Expression and Characterisation of Recombinant Molecules in Transgenic Soybean

Nicolau B. da Cunha, Andre M. Murad, Giovanni R. Vianna, Cintia Coelho and Elibio L. Rech*

Embrapa Genetic Resources and Biotechnology. Synthetic Biology and Nanotechnology Group. Parque Estação Biológica, Brasilia, DF, 70770-917, Brazil

Abstract: Seeds are organs specialised in accumulating proteins, and they may provide a potential economically viable platform for the large-scale production and storage of many molecules for pharmaceutical and other productive sectors. Soybean [*Glycine max* (L.) Merrill] has a high seed protein content and represents an excellent source of abundant and cheap biomass. Under greenhouse conditions and a daily photoperiod of 23 h of light, the soybean plant's vegetative growth can be significantly extended by inducing more than a ten-fold increase in seed production when compared with plants cultivated under field conditions. Some factors involved in the production of different recombinant proteins in soybean seeds are discussed in this review. These include transgenic system, regulatory sequences and the use of Mass Spectrometry as a new tool for molecular characterisation of seed produced recombinant proteins. The important intrinsic characteristics and possibility of genetically engineering soybean seeds, using current advances in recombinant DNA technology including metabolic engineering and synthetic biology, should form the foundation for large-scale and more precise genome modification, making this crop an important candidate as bioreactor for production of recombinant molecules.

Keywords: Molecular farming, Transgenic soybean, *Glycine max,* Plant seed storage proteins, Protein storage vacuoles.

INTRODUCTION

Grain legumes, such as peanut, bean, soybean, lentil, pea and cowpea represent a significant contribution to human dietary requirements and, along with cereals, compose a foundation source of protein and carbohydrates to fulfill large-scale food demands worldwide [1]. Among the legume seeds, soybean occupies an important position in the food/feed industry due to its high protein (and oil) content and its use as one of the most important oilseed crops [2]. The nutritional content of soybean grain is composed of approximately 40% protein and 20% oil. Soybean grain also contains functional secondary metabolites that are of medical and industrial importance, such as isoflavones, lecithins, tocopherols and saponines [3, 4]. These characteristics are intensely exploited by seed processing companies, resulting in the extensive worldwide acreage that is devoted to soybean cultivation.

The United States, Brazil and Argentina are responsible for 80% of the global production of soybean, which reached approximately 255,000 tonnes in 2011. Soybean production has almost doubled since 1990, mostly due to expansion of the area planted with soybean and the introduction of different advanced technologies, such as genomics and genetic engineering (transgenic plants), for soybean production [5]. Indeed, soybean has exhibited a consistent increase in total yield; the maximum yields in the United States increased from 740 kg ha^{-1} to 2986 kg ha^{-1} in just eight decades of cultivation [6]. The economic importance of soybeans in industrialised countries has led this crop to become a primary target for recombinant DNA technology. As a result genetic engineering of soybean is very well advanced [7, 8]. The herbicide-tolerant Roundup Ready soybean-derived cultivars are the most striking example of the success of this technology. It is estimated that 93% of U.S. soybean cultivation is devoted to genetically engineered soybeans with herbicide tolerance/insect resistance traits [6].

In 2010, the commercial cultivation of the transgenic soybean Cultivance™ was approved in Brazil. This herbicide-tolerant cultivar was developed through a partnership between the leading chemical company BASF and the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). Cultivance™ can tolerate commercial doses of the imidazolinone-class herbicide Imazapyr (Arsenal™), reaches agronomical yield requirements and meets the biosafety standards for environmental, agricultural and food/feed safety [1, 9-11]. The increasing capacity to genetically engineer soybean through recombinant DNA technology and an increased demand for production of biopharmaceutical molecules, have led to our evaluation of soybean seeds as potential bioreactors to produce recombinant proteins of therapeutic and industrial interest. Recent estimates indicated that the recombinant protein market reached US$ 100 billion in annual sales in 2010, with a projected increase of 8% per year for the next decade [4].

The final production costs of recombinant proteins may be considered as one of the key factors in the proof of concept with regard to development of an economically viable plant-based production platform. In addition, aspects such as expression system, yields, purification and scale up capacities are also important when choosing a suitable host as bioreactor for recombinant protein production [12-14].

SOYBEANS AS REACTORS OF RECOMBINANT PROTEINS

Soybeans seeds have a high endogenous protein content and produce abundant and cheap biomass, which makes them a promising bioreactor system for the production of large amounts of recombinant proteins at a reduced cost [15]. Soybean seeds also provide a biochemical environment that is optimised for the long-term storage of these molecules [16-20]. Moreover, the stable accumulation of recombinant proteins within soybean seeds may be maximised by localising the nascent polypeptides in various subcellular organelles of the seed, such as the vacuoles [7].

Other important advantages in the use of soybeans for recombinant protein production include easy scalability; a high protein quality, which includes most of the post-translational modifications required for biological activity; a reduced risk of pathogen contamination; the development of marketable formulations that require no further protein purification; the well-established practices of cultivation and harvest; and known processing and manufacturing protocols [4, 21, 22].

Soybeans are a particularly interesting host for the production of recombinant protein synthesis because of their scale-up potential.

*Address correspondence to this author at the Embrapa Genetic Resources and Biotechnology. Synthetic Biology and Nanotechnology Group. Parque Estação Biológica, Brasilia, DF, 70770-917, Brazil;
Tel/Fax: +55-61-34484694; E-mail: elibio.rech@embrapa.br

Transgenic soybeans provide a unique production facility because the seed yield can be considerably maximised under greenhouse containment through photoperiod manipulation. Thus, this large-scale production of massive seed biomass may not only reduce the final production costs of a given therapeutic protein but also provide a large amount of material that is only limited by the greenhouse area that is available for planting [6, 7, 19, 20, 23].

Recombinant DNA technology can also be used to manipulate soybean pathways, through metabolic engineering and synthetic biology, for the biosynthesis and manufacture of valuable molecules [24]. By the redirection and modulation of flux through one or more metabolic pathways, an increased concentration of an endogenous soybean compound or the production of a novel product can be achieved, resulting in the accumulation of different types of compounds, such as proteins, oils, pigments and antibiotics [25, 26]. This technology, therefore, has the potential to have a major impact on the phenotypic improvement of crops that are used as renewable feedstocks for the production of biofuels, such as soybeans.

Crops with high endogenous protein content have become interesting potential candidates for molecular farming. Therefore, soybeans, which naturally accumulate proteins and have a higher endogenous protein content than most other cultivated seed crops are a promising host system for the expression and production of recombinant proteins. The protein content of soybean can reach up to 40% of dry seed weight, whereas maize and rice, wheat, barley and canola reach protein levels of 8-10%, 12%, 13% and 22%, respectively [7, 27]. An additional advantage of soybean plants arises from the intrinsic protein/biomass ratio among the crops under evaluation for the production of recombinant proteins. Significantly high quantities of soybean biomass, including leaves and seeds, can be generated in less than 4 months at lower costs than any other grain legume, such as pea, bean or peanut. It is estimated that the total protein yield per acre that can be obtained for soybean is double the amount for maize or rice [4, 21], which makes soybean an economically attractive bioreactor.

One important factor that is responsible for the reduction of soybean production costs is the absence of nitrogen fertilisation in the proper cultivation of the crop. Soybean roots can capture atmospheric nitrogen that is fixed by symbiotic soil bacteria of the genera *Rhizobium* and *Bradyrhizobium* and use it in the biosynthesis of aminated molecules, such as plant proteins and nucleic acid bases [6]. Through a simple regimen of soil fertilisation and an increase in daylight, the metabolic flux of input energy can be redirected to expand the biomass yield of soybean leaves and seeds. Under containment conditions, soybean grain yields can be considerably maximised by delaying their flowering through a prolonged photoperiod. A tenfold increase in seed production (approximately 1,000 seeds per plant) can be achieved by simply using a daily photoperiod of 23 hours of light, resulting in a higher yield of the recombinant protein of interest [19, 20, 23, 28, 29] (Fig. **1**).

Fig. (1). Adult transgenic soybean plants (5 feet tall) expressing the human growth hormone (hGH) after 4 months of cultivation in a greenhouse under a daily photoperiod of 23 h.

Large and structurally complex proteins, such as whole secretory antibodies, can be expressed in soybean seeds as high-quality biosynthetic products [22, 30-32]. The enzymatic machinery of seed cells can promote the correct folding and most of the post-translational modifications required for biological function at very little expense [33-36]. The typical yield of recombinant proteins that are synthesised by plant cells is low, normally ranging from 0.001 to 1% total soluble protein (TSP). However, the accumulation of proteins in soybean seeds can overcome this major challenge of plant expression systems. The typically poor yields are frequently the result of low levels of transcript accumulation; protein instability after translation [13] and/or the intrinsic molecular, chemical and physical structure of the target protein.

Some of the reasons for the low levels of protein accumulation in leaf tissues may be associated with the high endogenous proteolytic activity of hydrolases and the interference caused by phenolic compounds, which contribute to protein instability and present difficulties for downstream processing. In contrast to the leaves, soybean seeds provide a biochemical environment that is optimised for the stable accumulation of large amounts of protein for many months, and even years, without any detectable degradation [18, 30].

The use of tissue-specific promoters with strong transcriptional activities is crucial for increasing gene transcription and restricting the localisation of the recombinant protein to the seeds. The use of different regulatory elements in the design of expression cassettes can lead to a significant increase in the total soluble proteins of the seeds, which, in turn, makes the scalability of this system economically viable for the recovery and manufacturing of recombinant products [13, 14, 37]. Several well-characterised seed-specific promoters have been used with a wide range of crops to specifically direct gene expression to the seeds. The most commonly adopted promoters for recombinant expression in the seeds of monocots and dicots are Δ-zein, glutenin, glutelin, γ-kaffirin, β-phaseolin and the α-subunit of β-conglycinin genes respectively of maize, wheat, rice, sorghum, the common bean and soybean [18-20, 38-43].

Additionally, the genome of the soybean plant has been fully sequenced and has a total of 46,430 predicted protein-coding genes [44]. This genome represents an essential framework for vast new experimental information, such as tissue-specific expression, which holds promise for the development of novel genetic engineering strategies. The first recombinant protein obtained in soybeans was accumulated using the constitutive CaMV 35S promoter. When a functional humanized anti-HSV glycoprotein B was produced in soybeans it was observed that the immunoglobulins had the same efficacy in preventing vaginal HSV-2 infection in mice as the same antibodies secreted by mammalian suspension cells [32]. This finding prompted the discussion on the use of soybeans for the development of inexpensive healthcare products because the functional quality of the products is equivalent to that of compounds produced using other systems.

Since the findings of Zeitlin *et al*. [32] were reported, diverse recombinant proteins with different biochemical characteristics, molecular weights and structures, and commercial purposes have been produced in genetically engineered soybeans at yields ranging from 0.1 to 4% of total soluble proteins in the seeds (Table **1**)

The α-subunit of the β-conglycinin promoter is particularly interesting for the accumulation of recombinant proteins in soybean seeds because it modulates the expression of β-conglycinin, which is the most abundant protein complex present in soybean seeds, attending for 25 to 45% of the protein content in the grain [54]. This 156 kDa trimeric protein (with α, α` and β subunits) belongs to the globulins 7S family of storage proteins and shares structural domains and post-translational processing with seed-specific rice glutelins [55].

Table 1.   Examples of Important Recombinant Proteins Produced in Transgenic Soybean Plants

Protein	Original source	Intended use	Reference
IgG (HSV)	Mammalian cell culture	Therapeutic	[32]
Δ-zein	Maize	Nutrition	[45]
Human growth hormone (hGH)	Human	Therapeutic	[19, 46]
K99 fimbriae subunit antigen	*E. coli*	Vaccine	[47]
Fibroblast growth factor	Human	Therapeutic	[48]
Heat labile toxin (LT) B subunit	*E. coli*	Vaccine	[17]
FanC antigen	*E. coli*	Vaccine	[49, 50]
Anti-hypertensive factor	Synthetic	Therapeutic	[3]
Lysophosphatidic acid acyltransferase	Yeast	Nutrition	[51]
APase	Arabidopsis	Agronomic	[52]
Aspartate kinase	Soybean	Nutrition	[53]
Proinsulin	Human	Therapeutic	[18]
Coagulation factor IX (FIX)	Human	Therapeutic	[20]

β-Conglycinin functions as a storage protein and is mobilized at high rates after seed germination and during embryo nutrition, when the photosynthetic activity of green tissues has not yet started [53].

The expression of the three genes that encodes each subunit follows a tissue-specific pattern and is strictly regulated during soybean embryogenesis [56]. The transcription of the α` subunit coding sequence in the seeds tends to occur earlier that the other two, frequently 12 days after pollination, but there are no significant differences in the accumulation levels among the three subunits [3, 54].

The α-subunit of the β-conglycinin promoter is regulated by complex interactions between activation sequences rich in GC, localized approximately 270 bp upstream of the transcription initiation site, with trans activators synthesized during the maturation of the cotyledons [3, 56, 57]. The efficiency of this promoter was previously demonstrated by the accumulation of a maize methionine-rich delta-zein and a synthetic anti-hypertensive, denominated novokinin, both in soybean seeds [3, 45].

Our research group has used the β-conglycinin promoter to express and accumulate a number of functional recombinant proteins in soybean seeds [1, 19, 20, 23, 58]. Some of these proteins have major economic and social relevance because of their proven effectiveness against important human diseases, such as *diabetes mellitus*, growth and blood disorders, several types of cancer and AIDS [59-63]. These molecules can be grouped in four different classes according to their structure and function: human therapeutic proteins, recombinant antibodies, recombinant microbicides and immunogenic antigens.

Human Therapeutic Proteins

We expressed three functional pharmaceutical proteins of mammalian origin in soybean seeds: two hormones, human proinsulin (used in the treatment of *diabetes mellitus*) and human growth hormone (a therapeutic molecule for the treatment of hypo-pituitary paediatric dwarfism), and one blood product, human coagulation factor IX (FIX), used to attenuate the symptoms of the second most frequent blood coagulation disorder, type-B haemophilia, which affects 1 in 30,000 men [18-20].

Recombinant proinsulin, human growth hormone and coagulation factor IX were specifically accumulated in the seeds and were not detected in the leaves, flowers or roots of transgenic soybean plants [19, 20]. We could also detect the proteins in transgenic seeds after 7 years of storage at room temperature with little or undetectable levels of degradation after compartmentalization in the cotyledonary protein storage vacuoles (PSVs) [18-20, 23]. Similar results have previously been reported in the literature, and have demonstrated full or partial biological activities, corroborating our hypothesis that accumulation in dry seeds allows long-term storage at ambient temperature without noticeable proteolysis or loss of function [8, 64, 65].

The accumulation levels of recombinant human growth hormone and coagulation factor IX in transgenic soybeans reached up to 2.9% and 0.23% of the total soluble protein content of the seeds, respectively, which corresponds to a production potential of 9 g of the hormone and 0.8 g of the coagulation factor per kg of harvested seeds [19, 20]. These expression levels are higher than the typical yields observed in the recombinant production of pharmaceuticals in leaves and seeds, usually less than 0.1% of total soluble proteins [7, 8, 38, 66-68], corroborating the efficiency of the expression strategy, which focused on the strong transcription rates associated with the β-conglycinin promoter and the post-translational stability provided by subcellular compartmentalization in the vacuoles [69].

Recombinant Antibodies

Plants are currently recognised as promising platforms for the production of recombinant antibodies [30, 68]. Many examples of the biosynthesis of "plantibodies" in leaves and seeds of different plant species are reported in the literature, including secretory antibodies, and derivatives such as minibodies, diabodies, Fab fragments, single chain fragment variables (scFvs), antibody-fusion complexes and camelid heavy-chain antibodies [15, 38, 66, 70, 71].

Soybean seeds have already been used as vehicles for the production of a humanized IgG1 anti-herpes simplex virus (HSV)-2 glycoprotein B, which prevented vaginal HSV-2 transmission in

mice and has reached advanced stages of development as a drug for topical use [32].

Our group expressed the 29kDa fragment of the monoclonal antibody IgM AcM 83D4, known as scFvDIR83D4, in transgenic soybean seeds [58]. This scFv is highly efficient in the recognition of the Tn antigen, a mucin protein with a truncated pattern of O-linked oligosaccharides that is secreted by tumoral cells in mammary glands [72]. The Tn antigen presents clinical utility as a tumor marker in the early stages of breast cancer and other types of tumor, such as ovary and colon cancers, without detectable reactivity with the cell surface of healthy tissues [72-75].

Transgenic soybeans expressed recombinant scFvDIR83D4 as 0.93% of total soluble proteins of the seeds, presenting a production potential of even 2.8 g of the antibody per kg of dry seeds. The post-translational stability of the protein was similar to that observed for proinsulin, human growth hormone and coagulation factor IX (in preparation).

Another humanized scFv accumulated in soybean seeds using the α-subunit of β-conglycinin promoter was the scFv anti-CD18 (36 kDa) [58], a fragment of the monoclonal anti-CD18 (IgG1) that specifically recognizes the neutrophil β-2 CD18 integrin, blocking the vascular endothelial adhesion and preventing infiltration, post-ischemic reperfusion and other associated cardiac injuries [76-78].

Analysis of mature seeds by immunocytochemistry demonstrated that the recombinant scFv anti-CD18 molecules were detected after five years of room temperature storage, mostly inside the lumen of the protein storage vacuoles, but at levels lower than 0.1% of total soluble protein (in preparation).

Recombinant Microbicides

Microbicides are molecules that can be topically applied to the vagina or rectum prior to intercourse to prevent the sexual transmission of human immunodeficiency virus (HIV) [79]. Among the microbicide candidate molecules, three algae lectins have proven their high efficiency against the HIV: cyanovirin-N (CV-N) isolated from the blue alga *Nostoc ellipsosporum*) [80]; scytovirin, from cyanobacterium *Scytonema varium* [81]; and griffithsin, from the red alga *Griffithsia* sp. [82].

The molecular mechanism that inhibits viral infection is similar in both lectins [63]. It is based on specific binding of the microbicides to N-linked glycans that are displayed on the surface of the HIV envelope glycoproteins, primarily gp120, but also gp160 and gp41, thereby avoiding the interaction of these viral proteins with membrane receptors CCR5 and CXCR4 present in the plasmatic membrane, blocking the entry of the virus into susceptible cells [83-85].

Recombinant molecules of cyanovirin-N capable of selectively binding to carbohydrate viral moieties under *in vitro* conditions were constitutively expressed in hydroponic culture of tobacco (0.85% of total soluble proteins) and as a fusion protein with the monoclonal antibody b12 [86, 87].

Considerable amounts of functional griffithsin were obtained in tobacco plants using a transient expression TMV-based vector system (1 g/kg of fresh weight), providing the purification of 60 g of the microbicide from 226.5 kg of processed leaves [63].

Our group generated independent transgenic soybean lines expressing cyanovirin-N, griffithsin and the scytovirin microbicides in the seeds [58]. Seeds of transgenic line 11 were able to accumulate high levels of cyanovirin-N (3% of total soluble proteins), and HIV inhibition preliminary assays with purified fractions of microbicide have shown the efficacy of the microbicide in preventing the *in vitro* infection of cultured CD4 cells susceptible to HIV (in preparation).

Immunogenic Antigens

Some of the most promising molecules for the immunotherapy of cancer could also be accumulated in the seeds of transgenic soybean plants using the α-subunit of the β-conglycinin promoter [58]. These proteins, denominated NY-ESO-1, PLAC and GAGE, are immunogenic antigens whose main therapeutic property is to elicit a spontaneous antibody and T-cell response in patients with testicular and other forms of tumours [88-91].

The NY-ESO-1 antigen is encoded by the *CTAG1* gene and is frequently expressed in some types of testicular cancer [92], whereas the *PLAC1* gene encodes a putative protein that is highly expressed in placental, testicular and lung tumours [90], and the *GAGE-1* gene is abundantly expressed in a variety of tumours, including melanomas, sarcomas, head, neck and bladder tumours, as well as those associated with non-small-cell lung cancer [91].

All antigens produced in soybean seeds could be identified by NanoLC-MS^E (in preparation), and the observed that the NY-ESO-1 antigen presented the correct N-terminal sequence and mono-isotopic molecular masses and positions [93]. However, these compounds accumulated at expression levels below 1% of the total soluble proteins. Currently, experiments are being carried out to enhance the expression levels and the downstream processing protocols. In addition, solubility of these molecules is being investigated in a variety of extraction buffers to improve the separation of the recombinant proteins from the endogenous proteins.

SUBCELLULAR COMPARTMENTALISATION OF PROTEINS IN PROTEIN STORAGE VACUOLES OF SOYBEAN SEEDS

The total yield of a given protein reflects not only the rate of its biosynthesis but also its degradation. Therefore, protein stability after translation must be considered when adopting a recombinant expression strategy. The compartmentalisation of the recombinant protein into different organelles not only enables the avoidance of proteolysis and unintended post-translational modifications, such as the addition of plant glycans that are absent in humans, but also decisively contributes to the attainment of the production yields that are required for commercialisation [8, 70, 94, 95].

A number of different organelles and compartments, such as nuclei, mitochondria, plastids, oil bodies and vacuolar compartments, are available within plant cells for protein targeting [96]. Each of these organelles and compartments provides different post-translational processing machineries and levels of protein stability [97, 98]. The recombinant proteins can also be secreted into the apoplast or retained in the endoplasmic reticulum (ER) [38]. The most studied molecular mechanism of subcellular targeting involves the recognition of different signal peptides, which are N- or C-terminal protein segments with 20 to 30 amino acids, by their respective RNA-protein complexes, which are called "signal recognition particles" (SRPs) [99, 100]. This hybrid complex binds specifically to one or more of the positively charged amino acids that flank the core of the signal peptide because the nascent protein is processed in the ribosome [101]. After binding, the SRP transports the whole protein to the membrane surface of a specific organelle, where there is a second intermediary protein complex that is associated with selective membrane pores, which is denominated the "translocation complex" or "translocon". This recognises both the signal peptide and its specific SRP and promotes a series of complex structural modifications that enable pore opening and the internalisation of the protein. Following protein internalisation, the SRP is detached, the signal peptide is cleaved and the target protein is released into the organelle [102, 103]. The secretory pathway, with its complex endomembrane net composed of extensions of the ER and the Golgi complex, is a common destination for the effective accumulation of proteins [104]. Because most post-transla-

tional processing occurs in the endomembrane system, this system must be a stable environment for the full activity of plant glycosidases and chaperones and, therefore, provides a high level of post-translational stability for recombinant proteins that are retained in the ER or Golgi apparatus [13].

The localisation of proteins to the ER using the C-terminal tetrapeptide H/KDEL is a commonly adopted strategy that considerably increases the final yields of many proteins that are expressed in plants. This procedure permits the protein relocation from the Golgi apparatus to the ER lumen, which allows high levels of accumulation and prevents the addition of most undesirable plant glycans [34].

Seed vacuoles are the intracellular endpoints of the plant secretory pathway and, thus, the major sites of protein accumulation in seeds. There are two main classes of seed temporal organelless: the lytic vacuoles (LVs) and the protein storage vacuoles (PSVs) [105]. The LVs constitute the acidic compartments, which are the equivalent of mammalian lysosomes, and function as the main plant cell degradation site [106]. The protein storage vacuoles are cisternae that are derived from the ER and are specialised in the storage of the large amounts of protein required for seed germination [107]. The cotyledonary protein storage vacuoles are therefore the preferred target sites for the long-term storage of recombinant proteins. The lumen of these organelles exhibits minimal protein degradation since there are substantially low concentrations of amino peptidases and large quantities of protease inhibitors inside these organelles [17-20, 107].

The abundance of cotyledonary protein storage vacuoles in soybeans makes it an important temporal organelle for the accumulation of high endogenous protein levels in seeds. These organelles are available for protein storage during the early stages of seed development and reach their maximal concentration and subcellular volume in the final stages of grain maturation [107]. Among the protein storage vacuoles signal peptides applied for molecular targeting, the α-coixin signal peptide from the monocot *Coix lacryma-jobi* was used to direct β-glucuronidase and the human growth hormone to the protein storage vacuoles of tobacco seeds [108]; and the β-conglycinin signal peptide from soybeans targeted the accumulation of a Δ-zein of maize and a synthetic anti-hypertensive factor at high levels in soybean seeds [3, 45]. Our group has used the α-subunit regulatory sequences of the soybean β-conglycinin and α-coixin genes to accumulate recombinant proteins with different molecular weights and structures in the protein storage vacuoles of transgenic soybean seeds. The transgenic plants were obtained following a previously developed method of genetic transformation [10]. This method utilises direct DNA transfer through biolistics using the multiple shoot induction of selected herbicide-tolerant meristematic cells [9, 10].

The accumulation kinetics of recombinant the human growth hormone and the coagulation factor IX in the protein storage vacuoles exhibited an increase throughout the first 4 weeks of the seed development cycle and a sensible turnover from the 6th week onwards, corresponding to the increasing availability of these vacuoles in the initial steps of seed maturation [107] (Fig. **2**).

The sequencing of the N-terminus of the recombinant human growth hormone and the coagulation factor IX by nanoscale liquid chromatography coupled with mass spectrometry (NanoLC-MS^E) demonstrated that the signal peptides were correctly removed and that the PSV targeting occurred with high specificity. There were no unintended post-translational modifications or protein degradation detected indicating that their localisation in the vacuoles maintained the protein structure and ensured their biological activity [19, 20].

Ultrastructural immunocytochemistry evaluation indicated that cyanovirin-N was also correctly targeted to the protein storage

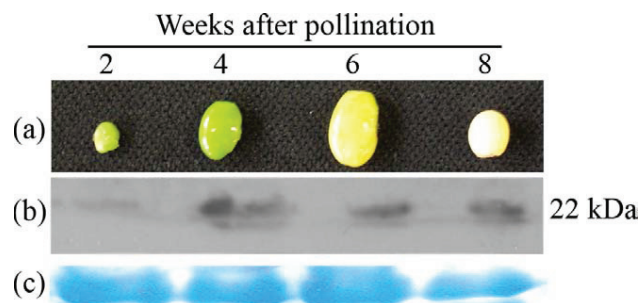vacuoles of mature soybean seeds (Fig. **3**), where it was identified and sequenced by NanoLC-MS^E.

Fig. (2). The human growth hormone expression kinetics. **(a)** Different phenological stages of R_1 soybean seeds at 2, 4, 6, and 8 weeks after pollination were used to measure the human growth hormone accumulation kinetics; **(b)** western blot analysis showing the accumulation of the human growth hormone protein during seed development; **(c)** SDS-PAGE loading controls of each extract of total soluble protein (~100 μg) were used to determine the uniformity of the electrophoresis samples.
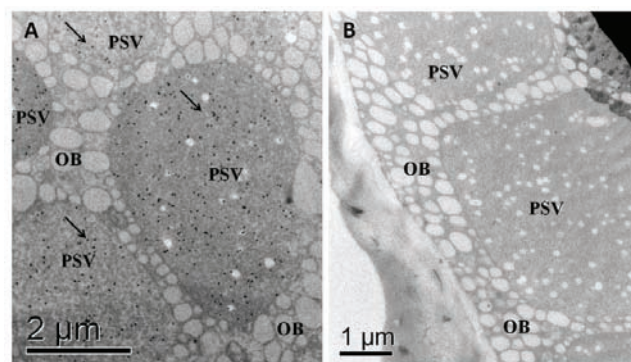

Fig. (3). Ultrastructural immunocytochemistry evaluation of recombinant cyanovirin-N in ultra-thin sections of the soybean cotyledon showing the subcellular accumulation of recombinant cyanovirin-N in the protein storage vacuoles of transgenic seeds (**A**) and non-transgenic controls (**B**). Arrows indicate the recombinant molecules, which are detected alongside the lumen of the protein storage vacuoles in the transgenic seeds. OB = oil bodies.

IDENTIFICATION, CHARACTERISATION AND QUANTIFICATION OF BIOMOLECULES BY MASS SPECTROMETRY IN SOYBEAN

Mass spectrometry (MS) has become a crucial and powerful tool in the pharmaceutical industry, particularly in the areas of new molecule discovery and molecule characterisation. As an analytical technique, MS allows for the measurement of mass, which is the inherent property of a molecule. In the last decade, MS has become the most potent instrument used, especially in the biological sciences, biochemistry and molecular biology, because it can identify and characterise biomolecules, such as proteins, by determining their structure and amino acid sequence. Through the use of MS techniques, it is now possible to identify proteins in complex samples from cells or tissues enabling the identification of the relationships between the physiology of an organism and its genome [93].

Utilizing nanoUPLC-MS^e analysis [93], we were able to detect and quantify different pharmaceutical recombinant proteins produced in soybean seed complex mixtures [19, 20]. (Fig. **4**) shows three recombinant molecules that have been quantified in transgenic soybean seeds: cyanovirin-N [80], human growth hormone [19] and NY-ESO-1 antigen [93].
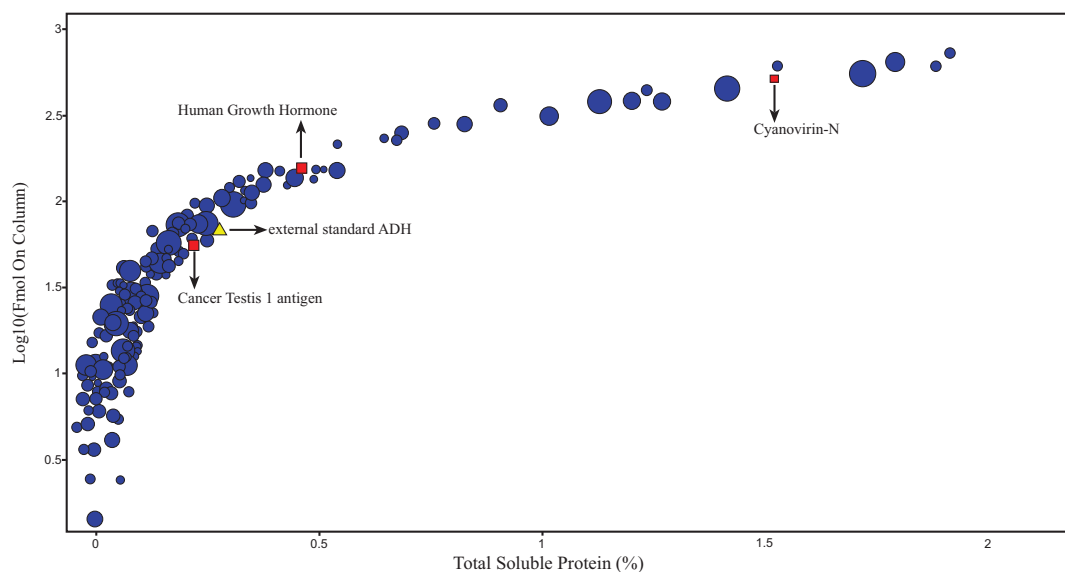
Fig. (4). Data-independent acquisition (MS^E) for the identification of the percentage of total soluble proteins. The femtomolar values that were obtained in the experiment have been logarithmically transformed. A relationship between the total femtomolar values of each protein identified can be performed to obtain the percentage. The blue circles represent the soybean seed proteins that were identified using the UniProt database sequences. The red squares represent three recombinant proteins (cyanovirin-N, human growth hormone and NY-ESO-1 antigen) expressed in the transgenic soybean seeds. The yellow triangle indicates the external standard, alcohol dehydrogenase (ADH), which was used for the quantification of the identified proteins. The size of the symbols represents the molecular weight of the identified protein.

In this example, the recombinant proteins were detected and quantified at the nanogram scale based on the stoichiometric ion intensity values of the minimum three prototypic peptides of the external standard, alcohol dehydrogenase (ADH), and the identified protein. A relationship between the total detected protein and the specific protein concentration was used, which allowed the percentage of the expressed protein in relation to the injected volume of total soluble proteins to be calculated. The dynamic range detection of this combination of techniques (nanoLC-ESI-Q-ToF-MS^E) resulted in a more accurate analysis of the soybean seed proteome and a more refined recombinant protein expression estimate than that obtained with traditional ELISA and western blot methods. The nanoLC-ESI-Q-ToF-MS^E method is also an important methodology for the detection of recombinant proteins with very low expression values or when other immunodetection methods are not possible due to the lack of specific antibodies, such as in the case of the NY-ESO-1 antigen [93]. In (Table **2**), the most abundant proteins in soybean seeds are shown.

The proteins shown in (Table **2**) are in line with the current literature on soybean seed proteins, which reports that glycinin and β-conglycinin are the most abundant storage proteins found in soybean seeds [109, 110]. Moreover, the nanoLC-MS method has been used for the optimised detection of proteins in laser-captured breast cancer tissues [111], the identification of low-abundance proteins in human plasma [112], the detection of differences in the expression of soybean plasma membrane proteins under osmotic stress [113], the regulation of stress in tomatoes that is induced by iron deficiency [114] and the detection of neuropeptides secreted in *Cancer borealis* [115], demonstrating the wide array of applications for this method. Furthermore, nanoLC-MS can be used for the post-translational characterisation of proteins, such as the identification of N-terminal peptide modifications in the chloroplast proteome [116], the analysis of human protein oxidations that result in function reduction/annulation [117] and the characterisation of the phosphorylation pattern of several phosphatase splice variants that are expressed in human cell lines [118, 119].

The current MS instruments not only enable the total characterisation of a protein, from sequence to post-translational modifica-

tions, but also provide insight into the protein tertiary structure, which is usually obtained through nuclear magnetic resonance and X-ray diffraction. Hybrid MS machines containing an ion mobility (IM-MS) device are capable of distinguishing separate molecules that have the same mass but differential spatial structures [120]. IM-MS experiments can be associated with protein structure prediction to obtain information on the conformation and stability of biomolecules [121] and for the characterisation of intact multi-protein complexes [122].

Aiming at future prospects in protein characterisation, the evolution of MS instruments not only allows the total characterisation of a protein, from sequence to post translational modification but now enters a novel area once only performed by nuclear magnet resonance and X-ray diffraction, giving a glimpse of protein tertiary structure. For example, the IM-MS of multi-charged equine apo-myoglobin (Fig. **5**) is used for calculation of the collision cross-section (CCS) of multi-charged ions in a IM-MS system [123].

An exponential curve can be established between the drift time and the mass-to-charge ratio and used to calculate the drift velocity at any point in the standard drift tubes, which can, in turn, be used to calculate CCS. For the equation details, please refer to Giles *et al.* [123]. The new MS instruments are focused on protein research but can also be used for other types of molecules, such as lipids, metabolites, drugs and carbohydrates. Consequently, the use of mass spectrometers in biochemistry and all other life sciences areas will continue to be expanded.

BIOPHARMACEUTICAL PRODUCTS AND THE NEW TECHNOLOGIES FOR GENE EXPRESSION MANIPULATION IN SOYBEANS

In order to achieve high yields of pharmaceuticals in soybean seeds it is necessary to overcome the usually low levels of recombinant biosynthesis in plant cells [7, 14, 95, 124]. Transcription activator-like effectors (TALENs) and zinc-finger nucleases (ZFNs) are important auxiliary technologies in plant metabolic engineering as they allow site-specific manipulation of genomes and the specific integration of target sequences in hot spot sites on the chromatin,

Table 2. **Identification and quantification of the most abundant proteins in soybean seeds and recombinant proteins accumulated in transgenic soybean seeds using tryptic digestion of total soluble protein and nanoLC-MS experiments. For a more detailed table, please refer to Murad *et al.* [93]**

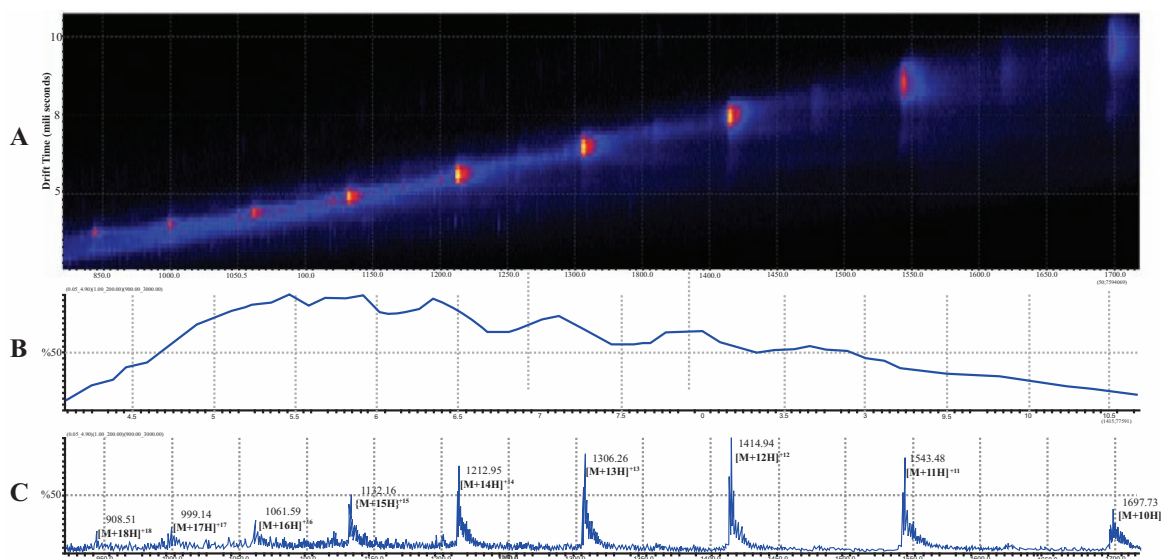| Accession	Description	Score	Average mass	Coverage (%)	fmol On Column	TSP%
CVN1_SOYBN	CYANOVIRIN-N	6686.03	11354.40	51.96	512.10	1.54
HGH1_SOYBN	human Growth hormone	3993.11	22357.18	47.12	145.18	0.43
CTG1B_HUMAN	Cancer testis antigen 1	4355.03	18277.56	30.56	59.29	0.17
Q3V5S6_SOYBN	Beta conglycinin alpha subunit OS Glycine max	78643.19	70569.34	53.72	2608.00	7.84
Q4LER6_SOYBN	Beta conglycinin alpha prime subunit OS Glycine max	51535.62	72513.18	61.03	1531.12	4.60
GLYG1_SOYBN	Glycinin G1 OS Glycine max GN GY1	63053.70	56333.70	49.9	1631.81	4.90
Q50JD8_SOYBN	Beta conglycinin beta subunit Fragment OS Glycine max	40740.82	48387.21	62.38	1720.33	5.17
GLYG2_SOYBN	Glycinin G2 OS Glycine max GN Gy2	44683.35	54961.11	60.21	1442.70	4.33
Q948X9_SOYBN	Beta conglycinin alpha subunit OS Glycine max	31870.91	72760.60	36.12	995.04	2.99
Q948Y0_SOYBN	Beta conglycinin alpha prime subunit OS Glycine max	37941.53	72423.18	61.03	988.28	2.97
C6T488_SOYBN	Putative uncharacterized protein OS Glycine max	74883.94	24403.62	49.77	2359.54	7.09
Q4LER5_SOYBN	Beta conglycinin alpha subunit Fragment OS Glycine max	79047.15	70348.96	53.81	795.86	2.39
LEC_SOYBN	Lectin OS Glycine max GN LE1	17424.92	30928.01	33.68	839.03	2.52
C6SWW4_SOYBN	Putative uncharacterized protein OS Glycine max	38844.21	22731.97	59.11	976.03	2.93
Q53WV6_SOYBN	Napin type 2S albumin 3 OS Glycine max	38269.60	19030.30	36.08	1134.86	3.41
Q70EM0_SOYBN	Dehydrin OS Glycine max GN lea D 11	21909.67	23787.83	60.62	630.36	1.89
IBBC2_SOYBN	Bowman Birk type proteinase inhibitor C II OS Glycine max	26035.52	9999.68	56.63	905.69	2.72
Q43709_SOYBN	Bowman Birk proteinase isoinhibitor D II OS Glycine max GN GB D II	27696.22	12351.73	56.86	512.99	1.54

Fig. (5). IM-MS of equine apomyoglobin (16,952 Da average mass). **A)** The distribution of multi-charged apomyoglobin peaks as a function of the drift time in milliseconds. The yellow and red areas represent the high and low signal intensities, respectively, of the multi-charged isotope distribution. **B)** Intensity signal (%) as a function of the drift time in milliseconds. **C)** Intensity signal (%) as a function of the mass-to-charge (m/z) ratio of each multi-charged peak. The drift time can be used for the calculation of the collision cross section (CCS) and can be correlated with pharmacological proteins, such as cyanovirin-N, NY-ESO-1 antigen and human growth hormone [93].

resulting in a sophisticated and precise control of gene expression [125, 126].

Several studies have already shown that the utilization of TALENs and zinc-finger nucleases in the expression strategies have become a choice for precise and efficient *in vivo* genome editing [127-130]. Another important aspect of the use of TALENS to achieve a high level of transgene expression in soybeans is that the DNA binding domain of TAL effectors can be linked to several different proteins, such as activator binding domains and chromatin remodeling proteins, therefore allowing genome activation and rapidly managing to overcome the issue of poorly accumulated proteins of interest [131, 132].

Our group used different plasmids carrying TALENs to obtain transgenic soybean lines accumulating human growth hormone, cyanovirin-N, griffithsin and NY-ESO-1 antigen coding sequences inserted in the genomic site of the native α-subunit of the β-conglycinin sequence. We believe that submitting the sequences of interest to control of the endogenous conglycinin tissue-specific promoter may significantly boost the recombinant accumulation of the pharmaceutical proteins [133, 134]. This expression strategy takes advantage of the high transcription rates associated with the α-subunit of the β-conglycinin promoter and simultaneously has the potential to provide proteome rebalancing in soybean seeds, with lower levels of β-conglycinin and enhanced content of foreign proteins [134-136].

Another important drawback associated with molecular pharming in seeds is the different post-translational processing promoted by mammalian and plant enzymatic machineries, notably the structurally diverse complex-type N-glycans added to the Golgi when the proteins bypass the secretory pathway [137, 138]. In plants, the N-acetylglucosamine of the core is substituted by an α1,3 fucose, instead of the mammalian α1,6-fucose; whilst the β-mannose of the core is substituted by a β1,2 xylose in place of a β1,4-N-acetylglucosamine present in mammals [138, 139]. Also, plant-derived recombinant proteins lack the sialic acid residues that are found in many native human glycoproteins [137, 140]. These differences can affect the distribution, activity and longevity of the recombinant proteins, and also result in allergenicity when administered to humans [38, 67, 141].

Gene silencing mediated by RNA interference (RNAi) allows the specific manipulation of seed glycosylation pathways in order to humanize the glycan structures of recombinant pharmaceutical glycoproteins [139]. This strategy emphasises the knock-down of genes that encode Golgi glycosyltransferases and the expression of newly synthesized mammalian galactosyl and sialyltransferases [140, 142].

We obtained different transgenic soybean lines using the RNAi technology aiming to knock down, simultaneously and specifically in the seeds, four endogenous glycosyltransferases genes: β(1,2)-xylosyltransferase, α(1,3)-fucosyltransferase, β(1,3)- galactosyl-transferase and α(1,4)-fucosyltransferase. The first plants obtained are currently being characterized, and the N-glycosylation of endogenous proteins accumulated in the seeds will be evaluated by nanoUPLC-MS^e.

CONCLUDING REMARKS

Seed-based bioreactors may constitute an important alternative for large-scale production of recombinant proteins. The results discussed in this review demonstrate that soybean seeds are capable of producing a different class of functional recombinant proteins correctly targeted to the protein storage vacuoles, at suitable levels to enter the purification processes. As endpoints of the plant secretory pathway, the protein storage vacuoles are the major sites for accumulation and long-term storage of proteins in seeds. The abundance of the temporal cotyledonary protein storage vacuoles in soybeans is important for accumulation of high endogenous protein levels.

The use of different coding sequences under control of the β-conglycinin seed-specific promoter and signal peptides was effective to direct recombinant-derived proteins to the protein storage vacuoles. Therefore, a better understanding of gene expression control is needed due to overcome the great challenge of using a soybean-based system for molecular farming. The maximization of the transgene expression levels is a key step to reach the commercial scale for viable production and subsequent recovery of recombinant proteins. In this regard, current advances in metabolic engineering and synthetic biology may lead to more precise genome editing and manipulation, which, in turn, should allow us to increase expression levels in plants.

We have also demonstrated here the use of nanoUPLC-MS^e analysis as an important tool for a comparative display of protein expression. It has the capability not only to provide in-depth characterisation of the protein content of seeds but also to qualify and quantify precisely the expression levels of a specific recombinant protein present in complex extracts with abundant endogenous storage proteins. Due to its high precision in characterizing protein structure, sequence and accumulation level, we believe that the utilization of nanoUPLC-MS^e is a crucial aspect to fulfill the required criteria for the commercialization of plant-derived pharmaceuticals.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Rech EL. Seeds, recombinant DNA and biodiversity. Seed Sci Res 2012; 22: 36-44.
[2]   Natarajan SS. Natural variability in abundance of prevalent soybean proteins. Regul Toxicol Pharmacol 2010; 58: S26-S29.
[3]   Yamada Y, Nishizawa K, Yokoo M, *et al.* Anti-hypertensive activity of genetically modified soybean seeds accumulating novokinin. Peptides 2008; 29: 331-37.
[4]   Bost K, Piller K. In: Sudaric A, Eds, Soybean - Molecular Aspects of Breeding. Rijeka, InTech Press. 2011; 3-19.
[5]   James C. A global overview of biotech (GM) crops: Adoption, impact and future prospects. GM Crops 2010; 1: 8-12.
[6]   Ainsworth EA, Yendrek CR, Skoneczka JA, Long SP. Accelerating yield potential in soybean: potential targets for biotechnological improvement. Plant Cell Environ 2012; 35: 38-52.
[7]   Boothe J, Nykiforuk C, Shen Y, *et al.* Seed-based expression systems for plant molecular farming. Plant Biotechnol J 2010; 8: 588-606.
[8]   Xu J, Ge X, Dolan MC. Towards high-yield production of pharmaceutical proteins with plant cell suspension cultures. Biotechnol Adv 2011; 29: 278-99.
[9]   Aragão FJL, Sarokin L, Vianna GR, Rech EL. Selection of transgenic meristematic cells utilizing a herbicidal molecule results in the recovery of fertile transgenic soybean [Glycine max (L.) Merril] plants at a high frequency. Theor Appl Genet 2000; 101: 1-6.
[10]   Rech EL, Vianna GR, Aragao FJL. High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. Nat Protocols 2008; 3: 410-18.
[11]   Vianna GR, Aragão FJL, Rech EL. A minimal DNA cassette as a vector for genetic transformation of soybean (Glycine max). Genet Mol Res 2011; 1: 382-290.

[12]   Karg SR, Kallio PT. The production of biopharmaceuticals in plant systems. Biotechnol Adv 2009; 27: 879-94.

[13]   Tremblay R, Wang D, Jevnikar AM, Ma S. Tobacco, a highly efficient green bioreactor for production of therapeutic proteins. Biotechnol Adv 2010; 28: 214-21.

[14]   Rybicki EP. Plant-made vaccines for humans and animals. Plant Biotechnol J 2010; 8: 620-37.

[15]   Obembe OO, Popoola JO, Leelavathi S, Reddy SV. Advances in plant molecular farming. Biotechnol Adv 2011; 29: 210-22.

[16]   Drakakaki G, Marcel S, Arcalis E, *et al.* The intracellular fate of a recombinant protein is tissue dependent. Plant Physiol 2006; 141: 578-86.

[17]   Moravec T, Schmidt MA, Herman EM, Woodford-Thomas T. Production of Escherichia coli heat labile toxin (LT) B subunit in soybean seed and analysis of its immunogenicity as an oral vaccine. Vaccine 2007; 25: 1647-57.

[18]   Cunha NB, Araújo ACG, Leite A, Murad AM, Vianna GR, Rech EL. Correct targeting of proinsulin in protein storage vacuoles of transgenic soybean seeds. Genet Mol Res 2010; 9: 1163 - 70.

[19]   Cunha NB, Murad AM, Cipriano TM, , *et al.* Expression of functional recombinant human growth hormone in transgenic soybean seeds. Transgenic Res 2011; 20: 811-26.

[20]   Cunha NB, Murad AM, Ramos GL, *et al.* Accumulation of functional recombinant human coagulation factor IX in transgenic soybean seeds. Transgenic Res 2011; 20: 841-55.

[21]   Paul M, Dolleweerd Cv, Drake PMW, *et al.* Molecular pharming: Future targets and aspirations. Hum Vaccin 2011; 7: 375-82.

[22]   Paul M, Ma JKC. Plant-made pharmaceuticals: Leading products and production platforms. Biotechnol Appl Biochem 2011; 58: 58-67.

[23]   Vianna GR, Cunha NB, Murad AM, Rech EL. Soybeans as bioreactors for biopharmaceuticals and industrial proteins. Genet Mol Res 2011; 10: 1733-52.

[24]   Capell T, Christou P. Progress in plant metabolic engineering. Curr Opin Biotechnol 2004; 15: 148-54.

[25]   Kinney AJ. Metabolic engineering in plants for human health and nutrition. Curr Opin Biotechnol 2006; 17: 130-38.

[26]   Ye VM, Bhatia SK. Metabolic engineering for the production of clinically important molecules: Omega-3 fatty acids, artemisinin, and taxol. Biotechnol J 2012; 7: 20-33.

[27]   Stöger E, Ma JKC, Fischer R, Christou P. Sowing the seeds of success: pharmaceutical proteins from plants. Curr Opin Biotechnol 2005; 16: 167-73.

[28]   Cavazzoni J, Volk T, Bugbee B, Dougher T. Phasic temperature and photoperiod control for soybean using a modified CROPGRO model. Life Support Biosph Sci: Int J of Earth Space 1999; 6: 273-8.

[29]   Kantolic AG, Slafer GA. Development and seed number in indeterminate soybean as affected by timing and duration of exposure to long photoperiods after flowering. Annals of Botany 2007; 99: 925-33.

[30]   Davies HM. Commercialization of whole-plant systems for bio-manufacturing of protein products: evolution and prospects. Plant Biotechnol J 2010; 8: 845-61.

[31]   De Muynck B, Navarre C, Boutry M. Production of antibodies in plants: status after twenty years. Plant Biotechnol J 2010; 8: 529-63.

[32]   Zeitlin L, Olmsted SS, Moench TR, *et al.* A humanized monoclonal antibody produced in transgenic plants for immunoprotection of the vagina against genital herpes. Nat Biotechnol 1998; 16: 1361-64.

[33]   Egelkrout E, Rajan V, Howard JA. Overproduction of recombinant proteins in plants. Plant Sci 2012; 184: 83-101.

[34]   Kawakatsu T, Takaiwa F. Cereal seed storage protein synthesis: fundamental processes for recombinant protein production in cereal grains. Plant Biotechnol J 2010; 8: 939-53.

[35]   Ghaderi D, Zhang M, Hurtado-Ziola N, Varki A. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. Biotechnol Genet Eng Rev 2012; 28: 147-75.

[36]   Pogue GP, Vojdani F, Palmer KE, *et al.* Production of pharmaceutical-grade recombinant aprotinin and a monoclonal antibody product using plant-based transient expression systems. Plant Biotechnol J 2010; 8: 638-54.

[37]   Weathers P, Towler M, Xu J. Bench to batch: advances in plant cell culture for producing useful products. Appl Microbiol Biotechnol 2010; 85: 1339-51.

[38]   Ma JKC, Drake PMW, Christou P. The production of recombinant pharmaceutical proteins in plants. Nat Rev Genet 2003; 4: 794-805.

[39]   Fiedler U, Conrad U. High-level production and long-term storage of engineered antibodies in transgenic tobacco seeds. Nat Biotech 1995; 13: 1090-93.

[40]   Tackaberry ES, Dudani AK, Prior F, *et al.* Development of bio-pharmaceuticals in plant expression systems: cloning, expression and immunological reactivity of human cytomegalovirus glycoprotein B (UL55) in seeds of transgenic tobacco. Vaccine 1999; 17: 3020-29.

[41]   Horvath H, Huang J, Wong O, *et al.* The production of recombinant proteins in transgenic barley grains. Proc Natl Acad Sci USA 2000; 97: 1914-19.

[42]   Stöger E, Vaquero C, Torres E, *et al.* Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. Plant Mol Biol 2000; 42: 583-90.

[43]   Jaeger GD, Scheffer S, Jacobs A, *et al.* Boosting heterologous protein production in transgenic dicotyledonous seeds using Phaseolus vulgaris regulatory sequences. Nat Biotech 2002; 20: 1265-68.

[44]   Schmutz J, Cannon SB, Schlueter J, *et al.* Genome sequence of the palaeopolyploid soybean. Nature 2010; 463: 178-83.

[45]   Kim W-S, Krishnan HB. Expression of an 11 kDa methionine-rich delta-zein in transgenic soybean results in the formation of two types of novel protein bodies in transitional cells situated between the vascular tissue and storage parenchyma cells. Plant Biotechnol J 2004; 2: 199-210.

[46]   Russell DA, Spatola LA, Dian T, *et al.* Host limits to accurate human growth hormone production in multiple plant systems. Biotechnol Bioeng 2005; 89: 775-82.

[47]   Piller K, Clemente T, Jun S, *et al.* Expression and immunogenicity of an Escherichia coli K99 fimbriae subunit antigen in soybean. Planta 2005; 222: 6-18.

[48]   Ding S-H, Huang L-Y, Wang Y-D, Sun H-C, Xiang ZH. High-level expression of basic fibroblast growth factor in transgenic soybean seeds and characterisation of its biological activity. Biotechnol Lett 2006; 28: 869-75.

[49]   Garg R, Tolbert M, Oakes J, Clemente T, Bost K, Piller K. Chloroplast targeting of FanC, the major antigenic subunit of Escherichia coli K99 fimbriae, in transgenic soybean. Plant Cell Rep 2007; 26: 1011-23.

[50]   Oakes JL, Bost KL, Piller KJ. Stability of a soybean seed-derived vaccine antigen following long-term storage, processing and transport in the absence of a cold chain. J Sci Food Agric 2009; 89: 2191-99.

[51]   Rao S, Hildebrand D. Changes in oil content of transgenic soybeans expressing the yeast SLC1 gene. Lipids 2009; 44: 945-51.

[52]   Wang X, Wang Y, Tian J, Lim BL, Yan X, Liao H. Overexpressing AtPAP15 enhances phosphorus efficiency in soybean. Plant Physiol 2009; 151: 233-40.

[53]   Qi Q, Huang J, Crowley J, *et al.* Metabolically engineered soybean seed with enhanced threonine levels: biochemical characterisation and seed-specific expression of lysine-insensitive variants of aspartate kinases from the enteric bacterium Xenorhabdus bovienii. Plant Biotechnol J 2010; 9: 193-204.

[54]   Wadahama H, Iwasaki K, Matsusaki M, , *et al.* Accumulation of β-conglycinin in soybean cotyledon through formation of disulfide bonds between α' and α subunits. Plant Physiol 2012; 158: 1395-405.

[55]   Katsube T, Kurisaka N, Ogawa M, *et al.* Accumulation of Soybean Glycinin and Its Assembly with the Glutelins in Rice 1999; 120: 1063-74.

[56]   Chen ZL, Schuler MA, Beachy RN. Functional analysis of regulatory elements in a plant embryo-specific gene. Proc Natl Acad Sci USA 1986; 83: 8560-64.

[57]   Doyle JJ, Schuler MA, Godette WD, Zenger V, Beachy RN, Slightom JL. The glycosylated seed storage proteins of Glycine max and Phaseolus vulgaris. Structural homologies of genes and proteins. J Biol Chem 1986; 261: 9228-38.

[58]   Vianna GR, Cunha NB, Rech EL. Expression and accumulation of heterologous molecules in the protein storage vacuoles of soybean seeds Protocol Exchange 1 February 2011

[59]   Wang CC, Tsou CL. The insulin A and B chains contain sufficient structural information to form the native molecule. Trends Biochem Sci 1991; 16: 279-81.

[60]   Kim T-G, Baek M-Y, Lee E-K, Kwon T-H, Yang M-S. Expression of human growth hormone in transgenic rice cell suspension culture. Plant Cell Rep 2008; 27: 885-91.

[61]   Goldenberg NA, Bombardier C, Hathaway WE, McFarland K, Jacobson L, Manco-Johnson MJ. Influence of factor IX on overall plasma coagulability and fibrinolytic potential as measured by global assay: monitoring in haemophilia B. Haemophilia 2008; 14: 68-77.

[62]   Caballero OL, Chen Y-T. Cancer/testis (CT) antigens: Potential targets for immunotherapy. Cancer Sci 2009; 100: 2014-21.

[63]   O'Keefe BR, Vojdani F, Buffa V, *et al*. Scaleable manufacture of HIV-1 entry inhibitor griffithsin and validation of its safety and efficacy as a topical microbicide component. Proceedings of the National Academy of Sciences 2009; 106: 6099-104.

[64]   Hood EE, Love R, Lane J, *et al*. Subcellular targeting is a key condition for high-level accumulation of cellulase protein in transgenic maize seed. Plant Biotechnol J 2007; 5: 709-19.

[65]   Miao CH, Thompson AR, Loeb K, Ye X. Long-Term and Therapeutic-Level Hepatic Gene Expression of Human Factor IX after Naked Plasmid Transfer *in vivo*. Mol Ther 2001; 3: 947-57.

[66]   Ma JKC, Chikwamba R, Sparrow P, Fischer R, Mahoney R, Twyman RM. Plant-derived pharmaceuticals - the road forward. Trends Plant Sci 2005; 10: 580-85.

[67]   Twyman RM, Stoger E, Schillberg S, Christou P, Fischer R. Molecular farming in plants: host systems and expression technology. Trends Biotechnol 2003; 21: 570-78.

[68]   Xu J, Dolan MC, Medrano G, Cramer CL, Weathers PJ. Green factory: Plants as bioproduction platforms for recombinant proteins. Biotechnol Adv 2012; 30: 1171-84.

[69]   Takaiwa F, Takagi H, Hirose S, Wakasa Y. Endosperm tissue is good production platform for artificial recombinant proteins in transgenic rice. Plant Biotechnol J 2007; 5: 84-92.

[70]   Sharma AK, Sharma MK. Plants as bioreactors: recent developments and emerging opportunities. Biotechnol Adv 2009; 27: 811-32.

[71]   Larrick JW, Yu L, Naftzger C, Jaiswal S, Wycoff K. Production of secretory IgA antibodies in plants. Biomol Eng 2001; 18: 87-94.

[72]   Petrakou E, Murray A, Price MR. Epitope Mapping of Anti-MUC1 Mucin Protein Core Monoclonal Antibodies. Tumor Biol 1998; 19: 21-29.

[73]   Charpin C, Pancino GF, Osinaga E, *et al*. Monoclonal antibody 83D4 immunoreactivity in human tissues: cellular distribution and microcytophotometric analysis of immunoprecipitates on tissue sections. Anticancer Res 1992; 12: 209-23.

[74]   Osinaga E, Bay S, Tello D, *et al*. Analysis of the fine specificity of Tn-binding proteins using synthetic glycopeptide epitopes and a biosensor based on surface plasmon resonance spectroscopy. FEBS Lett 2000; 469: 24-28.

[75]   Babino A, Oppezzo P, Bianco S, *et al*. Tn antigen is a precancerous biomarker in breast tissue and serum in n-nitrosomethylurea-induced rat mammary carcinogenesis. Int J Cancer 2000; 86: 753-59.

[76]   Arai M, Lefer DJ, Takehiko SO, Dipaula A, Aversano T, Becker LC. An anti-CD18 antibody limits infarct size and preserves left ventricular function in dogs with ischemia and 48-hour reperfusion. J Am Coll Cardiol 1996; 27: 1278-85.

[77]   Caldas C, Coelho V, Kalil J, Moro AM, Maranhão AQ, Brígido MM. Humanization of the anti-CD18 antibody 6.7: an unexpected effect of a framework residue in binding to antigen. Mol Immunol 2003; 39: 941-52.

[78]   Diacovo TG, Roth SJ, Buccola JM, Bainton DF, Springer TA. Neutrophil rolling, arrest, and transmigration across activated, surface-adherent platelets via sequential action of P-selectin and the beta 2-integrin CD11b/CD18 1996; 88: 146-57.

[79]   Balzarini J. Inhibition of HIV entry by carbohydrate-binding proteins. Antivir Res 2006; 71: 237-47.

[80]   Boyd MR, Gustafson KR, McMahon JB, *et al*. Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. Antimicrob Agents Chemother 1997; 41: 1521-30.

[81]   Bokesch HR, O'Keefe BR, McKee TC, *et al*. A potent novel anti-HIV protein from the cultured cyanobacterium Scytonema varium Biochemistry 2003; 42: 2578-84.

[82]   Mori T, O'Keefe BR, Sowder RC, *et al*. Isolation and Characterisation of Griffithsin, a Novel HIV-inactivating Protein, from the Red Alga Griffithsia sp. J Biol Chem 2005; 280: 9345-53.

[83]   Lotter-Stark HCT, Rybicki EP, Chikwamba RK. Plant made anti-HIV microbicides - A field of opportunity. Biotechnol Adv 2012; [Epub ahead of print].

[84]   Xiong C, O'Keefe BR, Botos I, Wlodawer A, McMahon JB. Overexpression and purification of scytovirin, a potent, novel anti-HIV protein from the cultured cyanobacterium Scytonema varium. Protein Expr Purif 2006; 46: 233-39.

[85]   Xiong S, Fan J, Kitazato K. The antiviral protein cyanovirin-N: the current state of its production and applications. Appl Microbiol Biotechnol 2010; 86: 805-12.

[86]   Sexton A, Drake PM, Mahmood N, Harman SJ, Shattock RJ, Ma JKC. Transgenic plant production of Cyanovirin-N, an HIV microbicide. FASEB J 2006; 20: 356-58.

[87]   Sexton A, Harman S, Shattock RJ, Ma JKC. Design, expression, and characterisation of a multivalent, combination HIV microbicide. FASEB J 2009; 23: 3590-600.

[88]   Chen YT, Scanlan MJ, Sahin U, *et al*. A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. proc Natl Acad Sci USA 1997; 94: 1914-18.

[89]   Cilensek ZM, Yehiely F, Kular RK, Deiss LP. A member of the GAGE family of tumor antigens is an anti-apoptotic gene that confers resistance to Fas/CD95/APO-1, interferon-g, taxol and g-irradiation. Cancer Biol Ther 2002; 1: 372-79.

[90]   Koslowski M, Sahin U, Mitnacht-Kraus R, Seitz G, Huber C, Türeci O. A placenta-specific gene ectopically activated in many human cancers is essentially involved in malignant cell processes. Cancer Res 2007; 67: 9528-34.

[91]   Nagel H, Laskawi R, Eiffert H, Schlott T. Analysis of the tumour suppressor genes, FHIT and WT-1, and the tumour rejection genes, BAGE, GAGE-1/2, HAGE, MAGE-1, and MAGE-3, in benign and malignant neoplasms of the salivary glands. Mol Pathol 2003; 56: 226-31.

[92]   Gnjatic S, Nishikawa H, Jungbluth AA, *et al*. In: Woude VGF, Fisher P, Tew K, Klein G, Eds, Adv Cancer Res. Waltham, Academic Press Press. 2006; 1-30.

[93]   Murad AM, Souza GHMF, Garcia JS, Rech EL. Detection and expression analysis of recombinant proteins in plant-derived complex mixtures using nanoUPLC-MSE. J Sep Sci 2011; 34: 2618-30.

[94]   Gleba Y, Klimyuk V, Marillonnet S. Magnifection - a new platform for expressing recombinant vaccines in plants. Vaccine 2005; 23: 2042-48.

[95]   Streatfield SJ. Approaches to achieve high-level heterologous protein production in plants. Plant Biotechnol J 2007; 5: 2-15.

[96]   Egea PF, Stroud RM, Walter P. Targeting proteins to membranes: structure of the signal recognition particle. Curr Opin Struct Biol 2005; 15: 213-20.

[97]   Okita TW, Rogers JC. Compartmentalization of proteins in the endomembrane system of plant cells. Annu Rev Plant Physiol Plant Mol Biol 1996; 47: 327-50.

[98]   Benchabane M, Goulet C, Rivard D, Faye L, Gomord V, Michaud D. Preventing unintended proteolysis in plant protein biofactories. Plant Biotechnol J 2008; 6: 633-48.

[99]   Shan S-o, Walter P. Co-translational protein targeting by the signal recognition particle. FEBS Lett 2005; 579: 921-26.

[100]   Nagai K, Oubridge C, Kuglstatter A, Menichelli E, Isel C, Jovine L. Structure, function and evolution of the signal recognition particle. EMBO J 2003; 22: 3479-85.

[101]   Bohnsack MT, Schleiff E. The evolution of protein targeting and translocation systems. Biochim Biophys Acta 2010; 1803: 1115-30.

[102]   Martoglio B, Dobberstein B. Signal sequences: more than just greasy peptides. Trends Cell Biol 1998; 8: 410-15.

[103]   Zimmermann R, Eyrisch S, Ahmad M, Helms V. Protein translocation across the ER membrane. Biochim Biophys Acta 2011; 1808: 912-24.

[104]   Hanton S, Matheson L, Chatre L, Rossi M, Brandizzi F. Post-Golgi protein traffic in the plant secretory pathway. Plant Cell Rep 2007; 26: 1431-38.

[105]   Herman EM, Larkins BA. Protein Storage Bodies and Vacuoles. Plant Cell 1999; 11: 601-13.

[106]   Jiang L, Phillips TE, Hamm CA, *et al*. The protein storage vacuole. J Cell Biol 2001; 155: 991-1002.

[107] Jolliffe NA, Craddock CP, Frigerio L. Pathways for protein transport to seed storage vacuoles. Biochem Soc Trans 2005; 33: 1016-18.

[108] Leite A, Kemper EL, da Silva MJ, *et al.* Expression of correctly processed human growth hormone in seeds of transgenic tobacco plants. Mol Breed 2000; 6: 47-53.

[109] Aghaei K, Ehsanpour A, Shah A, Komatsu S. Proteome analysis of soybean hypocotyl and root under salt stress. Amino Acids 2009; 36: 91-98.

[110] Shutov AD, Kakhovskaya IA, Bastrygina AS, Bulmaga VP, Horstmann C, Müntz K. Limited proteolysis of β-conglycinin and glycinin, the 7S and 11S storage globulins from soybean [Glycine Max (L.) Merr.]. Eur J Biochem 1996; 241: 221-28.

[111] Braakman RBH, Tilanus-Linthorst MMA, Liu NQ, *et al.* Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. J Proteomics 2012; 75: 2844-54.

[112] Cho SY, Lee EY, Lee JS, *et al.* Efficient prefractionation of low-abundance proteins in human plasma and construction of a two-dimensional map. Proteomics 2005; 5: 3386-96.

[113] Nouri M-Z, Komatsu S. Comparative analysis of soybean plasma membrane proteins under osmotic stress using gel-based and LC MS/MS-based proteomics approaches. Proteomics 2010; 10: 1930-45.

[114] Brumbarova T, Matros A, Mock H-P, Bauer P. A proteomic study showing differential regulation of stress, redox regulation and peroxidase proteins by iron supply and the transcription factor FER. Plant J 2008; 54: 321-34.

[115] Behrens HL, Chen R, Li L. Combining microdialysis, NanoLC-MS, and MALDI-TOF/TOF to detect neuropeptides secreted in the crab, Cancer borealis. Anal Chem 2008; 80: 6949-58.

[116] Zybailov B, Rutschow H, Friso G, *et al.* Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS One 2008; 3: 1994.

[117] Barnes S, Shonsey EM, Eliuk SM, *et al.* High-resolution mass spectrometry analysis of protein oxidations and resultant loss of function. Biochem Soc Trans 2008; 36: 1037-44.

[118] Bouché J-P, Froment C, Dozier C, *et al.* NanoLC-MS/MS analysis provides new insights into the phosphorylation pattern of Cdc25B *in vivo*: full overlap with sites of phosphorylation by Chk1 and Cdk1/cycB kinases *in vitro*. J Proteome Res 2008; 7: 1264-73.

[119] Unwin RD, Griffiths JR, Whetton AD. A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). Nat Protocols 2009; 4: 870-77.

[120] Kanu AB, Dwivedi P, Tam M, Matz L, Hill HH. Ion mobility–mass spectrometry. J Mass Spectrom 2008; 43: 1-22.

[121] Niñonuevo MR, Leary JA. Ion mobility mass spectrometry coupled with rapid protein threading predictor structure prediction and collision-induced dissociation for probing chemokine conformation and stability. Anal Chem 2012; 84: 3208-14.

[122] Bornschein R, Hyung S-J, Ruotolo B. Ion mobility-mass spectrometry reveals conformational changes in charge reduced multi-protein complexes. J Am Soc Mass Spectrom 2011; 22: 1690-98.

[123] Giles K, Wildgoose JL, Langridge DJ, Campuzano I. A method for direct measurement of ion mobilities using a travelling wave ion guide. Int J Mass Spectrom 2010; 298: 10-16.

[124] Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 2009; 27: 297-306.

[125] Bibikova M, Golic M, Golic KG, Carroll D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 2002; 161: 1169-75.

[126] de Souza N. Primer: genome editing with engineered nucleases. Nat Meth 2012; 9: 27.

[127] Christian M, Cermak T, Doyle EL, *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 2010; 186: 757-61.

[128] Morbitzer R, Römer P, Boch J, Lahaye T. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci USA 2010; 107: 21617-22.

[129] Miller JC, Tan S, Qiao G, *et al.* A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 2011; 29: 143-48.

[130] Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 2011; 29: 149-53.

[131] Li T, Huang S, Zhao X, *et al.* Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res 2011; 39: 6315-25.

[132] Mahfouz M, Li L, Piatek M, *et al.* Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. Plant Mol Biol 2012; 78: 311-21.

[133] Hagan ND, Upadhyaya N, Tabe LM, Higgins TJV. The redistribution of protein sulfur in transgenic rice expressing a gene for a foreign, sulfur-rich protein. The Plant J 2003; 34: 1-11.

[134] Goossens A, Van Montagu M, Angenon G. Co-introduction of an antisense gene for an endogenous seed storage protein can increase expression of a transgene in Arabidopsis thaliana seeds. FEBS Lett 1999; 456: 160-64.

[135] Schmidt MA, Herman EM. Proteome rebalancing in soybean seeds can be exploited to enhance foreign protein accumulation. Plant Biotechnol J 2008; 6: 832-42.

[136] Tada Y, Utsumi S, Takaiwa F. Foreign gene products can be enhanced by introduction into low storage protein mutants. Plant Biotechnol J 2003; 1: 411-22.

[137] Gomord V, Fitchette AC, Menu-Bouaouiche L, *et al.* Plant-specific glycosylation patterns in the context of therapeutic protein production. Plant Biotechnol J 2010; 8: 564-87.

[138] Gomord V, Sourrouille C, Fitchette AC, *et al.* Production and glycosylation of plant-made pharmaceuticals: the antibodies as a challenge. Plant Biotechnology J 2004; 2: 83-100.

[139] Gomord Vr, Chamberlain P, Jefferis R, Faye Lc. Biopharmaceutical production in plants: problems, solutions and opportunities. Trends Biotechnol 2005; 23: 559-65.

[140] Faye Lc, Boulaflous A, Benchabane M, Gomord Vr, Michaud D. Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. Vaccine 2005; 23: 1770-78.

[141] Paul M, Dolleweerd Cv, Drake PMW, *et al.* Molecular pharming: Future targets and aspirations. Human Vaccines & Immunotherapeutics 2011; 7: 375-82.

[142] Vézina L-P, Faye L, Lerouge P, *et al.* Transient co-expression for fast and high-yield production of antibodies with human-like N-glycans in plants. Plant Biotechnol J 2009; 7: 442-55.