Post-translational processing of two α-amylase inhibitors and an arcelin from the common bean, *Phaseolus vulgaris*

N. Martin Young[a,*], Pierre Thibault[a], David C. Watson[a], Maarten J. Chrispeels[b]

[a] *Institute for Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ont. K1A 0R6, Canada*
[b] *Department of Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0116, USA*

Received 18 January 1999

Abstract Mass spectrometric methods were used to investigate the proteolytic processing and glycopeptide structures of three seed defensive proteins from *Phaseolus vulgaris*. The proteins were the α-amylase inhibitors αAI-1 and αAI-2 and arcelin-5, all of which are related to the seed lectins, PHA-E and PHA-L. The mass data showed that the proteolytic cleavage required for activation of the amylase inhibitors is followed by loss of the terminal Asn residue in αAI-1, and in all three proteins, seven or more residues were clipped from the C-termini, in the manner of the seed lectins. In most instances, individual glycoforms could be assigned at each Asn site, due to the unique masses of the plant glycopeptides. It was found that αAI-1 and αAI-2 differed significantly in their glycosylation patterns, despite their high sequence homology. These data complement the previous X-ray studies of the α_1-amylase inhibitor and arcelin, where many of the C-terminal residues and glycopeptide residues could not be observed.

Key words: Arcelin; Amylase inhibitor; Mass spectrometry; Glycopeptide; Post-translational processing

1. Introduction

Seeds of legume plants contain in addition to their well-known lectins other proteins that show sequence homology to the lectins but do not bind carbohydrates. In the common bean, *Phaseolus vulgaris*, several of these lectin-like proteins have been identified [1], and have been shown to function as defensive proteins [2]. They include two inhibitors of insect gut α-amylases, αAI-1 and αAI-2, and a small group of proteins termed arcelins, that are only found in some wild accessions of the common bean. The major interest in these proteins is that they can be used to produce insect resistant transgenic plants [3]. The proteins show deletions in some of the loops that form the carbohydrate binding site in the homologous *P. vulgaris* lectins PHA-E and PHA-L [4,5], and the two amylase inhibitors also undergo an internal proteolytic cleavage that is essential for their activity [6]. X-ray crystallographic studies have been reported on the complex of αAI-1 with porcine pancreatic amylase [7], arcelin-5 [8] and arcelin-1 [9], and PHA-L [10].

In the crystal structures, many C-terminal amino acid residues and most of the sugar residues of the glycopeptides could not be located in the electron density maps. Previous mass spectrometric studies of legume lectins, including *P. vulgaris* hemagglutinins, have shown that they undergo extensive post-translational proteolysis at their C-termini [11,12]. We report mass spectrometric experiments that show most of the unlocatable C-terminal residues are in fact not present in the mature molecules, and that the processing at the internal cleavage site in the amylase inhibitors is more extensive than a simple scission of the polypeptide chain. The data also allowed assignment of specific glycoforms to the individual Asn sites, aided by the simpler structures of plant *N*-linked glycopeptides compared to their mammalian counterparts [13,14]. For vacuolar plant glycoproteins, the species found are high mannose forms of general formula $Man_{5-9}GlcNAc_2$ and complex ones, $Man_3GlcNAc_2Fuc_{0-1}Xyl_{0-1}$, all of which have unique molecular masses.

2. Materials and methods

2.1. Proteins and peptide preparations

The amylase inhibitors and arcelin-5 were prepared as previously described [15]. Initial ES/MS experiments showed significant amounts of a small (8–9 kDa) protein were present in some of the samples, which were removed by size exclusion chromatography on Superdex 75 or reverse phase HPLC. The latter procedure was also used to separate the two chains of the amylase inhibitors. Peptides were prepared from the individual chains as previously described [16], including cleavage at Asp residues by mild acid hydrolysis, further fragmentation with trypsin and separation of the peptide digests by reverse phase HPLC, including on-line LC/MS.

2.2. Mass spectrometry

Electrospray mass spectra were obtained with a PE-Sciex (Concord, Ont.) API 300 instrument, and MALDI-TOF spectra were obtained with a Perseptive Biosystems Elite-STR (Framingham, MA) using linear mode. The matrix α-cyano cinnaminic acid (Aldrich) was used for all MALDI-TOF analyses. Mass assignment was made using external calibration and mass accuracy was found to be within ±3 Da of the predicted molecular weight. Nanoelectrospray mass spectra were obtained using a modified MicroIonspray interface (PE-Sciex) comprising a small tee insert used to establish electrical contact between a fused silica transfer line (80 cm × 50 μm i.d.) and the emitter tip. Disposable nanoelectrospray emitters with no gold coating were made of a small length of tapered fused silica, 5.3 cm × 50 μm i.d. with a 5 μm tip diameter, similar to those described previously [17]. Samples dissolved in 0.1 M HCOOH were infused to the mass spectrometer at a flow rate of approximately 200 nl/min using a pressurized inlet device.

3. Results

3.1. αAI-1

The masses of the α and β chains of αAI-1 were measured by MALDI-TOF/MS (Fig. 1a) and by ES/MS of the separated chains. The major α-chain species had a mass of 11 646 Da. It has two glycosylation sites with reported major

*Corresponding author. Fax: (1) (613) 941-1327.
E-mail: martin.young@nrc.ca

Abbreviations: αAI-1 and αAI-2, α-amylase inhibitors 1 and 2; ES/MS, electrospray mass spectrometry; LC/MS, liquid chromatography-mass spectrometry; MALDI-TOF/MS, matrix assisted laser desorption ionisation-time of flight mass spectrometry

PII: S0014-5793(99)00212-4

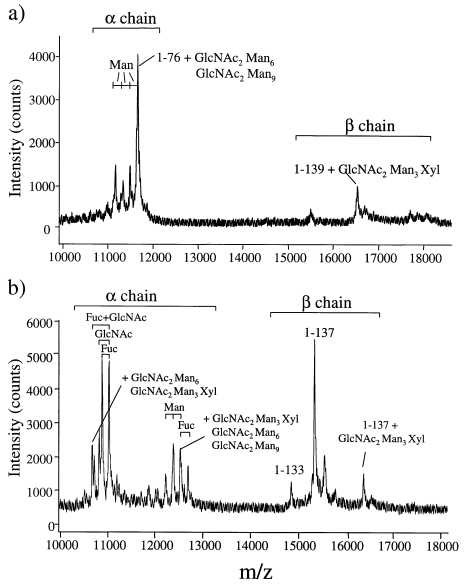

Fig. 1. MALDI-TOF mass spectra of (a) αAI-1 and (b) αAI-2.

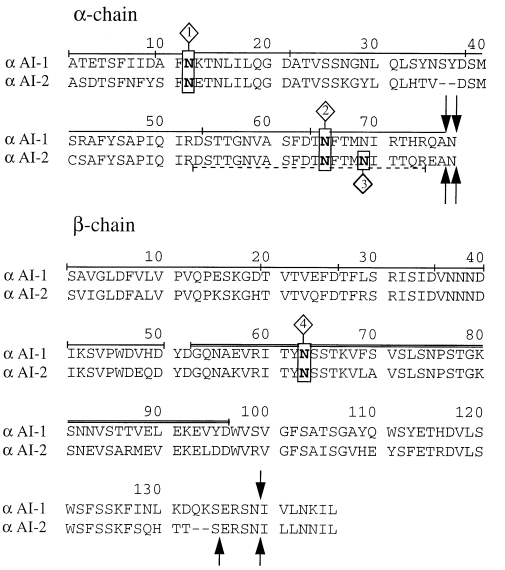

Fig. 2. Sequences of αAI-1 [20] and αAI-2 [22] with the sequenced peptides and the located cleavage (arrows) and glycopeptide sites (diamonds). Glycan 1 was predominantly $GlcNAc_2Man_6$, glycan 2 was predominantly $GlcNAc_2Man_9$, and glycans 3 and 4 were predominantly $GlcNAc_2Man_3Xyl$. Peptides from mild acid cleavage at Asp are shown with a single line above the αAI-1 sequence and tryptic sub-peptides with a double line. The tryptic peptide with two Asn sites from αAI-2, examined by tandem mass spectrometry, is indicated with a dashed underline.

glycoforms of $Man_6GlcNAc_2$ and $Man_9GlcNAc_2$, from chemical [18] or tandem mass spectrometric analyses [19]. Together with the reported 1–77 sequence [20] these give a predicted mass of 11 761 Da. The discrepancy of 115 Da suggested that the C-terminal Asn-77 had also been removed proteolytically. To confirm this and to assign the glycoforms to the respective sites, peptides were prepared by mild acid cleavage at Asp residues. Mass data and N-terminal sequencing confirmed the sequence of the α-chain and that Asn-77 was not present (Fig. 2). The peptide data also confirmed that Asn-12 carried the $Man_6GlcNAc_2$ and related glycoforms, and Asn-65 carried the $Man_9GlcNAc_2$ and its trimmed subspecies [19].

The β-chain mass was measured as 16 528 Da. Its dominant single glycan was reported to be $Man_3GlcNAc_2Xyl_1$ [18], hence the calculated mass was 17 319 Da. The discrepancy again suggested C-terminal processing, removing seven amino acids after an Asn residue, which would give a calculated mass of 16 527 Da. The β-chain sequence contains two possible Asn-Xaa-Ser/Thr sites for glycosylation, Asn-63 and Asn-83 (Fig. 2). LC/MS peptide mapping of a mild acid and trypsin double digest (data not shown) identified the site as Asn-63, which is homologous with a site in αAI-2 [1]. This site was identified in the subsequent X-ray structure [7] and the present mass spectrometry data confirmed the glycan was predominantly the $Man_3GlcNAc_2Xyl_1$ species [18].

3.2. αAI-2

Similar experiments to the above were carried out with the second *P. vulgaris* amylase inhibitor to see if it had undergone similar post-translational processing. A MALDI-TOF experiment with material purified by size exclusion chromatography showed two species of α-chain with two or three glycopeptide chains respectively (Fig. 1b), consistent with the additional site for *N*-glycosylation at Asn-69 compared to αAI-1. The mass of the larger species with two major glycans, 11 014 Da, is consistent with glycans of combined composition $Man_{11}GlcNAc_5Xyl_1Fuc_1$, calculated mass 11 010 Da. Tandem mass spectral analysis of a tryptic peptide 51–72, containing both the Asn-63 and Asn-69 sequons, showed that one of these carries a single GlcNac residue (data not shown). The other glycan on this glycopeptide was $GlcNAc_2Man_3XylFuc_{0-1}$. In the α-chain species with three glycans, the single GlcNac is replaced by $GlcNAc_2Man_9$. Homology to αAI-1 suggests this should be assigned to Asn-63. In contrast to the glycosylation of αAI-1, the predominant species of αAI-2 β-chain had a mass of 15 353 Da, consistent with a non-glycosylated β-chain truncated at the homologous position to αAI-1, calculated mass 15 351 Da.

3.3. Arcelin-5

This protein is monomeric [8], does not undergo internal cleavage, and its mode of toxic action is not known. There are

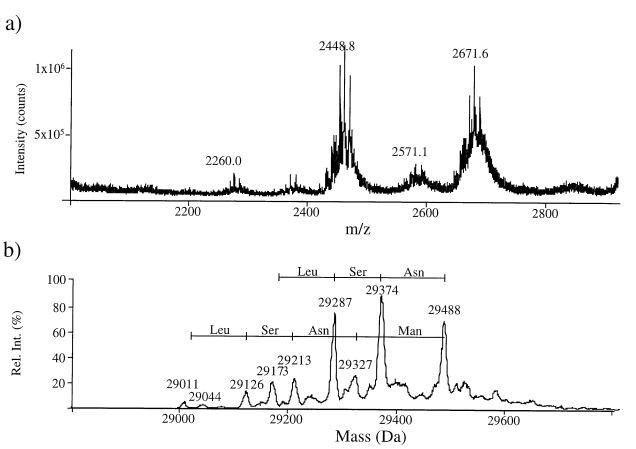

Fig. 3. Conventional ES/MS spectrum of arcelin-5 (a) and reconstructed molecular mass profile (b). The assigned C-terminal sequence from the two sets of peaks is described in Fig. 4.

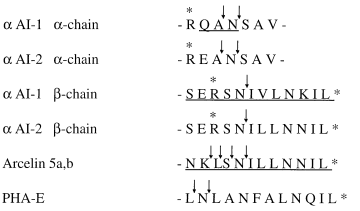

Fig. 4. Comparison of the C-terminal processing of the *P. vulgaris* hemagglutinin, PHA-E [11], with the processing of the three lectin-like proteins. The internal cleavage points are indicated with gaps in the α-chain sequences, residues not located in the crystal structure work [7–10] are underlined (a further seven residues were not found in the αAI-1 β-chain) and the asterisks indicate the C-terminal residues reported from protein sequencing [23,24].

two closely related forms, arcelin-5a with three potential *N*-glycosylation sites, and arcelin-5b with two sites [21], and it was suggested that only two sites on 5a were substituted [21]. However, several C-terminal residues were not located in the recent crystallographic work [8], pointing to possible C-terminal truncation, and only one glycosylation site was identified, at Asn-22. ES/MS measurements of the intact protein disclosed a complex set of components which could be resolved into two sets, each comprised of species with different degrees of C-terminal truncation (Fig. 3). The C-terminal sequence could thus be read directly and matched to the reported gene sequences which are identical in this region (Fig. 4). This showed that both species were truncated around the same point as the amylase inhibitors. The arcelin type was assigned as 5a, based on a peptide consistent with the 5a sequence for residues 44–49 [21] being found in tryptic peptide mapping by LC/MS. The mass spectrometry data suggest there are two glycans per chain, but exact attribution to individual glycoforms could not be made directly from these data.

4. Discussion

By means of relatively simple mass spectrometric experiments, it was possible to investigate simultaneously two aspects of the post-translational processing of these plant seed glycoproteins, viz. glycosylation and proteolytic C-terminal trimming. This was made possible by the simplicity and unique masses of each of the possible plant glycans. The results are complementary to the recent crystallographic studies of these proteins, which did not locate many residues involved in the post-translational modifications.

The proteolysis which leads to activation of αAI-1 is demonstrated by the present mass spectrometric data to be more complex than previously thought. Rather than a simple cleavage at the carboxyl side of Asn-77, presumably by the Asn-specific seed protease which seems to be generally important in legume lectin processing [10,11], the Asn-79 is completely removed, apparently by the action of a carboxypeptidase. In the crystal structure of the porcine amylase complex with this inhibitor [7], three residues were actually unassignable from the electron density in this key active site region. The cleavage occurs at a point which is close to that of a *cis* peptide bond in legume lectin and arcelin [8] structures, at Ala-Asp and Ala-Tyr sequences respectively. The cleavage may therefore reflect an alternative means to achieving the unusual conformations required for the active form of members of this protein family. Furthermore, 19 residues at the C-terminus of the β-chain of αAI-1 and 12 residues of arcelin were also not locatable from their electron density maps [7,8]. The mass data show that many of these residues are in fact not present in the mature forms of the proteins. The third protein, αAI-2, shows similar cleavages to αAI-1, but a somewhat different glycosylation pattern.

The mass spectrometry data for the αAI-1 and its peptides allowed further assignment of the reported glycoforms [16] to individual Asn residues in the α- and β-chains. Though the gene sequences of both inhibitors had been reported [20,22], the sequences were re-determined by protein methods recently [23,24]. The mass spectrometry data show that the latter experiments did not give a complete sequence as short peptides at each of the C-termini were overlooked (Fig. 4). Surprisingly the glycan assignments show that the homology of the Asn sequon locations in the two inhibitors does not lead to similar glycopeptides, because of differential glycan modification. The additional site in αAI-2 at Asn-67, which is remarkably close to the Asn-63 one, appears to influence the glycosylation of the Asn-63. In some molecules, it has no glycan, in others it carries a $GlcNAc_2Man_9$ and in a third set there appears to be a single GlcNAc. The simplest way this could arise is by the action of a seed endoglucosaminidase on the $GlcNAc_2Man_9$ species. In addition, the β-chain of αAI-2 was more rarely glycosylated at the homologous Asn sequon than αAI-1. A similar phenomenon has been observed in the bean storage protein phaseolin, which has glycans at Asn-252 and Asn-341. When both positions are glycosylated, both remain in the high mannose form. When only Asn-252 is glycosylated, the high mannose glycan is modified in the Golgi to become a complex glycan [25]. These positions are close together in the three-dimensional structure of phaseolin [26].

The C-terminal processing of several legume lectins has been shown to be attributable to the combined actions of an Asn-specific seed protease and a carboxypeptidase [11,12]. The lectin homologues studied here show very similar processing patterns to PHA-E when the various sequences and cleavage sites are aligned (Fig. 4). The C-terminal segment appears to be highly susceptible to proteolysis since lectins expressed in *Escherichia coli* also show C-terminal processing from endogenous proteases [11]. This may be true of many other proteins since accurate mass determination has only recently become possible and is not yet routine in many protein laboratories. However, in crystallographic studies when C-terminal residues are not detectable in the electron density, a check to see if they are actually present in the protein, whether recombinantly produced or not, is definitely indicated.

Acknowledgements: We thank Tomoko Hirama for assistance with the purification of αAI-2, Dr. José J. Pueyo for the purification of the first batch of αAI-1, and Dr. M.F. Grossi de Sa for a gift of arcelin-5 purified in her laboratory at Cenargen, EMBRAPA, Brazil. This work was supported in part by a grant to M.J.C. from the US Department of Energy, Energy Biosciences Program.

References

[1] Mirkov, T.E., Wahlstrom, J.M., Hagiwara, K., Finardi-Filho, F., Kjemtrup, S. and Chrispeels, M.J. (1994) Plant Mol. Biol. 26, 1103–1113.

[2] Chrispeels, M.J. and Raikhel, N.V. (1994) Plant Cell 3, 1–19.
[3] Shade, R.E., Schroeder, H.E., Pueyo, J.J., Tabe, L.M., Murdock, L.L., Higgins, T.J.V. and Chrispeels, M.J. (1994) BioTechnology 12, 793–796.
[4] Mirkov, T.E., Evans, S.V., Wahlstrom, J., Gomez, L., Young, N.M. and Chrispeels, M.J. (1995) Glycobiology 5, 45–50.
[5] Rougé, P., Barre, A., Causse, H., Chatelain, C. and Porthé, G. (1993) Biochem. Syst. Ecol. 21, 695–703.
[6] Pueyo, J.J., Hunt, D.C. and Chrispeels, M.J. (1993) Plant Physiol. 101, 1341–1348.
[7] Bompard-Gilles, C., Rousseau, P., Rougé, P. and Payan, F. (1996) Structure 4, 1441–1452.
[8] Hamelryck, T.W., Poortmans, F., Goossens, A., Angenon, G., Van Montagu, M., Wyns, L. and Loris, R. (1996) J. Biol. Chem. 271, 32796–32802.
[9] Mourey, L., Pédelacq, J.-D., Birck, C., Fabre, C., Rougé, P. and Samama, J.-P. (1998) J. Biol. Chem. 273, 12914–12922.
[10] Hamelryck, T.W., Dao-Thi, M.-H., Poortmans, F., Chrispeels, M.J., Wyns, L. and Loris, R. (1996) J. Biol. Chem. 271, 20479–20485.
[11] Young, N.M., Watson, D.C., Yaguchi, M., Adar, R., Arango, R., Rodriguez-Arango, E., Sharon, N., Blay, P.K.S. and Thibault, P. (1995) J. Biol. Chem. 270, 2563–2570.
[12] Young, N.M., Watson, D.C. and Thibault, P. (1995) Glycoconjug. J. 12, 135–141.
[13] Ashford, D.A., Dwek, R.A., Rademacher, T.W., Lis, H. and Sharon, N. (1991) Carbohydr. Res. 213, 215–227.
[14] Sturm, A. (1995) in: Glycoproteins (Montreuil, J., Schachter, H. and Vliegenthart, J.F.G., Eds.), pp. 521–541, Elsevier Science, Amsterdam.
[15] Moreno, J. and Chrispeels, M.J. (1989) Proc. Natl. Acad. Sci. USA 86, 7885–7889.
[16] Young, N.M., Johnston, R.A.Z. and Watson, D.C. (1991) Eur. J. Biochem. 196, 631–637.
[17] Bateman, K.P., White, R.L. and Thibault, P. (1997) Rapid Commun. Mass Spectrom. 11, 307–315.
[18] Yamaguchi, H., Funaoka, H. and Iwamoto, H. (1992) J. Biochem. (Tokyo) 111, 388–395.
[19] Bateman, K.P., White, R.L., Yaguchi, M. and Thibault, P. (1998) J. Chromatogr. A 794, 327–344.
[20] Hoffman, L.M. (1984) J. Mol. Appl. Genet. 2, 447–453.
[21] Goossens, A., Geremia, R., Bauw, G., Van Montagu, M. and Angenon, G. (1994) Eur. J. Biochem. 225, 787–795.
[22] Suzuki, K., Ishimoto, M. and Kitamura, K. (1994) Biochim. Biophys. Acta 1206, 289–291.
[23] Kasahara, K., Hayashi, K., Arakawa, T., Philo, J.S., Wen, J., Hara, S. and Yamaguchi, H. (1996) J. Biochem. (Tokyo) 120, 177–183.
[24] Nakaguchi, T., Arakawa, T., Philo, J.S., Wen, J., Ishimoto, M. and Yamaguchi, H. (1997) J. Biochem. (Tokyo) 121, 350–354.
[25] Sturm, A., van Kuik, J.A., Vliegenthart, J.F.G. and Chrispeels, M.J. (1987) J. Biol. Chem. 262, 13392–13403.
[26] Lawrence, M.C., Suzuki, E., Varghese, J.N., Davis, P.C., van Donkelaar, A., Tulloch, P.A. and Colman, P.M. (1990) EMBO J. 9, 9–15.