High-Level Protein Expression System Uses Self-Pollinating Crops As Hosts

Ning Huang

Bacterial, fungal, and cultured cell expression systems have been used to produce a variety of recombinant proteins, and they play an extremely important role in biopharmaceutical production. However, because of limitations of cost and scale, all those systems are facing challenges in meeting the high-volume demand of some proteins. Amgen's drug Enbrel (www.enbrel.com) is one recent example. Plants, with their potential to produce large volumes of recombinant proteins at low cost, are considered by some to be an excellent alternative to the common expression systems used to produce recombinant proteins (1).

Since 1997, Ventria Bioscience (www.ventria.com) has worked to develop its version of the concept, leading to the ExpressTec high-level protein expression system, which uses the self-pollinating crops of rice and barley. We have achieved recombinant protein expression at up to 1% of the rice grain's weight. For many candidate proteins, the expression level is between 0.1 and 1.0% of grain weight. Such high expression levels, coupled with the large-scale production of cereal grains, could make the transgenic production of large volumes of recombinant proteins and peptides a reality. Our system offers several advantages: high and stable expression levels of recombinant proteins, tissue-specific expression in the grain endosperm, rapid scalability to metric-ton quantities, prevention of gene flow with self-pollinating crops, low capital investment and production costs, and efficient processing and recovery.

Rice Life Cycle Suits Recombinant Protein Production: The rice plant's life cycle begins with seed germination, followed by a vegetative stage, a reproductive stage, and a ripening stage. The vegetative stage is about 60–100 days, during which the rice plant develops and generates tillers, stalks that sprout up from the base of the plant. A single rice plant could have as many as 100 tillers producing more than 10,000 rice grains, which is important for early generation seed increases and producing small quantities of recombinant protein rapidly for feasibility studies.

Transgenic rice grain expressing recombinant human lactoferrin

When the rice plant enters its reproductive stage, the panicle (a loosely branched flower cluster) grows. Over 99% of cultivated rice flowers are self-pollinated because of the floral architecture, which is not receptive to pollen from other plants. Also, rice pollen grains are relatively short-lived, losing their viability within five minutes of shedding from the anther, which leads to an extremely low rate of cross-pollination. No outcrossing rates beyond 10 meters have been observed. Thus the isolation distance for transgenic rice that would be required by regulatory agencies is significantly lower than that for cross-pollinating plants such as corn and tobacco.

PRODUCT FOCUS: RECOMBINANT PROTEINS

PROCESS FOCUS: PROTEIN PRODUCTION AND PROCESSING

WHO SHOULD READ: PROJECT MANAGERS; PROCESS AND PRODUCT DEVELOPMENT; BUSINESS DEVELOPMENT FROM ENGINEERS AND SCIENTISTS TO CHIEF TECHNICAL OFFICER

KEYWORDS: PROTEIN EXPRESSION, CEREAL GRAIN, BIOPROCESSING, INDUSTRIAL BIOTECH, NUTRACEUTICALS, ENZYMES, TRANSGENIC PLANTS

LEVEL: INTERMEDIATE

Rice grains ripen after pollination. The ripening stage takes 30–50 days. This is when starch and storage proteins accumulate, and the grains are desiccated after filling out. Recombinant proteins expressed in rice accumulate in the grains at this time as well. Rice is then harvested and stored after being dried to a moisture content of 14% or less. The grain can be stored for three to five years without losing viability if it is kept below 13.5% moisture content. Recombinant proteins in the rice grains are also stable in excess of three years.

Storage Protein Gene Expression Is the Foundation: A rice grain consists of an inedible outer husk (20% of grain weight) and the enclosed brown rice (80%), which in turn contains an embryo (2%), an aleurone layer (7%), and endosperm (91%). Aleurone is outer-layer protein-rich cells. The major component of brown rice is starch. Almost all of its 8% protein content comes in the form of four major types of storage proteins: glutelin (>80%), globulin (10%), prolamin (<5%), and albumin (5%). Stored in type I and type II protein bodies, rice proteins are usually encoded by large gene families. Both glutelin and globulin deposit into type II bodies. An estimated 20 genes code for glutelin in rice. On the other hand, the plant has only a single gene for globulin.

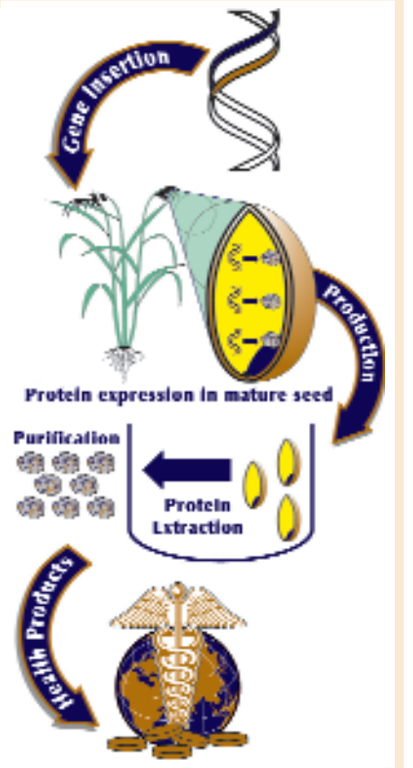
Rice storage protein genes begin to express five days after fertilization, peaking at 20–30 days depending on the environment and the variety of rice. Messenger RNA transcribed from the storage protein genes is differentially distributed to subdomains of the endoplasmic reticulum (ER). The proteins are synthesized and guided by a signal sequence into the ER through a translation/translocation process. During that process, the signal sequence is cleaved off to leave an intact mature protein. Storage proteins such as prolamin are retained inside the ER to later form a type I protein body within its membrane through budding. Glutelin and globulin deposit into a type II protein body, a single-membrane–bound organelle.

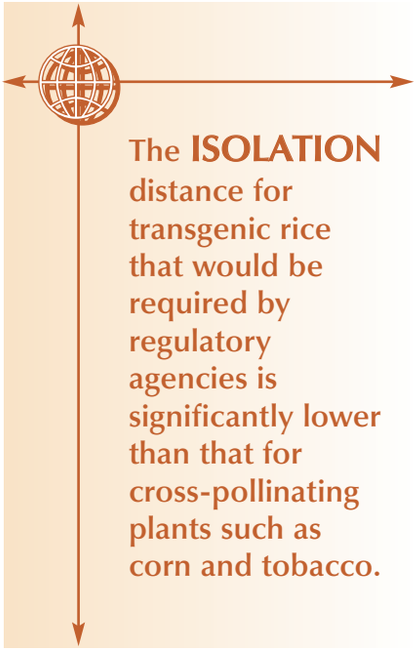
Evidence of high-level expression and accumulation of rice storage proteins in the grains led us to believe that recombinant protein could be expressed at similarly high levels using promoters and signal sequences from these storage proteins.

Glycosylation Pathway and Glycan Structure: In the ER, protein is attached with a 10-mannose chain through N-linked glycosylation at the amide residue of asparagine. Some mannose residues are removed from the chain later in the protein sorting pathway, where new sugars such as xylose, fucose, and glucose are added. A typical plant glycan carries one fucose and one xylose molecule. Recombinant proteins expressed in rice grains would bear the typical plant glycan structure. This has been confirmed by the glycan structure determination of human lactoferrin expressed in rice grains. Three major types of glycans are present in recombinant human lactoferrin, and all of them carry the typical core structure of the plant glycan with both xylose and fucose. None carries sialic acid as in the native form of human lactoferrin.

There has been recent attention to the importance of correct glycan structures on human therapeutic proteins. Clearly, certain protein targets require human glycans for optimal efficacy and stability when reintroduced into the human system. Alternatively, many other molecules may tolerate plant glycans with no effect on function. Regarding plant glycan immunogenicity issues, humans are exposed to plant glycan structures throughout their lives, especially through food ingestion. There should generally be little effect of plant glycan structures in plant produced proteins, espcially in oral and topical therapeutic and nutritional applications.

Figure 1: Schematic diagram of ExpressTec





Expression

Strong promoters are needed for high-level expression of recombinant proteins, so we tested various rice storage protein promoters along with a signal sequence. Genes of interest were placed downstream of the signal sequence, and the final product was expressed as a fusion protein. To select for strength, we tested six promoters isolated from rice storage

protein genes. All six constructs were identical except for the promoter sequences. The same signal sequence was included in each construct because protein stability would be lower in the cytosol (cytoplasmic fluid) than if the protein were deposited in an organelle such as the protein body. We used the gene for human lysozyme as our target gene.

Transgenic grains from each line were analyzed for the expression of recombinant human lysozyme. The glutelin 1 (Gt1) promoter caused higher expression levels of human lysozyme than the other five promoters, ranging from 4.5% to 40% of total soluble protein with an average of 14%. Expression levels from the globulin (Glb) promoter varied from 3% to 23% with an average of 13%, which was very close to that achieved with the Gt1 promoter. Three other promoters derived from rice glutelin genes were able to direct human lysozyme expression to about 10% of total soluble protein, all lower than that controlled by either Gt1 or Glb promoter.

Next, we compared the strength of the Gt1 and Glb promoters in expressing various genes and found that Gt1 was stronger. Thus we consider the Gt1 promoter to be the strongest promoter for recombinant protein expression in rice. Our expression system uses such a strong promoter along with a signal sequence, a codon-optimized gene of interest, and sometimes (if the gene of interest codes for a small peptide) an additional fusion partner.

One way to illustrate our level of recombinant protein expression is to compare the expression of the same protein by various plant species. Table 1 shows the expression of human lactoferrin in five plant hosts. Human lactoferrin is expressed in our rice grain at a level 10 times higher than in rice grain using other systems — and at least 30 times higher than in tobacco. Another protein we have tested is human lysozyme. The expression level of human lysozyme using our system has reached 1% of the rice grain weight or 40% of the total soluble protein, which is 400 times higher than that expressed in the tobacco plant (**2**). California rice yields are 7200–8000 kg of brown rice per hectare of cropland. With our expression system, that would suggest a potential yield of human lysozyme at 72–80 kg/ha.

Our expression of recombinant proteins is also highly tissue-specific. Recombinant protein expression under control of the Gt1 promoter is present only in the rice grain and not detectable in any other plant tissues such as the root, stem, young leaf, mature leaf, or anther.

We believe that the high level of protein expression possible with the ExpressTec system can be attributed to its capability for targeting recombinant protein into the protein bodies (organelles) where it is sequestered and protected from protease digestion. The signal sequence plays an important role in directing recombinant proteins into the ER and then through the various protein sorting systems within the rice cells to reach the protein body. Electron-microscopic evidence shows that rice endosperm possesses a highly adaptive capability to store recombinant proteins. A comparison study showed that the expression level of recombinant protein in barley grains with a signal sequence is at least 10 times higher than without the signal sequence (**3**). Evidently, recombinant proteins in cytosal are much more susceptible to protease digestion.

Production of Metabolites: Our system is not used only to produce recombinant proteins of pharmaceutical and nutraceutical importance, but also to produce secondary metabolites through metabolic engineering. One such product is matairesinol, a type of plant lignan (which are a group of phenypropanoid natural products, usually dimers or oligomers). Consumption of matairesinol is associated with reduced risk of breast cancer, prostate cancer, and colon cancer. To produce the substance in rice, four genes catalyzing the matairesinol synthesis pathway were expressed. Analysis indicated a significant elevation of matairesinol in the rice grains.

Recombinant Proteins Expressed in Rice Grains Are As Active As Native Proteins: We have produced human lactoferrin, human lysozyme, and many other proteins in rice grain (Figure 1). After purification, the biophysical and biochemical characteristics of those recombinant proteins were compared to those proteins derived from native sources. We have found that recombinant proteins and their native counterparts have the same molecular weight, N-terminal sequence, enzymatic activity, surface charge (an indication of same three-dimensional structure), and surface structure as indicated by reaction to specific antibodies. In addition, we found that they possess the same antibacterial activity, pH stability, and resistance level to protease activity. We have been unable to find any difference between recombinant proteins expressed in rice grains and their native counterpart except for the different glycosylation pattern.

Stable Recombinant Protein Expression Over Generations: High-level expression of recombinant proteins in plant cells may lead to gene silencing, so expression stability of recombinant proteins over generations is a concern. Our

Table 1: Expression of human lactoferrin in various plant species (TSP = total soluble protein)

Crop	Tissue	Expression Level (% TSP)	Expression Level (% Biomass)	Reference
Tomato	Fruit, Leaf	Detectable	Detectable	4
Tobacco	Culture cell	Detectable	Detectable	9
Tobacco	Leaf	0.8	Detectable	6
Tobacco	Leaf	0.3	Detectable	7
Corn	Grain	Detectable	Detectable	8
Potato	Tuber, Leaf	0.1	Detectable	9
Rice	Grain	Detectable	0.05	4
Rice (ExpressTec)	Grain	25	0.5	10

system seems to represent an exception to the rule. We have observed no gene silencing for the expression of human lactoferrin and lysozyme after testing two advanced lines (one expressing human lactoferrin and one expressing human lysozyme) for six generations. Expression of human lactoferrin remains at 0.5% of brown rice weight, and the expression of human lysozyme is at 0.6% of brown rice weight.

Why do we see such stable expression? Our hypothesis is that rice endosperm cells play the role of nutrient storage for seed germination and thus enter into a self-destruction stage (apoptosis) after starch and protein have accumulated. High-level expression of recombinant proteins (and their accumulation in the protein bodies) would not impose any physiological pressure to the cells. Thus, the rice plant does not need to activate a cell defense system such as methylation. On the other hand, the expression level will have a limit beyond which it could negatively affect seed development and germination. We have not yet discovered this limit but assume that it is somewhere between 1.0 and 1.5% of brown rice weight, depending on the protein expressed.

Storage and Processing

One advantage of using rice as the host for recombinant protein expression is that rice proteins are stable in stored grain. As mentioned above, rice grain can be stored for three to five years without losing viability if stored under ideal temperature and moisture conditions. We have tested the stability of human lysozyme and lactoferrin in rice grains stored for over 2.5 years and found no protein loss. Because of this high degree of stability, large amounts of rice can be stored to provide a continuous supply of grain for processing and manufacturing of pharmaceutical and nutraceutical products.

With human lactoferrin, simple extraction of brown rice flour recovers most human lactoferrin from the grain at a concentration of about 25% of total soluble protein. Filtered extract can be loaded directly onto a Sepharose column from Amersham Biosciences (www.amershambiosciences.com). Human lactoferrin eluted from that single column is greater than 90% pure, with a high recovery rate (>50%) of the final product in lyophilized powder.

Similarly, our lines of transgenic rice express human lysozyme as high as 50% of the total soluble protein, providing for a simple and efficient recovery process that can be scaled up easily. To extract recombinant proteins from rice grains, the rice is ground down to 20–100 mesh size, then the protein is extracted with a buffer containing salt. Most recombinant proteins are water soluble, so organic solvents and detergent extraction are not needed. Rice proteins are less soluble at low pH values. Therefore, if the recombinant protein can survive low pH conditions, it can be enriched by a low pH extraction step.

Economics

Rice production costs and yields vary from ecosystem to ecosystem and from country to country. Our cost estimates (in US dollars) for rice production are based on the world price of commodity rice, about $120/1000 kg of paddy rice. Of course, growing bioengineered rice to produce pharmaceutical and nutraceutical proteins would require higher capital investment than producing commodity rice. The higher input in producing plant-made pharmaceuticals includes more intensive care of the plant, segregation from other rice, dedicated equipment for planting, harvesting, drying, storage, and milling — as well as documentation and stringent regulatory compliance. We estimate the cost of rice production for pharmaceuticals at about three times the commodity rice production: $360/1000 kg paddy rice. With our system, the expression level of recombinant protein is about 0.5% of brown rice weight, which translates to $90/kg of recombinant protein from the rice grain.

In California, commodity rice production is on the high end of costs ($2000/ha or $800/acre), but the average yield (9659 kg/ha or 8500 lbs/acre) is also very high, resulting in reasonable rice production costs per kilogram ($207/1000 kg paddy rice). Increasing threefold that cost for production of rice-made pharmaceuticals would put it at about $621/1000 kg or about $155/kg of recombinant protein if expressed at 0.5%.

Cost of Processing: Cost associated with processing rice grain will largely depend on the final product and whether it is an injectible, oral, or skin-care therapeutic product, a nutraceutical, or an industrial purpose. It will cost much less to make a functional food than to prepare a high-purity pharmaceutical. To prepare an extract from rice flour expressing

MARTIN DEWIT (WWW.ISTOCKPHOTO.COM)

lactoferrin for functional food use, food industrial processing would be used at a cost of about $0.50 to 1.00/kg flour. Higher purity products are also very efficiently produced using a single column process. Greater than 90% pure human lactoferrin is produced and the process can be easily scaled up to meet increaing demand. We estimate that it would cost about $5 to $10 to generate one gram of human lactoferrin from rice flour ina cGMP facility operating at a scale of 600 kg/year.

By-Product Use: During the processing of rice flour, 98% of the biomass remains as a by-product. This cake is rich in starch and protein thus could provide an alternative biomass source for ethanol production. This revenue potential is not considered in the above economics.

CONSIDER THIS . . .

By exploiting the nature of cereal grains, we have developed the ExpressTec recombinant protein expression system for producing pharmaceutical, nutraceutical, and industrial enzymes. Our system could provide several advantages over others in expression level, scalability, cost of production, and processing ease. In addition, the system is more environmentally sound than many others because it uses a self-pollinating crop. Data show expression levels of recombinant protein as high as 1% of rice grain weight, which allows it to be easily processed and recovered. Considering all of the above, we believe ExpressTec deserves consideration for production of recombinant proteins and peptides requiring large volumes and low cost of production.

REFERENCES

1 Schwab, J. *A Strategic Evaluation of Transgenic Plant and Animal Biomanufacturing Systems.* Revelogic, Inc.: Fort Collins, CO, 2003.

2 Nakajima, H; et al. Fungal and Bacterial Disease Resistance in Transgenic Plants Expressing Human Lysozyme. *Plant Cell Report* 1997, 16; 674–679.

3 Horvath, H; et al. The Production of Recombinant Proteins in Transgenic Barley Grains. *Proc. Natl Acad. Sci. USA* 2000, 97; 1914–1919.

4 Anzai, H; Takaiwa, F; Katsumata, K. Production of Human Lactoferrin in Transgenic Plants. In *Lactoferrin: Structure, Function and Applications,* Shimazaki K; et al., Eds. Elsevier: Amsterdam, 2000; pp. 265–271.

5 Mitra, A; Zhang, Z. Expression of a Human Lactoferrin cDNA in Tobacco Cells Produces Antibacterial Protein(s). *Plant Physiol.* 1994, 106; 977–981.

6 Zhang, Z; et al. Expression of Human Lactoferrin Cdna Confers Resistance to Ralstonia Solanacearum in Transgenic Tabacco Plants. *Phytopathology* 1998, 88; 730–734.

7 Salmon, V, et al. Production of Human Lactoferrin in Transgenic Tobacco Plants. *Protein Expr. Purif.* 13: 127–35 (1998).

8 Samyn-Petit, B; et al. N-Glycosylation Potential of Maize: The Human Lactoferrin Used As a Model. *Glycoconj. J.* 2001, 18; 519–527.

9 Chong, DK; Langridge WH. Expression of Full-Length Bioactive Antimicrobial Human Lactoferrin in Potato Plants. *Transgenic Res.* 2000, 9; 71–78.

10 Nandi, S; et al. Expression of Human Lactoferrin in Transgenic Rice Grains for the Application in Infant Formula. *Plant Science* 2002, 163; 713–722.

Ning Huang *is vice president of research and development for Ventria Bioscience, 4110 North Freeway, Sacramento, CA 95834; 1-916-921-6148, nhuang@ventriabio.com, www.ventriabio.com.*