

Correct targeting of proinsulin in protein storage vacuoles of transgenic soybean seeds

N.B. Cunha¹, A.C.G. Araújo¹, A. Leite^{2†}, A.M. Murad¹, G.R. Vianna¹
and E.L. Rech¹

¹Embrapa Recursos Genéticos e Biotecnologia,
Parque Estação Biológica, Brasília, DF, Brasil

²Centro de Biologia Molecular e Engenharia Genética,
Universidade de Campinas, São Paulo, SP, Brasil

†In memoriam

Corresponding author: E.L. Rech

E-mail: rech@cenargen.embrapa.br

Genet. Mol. Res. 9 (2): 1163-1170 (2010)

Received February 24, 2010

Accepted March 26, 2010

Published June 22, 2010

DOI 10.4238/vol9-2gmr849

ABSTRACT. Soybean plants are promising bioreactors for the expression of biochemically complex proteins that cannot be produced in a safe and/or economically viable way in microorganisms, eukaryotic culture cells or secreted by transgenic animal glands. Soybeans present many desirable agronomic characteristics for high scale protein production, such as high productivity, short reproductive cycle, photoperiod sensitivity, and natural organs destined for protein accumulation in the seeds. The significant similarities between plant and human cells in terms of protein synthesis processes, folding, assembly, and post-translational processing are important for efficient accumulation of recombinant proteins. We obtained two transgenic lines using biolistics, incorporating the human proinsulin gene under control of the monocot tissue-specific promoter from sorghum γ -kafirin seed storage protein gene and the α -coixin cotyledonary vacuolar signal peptide from *Coix lacryma-jobi* (Poaceae). Transgenic plants expressed the proinsulin gene and accumulated the polypeptide in mature

seeds. Protein targeting to cotyledonary protein storage vacuoles was successfully achieved and confirmed with immunocytochemistry assays. The combination of different regulatory sequences was apparently responsible for high stability in protein accumulation, since human proinsulin was detected after seven years under room temperature storage conditions.

Key words: Human proinsulin; Protein storage vacuole; Protein stability; Molecular farming

INTRODUCTION

Human insulin, like many protein hormones, is synthesized as an 86-amino acid precursor prohormone of 9.5 kDa (proinsulin), followed by proteolytic cleavage to generate the active hormone (Wang and Tsou, 1991), which is a small molecule composed of 51 amino acids with a molecular weight of 5.8 kDa. Transgenic *Escherichia coli* cells were the first recombinant system utilized as an efficient bioreactor for insulin production (Goeddel et al., 1979); this was approved by the United States Department of Agriculture for commercial purposes under the name Humulin™, replacing a cattle-derived hormone.

Since then, there has been considerable progress in improving biosynthesis of recombinant insulin in transgenic organisms. Currently, many transgenic plants have been developed as efficient platforms for recombinant protein accumulation; they constitute one of the least expensive alternatives among all heterologous protein bioreactor systems. They can produce proteins at less than 10% of the cost of fermentation systems and are almost 50 times cheaper than prokaryotic bioreactors (Twyman et al., 2003).

Among the most important factors related to plant-associated cost reductions in platforms for protein expression, one can cite the short reproductive cycle of most crops, the elimination of expensive fermentation chambers, the possibility of high production scalability, well-known and established agriculture practices, and protein storage and compartmentalization in natural polypeptide storage organs, such as seeds and tubers (Kusnadi et al., 1997).

Another important factor that can be included among the advantages of vegetable bioreactors is the absence of prions, oncogenes and pathogens that affect humans, which results in considerably lower risks of product contamination when compared to recombinant proteins derived from microorganisms, mammal cells and transgenic-animal platforms (Giddings, 2001).

There is a high degree of similarity between human and higher plants in the steps of protein metabolism, such as synthesis, folding, assembly, and post-translational modifications in the endoplasmic reticulum and Golgi complex. This results in a potential for high-quality production of structurally complex proteins, such as complete and secretory antibodies (Schilberg et al., 2002; Ma et al., 2003).

Natural protein accumulation organelles in legume seeds can maximize polypeptide stability and increase yield accumulation (Fischer et al., 2004). The compartmentalization of large amounts of polypeptides in soybean seeds occurs in the cotyledonary protein storage vacuoles (PSVs) (Yoo and Chrispeels, 1980), constituting an excellent sub-cellular vehicle for long-term protein storage (Takaiwa et al., 2007). PSVs are endoplasmic reticulum-derived

cisterns that are specialized in stable protein storage and accumulation (Zheng et al., 1992); they are characterized by absence of amino peptidases, since these organelles do not undergo membrane fusion with lytic vacuoles, thus avoiding protein degradation.

The development of an efficient proinsulin production platform based on compartmentalization of insulin prohormone in seeds of transgenic soybean plants may not only guarantee lower risks in the therapeutic application of this recombinant protein but also has the potential to reduce production and commercialization costs. We utilized an expression cassette containing the proinsulin coding sequence under control of the monocot tissue-specific promoter from sorghum γ -kafirin seed storage protein gene (de Freitas et al., 1994) and the sequence of the cotyledonary PSV signal peptide from α -coixin (a *Coix lacryma-jobi* (Poaceae) seed protein). The plasmid (pRT-GK-PSC-PIh) containing the expression cassette was then introduced in the soybean plant genome via particle bombardment, with the objective of obtaining recombinant polypeptide in these sub-cellular compartments of seeds.

MATERIAL AND METHODS

The pRT-GK-PSC-PIh vector was introduced using biolistics in a co-transformation system with the *ahas* fragment at a 1:1 ratio to obtain transgenic soybeans. The vector contains the human proinsulin gene under control of the monocot tissue-specific promoter from sorghum γ -kafirin seed storage protein gene, the α -coixin cotyledonary PSV signal peptide from *C. lacryma-jobi* and the CaMV 35S terminator (Figure 1).

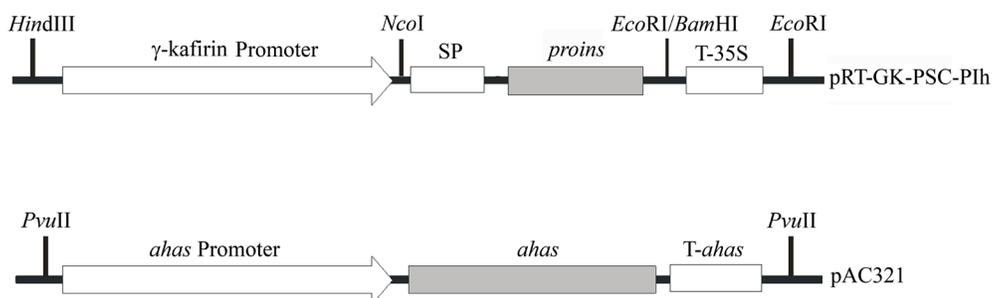


Figure 1. Schematic representation of the expression cassettes of the plasmid pRT-GK-PSC-PIh and the *ahas* fragment utilized for particle bombardment transformation of soybean embryos. The proinsulin gene fused to the α -coixin signal peptide is under control of the γ -kafirin promoter and the 3' region of the CaMV 35S gene. The *ahas* gene is controlled both by the *ahas* promoter and terminator.

The *ahas* fragment was utilized as a selective marker gene. The *ahas* cassette, containing the acetic-acid-hydroxy-synthase coding sequence, was previously isolated from *Arabidopsis thaliana* (Brassicaceae); this cassette confers tolerance to the herbicide imazapyr, under control of the *ahas* promoter and terminator. It was then removed from the 8672-bp plasmid vector pAC 321 (Rech et al., 2008), by digestion with *PvuII* at positions 530 and 6689 bp, and purified (Vianna et al., 2004) to generate a 6159-bp fragment (Figure 1).

Cultivar Conquista seed disinfestations, vegetable material preparation, and bombardment conditions were conducted according to Aragão et al. (2000) and Rech et al. (2008).

To identify the R_0 transgenic events and their progeny, genomic DNA was isolated

from foliar discs (Doyle and Doyle, 1987) and each polymerase chain reaction (PCR) was carried out in a 25- μ L volume with 10 mM Tris-HCl, pH 8.4, 50 mM KCl, 2 mM MgCl₂, 160 μ M of each dNTP, 200 nM of each primer (Cons 5 CCCTCGTACGCCTATGCAC and Cons 3 CTTATCTGGGAAGTACTCACAC), which anneals in the γ -kafirin promoter and CaMV 35S terminator, respectively, resulting in a 701-bp fragment, 2 U Taq polymerase (Invitrogen, USA), and 20 ng genomic plasmid. PCR mixes were covered with mineral oil, and the DNA was pre-denatured at 95°C for 5 min and amplified with 35 cycles of 95°C for 1 min, 55°C for 1 min, and 73°C for 1 min, with a final extension of 7 min at 72°C. Amplification reactions were resolved by electrophoresis on 1% agarose gels containing 0.5 μ g/mL ethidium bromide and visualized under UV light.

The R₀ plants were grown in a greenhouse and their seeds were harvested and stored at room temperature and examined seven years later to evaluate recombinant protein expression. Total soluble protein from soybean seeds was isolated by macerating the seeds in liquid N₂ and adding 300 μ L/mg fresh tissue in PBS buffer (10 mM Na₂HPO₄, 1.7 mM NaH₂PO₄, 140 mM NaCl and 2.7 mM KCl, pH 7.2) and quantified using the Bio-Rad protein assay kit™ (Bio-Rad, USA).

Samples (20 μ L) were incubated at 95°C for 10 min (1:1) with loading buffer (20% SDS, 0.5 M Tris-HCl, 10% β -mercaptoethanol, 20% glycerol, 0.04% bromophenol blue), and separated by SDS-PAGE (5% stacking gel and 15% resolving gel). Proteins were transferred to nitrocellulose membranes C-Extra™ (Hybond, USA) using a transblot SD-eleetrotransfer™ (Bio-Rad, USA) under constant 100 mA and 15 V for 50 min.

Transferred membranes were blocked with 50 mL 5% skim milk solution [1X TBS (0.02 M Tris base and 0.137 M NaCl, pH 7.6)] for 16 h at 4°C and incubated for 4 h at room temperature with rabbit polyclonal anti-human proinsulin at a dilution of 1:3000 - 4 ng/ μ L 10 mL blocking solution.

After washing three times with 1X PBS, blots were incubated in 10 mL blocking solution with goat anti-rabbit IgG conjugated with alkaline phosphatase at a dilution of 1:5000 (0.2 ng/ μ L). Blots were incubated with chemiluminescent substrate CSPD™ (Applied Biosystems) according to manufacturer instructions, placed in a film cassette and exposed to Kodak Standard™ film.

For immunolocalization of the recombinant proinsulin in cotyledons of transgenic plant lines, mature R₁ transgenic seeds were sliced (2 mm thick) and fixed (2% paraformaldehyde and 0.5% glutaraldehyde in 0.05 M cacodylate buffer, pH 7.2) for 4 h at 4°C. Fixed seed slices were washed three times in fixation buffer and dehydrated for 5 h in an ethanol series (30, 50, 70, 95, and 100%, 1 h each, at -20°C and partial vacuum).

Samples were infiltrated with increasing concentrations (30 to 100%) of ethanol-diluted LR White™ resin (SPI Supplies, USA) for 3 h, followed by 8-h incubation in pure LR White™ resin. Inclusion was performed by transferring the samples to 1.98% benzoyl peroxide in LR White™ resin and incubating at 4°C under UV light for 72 h. Ultra-thin sections (50 nm thick) were collected in 400-mesh nickel nets. The nets were incubated for 1 h at room temperature with 1X PBS-T (10 mM Na₂HPO₄, 1.7 mM KH₂PO₄, 140 mM NaCl, 2.7 mM KCl, 0.5% Tween 20, and 2% serum bovine albumin), and for 2 h with 5 ng/ μ L rabbit polyclonal anti-proinsulin diluted in 1X PBS-T. Samples were washed for 1 h in PBS, and incubated in gold conjugate protein A™ (SPI Supplies, USA) for 2 h at room temperature. After washing in PBS and drying for 24 h, samples were contrasted

with 1% uranile acetate in 0.1 M PBS, and observed under a Zeiss EM 900™ electron transmission microscope (Carl Zeiss, Germany).

RESULTS

In order to determine if recombinant proinsulin was correctly expressed in soybean seeds, the gene was cloned in a vector under control of the monocot tissue-specific promoter from sorghum γ -kafirin seed storage protein gene, the α -coixin cotyledonary protein body signal peptide from *C. lacryma-jobi* and the CaMV 35S terminator (Leite et al., 2000), generating the vector pRT-GK-PSC-PIh, which was used in co-transformation with the *ahas* fragment. Selection of the *in vitro* transformants was based on the expression of the *ahas* gene, which confers resistance to the herbicide imazapyr.

Two transgenic R₀ soybean lines containing transgenes, the proinsulin and the *ahas* gene were obtained; they were named 194 and 195. We also obtained transgenic plants that had only the *ahas* gene, but these were not considered here. The incorporation of the proinsulin gene in R₁ seeds of self-pollinated plants was detected by PCR (Figure 2). These seeds had been maintained under room temperature storage conditions since 2001.



Figure 2. Polymerase chain reaction showing the amplification of a 701-bp internal fragment of the proinsulin gene in R₁ soybean seeds. Lanes 1 and 2 = Transgenic seeds from lines 194 and 195, respectively; lane 3 = a non-transgenic seed; lane 4 = the water control, and lane 5 = the plasmid pRT-GK-PSC-PIh.

Expression of the proinsulin in R₁ seeds from transgenic lines 194 and 195 was confirmed by Western blot using total protein extracts obtained from PCR-positive soybean seeds (Figure 3).

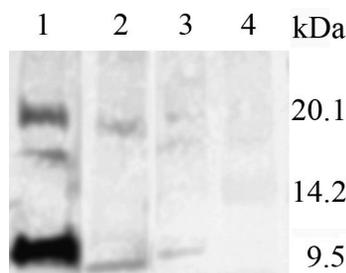


Figure 3. Western blot analysis of R₁ transgenic soybean seeds expressing recombinant proinsulin. Lane 1 = 40 ng standard proinsulin purified from *Escherichia coli*; lanes 2, 3 and 4 = total protein extracts (100 µg) from transgenic R₁ lines 194, 195 and a non-transgenic seed. All molecular weights were estimated using a Low Range™ Molecular Weight Marker (Sigma, USA).

The spatial localization of the heterologous proinsulin, along with sub-cellular targeting, promoted by the α -coixin signal peptide in mature soybean seeds, was verified by immunocytochemistry assays (Figure 4). Electron microscope images of 50-nm thick sections incubated with proinsulin-specific antibodies demonstrated that prohormone accumulation in soybean seeds of transgenic 194 and 195 lines was restricted to the PSV, as evidenced by detectable 20-nm gold particles. No significant gold particle accumulation was found in the apoplast, starch grains or the cytoplasm.

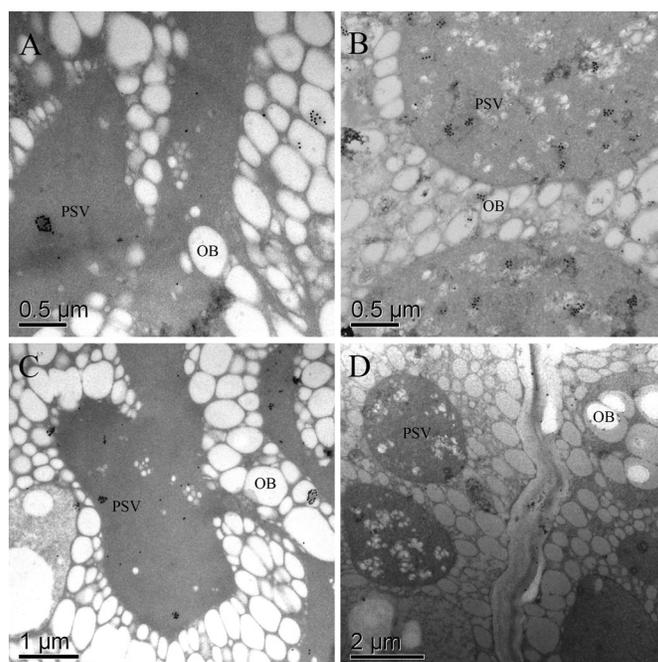


Figure 4. Ultra-thin soybean cotyledonary sections showing the sub-cellular accumulation of recombinant proinsulin in R_1 transgenic soybeans. PSV = protein storage vacuoles; OB = oil bodies. **A.** and **B.** Transgenic seeds from line 194. **C.** A transgenic seed from line 195. **D.** A non-transgenic seed.

DISCUSSION

In the genetic transformation experiments, we obtained two different transgenic soybean lines containing the proinsulin gene. We observed that transgene integration in the soybean genome was stable, since R_1 progenies from 194 and 195 line plants, generated in a greenhouse, inherited copies of the transgene.

Expression of proinsulin in transgenic mature seeds from the two different lines was detected by Western blot. Bands associated with recombinant proinsulin showed the expected molecular mass of approximately 9.5 kDa, when compared with purified protein produced in *E. coli*. Other bands of approximately 20 kDa were also detected both in the positive control and in the transgenic extracts, probably due to protein dimerization, resulting in doublet formation. The extract from transgenic line 194 also presented a slightly lower band than that shown by the positive control.

This is probably a consequence of improper processing of the signal peptide removal during protein targeting, which was also detected in the corresponding recombinant dimer (Hjerno, 2007). As the recombinant proteins were detected in the transgenic seeds and were absent in the non-transgenic extract, we concluded that the sorghum γ -kafirin tissue-specific promoter was efficient in driving proinsulin accumulation in soybean seeds.

Immunolocalization analysis of cotyledonary sections demonstrated that the α -coixin signal peptide properly targeted the nascent prohormone to dicot PSVs in transgenic seeds. Recombinant proinsulin molecules were detected all along the cistern lumen and also in the interior of seed oil bodies. No significant microparticle gold detection was observed in non-transgenic seeds.

This result corroborates the spatial and temporal transgene expression expected from the molecular strategy adopted in this study. We used a monocot tissue-specific seed γ -kafirin promoter and an α -coixin signal peptide for recombinant protein sub-cellular targeting, as was used for detecting human growth hormone expression in transgenic tobacco (Leite et al., 2000). The progenies of the two transgenic lines had been stored at room temperature since 2001, which demonstrates that proinsulin expression and accumulation in soybean seed PSV were efficient in promoting recombinant protein stability on a synthesis/accumulation level. We detected this human protein in seeds that had been stored at room temperature for seven years.

This molecular regulatory strategy for gene expression in seeds appears to minimize recombinant protein degradation and allow long-term and inexpensive storage. Seven years is far longer than has been reported for other seed expression systems (Fiedler and Conrad, 1995; Stöger et al., 2000; Larrick and Thomas, 2001; Ramirez et al., 2001). Other vegetable cell organelle systems have been used for recombinant protein accumulation (Staub et al., 2000). Notably, however, despite the relatively simple storage conditions used in this experiment, with storage at room temperature and absence of air humidity control, the recombinant proinsulin yields were not reduced to undetectable levels.

We conclude that transgenic soybean seeds can target proinsulin to sub-cellular PSVs and that prohormone accumulation was stable in mature seeds stored under relatively primitive conditions for at least seven years. We believe that higher yields of recombinant proteins could be obtained with this method if the proteins were extracted immediately after harvesting the seeds. We tried to germinate the seeds to prove this hypothesis, but the germination rate was zero because of the poor storage conditions.

Plant-produced recombinant proteins can greatly reduce production costs and are an attractive alternative to commonly used production methods based on microorganisms or cultured cells (Kusnadi et al., 1997; Ma et al., 2003; Twyman et al., 2003). Seed-based expression has some important advantages over leaf-produced heterologous proteins. The net protein content is much higher and the seeds can be stored for long periods of time without any detectable loss of activity (Stöger et al., 2000). Furthermore, the seed proteosome is relatively homogeneous when compared to the complex array of proteins found in leaves. Still, purification of a given protein from seeds can be cumbersome and inefficient.

Further studies on the expression of heterologous protein in soybean seeds will likely contribute to develop plant-based technologies for recombinant biopharmaceutical-protein production. Harnessing this potential will largely depend on further increasing the expression level as well as developing efficient purification methods. This will require developing technology and further studying the expression of recombinant protein in soybean seeds, which have potential as an effective and low-cost recombinant-protein production platform.

ACKNOWLEDGMENTS

The authors are grateful to Warley Almeida and Luis Carlos Lemos for technical assistance and to A.C.M.M. Gomes for scanning electron microscopy assistance. Research supported by Empresa Brasileira de Pesquisa Agropecuária (Embrapa) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERENCES

- Aragão FJL, Sarokin L, Vianna GR and Rech EL (2000). Selection of transgenic meristematic cells utilizing a herbicidal molecule results in the recovery of fertile transgenic soybean [*Glycine max* (L.) Merrill] plants at a high frequency. *Theor. Appl. Genet.* 101: 1-6.
- de Freitas FA, Yunes JA, da Silva MJ, Arruda P, et al. (1994). Structural characterization and promoter activity analysis of the gamma-kafirin gene from sorghum. *Mol. Gen. Genet.* 245: 177-186.
- Doyle JJ and Doyle JL (1987). A rapid DNA isolation procedure from small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11-15.
- Fiedler U and Conrad U (1995). High-level production and long-term storage of engineered antibodies in transgenic tobacco seeds. *Biotechnology* 13: 1090-1093.
- Fischer R, Stoger E, Schillberg S, Christou P, et al. (2004). Plant-based production of biopharmaceuticals. *Curr. Opin. Plant Biol.* 7: 152-158.
- Giddings G (2001). Transgenic plants as protein factories. *Curr. Opin. Biotechnol.* 12: 450-454.
- Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, et al. (1979). Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. U. S. A.* 76: 106-110.
- Hjerno K (2007). Protein identification by peptide mass fingerprinting. *Methods Mol. Biol.* 367: 61-75.
- Kusnadi AR, Nikolov ZL and Howard JA (1997). Production of recombinant proteins in transgenic plants: Practical considerations. *Biotechnol. Bioeng.* 56: 473-484.
- Larrick JW and Thomas DW (2001). Producing proteins in transgenic plants and animals. *Curr. Opin. Biotechnol.* 12: 411-418.
- Leite A, Kemper EL, Silva MJ, Luchesi AD, et al. (2000). Expression of correctly processed human growth hormone in seeds of transgenic tobacco plants. *Mol. Breed.* 6: 47-53.
- Ma JK, Drake PM and Christou P (2003). The production of recombinant pharmaceutical proteins in plants. *Nat. Rev. Genet.* 4: 794-805.
- Ramírez N, Oramas P, Ayala M, Rodríguez M, et al. (2001). Expression and long-term stability of a recombinant single-chain Fv antibody fragment in transgenic *Nicotiana tabacum* seeds. *Biotechnol. Lett.* 23: 47-49.
- Rech EL, Vianna GR and Aragao FJ (2008). High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. *Nat. Protoc.* 3: 410-418.
- Schillberg S, Emans N and Fischer R (2002). Antibody molecular farming in plants and plant cells. *Phytochem. Rev.* 1: 45-54.
- Staub JM, Garcia B, Graves J, Hajdukiewicz PT, et al. (2000). High-yield production of a human therapeutic protein in tobacco chloroplasts. *Nat. Biotechnol.* 18: 333-338.
- Stöger E, Vaquero C, Torres E, Sack M, et al. (2000). Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. *Plant Mol. Biol.* 42: 583-590.
- Takaiwa F, Takagi H, Hirose S and Wakasa Y (2007). Endosperm tissue is good production platform for artificial recombinant proteins in transgenic rice. *Plant Biotechnol. J.* 5: 84-92.
- Twyman RM, Stoger E, Schillberg S, Christou P, et al. (2003). Molecular farming in plants: host systems and expression technology. *Trends Biotechnol.* 21: 570-578.
- Vianna GR, Albino MMC, Dias BBA, de Mesquita Silva L, et al. (2004). Fragment DNA as vector for genetic transformation of bean (*Phaseolus vulgaris* L.). *Sci. Hortic.* 99: 371-378.
- Wang CC and Tsou CL (1991). The insulin A and B chains contain sufficient structural information to form the native molecule. *Trends Biochem. Sci.* 16: 279-281.
- Yoo BY and Chrispeels MJ (1980). The origin of protein bodies in developing soybean cotyledons: a proposal. *Protoplasma* 103: 201-204.
- Zheng Y, He M, Hao S and Huang B (1992). The ultrastructural evidence on the origin of protein bodies in the rough endoplasmic reticulum of developing cotyledons of soybean. *Ann. Bot.* 69: 377-383.