds of the molecules in the course of the removal of a C-terminal peptide, which re-orders the linear sequence [5, 6]. A less complex process occurs in the two-chain lectins of the Viciae tribe, which include the pea lectin, lentil lectin, favin and the various *Lathyrus* lectins. As exemplified by the pea lectin [7], these two-chain lectins are also cleaved into two chains with removal of an internal peptide and a C-terminal peptide from the proprotein. The internal cleavage sites are not as centrally located in the proprotein as those in Con A, and there is no peptide ligation. The resulting α- and β-chains have molecular masses of ~6000 and ~20 000 Da respectively. It has recently been shown that the single-chain lectins also undergo post-translational proteolysis, with up to 14 residues being removed from their C-terminal ends [2, 3, 8].

*To whom correspondence should be addressed

In most of the two-chain Viciae lectins, the exact sites of the proteolytic cleavages have not been assigned, largely due to the imperfect reliability of C-terminal identification methods. For example, the C-terminus of the pea lectin β-chain was reported to be Tyr179 on the basis of a tritium-labelling procedure [9], and Asn181 based on isolation of a peptide containing Tyr179 [10]. Recently, this problem was resolved in favour of the Asn181 site by characterization of C-terminal peptides from the pea isolectins, and a considerable degree of C-terminal heterogeneity was found in their α-chains [11]. This study also demonstrated that the two isolectins are products of the same gene.

Isolectin forms are also known for lentil lectin [12], but whether these are post-translational in origin as for the pea lectin or genetic in the manner of the *Lathyrus ochrus* lectin [13] was not known. Additionally, X-ray studies of four of the Viciae lectins are underway [14–17], and knowledge of the exact points of proteolytic cleavage is of interest for this structural work. We have therefore used electrospray mass spectrometry [1] to determine the molecular masses of the constituent chains of these lectins, from which the locations of the cleavage points and genetic diversity, if any, could be deduced. In three of these lectins, the processing led to 'ragged' ends on the α-chains, which cause the multiple isolectin forms observed in their iso-electropherograms.

Materials and methods

Proteins

The four Viciae lectins, namely the pea (*Pisum sativum* L.), lentil (*Lens culinaris*) and sweet pea (*Lathyrus odoratus* L.) lectins and favin (from *Vicia faba* L.), as well as Con A were all prepared by affinity chromatography on Sephadex G75 as previously described [18]. Trypsin (sequencing grade) was obtained from Boehringer Mannheim Canada Ltd (Laval, Canada). The lentil isolectin forms were separated by ion-exchange chromatography on DEAE-cellulose [12] or by chromatofocusing, on a 25 cm × 0.9 m column of Polybuffer exchanger (Pharmacia Biotech Inc, Montreal) equilibrated with 25 mM ethanolamine-HCl, pH 9.4, and eluted with 1:10 Polybuffer 96 adjusted to pH 7.0. The products were recovered from the Polybuffer by affinity chromatography as above. Iso-electric focusing was carried out with pre-cast pH 3–9 gels on a PhastGel instrument (Pharmacia Biotech Inc, Montreal) standardized with a wide-range reference protein mixture.

Mass spectrometry

Mass spectra were acquired using a SCIEX (Thornhill, Ont.,Canada) API III triple quadrupole mass spectrometer equipped with an atmospheric pressure ionization (API) source and an IonSpray® interface. Data acquisition and processing were achieved using a Quadra 950 computer and the computer program MacBioSpec (Sciex, Thornhill, Ontario) was used to support interpretation of the mass spectra. Electrospray mass spectra of proteins were obtained by injecting 2 μl of 1 mg ml^{-1} solutions into a stream of solvent (50% aqueous acetonitrile, 0.1% TFA) introduced to the mass spectrometer at a flow rate of 15 μl min^{-1}. The mass range was calibrated by injecting a solution of horse heart myoglobin (M_r:16951). Mass spectra of the lentil lectin samples were also obtained with a VG Quattro electrospray mass spectrometer and analysed with MassLynx software (Fisons Instruments, Altrincham, England).

Protein sequencing

The α-chains of the lentil lectin were separated from the β-chain by gel-filtration on Sephadex G100 in 2 M acetic acid. The product was dialysed and freeze-dried. Treatment of an aliquot of the α-chain (0.4 mg in 1 ml 0.1 M $NaHCO_3$) with succinic anhydride (1.5 mg) blocked the N-terminal residue and modified Lys 13. Trypsin treatment of the dialysed and freeze-dried product therefore caused cleavage at only Arg 21. This cleaved product mixture and the native α-chain were then sequenced by N-terminal automated amino-acid sequencing using an Applied Biosystems model 475A sequencing system (Foster City, CA) with on-line identification of the PTH derivatives.

Results

Viciae lectin mass spectra

The four two-chain lectins examined, the pea, lentil and *L. odoratus* lectins and favin, were from different genera in the Viciae tribe of the Leguminoseae. One of them, favin, bears a Fuc/Xyl/Man glycopeptide on its β-chain [19, 20] and there is similar but partial modification of the lentil lectin β-chain [19].

The molecular mass data are presented in Figs 1–5, and in Table 1. The Viciae lectins all showed β-chains of uniform polypeptide length. The pea lectin and favin β-chain molecular masses were in good agreement with the reported sequences and the latter's known pair of glycoforms [19], with the pea chain being 181 residues long in accordance with recent peptide work [11]. The molecular mass of the lentil lectin β-chain was in agreement with the recently reported revised 182-residue sequence [15], rather than the previously proposed 159 residue sequence [21]. The earlier sequence had been questioned [10]. Partial glycosylation of the β-chain of the lentil lectin has been reported [19], but no species of the appropriate mass was observed. The β-chain of the fourth lectin, *L. odoratus* lectin, has not been sequenced, but its smaller molecular mass (Table 1) suggests that

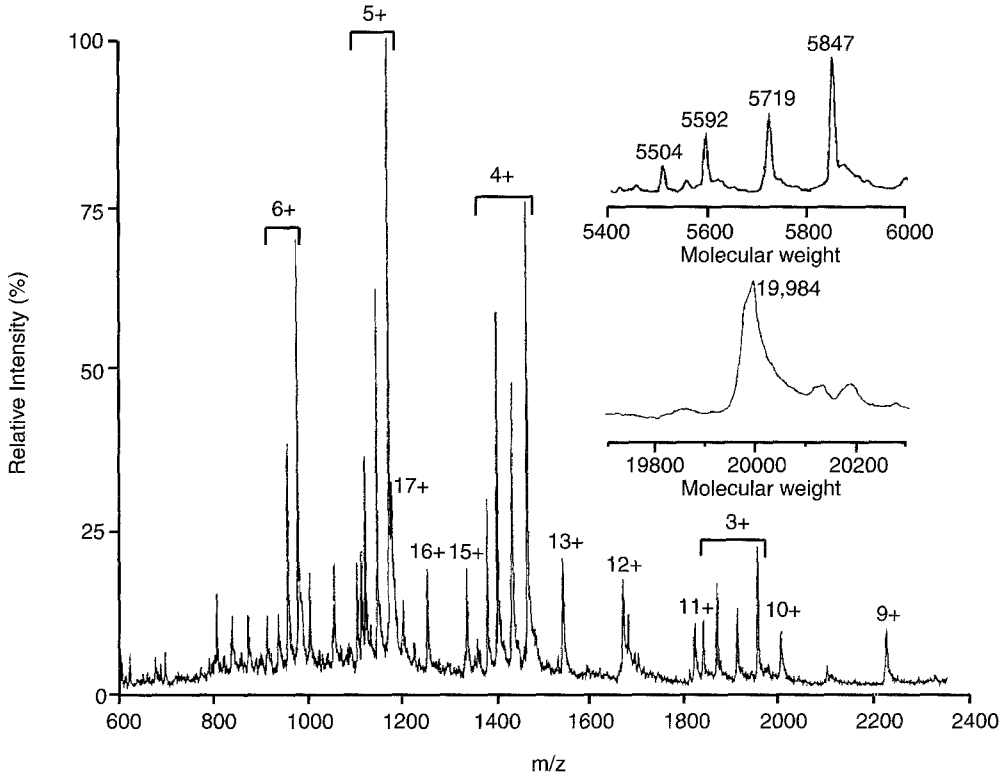

Figure 1. Mass spectrum of pea lectin, with its deconvolution (upper inset α-chains, and lower inset β-chain). The ion series 9+ to 17+ is from the β-chain and the series 3+ to 6+ is from the α-chains.

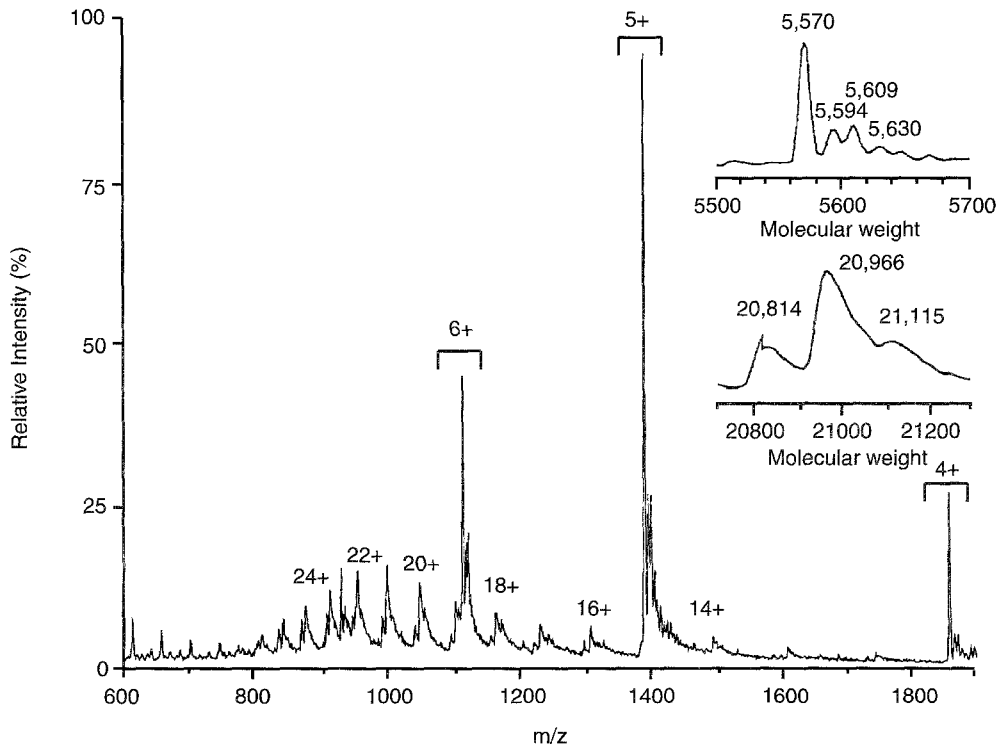

Figure 2. Mass spectrum of favin, with its deconvolution (upper inset α-chain, and lower inset β-chain). The ion series 13+ to 25+ is from the β-chain and the series 4+ to 6+ is from the α-chain.

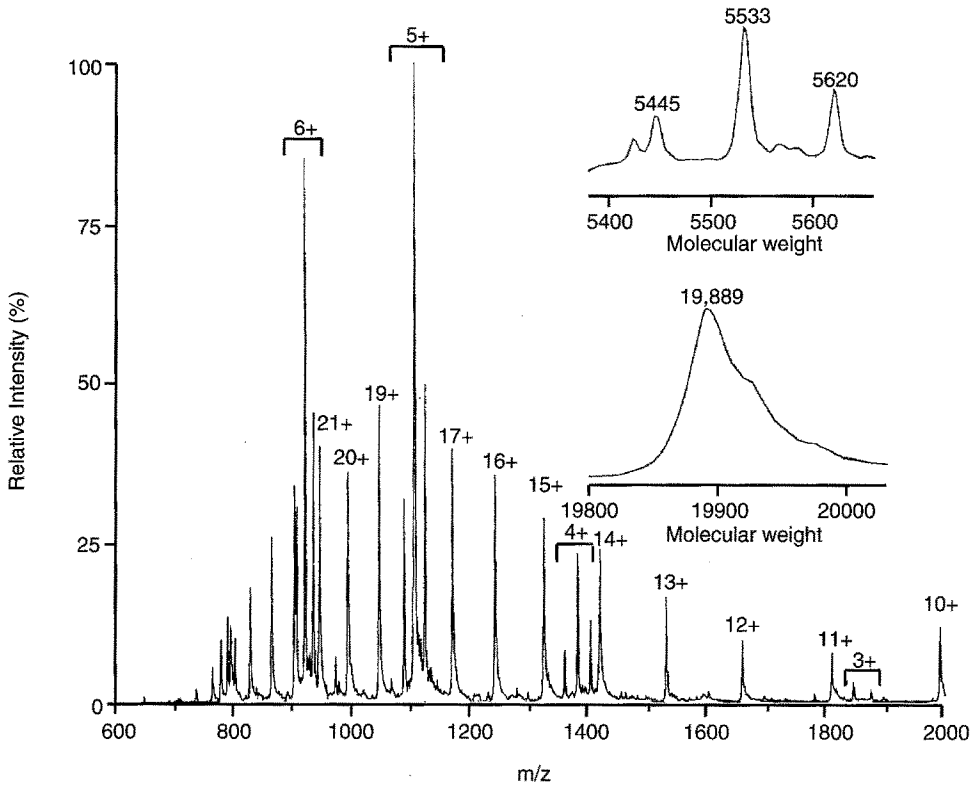

Figure 3. Mass spectrum of lentil lectin, with its deconvolution (upper inset α-chains, and lower inset β-chain). The ion series 10+ to 21+ is from the β-chain and the series 3+ to 6+ is from the α-chains.

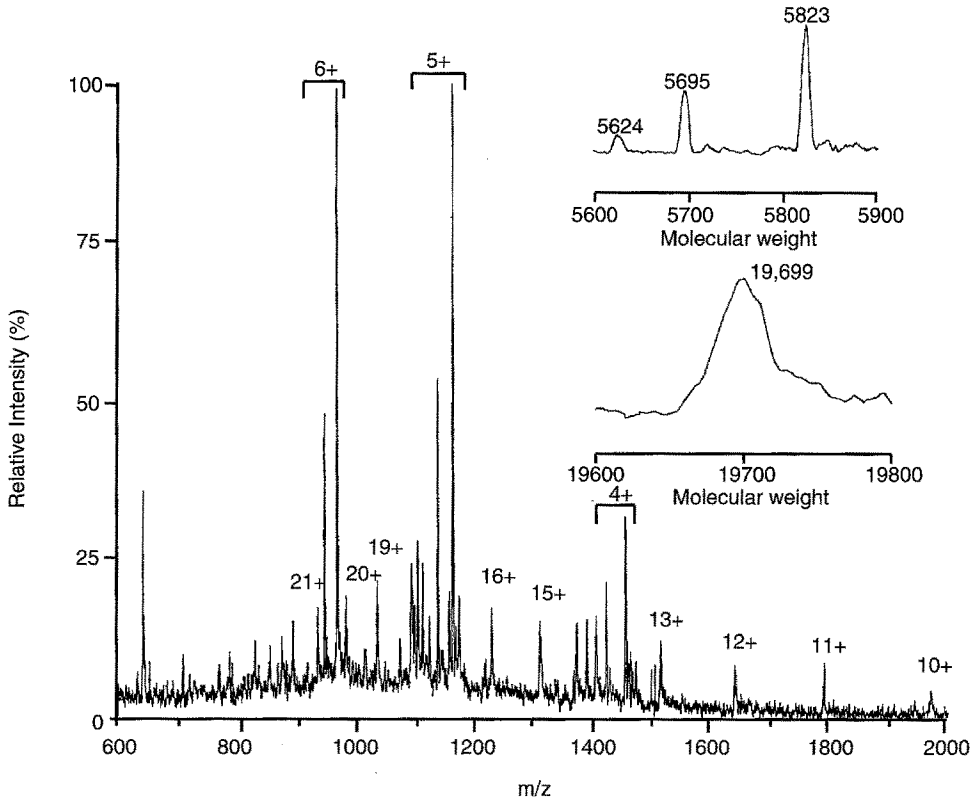

Figure 4. Mass spectrum of *L. odoratus* lectin, with its deconvolution (upper inset, α-chains and lower inset, β-chain). The ion series 10+ to 21+ is from the β-chain and the series 4+ to 6+ is from the α-chains.

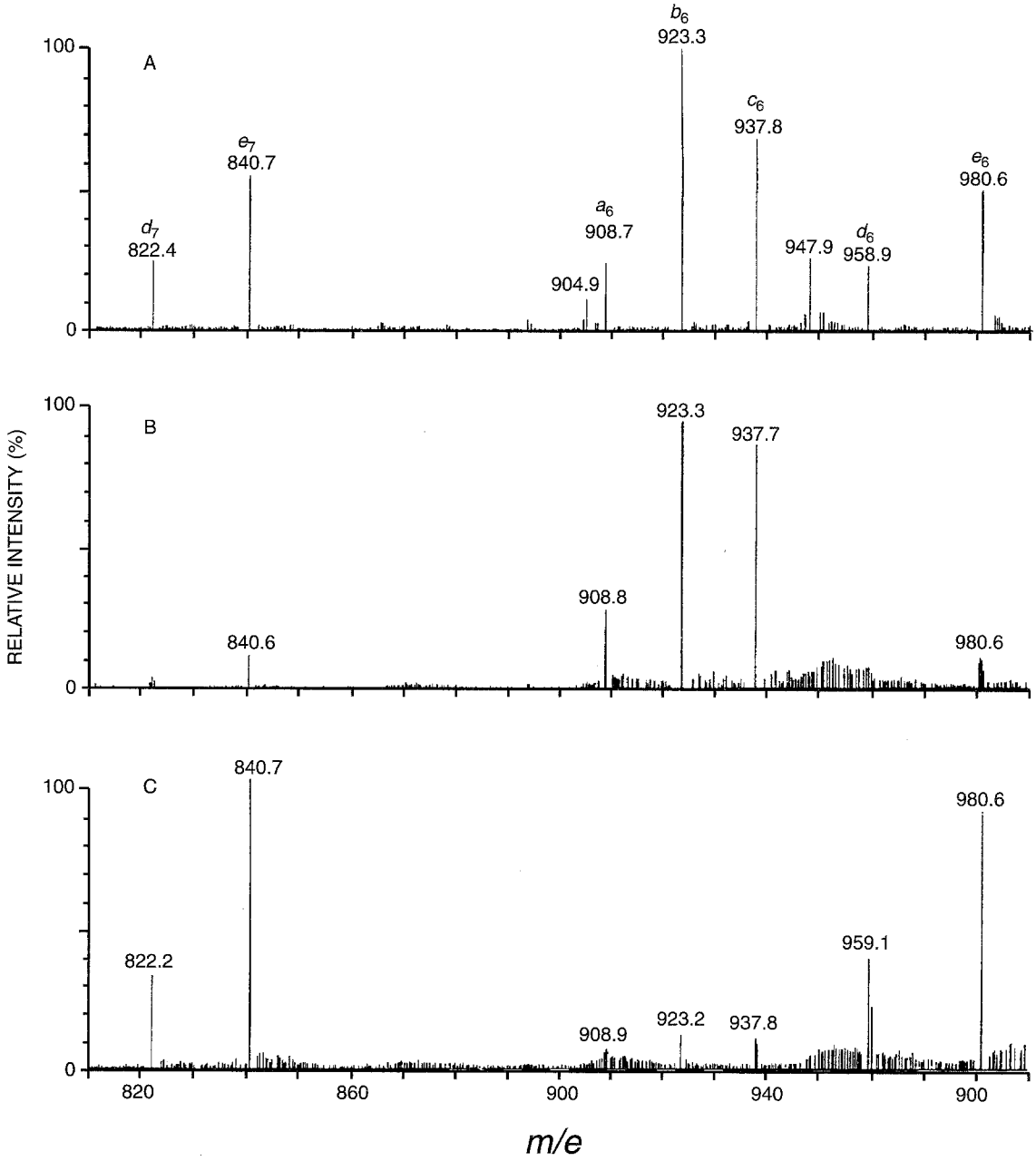

Figure 5. Mass spectra of lentil isolectins in α-chain region. (A) Lentil isolectin mixture; (B) lentil isolectin A; (C) lentil isolectin B. The ions a–e correspond to α-chain species 1–50 to 1–54 (Table 1) and the subscripts indicate the 6^+ and 7^+ charge states.

this chain is one or two residues shorter than the other β-chains.

In contrast to the uniformity of the β-chain species, three of the α-chains, pea, lentil and *L. odoratus*, were found to be mixtures arising from ragged C-terminal processing. The α-chain of favin appeared to be homogeneous and it is noteworthy that, unlike the others, it ends at an Asn residue (see Discussion below). The molecular mass data for the α-chains of the pea and the *L. odoratus* lectins and favin were in accord with the reported sequences [7, 20, 22] but the lentil lectin α-chain molecular mass differed by 166 Da from the most recent (52 residue) sequence [15]. The mass differences between the various species of lentil α-chain also could not be matched with the loss of any of the reported C-terminal residues [15].

Characterization of the lentil lectin α-chains

Because of the lack of agreement between its mass and sequence data, this chain was re-sequenced by N-terminal amino-acid sequencing. The sequence was obtained from

Table 1. Assignments and masses of two-chain lectins.

Lectin	*Assignment*	*Measured mass*	*Calculated mass*[a]	*Glycoform*
Pea lectin	α1–51	5504	5505.1	
	α1–52	5592	5592.1	
	α1–53	5719	5720.3	
	α1–54	5847	5848.5	
	β1–181	19984	19985.3	
Favin	α1–51	5570	5570.3	
	β1–182	20966	20965.3	$Man_2GlcNAc_2FucXyl$
	β1–182	20814	20819.2	$Man_2GlcNAc_2Xyl$
Lentil lectin	α1–50	5445	5448.0	
	α1–51	5533	5535.1	
	α1–52	5620	5622.2	
	α1–53	5748	5750.4	
	α1–54	5877	5878.5	
	β1–181	19889	19891.1	
L. odoratus lectin	α1–53	5695	5695.4	
	α1–54	5823	5823.5	
	β	19699	ND	
Con A	1–118	12938	12937.5	
	119–237	12680	12678.1	
	1–237	25604	25597.6	

[a]The sequence data used for the mass calculations were from cDNA for pea lectin [7], lentil lectin β-chain [15] and Con A [6], and from protein sequencing for favin [20] and the lentil (this paper) and *L. odoratus* lectin α-chains [22]. The glycan structures are from reference 19.
ND, sequence not determined.

runs on the intact α-chain and on a tryptic digest of succinylated α-chain, which gave a sequence beginning after Arg 21. This permitted the sequence to be determined up to residue 51. The next two residues were deduced from the mass spectra (Figs 3 and 5). The mass difference between the 1–52 and 1–53 species was 128 Da (Table 1) as was the difference between the 1–53 and 1–54 species. This mass corresponds to either Lys or Gln. Since no 8^+ ion was seen for 1–54 while both 1–53 and 1–54 gave 7^+ ions, residue 53 was assigned as Lys and 54 as Gln, which is the same as in the pea α-chain. The sequence is shown in Fig. 6 along with that previously reported and the pea lectin α-chain sequence [17]. The sequence changes found significantly increase the lentil α-chain's homology to the pea lectin and to the other α-chains.

Samples of the two isolectin forms were obtained by either ion-exchange [12] or chromatofocusing. The isolated lentil isolectins gave single bands on IEF (Fig. 7) but an intermediate composite band appeared in the unfractionated isolectin mixture. It has previously been shown [23] that the appearance of this mixed species of dimer is due to ampholyte-catalysed subunit exchange. Their mass spectra (Fig. 5) showed that the more acidic form had only α-chains that terminated before Lys53, while the more basic form had only α-chains with Lys 53.

```
                1         10         20
Lentil (a) VTSYTLNEVVPLKDVVPEWVRIGFSA
Lentil (b) ----------------------------
Pea        ------SD--S-----------------

                30         40         50
Lentil (a) TTGAEFAAHEVHSWSFHSELGGTSSSKQ
Lentil (b) --------Q-------N-Q--H--K-
Pea        -----Y-----L--------S-------
```

Figure 6. Sequence of lentil lectin α-chain determined in this paper (a), the previously reported sequence [15] (b), and the pea lectin α-chain sequence [17]. The two N-terminal sequencing runs of native α-chain and succinylated and trypsin-digested α-chain spanned residues 1–23 and 21–52 respectively and the final two residues were obtained by mass spectrometry.

Con A mass spectrum

Con A was also examined because of the similarities in its processing, i.e. excision of an interior peptide. Con A showed one single-chain form (Fig. 8; Table 1), together with the expected two half-molecules. No C-terminal heterogeneity was detected in any of these species. Contrary to a previous report [24], we were able to detect all three species and assign them in accordance with the gene sequence [6].

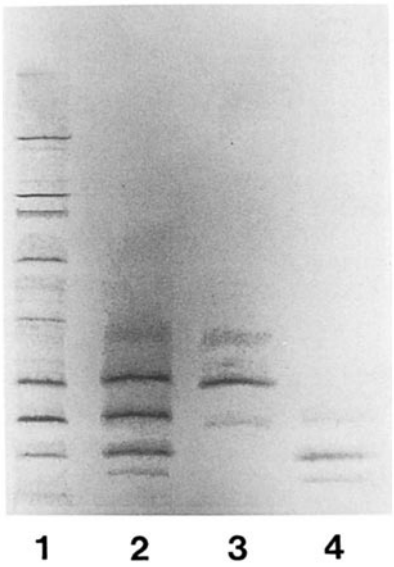

Figure 7. Iso-electric focusing of lentil isolectins. Lane 1, reference protein mixture, which includes three lentil isolectins with pIs of 8.15, 8.45 and 8.65; Lane 2, unfractionated lentil lectin; Lane 3, lentil isolectin A; Lane 4, lentil isolectin B.

Discussion

With one exception, the molecular mass data were in excellent agreement with the reported sequence data, allowing assignment of the various post-translational cleavage sites (Fig. 9). There was a large discrepancy between the molecular mass of the lentil lectin α-chain measured here and that calculated from the previously reported sequence, which was derived from cDNA and X-ray work [15]. This chain was therefore re-sequenced, and the new sequence data bring it into greater homology with the α-chains of the other Viciae lectins. The two lentil isolectin forms [12] were found to arise from species with or without the α-chain charged residue, Lys53, in the manner of the pea isolectins [11].

The agreement between the mass and sequence data strongly suggests that all four Viciae lectins are products of a single proprotein and hence probably a single gene in each species. The differences among isolectin forms arise from different degrees of post-translational processing. The *L. odoratus* lectin therefore resembles the *L. aphaca* and *L. articulatus* lectins rather than the

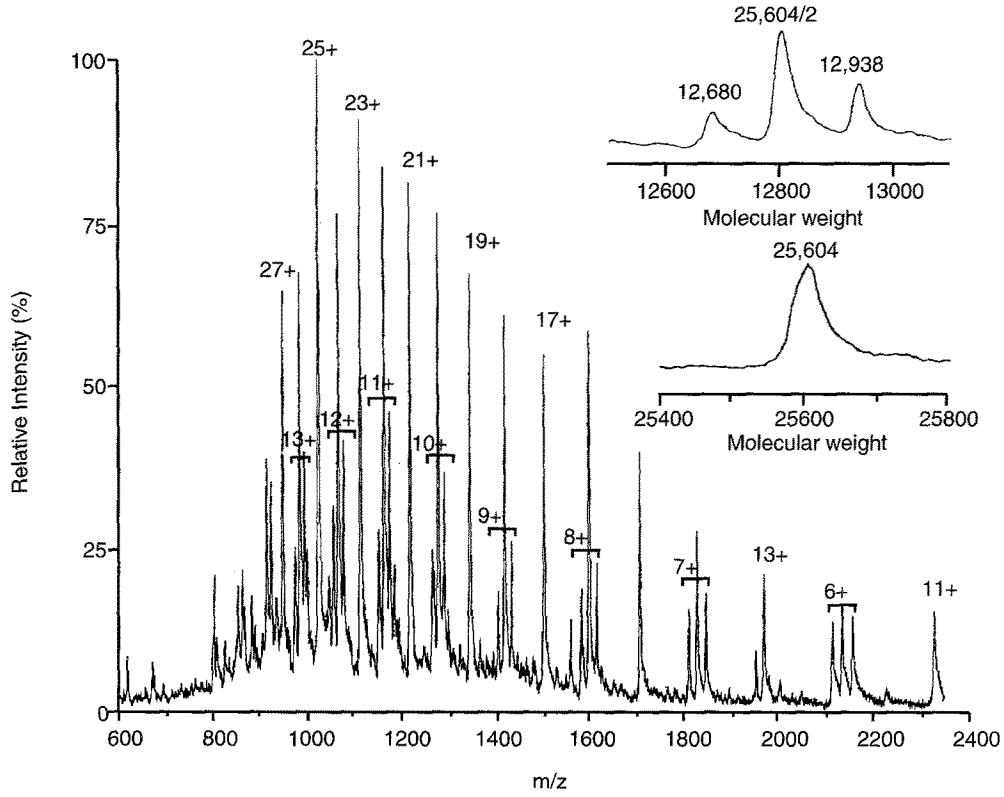

Figure 8. Mass spectrum of Con A, with its deconvolution (upper inset, the half-molecule fragments with a deconvolution artifact from the single-chain form in the centre, and lower inset, the single-chain form). The ion series 11+ to 27+ is from the single-chain form and the 6+ to 10+ is from the half-molecule fragments.

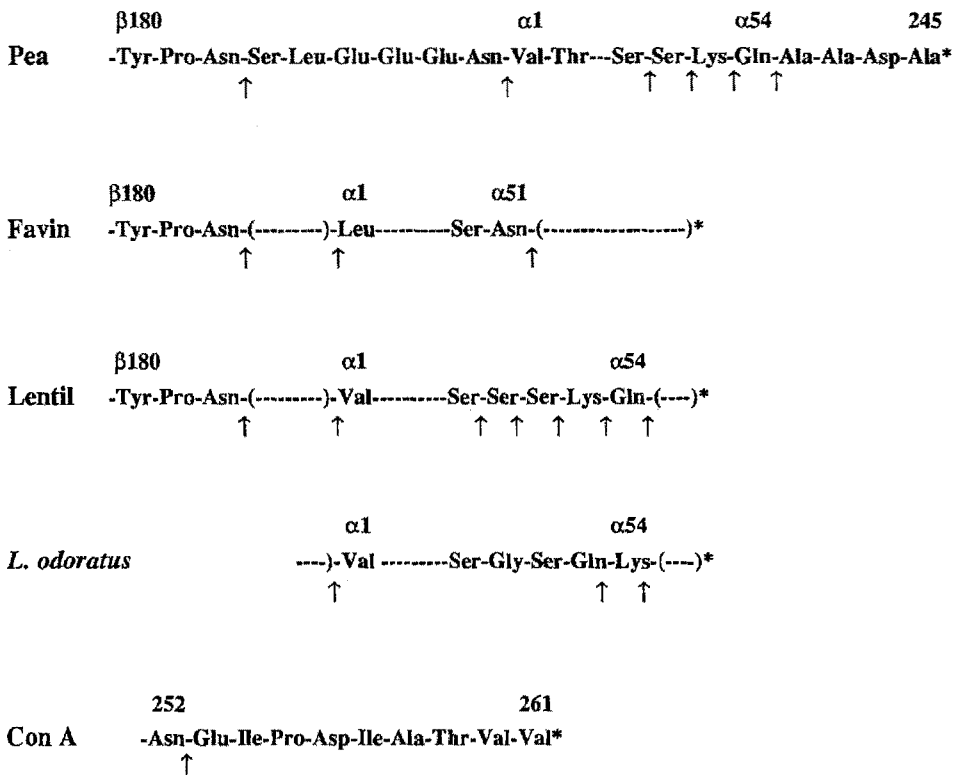

Figure 9. Post-translational cleavage sites in the lectin proproteins deduced from the mass spectra. The sequence data are from the cDNA results for the proproteins of pea lectin [7] and Con A [6], and the lentil lectin β-chain [15] and from protein sequencing for the lentil (this paper) and *L. odoratus* [22] α-chains, and favin [20]. Asterisks indicate the C-termini of the proprotein chains, and parentheses indicate segments of unknown sequence. Only the C-terminal processing site of con A is shown.

L. ochrus and *L. cicera* lectins, whose isolectins are genetic in origin [13]. The mass spectra showed that considerably more isolectin forms occur in the pea, lentil and *L. odoratus* lectins than their IEF patterns suggest. This is due to the ability of mass spectrometry to detect the loss of uncharged C-terminal amino acid residues from the chains. This method, being a direct measurement of the whole protein, is also considerably easier to use than the peptide mapping approach used previously with the pea lectin [11]. The families of α-chains detected, formed by sequential loss of C-terminal residues, resemble the families of truncated peptides created in the N-terminal sequencing method of Chait *et al.* [25].

Two types of proteolytic enzyme appear to be responsible for the various post-translational cleavages found, Asn-specific endoproteases similar to the *Canavalia ensiformis* enzyme [26] and carboxypeptidases. The processing of Con A, including its unusual polypeptide ligation [5, 6], is attributed to this Asn-specific protease, which has been purified [26]. By cleavage at two Asn residues, a central glycopeptide is excised. Cleavage at a third Asn site removes a nonapeptide from the C-terminus of the proprotein [6] and causes the incomplete ligation. In the mass spectrum of Con A, no heterogeneity at this point was detected, i.e. the fragment corresponding to residues 1–118 in the ligated form appeared homogeneous. The ligated form and the other fragment were also homogeneous. Therefore there appears to be no involvement of carboxypeptidases in Con A processing, unlike the case of the single-chain lectins [2, 3, 8].

The C-terminal residues of the β-chains of the pea and lentil lectins [7, 15] and of favin [20] are also Asn, as are the terminal residues of the excised internal pea peptide [7] and the favin α-chain [20]. The three lectins whose α-chains had ragged ends, however, do not have Asn residues in the C-terminal peptide segment. Hence two proteases may be active in the seeds of this group, an Asn-specific one and a carboxypeptidase which in the pea lectin removes four or more residues from the α-chain end of the precursor. But in both these lectins and Con A, chains that terminate at Asn due to the action of an Asn-specific protease do not undergo any further loss of residues by the action of carboxypeptidase. This suggests that the seed carboxypeptidases in these plants are incapable of cleaving off the terminal Asn residues.

This is in contrast to the behaviour of the single-chain lectins. It has recently been shown that lectins such as

the peanut and soybean agglutinins undergo extensive 'ragged' proteolytic trimming at their C-terminal ends [2, 8]. The *Erythrina corallodendron*, *Phaseolus vulgaris* and *Dolichos biflorus* lectins also are trimmed in this way [3]. The cleavages appear to be effected by both Asn-specific proteases and carboxypeptidases, and since chains terminating at Asn were not observed, these carboxypeptidases do not stop at Asn residues. It is also evident that the C-terminal proteolytic processing of the Viciae lectins does not have the structural significance that it has in the case of the soybean agglutinin, where it is essential for the formation of the tetramer structure [27], or in the *D. biflorus* lectin, where it is related to the number of active binding-sites [28, 29].

It is noteworthy that the cDNA sequences for Con A and the pea lectin show that their proproteins have shorter C-terminal extensions than those of the single-chain lectins. Con A has nine extra residues and the pea lectin has four, compared to their mature forms (Fig. 9), whereas the single-chain lectins can have more than a dozen extra residues [2, 3]. It is important to consider this difference when looking for any biological significance in the C-terminal processing. If these regions have functional roles, such as targeting of the proprotein to organelles in the manner of the barley lectin [30], then they are not necessarily the same in the two types of legume lectin.

Acknowledgements

We thank Rebecca To for excellent technical assistance, Don Krajcarski for help with the VG mass spectrometer, and Dr Makoto Yaguchi for facilitating this work.

References

1. Mann M, Meng CK, Fenn JB (1989) *Anal Chem* **61**: 1702–8.
2. Thibault P, Watson DC, Yaguchi M, Young NM (1993) In: *Techniques in Protein Chemistry* vol IV (Angeletti RH, ed.) pp. 91–98. San Diego, CA: Academic Press.
3. Young NM, Watson DC, Yaguchi M, Adar R, Arango R, Rodriguez-Arango E, Sharon N, Blay, PKS, Thibault P (1995) *J Biol Chem* **270**: 2563–70.
4. Young NM, Watson DC, Thibault P (1995) *Glycoconj J* **12**: 135–41.
5. Bowles DJ, Marcus SE, Pappin DJC, Findlay JBC, Eliopoulos E, Maycox PR, and Burgess J (1986) *J Cell Biol* **102**: 1284–97.
6. Carrington DM, Auffret A, Hanke DE (1985) *Nature* **313**: 64–67.
7. Higgins TJV, Chandler PM, Zurawski G, Button SC, Spencer D (1983) *J Biol Chem* **258**: 9544–49.
8. Mandal DK, Nieves E, Bhattacharyya L, Orr GA, Roboz J, Yu Q, Brewer CF (1994) *Eur J Biochem* **221**: 547–53.
9. Rini JM, Hofmann T, Carver JP (1986) *Biochem Cell Biol* **65**: 338–44.
10. Jackson GED, Young NM (1987) *Analyt Biochem* **162**: 251–56.
11. Hoedemaeker FJ, Richardson M, Diaz CL, de Pater BS, Kijne JW (1994) *Plant Mol Biol* **24**: 75–81.
12. Howard IK, Sage HJ, Stein MD, Young NM, Leon MA, Dyckes DF (1971) *J Biol Chem* **246**: 1590–95.
13. Rougé P, Borrebaek CAK, Richardson M, Yarwood A (1987) *Glycoconj J* **4**: 371–78.
14. Bourne Y, Nésa M-P, Rougé P, Mazurier J, Legrand D, Spik G, Montreuil J, Cambillau C (1992) *J Mol Biol* **227**: 938–41.
15. Loris R, Steyaert J, Maes D, Lisgarten J, Pickersgill R, Wyns L (1993) *Biochemistry* **32**: 8772–81; correction, 14229.
16. Reeke GN, Becker JW (1986) *Science* **234**: 1108–11.
17. Rini JM, Hardman KD, Einspahr H, Suddath FL, Carver JP (1993) *J Biol Chem* **268**: 10126–32.
18. Young NM, Williams RE, Roy C, Yaguchi M (1982) *Can J Biochem* **60**: 933–41.
19. Ashford DA, Dwek RA, Rademacher TW, Lis H, Sharon N (1991) *Carbohydr Res* **213**: 215–27.
20. Hopp TP, Hemperly JJ, Cunningham BA (1982) *J Biol Chem* **257**: 4473–83.
21. Foriers A, Lebrun E, Van Rapenbusch R, de Neve R, Strosberg AD (1981) *J Biol Chem* **256**: 5550–60.
22. Sletten K, Kolberg J, Michaelsen TE (1983) *FEBS Lett* **156**: 253–56.
23. Bhattacharyya L, Brewer CF (1990) *J Chromatog* **502**: 131–42.
24. Loo JA, Loo RRO, Light KJ, Edmonds CG, Smith RD (1992) *Analyt Chem* **64**: 81–88.
25. Chait BT, Wang R, Beavis RC, Kent SBH (1993) *Science* **262**: 89–92.
26. Abe Y, Shirane K, Yokosawa H, Matsushita H, Mitta M, Kato I, Ishii S-I (1993) *J Biol Chem* **268**: 3525–29.
27. Dessen A, Gupta D, Sabesan S, Brewer CF, Sacchettini JC (1995) *Biochemistry* **34**: 4933–42.
28. Etzler ME, Gupta S, Borrebaeck CAK (1981) *J Biol Chem* **256**: 2367–70.
29. Etzler ME (1994) *Biochemistry* **33**: 9778–83.
30. Chrispeels MJ, Raikhel NV (1992) *Cell* **68**: 613–16.