

## Molecular Strategies to Engineer Transgenic Rice Seed Compartments for Large-Scale Production of Plant-Made Pharmaceuticals

Trevor Greenham and Illimar Altosaar

### Abstract

The use of plants as bioreactors for the large-scale production of recombinant proteins has emerged as an exciting area of research. The current shortages in protein therapeutics due to the capacity and economic bottlenecks faced with modern protein production platforms (microbial, yeast, mammalian) has driven considerable attention towards molecular pharming. Utilizing plants for the large-scale production of recombinant proteins is estimated to be 2–10% the cost of microbial platforms, and up to 1,000-fold more cost effective than mammalian platforms (Twyman et al. *Trends Biotechnol* 21:570–578, 2003; Sharma and Sharma, *Biotechnol Adv* 27:811–832, 2009). In order to achieve an economically feasible plant production host, protein expression and accumulation must be optimized. The seed, and more specifically the rice seed has emerged as an ideal candidate in molecular pharming due to its low protease activity, low water content, stable protein storage environment, relatively high biomass, and the molecular tools available for manipulation (Lau and Sun, *Biotechnol Adv* 27:1015–1022, 2009).

**Key words:** Bioreactor, Recombinant protein, Pharmaceutical, Biologics, Transgenic rice, Molecular pharming

---

### 1. Introduction

Biopharmaceuticals are the fastest growing sector within the pharmaceutical industry, with a market value of \$120 billion (US dollars) in 2009 (1–5). These biologics are mainly produced using recombinant technology and established production platforms such as microbial, yeast, or mammalian cell cultures (6). The effectiveness of different platforms is judged primarily on protein yield, posttranslational modifications, ease of downstream purification and the capital requirements needed for commercialization. *E. coli* was the first large-scale protein production host and has several advantages such as cheap fermentation runs,

short generation times, and high titers of recombinant protein. Mammalian cultures (CHO cells predominantly) were introduced to overcome some of the shortfalls of the microbial expression platforms such as the formation of inclusion bodies upon high titers, difficulty in purification due to endogenous contaminants, and most importantly their lack of eukaryotic posttranslational modifications (glycosylation, acylation, disulphide bridge formation, etc.) which are often required for protein folding and function (7). CHO cells can produce recombinant proteins with similar glycoprofiles as native proteins and innovations in target-gene insertion, culture media manipulation and apoptosis inhibition have improved titers to over 5 g/L (8). Currently, CHO cells are the most utilized production platform despite their high infrastructure and process costs (9).

The biopharmaceutical industry has seen considerable growth over the past 5 years and has a forecasted growth between 7 and 15% during the next decade (10). This is attributed to the increasing population of elders concomitant with innovations in drug discovery. The growth is apparent when comparing past and present market demands for biopharmaceuticals. Insulin for example had a global demand of 5,000–6,000 kg/year in 2009 and is expected to reach 16,000 kg/year by 2012 (2). The growth of this industry has caused extreme shortages in manufacturing capacity. The high capital requirements related with the aforementioned platforms have put an inhibitory effect on supplying adequate amounts of biopharmaceuticals, prompting other production strategies to be investigated.

With the advent of plant transformation technology (11–13), plants have proven to be feasible bioreactors for the large-scale production of recombinant proteins (14). Due to their lengthy life cycles (3–6 months for rice), the true economic advantages of plants are obtained when dealing with large-scale applications. These advantages are in terms of production costs, scalability, product safety, ease of storage and distribution, none of which can be matched by any current bacterial or mammalian production platform (15–17). It is estimated that using achieved yields, enough hepatitis B-antigen to vaccinate all the infants worldwide could be produced on approximately 200 acres of land (18).

---

## 2. Rice Seed as an Expression Host

The seed has emerged as one of the most prominent plant organs for recombinant protein production. For plants, the seed evolved to serve as a storage organ for the long-term maintenance of nutritional components required for the germination of the new sporophytic generation (19). Rice seeds are composed of 7–8%

protein and 92–93% starch. It has been shown that throughout the dormancy period of the rice seed, its storage proteins remain intact and functional (20), meaning it should be a suitable area for the stable deposition of recombinant proteins. This is a considerable advantage over plant platforms that accumulate protein in leafy tissue (e.g., tobacco), as they are more prone to hydrolytic activity and have drastically shorter shelf life. Rice crops are self-pollinating, which ensures that no genetic material is gained or lost, and the gene coding for the protein of interest remains present in each new generation. This provides an advantage from a regulatory standpoint as it reduces the chances of environmental contamination. The seed is also being investigated as an edible vaccine (21–26). Edible vaccines show economic promise because it eliminates the need for downstream purification, reducing capital requirements needed for commercialization by over 60%.

Rice is the staple food worldwide, cultivated in over 100 countries on more than 150 million hectares of land (19). The familiarity with the agronomy and nutritional values of rice, along with GRAS (generally recognized as safe) designation by the Food and Drug Administration make it a strong candidate for the large-scale production of biopharmaceuticals.

---

### **3. Strategies to Boost Recombinant Protein Expression in Rice Seeds**

The endosperm is the main storage compartment for rice and accounts for over 80% of the total seed weight, thus it is the most enticing site for protein accumulation. The protein composition of the endosperm is composed of 60–70% glutelins, 25–30% prolamins, 5–10% globulins, and 0–5% others. The highest yields seen in rice seeds is a result of utilizing the parameters from natively high expressing seed storage proteins. Protein yields have gotten considerable attention as they play a key role in the capital performance of the production platform. The most common strategies to achieve higher yields of recombinant protein in rice seeds are by optimizing the promoter region, the untranslated regions, translation efficiency and subcellular localization/targeting of the target protein (Table 1).

#### **3.1. Promoters**

An early process in recombinant protein generation, transcription is one of the most important areas in determining protein accumulation. In industry, the DNA constructs used for transformation are synthesized *de novo* from oligonucleotides, providing complete control over its sequence.

Promoters are stretches of DNA found upstream of a gene's coding sequence that interacts with transcription factors and allows RNA polymerase II to bind. In plants, at least seven different

**Table 1**  
**Strategies used for expression of biomedical recombinant proteins in rice seeds**

Rice expression strategy	Target protein	Accumulation level	References
Glutelin B-1 promoter Glutelin B-1 signal peptide ER-retention signal KDEL Glutelin B-1 terminator	Allergenic T-cell epitope peptides	60 µg/grain	Takagi et al. (37)
Globulin promoter NosTer	Glucagon-like peptide-1 (GLP-1)	50 µg/20 mg seed	Sugita et al. (83)
Glutelin B-1 promoter Glutelin A-2 5'UTR N-terminus GluA-2 acidic subunit fusion C-terminus GluA-2 basic subunit fusion Glutelin B-1 terminator	Cedar Pollen Allergen Cry j1	15% total soluble protein	Yang et al. (91)
Glutelin-1 promoter Glutelin-1 signal peptide NosTer	Human GM-CSF	1.3% total soluble protein	Sardana et al. (34)
AGPase large subunit promoter Glutelin B-1 signal peptide ER-retention signal KDEL NosTer	7 Crp peptide	10 µg/grain	Takaiwa et al. (92)
Glutelin-13A promoter Glutelin-13A signal peptide NosTer	Human IGF-1	6.8 ± 0.5% total soluble protein	Xie et al. (93)
Bx17 HMW glutenin endosperm-specific promoter First intron of rice actin1 gene Heat-labile enterotoxin B subunit bacterial signal peptide ER-retention signal (SEKDEL) NosTer	Cholera Toxin B subunit	2.1% total soluble protein	Oszvald et al. (94)

Glutelin A-2 promoter N-terminus GluA-2 fusion NosTer	Type II-collagen tolerogenic peptide	1 µg/seed	Hashizume et al. (82)
Glutelin B-1 promoter Glutelin B-1 signal peptide NosTer	Luminal binding protein	15-fold increase relative to endogenous levels	Yasuda et al. (95)
Glutelin B-1 promoter Glutelin B-1 signal peptide ER-retention signal (KDEL) Glutelin B-1 3' UTR	<i>Ascaris suum</i> AsI4 fused to Cholera Toxin B subunit	1.5 µg/grain	Matsumoto et al. (96)
Glutelin-1 promoter Glutelin-1 signal peptide NosTer	Human Lactoferrin	1% seed dry weight	Zhang et al. (38)

transcription factors are required to transcribe a single gene and most of them are dependant on the nucleotide sequence of the promoter. The use of strong constitutive promoters such as the plant cauliflower mosaic virus 35S promoter (27), the rice ubiquitin (28), and actin promoters (29) have been used to drive expression in rice seeds. Although these promoters are known to be highly active in plants, they showed low expression in monocot seeds (under 5% total seed protein) (30, 31). Constitutive promoters also do not allow for much control over the deposition of the recombinant protein, which can negatively effect plant growth and development. Constitutive expression also reduces the opportunity to develop a more cost effective purification strategy that does not rely on prior art.

Seed-specific promoters, such as the ones driving the expression of the major storage proteins (glutelin, prolamin, globulin), have shown to provide higher seed expression levels than when using a strong constitutive promoter (32–34). Using high expressing seed-specific promoters is the simplest way to increase transgene expression. Replacing constitutive promoters with a seed-specific promoter can account for up to a 10-fold increase in seed expression (33). Dozens of rice seed-specific promoters have been characterized for their activity and among them the glutelin B1 (GluB1), glutelin B4 (GluB4), globulin-1 (Glb-1), and 10 and 16 kDa prolamin promoters have shown to give the highest seed expression levels ranging from 6 to 15% of total seed protein (1–2% total seed weight) (33, 35–38).

Engineering promoters, or adding synthetic components to promoters can increase seed-specific expression of the target gene. Conserved plant promoter elements are important regulators of transcription and should be considered when designing synthetic promoters. The most studied elements are the TATA consensus sequence, the transcription initiation site (TS), the 5' untranslated region (5'UTR), and the context of the translational start codon (39). The design of synthetic promoters relies heavily on additions and/or modification to the already high expressing seed promoters.

The most common sequences added to promoters are transcriptional enhancer domains (cis elements). Cis elements are what determine, either independently or synergistically, the spatial and temporal expression of the promoter. The strength of these motifs is dependant on their copy number and location (40). The conserved motifs found in the 5' flanking region of the 27 kDa gamma-zein genes have shown to be critical for endosperm-specific expression (41). These motifs include a 13 bp AACAA motif, a 7 bp GCN4 motif, a 96 bp GCAA motif, and a prolamin box (TG(T/A/C)AAA(G/T)) which are all found within 245 bp upstream of the TS site in glutelin genes and act in concert to direct the expression of their gene(s) (42). The rice 13 kDa prolamin gene has similar upstream combinations of ACGT, GCN4, and AACAA motifs

that direct its endosperm specific expression. These proximal elements when fused to a truncated CaMV35S promoter showed endosperm expression levels comparable to a seed-specific promoter, proving their efficacy to direct endosperm expression. Glutelin genes also contain distal motifs found from -245 bp to -1,302 that account for approximately a sixfold enhancement of their promoter activity (42). The GluA-2 promoter in rice for example has one distal and six proximal cis elements that are responsible for its high expression activity (42). Loss and gain of function experiments have shown that at least three of these cis elements are involved in the endosperm expression of the glutelin genes (42). Cis elements found in other monocot species (e.g., coconut, oats) can be used to direct transgene expression to the rice endosperm, though they have not proven to be more effective than the native seed storage promoter cis acting elements (43).

Inserting global regulatory sequences flanking the transgene can increase expression levels by several fold (2, 44). Some commonly used sequences are those required for stress response or matrix attachment. These sequences place the transgene in a surrounding locus that is suitable for the recruitment of transcription factors. Matrix attachment regions, for example, are AT-rich regions that reduce positional effects by forming chromatin loops, keeping the DNA proximal to the transgene loosely packed and available for transcription (45).

Transacting factors can be utilized to increase transgene expression by either directly interacting with cis elements within the promoter or interacting with other transcription factors, recruiting them to the promoter. Most transacting factors identified in rice are involved in stress response. Some of these include the iron deficiency regulator OSBZ8 which interacts with an abscisic acid response element, the dehydration response factors (DREBs) that interact with DRE elements (40), the iron deficiency IDEF1 which binds the cis element IDE1 (43), and infection response factor rice endosperm bZIP (REB). REB is a transcription factor that binds to promoters containing a REB-responsive element. Incorporation of a REB-responsive element (GCCACGT(A/C)AG) in promoter regions showed increased seed expression by two- to fourfold (46). The alpha-globulin gene promoter contains three of these repeats, each of which can bind REB (47). REB(s) can increase transcription synergistically as homodimers or heterodimers depending on binding site proximity and flanking motif configuration (39). In order for expression to be elevated by transgenic overexpression, the transacting factor(s) must be limiting natively in the plant.

The use of hybrid promoters or a combination of transcriptional elements from more than one promoter has shown to increase seed expression compared to when used independently (48). Combining elements from the wheat puroindoline-b gene promoter and elements from the rice glutelin-1 promoter

showed double the expression of a reporter gene in rice seed (49). The addition of repeated promoter elements, if spaced correctly, can increase seed expression levels while concomitantly reducing the possibility of recombination and positional/silencing effects (50).

### **3.2. Effects of Untranslated Regions**

#### **3.2.1. 5' Untranslated Region**

The 5'UTR is located just upstream of the translation initiation start site and plays an important role in translation (51, 52). Modifying the 5'UTR of a transgene can increase expression levels, though it is difficult to precisely evaluate their efficiency since they work in tandem with many other factors (53). A conserved 5'UTR from the soybean lipoxygenase genes, when fused upstream of a beta-glucuronidase start codon showed an 11-fold increase in expression in rice (54). The 5'UTR region from a tobacco alcohol dehydrogenase gene (NtADH) can act as a translational enhancer for rice transgenes (55). When used upstream of a reporter gene, the three regions of the 5'UTR (F, NF, 221) of NtADH when added to a CaMV35S promoter increased expression levels by 5-, 1.5-, and 2-fold in rice relative to the CaMV35S promoter alone (55). A 5'UTR from *Vigna radiata* aminocyclopropane-1-carboxylate synthase (VR-ACS1) gene has been used as a translational enhancer in plants. This 5'UTR region increased GUS expression by two- to fivefold relative to the control (56). A 66 bp chlorophyll a/b binding gene 5'UTR region has been reported to enhance translation up to eightfold (57). The use of a 5'UTR region from a rice polyubiquitin gene (*rubi3*) has shown its efficacy in enhancing monocot gene expression (58). Rubi3 is an abundant protein involved in protein degradation and cell cycle control. The 1,140 bp 5'UTR of this gene enhances its promoter activity by 20-fold in transient expression assays (59, 60). 5'UTR regions can sometimes be made more active by the use of introns. Intron mediated enhancement (IME) was first defined in 1990 when investigating two maize introns (61). The mechanisms of IME activity are poorly understood, but it has proven to be a useful tool to increase translation of rice transgenes. The highly expressed rice alpha-tubulin genes (*Ostua1*, *Ostua2* and *Ostua3*) have been isolated for expression analyses. Characterization of their upstream 5'UTR intron demonstrated that it is highly responsible for their expression (62). The 5'UTR intron in rice beta-tubulin gene *Ostub16* has also been used to enhance expression levels in rice (63). The *rubi3* 5'UTR contains an intron that increases its translation when present (58). Removing the intron from the *rubi3* 5'UTR showed that it accounted for a twofold enhancement in transcription activity and a 20-fold enhancement in transcript stability, implying their important role of 5'UTRs in mRNA processing and export (64).

#### **3.2.2. 3' Untranslated Region**

The 3' untranslated region (3'UTR), located just downstream of the transcription stop codon, is responsible for pre-mRNA 3'end formation (cleavage and addition of the poly(A) tract) and helps

stabilize the transcript. The poly(A) tail in particular plays an important role in determining transcript stability and function (65), and a poor 3'UTR can greatly reduce transcript stability (66). There are three primary elements that dictate the efficiency of poly(A) signals in plant 3'UTRs: the far upstream elements (FUE), the near upstream elements (NUE), and the cleavage site (67). Incorporating 3'UTRs harboring these elements can increase gene expression (68). Several studies have shown the efficacy of 3'UTRs for increasing expression levels in plants (69, 70). One study using rice as a host demonstrated that the rice glutelin, GluB-1 3'UTR, when used with a ubiquitin constitutive promoter to drive reporter gene expression, caused an increase in recombinant seed protein levels by 1.8- and 4-fold higher when compared to using a nopaline synthase terminator (19, 71). Very few plant 3'UTRs have been investigated for their impact on rice transgene expression, but it is apparent that the choice of a 3'UTR can significantly contribute to high mRNA stability and accumulation levels of a recombinant protein (72, 73).

### **3.3. Protein Accumulation and Storage**

Once a messenger RNA transcript for a recombinant protein is translated, focus shifts towards its stable accumulation. The two main strategies for the stable accumulation of a recombinant protein are targeting it to a subcellular compartment or using a fusion partner (3). Targeting recombinant proteins to subcellular compartments in plants can increase accumulation levels by several fold (2, 16, 19, 74, 75). The deposition of a recombinant protein impacts not only its accumulation levels but also its assembly and posttranslational modifications. The rationale behind fusion proteins is if a high-expressing plant protein can be stably accumulated, then fusing it with a recombinant protein should also result in the same high and stable accumulation (3). If the natively fused protein component resides in a certain compartment, then it can be assumed that is where the fusion partner will reside, although this is not always the case. In plant cells, an expressed protein without a signal sequence or fusion partner will accumulate in the cytoplasm (76). Otherwise, the protein can be targeted for chloroplast, mitochondrial, or secretory pathway (ER, protein bodies, protein storage vacuoles, apoplasts) deposition. For seed-specific localization, targeting a protein towards the secretory pathway via the ER has proven to be the preferred method. The ER has numerous chaperone proteins (BiP, PDI, calnexin, calreticulin) and an oxidizing environment that is suitable for most proteins (3, 76). Targeting a recombinant protein to the secretory pathway often results in over a tenfold increase in recombinant seed protein levels relative to cytoplasmic expression (3, 77–79).

The highest seed accumulation levels have been achieved by targeting recombinant proteins to the starchy endosperm. The endosperm is composed primarily of starch and storage proteins

found in the form of protein bodies (PB) (80). There are two classes of PBs in rice, PBI and PBII. PBII mainly store glutelins and globulins that get sequestered via the Golgi network or vesicular pathway, and PBI store prolamins via ER-derived deposition (19). Recombinant proteins can be tailored with N or C terminal propeptides, internal sequence motifs, signal peptides, or fusion partners to direct targeting to a specific subcellular domain (49). The most utilized targeting sequences are those found within the major seed storage proteins.

Several strategies have been used for the localization of a recombinant protein towards the PBIIs. The N-terminus portion of the 26 kDa rice globulin protein fused to a green fluorescent protein, under the control of a seed-specific promoter resulted in PBII localization (81). The rice glutelin A (GluA) protein has been used as a fusion partner for PBII localization of a type II-collagen tolerogenic peptide. This peptide, when fused to the C-terminus of the GluA-2 deposited mainly in PBIIs with trace amounts found in the ER, and had a seed expression level of 1 µg/seed (82). It is also a common approach to direct recombinant proteins to PBIIs by inserting them within the hyper variable region of the large acidic subunit of rice glutelin (37, 83, 84). The latter has shown accumulation levels of 20–50 µg per seed (83). The N-terminus from glycinin, a soybean seed storage protein can also direct proteins to PBII in rice seeds, showing the conservation in protein sorting mechanisms between plant species (85).

Targeting recombinant proteins to PBIs has proven to be a valid strategy for their high accumulation. PBIs are viewed as one of the most prominent areas for the deposition of a protein tailored for oral delivery. It has been determined both *in vitro* and *in vivo* that prolamins, the main component of PBIs, are highly resistant to gastrointestinal digestion due in part to their hydrophobic properties (43). Directing recombinant proteins towards PBIs usually involves a fusion partner with prolamins or a C-terminus ER-retention signal (86). A green fluorescent protein when fused to the C-terminus of rice prolamins localized within the endosperm's PBIs (87). PBI localization was also observed when a hybrid of seven human T-cell epitopes (7Crp) was fused with GluA-2. The localization was unexpected, as the GluA-2 natively resides in PBIIs, however, the PBI localization was attributed to a free cysteine residue found within the 7Crp that caused it to interact with prolamins via disulfide bridges (86). ER-retention signals (H/KDEL) are involved in the retrieval between the Golgi and ER and localize proteins to PBIs. The ER-retention signal showed an increase in accumulation levels by up to fourfold when compared to the similar construct lacking the C-terminal signal (37, 88). The expression of a cholera toxin B subunit containing the KDEL signal accumulated to level of 30 µg/seed in PBIs. A mature dust mite allergen (Der p 1) was expressed in rice under the control of the GluB-1 promoter, GluB-1 signal sequence, and a KDEL retention

signal. This resulted in PBI accumulation levels of 60  $\mu\text{g}/\text{seed}$ , whereas a yield of 9.5  $\mu\text{g}/\text{seed}$  was obtained when the signal sequence was used independently (86). Combining N-terminal signal peptides with C-terminal ER-retention signals will result in more efficient deposition of recombinant proteins into PBs (37). Targeting recombinant proteins to endosperm-specific organelles via the ER by signal sequence(s), retention signal(s), or fusion partner(s) increases seed accumulation levels by several fold.

There have been a few observations of targeting a recombinant protein within or to the surface of starch granules. Starch granules are indeed an attractive rice seed organelle for the deposition of recombinant proteins because they occupy over 90% of the seed endosperm and are a valuable commodity in several different industries. The promoter and signal sequence from puuroindoline-b, a protein natively found on the surface of starch granules in wheat, were found to direct the expression and localization of human lysozyme in rice seed (49). The localization of the recombinant protein was expected to be to the starch granules, however, this was not the case as analyses showed only PBI and PBII deposition. These results are estimated to be attributed to possible cryptic sequences within the human lysozyme that triggers it to undergo deposition via a rice default pathway, or perhaps lysozyme contains internal PB localization sequences (49). These localization results are similar to those found in maize when the promoter and signal peptide of the maize storage protein (zein) was fused to the bacterial protein LT-B (41). The LT-B protein was deposited solely within the maize starch granules, suggesting that it must have a native internal starch targeting signal, as a C-terminal PB signal peptide was used (41). Although starch granules are the major occupants in rice endosperm, the highest recombinant protein yields recorded remain to be achieved through such PB deposition strategies (Table 1).

### **3.4. Multigene Engineering**

Multigene engineering can be used for the production of multi-meric proteins, manipulation of metabolic pathways, or introduction of foreign metabolic pathways. The most prominent example of multigene engineering in rice is “Golden rice,” which involved the insertion of three genes (phytoene synthase, phytoene desaturase, and lycopene  $\beta$ -cyclase) that resulted in the production of  $\beta$ -carotene (43). The traditional methods used for multigene insertions are crossing two independent transgenic lines, or by performing sequential transformations (transforming a transgenic plant with another transgene). These processes, however, are time and labor intensive and are susceptible to segregation events in later generations due to the result of the different transgenes residing on different loci (86). To insert multiple genes in one generation and on the same locus, co-transformation methods have been developed (84, 87, 88). The most common co-transformation method for rice is using *Agrobacterium*-mediated transformations

with the transgenes on a single T-DNA or inserted using multiple T-DNA fragments (86). Particle bombardment has also shown efficacy in multigene engineering, as it has been reported that 11 different plasmids, each containing a different marker and transgene were found on the same locus and expressed in the subsequent generation (89). Co-transforming plants with five to ten transgenes is now possible with the innovations in T-DNA-based vectors and artificial chromosomes but there are still issues involved with transgene silencing and unpredictable interactions with the host proteome.

---

## 4. Conclusion

Plant molecular pharming has undergone considerable advancements in recombinant protein production and is recognized as a promising opportunity to meet the future demands for biopharmaceuticals. The drastic shortage of protein-based therapeutics has pushed collaborations between industry giants and molecular pharming start-ups as proof-of-concept trials have proven the large-scale efficacy of plants as production hosts (18). Molecular pharming is still in its infancy as there has yet to be a plant-made pharmaceutical approved for commercialization. There are 21 plant-derived pharmaceuticals currently in phases I and II of clinical trials and eleven plant-derived pharmaceuticals that are either in phases III and IV, or are currently being marketed as a medical device or fine chemical (3, 90). The seed, and more specifically the rice seed has emerged as an ideal candidate for the large-scale production of protein therapeutics due to its high protein content, low protease activity, low water content, stable protein storage environment, relatively high biomass, and molecular tools available for manipulation (3). Due to the preservation capabilities of seeds, recombinant seed banks could be generated, providing an economically feasible, timely scale-up of therapeutics in response to changing market demands. With the increasing demand for protein therapeutics and the era of follow-on biologics upon us, molecular pharming as an industry will see considerable growth and attention over the next decade. To fully utilize the advantages of seeds, advancements in protein expression and downstream purification will get the most consideration as they are the key factors that determine the efficacy of a platform's economic performance (50).

## References

1. Twyman RM et al (2003) Molecular farming in plants: host systems and expression technology. *Trends Biotechnol* 21:570–578
2. Sharma AK, Sharma MK (2009) Plants as bioreactors: recent developments and emerging opportunities. *Biotechnol Adv* 27: 811–832
3. Lau OS, Sun S (2009) Plant seeds as bioreactors for recombinant protein production. *Biotechnol Adv* 27:1015–1022

4. IMS Health (2010) Accurate global assessment of a dynamic and complex niche market. IMS global biologics perspectives. IMS Health Incorporated
5. Walsh G (2010) Biopharmaceutical benchmarks 2010. *Nat Biotechnol* 28:917–924
6. Boehm R (2007) Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. *Ann NY Acad Sci* 1102:121–134
7. Karg S, Kallio P (2009) The production of biopharmaceuticals in plant systems. *Biotechnol Adv* 27:879–894
8. Parl J et al (2010) Cellular engineering for the high-level production of recombinant proteins in mammalian cell systems. *Kor J Chem Eng* 27:1042–1048
9. Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* 11:1393–1398
10. Hiller A (2009) Fast growth foreseen for protein therapeutics. *GEN* 29:153–155
11. Cheng XY et al (1998) *Agrobacterium*-transformed rice plants expressing synthetic *cryIA(b)* and *cryIA(c)* genes are highly toxic to striped stem borer and yellow stem borer. *Proc Natl Acad Sci USA* 6:2767–2772
12. Ma JK et al (2003) The production of recombinant pharmaceuticals in plants. *Nat Rev Genet* 4:794–805
13. De Zoeten G et al (1989) The expression, localization, and effect of a human interferon in plants. *Virology* 172:213–222
14. Hood EE et al (1997) Commercial production of avidin from transgenic maize: characterization of transformant, production, processing, extraction and purification. *Mol Breed* 3:291–306
15. Twyman RM et al (2005) Transgenic plants in the biopharmaceutical market. *Expert Opin Emerg Drugs* 10:185–218
16. Boothe J et al (2010) Seed-based expression systems for plant molecular farming. *Plant Biotechnol J* 8:588–606
17. Houdebine LM (2009) Production of pharmaceutical proteins by transgenic animals. *Comp Immunol Microbiol Infect Dis* 32:107–121
18. Basaran P, Rodriguez-Cerezo E (2008) Plant molecular farming: opportunities and challenges. *Crit Rev Biotechnol* 28:153–172
19. Yang L et al (2008) Biopharming to increase bioactive peptides in rice seed. *J AOAC Int* 91:957–963
20. Golovina EA et al (2008) Sustained expression of human cytomegalovirus glycoprotein B (UL55) in the seeds of homozygous rice plants. *Mol Biotechnol* 40:1–12
21. Tackaberry ES et al (2008) Sustained expression of human cytomegalovirus glycoprotein B (UL55) in the seeds of homozygous rice plants. *Mol Biotechnol* 40:1–12
22. Alli Z et al (2002) Pharming vaccines for hepatitis and cytomegalovirus: towards the development of multivalent and subunit vaccines for oral delivery of antigens. *Phytochem Rev* 1:55–66
23. Ganz PR et al (1996) Expression of human blood proteins in transgenic plants: the cytokine GM-CSF as a model protein. In: *Transgenic plants: a production system for industrial and pharmaceutical proteins*, ch 7–4. pp 281–297
24. Ganz PR et al (1995) Plants as factories for producing human blood proteins. *Can Soc Transfus Med Bull* 7:112–121
25. Blais DR, Altosaar I (2006) Human CD14 expressed in seeds of transgenic tobacco displays similar proteolytic resistance and bioactivity with its mammalian-produced counterpart. *Transgenic Res* 15:151–164
26. Blais DR, Altosaar I (2007) Humanizing infant milk formula to decrease postnatal HIV transmission. *Trends Biotechnol* 25:376–384
27. Stoger E et al (2000) Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. *Plant Mol Biol* 4: 583–590
28. Wang J, Oard H (2003) Rice ubiquitin promoters: deletion analysis and potential usefulness in plant transformation systems. *Plant Cell Rep* 22:129–134
29. Huang Z et al (2006) Rapid, high-level production of hepatitis B core antigen in plant leaf and its immunogenicity in mice. *Vaccine* 24:2506–2513
30. Stoger E et al (2002) Practical considerations for pharmaceutical antibody production in different crop systems. *Mol Breed* 9:149–158
31. Torres E et al (1999) Rice cell culture as an alternative production system for functional diagnostic and therapeutic antibodies. *Transgenic Res* 8:441–449
32. Hood E et al (2003) Criteria for high-level expression of a fungal laccase gene in transgenic maize. *Plant Biotechnol J* 1:129–140
33. Qu L, Takaiwa F (2004) Evaluation of tissue specificity and expression strength of rice seed component gene promoters in transgenic rice. *Plant Biotechnol J* 2:113–125
34. Sardana R et al (2007) Biologically active human GM-CSF produced in the seeds of transgenic rice plants. *Transgenic Research* 16:713–721
35. Wu CY et al (1998) Promoters of rice seed storage protein genes direct endosperm-

- specific gene expression in transgenic rice. *Plant Cell Physiol* 39:885–889
36. Wakasa Y et al (2006) High accumulation of bioactive peptide in transgenic rice seeds by expression of introduced multiple genes. *Plant Biotechnol J* 4:499–510
  37. Takagi H et al (2005) Oral immunotherapy against pollen allergy using a seed-based peptide vaccine. *Plant Biotechnol J* 3:521–533
  38. Zhang D et al (2010) Expression, purification, and characterization of recombinant human transferring from rice (*Oryza sativa* L.). *Protein Expr Purif* 74:69–79
  39. Sawant S et al (2001) Designing of an artificial expression cassette for the high-level expression of transgenes in plants. *Theor Appl Genet* 102:635–644
  40. Rushton PJ et al (2002) Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen and wound-induced signaling. *Plant Cell* 14:749–762
  41. Chikwamba RK et al (2003) Localization of a bacterial protein in starch granules of transgenic maize kernels. *Proc Natl Acad Sci USA* 100:11127–11132
  42. Wu CY et al (1998) The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by Opaque-2 in transgenic rice plants. *Plant J* 14:673–683
  43. Ye X et al (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287:303–305
  44. Li X et al (2002) Influence of matrix attachment regions from maize on transgene expression level in tobacco. *Acta Botanica Sinica* 44:804–808
  45. Tetko I et al (2006) Spatiotemporal expression control correlates with intragenic matrix attachment regions (S/MARs) in *Arabidopsis thaliana*. *PLoS Comput Biol* 2:136–145
  46. Yang D et al (2001) Expression of the REB transcriptional activator in rice grains improves the yield of recombinant proteins whose genes are controlled by a Reb-responsive promoter. *Proc Natl Acad Sci USA* 98:11438–11443
  47. Nakase M et al (1997) Characterization of a novel rice bZIP protein which binds to the  $\alpha$ -globulin promoter. *Plant Mol Biol* 33:513–522
  48. Comai L et al (1990) Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements. *Plant Mol Biol* 15:373–381
  49. Hennegan K et al (2005) Improvement of human lysozyme expression in transgenic rice grain by combining wheat *Triticum aestivum* puroindoline b and rice *Oryza sativa* Gt1 promoters and signal peptides. *Transgenic Res* 14:583–592
  50. Streatfield S (2007) Approaches to achieve high-level heterologous protein production in plants. *Plant Biotechnol J* 5:2–15
  51. Fütterer J, Hohn T (1996) Translation in plants—rules and exceptions. *Plant Mol Biol* 32:159–189
  52. Kawaguchi R, Bailey-Serres J (2002) Regulation of translation initiation in plants. *Curr Opin Plant Biol* 5:460–465
  53. Mauro VP et al (2002) The ribosome filter hypothesis. *Proc Natl Acad Sci USA* 99:12031–12036
  54. Kato T et al (1991) A modified B-glucuronidase gene: sensitive detection of plant promoter activities in suspension cultured cells of tobacco and rice. *Plant Mol Biol Rep* 9:333–339
  55. Satoh J et al (2004) The 5'-untranslated region of the tobacco alcohol dehydrogenase gene functions as an effective translational enhancer in plant. *J Biosci Bioeng* 98:1–8
  56. Wever W et al (2010) The 5' untranslated region of the VR-ACSI mRNA acts as a strong translational enhancer in plants. *Transgenic Res* 19:667–674
  57. Danthinne X, Emmelo J (1990) Studies on the translational properties of STNV RNA non-coding regions. In: 42nd international symposium on crop protection. Gent, Belgium
  58. Sivamani E et al (2009) Sequence analysis of rice rubi3 promoter gene expression cassette for improved transgene expression. *Plant Sci* 177:549–556
  59. Sivamani E, Qu R (2006) Expression enhancement of a rice polyubiquitin gene promoter. *Plant Mol Biol* 60:225–239
  60. Lu J et al (2008) Activity of the 5' regulatory regions of the rice polyubiquitin rubi3 gene in transgenic rice plants as analyzed by both GUS and GFP reporter genes. *Plant Cell Rep* 27:1587–1600
  61. Mascarenhas D et al (1990) Intron-mediated enhancement of heterologous gene expression in maize. *Plant Mol Biol* 15:913–920
  62. Fiume E et al (2004) Introns are key regulatory elements of rice tubulin expression. *Planta* 218:693–703
  63. Giani S et al (2009) In transgenic rice, a- and b-tubulin regulatory sequences control GUS amount and distribution through intron mediated enhancement and intron dependant spatial expression. *Transgenic Res* 18:151–162
  64. Samadder P et al (2008) Transcriptional and post-transcriptional enhancement of gene

- expression by the 5'UTR intron of rice rubi3 gene in transgenic rice cells. *Mol Genet Genomics* 279:429–439
65. Hunt AG (2008) Messenger RNA 3' end formation in plants. *Nuclear pre-mRNA processing in plants*, *Curr Top Microb Immunol* 326, 151–177
  66. Green PJ (1993) Control of mRNA stability in higher plants. *Plant Physiol* 102:1065–1070
  67. Shen Y et al (2008) Genome level analysis of rice mRNA 3'-end processing signals and alternative polyadenylation. *Nucleic Acids Res* 36:3150–3161
  68. Dong H et al (2006) An exploration of 3'-end processing signals and their tissue distribution in *Oryza sativa*. *Gene* 389:107–113
  69. Ali S, Taylor W (2001) The 3' non-coding region of a C4 photosynthesis gene increases transgene expression when combined with heterologous promoters. *Plant Mol Biol* 46:325–333
  70. Knirsch L, Clerch LB (2000) A region in the 3' UTR of MnSOD RNA enhances translation of a heterologous RNA. *Biochem Biophys Res Commun* 272:164–168
  71. Yang L et al (2009) The 3'-untranslated region of rice glutelin GluB-1 affects accumulation of heterologous protein in transgenic rice. *Biotechnol Lett* 31:1625–1631
  72. Mishra S et al (2006) Ubiquitin fusion enhances cholera toxin B subunit expression in transgenic plants and the plant-expressed protein binds GM1 receptor more efficiently. *J Biotechnol* 127:95–108
  73. Kang TJ et al (2004) Modification of the cholera toxin B subunit coding sequence to enhance expression in plants. *Mol Breed* 13: 143–153
  74. Ma JK et al (2005) Plant-derived pharmaceuticals—the road forward. *Trends Plant Sci* 10:580–585
  75. Stoger E et al (2005) Sowing the seeds of success: pharmaceutical proteins from plants. *Curr Opin Biotechnol* 16:167–173
  76. Vitale A, Pedrazzini E (2005) Recombinant pharmaceuticals from plants: the plant endomembrane system as bioreactor. *Mol Interv* 5:216–225
  77. Avesani L et al (2003) Improved in planta expression of the human islet autoantigen glutamic acid decarboxylase (GAD65). *Transgenic Res* 12:203–212
  78. Garbarino JE et al (1995) Isolation of a polyubiquitin promoter and its expression in transgenic potato plants. *Plant Physiol* 109:1371–1378
  79. Benchabane M et al (2008) Preventing unintended proteolysis in plant protein biofactories. *Plant Biotechnol J* 6:633–648
  80. Jiang L, Sun SS (2002) Membrane anchors for vacuolar targeting: applications in plant bioreactors. *Trends Biotechnol* 20:99–102
  81. Kawagoe Y et al (2005) The critical role of disulfide bond formation in protein sorting in the endosperm of rice. *Plant Cell* 17:1141–1153
  82. Hashizume F et al (2008) Development and evaluation of transgenic rice seeds accumulating a type II-collagen tolerogenic peptide. *Transgenic Res* 17:1117–1129
  83. Sugita K et al (2005) Genetically modified rice seeds accumulating GLP-1 analogue stimulate insulin secretion from a mouse pancreatic beta-cell line. *FEBS Lett* 579:1085–1088
  84. Wu G et al (2005) Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat Biotechnol* 23:1013–1017
  85. Takaiwa F et al (2008) Co-expression of soybean glycinins A1aB1b and A3B4 enhances their accumulation levels in transgenic rice seed. *Plant Cell Physiol* 49:1589–1599
  86. Naqvi S et al (2009) When more is better: multigene engineering in plants. *Trends Plant Sci* 15:48–56
  87. Abbadi A et al (2004) Biosynthesis of very-long-chain polyunsaturated fatty acids in transgenic oilseeds: constraints on their accumulation. *Plant Cell* 16:2734–2748
  88. Wada N et al (2009) Bioactive beads-mediated transformation of rice with large DNA fragments containing *Aegilops tauschii* genes. *Plant Cell Rep* 28:759–768
  89. Chen L et al (1998) Expression and inheritance of multiple transgenes in rice plants. *Nat Biotechnol* 16:1059–1064
  90. Obembe O et al (2011) Advances in plant molecular farming. *Biotechnol Adv* 29: 210–222
  91. Yang L et al (2007) Development of transgenic rice seed accumulating a major Japanese cedar pollen allergen (Cry j 1) structurally disrupted for oral immunotherapy. *Plant Biotechnol J* 5:815–826
  92. Takaiwa F et al (2007) Endosperm tissue is good production platform for artificial recombinant proteins in transgenic rice. *Plant Biotechnol J* 5:84–92
  93. Xie T et al (2008) A biologically active rhIGF-1 fusion accumulated in transgenic rice seeds can reduce blood glucose in diabetic mice via oral delivery. *Peptides* 29:1862–1870

94. Oszvald M et al (2008) Expression of cholera toxin B subunit in transgenic rice endosperm. *Mol Biotechnol* 40:261–268
95. Yasuda H et al (2009) Overexpression of BiP has inhibitory effects on the accumulation of seed storage proteins in endosperm cells of rice. *Plant Cell Physiol* 50:1532–1543
96. Matsumoto Y et al (2009) Oral immunogenicity and protective efficacy in mice of transgenic rice plants producing a vaccine candidate antigen (As16) of *Ascaris suum* fused with cholera toxin B subunit. *Transgenic Res* 18:185–192