1

Protein Expression Systems: Why Soybean Seeds?

Kenneth Bost and Kenneth Piller
University of North Carolina at Charlotte and SoyMeds, Inc.
United States of America

1. Introduction

The global protein therapeutics market is approaching $100 billion in annual sales. Furthermore, the in vitro diagnostic market, which relies heavily on the use of recombinant proteins as analyte specific reagents, is approaching $50 billion in annual sales. Increases in each market sector are estimated to be 8% to 20% per year for the next decade. Therefore, the costs of protein-based therapeutics and diagnostics will continue to be a significant percentage of health care expenses for patients and for agricultural animals and pets.

A platform technology, which produces recombinant proteins at greatly reduced costs, provides inherent advantages, and allows low-tech sustainability of product lines, represents a competitive innovation which could create significant wealth in this industry. The use of soybean-derived proteins has the potential to provide such advantages. Since the use of recombinant proteins is widespread in human and animal therapeutics and diagnostics, there are thousands of potential applications for such a platform technology. While no one technology is optimal for the expression of every transgenic protein (Brondyk, 2009), the research that we have performed to date demonstrates the utility and feasibility of commercial applications for proteins made in transgenic soybean seeds.

Soybean seeds expressing transgenic proteins represent a novel, sustainable platform technology which overcomes some of the current limitations for producing recombinant proteins for diagnostics, therapeutics, and industrial applications. Advantages include low cost of glycoslyated protein production, greenhouse containment, highest protein/biomass ratio, marketable formulations which require no purification from soy, safety, accurate dosing, low cost of protein purification, low-tech sustainability of product lines, reduced risk of contamination, ease of scalability, minimal waste produced, as well as being a green technology. To our knowledge, no other protein expression technology compares.

Despite the fact that the present protein therapeutic and diagnostic markets are growing rapidly, some applications continue to be limited by the current technologies employed (Brondyk, 2009). For example the cost of production and purification of certain recombinant proteins makes their use in particular applications non-profitable. Some proteins cannot be expressed by any current technology and require expensive purification from human or animal tissues. A platform technology which could alleviate these concerns would create significant opportunities for novel product development, or expanded applications for existing products. In short, there is a need for alternative technologies for expressing and purifying recombinant proteins which can advance their applications. We propose that

soybean seeds expressing transgenic proteins represent a platform technology that can be a solution for many of these problems. Recent research provides a strong basis for this supposition. These results support the utility of this technology, have demonstrated its advantages, and suggest additional benefits that are currently being explored.

2. Utility of glycosylated proteins produced in transgenic soybean seeds

Roundup Ready soybeans were one of the first examples of a commercially viable transgenic plant (Padgette et al., 1995). These transgenic soybeans express a functional enzyme, 5-enolpyruvylshikimate-3-phosphate synthase, making the plants tolerant to the herbicide, Roundup™. Presently, approximately 90% of the soybeans farmed in the United States are Roundup Ready.

In addition to this value-added trait, there have been several recent successes in modifying soy crop lines aimed at imparting some commercial advantage. Transgenic soybean lines expressing the cry1A gene provide protection against Lepidopteran species (McPherson & MacRae, 2009). Transgenic soybean lines expressing an active APase, demonstrated increased phosphorus content when grown in soils with limited phosphate content (Wang et al., 2009). Alterations in seed oil content have been achieved using transgenic soybean plants expressing genes having lysophosphatidic acid acyltransferase activity (Rao & Hildebrand, 2009). Over-expression of an aspartate kinase in transgenic seeds allowed increased threonine levels to occur (Qi et al., 2010). While these examples are not a comprehensive listing of value-added traits that require the expression of a function enzymatic protein, they serve to demonstrate the amenability of soybean seeds to such transformations.

While value-added traits are being exploited, transgenic soybean seeds have also been used as bioreactors to express a variety of foreign proteins. For example, Zeitlin, et al. (Zeitlin et al., 1998) successfully expressed functional antibodies against herpes simplex virus-2 glycoprotein B in transgenic soybeans. More recently, proinsulin has been expressed and the storage vacuoles can accumulate mature polypeptide in these seed lines (Cunha et al., 2010).

In our laboratories, we have successfully expressed the subunit protein antigens, E. coli, FanC (Garg et al., 2007; Oakes, Bost, & Piller, 2009; Piller et al., 2005), non-toxic, mutant forms of several bacterial toxins, and potential immunomodulatory proteins in transgenic soybean seeds, and are evaluating their usefulness as therapeutics. We have also successfully expressed particular full-length human proteins, and ongoing studies are aimed at demonstrating their substantial equivalence for use as analyte specific reagents in diagnostic assays.

While this is not an exhaustive listing of the plant and foreign proteins which have been expressed in transgenic soybean seeds, these examples demonstrate the utility of this platform technology. There are several reasons for the success of such endeavors, but perhaps the most important lies in the biology of the soybean seed itself. One of the most important functions of the seed is to express and package proteins. This entails post-translational modifications, including glycosylation, and packaging which not only allows proper folding, but also provides an environment for stable, long-term protein storage. This conclusion is supported by the fact that many of the transgenic proteins expressed have enzymatic activity, ability to bind antigen, or the ability to be recognized by monoclonal antibodies specific for their native counterparts.

3. Soybean seeds represent the highest protein to biomass ratio

Soybean seeds, by weight, are 40% protein with approximately 20% oil, 35% carbohydrates, and 5% ash (Liu, 1999). Most of the normal soy seed proteins are heat-stable and desiccant-resistant, in keeping with the ability of soybeans to remain germinate-capable following years of storage in ambient conditions. Soybean plants can produce as much as twice the protein per acre of any other major crop (see: http://www.soyatech.com/soy_health.htm). Protein production by soy is also more efficient than animal-derived protein when factoring the acreage required for grazing or feeding. As we will discuss below, this fundamental characteristic of soybean seeds to produce and store large amounts of protein may be exploited as a platform technology for expression.

4. High yield translates into a potential for low cost protein production

Present and future use of recombinant proteins will be limited in large part by their cost of production. Presently, the expense of expressing and purifying some recombinant proteins prohibits or limits their practical or realistic use. This is true for some therapeutic, as well as diagnostic, applications in westernized societies, and such barriers are even greater for developing countries. Unless this economic burden can be overcome, barriers for product development will remain. For some applications, too much recombinant protein is needed such that the cost is prohibitive. For some applications, elaborate purification schemes make particular proteins unaffordable. For some applications, there is no source of some recombinant proteins, requiring isolation from human or animal tissues. A platform technology which could alleviate these concerns would create significant opportunities for novel product development, or expanded applications for existing products, and therefore, create significant wealth. Decisions to develop commercial products which include such proteins will depend largely on the practicality of having a cost-sustainable platform technology. Theoretically, expression of transgenic proteins in soybean seeds represents one of the most cost-efficient platforms, and recent work has demonstrated potential economic advantages.

Presently we (Garg, et al., 2007; Oakes, Bost, et al., 2009; Piller, et al., 2005), and others (Cunha, et al., 2010; Ding, Huang, Wang, Sun, & Xiang, 2006; Moravec, Schmidt, Herman, & Woodford-Thomas, 2007; Qi, et al., 2010; Rao & Hildebrand, 2009; Wang, et al., 2009; Zeitlin, et al., 1998), have developed stable soybean lines that express 1% to 4% of their total soluble protein as the transgenic protein. Since soybean seeds are 40% protein by weight, an average sized seed weighing approximately 150 milligrams represents approximately 2.4 milligrams of transgenic protein per seed at 4% expression (Table 1). As will be discussed below, soybeans can easily be converted into soy powder, and, one liter of this powder totals approximately 800 grams of seed material. At 4% expression, this one liter of soy powder contains approximately 12.8 grams of the unpurified, transgenic protein (Table 1). It is useful to compare this production with that for a liter of broth from bacterial (Zerbs, Frank, & Collart, 2009), yeast (Cregg et al., 2009), insect (Jarvis, 2009), mammalian (Geisse & Fux, 2009), or plant (Hellwig, Drossard, Twyman, & Fischer, 2004; Lienard, Sourrouille, Gomord, & Faye, 2007) cell culture.

Since retail costs of commercially available recombinant proteins can easily be hundreds to thousands of dollars per milligram, extrapolations presented in Table 2 are enlightening. One liter, or 800 grams, of soy powder could represent as much as 12.8 grams of transgenic protein

at an expression level of 4%. As shown in Table 2, potentially, as little as one liter of total soy protein could contain the equivalent of millions of dollars of unpurified transgenic protein.

Percent expression of the transgenic protein	Milligrams of transgenic protein per seed (150 milligrams)	Grams of transgenic protein per liter (800 grams of soy powder)
1%	0.6 milligrams	3.2 grams
2%	1.2 milligrams	6.4 grams
4%	2.4 milligrams	12.8 grams
8%	4.8 milligrams	25.6 grams

Table 1. Estimated amounts of transgenic protein per seed or per one liter of soy powder

Retail commercial cost of a theoretical recombinant protein per milligram	Linear extrapolation of commercial value for 12.8 grams of the theoretical recombinant protein
$10	$128,000
$100	$1,280,000
$1000	$12,800,000
$10,000	$128,000,000

Table 2. Extrapolation of the potential value for one liter of soy protein powder expressing a particular transgenic protein at a level of 4%

Further extrapolations can be made for bulk production of transgenic soybeans in secure greenhouses. Such greenhouses provide containment and controlled conditions which allow optimal growth for maximal yields. Using such conditions, it is not difficult to obtain 60 bushels of soybeans per greenhouse acre (see http://www.soystats.com). The industry standard for an average weight of a bushel of soybeans is 60 pounds, with an average quantity of 2500 seeds per pound. This computes to approximately 9 million transgenic soybean seeds per acre. At an average weight of 150 milligrams per seed, this represents approximately 1,350 kilograms of seeds or 540 kilograms of total soy protein. As shown in Table 3, at already achieved 4% expression levels, this calculates to 21.6 kilograms of transgenic protein per greenhouse acre.

Percentage expression of the transgenic protein	Average estimated quantity of soybean protein per greenhouse acre	Calculated quantity of transgenic protein per greenhouse acre
1%	540 kilograms total soy protein	5.4 kilograms of transgenic protein
2%	540 kilograms total soy protein	10.8 kilograms of transgenic protein
4%	540 kilograms total soy protein	21.6 kilograms of transgenic protein
8%	540 kilograms total soy protein	43.2 kilograms of transgenic protein

Table 3. Estimated quantities of transgenic protein per greenhouse acre based on percent expression

It should be noted that we have included 8% expression levels of the transgenic protein in Tables 1 and 3. While we have yet to achieve such levels in our laboratory, the use of transgenic soybeans to express foreign proteins is evolving [e.g. (Schmidt & Herman, 2008)]. Recent DNA sequencing of the soybean genome (Hyten et al., 2010; Schmutz et al., 2010), the development of better promoters, engineering of high protein-expressing seeds, and other coming advances, promise to further increase the efficiency of expression of transgenic proteins using this platform technology. Therefore it seems reasonable to conclude that future advances will only facilitate our ability to increase the level of protein expression.

While future advances promise even higher percentages of transgenic protein expression, current levels will permit contained greenhouse to produce bulk quantities of particular proteins. Using theoretical costs for a recombinant protein as before, Table 4 extrapolates the potential value for an acre of greenhouse grown soybeans at an expression level of 4%. Again, these numbers serve to underscore the potential for high capacity of transgenic protein production using a confined growth space.

Retail commercial cost of a theoretical recombinant protein per milligram	Calculated quantity of transgenic protein per greenhouse acre at 4% expression	Linear extrapolation for the potential value of 21.6 kilograms of the theoretical recombinant protein
$10	21.6 kilograms of transgenic protein	$216,000,000
$100	21.6 kilograms of transgenic protein	$2,160,000,000
$1,000	21.6 kilograms of transgenic protein	$21,600,000,000
$10,000	21.6 kilograms of transgenic protein	$216,000,000,000

Table 4. Extrapolation of the potential value for an acre of greenhouse soybeans expressing a particular transgenic protein at a level of 4%

5. Greenhouse containment for growing transgenic soybeans

As shown in Tables 1, 2, 3, and 4, it is clear that at expression levels already obtained (e.g. 1% - 4%) there will be little reason to grow such transgenic plants in open fields. At production levels of 3-10 kilograms per acre, and with the potential for 3 separate growing seasons per year, propagation in contained greenhouses would impart few limitations, even for bulk production.

Despite the fact the soybean plant is self pollinating, secure greenhouse growth would provide additional containment by eliminating transgenic seed escape (Traynor, 2001). As will be discussed below, processing of soybean seeds to soy powder can easily be accomplished in the greenhouse prior to removal of this non-germinating material for formulation and/or protein purification. Such containment procedures, therefore, provide management of these genetically modified crops from release into the environment.

Propagation of transgenic soybean plants in secure greenhouses also provides advantages in addition to containment. Therapeutic and diagnostic proteins must be produced with good manufacturing practices (GMP) as dictated by approval agencies. Greenhouse growth

allows for ease of standard operating procedures to be implemented with respect to growth conditions, disease and pathogen monitoring, harvesting, and quality control. Stated simply, there are numerous advantages for greenhouse growth, and very few reasons for proposing open field propagation of transgenic soybean plants, that are destined to produce proteins for therapeutic and diagnostics purposes.

6. Costs to grow an acre of transgenic soybeans in contained greenhouses

The efficiency with which soybean plants can be grown in the field or in greenhouses is well documented (Traynor, 2001). This understanding of maximizing crop yields serves to reduce the cost of producing seeds from transgenic plants. Current costs per acre to plant and harvest soybeans from open fields range from $300 to $600 per acre depending upon planting conditions, treatments, geographic area, etc. It has been estimated that greenhouse containment for soybean growth at Biosafety Level 2 (BSL-2) would increase this cost approximately 20 fold (Traynor, 2001). Despite this increased cost and the benefits that go with greenhouse production, the total expense for planting and harvesting remains a reasonable $6,000 to $12,000 per acre. Based on these estimates, it is possible to project a cost per milligram for production and harvest of soybeans expressing a transgenic protein using BSL-2 conditions (Table 5).

Percentage expression of the transgenic protein	Calculated quantity of transgenic protein per greenhouse acre	Costs per milligram of protein assuming $12,000 production costs
1%	5.4 kilograms of transgenic protein	$ 0.002 per milligram
2%	10.8 kilograms of transgenic protein	$ 0.001 per milligram
4%	21.6 kilograms of transgenic protein	$ 0.0005 per milligram
8%	43.2 kilograms of transgenic protein	$ 0.00025 per milligram

Table 5. Cost production projections per milligram of transgenic protein using BSL-2 greenhouse conditions and assuming $12,000 per acre total production costs

It is clear to see the potential for large scale production of transgenic proteins at a very low cost of production when one considers such calculations.

7. Formulations which require no purification of the transgenic protein from soy

The cost projections in Table 5 pertain only to the growth and harvesting of soybean seeds expressing the transgenic protein. As with all protein expression systems, additional costs are required for purification of the protein of interest. However, before we discuss purification costs and the advantages of soybean-derived proteins, an intriguing possibility is the use of formulations made from transgenic soybean seeds which would require no purification of the protein prior to its use.

Soybeans can be formulated into a variety of consumables. Simply grinding soybeans into a powder (e.g. soy powder) is sufficient for some agricultural purposes, including the addition to feedstocks. However soy powder is routinely solvent extracted to remove oils, followed by moist heat and drying to produce soybean meal, which can also be consumed. Other formulations include soy flour which is made from finely ground, defatted beans, followed by removal of solvents and drying. Perhaps the most recognizable formulation is soy milk where whole beans can be ground and mixed with heated water to produce this consumable product. These are just a few examples of the consumable formulations that can be made from soybeans, and serve as a starting point for discussing how such formulations might be useful in human therapeutics.

Oral therapies with recombinant proteins have been suggested over the years. Some advantages of such therapies include the ease of administration, patient compliance, safety, and no requirement for medical personnel, among others. Some problems with oral protein therapy include the degradation of proteins as they pass through the gastrointestinal tract and the limited bio-availability to the blood stream and other organs. Despite these limitations, the advantages of oral therapies have spawned continued interest in commercial development by attempting to overcome the shortcomings using technological advances (Karsdal et al., 2010).

Theoretically, formulations made from transgenic soybean seeds expressing a particular protein may represent one technological advance that can be applied to some oral therapeutics for human or agricultural use. For discussion, consider the advantages of expressing a theoretical therapeutic protein in transgenic soybean seeds. The first consideration is purification. Bacterial, yeast, insect, mammalian, or plant culture technologies all require purification since the broth in which they propagate contain enumerable substances which must be removed before the protein can be deemed sufficiently safe to be consumed. Alternatively, the equivalent soybean-derived therapeutic protein would demand no purification. Soy formulations (e.g. soy powder, soy flour, soy milk, etc.) are consumed daily by billions of individuals and agricultural animals with few undue effects, and many health benefits (Messina, 1999; Slavin, 1991). Therefore, unless there is some compelling reason to purify a protein from soy, there would be no reason to suggest that oral therapeutic formulations made from transgenic beans would not be safe for consumption. Stated simply, purification of soybean-derived therapeutic proteins from soy is not, necessarily, a prerequisite for their use.

Such a theoretical scenario becomes immediately untenable if the soybean-derived therapeutic protein is destroyed following formulation from beans to a consumable. Recent success in our laboratory (Oakes, Bost, et al., 2009), and by others (Robic, Farinas, Rech, & Miranda, 2010), has demonstrated that a variety of formulations contain intact transgenic protein. While such empirical, stability studies will need to be performed with each particular transgenic protein, there are enumerable formulation strategies which have already been industrialized for soy consumables. Gentle versus more aggressive manipulations of soybeans have already been defined for their ability to maintain endogenous protein stability. Many of these manipulations are less egregious than those used during purification of recombinant proteins from bacterial, yeast, insect, mammalian, or plant cell culture broth expression systems. With such a diversity of formulations strategies already available for soy, identifying one which maintains transgenic protein stability seems likely.

While individual soybean seeds from the same and different plants can vary somewhat in the level of transgenic protein expression, this variability is not a major concern when

developing therapeutic dosages. Soy formulations made from a lot of seeds can easily be blended into highly consistent mixtures. This is true for soy milk or soy powder formulations. Following homogenization, a given volume or weight would contain equivalent doses of the protein. Furthermore, it would be easy to standardize different lots to an equivalent dose by the addition of wild type soy milk or soy powder. Therefore the ability to formulate large lots of transgenic seeds into homogenous mixtures alleviates concerns about variability in dosing.

Labile proteins which are destroyed as they pass through the gastrointestinal tract make poor oral therapeutics. However formulations made from transgenic soybeans may have some advantages in protecting the therapeutic protein while passing through the gut. Soy milk formulations have inherent buffering capacity (Lutchman et al., 2006; Park, 1991), which helps to neutralize the acidic environment in the gastrointestinal tract. Such acid-neutralizing capacity of such formulations might aid in protein stability. Furthermore, current formulations made from transgenic soybean seeds would contain 1% to 4% of the therapeutic protein, leaving the remaining 96% to 99% protein as soy. These excess soy proteins in milk or powder formulations may facilitate the passage of proteins through the gut by competitively limiting protease activity. Taken together, the inherent properties of soy formulations would likely provide some protection against degradation of the therapeutic protein following oral treatment.

The composition of oral therapeutics is tightly regulated to assure that no significant level of toxins, harmful substances, or transmissible agents are present. Bacterial, yeast, insect, mammalian, or plant culture technologies all require purification since the broth in which they propagate contain enumerable substances which must be removed before the protein can be deemed sufficiently safe to be consumed. Alternatively, soy protein and soy milk formulations for infant, adult, and agricultural use are safe to consume and have significant nutritional benefit (Messina, 1999; Slavin, 1991). In fact, soy formulations are so safe that they are routinely fed to infants with little side effects (Badger, Ronis, Hakkak, Rowlands, & Korourian, 2002; Motil, 2000; Seppo et al., 2005). Such widespread use suggests that therapeutic-containing soy formulations would not pose any significant risk, supporting the notion that transgenic soybean formulations would not require purification of the protein prior to treatment. While glycosylation of soybean-derived transgenic proteins will likely differ from human or animal glycosylation patterns (Liu, 1999), this too is unlikely to be problematic. Glycosylation patterns of normal soy proteins will likely be similar to those glycoslyation patterns observed on the transgenic proteins since these post translational modifications will be carried out by the same cellular machinery. If true, then humans and animals have been exposed to these particular glycosylation patterns in consumable soy formulations for a long time without undue effect. Therefore there is little reason to believe that formulations derived from transgenic soybeans will contain any harmful substances which would preclude their use in oral therapies.

One recent study (Oakes, Piller, & Bost, 2009) has demonstrated the safety of oral soy formulations. While soy proteins themselves are considered safe, it was not altogether clear whether vaccine formulations which contain soy proteins combined with immune stimulating molecules would induce an anti-soy immune response. Using this mouse model, we found that no significant antibody response against soy could be detected, even when these animals were given soy protein formulations containing the oral adjuvant, cholera toxin. The likely reason for this outcome is that most animal chows contain soy protein as a significant ingredient. In fact, these mice were consumed soy in their food chow throughout

their life, as is the case for many agricultural animals and most humans. It is likely that these mice did not respond to soy proteins since it is a normal dietary constituent. This study (Oakes, Piller, et al., 2009), therefore, demonstrated the safety of oral soy formulations, even when combined with immune stimulating agents.

A final consideration for oral therapeutics is targeting the protein so that it is efficiently adsorbed from the gut, allowing therapeutic concentrations to be achieved in the target organ. There have been many strategies for targeting (Lambkin & Pinilla, 2002), including the use of fusion proteins which effectively bind particular cells and facilitate passage from the lumen or binding to a target organ. Unfortunately, fusion proteins can be large in size and complex in their folding making some systems inappropriate or incapable of expressing. As noted previously (see section 2), transgenic soybean seeds have been able to express large, complex proteins (Zeitlin, et al., 1998) which have binding characteristics. Therefore it seems likely that such fusion proteins which target cells and organs would be ideal candidates for this platform technology. Once again, there would seem to be little advantage of purifying such fusion proteins from soy formulations prior to their use since targeting would likely be unaffected by the presence of native soy proteins.

In summary, we have presented the intriguing possibility that it might be possible to use formulations made from transgenic soybean seeds containing a therapeutic protein without the need for prior purification. In fact, there may even be some significant advantages of soy formulations. Taken together, an intriguing question for some soybean-derived oral protein therapeutics might be: Why purify?

8. Soybean-derived proteins: the potential for simple, low cost protein production

For most therapeutic and diagnostic applications, GMP-certified proteins will require purification. Isolation schemes for such proteins can be fraught with technical challenges and regulatory oversight. While concerns about the presence of transmissible agents, trace contaminants, and necessary additives to increase protein stability continue to mount, technological advances which minimize any of these concerns would represent true breakthroughs. The use of soybean formulations to overcome present limitations, and possibly eliminate the need for purification of transgenic proteins for some applications, may represent one such solution.

When considering protein purification schemes, it is useful to examine the starting material from which the protein must be isolated. For those proteins which must be isolated from human or animal tissues or fluids, concerns about transmissible agents, complexity of the starting material, and variability of purity by suppliers exist. For bacterial expression, removal of toxins, and denaturation and renaturation requirements can be costly. For yeast, insect, mammalian, or plant cell culture broths, the complexity of the starting material and the high volume to protein biomass ratios can be problematic. Regardless of the starting material or expression system, a typical isolation of a desired protein requires removal of hundreds to thousands of irrelevant proteins and similar numbers of other contaminants.

Conversely, soybean seeds contain a very limited number of protein species (Mooney, Krishnan, & Thelen, 2004; Natarajan, 2010). Studies have estimated 40-50 different protein spots on 2D gel electrophoresis, with about half of these representing aggregates, subunits, or truncated forms of a few dominant soybean proteins (e.g. glycinin and conglycinin). The ability to remove such a limited spectrum of soy proteins from the desired transgenic

protein would not seem to be a difficult task. Furthermore, these endogenous soy proteins are the same for every transgenic soybean line regardless of the foreign protein being expressed. Therefore it seems logical that once purification schemes are established for reference proteins with different physical properties (e.g. proteins with acidic versus basic isoelectric points), it will be possible to quickly adapt these standard operating procedures to new transgenic proteins.

Due to the high quantity of transgenic protein per liter of soy powder (see Table 1), the starting biomass for protein purification is small and can be easily controlled. This reduces the need for concentration and further aids in the initial steps of the purification procedure.

In summary, the inherent simplicity of the soybean seed and its high protein to biomass ratio should lead to the development of straightforward and cost-effective standard operating procedures for the purification of transgenic proteins.

9. Low-tech propagation and production reduces costs

While technological advances have expanded the protein therapeutic and diagnostic markets, many of these achievements require significant facilities and expertise, and, therefore, increased costs. Production facilities are often highly specialized, and these buildings are expensive to maintain. Similar overhead is required for protein purification, where the low protein biomass of some expression systems dictates the quantities of material which must be processed. Contrast these requirements with those necessary for propagation of an acre of transgenic soybean plants. BSL-2 greenhouses are cheap and easy to maintain (Traynor, 2001), relative to specialized buildings for recombinant protein production. Technical support for the agricultural care of soy plants and greenhouse harvesting of beans requires substantially less training and expertise than maintaining prokaryote or eukaryote cell cultures. Furthermore, the starting material for protein purification (e.g. soy powder) is easy to obtain, and has extended pre-processing storage times. These characteristics are in stark contrast to the procedural and time constraints placed on harvesting cultured cells or media.

Stocks of prokaryotic and eukaryotic cell lines expressing a particular recombinant protein must be maintained, and this low temperature storage requires significant equipment and oversight. Conversely, founder seed stocks of stably expressing transgenic soybean lines can be stored for years at room temperature or with refrigeration, and retain their ability to germinate if kept dry. This simplistic, low-tech method of maintaining stocks and lines reflects a fundamental characteristic of soybean seeds, and takes advantage of their ability to remain viable for long periods of time under ambient conditions.

10. Reduced risk of contamination with toxins and transmissible agents

Regulatory agencies require lots of therapeutic and diagnostic reagent proteins to be screened for the absence of a variety of toxins and transmissible agents. These include molecules like lipopolysaccharide and other bacterial contaminants, especially when prokaryotic expression systems are being used. For proteins isolated from human or animal tissues or cell expression systems, screening for the absence of transmissible agents (e.g. HIV, Hepatitis B, etc.) is an important quality control. The number and type of screenings will likely increase as the presence of prion proteins and other agents continue to be added to the list of possible contaminants. Screening for the absence of each of these toxins or

transmissible agents adds to the cost of production and increases possible product liability when these expression systems are used.
Soybean-derived proteins pose little risk of spreading transmissible human or animal diseases. This is clear from the billions of infants, adults, and agricultural animals that routinely consume soy products. Screening for toxins or human or animal transmissible agents will likely be significantly less for lots of soybean-derived proteins. In fact, any contaminants present in soybean-derived therapeutics or diagnostics would most likely be introduced during the purification process. Ultimately, this will reduce the cost of production and limit product liability for soybean-derived proteins.

11. Waste disposal following production

An increasing concern and cost-consideration for protein expression technologies is the amount and composition of waste produced during production. Regulatory requirements for decontamination and disposal of such waste into the environment add to the expense of production. This expense not only applies to the toxic byproducts produced, but to the equipment which is required. It is often more cost-efficient to replace machinery than attempt to clean and reuse it. It is highly likely that such waste disposal costs will only increase in the future.
As transgenic soybean plants grow and mature, they proceed from vegetative stage to mature seed formation. As the vegetative plant dies, the seed matures, decreases its moisture content, and is then ready for harvesting. What remains after seeds have been extracted is non-germinating plant material. This material can easily be shredded, and for all intents and purposes is mulch. Stated simply, the environmental waste which requires decontamination prior to disposal for current expression systems will not be a significant cost consideration for soybean-derived proteins expressed using this platform technology.

12. Ease of scaling

Transgenic soybeans as a platform technology for protein expression provide a unique opportunity for scaling production. The only real limitation for dictating the magnitude of soybean-derived protein produced is the availability of BSL-2 greenhouse space. The production facilities available for expanding other culture-based expression systems are more costly and can be specialized. Such limitations threaten to increase future costs for conventional manufacturing. Alternatively, the relative low cost of greenhouse construction and maintenance, the lack of geographical limitations for facility locations, and the low-tech requirements for production and propagation provide a compelling rational for expanding platform expression technologies to include transgenic soybeans.

13. Long term soy powder storage: Uncoupling production and purification

Expiration dates for therapeutics and diagnostic reagents are required by regulatory agencies. The functional lifespan of proteins purified from any expression system depends largely upon the chemical nature of the particular protein itself, and upon the form in which that particular protein is prepared. In general, labile proteins have shorter half-lives, and lyophilized preparations have longer expiration dates than proteins in aqueous solutions. To date, these characteristics have been some of the predominant limitations that increase

costs and require product replacement. Unfortunately, once recombinant proteins are purified, they have the half-lives and expiration dates dictated by these chemical and physical properties.

Stockpiling purified proteins can be a risky proposition if such products cannot be used prior to their expiration dates. With most current expression technologies, the purification of a protein follows soon after the propagation of cells or cultures containing the desired protein. This necessitates coupling production to purification schemes in a limited time frame. Problems with production or purification can result in the lack of available product. Alternatively, an exaggerated estimate of product need may lead to overstocked, expired product.

Platform expression technologies which allow the processes of production and purification to be uncoupled would provide a significant advantage. Consider the example of transgenic soybeans expressing a desired protein. A manufacturer could produce an excess of transgenic seeds and convert them to powder. At this point in production, the desired amount of soy powder could be taken for protein purification, while excess soy powder could easily be stored for months to years until needed. If there were problems with protein purification, this excess soy powder would be immediately available. If product demand had initially been underestimated, the excess powder could be used for protein purification forthwith. Cultivating an excess of transgenic soy allows stockpiling of seed material which could be used immediately or well into the future. Such flexibility adds significant value, since this low-tech, long term storage of crude product allows for future use without delay. In this manner, product expiration dates do not necessarily have to dictate production schedules.

14. Stability of transgenic proteins in seeds and soy powder: Time

One function of the soybean seed is to package and store proteins. This allows seeds to be dormant for years in ambient, dry conditions and still maintain the ability to germinate. We anticipated that this fundamental property of soybean seeds would also allow transgenic proteins to be stable for years once expressed.

A recent publication demonstrates that this is in fact the case for one model protein (Oakes, Bost, et al., 2009). Seeds were stored at room temperature and at ambient conditions for more than 4 years, and demonstrated no detectable degradation of the transgenic protein expressed in these seed lines. We have yet to reach a time when seed-contained transgenic proteins have demonstrated significant degradation compared to their stability in newly harvested seeds if the aged seeds are kept dry. Similar results have been obtained when analyzing the stability of soy powder formulations contaning transgenic proteins. While this first publication demonstrates the results for one model protein, we have since conducted studies in our laboratories to show stability for several additional transgenic proteins that were stored under ambient conditions for years (data not shown). Such empirical studies will need to be performed with each new protein expressed. However, the work performed to date suggests that an extended stability over time of transgenic proteins expressed in soybean seeds is a fundamental characteristic of this platform technology.

15. Stability of transgenic proteins in seeds and soy powder: Formulations

In the same publication (Oakes, Bost, et al., 2009), we demonstrated the ability to formulate transgenic soybeans into milk and powders without the destruction of the expressed

protein. Varying combinations of heating, grinding, and solvent extractions were used to demonstrate the ability to formulate this model protein without degradation or destruction. It should be noted that the processes used in this work mimicked some industry standards, suggesting that manufacturing equipment and operating procedures already in place may be adapted for use. Furthermore, we have performed similar studies with other transgenic proteins and have had success in formulating them as well while maintaining stability (data not shown). Therefore, to date, a variety of procedures which formulate soybeans into consumables did not result in the destruction of the transgenic protein. This property may not be true for every soy extraction procedure, and may not be true for every foreign protein expressed. However with the diversity of extraction methods available, it should be possible to tailor manufacturing toward optimal formulations.

16. Stability of transgenic proteins in seeds and soy powder: No cold chain requirement

We have also demonstrated the ability of transgenic soy powder and milk to be shipped internationally at ambient temperatures without degradation of the protein (Oakes, Bost, et al., 2009). These studies not only demonstrate that refrigeration is unnecessary, but suggest that the process of production and protein purification can be uncoupled across continents. For example, production of transgenic seeds and processing to non-germinating powder might occur in one country. The soy powder could then be shipped internationally and stored until formulated or until the transgenic protein is isolated. Such a scenario would add manufacturing flexibility, and therefore value, for therapeutic and diagnostic proteins sent abroad. This should be especially true for emerging markets. The feasibility and success of such endeavors depends upon one fundamental property of the transgenic soybean seed; namely the ability to formulate, store, and ship material while maintaining protein stability without the requirement for cold storage.

17. A green technology

Safety concerns for recombinant protein expression technologies encompass not only product liabilities, but also the environmental impact of production. Oxygen consumption and the generation of toxic byproducts are becoming increasingly scrutinized when evaluating expression platforms. Perhaps one of the most attractive attributes of transgenic soybeans as a platform for protein expression is that of being a green technology. Growth of these plants consumes carbon dioxide and releases oxygen, which contrasts with most other prokaryotic and eukaryotic cell culture systems. Thus, the environmental-friendly production of proteins using this green technology is another inherent advantage of transgenic soybeans.

18. Future directions

Translation of transgenic soybeans as a platform technology toward commercialization is developing; however there are several hurdles which must be overcome. Standard operating procedures for growth, formulation, purification, and storage of soybean-derived powder, milk and purified proteins using GMP that are consistent with regulatory agency standards must be delineated. In addition, product data for stability, expiration dates, purity, content

and substantial equivalence must also be defined. Future success with these endeavors will be required to move this novel platform technology toward market acceptance, and ultimately to product commercialization.

19. Conclusion

Transgenic soybean seeds represent a unique platform for the expression and accumulation of a desired protein. The advantages of this protein expression system stem from basic seed biology and their ability to express high levels of glycosylated protein for storage over time. For those applications which require high protein expression in a form that is stable in the environment at very low cost, transgenic soybean seeds provide significant advantages.
As protein expression technologies expand and evolve, a consideration of the optimal platform for each protein produced seems sensible. Transgenic soybean seeds represent an emerging technology with significant advantages for expressing some recombinant proteins.

20. References

Badger, T. M., Ronis, M. J., Hakkak, R., Rowlands, J. C., & Korourian, S. (2002). The health consequences of early soy consumption. *J Nutr, 132*(3), 559S-565S.

Brondyk, W. H. (2009). Selecting an appropriate method for expressing a recombinant protein. *Methods Enzymol, 463*, 131-147.

Cregg, J. M., Tolstorukov, I., Kusari, A., Sunga, J., Madden, K., & Chappell, T. (2009). Expression in the yeast Pichia pastoris. *Methods Enzymol, 463*, 169-189.

Cunha, N. B., Araujo, A. C., Leite, A., Murad, A. M., Vianna, G. R., & Rech, E. L. (2010). Correct targeting of proinsulin in protein storage vacuoles of transgenic soybean seeds. *Genet Mol Res, 9*(2), 1163-1170.

Ding, S. H., Huang, L. Y., Wang, Y. D., Sun, H. C., & Xiang, Z. H. (2006). High-level expression of basic fibroblast growth factor in transgenic soybean seeds and characterization of its biological activity. *Biotechnol Lett, 28*(12), 869-875.

Garg, R., Tolbert, M., Oakes, J. L., Clemente, T. E., Bost, K. L., & Piller, K. J. (2007). Chloroplast targeting of FanC, the major antigenic subunit of Escherichia coli K99 fimbriae, in transgenic soybean. *Plant Cell Rep, 26*(7), 1011-1023.

Geisse, S., & Fux, C. (2009). Recombinant protein production by transient gene transfer into Mammalian cells. *Methods Enzymol, 463*, 223-238.

Hellwig, S., Drossard, J., Twyman, R. M., & Fischer, R. (2004). Plant cell cultures for the production of recombinant proteins. *Nat Biotechnol, 22*(11), 1415-1422.

Hyten, D. L., Cannon, S. B., Song, Q., Weeks, N., Fickus, E. W., Shoemaker, R. C., et al. (2010). High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics, 11*, 38.

Jarvis, D. L. (2009). Baculovirus-insect cell expression systems. *Methods Enzymol, 463*, 191-222.

Karsdal, M. A., Henriksen, K., Bay-Jensen, A. C., Molloy, B., Arnold, M., John, M. R., et al. (2010). Lessons Learned From the Development of Oral Calcitonin: The First Tablet Formulation of a Protein in Phase III Clinical Trials. *J Clin Pharmacol*.

Lambkin, I., & Pinilla, C. (2002). Targeting approaches to oral drug delivery. *Expert Opin Biol Ther, 2*(1), 67-73.

Lienard, D., Sourrouille, C., Gomord, V., & Faye, L. (2007). Pharming and transgenic plants. *Biotechnol Annu Rev, 13*, 115-147.

Liu, K. (1999). *Soybeans: Chemistry, Technology, and Utilization*. Gaithersburg, MD: Aspen Publishers, Inc.

Lutchman, D., Pillay, S., Naidoo, R., Shangase, N., Nayak, R., & Rughoobeer, A. (2006). Evaluation of the buffering capacity of powdered cow's, goat's and soy milk and non-prescription antacids in the treatment of non-ulcer dyspepsia. *S Afr Med J, 96*(1), 57-61.

McPherson, R. M., & MacRae, T. C. (2009). Evaluation of transgenic soybean exhibiting high expression of a synthetic Bacillus thuringiensis cry1A transgene for suppressing lepidopteran population densities and crop injury. *J Econ Entomol, 102*(4), 1640-1648.

Messina, M. J. (1999). Legumes and soybeans: overview of their nutritional profiles and health effects. *Am J Clin Nutr, 70*(3 Suppl), 439S-450S.

Mooney, B. P., Krishnan, H. B., & Thelen, J. J. (2004). High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. *Phytochemistry, 65*(12), 1733-1744.

Moravec, T., Schmidt, M. A., Herman, E. M., & Woodford-Thomas, T. (2007). Production of Escherichia coli heat labile toxin (LT) B subunit in soybean seed and analysis of its immunogenicity as an oral vaccine. *Vaccine, 25*(9), 1647-1657.

Motil, K. J. (2000). Infant feeding: a critical look at infant formulas. *Curr Opin Pediatr, 12*(5), 469-476.

Natarajan, S. S. (2010). Natural variability in abundance of prevalent soybean proteins. *Regul Toxicol Pharmacol*.

Oakes, J. L., Bost, K. L., & Piller, K. J. (2009). Stability of a soybean seed-derived vaccine antigen following long-term storage, processing and transport in the absence of a cold chain. *Journal of the Science of Food and Agriculture, 89*(13), 2191-2199.

Oakes, J. L., Piller, K. J., & Bost, K. L. (2009). An antibody response to cholera toxin, but not soy proteins, following oral administration of adjuvanted soybean formulations. *Food and Agricultural Immunology, 20*(4), 305-317.

Padgette, S. R., Kolacz, K. H., Delannay, X., Re, D. B., Lavallee, B. J., Tinius, C. N., et al. (1995). Development, Identification, and Characterization of a Glyphosate-Tolerant Soybean Line. *Crop Science, 35*(5), 1451-1461.

Park, Y. W. (1991). Relative buffering capacity of goat milk, cow milk, soy-based infant formulas and commercial nonprescription antacid drugs. *J Dairy Sci, 74*(10), 3326-3333.

Piller, K. J., Clemente, T. E., Jun, S. M., Petty, C. C., Sato, S., Pascual, D. W., et al. (2005). Expression and immunogenicity of an Escherichia coli K99 fimbriae subunit antigen in soybean. *Planta, 222*(1), 6-18.

Qi, Q., Huang, J., Crowley, J., Ruschke, L., Goldman, B. S., Wen, L., et al. (2010). Metabolically engineered soybean seed with enhanced threonine levels: biochemical characterization and seed-specific expression of lysine-insensitive variants of aspartate kinases from the enteric bacterium Xenorhabdus bovienii. *Plant Biotechnol J*.

Rao, S. S., & Hildebrand, D. (2009). Changes in oil content of transgenic soybeans expressing the yeast SLC1 gene. *Lipids, 44*(10), 945-951.

Robic, G., Farinas, C. S., Rech, E. L., & Miranda, E. A. (2010). Transgenic soybean seed as protein expression system: aqueous extraction of recombinant beta-glucuronidase. *Appl Biochem Biotechnol, 160*(4), 1157-1167.

Schmidt, M. A., & Herman, E. M. (2008). Proteome rebalancing in soybean seeds can be exploited to enhance foreign protein accumulation. *Plant Biotechnol J, 6*(8), 832-842.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature, 463*(7278), 178-183.

Seppo, L., Korpela, R., Lonnerdal, B., Metsaniitty, L., Juntunen-Backman, K., Klemola, T., et al. (2005). A follow-up study of nutrient intake, nutritional status, and growth in infants with cow milk allergy fed either a soy formula or an extensively hydrolyzed whey formula. *Am J Clin Nutr, 82*(1), 140-145.

Slavin, J. (1991). Nutritional benefits of soy protein and soy fiber. *J Am Diet Assoc, 91*(7), 816-819.

Traynor, P. L., Adair, D., Irwin, R. (2001). *A practical guide to containment: Greenhouse research with transgenic plants and microbes*. Blacksburg, VA: Information Systems for Biotechnology.

Wang, X., Wang, Y., Tian, J., Lim, B. L., Yan, X., & Liao, H. (2009). Overexpressing AtPAP15 enhances phosphorus efficiency in soybean. *Plant Physiol, 151*(1), 233-240.

Zeitlin, L., Olmsted, S. S., Moench, T. R., Co, M. S., Martinell, B. J., Paradkar, V. M., et al. (1998). A humanized monoclonal antibody produced in transgenic plants for immunoprotection of the vagina against genital herpes. *Nat Biotechnol, 16*(13), 1361-1364.

Zerbs, S., Frank, A. M., & Collart, F. R. (2009). Bacterial systems for production of heterologous proteins. *Methods Enzymol, 463*, 149-168.