

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number  
**WO 02/064814 A2**

(51) International Patent Classification<sup>7</sup>: **C12Q**

(21) International Application Number: PCT/US02/04921

(22) International Filing Date: 14 February 2002 (14.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/269,199 14 February 2001 (14.02.2001) US  
09/847,232 2 May 2001 (02.05.2001) US

(71) Applicant: **VENTRIA BIOSCIENCE** [US/US]; 4110  
North Freeway Boulevard, Sacramento, CA 95834-1219  
(US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors: **HUANG, Ning**; 1226 Caricia Drive, Davis, CA 95616 (US). **RODRIGUEZ, Raymond, L.**; 3017 Anza Avenue, Davis, CA 95616 (US). **HAGIE, Frank, E.**; 4400 Truxel Road, No. 191, Sacramento, CA 95834 (US).

(74) Agents: **ARNETT, Stephen, E.** et al.; Perkins Coie LLP, P.O. Box 2168, Menlo Park, CA 94026 (US).

(54) Title: EXPRESSION OF HUMAN MILK PROTEINS IN TRANSGENIC PLANTS

(57) Abstract: The invention is directed to food and food additive compositions comprising one or more human milk proteins produced in the seeds of a transgenic plant and methods of making the same. The invention is further directed to improved infant formula comprising such food supplement composition.



**WO 02/064814 A2**

## EXPRESSION OF HUMAN MILK PROTEINS IN TRANSGENIC PLANTS

### Field of the Invention

The present invention relates to human milk proteins produced in the seeds of  
5 transgenic plants, seed extracts containing the proteins, transgenic plants and seeds,  
and methods for producing and using the same.

### References

The following references are cited herein, and to the extent they may be pertinent to  
10 the practice of the invention, are incorporated herein by reference.

Alber and Kawasaki, *Mol. and Appl. Genet.* 1:419-434, 1982.

Aniansson, G. *et al.* (1990). "Anti-Adhesive Activity Of Human Casein Against  
Streptococcus Pneumoniae And Haemophilus Influenzae," *Microb Pathog*, 8(5):315-323.

Arnold R.R. *et al.*, *Infect Immun.* 28:893-898, 1980.

15 Ausubel F.M. *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley &  
Sons, New York, N.Y., 1993.

Bhan, M.K. *et al.*, *J Pediatr Gastroenterol Nutr* 7:208-213, 1988.

Beatty, K., *et al.*, "Kinetics of association of serine proteinases with native  
and oxidized alpha-1-proteinase inhibitor and alpha-1-antichymotrypsin," *J. Biol.*  
20 *Chem.* 255:3931-3934, 1980.

Boesman-Finkelstein M. and Finkelstein R.A. *FEBS Letters*, 144:1-5, 1982.

Bradford M. *Analytical Biochem.* 72:248-254, 1976.

Brandt, *et al.*, *Carlsberg Res. Commun.* 50:333-345, 1985.

Briggs, D, *et al.* (1981). Malting And Brewing Science - Volume 1, Malt And  
25 Sweet Wort (Chapman & Hall, New York, 2nd edition).

Briggs, D (1998). Malts And Malting (Blackie Academic And Professional, New  
York).

Bullen J.J. *et al.*, *Br. Med. J.* 1, 69-75, 1972.

Carrell, R.W. *et al.* "Structure and variation of human alpha-1-antitrypsin",  
30 *Nature* 298:329-334, 1983.

Castañón M.J. *et al.*, *Gene*, 66:223-234, 1988.

Chandan R.C., *J Dairy Sci*, 51:606-607, 1968.

Chong, D.K. and Langridge, W.H., *Transgenic Res* 9:71-8, 2000.

- Chowanadisai, W. and Lönnnerdal, B., "Alpha-1-antitrypsin and antichymotrypsin in human milk: origin, concentrations, and stability", *Am. J. Clin. Nutr.* 2001 (in press).
- Chrispeels K., *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53, 1991.
- 5 Clark *et al.*, *J. Biol. Chem.* 264:17544-17550, 1989.
- Davidson LA, *et al.* (1987). "Persistence Of Human Milk Proteins In The Breast-Fed Infant", *Acta Paediatr Scand* 76(5):733-740.
- Davidson, L.A., and Lönnnerdal, B. "Fecal alpha-1-antitrypsin in breast-fed infants is derived from human milk and is not indicative of enteric protein loss",
- 10 *Acta Paediatr. Scand.* 79:137-141, 1990.
- della-Cioppa *et al.*, *Plant Physiol.* 84:965-968, 1987.
- Dellaporta S.L. *et al.*, *Plant Mol. Biol. Rep.* 1:19-21, 1983.
- Depicker *et al.*, *Mol. Appl. Genet.* 1:561-573, 1982.
- Dewey K.G. *et al.*, *J Pediatrics* 126:696-702, 1995.
- 15 Dewey K.G. *et al.*, *Pediatrics* 89:1035-1041, 1992.
- Dewey K.G. *et al.*, *Am J Clin Nutr* 57:140-145, 1993.
- Ditta *et al.*, *Proc. Nat. Acad. Sci., U.S.A.* 77:7347-7351, 1980.
- Doncheck, J (1995). *Malts And Malting*. In: *Kirk-Othmer Encyclopedia Of Chemical Technology* (John Wiley & Sons, New York, 4th edition), volume 15, pp. 947-
- 20 962.
- Eley, E, (1990), *A Wandful World. Food Processing* August 1990, pp. 17-18
- Faure A. and Jollès P., *et al.*, *Comptes Rendus Hebdomadaires des Seances de L Academie des Sciences. D: Sciences Naturelles*, 271:1916-1918, 1970.
- Fujihara T. and Hayashi K., *Archives of Virology* 140:1469-1472, 1995.
- 25 Gastañaduy, A. *et al.*, *J Pediatr Gastroenterol Nutr* 11:240-6, 1990.
- Gelvin, S.B. *et al.*, eds. *PLANT MOLECULAR BIOLOGY MANUAL*, 1990.
- Gielen, *et al.*, *EMBO J.* 3:835-846, 1984.
- Goto, F *et al.*, *Nature Biotech.* 17:282-286, 1999.
- Grover M. *et al.*, *Acta Paediatrica* 86:315-316, 1997.
- 30 Hamosh, M, *et al.* (1999). "Protective Function Of Human Milk: The Fat Globule," *Semin Perinatol* 23(3):242-9.
- Harmsen M.C. *et al.*, *Journal of Infectious Diseases* 172:380-388, 1995.
- Hickenbottom, JW (1977). *Sweeteners In Biscuits And Crackers. The Bakers Digest*, December 1977, pp. 18-22.

- Hickenbottom, JW (1983). AIB Research Department Technical Bulletin, V(3), March 1983, pp. 1-8.
- Hickenbottom, JW (1996). *Processing, Types, And Uses Of Barley Malt Extracts And Syrups*. Cereal Foods World 41(10): 778-790.
- 5 Hickenbottom JW (1997). *Malts In Baking*. American Society Of Bakery Engineers, Technical Bulletin #238, August 1997, pp. 1010-1013.
- Huang N. *et al.*, Plant Mol. Biol., 23:737-747, 1993.
- Hough, JS, *et al.* (1982). ). Malting And Brewing Science - Volume 2, Hopped Wort And Beer (Chapman & Hall, New York, 2nd edition).
- 10 Huang N. *et al.*, J CAASS, 1:73-86, 1990a.
- Horvath H. *et al.*, Proc. Natl. Acad. Sci. USA, 97:1914-1919, 2000.
- Humphreys, D.P. and Glover, D.J. "Therapeutic antibody production technologies: molecules, applications, expression and purification," Curr Opin Drug Discov Devel Mar;4(2):172-85, 2001.
- 15 Huang N. *et al.*, Plant Mol. Biol. 14:655-668, 1990b.
- Huang, J., *et al.*, "Expression and purification of functional human alpha-1-antitrypsin from cultured plant cells", Biotechnol. Prog. 17:126-133, 2001.
- Jensen L.G. *et al.*, Proc. Natl. Acad. Sci. USA, 93:3487-3491, 1996.
- Jigami Y. *et al.*, Gene, 43:273-279, 1986.
- 20 Johnson, T. "Human alpha-1-proteinase inhibitor", Methods Enzymol. 80:756-757, 1981.
- Jollès P., LYSOZYMES--MODEL ENZYMES IN BIOCHEMISTRY AND BIOLOGY. Birkhäuser Verlag, Basel; Boston, 1996.
- Jouanin *et al.*, Mol. Gen. Genet. 201:370-374, 1985.
- 25 Kortt A.A. *et al.*, "Dimeric and trimeric antibodies: high avidity scFvs for cancer targeting," Biomol Eng 18:95-108, 2001.
- Kovar M.G. *et al.*, Pediatrics 74:615-638, 1984.
- Kunz C. *et al.*, Clin. Perinatol. 26(2):307-33, 1999.
- Lake, C, (1988). *From Strength To Strength*. Flood Flavorings, Ingredients,
- 30 Packaging And Processing, February 1988, pp. 55 -57.
- Langridge *et al.*, Planta 156:166-170, 1982.
- Larrick, J.W. *et al.*, "Production of antibodies in transgenic plants", Res Immunol 149:603-8, 1998.

- Lásztity, R. The chemistry of cereal proteins. CRC Press, Boca Raton, 1996.
- Lee-Huang S. *et al.*, *Proc. Natl. Acad. Sci. USA*, 96:2678-2681, 1999.
- Lindberg, T. "Protease inhibitors in human milk", *Pediatr. Res.* 13:969-972, 1979.
- Lollike K. *et al.*, *Leukemia*, 9:206-209, 1995a.
- Lönnerdal B., *Am L Clin Nutr*, 42:1299-1317, 1985.
- Lönnerdal, B. "Recombinant milk proteins- an opportunity and a challenge", *Am. J. Clin. Nutr.* 63:622S-626S, 1996.
- Maga E. *et al.*, *J. of Food Protection*, 61:52-56, 1998.
- Maga E. *et al.*, *Transgenic Research*, 3:36-42, 1994.
- Maga E. *et al.*, *Journal of Dairy Science*, 78:2645-2652, 1995.
- Maniatis, *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Edition, 1989.
- Matsumoto, A. *et al.*, *Plant Mol Bio* 27:1163-72, 1995.
- Maynard, J. and Georgiou, G. Antibody engineering, *Annu Rev Biomed Eng* 2:339-76, 2000.
- McBride and Summerfelt, *Plant Mol. Biol.* 14:269-276, 1990.
- Mcgilligan, KM, *et al.*, (1987), "Alpha-1-Antitrypsin Concentration In Human Milk," *Pediatr Res* 22(3):268-270.
- Mitra A. and Zhang Z., *Plant Physiol.* 106:977-981, 1994.
- McGilligan, K.M. *et al.*, "Alpha-1-antitrypsin concentration in human milk," *Pediatr. Res.* 22:268-270, 1987.
- Mitra, A, and Zhang, Z., *Plant Physiol* 106:977-81, 1994.
- Moe, T, *et al.* (1995). *Mashing Of Rice With Barley Malt Under Non-Conventional Process Conditions For Use In Food Processes*. *International Journal Of food Science And Technology* 29:635-649.
- Moore, K (1978). "Rice-Barley Malt Combo Improves Functionality And Nutrition Of Meats, Breads, And Baked Goods", *Food Product Development* 12(5):74.
- Moreira-Ludewig, R. and Healy, C., *J. Pharm. and Tox. Meth.*, 27:75-100, 1992.
- Motil K.J., *Curr Opin Pediatr* 12(5):469-76, 2000.
- Murphy MS (1998). "Growth Factors And The Gastrointestinal Tract", *Nutrition* 14(10):771-4.
- Nakajima H. *et al.*, *Plant Cell Reports*, 16:674-679, 1997.

- Newburg, DS (1999). "Human Milk Glycoconjugates That Inhibit Pathogens", *Curr Med Chem* 6(2):117-127.
- Newburg, DS, *et al.* (1998). "Role Of Human-Milk Lactadherin In Protection Against Symptomatic Rotavirus Infection", *Lancet* 351(1910):1160-4.
- 5 NIH publication, American Academy of Allergy and Immunology Committee on adverse reaction to food and National Institute of Allergy and Infectious Diseases, NIH, Bethesda, 1984.
- Peeters, K., *et al.*, "Production of antibodies and antibody fragments in plants", *Vaccine* 9:2756-61, 2001.
- 10 Peterson JA, *et al.* (1998). "Glycoproteins Of The Human Milk Fat Globule In The Protection Of The Breast-Fed Infant Against Infections", *Bio Neonate* 74(2):143-162.
- Piper, D.W. and Fenton, B.H. "pH stability and activity curves of pepsin with special reference to their clinical importance", *Gut* 6:506-508, 1965.
- Prosser, CG (1996). "Insulin-Like Growth Factors In Milk And Mammary Gland," *J Mammary Gland Bio Neoplasia* (3):297-306.
- 15 Raikhel N., *Plant Phys.* 100:1627-1632, 1992.
- Rey, M.W., *Nucleic Acid Res.* 18:5288, 1990.
- Romer *et al.*, *Biochem. Biophys. Res Commun.* 196:1414-1421, 1993.
- Rudloff, S. and Lönnerdal, B. "Solubility and digestibility of milk proteins in infant formulas exposed to different heat treatments", *J. Pediatr. Gastroenterol. Nutr.* 15:25-33, 20 1992.
- Saarinén K.M. *et al.*, *Adv Exp Med Biol* 478:121-30, 2000.
- Salmon V. *et al.*, *Protein Expression and purification* 9:203-210, 1997.
- Salmon, V. *et al.*, *Protein Expr Purif* 13:127-135, 1998.
- 25 Samaranayake Y.H. *et al.*, *Apmis*, 105:875-883, 1997.
- Sambrook J. *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL* (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989.
- Satue-Gracia M.T. *et al.*, *J Agric Food Chem* 48(10):4984-90, 2000.
- Schütte H. and Kula M.R., *Biotechnology and Applied Biochemistry*, 12:599-620, 30 1990.
- Sfat, M, *et al.* (1981). *Malts And Malting*. In: *Kirk-Othmer Encyclopedia Of Chemical Technology* (John Wiley & Sons, New York, 3rd edition), volume 14, pp. 810-823.
- Shah *et al.*, *Science* 233:478-481, 1986.

Shin, K, *et al.* (2000). "PCR Cloning And Baculovirus Expression Of Human Lactoperoxidase And Myeloperoxidase", *Biochem Biophys Res Commun* 27(13):831-836.

Shugar D., *Biochim. Biophys. Acta*, 8:680-68, 1952.

5 Symbicon AB (1995). "Human Milk Kappa-Casein And Inhibition Of Helicobacter pylori Adhesion To Human Gastric Mucosa", *J Pediatr Gastroenterol Nutr* 21(3):288-296.

Takai I. *et al.*, *J. Chrom. B, Biomedical Applications*, 685:21-25, 1996.

Tsuchiya K. *et al.*, *Applied Microbiology and Biotechnology*, 38:109-114, 1992.

Von Heijne *et al.*, *Plant Mol. Biol. Rep.* 9:104-126, 1991.

10 Wang C.S. and Kloer H.U., *Anal. Biochem.*, 139:224-227, 1984.

Wang C. *et al.*, *Comp. Biochem. Physiol.* 78B:575-580, 1984.

Ward P.P. *et al.*, *Bio/technology* 10:784-789, 1992.

Ye X. *et al.*, *Science* 287:303-305, 2000.

Yoshimura K. *et al.*, *Biochem. Biophys Res Com*, 150:794-801, 1988.

15

### **Background Of The Invention**

Milk proteins such as lactoferrin (LF), lysozyme (LZ), lactoperoxidase (LP), immunoglobulin-A (IgA), alpha-lactalbumin, beta-lactoglobulin, alpha-, beta- and kappa-caseins, serum albumin, lipase and others are known to have a number of nutritional and  
 20 other beneficial effects, particularly for infants. Breast feeding of fresh human milk has traditionally been considered the best means to provide nutrition to an infant. Although all physiological roles of human milk proteins have not yet been elucidated, evidence has been obtained that lysozyme, lactoferrin and other milk proteins control the microflora in the gut of infants (Lönnerdal, 1985). Breast milk has been suggested to contain many  
 25 immune factors that compensate for the undeveloped defense mechanisms of the gut of infants (Saarinen KM *et al.*, 2000). Several human milk proteins have been demonstrated to have beneficial physiologic effects in infants, particularly in the defense against infection and in the optimization of nutrient uptake.

However, many situations arise where the infant cannot be fed mother's milk and  
 30 synthetic infant milk formulas are used in the place of breast feeding (Motil KJ, 2000). Considerable effort has been made to improve synthetic infant milk formulas in order to closely simulate mother's milk.

The protein and non-protein composition of the human milk and cow milk is described by Kunz *et al.*, 1999. The relative concentrations of milk proteins vary

between human and cows' milk. For example, lactoferrin and lysozyme are present in a relatively high amount in human milk but in only low or trace amounts in cow's milk.

In general, synthetic infant formula is prepared using cow's milk that does not closely resemble the protein composition found in human milk. Accordingly, cow's milk based infant formula is typically supplemented with various human milk protein components. Typically, commercial infant formulas based on cow's milk contain approximately of 0.1 mg/mL lactoferrin whereas natural human breast milk contains an average concentration of 1.4 mg/mL. Soy-based infant formulas contain no added lactoferrin.

Although addition of recombinant human milk proteins to infant milk or milk formula has been proposed, *e.g.*, using transgenic cows or by addition of microbially produced human milk proteins to milk or milk formula, these approaches do not overcome the various problems of (i) allergies to cow's milk, (ii) the high cost of recombinant protein production and/or (iii) safety issues related to food products.

It would therefore be desirable to provide a plant-derived infant formula having beneficial levels of one or more proteins normally present in human milk, while largely avoiding costly recombinant protein production techniques and associated safety issues. More generally, it would be desirable to provide a nutritional food extract that may be readily and inexpensively obtained in large quantities, can be delivered by itself, as a nutraceutical or added to processed foods, for supplying one or more human milk proteins beneficial to human health.

### **Summary Of The Invention**

In one aspect, the invention includes a nutritionally enhanced food having one or more plant-derived food ingredients, and as an additive, a seed composition containing a flour, extract, or malt obtained from mature monocot seeds and one or more seed-produced human milk proteins in substantially unpurified form. The seed-produced protein(s) include lactoferrin, lysozyme, lactoferricin, epidermal growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, and/or alpha-1-antitrypsin, preferably at least lysozyme and/or lactoferrin.

The seed composition preferably comprises between 0.1 to 20% of the total solid weight of the food. The seed-produced human milk protein(s) are preferably present in an amount that is at least 50% of the amount of the protein(s) in human milk, on a weight/weight basis.

In one embodiment, the food is an infant formula, either in dry or liquid form. The milk proteins include at least lysozyme and lactoferrin, and the seed composition contains a seed extract or malt obtained from mature seeds of rice or barley. The lysozyme is preferably present in an amount between 0.03 to 0.3 g protein/liter formula, and lactoferrin, in an amount between 0.3 and 3 g protein/liter<sup>2</sup>.

In another aspect, the invention includes an ingestible monocot-seed composition containing a flour, extract, or malt obtained from mature monocot seeds and one or more seed-produced human milk proteins in substantially unpurified form. As above, the seed-produced protein(s) preferably include lactoferrin and/or lysozyme, but may alternatively or in addition, include lactoferricin, epidermal growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, IgA, and alpha-1-antitrypsin. The one or more milk proteins are in the composition extract in an amount greater than 1 mg/gram dry weight of extract.

The flour may be prepared by milling mature monocot seeds; the extract, by suspending milled flour in a buffered aqueous medium; and the malt, by (i) steeping barley seeds to a desired water content, (ii) germinating the steeped barley, (iii) drying the germinated seeds under conditions effective to stop germination, (iv) crushing the dried seeds, (v) optionally, adding crushed seeds from a non-barley monocot plant, (vi) forming a mixture of the crushed seeds in water, and (vii) malting the crushed seed mixture until a desired malt is achieved, where at least one of the barley or non-barley monocot seeds contain such milk protein(s).

Also disclosed is a monocot seed containing, in extractable form, one or more proteins normally present in human milk, where the human-milk protein(s) include lactoferrin, lysozyme, lactoferricin, EGF, IGF-I, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, and immunoglobulins, preferably at least lactoferrin and/or lysozyme. The milk protein preferably includes at least 0.25 weight percent of the total protein in the harvested mature seeds.

In a related aspect, the invention includes a method of producing an ingestible seed composition. In practicing the method, there is first obtained a monocot plant that has been stably transformed with a first chimeric gene having (i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter, (ii) operably linked to said transcriptional regulatory region, a leader DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and (iii) a protein-coding sequence encoding

a protein normally present in human milk. The transformed plant is cultivated under seed-maturation conditions, and the mature seeds harvested. From the harvested seeds is obtained a flour, extract, or malt composition containing the human milk protein in substantially unpurified form. The human milk protein(s) preferably constitute at least 0.1 percent of the total protein in the harvested mature seeds.

The flour may be prepared by milling mature monocot seeds; the extract, by suspending milled flour in a buffered aqueous medium; and the malt, by (i) steeping barley seeds to a desired water content, (ii) germinating the steeped barley, (iii) drying the germinated seeds under conditions effective to stop germination, (iv) crushing the dried seeds, (v) optionally, adding crushed seeds from a non-barley monocot plant, (vi) forming a mixture of the crushed seeds in water, and (vii) malting the crushed seed mixture until a desired malt is achieved, where at least one of the barley or non-barley monocot seeds contain such milk protein(s).

The monocot plant obtained may be further transformed with a second chimeric gene having (i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter, (ii) operably linked to said transcriptional regulatory region, a transit DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and (iii) a protein-coding sequence encoding a protein normally present in human breast milk other than that encoded by the first chimeric gene.

In a related aspect, the invention includes a transgenic monocot plant which is stably transformed with a first chimeric gene having (i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter, (ii) operably linked to said transcriptional regulatory region, a transit DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and (iii) a protein-coding sequence encoding a protein normally present in human milk.

Exemplary transcriptional regulatory regions in the chimeric gene are from the promoter of the group of genes: rice glutelins, rice globulins, oryzins, and prolamines, barley hordeins, wheat gliadins and glutenins, maize zeins and glutelins, oat glutelins, and sorghum kafirins, millet pennisetins, and rye secalins genes. The leader sequence is likewise from the group of genes: gene selected from the group of rice glutelins, rice globulins oryzins, and prolamines, barley hordeins, wheat gliadins and glutenins, maize

zeins and glutelins, oat glutelins, and sorghum kafirins, millet pennisetins, and rye secalins genes.

In one preferred embodiment, the transcriptional regulatory region in the chimeric gene is a rice glutelin Gt1 promoter, and the leader DNA sequence is a rice glutelin Gt1 signal sequence capable of targeting a linked polypeptide to a protein storage body. An exemplary glutelin Gt1 promoter and glutelin Gt1 signal sequence are included within the sequence identified by SEQ ID NO:15. In another preferred embodiment, the transcriptional regulatory region in the chimeric gene is a rice globulin Glb promoter, and the leader DNA sequence is a rice glutelin Gt1 signal sequence capable of targeting a linked polypeptide to a protein storage body. An exemplary globulin Glb promoter and glutelin Gt1 signal sequence are included within the sequence identified by SEQ ID NO:16.

The transformed monocot seed may further encode at least one transcription factors O2, PBF, and Reb, as exemplified by SEQ ID NOS: 31, 32, and 33, respectively, and preferably O2 and/or PBF.

The protein-coding sequence is the a coding sequence for a human milk protein selected from the group consisting of lactoferrin, lysozyme, lactoferricin, EGF, IGF-I, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, and immunoglobulins, preferably a sequence which has been codon-optimized for expression in monocots. Exemplary codon-optimized sequences for these proteins are represented by SEQ ID NOS: 1, 3, and 7-14.

The plant may be further stably transformed with a second chimeric gene having (i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter, (ii) operably linked to said transcriptional regulatory region, a transit DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and (iii) a protein-coding sequence encoding a protein normally present in human breast milk other than that encoded by the first chimeric gene.

In still another aspect, the invention includes a method of forming a malt syrup containing one or more human milk proteins. The method includes the steps of (i) steeping barley seeds to a desired water content, (ii) germinating the stepped barley, (iii) drying the germinated seeds, under conditions effective to stop germination, (iv) crushing the dried seeds, (v) optionally, adding crushed seeds from a non-barley monocot plant, (vi) forming a mixture of crushed seeds in water, and (vii) malting the

crushed seed mixture until a desired malt is achieved. At least one of the barley or non-barley monocot seeds are obtained from plants that have been stably transformed with a first chimeric gene having (i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter, (ii) operably linked to said transcriptional  
5 regulatory region, a transit DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and (iii) a protein-coding sequence encoding a protein normally present in human breast milk.

These and other objects and features of the invention will become more fully apparent when the following detailed description of the invention is read in conjunction  
10 with the accompanying drawings.

### **Brief Description of the Figures**

Figure 1 is a map of the pAPI159 expression construct that contains the human lysozyme coding sequence under the control of a Gt1 promoter and Gt1 signal  
15 sequence.

Figure 2 shows the results of Western blot analysis for the expression of recombinant human lysozyme in various tissues of rice plants, where lanes 1 and 15 are a human milk lysozyme standard; lane 2 is a broad range molecular weight marker from Sigma; lanes 3 and 4 represent mature seed tissue extracts; lanes 5 and 6 represent  
20 germinated seed extracts; lanes 7 and 8 represent root tissue extracts; lanes 9 and 10 represent extracts from young root tissue; lanes 11 and 12 represent leaf extracts; and lanes 13 and 14 represent extracts from young leaf; from untransformed ("U") or transgenic ("T") plants, respectively. The total loading protein amount was 40 µg per lane.

Figure 3 shows the effect of incubating recombinant human lysozyme from transgenic rice seed, a human lysozyme standard (30 µg/ml), a control (20 mM sodium phosphate, pH 7.0, 5 mM EDTA) or an untransformed rice extract on the growth of *E.coli* strain JM109. At the end of the incubation (for the time indicated), an aliquot of the mixture was plated on LB plates and colony forming units per ml (CFU/ml) was  
25 calculated.  
30

Figure 4 is a graph showing the specific activity of lysozyme, as determined by incubating an identical concentration of a human lysozyme standard, human lysozyme from transgenic rice (plant) and lysozyme from chicken egg white with a standard amount

of *M. luteus*, followed by evaluation of the reduction in the turbidity due to the activity of lysozyme over five minutes.

Figure 5A: Thermal stability of human lysozyme ("Hlys") and recombinant human lysozyme from transgenic rice ("rHLys"). Lysozyme was dissolved at 100 µg/ml in PBS.

- 5 The mixtures were subjected to different temperatures for different lengths of time. At the end of each heat treatment, the remaining lysozyme activity was assessed by activity assay.

Figure 5B: pH stability of Hlys and rHlys. Lysozyme was dissolved in different buffers at 100 µg/ml. The mixture was incubated at 37°C for 30 min. The lysozyme  
10 activity was determined by activity assay.

- Figure 6 presents the results of an analysis of lysozyme expression in transgenic rice grains over several generations. Proteins from 1 g of brown rice flour were extracted with 40 ml of extraction buffer containing 0.35 M NaCl in PBS. Extraction was conducted at room temperature for 1 h with shaking. Homogenate  
15 was centrifuged at 14,000 rpm for 15 min at 4°C. Protein supernatant was removed and diluted as needed for lysozyme turbidimetric activity assay. Extraction was repeated three times and standard deviation was shown as an error bar. Lysozyme yield was expressed as percentage of total soluble protein (%TSP).

- 20 Figure 7 is a restriction map of the pAPI164 plasmid that contains the human lactoferrin coding sequence under the control of a rice glutelin (Gt1) promoter, aGt1 signal peptide, and a nopaline synthase (NOS) terminator/polyadenylation site.

- Figure 8 shows the results of a SDS-PAGE analysis for human lactoferrin  
25 stained with Coomassie blue, where lane 1 is the molecular weight marker; lanes 2 - 5 are purified human derived lactoferrin (Sigma, USA); lanes 6 - 10 are single seed extracts from homozygous transgenic lines and lane 11 is a seed extract from non-transformed TP-309.

- Figure 9 shows the results of a Western blot analysis of various tissues of  
30 the transgenic rice plants, demonstrating the tissue specificity of rLF expression. Lane 1 is the molecular weight marker; lane 2 is human lactoferrin (Sigma, USA); lane 3 is an extract from leaf; lane 4 is an extract from sheath; lane 5 is an extract from root; lane 6 is an extract from seed and lane 7 is an extract from 5-day germinated seeds.

Figure 10 is a bar diagram illustrating the bactericidal effect of native human lactoferrin ("nHLF") and purified recombinant human lactoferrin produced by transgenic rice ("rHLF") on growth of *E. coli* (EPEC) after pepsin/pancreatic treatment.

Figure 11 is a graph illustrating pH-dependent iron release by native human lactoferrin ("nHLF") and purified recombinant human lactoferrin produced by transgenic rice seeds ("rHLF").

Figure 12 shows the binding and uptake of HLF to Caco-2 cells after *in vitro* digestion. Figure 12 A shows the determination of Dissociation constant. Figure 12B shows the number of binding sites for HLF on Caco-2 cells. Figure 12C shows the total uptake of HLF and Fe to Caco-2 cells within 24 h. Figure 12D shows degradation of HLF after uptake into Caco cells determined by the amount of free  $^{125}\text{I}$  in the cell fractions.

Figure 13 shows three AAT plasmids: pAPI255 containing Glb promoter, Glb signal peptide, codon-optimized AAT gene, Nos terminator and ampicillin resistance gene; pAPI250 containing Gt1 promoter, Gt1 signal peptide, codon-optimized AAT gene, Nos terminator and ampicillin resistance gene; and pAPI282 containing Bx7 promoter, Bx7 signal peptide, codon-optimized AAT gene, Nos terminator and ampicillin resistance gene.

Figure 14 shows Coomassie brilliant blue staining of aqueous phase extraction of transgenic rice cells expressing human AAT. Both untransformed and transgenic rice grains were ground with PBS. The resulting extract was spun at 14,000 rpm at 4° C for 10 min. Supernatant was collected and loaded onto a precast SDS-PAGE gel.

Figure 15 shows Western blot analysis of recombinant human AAT from transgenic rice grains. The extract from transgenic rice grain was separated by SDS-PAGE gel and then blotted onto a filter. The identification of AAT in rice grain was carried out by anti-AAT antibody by Western analysis.

Figure 16 shows Coomassie staining (Figure 16A) and western blot analysis (Figure 16B) of protein from transgenic rice grains expressing AAT. The activity of rAAT was demonstrated by a band shift assay. AAT samples from different sources were incubated with equal moles of porcine pancreatic elastase (PPE) at 37°C for 15 min. Negative control for band shift assay was prepared with the AAT samples incubated with equal volume of PPE added. Lane M is molecular weight markers. Lane 1a is purified AAT from human plasma. Lane 1b

is purified AAT from human plasma + PPE. Lane 2a is protein extract containing AAT from transgenic rice seed; Lane 2b is protein extract containing AAT from transgenic rice seed + PPE. Lane 3a is untransformed seed extract. Lane 3b is untransformed seed extract + PPE. A shifted band was shown in lane 1b, 2b and 3b in Figure 16A. The shifted band was confirmed to contain AAT entity by Western blot in Figure 16B.

Figures 17A-C are schematic representations of 3 plasmids containing the Reb coding sequence under the control of 3 different promoters. Figure 17A shows the globulin promoter (Glb), with the Reb gene and the Reb terminator. Figure 17B shows the actin promoter (Act), with the Reb gene and the Reb terminator. Figure 17C shows the native Reb promoter, with the Reb gene and the Reb terminator.

Figures 18A-B are schematic depictions of 2 plasmids which contain different transcription factor coding sequences under the control of the rice endosperm-specific glutelin promoter (Gt-1). Figure 18A shows plasmid pGT1-BPBF (API286) containing the Gt1 promoter, barley prolamin box binding factor (BPBF), Nos terminator and kanamycin resistance gene. Figure 18B shows pGT1-PBF (API285) containing the Gt1 promoter, the maize prolamin box binding factor (PBF), Nos terminator and kanamycin resistance gene.

Figure 19 illustrates the results of an analysis for the expression of recombinant human lysozyme in mature seed of  $T_0$  transgenic plants derived from progenitor cells transformed with constructs containing the human lysozyme gene expressed under the control of the Glb promoter and the Reb gene expressed under the control of its own promoter ("Native-Reb"). Seeds of 30 plants containing the Reb and lysozyme genes and seeds from 17 plants containing only the lysozyme gene were analyzed for lysozyme, with twenty individual seeds of each plant analyzed.

Figure 20 is a comparison of the codon-optimized epidermal growth factor sequence ("Egfactor") with a native epidermal growth factor sequence ("Native Gene"), aligned to show 53 codons in the mature sequences, with 27 (51%) codon changes and 30 (19%) nucleotides changes.

Figure 21 is a restriction map of the 4,143 bp plasmid, API270 (pGlb-EFG v2.1), showing an expression cassette for epidermal growth factor ("EGF"), and containing a Glb promoter, a Glb signal peptide, codon optimized EGF, a Nos

terminator and an ampicillin resistance selectable marker.

Figure 22 is a restriction map of the 3877 bp plasmid, API303 (pGt1-EGF v2.1), showing an expression cassette for epidermal growth factor (EGF), and containing a rice Gt1 promoter, a Gt1 signal peptide, codon optimized EGF, a Nos terminator and an ampicillin resistance selectable marker.

Figure 23 is a Western blot analysis of recombinant human EFG ("rhEGF") in transgenic rice seed. Lane 1 shows a broad range of molecular weight markers. Lane 2 shows rhEGF expressed in yeast, loaded at 125 ng. Lanes 2 to 6 show rhEGF expressed from different transgenic rice seeds. Lane 7 is from seeds of control untransformed TP 309.

Figure 24 is a comparison of the codon-optimized insulin-like growth factor I sequence ("Insgfact") with a native human insulin-like growth factor I sequence ("native Gene"), aligned to show 70 codons in the mature sequences, with 40 (57%) codon changes and 47 (22%) nucleotides changes.

Figure 25 is a restriction map of the 3928 bp plasmid, API304 (pGt1-IGF v2.1), showing an expression cassette for insulin-like growth factor I ("IGF"), and containing a rice Gt1 promoter, a Gt1 signal peptide, codon optimized IGF, a Nos terminator and an ampicillin resistance selectable marker.

Figure 26 is a restriction map of the 4194 bp plasmid, API271 (pGlb-IGF v2.1), showing an expression cassette for insulin-like growth factor I ("IGF"), and containing a Glb promoter, a Glb signal peptide, codon optimized IGF, a Nos terminator and an ampicillin resistance selectable marker.

Figure 27 is a Western blot analysis of recombinant human IGF-I ("rhIGF") expressed in transgenic rice seeds. Lane 1 shows a broad range of molecular weight markers. Lane 2 shows rhIGF expressed in yeast, loaded at 1 µg. Lanes 3-9 show rhIGF from different transgenic seeds. Lane 10 is from seeds of control untransformed TP 309.

Figure 28 is a restriction map of the 5250 bp plasmid, API321 (pGlb-gt1sig-Haptocorrin v 2.1), showing an expression cassette for haptocorrin, and containing a Glb promoter, a Gt1 signal peptide, codon optimized haptocorrin, a Nos terminator and an ampicillin resistance selectable marker.

Figure 29 is a restriction map of the 4948 bp plasmid, API320 (pGt1-Haptocorrin v 2.1), showing an expression cassette for haptocorrin, and containing a Gt1 promoter, a Gt1 signal peptide, codon optimized haptocorrin, a

Nos terminator and an ampicillin resistance selectable marker.

Figure 30 is a restriction map of the 4468 bp plasmid, API292 (pGlb-kcasein v2.1), showing an expression cassette for kappa-casein ("k-casein"), and containing a Glb promoter, a Glb signal peptide, a k-casein gene, a Nos terminator and an ampicillin resistance selectable marker.

Figure 31 is a restriction map of the 4204 bp plasmid, API297 (pGT1-kaapa-Casein v2.1), showing an expression cassette for kappa-casein, and containing a Gt1 promoter, a Gt1 signal peptide, mature kappa-casein polypeptide encoding gene, a Nos terminator and an ampicillin resistance selectable marker.

Figure 32 is a restriction map of the 4834 bp plasmid, API420 (pGT1-LAD), showing an expression cassette for lactahedrin, and containing a Gt1 promoter, a Gt1 signal peptide, lactohedrin gene, a Nos terminator and a kanamycin resistance selectable marker.

Figure 33 is a restriction map of the 5638 bp plasmid, API418 (pGT1-LPO-S), showing an expression cassette for lactoperoxidase (minus the propeptide), and containing a Gt1 promoter, a Gt1 signal peptide, lactoperoxidase gene without the propeptide, a Nos terminator and a kanamycin resistance selectable marker.

Figure 34 is a restriction map of the 5801 bp plasmid, API416 (pGT1-lactoperoxidase), showing an expression cassette for codon optimized human lactoperoxidase, and containing a rice Gt1 promoter, a Gt1 signal peptide, codon optimized lactoperoxidase, a Nos terminator and a kanamycin resistance selectable marker.

Figure 35 is a restriction map of the 4408 bp plasmid, API230 (pBX7-Lysozyme v2.1.1), showing an expression cassette for codon optimized lysozyme, and containing a BX-7 promoter, a Gt1 signal peptide, codon optimized lysozyme gene, a Nos terminator and an ampicillin resistance selectable marker.

Figures 36A-B represent schematic diagrams of the map of 2 plasmids, API254 (Figure 36A) and API264 (Figure 36B) containing heterologous protein coding sequences under the control of the rice endosperm-specific globulin promoter (Glb), the Glb signal peptide, and Nos terminator. API254 contains the lactoferrin coding sequence, and API264 contains the human lysozyme coding sequence.

Figure 37 is a restriction map of the 4271 bp plasmid, API225, showing an

expression cassette for codon optimized lysozyme, and containing a GT-3 promoter, a Gt1 signal peptide, codon optimized lysozyme, a Nos terminator and an ampicillin resistance selectable marker.

Figure 38 is a restriction map of the 4106 bp plasmid, API229, showing an expression cassette for codon optimized lysozyme, and containing a RP-6 promoter, a Gt1 signal peptide, codon optimized lysosyme, a Nos terminator and an ampicillin resistance selectable marker.

Figures 39A-B are a comparison of the expression of lysozyme under Gt1 or Glb promoter with Gt1 signal peptide or Glb signal peptide. FIG. 39A is a schematic representation of plasmid API159 that contains Gt1 promoter, Gt1 signal peptide, a lysozyme gene and Nos terminator; plasmid API 228 that contains Glb promoter, Gt1 signal peptide, a lysozyme gene and Nos terminator; and plasmid API264 that contains Glb promoter, Glb signal peptide, a lysozyme gene and Nos terminator. FIG. 39B shows the activities of lysozyme in lysozyme-positive seeds produced in transgenic rice plants transformed with API159, API228 and API264. The seeds from multiple lines of each construct were analyzed by the lysozyme activity assay. Individual seeds from each plant were analyzed. Seeds lacking detectable amounts of lysozyme were excluded. The activities of 20-lysozyme-positive seeds per plant, including both hemizygous and homozygous seeds were averaged. The average activities were plotted on the chart.

Figure 40 shows the expression time course of human lysozyme during endosperm development in transgenic line. Ten spikelets were harvested at 7, 14, 21, 28, 35, 42 and 49 days after pollination ("DAP") and analyzed by the lysozyme activity assay. The dark bars were from 159-1-53-16-1. The light bars were from 264-1-92-6-1.

Figure 41 is a bar graph comparing the level of lysozyme expression in transgenic T1 rice seeds under 7 different promoters: Gt1, Glb, Glub-2, Bx7, Gt3, Glub-1 and Rp6. All constructs contained a Gt1 signal peptide.

## **Detailed Description of the Invention**

### **I. Definitions**

Unless otherwise indicated, all terms used herein have the meanings given below, and are generally consistent with same meaning that the terms have to those

skilled in the art of the present invention. Practitioners are particularly directed to Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (Second Edition), Cold Spring Harbor Press, Plainview, N.Y. and Ausubel FM *et al.* (1993) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y., for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology,  
5 protocols, and reagents described, as these may vary.

All publications cited herein are expressly incorporated herein by reference for the purpose of describing and disclosing compositions and methodologies that might be used in connection with the invention.

10 The term "polypeptide" refers to a biopolymer compound made up of a single chain of amino acid residues linked by peptide bonds. The term "protein" as used herein may be synonymous with the term "polypeptide" or may refer, in addition, to a complex of two or more polypeptides.

The term "anti-microbial protein" refers to a protein that is anti-bacterial and can  
15 include acute phase proteins, cationic anti-microbial peptides and probiotic proteins. Such anti-microbial proteins are capable of inhibiting the growth of one or more of Gram-negative bacteria, Gram-positive bacteria, fungi (including yeast), parasites (including planaria and nematodes) and viruses. Typically, such anti-microbial peptides exhibit selective biological activity against such microbes over eukaryotic cells.

20 The term "anti-bacterial protein" refers to a protein that is bacteriostatic or bactericidal in nature.

The term "bacteriostatic protein" refers to a protein capable of inhibiting the growth of, but not capable of killing bacteria.

The term "bactericidal protein" refers to a protein capable of killing bacteria.

25 The term "vector" refers to a nucleic acid construct designed for transfer between different host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art. Accordingly,  
30 an "expression cassette" or "expression vector" is a nucleic acid construct generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a target cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette

portion of an expression vector includes, among other sequences, a nucleic acid sequence to be transcribed and a promoter.

The term "plasmid" refers to a circular double-stranded (ds) DNA construct used as a cloning vector, and which forms an extrachromosomal self-replicating genetic  
5 element in many bacteria and some eukaryotes.

The term "selectable marker-encoding nucleotide sequence" refers to a nucleotide sequence capable of expression in plant cells and where expression of the selectable marker confers to plant cells containing the expressed gene the ability to grow in the presence of a selective agent. As used herein, the term "Bar gene" refers to a  
10 nucleotide sequence encoding a phosphinothricin acetyltransferase enzyme that upon expression confers resistance to the herbicide glufosinate-ammonium ("Basta").

A "transcription regulatory region" or "promoter" refers to nucleic acid sequences that influence and/or promote initiation of transcription. Promoters are typically considered to include regulatory regions, such as enhancer or inducer elements. The  
15 promoter will generally be appropriate to the host cell in which the target gene is being expressed. The promoter, together with other transcriptional and translational regulatory nucleic acid sequences (also termed "control sequences"), is necessary to express any given gene. In general, the transcriptional and translational regulatory sequences include, but are not limited to, promoter sequences, ribosomal binding sites,  
20 transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences.

"Chimeric gene" or "heterologous nucleic acid construct", as defined herein refers to a construct which has been introduced into a host and may include parts of different genes of exogenous or autologous origin, including regulatory elements. A chimeric  
25 gene construct for plant/seed transformation is typically composed of a transcriptional regulatory region (promoter) operably linked to a heterologous protein coding sequence, or, in a selectable marker heterologous nucleic acid construct, to a selectable marker gene encoding a protein conferring antibiotic resistance to transformed plant cells. A typical chimeric gene of the present invention, includes a transcriptional regulatory region  
30 inducible during seed development, a protein coding sequence, and a terminator sequence. A chimeric gene construct may also include a second DNA sequence encoding a signal peptide if secretion of the target protein is desired.

A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory

leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate

5 translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading frame. However, "operably linked" elements, e.g., enhancers, do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with

10 conventional practice.

The term "gene" means the segment of DNA involved in producing a polypeptide chain, which may or may not include regions preceding and following the coding region, e.g. 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons).

15 The term "sequence identity" means nucleic acid or amino acid sequence identity in two or more aligned sequences, aligned using a sequence alignment program. The term "% homology" is used interchangeably herein with the term "% identity" and refers to the level of nucleic acid or amino acid sequence identity between two or more aligned sequences, when aligned using a sequence alignment program. For example,

20 70% homology means the same thing as 70% sequence identity determined by a defined algorithm, and accordingly a homologue of a given sequence has greater than 80% sequence identity over a length of the given sequence. Exemplary levels of sequence identity include, but are not limited to, 80, 85, 90 or 95% or more sequence identity to a given sequence, e.g., the coding sequence for lactoferrin, as described herein.

25 Exemplary computer programs which can be used to determine identity between two sequences include, but are not limited to, the suite of BLAST programs, e.g., BLASTN, BLASTX, and TBLASTX, BLASTP and TBLASTN, publicly available on the Internet at "www.ncbi.nlm.gov/BLAST/". See, also, Altschul, S.F. *et al.*, 1990 and Altschul, S.F. *et al.*, 1997.

30 Sequence searches are typically carried out using the BLASTN program when evaluating a given nucleic acid sequence relative to nucleic acid sequences in the GenBank DNA Sequences and other public databases. The BLASTX program is preferred for searching nucleic acid sequences which have been translated in all reading frames against amino acid sequences in the GenBank Protein Sequences and other

public databases. Both BLASTN and BLASTX are run using default parameters of an open gap penalty of 11.0, and an extended gap penalty of 1.0, and utilize the BLOSUM-62 matrix. [See, Altschul, *et al.*, 1997.]

5 A preferred alignment of selected sequences in order to determine "% identity" between two or more sequences, is performed using for example, the CLUSTAL-W program in MacVector version 6.5, operated with default parameters, including an open gap penalty of 10.0, an extended gap penalty of 0.1, and a BLOSUM 30 similarity matrix.

10 A nucleic acid sequence is considered to be "selectively hybridizable" to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Hybridization conditions are based on the melting temperature ( $T_m$ ) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about  $T_m - 5^\circ\text{C}$  ( $5^\circ$  below the  $T_m$  of the probe); "high stringency" at about  $5-10^\circ$  below the  $T_m$ ; "intermediate stringency" at about  $10-20^\circ$  below the  $T_m$  of the probe; and "low  
15 stringency" at about  $20-25^\circ$  below the  $T_m$ . Functionally, maximum stringency conditions may be used to identify sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify sequences having about 80% or more sequence identity with the probe.

Moderate and high stringency hybridization conditions are well known in the art  
20 (see, for example, Sambrook *et al.*, 1989, Chapters 9 and 11, and in Ausubel *et al.*, 1993, expressly incorporated by reference herein). An example of high stringency conditions includes hybridization at about  $42^\circ\text{C}$  in 50% formamide, 5X SSC, 5X Denhardt's solution, 0.5% SDS and 100  $\mu\text{g/ml}$  denatured carrier DNA followed by washing two times in 2X SSC and 0.5% SDS at room temperature and two additional times in 0.1X SSC and 0.5%  
25 SDS at  $42^\circ\text{C}$ .

As used herein, "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid sequence or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or  
30 express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention.

A plant cell, tissue, organ, or plant into which a heterologous nucleic acid construct comprising the coding sequence for an anti-microbial protein or peptide has been introduced is considered transformed, transfected, or transgenic. A transgenic or

transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of the coding sequence for an anti-microbial protein. Hence, a plant of the invention will include any plant which has  
5 a cell containing introduced nucleic acid sequences, regardless of whether the sequence was introduced into the plant directly through transformation means or introduced by generational transfer from a progenitor cell which originally received the construct by direct transformation.

The term "transgenic plant" refers to a plant that has incorporated exogenous  
10 nucleic acid sequences, *i.e.*, nucleic acid sequences which are not present in the native ("untransformed") plant or plant cell. Thus a plant having within its cells a heterologous polynucleotide is referred to herein as a "transgenic plant". The heterologous polynucleotide can be either stably integrated into the genome, or can be extra-chromosomal. Preferably, the polynucleotide of the present invention is stably integrated  
15 into the genome such that the polynucleotide is passed on to successive generations. The term "transgenic" as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or  
20 spontaneous mutation. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acids including those transgenics initially so altered as well as those created by sexual crosses or asexual reproduction of the initial transgenics.

Terms "transformed", "stably transformed" or "transgenic" with reference to a  
25 plant cell means the plant cell has a non-native (heterologous) nucleic acid sequence integrated into its genome which is maintained through two or more generations.

The term "expression" with respect to a protein or peptide refers to the process by which the protein or peptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation. The term "expression" may  
30 also be used with respect to the generation of RNA from a DNA sequence.

The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell where the nucleic acid sequence may be incorporated into the genome of the cell (for example,

chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA).

By "host cell" is meant a cell which contains a vector and supports the replication, and/or transcription or transcription and translation (expression) of the expression  
5 construct. Host cells for use in the present invention can be prokaryotic cells, such as *E. coli*, or eukaryotic cells such as yeast, plant, insect, amphibian, or mammalian cells. In general, host cells are monocotyledenous or dicotyledenous plant cells.

A "plant cell" refers to any cell derived from a plant, including undifferentiated tissue (e.g., callus) as well as plant seeds, pollen, progagules and embryos.

10 The term "mature plant" refers to a fully differentiated plant.

The terms "native" and "wild-type" relative to a given plant trait or phenotype refers to the form in which that trait or phenotype is found in the same variety of plant in nature.

The term "plant" includes reference to whole plants, plant organs (for example,  
15 leaves, stems, roots, etc.), seeds, and plant cells and progeny of same. Plant cell, as used herein includes, without limitation, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves roots shoots, gametophytes, sporophytes, pollen, and microspores. The class of plants that can be used in the methods of the present invention is generally as broad as the class of higher plants amenable to  
20 transformation techniques, including both monocotyledenous and dicotyledenous plants.

The term "seed" is meant to encompass all seed components, including, for example, the coleoptile and leaves, radicle and coleorhiza, scutulum, starchy endosperm, aleurone layer, pericarp and/or testa, either during seed maturation and seed germination.

25 The term "seed in a form for use as a food or food supplement" includes, but is not limited to, seed fractions such as de-hulled whole seed, flour (seed that has been de-hulled by milling and ground into a powder) a seed protein extract (where the protein fraction of the flour has been separated from the carbohydrate fraction) and/or a purified protein fraction derived from the transgenic grain.

30 The term "purifying" is used interchangeably with the term "isolating" and generally refers to the separation of a particular component from other components of the environment in which it was found or produced. For example, purifying a recombinant protein from plant cells in which it was produced typically means subjecting

transgenic protein containing plant material to biochemical purification and/or column chromatography.

The term "active" refers to any biological activity associated with a particular milk protein, such as the enzymatic activity associated with human lysozyme. It follows that  
5 the biological activity of a given milk protein refers to any biological activity typically attributed to that factor by those of skill in the art.

The term "human milke protein" or "proteins normally present in human milk" refers to one or more proteins, or biologically active fragments thereof, found in normal human milk, including, without limitation, of lactoferrin, lysozyme, lactoferricin, EGF, IGF-  
10 I, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, and immunoglobulins, and biologically active fragments thereof.

The term "nutritionally enhanced food" refers to a food, typically a processed food, to which a seed-produced human milk protein has been added, in an amount effective to confer some health benefit, such as improved gut health, resistance to pathogenic  
15 bacteria, or iron transport, to a human consuming the food.

"Plant-derived food ingredients" refers to plant-derived food stuff, typically monocot grain, but also including, separately, lectins, gums, sugars, plant-produced proteins and lipids, that may be blended or combined, alone or in combination with one or more plant-derived ingredients, to form an edible food.

20 "Monocot seed components" refers to carbohydrate, protein, and lipid components extractable from monocot seeds, typically mature monocot seeds.

"Malting-seed components" refers to seed-derived components, predominantly carbohydrate components, after conversion of complex carbohydrates to malt sugars by malting, *i.e.*, treating with malting enzymes such as a barley amylase and glucanases,  
25 under conditions effective to conversion seed-derived carbohydrates to malt sugars.

"Substantially unpurified form", as applied to human milk proteins in a seed extract means that the protein or proteins present in the extract are present in an amount less than 50% by weight, typically between 0.1 and 10 percent by weight.

"Seed maturation" or "grain development" refers to the period starting with  
30 fertilization in which metabolizable reserves, *e.g.*, sugars, oligosaccharides, starch, phenolics, amino acids, and proteins, are deposited, with and without vacuole targeting, to various tissues in the seed (grain), *e.g.*, endosperm, testa, aleurone layer, and scutellar epithelium, leading to grain enlargement, grain filling, and ending with grain desiccation.

"Inducible during seed maturation" refers to promoters which are turned on substantially (greater than 25%) during seed maturation.

"Heterologous DNA" or "foreign DNA" refers to DNA which has been introduced into plant cells from another source, or which is from a plant source, including the same  
5 plant source, but which is under the control of a promoter or terminator that does not normally regulate expression of the heterologous DNA.

"Heterologous protein" is a protein, including a polypeptide, encoded by a heterologous DNA.

A "signal/targeting/transport sequence" is an N- or C-terminal polypeptide  
10 sequence which is effective to localize the polypeptide or protein to which it is attached to a selected intracellular or extracellular region, including an intracellular vacuole or other protein storage body, chloroplast, mitochondria, or endoplasmic reticulum, or extracellular space or seed region, such as the endosperm, following secretion from the cell.

15 A "product" encoded by a DNA molecule includes, for example, RNA molecules and polypeptides.

A DNA sequence is "derived from" a gene if it corresponds in sequence to a segment or region of that gene. Segments of genes which may be derived from a gene include the promoter region, the 5' untranslated region, and the 3' untranslated region of  
20 the gene.

"Alpha-amylase" as used herein refers to an enzyme which principally breaks starch into dextrins.

"Beta-amylase" as used herein refers to an enzyme which converts starch and dextrins into maltose.

25 "Cereal adjuncts" as used herein refers to cereal grains, principally barley, wheat, rye, oats, maize, sorghum and rice, or processed whole or portions thereof, especially the starch fraction, which are added to the barley mash, which allows the barley enzymes to hydrolyze both the barley starch and the starch derived from the cereal adjunct.

"Transgenic cereal adjuncts" as used herein refers to transgenic cereal grains, principally  
30 barley, wheat, rye, oats, maize, sorghum and rice, and which is expressing a recombinant molecule in a grain part, principally the endosperm (starch) layer.

"Conversion" as used herein refers to the process of starch hydrolysis, usually catalyzed by acid or enzyme action, which produces dextrose, maltose, and higher polysaccharides from starch.

"Diastatic enzyme (amylolytic)" as used herein refers to an enzyme capable of causing the hydrolysis of starch.

"Diastatic malt flour" as used herein refers to enzyme active flour milled from germinated (malted) barley.

5 "Diastatic malt syrup" as used herein refers to enzyme active liquid malt syrup (barley and cereal adjuncts).

"Dry diastatic malt" as used herein refers to a blend of diastatic malted barley flour, wheat flour and dextrose with standardized enzyme levels at 20degrees and 60 degrees Lintner.

10 "Dry nondiastatic malt" as used herein refers to spray dried form of liquid nondiastatic malt extract or syrup.

"Lintner" as used herein refers to a laboratory measurement of enzyme activity strength. The higher the value, the higher activity.

15 "Dried malt" as used herein refers to the dried grain resulting from controlled germination of cereal grains, usually barley, but other cereals can be malted as well.

"Malt extract" as used herein refers to a viscous concentrate of the water extract of dried malt.

20 "Maltodextrin" as used herein refers to a purified, concentrated aqueous solution of nutritive saccharides, obtained from edible starch, or the dried product derived from the solution. Maltodextrins have a dextrose equivalent of less than 20 and are considered 'non-sweet soluble solids'. They are usually marketed dry, but may be obtained as a concentrated solution. Maltodextrins are usually offered as 10 to 14 D.E. products or as 15 to 19 D.E. versions. Another maltodextrin, with a D.E. of about 5, is sometimes manufactured, but currently not used widely. Composition of maltodextrins is  
25 roughly 65 to 80% higher saccharides, 4 to 9% pentasaccharides, 4 to 7% tetrasaccharides, and 5 to 9% trisaccharides. Traces of mono and disaccharides are present. They are usually used as bulking agents or viscosity builders, without sweetness.

30 "Malt syrup" as used herein refers to viscous concentrate of the water extract of dried 'malt' and other cereal grains.

"Malt" refers to a malt extract or malt syrup.

"Nondiastatic malt syrup" as used herein refers to liquid malt syrup (barley and cereal adjuncts) without enzyme activity.

"Transgenic malt extract" as used herein refers to a vicious concentrate of the water extract of dried malt which includes a recombinant protein, polypeptide and/or metabolite.

5 "Transgenic malt syrup" as used herein refers to vicious concentrate of the water extract of dried 'malt' and other cereal grains which includes a recombinant protein, polypeptide and/or metabolite.

## II. Milk Products-State of the Art/Issues

10 Human milk provided by healthy and well-nourished mothers is believed by pediatricians and nutritionists to be the optimal way to feed infants during the first six months of life. Breast milk not only provides the infant with a well-balanced supply of nutrients, but also a multitude of unique components that facilitate nutrient digestion and absorption, protect against microorganisms and promote growth and development. Human milk is a source of peptides, amino acids, and nitrogen and also contains whey  
15 proteins involved in the development of the immune response (e.g., immunoglobulins), together with other non-immunologic defense proteins (e.g., lactoferrin).

However, infant formulas are often used as a nutritional source for infants less than one year of age for a variety of reasons, e.g., insufficient milk production by or pathogenic infection of, the mother. Infant formulas and not standard cow's milk are  
20 used because (1) cow's milk has more than twice the protein of breast milk or infant formula and this protein may be hard for babies to digest; (2) the level of iron, zinc and vitamin C (which babies need in their diet) is low in cow's milk; and (3) the level of sodium level is three to four times that of breast milk and generally too high for infants less than a year old. A number of types of infant formulas which vary in caloric content,  
25 nutrient composition, digestibility, taste, and cost are available as an alternative to breast milk. Examples include standard cow milk-based formulas, soy protein formulas and formulas for premature infants or infants with special dietary needs due to allergies, etc.

During the last several decades, improved infant formulas have become available that are safe and contain nutrient concentrations similar to, or higher than, breast milk.  
30 However, breast-fed infants still have a lower prevalence of infection than formula-fed infants and when they become ill, the duration of both diarrhea and upper respiratory infections is shorter than in formula-fed infants. (See, e.g., Kovar *et al.*, 1984, and Dewey *et al.*, 1995). In addition, it has been reported that breast-fed infants have a different growth pattern than formula-fed infants (Dewey *et al.*, 1992; Dewey *et al.*, 1993),

and epidemiological studies show that they have a lower incidence of chronic diseases, such as diabetes and coronary heart disease.

It has been postulated that many of advantages to infants provided by mother's milk are effectuated through unique proteins present in breast milk, but not in baby formula (Lönnerdal, 1985). Human milk proteins are unique and even if the alternative protein sources used in infant formulas (e.g., skim milk, whey protein and soy isolates) mimic the amino acid concentration and ratio found in breast milk, the biological properties of human proteins cannot be readily copied.

Exemplary unique proteins present in human milk include lactoferrin and lysozyme. Lactoferrin is an iron-binding protein found in the granules of neutrophils which exerts an antimicrobial activity and lysozyme is a crystalline, basic protein present in saliva, tears, egg white, and many animal fluids, which functions as an antibacterial enzyme.

Improved food compositions containing human milk proteins are provided by the present invention. The human milk proteins are produced in the grain of transgenic plants and added to novel food compositions, infant formula being one example. Infant formula containing such recombinant human milk proteins are useful in supplementing or enhancing the diet of infants, particularly very-low-birth-weight infants.

Other foods that may be supplemented with human milk proteins include, but are not limited to foods where recombinant lactoferrin can be added and utilized as an iron supplement replacing the need for exogenously added iron in the final food formulation.

### III. Compositions Containing Human Milk Proteins

The present invention provides food supplement compositions (also termed "improved food compositions" comprising human milk proteins and methods of making such compositions. In practicing the invention, a human milk protein is produced in the seeds or grain of transgenic plants which express the nucleic acid coding sequence for the human milk protein and the transgenic grains added to a food such as an infant formula to result in an "improved food compositions". More specifically, the invention is based on the expression of human milk proteins, exemplified by human lactoferrin (hLF) and human lysozyme, under the control of a seed specific promoter in rice. The human protein produced by transgenic plants is compared to the native form of the same protein, information on the stability of the recombinant protein and the advantages of

using rice grain containing such recombinant human milk protein in infant formula and/or other foods, is further described below.

The invention relies on the use of heterologous nucleic acid constructs including the coding sequence for a commercially important milk protein or polypeptide of

5 nutritional and/or therapeutic value, exemplified herein by lactoferrin and lysozyme.

The exemplary milk proteins, lysozyme and lactoferrin are an integral part of the immune system of multicellular animals. They are found in epithelial secretions (tears, mucous, gastric juice) and blood plasma of mammals, birds, reptiles, amphibia, and a variety of invertebrates. They are also enriched in mammalian milk and avian eggs, 10 where they serve as primary antimicrobial proteins. Furthermore, lysozyme is a major component of the secretory granules of neutrophils and macrophages and is released at the site of infection in the earliest stages of the immune response. Lactoferrin is found at high concentrations within specific granules of polymorphonuclear leukocytes.

It has previously been demonstrated that lysozyme and lactoferrin are efficacious 15 in promoting resistance to infectious diseases in experimental animals and humans and that they play a role of primary defense proteins on epithelial surfaces in addition to being important determinants in the establishment of a healthy microflora within the digestive tract. These properties suggest that food supplements comprising lysozyme and/or lactoferrin will be beneficial to the overall health of infants.

20 The improved food compositions of the invention include milk proteins such as lactoferrin, and lysozyme, produced in the grain of transgenic plants, which are useful for improved nutrition. In one preferred approach, the improved food compositions are administered to an infant. Typically the improved food compositions, *e.g.*, infant formula contain one or more recombinant human milk proteins in an amount that corresponds to 25 the amount and proportions of the same human milk proteins found in endogenous human milk.

#### A. Lysozymes

Human milk lysozyme, called muramidase or peptidoglycan N-acetylmuramoyl- 30 hydrolase (EC 3.2.1.17) contains 130 amino acid residues and is a protein of 14.7 kDa in size. Human lysozyme is non-glycosylated and possesses unusual stability *in vitro* and *in vivo* due to its amino acid and secondary structure.

Lysozyme is one of the most abundant proteins present in human milk with a concentration of about 400 µg/ml. The concentration of lysozyme is approximately 0.13

µg/ml in cow's milk (almost 3000 times less than found in human milk), 0.25 µg/ml in goat's milk, 0.1 µg/ml in sheep's milk and almost absent in rodent's milk (Chandan RC, 1968). Lysozyme is also found in other mammalian secretions, such as tears and saliva.

The protective role of lysozyme has been observed to include lysis of microbial  
5 cell walls, adjuvant activity of the end products peptidoglycan lysis, direct immunomodulating effects on leukocytes, and neutralization of bacterial endotoxins. The bacteriostatic and bactericidal actions of lysozyme were originally discovered by Flemming in 1922 and have been studied in detail. Lysozyme is effective against both  
10 antimicrobial effects of lysozyme often act synergistically with other defense molecules, including immunoglobulin and lactoferrin. Furthermore, structural changes in the cell wall due to lysozyme render bacteria more susceptible to phagocytosis by macrophages and neutrophils.

The hydrolysis of microbial peptidoglycans results in the release of the cleavage  
15 product, muramyl dipeptide, which is a potent adjuvant and is the active component of Freund's complete adjuvant. Muramyl dipeptide enhances IgA production, macrophage activation, and rapid clearance of a variety of bacterial pathogens *in vivo*. Lysozyme itself is also immunomodulatory. It directly interacts with the cell membrane of phagocytes to increase their uptake of bacteria. Lysozyme also augments the  
20 proliferative response of mitogen stimulated lymphocytes to interleukin-2 and increases the rate of synthesis of IgG and IgM by more than 5-and 2- fold respectively. Furthermore, the immunomodulatory action of lysozyme is not dependent upon enzymatic activity and is retained following denaturation. When lysozyme is fed to mice, it increases the number of intraepithelial and mesenteric lymph node lymphocytes that  
25 display antigens.

Lysozymes act as enzymes that cleave peptidoglycans, and ubiquitous cell wall component of microorganisms, in particular bacteria. Specifically, lysozymes are 1,4- $\beta$ -acetylmuramidases that hydrolyze the glycoside bond between N-acetylmuramic acid and N-acetylglucosamine. Gram-positive bacteria are highly susceptible to lysozyme  
30 due to the polypeptidoglycan on the outside of the cell wall. Gram-negative strains have a single polypeptidoglycan layer covered by lipopolysaccharides and are therefore less susceptible to lysis by lysozyme, however, the sensitivity can be increased by the addition of EDTA (Schütte and Kula, 1990). Lysozyme also exhibits antiviral activity, as exemplified by the significant reduction in recurrent occurrences of genital and labial

herpes after oral treatment of patients with lysozyme (Jollès, 1996). More recently, lysozyme from chicken egg whites, human milk and human neutrophils has been shown to inhibit the growth of HIV-1 in an *in vitro* assay (Lee-Huang *et al.*, 1999). In addition, an anti-fungal activity has been demonstrated for lysozymes using oral isolates of *Candida albicans* (the most common fungal causative agent of oropharyngeal infection in humans; 5 (Samaranayake *et al.*, 1997). Lysozyme thus functions as a broad spectrum antimicrobial agent.

The ability of lysozyme to bind bacterial endotoxins, especially LPS, confers an important anti-microbial property to the molecule. Lysozyme binds electrostatically to the 10 lipid A component of bacterial endotoxins at a 1:3 molar ratio. The resulting conformational change in endotoxin keeps it from interacting with macrophage receptors and dampens the release of pro-inflammatory cytokines such as interleukin-1 (IL-1), interleukin-6 (IL-6), and tumor necrosis factor (TNF). Thus, lysozyme exhibits anti-inflammatory activity during pathogen challenges.

15 The current major commercial source for lysozyme is chicken egg whites. Sequence analysis shows that lysozyme from chicken egg whites exhibits only partial homology (60%) with that synthesized by humans. Chicken and human lysozyme do not cross-react with their respective antibodies (Faure *et al.*, 1970), indicating significant structural differences between these two lysozymes. Human lysozyme has been purified 20 from breast milk (Boesman-Finkelstein *et al.*, 1982; Wang *et al.*, 1984), neutrophils (Lollike *et al.*, 1995), and urine of hemodialysis patients (Takai *et al.*, 1996). Breast milk remains the main source for isolation of human lysozyme, but the supply is limited. Precautions are required for isolation of the enzyme from human sources to avoid contamination with viral and microbial pathogens.

25 Recombinant human lysozyme has been produced in the mammary gland of transgenic mice. The enzyme retained its antimicrobial activity, but the final concentration in the milk was low (Maga *et al.*, 1998; Maga *et al.*, 1994; Maga *et al.*, 1995). Human lysozyme has been expressed in *Aspergillus oryzae* (*A. oryzae*) (Tsuchiya *et al.*, 1992) yeast (*S. cerevisiae*; Castañón *et al.*, 1988; Jigami *et al.*, 1986; 30 and Yoshimura *et al.*, 1988) and in small amounts in tobacco leaves (Nakajima *et al.*, 1997). However, the expression level of recombinant human lysozyme in these organisms could be very low, and the cost of these forms may be prohibitive for food applications. In addition, human lysozyme produced in microorganisms may require extensive purification before it can be used in foods, particularly for infants and children.

In contrast to many other proteins, lysozyme is highly resistant to digestion in the gastrointestinal tract. *In vitro* studies have demonstrated that both molecules are resistant to hydrolysis by pepsin in the pH range found in the stomach. Furthermore, partial denaturation of lysozyme increases its bactericidal activity against some types of bacteria, and low pH, such as found in the stomach, increases the bactericidal effects of lysozyme. A proteolytic fragment (amino acids 98-112 of chicken egg white lysozyme) completely lacking enzymatic activity has been found to be the active bactericidal component of lysozyme. Additionally, a fragment of lactoferrin, known as lactoferricin, is formed by limited proteolytic digestion and has been shown to have extremely effective antibacterial activity.

The rice produced human lysozyme of the present invention exhibits acid pH resistance, as well as resistance to pepsin and pancreatin to make it resistant to digestion in the gastrointestinal tract. The excellent thermostability provides the feasibility to pasteurize products that include the recombinant human lysozyme.

#### B. Lactoferrin

Lactoferrin is an iron-binding protein found in the granules of neutrophils where it apparently exerts an antimicrobial activity by withholding iron from ingested bacteria and fungi; it also occurs in many secretions and exudates (milk, tears, mucus, saliva, bile, etc.). In addition to its role in iron transport, lactoferrin has bacteriostatic and bactericidal activities, in addition to playing a role as an anti-oxidant (Satue-Gracia *et al.*, 2000).

The mature lactoferrin (LF) polypeptide consists of 692 amino acids, consists of a single-chain polypeptide that is relatively resistant to proteolysis, is glycosylated at two sites (N138 and N478) and has a molecular weight of about 80 kD. Human lactoferrin (hLF) is found in human milk at high concentrations (at an average of 1-3 mg/ml), and at lower concentration (0.1-0.3 mg/ml), in exocrine fluids of glandular epithelium cells such as bile, tears, saliva etc.

The primary functions of lactoferrin have been described as iron regulation, immune modulation and protection from infectious microbes. Lactoferrin can bind two ferric ions and has been shown to have biological activities including bacteriostatic (Bullen *et al.*, 1972), bactericidal (Arnold, *et al.*, 1980) and growth factor activity *in vitro*. Further, lactoferrin can promote the growth of bacteria that are beneficial to the host organism by releasing iron in their presence. Additional studies have recently shown lactoferrin to have antiviral activity towards cytomegalovirus, herpes simplex virus, rotavirus and HIV

both *in vitro* and *in vivo*. (See, e.g., Fujihara *et al.*, 1995; Grover *et al.*, 1997; and Harmsen *et al.*, 1995.)

Lactoferrin, like transferrin, has a strong capacity to bind free iron under physiological conditions due to its tertiary structure, which consists of two globular lobes  
5 linked by an extended alpha-helix. The ability of lactoferrin to scavenge iron from the physiological environment can effectively inhibit the growth of "more than 90% of all microorganisms" by depriving them of a necessary component of their metabolism, which will inhibit their growth *in vivo* and *in vitro*.

Unrelated to iron binding, the bactericidal activity of lactoferrin stems from its  
10 ability to destabilize the outer membrane of gram-negative bacteria through the liberation of lipopolysaccharides that constitute the cell walls of the bacteria. Additionally, lactoferrin has recently been shown to bind to prions, a group of molecules common in *E. coli*, causing permeability changes in the cell wall. Studies in germfree piglets fed lactoferrin before being challenged with *E. coli* show significant decrease in mortality  
15 compared to the control group.

Recombinant LF (rLF) has been produced as a fusion protein in *Aspergillus oryzae* (Ward *et al.*, 1992) and in the baculovirus expression system (Salmon *et al.*, 1997). The *Aspergillus*-produced protein will require a high degree of purification as well as safety and toxicity testing prior to using it as a food additive (Lönnerdal, 1996).

20 Lactoferrin has also been expressed in tobacco (*Nicotiana tabacum* L. cv Bright Yellow) cell culture (Mitra and Zhang, 1994), tobacco plants (Salmon *et al.*, 1998) and potato (*Solanum tuberosum*) plants (Chong and Langridge, 2000). In tobacco cell culture the protein was truncated, whereas in tobacco and potato plants the rLF was processed correctly, but its expression level was very low (0.1% of total soluble protein) (Chong and  
25 Langridge, 2000). However, the expression level of recombinant human lactoferrin in these organisms could be very low, and the cost of these forms may be prohibitive for food applications. In addition, human lactoferrin produced in microorganisms may require extensive purification before it can be used in foods, particularly for infants and children.

30 In contrast to most other proteins, lactoferrin has also been shown to be resistant to proteolytic degradation *in vitro*, with trypsin and chymotrypsin remarkably ineffective in digesting lactoferrin, particularly in its iron-saturated form. Some large fragments of lactoferrin were formed, but proteolysis was clearly limited.

### C. Lactoperoxidase

Lactoperoxidase is an enzyme which catalyzes the conversion of hydrogen peroxide to water. This enzyme is found in human milk, and plays host defensive roles through antimicrobial activity. When hydrogen peroxide and thiocyanate are added to raw milk,  
5 the SCN's oxidized by the enzyme-hydrogen peroxide complex producing bactericidal compounds which destroy Gram-negative bacteria (Shin).

### D. Kappa-Casein

This group of proteins are readily digested and account for almost half of the protein  
10 content in human milk. They are important as nutritional protein for breast-fed infants. It has also been advocated that part of the antimicrobial activity of human milk resides in the caseins, most likely the glycosylated kappa-casein (Aniansson).

### E. Alpha-1-antitrypsin ("AAT")

15 AAT belongs to the class of serpin inhibitors, has a molecular mass of 52 kD, and contains about 15% carbohydrate (Carrell et al, 1983). Concentrations of AAT in human milk range from 0.1 to 0.4 mg/mL (Davidson and Lönnnerdal, 1979; McGilligan *et al.*, 1987). While the binding affinity of AAT is highest for human neutrophil elastase, it also has affinity for pancreatic proteases such as  
20 chymotrypsin and trypsin (Beatty *et al.*, 1980).

While milk proteins have been expressed in systems such as transgenic cows and *Aspergillus* (Lönnnerdal, 1996), transgenic rice provides a more attractive vehicle for the production of recombinant human AAT for food applications. High levels of expression are possible by using the combination of regulatory elements  
25 such as promoter, signal peptide, and terminator as disclosed herein. In addition, rice is often one of the first foods introduced to infants because of its nutritional value and low allergenicity. Safety concerns about microbial expression systems (e.g. *Aspergillus*) limit the feasibility of using proteins from such sources as food components in formula (Lönnnerdal, 1996). In addition, the cost of producing and  
30 purifying proteins from these other systems is often prohibitive for food applications. Thus, expression of recombinant human milk proteins in rice may be a safe and economically viable possibility for supplementing infant formula with such proteins. (See also Chowanadisai; Huang; Johnson; Lindberg; and Rudloff).

#### F. Lactadherin

Lactadherin is a protective glycoprotein present in human milk that helps protect breast-fed infants against infection by microorganisms. Protection against certain virus infections by human milk is also associated with lactadherin. (Newburg, 1999, 1998;

5 Peterson; Hamosh).

#### G. Epidermal Growth Factor and Insulin-like Growth Factor

Epidermal Growth Factor and Insulin-like Growth Factor-1 are two growth factors present in human milk. These molecules may stimulate growth and development of the

10 infant gastrointestinal tract. (Murphy; Prosser).

#### H. Immunoglobulins

Immunoglobulins present in human act to confer resistance to a variety of pathogens to which the mother may have been exposed. (See, for example,

15 Humphreys; Kortt; Larrick; Maynard; and Peeters).

### IV. Expression Vectors For Generation Of Transgenic Plants Expressing Human Milk Proteins

Expression vectors for use in the present invention are chimeric nucleic acid

20 constructs (or expression vectors or cassettes), designed for operation in plants, with associated upstream and downstream sequences.

In general, expression vectors for use in practicing the invention include the following operably linked components that constitute a chimeric gene: (i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter, (ii)

25 operably linked to said transcriptional regulatory region, a leader DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, such as the leader sequence for targeting to a protein-storage body, and (iii) a protein-coding sequence encoding a protein normally present in human milk.

30 The chimeric gene, in turn, is typically placed in a suitable plant-transformation vector having (i) companion sequences upstream and/or downstream of the chimeric genewhich are of plasmid or viral origin and provide necessary characteristics to the vector to permit the vector to move DNA from bacteria to the desired plant host; (ii) a selectable marker

sequence; and (iii) a transcriptional termination region generally at the opposite end of the vector from the transcription initiation regulatory region.

Exemplary methods for constructing chimeric genes and transformation vectors carrying the chimeric genes are given in the examples below.

5

#### A. Promoters

In one aspect of this embodiment, the expression construct includes a transcription regulatory region (promoter) which exhibits specifically upregulated activity during seed maturation. Examples of such promoters include the maturation-specific promoter region associated with one of the following maturation-specific monocot storage proteins: rice glutelins, oryzins, and prolamines, barley hordeins, wheat gliadins and glutenins, maize zeins and glutelins, oat glutelins, and sorghum kafirins, millet pennisetins, and rye secalins. Exemplary regulatory regions from these genes are exemplified by SEQ ID NOS: 15-23, as identified in the Description of the Sequences.

Of particular interest is the expression of the nucleic acid encoding a human milk protein from a transcription initiation region that is preferentially expressed in plant seed tissue. Examples of such seed preferential transcription initiation sequences include those sequences derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Exemplary preferred promoters include a glutelin (Gt-1) promoter, as exemplified by SEQ ID NO: 18, which effects gene expression in the outer layer of the endosperm and a globulin (Glb) promoter, as exemplified by SEQ ID NO: 16, which effects gene expression in the center of the endosperm. Promoter sequences for regulating transcription of gene coding sequences operably linked thereto include naturally-occurring promoters, or regions thereof capable of directing seed-specific transcription, and hybrid promoters, which combine elements of more than one promoter. Methods for construction such hybrid promoters are well known in the art.

In some cases, the promoter is derived from the same plant species as the plant cells into which the chimeric nucleic acid construct is to be introduced. Promoters for use in the invention are typically derived from cereals such as rice, barley, wheat, oat, rye, corn, millet, triticale or sorghum.

Alternatively, a seed-specific promoter from one type of monocot may be used regulate transcription of a nucleic acid coding sequence from a different monocot or a non-cereal monocot.

Numerous types of appropriate expression vectors, and suitable regulatory sequences are known in the art for a variety of plant host cells. The transcription regulatory or promoter region is chosen to be regulated in a manner allowing for induction under seed-maturation conditions. Examples of such promoters include those associated with the following monocot storage proteins: rice glutelins, oryzins, and prolamines, barley hordeins, wheat gliadins and glutelins, maize zeins and glutelins, oat glutelins, and sorghum kafirins, millet pennisetins, and rye secualins. Exemplary promoter sequences are identified herein as SEQ ID NOS: 15-23. Other promoters suitable for expression in maturing seeds include the barley endosperm-specific B1-hordein promoter (Brandt, A., *et al.*, (1985), Glub-2 promoter, Bx7 promoter, Gt3 promoter, Glub-1 promoter and Rp-6 promoter, particularly if these promoters are used in conjunction with transcription factors. The primary structure of a B1 hordein gene from barley is provided in Carlsberg Res. Commun. 50, 333-345.

15        B. Signal/targeting/transport Sequences

In addition to encoding the protein of interest, the expression cassette or heterologous nucleic acid construct may encode a signal/targeting/transport peptide that allows processing and translocation of the protein, as appropriate. Exemplary signal/targeting/transport sequences, particularly for targeting proteins to intracellular bodies, such as vacuoles, are signal/targeting sequences associated with the monocot maturation-specific genes: glutelins, prolamines, hordeins, gliadins, glutenins, zeins, albumin, globulin, ADP glucose pyrophosphorylase, starch synthase, branching enzyme, Em, and lea. Exemplary sequences encoding a leader sequence for protein storage body are identified herein as SEQ ID NOS: 24-30.

25        In one preferred embodiment, the method is directed toward the localization of recombinant milk protein expression in a given subcellular compartment, in particular a protein-storage body, but also including the mitochondrion, endoplasmic reticulum, vacuoles, chloroplast or other plastidic compartment. For example, when recombinant milk protein expression is targeted to plastids, such as chloroplasts, in order for expression to take place the construct also employ the use of sequences to direct the gene to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, when the gene of interest is not directly inserted into the plastid, the expression construct additionally contains a gene encoding a transit peptide to direct the gene of interest to the plastid.

The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne *et al.*, 1991; Clark *et al.*, 1989; della-Cioppa *et al.*, 1987; Romer *et al.*, 1993; and Shah *et al.*, 1986. Additional transit peptides for the translocation of the protein to the endoplasmic reticulum (ER) (Chrispeels, K., 1991), nuclear localization signals (Raikhel, 1992), or vacuole may also find use in the constructs of the present invention.

Another exemplary class of signal/targeting/transport sequences are sequences effective to promote secretion of heterologous protein from aleurone cells during seed germination, including the signal sequences associated with  $\alpha$ -amylase, protease, carboxypeptidase, endoprotease, ribonuclease, DNase/RNase, (1-3)- $\beta$ -glucanase, (1-3)(1-4)- $\beta$ -glucanase, esterase, acid phosphatase, pentosamine, endoxylanase,  $\beta$ -xylopyranosidase, arabinofuranosidase,  $\beta$ -glucosidase, (1-6)- $\beta$ -glucanase, peroxidase, and lysophospholipase.

Since many protein storage proteins are under the control of a maturation-specific promoter, and this promoter is operably linked to a leader sequence for targeting to a protein body, the promoter and leader sequence can be isolated from a single protein-storage gene, then operably linked to a milk-protein storage protein in the chimeric gene construction. One preferred and exemplary promoter-leader sequence is from the rice Gt1 gene, having an exemplary sequence identified by SEQ ID NO:15. Alternatively, the promoter and leader sequence may be derived from different genes. One preferred and exemplary promoter/leader sequence combination is the rice Glb promoter linked to the rice Gt1 leader sequence, as exemplified by SEQ ID NO: 16.

### C. Protein Coding Sequences

The construct also includes the nucleic acid coding sequence for a heterologous protein, under the control of a promoter, preferably a seed-specific promoter. In accordance with the present invention, polynucleotide sequences which encode human milk proteins such as lysozyme or lactoferrin, include splice variants, fragments of such human milk proteins, fusion proteins, modified forms or functional equivalents thereof, collectively referred to herein as "human milk protein-encoding nucleic acid sequences".

Such "human milk protein-encoding nucleic acid sequences" may be used in recombinant expression vectors (also termed heterologous nucleic acid constructs), that direct the expression of a human milk protein in appropriate host cells.

Due to the inherent degeneracy of the genetic code, a number of nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence may be generated and used to clone and express a given human milk protein, as exemplified herein by the codon optimized coding sequences used to practice the invention (further described below). Thus, for a given human milk protein-encoding nucleic acid sequence, it is appreciated that as a result of the degeneracy of the genetic code, a number of coding sequences can be produced that encode the same human milk protein amino acid sequence. For example, the triplet CGT encodes the amino acid arginine. Arginine is alternatively encoded by CGA, CGC, CGG, AGA, and AGG. Therefore such substitutions in the coding region fall within the range of sequence variants covered by the present invention. Any and all of these sequence variants can be utilized in the same way as described herein for a "reference" human milk protein-encoding nucleic acid sequence.

A "variant" human milk protein-encoding nucleic acid sequence may encode a "variant" human milk protein amino acid sequence which is altered by one or more amino acids from the native milk protein sequence, both of which are included within the scope of the invention.. Similarly, the term "modified form of", relative to a given human milk protein, means a derivative or variant form of a native human milk protein or the coding sequence therefor. That is, a "modified form of" a human milk protein has a derivative sequence containing at least one nucleic acid or amino acid substitution, deletion or insertion. The nucleic acid or amino acid substitution, insertion or deletion may occur at any residue within the sequence, as long as the encoded amino acid sequence maintains the biological activity of the native human milk protein, *e.g.*, the bactericidal effect of lysozyme.

A "variant" human milk protein-encoding nucleic acid sequence may encode a "variant" human milk protein sequence which contains amino acid insertions or deletions, or both. Furthermore, a variant human milk protein coding sequence may encode the same polypeptide as the reference polynucleotide or native sequence but, due to the degeneracy of the genetic code, has a nucleic acid coding sequence which is altered by one or more bases from the reference or native polynucleotide sequence.

The variant nucleic acid coding sequence may encode a variant amino acid sequence which contains a "conservative" substitution, wherein the substituted amino acid has structural or chemical properties similar to the amino acid which it replaces and physicochemical amino acid side chain properties and high substitution frequencies in

homologous proteins found in nature (as determined, e.g., by a standard Dayhoff frequency exchange matrix or BLOSUM matrix). In addition, or alternatively, the variant nucleic acid coding sequence may encode a variant amino acid sequence which contains a "non-conservative" substitution, wherein the substituted amino acid has dissimilar  
5 structural or chemical properties to the amino acid which it replaces.

Standard substitution classes include six classes of amino acids based on common side chain properties and highest frequency of substitution in homologous proteins in nature, as is generally known to those of skill in the art and may be employed to develop variant human milk protein-encoding nucleic acid sequences. A "variant"  
10 human milk protein-encoding nucleic acid sequence may encode a "variant" human milk protein sequence which contains a combination of any two or three of amino acid insertions, deletions, or substitution.

Human milk protein-encoding nucleotide sequences also include "allelic variants" defined as an alternate form of a polynucleotide sequence which may have a  
15 substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

The polynucleotides for use in practicing the invention include sequences which encode human milk proteins and splice variants thereof, sequences complementary to the protein coding sequence, and novel fragments of the polynucleotide. The  
20 polynucleotides may be in the form of RNA or in the form of DNA, and include messenger RNA, synthetic RNA and DNA, cDNA, and genomic DNA. The DNA may be double-stranded or single-stranded, and if single-stranded may be the coding strand or the non-coding (anti-sense, complementary) strand.

As will be understood by those of skill in the art, in some cases it may be  
25 advantageous to use a human milk protein-encoding nucleotide sequences possessing non-naturally occurring codons. Codons preferred by a particular eukaryotic host (Murray *et al.*, 1989) can be selected, for example, to increase the rate of human milk protein expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring  
30 sequence. Codon-optimized sequences for use in practicing the invention are further described below.

A human milk protein-encoding nucleotide sequence may be engineered in order to alter the human milk protein coding sequence for a variety of reasons, including but

not limited to, alterations which modify the cloning, processing and/or expression of the human milk protein by a cell.

Heterologous nucleic acid constructs may include the coding sequence for a given human milk protein, a variant, fragment or splice variant thereof: (i) in isolation; (ii) 5 in combination with additional coding sequences; such as fusion protein or signal peptide, in which the human milk protein coding sequence is the dominant coding sequence; (iii) in combination with non-coding sequences, such as introns and control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host; and/or (iv) in a vector 10 or host environment in which the human milk protein coding sequence is a heterologous gene.

Depending upon the intended use, an expression construct may contain the nucleic acid sequence which encodes the entire human milk protein, or a portion thereof. For example, where human milk protein sequences are used in constructs for use as a 15 probe, it may be advantageous to prepare constructs containing only a particular portion of the human milk protein encoding sequence, for example a sequence which is discovered to encode a highly conserved human milk protein region.

In one general embodiment, a human lysozyme amino acid sequence encoded by a human lysozyme-encoding nucleic acid sequence in an expression vector used to 20 practice the invention has at least 70%, preferably 80%, 85%, 90% or 95% or more sequence identity to the human lysozyme amino acid sequence presented as SEQ ID NO:2.

In another general embodiment, a human lactoferrin amino acid sequence encoded by a human lactoferrin-encoding nucleic acid sequence in an expression vector 25 used to practice the invention has at least 70%, preferably 80%, 85%, 90% or 95% or more sequence identity to the human lactoferrin amino acid sequence presented as SEQ ID NO:4.

#### D. Codon Optimization

30 It has been shown that production of recombinant protein in transgenic barley grain was enhanced by codon optimization of the gene (Horvath *et al.*, 2000; Jensen *et al.*, 1996). The intent of codon optimization was to change an A or T at the third position of the codons of G or C. This arrangement conforms more closely with codon usage in typical rice genes (Huang *et al.*, 1990a).

In order to obtain a high expression level for human lysozyme in rice cells, the coding sequence was codon optimized. The G + C content was thus increased from 46% to 68%. The codon optimized lysozyme coding sequence for use in practicing the invention is presented as SEQ ID NO:1.

5 Similarly, in order to obtain high level expression level of human lactoferrin (hLF) in rice cells, the native hLF coding sequence was codon optimized. Out of 693 codons used in the lactoferrin gene, 413 codons were changed by one or two nucleotides. The amino acid sequence of LF was unchanged. The codon optimized lactoferrin coding sequence for use in practicing the invention is presented as SEQ ID NO:3.

10 Codon optimized sequences for other human milk proteins are given as follows: for lactoferricin, (SEQ ID NO: 7; for EGF, SEQ ID NO: 8; for IGF-1, SEQ ID NO: 9; for lactohedrin, SEQ ID NO: 10; for kappa-casein, SEQ ID NO: 11; for haptocorrin, SEQ ID NO: 12; for lactoperoxidase, SEQ ID NO: 13; and for alpha-1-antitrypsin, SEQ ID NO: 14.

15 E. Transcription Factor Coding Sequences

In one embodiment of the invention, the transgenic plant is also transformed with the coding sequence of one or more transcription factors capable of stimulating the expression of a maturation-specific promoter. Specifically, the embodiment involves the use of the maize *Opaque 2* (O2) and prolamin box binding factor (PBF) together with the  
20 rice endosperm bZip (Reb) protein as transcriptional activators of monocot storage protein genes. Exemplary sequence for these three transcription factors are given identified below as SEQ ID NOS: 31-33. Transcription factor sequences and constructs applicable to the present invention are detailed in co-owned PCT application No. PCT/US01/14234, International Publication number WO 01/83792 A1, published  
25 November 8, 2001, which is incorporated herein by reference.

Transcription factors are capable of sequence-specific interaction with a gene sequence or gene regulatory sequence. The interaction may be direct sequence-specific binding in that the transcription factor directly contacts the gene or gene regulatory sequence or indirect sequence-specific binding mediated by interaction of the  
30 transcription factor with other proteins. In some cases, the binding and/or effect of a transcription factor is influenced (in an additive, synergistic or inhibitory manner) by another transcription factor. The gene or gene regulatory region and transcription factor may be derived from the same type (e.g., species or genus) of plant or a different type of plant. The binding of a transcription factor to a gene sequence or gene regulatory

sequence may be evaluated by a number of assays routinely employed by those of skill in the art, for example, sequence-specific binding may be evaluated directly using a label or through gel shift analysis.

As detailed in the cited PCT application, the transcription factor gene is  
5 introduced into the plant in a chimeric gene containing a suitable promoter, preferably a maturation-specific seed promoter operably linked to the transcription factor gene. Plants may be stably transformed with a chimeric gene containing the transcription factor by methods similar to those described with respect to the milk-protein gene(s). Plants  
10 stably transformed with both exogenous transcription factor(s) and milk-protein genes may be prepared by co-transforming plant cells or tissue with both gene constructs, selecting plant cells or tissue that have been co-transformed, and regenerating the transformed cells or tissue into plants. Alternatively, different plants may be separately transformed with exogenous transcription factor genes and milk-protein genes, then crossed to produce plant hybrids containing by added genes.

15

#### F. Additional Expression Vector Components

Expression vectors or heterologous nucleic acid constructs designed for operation in plants, comprise companion sequences upstream and downstream to the expression cassette. The companion sequences are of plasmid or viral origin and provide necessary  
20 characteristics to the vector to permit the vector to move DNA from bacteria to the plant host, such as, sequences containing an origin of replication and a selectable marker. Typical secondary hosts include bacteria and yeast.

In one embodiment, the secondary host is *E. coli*, the origin of replication is a colE1-type, and the selectable marker is a gene encoding ampicillin resistance. Such  
25 sequences are well known in the art and are commercially available as well (e.g., Clontech, Palo Alto, Calif.; Stratagene, La Jolla, CA).

The transcription termination region may be taken from a gene where it is normally associated with the transcriptional initiation region or may be taken from a different gene. Exemplary transcriptional termination regions include the NOS terminator from  
30 *Agrobacterium* Ti plasmid and the rice  $\alpha$ -amylase terminator.

Polyadenylation tails (Alber *et al.*, 1982) may also be added to the expression cassette to optimize high levels of transcription and proper transcription termination, respectively. Polyadenylation sequences include, but are not limited to, the *Agrobacterium*

octopine synthetase signal, Gielen, *et al.*, 1984 or the nopaline synthase of the same species, Depicker, *et al.*, 1982.

Suitable selectable markers for selection in plant cells include, but are not limited to, antibiotic resistance genes, such as, kanamycin (*npfII*), G418, bleomycin, hygromycin, chloramphenicol, ampicillin, tetracycline, and the like. Additional selectable markers include a bar gene which codes for bialaphos resistance; a mutant EPSP synthase gene which encodes glyphosate resistance; a nitrilase gene which confers resistance to bromoxynil; a mutant acetolactate synthase gene (ALS) which confers imidazolinone or sulphonylurea resistance; and a methotrexate resistant DHFR gene.

The particular marker gene employed is one which allows for selection of transformed cells as compared to cells lacking the DNA which has been introduced. Preferably, the selectable marker gene is one which facilitates selection at the tissue culture stage, e.g., a kanamycin, hygromycin or ampicillin resistance gene.

The vectors of the present invention may also be modified to include intermediate plant transformation plasmids that contain a region of homology to an *Agrobacterium tumefaciens* vector, a T-DNA border region from *Agrobacterium tumefaciens*, and chimeric genes or expression cassettes (described above). Further, the vectors of the invention may comprise a disarmed plant tumor inducing plasmid of *Agrobacterium tumefaciens*.

In general, a selected nucleic acid sequence is inserted into an appropriate restriction endonuclease site or sites in the vector. Standard methods for cutting, ligating and *E. coli* transformation, known to those of skill in the art, are used in constructing vectors for use in the present invention. (See generally, Maniatis, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2d Edition (1989); Ausubel, *et al.*, (c) 1987, 1988, 1989, 1990, 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y; and Gelvin, S.B., *et al.*, eds. PLANT MOLECULAR BIOLOGY MANUAL, (1990), all three of which are expressly incorporated by reference, herein.

#### V. Generation of Transgenic Plants

Plant cells or tissues are transformed with expression constructs (heterologous nucleic acid constructs, e.g., plasmid DNA into which the gene of interest has been inserted) using a variety of standard techniques. Effective introduction of vectors in order to facilitate enhanced plant gene expression is an important aspect of the invention. It is preferred that the vector sequences be stably integrated into the host genome.

The method used for transformation of host plant cells is not critical to the present invention. The transformation of the plant is preferably permanent, *i.e.* by integration of the introduced expression constructs into the host plant genome, so that the introduced constructs are passed onto successive plant generations. The skilled artisan will  
5 recognize that a wide variety of transformation techniques exist in the art, and new techniques are continually becoming available.

Any technique that is suitable for the target host plant may be employed within the scope of the present invention. For example, the constructs can be introduced in a variety of forms including, but not limited to, as a strand of DNA, in a plasmid, or in an  
10 artificial chromosome. The introduction of the constructs into the target plant cells can be accomplished by a variety of techniques, including, but not limited to calcium-phosphate-DNA co-precipitation, electroporation, microinjection, *Agrobacterium*-mediated transformation, liposome-mediated transformation, protoplast fusion or microprojectile bombardment. The skilled artisan can refer to the literature for details and select suitable  
15 techniques for use in the methods of the present invention. Exemplary methods for plant transformation are given in Example 2.

When *Agrobacterium* is used for plant cell transformation, a vector is introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA  
20 for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming  
25 host plant cells, the expression or transcription construct bordered by the T-DNA border region(s) is inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, examples of which are described in the literature, for example pRK2 or derivatives thereof. See, for example, Ditta *et al.*, 1980 and EPA 0 120 515, expressly incorporated by reference herein. Alternatively, one may insert the sequences to be  
30 expressed in plant cells into a vector containing separate replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium*. See, for example, McBride *et al.*, 1990, wherein the pRiHRI (Jouanin, *et al.*, 1985, origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

Included with the expression construct and the T-DNA is one or more selectable marker coding sequences which allow for selection of transformed *Agrobacterium* and transformed plant cells. A number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, kanamycin, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, with a particular marker preferred depending on the particular host and the manner of construction.

For *Agrobacterium*-mediated transformation of plant cells, explants are incubated with *Agrobacterium* for a time sufficient to result in infection, the bacteria killed, and the plant cells cultured in an appropriate selection medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of the recombinant protein produced by the plants.

There are a number of possible ways to obtain plant cells containing more than one expression construct. In one approach, plant cells are co-transformed with a first and second construct by inclusion of both expression constructs in a single transformation vector or by using separate vectors, one of which expresses desired genes. The second construct can be introduced into a plant that has already been transformed with the first expression construct, or alternatively, transformed plants, one having the first construct and one having the second construct, can be crossed to bring the constructs together in the same plant.

#### A. Plants

Host cells of the present invention include plant cells, both monocotyledenous and dicotyledenous. In one preferred embodiment, the plants used in the methods of the present invention are derived from monocots, particularly the members of the taxonomic family known as the *Gramineae*. This includes all members of the grass family of which the edible varieties are known as cereals. The cereals include a wide variety of species such as wheat (*Triticum* sps.), rice (*Oryza* sps.) barley (*Hordeum* sps.) oats, (*Avena* sps.) rye (*Secale* sps.), corn (maize) (*Zea* sps.) and millet (*Pennisetum* sps.). In practicing the present invention, preferred grains are rice, wheat, maize, barley, rye, triticale. Also preferred are dicots exemplified by soybean (*Glycine* spp.)

In order to produce transgenic plants that express human milk protein, monocot plant cells or tissues derived from them are transformed with an expression vector comprising the coding sequence for a human milk protein. Transgenic plant cells obtained as a result of such transformation express the coding sequence for a human milk protein, such as lysozyme or lactoferrin. The transgenic plant cells are cultured in medium containing the appropriate selection agent to identify and select for plant cells which express the heterologous nucleic acid sequence. After plant cells that express the heterologous nucleic acid sequence are selected, whole plants are regenerated from the selected transgenic plant cells. Techniques for regenerating whole plants from transformed plant cells are generally known in the art. Transgenic plant lines, e.g., rice, wheat, corn or barely, can be developed and genetic crosses carried out using conventional plant breeding techniques.

Production of recombinant proteins in monocot seeds, e.g., rice (*Oryza sativa* L.) seeds has the advantages that (a) high level expression make it an economically practical strategy, and (b) rice is a normal part of the diet of infants and children, has good nutritional value and low allergenicity. Thus, the use of rice as the basis for a food supplement is unlikely to introduce any risk and thereby eliminates the need for a high degree of purification when included in infant formula.

In addition, rice is the staple food crop of more than half the world's population. Recent reports on the production of provitamin A (beta-Carotene) in rice seeds exemplifies the need for value added food crops especially in the developing world (Ye *et al.*, 2000) where rice is used as major food crop.

#### VI. Detecting Expression of Recombinant Human Milk Proteins

Transformed plant cells are screened for the ability to be cultured in selective media having a threshold concentration of a selective agent. Plant cells that grow on or in the selective media are typically transferred to a fresh supply of the same media and cultured again. The explants are then cultured under regeneration conditions to produce regenerated plant shoots. After shoots form, the shoots are transferred to a selective rooting medium to provide a complete plantlet. The plantlet may then be grown to provide seed, cuttings, or the like for propagating the transformed plants. The method provides for efficient transformation of plant cells with expression of a gene of autologous or heterologous origin and regeneration of transgenic plants, which can produce a recombinant human milk protein.

The expression of the recombinant human milk protein may be confirmed using standard analytical techniques such as Western blot, ELISA, PCR, HPLC, NMR, or mass spectroscopy, together with assays for a biological activity specific to the particular protein being expressed.

5           Example 3 describes the characterization of human lysozyme produced in the seeds of transgenic rice plants. Analyses used to confirm that recombinant lysozyme produced in transgenic rice is essentially the same as the native form of the protein both in physical characteristics and biological activity included, SDS-PAGE, reverse IEF gel electrophoresis, Western blot analysis, enzyme linked immunosorbant assay (ELISA),  
10   enzymatic activity assay and bactericidal activity assay using indicator strains, *Micrococcus luteus* and *E.coli* strain JM109.

          Example 4 describes the characterization of human lactoferrin produced in the seeds of transgenic rice plants. Analyses used to confirm that recombinant lactoferrin produced in transgenic rice is essentially the same as the native form of the protein both  
15   in physical characteristics and biological activity included, Southern blot, Western blot, ELISA, N-Terminal Amino Acid Sequencing, analysis of glycosylation and determination of sugar content, a determination of the isoelectric point, pH dependent iron release of rLF, bacteriostatic activity assay of rLF using enteropathogenic *E. coli* as the indicator strain.

20           Example 5 details the characterization of alpha-1-antitrypsin produced by transgenic monocot plant cells. Example 6 details the characterization of other milk proteins also produced by monocot plant transformed with the chimeric genes of the invention.

## 25   VII. Preparation of Seed Composition and Processed foods

          The invention provides, in one embodiment, a seed composition containing a flour, extract, or malt obtained from mature monocot seeds and one or more seed-produced human milk proteins in substantially unpurified form. Where the milk protein is expressed at a level of between about 0.1 to 1 percent of the seed weight, the  
30   composition will contain the same or preferably a higher percentage of milk protein, e.g., 0.1 to 20 % of the composition depending on the composition added. In particular, a grain composition will yield an amount of milk protein that is comparable to that in the mature seed; the extract composition, by contrast, in which most of the starch has been removed, will typically show a severalfold increase in percentage of milk protein, e.g., 10-

40% of the total weight of the extract. The malt composition will contain an intermediate level, typically greater than grain, but less than extract.

In determining the amount of grain, extract, or malt composition to be added to a food, it is useful to determine the amount of any milk protein present (see Section VI  
5 above), and add an amount of composition which brings the final level of milk protein to a desired level in the food. For example, in infant formula, it may be desired to have a final concentration of lysozyme between 0.03 and 0.3 grams/liter of formula, and an amount of lactoferrin between about 0.3 to 3 grams/liter formula. Thus, if a seed composition is found to contain 10g/kg lysozyme, about 10 grams of the composition would be added to  
10 make up a liter of formula with a final lysozyme concentration of about 0.1 g/liter. Below are described methods for preparing each of the three types of milk-protein-containing seed compositions.

#### A. Flour composition

15 The flour composition is prepared by milling mature monocot plant seeds, using standard milling and, optionally, flour purification methods, e.g., in preparing refined flour. Briefly, mature seeds are dehusked, and the dehusked seeds then ground into a fine flour by conventional milling equipment.

The flour may be added to foods during food processing according to standard  
20 food processing methods. Preferably, the processing temperature does not lead to denaturation of the milk proteins, e.g., above 60°-70°C. The flour may also be used directly, either in capsule, tabletized, or powder form, as a nutraceutical composition. One preferred flour composition contains lactoferrin and/or lysozyme. The flour may alternatively, or in addition, include one or more of the other human milk proteins such as  
25 epidermal growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, and alpha-1-antitrypsin.

Flour containing two or more milk proteins may be prepared by combining flour from seeds that separately produce the different proteins, for example, equal amounts of a flour containing lysozyme and a flour containing lactoferrin. Alternatively, a multi-  
30 protein composition can be prepared as seed flour from plants, such as monocot plants co-transformed with chimeric genes expressing different milk proteins, e.g., lactoferrin and lysozyme.

## B. Extract Composition

An extract composition is prepared by milling seeds to form a flour, extracting the flour with an aqueous buffered solution, and optionally, further treating the extract to partially concentrate the extract and/or remove unwanted components. Details of exemplary methods for producing the extract composition are given in Example 9. Briefly, mature monocot seeds, such as rice seeds, are milled to a flour, and the flour then suspended in saline or in a buffer, such as Phosphate Buffered Saline ("PBS"), ammonium bicarbonate buffer, ammonium acetate buffer or Tris buffer. A volatile buffer or salt, such as ammonium bicarbonate or ammonium acetate may obviate the need for a salt-removing step, and thus simplify the extract processing method.

The flour suspension is incubated with shaking for a period typically between 30 minutes and 4 hours, at a temperature between 20-55°C. The resulting homogenate is clarified either by filtration or centrifugation. The clarified filtrate or supernatant may be further processed, for example by ultrafiltration or dialysis or both to remove contaminants such as lipids, sugars and salt. Finally, the material may be dried, e.g., by lyophilization, to form a dry cake or powder. The extract combines advantages of high milk-protein yields, essentially limiting losses associated with protein purification. At the same time, the milk proteins are in a form readily usable and available upon ingestion of the extract or food containing the extract. One particular advantage for use in infant formula or infant foods is the low amount of seed starch present in the extract. In particular, the extract may increase the concentration of recombinant protein from about 0.5% of total soluble protein ("TSP") in conventional approaches to over about 25% of TSP in the extract approach. Some of the present extract approach even reached 40% of TSP depending on the expression level of the recombinant protein in the seeds. In addition, the extract approach removes starch granules, which require high gelling temperature, for example above about 75°C. Consequently, the extract approach provides more flexibility in processing the rice grain and the recombinant proteins into food and nutritional drinks, particularly infant food and formula, because of the difficulty infants have in digesting undenatured seed starch. Undenatured starch granule cannot be digested by human gut without initial gelatinization, by for example high temperature.

The extract can be used as a nutraceutical for direct use, e.g., in capsule, tabletized or powder form, or as food additive in food processing. In one embodiment, the extract is added to an infant milk formula, in an amount typically between 0.1 to 10 percent by dry weight, preferably 1-5% by dry weight of the total formula weight. One

preferred infant formula contains both lactoferrin and lysozyme, preferably in an amount between 50-200% of the amount of human lactoferrin or lysozyme, respectively, of that found in normal human milk. As noted above, lactoferrin is present in a concentration of about 1 gram /liter human milk, and lysozyme, about 0.1/liter human milk. The extract  
5 may alternatively, or in addition, include one or more of the other human milk proteins including epidermal growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin and immunoglobulins. Similarly, for use as an additive to solid baby food, or to nutritional drinks, the extract is added in amounts preferably between about 0.1 to 10% of the food/drink material by dry weight.

10 As above, extract containing two or more milk proteins may be prepared by combining extracts from seeds that separately produce the different proteins, or by processing seeds from plants co-transformed with chimeric genes expressing different milk proteins, e.g., lactoferrin and lysozyme.

15 C. Malt composition

One technical challenge to commercialization of engineered monocot grains expressing human milk proteins is to formulate the transgenic grains into edible products without loss of bioavailable milk protein in the final product.

In accordance with another embodiment, the invention provides a malt extract or  
20 malt syrup ("malt") in which seed starches have been largely reduced to malt sugars, and the milk protein(s) are in an active, bioavailable form. A wide range of food products and/or food additive may be produced by varying the types of malt used, the mashing program and the ways in which the wort is subsequently handled. If materials other than barley malt are used in the mash (such as starch from other grains), the resulting product  
25 is classified as a malt syrup. Malt extracts, which may have a syrupy consistency or may be powders, are made by mashing ground malt, usually barley malt, in conventional brewery equipment, collecting the wort and concentrating it or drying it. Modern production of food malt extracts and malt syrups has evolved into three basic grain stages: steeping, germination, and drying of the germinated seed, followed by three more  
30 steps involving liquefaction of the germinated grain, mashing of the germinated grain, lautering (filtering), and evaporation. Many variations of malt extracts or syrups are possible. Flavor, color, solids, enzymatic activity, and protein are the basic characteristics that can be adjusted during production to provide malts specific for given

food applications. (See, generally, Eley; Hickenbottom, 1996, 1997a, 1997b, 1983; Lake; Moore; Moe; Sfat; Doncheck; Briggs, 1981, 1998; and Hough).

#### C1. Steeping

5 After the barley of choice has been cleaned of foreign material, it is graded to size and transferred to steep tanks equipped with water inlet and outlet pipes. Compressed air is fed from the tank bottom for vigorous aeration and mixing for the barley/water mixture. When the barley has reached a water content of 43-45%, steeping is stopped.

#### 10 C2. Germination

The steeped barley is moved to germination floors or rooms depending on the particular malt house's capabilities and allowed to germinate under controlled temperature, air, and moisture conditions. Total germination varies from four to seven days, depending on the barley type, density end use of the malt, and the controls or  
15 germination method used. All aspects of germination must be kept in constant balance to ensure proper kernel modification and yield.

Many enzymatic systems are activated during germination. Two of the systems are the oxidative and reductive systems involved with the respiration phase. Other enzymes break down the endosperm cell structure, which in itself is a measure of germination rate  
20 when the pentose production is evaluated. The proteolytic enzymes release or active beta-amylase and also work on the proteins present to render them soluble. In fact, about 40% of the total protein is made soluble in water. Optimum germination activates a balanced enzyme system, which hydrolyzes the starch present.

#### 25 C3. Kilning

Drying or kilning, when done at the proper time and optimum degree of starch modification, stops the germination. The heat also catalyzes additional reactions, notably flavor and color development. The heating step is carried out according to well known kilning conditions. When drying is complete, the sprouts and other extraneous materials  
30 are removed, and the kernels are then ready for further processing.

#### C4. Malt Extracts and Syrups

The malted barley (kernel) is coarsely ground in crushers and fed into mash tuns where it is mixed with water. During a series of time and temperature changes, some of

the starch is converted into fermentable sugars by action of the natural alpha- and beta-amylases, better known as the diastatic system. If cereal adjuncts are to be added, which result in malt syrups with mellow and sweeter flavors than the extracts, they are added at this stage usually derived from the cereal grains, corn and rice, although barley, 5 wheat, rye, millet and sorghum are sometimes used, derived from mature seeds that produce the desired recombinant milk proteins.

Once the mash batch has achieved the correct degree of hydrolysis, it is transferred to lauter tuns. The lauter tun has a slotted or false bottom a few inches above the real bottom to allow for filtration and is also equipped with some means of agitation. During 10 this extraction stage, the amylolytic enzymes liquefy additional insoluble starches, converting them to maltose and dextrans. At the same time, the proteolytic enzymes attach certain proteins converting them into simpler, soluble forms. After the appropriate conditions have been met, the liquid phase, or wort, is drawn from the lauter tuns into evaporators.

15 Evaporation of the wort is conducted under vacuum where it is converted into a syrup of about 80% solids. Depending on the temperatures used, malt extracts or syrups of high, medium, or zero enzymatic activity can be produced. Color and flavor also can be controlled during this stage. The finishing steps of filtering, cooling, and packaging complete the malt extract/syrup process.

20

#### C5. Transgenic Malt Extract

For a transgenic malt extract, the starting barley is a transgenic barley engineered to produce one or more human milk proteins in the endosperm either in grain maturation or in the malting process, or at both times. Malting and processing times and conditions 25 are adjusted so that the bioactivity of the target recombinant molecules is preserved and the bioavailability of the target recombinant molecule is maximized. The resulting malt extract is either consumed directly as a concentrate that is either consumed directly as a food, or is incorporated as an ingredient in a food mixture. Studies conducted in support of the present invention demonstrate that recombinant proteins retain activity after 30 malting for up to at least 288 hrs.

#### C6. Transgenic Malt Syrup

For a transgenic malt syrup, the starting barley can be a non-transgenic barley, or a transgenic barley, or a mixture of both. The barley is processed as described, except

that during the mashing process, a cereal adjunct is added in a form that it is converted during the mashing process with the concurrent retention and generation of bioavailability and bioactivity of the target recombinant molecule found within the transgenic cereal adjunct. The use of a transgenic cereal adjunct enables the production in the malt syrup  
5 of the target recombinant molecule expressed in the transgenic grain endosperm.

The malt extract or syrup may be used directly as a syrup, or added to processed foods or drinks, according to standard food processing procedures that employ grain extracts or syrups, *e.g.*, for sweetening. One preferred food is an infant formula containing between 0.1 to 10% malt (extract or syrup). The malt is also useful as a  
10 sweetener/nutritional additive in baby and adult foods, and nutritional drinks.

Preferred malt extracts or syrups contain lactoferrin and/or lysozyme. The malt may alternatively, or in addition, include one or more of the human milk proteins such as epidermal growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin and immunoglobulins. As above, malt  
15 containing two or more milk proteins may be prepared by combining or preparing malts from seeds that separately produce the different proteins, or by preparing a malt from the seeds of plants co-transformed with chimeric genes expressing different milk proteins, *e.g.*, lactoferrin and lysozyme.

From the foregoing, it can be appreciated how various objects and features of the  
20 invention are met. The production of high levels of human milk proteins in grains, exemplified herein by rice provides the distinct advantage that food supplements may be prepared with little or no purification. In a preferred approach, the human milk protein containing transgenic grain is ground (*e.g.*, into flour) and directly added to a food such as infant formula, without additional processing. Since the recombinant grain finds utility  
25 as a food or food supplement, the regulatory requirements for purity are not stringent.

Transgenic seeds are ideal bioreactors, combining low production costs and low or minimal downstream processing costs prior to use. Seed grain proteins can accumulate to 9-19% of grain weight (Lásztitym 1996); the endosperm proteins are synthesized during grain maturation and stored in protein bodies for use in the  
30 germination and seedling growth of the next plant generation; grains can be stored for years without loss of functionality, and therefore the downstream processing can be conducted independently of growing seasons.

The human milk protein-containing transgenic grains of the invention may be used directly as food, *e.g.*, rice, corn, wheat, barley, soybeans, etc. Alternatively, food

supplements are prepared from the human milk protein-containing transgenic grain. The results presented herein demonstrate that human milk proteins may be expressed at high levels in the seeds of transgenic plants, e.g., up to 0.25 to 1% of total seed dry weight. The production of high levels of human milk proteins in grains, exemplified herein by rice, provides the distinct advantage that food supplements may be prepared with little or no purification. In a preferred approach, the human milk protein containing transgenic grain is ground (e.g., into flour) and directly added to a food, or in the form of an extract or malt, such as for preparing a nutritionally enhanced infant formula, without additional processing. Since the recombinant grain finds utility as a food or food supplement, as a flour, extract or malt, the regulatory requirements for purity are not stringent. Accordingly, human milk protein-containing transgenic grains are ideal bioreactors, combining low production costs and low or minimal downstream processing costs prior to use.

The human milk protein-containing transgenic grains of the invention may be used directly as food, e.g., rice, corn, wheat, barley, soybeans, etc. Alternatively, food supplements are prepared from the human milk protein-containing transgenic grain. Where the transgenic seed is rice, the invention provides additional advantages in that: rice is consumed by a majority of the population in the world and is being generally regarded as safe for human consumption. Rice-based foods are considered hypoallergenic (NIH publication, 1984). In many countries, rice is the first solid food for infants and rice-based infant formulas are commercially available (Bhan *et al.*, 1988; Gastañaduy *et al.*, 1990). These make rice attractive as a "protein factory" to produce biomedical and nutraceuticals for human consumption. The cloning and expression of human proteins, for example, human milk proteins lysozyme and lactoferrin in rice grains has opened a new avenue for the bioproduction of other milk proteins.

All publications, patents and patent applications are herein expressly incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

The following examples illustrate but are not intended in any way to limit the invention.

## EXAMPLE 1

### Expression Vectors for Generation of Transgenic Plants

In general, expression vectors were constructed using standard molecular biological techniques as described in Ausubel *et al.*, 1987. The vectors contain a

- 5 heterologous protein coding sequence for lactoferrin or lysozyme under the control of a rice tissue-specific promoter, as further described below.

#### A. An Expression Vector For Human Lysozyme Expression In Transgenic Rice Cells

- 10 The synthesized lysozyme gene was cloned into an API base vector pAPI137 by conventional molecular cloning techniques (Sambrook *et al.*, 1989). Plasmid pAPI137 contains the *RAmy3D* promoter (Huang *et al.*, 1993), the codons for the *RAmy3D* signal peptide and the *RAmy3D* terminator. The *RAmy3D* promoter, isolated from the rice amylase gene family, is activated in rice calli by sugar starvation (Huang *et al.*, 1993).

- 15 The human lysozyme gene was placed between the sequences of the *RAmy3D* signal peptide and the *RAmy3D* terminator to give plasmid pAPI156 having a size of 4829 bp.

The promoter of the rice Glutelin 1 gene (*Gt-1*) and the nucleotide sequence of the signal peptide were cloned with two primers based on the published *Gt1* gene sequence (Okita *et al.* J Biol Chem 264: 12573-12581, 1989). The forward primer with

- 20 *HindIII* site was named MV-*Gt-1*-F1; 5'-

ATCGAAGCTTCATGAGTAATGTGTGAGCATTATGGGACCACG-3' (SEQ ID NO:5).

The reverse primer was named Xba-*Gt-1*-R1; 5'-

CTAGTCTAGACTCGAGCCACGGCCATGGGGCCGGCTAGGGAGCCATCGCACAAGAGGAA-3' (SEQ ID NO:6). Genomic DNA was isolated from leaves of rice variety M202

- 25 (Dellaporta *et al.*, 1983). The PCR product amplified from the genomic DNA was cloned into pCR 2.1 (Invitrogen, Carlsbad, CA). The resulting plasmid was named pCR*Gt-1* or pAPI134.

- To generate a *Gt-1* expression plasmid, pAPI134 was digested with *HindIII* and *XbaI*. The fragment containing the *Gt-1* promoter and *Gt-1* signal peptide was cloned  
30 into a pUC19 based plasmid containing the nopaline synthase 3' (*nos*) terminator. The resulting plasmid was named pAPI141 and contains the rice *Gt-1* promoter, the *Gt-1* signal peptide, a multiple cloning site and the *nos* terminator.

The synthesized human lysozyme gene "lys-ger" (by Operon Technologies, Inc., Alameda, CA) that was optimized based on the rice gene codon usage was digested with

*DraI* and *XhoI* and cloned into pAPI141 digested with *NaeI* and *XhoI* according to standard cloning techniques (Sambrook *et al.*, 1989). The resulting plasmid was called pAPI159 (Fig. 1) having a size of 4131bp.

5           B. An Expression Vector For Human Lactoferrin Expression In Transgenic Rice

The hLF gene (Rey, MW, 1990) was codon optimized and synthesized by Operon Technologies (CA, USA). The plasmid containing the codon-optimized gene was called Lac-ger. Lac-ger was digested with *SmaI/XhoI* and the fragment containing the lactoferrin gene was cloned into pAPI141 which was partially digested with *NaeI* and  
10 completely digested with *XhoI*. The resulting plasmid was named pAPI164. For expression of hLF in rice seeds, the codon optimized gene was operably linked to the rice endosperm specific glutelin (Gt1) promoter and NOS terminator (Fig. 7).

## EXAMPLE 2

15           Generation Of Transgenic Plant Cells Expressing Human Milk Proteins

The procedure of microprojectile-mediated rice transformation (US Patent 6,284,956) was followed. Calli was raised from TP309 mature rice seeds, with calli two to four mm in diameter selected and placed on N6 media supplemented with 0.3 M mannitol and 0.3 M sorbitol for 20 hours before bombardment. Biolistic bombardment  
20 was carried out with the biolistic PDC-1000/He system (Bio-Rad, USA). Plasmid carrying milk protein genes and pAPI76, a plasmid carrying hygromycin selectable marker gene were gold-coated and co-bombarded at a ratio of 6:1 with a helium pressure of 1100psi. Two day old bombarded calli were then transferred to N6 selection media supplemented with 20 mg/l hygromycin B and allowed to grow in the dark at 26°C for 45 days.

25           In order to develop transgenic rice plants, the selected calli were transferred to pre-regeneration and regeneration media. When regenerated plants became 1-3 cm in height, the plantlets were transferred to rooting media which consisted of half concentration of MS and 0.05 mg/l NAA. After two weeks, plantlets with developed roots and shoots were transferred to soil and kept under the cover of plastic container for a  
30 week. The plants were allowed to grow about 12 cm tall and shifted to the green house where they were grown up to maturity.

#### A. Generation of Human Lysozyme Expressing Transgenic Rice Cells and Plants

The synthetic human lysozyme (hLys) gene under the control of the *RAmy3D* promoter and terminator in the pAPI156 plasmid (example 1A) was used to generate sixty independent transformants by particle bombardment-mediated transformation.

5 Particle bombardment mediated transformation of rice was carried out as described above. Briefly, rice calli derived from TP309 were bombarded with gold particles coated with plasmids pAPI156 and pAPI76 in a ratio of 6:1 using the helium biolistic particle delivery system, PDS 1000 (Bio-Rad, CA). Transformed calli were selected in the presence of hygromycin B (35 mg/L) on N6 (Sigma, MO).

10 Selected cell lines were maintained in culture media with 3% sucrose (Huang *et al.*, 1993). Lysozyme expression was induced by sugar starvation. Briefly, AA medium (containing 3% sucrose) was removed by aspiration, followed by washing the cells three times with AA minus sucrose (AA-S). The cells were then incubated with AA-S at 40% (v/v) density for three and a half days to obtain the optimal level of lysozyme expression.

15 Transformants expressing lysozyme were identified by immunoblot analysis, turbidimetric rate determination with *Micrococcus lysodeikticus* or ELISA. Calli were ranked according to the expressed lysozyme level. Suspension cell cultures from the top lines were established following the procedure described previously (Huang *et al.*, 1993). The amount of total protein (Bradford assay) and lysozyme (ELISA) was evaluated in  
20 selected calli (Table 1).

Table 1. Expression Level Of Human Milk Lysozyme In Transformed Calli

| Cell line | Calli (g) | Total protein (µg) | Lysozyme (µg) | Lysozyme/protein (%) |
|-----------|-----------|--------------------|---------------|----------------------|
| 156-1     | 0.39      | 2626.5             | 65.7          | 2.5                  |
| 156-5     | 0.38      | 5510               | 68.9          | 1.25                 |
| 156-16    | 0.4       | 4815               | 120.4         | 2.5                  |
| 156-19    | 0.44      | 2440               | 30.5          | 1.25                 |
| 156-28    | 0.49      | 4910               | 24.6          | 0.5                  |
| 156-43    | 0.56      | 8150               | 101.9         | 1.25                 |
| 156-47    | 0.37      | 2472               | 6.2           | 0.25                 |

25 The synthetic human lysozyme (hLys) gene under the control of the *Gt1* promoter and Nos terminator in the pAPI159 plasmid (Fig. 1) was used to generate independent

transformants by particle bombardment-mediated transformation. Transformed calli were selected as described above, then transferred to pre-regeneration and regeneration media. When regenerated plants became 1-3 cm in height, the plantlets were transferred to rooting media which consisted of half concentration of MS and 0.05 mg/l NAA. After two weeks, plantlets with developed roots and shoots were transferred to soil and kept under the cover of plastic container for a week. The plants were allowed to grow about 12 cm tall and shifted to the green house where they were grown up to maturity (R<sub>0</sub> plants).

Screening for R<sub>0</sub> plants expressing human lysozyme. Individual rice endosperms or grains were ground with cold phosphate buffered-saline (PBS) with the addition of 0.35 M NaCl. Grinding was conducted with a pre-cooled mortar and pestle at 1 ml buffer/grain. Clear grain homogenate was obtained by subjecting the resulting grain extract to centrifugation at 14,000 rpm for 10 min at 40°C.

Embryos from individual R<sub>1</sub> seed (derived from R<sub>0</sub> plants) that showed a level of lysozyme expression that was greater than 10 µg/seed were saved and used to generate R<sub>1</sub> plants. Briefly, seeds were dissected into embryo and endosperm portions. The endosperm was ground and assayed for lysozyme expression (as further described below). Embryos were sterilized in 50% commercial bleach for 25 minutes and washed with sterile H<sub>2</sub>O three times for 5 minutes each. Sterilized embryos were placed in a tissue culture tube that contained MS solid medium. Embryos germinated and plantlets having about three inches shoots and healthy root systems were obtained in two weeks. The plantlets were then transferred to pots to obtain mature plants (R<sub>1</sub>).

A total of 197 embryos from 12 selected R<sub>0</sub> plants were germinated and 157 R<sub>1</sub> seedlings planted in the greenhouse for generation of R<sub>2</sub> grains. Individual R<sub>2</sub> grains (n=1502) from 109 R<sub>1</sub> fertile plants were screened for lysozyme expression by lysozyme activity assay in order to identify 42 homozygous plants.

Homozygous R<sub>1</sub> plants were identified by analyzing positive expressions of recombinant human lysozyme (rHlys) from a minimum of 20 individual R<sub>2</sub> grains. Homozygous lines derived from these plants were planted in a rice field in California. During growth, agronomic characteristics of both transgenic and non-transgenic plants, such as plant height, percentage of fertility, number of effective tillers, filled grains/panicle, non-filled grains/plant, time to maturity and 1000 grain weight were determined and compared. Plants with satisfactory agronomic traits were selected and rHlys expression levels were determined by lysozyme activity assay. Plants that met the

criteria for satisfactory agronomic traits and had more than 35 µg of rHlys/grain were advanced to next generations.

SDS-PAGE, electroblotting and Western blot analysis were carried out with 18% precast gel (Invitrogen, Carlsbad, CA) as described in Example 3. The primary rabbit  
5 polyclonal antibody against human lysozyme was purchased from Dako A/S (Denmark) and used at 1:5000. Lysozyme was quantified by a turbidimetric activity assay with *Micrococcus luteus* (Sigma) on 96-well microtiter plate as described in Example 3. Briefly, 250 µl of 0.015% *M. luteus* cell suspension was incubated with 10 µl of samples containing lysozyme with a concentration less than 2.4 µg/ml. The reaction was followed  
10 by the kinetic mode in Microplate Manager (Bio-Rad, CA) for 5 min at 450 nm. The concentration of lysozyme was then determined in reference to the standard curve.

The stable expression level of human lysozyme (rHlys) reached at least about 0.6% rHlys per brown rice weight amounting to 45% of the total soluble protein extract from rice grain. Figure 2 illustrates the seed specific expression of human lysozyme in  
15 transgenic plants. rHlys is only found in mature and germinated grain, but not in any other tissues tested. Fig. 6 shows the expression level of human lysozyme in powdered R3 seeds taken from transgenic rice plants

#### 20 B. Generation of Human Lysozyme-Expressing Transgenic Wheat Cells and Plants

Plasmid API159 (Figure 1) and API230 (Figure 35) were used to transform wheat cells substantially in the same manner as in transforming rice cells. Eight transgenic wheat lines were produced with API159, generating an expression level of about 150 to 300 µg of lysozyme per grain. Two transgenic wheat lines were produced with API230,  
25 yielding an expression level of about 50 to 120 µg of lysozyme per grain.

#### C. Generation of Human Lysozyme-Expressing Transgenic Barley Cells and Plants

The plasmid API159 was also used to transform barley cells substantially as  
30 described as transformation of rice cells. Five transgenic barley lines were produced, yielding about 3.9 to 12.3 µg of lysozyme per grain.

D. Generation of Human Lactoferrin Expressing Transgenic Rice Cells and Plants

The synthetic human lactoferrin gene under the control of the Gt1 promoter in the pAPI164 plasmid was used to generate over 100 independent transformants by particle bombardment-mediated transformation.

Particle bombardment mediated transformation of rice was carried out as described above. At least 20 R1 grains from each R0 plant were analyzed for rHLF expression. Individual R1 grains were cut into halves. The endospermic half was subjected to rHLF expression analysis by Western blot or ELISA and the corresponding positive embryonic half was germinated to generate R1 seedlings. The seedlings were transplanted to generate R2 grains. During the screening of R1 grains we observed that all the positive grains were opaque-pinkish in color in comparison to negative or control grains. The opaque-pinkish color in rice grains was then used to identify homozygous lines. A transgenic plant was considered to be homozygous and expressing rHLF if all grains from that plant were opaque-pinkish. Homozygous lines were then confirmed by ELISA analysis. Based on the expression analysis and agronomic characters, selected homozygous R2 lines were advanced to R3 and R6 generations.

**EXAMPLE 3**

20 Characterization of Recombinant Human Lysozyme (rLys) Produced by Transgenic Rice Cells and Plants

A. Southern blot analysis

About three grams of young leaves were collected and grounded with liquid nitrogen into a fine powder. The genomic DNA was isolated according to the procedure as described in Dellaporta *et al.*, 1983, and purified by phenol-chloroform extraction. Approximately 5 µg of DNA was then with *Hind*III and *Eco*RI, separated on a 1% agarose gel, blotted onto a Hybond<sup>+</sup> membrane (Amersham Pharmacia Biotech, Piscataway, NJ). The blot was probed with gel purified human Hlys gene and developed by ECL<sup>TM</sup> direct nucleic acid labeling and detection system (Amersham Pharmacia). By comparing to known amounts of the intact 1470 bp human lysozyme (Hlys) gene, the intact copy number of the transgenes, including promoter and Hlys gene, was estimated to vary from about 1 to about 6. No positive correlation between copy number of the rHlys transgene and amount of rHlys synthesized was discernible.

### B. SDS-PAGE and reverse IEF gel electrophoresis

Induced calli or harvested cells from suspension cell cultures were ground with cold phosphate buffered-saline (PBS) with a protease inhibitor cocktail (2 µg/ml aprotinin, 0.5 µg/ml leupeptin, 1 mM EDTA and 2 mM Pefabloc). The protease inhibitor cocktail was excluded from the buffer used subsequently during the purification of the enzyme, since the inhibitors did not increase the lysozyme expression yield. Grinding was conducted with a pre-chilled mortar and pestle at approximately 2 ml buffer/g calli or cells. A clear homogenate was obtained by subjecting the resulting extract to centrifugation at 16,000 x g for 10 minutes at 4 C.

SDS-PAGE was carried out using an 18% precast gel (Novex, CA). The resulting gel was stained with 0.1% Coomassie brilliant blue R-250 at 45% methanol and 10% glacial acetic acid for three hours. Gel destaining was conducted with 45% methanol and 10% glacial acetic acid until the desired background was reached.

Reverse IEF gel electrophoresis was carried out using a precast Novex pH 3-10 IEF gel according to the manufacturer's instructions (Novex, CA). About 30 µg of lysozyme was loaded onto the gel and electrophoresed at 100 V for 50 minutes followed by application of 200 V for 20 minutes. The gel was then fixed in 136 mM sulphosalicylic acid and 11.5% TCA for 30 minutes and stained in 0.1% Coomassie brilliant blue R-250, 40% ethanol, 10% glacial acetic acid for 30 minutes. The destaining solution contained 25% ethanol and 8% acetic acid.

### C. Western blot analysis

A SDS-PAGE gel was electroblotted to a 0.45 µm nitrocellulose membrane using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, CA) and subsequently subjected to immuno-blotting analysis. The blot was blocked with 5% non-fat dry milk in PBS, pH 7.4 for at least two hours followed by three washes with PBS, pH 7.4 for 10 minutes each. The primary rabbit polyclonal antibody against human lysozyme (Dako A/S, Denmark) was diluted at 1:2000 in the blocking buffer and the blot was incubated in the solution for at least one hour. The blot was then washed with PBS three times for 10 minutes each. The secondary goat anti-rabbit IgG (H+L)-alkaline phosphatase conjugate (Bio-Rad, CA) was diluted in the blocking buffer at 1:4000. The membrane was then incubated in the secondary antibody solution for one hour and then washed three times. Color development was initiated by adding the substrate system BCIP-NBT (Sigma) and

the process was stopped by rinsing the blot with H<sub>2</sub>O once the desirable intensity of the bands had been achieved.

#### D. Enzyme Linked Immunosorbant Assay (ELISA)

- 5 An indirect sandwich ELISA was developed to quantify total lysozyme expressed in rice calli or cells and used as an alternative assay to determine the lysozyme expression yield. A direct sandwich ELISA for lysozyme quantification has been previously reported (Lollike *et al.*, 1995, Taylor, 1992), however an alternate assay was developed as a key reagent used in the assay is no longer commercially available.
- 10 In carrying out the assay, rabbit anti-human lysozyme antibody (Dako D/K, Denmark) was used to coat a 96 well plate at 1:5000 diluted in PBS overnight at room temperature. After washing with PBS, the plate was blocked with 5% normal donkey serum (Jackson ImmunoResearch Laboratories, PA) in PBS for one hour. The plate was washed again with PBS. Lysozyme samples were diluted in 0.05% Tween in PBS and
- 15 captured by adding to the plate and incubating for one hour. After washing the plate with PBS, sheep anti-human lysozyme at 1:1000 diluted with 0.05% Tween in PBS was added and incubated for one hour. The plate was washed again with PBS. Peroxidase-conjugated affinipure donkey anti-sheep IgG (H+L) diluted in 0.05% Tween in PBS at 1:10,000 was added and incubated for one hour. After a final wash of the plate with
- 20 PBS, color was developed by incubating the plate with TMB substrate (Sigma, MO) for 5-15 minutes and the absorbance read at 655 nm.

#### E. Enzymatic Activity Assay For Lysozyme

- 25 A reliable and quantitative method was developed to analyze the expression level of enzymatically active lysozyme. The turbidimetric assay was developed using a 96-well microtiter plate format and based on the standard lysozyme assay that is carried out spectrophotometrically in cuvettes. A microtiter plate based method previously described for the detection of lysozyme release from human neutrophils had a detection range of 1-100 ng/ml (Moreira-Ludewig *et al.*, 1992). The assay conditions were modified to
- 30 maintain the linearity of detection up to 3.0 µg/ml.

The enzymatic activity of lysozyme was routinely determined by spectrophotometric monitoring of the decrease in turbidity at 450 nm of a suspension of *Micrococcus luteus* (*M. lysodeikticus*) cells (Shugar, 1952). Specifically, 250 µl of a 0.015% (w/v) *Micrococcus luteus* cell suspension was prepared in 66 mM potassium

phosphate, pH 6.24 (buffer A). Cell suspensions were equilibrated at room temperature and the reaction was initiated by adding 10  $\mu$ l samples containing lysozyme with concentrations from 0 to 2.4  $\mu$ g/ml. Lysozyme activity was determined in a kinetic mode for 5 minutes at 450 nm. The concentration of lysozyme was then calculated by

5 reference to the standard curve constructed with human milk-derived lysozyme.

The enzymatic activity of human milk lysozyme and the rice cell derived lysozyme of the invention was compared. As shown in Figure 4, the lysozyme effected reduction of the turbidity of *Micrococcus luteus* cell suspensions at 450 nm was very similar for lysozyme from the two sources, while buffer alone did not have any effect on the

10 reduction of turbidity.

Three selected suspension cell culture lines were induced to express lysozyme and the yield estimated in parallel by ELISA and the enzymatic activity assay described above (Table 2). T-test analysis showed that there was no significant difference between the lysozyme concentration measured by ELISA and enzymatic activity assay ( $p < 0.05$ ).

15 These results demonstrate that active recombinant human milk lysozyme is synthesized and maintained in rice callus cells and can be isolated without losing its activity.

Table 2. Comparison of Lysozyme Yields Estimates by Enzymatic Activity Assay and ELISA

20

| Cell line | Lysozyme yield by enzymatic activity assay (lysozyme/total protein $\mu$ g/mg) | Lysozyme yield by ELISA (lysozyme/total protein $\mu$ g/mg) |
|-----------|--------------------------------------------------------------------------------|-------------------------------------------------------------|
| 156-5     | 25.8 +/- 6.3                                                                   | 30.3 +/- 3.9                                                |
| 156-16    | 32.1 +/- 5.7                                                                   | 32.9 +/- 3.2                                                |
| 156-31    | 47.0 +/- 6.2                                                                   | 42.3 +/- 7.0                                                |

#### F. Recombinant Human Lysozyme Has Bactericidal Function

The sensitive lysis of *Micrococcus luteus* cells in a turbidimetric assay (Fig. 4) indicates that recombinant human lysozyme possesses enzymatic activity and functions

25 as a bactericide. To confirm this with a gram-negative bacterium, a bactericidal assay was carried out using an *E.coli* strain (JM109) as a test organism (Fig. 3).

In carrying out the assay, an aliquot of overnight JM109 culture was grown in LB medium until mid log phase. A standard inoculum of mid-log phase JM109 at  $2 \times 10^5$  CFU (colony forming units)/ml was used in the bactericidal assay. Buffer (20 mM Sodium

phosphate, pH 7.0, 0.5 mM EDTA) alone, buffer containing human milk lysozyme or rice seed derived lysozyme at about 30 µg/ml were sterilized by filtration. The mixture of cells and lysozyme solution was then incubated at 37° C for the specified length of time. One-fifth of the mixture volume was plated onto the LB agar plates and incubated overnight at 37 C in order to determine the number of colony forming units. At the concentration of 30 µg/ml, recombinant human lysozyme exhibited a similar bactericidal effect as lysozyme from human milk. There was no reduction of colony forming units using an extract from the non-transgenic control.

10           G. Purification of Lysozyme From Rice Calli, Suspension Cultures And Transgenic Rice Grains

Five rice calli lines expressing high levels of lysozyme were propagated and induced by sucrose starvation. The calli or cells were ground by a Tissuemizer in extraction buffer (PBS, 0.35 M NaCl) at 2 ml buffer/g of wet calli. The resulting tissue homogenate was centrifuged at 25,000 x g for 30 minutes at 4 C. The supernatant was removed and subjected to filtration through a pre-filter and then through a 0.45 µm nitrocellulose filter.

Approximately 1 liter of filtered supernatant from 500 grams of induced wet calli were then dialyzed against 50 mM sodium phosphate, pH 8.5 at 4 C overnight. The supernatant was loaded onto a 200 ml SP Sepharose fast flow column (XK26/40, Pharmacia) equilibrated with the loading buffer (50 mM sodium phosphate, pH 8.5) at a flow rate of four ml/min. The column was then washed with the same buffer until a baseline of A280 was achieved. Lysozyme was eluted by 0.2 M NaCl in the loading buffer and fractions containing lysozyme activity were pooled, concentrated and reapplied to a Sephacryl-100 column equilibrated and run with PBS at a flow rate of one ml/min. Proteins were eluted and separated by using PBS at a flow rate of one ml/min. Pure lysozyme fractions were identified by activity assay and total protein assay (Bradford) and the purity of lysozyme was confirmed by SDS-PAGE.

The five lines with the highest lysozyme expression level were selected and propagated continuously in petri dishes or shake flasks for lysozyme isolation and purification. A crude extract from rice callus contains both recombinant human lysozyme and large amounts of native rice proteins. Since the calculated pI of lysozyme is approximately 11, a strong cation exchange column, SP-Sepharose fast flow (Pharmacia), was chosen as the first column to separate the rice proteins from

recombinant human lysozyme. Most of the rice proteins did not bind to the column when equilibrated with 50 mM sodium phosphate, pH 8.5. The recombinant human lysozyme, on the other hand, bound to the column and was eluted by 0.2 M NaCl. Rice proteins that co-eluted with recombinant human lysozyme, were separated from lysozyme by gel  
5 filtration through a Sephacryl S-100 column and highly purified recombinant human lysozyme was obtained.

To purify human lysozyme from rice grains, R<sub>2</sub> rice seeds from transgenic plants were dehusked and milled to flour using conventional methods. Lysozyme was extracted by mixing the rice flour with 0.35 N NaCl in PBS at 100 grams/liter at room temperature  
10 for one hour. The resulting mixture was subjected to filtration through 3 µm of a pleated capsule, then through 1.2 µm of a serum capsule and finally through a Suporcap 50 capsule with a 0.8 µm glass filter on top of 0.45 µm filter (Pall, MI).

The clear rice extract (1 liter) was then dialyzed against 50 mm sodium phosphate, pH 8.5 at 4°C overnight and the dialyzed sample was loaded onto a cation  
15 exchange resin SP-Sepharose (Pharmacia Amersham), which was pre-conditioned with 50 mm sodium phosphate, pH 8.5 before loading. After loading, the column was washed with the same buffer until a base line A280 reading was achieved, then lysozyme was eluted with 0.2 N NaCl in 50 mm sodium phosphate, pH 8.5. Fractions containing lysozyme were pooled and reappplied to a Sephacryl S-100 column (Bio-Rad; equilibrated  
20 and run with PBS). Pure lysozyme was fractions were identified by enzymatic assay and total protein assay (Bradford). Finally the purity of lysozyme was confirmed by SDS-PAGE.

#### H. Attributes Of Recombinant Human Lysozyme Produced In Rice

##### 25 (i). N-Terminal Amino Acid Sequencing

Recombinant human lysozyme (rLys) isolated from rice cells as described above, was separated by 18% SDS-PAGE followed by electroblotting to a PVDF membrane (Bio-Rad, CA). The lysozyme band was identified by staining the membrane with 0.1% Coomassie Brilliant Blue R-250 in 40% methanol and 1% glacial acetic acid for 1 minute.  
30 The stained PVDF membrane was immediately destained in 50% methanol until the band was clearly visible. After the blot was thoroughly washed with H<sub>2</sub>O and air-dried, it was sequenced with a sequencer ABI 477 by Edman degradation chemistry at the Protein Structure Laboratory of the University of California at Davis. The results showed

that the rLys produced in transgenic rice seed had an identical N-terminal sequences to the human lysozyme, as follows:

Recombinant Lys-----LysValPheGluArg( )GluLeuAlaArgThr

Human Lys-----LysValPheGluArgCysGluLeuAlaArgThr

- 5       The blank parenthesis in recombinant lysozyme represents residue Cys which cannot be detected by the machine. This cycle was not defined, and could be due to the un-modified cysteine residue which cannot form a stable derivative in Edman degradation analysis.

- 10       Additionally, a number of structural and functional attributes of human lysozyme and recombinant lysozyme produced in rice were found to be the same, including molecular weight, pI, bactericidal effect with *E. coli*, thermal and pH stability and specific activity.

(ii). Thermal and pH stability of lysozyme

- 15       For biotechnological applications of the recombinant human lysozyme, its thermal and pH stability as well as its resistance to proteases is of decisive importance. A human lysozyme standard and lysozyme from rice were diluted to a final concentration of 50 µg/ml in PBS and subjected to the following thermal treatment in a sequential mode: (1): 62 °C for 15 minutes; (2): 72 °C for 20 seconds; (3): 85°C for 3 minutes and finally; (4): 100 °C for about 8 to about 20 seconds. Studies were conducted with 100 µl per tube and repeated three times. Aliquots were saved at the end of each treatment and the remaining lysozyme activity was measured by activity assay. The result showed that recombinant lysozyme exhibited the same degree of thermal stability in the temperature range from 62 °C to 100 °C as human lysozyme.

- 25       In another embodiment, approximately 50 µl of Hlys or rHlys was dissolved in PBS at 100 µg/ml and subjected to heat treatment. Four different temperatures of 65°C, 72°C, 85°C and 100°C were tested. With each temperature, 0 min, 0.33 min, 1.5 min, 3 min, 5 min and 15 min were selected to analyze the impact of incubation time on the stability of lysozyme (Fig 5A).

- 30       For studies on pH stability, lysozyme was dissolved in 0.9% NaCl at 100 µg/ml at pH 10, 9, 7.4, 5, 4, and 2. The solutions were incubated at 24 °C for one hour. Experiments were conducted with 200 µl per tube and repeated three times. Remaining lysozyme was detected by lysozyme activity assay.

For pH treatments at pH 2, 4 and 5, Hlys and rHlys was dissolved in PBS adjusted to the corresponding pHs with HCl at 100 µg/ml. For pH 9 and 10, lysozyme

was dissolved in TBS and 150 mM sodium carbonate/bicarbonate at 100 µg/ml, respectively. Approximately 100 µl of lysozyme solution was incubated at 37°C for 30 min. The lysozyme activity was assessed by activity assay (Fig 5B).

Both Hlys and rHlys displayed similar thermal and pH stability.

5 (iii). Determination of *in vitro* protease resistance of lysozyme

Lysozyme was dissolved in 0.9% NaCl at 100 µg/ml. The pH of the solution was reduced to 3, 4 and 5 with HCl. Pepsin (Sigma, MO) (pepsin: lysozyme = 1:22 (w/w)) was added and the solutions were incubated at 37°C for one hour. Then the pH of all treatments was raised to pH 7 with bicarbonate. Pancreatin (Sigma, MO) (pancreatin :  
10 lysozyme = 1:110 (w/w)) was added to the neutral solution and incubated at 37 C for two hours. The remaining lysozyme activity was measured by activity assay.

In *in vitro* digestion experiments with pepsin and pancreatin, the native and recombinant human lysozyme displayed very similar resistance to pepsin and pancreatin digestion. Under these conditions, human albumin was degraded as demonstrated by  
15 SDS-PAGE (data not shown).

(iv). Biochemical Characterization of Lysozyme

After recombinant human lysozyme was purified to near homogeneity, several biochemical characterizations were carried out to compare human milk lysozyme with recombinant human milk lysozyme derived from rice cells. The results summarized in  
20 Table 3 show that by SDS-PAGE, native human milk lysozyme and recombinant lysozyme migrated to the same position.

Nucleotides encoding the rice *Ramy3D* signal peptide were attached to the human lysozyme gene in the expression vector pAPI156. Determination of the N-terminal amino acid sequence of the purified recombinant human lysozyme revealed an  
25 N-terminal sequence identical with that of native human lysozyme, as detailed above. Rice cells thus cleave the correct peptide bond to remove the *RAmy3D* signal peptide, when it is attached in the human lysozyme precursor.

The overall charge of recombinant and native human lysozyme were compared by isoelectric-focusing (IEF) gel electrophoresis and pI values determined. Since  
30 lysozyme is a basic protein with a calculated pI of 10.20, the pI comparison studies were carried out by reverse IEF gel electrophoresis. Recombinant and native human lysozyme displayed identical pI, indicating the same overall charge (data not shown).

Recombinant human lysozyme derived from transgenic rice had a specific activity similar to the native lysozyme (200,000 units/mg (Sigma, MO), whereas, lysozyme from chicken egg whites had the expected 3-4 fold lower specific activity (Sigma, MO) (Fig. 4).

5 Table 3. Comparison of Biochemical Characteristics of Human Milk Lysozyme and Recombinant Lysozyme

| Lysozyme source | N-terminal sequence | Size (kDa) | Glycosylation | Specific activity (units/mg) | pI   |
|-----------------|---------------------|------------|---------------|------------------------------|------|
| Human milk      | KVFER C ELART       | 14         | No            | 201,526                      | 10.2 |
| rice            | KVFER(-)*ELART      | 14         | No            | 198,000                      | 10.2 |

\* This cycle was not defined, and could be due to the un-modified cysteine residue which cannot form a stable derivative in Edman degradation analysis.

The results described above demonstrate the ability to use rice cells as a production system to express human lysozyme from milk. Over 160 individual transformants were screened by immunoblot, enzymatic activity assay and ELISA. Yields of recombinant human milk lysozyme reached 4% of soluble cell proteins in culture cells and over 40% of soluble proteins in rice grains. Although the mechanism is not part of the invention, the high expression level may be explained by the utilization of the strong *RAmy3D* promoter (Huang *et al.*, 1993) in culture cell system and *Gt1* promoter in grain expression system and the codon-optimized gene.

The plant derived human milk lysozyme obtained by the methods of the present invention was identical to endogenous human lysozyme in electrophoretic mobility, molecular weight, overall surface charges and specific bactericidal activity.

#### EXAMPLE 4

##### Characterization of Recombinant Human Lactoferrin (rLF) Produced by Transgenic Rice

##### Plants

##### A. Southern blot analysis

About three grams of young leaf were collected and ground with liquid nitrogen into a very fine powder. The DNA was isolated according to the procedure as described in Dellaporta *et al.*, 1983, and purified by phenol-chloroform extraction. Approximately 5 µg of *EcoRI* and *HindIII* digested DNA from each line was used to make blot for

Southern analysis. The ECLTM direct nucleic acid labeling and detection system (Amersham, USA) was used for analysis.

The lactoferrin gene copy number was estimated to be from about 1 to about 10 as determined by Southern blot hybridization using *EcoRI* and *HindIII* digested genomic DNA. The API164-12-1 (R0) transgenic plant line was subjected to Southern analysis together with ten Western blot positive, field grown R1 lines. A typical Southern blot shows that there are at least three fragments above the original plasmid derived plant transformation unit (3156 bp). All the LF inserts appear to be inherited from the original R0 transgenic plant event to R5 generation.

10

#### B. Protein Isolation and Western blot

Rice seeds were ground with 1 ml of 0.35 N NaCl in phosphate buffer saline (PBS), pH 7.4 using an ice-cold mortar and pestle and the resulting homogenate was centrifuged at 15000 rpm for 15 min at 4°C. The supernatant was used as a protein extract and about 1/25 or 1/50 of the salt soluble content was loaded onto a 10% pre cast gel (Novex, USA) and electrophoresis was carried according to the manufacturer's instructions. For total protein detection, the polyacrylamide gel was stained with 0.1% Coomassie brilliant blue R-250 (dissolved in 45% methanol and 10% glacial acetic acid) for at least three hours and destained with 45% methanol and 10% glacial acetic acid until the desired background was achieved.

20

For Western blot analysis, SDS-PAGE gels were electroblotted onto a 0.45 µm nitrocellulose membrane with a Mini-Trans-Blot Electrophoretic Transfer Cell System (Bio-Rad, USA) and subsequently subjected to immuno-blotting analysis. The blot was blocked with 5% non-fat dry milk in PBS for at least two hours followed by three washes with PBS for 10 minutes each. The primary rabbit polyclonal antibody against hLF (Daka A/S, Denmark) was diluted at 1:2500 in the blocking buffer and the blot was incubated in the solution for one hour. The blot was washed with PBS for three times with 10 minutes each. The secondary goat anti-rabbit IgG (H+L)-alkaline phosphatase conjugated (Bio-Rad, USA) was diluted in the blocking buffer at 1:5000 ratio. The membrane was incubated in the secondary antibody solution for one hour and followed by three washes with PBS. Color development was initiated by adding the substrate system BCIP-NBT (Sigma, USA) and the process was stopped by rinsing the blot with H<sub>2</sub>O once the desirable intensity of the bands was achieved.

25

30

One hundred eight (108) R0 plants were grown to maturity, seeds were harvested from 56 fertile plants and individual seeds analyzed by Western blot to detect the expression of rLF. Coomassie blue staining was carried out to compare the mobility of rLF with native human lactoferrin (hLF) (Fig. 8), with 40 µg of total protein loaded onto each lane, along with 40 ng of native purified hLF per lane as the positive control.

Estimation of total rLF by ELISA indicated that from 93 µg to 130 µg rLF was expressed in transformed rice seeds. A typical Western blot analysis (Fig. 9) illustrates that both rLF and native hLF migrate at approximately the same rate with the molecular weight about 80 kDa, consistent with that determined by other researchers (Wang *et al.*, 1984).

### C. Protein purification

Rice seeds from R2 homozygous generation were dehusked and milled to flour conventionally. Recombinant lactoferrin was extracted by mixing the rice flour with 0.35 N NaCl in PBS at 100 g/l at room temperature for two hours. The resulting mixture was centrifuged at 15,000 rpm for one hour at 4°C. The collected supernatant was subjected to the following steps of filtration before loading onto a Sepharose column. First, the supernatant was run through a few layers of cheesecloth. Then the filtrate was passed sequentially through an 8µm paper, 1 µm paper and a 0.25µm nitrocellulose membrane. The clear protein solution was loaded onto a ConA Sepharose column (Pharmacia, XK 26) which had been equilibrated with 0.5 N NaCl in 20 mM Tris, pH 7.4 (binding buffer) at a flow rate at 4 ml/min. After the loading was complete, the column was washed with binding buffer until the baseline at A280nm was achieved. Lactoferrin was eluted with 0.1N mannoside in the binding buffer. Fractions containing lactoferrin were pooled and loaded onto a second column SP-Sepharose (Bio-Rad, USA) which has been equilibrated with 0.4 N NaCl in 50 mM sodium phosphate, pH 8.0 (binding buffer\_B) at the flow rate 4 ml/min. Then the column was washed with the binding buffer\_B until the baseline at A280 nm was obtained. Lactoferrin was eluted by 1 N NaCl in 50 mM sodium phosphate, pH 8.0 and the fractions containing LF were pooled and dialyzed against PBS. Finally the purity of LF was assessed by SDS-PAGE and stored at - 80°C.

In another embodiment, recombinant human lactoferrin (rHLF) was extracted by mixing rice flour with 0.35 M NaCl in PBS at 75 g/L at room temperature for 2.5 hours. The extract was passed through six layers of cheesecloth before centrifugation (10,000 g for 1 hour at 4°C). The supernatant was recovered and the NaCl concentration was

adjusted to 0.4 M (pH 8.0). After a second centrifugation at 10,000 g for 10 minutes at 4°C, the supernatant was collected and filtered through 0.45 µm nitrocellulose membrane. The filtrate was loaded onto a SP-Sepharose column (Bio-Rad, Hercules, CA) which had been equilibrated with 0.4 M NaCl in 50 mM sodium phosphate, pH 8.0 (binding buffer) at a flow rate of 4 ml/min. The column was washed with the binding buffer until baseline A280 was obtained. Lactoferrin was eluted by a linear gradient and dialyzed against PBS. The purified rHLF was analyzed by SDS-PAGE and stored at –80°C.

10           D. Enzyme Linked Immunosorbant Assay (ELISA)

ELISA was conducted using seed extracts, isolated as described above, with total protein assayed using the Bradford method (Bradford, M., 1976). The ELISA was based on a typical sandwich format generally known in the art. Briefly, 96 well plates were coated with rabbit anti-human lactoferrin antibody (Daka A/S, Denmark), then rLF and control samples were added to individual wells of the plate and incubated for 1 hour at 35°C. Rabbit anti-human lactoferrin horseradish peroxidase conjugate (Bioscience, USA) was then added to each well and incubated for 1 hour at 35°C, followed by addition of the tetramethylbenzidine substrate (Sigma, USA) and incubation for 3 minutes at room temperature. The reaction was stopped by adding 1N H<sub>2</sub>SO<sub>4</sub> to each well. The plates were read at dual wavelengths of 450 and 650 nm in a Microplate Reader (Bio-Rad, model 3550) and the data was processed by using Microplate Manager III (Bio-Rad). The results of an analysis of 10 homozygous selected lines showed that from 93 µg to 130 µg rLF was expressed per seed.

25           E. Selection of plants for advance generations

At least 20 - 40 seeds from 11 independent lines were analyzed. Individual R1 seeds were cut into half and endospermic halves were subjected to analysis by Western blot with the positive corresponding embryonic halves germinated on 3% sucrose medium with 0.7% agar. The seedlings were transplanted to the field for R1 generation. Out of 11 individual lines, 3 lines were expressed. A total of 38 plants were grown in the field derived from the 3 expressed mother lines. Based on the agronomic character (Table 4) of those 38 plants, 28 plants were selected.

It was observed that all the Western positive R1 seeds were opaque to pinkish in color in comparison to control seeds, so this criterion was applied in screening the R2

seeds. Mature R2 seeds were harvested at maturity and dehusked. The pinkish R2 seeds were confirmed by Western dot blot and ELISA as expressing rLF (data not presented). Finally 10 homozygous R2 lines were selected and grown in the field in order to advance the generation.

5

Table 4. Comparison Of Phenotypic Characteristics Of Native TP-309 And Transformed TP-309 Rice Seeds

| Source                      | Effective tiller | Blank grain | 1000 seed weight (g) | µg of rLF/seed |
|-----------------------------|------------------|-------------|----------------------|----------------|
| TP-309                      | 43               | 5.0         | 25                   |                |
| Homozygous transgenic lines | 42               | 19.7        | 20.2                 | 125            |

10 During R2 and R3 generation the percentage of blank seeds was higher in homozygous transgenic lines than in the non-transgenic control. This affected the 1000 seed weight. However, in the R4 generation no significant differences in phenotypic character were observed in homozygous transgenic lines when compared to non-transformed TP309 (Table 4).

15

#### F. Attributes Of Recombinant Human Lactoferrin Produced In Rice

Physical characterization of the rLF showed there was no significant difference between the rLF and a commercially available purified form of hLF based on N-terminal amino acid sequencing, and physical characteristics of rLF such as molecular weight as determined by MALDI-MS, HPLC profile of which showed a comparable peptide map, pH dependent iron release and bacteriostatic activity, using the analyses described below.

20

##### (i). N-Terminal Amino Acid Sequencing

Purified rLF from rice seeds was resolved by 10% SDS-PAGE, followed by electroblotting to PVDF membrane (Bio-Rad, USA). The target band was identified by staining the membrane with 0.1% Coomassie brilliant blue R-250 in 40% methanol and 1% glacial acetic acid for 1 minute. The stained PVDF membrane was immediately destained in 50% methanol until the band is clearly visible. The blot was thoroughly washed with ddH<sub>2</sub>O and air dried. Finally this sample was sent to the Protein Structure Laboratory in University of California at Davis (CA, USA) for sequencing analysis.

25

(ii). Detection of glycosylation and determination of sugar content

Glycosylation of the recombinant human lactoferrin produced in rice was analyzed by an immunoblot kit for glycoprotein detection (Bio-Rad, USA) per instructions from the manufacturer. An increase of molecular weight of lactoferrin due to carbohydrate content  
5 was determined by Matrix Assisted Laser Desorption Ionization-Mass spectrometry (MALDI-MS) (PE Applied Biosystems, Voyager System).

Recombinant lactoferrin produced in rice is glycosylated as evident from the binding to Con A resin, the positive staining by glycoprotein detection kit as well as the larger detected mass as compared to the calculated mass (76.2 kDa) based on the  
10 peptide backbone. MALDI-MS showed that seed derived recombinant lactoferrin has molecular weight of 78.5 kD while human milk lactoferrin is 80.6 kDa (Table 5). The difference could be due to the lesser degree of glycosylation in the rice seed-derived lactoferrin. Analysis shows that the purified rHLF contains xylose but lacks sialic acid, which is consistent with plant post-translational modification patterns (Matsumoto *et al.*,  
15 1995).

(iii). Determination of isoelectric point of Lactoferrin

Reverse isoelectric focusing (IEF) gel electrophoresis was carried out with a precast Novex IEF gel, pH 3-10 according to the manufacturer' instruction. About 30 µg of purified rLF was loaded and the running condition was 100 V for 50 minutes and 200 V  
20 for 20 minutes. The gel was then fixed in 136 mM sulphosalicylic acid and 11.5% TCA for 30 minutes, stained in 0.1% Coomassie brilliant blue R-250, 40% ethanol, 10% glacial acetic acid for 30 minutes and destained in a solution containing 25% ethanol and 8% acetic acid.

(iv). Comparison of physical characteristics of rLF with native hLF

25 The HPLC profile of native and rLF showed a comparable peptide map. This confirmed that LF from the two sources have an identical amino acid sequence (data not presented). Additional comparisons confirm that human lactoferrin produced in transgenic rice closely resembles native human lactoferrin, as evidenced by (1) the N-terminal sequence of purified rLF from homozygous R2 seeds and hLF (Dakao A/S,  
30 Denmark), which were shown to be identical (Table 5); (2) the isoelectric point (pI) of native and rice seed derived LF which is the same, indicating that they have similar surface charges (Table 5); (3) the pH dependent iron release of rLF which was shown to be closely related to that of native hLF (Fig. 11 and see section vii of example 4); and (4) the bacteriostatic activity of rHLf which was shown to be similar to that of native human

lactoferrin (nHLf) on enteropathogenic *E. coli* (EPEC; Fig. 10) and confirmed the presence of active recombinant LF in extracts derived from transformed rice seeds (see section ix of Example 4).

5 Table 5. Physical characterization data for human (hLF) and rice seed derived recombinant lactoferrin (rLF)

| LF source | Size (kDa) | N-terminal sequence               | pI  | Glycosylated | Sugar content (%) |
|-----------|------------|-----------------------------------|-----|--------------|-------------------|
| hLF       | 80.6       | GlyArgArgArgArgSerValGlnTrpCysAla | 8.2 | YES          | 5.5               |
| rLF       | 78.5       | GlyArgArgArgArgSerValGlnTrp( )Ala | 8.2 | YES          | 2.9               |

(v). Iron content and nutrient value determination of rice seeds

10 The iron content of R2 homozygous seeds was determined. Two grams of dry mature seeds from each transformed and non-transformed line were weighed and wet-ashed with HNO<sub>3</sub> and H<sub>2</sub>O<sub>2</sub> solution at 110°C (Goto *et al.*, 1999). The ash was dissolved in 1N HCl solution. The iron content was then measured by absorbance of Fe-O-phenanthroline at 510 nm, using a Sigma kit (Sigma, USA) per instructions of  
15 manufacturer.

The different values of nutrient facts of homozygous transgenic seeds and non transgenic seeds were measured by standard procedure at A & L Western Agricultural Laboratories (Modesto, CA, USA).

20 A comparative analysis of transgenic lactoferrin-expressing rice seeds with non transformed native Teiwei-309 showed that there is no significant difference between transformed and non transformed seeds in nutrient value with the exception that the concentration of iron is 50% greater (Table 6). The increased level of iron may be the reason for the opaqueness and pink coloration of the rLF expressing transgenic rice seeds.

25 In another embodiment, 0.2 grams of dried, dehusked grains expressing rHLF were wet-ashed with concentrated HNO<sub>3</sub> for two days and dissolved in 5 ml of DDI H<sub>2</sub>O. The iron contents of the samples were measured by flame atomic absorption spectrophotometry (Thermo Jarrel Ash SH4000, Franklin, MA). NIST liver was analyzed concurrently to verify the accuracy of the standard curve.

30 The iron content of transgenic rice grains was more than twice that of non-transformed TP309 grains, while there were no significant differences in other tested

nutrition factors between transformed and non-transformed grains (Table 7). This suggests that groups ingesting transgenic rice with rHLF will increase the iron intake.

The transgenic grains with increased iron content were opaque-pinkish in color. The opaque-pinkish color was observed inside as well as outside the rice endosperm.

- 5 This opaque-pinkish color, segregated in Mendelian fashion, was linked with expression of rHLF and was inherited through the R4 generation.

There was no difference noticed during the seed germination of transgenic seeds, the phenotype of R2 R3 and R4 plants was vigorous and the seed yield was similar to that of non-transgenic Teipei-309 plants (data not shown).

10

Table 6. Comparison of Nutrition Value (in mg) Per 100 Gram of Non Transformed and Transformed Rice Seeds

| Source                             | Carbohy-<br>drate | Protein | Fat | Ca | K   | Na  | Fe  | Water | Calories |
|------------------------------------|-------------------|---------|-----|----|-----|-----|-----|-------|----------|
| TP-309                             | 76.0              | 8.7     | 2.4 | 9  | 370 | <10 | 0.8 | 11.3  | 369      |
| Homozygous<br>transformed<br>lines | 75.7              | 8.7     | 2.2 | 8  | 330 | <10 | 1.2 | 11.8  | 367      |

- 15 Table 7. Comparison of Mineral Contents (in µg) Per Gram of rHLF-Transformed and Non-Transformed Rice Grains

| Source          | Cu  | Fe   | Mn   | Zn   |
|-----------------|-----|------|------|------|
| Non-transformed | 2.9 | 8.7  | 33.1 | 20.8 |
| Transformed     | 4.7 | 19.2 | 17.7 | 28.7 |

(vi). Tissue specificity and stability of rLF

- 20 An endosperm specific rice glutelin promoter was used to express recombinant lactoferrin in maturing or matured seeds. To confirm the tissue specificity of the expressed lactoferrin, protein was extracted from root, shoot, leaf beside mature seed and subjected to Western blot and the results indicated that there was no detectable expression of rLF except in the seed/endosperm (Fig. 9). Furthermore, the presence of
- 25 rLF in 5 day old germinated seeds showed the stability of stored rLF within the plant cell during germination.

(vii). Iron Saturation And pH Dependent Iron Release

Lactoferrin was incubated with 2M excess ferric iron ( $\text{FeCl}_3$  :NTA = 1:4) and sodium bicarbonate ( $\text{Fe}:\text{HCO}_3^- = 1:1$ ) for 2h at room temperature. Excess free iron was removed by using a PD-10 desalting column (Pharmacia, USA) and the iron saturation level was determined by the A280/A456 ratio. Both native hLF and rLF were completely saturated by iron. Holo hLF was incubated in buffers with a pH between 2 and 7.4, at room temperature for 24 h. Free iron released from hLF was removed and the iron saturation level was determined by A280/A456 ratio.

The results showed that iron release was similar for both hLF and rLF. Iron release began around pH 4 and was completed around pH 2 (Fig. 11). The iron binding was reversible since iron-desaturated rLF was re-saturated by raising the pH to 7 (data not shown). The similarity in pH dependent iron release of rLF to that of the hLF standard demonstrated that rLF is able to adapt the appropriate tertiary structure for proper iron binding and release (Salmon, Legrand *et al.* 1997).

(viii). Binding and uptake by Caco-2 cells

50,000 Caco-2 cells/well were seeded and grown in Minimum Essential Medium (GIBCO, Rockville, MD) containing 10% fetal bovine serum in 24 or 48 well tissue culture plates for 3 weeks. For binding studies, Caco-2 cells were incubated with varying concentrations (0-2  $\mu\text{M}$ ) of  $^{125}\text{I}$ -HLf in the presence or absence of 100-fold excess of unlabeled nHLf for 2 hours at 4 °C and cells were washed 5 times with ice-cold PBS. Cells were solubilized with 0.5 ml of 0.1% SDS and radioactivity was quantified in a gamma counter. For uptake studies, 0.4  $\mu\text{M}$  of  $^{125}\text{I}$ -HLf was incubated with Caco-2 cells for 0 to 24 hours at 37 °C and cells were washed, dissociated by the same way as in the binding study. 0.5 ml of 24% TCA solution was added to the dissociated cells and free iodine was removed by the centrifugation. Free and protein-bound  $^{125}\text{I}$  were quantified separately to evaluate how much of HLf was degraded in the cells. Receptor-binding of rHLf to the human intestinal Caco-2 cell line was saturable and specific, indicating that rHLf bound to the Lf receptor. The binding constant was similar for rHLf and nHLf, but the number of binding sites was slightly higher for rHLf, which may be due to the difference in glycosylation. Uptake of HLf by Caco-2 cells was identical for rHLf and nHLf.

(ix). In vitro Digestion: effect on antimicrobial activity and binding/uptake to Caco-2 cells

Lactoferrin is known to inhibit the growth of a variety of bacterial species based on its iron chelation and direct bactericidal properties. The anti-microbial effect of rLF

extracted from rice seeds was tested following treatment using an *in vitro* digestion model with an enzymatic system containing pepsin (an enzyme active in stomach) and pancreatin (an enzyme active in deodenum).

LF proteins were dissolved in PBS at 1mg/ml, and either left untreated, pepsin  
5 treated (0.08mg/ml at 37°C for 30 min), or pepsin/pancreatin treated (0.016 mg/ml at 37°C for 30 min). LF proteins were sterilized by passing through a membrane filter with a pore size of 0.2 µm [Rudloff, 1992]. The filter sterilized LF (0.5µg/ml) was incubated with 104 colony forming unit (CFU) enteropathogenic *E. coli* (EPEC)/µl in 100 µl sterile synthetic broth (1.7% : AOAC) containing 0.1% dextrose and 0.4 ppm ferrous sulfate at  
10 37°C for 12h and colony forming units (CFU) were determined.

Starting with an enteropathogenic *E. coli* (EPEC) concentration of 10<sup>4</sup> CFU (colony forming units), the untreated samples of rLF reached up to 10<sup>6.5</sup> CFU after 12 h of incubation at 37°C in comparison to hLF, which produced up to 10<sup>6</sup> CFU. An *in vitro* digestion model using an enzymatic system containing pepsin (enzyme active in  
15 stomach) and pancreatin (enzyme active in deodenum) with moderate shaking to imitate the transit of protein through infant gut [Rudloff, 1992] was used. rLf and nHLf were treated with active pepsin and pancreatic enzymes and exposed to 10<sup>4</sup> CFU EPEC cells for 12 h at 37°C (Fig. 10). Both the native human lactoferrin standard (nHLf) and the recombinant rice-derived lactoferrin (rLf) remained active in inhibiting growth of  
20 enteropathogenic *E. coli*, indicating that both nHLf and rHLf are resistant to protease digestion.

SDS-PAGE and ELISA revealed that nHLf and rHLf resist digestion by pepsin (at pH 3.8) and pancreatin, whereas human serum albumin is completely digested after *in vitro* digestion. Western blots revealed that immunoreactivity was also maintained after  
25 digestion. Although some smaller molecules were generated during digestion of HLf, most of the immunologically detectable HLf retained its intact size. More than 50 % of rHLf and nHLf was immunologically detectable by ELISA, but <sup>125</sup>I-HLf was around 40 % and <sup>59</sup>Fe-HLf was only 20 % detectable, indicating that ELISA detects small peptide fragments of HLf, which are removed by the PD-10 column and that about 50-60 % of Fe  
30 was released from detectable HLf after *in vitro* digestion. The iron-holding capacity was not significantly different.

The dissociation constant (Kd) and the number of binding sites for HLf to its receptor were determined from the binding study. Both Kd and the number of bindings sites were not significantly different between nHLf and rHLf after *in vitro* digestion (Figure

12A, 12B). Digestion did not appear to affect on the Kd but made the number of binding sites much lower. Total Lf uptake was not significantly different between nHLf and rHLf after *in vitro* digestion (Figure 12C), though uptake was about one third when compared with undigested nHLf. Total iron uptake from nHLf was twice as high as that from rHLf.

- 5 Percent degradation of HLf was similar regardless of digestion or not, and the native or recombinant form (Figure 12D).

(x). Thermal Stability: effect on antimicrobial activity and binding/uptake to Caco-2 cells

- 1.0 mg/ml of holo-HLf in PBS was treated by the following conditions: (a) 62 °C for 15 minutes, (b) 72 °C for 20 seconds, (c) 85 °C for 3 minutes, or (d) 100 °C for 8 seconds. Survival ratio of HLf determined by ELISA were more than 90% following treatment at 62 °C for 15 minutes, at 72 °C for 20 seconds, or at 85 °C for 3 minutes, but it was considerably lower after 100 °C for 8 seconds. This high temperature precipitated both types of HLf and only 10% of HLf was detectable by ELISA. More than 80% of iron was still bound to both rHLf and nHLf after all thermal treatments with the exception of 100 °C for 8 sec. In 10% of survived HLf after 100 °C for 8 sec, the iron saturation level of nHLf was above 80% whereas that of rHLf was only about 40%.

- SDS-PAGE and Western blots revealed no difference in immunoreactivity between nHLf and rHLf at 62 °C for 15 minutes, at 72 °C for 20 seconds, and at 85 °C for 3 minutes, but at 100 °C for 8 seconds, rHLf almost completely lost its immunological activity, whereas nHLf still maintained detectable immunoreactivity.

There was no significant difference in anti-microbial activity between nHLf and rHLf after heat-treatment. Anti-microbial activity of HLf was not affected by treatment at either 62 °C for 15 min, 72 °C for 20 sec or 85 °C for 3 min.

- The Kd and the number of binding site for nHLf and rHLf were not significantly different at 62 °C and 72 °C though there is a trend that nHLf is somewhat lower Kd and binding sites than rHLf. As the temperature was increased (such as 85 °C and 100 °C), more rHLf bound to Caco-2 cells, most likely by non-specific binding due to more rHLf being denatured than nHLf. Uptake properties were similar for nHLf and rHLf even in the group treated at 100 °C where uptake of both types of HLf was highest among all the thermal treatments. Free iodine levels in the cells were also evaluated since it reflects degradation of HLf. About 20% of HLf was degraded in the untreated sample. There was no significant difference between nHLf and rHLf. Interestingly, samples treated at 100 °C were degraded twice as much as untreated samples of nHLf and rHLf, which may

indicate that denaturation of HLf caused by heat treatment will make the protein more susceptible to proteases in the cells.

(xi). pH Stability: effect on antimicrobial activity and binding/uptake to Caco-2 cells

5 1.0 mg/ml of holo-HLf in PBS was adjusted to pH 2, 4, 6, or 7.4 by the addition of 1 M HCl and incubated for 1 h at room temperature. The pH was then adjusted to 7.0 with 1 M NaHCO<sub>3</sub>. Free iron released from HLf, was removed by a desalting column.

After low pH treatment, 100% of both nHLf and rHLf survived. The iron-holding capacity was maintained in all samples and the iron saturation level was above 95%.

10 SDS-PAGE and Western blots revealed that there was no difference between nHLf and rHLf for any of the treatments. A slightly smaller immunoreactive molecule (~70 kD) was detected after exposure of nHLf to pH 2 and 4 and of rHLf to pH 2.

Antimicrobial activities of nHLf and rHLf were stable after exposure to low pH in the range of pH 2.0 to 7.4. As the pH was lowered, the activity of rHLf appeared to be  
15 higher and constant, whereas nHLf did not show any pH dependency.

K<sub>d</sub> and the number of binding sites for nHLf were not significantly different from those for rHLf but a trend was always lower for nHLf within the range of pH 2.0 to 7.4, which is similar to control and thermal treatment samples. The K<sub>d</sub> and the number of binding sites for nHLf and rHLf were unaffected by pH treatment down to 2.0 for 1 hour.

20 Uptake properties were similar for nHLf and rHLf in the pH range of 2.0 to 7.4.

Degradation of HLf in Caco-2 cells was also evaluated and there was no significant difference between nHLf and rHLf.

## EXAMPLE 5

### 25 Generation and Characterization of Recombinant Human $\alpha$ -1-antitrypsin (AAT) Produced by Transgenic Rice Plants

#### A. Construction and expression of human AAT in rice cells

The construction and purification of functional recombinant human AAT were carried out as exemplified in previous examples. Briefly, codon-optimized  
30 AAT gene was cloned into an pAPI145 that contains the rice Gt1 promoter, Gt1 signal peptide, and Nos terminator, pAPI241 that contains Glb promoter, Glb signal peptide, and Nos terminator, and API280 that contains Bx7 promoter, Bx7 signal peptide, and Nos terminator, as exemplified in Example 1. The resulting plasmids were named pAPI250, API255 and pAPI282, respectively (Figure 13).

Transgenic plants expressing AAT were generated as above, and plant-generated recombinant AAT was characterized. To express AAT in culture cells, codon-optimized AAT gene was cloned into an expression cassette that contains the rice RAmy3D promoter, signal peptide, and terminator. Recombinant AAT  
5 expression was induced and secreted to the culture medium under the sugar starvation condition. Purification of rAAT was achieved through a scheme that consisted of an affinity column (Con A), anion exchange column (DEAE), and a hydrophobic interaction column (Octyl).

10        **B. SDS- Polyacrylamide Gel Electrophoresis (SDS-PAGE)**

AAT samples were ground with PBS with mortar and pestle. The resulting extract was spun and 20 microliters of supernatant loaded into a precast SDS-PAGE gel. The AAT protein was clearly visualized with Coomassie brilliant blue staining (Figure 14).

15

**C. Western Blot Analysis**

For immunoblotting analysis, gels were electroblotted to a 0.45 µm nitrocellulose membrane with a Mini Trans-blot Electrophoretic Transfer cell (Bio-Rad, USA) and subsequently subjected to immunoblotting analysis. Blots were  
20 blocked with 5% non-fat dry milk in PBS, pH 7.4 for at least two hours followed by three washes with PBS, pH 7.4 for 10 minutes each. The primary rabbit polyclonal antibody against human alpha-1-antitrypsin (Dako A/S, Denmark) was diluted to 1:2500 in the blocking buffer and the blot was incubated for at least one hour. The blot was then washed as described previously. The secondary antibody, goat anti-  
25 rabbit IgG (H+L)-alkaline phosphatase conjugated (Bio-Rad), was diluted in the blocking buffer at a dilution of 1:4000. The membrane was then incubated in the secondary antibody solution for one hour and followed by the same wash process. Color development was initiated by adding the substrate BCIP/NBT from Sigma.

The western result showed that AAT protein is clearly visualized and  
30 confirmed that AAT expressed and deposited in transgenic rice grain, has a molecular weight that is somewhat smaller than that of native AAT (Figure 15).

**D. ELISA**

Standards for this assay ranged from 1.25 - 20 ng/mL of AAT (Athena)

diluted in PBST. Nunc Immuno-plate Maxisorp 96-well plates (Nunc, Denmark) were coated for 16 h at 4°C by a 1:10,000 dilution of rabbit anti-human AAT in 0.05 M sodium bicarbonate, pH 9.6. The plates were washed 3 times with PBST (PBS, pH=7.4, 0.05% Tween-20) and subsequently incubated with sample for 1 h  
5 at room temperature while rocking. The plates were washed again 3 times with PBST, followed by incubation with a 1:50,000 dilution of goat anti-human AAT conjugated to HRP for 1 h at room temperature. The plates were washed 3 times with PBST, and bound antibody was detected by the HRP/ hydrogen peroxide catalyzed reaction of TMB. The reaction was stopped with 2 M sulfuric acid, and  
10 the plates were read on a microtiter plate reader at 450 nm, using 620 nm as a reference filter.

Recombinant AAT is 2.1 times more immunoreactive, when comparing equal concentrations as determined by the Lowry assay.

15 E. AAT activity assay

AAT activity was analyzed using a modified method published by Travis and Johnson (1981). In 96-well microtiter plates, 60 µL samples diluted in Tris buffer, (0.2 M Tris, pH 8.0) were added. In each well, 60 µL of elastase [0.01 mg/mL porcine pancreatic elastase (PPE) in Tris buffer] was also added. The  
20 plate was rocked for 5 min at room temperature to allow any available AAT to bind to the elastase. Another 120 µL of substrate solution (10 M N-Succinyl-AAA-p-nitroanilide in DMSO diluted in Tris buffer to give 0.33 M N-Succinyl-AAA-p-nitroanilide) was added, and the plate was rocked for 1-2 min at room temperature. The plate was immediately read on a microtiter plate at 405 nm.  
25 The plate was read again after 5 min, and the change in absorbance was calculated. AAT activity was determined using linear regression from a standard curve. The results show that AAT protein produced in rice grain has similar bioactivity as that of native AAT.

30 F. Band shift assay

The unique property of the covalent-linked complex formed between AAT and PPE permits an analysis of the activity of AAT by SDS-PAGE. Briefly, 20 µl of tested samples containing AAT from the screening or purification processes was incubated with 100 ng PPE at 37°C for 15 minutes. Five µl of SDS-loading dye

was added and the reaction mixture boiled for five minutes. The sample was then centrifuged and kept on ice until loaded onto a 10% precast SDS-PAGE gel. The resulting gel was stained with 0.1% Coomassie brilliant blue R-250 as described below. For immunodetection, a western blot analysis was carried as described  
5 above. Again the band shift assay indicated that AAT protein produced in rice grain has similar bioactivity as that of native AAT (Fig. 16A and 16B).

G. In vitro digestion. The digestion was carried out using a modified method of Rudloff and Lonnerdal (1992) was used after some modifications.  
10 Native and recombinant AAT were diluted in PBS or formula to 0.5 mg/mL. Hydrochloric acid (1 M) was added to all samples to adjust the pH 3, 4, and 5, then 2.5  $\mu$ L of 2% pepsin in 0.01 M HCl (3,100 U/mg solid) were added and all samples were placed in a shaking incubator for 30 or 60 min at 37°C. The pH was restored by drop-wise addition of 1 M NaHCO<sub>3</sub>, and 2.5  $\mu$ L of 0.4%  
15 pancreatin in 0.1 M NaHCO<sub>3</sub> were added. Samples were incubated for 1 or 2 hours at 37°C, and the reaction was halted by dilution 1:2 in sample buffer and boiling for 3 min. For samples subjected to pepsin digestion only, boiling was unnecessary since the pepsin was inactivated when the pH was raised above pH 6 with NaHCO<sub>3</sub> (Piper and Fenton, 1965). The enzyme: substrate ratio was  
20 approximately 1:20 for samples in buffer only and about 1:600 for samples in formula.

A significant amount of recombinant and native AAT survived the *in vitro* digestion, and both forms were more resistant to degradation than human serum albumin. Digestion with pepsin at pH 4 shows that 65% of recombinant AAT is  
25 detectable by ELISA after digestion, which is similar to 67% of native AAT surviving. The trypsin assay shows that much of the inhibitory properties of both forms are still intact, and the activity assay reveals that 63% and 59% of the activity of native and recombinant AAT remains, respectively. When exposed to both pepsin and pancreatin in buffer, native AAT resisted degradation when the  
30 pH of the pepsin incubation was pH 4 or higher. Under this condition, the recombinant form was less resistant, although a large part remained after pepsin digestion at pH 5 and pancreatin digestion. At pH 4, more of the recombinant protein was degraded, either due to pepsin activity or pH instability. AAT activity could not be determined after digestion by pepsin and pancreatin because of the

inactivation of pancreatin by boiling which also inactivates AAT activity. In formula, both forms appeared to be equally resistant to degradation. While both native and recombinant AAT were still present after pepsin digestion at pH 5 followed by pancreatin digestion, bands at about 33 kD (casein) are faint or missing. It is possible that other proteins in formula are preferentially cleaved, reducing the amount of AAT being digested.

H. Thermal Stability of recombinant human ATT. Both native and recombinant human AAT were diluted in phosphate buffered saline (PBS) or infant formula (Enfamil with Iron, Mead Johnson, Evansville, IL) to a concentration of 0.1 mg/mL. Samples, 100  $\mu$ L in capped, 10 x 75 mm glass tubes, were treated as follows: 60°C for 15 min, 72°C for 20 sec, 85°C for 3 min, and 137°C (temperature of oil bath) for 20 sec. The samples were allowed to cool to room temperature after heat treatment. For formula samples with bile extract added, 2.5  $\mu$ L of 12% porcine bile extract (Sigma) were added, then vortexed quickly, incubated at 37°C for 10 min, and vortexed again. All samples were diluted 1:10 in PBS and transferred to 1.5 mL tubes. Formula samples were centrifuged at 15,000g for 20 min to remove the insoluble fraction, and the supernatant was withdrawn after skimming off the fat. All samples were subsequently transferred to 1.5 mL tubes and analyzed.

The thermal stability of native AAT exceeded that of the recombinant form in buffer, but the recombinant AAT retained significant stability under most conditions. When heated in buffer only, SDS-PAGE and Western blots show that the two forms of AAT have similar structural stability. While the ELISA data show that the recombinant protein is less stable at the higher temperatures, the recombinant protein is similar to the native form under the other conditions. However, the functional stability of the recombinant protein may be affected. The thermal stability assay shows that the recombinant protein lost functional ability at several of the heat conditions, whereas the native protein was functional at all heat conditions except for at 62°C for 15 minutes. While the elastase-inhibiting properties of native AAT were about 90% after all heat treatments, 62 and 51% of the recombinant protein's activity remained after 85°C, 3 minutes, and 137°C, 20 seconds, respectively.

The heat treatments of native and recombinant AAT in formula affected the detection of the proteins, but the addition of bile extract following heat treatment restored antibody recognition of the recombinant form. While the Western blot data show less detectable protein only at 85°C, 3 min for the native AAT and at 5 72°C, 20 sec and 137°C, 20 sec for the recombinant AAT, the ELISA data shows less than 20% protein detected for both forms and for all heat conditions. When bile extracts were added to the heated formula samples, the ELISA data for the recombinant form showed that more than 50% was still detectable after heat treatment. The bile extract did affect detection of the native form by ELISA for 10 most of the heat treatments. The Western blots corroborated the ELISA data and showed that the bile extract may dissociate the recombinant AAT from other formula proteins, but it is not effective for native AAT at the higher temperatures.

I. pH stability of human AAT. Native and recombinant AAT were diluted in 15 PBS or formula to 0.1 mg/mL. The sample volume was 1 mL, and the pH of each sample was adjusted drop-wise with 1 M HCl. The range of pHs tested was from pH 2 to 8 for the samples in PBS and pH 2 to 7 for samples in formula. After a 1 hour incubation at room temperature, the pH was restored to pH 7 with 1M NaHCO<sub>3</sub>. Formula samples were centrifuged as exemplified in above Thermal 20 Stability section.

Both native and recombinant AAT appear resistant to low pH conditions in both PBS and formula. There were no differences between treatment groups and controls for pH 3 through 7, and controls or between the native and recombinant AAT according to SDS-PAGE, Western blots, and trypsin assay. However, the 25 elastase assay and ELISA data show that recombinant AAT is more affected by acidic conditions than the native form. In PBS, native AAT was more than 95% intact, while about 60-80% of the recombinant AAT activity was intact. Infant formula may have a stabilizing effect on the recombinant protein, since it was found to be as stable as the native form according to ELISA and the Western blot.

30 Native and recombinant AAT can withstand acidic and digestive conditions as assessed by SDS-PAGE, Western blots, ELISA and activity assay. Native AAT regains much of its structural and functional stability after treatment at acidic conditions followed by neutralization, whereas recombinant AAT shows some loss of activity at a wide pH range, which may reflect a different glycosylation pattern.

The conditions of the infant-modeled digestion, pH 5 during pepsin treatment, are not ideal for pepsin, which normally possesses full activity at pH 2. AAT has been detected in human infant feces, which supports the notion that it is capable of surviving digestion *in vivo*, particularly during the first three months of the infant's life. This evidence also supports the validity of the *in vitro* digestion system. It is likely that AAT possesses enough resistance to acidic and digestive conditions to allow a significant amount to survive and affect the digestion process.

Recombinant AAT remained functionally intact after being exposed to low pH, *in vitro* digestion, and several types of heat treatment. It is therefore possible that recombinant AAT may be added to infant formula, can tolerate some processing conditions, and remain intact in the gastrointestinal tract of infants. Thus, recombinant AAT may help protect other physiologically active proteins, such as lactoferrin and lysozyme, which also may be added in recombinant forms in the gut of formula-fed infants. In conclusion, addition of recombinant AAT together with other recombinant proteins may enhance their bioactivity and make the formula more similar to human milk.

J. Expression of AAT in transgenic wheat. The plasmid API282 containing the Bx7 promoter, Bx7 signal peptide and AAT gene, Nos terminator and ampicillin resistance gene was used to transform wheat cells, substantially as in the transformation of rice cells. Twenty one transgenic lines were produced. Expression of AAT was determined to be about 5 to 12 µg per grain of wheat seeds.

25

## EXAMPLE 6

### Generation and Characterization of Recombinant Proteins Produced by Transgenic Rice Plants

#### A. Generation of recombinant antibodies

Recombinant antibodies have been expressed in transgenic plants (for examples, see Peeters *et al.*, 2001; Giddings *et al.*, 2000; Larrick *et al.*, 1998). However, expression and production of recombinant antibodies in the seeds of transgenic plants have certain advantages. The production of high levels of antibodies in grains, for example rice grains, provides distinct advantage that food

supplements may be prepared with little or no purification, and other advantages that are illustrated herein the patent application.

In one embodiment, an expression vector is constructed as illustrated in Example 1 that includes codon optimized nucleotide sequences encoding  
5 functional components of an antibody. For example, the components can be a heavy chain, a light chain, a linker region or a J chain and a secretory component. The expression vector may also include a promoter, a signal/target/transport sequence or sequences and a terminal sequence or sequences. Preferred  
10 promoter, signal/target/transport sequence and terminal sequence are exemplified herein. For example, for expression of each functional component of an antibody in rice seeds, a codon-optimized component gene is operably linked to the rice endosperm specific glutelin (Gt1) promoter, a Gt1 signal peptide and NOS terminator to form a component expression vector.

Each component expression vector is introduced to rice cells and plants to  
15 generate antibody component-expressing transgenic rice cells and plants, as exemplified in Example 2. In one embodiment, the expression vectors containing antibody heavy chain, light chain, linker region or J chain, and a secretory component can be introduced individually. The plants expressing each individual component can be crossed to generate plants that express a functional antibody.

20 In another embodiment, the expression vectors containing functional components of an antibody can be introduced to the plant at the same time, using the transformation methods exemplified in Example 2, such as by co-bombardment. A plant that expresses functional antibody is selected for further propagation.

25 In another embodiment, the expression vector containing codon optimized nucleotide sequence encoding a single chain antibody is introduced to rice cells and plants to generate antibody expressing transgenic rice cells and plants, as exemplified in Example 2. The nucleotide sequence encoding a single chain antibody can be constructed as conventional in the art, for example Kortt *et al.*,  
30 2001, Maynard and Georgiou, 2000; Humphreys DP and Glover, 2001.

The plant-generated recombinant antibody can be isolated and purified as exemplified in the patent application.

### B. Generation of human EGF

The Epidermal Growth Factor (EGF) gene was codon optimized as shown in Figure 20, and synthesized by Operon Technologies (CA, USA) with a SEQ ID NO: 8. The gene was cloned into pAPI145 and pAPI241 respectively, as  
5 exemplified in Example 1. The resulting plasmids were named API270 (Figure 21) and API303 (Figure 22), respectively. For expression of EGF in rice seeds, the codon optimized gene was operably linked to the rice endosperm specific glutelin (Gt1) promoter, Gt1 signal peptide and NOS terminator in pAPT303, and to the rice endosperm specific globulin (Glb) promoter, Glb signal peptide and  
10 NOS terminator in API270. The transgenic plant expressing EGF was generated, and plant-generated recombinant EGF was detected, as shown in Figure 23 and as exemplified herein.

### C. Generation of human IGF

15 The Insulin-like Growth Factor (IGF) gene was codon optimized as shown in Figure 24, and synthesized by Operon Technologies (CA, USA) with SEQ ID NO: 9. The gene was cloned into pAPI145 and pAPI241 respectively, as exemplified in Example 1. The resulting plasmids were named API271 (Figure 26) and API304 (Figure 25), respectively. For expression of IGF in rice seeds, the  
20 codon optimized gene was operably linked to the rice endosperm specific glutelin (Gt1) promoter, Gt1 signal peptide and NOS terminator in pAPI304, and to the rice endosperm specific globulin (Glb) promoter, Glb signal peptide and NOS terminator in API271. The transgenic plant expressing IGF was generated, and plant-generated recombinant IGF was detected as shown in Figure 27 and as  
25 exemplified herein.

### D. Generation of Other Expression Plasmids

Other expression plasmids for use in transforming plants herein for the production of recombinant polypeptides in transgenic plants were made  
30 substantially as previously described. These plasmids are shown in Figure 28, showing API321, containing a Glb promoter, a Gt1 signal peptide, codon-optimized haptocorrin gene, Nos terminator, and an ampicillin resistance gene; Figure 29, showing API320, containing a Gt1 promoter, a Gt1 signal peptide, codon-optimized human haptocorrin gene, Nos terminator, and an ampicillin

resistance gene; Figure 30, showing API292, containing a Glb promoter, a Glb signal peptide, kappa-casein gene, Nos terminator, and an ampicillin resistance gene; Figure 31, showing API297, containing a Gt1 promoter, a Gt1 signal peptide, a gene encoding mature kappa-casein polypeptide, Nos terminator, and an ampicillin resistance gene; Figure 32, showing API420, containing a Gt1 promoter, a Gt1 signal peptide, lactohedrin gene, Nos terminator, and a kanamycin resistance gene; Figure 33, showing API418, containing a Gt1 promoter, a Gt1 signal peptide, lactoperoxidase gene minus the sequence encoding the propeptide, Nos terminator, and a kanamycin resistance gene; Figure 34, showing API416, containing a rice Gt1 promoter, a Gt1 signal peptide, codon-optimized lactoperoxidase gene, Nos terminator, and a kanamycin resistance gene; and Figure 35, showing API230, containing a Bx7 promoter, a Gt1 signal peptide, codon-optimized lysozyme gene, Nos terminator, and an ampicillin resistance gene; Figure 36A, showing API254, containing a Glb promoter, a Glb signal peptide, lactoferrin gene, Nos terminator, and an ampicillin resistance gene; Figure 36B, showing API264, containing a Glb promoter, a Glb signal peptide, human lysozyme gene, Nos terminator, and an ampicillin resistance gene; Figure 37, showing API225, containing a GT3 promoter, a Gt1 signal peptide, codon-optimized lysozyme gene, Nos terminator, and an ampicillin resistance gene; and Figure 38, showing API229, containing a RP-6 promoter, a Gt1 signal peptide, codon-optimized lysozyme gene, Nos terminator, and an ampicillin resistance gene.

## EXAMPLE 7

### 25 Comparison of Promoter Activity in the Expression of Lysozyme in Transgenic Rice

#### A. Comparison between Gt1 and Glb promoters and signal peptides

In earlier studies, inconsistencies were observed between promoter activity of *Glb* and *Gt1* from transient assay data and the protein accumulation level in transgenic plants bearing the same promoters with signal peptides. These unpublished studies suggested that post-translational regulation was involved in recombinant protein expression and accumulation in the endosperm. It was unknown whether the storage protein signal peptide played a role in recombinant protein expression level or whether heterologous proteins could be sent to the

protein bodies along the sorting pathways of native storage proteins. In order to improve the expression level of recombinant proteins in cereal crop seed, it is important to understand recombinant protein targeting and trafficking in the endosperm expression system. Hence, comparison was made between the rice storage protein promoters and signal peptides from the Glutelin-1 gene ("*Gt1*") and the globulin gene ("*Glb*") showed that both promoters and both signal peptides were capable of effecting expression of lysozyme.

(i). Storage proteins

Rice endosperm contains four main storage proteins: acid-soluble glutelin, alcohol-soluble prolamin, water-soluble albumin and salt-soluble globulin (Juliano BO. Polysaccharides, proteins, and lipids of rice. Am. Assoc. Cereal Chem., St. Paul, MN (1985)). They are targeted into two types of protein bodies in rice endosperm. Prolamin aggregates within the endoplasmic reticulum ("ER") lumen into regularly shaped vacuole called protein body type I. The formation of these protein bodies is dependent on the chaperone BiP80 in the ER. Glutelin is deposited into protein storage vacuoles (PSV) via the Golgi apparatus into irregularly shaped vacuole called type II protein body. The components in the protein body type II and its sorting pathway are not well known. The targeting locations and sorting pathway of globulin and albumin also remain unknown. It appears that once the signal sequence is removed in the ER, the sorting and trafficking depend on the targeting information within the polypeptides and chaperones in the ER. The sorting signals are divided into three categories: sequence-specific vacuole sorting signals (ssVSS), C-terminal vacuole sorting signal (ctVSS), and physical structure vacuole sorting signals (psVSS), as described in Frigerio L. *et al.*, Plant Physiol. 126: 167-175 (2001); Matsuoka K. *et al.*, J. Exp. Botany 50: 165-174 (1999) and Vitale A. & Raikhel N.V., Trends in Plant Science 4: 149-155 (1999).

(ii). Method

Two promoters storage protein genes, *Gt1* and *Glb*, and the corresponding glutelin-1 and globulin signal peptide coding sequences were used to express the human lysozyme protein in developing endosperm. In the three plasmids, pAPI264, pAPI159 and pAPI228, the human lysozyme gene was fused with the nucleotide sequences of the *Glb* promoter and globulin signal peptide coding sequences, the *Gt1* promoter and glutelin signal peptide coding sequences and

the combination of the *Glb* promoter with the glutelin (*GT1*) signal peptide coding sequences, respectively (Figure 39A). Lysozyme amounts of T1 seeds were determined for 23 independently transformant lines of pAPI264, 10 lines of pAPI159 and 7 lines of pAPI159. The transgenic lines of pAPI159, which  
 5 synthesized lysozyme using the *Gt1* promoter and the glutelin signal peptide, produced the enzyme in amounts ranging from 34.25  $\mu\text{g}$  to 297.23  $\mu\text{g}\cdot\text{mg}^{-1}$  total soluble protein (TSP) with an average of 133.76  $\mu\text{g}\cdot\text{mg}^{-1}$  TSP. Plants transformed with pAPI264 carrying the *Glb* promoter and the globulin signal peptide yielded between 4.09 and 63.64  $\mu\text{g}\cdot\text{mg}^{-1}$  TSP lysozyme with an average of 33.96  $\mu\text{g}\cdot\text{mg}^{-1}$   
 10 TSP, while lines of pAPI228, which combined the *Glb* promoter and the glutelin signal peptide, yielded between 8.9 and 203.46  $\mu\text{g}\cdot\text{mg}^{-1}$  TSP with an average of 87.70  $\mu\text{g}\cdot\text{mg}^{-1}$  TSP.

The lysozyme expression amounts achieved with the *Gt1* promoter+GT1 signal peptide was 3.94 fold higher than that with the *Glb* promoter+GLB signal  
 15 peptide, while the expression amounts of lysozyme obtained with the *Glb* promoter+GT1 signal peptide was intermediate but increased 2.58 fold over that produced with the GLB signal peptide (Fig. 39B). Apparently the GT1 signal peptide is more efficient than the GLB signal peptide at lysozyme expression and deposition in rice endosperm. This demonstrates the importance of choosing an  
 20 optimal signal peptide for the production of recombinant proteins in developing rice grains.

### (iii). Chimeric Gene Components

Time course of human lysozyme expression during rice endosperm development. We monitored lysozyme accumulation during endosperm  
 25 development of transgenic lines 159-1-53-16-1 and 264-92-6. Immature spikelets were harvested at 7, 14, 21, 28, 35, 42 and 49 days after pollination ("DAP"). The lysozyme amounts in the endosperms were measured by the activity assay. Lysozyme accumulation in the seeds of transgenic plant 159-1-53-16-1 began at 7  
 DAP and peaked at 21 DAP. Thereafter lysozyme content decreased until 35  
 30 DAP and then stabilized until seed maturity (Figure 40). Lysozyme accumulation in developing seeds of the transgenic plant 264-92-6 likewise began at 7 DAP, peaked at 28 DAP, after which lysozyme content steadily decreased through seed maturation (Figure 40). These results show that lysozyme accumulation in the two types of transgenic lines during endosperm development follows the same

pattern as that of the native globulin and glutelin storage proteins.

(iv). Subcellular localization of human lysozyme in transgenic seeds

In order to determine whether the recombinant lysozyme was targeted to protein bodies in the endosperm, we investigated its subcellular localization by immunofluorescence microscopy. Transgenic plant 264-92-6 synthesizing lysozyme with the *Glb* promoter and globulin signal peptide and transgenic plant 159-1-53-16-1, producing human lysozyme with the *Gt-1* promoter and the glutelin signal peptide were analyzed. Dual localization with either native glutelin or globulin was used to determine the site of lysozyme deposition.

Synthetic peptides derived from the amino acid sequences of rice glutelin and globulin were used to raise antibodies in rabbits. Antibody specificity was confirmed with Western blots of endosperm proteins. No cross-reaction of glutelin and globulin antibodies with other endosperm proteins was detected with the host TP309 or the transgenic lines 264-92-6 and 159-1-53-16-1. The human lysozyme specific antibody detected the 13 kD of lysozyme protein exclusively in the fractionated and total protein extracts.

Immature seeds from two transgenic lines, 159-1-53-16-1(T4) and 264-92-6(T2) and untransformed control, TP309, were harvested at 14 DAP and fixed and a comparable analysis was conducted. In transgenic line 264-92-6, strong immunofluorescence signals of lysozyme and native proteins were detected with fully overlapping pattern, both when lysozyme and globulin or lysozyme and glutelin were compared (data not shown). Merging the two separately recorded images produces a yellow pseudo color signal. A scan for green and red wavelength emission across the 5 protein bodies along the white line in the not quite perfectly aligned images identifies the co-localization of human lysozyme with globulin. The orange tinge of the protein bodies is due to the stronger emission of red fluorescence than green. The perfect image merger provides a bright yellow color, and the recording of the green and red fluorescence emission along the white line identifies the same 5 protein bodies. These results demonstrated that lysozyme was colocalized in protein bodies with the native storage proteins. The results also demonstrated that the storage proteins globulin and glutelin are localized in the same cell compartment, substantiating the indication that globulin and glutelin are targeted into the same type II protein bodies in rice endosperm. We conclude that lysozyme contains all sorting

information for protein body targeting, at least when co-expressed with rice storage proteins.

The localization patterns of lysozyme and native storage proteins in 159-1-53-16-1 are, however, quite different and more complex than those of transgenic line 264-92-6. In transgenic 159-1-53-16-1, lysozyme does not completely colocalize with the native storage proteins. Globulin localized preferentially in smaller, peripheral protein bodies in the younger cells of the cortical region from 14 DAP endosperm, while lysozyme localized preferentially in irregularly shaped protein bodies. However, lysozyme did colocalize more evenly with globulin in the older cells of the central region from the developing endosperm. Merging the two separately scanned images visualized green fluorescing, lysozyme-rich type II protein bodies and red fluorescing, smaller, globulin-rich protein bodies.

Recording of the red and green fluorescence emission along the white scanning line reveals that there is almost twice as much lysozyme in the large type II protein bodies as there is in the small protein bodies, while the globulin signal in the small protein bodies is 2-3 times observed in type II protein bodies. Thus, there appears to be a preferential targeting of the two proteins. In the central region of the endosperm, a more equal co-localization of lysozyme and globulin was observed, especially in the larger type II protein bodies, when judged by the intensity of green and red fluorescence, which provides the yellow color upon merging the two images. This is evident from the merged image scan at the two emission wavelengths. However, there are also small protein bodies containing a dominant portion of globulin or lysozyme in these cells.

Distinct patterns were also found in 159-1-53-16-1 when anti-glutelin antibody was co-incubated with anti-lysozyme antibody in the younger cells of the cortical region from mid-developing endosperm. Like globulin, most of the glutelin localized in the smaller, peripheral protein bodies in younger cells, while lysozyme localized in irregularly shaped protein bodies. Lysozyme partially colocalized with glutelin in the older cells from the center region of mid-developing endosperm. Merging the two images and scanning for fluorescence-intensity at the two wavelengths reveals co-localization of the two proteins in the large and small protein bodies, some being highly enriched in lysozyme and others in glutelin. A comparable distribution is observed in the cells of the central part of the endosperm. The results suggested that lysozyme distorted the native storage

protein targeting/sorting when under the control of the *Gt1* promoter/GT signal peptide, producing high lysozyme expression, but not when under the control of *Glb* promoter/ GLB signal peptide with lower lysozyme expression.

To determine if native protein accumulation was affected in the endosperm of the transgenic plants, we analyzed the amounts of glutelin, globulin and lysozyme proteins from two transgenic lines and TP309 by Western blotting. The results showed that glutelin protein was reduced in 159-1-53-16-1, but was increased in 264-92-6 in comparison to TP309. Amounts of globulin protein were reduced in both 264-92-6 and 159-1-53-16-1. This change is particularly significant in the transgenic line 159-1-53-16-1 with its higher lysozyme expression level. The results showed that globulin was more affected than glutelin, no matter which signal peptide was used.

Based on the results, we conclude that lysozyme was targeted to the protein bodies and that the signal peptide played an important role in lysozyme expression. The plants with high expression levels of the recombinant protein showed distorted native protein expression and trafficking.

Thus, the combination of the *Gt1* promoter and *Gt1* signal peptide was more effective than the combination of the *Glb* promoter and *Glb* signal peptide, with the combination of *Glb* promoter and *Gt1* signal peptide having intermediate level of activity. Results showed that the high level expression of recombinant protein distorted the trafficking and sorting of the native storage proteins and affected the native storage protein expression. Results also indicated that mature human lysozyme protein contains a determinant recognized in the plant cell for the protein storage vacuole (PSV) sorting following signal peptide cleavage, and that the lysozyme was sorted to Type II protein bodies.

#### B. Comparison of seven promoters and *Gt1* signal peptide in regulating the expression of lysozyme

Plasmids API159 (*Gt1* promoter) (Figure 1), API228 (*Glb* promoter) (Figure 39), API230 (Figure 35), API229 (RP-6 promoter) (Figure 38), API225 (*GT3* promoter), a plasmid carry the *Glub-2* promoter, and another plasmid carrying the *Club-1* promoter, were compared in their ability to effect the expression of lysozyme in transgenic rice T1 seeds. Results shown in Figure 41 indicate that for

expression of lysozyme, Gt1 was the strongest promoter, followed by Glb, Glub-2, Bx7, Gt3, Glub-1 and Rp6, in order of promoter strength.

### EXAMPLE 8

#### 5 Co-transformation of Heterologous Polypeptide and Reg Gene in Transgenic Rice Plants

##### A. Enhanced Lysozyme Expression in Transgenic Rice Seed Co-Transformed with Reb

Codon-optimized human lysozyme gene was linked to *Glb* promoter and  
10 *Glb* signal peptide to generate plasmid *Glb-Lys* (API264) as shown in Figure 36B, which was used to transform rice with or without *Native-Reb*, as previously described and as described in WO 01/83792. Normal plant phenotypes were obtained among transformants containing *Glb-Lys* alone or with *Native-Reb*. To determine the presence of *Reb* gene and *Glb-Lys* in the transgenic rice genome,  
15 one primer designed from vector sequence and another designed from the *Reb* gene 3' terminator were used to identify these lines. In this case, only the recombinant *Reb* gene could be amplified. PCR analysis confirmed the presence of transgenes in the rice genome. Ten of 11 plants from independent transformation events contained both *Reb* and the lysozyme transgenes. The  
20 REB protein of immature seeds from five randomly selected transgenic lines was detected by Western blotting. The expression level of the REB protein in transgenic lines ranged from 25% to 71% higher than that in untransformed TP309. This demonstrated that the transgenic *Reb* gene was active in transgenic plants.

25 Seeds of confirmed transgenic rice plants were harvested at maturity, and the lysozyme activity was analyzed. As shown in FIG. 19, lysozyme expression in the seeds from 30 independent transformation events containing both the *Native-Reg* and the *Glb-Lys* ranged from 30.57 to 279.61  $\mu\text{g/mg TSP}$  with an average of  $125.75 \pm 68.65 \mu\text{g/mg TSP}$ . Seeds of 17 transgenic events containing the *Glb-Lys* gene alone expressed lysozyme in amounts ranging from 7 to 76  $\mu\text{g/mg TSP}$   
30 with an average of  $33.95 \pm 20.55 \mu\text{g/mg TSP}$ . No lysozyme activity was detected in untransformed rice seeds. The results showed that the expression level of lysozyme increased an average of 3.7-fold when seeds were transgenic for both the *Reb* gene and *Glb-Lys*. Statistical analysis (t test) showed that the amount of

lysozyme in seeds from the plants transgenic for the *Reb* gene and *Glb-Lys* is significantly higher than that in the plants transgenic for *Glb-Lys* alone ( $p < 0.001$ ).

5     B. Enhanced Human Lysozyme Expression in Transgenic Rice Seed Co-  
Transformed with Maize Transcriptional Factor, Prolamin-Box Binding Factor (PBF)

Three transcriptional factors were tested; rice endosperm bZIP protein (REB), Opaque2 (O2) and PBF. The transcriptional factors and human lysozyme gene under the control of rice glutelin 1 (Gt1) or globulin (Glb) promoter were co-  
10   bombarded into rice callus. Transgenic  $R_1$  grains carrying both genes were obtained. The effect of transcriptional factors on the expression of human lysozyme were monitored. Under the control of Glb promoter, REB increased Lys expression by about 3-fold. REB showed no effect on a stronger promoter, Gt1. Transcription factor increased Lys expression, but not significantly. PBF  
15   increased Lys expression on average 1.5-fold over *Gt1-Lys* alone. The highest Lys-expressing lines were selected and advanced to  $R_2$  generation in the greenhouse. As shown in Table 8 below, Lys expression level from an  $R_2$  line, 265/159-41-5, was about 190  $\mu\text{g}$  per grain and 9.5 mg/gram of brown rice flour (equivalent to 0.8% grain weight). The level of expression was about 1.5-fold  
20   higher than that of the highest expression line without the transcription factor. In addition, data showed that PBF not only increased the expression of Lys, but also increased the expression of native storage proteins such as glutelin and globulin, and the protein related to protein trafficking. It implies that PBF can act on the promoters of multiple genes to increase the expression of those proteins in rice  
25   endosperm.

Table 8.

| Line Number | R <sub>1</sub> <sup>*</sup> (μg/grain) | R <sub>2</sub> <sup>*</sup> (μg/grain) | R <sub>2</sub> Lysozyme<br>(mg/g brown rice) | Homozygous   |
|-------------|----------------------------------------|----------------------------------------|----------------------------------------------|--------------|
| 285/159-41  | 150.23                                 | 190.00                                 | 9.5                                          | homozygous   |
| 285/159-43  | 114.30                                 | 155.00                                 | 9.0                                          | homozygous   |
| 285/159-81  | 103.58                                 | 175.59                                 |                                              | heterozygous |
| 285/159-286 | 152.07                                 | 180.00                                 |                                              | heterozygous |

\* The expression data were averaged from 10 seeds in R<sub>1</sub> and from 10 lines in R<sub>2</sub>.

5

## EXAMPLE 9

Production of rice extract containing recombinant proteins and its use in foodA. General procedure for production of rice extract

Transgenic rice containing heterologous polypeptides can be converted to rice extracts by either a dry milling or wet milling process. In the dry milling process, transgenic paddy rice seeds containing the heterologous polypeptides, such as recombinant human lysozyme or lactoferrin were dehusked with a dehusker. The rice was grounded into a fine flour through a dry milling process, for example, in one experiment, at speed 3 of a model 91 Kitchen Mill from K-TEC. Phosphate buffered saline ("PBS"), containing 0.135 N NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.7 mM KH<sub>2</sub>PO<sub>4</sub>, at pH 7.4, with or without additional NaCl, such as 0.35 N NaCl, was added to the rice flour. In some experiments, approximately 10 ml of extraction buffer was used for each 1 g of flour. In other experiments, the initial flour/buffer ratio varied over a range such as 1 g/40 ml to 1 g/10 ml. The mixture was incubated at room temperature with gentle shaking for 1 hr. In other experiments, the incubation temperature was lower or higher, such as from about 22°C to about 60°C, and the incubation time was longer or shorter, such as from about 10 minutes to about 24 hr. A Thermolyne VariMix platform mixer set at high speed was used to keep the particulates suspended.

In place of PBS, other buffers were used in some experiments, such as ammonium bicarbonate. In one embodiment, 10 liters of 0.5M ammonium bicarbonate was added to 1 kg of rice flour.

The resulting homogenate was clarified either by filtration or centrifugation. For the filtration method, the mixture was allowed to settle for about 30 minutes at room temperature, after which the homogenate was collected and filtered. Filters

in three different configurations were purchased from Pall Gemansciences and used. They were: a 3  $\mu$ m pleated capsule, a 1.2  $\mu$ m serum capsule and a Suporcap capsule 50 (0.2  $\mu$ m). For centrifugation, a Beckman J2-HC centrifuge was used and the mixture was centrifuged at 30,000 g at 4°C for about 1 hr. The  
5 supernatant was kept and the pellet was discarded.

In one embodiment, the filtrate and supernatant were further processed, for example by ultra-filtration or dialysis or both to remove components such as lipids, sugars and salt.

The filtrate from the above filtration procedure, which is also called the  
10 clarified extract, was then concentrated using a spiral wound tangential flow filter operated in a batch recirculation mode. In one embodiment, PES (polyethersulfone) 3000-4000 molecular weight cutoff membranes was used for this step. These final concentrated extracts were held overnight in a cold room.

The concentrated extracts were next dried to a powder by lyophilization.  
15 During loading of the lyophilizer trays, the extracts were not subjected to a final 0.2 or 0.45 micron depth filtration to minimize loss of target proteins. The lyophilized material was scraped from the lyophilizer trays and combined into a plastic bag. The dry material was compressed by drawing a vacuum on the bag and then the material was blended and the particle size reduced by hand-  
20 kneading it through the plastic.

The lyophilized materials were then suitable for use as an extract directly or in admixture with other food. In one experiment, the lyophilized materials were blended with various ingredients to produce control and test infant-formula. The ingredients were blended using a Hobart mixer (140 quart size) equipped with a  
25 paddle agitator. These final blends were packed in 1 kg double Mylar bags and the headspace was filled with nitrogen before sealing.

Table 9 shows the recovery of recombinant human lactoferrin from 105 kg transgenic rice flour during each extraction step. The amount of recombinant human LF present was determined quantitatively as described in Example 4.

Table 9.

|                           | O/Lactoferrin  |          |
|---------------------------|----------------|----------|
| Stage of process          | Lac mass       | % of max |
| Baseline extraction yield | 4.0 mg/g flour |          |
| Expected maximum          | 420 g          | 100 %    |
| Initial extract           | 338 g          | 80       |
| Clarified extract         | 373 g          | 89       |
| Concentrated extract      | 343 g          | 82       |
| Dried extract             | 340 g          | 81       |

Rice extract can also be produced using a wet milling procedure.

Transgenic paddy rice seeds containing recombinant human lysozyme were re-hydrated for a period of 0 to 288 hrs at 30°C. The rehydrated seeds were ground in PBS extraction buffer. The initial seed/buffer ratio varied over a range such as 1 g/40 ml to 1 g/10 ml. Table 10 shows recovery of human lysozyme from rice seeds soaked from 0 to 288 hrs.

Table 10.

| Rehydration time (hrs) | Lysozyme (ug/grain) | Recovery (%) |
|------------------------|---------------------|--------------|
| 0                      | 87                  | 100          |
| 48                     | 69                  | 79           |
| 60                     | 79                  | 91           |
| 168                    | 60                  | 69           |
| 216                    | 56                  | 64           |
| 288                    | 58                  | 67           |

10

Over 60% human lysozyme was recovered from the wet milling process. The result of the wet milling becomes initial extract which may be stored frozen until use. The processing of initial extract to obtain dried extract was the same as that described for dry milling in this section.

15

#### B. Concentration and diafiltration of recombinant lysozyme and control rice extracts.

The conditions used in concentration and diafiltration varied depending on volume, speed, cost, etc. These conditions are all routine in the art based on the description herein. The frozen initial extract was thawed in the coldroom (about 2-8°C) for six hours. The thawed material were clarified through a 0.45µm filter and concentrated using a 5000 Nominal Molecular Weight Cutoff membrane of Polyethersulfone.

20

90 ml of the filtrate of control extract was concentrated to 10 ml and additional 10 ml of deionized water was added to the concentrated filtrate. The diluted filtrate was diafiltered one more time using water. The precipitate started forming at 16 mS and increased as the ionic strength decreased. 1M ammonium bicarbonate was added to the retentate to add ionic strength. The haze decreased although did not disappear completely. The material was diafiltered multiple times, in one embodiment three times, with water and multiple times, in one embodiment three times, with 0.1 M ammonium bicarbonate. It was concentrated to 9 ml and the membrane was rinsed with 0.1 M ammonium bicarbonate. The concentrate was filtered through several 0.2  $\mu$ m button filters. In one embodiment, 2.3 ml of the filtrate was lyophilized as is; 2.3 ml of the filtrate was diluted to 12 ml with deionized water and lyophilized, and 2.0 ml of the filtrate was diluted to 25 ml with deionized water and lyophilized. All remained clear.

A total of 89 ml of the filtrate of rHLys extract was concentrated to 10 ml, and additional 10 ml of 0.1 M ammonium bicarbonate was added. The resulting mixture was concentrated back to 10 ml and another 10 ml of 0.1 M ammonium bicarbonate was added. The retentate started to haze up. The material was diafiltered multiple times, in one embodiment three times, with 0.1 M ammonium bicarbonate. It was concentrated to 9 ml and the membrane was rinsed with 0.1 M ammonium bicarbonate. The concentrate was filtered through several 0.45  $\mu$ m button filters. In one embodiment, 2.0 ml of the filtrate was lyophilized as is; 2.0 ml of the filtrate was diluted to 12 ml with deionized water where a haze formed, and lyophilized, and 2.0 ml of the filtrate was diluted to 12 ml with 0.1 M ammonium bicarbonate which remained clear, and lyophilized.

25

#### C. Comparison of trial extraction of recombinant lysozyme rice with PBS and ammonium bicarbonate

The conditions used in concentration and diafiltration varied depending on volume, speed, cost, etc. These conditions are all routine in the art based on the description herein. rHlys rice flour was mixed with extraction buffer at about 100 g/L for about 1 hour using a magnetic stir bar. In one 2 liter beaker, the extraction buffer was PBS, pH7.4 plus 0.35 M NaCl. In another 2 liter beaker, the extraction buffer was 0.5 M ammonium bicarbonate. A 15 cm buchner was pre-coated with about 6g of Cel-pure C300 before adding another 20g of Cel-pure C300. The

30

mixture was filtered at about 3-4 Hg. It was then washed twice with about 100 ml of respective extraction buffer. The extracted filtrate was collected and concentrated with ultra-filtration cartridges: 5K Regenerate Cellulose, 5K PES, and 1K Regenerated Cellulose. The concentrates were lyophilized and analyzed for rHlys contents. The ammonium bicarbonate and PBS, pH7.4 plus 0.35 M NaCl both extracted approximately the same amount of rHlys. There was little loss of lysozyme units in the permeate with any of the ultrafiltration units that were used.

Other extraction buffer can also be used to extract recombinant proteins expressed in transgenic rice grains, for example Tris buffer, ammonium acetate, depending on applications. For example, for using recombinant human LF for iron supplement, iron may be added to the extraction buffer and the buffer is set at a pH so that the apo-LF can pick up iron during the extraction process. Under this condition, LF can become saturated with iron (holo-LF). In another example, a buffer lacking of iron and a pH resulting in iron release from LF is used to produce apo-LF.

#### D. Production of rice extracts containing recombinant proteins

The conditions used in concentration and diafiltration varied depending on volume, speed, cost, etc. These conditions are all routine in the art based on the description herein. All equipment was soaked in hot 0.1M NaOH at a starting temperature of about 55°C. Rice flour was added to an about 250-500 gal. stainless steel tank containing 0.5M ammonium bicarbonate at about 95-105 g/L. It was mixed for about 60-80 minutes at about 9°C.

12 plates of 36 inch filter press C300 were pre-coated with about 3-6 kg Cel-pure C300. About 19-26 g/L of Cel-pure was added to the extract and mixed thoroughly. The mixture was pressed at a pressure of about 22 psi at a flow rate of about 82 liters/minute. The filtrate was collected into a 250 gal. stainless steel tank and washed with 0.5M ammonium bicarbonate. The press was blown dry. The process was carried out at about 10°C.

The 300 NMW cut-off membranes (Polysulfone) which had been cleaned and stored with 0.1M NaOH after control run was rinsed thoroughly with deionized water. The extract was concentrated and bumped to a 100 gal stainless steel tank. The membrane and the concentration tank were flushed with 0.1M

ammonium bicarbonate to recover all of the products. The product were covered with plastic and left in the 100 gal tank overnight at room temperature. The concentrate was filtered through spiral wound 1  $\mu$ m filter and into 5 gal poly container. The concentrate was lyophilized. About 81% of lactoferrin and about  
 5 58% of lysozyme was recovered from transgenic rice grains, respectively.

E. Blending of rice extract containing recombinant proteins into infant formula

The three types of lyophilized dry extract that contains rice proteins  
 10 (control) or rice proteins with lysozyme or lactoferrin were combined with standard infant formula. The blending was done such the final infant formula contained about 1 gram lactoferrin and 0.1 gram lysozyme per liter of infant formula. The ingredients were blended using a Hobart mixer (140 quart size) equipped with a paddle agitator. These final blends were packed in 1 kg double Mylar bags and  
 15 the headspace was filled with nitrogen before sealing.

Samples of infant formula containing human lysozyme and lactoferrin were quantified using procedures described in Example 3 and 4.

Table 11 shows human lysozyme and lactoferrin in infant formula.  
 20 Table 11.

| Infant Formula               | Lactoferrin (mg/ ml) | Lysozyme (mg/ml) |
|------------------------------|----------------------|------------------|
| With control rice extract    | 0.0                  | 0.0              |
| With transgenic rice extract | 1.03                 | 0.13             |

Using extract as a delivery method of recombinant protein has clear advantages over the purified form or in the whole grain. The conventional approach, such as in the whole grain form, has limitations such as protein stability  
 25 during high temperature and pressure processing. Furthermore, the purification approach is expensive. Therefore the extract approach 1) maintains a low cost compared to purification approach; 2) requires much smaller volume, for example about 1-10% of whole grain weight; 3) increases the concentration of recombinant protein from about 0.05-0.5% in whole grain form to about 10 to 20 % in the  
 30 extract form. Some extract form even reaches 40% depending on the expression level of recombinant protein. Therefore, the extract approach will allow broader

- application of the recombinant proteins compared to the whole grain approach. In addition, the extract approach removes starch granule, which requires high gelling temperature, for example about 75°C. Consequently, the extract approach provides more flexibility in processing the rice grain and the recombinant proteins
- 5 into food and diet, and the alike, without worrying about using high temperature to denature starch granule. The undenatured starch granule cannot be digested by human gut without gelatinization by for example high temperature.

#### Brief Description of the Sequences

10

| Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | SEQ ID NO |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <p>Codon optimized lysozyme coding sequence:</p> <p>AAAGTCTTCGAGCGGTGCGAGCTGGCCCGCACGCTCAAGCGGCTCGGCAT<br/> GGACGGCTACCGGGGCGATCAGCCTCGCCAACTGGATGTGCCTCGCCAAGT<br/> GGGAGTCGGGCTACAACACCCGCGCAACCACTACAACGCCGGCGACCGC<br/> TCCACCGACTACGGCATCTTCCAGATCAACTCCCGCTACTGGTGCAACGAC<br/> GGCAAGACGCCCCGGGGCCGTCAACGCCTGCCACCTCTCCTGCTCGGCCCT<br/> GCTGCAAGACAACATCGCCGACGCCGTGCGGTGCGCGAAGCGCGTCGTCC<br/> GCGACCCGCGAGGGCATCCGGGCCCTGGGTGGCCTGGCGCAACCGCTGCCA<br/> GAACCGGGACGTGCGCCAGTACGTCCAGGGCTGCGGCGTCTGA</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 1         |
| <p>Amino acid sequence based on codon optimized lysozyme coding sequence:</p> <p>KVFERCELARTLRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRST<br/> DYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGI<br/> RAWVAWRNRCQNRDVRQYVQCGV</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 2         |
| <p>Codon optimized lactoferrin coding sequence:</p> <p>GGGCGGCGGCGGCGCTCGGTGCAGTGGTGCGCCGTGTCCCAGC<br/> CCGAGGCGACCAAGTGCTTCCAGTGGCAGCGCAACATGCGGAAGGTGCGC<br/> GGCCCCGCCGGTCAGCTGCATCAAGCGGGACTCCCCATCCAATGCATCCAG<br/> GCCATCGCGGAGAACCAGCGCCGACGCGGTACCCCTGGACGGCGGGTTCAT<br/> CTACGAGGCGGGGCTCGCCCCGTACAAGCTCCGCCGGTGGCGGCGGAG<br/> GTGTACGGCACCGAGCGCCAGCCGCGCAGCACTACTACGCGGTGGCCGT<br/> CGTCAAGAAGGGCGGGTCTTCCAGCTCAACGAGCTGCAGGGCCTGAAGT<br/> CGTGCCACACGGGCCTCCGGCGGACGGCGGGCTGGAACGTGCCCATCGG<br/> CACCTGCGCCCCCTTCTGAAGTGGACCGGCCCGCCGGAGCCGATCGAGG<br/> CCGCCGTGGCCCGCTTCTCAGCGCCTCCTGCGTCCCCGGCGCCGACAAG<br/> GGCCAGTTCCCGAACCTCTGCCGGCTCTGCGCCGGGACGGGCGAGAACA<br/> GTGCGCCTTCTCCTCGCAGGAGCCGTACTTCTCCTACTCGGGCGCGTTCAA<br/> GTGCCTCCGCGACGGGGCCGGCGACGTGGCGTTTCATCCGCGAGTCCACCG<br/> TGTTTCGAGGACCTCTCCGACGAGGCGGAGCGGGACGAGTACGAGCTGCTG<br/> TGCCCCGACAACACCCGCAAGCCGGTGGACAAGTTCAAGGACTGCCACCTG<br/> GCGCGGGTGGCCTCGCACGCGGTGTCGCCCCGACGCTCAACGGCAAGGA<br/> GGACGCGATCTGGAACCTCCTCCGCCAGGCCAGGAGAAGTTTCGGCAAGG<br/> ACAAGTCCCCCAAGTTCCAGCTTTCGGGAGCCCCAGCGGCCAGAAGGACC<br/> TCCTCTTCAAGGACTCCGCGATCGGCTTCTCCCGCGTCCCCCGCGCATCG<br/> ACTCCGGCCTGTACCTCGGCTCCGGGTACTTCACCGCGATCCAGAACCTCC<br/> GGAAGAGCGAGGAGGAGGTGGCGGCGCGGCGGGCCCGCGTCTGTGGTG<br/> CGCCGTGGGCGAGCAGGAGCTGCGGAAGTGCAACAGTGGAGCGGCCTGA</p> | 3         |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| GCGAGGGGTGCGGTGACCTGCTCGTCCGCCAGCACCACCGAGGACTGCATC<br>GCGCTCGTCTCAAGGGGGAGGCCGACGCGATGAGCCTCGACGGGGGGTA<br>CGTCTACACCGCCGGCAAGTGCGGCCTGGTCCCGGTCTGGCGGAGAACT<br>ACAAGTCGCAGCAGTCCAGCGACCCGACCCGAACTGCGTGGACCGCCCC<br>GTCGAGGGGTACCTCGCCGTGGCCGTCTGCGCCGGTCCGACACCTCCCT<br>GACGTGGAACAGCGTCAAGGGCAAGAAGAGCTGCCACACCGCCGTGGACC<br>GCACCGCCGGCTGGAACATCCCGATGGGCCTCCTCTTCAACCAGACCGGCT<br>CCTGCAAGTTCGACGAGTACTTCTCCAGTCTGCGCCCCCGGCTCGGACC<br>CCCGCTCCAACCTGTGCGCCCTCTGCATCGGGGACGAGCAGGGCGAGAAC<br>AAGTGCGTGCCCAACAGCAACGAGCGGTACTACGGCTACACGGGGGCCTT<br>CCGCTGCCTGGCGGAGAACGCCGGGGACGTGCGGTTCTGTAAGGACGTGA<br>CCGTGCTGCAAAACACGGACGGGAACAACAACGAGGCGTGGGCGAAGGAC<br>CTCAAGCTCGCCGACTTCGCCCTGCTGTGCCTCGACGGCAAGCGCAAGCCC<br>GTCACCGAGGCGCGGTCTGCCACCTGGCGATGGCCCCCAACCACGCCGT<br>CGTCTCCCGCATGGACAAGGTCGAGCGCCTCAAGCAGGTGCTCCTGCACCA<br>GCAGGCCAAGTTCGGCCGGAACGGCAGCGACTGCCCGGACAAGTCTGCC<br>TGTTCCAGTCGGAGACCAAGAACCTCCTCTTCAACGACAACACCGAGTGCCT<br>GGCGCGCCTCCACGGCAAGACCACCTACGAGAAGTACCTCGGCCCGCAGT<br>ACGTCGCCCGGCATCACCAACCTCAAGAAGTGCTCCACCTCCCCCTCCTGG<br>AGGCGTGCGAGTTCCTCCGCAAGTGA |    |
| Amino acid sequence based on codon optimized lactoferrin coding sequence:<br>GRRRSVQWCAVSQPEATKCFQWQRNMRKVRGPPVSCIKRDSPIQCIQIAIEN<br>RADAVTLDDGGFIYEAGLAPYKLRPVAAEVYGTERRQPRTHYYAVAVVKKGGSFQL<br>NELQGLKSCHTGLRRTAGWNVPIGTLRPFLNWTGPPEPIEAARFFSASCVP<br>ADKGQFPNLCRLCAGTGENKCAFSSQEPYFSYSGAFKCLRDGAGDVAFIGEST<br>VFEDLSDEAERDEYELLCPDNTRKPVDFKDKCHLARVPSHAVVARSVNGKEDAI<br>WNLRLQAQEFKDKSPKFQFSGPSGQKDLLFKDSAIGFSRVPPRIDSGLYLG<br>SGYFTAIQNLKSEEEVAARRARVWCAVGEQELRKNQWSGLSEGSVTCSS<br>ASTTEDICIALVLKGEADAMSLDGGYVYTAGKCLVPVLAENYKSQSSDPDPN<br>CVDPRVEGYLAVAVRRSDTSLTWNSVKGKKSCHTAVDRTAGWNIPMGLLFNQ<br>TGSCKFDEYFSQSCAPGSDPRSNLCALCIGDEQGENKCVPSNERYYYGYTGAF<br>RCLAENAGDVAFVKDVTVLQNTDGNNEAWAKDLKLADFALLCLDGKRKPVTE<br>ARSCHLAMAPNHAVVSRMDKVERLKQVLLHQAKFGRNGSDCPDKFLFQSE<br>TKNLLFNDNTECLARLHGKTTYEKYLGPPYVAGITNLKKCSTSPLEACEFLRK                                                                                                                                                                                                                                                          | 4  |
| MV-Gt1-F1 primer:<br>5' ATC GAA GCT TCA TGA GTA ATG TGT GAG CAT TAT GGG ACC ACG 3'                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 5  |
| Xba-Gt1-R1 primer: 5' CTA GTC TAG ACT CGA GCC ATG GGG CCG GCT<br>AGG GAG CCA TCG CAC AAG AGG AA 3'                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 6  |
| Codon optimized lactoferrin coding sequence<br>ACCAAGTGCTTCCAGTGGCAGCGCAACATGCGGAAGGTGCGCGGCCCGCC<br>GGTCAGCTGCATCAAGCGGGAC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 7  |
| Codon optimized EGF coding sequence<br>AACTCCGACTCGGAGTGCCCCCTCTCCACGACGGTTACTGCCTCCACGAC<br>GGGGTCTGCATGTACATCGAGGCCCTCGACAAGTACGCCTGCAACTGCGTC<br>GTGGGCTACATCGGCGAGCGGTGCCAGTACCGCGACCTCAAGTGGTGGGA<br>GCTGCGCTGA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 8  |
| Codon optimized IGF-1 coding sequence<br>GGCCCGGAGACCCTCTGCGGCGCCGAGCTCGTGGACGCCCTCCAGTTCGT<br>GTGCGGCGACCGCGGCTTCTACTTCAACAAGCCGACCGGCTACGGCAGCA<br>GCAGCCGCCGCGCCCCGACACCGGCATCGTGGACGAGTGCTGCTTCCGC<br>AGCTGCGACCTCCGCCGCCTGGAGATGACTGCGCCCCGCTCAAGCCCGC<br>CAAGAGCGCCTGA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 9  |
| Codon optimized lactohedrin coding sequence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 10 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <p>CTGGACATCTGCTCGAAGAACCCGTGCCACAACGGCGGGCTCTGCGAGGA<br/> GATCAGCCAGGAGGTGCGGGGCGACGTGTTCCCTCGTACACCTGCACCT<br/> GCCTGAAGGGGTACGCCGGGAACCACTGCGAGACGAAGTGCGTGGAGCCC<br/> CTGGGGATGGAGAACGGCAACATCGCCAACTCCCAGATCGCCGCCTCCTCC<br/> GTGCGGGTGACCTTCCTCGGCCTCCAGCACTGGGTCCCGGAGCTGGCCCCG<br/> GCTCAACCGGGCGGGCATGGTGAACGCGTGGACCCCTCGTCCAACGACG<br/> ACAACCGTGGATCCAAGTGAACCTGCTCCGCCGATGTGGGTACCCGGCG<br/> TGGTCACCCAAGGCGCCAGCCGCCTGGCCAGCCACGAGTACCTCAAGGCC<br/> TTCAAGGTGCGCTACAGCCTCAACGGCCACGAGTTCGACTTCATCCACGAC<br/> GTCAACAAGAAGCACAAGGAGTTCTGTTGGGCAACTGGAACAAGAACGCGGT<br/> CACGTGAACCTCTTCGAGACCCCCGTGAGAGGCCAGTACGTCCGCCTCTAC<br/> CCCACGAGCTGCCACACCGCCTGCACGCTCCGCTTCGAGCTGCTGGGGTG<br/> CGAGCTGAACGGGTGCGCGAACCCGCTGGGGCTCAAGAACAACAGCATCC<br/> CCGACAAGCAGATCACGGCCTCGTCGTGTACAAGACCTGGGGCCTGCACC<br/> TCTTCTCGTGAACCCGAGCTACGCCCGGCTGGACAAGCAGGGCAACTTCA<br/> ACGCCTGGGTGCGCGGGAGCTACGGGAACGACCAGTGGCTCCAGGTGGAC<br/> CTCGGCAGCTCCAAGGAGGTACCCGGCATCATCACGAGGGGGGCCGCAA<br/> CTTCGGCTCCGTGCAGTTCGTGGCCTCCTACAAGGTGGCCTACTCGAACGA<br/> CAGCGCCAACCTGGACCGAGTACCAGGACCCGCGCACCGGGTCCAGCAAGA<br/> TCTTCCCCGGCAACTGGGACAACCACAGCCACAAGAAGAACCTGTTTCGAGA<br/> CCCCATCCTCGCCCGGTACGTCCGCATCCTCCCCGTGCTTGGCACAACC<br/> GGATCGCGCTCCGGCTGGAGCTCCTCGGCTGTGA</p>                                                                      |    |
| <p>Codon optimized kappa-casein coding sequence<br/> GAGGTCCAAAACCAGAAGCAGCCCGCCTGCCACGAGAACGACGAGCGCCC<br/> CTTCTACCAGAAGACCGCACCCCTACGTCCCGATGTACTACGTCCCGAACAG<br/> CTACCCCTACTACGGTACGAACCTGTACCAGCGCCCGCCGGCCATCGCTAT<br/> CAACAACCCCTACGTCCCCCGGACCTACTACGCGAACC CGCGCTGGTGC<br/> GGCCCCACGCGCAGATCCCGCAGCGGCAGTACCTGCCAAACAGCCACCCC<br/> CCCACCGTGGTGCGGCGGCCCAACCTCCACCCGAGCTTCATCGCTATCCCC<br/> CCCAAGAAGATCCAGGACAAGATCATCCTCCGACCATCAACACCATCGCCA<br/> CCGTGGAGCCGACGCCAGCCCCCGCGACCGAGCCACGGTGGACAGCGT<br/> CGTGACCCAGAGGCGTTCTCCGAATCGATCATCACCTCCACCCCCGAGAC<br/> CACCACGGTGGCCGTACGCGCGCGACGGCATGA</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 11 |
| <p>Codon optimized haptocorrin coding sequence<br/> GAGATCTGCGAGGTCTCCGAGGAGAACTACATCCGCCTCAAGCCCCTCCTG<br/> AACACCATGATCCAGAGCAACTACAACCGGGGCACGTGCGCCGTGAACGTC<br/> GTGCTCTCCCTGAAGCTCGTGGGCATCCAGATCCAGACCCTCATGCAGAAG<br/> ATGATCCAGCAGATCAAGTACAACGTGAAGAGCCGCCTCTCGGACGTGTCC<br/> AGCGGCGAGCTGGCGCTCATCATCCTCGCGCTCGGCGTGTGCCGGAACGC<br/> GGAGGAGAACCTCATCTACGACTACCACCTACGGACAAGCTGGAGAACAA<br/> GTTCCAGGCCGAGATCGAGAACATGGAGGCCACAAACGGCACCCCCGCTGA<br/> CCAACTACTACAGCTCAGCCTGGACGTCTCGCGCTCTGCCTGTTCAACG<br/> GAACTACTCCACCGCCGAGGTGGTCAACCACTTCACCCCGAGAACAAGA<br/> ACTACTACTTCGGCTCGCAGTTCTCCGTGGACACCGGGGCCATGGCCGTCC<br/> TGGCCCTCACCTGCGTGAAGAAGTCCCTCATCAACGGCCAGATCAAGGCCG<br/> ACGAGGGCTCCCTGAAGAACATCTCGATCTACACCAAGAGCCTCGTGGAGA<br/> AGATCCTCAGCGAGAAGAAGGAGAACGGGCTGATCGGCAACACCTTCTCGA<br/> CCGGCGAGGCGATGCAGGCCCTGTTCTGTGAGCAGCGACTACTACAACGAG<br/> AACGACTGGAACCTGCCAGCAGACCCTCAACACGGTCCTGACCGAGATCAGC<br/> CAGGGCGCGTTTCAGCAACCCCAACGCCCGCCGAGGTCTGCGGCCCT<br/> GATGGGCAAGACCTTCTCGACATCAACAAGGACAGCTCCTGCGTGTCCGC<br/> GAGCGGCAACTTCAACATCTCCGCCGACGAGCCGATCACGGTGACGCCGC<br/> CCGACAGCCAGTCGTACATCTCCGTGAACCTACAGCGTGCGGATCAACGAGA<br/> CCTACTTCACGAACGTGACGGTCTCAACGGCTCGGTCTTCTGAGCGTGA<br/> TGGAGAAGGCGCAGAAGATGAACGACACGATCTTCGGCTTACGATGGAGG<br/> AGCGCAGCTGGGGCCCCTACATCACCTGCATCCAGGGCCTCTGCGCCAACA</p> | 12 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| ACAACGACCGCACCTACTGGGAGCTGCTGAGCGGCGGCGAGCCGCTGAGC<br>CAGGGGGCCGCGCAGCTACGTGGTCCGCAACGGCGAGAACCTGGAGGTCCG<br>GTGGAGCAAGTACTGA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |    |
| <p>Codon optimized lactoperoxidase coding sequence</p> <p>CAAACGACCCGGACGTCCGGCGATCTCCGACACGGTCTCGCAGGCCAAGGT<br/>GCAAGTCAACAAGGCATTCTGGATTTCGCGCACGCGGCTGAAGACCGCGAT<br/>GTCGTCCGAGACCCCGACGAGCCGGCAGCTGAGCGAGTACCTCAAGCACG<br/>CGAAGGGGCGGACGCGCACCGCCATCCGCAATGGCCAAGTGTGGGAGGAA<br/>TCCCTGAAGCGGCTGCGGCAGAAGGCGTCGCTACCAACGTGACCGACCC<br/>GTCCCTCGACCTGACCAGCCTCTCCCTGGAGGTGCGCTGCGGCGCCCCGG<br/>CGCCCGTCTGTGCGCTGCGACCCCTGCTCGCCATACCGCACGATCACGGGC<br/>GACTGCAACAACCGGCGGAAGCCGGCACTGGGGGCTGCGAACCGCGCCCT<br/>CGCGCGCTGGCTCCCCGCGGAGTACGAGGACGGCCTCAGCCTCCCCTTCG<br/>GTTGGACCCCGGCAAGACGCGCAACGGCTTCCCGCTCCCGCTCGCTCGC<br/>GAGGTGAGCAACAAGATCGTCGGTTACCTGAACGAGGAGGGGGTCTCTGAC<br/>CAAAACCGCTCCCTCCTCTTCATGCAGTGGGGGCGAGATCGTGGACCACGAC<br/>CTGGACTTCGCCCCGGACACGGAGCTGGGCTCCAGCGAGTACAGCAAGAC<br/>CCAGTGCAGCAATACTGCATCCAGGGCGACAACCTGCTTCCCGATCATGTT<br/>CCCCCGAACGACCCGAAGGCGGGCACCCAGGGCAAGTGCATGCCGTTCT<br/>TCCGGGCGAGGCTTCGTCTGCCCGACCCCCCGTACAAGTCCCTCGCGCGC<br/>GAGCAGATCAACGCGCTCACGTCTTCTCGACGCCAGCTTCGTCTACAGC<br/>AGCGAGCCGTCCCTCGCCAGCCGCTCCGCAACCTCAGCAGCCCCCTCGG<br/>CCTCATGGCGGTCAACCAGGAGGTGTGCGACCACGGCCTCCCATACCTGCC<br/>GTACGACAGCAAGAAGCCGTCCCCCTGCGAGTTCATCAACACCACCGCGCG<br/>CGTCCCCTGCTTCCCTCGCCGGCGATTTCGCGGGCGAGCGAGCACATCCTCC<br/>TCGCCACGAGCACACCCTGTTCTCCGCGAGCACAACCGCCTCGCCCGG<br/>GAGCTGAAGCGCCTCAACCCGCAAGTGGGACGGCGAGAAGCTCTACCAGGA<br/>GGCCCGGAAGATCCTCGGCGCTTTCGTCCAGATCATCACCTTCCGGGACTA<br/>CCTCCCCATCCTGCTCGGTGACCACATGCAGAAGTGGATCCCCCCTACCA<br/>AGGCTACTCCGAGAGCGTGGACCCGCGCATCTCCAACGTCTTCACGTTTCGC<br/>GTTCCGCTTCGGGACCTGGAGGTGCCGTGCTCGATGTTCCGCCTCGACGA<br/>GAACTACCAGCCCTGGGGCCAGAGCCGGAGCTGCCGCTCCACACCCTGT<br/>TCTTCAACACCTGGCGGATGGTCAAGGACGGCGGCATCGACCCGCTCGTG<br/>GCGGGCTCCTGGCTAAGAAGTCAAGCTCATGAAGCAGAACAAAGATGATGA<br/>CCGGCGAGCTGCGCAACAAGCTGTTCCAGCCCACCCACCGCATCCACGGG<br/>TTCGACCTGGCTGCAATCAACACCCAGCGGTGCCGCGACCAACGGCCAGCC<br/>CGGCTACAACCTCGTGGCGCGCGTTCTGCGACCTCTCCCAGCCACAGACGCT<br/>GGAGGAGCTCAACACCGTGCTCAAGAGCAAGATGCTCGCCAAGAAGCTGCT<br/>CGGGCTCTACGGCACGCCCCGACAACATCGACATCTGGATCGGGGCCATCG<br/>CGGAGCCGCTCGTGGAGCGCGGGCGCGTCCGGCCGCTGCTCGCGTGCCT<br/>CCTGGGCAAGCAATTCCAACAGATCCGCGACGGGGACCGGTTCTGGTGGG<br/>AGAACCCCGGCGTGTTACCAACGAGCAGAAGGATTGCTCCAAAAGATGA<br/>GCTTCTCCCGCCTGGTGTGCGACAACCCGCGATACCAAGGTCCCGCGCG<br/>ACCCATTCTGGGCCAACTCCTACCCGTACGACTTCGTGGACTGCTCCGCCA<br/>TCGACAAGCTCGACCTGTCCCCCTGGGCATCGGTGAAGAACTGA</p> | 13 |
| <p>Codon optimized alpha-1-antitrypsin coding sequence</p> <p>GAGGACCCGCGAGGGCGACGCGCCGAGGAAGACCGACACCAGCCACCACGA<br/>CCAGGACCACCCGACGTTCAACAAGATCACCCCGAATTTGGCCGAATTCGC<br/>CTTCAGCCTGTACCGCCAGCTCGCGCACCAAGTCCAACCTCCACCAACATCTT<br/>TTCAGCCCGGTGAGCATCGCCACCGCCTTCGCCATGCTGTCCCTGGGTACC<br/>AAGGCGGACACCCACGACGAGATCCTCGAAGGGCTGAACCTCAACCTGACG<br/>GAGATCCCGGAGGGCGCAGATCCACGAGGGCTTCAGGAGCTGCTCAGGAC<br/>GCTCAACCAGCCGGACTCCCAGCTCCAGCTCACCAACGGGAACGGGCTCTT<br/>CCTGTCCGAGGGCCTCAAGCTCGTCGATAAGTTCTGGAGGACGTGAAGAA<br/>GCTCTACCACTCCGAGGGCTTACCGTCAACTTCGGGGACACCGAGGAGGC<br/>CAAGAAGCAGATCAACGACTACGTGAGAAGGGGACCCAGGGCAAGATCGT</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 14 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <p>GGACCTGGTCAAGGAATTGGACAGGGACACCGTCTTCGCGCTCGTCAACTA<br/>CATCTTCTTCAAGGGCAAGTGGGAGCGCCCGTTTCGAGGTGAAGGACACCGA<br/>GGAGGAGGACTTCCACGTCGACCAGGTCAACACCGTCAAGGTCCCGATGAT<br/>GAAGAGGCTCGGCATGTTCAACATCCAGCACTGCAAGAAGCTCTCCAGCTG<br/>GGTGCTCCTCATGAAGTACCTGGGGAACGCCACCGCCATCTTCTCCTGCC<br/>GGACGAGGGCAAGCTCCAGCACCTGGAGAACGAGCTGACGCACGACATCA<br/>TCACGAAGTTCTGGAGAACGAGGACAGGCGCTCCGCTAGCCTCCACCTCC<br/>CGAAGCTGAGCATCACCGGCACGTACGACCTGAAGAGCGTGCTGGGCCAG<br/>CTGGGCATCAGGAAGGTCTTCAGCAACGGCGCGGACCTCTCCGGCGTGAC<br/>GGAGGAGGCCCCCTGAAGCTCTCCAAGGCCGTGCACAAGGCGGTGCTCA<br/>CGATCGACGAGAAGGGGACGGAAGCTGCCGGGGCCATGTTCTGGAGGCC<br/>ATCCCCGTGTCCATCCCGCCCGAGGTCAAGTTCAACAAGCCCTTCGTCTTCC<br/>TGATGATCGAGCAGAACACGAAGAGCCCCCTTTCATGGGGAAGGTCGTCA<br/>ACCCACGCAGAAGTGA<br/>Codon optimized immunoglobulin-A coding sequence</p>                                                                                                                                                                                                                                                                                                                                                                                                                          |    |
| <p>Rice Gt1 promoter and Gt1 leader coding sequence<br/>CATGAGTAATGTGTGAGCATTATGGGACCACGAAATAAAAAGAACATTTTGAT<br/>GAGTCGTGTATCCTCGATGAGCCTCAAAAGTTCTCTACCCCGGATAAGAAA<br/>CCCTTAAGCAATGTGCAAAGTTTGCATTCTCCACTGACATAATGCAAAATAAG<br/>ATATCATCGATGACATAGCAACTCATGCATCATATCATGCCTCTCTCAACCTA<br/>TTCATTCTACTCATCTACATAAGTATCTTCAGCTAAATGTTAGAACATAAACC<br/>CATAAGTCACGTTTGATGAGTATTAGGCGTGACACATGACAAATCACAGAT<br/>CAAGCAAGATAAAGCAAAATGATGTGTACATAAAACTCCAGAGCTATATGTCA<br/>TATTGCAAAAAGAGGAGAGCTTATAAGACAAGGCATGACTCACAAAATTCA<br/>CTTGCCCTTTCGTGTCAAAAAGAGGAGGGCTTTACATTATCCATGTCAATTGC<br/>AAAAGAAAGAGAGAAAGAACAACACAATGCTGCGTCAATTATACATATCTGTA<br/>TGTCATCATTATTATCCACCTTTCGTGTACCACACTTCATATATCATAAGA<br/>GTCACCTTCACGTCTGGACATTAACAAACTCTATCTTAACATTTAGATGCAAGA<br/>GCCTTTATCTCACTATAAATGCACGATGATTTCTCATTGTTTCTCACAAAAG<br/>CGGCCGCTTCATTAGTCCTACAACAACATGGCATCCATAAATCGCCCCATAG<br/>TTTTCTTCACAGTTTGCTTGTTCTCTTGTGCGATGGCTCCCTAGCC<br/>Codon optimized alpha-1-antitrypsin coding sequence</p>                                                                                                                                                                                                                                                 | 15 |
| <p>Rice G1b promoter and Gt1 leader coding sequence<br/>CTGCAGGGAGGAGAGGGGAGAGATGGTGAGAGAGGAGGAAGAAGAGGAG<br/>GGGTGACAATGATATGTGGGGCATGTGGGCACCCAATTTTTTAATTCATTCT<br/>TTTGTTGAACTGACATGTGGGTCCCATGAGATTTATTATTTTCGGATCGAA<br/>TCGCCACGTAAGCGCTACGTCAATGCTACGTACAGATGAAGACCGAGTCAAT<br/>TAGCCACGTAAGCGCCACGTACGCCAAAACCACCATCCAACCGCCGAGGG<br/>ACCTCATCTGCACTGGTTTTGATAGTTGAGGGACCCGTTGTATCTGGTTTTTC<br/>GATTGAAGGACGAAAATCAAATTTGTTGACAAGTTAAGGGACCTTAAATGAA<br/>CTTATTCCATTTCAAAATATTCTGTGAGCCATATATACCGTGGGCTTCCAATC<br/>CTCCTCAAATTAAGGGCCTTTTTAAATAGATAATTGCCTTCTTTCAGTCAC<br/>CCATAAAAGTACAAAACCTACTACCAACAAGCAACATGCGCAGTTACACACATT<br/>TTCTGCACATTTCCGCCACGTACAAAAGAGCTAAGAGTTATCCCTAGGACAA<br/>TCTCATTAGTGTAGATACATCCATTAATCTTTTATCAGAGGCAAACGTAAAGC<br/>CGCTCTTTATGACAAAAATAGGTGACACAAAAGTGTTATCTGCCACATACATA<br/>ACTTCAGAAATTACCCAACACCAAGAGAAAAATAAAAAAAATCTTTTTGCAA<br/>GCTCCAAATCTTGGAAACCTTTTTCACTCTTTCAGCATTGTACTCTTGCTCT<br/>TTTTCCAACCGATCCATGTCACCCTCAAGCTTCTACTTGATCTACACGAAGCT<br/>CACCGTGACACAACCATGGCCACAAAACCTATAAAACCCCATCCGATCG<br/>CCATCATCTCATCATCAGTTCATTACCAACAAAACAAAAGAGGAAAAAACAT<br/>ATACACTTCTAGTGATTGTCTGATTGATCATCAATCTAGAGGGCGGCCGATG<br/>GCTAGCAAGGTCGTCTTCTTCGCGGCGGCGCTCATGGCGGCCATGGTGGC<br/>CATCTCCGGC</p> | 16 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Rice Gt1 promoter and Gt1 leader coding sequence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |     |
| Other monocot maturation specific promoter sequences                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 17- |
| Seq #17 Bx7 promoter seq                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 23  |
| <p>CTGCAGGCCAGGGAAAGACAATGGACATGCAAAGAGGTAGGGGCAGGGAA<br/> GAAACACTTGGAGATCATAGAAGAACATAAGAGGTTAAACATAGGAGGGCAT<br/> AATGGACAATTAATCTACATTAATTGAACTCATTGGGAAGTAAACAAAATC<br/> CATATTCTGGTGTAATCAAACATTTGACGCGGATTACTAAGATCCTATGT<br/> TAATTTTAGACATGACTGGCCAAAGGTTTCAGTTAGTTCATTTGTCACGGAAA<br/> GGTGTTCATAGTCCAAACTCTACCAACTTTTTGACGTCATAGCATAG<br/> ATAGATGTTGTGAGTCATTGGATAGATATTGTGAGTCAGCATGGATTTGTGT<br/> GCCTGGAATCCAATAATGACAAGCAACAAACCTGAAATGGGCTTTAGG<br/> AGAGATGGTTTATCAATTTACATGTTCCATGCAGGCTACCTTCCACTACTCGA<br/> CATGGTTAGAAGTTTGTAGTGCCGCATATTTGCGGAAGCAATGGCACTACTC<br/> GACATGGTTAGAAGTTTGTAGTGCCGCATATTTGCGGAAGCAATGGCTAACA<br/> GATACATATTCTGCCAAACCCCAAGAAGGATAATCACTCCTCTTAGATAAAAA<br/> GAACAGACCAATGTACAAACATCCACACTTCTGCAAACAATACACCAGAACT<br/> AGGATTAAGCCCATTACGTGGCTTTAGCAGACCGTCCAAAAATCTGTTTTGC<br/> AAGCACCAATTGCTCCTTACTTATCCAGCTTCTTTTGTGTTGGCAAACCTGCCC<br/> TTTTCCAACCGATTTTGTTCCTTCTCACGCTTCTTCATAGGCTAACTAACCT<br/> CGGCGTGCACACAACCATGTCCTGAACCTTCACCTCGTCCCTATAAAAGCCC<br/> ATCCAACCTTACAATCTCATCATCACCCACAACACCGAGCACCCCAATCTAC<br/> AGATCAATTCACTGACAGTTCACTGATCTAGA</p>                                            |     |
| Seq #18 Glub-2 promoter seq                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |     |
| <p>CTGCAGTAATGGATACCTAGTAGCAAGCTAGCTTAAACAAATCTAAATCCAA<br/> TCTGTTTCGTAAACGTTTTCTCGATCGCAATTTTGATCAAACTATTGAAACC<br/> TCAATTAACCATTCAAAATTTTAATATACCCAACAAGAGCGTCCAAACCAA<br/> ATATGTAAATATGGATGTCATGATAATTGACTTATGACAATGTGATTATTTTAT<br/> CAAGTCTTTAAATCATTAAATCTAGTTGAAGGTTTATGTTTTCTTATGCTAAAG<br/> GGTTATGTTTATATAAGAATATTAAGAGCAAATTGCAATAGATCAACACAAC<br/> AAATTTGAATGTTTCCAGATGTGTAAAAATATCCAAATTAATTGTTTTAAATA<br/> GTTTTAAGAAGGATCTGATATGCAAGTTTGATAGTTAGTAACTGCAAAAGGG<br/> CTTATTACATGGAAAATTCCTTATTGAATATGTTTCATTGACTGGTTTATTTTA<br/> CATGACAACAAAGTTACTAGTATGTCAATAAAAAAATACAAGGTTACTTGTCA<br/> ATTGTATTGTGCCAAGTAAAGATGACAACAAACATACAAATTTATTTGTTCTTT<br/> TATAGAAACACCTAACTTATCAAGGATAGTTGGCCACGCAAAAATGACAACAT<br/> ACTTTACAATTGTATCATCATAAAGATCTTATCAAGTATAAGAACTTTATGGTG<br/> ACATAAAAAATAATCACAAGGGCAAGACACATACTAAAAGTATGGACAGAAAT<br/> TTCTTAACAAACTCCATTTGTTTTGTATCCAAAAGCATAAGAAATGAGTCATG<br/> GCTGAGTCATGATATGTAGTTCAATCTTGCAAAATTGCCTTTTTGTTAAGTATT<br/> GTTTTAACACTACAAGTCACATATTGTCTATACTTGCAACAAACACTATTACC<br/> GTGTATCCCAAGTGGCCTTTTCATTGCTATATAAACTAGCTTGATCGGTCTTT<br/> CAACTCACATCAATTAGCTTAAGTTTCCATTAGCAACTGCTAATAGCT</p> |     |
| Seq#19 Gt3 promoter seq                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |     |
| <p>CTGCAGTGTAAGTGTAGCTTCTTATAGCTTAGTGCTTTACTATCTTCACAAGC<br/> ACATGCTATAGTATTGTTCCAAGATGAAAGAATAATTCATCCTTGCTACCAAC<br/> TTGCATGATATTATTTGTGAATATCCTATCTCTTGGCTTATAATGAAATGTG<br/> CTGCTGGGTTATTCTGACCATGGTATTTGAGAGCCTTTGTATAGCTGAAACC<br/> AACGTATATCGAGCATGGAACAGAGAACAAAATGCAAGGATTTTTTTATTCTG<br/> GTTTCATGCCCTGGATGGGTTAATATCGTGATCATCAAAAAGATATGCATAAA<br/> ATTAAGTAATAAATTTGCTCATAAGAAACCAAAACCAAAAGCACATATGTCC<br/> TAAACAACTGCATTTTGTGTCATGTAGCAATACAAGAGATAATATATGAC</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |     |

GTGGTTATGACTTATTCACCTTTTTGTGACTCCAAAATGTAGTAGGTCTAACTG  
 ATTGTTTAAAGTGATGTCTTACTGTAGAAGTTTCATCCCAAAAGCAATCACTA  
 AAGCAACACACACGTATAGTCCACCTTCACGTAATTCTTTGTGGAAGATAACA  
 AGAAGGCTCACTGAAAAATAAAAGCAAAGAAAAGGATATCAAACAGACCATT  
 GTGCATCCCATTGATCCTTGTATGTCTATTTATCTATCCTCCTTTTGTGTACCT  
 TACTTCTATCTAGTGAGTCACTTCATATGTGGACATTAACAAACTCTATCTTAA  
 CATCTAGTCGATCACTACTTTACTTCACTATAAAAAGGACCAACATATATCATC  
 CATTTCTCACAAAAGCATTGAGTTCAGTCCCAAAAATCTAGA

Seq #20 Glub-1 promoter seq

CTGCAGAGATATGGATTTTCTAAGATTAATTGATTCTCTGTCTAAAGAAAAA  
 AGTATTATTGAATTAATGAAAAAGAAAAAGGAAAAAGGGGATGGCTTCTG  
 CTTTTTGGGCTGAAGGCGGCGTGTGGCCAGCGTGCTGCGTGCGGACAGCG  
 AGCGAACACACGACGGAGCAGCTACGACGAACGGGGGACCGAGTGAGCCG  
 GACGAGGATGTGGCCTAGGACGAGTGACACAAGGCTAGTGGACTCGGTCCC  
 CGCGCGGTATCCCGAGTGGTCCACTGTCTGCAAACACGATTCACATAGAGC  
 GGGCAGACGCGGGAGCCGTCTAGGTGCACCGGAAGCAAATCCGTCGCCT  
 GGGTGGATTTGAGTGACACGGCCACGTGTAGCCTCACAGCTCTCCGTGGT  
 CAGATGTGTAAAATTATCATAATATGTGTTTTTCAAATAGTTAAATAATATAT  
 AGGCAAGTTATATGGGTCAATAAGCAGTAAAAAGGCTTATGACATGGTAAAA  
 TTACTTACACCAATATGCCTTACTGTCTGATATATTTTACATGACAACAAAGT  
 ACAAGTACGTCATTTAAAAAATACAAGTTACTTATCAATTGTAGTGTATCAAGTA  
 AATGACAACAAACCTACAAATTTGCTATTTTGAAGGAACACTTAAAAAATCA  
 ATAGGCAAGTTATATAGTCAATAAACTGCAAGAAGGCTTATGACATGGAAAAA  
 TTACATACACCAATATGCTTTATTGTCCGGTATATTTTACAAGACAACAAAGT  
 ATAAGTATGTCATTTAAAAAATACAAGTTACTTATCAATTGTCAAGTAAATGAAA  
 ACAAACCTACAAATTTGTTATTTTGAAGGAACACCTAAATTATCAAATATAGCT  
 TGCTACGCAAAATGACAACATGCTTACAAGTTATTATCATCTTAAAGTTAGAC  
 TCATCTTCTCAAGCATAAGAGCTTTATGGTGCAAAAACAAATATAATGACAAG  
 GCAAAGATACATACATATTAAGAGTATGGACAGACATTTCTTTAACAAACTCC  
 ATTTGTATTACTCCAAAAGCACCAGAAGTTTGTATGGCTGAGTCATGAAATG  
 TATAGTTCAATCTTGCAAAGTTGCCTTTTCTTTTGTACTGTGTTTTAACTAC  
 AAGCCATATATTGTCTGTACGTGCAACAACTATATCACCATGTATCCCAAGA  
 TGCTTTTTTATTGCTATATAAACTAGCTTGGTCTGTCTTTGAACTCACATCAAT  
 TAGCTTAAGTTCCATAAGCAAGTACAAATAGCTCTAGA

Seq #21 Rice proalmin promoter seq

CTGCAGCATCGGCTTAGGTGTAGCAACACGACTTTATTATTATTATTATT  
 ATTATTATTATTTTACAAAAATATAAAATAGATCAGTCCCTCACCACAAGTAGA  
 GCAAGTTGGTGAGTTATTGTAAAGTTCTACAAAGCTAATTTAAAAGTTATTGC  
 ATTAACCTATTTTATATTACAAACAAGAGTGTCAATGGAACAATGAAAACCAT  
 ATGACATACTATAATTTTGTATTTTATTATTGAAATTATATAATTCAAAGAGAATA  
 AATCCACATAGCCGTAAAGTTCTACATGTGGTGCATTACCAAAATATATATAG  
 CTTACAAAACATGACAAGCTTAGTTTGAATAATTGCAATCCTTATCACATTGA  
 CACATAAAGTGAGTGATGAGTCATAATATTATTTTCTTGCTACCCATCATGT  
 ATATATGATAGCCACAAAGTTACTTTGATGATGATATCAAAGAACATTTTATG  
 GTGCACCTAACAGAATATCCAAATAATATGACTCACTTAGATCATAATAGAGC  
 ATCAAGTAAACTAACACTCTAAAGCAACCGATGGGAAAGCATCTATAAATAG  
 ACAAGCACAATGAAATCCTCATCATCCTTACCACAATTCAAATATTATAGT  
 TGAAGCATAGTAGTAGAATCCAACAACAATCTAGAG

Seq #22 Rice cysteine peptidase promoter seq

CCAGGCTTCATCCTAACCATTACAGGCAAGATGTTGTATGAAGAAGGGCGAA  
 CATGCAGATTGTTAAACTGACACGTGATGGACAAGAATGACCGATTGGTGAC

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| <p>CGGTCTGACAATGGTCATGTCGTCAGCAGACAGCCATCTCCACGTCGCGC<br/> CTGCTTCCGGTGAAAGTGAGGTAGGTATGGGCCGTCCCGTCAGAAGGTGA<br/> TTCGGATGGCAGCGATACAAATCTCCGTCCATTAATGAAGAGAAGTCAAGTT<br/> GAAAGAAAGGGAGGGAGAGATGGTGCATGTGGGATCCCCTTGGGATATAAA<br/> AGGAGGACCTTGCCCACTTAGAAAGGAGAGGAGAAAGCAATCCCAGAAGAA<br/> TCGGGGGCTGACTGGCACTTTGTAGCTTCTTCATACGCGAATCCACCAAAAC<br/> ACAGGAGTAGGGTATTACGCTTCTCAGCGGCCCGAACCTGTATACATCGCC<br/> CGTGTCTTGTGTGTTTCCGCTCTTGCGAACCTTCCACAGATTGGGAGCTTAG<br/> AACCTACCCAGGGCCCCCGGCCGAACCTGGCAAAGGGGGGCTGCGCGGT<br/> CTCCCGGTGAGGAGCCCCACGCTCCGTCACTTCTAAATTACCCGATGAGAA<br/> AGGGAGGGGGGGGGGGGAAATCTGCCTTGTTTATTTACGATCCAACGGATT<br/> TGGTCGACACCGATGAGGTGTCTTACCAGTTACCACGAGCTAGATTATAGTA<br/> CTAATTACTTGAGGATTTCGGTTCCTAATTTTTTACCCGATCGACTTCGCCATG<br/> GAAAATTTTTTATTCGGGGGAGAAATATCCACCCTGTTTCGCTCCTAATTAAGA<br/> TAGGAATTGTTACGATTAGCAACCTAATTCAGATCAGAATTGTTAGTTAGCGG<br/> CGTTGGATCCCTCACCTCATCCCATCCCAATTCCCAAACCCAAACTCCTCTT<br/> CCAGTCGCCGACCCAAACACGCATCCGCCGCTATAAATCCCACCCGCGATC<br/> GAGCCTATCAAGCCCCAAAAAACCACAAACCAAGGAAGGAAAAAAGAG<br/> GAGGAAAAGAAAAGAGGAGGAAAGCGAAGAGGTTGGAGAGAGACGCTCGT<br/> CTCCACGTCGCCGCC</p> <p>Seq #23 Barley D-Hordian promoter</p> <p>CTTCGAGTGCCCGCCGATTTGCCAGCAATGGCTAACAGACACATATTCTGCC<br/> AAAACCCAGAACAAATAATCACTTCTCGTAGATGAAGAGAACAGACCAAGAT<br/> ACAAACGTCCACGCTTACGCAAACAGTACCCAGAACTAGGATTAAGCCGAT<br/> TACGCGGCTTTAGCAGACCGTCCAAAAAACTGTTTTGCAAAGCTCCAATTC<br/> CTCCTTGCTTATCCAATTTCTTTTGTGTTGGCAAACCTGCACTTGTCCAACCGA<br/> TTTTGTTCTTCCCGTGTTTCTTCTTAGGCTAACTAACACAGCCGTGCACATAG<br/> CCATGGTCCGGAATCTTCACCTCGTCCCTATAAAGCCCAGCCAATCTCCAC<br/> AATCTCATCATACCGAGAACACCGAGAACCAAAAACCTAGAGATCAATTCA<br/> TTGACAGTCCACCG</p> <p>Rice Glb promoter and Gt1 leader coding sequence</p> |                   |
| <p>Other storage body leader sequences</p> <p>Bx7 #24 bx7 signal peptide seq</p> <p>ATGGCTAAGCGCCTGGTCCTCTTTGCGGCAGTAGTCGTCGCCCTCGTGGCT<br/> CTCACCGCC</p> <p>Seq #25 Glub-2 signal peptide seq</p> <p>ATGGCAACTACCATTTTCTCTCGTTTTTCTATATACTTTTGTGCTATGCTATTA<br/> TGCCAGGGTTCTATGGCC</p> <p>Seq #26 Gt3 signal peptide seq</p> <p>ATGTGGACATTAACAAACTCTATCTTAACATCTAGTCGATCACTACTTTACTTC<br/> ACTATAAAAGGACCAACATATATCATCCATT</p> <p>Seq #27 Glub-1 signal peptide seq</p> <p>ATGGCGAGTTCCGTTTTCTCTCGGTTTTCTATATACTTTTGTGTTCTTCTATTA<br/> TGCCATGGTTCTATGGCC</p> <p>Seq #28 proalmin signal peptide seq</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <p>24-<br/>30</p> |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <p>ATGAAGATCATTTTCGTATTTGCTCTCCTTGCTATTGTTGCATGCAACGCTTC<br/>TGCACGGTTTGATGCT</p> <p>Seq #29 Rice cysteine peptidase signal peptide seq</p> <p>ATGGCCGCCCCGCGCCGCGCCGCGCGCTTCCTGCTGCTGCTCATCGTCGT<br/>TGGTCACCGCGCC</p> <p>Seq #30 D- Hodian signal peptide<br/>ATGGCTAAGCGGCTGGTCCTCTTTGTGGCGGTAATCGTCGCCCTCGTGGCT<br/>CTCACCACCGCCOther monocot maturation specific promoter sequences</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |    |
| <p>O2 transcription factor sequence</p> <p>ATGGAGCACGTCATCTCAATGGAGGAGATCCTCGGGCCCTTCTGGGAGCTG<br/>CTACCACCGCCAGCGCCAGAGCCAGAGCGAGAGCAGCCTCCGGTAACCGG<br/>CATCGTCGTCGGCAGTGTATAGACGTTGCTGCTGCTGGTCATGGTGACGG<br/>GGACATGATGGATCAGCAGCACGCCACAGAGTGGACCTTTGAGAGGTTACT<br/>AGAAGAGGAGGCTCTGACGACAAGCACACCGCCGCGCGGTGGTGGTGGTGC<br/>CGAACTCTTGTTGCTCAGGCGCCCTAAATGCTGACCGGCCGCGCGGTGATGG<br/>AAGAGGCGGTAACATATGGCGCCTGCGGCGGTGAGTAGTGCCGTAGTAGGT<br/>GACCCCATGGAGTACAATGCCATACTGAGGAGGAAGCTGGAGGAGGACCTC<br/>GAGGCCTTCAAAATGTGGAGGGCGGCCTCCAGTGTTGTGACCTCAGATCAA<br/>CGTTCTCAAGGCTCAAACAATCACACTGGAGGTAGCAGCATCAGGAATAATC<br/>CAGTGCAGAACAAGCTGATGAACGGCGAAGATCCAATCAACAATAACCACG<br/>CTCAAACCTGCAGGCCCTTGGCGTGAGGCTTGCTACTAGCTCTTCCTCGAGAG<br/>ATCCTTCACCATCAGACGAAGACATGGACGGAGAAGTAGAGATTCTGGGGT<br/>TCAAGATGCCTACCGAGGAAAGAGTGAGGAAAAGAAAGGAATCCAATAGAG<br/>AATCAGCCAGACGCTCGAGATACAGGAAAGCCGCTCACCTGAAAGAACTGG<br/>AAGACCAGGTAGCACAGCTAAAAGCCGAGAATTCTTGCCTGCTGAGGCGCA<br/>TTGCCGCTCTGAACCAGAAGTACAACGACGCTAACGTCGACAACAGGGTGC<br/>TGAGAGCGGACATGGAGACCCTAAGAGCTAAGGTGAAGATGGGAGAGGACT<br/>CTCTGAAGCGGGTGATAGAGATGAGCTCATCAGTGCCGTCGTCCATGCCCA<br/>TCTCGGCGCCGACCCCCAGCTCCGACGCTCCAGTGCCGCGCGCCGCTATC<br/>CGAGACAGCATCGTCGGCTACTTCTCCGCCACAGCCGACAGACGATGCT<br/>TCGGTCGGCAACGTTTTCTTGCGACTGCAAGCTCATCAAGAGCCTGCATCC<br/>ATGGTCGTCGGTGGAACCTCTGAGCGCCACAGAGATGAACCGAGTAGCAGCA<br/>GCCACGCATTGCGCGGGGGCCATGGAGCACATCCAGACGGCGATGGGATC<br/>CATGCCGCGACCTCCGCTCCGGATCTACACCGCCGCGCGCAGGATTATGA<br/>GCTGCTGGGTCCAAATGGGGCCATACACATGGACATGTATTAG</p> | 31 |
| <p>PBF transcription factor sequence</p> <p>ATGGACATGATCTCCGGCAGCACTGCAGCAACATCAACACCCCACAACAAC<br/>CAACAGGCGGTGATGTTGTCATCCCCATTATAAAGGAGGAAGCTAGGGAC<br/>CCAAAGCAGACACGAGCCATGCCCCAAATAGGTGGCAGTGGGGAGCGTAA<br/>GCCGAGGCCGCAACTACCTGAGGCGCTCAAGTGCCCACGCTGCGACTCCA<br/>ACAACACCAAGTTTTGCTACTACAACAATTATAGCATGTCACAACCACGCTAC<br/>TTTTGCAAGGCTTGCCGCGCTATTGGACACATGGTGGTACCCTCCGCAAT<br/>GTCCCCATTGGTGGTGGGTGTCGCAAGAACAACATGCCTCTAGATTTGTCT<br/>TGGGCTCTCACACCTCATCGTCCTCATCTGCTACCTATGCACCATTATCCCC<br/>TAGCACCACGCTAGCTCTAGCAATATGAGCATCAACAACATATGATGATG<br/>GTGCCTAACATGACGATGCCTACCCCAACGACAATGGGCTTATTCCCTAATG<br/>TGCTCCCAACACTTATGCCGACAGGTGGAGGCGGGGGCTTTGACTTCACTA<br/>TGGACAACCAACATAGATCATTGTCCTTACACCAATGTCTCTACCTAGCCA<br/>GGGGCCAGTGCCATGCTGGCTGCAGGAGGGAGTGAGGCAACACCGTCTT<br/>TCCTAGAGATGCTGAGAGGAGGGATTTTTCATGGTAGTAGTAGCTATAACAC<br/>AAGTCTCACGATGAGTGGTGGCAACAATGGAATGGACAAGCCATTTTCGCTG<br/>CCATCATATGGTGCAATGTGCACAAATGGGTGAGTGGCTCAACCACTAATG</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 32 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| ATGCCAGACAACTGGTGGGGCCTCAGCAGGATAACAAGGCCATCATGAAGA<br>GCAGTAATAACAACAATGGTGTATCATTGTTGAACCTCTACTGGAACAAGCA<br>CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAG<br>GGACAATAA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |    |
| <p>Reb transcription factor sequence</p> <p>ATGGAGCGGGTGTCTCCGTGGAGGAGATCTCCGACCCATTCTGGGTCCCCG<br/>CCTCCGCCGCCGAGTCGGCGGGCGGGCCAGCAGCAGGGCGGGCGGGCG<br/>GCGTGGCTTCGGGAGGTGGTGGTGGTGTAGCGGGGGCGGGCGGGCGGGCGG<br/>GAACGCGATGAACCGGTGCCCGTCGGAGTGGTACTTCCAGAAGTTTCTGGA<br/>GGAGGCGGTGCTCGATAGCCCCGTCCCGAACCCCTAGCCCGAGGGCCGAAG<br/>CGGGAGGGATCAGGGGCGCAGGAGGGGTGGTGCCGGTTCGATGTTAAGCA<br/>GCCGAGCTCTCGGCGGGCGGCGACGACGAGCGCGGTGGTGGACCCCGTG<br/>GAGTACAACGCGATGCTGAAGCAGAAGCTGGAGAAGGACCTCGCCGCGGT<br/>CGCCATGTGGAGGGTACAGCCATTCTCCCCCCTCTAGTACTCGAGAGCTT<br/>ACTGAGATCGGCAATGCTAGCTACTGTTTGCATCGAATGTTTATAGGTATTTA<br/>GATCGGGCATTCTATAGACCAATGGCGTCCATGGTCTTGCAATGCGCTCTG<br/>TTGAGTGTCCGGTGGTTGGTTCGACTCATAGTATGTAGGGTTGTGCGTATGTA<br/>CAAACGGAAGCTTCATAGACCTCGGTATTGAGATTGCGATATCGATGCAACC<br/>TGCGAATTGGCGATGTAATCAGTCATATTCTTACTAACTGCGAGACAGTGG<br/>TTTGTGTTGCAATTGCAATATTTTGTATGGGGCTGCTTAACTGTCATTGCCT<br/>TTTTAGATTGGCAATATGTGACTTTATGCAAGTATTTGATTGGGCGGATCCAG<br/>GAACAAAAAGTTGGGGGGATTCAACATACCGAGTACACTGGCATAAACACAT<br/>CATCTCAGTATTAACATATGCTAAAATGCTATTAAGAGACCTTTAGCACCTCT<br/>TATCTTATCAACCATGGTGAAAAAATTGAAGGGGGGACTCAGGGGGGGTATCC<br/>ATGGGTCCGATGGGTGCAGGGGGGACTGAGTCCCCCTGCACCCACGTTG<br/>AATCCGCCCTGGCATGCGTATAAGCTGTCACAGCCATTTCTAGGTGCTTGTG<br/>CTTAGTTGGGTGATGTCAGCTTAATTTGTCTTTTCTATGTCGTCATCGATTTT<br/>CTAAGAAACGAAAAATAGCCTATTTATGTGCTCCAGAATTTGATGATCCCTGG<br/>CCCTTCATTTGCTGAAATTAGCCTATTTGTTGGTTGCCCTTCAGTTTTTTCCC<br/>AGCTTATGTTGTTGCAATGTGTGGCTATGCCTCGTTTTGTGCCCTATAATTTA<br/>TTATTTGCAATTCATTTTTGTACATGACTTAAATGACACTAGAGCAACATGCA<br/>CTGATTGGTTATCCTATAATCATTTATGTAGTTCTGTTCATTTTATCATGCTAG<br/>CTCATGTCATTTTCATCTTCAGGCCTCTGGCACAGTTCACCTGAGCGTCCT<br/>GGAGCTGGTTCATCCTTGCTGAATGCAGATGTTTCACACATAGGCGCTCCTA<br/>ATTCCATCGGAGGTACTTATCTTATCTGGTTACATTTTCAGATTGTTATGAAA<br/>CTACCCAAATATCCTGCACAATTGCATGGGATTAAATTTTAGTTTCTTTGAAAT<br/>AGAAGTAGAGTTGTATTGCTGTCACGTCATCAAAATAGTTCTGAAGCTATGAAT<br/>AAATAAGTTCGCGATTTGTTAGTGATTCTTTGAACATTAGAATTGTTATGCTTA<br/>AGTAGATAGGGTTATGTTTGTGTTGGAGTTCCCTTAAATCATTTTCATTGCTGAC<br/>TGCCAGCTGGCAGGAGCATTTGTTGTGCTTGCCCTTGACCATGAATGAAGACCTTC<br/>CTGTTCTGAGTGCTCACAAGAAAACATATTTTGATTAATGCACCTTGAATCCT<br/>TAGGATCTTGCAAAGATGGGCACTTAGCTTTAGAATTGAGTAGTACTTAAATA<br/>GCTGTTGTTATCATGATTTGTCCTGTAGTGAAATGTCGACAAAACAGGAATG<br/>CTACTTTTGACTTCTGATATTTTCATGCCTGGCTTTACTTATGCTCTGTTTGGAA<br/>CATGGGCACATATCAGGCAATGCTACTCCAGTTCAAACATGCTAAGTGGCC<br/>CAAGTGGGGGATCGGGCTCACAGTTGGTACAGAATGTTGATGTCCTTGTAAG<br/>GCAGCCCACCAGCTCTTCATCAAGGGAGCAGTCAGATGATGATGACATGAA<br/>GGGAGAAGCTGAGACCACTGGAACCTGCAAGACCTGCTGATCAAAGATTACA<br/>ACGAAGGTGATCATTGCTTCCCTTGTAAATATAGATTCTGTACATAATTAA<br/>CCTACCTCGTCATGCATGCATGTGTCCTATTTTCACCTTAGCCCTTTCAGTTG<br/>GATTTCCACTTTTCATCCGGTAGCCTTTCAGTTTCTTATGTCATCGCATATATG<br/>ATCTTTTACCTACCATATTAGTTCTCTGTGTGCCATACTCAGTGCTTAGTGTC<br/>TCGAGCAAGAGAGGAATTTGTATGGCTATTACACGTAGCACTTTGCTCTCTA<br/>CTTGTTTATTGACATAAGCAATTTGGGATGAATTAATCTGAGTTCACATCAT<br/>ATTCCTTATGTCACAAGTTTCTGAAACCGATTGTATCTAGTATCTGGTTGATG<br/>CACCCCATCTTGATTGCAAAATCAAAGTTATACTCCCTAGAGAGCTTTACC</p> | 33 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| TTTCATAAAGCAATTACCCCAATAAACCACGGATTTGATAGCTATTGACTATG<br>ATTACCAGAATTCATTTGGCAGCTATTTTCTCAATTTAAGTTTGGTATTAGTCT<br>CAGTTGGCTGTAAAATAATGTCACGGTAGGGTACATGTATGTGCAGCATACA<br>AGGTATGGGTGAGTTATGATATGGACAGTGTGTACACCCACATTTGCTCAC<br>TAAATCAAAATATTCAAACGTCACGTGATGATATGGTGGATTGCATTATACC<br>TTGTATTGTTTATTATGTTACTTGTGCTAGACAATAATATAGGCTGTTCTTTTG<br>GGTGATTTTGTATGAAGATGTTGAGCAAGCACTTCTCGATATAATGCTAGTTT<br>TGTTGACCTGTTCCAGGAAGCAATCCAATCGGGAGTCAGCCAGGCGCTCAA<br>GAAGCAGAAAGGCAGCTCACTTGAATGAGCTGGAGGCACAGGTGTGATAGT<br>TCACATAGTTATTTTCGATAAGACATAAAATCCTAAATTACTGGCTACTGACTT<br>CAGTTATGGATTTACTTGTTACAGGTATCGCAATTAAGAGTCGAGAACTCCTC<br>GCTGTAAAGGCGTCTTGCTGATGTTAACCAGAAGTACAATGATGCTGCTGTT<br>GACAATAGAGTGCTAAAAGCAGATGTTGAGACCTTGAGAGCAAAGGTATGCT<br>ATATATGCCTTTTGCAATATGCATCCCATGGATTGCTACTTTGGCTTGTTTCA<br>AACTTTCAACGTGACTTGTGTACCCTGTTATTAGAAGAATAATCCCGCCTACC<br>ATTATACTCTATAAATCACCATTTGGCCAGTCCAAACATGATTATTAAATCAG<br>GTCAATCTGAACATTGAAATGTATCAAAAATTCGCAGGTGAAGATGGCAGAG<br>GACTCGGTGAAGCGGGTGACAGGCATGAACGCGTTGTTCCCGCCGCTTCT<br>GATATGTCATCCCTCAGCATGCCATTCAACAGCTCCCCATCTGAAGCAACGT<br>CAGACGCTGCTGTTCCCATCCAAGATGACCCGAACAATTACTTCGCTACTAA<br>CAACGACATCGGAGGTAACAACAACACTACATGCCCGACATACCTTCTTCGGCT<br>CAGGAGGACGAGGACTTCGTCAATGGCGCTCTGGCTGCCGGCAAGATTGG<br>CCGGCCAGCCTCGCTGCAGCGGGTGCGAGCCTGGAGCATCTCCAGAAGA<br>GGATGTGCGGTGGGCCGGCTTCGTCTGGGTGACGTCCTGA |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

## WHAT IS CLAIMED IS:

1. A nutritionally enhanced food comprising  
one or more plant-derived food ingredients, and  
5 as an additive, a seed composition containing a flour, extract, or malt obtained  
from mature monocot seeds and one or more seed-produced human milk proteins in  
substantially unpurified form.
2. The food of claim 1, wherein the seed-produced protein(s) present in the food  
10 are selected from the group consisting of lactoferrin, lysozyme, lactoferricin, epidermal  
growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein, haptocorrin,  
lactoperoxidase, and alpha-1-antitrypsin.
3. The food of claim 2, wherein said the seed-produced proteins are lysozyme  
15 and lactoferrin.
4. The food of claim 3, which is an infant formula, wherein said the seed-  
produced proteins are lysozyme, in an amount between 0.03 to 0.3 g protein/liter  
formula, and lactoferrin, in an amount between 0.3 to 3 g/liter formula.  
20
5. The food of claim 1, wherein the seed composition comprises between 0.1 to  
20% of the total solid weight of the food.
6. The food of claim 5, which is an infant milk formula, and the seed composition  
25 contains a seed extract or malt obtained from mature seeds of rice or barley.
7. An ingestible monocot-seed composition containing a flour, extract, or malt  
obtained from mature monocot seeds and one or more seed-produced human milk  
proteins in substantially unpurified form.  
30
8. The composition of claim 7, wherein the seed-produced protein(s) present in  
the extract are selected from the group consisting of lactoferrin, lysozyme, lactoferricin,  
epidermal growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein,  
haptocorrin, lactoperoxidase, and alpha-1-antitrypsin.

9. The composition of claim 8, said the seed-produced proteins are lysozyme and lactoferrin.

5           10. The composition of claim 7, wherein at least one of human proteins is present in the added extract in an amount greater than 1 mg/gram dry weight of extract.

          11. The composition of claim 1, wherein

          (a) the flour is prepared by milling mature monocot seeds;

10           (b) the extract is prepared by suspending milled flour in a buffered aqueous medium; and

          (c) the malt is prepared by (i) steeping barley seeds to a desired water content, (ii) germinating the stepped barley, (iii) drying the germinated seeds, under conditions effective to stop germination, (iv) crushing the dried seeds, (v) optionally, adding crushed  
15 seeds from a non-barley monocot plant, (vi) forming a mixture of crushed seeds in water, and (vii) malting the crushed seed mixture until a desired malt is achieved, where at least one of the barley or non-barley monocot seeds contain such milk protein(s).

          12. The composition of claim 11, wherein step (v) includes adding to the crushed  
20 dried barley seeds, mature rice transgenic seeds that produce a milk protein.

          13. A monocot seed containing, in extractable form, one or more proteins normally present in human milk.

25           14. The seed of claim 13, wherein the human-milk protein(s) are selected from the group consisting of lactoferrin, lysozyme, lactoferricin, EGF, KGF, IGF-I, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, and immunoglobulins.

          15. The seed of claim 14, wherein the human milk proteins are lysozyme and/or  
30 lactoferrin.

          16. The seed of claim 13, which contains two or more proteins normally found in human milk.

17. The seeds of claim 13, wherein the milk proteins constitute at least 0.25 weight percent of the total protein in the harvested mature seeds.

18. A method of producing an ingestible seed composition, comprising

- (a) obtaining a monocot plant that has been stably transformed with a first  
5 chimeric gene having (i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter, (ii) operably linked to said transcriptional regulatory region, a leader DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and (iii) a protein-coding sequence encoding a protein normally present in human milk;  
10 (b) cultivating the transformed plant under seed-maturation conditions;  
(c) harvesting mature seeds from the cultivated plant; and  
(d) extracting from the harvested seeds, a flour, extract, or malt composition containing the human milk protein in substantially unpurified form.

15 19. The method of claim 18, wherein the human milk protein(s) constitute at least 0.25 weight percent of the total protein in the harvested mature seeds.

20. The method of claim 18, wherein

- (a) the flour is prepared by milling mature monocot seeds;  
20 (b) the extract is prepared by suspending milled flour in a buffered aqueous medium; and  
(c) the malt is prepared by (i) steeping barley seeds to a desired water content, (ii) germinating the stepped barley, (iii) drying the germinated seeds, under conditions effective to stop germination, (iv) crushing the dried seeds, and (v) after mixing the  
25 crushed seeds with water, malting the crushed seed mixture until a desired malt is achieved.

21. The method of claim 20, wherein the malt is further prepared by adding to the crushed dried seeds, mature non-barley transgenic monocot seeds that produce a milk  
30 protein.

22. The method of claim 18, wherein the monocot plant obtained is further transformed with a second chimeric gene having (i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter, (ii) operably linked to

said transcriptional regulatory region, a transit DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and (iii) a protein-coding sequence encoding a protein normally present in human breast milk other than that encoded by the first chimeric gene.

5

23. A transgenic monocot plant which is stably transformed with a first chimeric gene having

(i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter,

10 (ii) operably linked to said transcriptional regulatory region, a transit DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and

(iii) a protein-coding sequence encoding a protein normally present in human milk.

15

24. The plant of claim 23, wherein the transcriptional regulatory region in the chimeric gene is from the promoter of a gene selected from the group of rice glutelins, rice globulins, oryzins, and prolamines, barley hordeins, wheat gliadins and glutenins, maize zeins and glutelins, oat glutelins, and sorghum kafirins, millet pennisetins, and rye  
20 secalins genes.

25. The plant of claim 24, wherein the leader sequence in the chimeric gene is from the gene selected from the group of rice glutelins, rice globulins oryzins, and prolamines, barley hordeins, wheat gliadins and glutenins, maize zeins and glutelins, oat  
25 glutelins, and sorghum kafirins, millet pennisetins, and rye secalins genes.

26. The plant of claim 25, wherein the transcriptional regulatory region in the chimeric gene is a rice glutelin Gt1 promoter, and the leader DNA sequence is a rice glutelin Gt1 signal sequence capable of targeting a linked polypeptide to a protein  
30 storage body.

27. The plant of claim 26, wherein glutelin Gt1 promoter and glutelin Gt1 signal sequence are included within the sequence identified by SEQ ID NO:15.

28. The plant of claim 25, wherein the transcriptional regulatory region in the chimeric gene is a rice globulin Glb promoter, and the leader DNA sequence is a rice glutelin Gt1 signal sequence capable of targeting a linked polypeptide to a protein storage body.

5

29. The plant of claim 28, wherein the globulin Glb promoter and glutelin Gt1 signal sequence are included within the sequence identified by SEQ ID NO:16.

30. The plant of claim 23, wherein the transformed monocot plant further comprises a nucleic acid that encodes at least one transcription factor selected from the group consisting of Reb, O2 and PBF, and an active fragment thereof.

31. The plant of claim 30, wherein the transcription factor is O2 and/or PBF.

32. The plant of claim 18, wherein the protein-coding sequence is the a coding sequence for a human milk protein selected from the group consisting of lactoferrin, lysozyme, lactoferricin, EGF, IGF-I, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, and immunoglobulins.

33. The plant of claim 32, wherein the protein coding sequence is selected from the group of codon-optimized sequences identified by SEQ ID NOS: 1, 3, and 7-14.

34. The plant of claim 18, which is further stably transformed with a second chimeric gene having

- (i) a transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter,
- (ii) operably linked to said transcriptional regulatory region, a transit DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and
- (iii) a protein-coding sequence encoding a protein normally present in human breast milk other than that encoded by the first chimeric gene.

35. A method of forming a malt syrup containing one or more human milk proteins, comprising

- (i) steeping barley seeds to a desired water content,
- (ii) germinating the steeped barley,
- (iii) drying the germinated seeds, under conditions effective to stop germination,
- (iv) crushing the dried seeds,
- 5 (v) optionally, adding crushed seeds from a non-barley monocot plant, and
- (vi) forming a mixture of crushed seeds in water, and
- (vii) malting the crushed seed mixture until a desired malt is achieved.

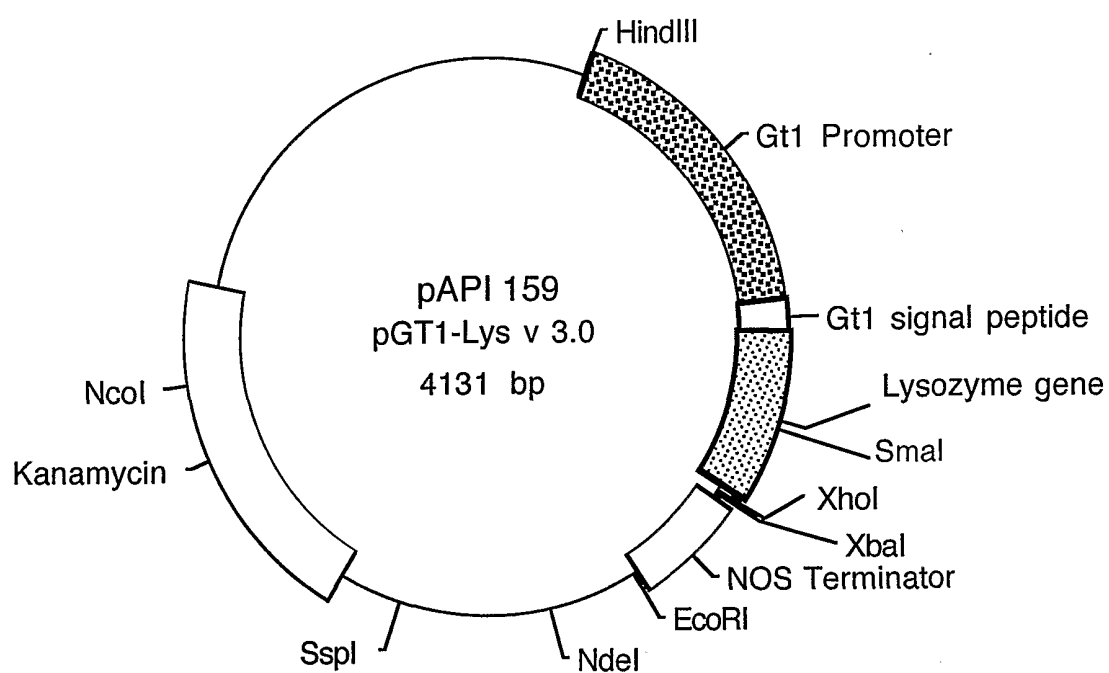
where at least one of the barley or non-barley monocot seeds are obtained from plants that have been stably transformed with a first chimeric gene having (i) a  
10 transcriptional regulatory region from a monocot gene having a seed maturation-specific promoter, (ii) operably linked to said transcriptional regulatory region, a transit DNA sequence encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to an endosperm-cell organelle, and (iii) a protein-coding sequence encoding a protein normally present in human breast milk.

15

36. The method of claim 35, wherein step (v) includes adding to the crushed dried barley seeds, mature rice transgenic seeds that produce a milk protein.

37. The method of claim 18, wherein said extracting includes crushing the seeds  
20 and suspending the crushed seeds in an aqueous medium buffered to physiological pH.

1/45

**Fig. 1**

2/45

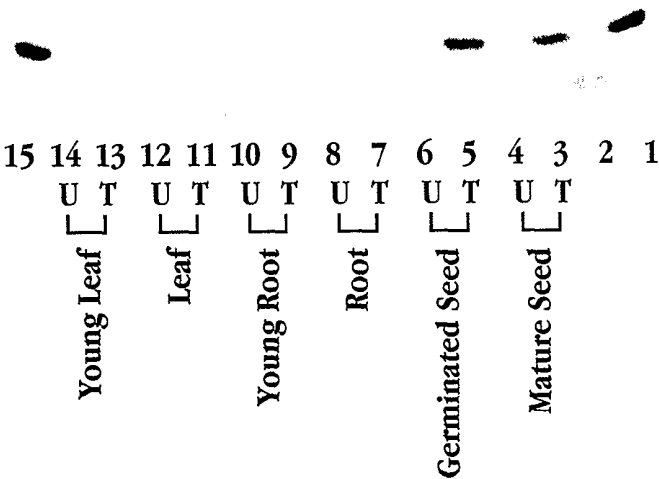
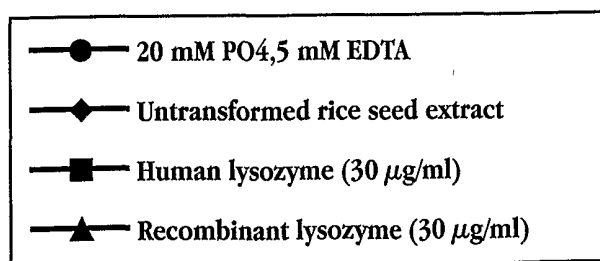
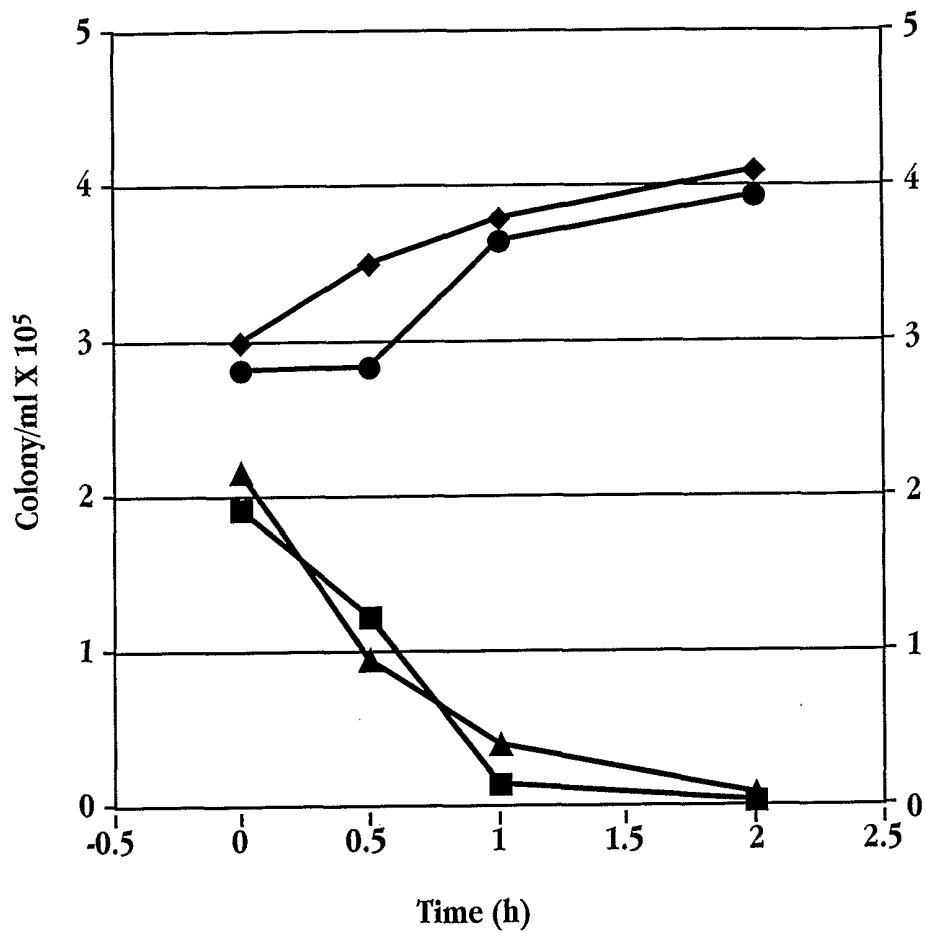
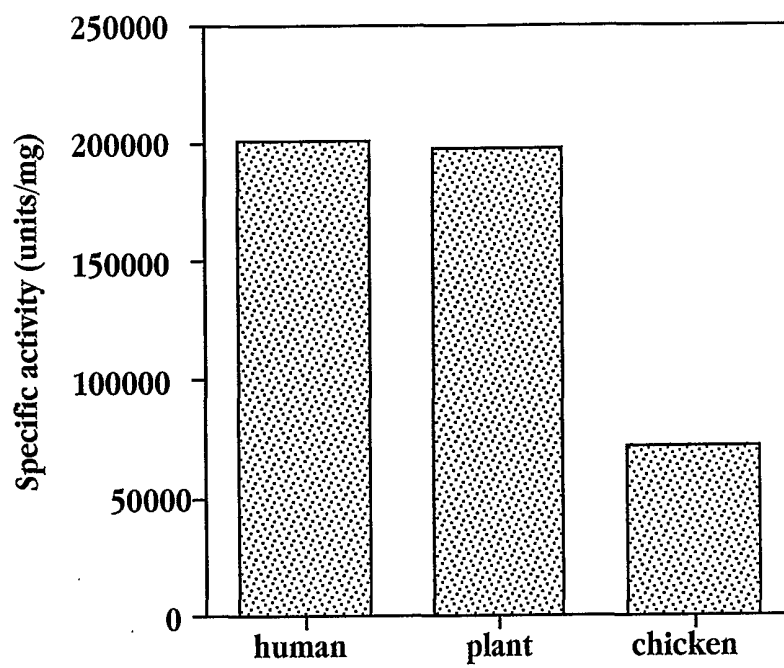


Fig. 2

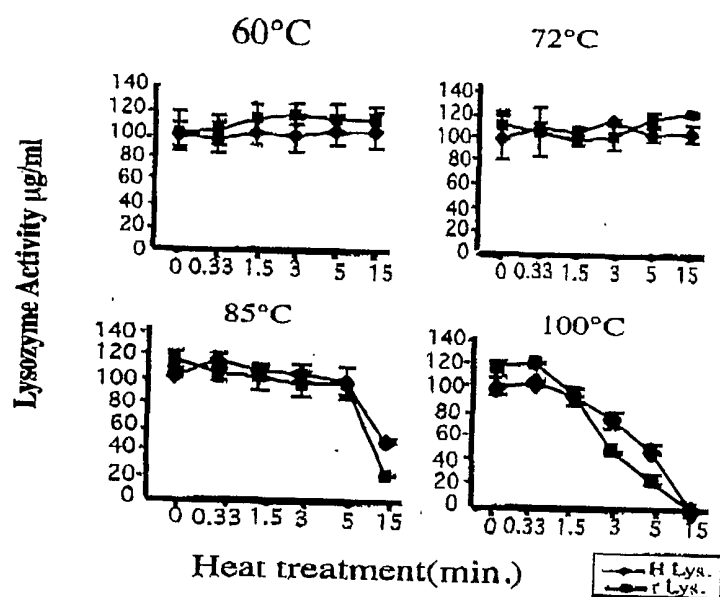
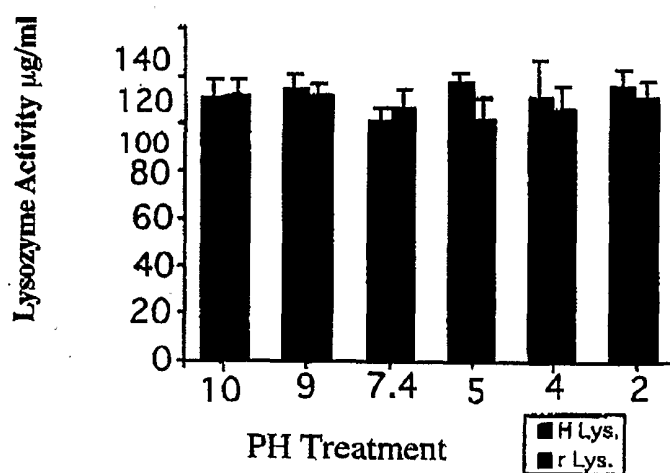
3/45

**Fig. 3**

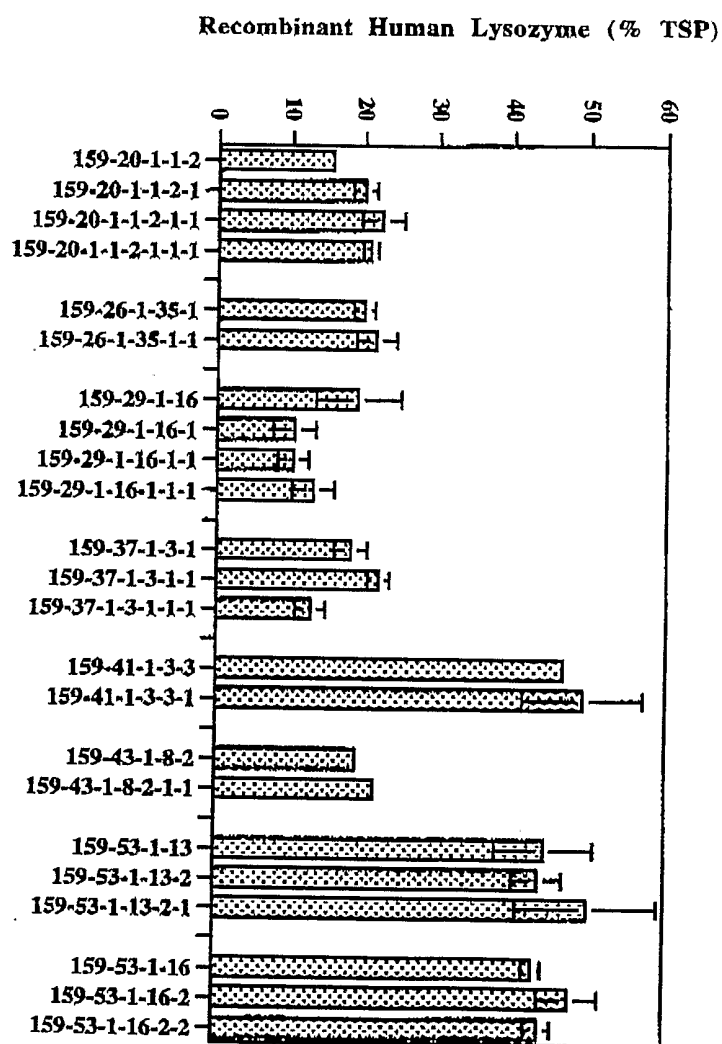
4/45

**Fig. 4**

5/45

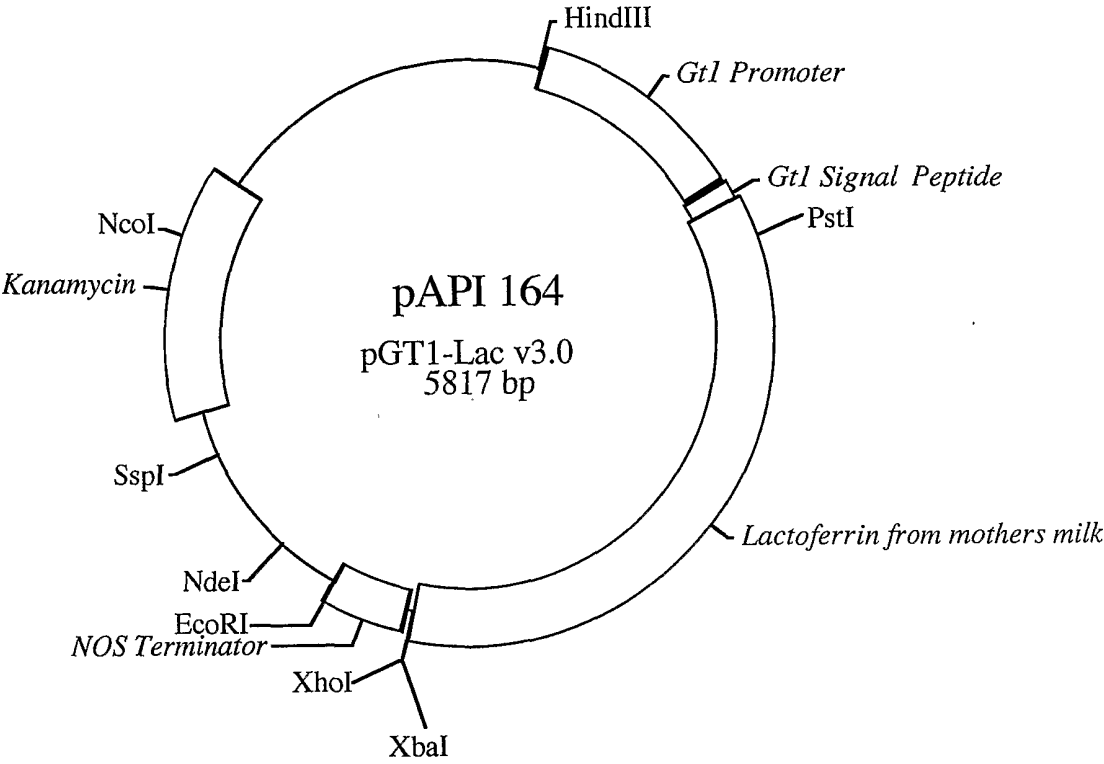
**Fig. 5A****Fig. 5B****Fig. 5C****Fig. 5D****Fig. 5E**

6/45



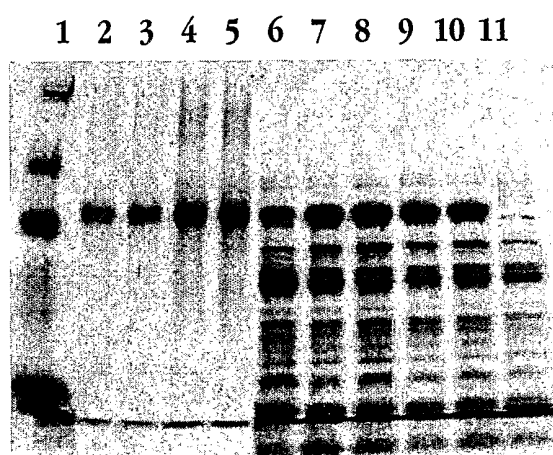
Transgenic Lines

**Fig. 6**



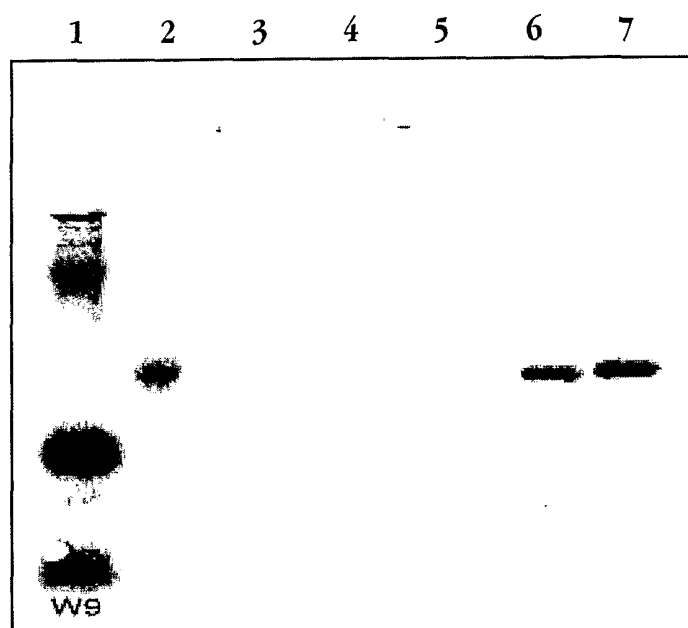
**Fig. 7**

8/45

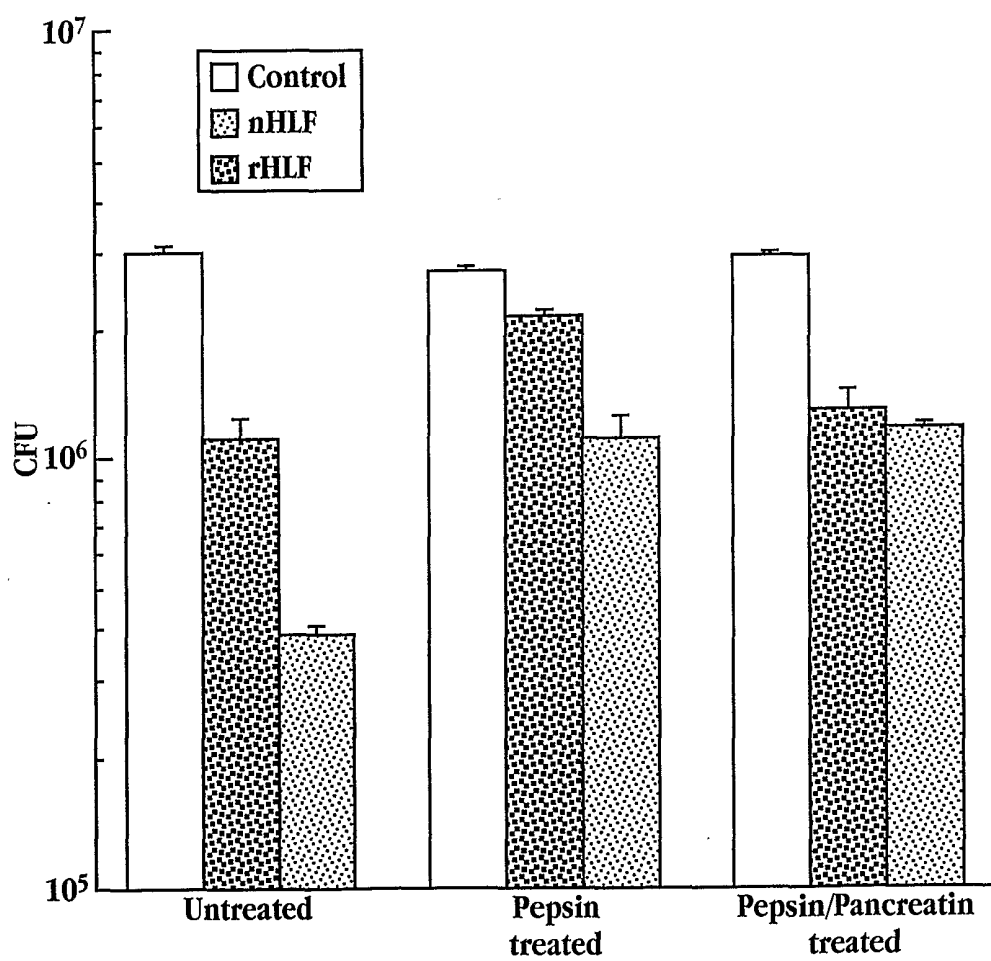


**Fig. 8**

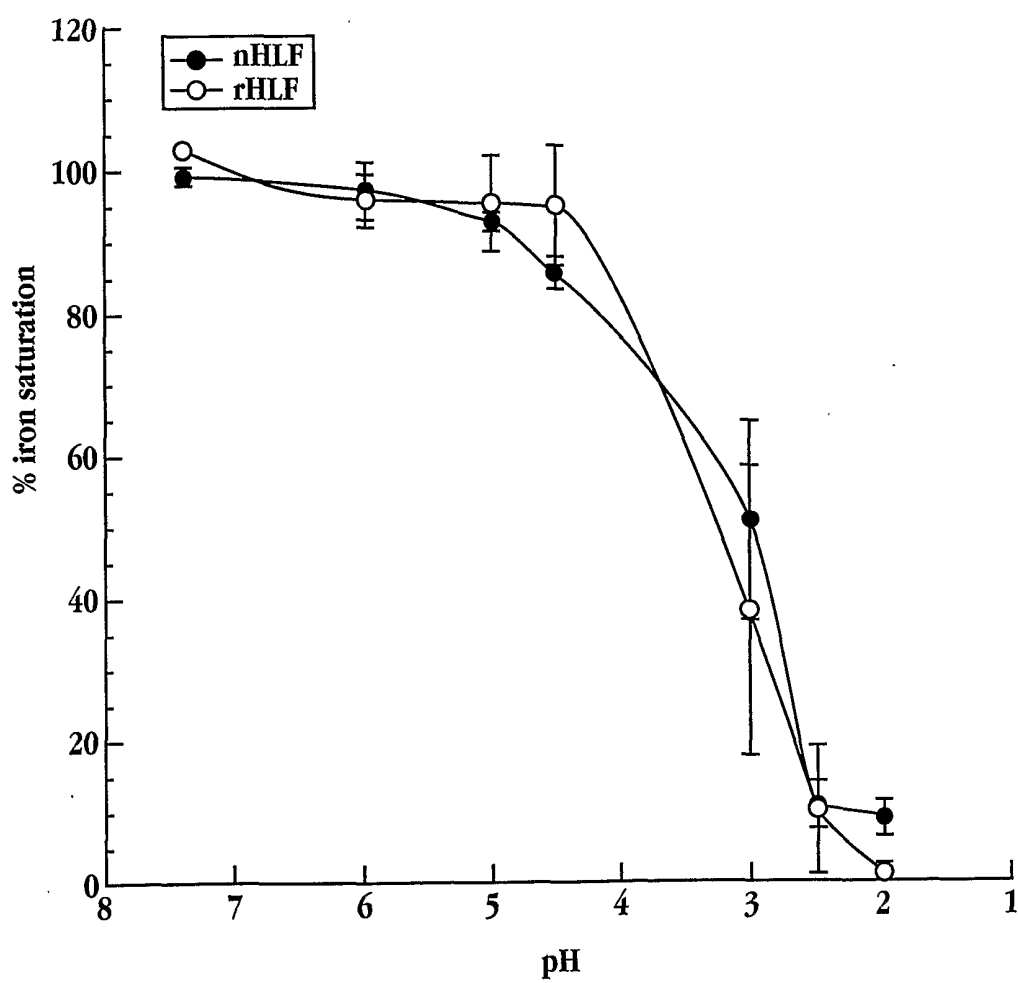
9/45

**Fig. 9**

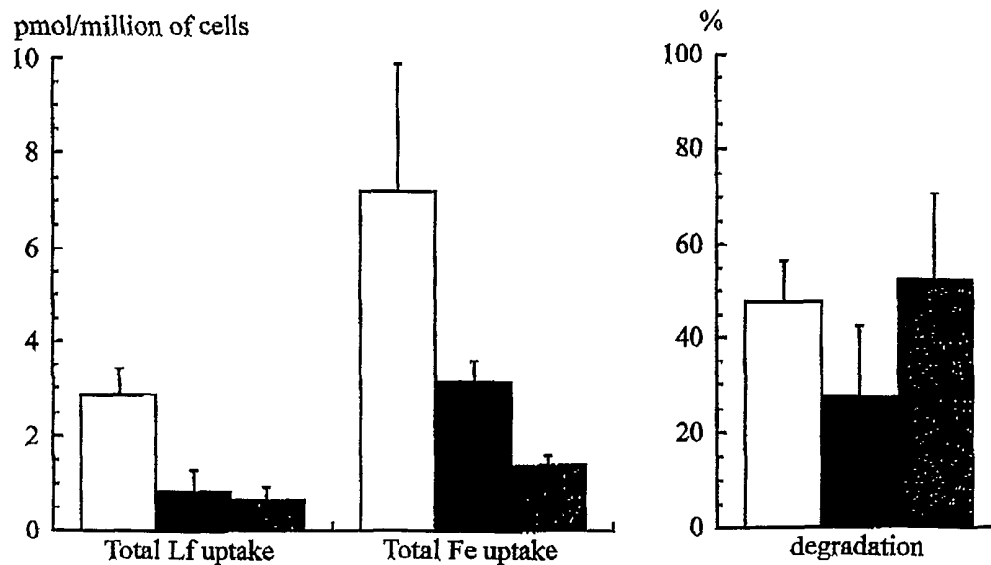
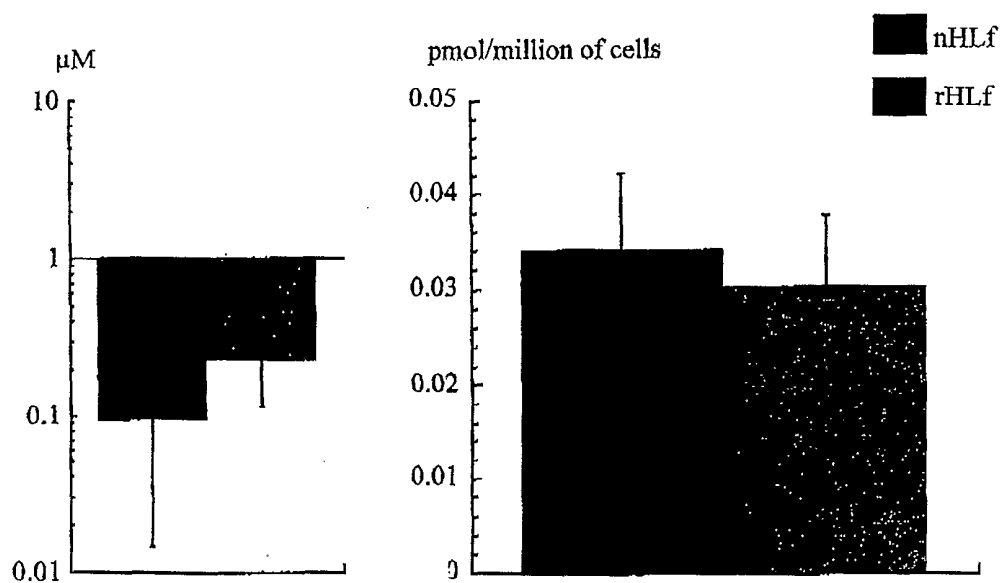
10/45

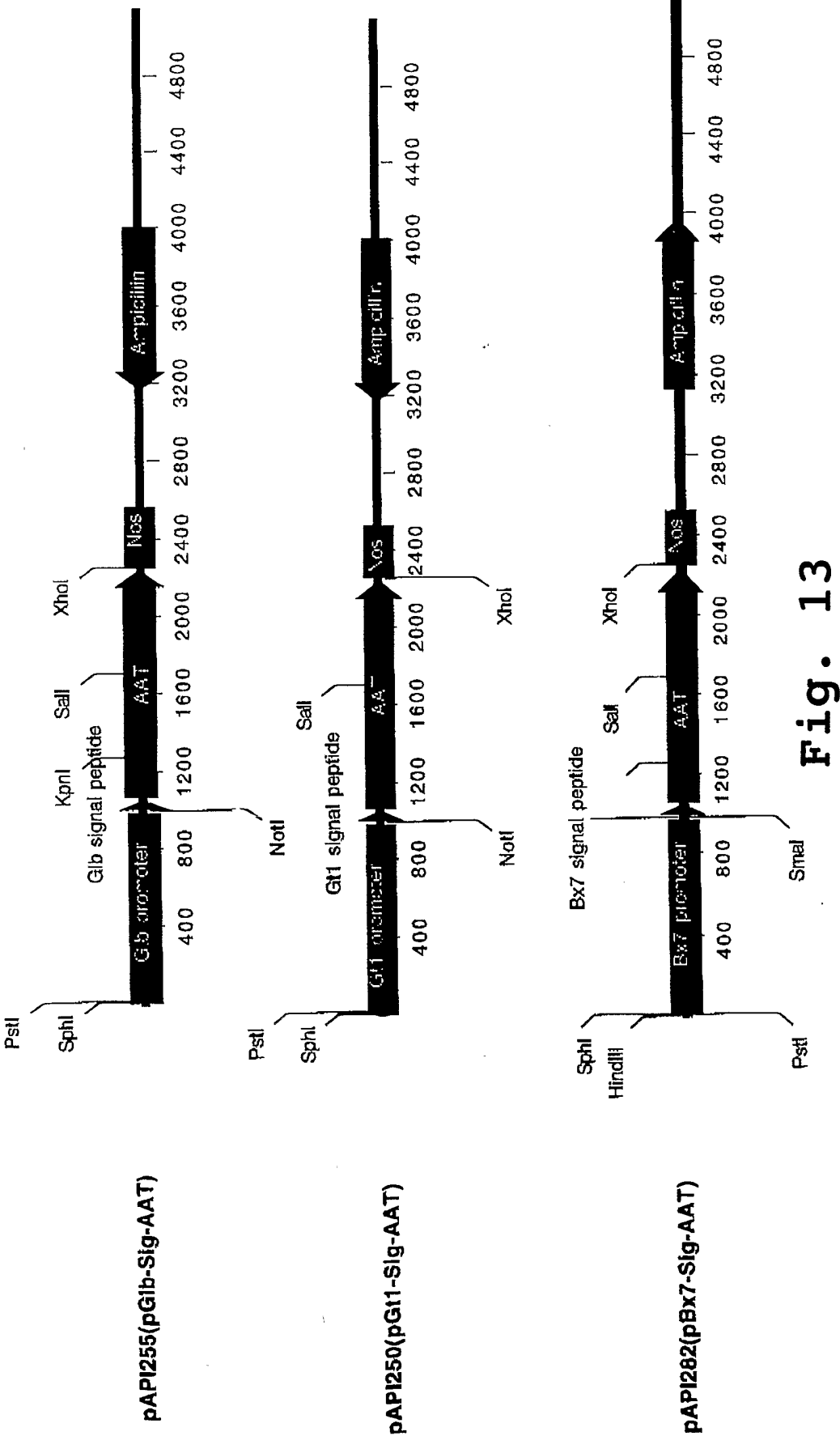
**Fig. 10**

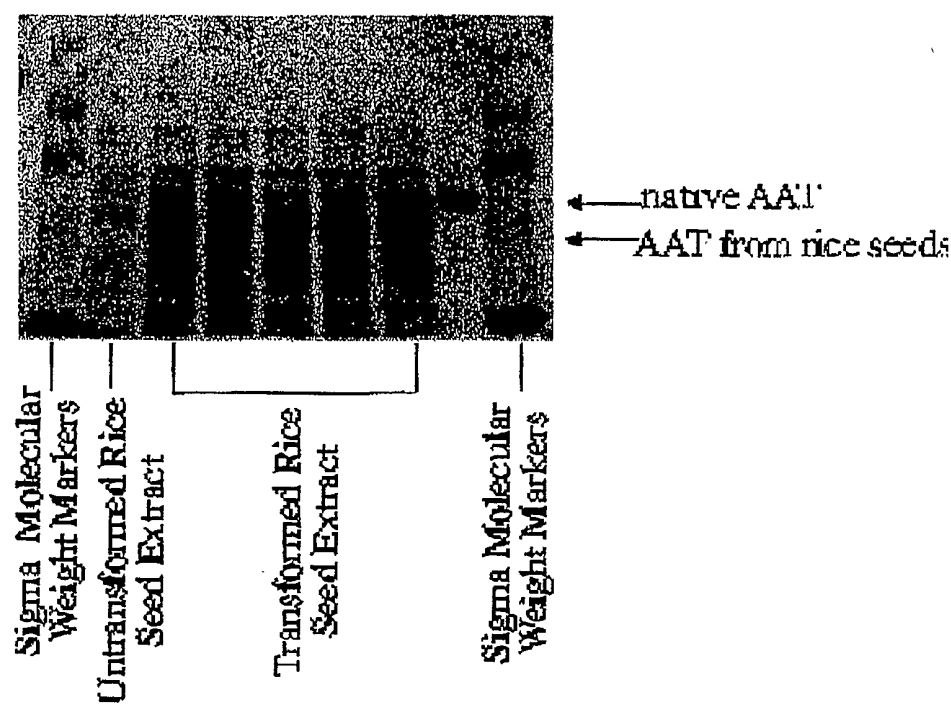
11/45

**Fig. 11**

12/45

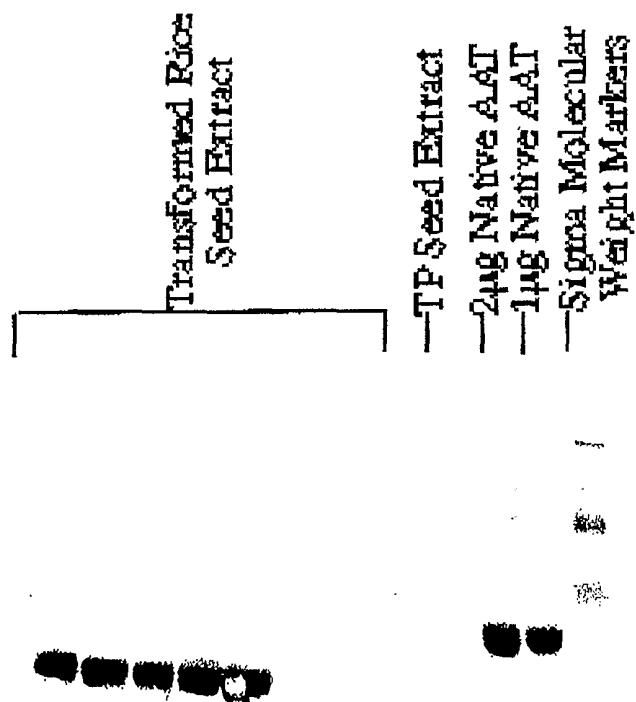
**Fig. 12A**      **Fig. 12B****Fig. 12C**      **Fig. 12D**





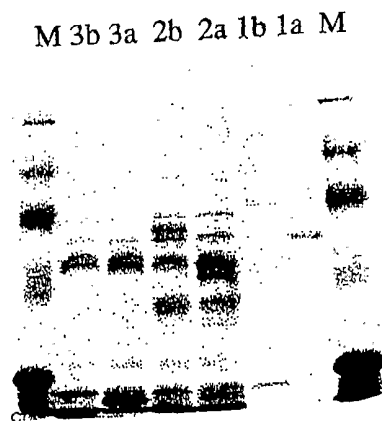
**Fig. 14**

15/45

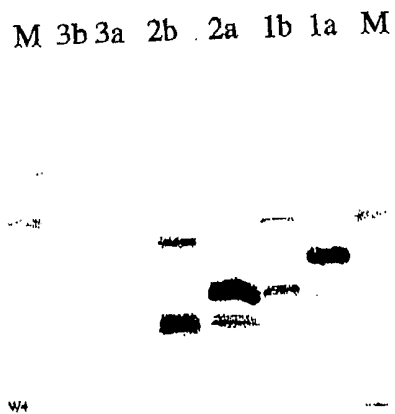
**Fig. 15**

16/45

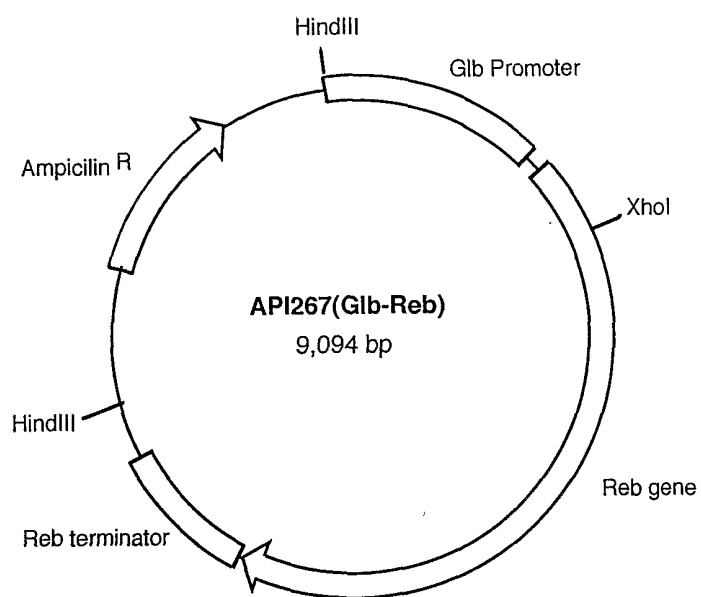
**Fig. 16A**



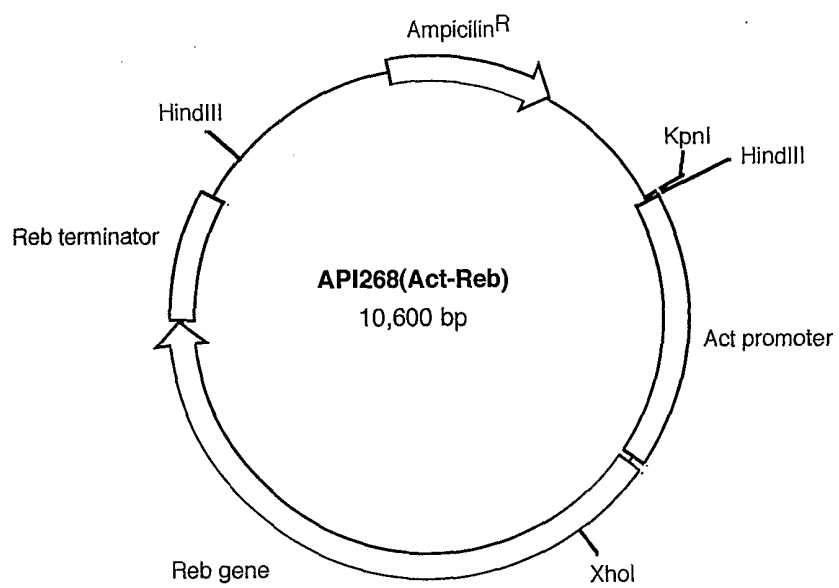
**Fig. 16B**

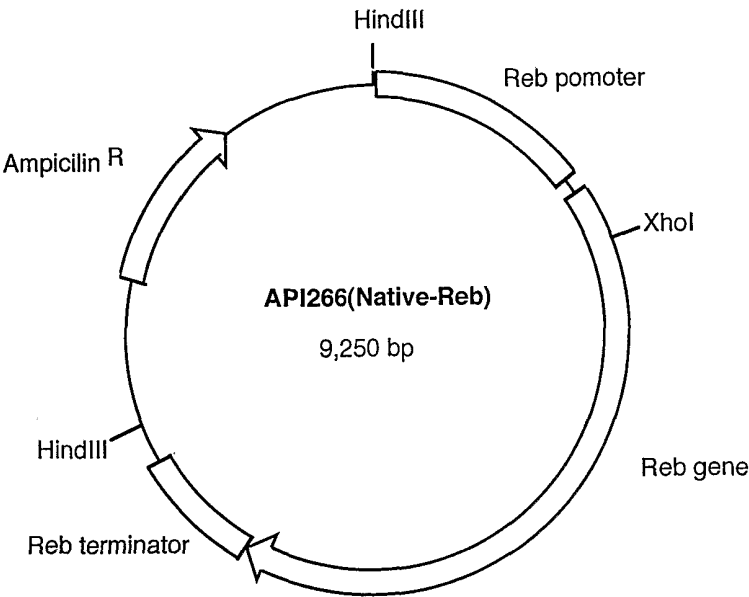


17/45

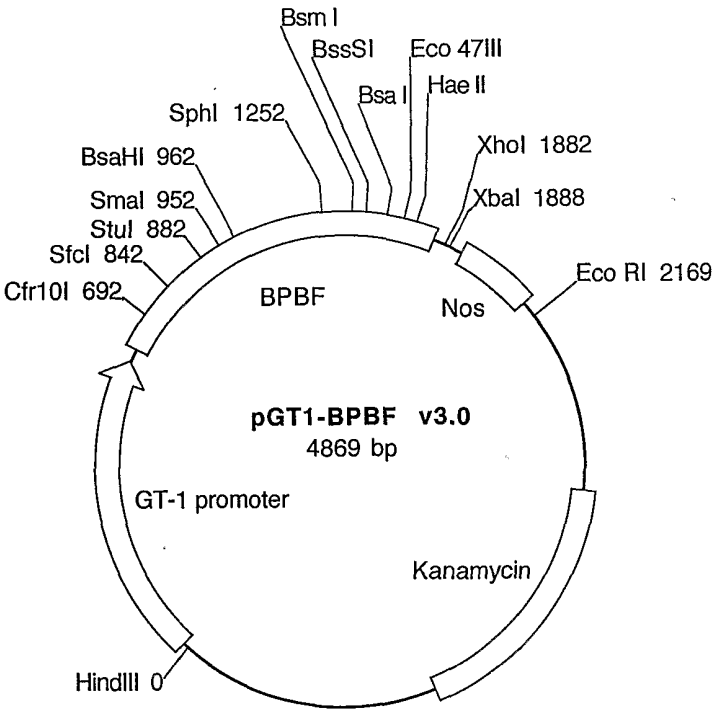
**Fig. 17A**

18/45

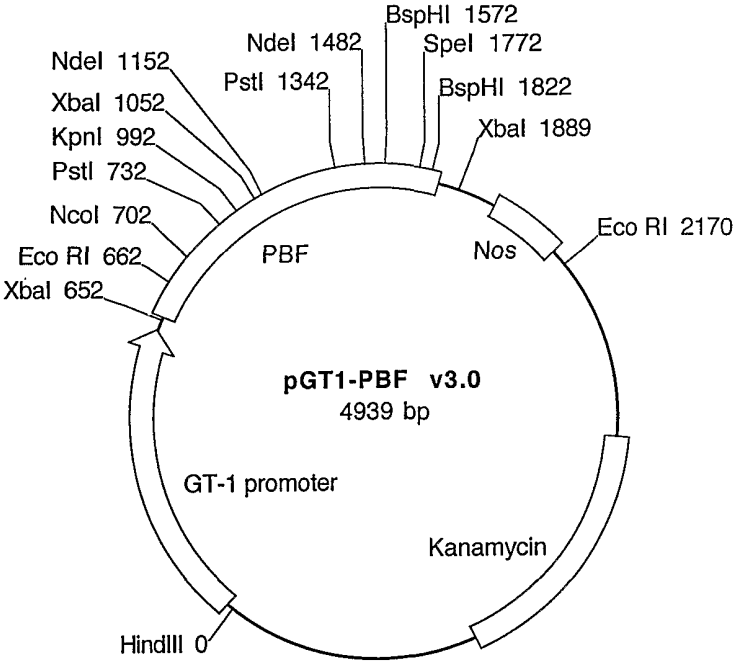
**Fig. 17B**



**Fig. 17C**

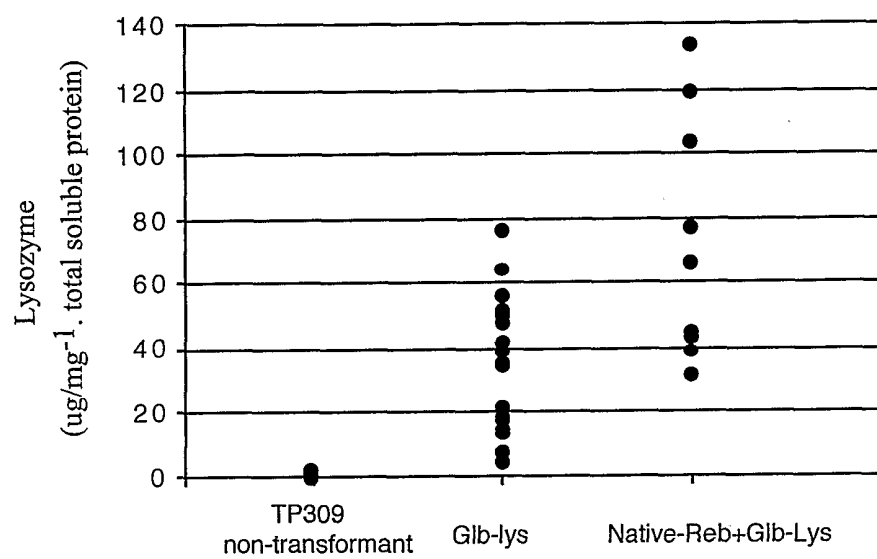


**Fig. 18A**



**Fig. 18B**

22/45

**Fig. 19**

23/45

Sequence Range Egfactor: 4 to 165

|          |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|          |     | 10  |     | 20  |     | 30  |     | 40  |     | 50  |     |     |     |     |     |     |     |     |
| Egfactor | AAC | TCC | GAC | TCG | GAG | TGC | CCC | CTC | TCC | CAC | GAC | GGT | TAC | TGC | CTC | CAC | GAC | GGG |
|          | N   | S   | D   | S   | E   | C   | P   | L   | S   | H   | D   | G   | Y   | C   | L   | H   | D   | G>  |

|             |     |      |     |      |     |      |     |      |     |      |     |     |     |     |     |     |     |     |
|-------------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|
|             |     | 3360 |     | 3370 |     | 3380 |     | 3390 |     | 3400 |     |     |     |     |     |     |     |     |
| Native Gene | AAT | GGT  | GAC | TCT  | GAA | TGT  | CCC | CTG  | TCC | CAC  | GAT | GGG | TAC | TGC | CTC | CAT | GAT | GGT |
|             |     |      |     |      |     |      |     |      |     |      |     |     |     |     |     |     |     |     |
| Egfactor    | AAC | TCC  | GAC | TCG  | GAG | TGC  | CCC | CTC  | TCC | CAC  | GAC | GGT | TAC | TGC | CTC | CAC | GAC | GGG |

|          |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|          | 60  |     | 70  |     | 80  |     | 90  |     | 100 |     | 110 |     |     |     |     |     |     |     |
| Egfactor | GTC | TGC | ATG | TAC | ATC | GAG | GCC | CTC | GAC | AAG | TAC | GCC | TGC | AAC | TGC | GTC | GTG | GGC |
|          | V   | C   | M   | Y   | I   | E   | A   | L   | D   | K   | Y   | A   | C   | N   | C   | V   | V   | G>  |

|             |      |     |      |     |      |     |      |     |      |     |     |     |     |     |     |     |     |     |
|-------------|------|-----|------|-----|------|-----|------|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|             | 3410 |     | 3420 |     | 3430 |     | 3440 |     | 3450 |     |     |     |     |     |     |     |     |     |
| Native Gene | GTG  | TGC | ATG  | TAT | ATT  | GAA | GCA  | TTG | GAC  | AAG | TAT | GCA | TGC | AAC | TGT | GTT | GTT | GGC |
|             |      |     |      |     |      |     |      |     |      |     |     |     |     |     |     |     |     |     |
| Egfactor    | GTC  | TGC | ATG  | TAC | ATC  | GAG | GCC  | CTC | GAC  | AAG | TAC | GCC | TGC | AAC | TGC | GTC | GTG | GGC |

|          |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|          | 120 |     | 130 |     | 140 |     | 150 |     | 160 |     |     |     |     |     |     |     |     |     |
| Egfactor | TAC | ATC | GGC | GAG | CGG | TGC | CAG | TAC | CGC | GAC | CTC | AAG | TGG | TGG | GAG | CTG | CGC | TGA |
|          | Y   | I   | G   | E   | R   | C   | Q   | Y   | R   | D   | L   | K   | W   | W   | E   | L   | R   | *>  |

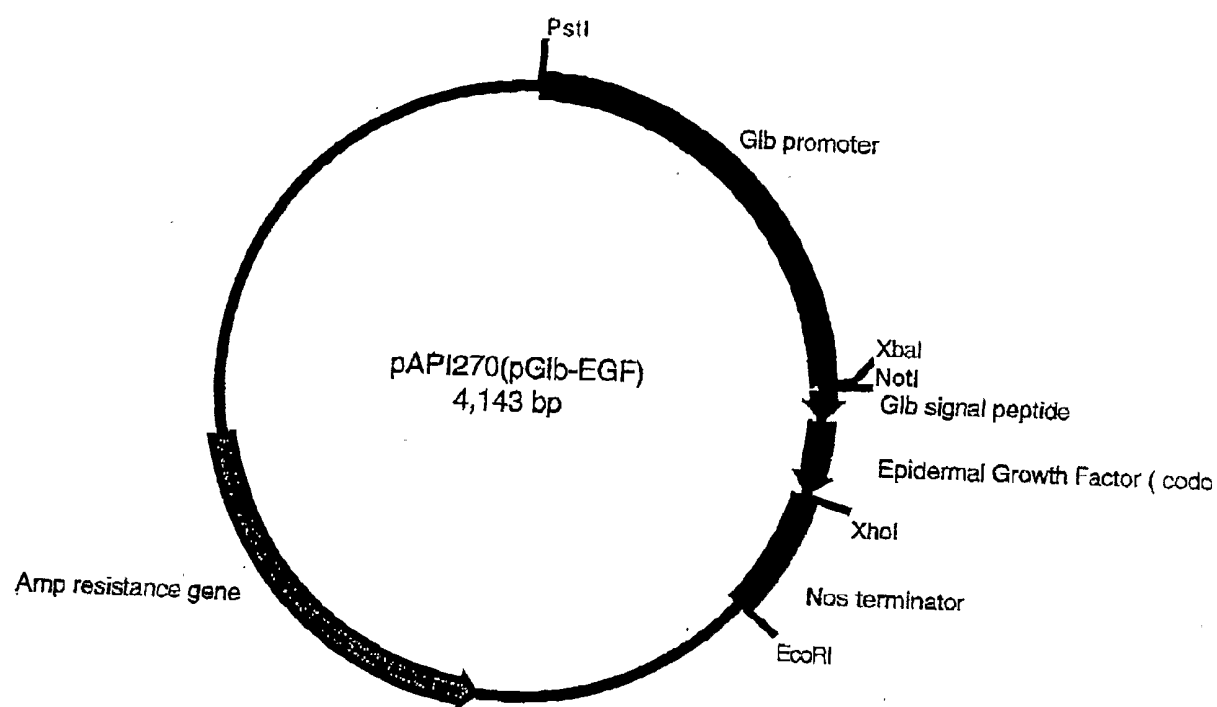
|             |      |     |      |     |      |     |      |     |      |     |     |     |     |     |     |     |     |
|-------------|------|-----|------|-----|------|-----|------|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|
|             | 3460 |     | 3470 |     | 3480 |     | 3490 |     | 3500 |     |     |     |     |     |     |     |     |
| Native Gene | TAC  | ATC | GGG  | GAG | CGA  | TGT | CAG  | TAC | CGA  | GAC | CTG | AAG | TGG | TGG | GAA | CTG | CGC |
|             |      |     |      |     |      |     |      |     |      |     |     |     |     |     |     |     |     |
| Egfactor    | TAC  | ATC | GGC  | GAG | CGG  | TGC | CAG  | TAC | CGC  | GAC | CTC | AAG | TGG | TGG | GAG | CTG | CGC |

## Epidermal Growth Factor

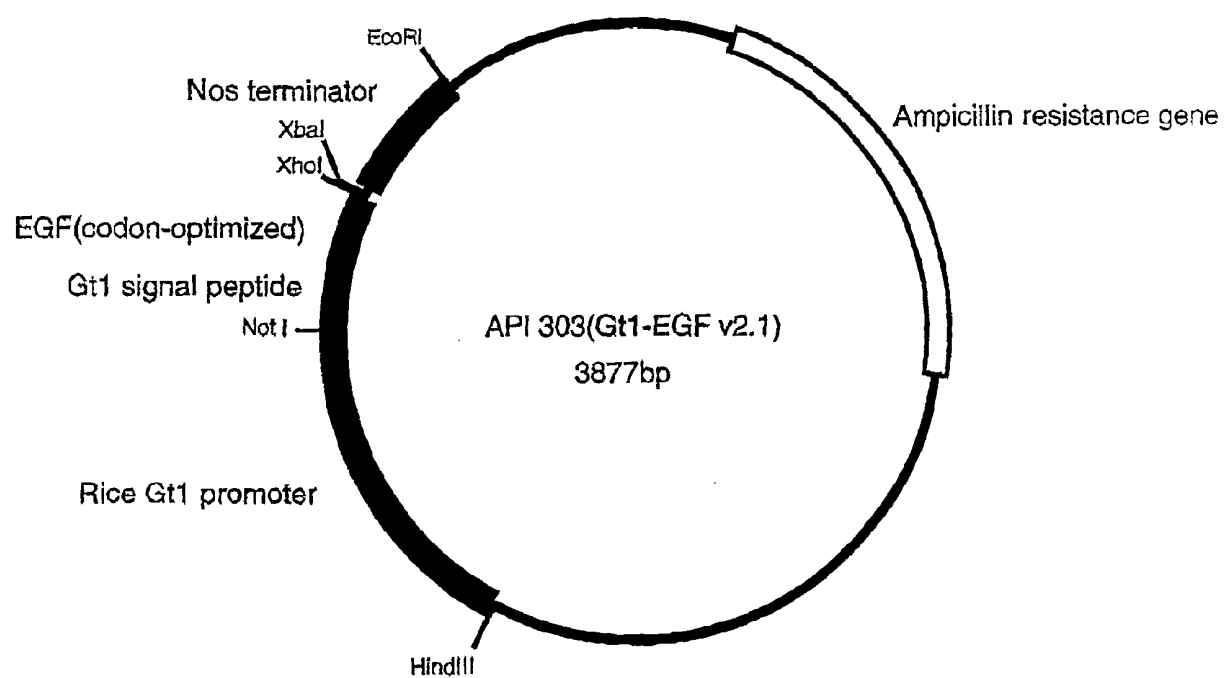
Number of codons in mature peptide: 53  
 Number of codons changed: 27 (51%)  
 Number of nucleotides changed: 30 (19%)

Fig. 20

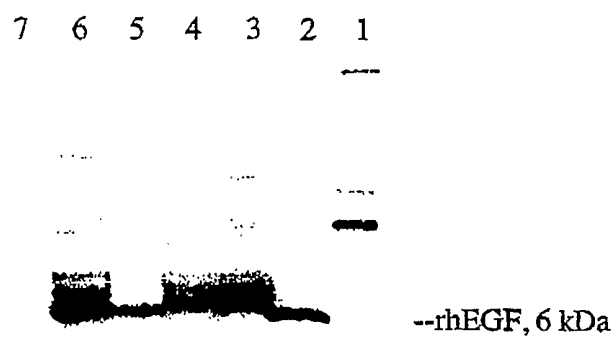
24/45

**Fig. 21**

25/45

**Fig. 22**

26/45

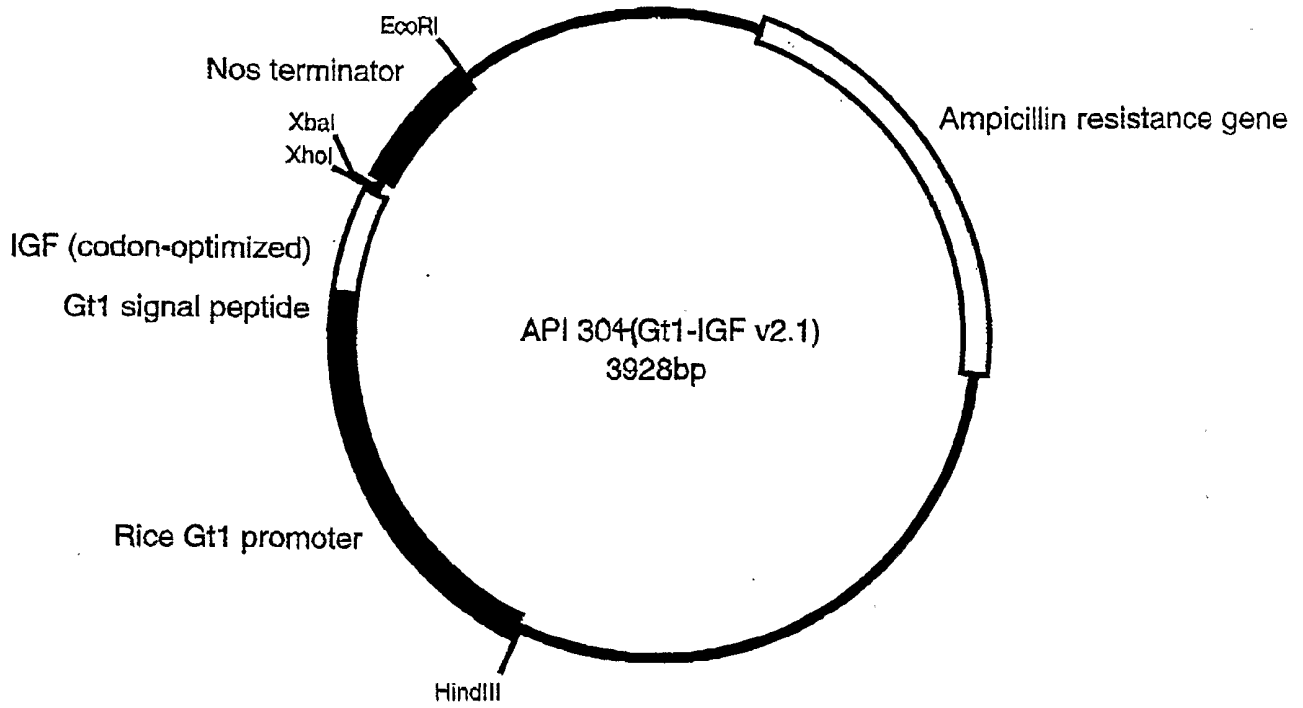


**Fig. 23**

Sequence Range: 4 to 216

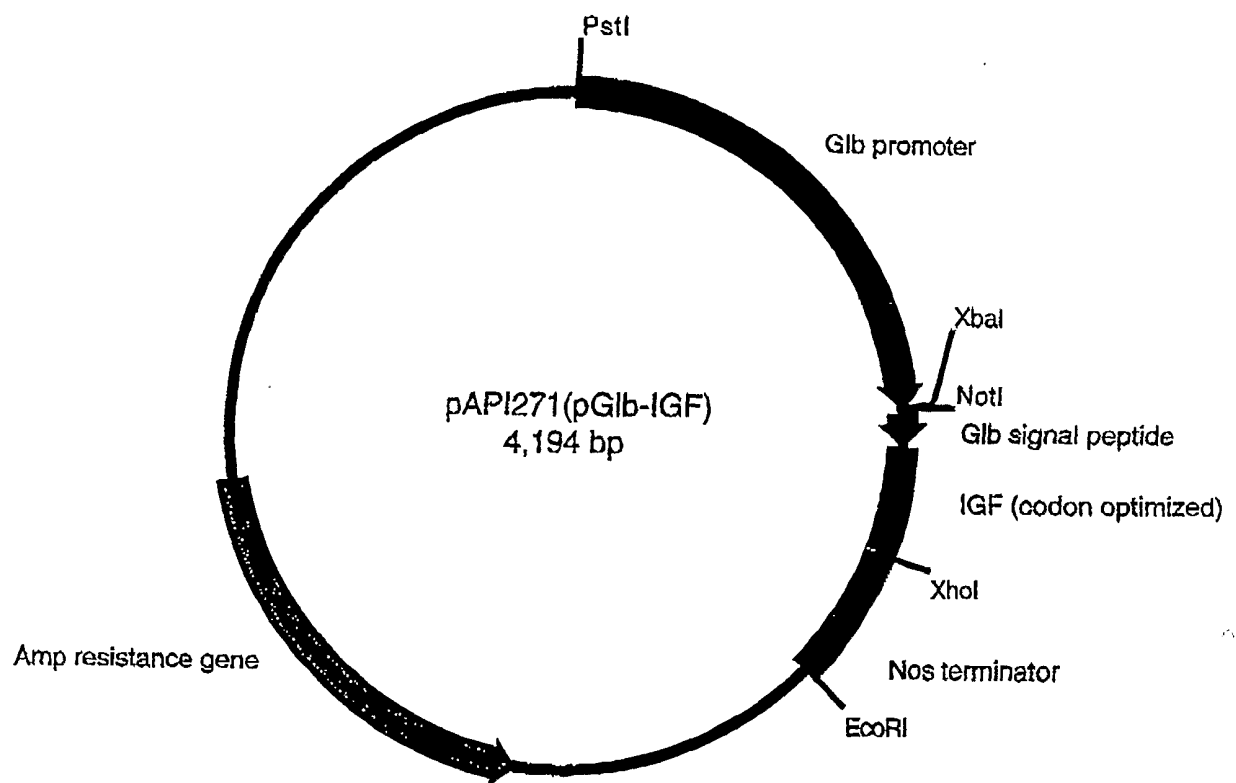
**Fig. 24**

28/45



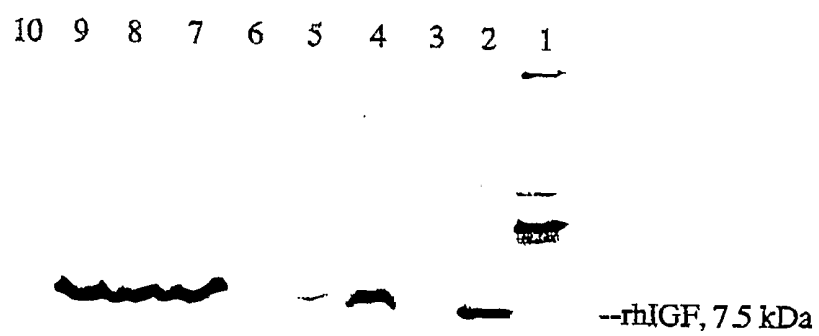
**Fig. 25**

29/45

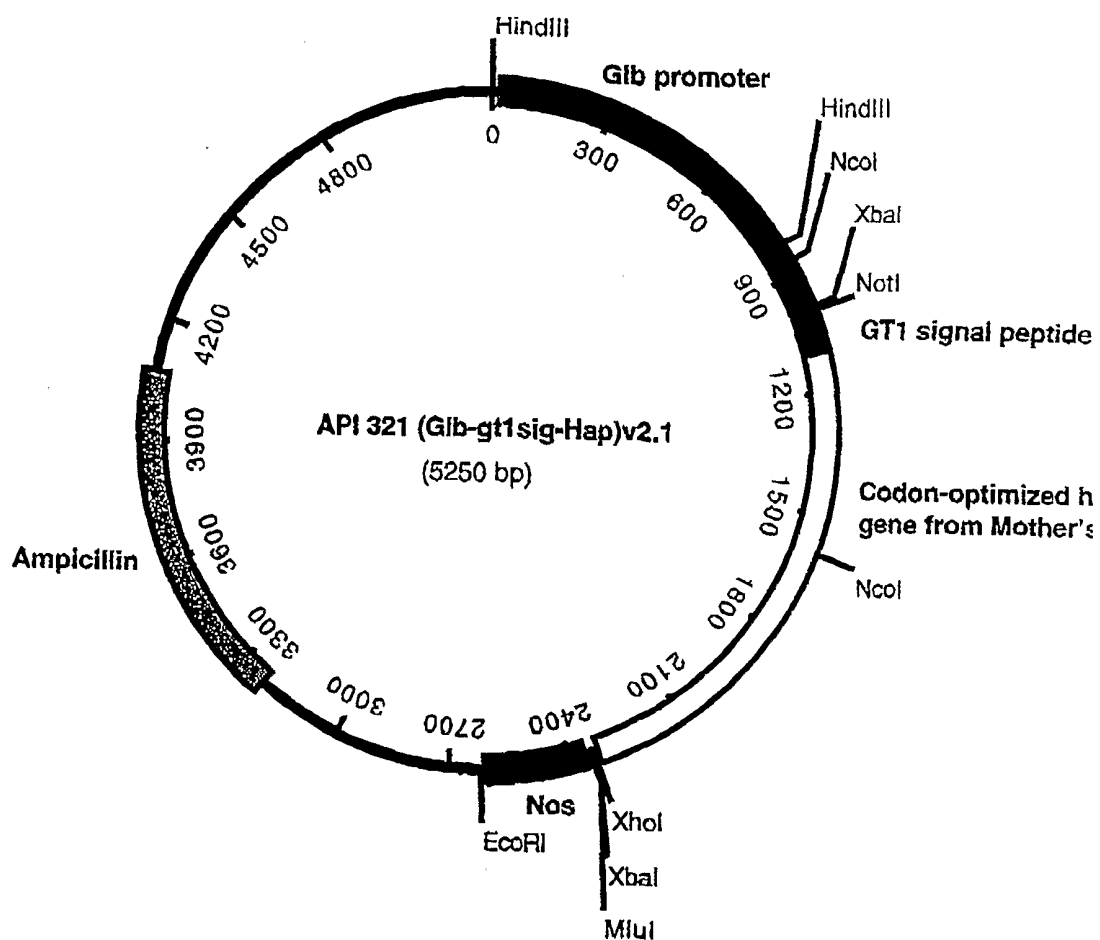


**Fig. 26**

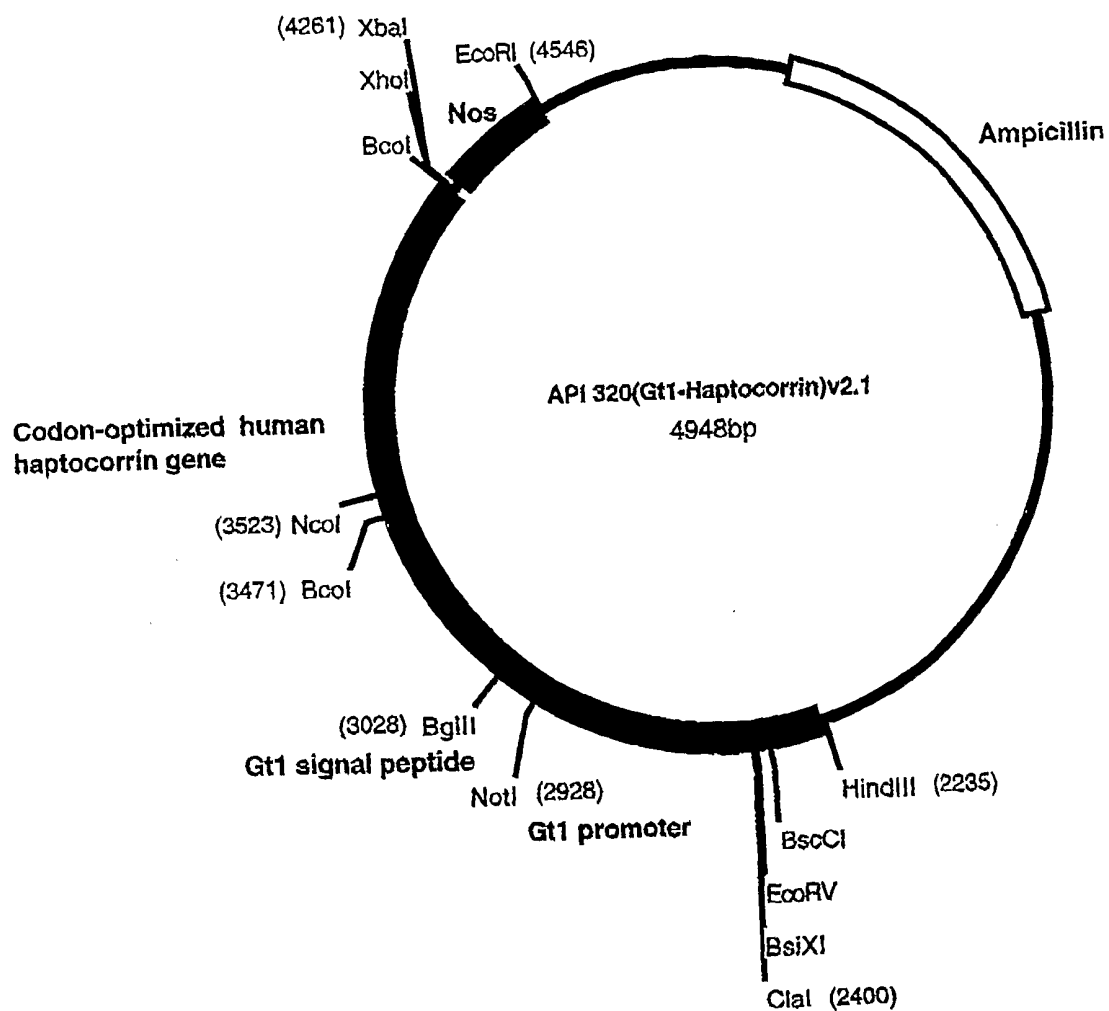
30/45

**Fig. 27**

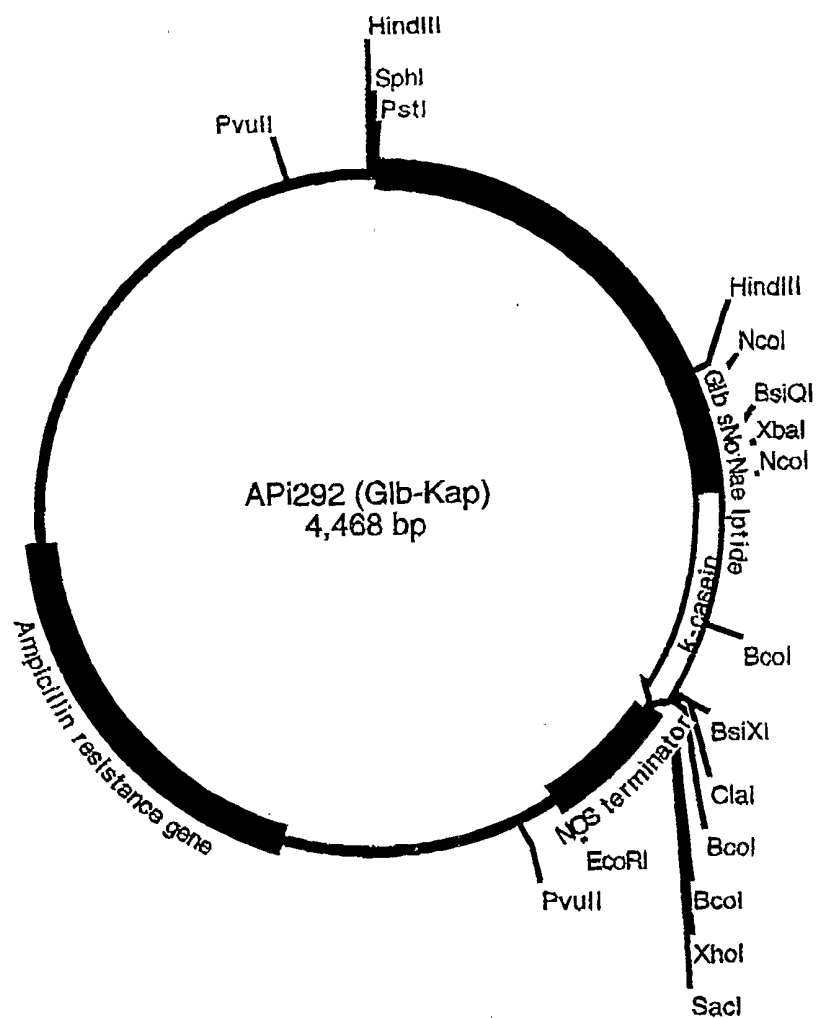
31/45

**Fig. 28**

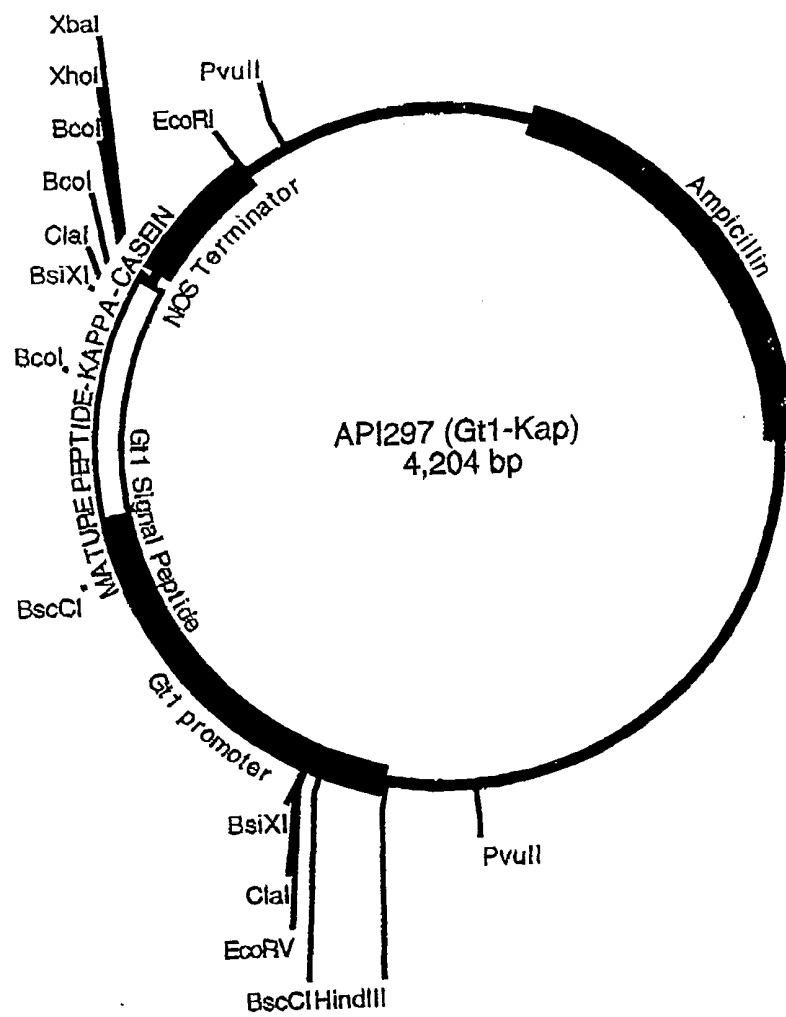
32/45

**Fig. 29**

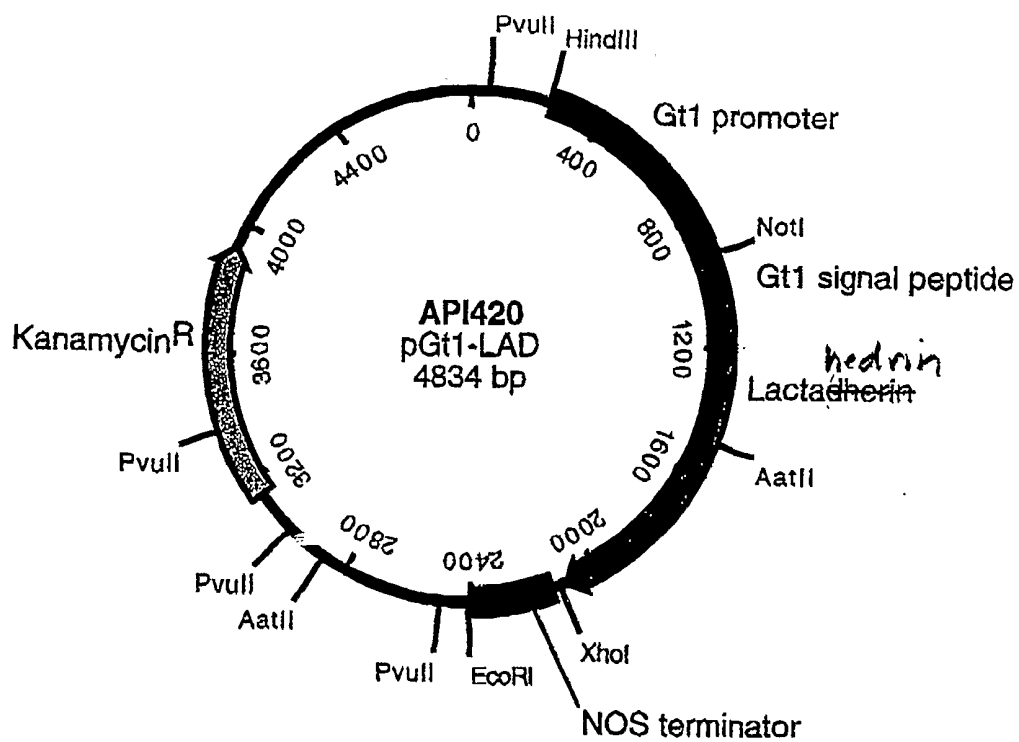
33/45

**Fig. 30**

34/45

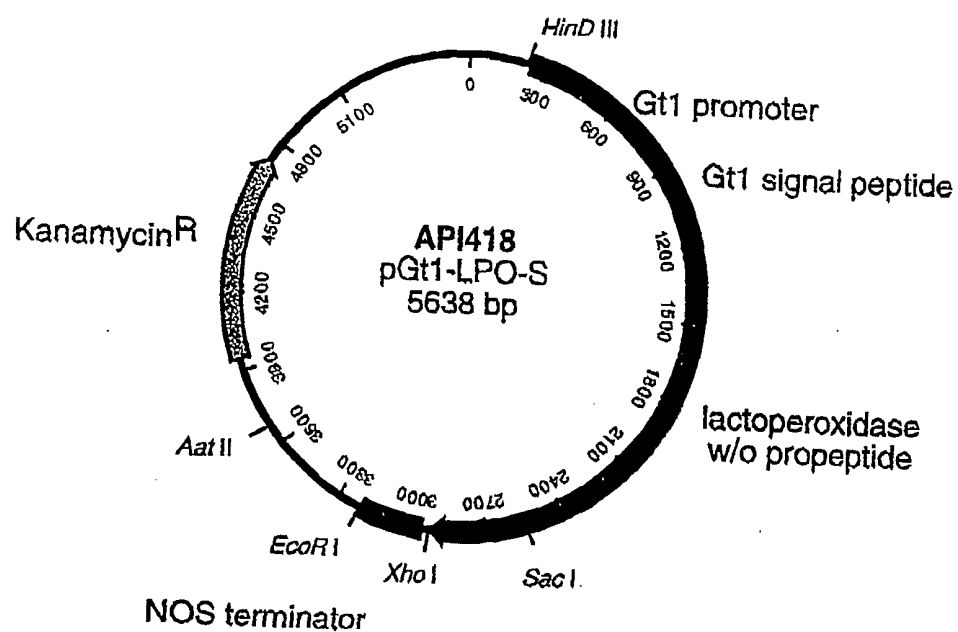
**Fig. 31**

35/45

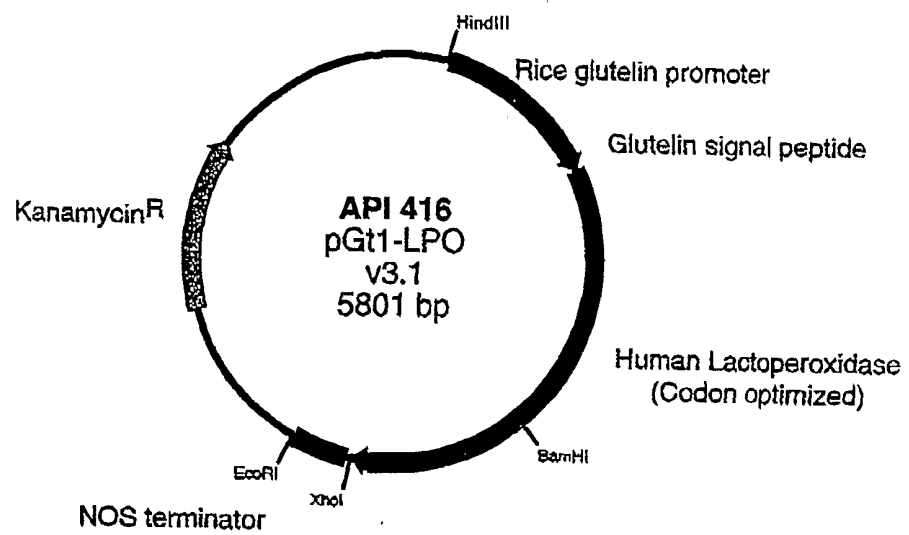


**Fig. 32**

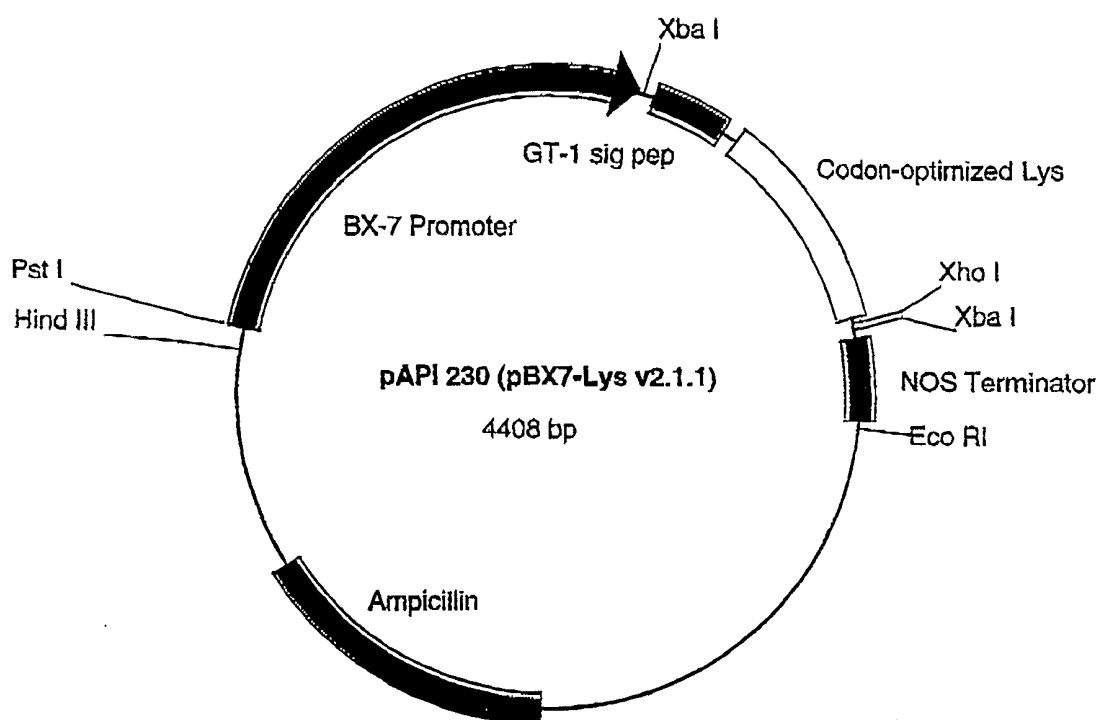
36/45

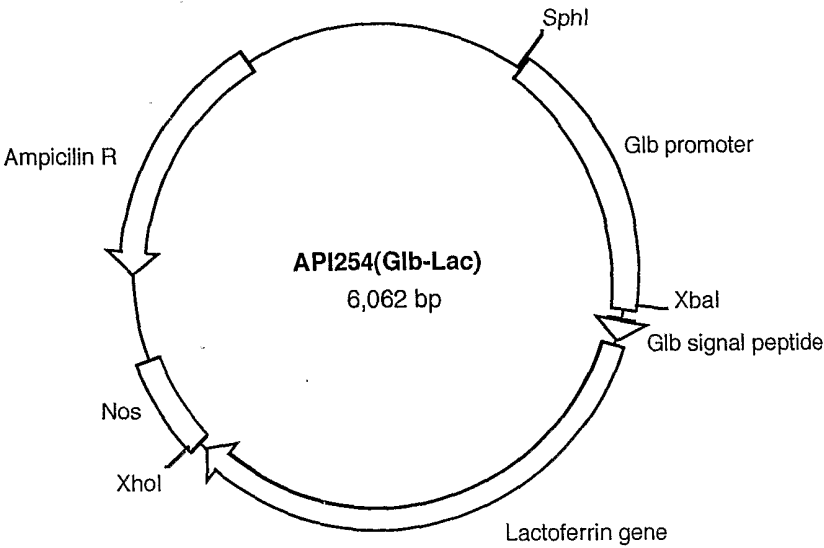
**Fig. 33**

37/45

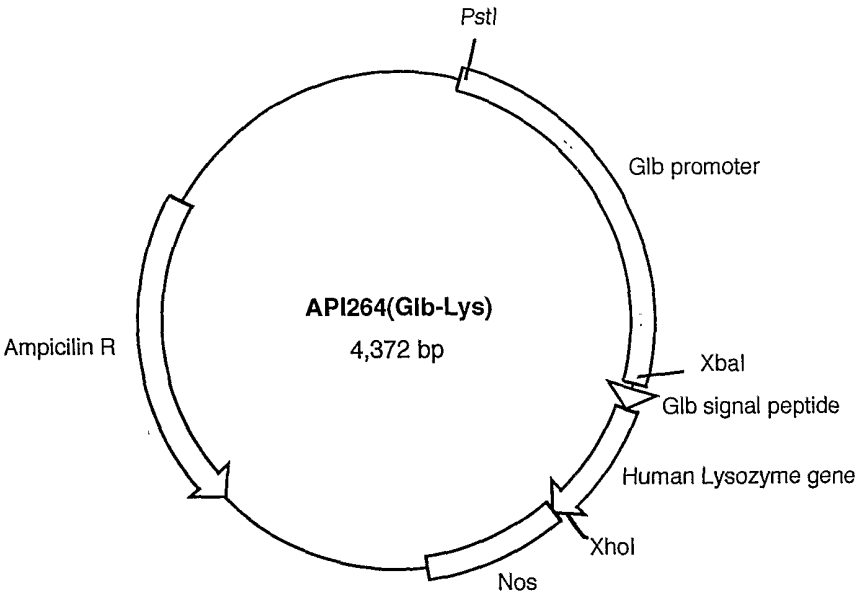
**Fig. 34**

38/45

**Fig. 35**

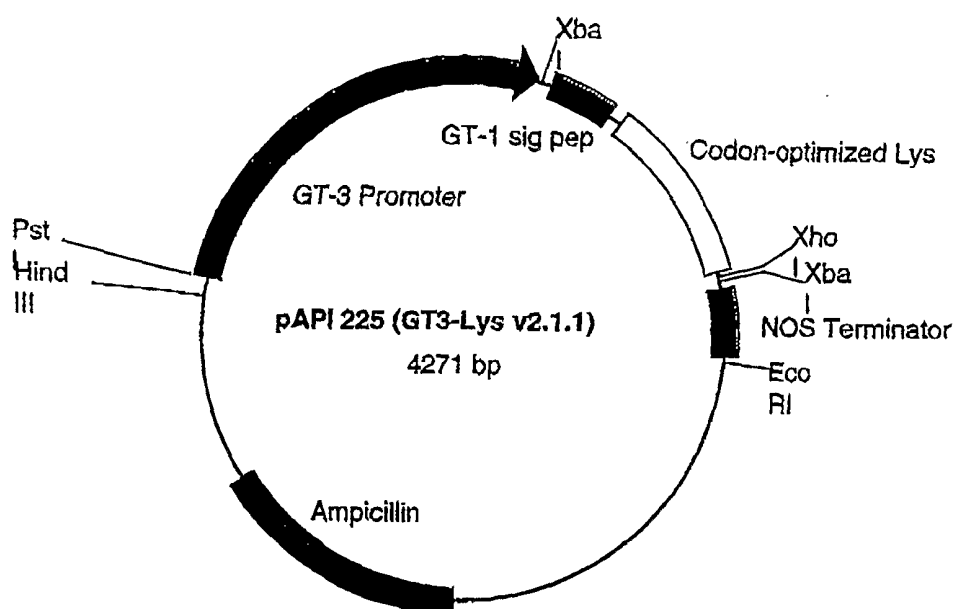


**Fig. 36A**

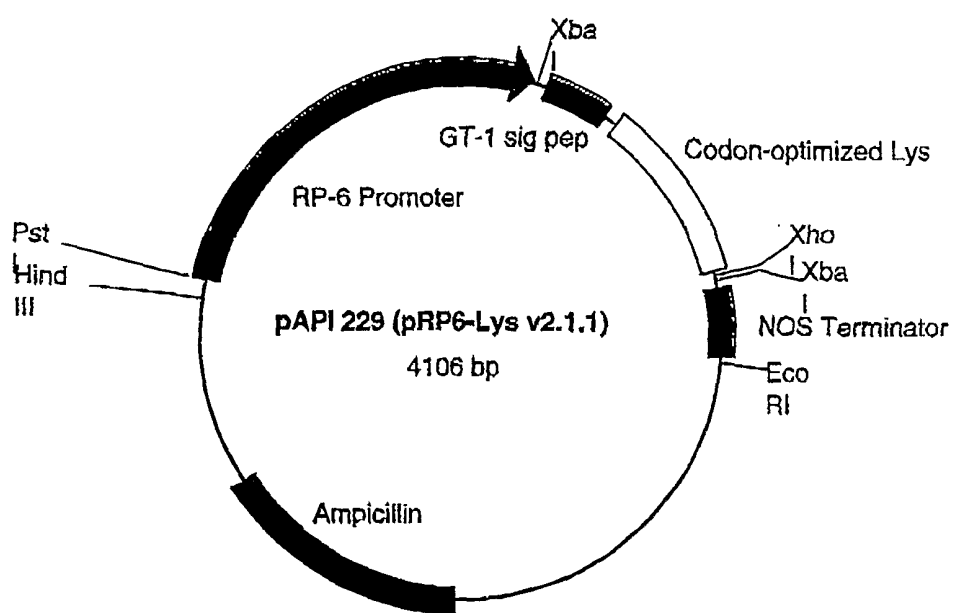


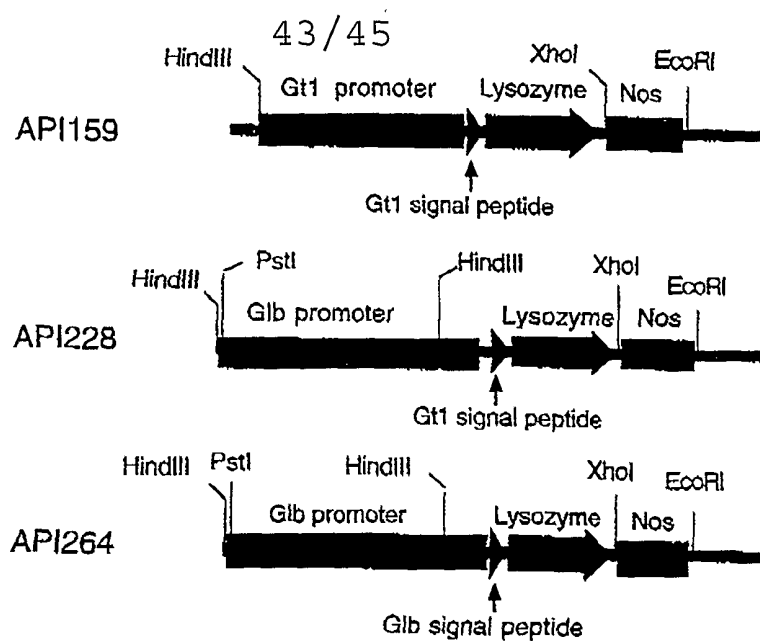
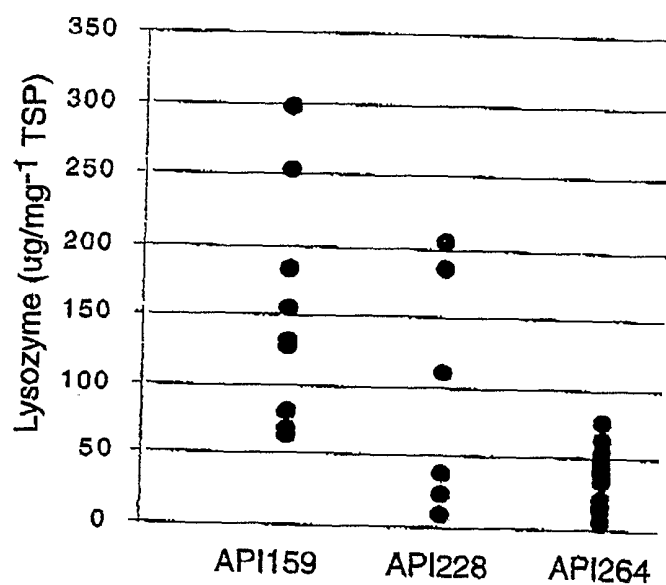
**Fig. 36B**

41/45

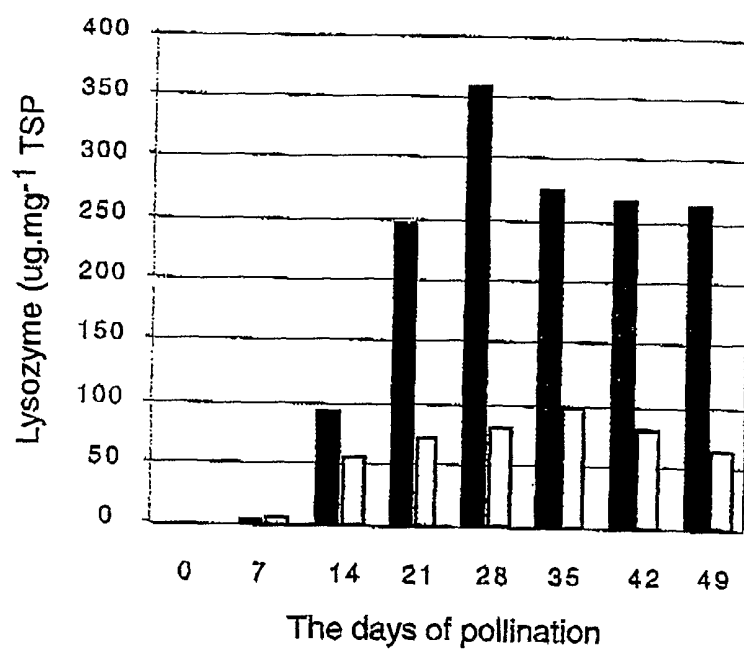
**Fig. 37**

42/45

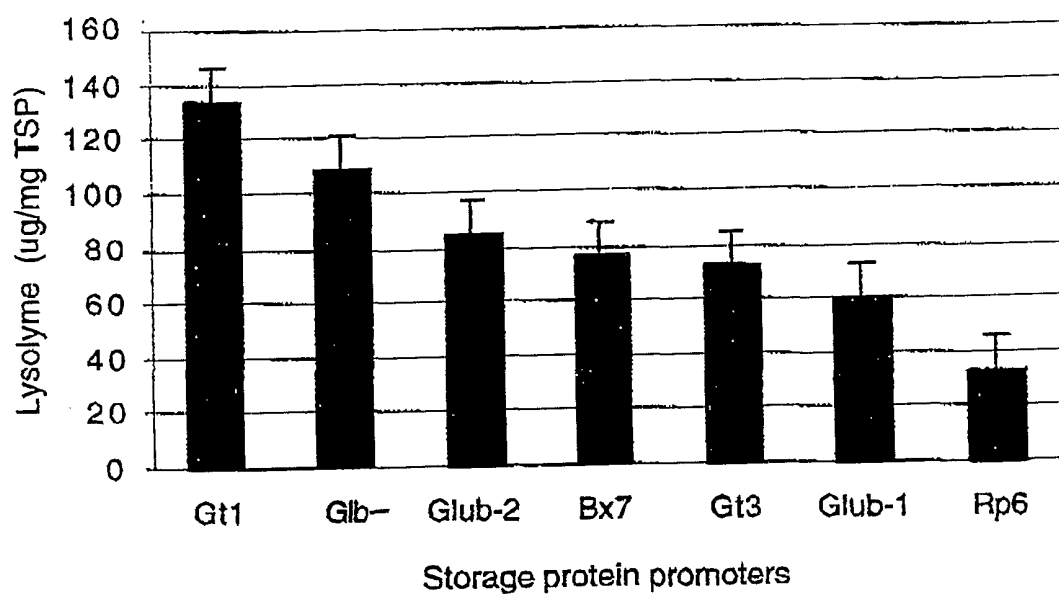
**Fig. 38**

**Fig. 39A****Fig. 39B**

44/45

**Fig. 40**

45/45

**Fig. 41**

## SEQUENCE LISTING

&lt;110&gt; Ventria Bioscience

<120> Expression of Human Milk Proteins in  
Transgenic Plants

&lt;130&gt; 50665-8022.WO00

&lt;140&gt; Not Yet Assigned

&lt;141&gt; Filed Herewith

&lt;150&gt; US 60/269,199

&lt;151&gt; 2001-02-14

&lt;150&gt; US 09/847,232

&lt;151&gt; 2001-05-02

&lt;160&gt; 33

&lt;170&gt; FastSEQ for Windows Version 4.0

&lt;210&gt; 1

&lt;211&gt; 393

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 1

```

aaagtcttcg agcgggtgcga gctggcccg acgtcaagc ggctcgcat ggacggctac      60
cggggcatca gcctcgccaa ctggatgtgc ctgcgaagt gggagtcggg ctacaacacc      120
cgcgcaacca actacaacgc cggcgaccgc tccaccgact acggcatctt ccagatcaac      180
tcccgctact ggtgcaacga cggcaagacg cccggggccg tcaacgcctg ccacctctcc      240
tgctcggccc tgctgcaaga caacatcgcc gacgccgctg cgtgcgcgaa gcgcgtcgtc      300
cgcgacccgc agggcatccg ggctgggtg gcctggcgca accgctgcc gaaccgggac      360
gtgcgccagt acgtccaggg ctgcgcgctc tga                                393

```

&lt;210&gt; 2

&lt;211&gt; 130

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 2

```

Lys Val Phe Glu Arg Cys Glu Leu Ala Arg Thr Leu Lys Arg Leu Gly
 1             5             10             15
Met Asp Gly Tyr Arg Gly Ile Ser Leu Ala Asn Trp Met Cys Leu Ala
 20             25             30
Lys Trp Glu Ser Gly Tyr Asn Thr Arg Ala Thr Asn Tyr Asn Ala Gly
 35             40             45
Asp Arg Ser Thr Asp Tyr Gly Ile Phe Gln Ile Asn Ser Arg Tyr Trp
 50             55             60
Cys Asn Asp Gly Lys Thr Pro Gly Ala Val Asn Ala Cys His Leu Ser
 65             70             75             80
Cys Ser Ala Leu Leu Gln Asp Asn Ile Ala Asp Ala Val Ala Cys Ala
 85             90             95
Lys Arg Val Val Arg Asp Pro Gln Gly Ile Arg Ala Trp Val Ala Trp
100            105            110
Arg Asn Arg Cys Gln Asn Arg Asp Val Arg Gln Tyr Val Gln Gly Cys
115            120            125
Gly Val
130

```

&lt;210&gt; 3

&lt;211&gt; 2079

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 3

```

gggcggcgccg ggcgctcggg gcagtgggtgc gccgtgtccc agcccgaggc gaccaagtgc      60
ttccagtggc agcgcaacat gcggaagggtg cggggcccgc cggtcagctg catcaagcgg      120
gactccccc tccaatgcat ccaggccatc gcggaagaacc gcgccgacgc ggtcaccctg      180
gacggcgggg tcatctacga ggcggggctc gcccgtaca agctccgccc ggtggcgggc      240
gaggtgtacg gcaccgagcg ccagccgcgc acgcactact acgcggtggc cgtcgtcaag      300
aaggcggggt ccttccagct caacgagctg cagggcctga agtcgtgcca cacgggcctc      360
cggcgagcgg cgggctggaa cgtgccatc ggcaccctgc gccccttctt gaactggacc      420
ggcccgccgg agccgatcga ggccgcgtg gcccgttctt tcagcgcctc ctgctgcccc      480
ggcgccgaca agggccagtt cccgaacctc tgccggctct gcgccgggac gggcgagaac      540
aagtgcgcct tctcctcgca ggagccgtac ttctcctact cgggcgcgtt caagtgcctc      600
cgcgacgggg ccggcgacgt ggcgttcac cgcgagtcca ccgtgttcga ggacctctcc      660
gacgagggcg agcgggacga gtacgagctg ctgtgccccg acaacaccgg caagccggtg      720
gacaagttca aggactgcca cctggcgcg gtgcccctgc acgcggtcgt cgcccgacg      780
gtcaacggca aggaggacgc gatctggaac ctccctccgc agggccagga gaagttcggc      840
aaggacaagt cccccaagtt ccagctcttc gggagcccca gcggccagaa ggacctctc      900
ttcaaggact ccgcatcg cttctccgc gtccccccgc gcatcgactc cggcctgtac      960
ctcggctccg ggtacttcac cgcgatccag aacctccgga agagcgagga ggaggtggcg      1020
gcgcggcggg cccgcgtcgt gtgtgcgcc gtggcgagc aggagctgcg gaagtgaac      1080
cagtggagcg gcctgagcga ggggtcgtg acctgctcgt ccgccagcac caccgaggac      1140
tgcatcgcg tcgtcctcaa gggggaggcc gacgcgatga gcctcgacgg ggggtacgtc      1200
tacaccgccc gcaagtgcgg cctggtcccg gtccctggcg agaactaaa gtcgcagcag      1260
tccagcgacc ccgaccgaa ctgctgggac cgcgccgtcg agggctacct cgcctggcc      1320
gtcgtgcgcc ggtccgacac ctccctgacg tggaaacagcg tcaagggcaa gaagagctgc      1380
cacaccgccc tggaccgcac cgccggctgg aacatcccga tgggcctcct cttcaaccag      1440
accggctcct gcaagttcga cgagtacttc tccagtcct gcgcccccg ctcggacccc      1500
cgctccaacc tgtgcgccct ctgcatcggg gacgagcagg gcgagaacaa gtgctgccc      1560
aacagcaacg agcggacta cggctacacg ggggccttcc gctgcctggc ggagaacgcc      1620
ggggacgtcg cgttcgtgaa ggacgtgacc gtgctgcaaa acacggacgg gaacaacaac      1680
gaggcgtggg cgaaggacct caagctcgcc gacttcgccc tgctgtgcct cgacggcaag      1740
cgcaagcccc tcaccgagcg cgggtcctgc cacttgogga tggcccccaa ccacgcgctc      1800
gtctcccgca tggacaaggt cgagcgctc aagcaggtgc tcctgcacca gcaggccaag      1860
ttcggccgga acggcagcga ctgcccggac aagttctgcc tgttccagtc ggagaccaag      1920
aacctcctct tcaacgacaa caccgagtgc ctggcgcgcc tccacggcaa gaccacctac      1980
gagaagtacc tcggcccgca gtacgtcgcc ggcataacca acctcaagaa gtgctccacc      2040
tccccctcc tggaggcgtg cgagttcctc cgcaagtga      2079

```

&lt;210&gt; 4

&lt;211&gt; 690

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 4

```

Gly Arg Arg Arg Arg Ser Val Gln Trp Cys Ala Val Ser Gln Pro Glu
 1          5          10          15
Ala Thr Lys Cys Phe Gln Trp Gln Arg Asn Met Arg Lys Val Arg Gly
 20          25          30
Pro Pro Val Ser Cys Ile Lys Arg Asp Ser Pro Ile Gln Cys Ile Gln
 35          40          45
Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Thr Leu Asp Gly Gly Phe
 50          55          60
Ile Tyr Glu Ala Gly Leu Ala Pro Tyr Lys Leu Arg Pro Val Ala Ala
 65          70          75          80
Glu Val Tyr Gly Thr Glu Arg Gln Pro Arg Thr His Tyr Tyr Ala Val
 85          90          95
Ala Val Val Lys Lys Gly Gly Ser Phe Gln Leu Asn Glu Leu Gln Gly
 100         105         110
Leu Lys Ser Cys His Thr Gly Leu Arg Arg Thr Ala Gly Trp Asn Val
 115         120         125
Pro Ile Gly Thr Leu Arg Pro Phe Leu Asn Trp Thr Gly Pro Pro Glu
 130         135         140
Pro Ile Glu Ala Ala Val Ala Arg Phe Phe Ser Ala Ser Cys Val Pro
 145         150         155         160

```

Gly Ala Asp Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu Cys Ala Gly  
 165 170 175  
 Thr Gly Glu Asn Lys Cys Ala Phe Ser Ser Gln Glu Pro Tyr Phe Ser  
 180 185 190  
 Tyr Ser Gly Ala Phe Lys Cys Leu Arg Asp Gly Ala Gly Asp Val Ala  
 195 200 205  
 Phe Ile Arg Glu Ser Thr Val Phe Glu Asp Leu Ser Asp Glu Ala Glu  
 210 215 220  
 Arg Asp Glu Tyr Glu Leu Leu Cys Pro Asp Asn Thr Arg Lys Pro Val  
 225 230 235 240  
 Asp Lys Phe Lys Asp Cys His Leu Ala Arg Val Pro Ser His Ala Val  
 245 250 255  
 Val Ala Arg Ser Val Asn Gly Lys Glu Asp Ala Ile Trp Asn Leu Leu  
 260 265 270  
 Arg Gln Ala Gln Glu Lys Phe Gly Lys Asp Lys Ser Pro Lys Phe Gln  
 275 280 285  
 Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu Leu Phe Lys Asp Ser  
 290 295 300  
 Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser Gly Leu Gly  
 305 310 315 320  
 Ser Gly Tyr Phe Thr Ala Ile Gln Asn Leu Arg Lys Ser Glu Glu Glu  
 325 330 335  
 Val Ala Ala Arg Arg Ala Arg Val Val Trp Cys Ala Val Gly Glu Gln  
 340 345 350  
 Glu Leu Arg Lys Cys Asn Gln Trp Ser Gly Leu Ser Glu Gly Ser Val  
 355 360 365  
 Thr Cys Ser Ser Ala Ser Thr Thr Glu Asp Cys Ile Ala Leu Val Leu  
 370 375 380  
 Lys Gly Glu Ala Asp Ala Met Ser Leu Asp Gly Gly Tyr Val Tyr Thr  
 385 390 395 400  
 Ala Gly Lys Cys Gly Leu Val Pro Val Leu Ala Glu Asn Tyr Lys Ser  
 405 410 415  
 Gln Gln Ser Ser Asp Pro Asp Pro Asn Cys Val Asp Arg Pro Val Glu  
 420 425 430  
 Gly Tyr Leu Ala Val Ala Val Val Arg Arg Ser Asp Thr Ser Leu Thr  
 435 440 445  
 Trp Asn Ser Val Lys Gly Lys Lys Ser Cys His Thr Ala Val Asp Arg  
 450 455 460  
 Thr Ala Gly Trp Asn Ile Pro Met Gly Leu Leu Phe Asn Gln Thr Gly  
 465 470 475 480  
 Ser Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser Cys Ala Pro Gly Ser  
 485 490 495  
 Asp Pro Arg Ser Asn Leu Cys Ala Leu Cys Ile Gly Asp Glu Gln Gly  
 500 505 510  
 Glu Asn Lys Cys Val Pro Asn Ser Asn Glu Arg Tyr Tyr Gly Tyr Thr  
 515 520 525  
 Gly Ala Phe Arg Cys Leu Ala Glu Asn Ala Gly Asp Val Ala Phe Val  
 530 535 540  
 Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly Asn Asn Asn Glu Ala  
 545 550 555 560  
 Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala Leu Leu Cys Leu Asp  
 565 570 575  
 Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser Cys His Leu Ala Met  
 580 585 590  
 Ala Pro Asn His Ala Val Val Ser Arg Met Asp Lys Val Glu Arg Leu  
 595 600 605  
 Lys Gln Val Leu Leu His Gln Ala Lys Phe Gly Arg Asn Gly Ser  
 610 615 620  
 Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser Glu Thr Lys Asn Leu  
 625 630 635 640  
 Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg Leu His Gly Lys Thr  
 645 650 655  
 Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val Ala Gly Ile Thr Asn  
 660 665 670  
 Leu Lys Lys Cys Ser Thr Ser Pro Leu Leu Glu Ala Cys Glu Phe Leu

675  
 Arg Lys  
 690

680  
 685

<210> 5  
 <211> 42  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> primer

<400> 5  
 atcgaagctt catgagtaat gtgtgagcat tatgggacca cg 42

<210> 6  
 <211> 53  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> primer

<400> 6  
 ctagtctaga ctcgagccat ggggccggct agggagccat cgcacaagag gaa 53

<210> 7  
 <211> 72  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> codon optimized lactoferricin coding sequence based on Homo sapiens sequence

<400> 7  
 accaagtgtc tccagtggca gcgcaacatg cggaagggtgc gcgggcccgcc ggtcagctgc 60  
 atcaagcggg ac 72

<210> 8  
 <211> 162  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> codon optimized EGF coding sequence based on Homo sapiens sequence

<400> 8  
 aactccgact cggagtggcc cctctccac gacggttact gcctccacga cgggggtctgc 60  
 atgtacatcg aggcctcga caagtacgcc tgcaactgcg tcgtgggcta catcggcgag 120  
 cggtgccagt accgcgacct caagtgtgg gagctgcgct ga 162

<210> 9  
 <211> 213  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> codon optimized IGF-1 coding sequence based on Homo sapiens sequence

<400> 9  
 ggcccggaga ccctctgcgg cgccgagctc gtggacgccc tccagttcgt gtgcggcgac 60  
 cgcggtcttct acttcaacaa gccgaccggc tacggcagca gcagccgccc cgccccgcag 120  
 accggcatcg tggacgagtg ctgcttcgcg agctgcgacc tccgccgcct ggagatgtac 180

tgcgccccgc tcaagccccgc caagagcgcc tga

213

<210> 10

<211> 1095

<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized lactohedrin coding sequence based on Homo sapiens sequence

<400> 10

|            |             |            |            |             |             |      |
|------------|-------------|------------|------------|-------------|-------------|------|
| ctggacatct | gctcgaagaa  | cccgtgccac | aacggcgggc | tctgcgagga  | gatcagccag  | 60   |
| gaggtgcggg | gcgacgtgtt  | cccctcgtac | acctgcacct | gcctgaaggg  | ctacgccggg  | 120  |
| aaccactgcg | agacgaagtg  | cgtggagccc | ctggggatgg | agaacggcaa  | catcgccaac  | 180  |
| tcccagatcg | ccgcctcctc  | cgtgcgggtg | accttcctcg | gcctccagca  | ctgggtcccg  | 240  |
| gagctggccc | ggctcaaccg  | ggcgggcatg | gtgaacgcgt | ggacccccctc | gtccaacgac  | 300  |
| gacaaccctg | ggatccaagt  | gaacctgtc  | cgccgcatgt | gggtcaccgg  | cgtggtcacc  | 360  |
| caaggcgcca | gccgcctggc  | cagccacgag | tacctcaagg | ccttcaaggt  | cgctacagc   | 420  |
| ctcaacggcc | acgagttcga  | cttcatccac | gacgtcaaca | agaagcaca   | ggagttcgtg  | 480  |
| ggcaactgga | acaagaacgc  | ggtccacgtg | aacctcttcg | agacccccgt  | cgaggcccag  | 540  |
| tacgtccgcc | tctaccccc   | gagctgccac | accgcctgca | cgctccgctt  | cgagctgctg  | 600  |
| gggtgcgagc | tgaacgggtg  | cgcaaccgc  | ctggggctca | agaacaacag  | catccccgac  | 660  |
| aagcagatca | cggcctcgtc  | gtcgtacaag | acctggggcc | tgcacctctt  | ctcgtggaac  | 720  |
| ccgagctacg | cccggctgga  | caagcagggc | aacttcaacg | cctgggtcgc  | cgggagctac  | 780  |
| gggaacgacc | agtggctcca  | ggtggacctc | ggcagctcca | aggaggtcac  | cggcattcatc | 840  |
| acgcaggggg | cccgcgaactt | cggtccgtg  | cagttcgtgg | cctcctacaa  | ggtggcctac  | 900  |
| togaacgaca | gcgccaactg  | gaccgagtac | caggacccgc | gcaccgggtc  | cagcaagatc  | 960  |
| ttccccggca | actgggacaa  | ccacagccac | aagaagaacc | tgttcgagac  | ccccatcctc  | 1020 |
| gcccggtacg | tccgcatcct  | ccccgtcgct | tggcacaacc | ggatcgcgct  | ccggctggag  | 1080 |
| ctcctcggtc | gctga       |            |            |             |             | 1095 |

<210> 11

<211> 489

<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized kappa-casein coding sequence based on Homo sapiens sequence

<400> 11

|            |            |            |            |            |             |     |
|------------|------------|------------|------------|------------|-------------|-----|
| gaggtccaaa | accagaagca | gcccgcctgc | cacgagaacg | acgagcgccc | cttctaccag  | 60  |
| aagaccgcac | cctacgtccc | gatgtactac | gtcccgaaca | gctaccctta | ctacgggtacg | 120 |
| aacctgtacc | agcgccgccc | ggccatcgct | atcaacaacc | cctacgtccc | ccggacctac  | 180 |
| tacgcgaacc | cggccgtggt | gcgccccac  | gcgagatcc  | cgacgcggca | gtaoctgcca  | 240 |
| aacagccacc | ccccaccgt  | ggtgcgggcg | cccaacctcc | acccgagctt | catcgctatc  | 300 |
| ccccccaaga | agatccagga | caagatcatc | atccccacca | tcaacaccat | cgccaccgtg  | 360 |
| gagccgacgc | cagccccgc  | gaccgagccc | acggtggaca | gcgtcgtgac | cccagaggcg  | 420 |
| ttctccgaat | cgatcatcac | ctccaccccc | gagaccacca | cggtggccgt | caagccggcg  | 480 |
| acggcatga  |            |            |            |            |             | 489 |

<210> 12

<211> 1233

<212> DNA

<213> Artificial Sequence

<220>

<223> codon optimized haptocorrin coding sequence based on Homo sapiens sequence

<400> 12

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| gagatctgcg | aggtctccga | ggagaactac | atccgcctca | agccccctct | gaacaccatg | 60  |
| atccagagca | actacaaccg | gggcacgtcg | gccgtgaacg | tcgtgctctc | cctgaagctc | 120 |
| gtgggcatcc | agatccagac | cctcatgcag | aagatgatcc | agcagatcaa | gtacaacgtg | 180 |

|            |            |            |            |            |             |      |
|------------|------------|------------|------------|------------|-------------|------|
| aagagccgcc | tctcggacgt | gtccagcggc | gagctggcgc | tcatcatcct | cgcgctcggc  | 240  |
| gtgtgcggga | acgcggagga | gaacctcatc | tacgactacc | acctcacgga | caagctggag  | 300  |
| aacaagttcc | aggccgagat | cgagaacatg | gaggcccaca | acggcacccc | gctgaccaac  | 360  |
| tactaccagc | tcagcctgga | cgtcctcgcg | ctctgcctgt | tcaacgggaa | ctactccacc  | 420  |
| gccgaggtgg | tcaaccactt | cacccccgag | aacaagaact | actacttcgg | ctcgcagttc  | 480  |
| tccgtggaca | ccggggccat | ggcgtccctg | gccctcacct | gcgtgaagaa | gtccctcatc  | 540  |
| aacggccaga | tcaaggccga | cgagggtccc | ctgaagaaca | tctcgatcta | caccaagagc  | 600  |
| ctcgtggaga | agatcctcag | cgagaagaag | gagaacgggc | tgatcggcaa | caccttctcg  | 660  |
| accggcgagg | cgatgcaggc | cctgttcctg | agcagcgact | actacaacga | gaacgactgg  | 720  |
| aactgccagc | agaccctcaa | cacggtcctg | accgagatca | gccagggcgc | gttcagcaac  | 780  |
| cccaacgccg | ccgcccaggt | cctgccggcc | ctgatgggca | agaccttcct | cgacatcaac  | 840  |
| aaggacagct | cctgcgtgtc | cgcgagcggc | aacttcaaca | tctccgccga | cgagccgatac | 900  |
| acggtgacgc | cgcccagacg | ccagtcgtac | atctccgtga | actacagcgt | gcggatcaac  | 960  |
| gagacctact | tcacgaacgt | gacggtcctc | aacggctcgg | tcttcctgag | cgtgatggag  | 1020 |
| aaggcgacga | agatgaacga | cacgatcttc | ggcttcacga | tggaggagcg | cagctggggc  | 1080 |
| ccctacatca | cctgcatcca | gggcctctgc | gccacaaca  | acgaccgcac | ctactgggag  | 1140 |
| ctgctgagcg | gcggcgagcc | gctgagccag | ggggccggca | gctacgtggt | ccgcaacggc  | 1200 |
| gagaacctgg | aggtccggtg | gagcaagtac | tga        |            |             | 1233 |

&lt;210&gt; 13

&lt;211&gt; 2061

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; codon optimized lactoperoxidase coding sequence based on Homo sapiens sequence

&lt;400&gt; 13

|            |            |             |            |             |             |      |
|------------|------------|-------------|------------|-------------|-------------|------|
| caaacgaccc | ggacgtcggc | gatctccgac  | acggtctcgc | aggccaaggt  | gcaagtcaac  | 60   |
| aaggcattcc | tggattcgcg | cacgcggctg  | aagaccgcga | tgtcgtccga  | gaccccgacg  | 120  |
| agccggcagc | tgagcgagta | cctcaagcac  | gcgaaggggc | ggacgcgcac  | cgccatccgc  | 180  |
| aatggccaag | tgtgggagga | atccctgaag  | cggtctcggc | agaaggcgtc  | gctcaccaac  | 240  |
| gtgaccgacc | cgtccctcga | cctgaccagc  | ctctccctgg | aggtcggctg  | cggcgccccg  | 300  |
| gcgcccgtcg | tgcgctgcga | cccctgctcg  | ccataccgca | cgatcacggg  | cgactgcaac  | 360  |
| aaccggcgga | agccggcact | gggggctgcg  | aaccgcgccc | tgcgcgctg   | gctccccgcc  | 420  |
| gagtacgagg | acggcctcag | cctccccctc  | ggttggaacc | ccggcaagac  | gcgcaacggc  | 480  |
| ttcccgcctc | cgctcgctcg | cgaggtcagc  | aacaagatcg | tccgttacct  | gaacgaggag  | 540  |
| ggggctcctg | accaaaccg  | ctccctcctc  | ttcatgcagt | gggggcagat  | cgtggaccac  | 600  |
| gacctggact | tgcgcccgga | cacggagctg  | ggctccagcg | agtacagcaa  | gacccagtg   | 660  |
| gacgaatact | gcattccagg | cgacaactgc  | ttcccgatca | tgttcccccc  | gaacgacccg  | 720  |
| aaggcgggca | cccaggggca | gtgcatgccg  | ttcttccggg | caggcttcgt  | ctgcccgaac  | 780  |
| cccccgtaga | agtccctcgc | gcgcgagcag  | atcaacgcgc | tcacgtcctt  | cctcgacgcc  | 840  |
| agcttcgtct | acagcagcga | gccgtccctc  | gccagccgcc | tccgcaacct  | cagcagcccc  | 900  |
| ctcggcctca | tggcggtcaa | ccaggaggtg  | tccgaccaag | gcctcccata  | cctgccgtac  | 960  |
| gacagcaaga | agccgtcccc | ctgcgagttc  | atcaacacca | ccgcgcgcgt  | cccgtgcttc  | 1020 |
| ctcgcggcgc | attcgcgggc | gagcagacac  | atcctcctcg | ccacgagcca  | cacctgttcc  | 1080 |
| ctccgcgagc | acaaccgcct | cgcccgggag  | ctgaagcgcc | tcaaccgcga  | gtgggacggc  | 1140 |
| gagaagctct | accaggaggc | ccggaagatc  | ctcggcgctt | tccgtccagat | catcaccttc  | 1200 |
| cgggactacc | tccccatcct | gctcgggtgac | cacatgcaga | agtggatccc  | cccctaccaa  | 1260 |
| ggctactccg | agagcgtgga | cccgcgcate  | tccaacgtct | tcacgttcgc  | gttccgcttc  | 1320 |
| gggcaacctg | aggtgccgtc | gtcgatgttc  | cgctcgcagc | agaactacca  | gccctggggc  | 1380 |
| ccagagccgg | agctgccgct | ccacaccctg  | ttcttcaaca | cctggcggat  | ggtcaaggac  | 1440 |
| ggcggcacgc | acccgctcgt | gcgcgggctc  | ctggctaaga | agtcgaagct  | catgaagcag  | 1500 |
| aacaagatga | tgaccggcga | gctgcgcaac  | aagctgttcc | agcccaacca  | ccgcattccac | 1560 |
| gggttcgacc | tggctgcaat | caacacccag  | cggtgcgcgc | accaaggcca  | gcccggctac  | 1620 |
| aactcgtggc | gcgcgttctg | cgacctctcc  | cagccacaga | cgctggagga  | gctcaaacac  | 1680 |
| gtgctcaaga | gcgaagatgc | cgccaagaag  | ctgctcgggc | tctacggcac  | gcccgcacac  | 1740 |
| atcgacatct | ggatcggggc | catcgcgagc  | ccgctcgtgg | agcgcggggc  | cgctcgcccg  | 1800 |
| ctgctcgcgt | gcctcctggg | caagcaattc  | caacagatcc | gcgacggggg  | ccggttctgg  | 1860 |
| tgggagaacc | ccggcgtggt | caccaacgag  | cagaaggatt | cgctccaaaa  | gatgagcttc  | 1920 |
| tcccgcttgg | tgtgcgacaa | cacccgcata  | accaagggtc | cgcgcgaccc  | attctggggc  | 1980 |
| aactcctacc | cgtacgactt | cgtggactgc  | tccgccatcg | acaagctcga  | cctgtccccc  | 2040 |
| tgggcatcgg | tgaagaactg | a           |            |             |             | 2061 |

<210> 14  
 <211> 1185  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> codon optimized alpha-1-antitrypsin coding sequence based on Homo sapiens sequence

<400> 14  
 gaggacccgc agggcgacgc cgcccagaag accgacacca gccaccacga ccaggaccac 60  
 cogacgttca acaagatcac cccgaatttg gccgaattcg ccttcagcct gtaccgccag 120  
 ctgcgcgacc agtccaactc caccaacatc ttcttcagcc cggtagagcat cgccaccgcc 180  
 ttgcgcatgc tgtccctggg taccaaggcg gacacccacg acgagatcct cgaagggctg 240  
 aacttcaacc tgacggagat cccggaggcg cagatccacg agggcttcca ggagctgctc 300  
 aggacgctca accagccgga ctcccagctc cagctcacca ccggcaacgg gctcttcctg 360  
 tccgagggcc tcaagctcgt cgataagtct ctggaggacg tgaagaagct ctaccactcc 420  
 gaggcgttca ccgtcaactt cggggacacc gaggaggcca agaagcagat caacgactac 480  
 gtgcgagaagg ggacccaggg caagatcgtg gacctggtca aggaattgga cagggaacac 540  
 gtcttcgcgc tcgtcaacta catcttcttc aaggcgcaagt gggagcgccc gttcgagggtg 600  
 aaggacaccg agggaggagga cttccacgtc gaccagggtca ccaccgtcaa ggtcccgatg 660  
 atgaagaggc tcggcatgtt caacatccag cactgcaaga agctctccag ctgggtgctc 720  
 ctcatgaagt acctggggaa cgccaccgcc atcttcttcc tgccggacga gggcaagctc 780  
 cagcacctgg agaacgagct gacgcacgac atcatcacga agttcctgga gaacgaggac 840  
 aggcgctccg ctacgctcca cctcccgaag ctgagcatca ccggcacgta cgacctgaag 900  
 agcgtgctgg gccagctggg catcacgaag gtcttcagca acggcgcgga cctctcgggc 960  
 gtgacggagg agggccccct gaagctctcc aaggccgtgc acaaggcggg gctcacgatac 1020  
 gacgagaagg ggacggaagc tgccggggcc atgttctctg aggccatccc cgtgtccatc 1080  
 ccgcccagag tcaagttcaa caagcccttc gtcttctctg tgatcgagca gaacacgaag 1140  
 agccccctct tcatggggaa ggtcgtcaac ccacgcgaga agtga 1185

<210> 15  
 <211> 786  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Rice Gt1 promoter and Gt1 leader coding sequence

<400> 15  
 catgagtaat gtgtgagcat tatgggacca cgaaataaaa agaacatttt gatgagtcgt 60  
 gtatcctcga tgagcctcaa aagttctctc accccggata agaaaccctt aagcaatgtg 120  
 caaagtttgc attctccact gacataatgc aaaataagat atcatcgatg acatagcaac 180  
 tcatgcatca tatcatgcct ctctcaacct attcattcct actcatctac ataagtatct 240  
 tcagctaaat gttagaacat aaaccataaa gtcacgtttg atgagtatta ggcgtgacac 300  
 atgacaaatc acagactcaa gcaagataaa gcaaaatgat gtgtacataa aactccagag 360  
 ctatatgtca tattgcaaaa agaggagagc ttataagaca aggcattgact cacaaaaatt 420  
 cacttgccctt tcgtgtcaaa aagaggaggg ctttacatta tccatgtcat attgcaaaag 480  
 aaagagagaa agaacaacac aatgctgcgt caattataca tatctgtatg tccatcatta 540  
 ttcattccacc tttcgtgtac cacacttcat atatcataag agtcacttca cgtctggaca 600  
 ttaacaaact ctatcttaac atttagatgc aagagccttt atctcactat aaatgcacga 660  
 tgattttctca ttgtttctca caaaaagcgg ccgcttcatt agtcctacaa caacatggca 720  
 tccataaata gccccatagt tttcttcaca gtttgcttgt tcctcttggt cgatggctcc 780  
 ctagcc 786

<210> 16  
 <211> 1055  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Rice Glb promoter and Gt1 leader coding sequence

<400> 16  
 ctgcaggggag gagagggggag agatggtgag agaggaggaa gaagaggagg ggtgacaatg 60

|            |             |             |            |            |            |      |
|------------|-------------|-------------|------------|------------|------------|------|
| atatgtgggg | catgtgggca  | cccaatTTTT  | taattcattc | ttttgttgaa | actgacatgt | 120  |
| gggtcccatg | agatttatta  | TTTTtcggat  | cgaatcgcca | cgtaagcgct | acgtcaatgc | 180  |
| tacgtcagat | gaagaccgag  | tcaaatttagc | cacgtaagcg | ccacgtcagc | caaaaccacc | 240  |
| atccaaaccg | ccgaggggacc | tcatctgcac  | tggTTTTgat | agttgaggga | cccgttgtat | 300  |
| ctggtTTTTc | gattgaagga  | cgaaaatcaa  | atttgttgac | aagttaaggg | accttaaagt | 360  |
| aacttattcc | atttcaaaat  | attctgtgag  | ccatatatac | cgtgggcttc | caatcctcct | 420  |
| caaattaaag | ggcctTTTTa  | aaatagataa  | ttgccttctt | tcagtcaccc | ataaaagtac | 480  |
| aaaactacta | ccaacaagca  | acatgcgcag  | ttacacacat | tttctgcaca | tttccgccac | 540  |
| gtcacaaaga | gctaagagtt  | atccctagga  | caatctcatt | agtgtagata | catccattaa | 600  |
| tcttttatca | gaggcaaacy  | taaagccgct  | ctttatgaca | aaaataggtg | acacaaaagt | 660  |
| gttatctgcc | acatacataa  | cttcagaaat  | tacccaacac | caagagaaaa | ataaaaaaaa | 720  |
| atctttttgc | aagctccaaa  | tcttggaaac  | ctttttcact | ctttgcagca | ttgtactctt | 780  |
| gctctttttc | caaccgatcc  | atgtcacccct | caagcttcta | cttgatctac | acgaagctca | 840  |
| ccgtgcacac | aacctatggc  | acaaaaaccc  | tataaaaccc | catccgatcg | ccatcatctc | 900  |
| atcatcagtt | cattaccaac  | aaacaaaaga  | ggaaaaaaa  | catatacact | tctagtgtat | 960  |
| gtctgattga | tcatcaatct  | agaggcgcc   | gcatggctag | caaggtcgtc | ttcttcgcgg | 1020 |
| cggcgctcat | ggcggccatg  | gtggccatct  | ccggc      |            |            | 1055 |

&lt;210&gt; 17

&lt;211&gt; 976

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Bx7 promoter

&lt;400&gt; 17

|             |             |             |             |             |             |     |
|-------------|-------------|-------------|-------------|-------------|-------------|-----|
| ctgcaggcca  | gggaaagaca  | atggacatgc  | aaagaggtag  | gggcaggga   | gaaacacttg  | 60  |
| gagatcatag  | aagaacataa  | gaggttaaacy | ataggagggc  | ataatggaca  | attaaatcta  | 120 |
| cattaattga  | actcatttgg  | gaagtaaaacy | aaatccatat  | tctggtgtaa  | atcaaactat  | 180 |
| ttgacgcgga  | tttactaaga  | tcctatgtta  | atttttagaca | tgactggcca  | aaggttttcag | 240 |
| ttagtttcatt | tgtcacggaa  | aggtgttttc  | ataagtccaa  | aactctacca  | acttttttgc  | 300 |
| acgtcatagc  | atagatagat  | gttgtgagtc  | attggataga  | tattgtgagt  | cagcatggat  | 360 |
| ttgtgttgcc  | tggaaatcca  | actaaatgac  | aagcaacaaa  | acctgaaatg  | ggcttttagga | 420 |
| gagatggttt  | atcaattttac | atgttccatg  | caggctacct  | tccactactc  | gacatggtta  | 480 |
| gaagttttga  | gtgccgcata  | tttgcggaag  | caatggcact  | actcgacatg  | gttagaagtt  | 540 |
| ttgagtgcgc  | catattttgcg | gaagcaatgg  | ctaacagata  | catattctgc  | caaaccctaa  | 600 |
| gaaggataat  | cactcctctt  | agataaaaaag | aacagaccaa  | tgtacaaaca  | tccacacttc  | 660 |
| tgcaaaacaat | acaccagaac  | taggattaag  | cccattacgt  | ggcttttagca | gaccgtccaa  | 720 |
| aaatctgttt  | tgcaagcacc  | aattgtctct  | tacttatcca  | gcttcttttg  | tgttggcaaa  | 780 |
| ctgccccttt  | ccaaccgatt  | ttgtttcttc  | tcacgcttcc  | ttcataggct  | aaactaacct  | 840 |
| cggcgctgcac | acaaccatgt  | cctgaacctt  | cacctcgctc  | ctataaaaagc | ccatccaacc  | 900 |
| ttacaatctc  | atcatcacc   | acaacaccga  | gcaccccaat  | ctacagatca  | attcactgac  | 960 |
| agttcactga  | tctaga      |             |             |             |             | 976 |

&lt;210&gt; 18

&lt;211&gt; 1009

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Glub-2 promoter

&lt;400&gt; 18

|            |            |            |            |            |             |     |
|------------|------------|------------|------------|------------|-------------|-----|
| ctgcagtaat | ggatacctag | tagcaagcta | gcttaaaca  | atctaaatto | caatctgttc  | 60  |
| gtaaacygtt | tctcgatcgc | aattttgatc | aaaactattg | aaaacctcaa | ttaaaccatt  | 120 |
| caaaatTTTT | aatataccca | acaagagcgt | ccaaaccaa  | tatgtaaata | tggtatgtcat | 180 |
| gataattgac | ttatgacaat | gtgattattt | catcaagtct | ttaaatcatt | aattctagt   | 240 |
| gaagggttat | gttttcttat | gctaaagggt | tatgtttata | taagaatatt | aaagagcaaa  | 300 |
| ttgcaataga | tcaacacaac | aaatttgaat | gtttccagat | gtgtaaaaat | atccaaatta  | 360 |
| attgttttaa | aatagtttta | agaaggatct | gatatgcaag | tttgatagtt | agtaaactgc  | 420 |
| aaaagggtct | attacatgga | aaattcctta | ttgaatatgt | ttcattgact | ggtttatttt  | 480 |
| acatgacaac | aaagttacta | gtatgtcaat | aaaaaaatac | aaggttactt | gtcaattgta  | 540 |
| ttgtgccaa  | taaagatgac | aacaaacata | caaattttat | tgttctttta | tagaaacacc  | 600 |
| taacttatca | aggatagttg | gccacgcaaa | aatgacaaca | tactttacaa | ttgtatcatc  | 660 |

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| ataaagatct | tatcaagtat | aagaacttta | tggtgacata | aaaaataatc | acaagggcaa | 720  |
| gacacatact | aaaagtatgg | acagaaatth | cttaacaaac | tccatttggt | ttgtatccaa | 780  |
| aagcataaga | aatgagtcac | ggctgagtca | tgatagttag | ttcaatcttg | caaaattgcc | 840  |
| tttttggtta | gtattgtttt | aacactacaa | gtcacatatt | gtctatactt | gcaacaaaca | 900  |
| ctattaccgt | gtatcccaag | tgcccttttc | attgctatat | aaactagctt | gatcgggtct | 960  |
| tcaactcaca | tcaattagct | taagtttcca | ttagcaactg | ctaattagct |            | 1009 |

&lt;210&gt; 19

&lt;211&gt; 839

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Gt3 promoter

&lt;400&gt; 19

|            |            |             |             |            |            |     |
|------------|------------|-------------|-------------|------------|------------|-----|
| ctgcagtgtg | agtgtagctt | cttatagctt  | agtgcctttac | tatcttcaca | agcacatgct | 60  |
| atagtattgt | tccaagatga | aagaataatt  | catccttgct  | accaacttgc | atgatattat | 120 |
| atttgtgaat | atcctatctc | ttggcttata  | atgaaatgtg  | ctgctgggtt | attctgacca | 180 |
| tggtatttga | gagcctttgt | atagctgaaa  | ccaacgtata  | tcgagcatgg | aacagagaac | 240 |
| aaaatgcaag | gattttttta | ttctgggtca  | tgccctggat  | gggttaatat | cgtgatcatc | 300 |
| aaaaaagata | tgcatataat | ttaaagtaata | aatttgctca  | taagaaacca | aaacccaaag | 360 |
| cacatatgtc | ctaaacaaac | tgcatctttgt | ttgtcatgta  | gcaatacaag | agataatata | 420 |
| tgacgtgggt | atgacttatt | cactttttgt  | gactccaaaa  | tgtagtaggt | ctaactgatt | 480 |
| gtttaaagtg | atgtcttact | gtagaagttt  | catcccaaaa  | gcaatcacta | aagcaacaca | 540 |
| cacgtatagt | ccaccttcac | gtaattcttt  | gtggaagata  | acaagaaggc | tactgaaaaa | 600 |
| ataaaagcaa | agaaaaggat | atcaaacaga  | ccattgtgca  | tcccattgat | ccttgatgtg | 660 |
| ctatttatct | atcctccttt | tgtgtacctt  | acttctatct  | agtgagtcac | ttcatatgtg | 720 |
| gacattaaca | aactctatct | taacatctag  | tcgatcacta  | ctttacttca | ctataaaagg | 780 |
| accaacatat | atcatccatt | tctcacaana  | gcattgagtt  | cagtcccaca | aaatctaga  | 839 |

&lt;210&gt; 20

&lt;211&gt; 1302

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Glub-1 promoter

&lt;400&gt; 20

|            |             |             |            |             |            |      |
|------------|-------------|-------------|------------|-------------|------------|------|
| ctgcagagat | atggattttt  | taagattaat  | tgattctctg | tctaaagaaa  | aaaagtatta | 60   |
| ttgaattaaa | tggaanaaga  | aaaaggaaaa  | aggggatggc | ttctgctttt  | tggtctgaag | 120  |
| gcggcgtgtg | gccagcgtgc  | tgctgtcgga  | cagcgagcga | acacacgacg  | gagcagctac | 180  |
| gacgaacggg | ggaccgagtg  | gaccggacga  | ggatgtggcc | taggacgagt  | gcacaaggct | 240  |
| agtggactcg | gtccccgcgc  | ggtatcccg   | gtggtccact | gtctgcaaac  | acgattcaca | 300  |
| tagagcgggc | agacgcggga  | gccgtccctag | gtgcacggga | agcaaatccg  | tcgcctgggt | 360  |
| ggatttgagt | gacacggccc  | acgtgtagcc  | tcacagctct | ccgtgggtcag | atgtgtaaaa | 420  |
| ttatcataat | atgtgttttt  | caaatagtta  | aataatatat | ataggcaagt  | tatatgggtc | 480  |
| aataagcagt | aaaaaggctt  | atgacatgg   | aaaattactt | acaccaatat  | gccttactgt | 540  |
| ctgatataat | ttacatgaca  | acaaagtta   | aagtacgtca | tttaaaaata  | caagttactt | 600  |
| atcaattgta | gtgtatcaag  | taaatgacaa  | caaacttaca | aatttgctat  | tttgaaggaa | 660  |
| cacttaaaaa | aatcaatagg  | caagttatat  | agtcaataaa | ctgcaagaag  | gcttatgaca | 720  |
| tggaanaaat | acatacacca  | atatgcttta  | ttgtccggta | tatttttaca  | gacaacaaag | 780  |
| ttataagtat | gtcattttta  | aatacaagtt  | acttatcaat | tgtcaagtaa  | atgaaaacaa | 840  |
| acctacaaat | ttgttatatt  | gaaggaacac  | ctaaattatc | aaatatagct  | tgctacgcaa | 900  |
| aatgacaaca | tgcttacaag  | ttattatcat  | cttaaagtta | gactcatott  | ctcaagcata | 960  |
| agagctttat | ggtgcaaaaa  | caaataaat   | gacaaggcaa | agatacatat  | atattaagag | 1020 |
| tatggacaga | catcttcttt  | acaaactcca  | tttgtattac | tccaaaagca  | ccagaagttt | 1080 |
| gtcatggctg | agtcataaaa  | tgtatagttc  | aatcttgcaa | agttgctttt  | ccttttgtag | 1140 |
| tgtgttttaa | cactacaagc  | catatattgt  | ctgtacgtgc | aacaaactat  | atcaccatgt | 1200 |
| atcccaagat | gcttttttat  | tgctatataa  | actagcttgg | tctgtctttg  | aactcacatc | 1260 |
| aattagctta | agttttccata | agcaagtaca  | aatagctcta | ga          |            | 1302 |

&lt;210&gt; 21

&lt;211&gt; 675

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Rice proalmin promoter

&lt;400&gt; 21

|             |            |            |            |            |             |     |
|-------------|------------|------------|------------|------------|-------------|-----|
| ctgcagcatc  | ggcttaggtg | tagcaacacg | actttattat | tattattatt | attattatta  | 60  |
| ttatttttaca | aaaatataaa | atagatcagt | ccctcaccac | aagtagagca | agttggtgag  | 120 |
| ttattgtaaa  | gttctacaaa | gctaatttaa | aagttattgc | attaacttat | ttcatattac  | 180 |
| aaacaagagt  | gtcaatggaa | caatgaaaac | catatgacat | actataattt | tgttttttatt | 240 |
| attgaaatta  | tataattcaa | agagaataaa | tccacatagc | cgtaaagttc | tacatgtggt  | 300 |
| gcattaccaa  | aatatatata | gcttacaaaa | catgacaagc | ttagtttgaa | aaattgcaat  | 360 |
| ccttatcaca  | ttgacacata | aagtgagtga | tgagtcataa | tattattttt | cttgctaccc  | 420 |
| atcatgtata  | tatgatagcc | acaaagttac | tttgatgatg | atatcaaaga | acatttttag  | 480 |
| gtgcacctaa  | cagaatatcc | aaataatatg | actcacttag | atcataatag | agcatcaagt  | 540 |
| aaaactaaca  | ctctaaagca | accgatggga | aagcatctat | aaatagacaa | gcacaatgaa  | 600 |
| aatcctcatc  | atccttcacc | acaattcaaa | tattatagtt | gaagcatagt | agtagaatcc  | 660 |
| aacaacaatc  | tagag      |            |            |            |             | 675 |

&lt;210&gt; 22

&lt;211&gt; 1098

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Rice cysteine peptidase promoter

&lt;400&gt; 22

|            |            |             |            |            |            |      |
|------------|------------|-------------|------------|------------|------------|------|
| ccaggcttca | tcctaaccat | tacaggcaag  | atgttgatg  | aagaagggcg | aacatgcaga | 60   |
| ttgttaaact | gacacgtgat | ggacaagaat  | gaccgattgg | tgaccggtct | gacaatggtc | 120  |
| atgtcgtcag | cagacagcca | tctcccacgt  | cgcgcctgct | tccggtgaaa | gtggaggtag | 180  |
| gtatggggcg | tcccgtcaga | aggtgattcg  | gatggcagcg | atacaaatct | ccgtccatta | 240  |
| atgaagagaa | gtcaagttga | aagaaaggga  | gggagagatg | gtgcatgtgg | gatccccctg | 300  |
| ggatataaaa | ggaggacctt | gccacttag   | aaaggagagg | agaaagcaat | cccagaagaa | 360  |
| tcgggggctg | actggcactt | tgtagcttct  | tcatacgcg  | atccaccaa  | acacaggagt | 420  |
| agggtattac | gcttctcagc | ggcccgaacc  | tgtatacatc | gcccgtgtct | tgtgtgtttc | 480  |
| cgtctttgcg | aaccttccac | agattgggag  | cttagaacct | caccagggc  | ccccggccga | 540  |
| actggcaaag | gggggacctg | gcggtctccc  | ggtgaggagc | cccacgctcc | gtcagttcta | 600  |
| aattaccgga | tgagaaagg  | aggggggggg  | gggaaatctg | ccttgtttat | ttacgatcca | 660  |
| acggatttgg | tcgacaccga | tgagggtgtct | taccagttac | cacgagctag | attatagtac | 720  |
| taattacttg | aggattcggg | tcctaatttt  | ttacccgatc | gacttcgcca | tggaaaattt | 780  |
| tttattcggg | ggagaatatc | caccctgttt  | cgtcccta   | taagatagga | attgttacga | 840  |
| ttagcaacct | aattcagatc | agaattgtta  | gttagcggcg | ttggatccct | cacctcatcc | 900  |
| catcccaatt | cccaaaccga | aactcctctt  | ccagtcgccg | acccaaacac | gcacccgccc | 960  |
| cctataaatc | ccaccgcat  | cgagcctatc  | aagcccaaaa | aaccacaaac | caaacgaaga | 1020 |
| aggaaaaaaa | aaggaggaaa | agaaaagagg  | aggaaagcga | agaggttgga | gagagacgct | 1080 |
| cgtctccacg | tcgccgcc   |             |            |            |            | 1098 |

&lt;210&gt; 23

&lt;211&gt; 432

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Barley D-Hordian promoter

&lt;400&gt; 23

|            |            |             |            |             |            |     |
|------------|------------|-------------|------------|-------------|------------|-----|
| cttcgagtgc | ccgccgattt | gccagcaatg  | gctaacagac | acatattctg  | ccaaaacccc | 60  |
| agaacaataa | tcactttctg | tagatgaaga  | gaacagacca | agatacaaac  | gtccacgctt | 120 |
| cagcaaacag | taccccagaa | ctaggattaa  | gccgattacg | cggcttttagc | agaccgtcca | 180 |
| aaaaaactgt | tttgcaaagc | tccaattcct  | ccttgcttat | ccaatttctt  | ttgtgttggc | 240 |
| aaactgcact | tgtccaaccg | atthttgttct | tcccgtgttt | cttcttaggc  | taactaacac | 300 |
| agccgtgcac | atagccatgg | tccggaatct  | tcacctcgtc | cctataaaaag | cccagccaat | 360 |
| ctccacaatc | tcatcatcac | cgagaacacc  | gagaaccaca | aaactagaga  | tcaattcatt | 420 |

gacagtccac cg

432

&lt;210&gt; 24

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; bx7 signal peptide sequence

&lt;400&gt; 24

atggctaagc gcctggtcct ctttgcgga gtagtcgtcg ccctcgtggc tctcaccgcc 60

&lt;210&gt; 25

&lt;211&gt; 72

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Glub-2 signal peptide sequence

&lt;400&gt; 25

atggcaacta ccattttctc tcgtttttct atatactttt gtgctatgct attatgccag 60  
ggttctatgg cc 72

&lt;210&gt; 26

&lt;211&gt; 85

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Gt3 signal peptide sequence

&lt;400&gt; 26

atgtggacat taacaaactc tatcttaaca tctagtcgat cactacttta cttcactata 60  
aaaggaccaa catatatcat ccatt 85

&lt;210&gt; 27

&lt;211&gt; 72

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Glub-2 signal peptide sequence

&lt;400&gt; 27

atggcgagtt ccgttttctc tcggttttct atatactttt gtgttcttct attatgccat 60  
ggttctatgg cc 72

&lt;210&gt; 28

&lt;211&gt; 69

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; proalmin signal peptide sequence

&lt;400&gt; 28

atgaagatca ttttcgtatt tgctctcctt gctattgttg catgcaacgc ttctgcacgg 60  
tttgatgct 69

&lt;210&gt; 29

&lt;211&gt; 63

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Rice cysteine peptidase signal peptide sequence

&lt;400&gt; 29

|                                                        |    |
|--------------------------------------------------------|----|
| atggccgccc gcgccgcccg cgccgcgttc ctgctgctgc tcatcgctcg | 60 |
| tggtcaccgc gcc                                         | 63 |

&lt;210&gt; 30

&lt;211&gt; 63

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; D-Hordian signal peptide sequence

&lt;400&gt; 30

|                                                          |    |
|----------------------------------------------------------|----|
| atggctaagc ggctggctcct ctttgtggcg gtaatcgctg ccctcgctggc | 60 |
| tctcaccacc gcc                                           | 63 |

&lt;210&gt; 31

&lt;211&gt; 1314

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; O2 transcription factor sequence

&lt;400&gt; 31

|                                                                    |      |
|--------------------------------------------------------------------|------|
| atggagcacg tcatctcaat ggaggagatc ctcgggcccct tctgggagct gctaccaccg | 60   |
| ccagcgccag agccagagcg agagcagcct ccggtaaccg gcatcgctcg cggcagtgtc  | 120  |
| atagacgttg ctgctgctgg tcatgggtgac ggggacatga tggatcagca gcacgccaca | 180  |
| gagtggacct ttgagagggt actagaagag gaggctctga cgacaagcac accgcccgcg  | 240  |
| gtgggtgggtg tgccgaactc ttgttgctca ggcgccctaa atgctgaccg gccgcccgtg | 300  |
| atggaagagg cggtaactat ggcgccctgcg gcggtgagta gtgccgtagt aggtgacccc | 360  |
| atggagtaca atgccatact gaggaggaag ctggaggagg acctcgaggc cttcaaaatg  | 420  |
| tgaggggcgg cctccagtgt tgtgacctca gatcaacgtt ctcaaggctc aaacaatcac  | 480  |
| actggaggta gcagcatcag gaataatcca gtgcagaaca agctgatgaa cggcgaagat  | 540  |
| ccaatcaaca ataaccacgc tcaaactgca ggccttggcg tgaggcttgc tactagctct  | 600  |
| tctctgagag atccttcacc atcagacgaa gacatggacg gagaagtaga gattctgggg  | 660  |
| ttcaagatgc ctaccgagga aagagtgagg aaaagaaagg aatccaatag agaatcagcc  | 720  |
| agacgctcga gatacaggaa agccgctcac ctgaaagaac tggaagacca ggtagcacag  | 780  |
| ctaaaagccg agaattcttg cctgctgagg cgcattgccg ctctgaacca gaagtacaac  | 840  |
| gacgctaacg tcgacaacag ggtgctgaga gcggacatgg agaccctaag agctaagggtg | 900  |
| aagatgggag aggactctct gaagcgggtg atagagatga gctcatcagt gccgtcgctc  | 960  |
| atgcccatct cggcgccgac ccccagctcc gacgctccag tgccgcccgc gcctatccga  | 1020 |
| gacagcatcg tcggctactt ctccgccaca gccgcagacg acgatgcttc ggtcggcaac  | 1080 |
| ggtttcttgc gactgaacg tcatcaagag cctgcattca tggctcgctcg tggaaactctg | 1140 |
| agcgccacag agatgaaccg agtagcagca gccacgcatt gcgcgggggc catggagcac  | 1200 |
| atccagacgg cgatgggatc catgcccgcg acctccgcct ccggatctac accgcccgcg  | 1260 |
| caggattatg agctgctggg tccaaatggg gccatacaca tggacatgta ttag        | 1314 |

&lt;210&gt; 32

&lt;211&gt; 987

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; PBF transcription factor sequence

&lt;400&gt; 32

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| atggacatga tctccggcag cactgcagca acatcaacac cccacaacaa ccaacaggcg  | 60  |
| gtgatgttgt catcccccat tataaaggag gaagctaggg acccaaagca gacacgagcc  | 120 |
| atgccccaaa taggtggcag tggggagcgt aagccgaggc cgcaactacc tgaggcgctc  | 180 |
| aagtgccccac gctgcgactc caacaacacc aagttttgct actacaacaa ttatagcatg | 240 |

|            |             |             |            |            |            |     |
|------------|-------------|-------------|------------|------------|------------|-----|
| tcacaaccac | gctactttttg | caaggctttgc | cgccgctatt | ggacacatgg | tggtaccctc | 300 |
| cgcaatgtcc | ccattggtgg  | tggtgtgcgc  | aagaacaaac | atgcctctag | atttgtcttg | 360 |
| ggctctcaca | cctcatcgtc  | ctcatctgct  | acctatgcac | cattatcccc | tagcaccaac | 420 |
| gctagctcta | gcaatatgag  | catcaacaaa  | catatgatga | tggtgcctaa | catgacgatg | 480 |
| cctaccccaa | cgacaatggg  | cttattccct  | aatgtgctcc | caacacttat | gccgacaggt | 540 |
| ggaggcgggg | gctttgactt  | cactatggac  | aaccaacata | gatcattgtc | cttcacacca | 600 |
| atgtctctac | ctagccaggg  | gccagtgcct  | atgctggctg | caggaggag  | tgaggcaaca | 660 |
| ccgtctttcc | tagagatgct  | gagaggaggg  | atttttcatg | gtagtagtag | ctataacaca | 720 |
| agtctcacga | tgagtgggtg  | caacaatgga  | atggacaagc | cattttcgct | gccatcatat | 780 |
| ggtgcaatgt | gcacaaatgg  | gttgagtggc  | tcaaccacta | atgatgccag | acaactggtg | 840 |
| gggcctcagc | aggataacaa  | ggccatcatg  | aagagcagta | ataacaacaa | tggtgtatca | 900 |
| ttgttgaacc | tctactggaa  | caagcacaa   | aacaacaaca | acaacaacaa | caacaacaac | 960 |
| aacaacaaca | acaacaagg   | acaataa     |            |            |            | 987 |

&lt;210&gt; 33

&lt;211&gt; 3902

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Reb transcription factor sequence

&lt;400&gt; 33

|             |             |             |             |             |            |      |
|-------------|-------------|-------------|-------------|-------------|------------|------|
| atggagcggg  | tggtctccgt  | ggaggagatc  | tccgacccat  | tctgggtccc  | gcctccgccc | 60   |
| ccgcagtcgg  | cggcgccggc  | ccagcagcag  | ggcgccggcg  | gcgtggcttc  | gggagggtgt | 120  |
| ggtggtgtag  | cggggggcgg  | cggcgccggg  | aacgcgatga  | accggtgccc  | gtcggagtgg | 180  |
| tacttccaga  | agtttctgga  | ggaggcgggtg | ctcgatagcc  | ccgtcccgaa  | ccctagcccc | 240  |
| agggccgaag  | cgggagggat  | caggggcgca  | ggaggggtgg  | tgccggtcga  | tgtaagcag  | 300  |
| ccgcagctct  | cggcgccggc  | gacgacgagc  | gcgggtgggtg | accccggtga  | gtacaacgcg | 360  |
| atgctgaagc  | agaagctgga  | gaaggacctc  | gccgcggctc  | ccatgtggag  | ggtacagcca | 420  |
| ttctcccccc  | ctctagtact  | cgagagctta  | ctgagatcgg  | caatgctagc  | tactgtttgc | 480  |
| atcgaatggt  | tataggtatt  | tagatcgggc  | atttctatag  | accaatggcg  | tccatggtct | 540  |
| tgcaatgcgc  | tctgttgagt  | gtcgggtggt  | ggttcgactc  | atagtatgta  | gggttggtgc | 600  |
| tatgtacaaa  | cgggaagcttc | atagacctcg  | gtattgagat  | tgcgatatcg  | atgcaacctg | 660  |
| cgaattggcg  | atgtaatcag  | tcatattctt  | actaaactgc  | gagacagtgg  | ttgttttgca | 720  |
| attgcaatat  | ttttgtatgg  | ggctgcttaa  | actgtcattg  | ccttttttaga | ttggcaatat | 780  |
| gtgactttat  | gcaagtattt  | gattgggcgg  | atccaggaac  | aaaaagttgg  | ggggattcaa | 840  |
| cataccgagt  | acactggcat  | aaacacatca  | tctcagtatt  | aaactatgct  | aaaatgctat | 900  |
| taagagacct  | ttagcacctc  | ttatcttatc  | aaccatgggtg | aaaaaattga  | aggggggact | 960  |
| caggggggta  | tccatgggtc  | cgatgggtgc  | aggggggact  | gagtcccccc  | tgacccaccg | 1020 |
| ttgaatccgc  | ctggcatgc   | gtataagctg  | tcacagccat  | ttctaggtgc  | ttgtgcttag | 1080 |
| ttgggtgatg  | tcagcttaat  | ttgtcttttc  | tatgtcgtca  | tcgattttct  | aagaaacgaa | 1140 |
| aaatagccta  | tttatgtgct  | ccagaatttg  | atgatccctg  | gcccttcatt  | tgctgaaatt | 1200 |
| agcctatttg  | ttggttgccc  | ttcagttttt  | tcccagctta  | tggtgttgca  | atgtgtggct | 1260 |
| atgcctcggt  | ttgtgcccta  | taatttatta  | tttgcaattc  | atttttgtac  | atgacttaaa | 1320 |
| atgacactag  | agcaacatgc  | actgattggg  | tatcctataa  | tcattttatg  | agttctgttc | 1380 |
| attttatcat  | gctagctcat  | gtcattttca  | tcttcaggcc  | tctggcacag  | ttccacctga | 1440 |
| gcgtcctgga  | gctggttcat  | ccttgctgaa  | tgacagatgt  | tcacacatag  | gcgtccttaa | 1500 |
| ttccatcgga  | ggtacttatc  | ttatctggtt  | acattttcag  | attgttatga  | aactacccaa | 1560 |
| atatcctgca  | caattgcatg  | ggattaaatt  | ttagtttctt  | tgaaatagaa  | gtagagttgt | 1620 |
| attgctgtca  | cgtcatcaaa  | tagttctgaa  | gctatgaata  | aataagttcc  | gcatttggtt | 1680 |
| gtgattcttt  | gaacattaga  | attgttatgc  | ttaagtagat  | agggttatgt  | ttgtttggag | 1740 |
| ttcccttaaa  | tcattttcatt | gctgactgcc  | agctggcagg  | agcattttgt  | gttgccctga | 1800 |
| ccatgaatga  | agaccttcct  | gttctgagtg  | ctcacaagaa  | aacatatttt  | gattaatgca | 1860 |
| ccttgaatcc  | ttaggatctt  | gcaaagatgg  | gcacttagct  | ttagaattga  | gtagtactta | 1920 |
| aatagctggt  | gttatcatga  | tttgtcctgt  | agtgaatgt   | cgacaaaaca  | ggaatgctac | 1980 |
| ttttgacttc  | tgatatttca  | tgcttggtct  | tacttatgct  | ctgtttggaa  | catgggcaca | 2040 |
| tatcaggcaa  | tgctactcca  | gttcaaaaaca | tgctaagtgg  | cccaagtggg  | ggatcgggct | 2100 |
| cacagttggt  | acagaatggt  | gatgtccttg  | taaagcagcc  | caccagctct  | tcatcaaggg | 2160 |
| agcagtcaga  | tgatgatgac  | atgaaggag   | aagctgagac  | cactggaact  | gcaagacctg | 2220 |
| ctgatcaaag  | attacaacga  | aggtgatcat  | tcattgcttc  | cttgtaatat  | agattctgta | 2280 |
| cataattaac  | ctacctcgtc  | atgcatgcat  | gtgtcctatt  | ttcaccttag  | ccctttcagt | 2340 |
| tggaatttcca | ctttcatccg  | gtagcctttc  | agtttcctat  | tgcatcgcat  | atatgatctt | 2400 |
| ttacctacca  | tattagttct  | ctgtgtgcca  | tactcagtgc  | ttagtgtctc  | gagcaagaga | 2460 |
| ggaatttgta  | tggctattac  | acgtagcact  | ttgctctcta  | cttgtttatt  | gacataagca | 2520 |

|            |            |            |             |            |             |      |
|------------|------------|------------|-------------|------------|-------------|------|
| atttgggatg | aattaaatct | gagttcacat | catattcctt  | atgtcacaag | tttctgaaac  | 2580 |
| cgattgtatc | tagtatctgg | ttgatgcacc | cccatcttgg  | atttgcaa   | caaagttata  | 2640 |
| ctccctagag | agctttacct | ttcataaagc | aattacccca  | ataaaccacg | gattttgatag | 2700 |
| ctattgacta | tgattaccag | aattcatttg | gcagctat    | totcaattta | agtttgggtat | 2760 |
| tagtctcagt | tggctgtaaa | ataatgtcac | ggtagggtac  | atgtatgtgc | agcatacaag  | 2820 |
| gtatgggtga | gttatgat   | ggacagtgtg | tacacccac   | atttgctcac | taaaatcaaa  | 2880 |
| atattcaaac | gtcacgtgat | gatatgggtg | attgcattat  | accttgtatt | gtttattatg  | 2940 |
| ttacttgtgc | tagacaataa | tataggctgt | tcttttgggt  | gattttgtat | gaagatgttg  | 3000 |
| agcaagcact | tctcgatata | atgctagttt | tgttgacctg  | ttccaggaag | caatccaatc  | 3060 |
| gggagtcagc | caggcgctca | agaagcagaa | aggcagctca  | cttgaatgag | ctggaggcac  | 3120 |
| aggtgtgata | gttcacatag | ttattttcga | taagacataa  | aatcctaaat | tactggctac  | 3180 |
| tgacttcagt | tatggattta | cttgttacag | gtatcgcaat  | taagagtcga | gaactcctcg  | 3240 |
| ctgttaaggc | gtcttgctga | tgtaaccag  | aagtacaatg  | atgctgctgt | tgacaataga  | 3300 |
| gtgctaaaag | cagatgttga | gaccttgaga | gcaaagggtat | gctatatatg | ccttttgcaa  | 3360 |
| tatgcatccc | atggattgct | actttggctt | gtttcaaact  | ttcaacgtga | cttgtgtacc  | 3420 |
| ctgttattag | aagaataatc | ccgcctacca | ttatactcta  | taaatcacca | tttggccagt  | 3480 |
| ccaaacatga | ttattaaatc | aggtcaatct | gaacattgaa  | atgtatcaaa | aattcgcagg  | 3540 |
| tgaagatggc | agaggactcg | gtgaagcggg | tgacaggcat  | gaacgcgttg | tttccgcgcg  | 3600 |
| cttctgat   | gtcatccctc | agcatgccat | tcaacagctc  | cccatctgaa | gcaacgtcag  | 3660 |
| acgtgctgt  | tcccatccaa | gatgacccga | acaattactt  | cgctactaac | aacgacatcg  | 3720 |
| gaggtaacaa | caactacatg | cccgacatac | cttcttcggc  | tcaggaggac | gaggacttcg  | 3780 |
| tcaatggcgc | tctggctgcc | ggcaagattg | gccggccagc  | ctcgctgcag | cgggtggcga  | 3840 |
| gcctggagca | tctccagaag | aggatgtgcg | gtgggccggc  | ttcgtctggg | tcgacgtcct  | 3900 |
| ga         |            |            |             |            |             | 3902 |