



Review

Soybeans as bioreactors for biopharmaceuticals and industrial proteins

G.R. Vianna, N.B. Cunha, A.M. Murad and E.L. Rech

Laboratório de Transferência e Expressão de Genes,
Embrapa Recursos Genéticos e Biotecnologia,
Parque Estação Biológica, Brasília, DF, Brasil

Corresponding author: E.L. Rech
E-mail: rech@cenargen.embrapa.br

Genet. Mol. Res. 10 (3): 1733-1752 (2011)
Received May 6, 2011
Accepted August 4, 2011
Published August 18, 2011
DOI <http://dx.doi.org/10.4238/vol10-3gmr1476>

ABSTRACT. Plants present various advantages for the production of biomolecules, including low risk of contamination with prions, viruses and other pathogens, scalability, low production costs, and available agronomical systems. Plants are also versatile vehicles for the production of recombinant molecules because they allow protein expression in various organs, such as tubers and seeds, which naturally accumulate large amounts of protein. Among crop plants, soybean is an excellent protein producer. Soybean plants are also a good source of abundant and cheap biomass and can be cultivated under controlled greenhouse conditions. Under containment, the plant cycle can be manipulated and the final seed yield can be maximized for large-scale protein production within a small and controlled area. Exploitation of specific regulatory sequences capable of directing and accumulating recombinant proteins in protein storage vacuoles in soybean seeds, associated with recently developed biological research tools and purification systems, has great potential to accelerate preliminary characterization of plant-

derived biopharmaceuticals and industrial macromolecules. This is an important step in the development of genetically engineered products that are inexpensive and safe for medicinal, food and other uses.

Key words: Plant farming; Transgenic soybean; *Glycine max*; Plant seed storage proteins; Protein storage vacuoles

INTRODUCTION

Proteins are synthesized as part of the natural metabolism of all life forms and are central to the innumerable cellular processes that are as diverse as cell signalling and replication, reaction catalysis, immune responses, and the maintenance of the complex structural elements of many typical organelles. Native and recombinant proteins are widely used in research as laboratory reagents, in medicine as sophisticated diagnostic tools and in the food and nutrition industry. Recombinant proteins also play key roles in important sectors of the agriculture and the biopharmaceutical industries.

The commercial production of native proteins from their natural organism sources frequently requires difficult and expensive extraction processes, which can decisively contribute to drawbacks in terms of costs and production scalability. Furthermore, when a protein of interest is a biopharmaceutical product, the purification of the molecule from its natural source may represent a considerable risk for pathogen contamination and disease transmission.

Since the development of recombinant DNA technology in the early 1970s, the commercial production of recombinant proteins has traditionally relied on microbial fermentation systems and transgenic mammalian cell culture. Although these systems are highly efficient in terms of productivity, they present some disadvantages in terms of authenticity, safety and production costs. These important limitations have prompted the development of alternative production platforms in the past two decades based on the efficient scale-up, cost-effectiveness, high product quality, and low contamination risk presented by vegetable bioreactors, such as plant cell suspension, virus-infected plants and whole transgenic plants (Fischer et al., 2004).

Plant bioreactors are potentially the most economical expression systems for the large-scale production of many molecules of pharmaceutical or industrial interest. Since 1986, when the first relevant plant-made biopharmaceutical - human growth hormone (hGH) - was synthesized by transgenic tobacco plants, more than 100 different proteins have successfully been expressed in different transgenic crops; these crops primarily include tobacco suspension cells, cereal and legume seeds, oilseeds, potato tubers, leafy crops, such as lettuce and spinach, and edible fruits, such as tomatoes and bananas (Spök et al., 2008).

In 2010, approximately 30 plant-made pharmaceuticals (PMPs), including vaccines, antibodies, nutraceuticals, and therapeutic human proteins have entered the final stages of clinical trials prior to commercialization. Four PMPs have already been launched on the market, in addition to three previously commercialized proteins derived from maize, bovine trypsin, avidin, and β -glucuronidase, all of which have been utilized in research laboratories worldwide since 1998 (Table 1) (Ma et al., 2003; Obembe et al., 2011).

Table 1. Plant-made pharmaceuticals (PMPs) currently approved for commercialization.

Product	Trade name	Crop	Original organism	Commercial purpose	Company
Aprotinin	AproliZean	Maize	Cow	Research	ProdiGene
Aprotinin	Apronexin	Tobacco	Cow	Research	Kentucky BioProcessing, LLC (formerly manufactured by Large Scale Biology) (www.kbpllc.com/)
Avidin	Recombinant Avidin	Maize	Chicken	Research and diagnose reagent	ProdiGene
β -glucuronidase	NA	Maize	Bacteria	Research and diagnose reagent	ProdiGene
Trypsin	TrypZean	Maize	Cow	Research	ProdiGene
Lactoferrin	NA	Rice	Human	Research	Ventria Bioscience (www.ventria.com/)
Lysozyme	NA	Rice	Human	Research	Ventria Bioscience (www.ventria.com/)

NA = not applicable. Adapted from Spök and Karner, 2008.

Seed-based platforms are particularly interesting for accumulating recombinant proteins because they present a high endogenous protein content and provide a biochemical environment that is optimized for the long-term storage of these molecules. Moreover, stable accumulation of recombinant proteins within seeds may be maximized by localizing the nascent polypeptides to the endoplasmic reticulum (ER) lumen or targeting them to various subcellular organelles of the seed, such as the vacuoles. The vacuoles are typically involved in the stable compartmentalization of large amounts of protein that can be utilized as aminated compounds during seed germination and can provide seedlings with nutrients.

Various cereals (primarily maize, rice and barley) and grain legumes (pea and common bean) have been explored as hosts for diverse biopharmaceuticals. Among them, soybeans [*Glycine max* (L.) Merrill] have the highest seed protein content, mostly because of the abundance of soybean protein storage vacuoles (PSVs), which are seed-specific organelles specialized for protein accumulation that serve as the endpoints of the plant secretory pathway (Boothe et al., 2011).

Soybean plants also represent a good source of abundant and cheap biomass, especially if they are cultivated under controlled greenhouse conditions. Under containment, the plant cycle can be manipulated and the final seed yield can be maximized for large-scale protein production in a restricted area.

The utilization of dicot seed-specific promoters with strong transcriptional activities is crucial for the design of a successful strategy for recombinant protein expression and accumulation in soybean seeds. Furthermore, recent studies have demonstrated that the fusion of a PSV signal peptide upstream of the protein coding sequence is a promising approach to avoiding post-translational degradation and increasing the target protein yield (Cunha et al., 2011a,b).

EXPRESSION SYSTEMS FOR RECOMBINANT PROTEIN PRODUCTION

Although most genes can be expressed in many different recombinant systems, it is crucial to determine the platform that offers the most advantageous conditions for the synthesis of a particular protein (Daniell et al., 2001). To achieve high expression levels of biologically functional protein at the lowest possible cost, a variety of conditions must be simultane-

ously satisfied. The most important of these conditions is the most suitable combination of host organism and target protein (Streatfield, 2007).

There are a wide variety of available expression systems for recombinant protein production, each with intrinsic advantages and disadvantages. The nature and the particular properties of the different platforms can be explored in a complementary manner to increase the potential capacity to commercially produce a given protein (Kawakatsu and Takaiwa, 2010). The aspects of protein quality and functionality speed of production and product yield must be considered when choosing a suitable recombinant host for protein synthesis (Demain and Vaishnav, 2009).

Prokaryotic systems based on microbe fermentation are the workhorses of the biopharmaceutical and enzyme industries. Generally, these systems allow high product yields at considerably low costs and are characterized by rapid growth and expression and ease of cultivation. Other advantages include recent progress in the understanding of the major genetic processes of diverse microorganisms and the complete sequencing/characterization of many bacterial genomes, which permit the utilization of genetic tools to accurately modify microbial genomes, control promoter activity, modulate plasmid copy number, and manipulate many important metabolic processes and pathways (Demain and Vaishnav, 2009).

These microbe properties for recombinant protein production have allowed the replacement of many natural sources of native proteins and enabled the production of engineered enzymes with modified properties. Furthermore, this platform has been applied to the modification of fibers and lipids and for more efficient ethanol and paper production and for several other applications (Streatfield, 2007).

Escherichia coli is one of the earliest and most widely explored hosts for the production of recombinant proteins. Transgenic *E. coli* cells were primarily utilized for production of human insulin in the late 1970s (Goeddel et al., 1979) and were approved by the United States Department of Agriculture (USDA) for commercial purposes under the name Humulin™ to replace a less safe cattle-derived hormone.

In addition to facilitating high levels of expression for many different proteins, the *E. coli* system presents some important drawbacks, especially if the protein of interest must be glycosylated and terminally processed. Prokaryotic cells lack the machinery to perform complex protein folding and assembly or post-translational modifications, such as N- or O-glycosylation, phosphorylation, lipid acylation, disulfide bond formation, and removal of the N-terminal methionine. These factors complicate the synthesis of complex eukaryotic proteins, which in many cases can also be toxic to the bacterial cell (Karg and Kallio, 2009).

Other disadvantages of prokaryotic systems include toxicity that results from acetate accumulation in cultures with high bacterial cell density, as well as the production of proteins such as inclusion bodies, which are often non-functional, insoluble and require difficult refolding, and a high level of protein degradation mediated by endogenous proteases (Demain and Vaishnav, 2009).

One strategy for minimizing these problems is to secrete the recombinant proteins into the periplasmic space or into the culture medium in which microbes are grown, an approach which has been commonly and successfully applied to systems based on *E. coli* cells and species of the genus *Bacillus* (*B. megaterium*, *B. subtilis*, *B. licheniformis*, and *B. brevis*), *Ralstonia* (*R. eutropha*), *Pseudomonas* (*P. fluorescens*), and *Staphylococcus* (*S. carnosus*) (Demain and Vaishnav, 2009).

Yeasts have a long history in the fermentation industry and are frequently utilized to produce recombinant proteins that are not well-suited for synthesis by bacteria. Strains of the two most widely utilized yeast species in biotechnology, *Saccharomyces cerevisiae* and *Pichia*

pastoris, are able to perform many post-translational modifications of abundantly expressed proteins at lower costs compared to insect and mammalian cells. Yeasts are fast and easy to grow, are genetically well characterized and can secrete nascent proteins into the extracellular medium after the proper removal of a specific signal peptide (Boothe et al., 2010). However, the glycosylation patterns of many mammalian recombinant proteins produced by yeast, primarily antibodies, are often distinct from their native forms. Thus, O-linked oligosaccharides added to the protein surfaces by yeast enzymes present only mannose and lack mammalian sialylated O-linked chains. In addition, the activity of yeast-derived recombinant proteins can be considerably altered by over-glycosylation of N-linked sites, which can also result in undesirable immunogenicity of yeast-derived proteins (Gomord et al., 2010).

Foreign genes can also be stably integrated into the filamentous chromosomes of fungi as tandem repeats (as many as 100 copies have already been observed), and the transgenic fungal cells can secrete complete and fully active proteins with complex post-translational modifications and correct folding. In addition, some important proteins have already been expressed at high levels by species of the *Aspergillus*, *Acremonium* and *Chrysosporium* genera. The secreted yields of many recombinant proteins produced in fungal cells are lower compared to levels observed in other platforms, mostly because of transcription limitations, low stability of the corresponding mRNAs, and the abundance of naturally expressed fungal proteases (Demain and Vaishnav, 2009).

More complex post-translational modifications can be obtained using insect cells that are genetically transformed with circular double-stranded DNA baculovirus and by mammalian cell suspensions. In particular, immortalized Chinese hamster ovary cells (CHOs) have been employed to produce typical mammalian molecules, such as monoclonal antibodies, erythropoietin and tissue plasminogen activator (tPA) (Daniell et al., 2001; Ma et al., 2003). Both systems contain the complex enzymatic mechanisms that are capable of phosphorylating tyrosine, threonine and serine hydroxyl groups, adding fatty acids to lipoproteins, folding proteins and establishing disulfide bonds. In addition, protein secretion and targeting do not present a problem when proper secretion signals are utilized. In most cases, both insect and mammalian cells provide high expression levels and productivity, usually reaching 20–60 pg recombinant protein/cell/day (Demain and Vaishnav, 2009).

The major disadvantage of these recombinant platforms is the very high overall production cost associated with the media and manufacturing process. In the 1990s, one gram of recombinant hGH produced in mammalian cell culture was estimated to cost US\$35,000, which was approximately one hundred times more expensive than the same amount synthesized by transgenic *E. coli* (Demain and Vaishnav, 2009). Another major concern is the risk of contamination of biopharmaceuticals derived from mammalian suspension cells with human pathogens, which contributes to considerably elevated production costs by requiring drug purification and sanitary approval processes (Houdebine, 2009).

Whole transgenic animals have emerged as promising systems for the production of high-quality proteins that are secreted into the milk, blood, urine, seminal plasma, egg white, and silk worm fibers. The secretion of biopharmaceuticals into the milk of transgenic lactating females is possible when the genes of interest are controlled by tissue-specific promoters that restrict protein expression to cells of the mammary gland (Larrick and Thomas, 2001).

Considerably high levels of diverse biopharmaceuticals have been obtained using different animals, such as transgenic mice, goats, cows, pigs, sheep, rabbits, and even fish (Obembe et al., 2011). A single transgenic cow secreting 5 g hGH per liter of milk has the po-

tential to annually produce 25 kg of the recombinant protein (Salamone et al., 2006; Demain and Vaishnav, 2009), indicating that a herd of five animals could produce an amount sufficient to readily meet the annual demand of a large country, such as Brazil.

Some drugs that are important for human health have already been produced using animals. The length of time required to assess production levels, assuming a successful gestation rate, and the long periods between lactation cycles constitute important challenges for heterologous protein production in transgenic animals (Ma et al., 2003). Furthermore, the cost of maintenance of a single transgenic cow under Good Agricultural Practices is approximately US\$10,000 a year, resulting in prohibitively high overall costs for many of the biopharmaceuticals produced under these conditions (Demain and Vaishnav, 2009).

In addition, most transgenic animals present a low capacity to scale up production and have intrinsic risks for transmission of pathogenic viruses, prions and oncogenes carried in milk, urine, semen, and blood.

Over the past two decades, the alternative production of recombinant protein has turned to more effective and safer expression systems that are conducive to post-translational processing of these molecules at comparatively low production costs (Houdebine, 2009).

When dealing with large-scale production, transgenic plants are the most economical system for accurate expression of complex industrial proteins and recombinant drugs. Plant cells combine the potential for full-post-translational modification and correct folding with the simple growth requirements for water, minerals and light. Moreover, plants present a very high scalability when cultivated either in fields or greenhouses, and the technology for large-scale harvesting, processing and storage of whole plants and vegetable organs is already available for most cultivated crops, including soybeans.

The low risk of contamination with prions, viruses and other pathogens offers a major advantage of PMP. Because plant pathogenic organisms cannot cause human disease, this opens the possibility of exploiting plants and edible fruits, such as bananas and papaya, as good candidates for the production of vaccines and orally administered antigens (Ma et al., 2003; Boothe et al., 2010).

The major challenge of heterologous protein expression in plant systems is in overcoming the usually low yields of recombinant proteins that are synthesized by vegetable cells, normally ranging from 0.001 to 1% TSP (total soluble proteins), that results from low levels of transcript expression and recombinant protein instability (Tremblay et al., 2010).

Another disadvantage of plant systems lies in minor differences between the N-glycosylation patterns of plant versus mammalian-derived polypeptides. Such alterations can cause structural instability, inactivation, immunogenicity, and allergenicity of glycosylated plant-derived pharmaceuticals. Typical plant glycans, primarily β -1,2-xylose and α -1,3-fucose, are frequently added to the N- or C-terminal of proteins that pass through the Golgi along the secretory pathway (Sethuraman and Stadheim, 2006). Localizing proteins to the ER lumen and inactivating the N-glycosylation sites or knocking down Golgi glycosyltransferases can satisfactorily minimize these inconveniences (Gomord et al., 2004).

EXPLOITING THE ADVANTAGES THAT PLANTS OFFER FOR PROTEIN PRODUCTION

There are three strategies for recombinant protein production in plant-based systems:

1) the utilization of cell-cultured based systems that are equivalent to mammalian, microbial and insect cell systems; 2) the transient expression of foreign genes in plant tissues that are transformed by either agroinjection or by viral infection, and 3) the development of whole transgenic plants whose nuclear, mitochondrial or plastid genomes have been transformed with stably integrated transgenes (Kawakatsu and Takaiwa, 2010).

Under special cultivation conditions in simple synthetic liquid media using conventional bioreactors, undifferentiated plant cells can be separated and propagated as specialized factories for the production of secondary metabolites and, more recently, human-like therapeutic proteins (Xu et al., 2011).

The production of biopharmaceuticals in plant cell suspensions is cheaper, safer and faster than production in other eukaryotic suspension cell systems. Most stable productive lines can be generated within only a few months because they originate from cells that can double their concentration in just a single day and do not need to be regenerated and characterized as whole transgenic plants.

The system permits a high degree of containment and facilitates simple downstream processing with minimal heterogeneity of N-glycosylation due to the uniformity of the cells utilized in single cultures, which are usually composed of cells of *Arabidopsis thaliana*, tobacco, carrot, and rice (Obembe et al., 2011). Furthermore, no systemic post-transcriptional gene silencing is required due to the absence of plasmodesmata and the vascular system in undifferentiated culture cells (Xu et al., 2011).

A highly effective injectable vaccine against Newcastle disease virus, developed by Dow AgroSciences (<http://www.dowagro.com/>) for poultry immunization, can be obtained with 99% purity from tobacco suspension cells and, in 2007, became the first PMP approved for commercialization by the USDA. Another successful example is the Pfizer/Protalix (<http://www.protalix.com>) carrot cell-based glucocerebrosidase, a therapeutic protein used to treat a lysosomal storage disorder known as Gaucher's disease, that is currently in the pre-clinical pipeline and has received a Phase III Fast Track Designation from the Food and Drug Administration in 2009 (Xu et al., 2011).

Transient expression systems are very useful for research and are now being routinely used for the production of many valuable proteins at very low cost. These systems allow high throughput production and straightforward manipulation, permitting the rapid validation of expression constructs and the production of large amounts of recombinant protein within a few weeks. As a direct consequence, the protein yields from transient expression in plants are normally 10-fold higher than yields observed in other recombinant plant systems (Gleba et al., 2005).

Thus, some molecules that have already been successfully obtained from infected tobacco leaves are hGH, seven single-chain antibodies, alpha interferon, and some bacterial antigens, which present yields ranging from 1.2 to 5.4 g/kg fresh leaf biomass (Gleba et al., 2005). In addition, LargeScale Biology Corporation is currently evaluating the commercial production of an idiotype vaccine against B-cell non-Hodgkin's lymphoma using Magnifecation technology (Tremblay et al., 2010).

However, transiently expressed proteins must be processed immediately after harvest to prevent tissue degradation and protein turnover. The induction of expression is also technically demanding because agrobacterium inoculation must be performed mechanically or sprayed, and the efficacy of viral strains to promote proper infection must be established (Ma et al., 2003; Stoger et al., 2005; Obembe et al., 2011).

These limitations can be minimized when the host system provides a steady expression level of the transgenes mediated by stable nuclear or plastid transformation.

The chloroplast genome is relatively simple and contains smaller molecules of double-stranded circular DNA (approximately 200 kb) compared to the nuclear genome. The genome is presented in as many as 100 copies in a single plastid. A mature leaf contains up to 100 plastids, resulting in approximately 10,000 genome copies per cell. This characteristic allows chloroplasts to sustain a very high number of functional transgene copies delivered using biolistics and integrated into the genome by direct homologous recombination into spacer regions of a specific locus. The high transgene copy number frequently provides recombinant expression levels exceeding 25% TSP. Recently, the outstanding level of 70% TSP was achieved in transplastomic tobacco plants producing a functional antibiotic against bacteria from the genus *Streptococcus* (the highest yield of a recombinant protein expressed in a vegetable system) (Gleba et al., 2005).

A desirable characteristic is that chloroplasts are devoid of pollen and are inheritable through the maternal tissue, thus providing natural containment of transgene flow by outcrossing (Obembe et al., 2011).

Expression systems that utilize plastid transformation seem to be well-adapted for the accumulation of simple proteins that do not require complex post-translational modifications, especially glycosylation, as the chloroplast processing machinery closely resembles that of bacteria. This can be a crucial limitation when the protein of interest must be glycosylated in order to be biologically active. Moreover, proteins that are overexpressed in chloroplasts may form inclusion bodies and must be refolded after solubilization, which can contribute to an increase in overall production costs (Tremblay et al., 2010).

Although plastid transformation offers many practical advantages for molecular farming, it is limited to a few species, such as eggplant, tomato, lettuce, and soybean. Tobacco is the crop that is most routinely utilized for the expression of foreign genes in chloroplasts, but it is inedible and rich in toxic alkaloids, which can restrict its utility as a bioreactor for biopharmaceuticals (Obembe et al., 2011).

Stable nuclear transformation is the most common approach utilized in plant molecular farming and is applicable to a wide range of vegetable species, from cereals to grain legumes, including edible fruits, leafy crops, and oilseeds. This method permits the stable integration of recombinant genes into the host nuclear genome and allows for the continual production of recombinant protein as a conferrable, inheritable trait.

Apart from acquiring *in vitro* chimeric plants (organisms sharing transgenic and non-transgenic tissues derived from transformed and non-transformed germinative cells), stably transformed plants obtained in the laboratory can be repeatedly cultivated in fields or greenhouses to generate transgenic expression lines and can be utilized to transfer a desirable trait to conventional plants through controlled breeding (Lee and Natesan, 2006).

Because the most important factor related to the development of plant biotechnology products is ensuring competitive production costs and large-scale manufacturing, boosting heterologous expression in transgenic plants is the best way to achieve the production yields required for commercialization (Streatfield, 2007).

The expression levels of the transgenes integrated in the plant chromosomes are directly related to the transcriptional efficacy of the regulatory sequences provided by the recombinant expression cassette utilized for genetic transformation. Constitutive promoters

allow nonspecific expression in all plant tissues and organs and are routinely utilized for recombinant protein production in plants. The cauliflower mosaic virus 35S RNA (CaMV35S) and the maize ubiquitin-1 (*ubi-1*) constitutive promoters are commonly used to achieve high transcriptional levels in dicots and monocots, respectively (Obembe et al., 2011).

Although leafy crops have a considerably high biomass, proteins accumulated in the leaves are susceptible to degradation mediated by the endogenous proteolytic activity of hydrolases and interference from phenolic compounds naturally found in these tissues, which may reduce the final yields and present difficulties during downstream processing. To minimize protein degradation, leaves must be frozen or desiccated before harvesting and processing. In addition, because the transgenes can be expressed in the pollen and roots, constitutively expressing plants are frequently the subjects of biosafety concerns and issues, especially if they are cultivated in the field (Stoger et al., 2005; Lee and Natesan, 2006).

One of the most important advantages of plant bioreactors is that, in contrast to leaves, recombinant biopharmaceuticals and industrial proteins may be expressed in multiple plant organs that naturally accumulate high amounts of stored protein, such as tubers and especially seeds. To restrict the expression of transgenes to the seeds, well-characterized tissue-specific promoters with strong transcriptional activity can be utilized to not only boost heterologous expression but also to significantly increase the final protein yields.

Another important characteristic favoring transgene expression in seeds is that these organs have evolved as specialized compartments to store proteins for embryo nutrition after germination and during the initial stages of physiological seedling development. For this reason, seeds offer an environment with metabolic adaptations that permit the stable long-term storage of proteins, reducing the requirement for sophisticated and expensive conditions for appropriate storage.

As a result, by stockpiling harvested seeds, the agronomic production and downstream processing can be disconnected, and the recombinant protein may be purified and made available for commercialization according to the demands of the industry (Gleba et al., 2005).

Recombinant protein production in seeds can in fact take advantage of the abundance of oil bodies in oilseed rape and safflower. The company SemBioSys Genetics has developed a production platform in which target recombinant proteins are expressed as a fusion with a hydrophobic protein called oleosin, which can be recovered from oil bodies by a simple extraction procedure and separated by endoprotease digestion (Stoger et al., 2005).

Similarly, because the purification costs of a particular protein are inversely proportional to the final concentration in the plant biomass, the accumulation of high yields of recombinant protein in the restricted volume of most crop seeds provides a compact compartmentalization biomass, which can considerably reduce the overall production costs (Boothe et al., 2010).

In general, crops that have higher protein content are more cost-effective for molecular farming. Among all recombinant systems that utilize seeds, soybeans present an exceptionally high endogenous protein content, which can reach up to 40% of the dry seed weight, compared to 8-10% in maize and rice and 1-2% in leaf tissues of tobacco.

Soybeans are not only an outstanding source of protein for food and feed but are also a source of abundant oil. Thus, soybeans occupy a unique position as a premier target for genetic engineering among all grain legumes.

The extensive acreage devoted to the cultivation of soybeans in countries, such as the

United States, Brazil, China, and Argentina, and the scale of global transactions involving soybean agribusinesses reflect the economic importance of this crop to the industrialized world.

High quantities of soybean biomass (including leaves and seeds) can be obtained in less than four months at lower costs than other grain legumes, such as peas, beans, peanuts, and even most cereals. Soybeans are very economical because nitrogen fertilization is unnecessary once soybean roots can establish symbiotic associations with soil bacteria of the *Rhizobium* and *Bradyrhizobium* genera, which are capable of fixing atmospheric nitrogen for further utilization in the biosynthesis of aminated molecules, such as plant proteins and nucleic acid bases.

Under containment conditions, soybean grain yields can be considerably maximized by exploiting the plant's photoperiod sensitivity. By simply utilizing a daily photoperiod of 23 h of light, the plant vegetative growth can be extended and the flowering period delayed (Cavazzoni et al., 1999; Kantolic and Slafer, 2007), causing a 10-fold increase in seed production (approximately 1000 seeds per plant) compared to plants cultivated in the field (Figure 1).



Figure 1. Transgenic soybean plants cultivated in a greenhouse express different biopharmaceuticals, such as hGH, hFIX and the CV-N. High leaf and seed biomass were obtained by submitting the plants to a daily photoperiod of 23 h. The average grain yield achieved under these conditions was 1000 seeds per plant, approximately 10 times higher than in open-field cultivation.

Another important characteristic of soybeans is the low out-crossing rate of less than 1%, which is rare at distances greater than 6 m (Abud et al., 2007), reducing the likelihood of gene escape even with cultivation in contention.

Exploiting the photoperiod sensitivity together with the intrinsic high seed protein content makes soybean plants an attractive system for the production of biopharmaceuticals and industrial proteins. In addition, expressing transgenes specifically in the seeds of autogamous plants, such as soybeans, under containment conditions can minimize the occurrence of horizontal gene flux, which is desirable for avoiding pollen contamination and further bio-safety concerns (Lee and Natesan, 2006).

SUBCELLULAR COMPARTMENTALIZATION OF MOLECULES IN PSVs OF TRANSGENIC SEEDS

One of the most important aspects of a recombinant expression strategy is protein stability after translation. The total yield of a given protein reflects not only the rate of biosynthesis but also that of degradation. Protein targeting to different subcellular compartments is an essential stage when the goal is high-level protein accumulation and can also influence glycosylation and protein processing (Gleba et al., 2005).

Plant cells provide different organelles and compartments for protein targeting, such as the nucleus, the mitochondria, the plastids, the oil bodies, and the vacuolar compartments. Proteins can also be secreted into the apoplast or retained in the ER.

Protein targeting usually involves the enzymatic recognition and cleavage of N- or C-terminal signal peptides after the protein reaches its intracellular destination. In the absence of a signal peptide sequence, the protein accumulates a highly unstable and degradative environment in the cytoplasm, which may not be suitable for many proteins (Obembe et al., 2011).

The secretory pathway is a common destination for effective accumulation of proteins that require glycosylation to be fully active. Most post-translational processing occurs in the endomembrane system, including chaperone-mediated folding in the ER and glycosylation in the Golgi. If no additional signal peptide is added to the proteins, they are secreted into the apoplast, a suitable place for stable accumulation of proteins (Ma et al., 2003).

Localizing proteins to the ER is a commonly adopted strategy to considerably increase the final yields of many proteins expressed in plants. The C-terminal tetrapeptide H/KDEL is routinely utilized to direct proteins into the secretory pathway, causing them to relocate from the Golgi apparatus to the ER lumen, an oxidizing environment without proteases and containing an abundance of chaperones. This not only allows proteins to accumulate at high levels but also prevents the addition of most undesirable plant glycans, an important criterium for antibody production in plants (Obembe et al., 2011).

Seed vacuoles are the major sites of protein accumulation in seeds and are also the intracellular endpoints of the plant secretory pathway. They can be divided into two main categories: i) lytic vacuoles, the acidic compartments rich in hydrolases that are the equivalent of mammalian lysosomes, and ii) PSVs, the ER-derived cisternae specialized in storing the large amounts of protein required for seed germination (Jolliffe et al., 2005).

Cotyledonary PSVs are the preferable subcellular target for the long-term storage of recombinant proteins because they provide an oxidizing environment with low concentrations of amino peptidases and large quantities of protease inhibitors and ribosome inactivating proteins, thus minimizing protein degradation (Jolliffe et al., 2005; Cunha et al., 2010, 2011a,b).

The β -conglycinin signal peptide from soybeans and the cotyledonary α -coixin signal peptide from the monocot *Coix lacryma-jobi* are promising targeting signals for seed expression strategies.

Several molecules with different molecular weights and structures have been stably accumulated in the PSVs from transgenic soybean seeds using the α' subunit regulatory sequences of the soybean β -conglycinin and α -coixin, including hGH (2.9% TSP), human coagulation factor IX (FIX; 0.23% TSP), the mammary cancer marker scFvDIR83D4 (0.93% TSP), the anti-CD18 monoclonal antibody used to treat post-ischemic reperfusion injuries, the potent microbicides Cyanovirin-N (CV-N), scytovirin and griffithsin against-HIV, the immu-

nogenic antigens CTAG, GAGE-2B and PLAC, against different forms of cancer, and human proinsulin (Cunha et al., 2010, 2011a,b; Vianna et al., 2011a).

Some of these biopharmaceuticals have been effectively detected in the PSVs of transgenic seeds after six years of storage at room temperature, a far longer period than previously reported in the literature, and have demonstrated full or partial biological activities (Figure 2).

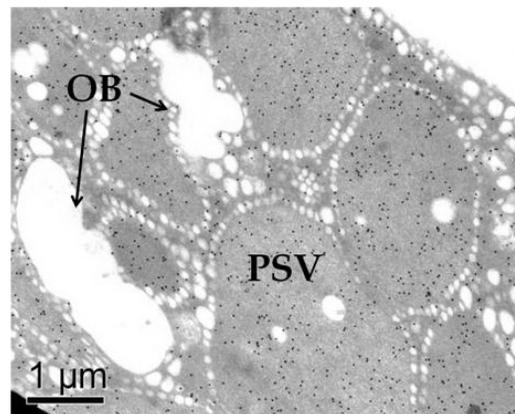


Figure 2. Ultrastructural immunocytochemistry showing the subcellular accumulation of recombinant hFIX in protein storage vacuoles (PSV) of transgenic soybean seeds and the absence of recombinant protein in oil bodies (OB).

Sequencing the N-termini of recombinant hGH and FIX by nanoscale liquid chromatography coupled with mass spectrometry (MS) has demonstrated that the signal peptides were correctly removed and that PSV targeting occurred with high specificity. It was also possible to evaluate the kinetics of protein accumulation in the PSVs during seed maturation, which indicated increasing levels later in physiological seed development, coinciding with the availability of a high number of PSVs in dry seeds (Cunha et al., 2011a,b).

SOYBEAN TRANSFORMATION

When choosing a plant-based bioreactor system, it is crucial to establish an effective genetic transformation method to obtain stable transgenic lines expressing the molecule of interest. Since Hinchey et al. (1988) and Christou et al. (1988) published the first two effective methods for soybean transformation (utilizing *Agrobacterium* and biolistics, respectively), much effort by several groups has been made to improve the system. Currently, there are several choices of methods for introducing foreign DNA into soybean plants by *Agrobacterium* or biolistics.

As an additional and interesting option, Aragão et al. (2000) have developed a protocol that is cultivar-independent and capable of generating transgenic soybeans at high frequency, up to 20.1% of frequency transformation to a Conquista Brazilian cultivar. The system introduces the transgene into the apical meristem of mature soybean seeds using the biolistic method, which is based on delivery of microparticle coated-DNA into the apical meristematic region of the embryonic axes (Christou et al., 1988). A mutant *ahas* gene that confers tolerance to imazapyr, a herbicidal molecule of the imidazolinone class capable of systemically translo-

cating to and concentrating in the apical meristematic region of the plant, is introduced along with the transgene. The biochemical mechanism of imazapyr, by inhibiting the enzymatic activity of acetohydroxyacid synthase (AHAS; EC 4.1.3.18), catalyses the initial stage in the biosynthesis of isoleucine, leucine and valine (Aragão et al., 2000). The non-transgenic plants die due to deficient production of these amino acids.

The *ahas* selectable marker gene, previously isolated from *Arabidopsis thaliana*, contains a mutation at position 653 bp, resulting in a serine to asparagine substitution that confers imidazolinone-specific resistance (Aragão et al., 2000). This selectable marker system, combined with the use of a multiple-shooting induction cytokinin BAP (benzylaminopurine), has resulted in a very effective protocol for obtaining fertile transgenic plants (Aragão et al., 2000; Rech et al., 2008; Vianna et al., 2011b).

In addition, this protocol has been used to generate stable transgenic soybean lines expressing pharmaceutical molecules as diverse as hGH (Cunha et al., 2011a), human FIX (Cunha et al., 2011b), human insulin (Cunha et al., 2010), and other molecules (Vianna et al., 2011a).

ATTEMPTS IN METABOLIC ENGINEERING AND SYNTHETIC BIOLOGY

For many therapeutic and industrial proteins, glycosylation is an essential post-translational modification required for proper biological function. The correct addition of superficial glycans by the enzymatic machinery of the host organism directly influences protein stability, solubility, folding and, consequently, *in vivo* activity.

While N-glycan addition in the ER is conserved among eukaryotes, glycosylation in the Golgi is kingdom-specific and results in superficial linkage of a variety of different carbohydrate structures to various glycoproteins. Although plant expression systems provide the correct glycosylation of many proteins, especially those that naturally occur in bacteria, yeast and insects, the biosynthesis of mammalian proteins in vegetable bioreactors frequently requires more attention when complex glycosylation is required (Lienard et al., 2007).

Plants and mammals glycosylate proteins on the same common Asn residues and the high-mannose type N-glycans are structurally identical in both systems. However, the complex N-glycans added by the plant and mammal machineries are quite different from one another. Usually, the mammalian glycans α -1,6-fucose and β -1,4-N-acetylglucosamine are respectively substituted by the plant glycans α -1,3-fucose and a bisecting β -1,2-xylose during the exchange of proximal N-acetylglucosamine and the β -mannose in the Golgi of plant cells (Gomord and Faye, 2004).

Another important difference in the processing of the terminal N-acetylglucosamines in plants is the attachment of two types of glycans: the β -1,3-galactose and the α -1,4-fucose, which lack the mammalian β -1,4-galactose and sialic acid, molecules that are not detected in proteins derived from plants. This can cause the rapid elimination of PMPs from the blood and immunogenicity in humans and can have major negative impacts on the specificity of most IgE- and IgG-antibodies produced in plants (Gomord et al., 2010).

Several strategies to humanize the structure of plant N-linked glycans can be utilized to allow the production of recombinant therapeutic glycoproteins in plants. Strategies are primarily focused on the control of plant-specific N-glycan maturation and on preventing the addition of immunogenic glycoepitopes onto PMPs. One of the most promising approaches is to directly mutate the Asn or Ser/Thr residues that are the major sites of superficial N-

glycosylation. This approach is well-suited for the production of antigens and simple PMPs that do not have *in vivo* activity or high longevity, but is not recommended for more complex glycoproteins (Gomord et al., 2010).

In addition, the inhibition of plant-specific Golgi glycosyltransferases can successfully prevent the addition of glyco-epitopes to recombinant PMPs. This can be achieved by the inactivation of the α -1,3-fucosyltransferase and β -1,2-xylosyltransferase genes using insertional mutations or by exploiting the RNA interference method, a highly effective approach for homogeneous glycosylation of humanized antibodies in plants (Sethuraman and Stadheim, 2006).

The combined knockout of four different glycosyl transferases in transgenic soybean plants developed by Embrapa is currently being evaluated as an efficient tool for the humanization of glycan-epitopes in recombinant proteins accumulated in seeds.

Another important point is to exploit the structural similarity of the high-mannose N-glycans added in ER lumen of eukaryotic cells. The addition of N- or C-terminal signal peptides (such as the KDEL) to retain proteins in the ER, avoiding processing in the Golgi, can not only improve the stability but can also influence the addition of exclusively non-immunogenic high-mannose type N-glycans to the protein surface (Sethuraman and Stadheim, 2006).

The remodeling of biosynthetic pathways to obtain specific metabolic products, normally by adding extra-enzymatic reaction stages, is another attractive strategy to humanize plant N-glycans. The most obvious way to do this is to express mammalian glycosyltransferases in plants, a process that triggers competition between the newly synthesized enzymes and the endogenous machinery for N-glycan processing in the Golgi apparatus. For example, the addition of galactose residues onto the terminal N-acetylglucosamine residues of plant N-glycans by recombinant human β -1,4-galactosyltransferase has consistently resulted in a glycan pattern in plant cells similar to that in mammalian cells (Gomord et al., 2004).

In addition, the production of sialylated N-glycans is feasible in plants because many Golgi complex biosynthetic pathways have literally been re-designed by expression of mammalian α -2,6-sialyltransferase, the human CMP-N-acetylneuraminic acid synthetase, the CMP-sialic acid transporter and two N-acetylneuraminic acid synthesizing enzymes (Gomord et al., 2010).

The modification of metabolic pathways through the use of recombinant DNA technology can also be applied for phenotypic improvements of important crops utilized as renewable feedstocks for the production of biofuels.

To fulfill the exacting demands of the fuel industry, a high-quality biodiesel must have a high oleic acid concentration and low saturated fatty acid content, providing more stability at high frying temperatures during processing, and a longer shelf life than the conventional commodity oil.

Although significant efforts have led to some success in the introgression of these desirable characteristics in elite varieties of maize, canola and soybeans, by exploiting the endogenous germplasm variation of these plants, conventional breeding for oil content modification has some drawbacks. This is especially in terms of environmental influences over the new traits, as the link between the novel fatty acid profile to a reduced oil yield and that the oleic acid concentration in the seeds tends to be a multi-loci inheritance and is thus difficult to manipulate (Graef et al., 2009).

In contrast, maximization of the oleic acid and reduction in trans-fat content of oil can be achieved by altering key stages in the well-known fatty acid pathway in oleaginous crops, notably in soybeans. Because these reactions are often catalyzed by important desaturases and

esterases, which directly or indirectly lead to the accumulation of the final metabolic products, they naturally have been converted to the main targets for overturning or boosting heterologous expression (Buhr et al., 2002).

Soybean seeds are constituted of 18% oil with a high proportion of polyunsaturated fatty acids. Thus, soybeans constitute one of the major targets for the application of synthetic biology tools aimed at the large-scale production of biodiesel with high fuel quality (Graef et al., 2009).

In soybean seeds, oleic acid is converted to linoleic acid in a single-desaturation step catalyzed by a $\Delta 12$ desaturase, encoded by the *FAD2-1* and *2* gene families. In addition, the concentration of undesirable seed palmitic acid is directly influenced by the activity of a palmitoyl thioesterase encoded by the *FatB* gene (Tang et al., 2005).

A transgenic variety of soybean developed by DuPont, called Plenish™, exhibits down-regulated expression of both the *FAD2-1* and *FatB* genes, resulting in a dramatic increase in oleic acid concentration, which can reach up to 85% of the total oil content, and a considerable reduction in the trans-fat level, mainly in palmitic acid (accounting for less than 5% of the total content). This is a good candidate for the biodiesel and consumer food industries, which have received USDA regulatory approval for cultivation in the US in 2010 (Graef et al., 2009).

New attempts to improve the oil quality from soybean seeds have been carried out by Embrapa. These approaches have utilized RNA interference mediated by strong constitutive promoters to achieve simultaneous silencing of the *FAD2-1*, *FAD2-2* and *FatB* genes. This is expected to result in increased accumulation of oleic acids and a reduction in trans-fat concentrations to the minimal detectable yields in the seeds.

PURIFICATION AND CHARACTERIZATION APPROACHES

The purification and characterization of recombinant proteins in soybean seeds is one of the most important and difficult tasks when using the soybean plant as bioreactor. To understand the complexity, we will describe briefly the composition of soybean seeds, which are mainly composed of protein (37%), carbohydrate (30%), oil (20%) and have low polyphenol content (0.02%). The protein content of the seeds is 10% albumin, with a molecular mass less than 20 kDa, and 90% globulins, which are separated into four groups according the sedimentation characteristic: 2S, 7S, 11S, and 15S. The 2S group corresponds to 22% of the globulins and is composed of β -conglycinin proteins of 8 to 50 kDa. The 7S fraction corresponds to 37% of the globulins and is composed mainly of the glycoprotein β -conglycinin. This protein consists of a combination of three subunits: α' (57-72 kDa), α (57-68 kDa) and β (42-52 kDa), with isoelectric points of: α' and α (4.9 to 5.2) and β (5.7 to 6), leading into the formation of seven isomeric units with molecular masses between 130 and 180 kDa and an isoelectric point around 4.9. Other proteins of the 7S group are lectin (hemagglutinin, 120 kDa), α -amylase (62 kDa), lipoxygenase (102 kDa), and cytochrome c (12 kDa) (Nielson, 1985). The 11S set is essentially composed of glycinin and constitutes 31% of the globulins. Native glycinin has a molecular mass between 300 and 360 kDa and is composed of a combination of five subunits (G1 to G5). Each unit can be disassociated by reducing agents into an acidic polypeptide chain (A; 31-45 kDa) and a basic (B; 18-20 kDa). The five subunits are a combination of the following proteins: G1 (A1aB2), G2 (A1bB1b), G3 (A2B1a), G4 (A3B4), and G5 (A5A4B3) (Staswick et al., 1981). The isoelectric points of the glycinins vary between 4.5 and 4.8. For the latter, the 15S fraction corresponds to 1% of the globulins, has a high molecular mass (700

kDa) and has an isoelectric point of 4.5. Furthermore, the known proteins in soybean seeds with high allergenic potential are mainly composed of a 34-kDa maturing seed protein (also known as P34, Gly m Bd 30 K or Gly m 1) (Sewekow et al., 2008), the 23-kDa protein Gly m Bd 28 K (Xiang et al., 2004), β -conglycinin (Krishnan et al., 2009), the glycinin Gly m 6 (Holzhauser et al., 2009), lectin and lipoxygenase (L'Hocine and Boye, 2007), and a Kunitz-type protease inhibitor (Roychaudhuri et al., 2004). There are also some allergenic proteins in seed shell, such as Gly m 1.0101, Gly m 1.0102, Gly m 2, Gly m 3, and SAM 22 (Gly m 4). For a complete list, check the Allergome project for updated information (Mari et al., 2009).

Soybean proteins have been purified and studied for so long that we now have an idea of how to begin the purification and what must be avoided. First, seeds must be grinded preferably into a fine powder similar to flour. This will increase the contact of solvent with the seed material, increasing protein recovery. The removal of fat content is also preferable, using petroleum ether or hexane prior to aqueous extraction. We must keep in mind that these soy seed proteins present the major difficulty in purification of recombinant molecules. When aqueous extraction with neutral pH is performed, high sample viscosity is evidenced. It is caused by the interactions between β -conglycinin and glycinin through disulfide bonds, forming a type of polymer. If possible, the use of mercaptoethanol or cysteine in the extraction buffer is encouraged to diminish viscosity. Also, the low ionic force used in the extraction solvent may cause dimerization in these proteins (Smith and Circle, 1972). This may be avoided by increasing the ionic force using NaCl. However, depending on the concentration, this may also affect chromatography, especially ion exchange. The choice of buffer or solution for extraction of soluble proteins from seeds is dependent on the protein target for purification. However, the use of water of pH 7-10 and solutions with 0.5-2.0 M NaCl are the most effective for solubilizing most soy proteins (Smith et al., 1938; Smith and Circle, 1938). Using solutions or buffers with low pH (<5) will affect the TSP, but the yield can be increased with the addition of 0.1 M monovalent ions (Smith et al., 1938). Recently, Robić et al. (2010) tested the efficacy of recombinant β -glucuronidase (rGUS) recovery produced in transgenic soybean using 50 mM citrate-phosphate buffers with pH from 3.4 to 7.6 containing NaCl at concentrations up to 0.2 M. The authors found that a high efficiency rGUS extraction (high-specific activity) was observed at pH of approximately 5.5 with no addition of salt. In addition, the extraction of native proteins, phenolic compounds, and soluble carbohydrates was minimized in these conditions. This extraction condition reduces the native protein content in the extract, thus minimizing its complexity and streamlining recombinant protein isolation by decreasing downstream processing. One point to consider is that because β -conglycinin and glycinin represent 90% of all soybean proteins, the extraction should be carried out at a pH between 4.5 and 5.5.

Protein purification relies on classical liquid chromatography methods derived from study of chemistry Nobel Prize recipient Archer J.P. Martin (1952). With minor changes over the years, resins and materials have evolved, but the concept of protein chromatography remains the same. The chromatography techniques include: affinity, ion exchange, hydrophobic interaction, and size exclusion. Affinity chromatography is a separation technique based upon molecular conformation, using specific resins with ligands attached to their surfaces that are specific for the compounds to be separated. The ion exchange technique separates molecules according to the nature and amount of their ionic charge. The column to be used is selected according to its type and strength of the charge. Anion exchange resins have a positive charge and are used to retain and separate negatively charged

compounds, while cation exchange resins have a negative charge and are used to separate positively charged molecules. Hydrophobic interaction relies on interaction between exposed hydrophobic residues of a protein to a resin coupled with a hydrophobic molecule, such as butyl or phenyl. Size-exclusion chromatography is also known as gel filtration chromatography and separates molecules according to their size. Smaller molecules are able to enter the pores of the media and, therefore, are trapped and removed from the flow of the mobile phase. Large proteins cannot enter the resin pores and are thus eluted before small proteins. The choice of purification strategy depends on several questions, such as: What is the intended use of the protein? What has to be removed completely? What will the final scale of purification be? Preparative chromatography aims to purify large amounts of target proteins. Analytical chromatography aims to purify small amounts of protein with a high degree of purity for the purpose of characterization. The addition of purification steps reduces the initial quantity because of natural losses during the process, so a simple strategy is an effective way to carry out the purification. A three-phase strategy is a good way to begin. The following specific objectives are assigned to each stage within the procedure: capture, intermediate purification and polishing. In the capture phase, the objectives are to isolate, concentrate and stabilize the target product. This is usually accomplished using resins with a high pore size ($>45\ \mu\text{m}$) and capable of handling samples obtained directly from extraction. During the intermediate purification phase, the objective is to remove most of the bulk impurities, such as other proteins and nucleic acids, endotoxins and viruses. The use of resins with pore size between 15 to 45 μm is recommended. In the polishing phase, the objective is to achieve high purity by removing any remaining trace impurities or closely related substances, generally using resins with pore size lower than 15 μm . Using these approaches, some proteins have been purified from transgenic soybean. Zeitlin et al. (1998) have purified a humanized monoclonal antibody for immunoprotection of the vagina against genital herpes by sequential chromatographic steps, namely protein A, cation exchange, and ion exchange, with a final storage solution of 20 mM Tris-HCl, pH 8.0. Philip et al. (2001) expressed and purified bovine β -casein from soybean seeds using denatured casein antibody coupled to CNBr-activated Sepharose 4B. These authors removed impurities by two cycles of washing with 0.1 M sodium acetate with 1 M NaCl, pH 4.8, and 0.1 M sodium carbonate with 1 M NaCl, pH 7.6. Bound protein was eluted with 0.5 M acetic acid, pH 2.5, and pooled protein fractions were adjusted immediately to pH 7.0 by dropwise addition of 2.5 M ammonium hydroxide. Robić et al. (2006) purified rGUS using only two steps of chromatography, yielding up to 93% purity. The authors used anionic exchange DEAE-Sepharose followed by HIC chromatography on Phenyl-Sepharose. After the purification procedure, characterization of the molecule must be carried out. This involves enzymatic assays, if molecule is an enzyme, and the determination of protein purity and sequence. For these purposes, MS is the most versatile and broad analytical tool available in biochemistry for determining the mass in Daltons of recombinant proteins. It can be also used for sequence determination by *de novo* sequencing (Shevchenko et al., 2006), and for determination of post-translational modifications, such as phosphorylation (Hunter and Games, 1994), glycosylation (Wilson et al., 2009) and protein oxidation (Barnes et al., 2008). In addition to MS methods, N-terminal sequencing can be verified by the EDMAN degradation technique when it cannot be sequenced by MS due to ionization problems or a long N-terminal fragment (Edman, 1949).

CONCLUSION

There is high demand for the production of a large number of recombinant proteins both for scientific evaluation and product development. In addition, due in large part to production costs, current therapies that utilize recombinant proteins cause significant pressure to health systems in developing and developed countries. In developing countries, cost is the main constraint for much of the population. The utilization of plants offers an additional option to bacteria and mammalian cell-based recombinant protein production systems. In some cases, the utilization of plants might determine whether an effective cancer antigen, a microbicide against HIV, a biopolymer may or may not be economically viable to reach the market. Seeds are very efficient production systems. In addition, seeds have specialized compartments to accumulate proteins. These characteristics might render seeds to produce a very efficient system for the large-scale production of recombinant proteins at an economically viable cost. The current tools, including synthetic biology and metabolic engineering, should expand even more the capacity to manipulate seeds for the production of recombinant proteins. The studies presented in this article indicate that seeds are capable of producing recombinant proteins with different biochemical characteristics at high levels. Specific bioassays have demonstrated the functional activity of the recombinant proteins produced. Currently, a suitable regulatory framework has been proposed which in turn, should allow plant-derived recombinant proteins to be commercialized.

ACKNOWLEDGMENTS

Research supported by Brazilian Agricultural Research Corporation, National Council for Scientific and Technological Development (CNPq) and Fundação de Apoio à Pesquisa-DF (FAP-DF).

REFERENCES

- Abud S, de Souza PI, Vianna GR, Leonardecz E, et al. (2007). Gene flow from transgenic to nontransgenic soybean plants in the Cerrado region of Brazil. *Genet. Mol. Res.* 6: 445-452.
- 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.
- Barnes S, Shonsey EM, Eliuk SM, Stella D, et al. (2008). High-resolution mass spectrometry analysis of protein oxidations and resultant loss of function. *Biochem. Soc. Trans.* 36: 1037-1044.
- Boothe J, Nykiforuk C, Shen Y, Zaplachinski S, et al. (2010). Seed-based expression systems for plant molecular farming. *Plant Biotechnol. J.* 8: 588-606.
- Buhr T, Sato S, Ebrahim F, Xing A, et al. (2002). Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. *Plant J.* 30: 155-163.
- Cavazzoni J, Volk T, Bugbee B and Dougher T (1999). Phasic temperature and photoperiod control for soybean using a modified CROPGRO model. *Life Support. Biosph. Sci.* 6: 273-278.
- Christou P, McCabe DE and Swain WF (1988). Stable transformation of soybean callus by DNA-coated gold particles. *Plant Physiol.* 87: 671-674.
- Cunha NB, Araujo AC, Leite A, Murad AM, et al. (2010). Correct targeting of proinsulin in protein storage vacuoles of transgenic soybean seeds. *Genet. Mol. Res.* 9: 1163-1170.
- Cunha NB, Murad AM, Cipriano TM, Araujo AC, et al. (2011a). Expression of functional recombinant human growth hormone in transgenic soybean seeds. *Transgenic Res.* 20: 811-826.
- Cunha NB, Murad AM, Ramos GL, Maranhao AQ, et al. (2011b). Accumulation of functional recombinant human coagulation factor IX in transgenic soybean seeds. *Transgenic Res.* 20: 841-855.

- Daniell H, Streatfield SJ and Wycoff K (2001). Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants. *Trends Plant Sci.* 6: 219-226.
- Demain AL and Vaishnav P (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* 27: 297-306.
- Edman P (1949). A method for the determination of amino acid sequence in peptides. *Arch. Biochem.* 22: 475.
- Fischer R, Stoger E, Schillberg S, Christou P, et al. (2004). Plant-based production of biopharmaceuticals. *Curr. Opin. Plant Biol.* 7: 152-158.
- Gleba Y, Klimyuk V and Marillonnet S (2005). Magniffection-a new platform for expressing recombinant vaccines in plants. *Vaccine* 23: 2042-2048.
- 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.
- Gomord V and Faye L (2004). Posttranslational modification of therapeutic proteins in plants. *Curr. Opin. Plant Biol.* 7: 171-181.
- Gomord V, Sourrouille C, Fitchette AC, Bardor M, et al. (2004). Production and glycosylation of plant-made pharmaceuticals: the antibodies as a challenge. *Plant Biotechnol. J.* 2: 83-100.
- Gomord V, Fitchette AC, Menu-Bouaouiche L, Saint-Jore-Dupas C, et al. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol. J.* 8: 564-587.
- Graef G, LaVallee BJ, Tenopir P, Tat M, et al. (2009). A high-oleic-acid and low-palmitic-acid soybean: agronomic performance and evaluation as a feedstock for biodiesel. *Plant Biotechnol. J.* 7: 411-421.
- Hinchee MAW, Connor-Ward DV, Newell CA, McDonnell RE, et al. (1988). Production of transgenic soybean plants using *Agrobacterium*-mediated DNA transfer. *Nat. Biotechnol.* 6: 915-922.
- Holzhauser T, Wackermann O, Ballmer-Weber BK, Bindslev-Jensen C, et al. (2009). Soybean (*Glycine max*) allergy in Europe: Gly m 5 (beta-conglycinin) and Gly m 6 (glycinin) are potential diagnostic markers for severe allergic reactions to soy. *J. Allergy Clin. Immunol.* 123: 452-458.
- Houdebine LM (2009). Production of pharmaceutical proteins by transgenic animals. *Comp. Immunol. Microbiol. Infect. Dis.* 32: 107-121.
- Hunter AP and Games DE (1994). Chromatographic and mass spectrometric methods for the identification of phosphorylation sites in phosphoproteins. *Rapid. Commun. Mass Spectrom.* 8: 559-570.
- Jolliffe NA, Craddock CP and Frigerio L (2005). Pathways for protein transport to seed storage vacuoles. *Biochem. Soc. Trans.* 33: 1016-1018.
- Kantolic AG and Slafer GA (2007). Development and seed number in indeterminate soybean as affected by timing and duration of exposure to long photoperiods after flowering. *Ann. Bot.* 99: 925-933.
- Karg SR and Kallio PT (2009). The production of biopharmaceuticals in plant systems. *Biotechnol. Adv.* 27: 879-894.
- Kawakatsu T and Takaiwa F (2010). Cereal seed storage protein synthesis: fundamental processes for recombinant protein production in cereal grains. *Plant Biotechnol. J.* 8: 939-953.
- Krishnan HB, Kim WS, Jang S and Kerley MS (2009). All three subunits of soybean beta-conglycinin are potential food allergens. *J. Agric. Food Chem.* 57: 938-943.
- L'Hocine L and Boye JI (2007). Allergenicity of soybean: new developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. *Crit. Rev. Food Sci. Nutr.* 47: 127-143.
- Larrick JW and Thomas DW (2001). Producing proteins in transgenic plants and animals. *Curr. Opin. Biotechnol.* 12: 411-418.
- Lee D and Natesan E (2006). Evaluating genetic containment strategies for transgenic plants. *Trends Biotechnol.* 24: 109-114.
- Lienard D, Sourrouille C, Gomord V and Faye L (2007). Pharming and transgenic plants. *Biotechnol. Annu. Rev.* 13: 115-147.
- Ma JK, Drake PM and Christou P (2003). The production of recombinant pharmaceutical proteins in plants. *Nat. Rev. Genet.* 4: 794-805.
- Mari A, Rasi C, Palazzo P and Scala E (2009). Allergen databases: current status and perspectives. *Curr. Allergy Asthma Rep.* 9: 376-383.
- Nielson NC (1985). Structure of Soy Proteins. In: *New Protein Foods* (Altschul AM and Wilcke HL, eds.). Academic Press, New York.
- Obembe OO, Popoola JO, Leelavathi S and Reddy SV (2011). Advances in plant molecular farming. *Biotechnol. Adv.* 29: 210-222.
- Philip R, Darnowski DW, Maughan PJ and Vodkin LO (2001). Processing and localization of bovine beta-casein expressed in transgenic soybean seeds under control of a soybean lectin expression cassette. *Plant Sci.* 161: 323-335.
- 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.
- Robić G, Farinas CS, Rech EL, Bueno SMA, et al. (2006). Downstream process engineering evaluation of transgenic soybean seeds as host for recombinant protein production. *Biochem. Eng. J.* 32: 7-12.

- Robić G, Farinas CS, Rech EL and Miranda EA (2010). Transgenic soybean seed as protein expression system: aqueous extraction of recombinant beta-glucuronidase. *Appl. Biochem. Biotechnol.* 160: 1157-1167.
- Roychaudhuri R, Sarath G, Zeece M and Markwell J (2004). Stability of the allergenic soybean Kunitz trypsin inhibitor. *Biochim. Biophys. Acta* 1699: 207-212.
- Salamone D, Baranao L, Santos C, Bussmann L, et al. (2006). High level expression of bioactive recombinant human growth hormone in the milk of a cloned transgenic cow. *J. Biotechnol.* 124: 469-472.
- Sethuraman N and Stadheim TA (2006). Challenges in therapeutic glycoprotein production. *Curr. Opin. Biotechnol.* 17: 341-346.
- Sewekow E, Kessler LC, Seidel-Morgenstern A and Rothkotter HJ (2008). Isolation of soybean protein P34 from oil bodies using hydrophobic interaction chromatography. *BMC Biotechnol.* 8: 27.
- Shevchenko A, Tomas H, Havlis J, Olsen JV, et al. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1: 2856-2860.
- Smith AK and Circle SJ (1938). Peptization of soybean proteins extraction of nitrogenous constituents from oil-bases with and without added slats. *Ind. Eng. Chem.* 30: 1414-1418.
- Smith AK and Circle SJ (1972). Soybeans: Chemistry and Technology. 2nd edn. The AVI Publishing Company, New York.
- Smith AK, Circle SJ and Brother GH (1938). Peptization of soybean proteins. The effect of neutral salts on the quantity of nitrogenous constituents extracted from oil-free meal. *J. Am. Chem. Soc.* 60: 1316-1320.
- Spök A and Karner S (2008). Plant Molecular Farming. Opportunities and Challenges. Final Report. The Institute for Prospective Technological Studies, Serville.
- Spök A, Twyman RM, Fischer R, Ma JK, et al. (2008). Evolution of a regulatory framework for pharmaceuticals derived from genetically modified plants. *Trends Biotechnol.* 26: 506-517.
- Staswick PE, Hermodson MA and Nielsen NC (1981). Identification of the acidic and basic subunit complexes of glycinin. *J. Biol. Chem.* 256: 8752-8755.
- Stoger E, Ma JK, Fischer R and Christou P (2005). Sowing the seeds of success: pharmaceutical proteins from plants. *Curr. Opin. Biotechnol.* 16: 167-173.
- Streatfield SJ (2007). Approaches to achieve high-level heterologous protein production in plants. *Plant Biotechnol. J.* 5: 2-15.
- Tang GQ, Novitzky WP, Carol GH, Huber SC, et al. (2005). Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. *Plant J.* 44: 433-446.
- Tremblay R, Wang D, Jevnikar AM and Ma S (2010). Tobacco, a highly efficient green bioreactor for production of therapeutic proteins. *Biotechnol. Adv.* 28: 214-221.
- Vianna GR, Cunha N and Rech E (2011a). Expression and accumulation of heterologous molecules in the protein storage vacuoles of soybean seeds. Protocol Exchange. Available at [<http://www.nature.com/protocolexchange/protocols/2012>]. Accessed May 10, 2011.
- Vianna GR, Aragao FJ and Rech EL (2011b). A minimal DNA cassette as a vector for genetic transformation of soybean (*Glycine max*). *Genet. Mol. Res.* 10: 382-390.
- Wilson N, Simpson R and Cooper-Liddell C (2009). Introductory glycosylation analysis using SDS-PAGE and peptide mass fingerprinting. *Methods Mol. Biol.* 534: 205-212.
- Xiang P, Haas EJ, Zeece MG, Markwell J, et al. (2004). C-Terminal 23 kDa polypeptide of soybean Gly m Bd 28 K is a potential allergen. *Planta* 220: 56-63.
- Xu J, Ge X and Dolan MC (2011). Towards high-yield production of pharmaceutical proteins with plant cell suspension cultures. *Biotechnol. Adv.* 29: 278-299.
- Zeitlin L, Olmsted SS, Moench TR, Co MS, et al. (1998). A humanized monoclonal antibody produced in transgenic plants for immunoprotection of the vagina against genital herpes. *Nat. Biotechnol.* 16: 1361-1364.