

Short sequence-paper

Cloning, sequencing and expression in the seeds and radicles of two *Lupinus albus* conglutin γ genes¹

Alessio Scarafoni ^{a,*}, Annalisa Di Cataldo ^a, Tatiana D. Vassilevskaia ^b,
Evguenia P. Bekman ^b, Claudina Rodrigues-Pousada ^{2,b}, Fabrizio Cecilianic,
Marcello Duranti ^a

^a Dipartimento di Scienze Molecolari Agroalimentari, Università degli Studi di Milano, Via Celoria 2, 20133 Milan, Italy

^b Instituto Gulbenkian de Ciência, Apartado 14, 2781 Oeiras, Portugal

^c Istituto di Fisiologia Veterinaria e Biochimica, Università degli Studi di Milano, Via Celoria 10, 20133 Milan, Italy

Received 29 August 2000; received in revised form 14 February 2001; accepted 4 April 2001

Abstract

Two genes encoding conglutin γ have been isolated from a *Lupinus albus* genomic library and sequenced. The expression of conglutin γ was studied by partial amino acid sequencing of the mature seed protein and by nucleotide sequencing of reverse transcriptase–polymerase chain reaction products from various tissues during the plant life cycle. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Conglutin γ ; Leguminous plant; Germinating seed; Young root; *Lupinus albus*

A number of major seed proteins have no storage role and in many cases no other evident biological function has been established so far. Conglutin γ (C γ) from *Lupinus* spp. is an example of these proteins [1]. C γ has previously been referred to as a storage protein, but due to several unusual features, this role has been reconsidered. Indeed, despite its high expression in the developing lupin seeds (up to 5% of the seed proteins) and deposition in the storage vacuoles [2], C γ dramatically differs from the canonical storage proteins because (i) it remains undegraded at germination [3,4]; (ii) it has been located in the extra-cellular spaces of germinating cotyledons [5] and in the young roots [6]; (iii) it is selectively secreted in high amounts upon heat treatment of lupin mature seeds [7]. C γ is a tetrameric protein, each monomer consisting of two disulphide-linked polypeptides of 30 and 17 kDa [8]; the large chain was shown to be glycosylated [9]. Significant sequence homology of *Lupinus angustifolius* C γ (LangC γ) [10] has only been found with three proteins,

all belonging to the plant kingdom. One is the 7S basic globulin (Bg7S) from soybean [11], which has been claimed to be part of the plant cellular signal transduction system due to its protein kinase activity [12], which is stimulated by leginsulin, a hormone-like peptide isolated from soybean radicles [13]. The second one is the extra-cellular dermal glycoprotein (EDGP) from carrot, which is expressed in response to wounding [14]. The third protein is a basic protein from cotton seeds showing a weak anti-fungal activity [15]. None of these activities have been identified for lupin C γ so far. *Lupinus albus* C γ (LalbC γ) has extensively been characterised at molecular level [16,17], but its nucleotide or amino acid sequences have as yet not been determined. In the present paper we describe the cloning and sequencing of two C γ genes from *L. albus* and report the expression of the protein in developing seeds and in the young roots of germinating seeds. These findings differ from those of Gayler and co-workers who detected only one version of the C γ gene in the *L. angustifolius* genome, which is almost exclusively expressed in developing seed cotyledons [18].

A genomic λ -Dash library constructed from *L. albus* nuclear DNA [19] was screened by using a probe generated by polymerase chain reaction (PCR) of genomic DNA isolated from *L. albus*, var. Multitalia, using the primers 5'-AAACCTCCTTGTCTACCC-3' (sense) and

* Corresponding author. Fax: +39-2-7063-3062;
E-mail: alessio.scarafoni@unimi.it

¹ EMBL accession numbers: AJ297490 and AJ297568.

² Present address: ITQB, Stress and genomics, Apartado 127, 2781-901 Oeiras, Portugal.

Cy32	YHNSQPTSSSKPNLLVLPIQQDASTKLHWGNILKRTPLMQVPVLLDLNGKHLWVTCSQ	58
Cy48	YHNSQPTSGSKPNLLVLPIQQDASTSLRWGNIHRRTPLMQVPLLLDLNGKHLWVTCSY	58
LangCy	YHNSQPTSS-KPNLLVLVQEDASTGLHWANIHKRTPLMQVPLLLDLNGKHLWVTCSQ	57
	***** :***** :***** :***** :***** :***** :*****	
Cy32	HYSSSTYQAPFCHSTQCSRANTHCFTCTDSTTSRPGCHNNTCGLISSNPVTQESGLG	116
Cy48	HYSSSTYQAPFCHSTQCSRANSHHCFTCTDSATSRPGCHNNTCALMSSNPVTQEAGFG	116
LangCy	HYSSSTYQAPFCHSTQCSRANTHCFTCTDSTTTTRPGCHNNTCGLSSNPVTQESGLG	115
	***** :***** :***** :***** :***** :***** :*****	
Cy32	ELAQDVLALHSTHGSKLGSVKIPQFLFSCAPTFLTQKGLPNNVQGALGLGHAPISLP	174
Cy48	ELAQDVLAIHSTHGSKLGPVRLQFLFSCAPSFQAQKGLPNNVQGALGLGHAPISLQ	174
LangCy	ELAQDVLAIHSTHGSKLGPVVKVPQFLFSCAPSFQAQKGLPNNVQGALGLQAPISLQ	173
	***** :***** :***** :***** :***** :***** :*****	
Cy32	NQLFSHFGLKRQFTMCLSSYPTSNAILFGDINDPNNNNYIHNSLDVLHDMVYTPITI	232
Cy48	NQLFSHFGLKRQFAMCLSRYPSTNGAILFGDIYDLN--YIHNSIDVLIDMVYTPILRI	231
LangCy	NQLFSHFGLKRQFSVCLSRYPSTNGAILFGDINDPNNNNYIHNSLDVLHDLVYTPITI	231
	***** :***** :***** :***** :***** :***** :*****	
Cy32	SKQGEYFIQVSAIRVNKHMVIPTKNPSMFPSSSSSYHESSEIGGAMITTTNPYTVLR	290
Cy48	SQQGEYFMQVNAIRVNKHMVPTKNPSML-----SSYHGDSRIGGAMITTTNPYTIHL	284
LangCy	SKQGEYFIQVNAIRVNKHLVIPTKNP--FISPSSTSYHSGEIGGALITTHPYTVLS	287
	* ***** :***** :***** :***** :***** :***** :*****	
Cy32	HSIFEVFTQVFANNVPKQAQVKAVGPFGLCYDTKKISGGVPSVDLIMDKSDVVRISG	348
Cy48	HSIFEVFTQVFANNMPKEAQVESVGPFGLCYDSRKLGGIPSEVFVMDSHDDVWRISD	342
LangCy	HSIFEVFTQVFANNMPKQAQVKAVGPFGLCYDSRKISGGAPSVDLILDKNDVWRIS	345
	***** :***** :***** :***** :***** :***** :*****	
Cy32	ENLMVQAQDGVSCGLGFDGGVHTRAGIALGTHQLEENLVVFDLARSRVGFNTNSLKSS	406
Cy48	ENLMVQAQNGVSCGLGFDGGMHTRTEIVLGTHQLEENMVVFDLERSRVGFNSNSLKSS	400
LangCy	ENFMVQAQDGVSCGLGFDGGVHARAGIALGAHLEENLVVFDLERSRVGFNSNSLKSS	403
	** ***** :***** :***** :***** :***** :***** :*****	
Cy32	HGKSCSNLFDLNNP	420
Cy48	HGKTCANIFDLNNA	414
LangCy	YGKTCSNLFDLNNP	417
	:***** :***** :*****	

Fig. 1. Alignment and comparison of the deduced amino acid sequences of conglutin γ from *L. albus* (Cy32 and Cy48) and *L. angustifolius* (LangCy). Asterisks indicate conserved residues; colons indicate conservative substitutions. Hyphens are inserted to optimise the alignments. The arrow indicates the processing site of the precursor.

5'-AAACCCAACCTTGACC-3' (antisense). These primers corresponded to the most conserved nucleotide sequences closest to the start and stop codons, respectively, when LangCy (accession number: L39786; from position 2059 to 2077 and from 3195 to 3211) and Bg7S (accession number: X16107; from position 1521 to 1539 and from 2630 to 2046) genes were compared using FASTA3 [20]. PCR conditions were as follows: 50 ng DNA, 10 mM Tris-HCl (pH 8.3), 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.2 mM dNTPs, 1 μ M each primer, 0.5 U *Taq* polymerase (Roche). After 30 amplification cycles (94°C for 1 min, 52°C for 1 min, 72°C for 1.5 min) the mixture was separated by electrophoresis on a 0.8% agarose gel. The amplified product, which migrated as a single band of approximately 1150 bp, was manually sequenced in order to confirm its correct composition and used as a probe to screen the library. The hybridisation of the replica filters

was performed overnight at 42°C in 40% formamide, 3 $\times$ SSC, 1 $\times$ Denhardt solution, 0.1 mg/ml herring sperm DNA. The last post-hybridisation washing was carried out at 60°C with 0.5 $\times$ SSC. A total of 56 plaques hybridised with the probe, but only two, designated λ LalB32 and λ LalB48, respectively, were positive to the PCR amplification described above. After Southern analysis, a 4.5 kb *Xba*I-*Xba*I fragment of the first clone and a 4.8 kb *Hind*III-*Hind*III fragment of the second clone were subcloned in pSE280 plasmid (Invitrogen) creating pPR12 and pAS1B, respectively. The two plasmids were sequenced at least twice in both directions by the dideoxy method [21] using specifically designed internal primers. With this approach the sequences of 2482 bp of pPR12 and 2249 bp of pAS1B were determined. Neither gene contained intron sequences nor alternative open reading frames.

The deduced amino acid sequences, identified as Cy32

and C γ 48, respectively, were aligned and compared with LangC γ sequence in Fig. 1. The two sequences corresponded to the C γ precursor [10]. C γ 32 precursor was 420 amino acids long (calculated M_r 45 608 Da), whereas C γ 48 precursor had 414 residues (calculated M_r 45 732 Da). The similarities with LangC γ are 94% and 90%, respectively. Only one glycosylation consensus sequence was found at position 98 in both proteins and the position of the 12 cysteine and the three tryptophan residues are strictly conserved. No specific functional consensus motifs were detected when the sequences were searched using the PROSITE database [22] and only a putative *Arabidopsis thaliana* protein (accession number: AC005278) was found as a further similar sequence (searching SWISSPROT and EMBL databases) in addition to BG7S and EDPG proteins.

The C γ 32 and C γ 48 sequences showed their major differences near the site of post-translational proteolytic cleavage between the large and the small chains (Fig. 1). In particular, an extra amino acid stretch composed of a proline and four serine residues is present at the C-terminus of the large chain of the C γ 32 sequence. Moreover, the N-terminal sequences of the small chains are SYHESSE and SYHGDSR in C γ 32 and C γ 48, respectively. The different number of acidic and basic residues of the two proteins leads to significantly different predicted pI s, i.e., 7.8 and 6.6 for C γ 32 and C γ 48, respectively.

The C γ 32 sequence perfectly matched the experimentally determined N-terminal amino acid sequences of both the 30 and 17 kDa chains of the protein extracted from mature dry seeds [7,16]. Also, the amino acid sequences of some internal peptides isolated, as described in [7,16], after trypsin digestion and reverse-phase high-performance liquid chromatography of the seed protein, confirmed such identity. The sequences of these peptides are shown in Fig. 2. Even the mentioned predicted pI of the C γ 32 sequence, namely 7.8, corresponded to that of the protein extracted from the seeds, i.e., 7.9 [8]. Furthermore, C γ 32 sequence shares the highest similarity with LangC γ , which has been found to be exclusively expressed and accumulated in the seeds [10]. All these data strongly supported the conclusion that the gene encoding C γ 32 is expressed during seed development and that the protein accumulated in the mature seeds.

Aiming to further investigate the expression of C γ , total RNAs were isolated, according to [23], from dry cotyledons and from leaves, cotyledons and roots of seedlings at different germination times. Northern analysis did not show any positive signal, except in the case of the RNA extracted from developing seeds used as a positive control (not shown), a result which is in agreement with previous findings [18]. Reverse transcription and PCR (RT-PCR) was used as a more sensitive technique to study C γ gene expression. The reverse transcription reaction was performed using the Superscript II kit (Gibco BRL) with 250 ng of random primers and following the manufactur-



Fig. 2. Amino acid sequencing of peptides isolated after trypsin digestion of C γ extracted from *L. albus* mature dry seeds. (a) N-terminal sequence of seed C γ large polypeptide chain from [7]; (b) this work; (c) N-terminal sequence of seed C γ small polypeptide chain from [16]. The sequences are compared with the same region of C γ 32 and C γ 48 sequences. Hyphens indicate identity between amino acids. The residues are indicated only when different from the reference sequence. The numbers indicate the corresponding position from the N-terminal residue in the C γ 32 sequence. X, not determined. Since PHT-cysteines are unstable in Edman conditions, cysteine residues have arbitrarily been assigned.

er's instructions. The reaction was carried out either with and without reverse transcriptase. With this experimental approach we were not able to amplify any full-length product. However, by using the primers 5'-GTTCTCATTGACATGGTGTAC-3' (sense) and 5'-AAACCCAACCTTGACC-3' (antisense), whose sequences are present in both genes and correspond to the central region of the C γ precursor (positions 760 and 1293 of the LalbA32 coding sequence), a product of about 530 bp was amplified in most samples. In the absence of RT no products were observed. To confirm these results, hybridisation of these fragments with the probe used to screen the genomic library was performed. As shown in Fig. 3A, all samples, except the leaves, gave positive signals, indicating that C γ is expressed also in plant tissues other than developing seeds. Nucleotide sequencing of selected regions of the products amplified from roots and cotyledons after 3-day germination and from dry seeds was carried out. Part of the nucleotide sequences of these fragments, along with their translation, is given in Fig. 3B; all of them appear

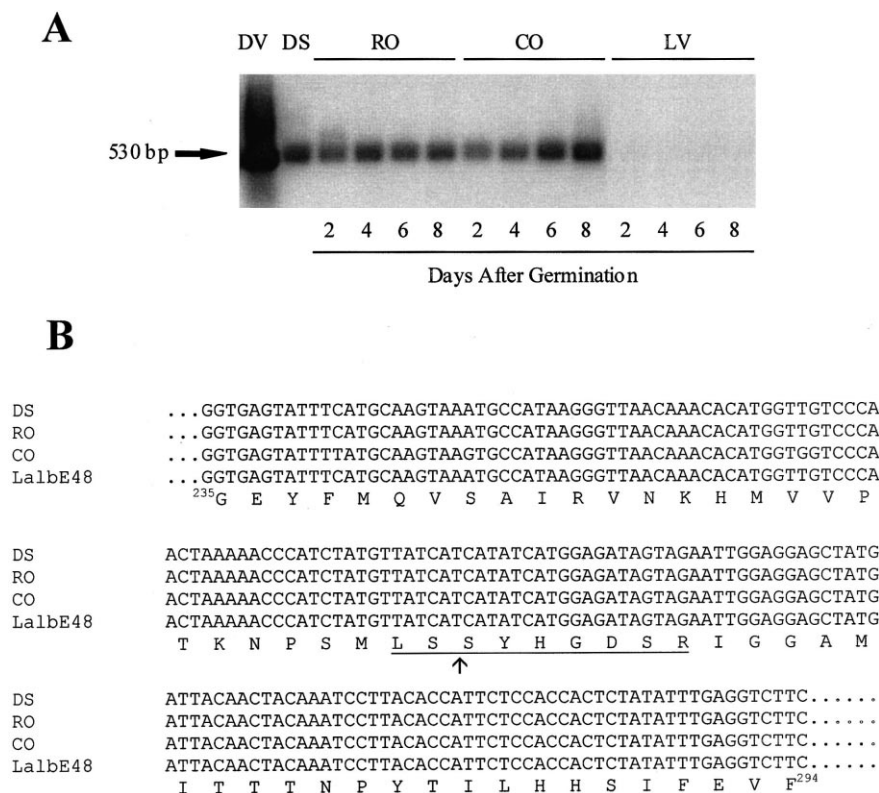


Fig. 3. Expression of conglutin γ in various *L. albus* tissues. (A) Southern blotting analysis of RT-PCR products. The hybridisation was carried out with the probe used to screen the genomic library. See text for details. The following samples were analysed: developing seeds (DV); dry seed cotyledons (DS); roots (RO), cotyledons (CO) and leaves (LV) of seedlings at different germination times. (B) Part of the nucleotide sequences of the RT-PCR product obtained from dry seed cotyledons (DS) and from 3-day-old roots (RO) and cotyledons (CO). The sequences are compared with the same region of the LalbE48 genomic sequence, which contains one of the two conglutin γ genes described in this work. The deduced amino acid sequence is also reported. The region between the large and the small polypeptide chains of the C γ precursor is underlined. The arrow indicates the proteolytic processing site between the two chains.

to be identical to the LalbE48 sequence. In particular, the lack of the serine stretch at the C-terminus of the large polypeptide chain and the sequence SYHGDSR at the N-terminus of the small polypeptide chain, typical of the LalbE48 clone, could be observed.

The results obtained from the dry cotyledon sample are quite surprising. Indeed, we could repeatedly detect only the mRNA corresponding to the LalbE48 clone, but the amino acid sequence of the protein accumulated in the mature seed was that of the LalbA32 clone as already discussed. One plausible explanation is that the two genes are expressed at different stages of seed development and only one of the proteins accumulated significantly. This point deserves further investigation. It is, however, possible that the two genes have originated from a common ancestor which has been duplicated and thus the respective functions have undergone specialisation. Experiments are in progress to verify whether the expression is modulated by specific physiological or pathological conditions.

This work was funded by grants MURST 60% (1998/1999 and 1999/2000) and by the grant Praxis/2/2.1/BIO/1146/95 awarded by FCT to C.R.-P.

References

- [1] M.-A. Esnault, J. Citharel, D. Thomas, P. Guegan, A. Cavalier, Behaviour of conglutin γ , a major lupin seed protein during germination, *Plant Physiol. Biochem.* 34 (1996) 101–109.
- [2] M. Duranti, F. Faoro, N. Harris, Immunocytochemical localisation of conglutin γ and legumin-like globulin in developing and mature seeds of *Lupinus albus*, *Protoplasma* 161 (1991) 104–110.
- [3] M. Duranti, E. Cucchetti, P. Cerletti, Changes in composition and subunits in the storage proteins of germinating lupin seeds, *J. Agric. Food Chem.* 32 (1984) 490–493.
- [4] J. Citharel, D. Delamarre, The γ -conglutin of *L. albus* L. var. Lutop (Leguminosae). Synthesis during seed maturation, *Plant Physiol. Biochem.* 27 (1989) 211–218.
- [5] M. Duranti, F. Faoro, N. Harris, The unusual extracellular localization of conglutin γ in germinating *Lupinus albus* seeds rules out its role as a storage protein, *J. Plant Physiol.* 143 (1994) 711–716.
- [6] M. Duranti, F. Ferrari, C. Gius, R. Carzaniga, F. Faoro, Biochemical and immunocytochemical identification of conglutin γ , a lupin seed lectin, in the roots of early germinating seedlings, *Protoplasma* 194 (1996) 208–214.
- [7] M. Duranti, A. Scarafoni, C. Gius, A. Negri, N. Faoro, Heat-induced synthesis and tunicamycin-sensitive secretion of the putative storage glycoprotein conglutin γ from mature lupin seeds, *Eur. J. Biochem.* 222 (1994) 387–393.
- [8] P. Restani, M. Duranti, P. Cerletti, P. Simonetti, Subunit composition

- tion of the seed globulins of *Lupinus albus*, *Phytochemistry* 20 (1981) 2077–2083.
- [9] M. Duranti, P. Restani, M. Poniatowska, P. Cerletti, The seed globulins of *Lupinus albus*, *Phytochemistry* 20 (1981) 2071–2075.
- [10] S. Kolivas, K.R. Gayler, Structure of the cDNA coding for conglutin γ , a sulphur-rich protein from *Lupinus angustifolius*, *Plant Mol. Biol.* 21 (1993) 397–401.
- [11] H. Kagawa, H. Hirano, Sequence of a cDNA encoding soybean basic 7S globulin, *Nucleic Acids Res.* 17 (1989) 8868.
- [12] S. Komatsu, O. Koshio, H. Hirano, Protein kinase activity and insulin-binding activity in plant basic 7S globulin, *Biosci. Biotechnol. Biochem.* 58 (1994) 1705–1706.
- [13] Y. Watanabe, S.F. Barbashov, S. Komatsu, A.M. Hemmings, M. Miyagi, A. Tsunasawa, H. Hirano, A peptide that stimulates phosphorylation of the plant insulin-binding protein: isolation, primary structure and cDNA cloning, *Eur. J. Biochem.* 224 (1994) 167–172.
- [14] S. Satoh, A. Sturm, T. Fujii, M.J. Chrispeels, cDNA cloning of an extracellular dermal glycoprotein of carrot and its expression in response to wounding, *Planta* 188 (1992) 432–438.
- [15] R.P.-T. Chung, G.M. Neumann, G.M. Polya, Purification and characterization of basic proteins with in vitro antifungal activity from seeds of cotton, *Gossypium hirsutum*, *Plant Sci.* 127 (1997) 1–16.
- [16] M. Duranti, C. Gius, F. Sessa, G. Vecchio, The saccharide chain of lupin seed conglutin γ is not responsible for the protection of the native protein from degradation by trypsin, but facilitate the refolding of the acid-treated protein to the resistant conformation, *Eur. J. Biochem.* 230 (1995) 886–891.
- [17] M. Duranti, F. Sessa, A. Scarafoni, T. Bellini, F. Dallochio, Thermal stabilities of lupin seed conglutin γ protomer and tetramer, *J. Agric. Food Chem.* 48 (2000) 1118–1123.
- [18] S.C. Ilgoutz, N. Knittel, J. Min Lin, S. Sterle, K.R. Gayler, Transcription of genes for conglutin γ and a leginsulin-like protein in narrow-leaved lupin, *Plant Mol. Biol.* 34 (1997) 613–627.
- [19] T.D. Vassilevskaia, C. Pinto Ricardo, C. Rodrigues-Pousada, Molecular cloning and sequencing analysis of a β -tubulin gene from *Lupinus albus*, *Plant Mol. Biol.* 22 (1993) 715–718.
- [20] W.R. Pearson, D.J. Lipman, Improved tools for biological sequence analysis, *Proc. Natl. Acad. Sci. USA* 85 (1988) 2444–2448.
- [21] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. USA* 74 (1977) 5463–5467.
- [22] A. Bairoch, P. Bucher, K. Hofmann, The Prosite database, its status in 1997, *Nucleic Acids Res.* 25 (1997) 217–221.
- [23] T.D. Vassilevskaia, E. Beckman, P. Jackson, C. Pinto Ricardo, C. Rodrigues-Pousada, Developmental expression and regulation by light of two closely related β -tubulin genes in *Lupinus albus*, *Plant Mol. Biol.* 32 (1996) 1185–1189.