Glycoconjugate Journal (1995) **12**: 135–141

Mass spectrometric analysis of genetic and post-translational heterogeneity in the lectins Jacalin and *Maclura pomifera* agglutinin

N. MARTIN YOUNG*, DAVID C. WATSON[1] AND PIERRE THIBAULT[2]
[1]*Institute for Biological Sciences, National Research Council of Canada, Ottawa, Canada K1A OR6*
[2]*Institute for Marine Biosciences, National Research Council of Canada, Halifax, Canada B3H 3Z1*

Received 3 August 1994, revised 7 October 1994

Jacalin and *M. pomifera* agglutinin are T-antigen specific lectins with $\alpha_4\beta_4$ structures that show far greater microheterogeneity than plant lectins from other families, due to multiple genetic isoforms and post-translational processing. Electrospray mass spectrometry and combined liquid chromatography-electrospray mass spectrometry were used to characterize the various forms. For both lectins, the mass data were consistent with previous protein sequencing of the major α-chain species of 133 residues and three β-chain species of 20 or 21 residues. In addition, for jacalin the mass of one minor α-chain species was consistent with a second of the four reported gene sequences. However, the glycopeptide α-chain form and one β-chain form did not match any of the genes, suggesting a fifth gene remains to be found. For *M. pomifera* agglutinin, three more β-chain forms were found, but all six could arise from only two genes, with additional post-translational proteolysis and post-translational substitution with an unidentified component of 106 Da creating the set of six forms. Only two α-chain forms were found also, with no glycosylation.

Keywords: electrospray mass spectrometry, lectin, microheterogeneity, post-translational processing, sequencing

Introduction

Lectins from different plant families show a wide diversity of protein structural types, ranging from the β-sheet metalloproteins such as concanavalin A [1] to the small disulphide-rich domains of wheat germ agglutinin [2]. Another distinctive and highly unusual polypeptide structure has been found in two seed lectins from plants of the Moraceae family, specific for the T-antigen. Jacalin, the lectin from *Artocarpus integrifolia*, and *Maclura pomifera* agglutinin (MPA) have subunits composed of 133-residue α-chains non-covalently associated with β-chains that are only 20 or 21 residues long [3], to form tetramers with $\alpha_4\beta_4$ overall structures [4, 5]. These lectins also differ from lectins of legumes and other plant families in showing a high degree of microheterogeneity. Protein sequence work [4–6] showed that both types of chain occurred in multiple genetic forms, although for the α-chains there was one dominant form, and the amounts of the three β-chain forms varied. In addition to the genetic variation, isoforms due to two types of post-translational modification were also found. Some of the jacalin α-chains were glycosylated, while one of the three MPA β-chains had one less residue at the N-terminus, presumably due to post-translational proteolysis. The jacalin glycopeptide has a $GlcNAc_2Man_3Fuc_1Xyl_1$ structure [7], which also occurs in other plant lectins [8].

Recent cloning work on jacalin [9] has shown that the α- and β-chains arise from a common precursor protein. This is comprised of five segments, namely a 21-residue leader sequence, followed by a 39-residue propeptide, the β-chain, a four residue linker and finally the α-chain. Four clones were isolated with one to eight sequence differences among their α-chain segments. It is evident from the variation in the amounts of the β-chain isomers [3] and the dominance of one α-chain type that these genetic variants are not equally expressed.

Jacalin and MPA therefore occur as heterogeneous populations of tetramers, unlike the better known plant lectins such as concanavalin A. The widely varying levels of the various polypeptide species make their complete characterization very difficult. Nor has it been possible to purify to homogeneity any of the individual α-chain forms [5], although the β-chains can be separated by reverse-phase HPLC [3]. We have therefore used electrospray mass spectrometry [10] to investigate the multiple forms of both lectins more completely, including examination of the various β-chains by combined liquid chromatography-mass spectrometry. The electrospray technique is capable of measuring protein molecular masses to 0.01–0.02% and can resolve components in protein mixtures in the size

* To whom correspondence should be addressed.

range of the α-chains, ~14 600 Da, that differ by as little as 15 Da. When the mass results are compared with the reported sequences, they suggest that there are further genes for jacalin beyond the four described so far. Additionally, three more β-chain forms were found in MPA, arising from post-translational processing.

Materials and methods

Proteins Jacalin was obtained from Pierce Chemical Co. (Rockford IL) and MPA was purified as previously reported [3]. Samples were extensively dialysed against 5% acetic acid, or passed through a Sephadex G25 column in the same solvent, to remove metal ions. The concentrations used were 2–2.5 mg ml^{-1}. Distilled and deionized (18 MΩ) water (Millipore Q water system, Millipore Inc.) was used in the preparation of the samples and mobile phases.

Mass spectrometry All mass spectra were acquired on a SCIEX (Thornhill, Ontario, 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 Macintosh Quadra 950 computer. Ionspray mass spectra of proteins were obtained by injecting 0.5 μl of 5 mg ml^{-1} solutions into a stream of solvent (50% acetonitrile, 0.1% TFA) introduced to the mass spectrometer at a flow rate of 10 μl min^{-1}. The mass range was calibrated by injecting a solution of horse heart myoglobin (Mr: 16951).

For LC-MS analyses, an HP1090L liquid chromatograph (Hewlett Packard), equipped with a tertiary DR5 solvent delivery system and a 2.1 mm I.D. × 25 cm Vydac 214TP52 column (Vydac Separation Group, Hisperia, CA), was coupled directly to the mass spectrometer via the IonSpray® interface. Injections of 20 μl of the 2 mg ml^{-1} protein solutions were made and the effluent from the column was split such that a flow rate of approximately 15 μl min^{-1} was introduced to the mass spectrometer. The voltage on the spray needle was maintained at 5 kV and high purity air was used as the nebulizing gas at an operating pressure of 50 psi. The LC-MS analyses were performed in full-scan mode (*m/z* 600–1500), using dwell times of 4 ms per Da.

Fragment ion spectra (MS-MS), of precursors selected via *m/z* values in the first quadrupole, were obtained using collisional activation by argon in the second (RF-only) quadrupole. A retrofit RF-only quadrupole cell (API/III+, SCIEX, Thornhill Ont.) was used with a gas thickness of 3.5×10^{15} atoms cm^{-2} and collision energies of 75 eV (laboratory frame of reference). The third quadrupole was scanned over the mass range of *m/z* 50–1500 in steps of 0.2 *m/z* units, with a dwell time of 2 ms per step.

Results

Jacalin Direct injection of both jacalin and MPA initially gave mass spectra that were dominated by the small 20–21 residue β-chains. Mass spectra of jacalin predominantly showing the α-chain peaks were obtained from samples

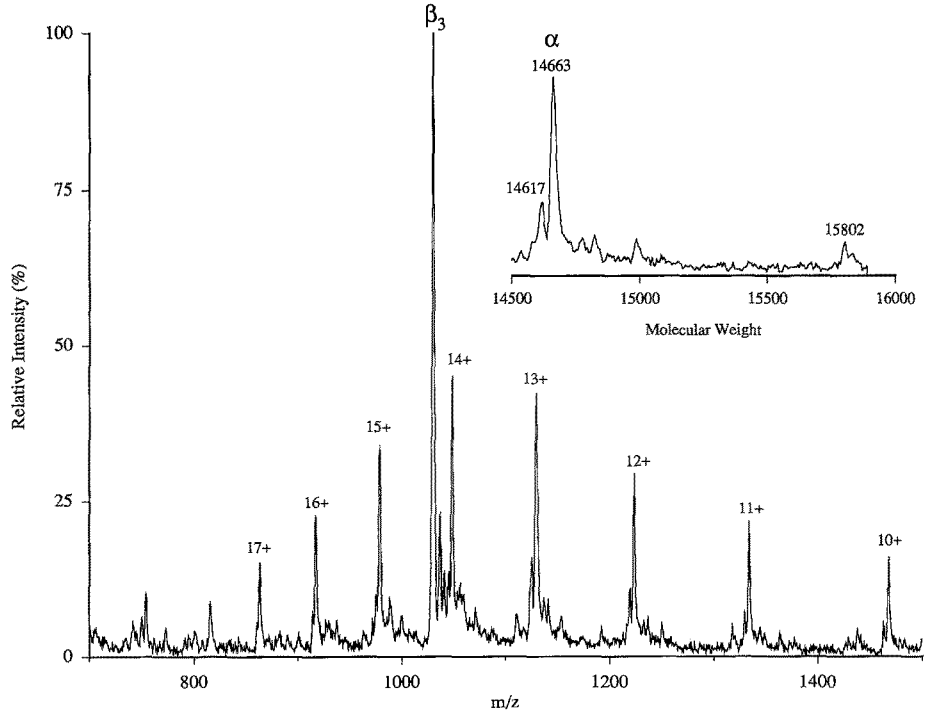

Figure 1. Electrospray mass spectrum of jacalin, with deconvolution of α-chain species (inset).

purged by centrifugal diafiltration (Amicon Centricon 10 units) with 15% acetic acid to remove the β-chains. In the mass spectrum of jacalin (Fig. 1) the β-chains occur mainly as the doubly-protonated forms (*m/z* 1029–1033), and are poorly resolved due to the proximity of their *m/z* values and the overlaying α-chain peaks. The deconvoluted spectrum (Fig. 1 inset) shows a predominant α-chain species at 14 663 ±2 Da, consistent with the reported sequence (Table 1), and a satellite species at 14 617 Da. The glycoprotein form gave a relatively weak peak compared with the intensity of the band seen in SDS-PAGE (data not shown). The mass of this species was 15 802 ±2 Da.

To characterize the jacalin components further, LC/MS experiments were carried out, using conditions previously found for the separation of the individual β-chains and recovery of the dominant α-chain species. The profile is shown in Fig. 2, along with the mass spectra of the individual components. The data for the individual β-chains agreed with the previously reported sequences (Table 1). An additional β-chain component with a molecular weight of 2260 co-eluted with β_1 at 21.6 min (Fig. 2C). This peptide is 189 higher than β_2 and could correspond to a processing variant of it with the first two residues of the linker segment, Thr-Ser [9] at its C-terminal end. The main α-chain peak, though more homogeneous, still showed the satellite species seen in the native protein spectrum.

Table 1. Assignments and masses of the two chains of the lectins.

Lectin	*Assignment*[a]	*Measured mass*	*Calculated mass*[b]	*Elution time (min)*
Jacalin	β_1	2061.7	2062.2[c]	18.9
	β_{2a}	2071.6	2072.2	21.6
	β_{2b}	2260.0	2260.4	21.6
	β_3	2057.8	2058.2	23.5
	α	14 663	14 662.6[c]	43.5
	α	14 617	14 619[d]	
	α	15 802	(glycoprotein)	
MPA	β_{1a}	2196.0	2196.4	21.0
	β_{1b}	2082.0	2082.3	21.0
	β_{1c}	2302.0		21.8
	β_{2a}	2256.0	2255.5	20.4
	β_{2b}	2141.5	2141.4	20.4
	β_{2c}	2361.5		21.0
	α	14 742	14 741.7	
	α	14 657	14 656.7	

[a] See Fig. 4 for the assignment of the β-chain subspecies.
[b] From the sequence data of references 1, 2, 4 and 7; all masses correspond to isotope-average values.
[c] Clone pSKcJA15 [9].
[d] Clone pSKcJA17 [9].

M. pomifera agglutinin In the mass spectrum of MPA (Fig. 3) the α and β species were much better resolved, partly due to the higher ionization states, $[MH_3]^{3+}$ and $[MH_4]^{4+}$, given by the MPA β-chains. A total of six β-chain species were found (Fig. 3 left inset) and these are attributable to two processing variants of each of two genetic forms, plus two species that are ~106 Da higher in mass (Table 1 and Fig. 4). The latter species appear to be variants of the β_1 and β_2 gene forms with an additional component. The MS/MS spectra obtained from the triply protonated precursor ions of β_{2a}, β_{2b} and β_{2c} are presented in Fig. 5. The spectra are mostly dominated by fragment ions of the b- and y-type formed by cleavages of the peptide bonds, with protons strongly anchored on basic residues located near the N-termini [11]. Such fragmentations give rise to abundant b_n^{2+} and a_n^{2+} fragment ions as observed in the *m/z* 400–700 region (Fig. 5). Interestingly, the MS/MS spectrum of the $[MH_3]^{3+}$ ion from β_{2c} showed similar fragment ions to those from β_{2a}. Furthermore, abundant ions were observed for a fragment at *m/z* 107 and for the loss of a 106 Da moiety from the precursor ion. This suggests that one of the amino acids of the β_2 chain is modified with a labile group of mass 106 Da which is readily cleaved upon collisional activation. N-terminal sequencing of HPLC peaks containing the β_{1c} and β_{2c} forms (data not shown) gave only the sequences expected for β_1 and β_2 peptides. The mass of 106 Da does not correspond to any of the twenty amino acids nor to any commonly found substituent, and apparently represents a labile adduct on one of the amino-acids which we have not been able to identify. As with jacalin, in MPA there was one dominant α-chain species whose mass, 14 742 Da, was in agreement with the reported sequence (Table 1), and a minor species around 14 657 Da.

An LC/MS experiment with MPA (Fig. 6) allowed correlation of the six β-chain species with the various HPLC peaks, and indicated trace amounts of additional β-chain forms. The MPA α-chains are not recoverable under these HPLC conditions [3].

Discussion

In the *Moraceae* lectins, the two chains are now known to be derived from a pro-protein precursor [9] by post-translational proteolytic processes that resemble those found in other two-chain proteins such as the pea lectin [12] and concanavalin A [13]. It would be expected therefore that the number and relative amounts of α-chain genetic species would be similar to the number of β-chain species. This was, however, not immediately apparent in the previous protein sequencing nor in the present mass spectrometry studies, since there was one dominant α-chain species in both jacalin and MPA, in contrast to the three jacalin and two MPA β-chain genes required to account for the various β species that were found (Fig. 4).

The jacalin β-chain species are not present in equal amounts, the ratio of the three species being approx. 1:2:4. On SDS-PAGE, the apparent ratio of the main jacalin α-chain species to the next most abundant, the glycoprotein form, is

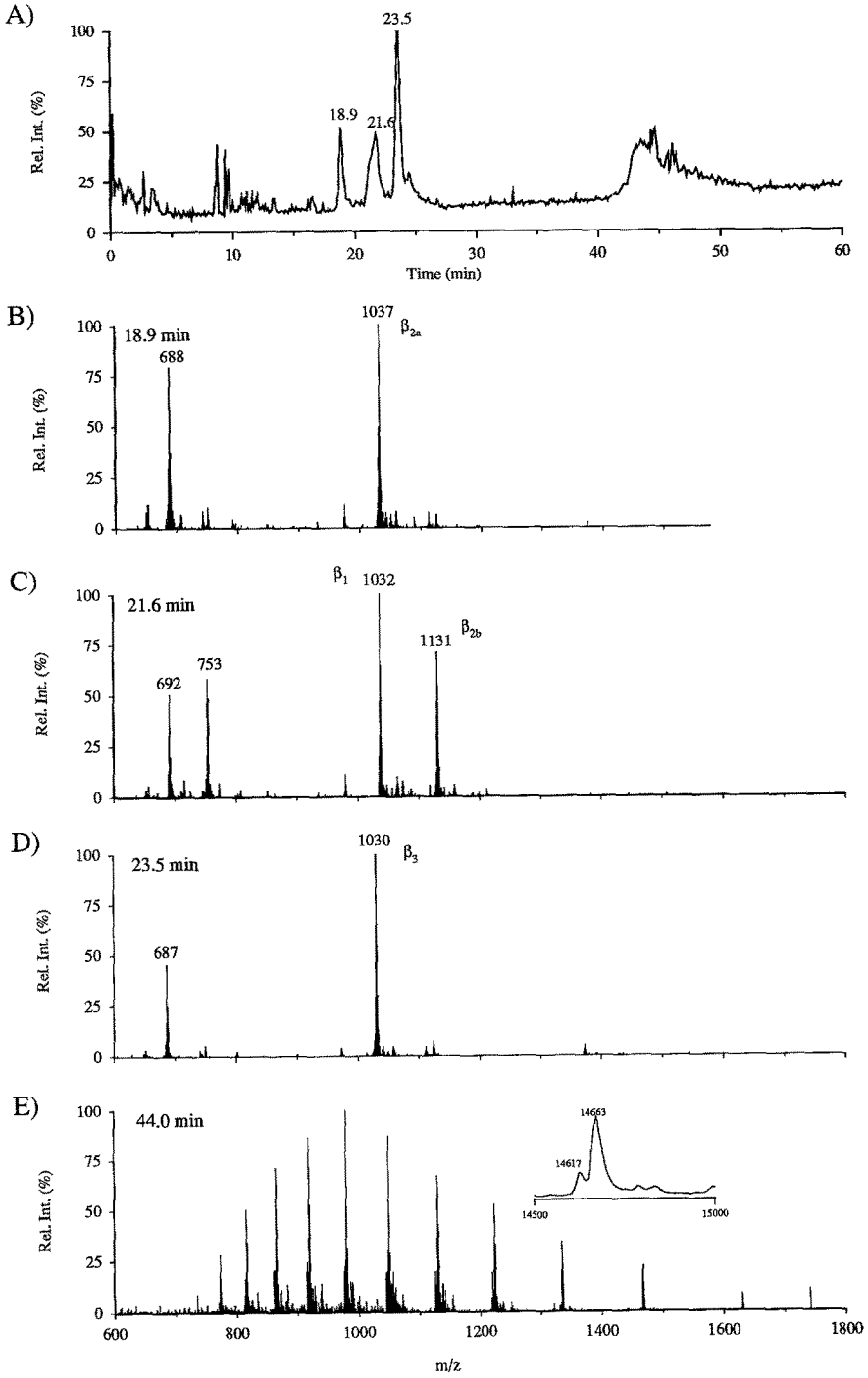

Figure 2. LC/MS of jacalin (A) and mass spectra of the three β chain peaks (B–D) and the α-chain peak (E).

approximately 3:1. A third species of mass 14 617 Da was present in a smaller amount in the mass spectrum. Thus each β-chain species can be roughly matched in amount with an α-chain one, and these data are therefore consistent with α/β proprotein precursors for the jacalin isoforms. However, it is clear that the four gene products indicated by the cloning [9] are not present in equal amounts. The most abundant β-chain form, β_3, and the main α-chain have sequences and masses that are consistent with the sequence of clone pSKcJA17, which must therefore represent the dominant proprotein. The minor α-chain form of mass 14 617 and the β-chain present in smallest amount, β_1, are assigned to the clone pSKcJA15. A third clone, pSKcJA3, gives a β-chain sequence identical to the β_3 one except for the first residue which is Asn rather than Asp. A small amount of this might be overlooked in the background present in the first sequencer cycle for the β_3. Its α-chain sequence also has only one difference, Val to Ile at residue 98, and the mass difference of 14 Da would make this

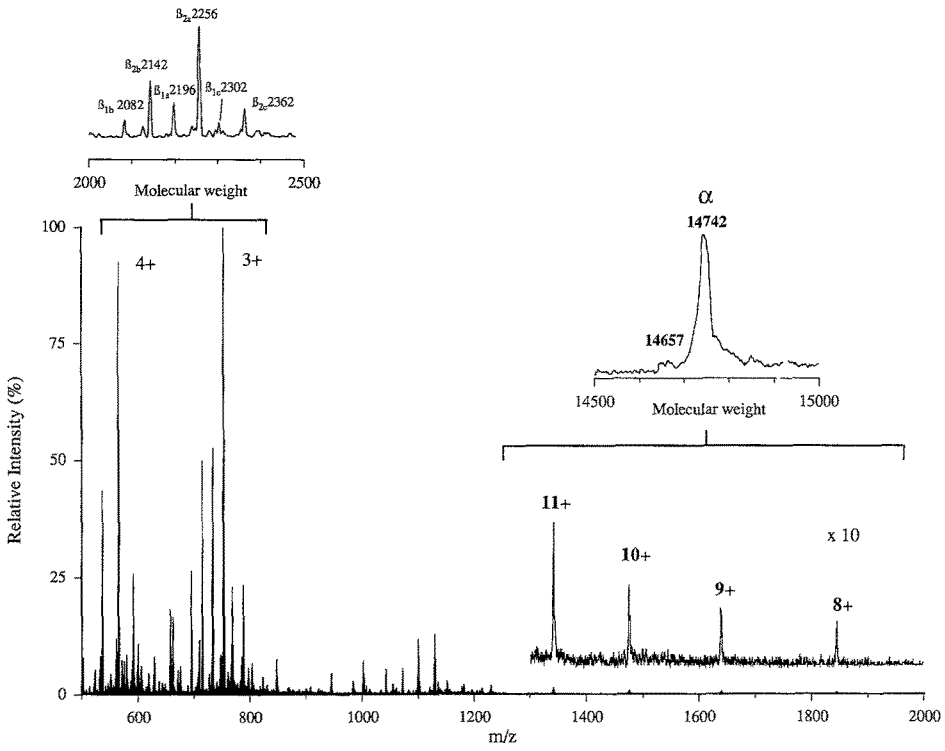

Figure 3. Electrospray mass spectrum of MPA, with deconvolution of the β chains (left top) and the α-chains (right inset).

```
MPA
          1                    21
        -NGPNGKSQSIIVGPWGDRVTN-
β1a     ↑                    ↑
β1b       ↑                  ↑
β1c     ↑                    ↑ +106 Da

        -NGRNGKSQSIIVGPWGDRVTN-
β2a     ↑                    ↑
β2b       ↑                  ↑
β2c     ↑                    ↑ +106 Da
```

```
JACALIN
                                        Clone:-
          1                  20
        -NNEQSGKSQTVIVGSWGAKVSTSSN-     pSKcJA15
β1      ↑                    ↑

        -NEQSGKSQTVIVGPWGAQVSTS-            -
β2a     ↑                    ↑
β2b     ↑                      ↑

        -NDEQSGISQTVIVGPWGAKVSTSSN-     pSKcJA17
β3      ↑                    ↑
```

Figure 4. Assignments of cleavage sites and 106 Da modification giving rise to the three sub-species from each of the two genetic forms of the MPA β-chain, and to the four jacalin β-chain species. The jacalin genes are also indicated [9], and the residues that precede and follow the β-chain in the proprotein sequences where known.

species difficult to resolve in the mass spectrum if it were present in small amounts. The evidence for expression of this third gene is therefore equivocal. The sequence of the fourth clone, pSKcJA1, cannot be correlated with any of the mass and protein sequencing data. It would give a β-chain with Glu at residue 2 and this was not seen in the sequencing or LC/MS. The α-chain has many more sequence differences from the main one than the other two, but its predicted mass at 14 659 is also within the limit of resolution from the main species. It appears from the mass and protein sequencing data that this protein form is not present in the jacalin sample.

The four clones do not provide a candidate sequence for the glycosylated form of jacalin, nor for β_2. Subtracting the mass for the known carbohydrate structure [7] of 1171 Da indicates the α-chain that carries it has a polypeptide mass of 14 631. The amount of the glycoform seen on SDS-PAGE is consistent with it being associated with β_2. The pSKcJA1 protein cannot form the glycoprotein as it has a Lys residue at position 74 in the α-chain, rather than the Asn that is *N*-glycosylated in the glycoform. Furthermore, pSKcJA1 is not likely to form the second minor glycoform found by Ruffet *et al.* [5]. While it does have the Lys to Thr replacement at residue 45 needed to form the Asn-X-Thr recognition sequence, it has as well, at residue 49, a Thr to Lys change which would have been detected by Ruffet *et al.* [5] due to the additional tryptic peptide it would give. It appears therefore that at least one additional gene for the main glycoform with β_2 is still to be found, and possibly a second one for the minor glycoform. From the

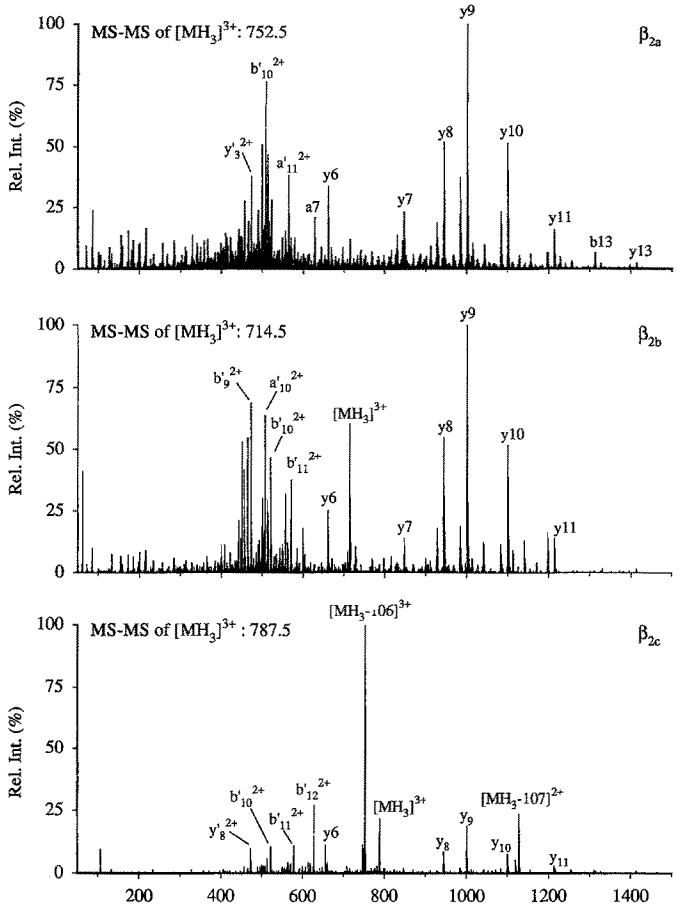

Figure 5. MS/MS analyses of the β_{2a} (A), β_{2b} (B) and β_{2c} (C) chains from MPA.

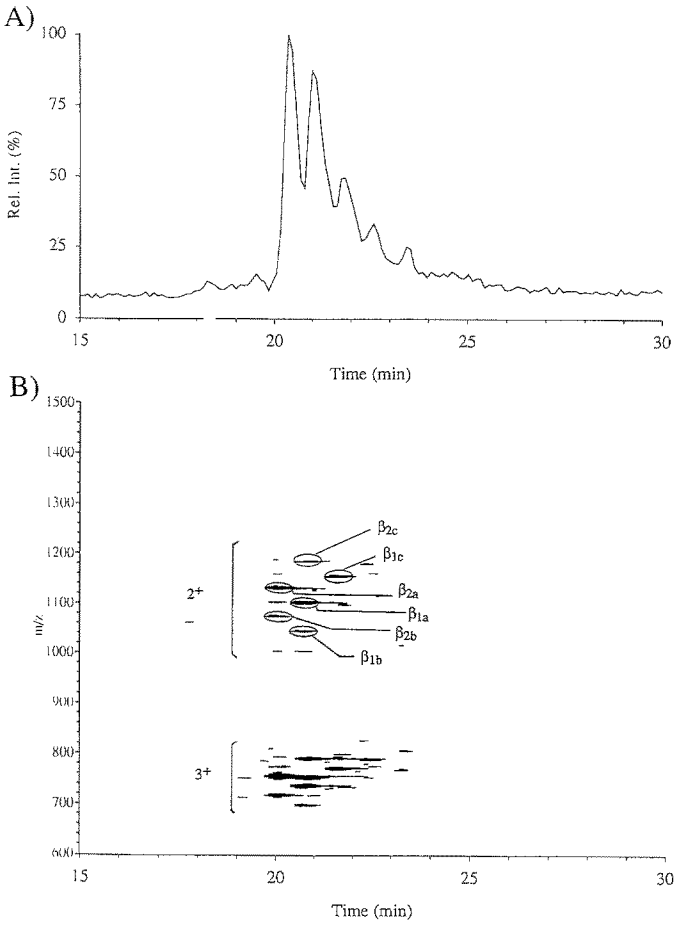

Figure 6. LC/MS of MPA (A) and contour plot of the β-chain region (B).

sequence data on the partially purified glycoform of jacalin [5], one combination of the major genetic variations found by both sequencing laboratories [5, 6] that would give the mass estimated for the main glycoform would be Ile113 → Val, Val31 → Ile and Met66 → Val. This represents a sequence that is intermediate between pSKcJA17 and pSKcJA1. The β_2 sequence is also intermediate between those of the two clones.

In addition to these various genetic forms, the mass spectrometric data also showed that jacalin has some further variation introduced during its post-translational proteolytic processing. A form of the β_2 chain with an additional two residues was found, and the mass difference corresponds to the first two of the four residues that connect the β and α chains in the proprotein, Thr-Ser-Ser-Asn [9]. This type of processing heterogeneity is even more common in MPA.

Previously three β-chain forms were found in MPA, but their peaks on HPLC were not well-shaped [3]. The mass spectrometry and LC/MS showed these components were mixtures and six species have now been identified. However, these are attributable to processing variants of the two genetic types previously found (Fig. 4), and no more genes are as yet indicated. Hence, MPA appears to have fewer genetic forms than jacalin, probably only two. In agreement with this, the IEF pattern of MPA is simpler than that of jacalin, though it still contains at least nine bands. Two of the additional β species that were found had masses which were 106 Da higher than the parent β_1 and β_2 21-residue forms. We have not been able to determine what substituent is responsible for the formation of these species. It is evidently a labile group, being easily cleaved in the MS/MS fragmentation and not giving a novel derivative in N-terminal sequencing. The sequence differences between jacalin and MPA (Fig. 4) are conservative changes with one exception, the replacement of Ala17 by Asp. Hence derivatization of this residue may be the cause of the 106 Da change.

One dominant α-chain peak was seen in the MPA mass spectrum, but in the protein sequencing there was evidence of genetic variation at seven positions. If all these changes occurred together, they would give an α-chain of 14 657 Da, and a small peak of this mass was seen in the mass spectrum (Fig. 3). But its relative amount compared with the main peak is not comparable to the relative levels of β_1 and β_2. It is possible that the main peak contains two species however, since many combinations of the sequence differences would give mass changes of 14 Da or less, which is at the 0.1% limit of resolution for the protein.

It is noteworthy that two of the three cleavages of the jacalin proprotein required to form the two-chain product occur at Asn residues. These sites are at the end of the propeptide segment and at the end of the four residue β-α linker [9]. In addition, processing in the MPA β-chains also occurs at Asn, at position 21 and in β_{1b} and β_{2b} at position 1. Cleavages at Asn residues govern the processing of the precursor forms of concanavalin A [13] and the pea lectin [12] to their mature forms, and for concanavalin A this processing is essential for carbohydrate-binding activity. The Asn-specific protease in jackbeans that is responsible for the concanavalin A processing has been purified and characterized recently [14]. The data on the *Moraceae* lectins suggest that a similar enzyme may be involved in the processing of these non-legume lectins.

In conclusion, the electrospray mass spectrometry allowed additional genetic and post-translational forms to be characterized for these highly polymorphic proteins, as well as their correlation with the reported genes for jacalin. These results will be of value for the X-ray crystallographic studies presently underway on both jacalin [15] and MPA [4], since an understanding of their sequence heterogeneity will facilitate interpretation of the electron density maps.

Acknowledgements

The authors thank R. A. Z. Johnston for excellent technical assistance and P. K. S. Blay for preparing the figures.

References

1. Becker JW, Reeke GN, Wang JL, Cunningham BA, Edelman GM (1975) *J Biol Chem* **259**:1513–24.
2. Wright CS (1977) *J Mol Biol* **141**:439–57.
3. Young NM, Johnston RAZ, Szabo AG, Watson DC (1989) *Arch Biochem Biophys* **270**:596–603.
4. Lee X, Johnston RAZ, Rose DR, Young NM (1989) *J Mol Biol* **210**:685–86.
5. Ruffet E, Paquet N, Frutiger S, Hughes GJ, Jaton J-C (1992) *Biochem J* **286**:131–34.
6. Young NM, Johnston RAZ, Watson DC (1991) *FEBS Letts* **282**:382–84.
7. Capon C, Piller F, Wieruszeski J-M, Leroy Y, Fournet B (1990) *Carbohydr Res* **199**:121–27.
8. Ashford DA, Dwek RA, Rademacher TW, Lis H, Sharon N (1991) *Carbohydr Res* **213**:215–27.
9. Yang H, Czapla TH (1993) *J Biol Chem* **268**:5905–10.
10. Smith RD, Loo JA, Edmonds CG, Barinaga CJ, Udseth HR (1990) *Anal Chem* **62**:882–89.
11. Tang X-J, Thibault P, and Boyd RK (1993) *Anal Chem* **65**:2824–34.
12. Higgins TJV, Chandler PM, Zurawski E, Button SC, Spencer D (1983) *J Biol Chem* **258**:9544–49.
13. Carrington DM, Auffret A, Hanks DE (1985) *Nature* **313**:64–67.
14. Abe Y, Shirane K, Yokosawa H, Matsushita H, Mitta M, Kato I, Ishii S-I (1993) *J Biol Chem* **268**: 3525–29.
15. Banerjee R, Dhanaraj V, Mahanta SK, Surolia A, Vijayan M (1991) *J Mol Biol* **221**: 773–76.