

# The Molecular Basis for N-Glycosylation in the 11S Globulin (Legumin) of Lupin Seed

M. Duranti,<sup>1,2</sup> C. Horstmann,<sup>3</sup> J. Gilroy,<sup>4</sup> and R. R. D. Croy<sup>4</sup>

Received September 27, 1994

Ion exchange-HPLC under denaturing conditions was used to purify to homogeneity the major  $M_r$  44,000  $\alpha$  subunit of lupin seed (*Lupinus albus*, L.) 11S storage globulin (legumin). The carboxymethylated subunit was digested with trypsin and the peptide fragments separated by reverse phase HPLC. Only one glycosylated peptide reacting with concanavalin A was identified by dot-blotting. Its amino acid sequence allowed the location of this peptide within a highly conserved region in proximity to the N-terminus of the  $\alpha$  subunits of the 11S globulins from other seeds. The unique presence of a serine residue in a sequence N-X-S of lupin 11S globulin, compared with all other 11S proteins, allows it to be the only protein of this class to bear covalently linked carbohydrate.

**KEY WORDS:** Lupin; 11S globulin; legumin; peptide mapping; amino acid sequence; glycosylation.

## 1. INTRODUCTION

Legume storage proteins represent a large group of proteins most of which belong to the 11S (legumin) and 7S globulin (vicilin) families, so named according to their sedimentation behavior (for a review see Casey *et al.*, 1986). These two major protein classes were found to share regions of amino acid sequence conservation reflecting a common ancestral origin (Borroto and Dure, 1987), and consequently a number of secondary structural features (Argos *et al.*, 1985). However, relevant differences in each protein family among species (Casey *et al.*, 1986) and protein polymorphism within a single species (Nielsen *et al.*, 1989) have been described. The heterogeneity arises from a combination of two factors, namely the multigene

origin of each storage globulin (Casey *et al.*, 1986) and the posttranslational modifications of relatively few expression products (Wright, 1986). A remarkable difference between 11S and 7S globulins is that the latter are usually glycosylated (Casey *et al.*, 1986), whereas the 11S globulins only rarely are. One example of a glycosylated protein of this class is lupin 11S globulin (Duranti *et al.*, 1988).

Lupin legumin, like all other legumins so far characterized (Casey *et al.*, 1986), represents the product of genes encoding contiguous precursors proteolytically processed posttranslationally to  $\alpha + \beta$  polypeptides. In lupin 11S globulin, four precursor polypeptides of  $M_r$  ranging from 67,000 to 72,000 yield a number of  $\alpha$  polypeptides of different sizes (Duranti *et al.*, 1992). Indeed the lupin mature 11S globulin consists of various  $\alpha$  chains ranging in apparent  $M_r$  from about 40,000 to 52,000 that are disulfide-bridged to  $\beta$  chains of  $M_r$  21,000 (Duranti *et al.*, 1988). No amino acid sequence of the 11S globulin from lupin seeds is available, except for the N-terminal region of the  $\beta$  chain, which was shown to be highly homologous to that of all other 11S globulins sequenced

<sup>1</sup> Dipartimento di Scienze Molecolari Agroalimentari e Centro Interuniversitario per lo Studio delle Macromolecole Informazionali, Università Statale di Milano, 20133-Milan, Italy.

<sup>2</sup> To whom correspondence should be addressed.

<sup>3</sup> On leave of absence from Institut für Genetik und Kulturpflanzenforschung, Gatersleben, Germany.

<sup>4</sup> Department of Biological Sciences, University of Durham, England.

(Duranti, 1993). All the  $\alpha$  subunits contain covalently linked carbohydrate (Duranti *et al.*, 1988).

With the aim of providing a molecular basis for lupin 11S globulin glycosylation, we purified to homogeneity the major  $\alpha$  chain, identified the glycosylated tryptic fragment, and sequenced it.

## 2. METHODS

Globulin 8, the major 11S storage globulin of the lupin seed, was purified from seeds of *Lupinus albus* L. cv. Multitalia as described by Duranti *et al.* (1988). The purified globulin (80 mg) was denatured with recrystallized 6 M urea in 0.1 M Tris-H<sub>2</sub>SO<sub>4</sub>, pH 8.6. Aliquots were loaded on a MonoQ HR 5/5 column (Pharmacia) equilibrated with the same buffer as above. Peaks were eluted with a linear gradient of Na<sub>2</sub>SO<sub>4</sub> from 0 to 0.3 M in 45 min at a flow rate of 0.25 ml/min. The two major peaks were reduced with the same buffer as above containing 10 mM dithiothreitol and carboxymethylated with 60 mM iodoacetamide for 1 hr at room temperature in the dark, dialyzed against 10 mM ammonium bicarbonate, pH 8.0, and freeze-dried. Each peak (about 10 mg), resuspended in 0.015 M Tris-H<sub>2</sub>SO<sub>4</sub>, pH 7.5, containing 6 M urea, was

rechromatographed on MonoQ HR 5/5. Peaks were eluted as above with a 30-min linear gradient of Na<sub>2</sub>SO<sub>4</sub>. Pooled fractions were dialyzed as above, freeze-dried, and then submitted to SDS-PAGE for polypeptide identification (Fig. 1) before digestion with trypsin (Fig. 2).

The freeze-dried polypeptide fragments were resuspended in 0.1% trifluoroacetic acid in milliQ water and submitted to reverse phase (RP) HPLC on a delta Pack C-18 resin, 300-Å, 3.9 × 150 mm column of a Model 510 Waters HPLC. Peaks were eluted with a linear gradient of acetonitrile from 0 to 60% in 60 min at a flow rate of 0.8 ml/min. The eluate was monitored at 216 nm.

Aliquots of each peak obtained from the RP HPLC separation were blotted onto a nitrocellulose (NC) membrane using a BioDot apparatus (BioRad). The membrane was blocked with 1% (w/v) gelatin (Sigma) in 10 mM sodium-potassium phosphate buffer, pH 7.4, containing 27 mM KCl and 137 mM NaCl and reacted with concanavalin A (Con A) and peroxidase essentially as described by Kuo-Liang Hsi *et al.* (1991).

The N-terminal amino acid sequence analysis was performed on 20–50 µg of sample in a Model 477A pulsed-liquid microsequencer (Applied Biosystems) equipped with on-line phenylthiohydantoin analyzer, Model 120A (Applied Biosystems) essentially according to Hunkapiller and Hood (1983). Amino acid sequence comparisons were carried out with PC/GENE 6.7 software (Intelligenetics, Mountain View, CA).

## 3. RESULTS AND DISCUSSION

The MonoQ-HPLC separations before and after carboxymethylation of the cysteine residues of the subunits gave a number of unresolved peaks (data not shown) as a result of the heterogeneity of the 11S globulin (Duranti *et al.*, 1988). The final preparations consisted of various  $\alpha$  and one size class of  $\beta$  chains (Fig. 1). While the  $\alpha$  chain of  $M_r$  44,000 size appeared to be homogeneous, the  $M_r$  52,000 one was contaminated by various bands which made up about 25% of the total stained polypeptide bands. The lack of homogeneity of some "purified" pea 11S globulin subunits has already been described by Bacon *et al.*, (1987). The band of  $M_r$  44,000 was therefore selected for tryptic digestion and amino acid sequencing of the separated peptides.

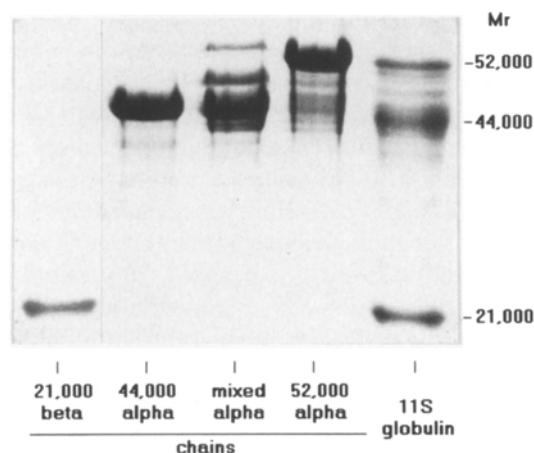


**Fig. 1.** SDS-PAGE pattern of the major carboxymethylated 11S globulin subunits separated by mono-Q HPLC. The polypeptides obtained by ion exchange HPLC were submitted to SDS-PAGE on 12% polyacrylamide gels under reducing conditions essentially as described by Laemmli (1970). From 5 to 10 µg protein was loaded in each lane. The protein bands were stained with Coomassie brilliant blue R250 and scanned on a Cream Image Processing System (KemEnTech, Copenhagen). Apparent molecular weights were determined by comparison with the marker proteins in the Pharmacia kit (not shown).

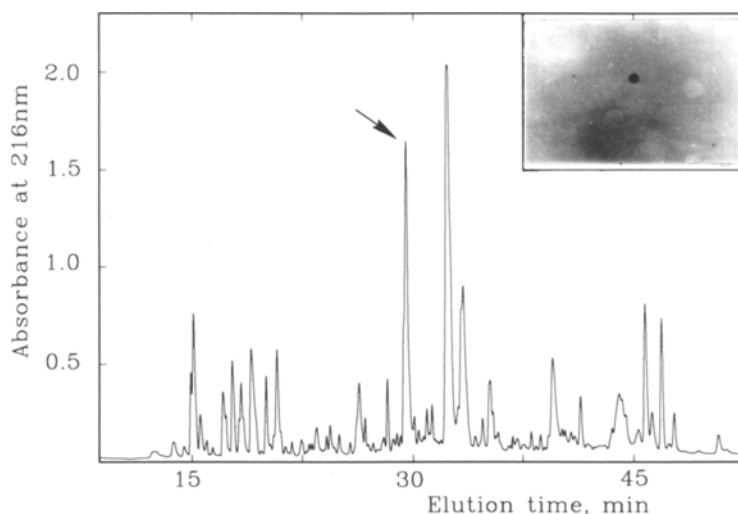


**Fig. 2.** RP HPLC separation of the tryptic peptides of lupin 11S globulin  $M_r$  44,000  $\alpha$  subunit. The  $M_r$  44,000  $\alpha$  polypeptide chain was incubated with trypsin in the ratio of 1:50 w/w in 25 mM  $(\text{NH}_4)_2\text{CO}_3$ , pH 7.5. The incubation was carried out at 37°C for 14–16 h. After digestion the solution was freeze-dried and then submitted to RP HPLC as described in Section 2. The elution profile of the tryptic digest of the  $M_r$  44,000  $\alpha$  chain is shown. The arrow indicates the glycosylated peptide. Inset: Reactivity of the glycosylated peptide with concanavalin A on a nitrocellulose membrane; experimental details in Section 2.

The elution profile of the tryptic fragments of this polypeptide separated by RP HPLC is shown in Fig. 2. Each peak was analyzed for the presence of carbohydrate by dot-blotting and reaction with Con A (Fig. 2, inset). The only peak which reacted with the lectin is marked with an arrow in Fig. 2.

The glycosylated peptide fragment from the  $M_r$  44,000  $\alpha$  chain was submitted to amino acid sequence analysis. The sequence obtained is shown in Fig. 3 and compared with the homologous

|                 |    |                                                |
|-----------------|----|------------------------------------------------|
| <b>L. albus</b> |    | VESEGGY <b>Y</b> TETW <b>X</b> S <b>R</b> PELR |
| <b>Glc4</b>     | 22 | VESEGGL <b>I</b> Q <b>T</b> WNS <b>O</b> HPELK |
| <b>Glc5</b>     | 22 | VESEGGL <b>I</b> ETWNS <b>O</b> HPELQ          |
| <b>Vic4</b>     | 24 | VESE <b>A</b> GLTETWN <b>P</b> NHPELR          |
| <b>LegA</b>     | 23 | IESEGGL <b>I</b> ETWNP <b>N</b> KKQFR          |
| <b>LegJ</b>     | 24 | VESE <b>A</b> GLTETWN <b>P</b> NHPELK          |

**Fig. 3.** Alignment of the amino acid sequence of the glycosylated region of lupin 11S globulin  $M_r$  44,000  $\alpha$  chain with those of other 11S proteins. Numbers indicate the positions from the N-terminus. Black boxes and asterisks indicate identical amino acid residues and the glycosylation consensus sequence, respectively. Glc4 and Glc5: Soybean (*Glycine max*) A4 and A3 subunits, respectively (Momma *et al.*, 1985). Vic4: Broad bean (*Vicia faba*) type A legumin (Baumlein *et al.*, 1986). LegA and LegJ: Pea (*Pisum sativum*) legumins (Lycett *et al.*, 1984; Gatehouse *et al.*, 1988).

regions of other unglycosylated 11S proteins (legumins). The region was unequivocally located in the proximity of the N-terminus of the  $\alpha$  chains of those polypeptides which have been fully sequenced and it appeared highly conserved also in lupin. The presence of an undetermined amino acid followed by two serine residues strongly suggests it consists of a carbohydrate-linked asparagine which is undetectable by the sequencer. In most other 11S globulins an asparagine is present (Fig. 3). With this assumption, the most remarkable difference among the sequences shown is the exclusive presence of an N-glycosylation consensus sequence, i.e., N-X-S, in the lupin 11S globulin. Indeed, lupin 11S globulin is the only protein of this class to bear covalently linked carbohydrate (Duranti *et al.*, 1988). The highest similarity (about 80%) was found between the lupin 11S peptide sequence and the soybean glycinins A3–B4 and A5, A4–B3, for which the serine residue immediately after the putative asparagine is also conserved, but not the serine in the third position (Fig. 3).

A significant amount of variation occurs in the storage globulins of most legume seeds. This is a reflection of the multigene families encoding legumin- and vicilin-type storage proteins and each

comprising up to ten or more genes (Casey *et al.*, 1986). This has arisen through gene duplication and mutation and translocation as well as the possible involvement of transposon-mediated deletion/insertion (Lycett *et al.*, 1985; Shirsat, 1988). One common mechanism appears to be through accumulation of single point mutations leading to conservative or nonconservative amino acid changes. Such changes account for the charge isomers of the constituent subunits (Krishna *et al.*, 1979), the presence or absence of posttranslational proteolytic cleavage sites (Gatehouse *et al.*, 1981), or, as we have demonstrated here in the 11S globulin of lupin, the presence or absence of a potential glycosylation site. The unique situation that a point mutation has realized a potential glycosylation site which is accessible on the nascent polypeptide chain and recognized by the plant N-glycosylation system is remarkable.

In this work we have presented the amino acid sequence of the lupin 11S globulin  $\alpha$  subunit glycosylated region and have shown that a single amino acid substitution dramatically affects the composition of this molecule, allowing about 1% (w/w) carbohydrate to be covalently linked to the polypeptide chain (Duranti *et al.*, 1981). Whether such a peculiarity of lupin 11S globulin has also structural and functional implications on the native protein is still to be demonstrated, though unusual properties in the folding and oligomeric assembly and in the proteolytic processing of this protein during seed development have already been noted (Duranti *et al.*, 1988, 1992; Duranti and Baffigo, 1993; Gayler *et al.*, 1984). In addition, an uneven glycosylation of the subunits can be a further cause of heterogeneity of this protein. It is possible that minor advantages such as nutritional benefit to the seedling or increased 11S mobilization during germination may have selected for and maintained the presence of the glycosylated 11S globulin in lupins. However, any apparent physiological advantage of the glycosylated protein over the unglycosylated (usual) variants found in other legumes would be difficult to assess between different legumes and may necessitate the use of transgenic lupins with unglycosylated 11S globulins.

## ACKNOWLEDGMENTS

Research supported by National Research Council of Italy, Special Project RAISA, Subproject No. 2, Paper No. 1971. The stay of C.H. at the Department of Agrifood Molecular Sciences, Milan, was supported by a Senior Scientist Fellowship, RAISA, National Research Council of Italy.

## REFERENCES

- Argos, P., Narayana, S. V. L., and Nielsen, N. C. (1985). *EMBO J.* **4**, 1111–1117.
- Bacon, J. R., Lambert, N., Phalp, M., Plumb, G. W., and Wright, D. J. (1987). *Anal. Biochem.* **160**, 202–210.
- Baumlein, A., Wobus, U., Pustell, J., and Kafatos, F. C. (1986). *Nucleic Acid Res.* **14**, 2707–2720.
- Boroto, K., and Dure III, L. (1987). *Plant Mol. Biol.* **8**, 113–131.
- Casey, R., Domoney, C., and Ellis, N. (1986). *Oxford Surv. Plant Mol. Cell Biol.* **3**, 1–95.
- Duranti, M. (1993). In *Food Proteins—Structure and Functionality* (Schwenke, K. D., and Mothes, R., eds.), VHC, Weinheim, pp. 129–136.
- Duranti, M., and Baffigo, M. (1993). *It. Biochem. Soc. Trans.* **4**, 295.
- Duranti, M., Restani, P., Poniatowska, T., and Cerletti, P. (1981). *Phytochemistry* **20**, 2071–2075.
- Duranti, M., Guerrieri, N., Takahashi, T., and Cerletti, P. (1988). *Phytochemistry* **27**, 15–23.
- Duranti, M., Guerrieri, N., Cerletti, P., and Vecchio, G. (1992). *Eur. J. Biochem.* **206**, 941–947.
- Gatehouse, J. A., Croy, R. R. D., Morton, H., Tyler, M., and Boulter, D. (1981). *Eur. J. Biochem.* **118**, 627–633.
- Gatehouse, J. A., Bown, D., Gilroy, J., Levasseur, M., Castleton, J., and Ellis, T. H. N. (1988). *Biochem. J.* **250**, 15–24.
- Gayler, K. R., Boadle, B. G., Snook, M., and Johnson, E. D. (1984). *Biochem. J.* **221**, 333–341.
- Hunkapiller, M. W., and Hood, L. E. (1983). *Meth. Enzymol.* **91**, 486–493.
- Krishna, T. G., Croy, R. R. D., and Boulter, D. (1979). *Phytochemistry* **18**, 1879–1880.
- Kuo-Liang Hsi, Ling Chen, Hawke, D. H., Zieske, L. R., and Pau-Miau Yuan (1991). *Anal. Biochem.* **198**, 238–245.
- Laemmli, U. K. (1970). *Nature* **227**, 680–685.
- Lycett, G. W., Croy, R. R. D., Shirsat, A. H., and Boulter, D. (1984). *Nucleic Acid Res.* **12**, 4493–4506.
- Lycett, G. W., Croy, R. R. D., Shirsat, A. H., Richards, D. M., and Boulter, D. (1985). *Nucleic Acid Res.* **13**, 6733–6743.
- Momma, T., Negoro, T., Hirano, H., Matsumoto, A., Uda, K., and Fukazawa, C. (1985). *Eur. J. Biochem.* **149**, 491–496.
- Nielsen, N. C., Dickinson, C. D., Cho, T. J., Thanh, V. H., Scallan, B. J., Fischer, R. L., Sims, T. L., Draws, G. N. and Goldberg, R. B. (1989). *Plant Cell* **1**, 313–328.
- Shirsat, A. H. (1988). *Mol. Gen. Genet.* **212**, 129–133.
- Wright, D. J. (1986). In *Development in Food Proteins—5* (Hudson, B. J. F. ed.) Elsevier Applied Science, London, pp. 81–157.