



Identification of cassava root protein genes

C.R.B. DE SOUZA, L.J.C.B. CARVALHO, E.R.P. DE ALMEIDA, and
E.S. GANDER*

*Embrapa – Recursos Genéticos e Biotecnologia-CENARGEN, Parque Estação Biológica,
PqEB – Final Av. W5 Norte, 70770-900 Brasília-DF, Brazil (*author for correspondence;
egander@cenargen.embrapa.br)*

Abstract. The protein population of cassava root layers was characterized by SDS-PAGE and bidimensional polyacrylamide gel electrophoresis. SDS-Page revealed the presence of a protein population in the molecular weight range between 94 and 20 kDa. The expression pattern of these proteins was well-defined within the different layers. Partial protein sequence analyses and preliminary results on the layer-specific expression pattern obtained with Northern analyses are presented.

Key words: Cassava (*Manihot esculenta* L.), Genetic engineering, Nutrition, Proteins, Roots

Introduction

Cassava originated in South America and reached Africa and Asia during the 16th and 17th centuries. Now cassava is one of the most important food crops in the tropics and rates, with respect to furnished calories, in fourth place, after rice, maize and sugar cane [1].

In developing countries, cassava roots are very often the sole source of calories. Cassava roots contain about 85% starch and only about 1–2% protein [2]. The efforts of cassava breeders have been directed towards the identification of disease resistant varieties and those with longer leaf senescence periods. Reduced cyanide content and varieties with improved protein content have also been studied. These characteristics are, in addition to varieties with particular starch compositions, of industrial interest. While all these goals can be obtained with classical plant breeding techniques, recombinant DNA technology offers an alternate approach circumventing the natural barriers of sexual compatibility.

The introduction of transgenes has been reported for a great spectrum of agriculturally important crops. However, cassava, until recently, was not easily transformable. The first report on successful cassava transformation was published only in 1996 [3]. Besides the necessity to dominate transformation and regeneration of the host crops, another important requirement is

the availability of suitable transgenes coding for agriculturally and socio-economically important traits and tissue and/or development-specific promoters that guarantee the transgene's correct expression.

Seed-specific genes from most economically important mono- and dicots have been isolated and characterized extensively. For storage protein genes of underground organs, a considerable body of information is also available. As long ago as 1986, Rosahl et al. [4] reported the isolation and characterization of a patatin gene from *Solanum tuberosum* and recently a storage protein gene from taro (*Colocasia esculenta* L. Schott) has been characterized physicochemically and biologically [5, 6].

Although the storage root of cassava is a primary source of food, few studies on gene expression in this organ have been reported [7]. Recently, Cabral et al. [8] initiated a survey on cassava root formation and protein profiles of roots at various stages of development.

Therefore, this research was done to ascertain the physicochemical and biological characterization of the protein population of various, anatomically distinct root layers, to enable isolation of their genes and identification of specific promoters and enhancers able to drive efficient storage root specific expression.

Material and methods

Plant material

In this study, a cultivar, 'IAC 12-829', descendent from an industrial breeding program, was used. This cultivar was originally obtained from the Agronomy Institute Campinas, São Paulo, Brazil and was planted in CENARGEN's experimental plots.

Fresh roots from 10 month old plants with a root diameter of not less than 10 cm were harvested and iced, then immediately dissected into the morphological layers L1 to L5 [8] depicted in Figure 1. The different materials were stored in liquid N₂ until further use.

Protein extraction and analysis

Extraction of saline buffer-soluble proteins

Root tissues (200 mg) were ground in liquid N₂. Proteins were extracted with 0.5 ml of extraction buffer (50 mM Na₂PO₄ pH: 7.0, 250 mM NaCl, 2 mM PMSF, 0.13 mg/ml leupeptin and 1% β -mercaptoethanol).

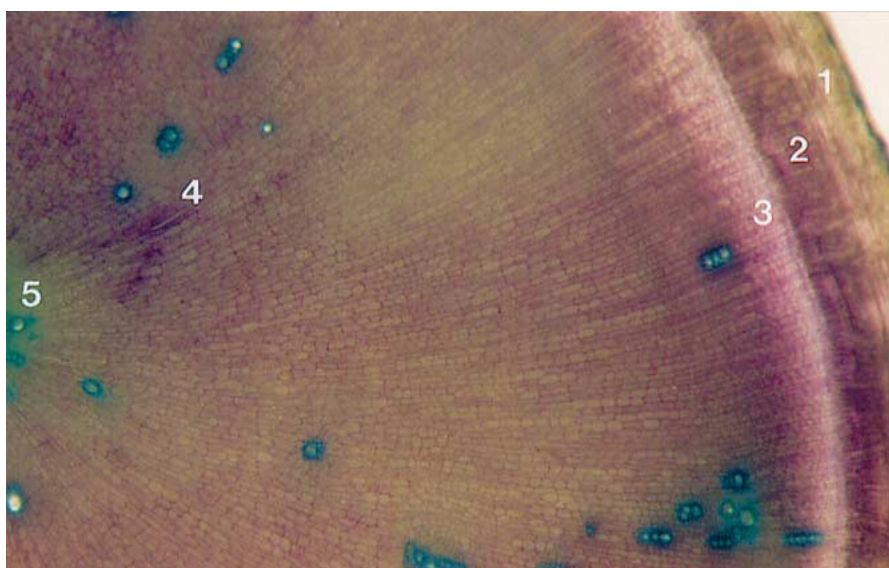


Figure 1. Transverse section through a cassava root. 1 to 5: morphologically different layers consisting predominantly of the following components: L1: peridermal tissue; L2: cambium and secondary phloem; L3: parenchymatic cells derived from secondary xylem during early stages of development; L4: fully differentiated secondary xylem; and L5: primary xylem.

Prolamin extraction

Alcohol-soluble proteins were obtained from cassava root layers by extraction with a buffer containing 55% isopropanol and 2% β -mercaptoethanol as described by Ottoboni et al. [9]. Protein extracts were loaded onto a 15% SDS-polyacrylamide gel and/or separated by two dimensional electrophoresis according to the methods of Laemmli [10] and O'Farrel [11], respectively.

Protein sequencing

After staining with Coomassie-Brilliant-Blue, the protein spots were excised from the bidimensional gels and destained. Sequence analyses were performed at the Harvard Microchemistry Facility by microcapillary reverse-phase HPLC nano-electrospray tandem mass spectrometry (μ LC/MS/MS) on a Finnigan LCQ quadrupole ion trap mass spectrometer.

Polymerase Chain Reaction (PCR) amplification

Based on the amino acid sequences obtained, degenerated primer sequences were designed to isolate cDNA sequences coding for the two major prolamins by PCR.

The following primer sequences were synthesized:

Pt2L4: 5' GAA/G ACI CCI AAA/G GAA/G GAA/G CC 3' and

Pt3L4: 5' GAA/G GCI GAA/G GAA/G GAA/G GCI GC 3'

PCR assays were performed according to Gibco BRL's instructions for the SuperScript Preamplification System Kit. For first strand cDNA synthesis, samples containing 3.0 μ g of total RNA from root layers L3 and L4 and cassava leaves (control) were used as templates. Amplification reactions were performed using the primers Pt2L4, Pt3L4 and oligo-dT. PCR products were cloned into the pGEMT vector (Promega) and sequenced as described.

DNA sequencing

Dideoxy-sequencing was performed in CENARGEN's sequencing lab on an automatic ABI-377 DNA sequencer (Perkin-Elmer) using universal forward and reverse M13 primers. The sequences corresponding to the PCR products were analyzed using the NCBI BLAST-program.

RNA blot hybridization

Total RNA was isolated from cassava roots, seeds and leaves as described by Jones et al. [12]. In short, 3 g of either roots, leaves or seeds were ground in liquid N₂, extracted with an SDS- and polyvinylpyrrolidone-containing neutral Tris-HCl-EDTA buffer, extracted with phenol-chloroform-isoamylalcohol followed by alcohol-precipitation from the aqueous phase and LiCl precipitation in order to get rid of the DNA. The RNAs were analyzed on formaldehyde-agarose gels and transferred to a Hybond-N membrane. The blot was probed with a radio-labeled 220 bp PCR product resulting from primer Pt2L4. Pre- and hybridization conditions were: 5X SSPE, 5X Denhardt's solution, 0.5% SDS, 0.04 mg/ml salmon sperm DNA at 65 °C overnight. The membrane was washed in 2X SSPE, 0.1% SDS and 1X SSPE, 0.1% SDS and 0.1X SSPE, 0.1% SDS at 65 °C for 20 minutes each.

Results

Protein patterns in the different root layers

In Figure 2A, the electrophoretic analyses of saline buffer-soluble extracts from the four root layers are shown. The molecular masses of the protein population in all layers analyzed spanned from 20 to about 100 kDa. However, although the qualitative pattern in the different layers remained largely the same, quantitative variations were obvious. Most strikingly, a 20 kDa protein, Pt1L4, seemed to increase from the peripheral layers towards the center while a group of about 10 proteins in the M_R range of 40 to 70 kDa exhibited

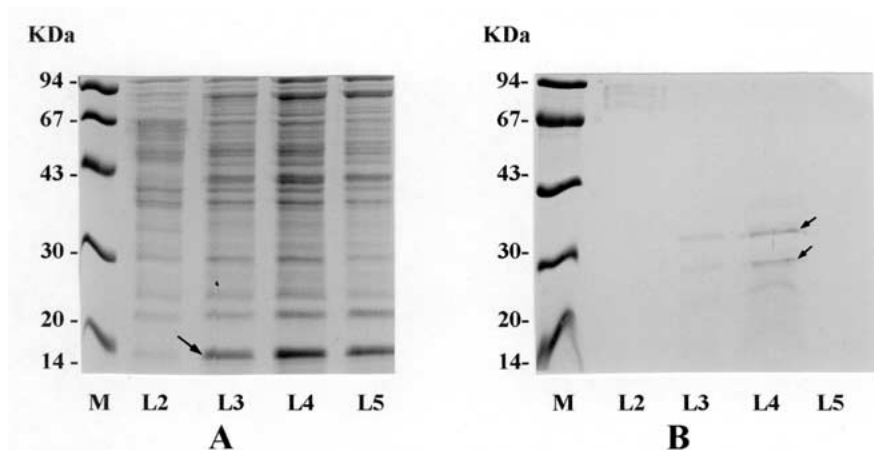


Figure 2. SDS-PAGE of saline soluble proteins (A) and prolamins (B) from different layers of cassava roots. M: protein molecular weight markers. The arrow in 'A' highlights the Pt1L4 protein; the arrows in 'B' point to the Pt2L4 and Pt3L4 prolamins used for sequencing.

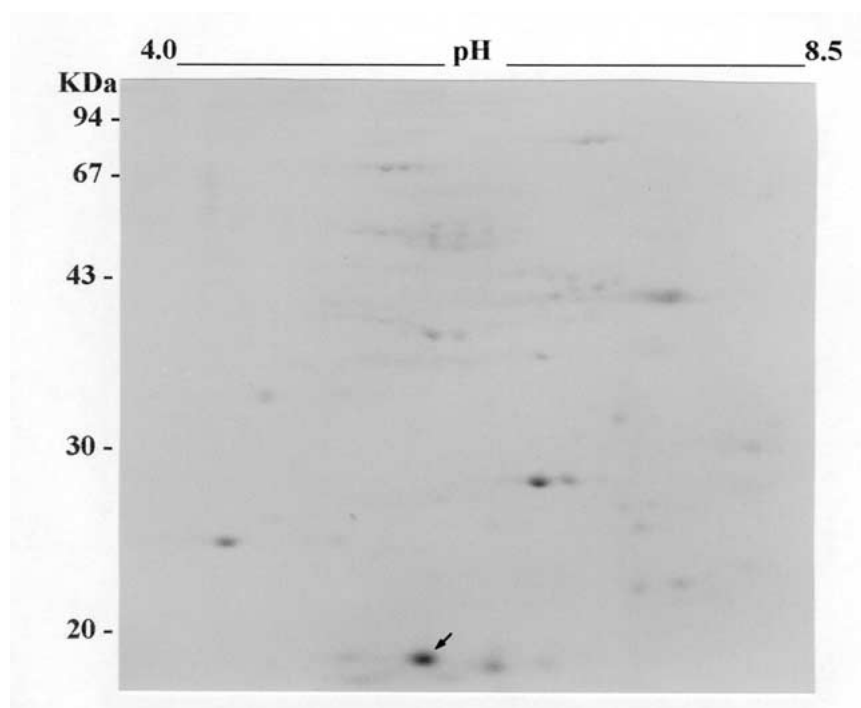


Figure 3. Two-dimensional electrophoresis of saline soluble proteins from layer L4 of cassava root. The arrow points to the Pt1L4 protein.

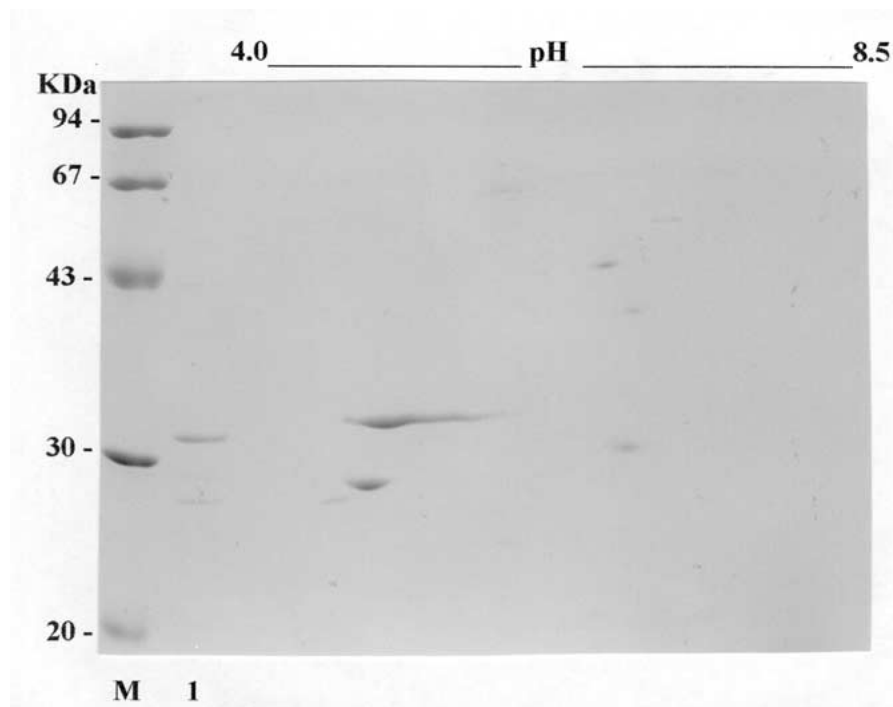


Figure 4. Two-dimensional electrophoresis of the prolamins extracted from layer 4 of cassava roots. 1: prolamins extracted from L4; M: protein molecular weight markers.

Pt1L4: K E T P E S H V F K.....A S M E N G V L T V T V P K E E I K K A E V K
 Pt2L4: K E T P K E E P A A E T V V
 Pt3L4: E A E E A A K E P A E A A E

Figure 5. Partial amino acid sequences of the proteins present in layer 4 of cassava roots.

the reverse pattern; they were abundant in the outer layers and very weakly represented in the center. Saline buffer-soluble extracts from layer L4 were analyzed by two dimensional gel electrophoresis. A representative result is shown in Figure 3. It showed that Pt1L4 had an isoelectric point of 6.5.

In Figure 2B the pattern of prolamins, i.e. the alcohol-soluble protein fraction according to Osborne [13], is shown. This class of proteins seemed to be present in substantially lower overall quantities and preferentially in layer 4. This layer contained two, more prominent prolamins with M_R 's of 32 and 28 kDa (designed Pt2L4 and Pt3L4, respectively) ranging in pI values from 5.0 to 6.0 (Figure 4).

Pisum sativum (P19243)

1 mslipsffsg rrsnvfdpfs ldvwdplkdf pfsnsspsas fprenpafvs trvdwketpe
 61 ahvfkadlpg lkkeevkvev eddrvlqisg ersvekedkn dewhrverss gkflrrfrlp
 121 enakmdkvka **smengvltvt vpkeeikkae vksieisg**

Glycine max (P05478)

1 mslipnffgg rrrnvfdpfs ldvwdpfdkf pfpntlssas fpefsrensa fvstrvdwke
 61 tpeahvfkad ipglkkeevk vqieddkvlq isgernveke dkndtwhrve rsgkfmrfr
 121 rlpnakveq **vkasmengvl tvtpkeevk** kpdvkaeis g

Oryza sativa (P27777)

1 mslvrrsnvf dpfslldwdp fdsvfrsvvp atsdndtaaf anaridw**ket peshvfkadl**
 61 pgvkkeevk eveegnlvi sgqrskedk kndkwhrver ssgkfmrfr lpenakvdqv
 121 kaglengvlt vtpkaevkk pevkaeisg

Helianthus annuus (P30693)

1 msiipsffts krsnifdpfs ldtwdpfqgi istepareta aivnaridw **ketpeahvika**
 61 dlpgmkkeev kvevedgrvl qisgercreq eekddtwhrv erssgkfrr frlpnakmd
 121 evkammengv ltvvvpkeee ekkpmvkaid isg

The bold characters indicate the homologies between the Pt1L4 sequence and pea-, soy-, rice and sunflower heat-shock proteins. Numbers in parenthesis refer to the Swissprot accessions.

Figure 6. Sequence comparison of Pt1L4 with heat shock class I proteins from various species.

Partial sequencing of proteins Pt1L4, Pt2L4, and Pt3L4

In order to construct molecular probes that could be used for the isolation of potentially interesting storage root genes, the partial amino acid sequences of the principal proteins were determined. In order to achieve this objective, the proteins were, after 2-D electrophoresis, excised, destained and proteolytic peptide fragments were sequenced at the Harvard Microchemistry Facility, Boston, USA. Figure 5 data show the amino acid sequence of peptide fragments from the three principal proteins of layer 4. The partial sequence of Pt1L4 was homologous to the 18 kDa small heat shock proteins from peas (*Pisum sativum*) [14], soybeans (*Glycine max*) [15], rice (*Oryza sativa*) [16] and sunflowers (*Helianthus annuus*) [17] (Figure 6). Through partial sequencing of the Pt2L4 and Pt3L4 proteins, about 14 amino acids residues of each protein were identified; no sequence in the NCBI databases matched these data.

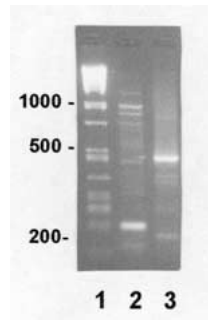


Figure 7. Analyses of PCR amplifications using primers Pt2L4 (lane 2) and Pt3L4 (lane 3) from cassava root. Lane 1: bp-markers.

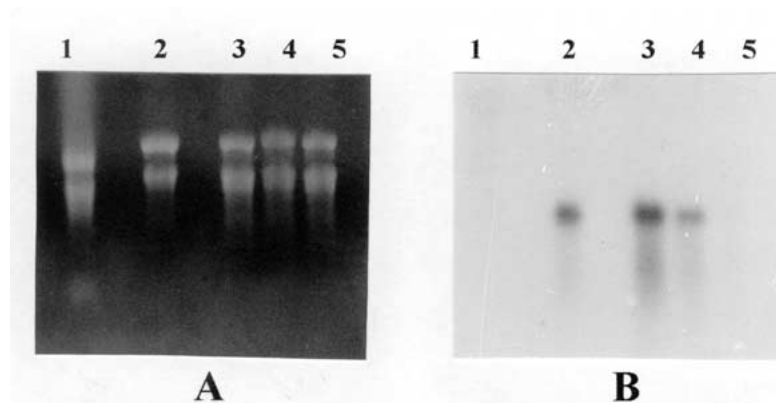


Figure 8. Pt2L4 gene expression in root layers L2, L3, and L4, seeds and leaves of cassava. Samples containing 5 μ g of total RNA were separated on a formaldehyde-agarose gel (A) and transferred to a Hybond-N membrane. The blot was probed with a radio-labeled 220 bp PCR product from Pt2L4(B). Lane 1: RNA from leaves; 2: RNA from seeds; 3 to 5: RNA from root layers L4, L3, and L2, respectively.

PCR amplification of a 220 bp DNA fragment of the Pt2L4 protein

Since heat shock genes are extremely well conserved and already well characterized, research focused on the nucleotide sequences of Pt2L4 and Pt3L4. In Figure 7, the amplification products obtained from RNA extracted from layer L4 with the primers described in the Materials and methods are shown. No amplification was detected from the PCR reaction conducted on leaf RNA (data not shown). Sequencing of the products revealed that only the 220 bp fragment obtained with primer Pt2L4 (lane 2 in Figure 7) corresponded to the partial amino acid sequence previously determined. Furthermore, no homologies with any other proteins available in data banks were detected.

ted. Sequencing analysis of the amplified fragment shown in lane 3 did not correspond to any of the peptide fragment sequences.

Organ-specific expression of the Pt2L4 gene

In Figure 8, a Northern analysis of cassava total RNAs from different organs, probed with the Pt2L4 220 bp fragment, is shown. A clear 1.3 kb hybridization signal is detected in layers L3 and L4 of the root and in RNA from the seeds. Typically the expression in roots increases in layers closer to the center; no hybridization signals could be detected in leaf RNA.

Conclusions

In cassava roots suited for commercialization, the protein concentration does not exceed 2%. The protein pattern in the root layers is fairly identical, with the exception of a 20 kDa protein, which is predominant in the inner root layers. A second predominant group of proteins in the molecular weight range of 40 to 70 kDa shows a reversed pattern of expression.

The comparison of partial amino acid sequences with profiles in data banks revealed that the 20 kDa protein might be a member of the small heat-shock family. It has been reported that this class of proteins is involved in water stress tolerance [18] and it might well be that in cassava, which is frequently planted in arid regions of Africa and South America, the 20 kDa protein is committed to a similar task.

Alcohol-soluble proteins were analyzed on 2-D gels. Two proteins with pI's between 5.5 and 6.0 and M_R 's of 28 and 32 kDa accounted for about 90% of the prolamin population. The expression pattern of the 32 kDa protein was investigated and, surprisingly, RNA's coding for this protein is only expressed in root layer 4 and in seeds but not in leaves. This might be an indication that this protein is a storage protein or could be specific for tissues common in roots and seeds. Future experiments will focus on the isolation of genomic sequences coding for these proteins and on the characterization of their regulatory elements.

Acknowledgments

CRBS and ESG hold a grant from the Brazilian National Research Council (CNPq) and are grateful for this organ's support. We thank the staff of CEN-ARGEN's sequencing laboratory for their efforts. This research was funded

in parts by EMBRAPA-CENARGEN and the Foundation for Science Support of the Federal District of Brasilia (FAP-DF).

References

1. Puonti-Kaerlas J (1998) Cassava biotechnology. *Biotech, Genet Enging Rev* 15: 329–364.
2. Cock JH (1985) Cassava: A basic energy source in the tropics. In Cock JH, Reyes JA (eds), *Cassava: Research, Production and Utilization*. UNDP CIAT.
3. Li HQ, Sautter C, Potrykus I, Puonti-Kaerlas J (1996) Genetic transformation of cassava (*Manihot esculenta* Crantz). *Nature Biotech* 14: 736–740.
4. Rosahl S, Schmidt R, Schell J, Willmitzer L (1986) Isolation and characterization of a gene from *Solanum tuberosum* encoding patatin, the major storage protein of potato tubers. *Mol Gen Genet* 203: 230–236.
5. de Castro LA, Carneiro M, de C Neshich D, de Paiva GR (1992) Spatial and temporal gene expression patterns occur during corm development. *Plant Cell* 4(12): 1549–1559.
6. Bezerra IC, de Castro LA, Neshich G, de Almeida ER, de Sá MF, Mello LV, Monte-Neshich DC (1995) A corm-specific gene encodes tarin, a major globulin of taro (*Colocasia esculenta* L. Schott). *Plant Mol Biol* 28(1): 137–144.
7. Salehuzzaman SNIM, Jacobsen E, Visser RGF (1992) Cloning, partial sequencing and expression of a cDNA coding for branching enzyme in cassava. *Plant Molec Biol* 20: 809–819.
8. Cabral GB, Carvalho LJC (2000) Cassava Biotechnology. IVth International Scientific Meeting – CBN. In Carvalho LJC, Thro AM, Vilarinhos AD (eds), *The Formation of Storage Root in Cassava*. Published by Embrapa-Recursos Genéticos e Biotecnologia. Brasília-DF, Brazil, pp 345–356.
9. Ottoboni LMM, Leite A, Targon MLPN, Silva MJ, Arruda P (1990) Heterogeneity of coix, maize and teosinte prolamins detected by isoelectric focusing. *Revista Brasileira de Genética* 13: 313–322.
10. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685.
11. O'Farrell PH (1975) High resolution two dimensional electrophoresis of proteins. *J Biol Chem* 250: 4007–4021.
12. Jones JDG, Dunsumuir P, Bedbrook J (1985) High level expression of introduced chimaeric genes in regenerated transformed plants. *EMBO J* 4: 2411–2418.
13. Osborne TB (1908) Our present knowledge of plant proteins. *Science* 28: 412–427.
14. Lauzon LM, Helm KW, Vierling E (1990) A cDNA clone from *Pisum sativum* encoding a low molecular weight heat shock protein. *Nucleic Acids Res* 18(14): 4274.
15. Raschke E, Baumann G, Schoffl F (1988) Nucleotide sequence analysis of soybean small heat shock protein genes belonging to two different multigene families. *J Mol Biol* 199(4): 549–557.
16. Tseng TS, Yeh KW, Yeh CH, Chang FC, Chen YM, Lin CY (1992) Two rice (*Oryza sativa*) full-length cDNA clones encoding low-molecular-weight heat-shock proteins. *Plant Mol Biol* 18(5): 963–965.
17. Almoguera C, Jordano J (1992) Developmental and environmental concurrent expression of sunflower dry-seed-stored low-molecular-weight heat-shock protein and Lea mRNAs. *Plant Mol Biol* 19(5): 781–792.

18. Coca MA, Almoguera C, Jordano J (1994) Expression of sunflower low-molecular-weight heat-shock protein during embryogenesis and persistence after germination: localization and possible function implications. *Plant Mol Biol* 25: 479–492.

