TITLE OF THE INVENTION

METHOD FOR PRODUCING PROTEINS IN PICHIA PASTORIS THAT LACK DETECTABLE CROSS BINDING ACTIVITY TO ANTIBODIES AGAINST HOST CELL ANTIGENS

BACKGROUND OF THE INVENTION

(1) Field of the Invention

The present invention relates to methods for producing protein and glycoproteins in *Pichia pastoris* that lack detectable cross binding activity to antibodies made against host cell antigens. In particular, the present invention relates to using recombinant *Pichia pastoris* strains that do not display a β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan and do not display at least one activity selected from the group consisting of β-mannosyltransferase 1, 3, and 4 activity with respect to an *N*-glycan or *O*-glycan. These recombinant *Pichia pastoris* strains can produce proteins and glycoproteins that lack detectable α-mannosidase resistant β-mannose residues thereon. The present invention further relates to methods for producing bi-sialylated human erythropoietin in *Pichia pastoris* that lack detectable cross binding activity to antibodies against host cell antigens.

(2) Description of Related Art

The ability to produce recombinant human proteins has led to major advances in human health care and remains an active area of drug discovery. Many therapeutic proteins require the posttranslational addition of glycans to specific asparagine residues (*N*-glycosylation) of the protein to ensure proper structure-function activity and subsequent stability in human serum. For therapeutic use in humans, glycoproteins require human-like *N*-glycosylation. Mammalian cell lines (e.g., CHO cells, human retinal cells) that can mimic human-like glycoprotein processing have several drawbacks including low protein titers, long fermentation times, heterogeneous products, and continued viral containment. It is therefore desirable to use an expression system that not only produces high protein titers with short fermentation times, but can also produce human-like glycoproteins.

Fungal hosts such as the methylotrophic yeast *Pichia pastoris* have distinct advantages for therapeutic protein expression, for example, they do not secrete high amounts of endogenous proteins, strong inducible promoters for producing heterologous proteins are available, they can be grown in defined chemical media and without the use of animal sera, and they can produce high titers of recombinant proteins (Cregg *et al.*, FEMS Microbiol. Rev. 24: 45-66 (2000)). However, glycosylated proteins expressed in *P. pastoris* generally contain additional mannose sugars resulting in "high mannose" glycans, as well as mannosylphosphate groups which impart a negative charge onto glycoproteins. Glycoproteins with either high mannose

glycans or charged mannans present the risk of eliciting an unwanted immune response in humans (Takeuchi, Trends in Glycosci. Glycotechnol. 9:S29-S35 (1997); Rosenfeld and Ballou, J. Biol. Chem. 249: 2319-2321 (1974)). Accordingly, it is desirable to produce therapeutic glycoproteins in fungal host cells wherein the pattern of glycosylation on the glycoprotein is identical to or similar to that which occurs on glycoproteins produced in humans and which do not have detectable β-mannosylation.

As evidenced by the presence of protective antibodies in uninfected individuals, β-linked mannans are likely to be immunogenic or adversely affect the individual administered a therapeutic protein or glycoprotein comprising β-linked mannans. Additionally, exposed mannose groups on therapeutic proteins are rapidly cleared by mannose receptors on macrophage cells, resulting in low drug efficacy. Thus, the presence of β-linked mannose residues on *N*- or *O*-linked glycans of heterologous therapeutic proteins expressed in a fungal host, for example, *P. pastoris*, is not desirable given their immunogenic potential and their ability to bind to clearance factors.

Glycoproteins made in *P. pastoris* have been reported to contain β-linked mannose residues. In 2003, Trimble *et al.* (Glycobiol. 14: 265-274, Epub Dec 23) reported the presence of β-1,2-linked mannose residues in the recombinant human bile salt-stimulated lipase (hBSSL) expressed in *P. pastoris*. The genes encoding several β-mannosyltransferases have been identified in Pichia pastoris and Candida albicans (*See* U.S. Patent No. 7,465,577 and Mille *et al.*, J. Biol. Chem. 283: 9724-9736 (2008)).

In light of the above, there is a need to provide methods for making recombinant therapeutic proteins or glycoproteins in methylotrophic yeast such as *Pichia pastoris* that lack eptitopes that might elicit an adverse reaction in an individual administered the recombinant therapeutic protein or glycoprotein. A method for determining whether a recombinant therapeutic protein or glycoprotein provides a risk of eliciting an adverse reaction when administered to an individual is to contact the recombinant therapeutic protein or glycoprotein to an antibody prepared against total host cell antigens. This is of particular concern for proteins or glycoproteins intended for chronic administration. The lack of cross binding to the antibody indicates that the recombinant therapeutic protein or glycoprotein lacks detectable cross binding activity to the antibody and is unlikely to elicit an adverse reaction when administered to an individual. Thus, there is a need for methods for producing a recombinant therapeutic protein or glycoprotein that lacks detectable cross binding activity to the antibody and is unlikely to elicit an adverse reaction when administered to an individual.

BRIEF SUMMARY OF THE INVENTION

The present invention provides methods for producing protein and glycoproteins in methylotrophic yeast such as *Pichia pastoris* that lack detectable cross binding activity to

antibodies made against host cell antigens. In particular, the present invention provides methods using recombinant methylotrophic yeast such as *Pichia pastoris* strains, which do not display β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan and do not display at least one activity with respect to an *N*-glycan or *O*-glycan selected from β-mannosyltransferase 1, β-mannosyltransferase 3, and β-mannosyltransferase 4 to produce recombinant proteins and glycoproteins. In one aspect, the host cell is a *Pichia pastoris* strain in which the *BMT2* gene encoding β-mannosyltransferase 2 and at least one gene encoding a β-mannosyltransferase selected from β-mannosyltransferase 1, 3, and 4 (genes *BMT1*, *BMT3*, *BMT4*, respectively) have been deleted or disrupted or mutated to produce an inactive β-mannosyltransferase to produce recombinant proteins and glycoproteins. In other aspects, the activity of one or more of the β-mannosyltransferase 1, β-mannosyltransferase 3, and β-mannosyltransferase 4 is abrogated using β-mannosyltransferase inhibitors which includes but is not limited to chemical compounds, antisense DNA to one or more mRNA encoding a β-mannosyltransferase, siRNA to one or more mRNA encoding a β-mannosyltransferase.

These recombinant *Pichia pastoris* strains can produce proteins and glycoproteins that lack detectable α-mannosidase resistant β-mannose residues thereon. The present invention further provides methods for producing bi-sialylated human erythropoietin in *Pichia pastoris* that lack detectable cross binding activity to antibodies against host cell antigens. The methods and host cells enable recombinant therapeutic proteins and glycoproteins to be produced that have a reduced risk of eliciting an adverse reaction in an individual administered the recombinant therapeutic proteins and glycoproteins compared to the same being produced in strains not modified as disclosed herein. The methods and host cells are also useful for producing recombinant proteins or glycoproteins that have a lower potential for binding clearance factors.

In one aspect, the present invention provides a recombinant methylotrophic yeast such as *Pichia pastoris* host cell that does not display β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan and does not display at least one activity with respect to an *N*-glycan or *O*-glycan selected from β-mannosyltransferase 1 activity and β-mannosyltransferase 3 activity and which includes a nucleic acid molecule encoding the recombinant glycoprotein. In further embodiments, the host cell does not display β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan. In further embodiments, the host cell further does not display β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan.

In another aspect, the present invention provides a recombinant methylotrophic yeast such as *Pichia pastoris* host cell that has a deletion or disruption of the gene encoding β-mannosyltransferase 2 activity and a deletion or disruption of at least one gene selected a gene encoding a β-mannosyltransferase 1 activity and a β-mannosyltransferase 3 activity and which includes a nucleic acid molecule encoding the recombinant glycoprotein. In further

embodiments, the host cell has a deletion or disruption of the genes encoding a β-mannosyltransferase 2 activity, a β-mannosyltransferase 1 activity, and a β-mannosyltransferase 3 activity. In further embodiments, the host cell has a deletion or disruption of the gene encoding a β-mannosyltransferase 4 activity.

In another aspect, the present invention provides a recombinant *Pichia pastoris* host cell in which the β-mannosyltransferase 2 (*BMT2*) gene and at least one gene selected from β-mannosyltransferase 1 (*BMT1*) and β-mannosyltransferase 3 (*BMT3*) has been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein. In further embodiments, the β-mannosyltransferase 2 (*BMT2*), β-mannosyltransferase 1 (*BMT1*), and β-mannosyltransferase 3 (*BMT3*) genes are deleted. In further embodiments, the host cell further includes a deletion or disruption of the β-mannosyltransferase 4 (*BMT4*) gene.

In another aspect, the present invention provides a method for producing a recombinant glycoprotein in methylotrophic yeast such as *Pichia pastoris* that lacks detectable cross binding activity with antibodies made against host cell antigens, comprising providing a recombinant host cell that does not display a β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan and does not display at least one activity with respect to an *N*-glycan or *O*-glycan selected from β-mannosyltransferase 1 and β-mannosyltransferase 3 and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein; growing the host cell in a medium under conditions effective for expressing the recombinant glycoprotein; and recovering the recombinant glycoprotein from the medium to produce the recombinant glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens. In further embodiments, the host cell does not display β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan. In further embodiments, the host cell further does not display β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan.

In another aspect, the present invention provides a method for producing a recombinant glycoprotein in methylotrophic yeast such as *Pichia pastoris* that lacks detectable cross binding activity with antibodies made against host cell antigens, comprising providing a recombinant host cell that has a deletion or disruption of the gene encoding a β-mannosyltransferase 2 activity and a deletion or disruption of at least one gene encoding an activity selected from β-mannosyltransferase 1 activity and β-mannosyltransferase 3 activity and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein; growing the host cell in a medium under conditions effective for expressing the recombinant glycoprotein; and recovering the recombinant glycoprotein from the medium to produce the recombinant glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens. In further embodiments, the host cell has a deletion or disruption of

the genes encoding a β-mannosyltransferase 2 activity, a β-mannosyltransferase 1 activity, and a β-mannosyltransferase 3 activity. In further embodiments, the host cell has a deletion or disruption of the gene encoding a β-mannosyltransferase 4 activity.

In another aspect, the present invention provides a method for producing a recombinant glycoprotein in *Pichia pastoris* that lacks detectable cross binding activity with antibodies made against host cell antigens, comprising providing a recombinant *Pichia pastoris* host cell in which the β-mannosyltransferase 2 (*BMT2*) gene and at least one gene selected from β-mannosyltransferase 1 (*BMT1*) and β-mannosyltransferase 3 (*BMT3*) has been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein; growing the host cell in a medium under conditions effective for expressing the recombinant glycoprotein; and recovering the recombinant glycoprotein from the medium to produce the recombinant glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens. In further embodiments, the β-mannosyltransferase 2 (*BMT2*), β-mannosyltransferase 1 (*BMT1*), and β-mannosyltransferase 3 (*BMT3*) genes have been deleted or disrupted. In further embodiments, the host cell further includes a deletion or disruption of the β-mannosyltransferase (*BMT4*) gene.

In general, the detectable cross binding activity with antibodies made against host cell antigens is determined in an assay such as sandwich ELISA or a Western blot. The method is particularly useful for producing therapeutic proteins or glycoproteins. Examples of therapeutic proteins or glycoproteins include but are not limited to erythropoietin (EPO); cytokines such as interferon α, interferon β, interferon γ, and interferon ω; and granulocyte-colony stimulating factor (GCSF); GM-CSF; coagulation factors such as factor VIII, factor IX, and human protein C; antithrombin III; thrombin,; soluble IgE receptor α-chain; immunoglobulins such as IgG, IgG fragments, IgG fusions, and IgM; immunoadhesions and other Fc fusion proteins such as soluble TNF receptor-Fc fusion proteins; RAGE-Fc fusion proteins; interleukins; urokinase; chymase; and urea trypsin inhibitor; IGF-binding protein; epidermal growth factor; growth hormone-releasing factor; annexin V fusion protein; angiostatin; vascular endothelial growth factor-2; myeloid progenitor inhibitory factor-1; osteoprotegerin; α-1-antitrypsin; α-feto proteins; DNase II; kringle 3 of human plasminogen; glucocerebrosidase; TNF binding protein 1; follicle stimulating hormone; cytotoxic T lymphocyte associated antigen 4 – Ig; transmembrane activator and calcium modulator and cyclophilin ligand; glucagon like protein 1; and IL-2 receptor agonist

In particular embodiments of the host cell or method, the codons of the nucleic acid sequence of the nucleic acid molecule encoding the recombinant protein or glycoprotein is optimized for expression in *Pichia pastoris*.

In a further still embodiment of the host cell or method, the host cell is genetically engineered to produce glycoproteins that have human-like *N*-glycans.

In a further embodiment of the host cell or method, the host cell further does not display α1,6-mannosyltransferase activity with respect to the *N*-glycan on a glycoprotein and includes an α1,2-mannosidase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target α1,2-mannosidase activity to the ER or Golgi apparatus of the host cell.

In a further still embodiment of the host cell or method, the host cell further includes a GlcNAc transferase I catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain of and selected to target GlcNAc transferase I activity to the ER or Golgi apparatus of the host cell.

In a further still embodiment of the host cell or method, the host cell further includes a mannosidase II catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target mannosidase II activity to the ER or Golgi apparatus of the host cell.

In a further still embodiment of the host cell or method, the host cell further includes a GlcNAc transferase II catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target GlcNAc transferase II activity to the ER or Golgi apparatus of the host cell.

In a further still embodiment of the host cell or method the host cell further includes a galactosyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target galactosyltransferase activity to the ER or Golgi apparatus of the host cell.

In a further still embodiment of the host cell or method, the host cell further includes a sialyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target sialyltransferase activity to the ER or Golgi apparatus of the host cell.

In a further still embodiment of the host cell or method, the host cell further includes a fucosyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target fucosyltransferase activity to the ER or Golgi apparatus of the host cell.

In a further still embodiment of the host cell or method, the host cell further includes one or more GlcNAc transferases selected from the group consisting of GnTIII, GnTIV, GnTV, GnTVI, and GnTIX.

In a further still embodiment of the host cell or method, the host cell is genetically engineered to produce glycoproteins that have predominantly an *N*-glycan selected from $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GalGlcNAcMan_5GlcNAc_2$, $NANAGalGlcNAcMan_5GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$, $GlcNAc_{(1-4)}Man_3GlcNAc_2$, $Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$, and $NANA_{(1-4)}Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$, wherein

the subscript indicates the number of the particular sugar residues on the *N*-glycan structure. Examples of *N*-glycan structures include but are not limited to $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$, $GlcNAc_2Man_3GlcNAc_2$, $GlcNAc_3Man_3GlcNAc_2$, $GlcNAc_4Man_3GlcNAc_2$, $GalGlcNAc_2Man_3GlcNAc_2$, $Gal_2GlcNAc_2Man_3GlcNAc_2$, $Gal_2GlcNAc_3Man_3GlcNAc_2$, $Gal_2GlcNAc_4Man_3GlcNAc_2$, $Gal_3GlcNAc_3Man_3GlcNAc_2$, $Gal_3GlcNAc_4Man_3GlcNAc_2$, $Gal_4GlcNAc_4Man_3GlcNAc_2$, $NANAGal_2GlcNAc_2Man_3GlcNAc_2$, $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$, $NANA_3Gal_3GlcNAc_3Man_3GlcNAc_2$, and $NANA_4Gal_4GlcNAc_4Man_3GlcNAc_2$.

Further provided are compositions, which comprise one or more recombinant glycoproteins obtained by the above method using any one of the above host cells.

In a further aspect, the present invention provides a recombinant methylotrophic yeast such as *Pichia pastoris* host cell that does not display β-mannosyltransferase 2 activity and at least one activity selected from β-mannosyltransferase 1 activity and β-mannosyltransferase 3 activity and which includes two or more nucleic acid molecules, each encoding a fusion protein comprising a mature human erythropoietin fused to a signal peptide that targets the ER and which is removed when the fusion protein is in the ER. In particular embodiments, the host cell further does not display β-mannosyltransferase 4 activity.

In a further aspect, the present invention provides a recombinant methylotrophic yeast such as *Pichia pastoris* host cell that has a deletion or disruption of the genes encoding β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, and β-mannosyltransferase 3 activity and which includes two or more nucleic acid molecules, each encoding a fusion protein comprising a mature human erythropoietin fused to a signal peptide that targets the ER and which is removed when the fusion protein is in the ER. In particular embodiments, the host cell further includes a deletion or disruption of the gene encoding β-mannosyltransferase 4 activity.

In a further aspect, the present invention provides a recombinant *Pichia pastoris* host cell that has a deletion or disruption of the β-mannosyltransferase 2 (*BMT2*) gene and at least one gene selected from a β-mannosyltransferase 1 (*BMT1*) and β-mannosyltransferase 3 (*BMT3*) gene and which includes two or more nucleic acid molecules, each encoding a fusion protein comprising a mature human erythropoietin fused to a signal peptide that targets the ER and which is removed when the fusion protein is in the ER. In particular embodiments, the host cell further includes a deletion or disruption of the gene encoding β-mannosyltransferase 4 (*BMT4*) gene.

In a further still aspect, the present invention provides a method for producing a mature human erythropoietin in methylotrophic yeast such as *Pichia pastoris* comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens, comprising: providing a recombinant host cell that does not display β-mannosyltransferase 2 activity with respect to an *N*-

glycan or *O*-glycan and does not display at least one activity with respect to an *N*-glycan or *O*-glycan selected from β-mannosyltransferase 1 activity and β-mannosyltransferase 3 activity and is genetically engineered to produce sialic acid-terminated biantennary *N*-glycans and which includes two or more nucleic acid molecules, each encoding a fusion protein comprising a mature human erythropoietin fused to a signal peptide that targets the ER and which is removed when the fusion protein is in the ER; growing the host cell in a medium under conditions effective for expressing and processing the first and second fusion proteins; and recovering the mature human erythropoietin from the medium to produce the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens. In further embodiments, the host cell does not display β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan. In further embodiments, the host cell further does not display β-mannosyltransferase 4 activity.

In a further still aspect, the present invention provides a method for producing a mature human erythropoietin in methylotrophic yeast such as *Pichia pastoris* comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens, comprising: providing a recombinant host cell genetically engineered to produce sialic acid-terminated biantennary *N*-glycans and in which the gene encoding a β-mannosyltransferase 2 activity and at least one gene encoding an activity selected from a β-mannosyltransferase 1 activity and a β-mannosyltransferase 3 activity has been deleted or disrupted and which includes two or more nucleic acid molecules, each encoding a fusion protein comprising a mature human erythropoietin fused to a signal peptide that targets the ER and which is removed when the fusion protein is in the ER; growing the host cell in a medium under conditions effective for expressing and processing the first and second fusion proteins; and recovering the mature human erythropoietin from the medium to produce the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens. In further embodiments, the host comprises a deletion or disruption of the genes encoding a β-mannosyltransferase 2 activity, a β-mannosyltransferase 1 activity, and a β-mannosyltransferase 3 activity have been deleted or disrupted. In further embodiments, the host cell further includes a deletion or disruption of a gene encoding a β-mannosyltransferase 3 activity.

In a further still aspect, the present invention provides a method for producing a mature human erythropoietin in *Pichia pastoris* comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens, comprising: providing a recombinant *Pichia pastoris* host cell genetically engineered to produce sialic acid-terminated biantennary *N*-glycans and in which the

β-mannosyltransferase 2 (*BMT2*) gene and at least one gene selected from a β-mannosyltransferase 1 (*BMT1*) and β-mannosyltransferase 3 (*BMT3*) gene has been deleted or disrupted and which includes two or more nucleic acid molecules, each encoding a fusion protein comprising a mature human erythropoietin fused to a signal peptide that targets the ER and which is removed when the fusion protein is in the ER; growing the host cell in a medium under conditions effective for expressing and processing the first and second fusion proteins; and recovering the mature human erythropoietin from the medium to produce the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens. In further embodiments, the host comprises a deletion or disruption of the β-mannosyltransferase 2 (BMT2) gene, β-mannosyltransferase 1 (*BMT1*) gene, and β-mannosyltransferase 3 (BMT3) gene have been deleted or disrupted. In further embodiments, the host cell further includes a deletion or disruption of a β-mannosyltransferase 3 gene (*BMT4*).

In particular embodiments of the host cell or method, the signal peptide fused to the *N*-terminus of the erythropoietin is a *S. cerevisiae* αMATpre signal peptide or a chicken lysozyme signal peptide.

In further embodiments of the host cell or method, at least one nucleic acid molecule encodes a fusion protein wherein the erythropoietin is fused to the *S. cerevisiae* αMATpre signal peptide and at least one nucleic acid molecule encodes a fusion protein wherein the erythropoietin is fused to the *S. cerevisiae* αMATpre signal peptide a chicken lysozyme signal peptide.

In further embodiments of the host cell or method, the codons of the nucleic acid sequence of the nucleic acid molecule encoding the erythropoietin is optimized for expression in *Pichia pastoris*.

In further embodiments of the method, recovering the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens from the medium includes a cation exchange chromatography step.

In further embodiments of the method, recovering the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens from the medium includes a hydroxyapatite chromatography step.

In further embodiments of the method, recovering the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens from the medium includes an anion exchange chromatography step.

In further embodiments of the method, recovering the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens from the medium includes a cation exchange chromatography step followed by a hydroxyapatite chromatography step, which is optionally followed by an anion exchange chromatography step.

The present invention further provides a composition comprising a mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens obtained from the above method using the above host cells and a pharmaceutically acceptable salt. In particular embodiments, about 50 to 60% of the *N*-glycans comprise sialic acid residues on both antennae; in further embodiments, greater than 70% of the *N*-glycans comprise sialic acid residues on both antennae; in further embodiments, greater than 80% of the *N*-glycans comprise sialic acid residues on both antennae. In further aspects, less than 30% of the *N*-glycans are neutral *N*-glycans (i.e., are not sialylated on at least one terminus at the non-reducing end of the *N*-glycan). In further still aspects, less than 20% of the *N*-glycans are neutral *N*-glycans. In particular aspects, about 99% of the *N*-glycans contain one ore more sialic acid residues and less than 1% of the *N*-glycans are neutral *N*-glycans. In further aspects, compositions are provided wherein there is 4.5 moles or more of sialic acid per mole of rhEPO. In further aspects, compositions are provided wherein there is at least 5.0 moles of sialic acid per mole of rhEPO.

In further embodiments of the composition, the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens is conjugated to a hydrophilic polymer, which in particular aspects is a polyethylene glycol polymer. In particular embodiments, the polyethylene glycol polymer is conjugated to the N-terminus of the mature human erythropoietin comprising predominantly sialic acid-terminated bi-antennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens.

Definitions

As used herein, the terms "*N*-glycan" and "glycoform" are used interchangeably and refer to an *N*-linked oligosaccharide, *e.g.*, one that is attached by an asparagine-*N*-acetylglucosamine linkage to an asparagine residue of a polypeptide. *N*-linked glycoproteins contain an *N*-acetylglucosamine residue linked to the amide nitrogen of an asparagine residue in the protein. The predominant sugars found on glycoproteins are glucose, galactose, mannose, fucose, *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc) and sialic acid (*e.g.*, *N*-acetyl-neuraminic acid (NANA)). The processing of the sugar groups occurs co-translationally

in the lumen of the ER and continues post-translationally in the Golgi apparatus for *N*-linked glycoproteins.

N-glycans have a common pentasaccharide core of $Man_3GlcNAc_2$ ("Man" refers to mannose; "Glc" refers to glucose; and "NAc" refers to *N*-acetyl; GlcNAc refers to *N*-acetylglucosamine). *N*-glycans differ with respect to the number of branches (antennae) comprising peripheral sugars (*e.g.*, GlcNAc, galactose, fucose and sialic acid) that are added to the $Man_3GlcNAc_2$ ("Man3") core structure which is also referred to as the "trimannose core", the "pentasaccharide core" or the "paucimannose core". *N*-glycans are classified according to their branched constituents (*e.g.*, high mannose, complex or hybrid). A "high mannose" type *N*-glycan has five or more mannose residues. A "complex" type *N*-glycan typically has at least one GlcNAc attached to the 1,3 mannose arm and at least one GlcNAc attached to the 1,6 mannose arm of a "trimannose" core. Complex *N*-glycans may also have galactose ("Gal") or *N*-acetylgalactosamine ("GalNAc") residues that are optionally modified with sialic acid or derivatives (*e.g.*, "NANA" or "NeuAc", where "Neu" refers to neuraminic acid and "Ac" refers to acetyl). Complex *N*-glycans may also have intrachain substitutions comprising "bisecting" GlcNAc and core fucose ("Fuc"). Complex *N*-glycans may also have multiple antennae on the "trimannose core," often referred to as "multiple antennary glycans." A "hybrid" N-glycan has at least one GlcNAc on the terminal of the 1,3 mannose arm of the trimannose core and zero or more mannoses on the 1,6 mannose arm of the trimannose core. The various N-glycans are also referred to as "glycoforms."

Abbreviations used herein are of common usage in the art, see, *e.g.*, abbreviations of sugars, above. Other common abbreviations include "PNGase", or "glycanase" or "glucosidase" which all refer to peptide N-glycosidase F (EC 3.2.2.18).

The term "recombinant host cell" ("expression host cell", "expression host system", "expression system" or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A recombinant host cell may be an isolated cell or cell line grown in culture or may be a cell which resides in a living tissue or organism. Preferred host cells are yeasts and fungi.

A host cell that "does not display" an enzyme activity refers to a host cell in which the enzyme activity has been abrogated or disrupted. For example, the enzyme activity can be abrogated or disrupted by deleting or disrupting the gene encoding the enzyme activity (included deleting or disrupting the upstream or downstream regulatory sequences controlling expression of the gene; the enzyme activity can be abrogated or disrupted by mutating the gene encoding the

enzyme activity to render the enzyme activity encoded gene non-functional; the enzyme activity can be abrogated or disrupted by use of a chemical, peptide, or protein inhibitor of the enzyme activity; the enzyme activity can be abrogated or disrupted by use of nucleic acid-based expression inhibitors such as antisense DNA and siRNA; and, the enzyme activity can be abrogated or disrupted by use of transcription inhibitors or inhibitors of the expression or activity of regulatory factors that control or regulate expression of the gene encoding the enzyme activity.

When referring to "mole percent" of a glycan present in a preparation of a glycoprotein, the term means the molar percent of a particular glycan present in the pool of N-linked oligosaccharides released when the protein preparation is treated with PNG'ase and then quantified by a method that is not affected by glycoform composition, (for instance, labeling a PNG'ase released glycan pool with a fluorescent tag such as 2-aminobenzamide and then separating by high performance liquid chromatography or capillary electrophoresis and then quantifying glycans by fluorescence intensity). For example, 50 mole percent $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$ means that 50 percent of the released glycans are $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$ and the remaining 50 percent are comprised of other N-linked oligosaccharides. In embodiments, the mole percent of a particular glycan in a preparation of glycoprotein will be between 20% and 100%, preferably above 25%, 30%, 35%, 40% or 45%, more preferably above 50%, 55%, 60%, 65% or 70% and most preferably above 75%, 80% 85%, 90% or 95%.

As used herein, the term "predominantly" or variations such as "the predominant" or "which is predominant" will be understood to mean the glycan species that has the highest mole percent (%) of total *N*-glycans after the glycoprotein has been treated with PNGase and released glycans analyzed by mass spectroscopy, for example, MALDI-TOF MS. In other words, the phrase "predominantly" is defined as an individual entity, such as a specific glycoform, is present in greater mole percent than any other individual entity. For example, if a composition consists of species A in 40 mole percent, species B in 35 mole percent and species C in 25 mole percent, the composition comprises predominantly species A.

The term "therapeutically effective amount" refers to an amount of the recombinant erythropoietin of the invention which gives an increase in hematocrit that provides benefit to a patient. The amount will vary from one individual to another and will depend upon a number of factors, including the overall physical condition of the patient and the underlying cause of anemia. For example, a therapeutically effective amount of erythropoietin of the present invention for a patient suffering from chronic renal failure can be in the range of 20 to 300 units/kg or 0.5ug/kg to 500ug/kg based on therapeutic indication. The term "unit" refers to units commonly known in the art for assessing the activity of erythropoietin compositions. A milligram of pure erythropoietin is approximately equivalent to 150,000 units. A dosing schedule can be from about three times per week to about once every four or six weeks. The

actual schedule will depend on a number of factors including the type of erythropoietin administered to a patient (EPO or PEGylated-EPO) and the response of the individual patient. The higher dose ranges are not typically used in anemia applications but can be useful on other therapeutic applications. The means of achieving and establishing an appropriate dose of erythropoietin for a patient is well known and commonly practiced in the art.

Variations in the amount given and dosing schedule from patient to patient are including by reference to the term "about" in conjunction with an amount or schedule. The amount of erythropoietin used for therapy gives an acceptable rate of hematocrit increase and maintains the hematocrit at a beneficial level (for example, usually at least about 30% and typically in a range of 30% to 36%). A therapeutically effective amount of the present compositions may be readily ascertained by one skilled in the art using publicly available materials and procedures. Additionally, iron may be given to the patient to maintain increased erythropoiesis during therapy. The amount to be given may be readily determined by methods commonly used by those skilled in the art.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 A-J shows the genealogy of *P. pastoris* strain **YGLY3159** (**Figure 1E**) and strains **YGLY7113** to **YGLY7122** (**Figure 1I**) beginning from wild-type strain **NRRL-Y11430** (**Figure 1A**).

Figure 2 shows a map of plasmid pGLY6. Plasmid pGLY6 is an integration vector that targets the *URA5* locus and contains a nucleic acid molecule comprising the *S. cerevisiae* invertase gene or transcription unit (ScSUC2) flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *P. pastoris URA5* gene (PpURA5-5') and on the other side by a nucleic acid molecule comprising the a nucleotide sequence from the 3' region of the *P. pastoris URA5* gene (PpURA5-3').

Figure 3 shows a map of plasmidpGLY40. Plasmid pGLY40 is an integration vector that targets the *OCH1* locus and contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by nucleic acid molecules comprising *lacZ* repeats (lacZ repeat) which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *OCH1* gene (PpOCH1-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *OCH1* gene (PpOCH1-3').

Figure 4 shows a map of plasmid pGLY43a. Plasmid pGLY43a is an integration vector that targets the *BMT2* locus and contains a nucleic acid molecule comprising the *K. lactis* UDP-*N*-acetylglucosamine (UDP-GlcNAc) transporter gene or transcription unit (KlGlcNAc Transp.) adjacent to a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by nucleic acid molecules comprising *lacZ* repeats (lacZ

repeat). The adjacent genes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *BMT2* gene (PpPBS2-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *BMT2* gene (PpPBS2-3').

Figure 5 shows a map of plasmid pGLY48. Plasmid pGLY48 is an integration vector that targets the *MNN4L1* locus and contains an expression cassette comprising a nucleic acid molecule encoding the mouse homologue of the UDP-GlcNAc transporter (MmGlcNAc Transp.) open reading frame (ORF) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris GAPDH* promoter (PpGAPDH Prom) and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* termination sequence (ScCYC TT) adjacent to a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by *lacZ* repeats (lacZ repeat) and in which the expression cassettes together are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *P. Pastoris MNN4L1* gene (PpMNN4L1-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MNN4L1* gene (PpMNN4L1-3').

Figure 6 shows as map of plasmid pGLY45. Plasmid pGLY45 is an integration vector that targets the *PNO1/MNN4* loci contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by nucleic acid molecules comprising *lacZ* repeats (lacZ repeat) which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *PNO1* gene (PpPNO1-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MNN4* gene (PpMNN4-3').

Figure 7 shows a map of plasmid pGLY247. Plasmid pGLY247 is an integration vector that targets the *MET16* locus and contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by nucleic acid molecules comprising *lacZ* repeats (lacZ repeat) which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *MET16* gene (PpMET16-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MET16* gene (PpMET16-3').

Figure 8 shows a map of plasmid pGLY248. Plasmid pGLY248 is an integration vector that targets the *URA5* locus and contains a nucleic acid molecule comprising the *P. pastoris MET16* gene or transcription unit (PpMET16) flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *URA5* gene (PpURA5-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *URA5* gene (PpURA5-3').

Figure 9 shows a map of plasmid pGLY582. Plasmid pGLY582 is an integration vector that targets the *HIS1* locus and contains in tandem four expression cassettes encoding (1)

the *S. cerevisiae* UDP-glucose epimerase (ScGAL10), (2) the human galactosyltransferase I (hGalT) catalytic domain fused at the N-terminus to the *S. cerevisiae KRE2-s* leader peptide (33), (3) the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by lacZ repeats (lacZ repeat), and (4) the *D. melanogaster* UDP-galactose transporter (DmUGT). All flanked by the 5' region of the *HIS1* gene (PpHIS1-5') and the 3' region of the *HIS1* gene (PpHIS1-3'). PMA1 is the *P. pastoris PMA1* promoter; PpPMA1 TT is the *P. pastoris PMA1* termination sequence; GAPDH is the *P. pastoris GADPH* promoter and ScCYC TT is the *S. cerevisiae CYC* termination sequence; PpOCH1 Prom is the *P. pastoris OCH1* promoter and PpALG12 TT is the *P. pastoris ALG12* termination sequence.

Figure 10 shows a map of plasmid pGLY167b. Plasmid pGLY167b is an integration vector that targets the *ARG1* locus and contains in tandem three expression cassettes encoding (1) the *D. melanogaster* mannosidase II catalytic domain (codon optimized) fused at the *N*-terminus to *S. cerevisiae MNN2* leader peptide (CO-KD53), (2) the *P. pastoris HIS1* gene or transcription unit, and (3) the rat *N*-acetylglucosamine (GlcNAc) transferase II catalytic domain (codon optimized) fused at the *N*-terminus to *S. cerevisiae MNN2* leader peptide (CO-TC54). All flanked by the 5' region of the *ARG1* gene (PpARG1-5') and the 3' region of the *ARG1* gene (PpARG1-3'). PpPMA1 prom is the *P. pastoris PMA1* promoter; PpPMA1 TT is the *P. pastoris PMA1* termination sequence; PpGAPDH is the *P. pastoris GADPH* promoter; ScCYC TT is the *S. cerevisiae CYC* termination sequence; PpOCH1 Prom is the *P. pastoris OCH1* promoter; and PpALG12 TT is the *P. pastoris ALG12* termination sequence.

Figure 11 shows a map of plasmid pGLY1430. Plasmid pGLY1430 is a KINKO integration vector that targets the *ADE1* locus without disrupting expression of the locus and contains in tandem four expression cassettes encoding (1) the human GlcNAc transferase I catalytic domain (codon optimized) fused at the *N*-terminus to *P. pastoris SEC12* leader peptide (CO-NA10), (2) mouse homologue of the UDP-GlcNAc transporter (MmTr), (3) the mouse mannosidase IA catalytic domain (FB) fused at the *N*-terminus to *S. cerevisiae SEC12* leader peptide (FB8), and (4) the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by lacZ repeats (lacZ). All flanked by the 5' region of the *ADE1* gene and ORF (ADE1 5' and ORF) and the 3' region of the *ADE1* gene (PpADE1-3'). PpPMA1 prom is the *P. pastoris PMA1* promoter; PpPMA1 TT is the *P. pastoris PMA1* termination sequence; SEC4 is the *P. pastoris SEC4* promoter; OCH1 TT is the *P. pastoris OCH1* termination sequence; ScCYC TT is the *S. cerevisiae CYC* termination sequence; PpOCH1 Prom is the *P. pastoris OCH1* promoter; PpALG3 TT is the *P. pastoris ALG3* termination sequence; and PpGAPDH is the *P. pastoris GADPH* promoter.

Figure 12 shows a map of plasmid pGFI165. Plasmid pGFI165 is a KINKO integration vector that targets the *PRO1* locus without disrupting expression of the locus and contains expression cassettes encoding (1) the *T. reesei* α-1,2-mannosidase catalytic domain

fused at the *N*-terminus to *S. cerevisiae* αMATpre signal peptide (aMATTrMan) to target the chimeric protein to the secretory pathway and secretion from the cell and (2) the *P. pastoris URA5* gene or transcription unit flanked by lacZ repeats (lacZ repeat). All flanked by the 5' region of the *PRO1* gene and ORF (5'PRO1orf) and the 3' region of the *PRO1* gene (3'PRO). ScCYC TT is the *S. cerevisiae CYC* termination sequence; PpALG3 TT is the *P. pastoris ALG3* termination sequence; and PpGAPDH is the *P. pastoris GADPH* promoter.

Figure 13 shows a map of plasmid pGLY2088. Plasmid pGLY2088 is an integration vector that targets the *TRP2* or AOX1 locus and contains expression cassettes encoding (1) mature human erythropoetin (co-hEPO) codon optimized fused at the *N*-terminus to a *S. cerevisiae* αMATpre signal peptide (alpha MF-pre) to target the chimeric protein to the secretory pathway and secretion from the cell and (2) the zeocin resistance protein (ZeocinR). The cassettes are flanked on one end with the *P. pastoris AOX1* promoter (PpAOX1 Prom) and on the other end with the *P. pastoris TRP2* gene or transcription unit (PpTRP2). ScCYC TT is the *S. cerevisiae CYC* termination sequence and ScTEF Prom is the *S. cerevisiae TEF1* promoter.

Figure 14 shows a map of plasmid pGLY2456. Plasmid pGLY2456 is a KINKO integration vector that targets the *TRP2* locus without disrupting expression of the locus and contains six expression cassettes encoding (1) the mouse CMP-sialic acid transporter codon optimized (CO mCMP-Sia Transp), (2) the human UDP-GlcNAc 2-epimerase/*N*-acetylmannosamine kinase codon optimized (CO hGNE), (3) the *Pichia pastoris ARG1* gene or transcription unit, (4) the human CMP-sialic acid synthase codon optimized (CO hCMP-NANA S), (5) the human N-acetylneuraminate-9-phosphate synthase codon optimized (CO hSIAP S), and, (6) the mouse a-2,6-sialyltransferase catalytic domain codon optimized fused at the *N*-terminus to *S. cerevisiae KRE2* leader peptide (comST6-33). All flanked by the 5' region of the *TRP2* gene and ORF (PpTRP2 5') and the 3' region of the *TRP2* gene (PpTRP2-3'). PpPMA1 prom is the *P. pastoris PMA1* promoter; PpPMA1 TT is the *P. pastoris PMA1* termination sequence; CYC TT is the *S. cerevisiae CYC* termination sequence; PpTEF Prom is the *P. pastoris TEF1* promoter; PpTEF TT is the *P. pastoris TEF1* termination sequence; PpALG3 TT is the *P. pastoris ALG3* termination sequence; and pGAP is the *P. pastoris GAPDH* promoter.

Figure 15 shows a map of plasmid pGLY3411 (pSH1092). Plasmid pGLY3411 (pSH1092) is an integration vector that contains the expression cassette comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by *lacZ* repeats (lacZ repeat) flanked on one side with the 5' nucleotide sequence of the *P. pastoris BMT4* gene (PpPBS4 5') and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT4* gene (PpPBS4 3').

Figure 16 shows a map of plasmid pGLY3430 (pSH1115). Plasmid pGLY3430 (pSH1115) is an integration vector that contains an expression cassette comprising a nucleic acid molecule encoding the Nourseothricin resistance ORF (NAT) operably linked to the *Ashbya gossypii TEF1* promoter (PTEF) and *Ashbya gossypii TEF1* termination sequence (TTEF)

flanked one side with the 5' nucleotide sequence of the *P. pastoris BMT1* gene (PBS1 5') and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT1* gene (PBS1 3').

Figure 17 shows a map of plasmid pGLY4472 (pSH1186). Plasmid pGLY4472 (pSH1186) contains an expression cassette comprising a nucleic acid molecule encoding the *E. coli* hygromycin B phosphotransferase gene ORF (Hyg) operably linked to the *Ashbya gossypii TEF1* promoter (pTEF) and *Ashbya gossypii TEF1* termination sequence (TRFtt) flanked one side with the 5' nucleotide sequence of the *P. pastoris BMT3* gene (PpPBS3 5') and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT3* gene (PpPBS3 3').

Figure 18 shows a map of plasmid pGLY2057. Plasmid pGLY2057 is an integration plasmid that targets the *ADE2* locus and contains an expression cassette encoding the *P.pastoris URA5* gene or transcription unit (PpURA5) flanked by *lacZ repeats* (lacZ repeat). The expression cassette is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *ADE2* gene (PpADE2-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *ADE2* gene (PpADE2-3').

Figure 19 shows a map of plasmid pGLY2680. Plasmid pGLY2680 is an integration vector that can target the *TRP2* or *AOX1* locus and contains expression cassettes encoding (1) the human mature erythropoetin codon optimized (co-hEPO) fused at the *N*-terminus to chicken lysozyme signal peptide (chicken Lysozyme ss) and (2) the *P. pastoris ADE2* gene without a promoter (PpADE2). The cassettes are flanked on one end with the *P. pastoris AOX1* promoter (PpAOX1 Prom) and on the other end with the *P. pastoris TRP2* gene or transcription unit (PpTRP2). ScCYC TT is the *S. cerevisiae CYC* termination sequence.

Figure 20 shows a map of plasmid pGLY2713. Plasmid pGLY2713 is an integration vector containing the *P. pastoris PNO1* ORF (PpPNO1 ORF) adjacent to the expression cassette comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by *lacZ* repeats (lacZ repeat) and flanked on one side with the 5' nucleotide sequence of the *P. pastoris PEP4* gene (PpPEP4 5') and on the other side with the 3' nucleotide sequence of the *P. pastoris PEP4* gene (PpPEP4 3').

Figure 21 shows a schematic diagram illustrating fermentation process flow.

Figure 22 shows that rhEPO produced in strain **YGLY3159** has cross binding activity to anti-HCA antibodies. Left panel shows a Commassie Blue stained 4-20% SDS-PAGE gel showing the position of the rhEPO and right panel shows a Western blot of a similar gel probed with rabbit anti-HCA antibodies (SL rProA purified rabbit: 9161) at 1:3,000 dilution. Bound anti-HCA antibody was detected using goat anti-rabbit antibody conjugated to horseradish peroxidase (HRP) at a 1:5,000 dilution in PBS. Detection of bound secondary antibody used the substrate 3'3 diaminobenzidine (DAB).

Figure 23 shows that the cross-bind activity of the rhEPO produced in strain **YGLY3159** to anti-HCA antibodies is not detected when the rhEPO is deglycosylated using PNAGase F. Left panel shows a Commassie Blue stained 4-20% SDS-PAGE gel showing the position of the glycosylated and deglycosylated forms of rhEPO and right panel shows a Western blot of a similar gel probed with anti-HCA antibodies as in **Figure 22**.

Figure 24 shows that a recombinant antibody (rhIgG) produced in wild-type *P. pastoris* and a glycoengineered *P. pastoris* GS2.0 strain in which the *BMT2* gene has been disrupted or deleted showed cross binding activity to anti-HCA antibodies. Left panel shows a Commassie Blue stained 4-20% SDS-PAGE gel and the right panel shows a Western blot of a similar gel probed with anti-HCA antibodies as in **Figure 22**. GS 2.0 is a *P. pastoris* strain that produces glycoproteins that have predominantly $Man_5GlcNAc_2$ *N*-glycans. The shown GS 2.0 strain produced rhIgG with about 5% $Man_9GlcNAc_2$ *N*-glycans. WT is wild type *P. pastoris*.

Figure 25 compares cross binding activity of rhEPO produced in strain **YGLY3159** to other glycosylated proteins containing complex glycosylation patterns but not produced in *P. pastoris* to anti-HCA antibody. Upper panel shows a Commassie Blue stained 4-20% SDS-PAGE gel showing the position of the glycosylated and deglycosylated forms of rhEPO produced in *P. pastoris* and of recombinant human fetuin, asialofetuin (human fetuin with terminal sialic acid residues removed), human serum albumin (HSA), and recombinant LEUKINE produced in *S. cerevisiae* and the lower panel shows a Western blot of a similar gel probed with anti-HCA antibodies as in **Figure 22**. S30S pools are rhEPO purified by cation exchange chromatography.

Figure 26 shows that rhEPO produced in strain **YGLY3159** and purified by hydroxyapatite chromatography still has cross binding activity to anti-HCA antibodies. Left panel shows a Commassie Blue stained 4-20% SDS-PAGE gel of chromatography elution pools 1, 2, and 3 showing the position of the rhEPO (reduced or non-reduced) and right panel shows a Western blot of a similar gel probed with anti-HCA antibodies as in **Figure 22**. Below the panels is shown the results of an HPLC analysis of N-glycans in pools 1, 2, and 3.

Figure 27A shows a chromatogram of Q SEPHAROSE FF anion chromatography purification of rhEPO produced in strain **YGLY3159** from hydroxyapatite pool 1.

Figure 27B shows a sandwich ELISA showing that the Q SEPHAROSE FF pool containing rhEPO from the Q SEPHAROSE FF anion chromatography has no detectable cross binding activity to anti-HCA antibodies whereas the flow through contained cross binding activity to anti-HCA antibodies. The capture antibody was anti-hEPO antibody and cross binding activity was detected with rabbit anti-HCA antibody at a 1:800 starting dilution in PBS which was then serially diluted 1:1 in PBS across a row ending with the 11^{th} well at a 1:819,200 dilution (well 12: negative control). Bound anti-HCA antibody was detected using goat anti-

rabbit antibody conjugated to alkaline phosphatase (AP) at a 1:10,000 dilution in PBS. Detection of bound secondary antibody used the substrate 4-Methylumbelliferyl phosphate (4-MUPS).

Figure 28 shows that rhEPO produced in strains **YGLY6661** (*Δbmt2*, *Δbmt4*, and *Δbmt1*) and **YGLY7013** (*Δbmt2* and *Δbmt4*) and captured by Blue SEPHAROSE 6 FF chromatography (Blue pools) still has cross binding activity to anti-HCA antibodies. Left panel shows a Commassie Blue stained 4-20% SDS-PAGE gel of the Blue pools with (+) and without (-) PNGase F treatment. The center panel shows a Western blot of a similar gel probed with anti-hEPO antibodies conjugated to HRP at a 1:1,000 dilution and DAB as the substrate. The right panel shows a Western blot of a similar gel probed with anti-HCA antibodies as in **Figure 22**.

Figure 29 shows in a sandwich ELISA to detect cross binding activity to anti-HCA antibodies that rhEPO produced in strains **YGLY6661** (*Δbmt2*, *Δbmt4*, and *Δbmt1*) and **YGLY7013** (*Δbmt2* and *Δbmt4*) and captured by Blue SEPHAROSE 6 FF chromatography (Blue pools) still has cross binding activity to anti-HCA antibodies. The ELISA was performed as in **Figure 27B**.

Figure 30 shows sandwich ELISAs used to detect cross binding activity to anti-HCA antibodies of rhEPO produced in strains **YGLY6661** (*Δbmt2*, *Δbmt4*, and *Δbmt1*) and **YGLY7013** (*Δbmt2* and *Δbmt4*), captured by Blue SEPHAROSE 6 FF chromatography, and purified by hydroxyapatite chromatography (HA pool 1). rhEPO in HA pool 1 from strain **YGLY6661** had no detectable cross binding activity to anti-HCA antibodies. The ELISAs were performed as in **Figure 27B**.

Figure 31 shows that rhEPO produced in strain **YGLY6661** (*Δbmt2*, *Δbmt4*, and *Δbmt1*) and captured by Blue SEPHAROSE 6 FF chromatography (Blue pools) still has cross binding activity to anti-HCA antibodies. Left panel shows a Commassie Blue stained 4-20% SDS-PAGE gel of the Blue pools with (+) and without (-) PNGase F treatment. The center panel shows a Western blot of a similar gel probed with anti-hEPO antibodies conjugated to HRP at a 1:1,000 dilution and DAB as the substrate. The right panel shows a Western blot of a similar gel probed with anti-HCA antibodies as in **Figure 22**.

Figure 32A shows a Commassie Blue stained 4-20% SDS-PAGE gel of the Blue Sepaharose 6 FF capture pools (Blue pools) prepared from strains **YGLY7361-7366** (all *Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*) with (+) and without (-) PNGase F treatment. The strains were grown in 500 mL SixFors fermentors.

Figure 32B shows a Commassie Blue stained 4-20% SDS-PAGE gel of the Blue Sepaharose 6 FF capture pools (Blue pools) prepared from strains **YGLY7393-7398** (all *Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*) with (+) and without (-) PNGase F treatment. The strains were grown in 500 mL SixFors fermentors.

Figure 33 shows the results of sandwich ELISAs used to detect cross binding activity to anti-HCA antibodies of rhEPO produced in strains **YGLY7361-7366** (all *Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*) and **YGLY7393-7398** (all *Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*) and captured by Blue SEPHAROSE 6 FF chromatography (Blue pools). Only rhEPO in the Blue pools from strain **YGLY7363** and **YGLY7365** had detectable cross binding activity to anti-HCA antibodies. The ELISAs were performed as in **Figure 27B**.

Figure 34 shows in chart form the results from HPLC analysis of the *N*-glycans on the rhEPO in the Blue pools prepared from strains **YGLY7361-7366** and **YGLY7393-7398** (all *Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*). "Bi" refers to *N*-glycans in which both arms of the biantennary *N*-glycan are sialylated. "Mono" refers to *N*-glycans in which only one arm of the biantennary *N*-glycan is sialylated. "Neutral" refers to *N*-glycans that are not sialylated.

Figure 35A shows a Commassie Blue stained 4-20% SDS-PAGE gel of the Blue SEPHAROSE 6 FF chromatography (Blue pools) and hydroxyapatite purification pools (HA pool 1s) prepared from strains **YGLY7362**, **YGLY7366**, **YGLY7396**, and **YGLY7398** (all *Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*), and **YGLY3159** (*Δbmt2*).

Figure 35B shows a Western blot of a 4-20% SDS-PAGE gel of the Blue SEPHAROSE 6 FF chromatography (Blue pools) and hydroxyapatite purification pools (HA pool 1s) prepared from strains **YGLY7362**, **YGLY7366**, **YGLY7396**, and **YGLY7398** (all *Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*), and **YGLY3159** (*Δbmt2*) and probed with anti-HCA antibodies as in **Figure 22**.

Figure 36 shows that rhEPO produced in strain **YGLY7398** (*Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*) and captured by Blue SEPHAROSE 6 FF chromatography (Blue pools) and purified by hydroxyapatite chromatography (HA pool 1s) had no detectable cross binding activity to anti-HCA antibodies. Left panel shows a Commassie Blue stained 4-20% SDS-PAGE gel of the Blue pool and HA pool 1 prepared from strain **YGLY7398** compared to rhEPO prepared from strain **YGLY3159**. The center panel shows a Western blot of a similar gel probed with anti-HCA antibodies as in **Figure 22**. The center panel shows a Western blot of a similar gel probed with anti-HCA antibodies as in **Figure 22** except anti-HCA antibodies were from another antibody preparation (GiF polyclonal rabbit::6316 at 1:2,000).

Figure 37 shows the results of sandwich ELISAs used to detect cross binding activity to anti-HCA antibodies of rhEPO produced in strains **YGLY7113-7122** (*Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*) and captured by Blue SEPHAROSE 6 FF chromatography (Blue pools). Strain **YGLY7118** showed very low detectable cross binding activity to anti-HCA antibodies. None of the other strains showed any detectable cross binding activity to anti-HCA antibodies. The ELISAs were performed as in **Figure 27B**.

Figure 38 shows in chart form the results from HPLC analysis of the *N*-glycans on the rhEPO in the Blue pools prepared from strains **YGLY7113-7122** (all *Δbmt2*, *Δbmt4*,

Δbmt1, and *Δbmt3*). "Bi" refers to *N*-glycans in which both arms of the biantennary *N*-glycan are sialylated. "Mono" refers to *N*-glycans in which only one arm of the biantennary *N*-glycan is sialylated. "Neutral" refers to *N*-glycans that are not sialylated.

Figure 39A shows a Commassie Blue stained 4-20% SDS-PAGE gel of the Blue SEPHAROSE 6 FF chromatography (Blue pools) and hydroxyapatite purification pools (HA pool 1s) prepared from strains **YGLY7115**, **YGLY7117**, **YGLY7394**, **YGLY7395**, and **YGLY7120** (all *Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*), and **YGLY3159** (*Δbmt2*).

Figure 39B shows a Western blot of a 4-20% SDS-PAGE gel of the Blue SEPHAROSE 6 FF chromatography (Blue pools) and hydroxyapatite purification pools (HA pool 1s) prepared from strains **YGLY7115**, **YGLY7117**, **YGLY7394**, **YGLY7395**, and **YGLY7120** (all *Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*), and **YGLY3159** (*Δbmt2*) and probed with anti-HCA antibodies as in **Figure 22**.

Figure 40A shows an HPLC trace of the *N*-glycans from rhEPO produced in **YGLY3159** (*Δbmt2*) and purified by hydroxyapatite column chromatography (i.e., analysis of HA pool 1).

Figure 40B shows an HPLC trace of the *N*-glycans from rhEPO produced in **YGLY7117** (*Δbmt2*, *Δbmt4*, *Δbmt1*, and *Δbmt3*) and purified by hydroxyapatite column chromatography (i.e., analysis of HA pool 1).

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides methods for producing proteins and glycoproteins in methylotrophic yeast such as *Pichia pastoris* that lack detectable cross binding to antibodies made against host cell antigens. Host cell antigens can also include residual host cell protein and cell wall contaminants that may carry over to recombinant protein compositions that can be immunogenic and which can alter therapeutic efficacy or safety of a therapeutic protein. A composition that has cross-reactivity with antibodies made against host cell antigens means that the composition contains some contaminating host cell material, usually *N*-glycans with phosphomannose residues or β-mannose residues or the like. Wild-type strains of *Pichia pastoris* will produce glycoproteins that have these *N*-glycan structures. Antibody preparations made against total host cell proteins would be expected to include antibodies against these structures. Proteins that do not contain *N*-glycans, however, might also include contaminating material (proteins or the like) that will cross-react with antibodies made against the host cell.

The methods and host cells enable recombinant therapeutic proteins and glycoproteins to be produced that have a reduced risk of eliciting an adverse reaction in an individual administered the recombinant therapeutic proteins and glycoproteins compared to the same being produced in strains not modified as disclosed herein. An adverse reaction includes eliciting an unwanted immune response in the individual or an unwanted or inappropriate

binding to, congregating in, or interaction with a site in the individual that in general adversely affects the health of the individual. The risk of eliciting an adverse reaction in an individual being administered the therapeutic protein or glycoprotein is of particular concern for proteins or glycoproteins intended to be administered to the individual chronically (e.g., therapies intended to be conducted over an extended time period). The recombinant therapeutic proteins or glycoproteins produced according to the methods herein have no detectable cross binding activity to antibodies against host cell antigens and thus, present a reduced risk of eliciting an adverse reaction in an individual administered the recombinant proteins or glycoproteins. The methods and host cells are also useful for producing recombinant proteins or glycoproteins that have a lower potential for binding clearance factors.

The inventors have found that particular glycoproteins that are produced in some strains of *Pichia pastoris* can have *N*- or *O*-glycans thereon in which one or more of the mannose residues thereon are in a β1,2-linkage. Glycoproteins intended for therapeutic uses and which have one or more β1,2-linked mannose residues thereon provide a risk of being capable of eliciting an undesirable immune response in the individual being administered the glycoprotein. These β-linked mannose residues can be detected using antibodies made against total host cell antigens. Because it cannot be predicted which therapeutic glycoproteins will have *N*- or *O*-glycans comprising one or more β1,2-linked mannose residues and whether a therapeutic glycoprotein that does have *N*- or *O*-glycans comprising β1,2-linked mannose residues thereon will produce an unwanted immunogenic response in the individual receiving the glycoprotein, it is desirable to produce therapeutic glycoproteins in *Pichia pastoris* strains that have been genetically engineered to that lack detectable cross binding to antibodies made against host cell antigens. Such strains can be produced by deleting or disrupting the activities of at least three of the four known β-mannosyltransferases (Bmtp) in the *Pichia pastoris* β-mannosyltransferase (*BMT*) gene family. As shown herein, *Pichia pastoris* strains that include a deletion or disruption of at least three of the these *BMT* genes provides a *Pichia pastoris* strain that can produce proteins or glycoproteins that lack detectable cross binding to antibodies made against host cell antigens. These strains are useful producing therapeutic proteins and glycoproteins. The presence of β-mannose structures on N- and/or O-glycans have been demonstrated to elicit an immune response.

Identification of the β-mannosyltransferase genes in *Pichia pastoris* and *Candida albicans* was reported in U.S. Patent No. 7,465,577 and Mille *et al.*, J. Biol. Chem. 283: 9724-9736 (2008), which disclosed that β-mannosylation was effected by a β-mannosyltransferase that was designated *AMR2* or *BMT2* and that disruption or deletion of the gene in *Pichia pastoris* resulted a recombinant host that was capable of producing glycoproteins with reduced β-mannosylation. The patent also disclosed three homologues of the gene, *BMT1*, *BMT3*, and *BMT4*. However, when investigating the source of cross binding activity of some glycoprotein

preparations to antibodies made against host cell antigens, the inventors discovered that the cross binding activity was a consequence of residual β-mannosylation persisting in some strains of recombinant *P. pastoris* host cells in which the *BMT2* gene had been disrupted or deleted. Thus, heterologous glycoproteins produced in these recombinant host cells have *N*-glycans that still contained β-mannose residues. These β-mannose residues were detectable in ELISAs and Western blots of the heterologous glycoproteins obtained from cultures of these recombinant host cells probed with antibodies made against host cell antigens (HCA). Anti-HCA antibodies are polyclonal antibodies raised against a wild-type *Pichia pastoris* strain or a NORF strain: a recombinant host cell that is constructed in the same manner as the recombinant host cell that produces the heterologous glycoprotein except that the open reading frame (ORF) encoding the heterologous protein has been omitted. For therapeutic glycoproteins produced in *Pichia pastoris*, these residual β-mannose residues present the risk of eliciting an immune response in some individuals that receive the therapeutic protein in a treatment for a disease or disorder. The present invention provides a method for producing glycoproteins in *Pichia pastoris* that do not contain any detectable β-mannosylation and as such do not cross bind to antibodies made against host cell antigens.

BMT1, *BMT2*, and *BMT3* demonstrate a high degree of sequence homology while *BMT4* is homologous to a lower extent and is thought to be a capping alpha-mannosyltransferase. However, all four members of the *BMT* family appear to be involved in synthesis of *N*- and/or *O*-glycans having β-linked mannose structures. Although a MALDI-TOF of *N*-glycans from a test protein produced in a *Pichia pastoris* strain in which the *BMT2* gene has been deleted might fail to detect β-mannosylation, the sensitive antibody-based assays herein were able to detect β-mannosylation in *Δbmt2* strains. Thus, the anti-HCA antibody-based detection methods taught herein showed that deletion or disruption of also the *BMT1* and *BMT3* genes and optionally the *BMT4* gene was needed to remove all detectable β-mannose structures. Deleting or disrupting the genes encoding the three β-mannosyltransferases can be achieved by (1) complete or partial knock-out of the gene (including the promoter sequences, open reading frame (ORF) and/or the transcription terminator sequences); (2) introduction of a frame-shift in the ORF; (3) inactivation or regulation of the promoter; (4) knock-down of message by siRNA or antisense RNA; (5) or the use of chemical inhibitors. The result is the production of a host cell that is capable of producing a glycoprotein that lacks detectable cross binding activity to anti-HCA antibodies.

To exemplify the methods for producing a glycoprotein that lacks detectable cross binding activity to anti-HCA antibodies, a strain of *Pichia pastoris*, which had been genetically engineered to lack *BMT2* expression or activity and to be capable of producing recombinant mature human erythropoietin (EPO) with sialic acid-terminated bi-antennary *N*-glycans, was further genetically engineered to lack expression of the *BMT1* and/or *BMT3* and/or *BMT4* genes. The strain in which only expression of the *BMT2* gene had been disrupted produced recombinant

mature human EPO having some detectable cross binding activity to anti-HCA antibodies. The detectable cross binding activity was found to be due to the presence of β-linked mannose residues on the EPO molecule (*See* **Figures 22-27B**, Example 6). When the genes encoding *BMT1*and *BMT4* were disrupted or deleted in the strain, the EPO produced still had detectable cross binding activity to anti-HCA antibodies (*See* **Figures 28-31**). However, when the *BMT1*, *BMT2*, *BMT3*, and *BMT4* genes were disrupted or deleted, most of the strains produced glycosylated recombinant human EPO that lacked detectable cross binding activity to anti-HCA antibodies and thus lacked detectable β-mannose residues (*See* **Figures 33** and **35B** for example).

Thus, the present invention further provides a method for producing a recombinant protein or glycoprotein that lacks detectable cross binding activity to antibodies made against host cell antigens that involves constructing host cells intended to be used to produce the recombinant protein to further not display various combinations β-mannosyltransferase activities. By way of example, a host cell is constructed that does not display β-mannosylttransferase 2 activity with respect to an *N*-glycan or *O*-glycan. The host cell lacking display β-mannosyltransferase 2 activity is used to produce the recombinant protein or glycoprotein, which is then evaluated by Western blot or ELISA using an antibody that has been made against a NORF version of the strain. A NORF strain is a strain the same as the host strain except it lacks the open reading frame encoding the recombinant glycoprotein. If the recombinant protein or glycoprotein produced by the host cell lacks detectable binding to the antibody made against host cell antigens, then the host cell is useful for producing the recombinant protein or glycoprotein that lacks cross binding activity to the antibodies against host cell antigens.

However, if detectable cross binding activity is detected, then the host cell is further manipulated to not display β-mannosyltransferase 1, β-mannosyltransferase 3, or β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan. For example, the host cell that lacks β-mannosyltransferase 2 activity is further manipulated to lack β-mannosyltransferase 1 actvity. The host cell is used to produce the recombinant protein or glycoprotein, which is then evaluated by Western blot or ELISA using an antibody that has been made against a NORF version of the strain. If the recombinant protein or glycoprotein produced by the host cell lacks detectable binding to the antibody made against host cell antigens, then the host cell is useful for producing the recombinant protein or glycoprotein that lacks cross binding activity to the antibodies against host cell antigens.

However, if detectable cross binding activity is detected, then the host cell is further manipulated to not display β-mannosyltransferase 3 activity or β-mannosyltransferase 4 activity. For example, the host cell that lacks β-mannosyltransferase 2 activity and β-mannosyltransferase 1 activity is further manipulated to lack β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan. The host cell is used to produce the protein or

recombinant glycoprotein, which is then evaluated by Western blot or ELISA using an antibody that has been made against a NORF version of the strain. If the recombinant protein or glycoprotein produced by the host cell lacks detectable binding to the antibody made against host cell antigens, then the host cell is useful for producing the recombinant protein or glycoprotein that lacks cross binding activity to the antibodies against host cell antigens.

However, if detectable cross binding activity is detected, then the strain is further manipulated to not display β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan. The host cell is used to produce the recombinant protein or glycoprotein, which is then evaluated by Western blot or ELISA using an antibody that has been made against a NORF version of the strain to confirm that the recombinant protein or glycoprotein lacks detectable binding to the antibody made against host cell antigens.

By way of a further example, a *Pichia pastoris* host cell is constructed in which various combinations of *BMT* genes are deleted or disrupted in. By way of example, a *Pichia pastoris* host cell is constructed that has a disruption or deletion of the *BMT2* gene. The *Δbmt2* host cell is used to produce the recombinant protein or glycoprotein, which is then evaluated by Western blot or ELISA using an antibody that has been made against a NORF version of the strain. A NORF strain is a strain the same as the host strain except it lacks the open reading frame encoding the recombinant glycoprotein. If the recombinant protein or glycoprotein produced by the *Δbmt2* host cell lacks detectable binding to the antibody made against host cell antigens, then the *BMT2* deletion or disruption is sufficient to enable the host cell to produce the recombinant protein or glycoprotein that lacks cross binding activity to the antibodies against host cell antigens.

However, if detectable cross binding activity is detected, then the host cell is further manipulated to have a deletion of the *BMT1*, *BMT3*, or *BMT4* genes. For example, the host cell that has a disruption or deletion of the *BMT2* gene is further manipulated to have a deletion or disruption of the *BMT1* gene. The *Δbmt2 Δbmt1* host cell is used to produce the recombinant protein or glycoprotein, which is then evaluated by Western blot or ELISA using an antibody that has been made against a NORF version of the strain. If the recombinant protein or glycoprotein produced by the *Δbmt2 Δbmt1* host cell lacks detectable binding to the antibody made against host cell antigens, then the *BMT1* and *BMT2* deletions or disruptions are sufficient to enable the host cell to produce the recombinant protein or glycoprotein that lacks cross binding activity to the antibodies against host cell antigens.

However, if detectable cross binding activity is detected, then the host cell is further manipulated to have a deletion of the *BMT3* or *BMT4* genes. For example, the host cell that has a disruption or deletion of the *BMT1* and *BMT2* gene is further manipulated to have a deletion or disruption of the *BMT3* gene. The *Δbmt2 Δbmt1 Δbmt3* host cell is used to produce the protein or recombinant glycoprotein, which is then evaluated by Western blot or ELISA using

an antibody that has been made against a NORF version of the host cell. If the recombinant protein or glycoprotein produced by the *Δbmt2 Δbmt1 Δbmt3* host cell lacks detectable binding to the antibody made against host cell antigens, then the *BMT1*, *BMT2*, and *BMT3* deletions or disruptions are sufficient to enable the host cell to produce the recombinant protein or glycoprotein that lacks cross binding activity to the antibodies against host cell antigens.

However, if detectable cross binding activity is detected, then the host cell is further manipulated to have a deletion of the *BMT4* gene. The *Δbmt2 Δbmt1 Δbmt3 Δbmt4* host cell is used to produce the recombinant protein or glycoprotein, which is then evaluated by Western blot or ELISA using an antibody that has been made against a NORF version of the strain to confirm that the recombinant protein or glycoprotein lacks detectable binding to the antibody made against host cell antigens.

The present invention further provides a recombinant methylotrophic yeast host cells such as *Pichia pastoris* host cell in which the host cell does not display a β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan and does not display at least one of a β-mannosyltransferase 1 activity or a β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein. In further embodiments, the host cell does not display β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan. In a further aspect, the present invention provides a recombinant host cell that does not display a β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, β-mannosyltransferase 3 activity, and β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein.

The present invention further provides a general method for producing a recombinant protein or glycoprotein that lacks detectable cross binding activity to anti-host cell antigen antibodies comprising providing a recombinant methylotrophic yeast such as *Pichia pastoris* host cell does not display a β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan and does not display at least one activity selected from β-mannosyltransferase 1 activity and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein; growing the host cell in a medium under conditions effective for expressing the recombinant protein or glycoprotein; and recovering the recombinant protein or glycoprotein from the medium to produce the recombinant protein or glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens. In further embodiments, the host cell lacks β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan.

In a further aspect, the present invention provides a general method for producing a recombinant protein or glycoprotein that lacks detectable cross binding activity to anti-host cell antigen antibodies comprising providing a recombinant methylotrophic yeast such as *Pichia pastoris* host cell that does not display β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, β-mannosyltransferase 3 activity, and β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein; growing the host cell in a medium under conditions effective for expressing the recombinant protein or glycoprotein; and recovering the recombinant protein or glycoprotein from the medium to produce the recombinant protein or glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens.

The present invention further provides a recombinant methylotrophic yeast host cells such as *Pichia pastoris* host cell in which the gene encoding a β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan has been deleted or disrupted and at least one gene encoding a β-mannosyltransferase 1 activity or β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan has been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein. In further embodiments, the genes encoding a β-mannosyltransferase 2 activity, a β-mannosyltransferase 1 activity, and a β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan have been deleted or disrupted. In a further aspect, the present invention provides a recombinant host cell the genes encoding a β-mannosyltransferase 2 activity, a β-mannosyltransferase 1 activity, a β-mannosyltransferase 3 activity, and β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan have been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein.

The present invention further provides a general method for producing a recombinant protein or glycoprotein that lacks detectable cross binding activity to anti-host cell antigen antibodies comprising providing a recombinant methylotrophic yeast such as *Pichia pastoris* host cell in which the gene encoding a β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan has been deleted or disrupted and at least one gene encoding an activity selected from β-mannosyltransferase 1 activity and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan has been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein; growing the host cell in a medium under conditions effective for expressing the recombinant protein or glycoprotein; and recovering the recombinant protein or glycoprotein from the medium to produce the recombinant protein or glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens. In further embodiments, the genes encoding a β-mannosyltransferase 2 activity, a β-mannosyltransferase 1 activity, and a β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan have been deleted or disrupted.

In a further aspect, the present invention provides a general method for producing a recombinant protein or glycoprotein that lacks detectable cross binding activity to anti-host cell antigen antibodies comprising providing a recombinant methylotrophic yeast such as *Pichia pastoris* host cell in which the genes encoding a β-mannosyltransferase 2 activity, a β-mannosyltransferase 1 activity,a β-mannosyltransferase 3 activity, and a β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan have been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein; growing the host cell in a medium under conditions effective for expressing the recombinant protein or glycoprotein; and recovering the recombinant protein or glycoprotein from the medium to produce the recombinant protein or glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens.

The present invention further provides a recombinant *Pichia pastoris* host cell in which the *BMT2* gene and at least one of *BMT1* gene and *BMT3* gene have been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein. In further embodiments, the *BMT2* gene, *BMT1* gene, and *BMT3* gene have been deleted or disrupted. In a further aspect, the present invention provides a recombinant *Pichia pastoris* host cell in which the *BMT1* gene, *BMT2* gene, *BMT3* gene, and *BMT4* gene have been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein.

The present invention further provides a general method for producing a recombinant protein or glycoprotein that lacks detectable cross binding activity to anti-host cell antigen antibodies comprising providing a recombinant *Pichia pastoris* host cell in which the *BMT2* gene and at least one of the *BMT1* gene and the *BMT3* gene have been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein; growing the host cell in a medium under conditions effective for expressing the recombinant protein or glycoprotein; and recovering the recombinant protein or glycoprotein from the medium to produce the recombinant protein or glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens. In further embodiments, the *BMT2* gene, *BMT1* gene, and *BMT3* gene have been deleted or disrupted.

In a further aspect, the present invention provides a general method for producing a recombinant protein or glycoprotein that lack detectable cross binding activity to anti-host cell antigen antibodies comprising providing a recombinant *Pichia pastoris* host cell in which the *BMT1* gene, *BMT2* gene, *BMT3* gene, and *BMT4* gene have been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein; growing the host cell in a medium under conditions effective for expressing the recombinant protein or glycoprotein; and recovering the recombinant protein or glycoprotein from the medium to

produce the recombinant protein or glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens.

The present invention further provides a recombinant *Pichia pastoris* host cell in which the *BMT2* gene and at least one of the *BMT1* gene and the *BMT3* gene have been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein. In further embodiments, the *BMT2* gene, *BMT1* gene, and *BMT3* gene have been deleted or disrupted. In a further aspect, the present invention provides a recombinant *Pichia pastoris* host cell in which the *BMT1* gene, *BMT2* gene, *BMT3* gene, and *BMT4* gene have been deleted or disrupted and which includes a nucleic acid molecule encoding the recombinant protein or glycoprotein.

In general, the recombinant protein or glycoprotein is a therapeutic glycoprotein. Examples of therapeutic glycoproteins contemplated, include but are not limited to erythropoietin (EPO); cytokines such as interferon α, interferon β, interferon γ, and interferon ω; and granulocyte-colony stimulating factor (GCSF); GM-CSF; coagulation factors such as factor VIII, factor IX, and human protein C; antithrombin III; thrombin,; soluble IgE receptor α-chain; immunoglobulins such as IgG, IgG fragments, IgG fusions, and IgM; immunoadhesions and other Fc fusion proteins such as soluble TNF receptor-Fc fusion proteins; RAGE-Fc fusion proteins; interleukins; urokinase; chymase; and urea trypsin inhibitor; IGF-binding protein; epidermal growth factor; growth hormone-releasing factor; annexin V fusion protein; angiostatin; vascular endothelial growth factor-2; myeloid progenitor inhibitory factor-1; osteoprotegerin; α-1-antitrypsin; α-feto proteins; DNase II; kringle 3 of human plasminogen; glucocerebrosidase; TNF binding protein 1; follicle stimulating hormone; cytotoxic T lymphocyte associated antigen 4 – Ig; transmembrane activator and calcium modulator and cyclophilin ligand; glucagon like protein 1; and IL-2 receptor agonist.

In particular aspects of the invention, the nucleic acid molecule encoding the recombinant protein or glycoprotein is codon-optimized to enhance expression of the recombinant protein or glycoprotein in the host cell. For example, as shown in the examples, the nucleic acid molecule encoding the human mature form of erythropoietin was codon-optimized for enhanced expression of the erythropoietin in a methylotrophic yeast such as *Pichia pastoris* strain that had been genetically engineered to produce an erythropoietin variant comprising bi-antennary *N*-glycans in which the predominant glycoform comprised both antennae terminally sialylated.

The present invention further provides compositions comprising one or more proteins or glycoproteins lacking detectable cross-binding to antibodies against host cell antigens produced using the methods herein and in the host cells described herein. The compositions can further include pharmaceutically acceptable carriers and salts.

Suitable host cells include any host cell that includes homologues of the *Pichia pastoris BMT1*, *BMT2*, *BMT3*, and/or *BMT4* genes. Currently, examples of such host cells include *Candida albicans* and the methylotrophic yeast *Pichia pastoris*. Thus, in particular aspects of the invention, the host cell is a methylotrophic yeast such as *Pichia pastoris* and mutants thereof and genetically engineered variants thereof. Methylotrophic yeast such as *Pichia pastoris* that are contemplated for use in the present invention can be genetically modified so that they express glycoproteins in which the glycosylation pattern is human-like or humanized. In this manner, glycoprotein compositions can be produced in which a specific desired glycoform is predominant in the composition. Such can be achieved by eliminating selected endogenous glycosylation enzymes and/or genetically engineering the host cells and/or supplying exogenous enzymes to mimic all or part of the mammalian glycosylation pathway as described in US 2004/0018590. If desired, additional genetic engineering of the glycosylation can be performed, such that the glycoprotein can be produced with or without core fucosylation. Use of lower eukaryotic host cells is further advantageous in that these cells are able to produce highly homogenous compositions of glycoprotein, such that the predominant glycoform of the glycoprotein may be present as greater than thirty mole percent of the glycoprotein in the composition. In particular aspects, the predominant glycoform may be present in greater than forty mole percent, fifty mole percent, sixty mole percent, seventy mole percent and, most preferably, greater than eighty mole percent of the glycoprotein present in the composition. Such can be achieved by eliminating selected endogenous glycosylation enzymes and/or supplying exogenous enzymes as described by Gerngross *et al.*, U.S. Patent No. 7,029,872 and U.S. Patent No. 7,449,308. For example, a host cell can be selected or engineered to be depleted in 1,6-mannosyl transferase activities, which would otherwise add mannose residues onto the *N*-glycan on a glycoprotein.

In one embodiment, the host cell further includes an α1,2-mannosidase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target the α1,2-mannosidase activity to the ER or Golgi apparatus of the host cell. Passage of a recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $Man_5GlcNAc_2$ glycoform, for example, a recombinant glycoprotein composition comprising predominantly a $Man_5GlcNAc_2$ glycoform. For example, U.S. Patent No. 7,029,872, U.S. Patent No. 7,449,308, and U.S. Published Patent Application No. 2005/0170452 disclose lower eukaryote host cells capable of producing a glycoprotein comprising a $Man_5GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes an *N*-acetylglucosaminyltransferase I (GlcNAc transferase I or GnT I) catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target GlcNAc transferase I activity to the ER or Golgi apparatus of the host cell. Passage of the

recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GlcNAcMan_5GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a $GlcNAcMan_5GlcNAc_2$ glycoform. U.S. Patent No, 7,029,872, U.S. Patent No. 7,449,308, and U.S. Published Patent Application No. 2005/0170452 disclose lower eukaryote host cells capable of producing a glycoprotein comprising a $GlcNAcMan_5GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexaminidase to produce a recombinant glycoprotein comprising a $Man_5GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes a mannosidase II catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target mannosidase II activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GlcNAcMan_3GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a $GlcNAcMan_3GlcNAc_2$ glycoform. U.S. Patent No, 7,029,872 and U.S. Published Patent Application No. 2004/0230042 discloses lower eukaryote host cells that express mannosidase II enzymes and are capable of producing glycoproteins having predominantly a $GlcNAc_2Man_3GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexaminidase to produce a recombinant glycoprotein comprising a $Man_3GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes *N*-acetylglucosaminyltransferase II (GlcNAc transferase II or GnT II) catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target GlcNAc transferase II activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GlcNAc_2Man_3GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a $GlcNAc_2Man_3GlcNAc_2$ glycoform. U.S. Patent No, 7,029,872 and U.S. Published Patent Application Nos. 2004/0018590 and 2005/0170452 disclose lower eukaryote host cells capable of producing a glycoprotein comprising a $GlcNAc_2Man_3GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexaminidase to produce a recombinant glycoprotein comprising a $Man_3GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes a galactosyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target galactosyltransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a

$GalGlcNAc_2Man_3GlcNAc_2$ or $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform, or mixture thereof for example a recombinant glycoprotein composition comprising predominantly a $GalGlcNAc_2Man_3GlcNAc_2$ glycoform or $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or mixture thereof. U.S. Patent No, 7,029,872 and U.S. Published Patent Application No. 2006/0040353 discloses lower eukaryote host cells capable of producing a glycoprotein comprising a $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a galactosidase to produce a recombinant glycoprotein comprising a $GlcNAc_2Man_3GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a $GlcNAc_2Man_3GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes a sialyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target sialytransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising predominantly a $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or $NANAGal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or mixture thereof. For lower eukaryote host cells such as yeast and filamentous fungi, it is useful that the host cell further include a means for providing CMP-sialic acid for transfer to the *N*-glycan. U.S. Published Patent Application No. 2005/0260729 discloses a method for genetically engineering lower eukaryotes to have a CMP-sialic acid synthesis pathway and U.S. Published Patent Application No. 2006/0286637 discloses a method for genetically engineering lower eukaryotes to produce sialylated glycoproteins. The glycoprotein produced in the above cells can be treated *in vitro* with a neuraminidase to produce a recombinant glycoprotein comprising predominantly a $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or $GalGlcNAc_2Man_3GlcNAc_2$ glycoform or mixture thereof.

Any one of the preceding host cells can further include one or more GlcNAc transferase selected from the group consisting of GnT III, GnT IV, GnT V, GnT VI, and GnT IX to produce glycoproteins having bisected (GnT III) and/or multiantennary (GnT IV, V, VI, and IX) *N*-glycan structures such as disclosed in U.S. Published Patent Application Nos. 2004/074458 and 2007/0037248.

In further embodiments, the host cell that produces glycoproteins that have predominantly $GlcNAcMan_5GlcNAc_2$ *N*-glycans further includes a galactosyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target Galactosyltransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising predominantly the $GalGlcNAcMan_5GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell that produced glycoproteins that have predominantly the $GalGlcNAcMan_5GlcNAc_2$ *N*-glycans further includes a sialyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target sialytransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $NANAGalGlcNAcMan_5GlcNAc_2$ glycoform.

In further aspects, any one of the aforementioned host cells, the host cell is further modified to include a fucosyltransferase and a pathway for producing fucose and transporting fucose into the ER or Golgi. Examples of methods for modifying *Pichia pastoris* to render it capable of producing glycoproteins in which one or more of the *N*-glycans thereon are fucosylated are disclosed in PCT International Application No. PCT/US2008/002787. In particular aspects of the invention, the *Pichia pastoris* host cell is further modified to include a fucosylation pathway comprising a GDP-mannose-4,6-dehydratase, GDP-keto-deoxy-mannose-epimerase/GDP-keto-deoxy-galactose-reductase, GDP-fucose transporter, and a fucosyltransferase. In particular aspects, the fucosyltransferase is selected from the group consisting of fucosyltransferase is selected from the group consisting of α1,2-fucosyltransferase, α1,3-fucosyltransferase, α1,4-fucosyltransferase, and α1,6-fucosyltransferase.

Various of the preceding host cells further include one or more sugar transporters such as UDP-GlcNAc transporters (for example, *Kluyveromyces lactis* and *Mus musculus* UDP-GlcNAc transporters), UDP-galactose transporters (for example, *Drosophila melanogaster* UDP-galactose transporter), and CMP-sialic acid transporter (for example, human sialic acid transporter). Because lower eukaryote host cells such as yeast and filamentous fungi lack the above transporters, it is preferable that lower eukaryote host cells such as yeast and filamentous fungi be genetically engineered to include the above transporters.

Host cells further include *Pichia pastoris* that are genetically engineered to eliminate glycoproteins having phosphomannose residues by deleting or disrupting one or both of the phosphomannosyl transferase genes *PNO1* and *MNN4B* (*See* for example, U.S. Patent Nos. 7,198,921 and 7,259,007), which in further aspects can also include deleting or disrupting the *MNN4A* gene. Disruption includes disrupting the open reading frame encoding the particular enzymes or disrupting expression of the open reading frame or abrogating translation of RNAs encoding one or more of the β-mannosyltransferases and/or phosphomannosyltransferases using interfering RNA, antisense RNA, or the like. The host cells can further include any one of the aforementioned host cells modified to produce particular *N*-glycan structures.

Host cells further include lower eukaryote cells (e.g., yeast such as *Pichia pastoris*) that are genetically modified to control *O*-glycosylation of the glycoprotein by deleting or disrupting one or more of the protein *O*-mannosyltransferase (Dol-P-Man:Protein (Ser/Thr)

Mannosyl Transferase genes) (*PMT*s) (*See* U.S. Patent No. 5,714,377) or grown in the presence of Pmtp inhibitors and/or an alpha-mannosidase as disclosed in Published International Application No. WO 2007061631, or both. Disruption includes disrupting the open reading frame encoding the Pmtp or disrupting expression of the open reading frame or abrogating translation of RNAs encoding one or more of the Pmtps using interfering RNA, antisense RNA, or the like. The host cells can further include any one of the aforementioned host cells modified to produce particular *N*-glycan structures..

Pmtp inhibitors include but are not limited to a benzylidene thiazolidinediones. Examples of benzylidene thiazolidinediones that can be used are 5-[[3,4-bis(phenylmethoxy) phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic Acid; 5-[[3-(1-Phenylethoxy)-4-(2-phenylethoxy)]phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic Acid; and 5-[[3-(1-Phenyl-2-hydroxy)ethoxy)-4-(2-phenylethoxy)]phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic Acid.

In particular embodiments, the function or expression of at least one endogenous *PMT* gene is reduced, disrupted, or deleted. For example, in particular embodiments the function or expression of at least one endogenous *PMT* gene selected from the group consisting of the *PMT1*, *PMT2*, *PMT3*, and *PMT4* genes is reduced, disrupted, or deleted; or the host cells are cultivated in the presence of one or more PMT inhibitors. In further embodiments, the host cells include one or more *PMT* gene deletions or disruptions and the host cells are cultivated in the presence of one or more Pmtp inhibitors. In particular aspects of these embodiments, the host cells also express a secreted alpha-1,2-mannosidase.

PMT deletions or disruptions and/or Pmtp inhibitors control *O*-glycosylation by reducing *O*-glycosylation occupancy; that is by reducing the total number of *O*-glycosylation sites on the glycoprotein that are glycosylated. The further addition of an alpha-1,2-mannsodase that is secreted by the cell controls *O*-glycosylation by reducing the mannose chain length of the *O*-glycans that are on the glycoprotein. Thus, combining *PMT* deletions or disruptions and/or Pmtp inhibitors with expression of a secreted alpha-1,2-mannosidase controls *O*-glycosylation by reducing occupancy and chain length. In particular circumstances, the particular combination of *PMT* deletions or disruptions, Pmtp inhibitors, and alpha-1,2-mannosidase is determined empirically as particular heterologous glycoproteins (antibodies, for example) may be expressed and transported through the Golgi apparatus with different degrees of efficiency and thus may require a particular combination of *PMT* deletions or disruptions, Pmtp inhibitors, and alpha-1,2-mannosidase. In another aspect, genes encoding one or more endogenous mannosyltransferase enzymes are deleted. This deletion(s) can be in combination with providing the secreted alpha-1,2-mannosidase and/or *PMT* inhibitors or can be in lieu of providing the secreted alpha-1,2-mannosidase and/or *PMT* inhibitors.

Thus, the control of *O*-glycosylation can be useful for producing particular glycoproteins in the host cells disclosed herein in better total yield or in yield of properly assembled glycoprotein. The reduction or elimination of *O*-glycosylation appears to have a beneficial effect on the assembly and transport of glycoproteins such as whole antibodies as they traverse the secretory pathway and are transported to the cell surface. Thus, in cells in which *O*-glycosylation is controlled, the yield of properly assembled glycoproteins such as antibody fragments is increased over the yield obtained in host cells in which *O*-glycosylation is not controlled.

Yield of glycoprotein can in some situations be improved by overexpressing nucleic acid molecules encoding mammalian or human chaperone proteins or replacing the genes encoding one or more endogenous chaperone proteins with nucleic acid molecules encoding one or more mammalian or human chaperone proteins. In addition, the expression of mammalian or human chaperone proteins in the host cell also appears to control *O*-glycosylation in the cell. Thus, further included are the host cells herein wherein the function of at least one endogenous gene encoding a chaperone protein has been reduced or eliminated, and a vector encoding at least one mammalian or human homolog of the chaperone protein is expressed in the host cell. Also included are host cells in which the endogenous host cell chaperones and the mammalian or human chaperone proteins are expressed. In further aspects, the lower eukaryotic host cell is a yeast or filamentous fungi host cell. Examples of the use of chaperones of host cells in which human chaperone proteins are introduced to improve the yield and reduce or control *O*-glycosylation of recombinant proteins has been disclosed in PCT International Application No. PCT/US2009/033507. Like above, further included are lower eukaryotic host cells wherein, in addition to replacing the genes encoding one or more of the endogenous chaperone proteins with nucleic acid molecules encoding one or more mammalian or human chaperone proteins or overexpressing one or more mammalian or human chaperone proteins as described above, the function or expression of at least one endogenous gene encoding a protein *O*-mannosyltransferase (PMT) protein is reduced, disrupted, or deleted. In particular embodiments, the function of at least one endogenous *PMT* gene selected from the group consisting of the *PMT1*, *PMT2*, *PMT3*, and *PMT4* genes is reduced, disrupted, or deleted.

Therefore, the methods disclose herein can use any host cell that has been genetically modified to produce glycoproteins wherein the predominant *N*-glycan is selected from the group consisting of complex *N*-glycans, hybrid *N*-glycans, and high mannose *N*-glycans wherein complex *N*-glycans are selected from the group consisting of $Man_3GlcNAc_2$, $GlcNAC_{(1-4)}Man_3GlcNAc_2$, $Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$, and $NANA_{(1-4)}Gal_{(1-4)}Man_3GlcNAc_2$; hybrid *N*-glycans are selected from the group consisting of $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GalGlcNAcMan_5GlcNAc_2$, and $NANAGalGlcNAcMan_5GlcNAc_2$; and high mannose *N*-glycans are selected from the group consisting of $Man_6GlcNAc_2$,

$Man_7GlcNAc_2$, $Man_8GlcNAc_2$, and $Man_9GlcNAc_2$. Examples of *N*-glycan structures include but are not limited to $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$, $GlcNAc_2Man_3GlcNAc_2$, $GlcNAc_3Man_3GlcNAc_2$, $GlcNAc_4Man_3GlcNAc_2$, $GalGlcNAc_2Man_3GlcNAc_2$, $Gal_2GlcNAc_2Man_3GlcNAc_2$, $Gal_2GlcNAc_3Man_3GlcNAc_2$, $Gal_2GlcNAc_4Man_3GlcNAc_2$, $Gal_3GlcNAc_3Man_3GlcNAc_2$, $Gal_3GlcNAc_4Man_3GlcNAc_2$, $Gal_4GlcNAc_4Man_3GlcNAc_2$, $NANAGal_2GlcNAc_2Man_3GlcNAc_2$, $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$, $NANA_3Gal_3GlcNAc_3Man_3GlcNAc_2$, and $NANA_4Gal_4GlcNAc_4Man_3GlcNAc_2$.

Yeast selectable markers that can be used to construct the recombinant host cells include drug resistance markers and genetic functions which allow the yeast host cell to synthesize essential cellular nutrients, e.g. amino acids. Drug resistance markers which are commonly used in yeast include chloramphenicol, kanamycin, methotrexate, G418 (geneticin), Zeocin, and the like. Genetic functions which allow the yeast host cell to synthesize essential cellular nutrients are used with available yeast strains having auxotrophic mutations in the corresponding genomic function. Common yeast selectable markers provide genetic functions for synthesizing leucine (*LEU2*), tryptophan (*TRP1* and *TRP2*), proline (PRO1), uracil (*URA3*, *URA5, URA6*), histidine (*HIS3*), lysine (*LYS2*), adenine (*ADE1* or *ADE2*), and the like. Other yeast selectable markers include the *ARR3* gene from *S. cerevisiae*, which confers arsenite resistance to yeast cells that are grown in the presence of arsenite (Bobrowicz *et al.*, Yeast, 13:819-828 (1997); Wysocki *et al.*, J. Biol. Chem. 272:30061-30066 (1997)). A number of suitable integration sites include those enumerated in U.S. Patent No. 7,479,389 and include homologs to loci known for *Saccharomyces cerevisiae* and other yeast or fungi. Methods for integrating vectors into yeast are well known (*See* for example, U.S. Patent No. 7,479,389, U.S. Patent No. 7,514,253, U.S. Published Application No. 2009012400, and WO2009/085135). Examples of insertion sites include, but are not limited to, *Pichia ADE* genes; *Pichia TRP* (including *TRP1* through *TRP2*) genes; *Pichia MCA* genes; *Pichia CYM* genes; *Pichia PEP* genes; *Pichia PRB* genes; and *Pichia LEU* genes. The *Pichia ADE1* and *ARG4* genes have been described in Lin Cereghino *et al.*, Gene 263:159-169 (2001) and U.S. Patent No. 4,818,700, the *HIS3* and *TRP1* genes have been described in Cosano *et al.*, Yeast 14:861-867 (1998), *HIS4* has been described in GenBank Accession No. X56180.

The present invention further provides a method for producing a mature human erythropoietin in methylotrophic yeast such as *Pichia pastoris* comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens. The method comprises providing a recombinant *Pichia pastoris* host cell genetically engineered to produce sialic acid-terminated biantennary *N*-glycans and in which at least the *BMT1*, *BMT2*, and *BMT3* genes have been deleted or disrupted and which includes two or more nucleic acid molecules, each encoding a fusion protein

comprising a mature human erythropoietin EPO fused to a signal peptide that targets the ER or Golgi apparatus and which is removed when the fusion protein is in the ER or Golgi apparatus; growing the host cell in a medium under conditions effective for expressing and processing the first and second fusion proteins; and recovering the mature human erythropoietin from the medium to produce the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens.

In particular aspects, the nucleic acid molecule encoding the mature human erythropoietin is codon-optimized for optimal expression in the methylotrophic yeast such as *Pichia pastoris*. As shown in the examples, the mature human erythropoietin is encoded as a fusion protein in which the EPO is fused at the *N*-terminus of the mature form of the erythropoietin to the *C*-terminus of a signal peptide that targets the fusion protein to the secretory pathway for processing, including glycosylation. Examples of signal peptides include but are not limited to the *S. cerevisiae* αMATpre signal peptide or a chicken lysozyme signal peptide. Other signal sequences can be used instead of those disclosed herein, for example, the *Aspergillus niger* α-amylase signal peptide and human serum albumin (HSA) signal peptide. In one embodiment, a first nucleic acid molecule encodes a fusion protein wherein the mature erythropoietin is fused to the *S. cerevisiae* αMATpre signal peptide and second nucleic acid molecule encodes a fusion protein wherein the mature erythropoietin is fused to the *S. cerevisiae* αMATpre signal peptide a chicken lysozyme signal peptide. The signal peptide can be fused to the mature human erythropoietin by a linker peptide that can contain one or more protease cleavage sites.

In further aspects, the host cell includes between two and twelve copies of the expression cassettes encoding the fusion protein comprising the mature human erythropoietin. In some aspects, the host cell includes about eight to eleven copies of the expression cassettes encoding the fusion protein comprising the mature human erythropoietin. In other aspects, the host cell includes about three to four copies of the first nucleic acid and five to seven copies of the second nucleic acid.

The host cell is genetically engineered to produce sialic acid-terminated biantennary *N*-glycans and in which at least the *BMT1*, *BMT2*, and *BMT3* genes have been deleted or disrupted. Such a host cell further includes at least a deletion or disruption of the *OCH1*, *PNO1*, *MNN4*, and *MNN4L1* genes. The host cell further includes one or more nucleic acid molecules encoding at least the following chimeric glycosylation enzymes: α1,2-mannosidase catalytic domain fused to a cellular targeting peptide that targets the catalytic domain to the ER or Golgi apparatus of the host cell; GlcNAc transferase I catalytic domain fused to a cellular targeting peptide that targets the catalytic domain to the ER or Golgi apparatus of the host cell; mannosidase II catalytic domain fused to a cellular targeting peptide that targets

the catalytic domain to the ER or Golgi apparatus of the host cell; GlcNAc transferase II catalytic domain fused to a cellular targeting peptide that targets the catalytic domain to the ER or Golgi apparatus of the host cell; β1,4-galactosyltransferase catalytic domain fused to a cellular targeting peptide that targets the catalytic domain to the ER or Golgi apparatus of the host cell; and α1,2-sialyltransferase catalytic domain fused to a cellular targeting peptide that targets the catalytic domain to the ER or Golgi apparatus of the host cell. These glycosylation enzymes are selected to be active at the location in the ER or Golgi apparatus to which they are targeted. Methods for selecting glycosylation enzymes and targeting the enzymes to particular regions of the ER or Golgi apparatus for optimal activity have been described in U.S. Patent No. 7,029,872 and 7,449,308 and in Published U.S. Application Nos. 2006/0040353 and 2006/0286637. The host cells are further modified to include the enzymes of a pathway as disclosed in Published U.S. Application No. and 2006/0286637to produce CMP-sialic acid and to include GlcNAc and galactose transporters and a UDP-galactose-4-epimerase. Finally, the host further includes a nucleic acid molecule encoding a fungal α1,2-mannosidase catalytic domain fused to a cellular targeting peptide that targets the catalytic domain to the secretory pathway for secretion and which effects a reduction in *O*-glycan occupancy and chain length.

Detection of detectable cross binding activity with antibodies made against host cell antigens can be determined in a sandwich ELISA or in a Western blot.

In further aspects, recovering the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens includes a cation exchange chromatography step and/or a hydroxyapatite chromatography step and/or an anion exchange chromatography step. In one embodiment, the recovering the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens comprises a cation exchange chromatography step followed by a hydroxyapatite chromatography step. Optionally, recovery of the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens includes an anion chromatography step.

Further provided is a composition comprising a mature human erythropoietin comprising predominantly sialic acid-terminated bi-antennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens obtained as disclosed herein and a pharmaceutically acceptable salt. In particular embodiments, about 50 to 60% of the *N*-glycans comprise sialic acid residues on both antennae; in further embodiments, greater than 70% of the *N*-glycans comprise sialic acid residues on both antennae. In further aspects, less than 30% of the *N*-glycans are neutral *N*-glycans (i.e., are not sialylated on at least

one terminus at the non-reducing end of the *N*-glycan). In further still aspects, less than 20% of the *N*-glycans are neutral *N*-glycans.

In particular aspects, the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens is conjugated to a hydrophilic polymer, which is particular embodiments is a polyethylene glycol polymer. Examples of mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans conjugated to polyethylene glycol polymers has been described in commonly-owned U.S. Published Application No. 2008/0139470.

The polyethylene glycol polymer (PEG) group may be of any convenient molecular weight and may be linear or branched. The average molecular weight of the PEG will preferably range from about 2 kiloDalton ("kDa") to about 100 kDa, more preferably from about 5 kDa to about 60 kDa, more preferably from about 20 kDa to about 50 kDa; most preferably from about 30 kDa to about 40 kDa. These PEGs can be supplied from any commericial vendors including NOF Corporation (Tokyo, Japan), Dow Pharma (ChiroTech Technology, Cambridge,UK), Nektar (San Carlos, CA) and SunBio (Anyang City, South Korea). Suitable PEG moieties include, for example, 40 kDa methoxy poly(ethylene glycol) propionaldehyde; 60 kDa methoxy poly(ethylene glycol) propionaldehyde; 31 kDa alpha-methyl-w-(3-oxopropoxy), polyoxyethylene; 30 kDa PEG: 30kDa Methoxy poly(ethylene glycol) propionaldehyde and 45 kDa 2,3-Bis(methylpolyoxyethylene-oxy)-1-[(3-oxopropyl) polyoxyethylene-oxy]-propane. The PEG groups will generally be attached to the erythropoeitin via acylation or reductive amination through a reactive group on the PEG moiety (e.g., an aldehyde, amino, thiol, or ester group) to a reactive group on the protein or polypeptide of interest (e.g., an aldehyde, amino, or ester group). For example, the PEG moiety may be linked to the *N*-terminal amino acid residue of erythropoietin, either directly or through a linker.

A useful strategy for the PEGylation of synthetic peptides consists of combining, through forming a conjugate linkage in solution, a peptide and a PEG moiety, each bearing a special functionality that is mutually reactive toward the other. The peptides can be easily prepared with conventional solid phase synthesis (*See*, for example, Example 4). The peptides are "preactivated" with an appropriate functional group at a specific site. The precursors are purified and fully characterized prior to reacting with the PEG moiety. Ligation of the peptide with PEG usually takes place in aqueous phase and can be easily monitored by reverse phase analytical HPLC. The PEGylated peptides can be easily purified by preparative HPLC and characterized by analytical HPLC, amino acid analysis and laser desorption mass spectrometry.

The following examples are intended to promote a further understanding of the present invention.

EXAMPLE 1

Genetically engineered *Pichia pastoris* strain **YGLY3159** is a strain that produces recombinant human erythropoietin with sialylated *N*-glycans (rhEPO). Construction of the strain has been described in U.S. Published Application No. 20080139470 and is illustrated schematically in **Figure 1**. Briefly, the strain was constructed as follows.

The strain **YGLY3159** was constructed from wild-type *Pichia pastoris* strain **NRRL-Y 11430** using methods described earlier (*See* for example, U.S. Patent No. 7,449,308; U.S. Patent No. 7,479,389; U.S. Published Application No. 20090124000; Published PCT Application No. WO2009085135; Nett and Gerngross, Yeast 20:1279 (2003); Choi *et al.*, Proc. Natl. Acad. Sci. USA 100:5022 (2003); Hamilton *et al.*, Science 301:1244 (2003)). All plasmids were made in a pUC19 plasmid using standard molecular biology procedures. For nucleotide sequences that were optimized for expression in *P. pastoris*, the native nucleotide sequences were analyzed by the GENEOPTIMIZER software (GeneArt, Regensburg, Germany) and the results used to generate nucleotide sequences in which the codons were optimized for *P. pastoris* expression. Yeast strains were transformed by electroporation (using standard techniques as recommended by the manufacturer of the electroporator BioRad).

Plasmid pGLY6 (**Figure 2**) is an integration vector that targets the *URA5* locus contains a nucleic acid molecule comprising the *S. cerevisiae* invertase gene or transcription unit (*ScSUC2*; SEQ ID NO:1) flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *P. pastoris URA5* gene (SEQ ID NO:59) and on the other side by a nucleic acid molecule comprising the a nucleotide sequence from the 3' region of the *P.pastoris URA5* gene (SEQ ID NO:60). Plasmid pGLY6 was linearized and the linearized plasmid transformed into wild-type strain **NRRL-Y 11430** to produce a number of strains in which the *ScSUC2* gene was inserted into the *URA5* locus by double-crossover homologous recombination. Strain **YGLY1-3** was selected from the strains produced and is auxotrophic for uracil.

Plasmid pGLY40 (**Figure 3**) is an integration vector that targets the *OCH1* locus and contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (SEQ ID NO:61) flanked by nucleic acid molecules comprising *lacZ* repeats (SEQ ID NO:62) which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *OCH1* gene (SEQ ID NO:64) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *OCH1* gene (SEQ ID NO:65). Plasmid pGLY40 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY1-3** to produce to produce a number of strains in which the *URA5* gene flanked by the *lacZ* repeats has been inserted into the *OCH1* locus by double-crossover homologous recombination. Strain **YGLY2-3** was selected from the strains produced and is prototrophic for *URA5*. Strain **YGLY2-3** was counterselected in the presence of 5-fluoroorotic acid (5-FOA) to

produce a number of strains in which the *URA5* gene has been lost and only the *lacZ* repeats remain in the *OCH1* locus. This renders the strain auxotrophic for uracil. Strain **YGLY4-3** was selected.

Plasmid pGLY43a (**Figure 4**) is an integration vector that targets the *BMT2* locus and contains a nucleic acid molecule comprising the *K. lactis* UDP-*N*-acetylglucosamine (UDP-GlcNAc) transporter gene or transcription unit (*KlMNN2-2*, SEQ ID NO:3) adjacent to a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats. The adjacent genes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *BMT2* gene (SEQ ID NO: 66) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *BMT2* gene (SEQ ID NO:67). Plasmid pGLY43a was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY4-3** to produce to produce a number of strains in which the *KlMNN2-2* gene and *URA5* gene flanked by the *lacZ* repeats has been inserted into the *BMT2* locus by double-crossover homologous recombination. The *BMT2* gene has been disclosed in Mille *et al.*, J. Biol. Chem. 283: 9724-9736 (2008) and U.S. Patent No.7,465,557. Strain **YGLY6-3** was selected from the strains produced and is prototrophic for uracil. Strain **YGLY6-3** was counterselected in the presence of 5-FOA to produce strains in which the *URA5* gene has been lost and only the *lacZ* repeats remain. This renders the strain auxotrophic for uracil. Strain **YGLY8-3** was selected.

Plasmid pGLY48 (**Figure 5**) is an integration vector that targets the *MNN4L1* locus and contains an expression cassette comprising a nucleic acid molecule encoding the mouse homologue of the UDP-GlcNAc transporter (SEQ ID NO:17) open reading frame (ORF) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris GAPDH* promoter (SEQ ID NO:53) and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* termination sequences (SEQ ID NO:56) adjacent to a nucleic acid molecule comprising the *P. pastoris URA5* gene flanked by *lacZ* repeats and in which the expression cassettes together are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *P. Pastoris MNN4L1* gene (SEQ ID NO:76) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MNN4L1* gene (SEQ ID NO:77). Plasmid pGLY48 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY8-3** to produce a number of strains in which the expression cassette encoding the mouse UDP-GlcNAc transporter and the *URA5* gene have been inserted into the *MNN4L1* locus by double-crossover homologous recombination. The *MNN4L1* gene (also referred to as *MNN4B*) has been disclosed in U.S. Patent No. 7,259,007. Strain **YGLY10-3** was selected from the strains produced and then counterselected in the presence of 5-FOA to produce a number of strains in which the *URA5* gene has been lost and only the *lacZ* repeats remain. Strain **YGLY12-3** was selected.

Plasmid pGLY45 (**Figure 6**) is an integration vector that targets the *PNO1/MNN4* loci contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *PNO1* gene (SEQ ID NO:74) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MNN4* gene (SEQ ID NO:75). Plasmid pGLY45 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY12-3** to produce to produce a number of strains in which the *URA5* gene flanked by the *lacZ* repeats has been inserted into the *MNN4* loci by double-crossover homologous recombination. The *PNO1* gene has been disclosed in U.S. Patent No. 7,198,921 and the *MNN4* gene (also referred to as *MNN4B*) has been disclosed in U.S. Patent No. 7,259,007. Strain **YGLY14-3** was selected from the strains produced and then counterselected in the presence of 5-FOA to produce a number of strains in which the *URA5* gene has been lost and only the *lacZ* repeats remain. Strain **YGLY16-3** was selected.

Plasmid pGLY247 (**Figure 7**) is an integration vector that targets the *MET16* locus and contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *MET16* gene (SEQ ID NO:84) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MET16* gene (SEQ ID NO:85). Plasmid pGLY247 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY16-3** to produce a number of strains in which the *URA5* flanked by the *lacZ* repeats has been inserted into the *MET16* locus by double-crossover homologous recombination. Strain **YGLY20-3** was selected.

Plasmid pGLY248 (**Figure 8**) is an integration vector that targets the *URA5* locus and contains a nucleic acid molecule comprising the *P. pastoris MET16* gene (SEQ ID NO:86) flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *URA5* gene (SEQ ID NO:59) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *URA5* gene (SEQ ID NO:60). Plasmid pGLY248 was linearized and the linearized plasmid transformed into strain **YGLY20-3** to produce a number of strains in which the *ScSUC2* gene inserted into the *URA5* locus has been replaced with the *MET16* gene by double-crossover homologous recombination. Strain **YGLY22-3** was selected and then counterselected in the presence of 5-FOA to produce a number of strains in which the *URA5* gene inserted into the *MET16* locus has been lost and only the *lacZ* repeats remain. Strain **YGLY24-3** was selected.

Plasmid pGLY582 (**Figure 9**) is an integration vector that targets the *HIS1* locus and contains in tandem four expression cassettes encoding (1) the *S. cerevisiae* UDP-glucose

epimerase (*ScGAL10*), (2) the human galactosyltransferase I (*hGalT*) catalytic domain fused at the N-terminus to the *S. cerevisiae KRE2-s* leader peptide (33) to target the chimeric enzyme to the ER or Golgi, (3) the *P. pastoris URA5* gene or transcription unit flanked by lacZ repeats, and (4) the *D. melanogaster* UDP-galactose transporter (*DmUGT*). The expression cassette encoding the *ScGAL10* comprises a nucleic acid molecule encoding the *ScGAL10* ORF (SEQ ID NO:21) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris PMA1* promoter (SEQ ID NO:45) and operably linked at the 3' end to a nucleic acid molecule comprising the *P. pastoris PMA1* transcription termination sequence (SEQ ID NO:46). The expression cassette encoding the chimeric galactosyltransferase I comprises a nucleic acid molecule encoding the *hGalT* catalytic domain codon optimized for expression in *P. pastoris* (SEQ ID NO:23) fused at the 5' end to a nucleic acid molecule encoding the *KRE2-s* leader 33 (SEQ ID NO:13), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris GAPDH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* transcription termination sequence. The *URA5* expression cassette comprises a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats. The expression cassette encoding the *DmUGT* comprises a nucleic acid molecule encoding the *DmUGT* ORF (SEQ ID NO:19) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris OCH1* promoter (SEQ ID NO:47) and operably linked at the 3' end to a nucleic acid molecule comprising the *P. pastoris ALG12* transcription termination sequence (SEQ ID NO:48). The four tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *HIS1* gene (SEQ ID NO:87) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *HIS1* gene (SEQ ID NO:88). Plasmid pGLY582 was linearized and the linearized plasmid transformed into strain **YGLY24-3** to produce a number of strains in which the four tandem expression cassette have been inserted into the *HIS1* locus by homologous recombination. Strain **YGLY58** was selected and is auxotrophic for histidine and prototrophic for uridine.

Plasmid pGLY167b (**Figure 10**) is an integration vector that targets the *ARG1* locus and contains in tandem three expression cassettes encoding (1) the *D. melanogaster* mannosidase II catalytic domain (KD) fused at the *N*-terminus to *S. cerevisiae MNN2* leader peptide (53) to target the chimeric enzyme to the ER or Golgi, (2) the *P. pastoris HIS1* gene or transcription unit, and (3) the rat *N*-acetylglucosamine (GlcNAc) transferase II catalytic domain (TC) fused at the *N*-terminus to *S. cerevisiae MNN2* leader peptide (54) to target the chimeric enzyme to the ER or Golgi. The expression cassette encoding the KD53 comprises a nucleic acid molecule encoding the *D. melanogaster* mannosidase II catalytic domain codon-optimized for expression in *P. pastoris* (SEQ ID NO:33) fused at the 5' end to a nucleic acid molecule encoding the *MNN2* leader 53 (SEQ ID NO:5), which is operably linked at the 5' end to a nucleic acid

molecule comprising the *P. pastoris GAPDH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* transcription termination sequence. The *HIS1* expression cassette comprises a nucleic acid molecule comprising the *P. pastoris HIS1* gene or transcription unit (SEQ ID NO:89). The expression cassette encoding the TC54 comprises a nucleic acid molecule encoding the rat GlcNAc transferase II catalytic domain codon-optimized for expression in *P. pastoris* (SEQ ID NO:31) fused at the 5' end to a nucleic acid molecule encoding the *MNN2* leader 54 (SEQ ID NO:7), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris PMA1* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris PMA1* transcription termination sequence. The three tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *ARG1* gene (SEQ ID NO:79) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *ARG1* gene (SEQ ID NO:80). Plasmid pGLY167b was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY58** to produce a number of strains (in which the three tandem expression cassette have been inserted into the *ARG1* locus by double-crossover homologous recombination. The strain **YGLY73** was selected from the strains produced and is auxotrophic for arginine and prototrophic for uridine and histidine. The strain was then counterselected in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strain **YGLY1272** was selected.

Plasmid pGLY1430 (**Figure 11**) is a KINKO integration vector that targets the *ADE1* locus without disrupting expression of the locus and contains in tandem four expression cassettes encoding (1) the human GlcNAc transferase I catalytic domain (NA) fused at the *N*-terminus to *P. pastoris SEC12* leader peptide (10) to target the chimeric enzyme to the ER or Golgi, (2) mouse homologue of the UDP-GlcNAc transporter (MmTr), (3) the mouse mannosidase IA catalytic domain (FB) fused at the *N*-terminus to *S. cerevisiae SEC12* leader peptide (8) to target the chimeric enzyme to the ER or Golgi, and (4) the *P. pastoris URA5* gene or transcription unit. KINKO (Knock-In with little or No Knock-Out) integration vectors enable insertion of heterologous DNA into a targeted locus without disrupting expression of the gene at the targeted locus and have been described in U.S. Published Application No. 20090124000. The expression cassette encoding the NA10 comprises a nucleic acid molecule encoding the human GlcNAc transferase I catalytic domain codon-optimized for expression in *P. pastoris* (SEQ ID NO:25) fused at the 5' end to a nucleic acid molecule encoding the *SEC12* leader 10 (SEQ ID NO:11), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris PMA1* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris PMA1* transcription termination sequence. The expression cassette encoding MmTr comprises a nucleic acid molecule encoding the mouse homologue of the UDP-GlcNAc transporter ORF operably linked at the 5' end to a nucleic acid molecule comprising the *P.*

pastoris SEC4 promoter (SEQ ID NO:49) and at the 3' end to a nucleic acid molecule comprising the *P. pastoris OCH1* termination sequences (SEQ ID NO:50). The expression cassette encoding the FB8 comprises a nucleic acid molecule encoding the mouse mannosidase IA catalytic domain (SEQ ID NO:27) fused at the 5' end to a nucleic acid molecule encoding the *SEC12-m* leader 8 (SEQ ID NO:15), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris GADPH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* transcription termination sequence. The *URA5* expression cassette comprises a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats. The four tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and complete ORF of the *ADE1* gene (SEQ ID NO:82) followed by a *P. pastoris ALG3* termination sequence (SEQ ID NO:54) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *ADE1* gene (SEQ ID NO:83). Plasmid pGLY1430 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY1272** to produce a number of strains in which the four tandem expression cassette have been inserted into the *ADE1* locus immediately following the *ADE1* ORF by double-crossover homologous recombination. The strain **YGLY1305** was selected from the strains produced and is auxotrophic for arginine and now prototrophic for uridine, histidine, and adenine. The strain was then counterselected in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strain **YGLY1461** was selected and is capable of making glycoproteins that have predominantly galactose terminated *N*-glcyans.

Plasmid pGFI165 (**Figure 12**) is a KINKO integration vector that targets the *PRO1* locus without disrupting expression of the locus and contains expression cassettes encoding (1) the *T. reesei* α-1,2-mannosidase catalytic domain fused at the *N*-terminus to *S. cerevisiae* αMATpre signal peptide (aMATTrMan) to target the chimeric protein to the secretory pathway and secretion from the cell and (2) the *P. pastoris URA5* gene or transcription unit. The expression cassette encoding the aMATTrMan comprises a nucleic acid molecule encoding the *T. reesei* catalytic domain (SEQ ID NO:29) fused at the 5' end to a nucleic acid molecule encoding the *S. cerevisiae* αMATpre signal peptide (SEQ ID NO:9), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris GAPDH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* transcription termination sequence. The *URA5* expression cassette comprises a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats. The two tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and complete ORF of the *PRO1* gene (SEQ ID NO:90) followed by a *P. pastoris ALG3* termination sequence and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *PRO1* gene (SEQ ID

NO:91). Plasmid pGFI165 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY1461** to produce a number of strains in which the two expression cassette have been inserted into the *PRO1* locus immediately following the *PRO1* ORF by double-crossover homologous recombination. The strain **YGLY1703** was selected from the strains produced and is auxotrophic for arginine and prototrophic for uridine, histidine, adenine, and proline. This strain is capable of producing glycoproteins that have reduced *O*-glycosylation (*See* Published U.S. Application No. 20090170159).

Plasmid pGLY2088 (**Figure 13**) is an integration vector that targets the *TRP2* or AOX1 locus and contains expression cassettes encoding (1) mature human erythropoetin (EPO) fused at the *N*-terminus to a *S. cerevisiae* αMATpre signal peptide (alpha MF-pre) to target the chimeric protein to the secretory pathway and secretion from the cell and (2) the zeocin resistance protein (Sh ble or Zeocin^R). The expression cassette encoding the EPO comprises a nucleic acid molecule encoding the mature human EPO codon-optimized for expression in *P. pastoris* (SEQ ID NO:92) fused at the 5' end to a nucleic acid molecule encoding the *S. cerevisiae* αMATpre signal peptide, which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris AOX1* promoter (SEQ ID NO:55) and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* transcription termination sequence. The Zeocin^R expression cassette comprises a nucleic acid molecule encoding the Sh ble ORF (SEQ ID NO:58) operably linked at the 5' end to the *S. cerevisiae TEF1* promoter (SEQ ID NO:57) and at the 3' end to the *S. cerevisiae CYC* termination sequence. The two tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence comprising the *TRP2* gene (SEQ ID NO:78). Plasmid pGLY2088 was linearized at the *Pme*I site and transformed into strain **YGLY1703** to produce a number of strains in which the two expression cassette have been inserted into the *AOX1* locus by roll in single-crossover homologous recombination, which results in multiple copies of the EPO expression cassette inserted into the *AOX1* locus without disrupting the AOX1 locus. The strain **YGLY2849** was selected from the strains produced and is auxotrophic for arginine and now prototrophic for uridine, histidine, adenine, and proline. The strain contains about three to four copies of the EPO expression cassette as determined by measuring the intensity of sequencing data of DNA isolated from the strain. During processing of the chimeric EPO in the ER and Golgi, the leader peptide is removed. Thus, the rhEPO produced is the mature form of the EPO.

Plasmid pGLY2456 (**Figure 14**) is a KINKO integration vector that targets the *TRP2* locus without disrupting expression of the locus and contains six expression cassettes encoding (1) the mouse CMP-sialic acid transporter (mCMP-Sia Transp), (2) the human UDP-GlcNAc 2-epimerase/*N*-acetylmannosamine kinase (hGNE), (3) the *Pichia pastoris ARG1* gene or transcription unit, (4) the human CMP-sialic acid synthase (hCMP-NANA), (5) the human N-acetylneuraminate-9-phosphate synthase (hSIAP S), (6) the mouse a-2,6-sialyltransferase

catalytic domain (mST6) fused at the *N*-terminus to *S. cerevisiae KRE2* leader peptide (33) to target the chimeric enzyme to the ER or Golgi, and the *P. pastoris ARG1* gene or transcription unit. The expression cassette encoding the mouse CMP-sialic acid Transporter comprises a nucleic acid molecule encoding the mCMP Sia Transp ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:35), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris PMA1* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris PMA1* transcription termination sequence. The expression cassette encoding the human UDP-GlcNAc 2-epimerase/*N*-acetylmannosamine kinase comprises a nucleic acid molecule encoding the hGNE ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:37), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris GAPDH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* transcription termination sequence. The expression cassette encoding the *P. pastoris ARG1* gene comprises (SEQ ID NO:81). The expression cassette encoding the human CMP-sialic acid synthase comprises a nucleic acid molecule encoding the hCMP-NANA S ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:39), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris GPDAH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* transcription termination sequence. The expression cassette encoding the human N-acetylneuraminate-9-phosphate synthase comprises a nucleic acid molecule encoding the hSIAP S ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:41), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris PMA1* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris PMA1* transcription termination sequence. The expression cassette encoding the chimeric mouse a-2,6-sialyltransferase comprises a nucleic acid molecule encoding the mST6 catalytic domain codon optimized for expression in *P. pastoris* (SEQ ID NO:43) fused at the 5' end to a nucleic acid molecule encoding the *S. cerevisiae KRE2* signal peptide, which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris TEF* promoter (SEQ ID NO:51) and at the 3' end to a nucleic acid molecule comprising the *P. pastoris TEF* transcription termination sequence (SEQ ID NO:52). The six tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the ORF encoding Trp2p ending at the stop codon (SEQ ID NO:98) followed by a *P. pastoris ALG3* termination sequence and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *TRP2* gene (SEQ ID NO:99). Plasmid pGLY2456 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY2849** to produce a number of strains in which the six expression cassette have been inserted into the *TRP2* locus immediately following the *TRP2* ORF by double-crossover homologous recombination. The strain **YGLY3159** was selected from the strains produced and is now prototrophic for uridine, histidine, adenine, proline, arginine, and tryptophan. The strain is resistant to Zeocin and

contains about three to four copies of the EPO expression cassette. The strain produced rhEPO; however, using the methods in Example 5, the rhEPO has cross-reactivity binding to antibodies made against HCA (*See* Example 6).

While the various expression cassettes were integrated into particular loci of the *Pichia pastoris* genome in the examples herein, it is understood that the operation of the invention is independent of the loci used for integration. Loci other than those disclosed herein can be used for integration of the expression cassettes. Suitable integration sites include those enumerated in U.S. Published Application No. 20070072262 and include homologs to loci known for *Saccharomyces cerevisiae* and other yeast or fungi.

EXAMPLE 2

Strain **YGLY3159** in Example 1 was further genetically engineered to disrupt the *BMT1*, *BMT3*, and *BMT4* genes as follows.

Strain **YGLY3159** was counterselected in the presence of 5-FOA to produce strain **YGLY3225**, which is now auxotrophic for uridine.

Plasmid pGLY3411 (pSH1092) (**Figure 15**) is an integration vector that contains the expression cassette comprising the *P. pastoris URA5* gene flanked by *lacZ* repeats flanked on one side with the 5' nucleotide sequence of the *P. pastoris BMT4* gene (SEQ ID NO:72) and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT4* gene (SEQ ID NO:73). Plasmid pGLY3411 was linearized and the linearized plasmid transformed into **YGLY3159** to produce a number of strains in which the *URA5* expression cassette has been inserted into the *BMT4* locus by double-crossover homologous recombination. Strain **YGLY4439** was selected from the strains produced and is prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan. The strain is resistant to Zeocin and contains about three to four copies of the rhEPO expression cassette. The strain has a disruption or deletion of the *BMT2* and *BMT4* genes.

Plasmid pGLY3430 (pSH1115) (**Figure 16**) is an integration vector that contains an expression cassette comprising a nucleic acid molecule encoding the Nourseothricin resistance (NAT^R) ORF (originally from pAG25 from EROSCARF, Scientific Research and Development GmbH, Daimlerstrasse 13a, D-61352 Bad Homburg, Germany, *See* Goldstein *et al.*, Yeast 15: 1541 (1999)) ORF (SEQ ID NO:102) operably linked to the *Ashbya gossypii TEF1* promoter (SEQ ID NO:105) and *Ashbya gossypii TEF1* termination sequences (SEQ ID NO:106) flanked one side with the 5' nucleotide sequence of the *P. pastoris BMT1* gene (SEQ ID NO:68) and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT1* gene (SEQ ID NO:69). Plasmid pGLY3430 was linearized and the linearized plasmid transformed into strain **YGLY4439** to produce a number of strains in which the NAT^R expression cassette has been inserted into the *BMT1* locus by double-crossover homologous recombination. The strain **YGLY6661** was selected from the strains produced and is prototrophic for uracil, adenine,

histidine, proline, arginine, and tryptophan. The strain is resistant to Zeocin and Nourseothricin and contains about three to four copies of the EPO expression cassette. The strain has a disruption or deletion of the *BMT1*, *BMT2*, and *BMT4* genes. Strain **YGLY7013** was selected as well; however, this strain had only a partial disruption of the *BMT1* gene. This strain was designated as having a disruption or deletion of the *BMT1, BMT2* and *BMT4* genes.

Plasmid pGLY4472 (pSH1186) (**Figure 17**) is an integration vector that contains an expression cassette comprising a nucleic acid molecule encoding the *E. coli* hygromycin B phosphotransferase gene (Hyg^R) ORF (SEQ ID NO:103) operably linked to the *Ashbya gossypii TEF1* promoter (SEQ ID NO:105) and *Ashbya gossypii TEF1* termination sequences (SEQ ID NO:106) flanked one side with the 5' nucleotide sequence of the *P. pastoris BMT3* gene (SEQ ID NO:70) and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT3* gene (SEQ ID NO:71). Plasmid pGLY3430 was linearized and the linearized plasmid transformed into strain **YGLY6661** to produce a number of strains in which the Hyg^R expression cassette has been inserted into the *BMT3* locus by double-crossover homologous recombination. Strains **YGLY7361** to **YGLY7366** and strains **YGLY7393** to **YGLY7398** were selected from the strains produced and are prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan. The strains are resistant to Zeocin, Nourseothricin, and Hygromycin and contain about three to four copies of the EPO expression cassette. The strains have disruptions or deletions of the *BMT1*, *BMT2*, *BMT3*, and *BMT4* genes and produce rhEPO lacking cross-reactivity binding to antibodies made against host cell antigen (HCA).

EXAMPLE 3

Strain **YGLY3159** in Example 1 was further genetically engineered to produce strains in which the *BMT1*, *BMT3*, and *BMT4* genes have been disrupted or deleted and to include several copies of an expression cassette encoding mature human EPO fused to the chicken lysozyme leader peptide. Briefly, construction of these strains from **YGLY3159** is shown in **Figure 1** and briefly described as follows.

Strain **YGLY3159** was counterselected in the presence of 5-FOA to produce strain **YGLY3225**, which is now auxotrophic for uridine.

Plasmid pGLY2057 (**Figure 18**) is an integration vector that targets the *ADE2* locus and contains an expression cassette encoding the *P.pastoris URA5* gene flanked by *lacZ repeats*. The expression cassette is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *ADE2* gene (SEQ ID NO:100) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *ADE2* gene (SEQ ID NO:101). Plasmid pGLY2057 was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY3225** to produce a number of strains in which the *URA5* cassette has been inserted into the *ADE2* locus by double-crossover homologous recombination. Strain

YGLY3229 was selected from the strains produced and is auxotrophic for adenine and prototrophic for uridine, histidine, proline, arginine, and tryptophan. The strain is resistant to Zeocin and contains about three to four copies of the EPO expression cassette.

Plasmid pGLY2680 (**Figure 19**) is an integration vector that can target the *TRP2* or *AOX1* locus and contains expression cassettes encoding (1) a chimeric EPO comprising the human mature erythropoetin (EPO) fused at the *N*-terminus to chicken lysozyme signal peptide to target the chimeric protein to the secretory pathway and secretion from the cell and (2) the *P. pastoris ADE2* gene without a promoter. The *ADE2* gene is poorly transcribed from a cryptic promoter. Thus, selection of *ade2Δ* yeast strains transformed with the vector in medium not supplemented with adenine requires multiple copies of the vector to be integrated into the genome to render the recombinant prototrophic for adenine. Since the vector further includes the EPO expression cassette, the recombinant yeast will also include multiple copies of the EPO cassette integrated into the genome. This vector and method has been described in Published PCT Application WO2009085135. The DNA sequence encoding the chicken lysozyme signal peptide is shown in SEQ ID NO:94, the codon-optimized ORF encoding the mature human EPO is shown in SEQ ID NO:92, and the *P. pastoris ADE2* gene without its promoter but including its termination sequences is shown in SEQ ID NO:96. The chimeric EPO is operably linked to the AOX1 promoter and *S. cerevisiae CYC* termination sequences. The two tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence comprising the *TRP2* gene.

Plasmid pGLY2680 was linearized at the *Pme*I site and transformed into **YGLY3229** to produce a number of strains in which the two expression cassette have been inserted into the *AOX1* locus by roll in single-crossover homologous recombination, which results in multiple copies of the EPO expression cassette inserted into the *AOX1* locus without disrupting the AOX1 locus. Strain **YGLY4209** was selected from the strains produced. This strain there are about 5-7 copies of the EPO expression cassette as determined by measuring the intensity of sequencing data of DNA isolated from the strain inserted into the locus. The strain is prototrophic for adenine, uridine, histidine, proline, arginine, and tryptophan. The strain contains in total about eight to eleven copies of EPO expression cassettes. During processing of the chimeric EPO in the ER and Golgi, the leader peptide is removed. Thus, the rhEPO produced is the mature form of the EPO.

Strain **YGLY4209** was counterselected in the presence of 5'-FOA to produce a number of strains that were auxotrophic for uracil. From the transformants produced, strain **YGLY4244** was selected.

Plasmid pGLY2713 (**Figure 20**), an integration vector that targets the *P. pastoris PEP4* gene (SEQ ID NO:104), contains the *P. pastoris PNO1* ORF adjacent to the expression cassette comprising the *P. pastoris URA5* gene flanked by *lacZ* repeats and flanked on one side

with the 5' nucleotide sequence of the *P. pastoris PEP4* gene and on the other side with the 3' nucleotide sequence of the *P. pastoris PEP4* gene. Plasmid pGLY2713 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY4244** to produce a number of strains in which the *PNO1* ORF and *URA5* expression cassette have been inserted into the *PEP4* locus by double-crossover homologous recombination. Strain **YGLY5053** was selected from the strains produced and counterselected in the presence of 5-FOA to produce a number of strains in which the *URA5* has been lost from the genome. Strain **YGLY5597** was selected from the strains produced and is prototrophic for adenine, histidine, proline, arginine, and tryptophan. The strain is resistant to Zeocin and contains about eight to eleven copies of the rhEPO expression cassette.

Plasmid pGLY3411 (pSH1092) (**Figure 15**) is an integration vector that contains the expression cassette comprising the *P. pastoris URA5* gene flanked by *lacZ* repeats flanked on one side with the 5' nucleotide sequence of the *P. pastoris BMT4* gene (SEQ ID NO:72) and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT4* gene (SEQ ID NO:73). Plasmid pGLY3411 was linearized and the linearized plasmid transformed into strain **YGLY5597** to produce a number of strains in which the *URA5* expression cassette has been inserted into the *BMT4* locus by double-crossover homologous recombination. The strain **YGLY5618** was selected from the strains produced and is prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan. The strain is resistant to Zeocin and Nourseothricin and contains about eight to eleven copies of the rhEPO expression cassette. The strain has disruptions of the *BMT2* and *BMT4* genes.

Plasmid pGLY3430 (pSH1115) (**Figure 16**) is an integration vector that contains an expression cassette comprising a nucleic acid molecule encoding the Nourseothricin resistance (NAT^R) ORF (originally from pAG25 from EROSCARF, Scientific Research and Development GmbH, Daimlerstrasse 13a, D-61352 Bad Homburg, Germany, *See* Goldstein *et al.*, Yeast 15: 1541 (1999)) ORF (SEQ ID NO:102) operably linked to the *Ashbya gossypii TEF1* promoter and *Ashbya gossypii TEF1* termination sequences flanked one side with the 5' nucleotide sequence of the *P. pastoris BMT1* gene (SEQ ID NO:68) and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT1* gene (SEQ ID NO:69). Plasmid pGLY3430 was linearized and the linearized plasmid transformed into strain **YGLY5618** to produce a number of strains in which the NAT^R expression cassette has been inserted into the *BMT1* locus by double-crossover homologous recombination. The strain **YGLY7110** was selected from the strains produced and is prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan. The strain is resistant to Zeocin and Nourseothricin and contains about eight to eleven copies of the rhEPO expression cassette. The strain has disruptions of the *BMT1*, *BMT2*, and *BMT4* genes.

Plasmid pGLY4472 (pSH1186) (**Figure 17**) is an integration vector that contains an expression cassette comprising a nucleic acid molecule encoding the *E. coli* hygromycin B

phosphotransferase gene (*Hyg*^{R}) ORF (SEQ ID NO:103) operably linked to the *Ashbya gossypii TEF1* promoter and *Ashbya gossypii TEF1* termination sequences flanked one side with the 5' nucleotide sequence of the *P. pastoris BMT3* gene (SEQ ID NO:70) and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT3* gene (SEQ ID NO:71). Plasmid pGLY3430 was linearized and the linearized plasmid transformed into strain **YGLY7110** to produce a number of strains in which the *Hyg*^{R} expression cassette has been inserted into the *BMT3* locus by double-crossover homologous recombination. Strains **YGLY7113** to **YGLY7122** were selected from the strains produced and are prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan. The strains are resistant to Zeocin, Nourseothricin, and Hygromycin and contain about eight to eleven copies of the EPO expression cassette. The strains have disruptions of the *BMT1*, *BMT2*, *BMT3*, and *BMT4* genes and produce rhEPO lacking detectable cross-reactivity binding to antibodies made against HCA.

EXAMPLE 4

Several of the strains in Examples 1 to 3 were used to produce rhEPO as described below and shown schematically in **Figure 21**. Briefly, production begins by inoculating shake flasks containing culture media with cells from the working cell bank and proceeds through a series of inoculations, incubations, and transfers of the expanding cultures into vessels of increasing size until sufficient biomass is available to inoculate the production bioreactor. Glycerol is the primary carbon source during batch phase, then culture growth is maintained through feeding of glycerol and salts. When the glycerol is depleted, cells are induced to express rhEPO protein by switching to a methanol feed. Inhibitors are added at induction to minimize *O*-glycosylation (e.g., PMTi 3, 5-[[3-(1-Phenylethoxy)-4-(2-phenylethoxy)]phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic Acid, (*See* Published PCT Application No. WO 2007061631)) and to minimize proteolysis. Inhibitors of proteolysis are added again at the end of the phase to minimize proteolysis. The culture is cooled to about 4° C and harvested.

Laboratory scale cultivation of the strains was conducted in 500 mL SixFors and 3L fermentors using in general the following procedures. Bioreactor Screenings (SIXFORS) are done in 0.5 L vessels (Sixfors multi-fermentation system, ATR Biotech, Laurel, MD) under the following conditions: pH at 6.5, 24°C, 0.3 SLPM, and an initial stirrer speed of 550 rpm with an initial working volume of 350 mL (330 mL BMGY medium and 20mL inoculum). IRIS multi-fermenter software (ATR Biotech, Laurel, MD) is used to linearly increase the stirrer speed from 550 rpm to 1200 rpm over 10 hours, one hour after inoculation. Seed cultures (200 mL of BMGY in a 1 L baffled flask) are inoculated directly from agar plates. The seed flasks are incubated for 72 hours at 24°C to reach optical densities (OD_{600}) between 95 and 100. The fermenters are inoculated with 200 mL stationary phase flask cultures that were concentrated to

20 mL by centrifugation. The batch phase ended on completion of the initial charge glycerol (18-24h) fermentation and are followed by a second batch phase that is initiated by the addition of 17 mL of glycerol feed solution (50% [w/w] glycerol, 5 mg/L Biotin, 12.5 mL/L PMTi salts (65 g/L $FeSO_4.7H_2O$, 20 g/L $ZnCl_2$, 9 g/L H_2SO_4, 6 g/L $CuSO_4.5H_2O$, 5 g/L H_2SO_4, 3 g/L $MnSO_4 \cdot 7H_2O$, 500 mg/L $CoCl_2 \cdot 6H_2O$, 200 mg/L $NaMoO_4 \cdot 2H_2O$, 200 mg/L biotin, 80 mg/L NaI, 20 mg/L H_3BO_4)). Upon completion of the second batch phase, as signaled by a spike in dissolved oxygen, the induction phase is initiated by feeding a methanol feed solution (100% MeOH 5 mg/L biotin, 12.5 mL/L PMTi) at 0.6 g/h for 32-40 hours. The cultivation is harvested by centrifugation.

Bioreactor cultivations (3L) are done in 3L (Applikon, Foster City, CA) and 15L (Applikon, Foster City, CA) glass bioreactors and a 40L (Applikon, Foster City, CA) stainless steel, steam in place bioreactor. Seed cultures are prepared by inoculating BMGY media directly with frozen stock vials at a 1% volumetric ratio. Seed flasks are incubated at 24°C for 48 hours to obtain an optical density (OD_{600}) of 20±5 to ensure that cells are growing exponentially upon transfer. The cultivation medium contained 40 g glycerol, 18.2 g sorbitol, 2.3 g K_2HPO_4, 11.9 g KH_2PO_4, 10 g yeast extract (BD, Franklin Lakes, NJ), 20 g peptone (BD, Franklin Lakes, NJ), 4 x 10^{-3} g biotin and 13.4 g Yeast Nitrogen Base (BD, Franklin Lakes, NJ) per liter. The bioreactor is inoculated with a 10% volumetric ratio of seed to initial media. Cultivations are done in fed-batch mode under the following conditions: temperature set at 24±0.5°C, pH controlled at to 6.5±0.1 with NH_4OH, dissolved oxygen was maintained at 1.7±0.1 mg/L by cascading agitation rate on the addition of O_2. The airflow rate is maintained at 0.7 vvm. After depletion of the initial charge glycerol (40 g/L), a 50% glycerol solution containing 12.5 mL/L of PTM1 salts is fed exponentially at 50% of the maximum growth rate for eight hours until 250 g/L of wet cell weight was reached. Induction is initiated after a 30 minute starvation phase when methanol was fed exponentially to maintain a specific growth rate of 0.01 h^{-1}. When an oxygen uptake rate of 150 mM/L/h is reached the methanol feed rate is kept constant to avoid oxygen limitation. The cultivation is harvested by centrifugation.

After clarification by centrifugation and microfiltration, the filtrate is concentrated 10X by ultrafiltration and the rhEPO protein is purified through a sequence of two chromatography steps using a blue dye-affinity and hydroxyapatite.

Primary clarification is performed by centrifugation. The whole cell broth is transferred into 1000 mL centrifuge bottles and centrifuged at 4° C for 15 minutes at 13,000 x g. An ultrafiltration step can be employed for larger fermentors (10 L to 40 L and larger). This step can be performed utilizing Sartorious flat sheets with a pore size of 10K to a five-fold concentration.

A capture step is performed using Blue SEPHAROSE 6 Fast Flow (Pseudo-Affinity) Chromatography. A Blue SEPHAROSE 6 fast Flow (FF) column (GE Healthcare) is

equilibrated with 50 mM MOPS, pH 7.0. The culture supernatant is adjusted to 100 mM NaCl and passed through dead-end filter (Whatman, Polycap TC) before loading to the column. The residence time is maintained to about 10 minutes with a 3 column volumes (CV) wash after loading. The elution is step elution of 4 CV with 1 M NaCl in 50 mM MOPS, pH 7.0. EPO elutes at the 1 M NaCl.

An intermediate step is performed using hydroxyapatite (HA) chromatography. A Macro-prep ceramic hydroxyapatite Type I 40 μm (Bio-Rad) is used after the capture step. This column is equilibrated with equilibration solution: 50mM MOPS, pH 7.0 containing 1 M NaCl and 10 mM $CaCl_2$. About 10 mM $CaCl_2$ is added to the pooled rhEPO from the blue column before loading. The column wash is executed with 3 CV of equilibration solution followed by step elution of 10 CV at 12.5 mM Na phosphate in MPOS, pH 7.0 to provide HA pool 1 containing the rhEPO.

A cation exchange chromatography step can be used to further purify the rhEPO. The pooled sample after hydroxyapatite chromatography step (e.g., HA pool 1) is dialyzed against 50 mM Na acetate, pH 5.0 overnight at 4° C and a Source 30S column or Poros cation exchange column (GE Healthcare) is equilibrated with the same buffer. The dialyzed sample is applied to the column and a 10 CV linear gradient from 0 to 750 mM NaCl is applied with rhEPO eluting between 350 to 500 mM NaCl to provide the rhEPO.

The *N* terminus of the purified rhEPO molecule can be conjugated to 40-kDa linear polyethylene glycol (PEG) via reductive amination (PEGylation). The activated PEG is added to the rhEPO sample (conc. about 1 mg/mL) in 50mM Sodium acetate buffer at pH 5.2 at a protein:PEG ratio of 1:10. The reaction is carried out at room temperature under reducing conditions by adding 10mM sodium cyanoborohydride to the reaction mixture with overnight stirring. The reaction is stopped by adding 10mM Tris, pH 6.0.

The mono-PEGylated rhEPO product is purified using a cation-exchange chromatography step before diafiltration into the final formulation buffer (20 mM sodium phosphate, 120 mM sodium chloride, 0.005% Polysorbate 20 (w/v), pH 7.0).

The final product is diluted to a concentration suitable for filling and sterile filtered into the drug substance storage container. The PEGylated rhEPO can be stored at 2-8° C until filling, at which time it is aseptically filled into glass vials that are then sealed with a rubber stopper and aluminum cap.

Commercial formulations of proteins are known and may be used. Examples include but are not limited to ARANESP®: Polysorbate solution: Each 1 mL contains 0.05 mg polysorbate 80, and is formulated at pH 6.2 ± 0.2 with 2.12 mg sodium phosphate monobasic monohydrate, 0.66 mg sodium phosphate dibasic anhydrous, and 8.18 mg sodium chloride in water for injection, USP (to 1 mL). Albumin solution: Each 1 mL contains 2.5 mg albumin (human), and is formulated at pH 6.0 ± 0.3 with 2.23 mg sodium phosphate monobasic

monohydrate, 0.53 mg sodium phosphate dibasic anhydrous, and 8.18 mg sodium chloride in water for injection, USP (to 1 mL). EPOGEN® is formulated as a sterile, colorless liquid in an isotonic sodium chloride/sodium citrate buffered solution or a sodium chloride/sodium phosphate buffered solution for intravenous (IV) or subcutaneous (SC) administration. Single-dose, Preservative-free Vial: Each 1 mL of solution contains 2000, 3000, 4000 or 10,000 Units of Epoetin alfa, 2.5 mg Albumin (Human), 5.8 mg sodium citrate, 5.8 mg sodium chloride, and 0.06 mg citric acid in water for injection, USP (pH 6.9 ± 0.3). This formulation contains no preservative. Preserved vials contain 1% benzyl alcohol.

EXAMPLE 5

Methods used for analyzing the presence or absence of host cell antigen (HCA) included Western blot analysis and sandwich enzyme-linked immunosorbent assay (ELISA).

Host cell Antigen (HCA) antibody was prepared in rabbits using the supernatant from NORF strain cultures. The NORF strain is genetically the same as **YGLY3159** except that it lacks the ORF encoding the human mature EPO. NORF strain fermentation supernatant prepared in complete Freund's adjuvant was injected into rabbits, which were then boosted three times with fermentation supernatant prepared in Incomplete Freund's adjuvant. After 45 days, the rabbits were bled and polyclonal antibodies to HCA were prepared using standard methods, for example, rabbit polyclonal IgG 9161 F072208-S, which was SLr Protein A purified, and GiF2 polyclonal rabbit::6316 whole rabbit serum. The GIF2 antibody was not protein A purified.

Western Blots for detecting *P. Pastoris* HCA were performed as follows. Purified PEGylated or non-PEGylated rhEPO-containing samples were reduced in sample loading buffer, of which 1 µL was then applied to the wells of 4-20% polyacrylamide SDS Tris-HCl (4-20% SDS-PAGE) gels (Bio RAD) and electrophoresed at 150V for about 60 minutes. The resolved proteins were electrotransferred to nitrocellulose membranes at 100V for about 60 minutes. After transfer, the membranes were blocked for one hour with 1% Blocking Solution (Roche Diagnostics). After blocking, the membranes were probed with the rabbit anti-HCA polyclonal antibody (primary antibody) diluted 1:3000. Afterwards, the membranes were washed and detection of the rabbit anti-HCA antibody was with the secondary antibody, goat-anti-Rabbit IgG (H+L) (Pierce #31460, Lot # H51015156) conjugated to horseradish peroxidase (HRP), at a 1:5000 dilution. After washing the membranes, detection of bound secondary antibody was using 3,3' Diaminobenzidine (DAB). For detecting EPO protein, the primary antibody was EPO (B-4) HRP-conjugated antibody used at a 1:1000 dilution (SC5290 Lot# A0507, Santa Cruz Biotechnology). A secondary antibody was not used. Routinely, the EPO samples were electrophoresed in parallel with rhEPO samples that had been deglycosylated with PNGaseF treatment. Deglycosylation was performed with 50 uL samples to which 1 µL of PNGaseF enzyme at 500 units/uL was added. After incubation at 37 °C for two hours, the samples were

reduced in sample loading buffer and 1 μL aliquots were removed and applied to the SDS gels as above.

Sandwich ELISAs for detecting *P. Pastoris* HCA were performed as follows. The wells of 96 well ELISA plates were coated with 1μg/well of mouse anti-hEPO monoclonal antibody. The wells were then blocked for 30 minutes with phosphate-buffered saline (PBS). About 100 μL of purified non-PEGylated rhEPO-containing samples concentrated to about 200 ng/mL were added to the wells. Primary detection used the rabbit anti-HCA polyclonal antibody at a 1:800 starting dilution in PBS which was then serially diluted 1:1 in PBS across a row ending with the 11^{th} well at a 1:819,200 dilution. The 12^{th} well served as a negative control. The standard for the ELISA was rhEPO purified from YGLY3159. After 60 minutes, the wells were washed with PBS three times. Detection of the rabbit anti-HCA antibody used goat anti-rabbit antibody conjugated to alkaline phosphatase (AP) at a 1:10,000 dilution in PBS. After 60 minutes the wells were washed three times with PBS and detection of bound secondary antibody used 4-Methylumbelliferyl phosphate (4-MUPS). The ELISA plates were read using a Tecan Genios Multidetection Microplate Reader at 340 nm excitation wavelength and 465 nm emission wavelength.

EXAMPLE 6

This example shows that **YGLY3159** produces rhEPO with cross binding activity (CBA) with anti-HCA antibody and that the cross-binding activity was due to the presence of β-1,2-mannose residues (α-1,2- mannosidase resistant) on at least a portion of the *N*-glycans on the rhEPO even though the rhEPO had been produced in strain in which the β-1,2-mannosyltransferase gene *BMT2* had been deleted or disrupted.

rhEPO was recovered by a three-step chromatographic separation from the fermentation supernatant of glyco-engineered *P. pastoris* production strain **YGLY 3159** showed about 95% protein purity as determined by SDS-PAGE, RP-HPLC, and SEC-HPLC. Mono-PEGylated rhEPO was separated by cation-exchange chromatographic step from its hyper and un-PEGylated conjugates with about 96% purity as determined by SDS-PAGE gel. However, antibody against HCA of the **YGLY3159** strain detected a glycoprotein in rhEPO preparations produced from the strain that co-migrated with rhEPO on Western blots. **Figure 22** which shows that anti-HCA antibody identified a protein that co-migrates with rhEPO on 4-20% SDS-PAGE gels. Removal of sialic acid from rhEPO did not abolish the cross-binding activity; however, removal of the entire N-glycan from rhEPO using PNGase F produced a deglycosylated form of rhEPO that was not detectable in Western blots probed with anti-HCA antibody. This is shown in **Figure 23** which shows that only the deglycosylated form of rHEPO lacked cross-binding activity with the anti-HCA antibody.

To determine wither the cross-binding activity was rhEPO specific or could be identified in purified glycoprotein preparations from other recombinant *P. pastoris* strains, an glycoproteins produced in other strains were isolated, resolved by 4-20% SDS-PAGE gels, and the gels transferred to nitrocellulose membranes. In the case of a recombinant human whole antibody (rhIgG) produced in a recombinant *P. pastoris*, cross-binding activity was detected in protein preparations produced in wild-type *P. pastoris* (hypermannosylated from both *N* and *O*-glycocylated region) and in a recombinant GS2.0 strain that makes predominantly $Man_5GlcNAc_2$ *N*-glycans but also contained detectable $Man_9GlcNAc_2$ *N*-glycans that were α-1,2-mannosidase resistant (**Figure 24**, arrow). However, the rhIgG preparations from wild-type *P. pastoris* contained cross-binding activity with an apparent molecular weight greater than that of rhIgG suggesting that the preparations contained contaminating host cell glycoproteins. The cross-binding activity was not removed by PNGase F digestion (circled in **Figure 24**).

Figure 25 shows that glycosylated rhEPO produced in **YGLY3159** had cross binding activity to anti-HCA antibody but that human fetuin, human asialofetuin, human serum albumin (HSA), and LEUKINE (a recombinant human granulocyte macrophage colony stimulating factor (rhu GM-CSF) produced in *S. cerevisiae*) had no cross-binding activity to anti-HCA antibody. Fetuins are heavily glycosylated blood glycoproteins that are made in the liver and secreted into the blood stream. They belong to a large group of binding proteins mediating the transport and availability of a wide variety of cargo substances in the blood stream. The best known representative of these carrier proteins is serum albumin, the most abundant protein in the blood plasma of adult animals. Fetuin is more abundant in fetal blood, hence the name "fetuin" (from Latin, fetus). Fetal calf serum contains more fetuin than albumin while adult serum contains more albumin than fetuin. Asialofetuins are fetuins which the terminal sialic acid from *N*- and *O*-glycans are removed by mild hydrolysis or neuraminidase treatment. Currently, there are no reports of β-linked mannoses in *S cerevisiae*. HSA is not a glycosylated protein.

Lab scale data demonstrated that the intermediate chromatographic step purification of rhEPO from Blue SEPHAROSE 6 FF capture pool using hydroxy apatite (HA) type I 40 μm resin can separate rhEPO that has nearly undetectable cross-binding activity (HA pool 1) from rhEPO that had high-mannose-type N-glycans (HA pools 2 and 3). HA pool 1 contained about 90.40 % bisialylated *N*-glycans (the desired *N*-glycan form) and less than 3.5% neutral *N*-glycans. In contrast, linear gradient elution from 0 to 100mM sodium phosphate showed that later elution fractions (HA pools 2 and 3) contained high mannose-type *N*-glycans and increased cross binding activity to anti-HCA antibody in Western blots. This can be seen in the HPLC *N*-glycan analysis and Western blots of 4-20% SDS-PAGE gels shown in **Figure 26**

Anion column chromatography using Q SEPHAROSE FF or Source 30Q anion resins were also tested. The HA pools 1-3 were combined and dialyzed against 50 mM Na acetate, pH 5.0 overnight at 4° C. The dialyzed sample was applied to the column and a 10 CV

linear gradient from 0 to 750 mM NaCl was applied with rhEPO eluting between 350 to 500 mM NaCl to provide the rhEPO. **Figure 27A** shows an example of a Q SEPHAROSE FF purification of rhEPO. Data showed that high mannose type glycans ($Man_{6,7,8,9,>9}$, mostly α 1,2 mannosidase resistant) that show corresponding higher cross-binding activity activity did not bind to the anion exchange resins when bound and unbound material was analyzed in a sandwich ELISA (**Figure 27B**). Table 1 shows the results of HPLC analysis of the *N*-glycan content of the rhEPO in the bound fraction (Q SEPHAROSE FF pool 1) and flow-through fraction (Q SEPHAROSE FF Flow Through). Table 2 shows the N-glycan content of the neutral N-glycans shown in Table 1.

Table 1

Q Sepharose FF – Purification of rhEPO
***N*-Glycan HPLC Analysis**

Sample	% Neutral	% Mono Sialylated	% Bi Sialylated
Input (HA pools)	11.04	13.66	75.30
Q SEPHAROSE FF pool 1	3.01	6.47	90.52
Q SEPHAROSE FF Flow Through	26.71	21.19	52.10

Table 2

Q Sepharose FF – Purification of rhEPO
% Neutral *N*-Glycan Profile

Sample	%G2	% Man_5	% Man_{6-8}	% Man_9
Input (HA pools)	3.1	2.8	2.4	2.74
Q SEPHAROSE FF pool 1	3.01	ND	ND	ND
Q SEPHAROSE FF Flow Through	4.2	8.0	3.9	10.61

ND – not detected
G2 – N-glycan structure is $Gal_2GlcNAc_2Man_3GlcNAc_2$

The figures and tables show that rhEPO with undetectable cross-binding activity to anti-HCA antibodies and good protein and glycan quality can therefore be bound/eluted from anion exchange resins. These data also suggested that the family of fungal genes involved in biosynthesis of β-1,2-linked oligomannosides (*BMT1*, *BMT2*, *BMT3*, *BMT4*) was responsible for the low level cross-binding impurities in the rhEPO preparations.

Therefore, when viewed as a whole, the results suggested that the cross-binding activity to anti-HCA antibodies was not specific to rhEPO but was due to α-1,2-mannosidase resistant *N*-glycans on the glycoproteins. **YGLY 3159** had been generated by knocking out five endogenous glycosylation genes and introducing 15 heterologous genes. **YGLY3159** is *bmt2Δ*

knockout strain. NMR spectroscopy studies suggest that *bmt2Δ* knockout strains can produce glycoproteins with varying amounts of residual β-1,2-mannose *N*-glycans. Since **YGLY 3159** is *bmt2Δ*, it was postulated that *BMT1* and *BMT3* were responsible for the residual low level β-1,2-mannose transfer on core *N*-glycans.

While a combination of chromatography steps to purify the rhEPO can produce rhEPO preparations free of detectable cross-binding activity to anti-HCA antibodies, it would be particularly desirable to genetically modify the *P. pastoris* host strains to reduce or eliminate detectable cross-binding activity to anti-HCA antibodies in the strains. This minimizes the risk of possible contamination of the rhEPO preparations with cross-binding activity due to variability during the purification. In addition, because each purification step can result in a loss of rhEPO, the genetically modified P. pastoris strains can reduce the number of purification steps and thus reduce the amount of rhEPO lost during the steps eliminated. Therefore, expression of the four *BMT* genes were serially deleted or disrupted to identify strains that did not produce detectable cross-binding activity to anti-HCA antibodies.

EXAMPLE 7

In order to reduce the presence of β-linked mannose type *N*-glycans to undetectable levels, the *BMT1* and *BMT4* genes were disrupted and the rhEPO analyzed for the presence of α-1,2-mannosidase resistant *N*-glycans.

Strains **YGLY6661** and **YGLY7013** were constructed as described Example 2 and analyzed for the presence of α-1,2-mannosidase resistant *N*-glycans using anti-HCA antibodies. Strain YGLY7013 was *bmt2Δ* and *bmt4Δ* and strain YGLY6661 was *bmt2Δ*, *bmt4Δ*, and *bmt1Δ*. rhEPO produced from the strains were subjected Blue SEPHAROSE 6FF chromatography and aliquots of the Blue SEPHAROSE 6FF capture pool were treated with PNGase F vel non. The treated and untreated aliquots were electrophoresed on SDS-PAGE, the gels transferred to nitrocellulose membranes, and the membranes probed with anti-EPO antibody or anti-HCA antibodies. **Figure 28** shows in Western blots of 4-20% SDS-PAGE gels of aliquots of Blue SEPHAROSE 6 FF capture pools that rhEPO produced in either strain still had α-1,2-mannosidase resistant *N*-glycans which cross-reacted with anti-HCA antibodies. Tables 3 and 4 show the distribution of *N*-glycan species in rhEPO in Blue Sepaharose 6 FF capture pools from both fermentation and SixFors cultures. As shown in the tables, both strains produced a substantial amount of neutral *N*-glycans of which a portion was resistant to *in vitro* α1,2-mannosidase digestion.

Table 3

Week 44 – rhEPO – Blue SEPHAROSE 6 FF Capture Pool - Fermentation				
Pools	% Bi-	% Mono	%	% Neutral

	Sialylated	Sialylated	Neutral	% G2	% M5	% M6-M8	% M9+
F074411 **(YGLY 6661)**	52.98	34.08	12.94	1.7	2.63	3.65	4.96
F074411 **(YGLY 6661)** α1,2 Mannosidase	53.42	34.32	12.26	1.9	5.05	3.4	1.91
F074410 **(YGLY 7013)**	25.10	47.00	27.90	12.99	2.22	5.67	7.02
F074410 **(YGLY 7013)** α1,2 Mannosidase	26.34	49.03	26.34	13.14	5.39	4.68	1.42

G2 – $Gal_2GlcNAc_2Man_3GlcNAc_2$ *N*-glycans
M6-M8 – $Man_6GlcNAc_2$ to $Man_8GlcNAc_2$ *N*-glycans
M9+ - $Man_9GlcNAc_2$ and lager *N*-glycans

Table 4

Week 41 – rhEPO – Blue SEPHAROSE 6 FF Capture Pool - SixFors

Pools	% Bi-Sialylated	% Mono Sialylated	% Neutral	% Neutral				
				% G2	% M5	% M6-M8	% M9	%M9+
X074128 **(YGLY 6661)**	43.49	39.24	17.27	1.7	7.8	6.69	0.45	0.63
X074128 **(YGLY 6661)** α1,2 Mannosidase	42.52	39.26	18.22	1.2	11.84	5.02	0.1	0.06
X074131 **(YGLY 7013)**	66.90	18.83	14.27	1.84	8.36	2.82	0.66	0.59
X074131 **(YGLY 7013)** α1,2 Mannosidase	64.81	19.70	15.49	1.06	13.1	0.77	0.56	0

G2 – $Gal_2GlcNAc_2Man_3GlcNAc_2$ *N*-glycans
M6-M8 – $Man_6GlcNAc_2$ to $Man_8GlcNAc_2$ *N*-glycans
M9 - $Man_9GlcNAc_2$ *N*-glycans
M9+ - $Man_9GlcNAc_2$ and lager *N*-glycans

A sandwich ELISA of rhEPO in the Blue SEPHAROSE 6 FF capture pools made from both strains compared to **YGLY3159** showed that both strains had cross-binding activity to anti-HCA antibody (**Figure 29**). Further purifying the rhEPO by hydroxyapatite (HA) chromatography and analyzing the samples by sandwich ELISA showed that the HA pool 1 containing rhEPO produced from **YGLY6661** (*bmt2Δ*, *bmt4Δ*, and *bmt1Δ*) appeared to lack detectable cross-binding activity to anti-HCA antibody but that rhEPO produced in **YGLY7013** (*bmt2Δ* and *bmt4Δ*) still had detectable cross-binding activity to anti-HCA antibody (**Figure 30**). The results suggested that deleting the *BMT2* and *BMT1* genes was not sufficient to remove all detectable cross-binding activity. The results also show that hydroxyapatite chromatography can remove detectable cross-binding activity in the HA pool 1. **Figure 31** is a Western blot of 4-20% SDS-PAGE gels showing that rhEPO in another Blue SEPHAROSE 6 FF capture pool prepared from strain **YGLY6661** continued to have cross-binding activity to anti-HCA antibody and that the cross-binding activity could be still be rendered undetectable by deglycosylating the rhEPO. The result indicated that to produce rhEPO that had no detectable cross-binding activity to anti-HCA antibodies, expression of the *BMT3* gene needed to be abrogated by disruption or deletion.

EXAMPLE 8

In order to more effectively achieve the elimination of detectable β-linked mannose type glycans, all four *BMT* genes involved in -mannosyltransferase pathway were disrupted. Strains **YGLY7361-7366** and **YGLY7393-7398** (Example 2) were evaluated for ability to produce rhEPO lacking detectable cross-binding activity to anti-HCA antibody.

Various **YGLY7361-7366** and **YGLY7393-7398** strains in which all four *BMT* genes involved in the β-mannosyltransferase pathway were disrupted were grown in 500 mL SixFors fermentors and then processed for rhEPO through Blue SEPHAROSE 6 FF pools (Blue pools). Aliquots from the Blue pools were analyzed by 4-20% SDS-PAGE. **Figure 32** shows Commassie blue stained 4-20% SDS-PAGE gels of the Blue pools from the various strains with and without PNGase F treatment. The gels show that all of the tested strains produced glycosylated rhEPO. Several of the strains were evaluated for cross-binding activity to anti-HCA antibody by sandwich ELISA. **Figure 33** shows that most of the strains lacked detectable cross-binding activity activity to anti-HCA antibody. However, strains **YGLY7363** and **YGLY7365** had detectable cross-binding activity activity to anti-HCA antibody. Reconfirmation of **YGLY7365** by PCR indicated that this strain was not a complete knock-out of the *BMT3* gene, explaining the relatively high binding observed with the anti-HCA antibody present in the ELISA (**Figure 33**). HPLC N-glycan analysis of strains **YGLY7361-7366** is shown in Table 5 and strains **YGLY7393-7398** are shown in Table 6. The data in the tables are graphically presented in **Figure 34**.

Table 5

Week 46a – SixFors – *Δbmt1-4* strains – Blue pools

Pools	% Bi-Sialylated	% Mono Sialylated	% Neutral	% Neutral			
				% G2	% M5	% M6-M8	% M9+
X074613 **(YGLY 7361)**	22.10	47.83	30.07	13.85	2.53	6.77	6.92
X074613 **(YGLY 7361)** α1,2 Mannosidase	21.55	48.18	30.27	13.68	4.53	5.47	6.59
X074614 **(YGLY 7362)**	67.36	24.69	7.95	0.6	4.36	2.8	0.19
X074614 **(YGLY 7362)** α1,2 Mannosidase	66.21	25.25	8.54	1.1	6.7	0.68	0.06
*X074615 **(YGLY 7363)**	49.40	39.17	11.43	0.8	4.42	5.91	0.3
X074615 **(YGLY 7363)** α1,2 Mannosidase	48.52	39.20	12.28	0.4	7.2	4.68	ND
X074616 **(YGLY 7366)**	55.99	33.85	10.16	0.8	3.73	4.94	0.69
X074616 **(YGLY 7366)** α1,2 Mannosidase	55.44	34.24	10.32	1.9	7.2	1.02	0.2
*X074617 **(YGLY 7365)**	43.22	42.10	14.68	5.37	5.88	3.03	0.4
X074617 **(YGLY 7365)** α1,2 Mannosidase	42.70	42.40	14.90	4.5	8.4	2.0	ND
X074618 **(YGLY 7364)**	48.18	38.44	13.38	0.7	6.56	5.76	0.36
X074618 **(YGLY 7364)** α1,2 Mannosidase	47.52	39.75	12.73	0.4	6.74	5.09	0.5

G2 – $Gal_2GlcNAc_2Man_3GlcNAc_2$ *N*-glycans

M6-M8 – $Man_6GlcNAc_2$ to $Man_8GlcNAc_2$ *N*-glycans

M9+ - $Man_9GlcNAc_2$ and lager *N*-glycans
*Showed cross-binding activity to anti-HCA antibody

Table 6

Week 46a – SixFors – *Δbmt1-4* strains – Blue pools

Pools	% Bi-Sialylated	% Mono Sialylated	% Neutral	% Neutral				
				% G2	% M5	% M6-M8	% M9	%Hyb
X074637 **(YGLY 7393)**	51.04	35.45	13.51	2.3	6.4	1.41	1.2	2.2
X074637 **(YGLY 7393)** α1,2 Mannosidase	50.33	35.44	14.23	2.5	9.14	ND	ND	2.54
X074638 **(YGLY 7394)**	63.56	25.65	10.79	1.1	6.6	1.9	0.4	0.79
X074638 **(YGLY 7394)** α1,2 Mannosidase	62.88	25.75	11.37	1.2	8.96	ND	ND	1.21
X074639 **(YGLY 7395)**	56.05	31.43	12.52	1.9	5.1	2.2	1.9	1.4
X074639 **(YGLY 7395)** α1,2 Mannosidase	56.27	31.81	11.92	1.9	8.43	ND	ND	1.59
X074640 **(YGLY 7396)**	50.42	36.71	12.87	3.2	6.7	1.27	0.3	1.4
X074640 **(YGLY 7396)** α1,2 Mannosidase	49.94	36.86	13.20	3.2	8.12	ND	ND	1.88
X074641 **(YGLY 7397)**	49.32	36.07	14.61	2.6	7.0	2.4	0.5	2.11
X074641 **(YGLY 7397)** α1,2 Mannosidase	48.72	35.86	15.42	2.7	10.24	ND	ND	2.48
X074642 **(YGLY 7398)**	65.74	22.61	11.65	0.8	7.7	1.97	0.43	3.71
X074642	64.99	22.87	12.14	1.0	10.02	ND	ND	1.12

(YGLY 7398) α1,2 Mannosidase								
G2 – $Gal_2GlcNAc_2Man_3GlcNAc_2$ *N*-glycans M6-M8 – $Man_6GlcNAc_2$ to $Man_8GlcNAc_2$ *N*-glycans M9 - $Man_9GlcNAc_2$ *N*-glycans M9+ - $Man_9GlcNAc_2$ and lager *N*-glycans Hyb – hybrid *N*-glycans								

Strains **YGLY7362, 7366, 7396,** and **7398** were cultivated in 3L fermentors and processed through Blue SEPHAROSE 6 FF chromatography followed by hydroxyapatite (HA) chromatography. Aliquots from both the Blue pools and the HA pools were reduced and analyzed by 4-20% SDS-PAGE. Corresponding pools for **YGLY3159** were included as positive controls. **Figure 35A** shows a Commassie blue stained 4-20% SDS-PAGE showing that both the Blue pools (left half of gel) and HA pools (right half of gel) produced rhEPO. **Figure 35B** shows a Western blot of the same samples probed with anti-HCA antibodies. None of the tested strains had any detectable cross-binding activity to anti-HCA antibodies in either the Blue pool or the HA pool 1.

Figure 36 analyzes the Blue pool and HA pool 1 for rhEPO isolated from 500 mL SixFors cultures of **YGLY7398** for cross-binding activity to anti-HCA antibodies. The right-most panel shows a Western blot probed with another anti-HCA preparation: GiF2 polyclonal rabbit::6316. This antibody produced the same results as produced using the F072208-S antibody, which had been used to produce the ELISAs and Western blots shown herein. The 6316 antibody shows that the cross-binding activity is not antibody specific.

These results show that deleting or disrupting all four *BMT* genes can result in strains that do not produce detectable cross-binding activity to anti-HCA antibodies in either the rhEPO after the preliminary Blue SEPHAROSE 6 FF capture step or the intermediate hydroxyapatite step using Type I 40 μM hydroxyapatite. These strains minimize the risk that rhEPO preparations will be made that contain cross-binding activity to anti-HCA antibodies. This enables the production of rhEPO with less risk of inducing an adverse immune response in the individual receiving the rhEPO.

EXAMPLE 9

A comparison of the pharmacokinetics of the rhEPO produced in the strains produced in Example 2 with all four *BMT* genes disrupted or deleted and PEGylated was compared to PEGylated rhEPO produced from strain **YGLY3159**. The comparison showed that the PEGylated EPO had a reduced *in vivo* half-life and lower *in vivo* potency (*See* Tables 7 and 8). The rhEPO produced in the strains produced in Example 2 with no detectable cross-binding activity to anti-HCA antibodies had pharmacokinetics generally similar to that of EPOGEN and

not the higher pharmacokinetics of ARANESP. The reduced pharmacokinetics was found to be a function of the amount of bi-sialylated biantennary *N*-glycans. Higher levels of bi-sialylated biantennary *N*-glycan on the rhEPO was correlated with higher pharmacokinetics. These results are consistent with published data showing that longer half life is correlated with greater sialic acid content in recombinant human erythropoietin produced in CHO cells (Egrie *et al*, Exp. Hematol. 31: 290-299 (2003)).

Table 7		
PK of rhEPO from YGLY3159 (CBA) vs YGLY7398 (no CBA)		
	YGLY3159	**YGLY7398**
T1/2 (hr)	20.9±2	13±2
CBA - cross-binding activity		

Table 8		
rhEPO source	Relative Potency (Reticulocyte Production)	95% Confidence Interval
YGLY3159 vs **YGLY7398**	0.82	(0.68, 1.00)

EXAMPLE 10

In order to effectively achieve the elimination of detectable β-linked mannose type glycans and produce a strain that produces rhEPO with higher pharmacokinetics, strains **YGLY7113-7122** described in Example 3 were made and evaluated for ability to produce rhEPO lacking detectable cross-binding activity to anti-HCA antibody. These strains were modified to also express human mature EPO as a fusion protein fused to the chicken lysozyme leader sequence. Thus, these strains express both human mature EPO fused to the *S. cerevisiae* αMATpre signal peptide and the human mature EPO as a fusion protein fused to the chicken lysozyme leader sequence.

Various **YGLY7113-YGLY7122** strains in which all four *BMT* genes involved in the β-mannosyltransferase pathway were disrupted and expressing the were grown in 500 mL SixFors fermentors and then processed for rhEPO through Blue SEPHAROSE 6 FF pools (Blue pools). Aliquots of the Blue pools for several strains were analyzed by sandwich ELISA using anti-HCA antibodies. **Figure 37** shows that **YGLY7118** had very low cross-binding activity to anti-HCA antibody but all of the other strains showed no detectable cross-binding activity to anti-HCA antibodies. HPLC N-glycan analysis of strains **YGLY7113-7117** is shown in Table 9 and

strains **YGLY7118-7122** are shown in Table 10. The tables are graphically presented in **Figure 38**.

Table 9

Week 48 – SixFors – *Δbmt1-4* strains – Blue pools

Pools	% Bi-Sialylated	% Mono Sialylated	% Neutral	% Neutral				
				% G2	% M5	% M6-M8	% M9	%Hyb
X074814 **(YGLY 7113)**	70.23	9.97	19.80	0.2	9.2	6.85	2.05	1.5
X074814 **(YGLY 7113)** α1,2 Mannosidase	68.96	10.56	20.48	0.3	18.4	ND	ND	1.78
X074815 **(YGLY 7115)**	62.61	14.01	23.38	0.5	7.15	10.37	3.96	1.4
X074815 **(YGLY 7115)** α1,2 Mannosidase	61.77	13.95	24.28	0.1	22.22	ND	ND	1.96
X074816 **(YGLY 7114)**	67.64	8.22	24.14	0.2	4.2	11.41	6.33	2.0
X074816 **(YGLY 7114)** α1,2 Mannosidase	65.92	8.32	25.76	0.2	23.35	ND	ND	2.21
X074817 **(YGLY 7116)**	66.46	8.06	25.48	4.73	5.38	6.94	7.23	1.2
X074817 **(YGLY 7116)** α1,2 Mannosidase	65.54	8.69	25.77	0.5	23.8	ND	ND	1.47
X074818 **(YGLY 7117)**	70.06	11.09	18.85	0.6	8.59	6.0	2.21	1.45
X074818 **(YGLY 7117)** α1,2 Mannosidase	68.67	11.42	19.91	0.4	17.5	ND	ND	2.01

G2 – $Gal_2GlcNAc_2Man_3GlcNAc_2$ *N*-glycans
M6-M8 – $Man_6GlcNAc_2$ to $Man_8GlcNAc_2$ *N*-glycans
M9 - $Man_9GlcNAc_2$ *N*-glycans

M9+ - $Man_9GlcNAc_2$ and larger *N*-glycans
Hyb – hybrid *N*-glycans

Table 10

Week 48 – SixFors – *Δbmt1-4* strains – Blue pools

Pools	% Bi-Sialylated	% Mono Sialylated	% Neutral	% Neutral				
				% G2	% M5	% M6-M8	% M9	%Hyb
X074819 **(YGLY 7119)**	58.12	27.10	14.78	0.7	6.83	4.98	1.17	1.1
X074819 **(YGLY 7119)** α1,2 Mannosidase	57.03	26.87	16.10	0.45	14.55	ND	ND	1.1
X074820 **(YGLY 7120)**	73.60	10.84	15.56	0.89	8.6	3.75	1.64	0.68
X074820 **(YGLY 7120)** α1,2 Mannosidase	72.43	11.13	16.44	0.7	15.63	ND	ND	0.11
X074821 **(YGLY 7121)**	59.41	19.85	20.74	0.8	3.04	10.7	5.55	0.65
X074821 **(YGLY 7121)** α1,2 Mannosidase	58.39	20.00	21.6	0.4	20.17	ND	ND	1.04
X074822 **(YGLY 7122)**	57.43	24.16	18.41	1.37	10.89	4.95	0.4	0.8
X074822 **(YGLY 7122)** α1,2 Mannosidase	55.77	24.44	19.79	1.8	17.28	ND	ND	0.71
X074824 **(YGLY 7118)**	55.56	21.47	22.97	0.33	2.98	11.85	6.59	1.22
X074824 **(YGLY 7118)** α1,2 Mannosidase	54.68	21.67	23.65	0.4	22.5	ND	ND	0.75

G2 – $Gal_2GlcNAc_2Man_3GlcNAc_2$ *N*-glycans
M6-M8 – $Man_6GlcNAc_2$ to $Man_8GlcNAc_2$ *N*-glycans
M9 - $Man_9GlcNAc_2$ *N*-glycans

M9+ - $Man_9GlcNAc_2$ and larger *N*-glycans
Hyb – hybrid *N*-glycans

Strains **YGLY7115**, **7117**, and **7120** were cultivated in 3L fermentors and processed through Blue SEPHAROSE 6 FF chromatography followed by hydroxyapatite (HA) chromatography. Aliquots from both the Blue pools and the HA pools were reduced and analyzed by 4-20% SDS-PAGE. Corresponding pools for **YGLY3159** were included as positive controls. Corresponding pools for **YGLY7395** were included as negative controls. **Figure 39A** shows a Commassie blue stained 4-20% SDS-PAGE showing that both the Blue pools (left half of gel) and HA pools (right half of gel) produced rhEPO. **Figure 39B** shows a Western blot of the same samples probed with anti-HCA antibodies. None of the tested strains had any detectable cross-binding activity to anti-HCA antibodies in either the Blue pool or the HA pool 1.

These results also show that deleting or disrupting all four *BMT* genes can result in strains that do not produce detectable cross-binding activity to anti-HCA antibodies in either the rhEPO after the preliminary Blue SEPHAROSE 6 FF capture step or the intermediate hydroxyapatite step using Type I 40 μM hydroxyapatite. These strains minimize the risk that rhEPO preparations will be made that contain cross-binding activity to anti-HCA antibodies. This enables the production of rhEPO with less risk of inducing an adverse immune response in the individual receiving the rhEPO.

EXAMPLE 11

The blue pools containing rhEPO produced by **YGLY7117** were further subjected to hydroxyapatite column chromatography and the rhEPO in the HA pools were analyzed for sialylation content. **Figure 40A** and **Figure 40B** show HPLC traces of the *N*-glycans from rhEPO produced in **YGLY3159** compared to the *N*-glycans from rhEPO produced in **YGLY7117**, respectively. The figures also show that the hydroxyapatite column removed additional contaminants; thus, in this analysis the sialylation content of the rhEPO produced by YGLY7117 was about 99% (neutral *N*-glycans were about 1%) of which about 89% was A2 or bisialylated and about 10% was A1 or monosialylated.

Sialylation analysis of rhEPO produced in YGLY7117 following PEGylation according to the process in Example 3 was similar to the amount of sialylation prior to PEGylation; however, the amount of sialylation can vary to a limited extent depending for example, on what modifications were made to the growing conditions, e.g., medium compositions, feeding rate, etc (*See* Table 11). Thus, the methods herein produce rhEPO compositions having at least about 75% A2 sialylation or between about 75 and 89% A2 sialylation. Thus, the total sialic acid content is at least 4.5 moles sialic acid per mole of rhEPO, more specifically, from about 4.6 to 5.7 mole of sialic acid per mole of rhEPO.

Table 11				
	BPP (2000L) (n=3)	FPP (800L) (n=2)	Avecia (15L) (n=2)	Avecia (100L) (n=1)
Purity by SDS PAGE (EPO related) *(≥ 95.0%)*	99.5 ± 0.4%	99.4 ± 0.0%	99.4 ± 0.1%	99.4%
Integrity by SDS PAGE (Mono-PEG) *(≥ 80.0%)*	96.8 ± 0.7%	96.0 ± 2.2%	95.2 ± 2.0%	97.7%
Total sialic acid *(≥ 4.5 mol SA / mol protein)*	5.0-5.7	4.6-4.7	5.1-5.2	5.2
N-Linked glycan by CE *(70-85 % A2)*	75.2-80.2%	74.2-77.8%	80.9-88.7%	83.9%
A2 – bi-sialylated CE – capillary electrophoresis SA – sialic acid BPP- Biologics Pilot Plant FPP – Fermentation Pilot Plant				

A comparison of the pharmacokinetics of the rhEPO produced in the **YGLY7117** produced in Example 3 with all four *BMT* genes disrupted or deleted and PEGylated was compared to PEGylated rhEPO produced from strain **YGLY3159**. The comparison showed that the PEGylated rhEPO produced in strain **YGLY7117** had *in vivo* half-life and *in vivo* potency similar to that of **YGLY3159** and ARANESP (*See* Tables 12 and 13).

Table 12		
PK of rhEPO from **YGLY3159** (CBA) vs **YGLY7117** (no CBA)		
	YGLY3159	**YGLY7117**
T1/2 (hr)	20.9±2	20.6±4
CBA - cross-binding activity		

Table 13		
rhEPO source	Relative Potency (Reticulocyte Production)	95% Confidence Interval
YGLY3159 vs **YGLY7117**	0.94	(0.77, 1.14)

SEQUENCES

Sequences that were used to produce some of the strains disclosed in Examples 1-11 are provided in Table 14.

Table 14

SEQ ID NO:	Description	Sequence
1	*S. cerevisiae* invertase gene (ScSUC2)	AGGCCTCGCAACAACCTATAATTGAGTTAAGTGCCTTTCCAAGCT AAAAAGTTTGAGGTTATAGGGGCTTAGCATCCACACGTCACAAT CTCGGGTATCGAGTATAGTATGTAGAATTACGGCAGGAGGTTTC CCAATGAACAAAGGACAGGGGCACGGTGAGCTGTCGAAGGTATC CATTTTATCATGTTTCGTTTGTACAAGCACGACATACTAAGACAT TTACCGTATGGGAGTTGTTGTCCTAGCGTAGTTCTCGCTCCCCCA GCAAAGCTCAAAAAAGTACGTCATTTAGAATAGTTTGTGAGCAA ATTACCAGTCGGTATGCTACGTTAGAAAGGCCCACAGTATTCTTC TACCAAAGGCGTGCCTTTGTTGAACTCGATCCATTATGAGGGCTT CCATTATTCCCCGCATTTTTATTACTCTGAACAGGAATAAAAAGA AAAAACCCAGTTTAGGAAATTATCCGGGGGCGAAGAAATACGCG TAGCGTTAATCGACCCCACGTCCAGGGTTTTTCCATGGAGGTTTC TGGAAAAACTGACGAGGAATGTGATTATAAATCCCTTTATGTGA TGTCTAAGACTTTTAAGGTACGCCCGATGTTTGCCTATTACCATC ATAGAGACGTTTCTTTTCGAGGAATGCTTAAACGACTTTGTTTGA CAAAAATGTTGCCTAAGGGCTCTATAGTAAACCATTTGGAAGAA AGATTTGACGACTTTTTTTTTTGGATTTCGATCCTATAATCCTTC CTCCTGAAAAGAAACATATAAATAGATATGTATTATTCTTCAAAA CATTCTCTTGTTCTTGTGCTTTTTTTTTACCATATATCTTACTTTTT TTTTTCTCTCAGAGAAACAAGCAAAACAAAAAGCTTTTCTTTTCA CTAACGTATATG**ATG**CTTTTGCAAGCTTTCCTTTTCCTTTTGGCTG GTTTTGCAGCCAAAATATCTGCATCAATGACAAACGAAACTAGC GATAGACCTTTGGTCCACTTCACACCCAACAAGGGCTGGATGAA TGACCCAAATGGGTTGTGGTACGATGAAAAAGATGCCAAATGGC ATCTGTACTTTCAATACAACCCAAATGACACCGTATGGGGTACG CCATTGTTTTGGGGCCATGCTACTTCCGATGATTTGACTAATTGG GAAGATCAACCCATTGCTATCGCTCCCAAGCGTAACGATTCAGG TGCTTTCTCTGGCTCCATGGTGGTTGATTACAACAACACGAGTGG GTTTTTCAATGATACTATTGATCCAAGACAAAGATGCGTTGCGAT TTGGACTTATAACACTCCTGAAAGTGAAGAGCAATACATTAGCT ATTCTCTTGATGGTGGTTACACTTTTACTGAATACCAAAAGAACC CTGTTTTAGCTGCCAACTCCACTCAATTCAGAGATCCAAAGGTGT TCTGGTATGAACCTTCTCAAAAATGGATTATGACGGCTGCCAAAT CACAAGACTACAAAATTGAAATTTACTCCTCTGATGACTTGAAGT CCTGGAAGCTAGAATCTGCATTTGCCAATGAAGGTTTCTTAGGCT ACCAATACGAATGTCCAGGTTTGATTGAAGTCCCAACTGAGCAA GATCCTTCCAAATCTTATTGGGTCATGTTTATTTCTATCAACCCA GGTGCACCTGCTGGCGGTTCCTTCAACCAATATTTTGTTGGATCC TTCAATGGTACTCATTTTGAAGCGTTTGACAATCAATCTAGAGTG

		GTAGATTTTGGTAAGGACTACTATGCCTTGCAAACTTTCTTCAAC ACTGACCCAACCTACGGTTCAGCATTAGGTATTGCCTGGGCTTCA AACTGGGAGTACAGTGCCTTTGTCCCAACTAACCCATGGAGATC ATCCATGTCTTTGGTCCGCAAGTTTTCTTTGAACACTGAATATCA AGCTAATCCAGAGACTGAATTGATCAATTTGAAAGCCGAACCAA TATTGAACATTAGTAATGCTGGTCCCTGGTCTCGTTTTGCTACTA ACACAACTCTAACTAAGGCCAATTCTTACAATGTCGATTTGAGCA ACTCGACTGGTACCCTAGAGTTTGAGTTGGTTTACGCTGTTAACA CCACACAAACCATATCCAAATCCGTCTTTGCCGACTTATCACTTT GGTTCAAGGGTTTAGAAGATCCTGAAGAATATTTGAGAATGGGT TTTGAAGTCAGTGCTTCTTCCTTCTTTTTGGACCGTGGTAACTCTA AGGTCAAGTTTGTCAAGGAGAACCCATATTTCACAAACAGAATG TCTGTCAACAACCAACCATTCAAGTCTGAGAACGACCTAAGTTA CTATAAAGTGTACGGCCTACTGGATCAAAACATCTTGGAATTGTA CTTCAACGATGGAGATGTGGTTTCTACAAATACCTACTTCATGAC CACCGGTAACGCTCTAGGATCTGTGAACATGACCACTGGTGTCG ATAATTTGTTCTACATTGACAAGTTCCAAGTAAGGGAAGTAAAA **TAG**AGGTTATAAAACTTATTGTCTTTTTTATTTTTTTCAAAAGCCA TTCTAAAGGGCTTTAGCTAACGAGTGACGAATGTAAAACTTTATG ATTTCAAAGAATACCTCCAAACCATTGAAAATGTATTTTTATTTT TATTTTCTCCCGACCCCAGTTACCTGGAATTTGTTCTTTATGTACT TTATATAAGTATAATTCTCTTAAAAATTTTTACTACTTTGCAATAG ACATCATTTTTTCACGTAATAAACCCACAATCGTAATGTAGTTGC CTTACACTACTAGGATGGACCTTTTTGCCTTTATCTGTTTTGTTAC TGACACAATGAAACCGGGTAAAGTATTAGTTATGTGAAAATTTA AAAGCATTAAGTAGAAGTATACCATATTGTAAAAAAAAAAAAGCG TTGTCTTCTACGTAAAAGTGTTCTCAAAAAGAAGTAGTGAGGGA AATGGATACCAAGCTATCTGTAACAGGAGCTAAAAAATCTCAGG GAAAAGCTTCTGGTTTGGGAAACGGTCGAC
2	*S. cerevisiae* invertase (ScSUC2)	MLLQAFLFLLAGFAAKISASMTNETSDRPLVHFTPNKGWMNDPNGL WYDEKDAKWHLYFQYNPNDTVWGTPLFWGHATSDDLTNWEDQPI AIAPKRNDSGAFSGSMVVDYNNTSGFFNDTIDPRQRCVAIWTYNTP ESEEQYISYSLDGGYTFTEYQKNPVLAANSTQFRDPKVFWYEPSQK WIMTAAKSQDYKIEIYSSDDLKSWKLESAFANEGFLGYQYECPGLIE VPTEQDPSKSYWVMFISINPGAPAGGSFNQYFVGSFNGTHFEAFDNQ SRVVDFGKDYYALQTFFNTDPTYGSALGIAWASNWEYSAFVPTNP WRSSMSLVRKFSLNTEYQANPETELINLKAEPILNISNAGPWSRFAT NTTLTKANSYNVDLSNSTGTLEFELVYAVNTTQTISKSVFADLSLWF KGLEDPEEYLRMGFEVSASSFFLDRGNSKVKFVKENPYFTNRMSVN NQPFKSENDLSYYKVYGLLDQNILELYFNDGDVVSTNTYFMTTGNA LGSVNMTTGVDNLFYIDKFQVREVK
3	*K. lactis* UDP-GlcNAc transporter gene (KIMNN2-2)	AAACGTAACGCCTGGCACTCTATTTTCTCAAACTTCTGGGACGGA AGAGCTAAATATTGTGTTGCTTGAACAAACCCAAAAAAACAAAA AAATGAACAAACTAAAACTACACCTAAATAAACCGTGTGTAAAA CGTAGTACCATATTACTAGAAAAGATCACAAGTGTATCACACAT GTGCATCTCATATTACATCTTTTATCCAATCCATTCTCTCTATCCC GTCTGTTCCTGTCAGATTCTTTTTCCATAAAAAGAAGAAGACCCC GAATCTCACCGGTACAATGCAAAACTGCTGAAAAAAAAAGAAA GTTCACTGGATACGGGAACAGTGCCAGTAGGCTTCACCACATGG

		ACAAAACAATTGACGATAAAATAAGCAGGTGAGCTTCTTTTTCA AGTCACGATCCCTTTATGTCTCAGAAACAATATATACAAGCTAA ACCCTTTTGAACCAGTTCTCTCTTCATAGTTATGTTCACATAAATT GCGGGAACAAGACTCCGCTGGCTGTCAGGTACACGTTGTAACGT TTTCGTCCGCCCAATTATTAGCACAACATTGGCAAAAAGAAAAA CTGCTCGTTTTCTCTACAGGTAAATTACAATTTTTTTCAGTAATTT TCGCTGAAAAATTTAAAGGGCAGGAAAAAAAGACGATCTCGACT TTGCATAGATGCAAGAACTGTGGTCAAAACTTGAAATAGTAATT TTGCTGTGCGTGAACTAATAAATATATATATATATATATATAT ATTTGTGTATTTTGTATATGTAATTGTGCACGTCTTGGCTATTGGA TATAAGATTTTCGCGGGTTGATGACATAGAGCGTGTACTACTGTA ATAGTTGTATATTCAAAAGCTGCTGCGTGGAGAAAGACTAAAAT AGATAAAAAGCACACATTTTGACTTCGGTACCGTCAACTTAGTG GGACAGTCTTTTATATTTGGTGTAAGCTCATTTCTGGTACTATTCG AAACAGAACAGTGTTTTCTGTATTACCGTCCAATCGTTTGTC**ATG** AGTTTTGTATTGATTTTGTCGTTAGTGTTCGGAGGATGTTGTTCCA ATGTGATTAGTTTCGAGCACATGGTGCAAGGCAGCAATATAAAT TTGGGAAATATTGTTACATTCACTCAATTCGTGTCTGTGACGCTA ATTCAGTTGCCCAATGCTTTGGACTTCTCTCACTTTCCGTTTAGGT TGCGACCTAGACACATTCCTCTTAAGATCCATATGTTAGCTGTGT TTTTGTTCTTTACCAGTTCAGTCGCCAATAACAGTGTGTTTAAATT TGACATTTCCGTTCCGATTCATATTATCATTAGATTTTCAGGTACC ACTTTGACGATGATAATAGGTTGGGCTGTTTGTAATAAGAGGTAC TCCAAACTTCAGGTGCAATCTGCCATCATTATGACGCTTGGTGCG ATTGTCGCATCATTATACCGTGACAAAGAATTTTCAATGGACAGT TTAAAGTTGAATACGGATTCAGTGGGTATGACCCAAAAATCTAT GTTTGGTATCTTTGTTGTGCTAGTGGCCACTGCCTTGATGTCATTG TTGTCGTTGCTCAACGAATGGACGTATAACAAGTACGGGAAACA TTGGAAAGAAACTTTGTTCTATTCGCATTTCTTGGCTCTACCGTT GTTTATGTTGGGGTACACAAGGCTCAGAGACGAATTCAGAGACC TCTTAATTTCCTCAGACTCAATGGATATTCCTATTGTTAAATTACC AATTGCTACGAAACTTTTCATGCTAATAGCAAATAACGTGACCC AGTTCATTTGTATCAAAGGTGTTAACATGCTAGCTAGTAACACGG ATGCTTTGACACTTTCTGTCGTGCTTCTAGTGCGTAAATTTGTTAG TCTTTTACTCAGTGTCTACATCTACAAGAACGTCCTATCCGTGAC TGCATACCTAGGGACCATCACCGTGTTCCTGGGAGCTGGTTTGTA TTCATATGGTTCGGTCAAAACTGCACTGCCTCGC**TGA**AACAATCC ACGTCTGTATGATACTCGTTTCAGAATTTTTTTGATTTTCTGCCGG ATATGGTTTCTCATCTTTACAATCGCATTCTTAATTATACCAGAA CGTAATTCAATGATCCCAGTGACTCGTAACTCTTATATGTCAATT TAAGC
4	*K. lactis* UDP-GlcNAc transporter (KlMNN2-2)	MSFVLILSLVFGGCCSNVISFEHMVQGSNINLGNIVTFTQFVSVTLIQL PNALDFSHFPFRLRPRHIPLKIHMLAVFLFFTSSVANNSVFKFDISVPI HIIIRFSGTTLTMIIGWAVCNKRYSKLQVQSAIIMTLGAIVASLYRDKE FSMDSLKLNTDSVGMTQKSMFGIFVVLVATALMSLLSLLNEWTYNK YGKHWKETLFYSHFLALPLFMLGYTRLRDEFRDLLISSDSMDIPIVKL PIATKLFMLIANNVTQFICIKGVNMLASNTDALTLSVVLLVRKFVSLL LSVYIYKNVLSVTAYLGTITVFLGAGLYSYGSVKTALPR
5	DNA	ATGCTGCTTACCAAAAGGTTTTCAAAGCTGTTCAAGCTGACGTTC

	encodes Mnn2 leader (53)	ATAGTTTTGATATTGTGCGGGCTGTTCGTCATTACAAACAAATAC ATGGATGAGAACACGTCG
6	Mnn2 leader (53)	MLLTKRFSKLFKLTFIVLILCGLFVITNKYMDENTS
7	DNA encodes Mnn2 leader (54) The last 9 nucleotides are the linker containing the AscI restriction site)	ATGCTGCTTACCAAAAGGTTTTCAAAGCTGTTCAAGCTGACGTTC ATAGTTTTGATATTGTGCGGGCTGTTCGTCATTACAAACAAATAC ATGGATGAGAACACGTCGGTCAAGGAGTACAAGGAGTACTTAGA CAGATATGTCCAGAGTTACTCCAATAAGTATTCATCTTCCTCAGA CGCCGCCAGCGCTGACGATTCAACCCCATTGAGGGACAATGATG AGGCAGGCAATGAAAAGTTGAAAAGCTTCTACAACAACGTTTTC AACTTTCTAATGGTTGATTCGCCCGGGCGCGCC
8	Mnn2 leader (54)	MLLTKRFSKLFKLTFIVLILCGLFVITNKYMDENTSVKEYKEYLDRY VQSYSNKYSSSSD AASADDSTPLRDNDEAGNEKLKSFYNNVFNFLMVDSPGRA
9	DNA encodes *S. cerevisiae* Mating Factor pre signal sequence	ATG AGA TTC CCA TCC ATC TTC ACT GCT GTT TTG TTC GCT GCT TCT TCT GCT TTG GCT
10	*S. cerevisiae* Mating Factor pre signal sequence	MRFPSIFTAVLFAASSALA
11	DNA encodes Pp SEC12 (10) The last 9 nucleotides are the linker containing the AscI restriction site used for fusion to proteins of interest.	ATGCCCAGAAAAATATTTAACTACTTCATTTTGACTGTATTCATG GCAATTCTTGCTATTGTTTTACAATGGTCTATAGAGAATGGACAT GGGCGCGCC
12	Pp SEC12	MPRKIFNYFILTVFMAILAIVLQWSIENGHGRA

	(10)	
13	DNA encodes ScMnt1 (Kre2) (33)	ATGGCCCTCTTTCTCAGTAAGAGACTGTTGAGATTTACCGTCATT GCAGGTGCGGTTATTGTTCTCCTCCTAACATTGAATTCCAACAGT AGAACTCAGCAATATATTCCGAGTTCCATCTCCGCTGCATTTGAT TTTACCTCAGGATCTATATCCCCTGAACAACAAGTCATCGGGCGC GCC
14	ScMnt1 (Kre2) (33)	MALFLSKRLLRFTVIAGAVIVLLLTLNSNSRTQQYIPSSISAAFDFTSG SISPEQQVIGRA
15	DNA encodes ScSEC12 (8) The last 9 nucleotides are the linker containing the AscI restriction site used for fusion to proteins of interest	ATGAACACTATCCACATAATAAAATTACCGCTTAACTACGCCAA CTACACCTCAATGAAACAAAAAATCTCTAAATTTTTCACCAACTT CATCCTTATTGTGCTGCTTTCTTACATTTTACAGTTCTCCTATAAG CACAATTTGCATTCCATGCTTTTCAATTACGCGAAGGACAATTTT CTAACGAAAAGAGACACCATCTCTTCGCCCTACGTAGTTGATGA AGACTTACATCAAACAACTTTGTTTGGCAACCACGGTACAAAAA CATCTGTACCTAGCGTAGATTCCATAAAAGTGCATGGCGTGGGG CGCGCC
16	ScSEC12 (8)	MNTIHIIKLPLNYANYTSMKQKISKFFTNFILIVLLSYILQFSYKHNLH SMLFNYAKDNFLTKRDTISSPYVVDEDLHQTTLFGNHGTKTSVPSV DSIKVHGVGRA
17	DNA encodes MmSLC35 A3 UDP-GlcNAc transporter	ATGTCTGCCAACCTAAAATATCTTTCCTTGGGAATTTTGGTGTTT CAGACTACCAGTCTGGTTCTAACGATGCGGTATTCTAGGACTTTA AAAGAGGAGGGGCCTCGTTATCTGTCTTCTACAGCAGTGGTTGTG GCTGAATTTTTGAAGATAATGGCCTGCATCTTTTTAGTCTACAAA GACAGTAAGTGTAGTGTGAGAGCACTGAATAGAGTACTGCATGA TGAAATTCTTAATAAGCCCATGGAAACCCTGAAGCTCGCTATCCC GTCAGGGATATATACTCTTCAGAACAACTTACTCTATGTGGCACT GTCAAACCTAGATGCAGCCACTTACCAGGTTACATATCAGTTGA AAATACTTACAACAGCATTATTTTCTGTGTCTATGCTTGGTAAAA AATTAGGTGTGTACCAGTGGCTCTCCCTAGTAATTCTGATGGCAG GAGTTGCTTTTGTACAGTGGCCTTCAGATTCTCAAGAGCTGAACT CTAAGGACCTTTCAACAGGCTCACAGTTTGTAGGCCTCATGGCA GTTCTCACAGCCTGTTTTTCAAGTGGCTTTGCTGGAGTTTATTTTG AGAAAATCTTAAAAGAAACAAAACAGTCAGTATGGATAAGGAA CATTCAACTTGGTTTCTTTGGAAGTATATTTGGATTAATGGGTGT ATACGTTTATGATGGAGAATTGGTCTCAAAGAATGGATTTTTTCA GGGATATAATCAACTGACGTGGATAGTTGTTGCTCTGCAGGCACT TGGAGGCCTTGTAATAGCTGCTGTCATCAAATATGCAGATAACAT TTTAAAAGGATTTGCGACCTCCTTATCCATAATATTGTCAACAAT AATATCTTATTTTTGGTTGCAAGATTTTGTGCCAACCAGTGTCTTT TTCCTTGGAGCCATCCTTGTAATAGCAGCTACTTTCTTGTATGGTT ACGATCCCAAACCTGCAGGAAATCCCACTAAAGCATAG
18	MmSLC35	MSANLKYLSLGILVFQTTSLVLTMRYSRTLKEEGPRYLSSTAVVVAE

	A3 UDP-GlcNAc transporter	FLKIMACIFLVYKDSKCSVRALNRVLHDEILNKPMETLKLAIPSGIYT LQNNLLYVALSNLDAATYQVTYQLKILTTALFSVSMLGKKLGVYQ WLSLVILMAGVAFVQWPSDSQELNSKDLSTGSQFVGLMAVLTACFS SGFAGVYFEKILKETKQSVWIRNIQLGFFGSIFGLMGVYVVYDGELVS KNGFFQGYNQLTWIVVALQALGGLVIAAVIKYADNILKGFATSLSIIL STIISYFWLQDFVPTSVFFLGAILVIAATFLYGYDPKPAGNPTKA
19	DNA encodes DmUGT	ATGAATAGCATACACATGAACGCCAATACGCTGAAGTACATCAG CCTGCTGACGCTGACCCTGCAGAATGCCATCCTGGGCCTCAGCAT GCGCTACGCCCGCACCCGGCCAGGCGACATCTTCCTCAGCTCCA CGGCCGTACTCATGGCAGAGTTCGCCAAACTGATCACGTGCCTG TTCCTGGTCTTCAACGAGGAGGGCAAGGATGCCCAGAAGTTTGT ACGCTCGCTGCACAAGACCATCATTGCGAATCCCATGGACACGC TGAAGGTGTGCGTCCCCTCGCTGGTCTATATCGTTCAAAACAATC TGCTGTACGTCTCTGCCTCCCATTTGGATGCGGCCACCTACCAGG TGACGTACCAGCTGAAGATTCTCACCACGGCCATGTTCGCGGTTG TCATTCTGCGCCGCAAGCTGCTGAACACGCAGTGGGGTGCGCTG CTGCTCCTGGTGATGGGCATCGTCCTGGTGCAGTTGGCCCAAACG GAGGGTCCGACGAGTGGCTCAGCCGGTGGTGCCGCAGCTGCAGC CACGGCCGCCTCCTCTGGCGGTGCTCCCGAGCAGAACAGGATGC TCGGACTGTGGGCCGCACTGGGCGCCTGCTTCCTCTCCGGATTCG CGGGCATCTACTTTGAGAAGATCCTCAAGGGTGCCGAGATCTCC GTGTGGATGCGGAATGTGCAGTTGAGTCTGCTCAGCATTCCCTTC GGCCTGCTCACCTGTTTCGTTAACGACGGCAGTAGGATCTTCGAC CAGGGATTCTTCAAGGGCTACGATCTGTTTGTCTGGTACCTGGTC CTGCTGCAGGCCGGCGGTGGATTGATCGTTGCCGTGGTGGTCAA GTACGCGGATAACATTCTCAAGGGCTTCGCCACCTCGCTGGCCAT CATCATCTCGTGCGTGGCCTCCATATACATCTTCGACTTCAATCT CACGCTGCAGTTCAGCTTCGGAGCTGGCCTGGTCATCGCCTCCAT ATTTCTCTACGGCTACGATCCGGCCAGGTCGGCGCCGAAGCCAA CTATGCATGGTCCTGGCGGCGATGAGGAGAAGCTGCTGCCGCGC GTCTAG
20	DmUGT	MNSIHMNANTLKYISLLTLTLQNAILGLSMRYARTRPGDIFLSSTAVL MAEFAKLITCLFLVFNEEGKDAQKFVRSLHKTIIANPMDTLKVCVPS LVYIVQNNLLYVSASHLDAATYQVTYQLKILTTAMFAVVILRRKLL NTQWGALLLLVMGIVLVQLAQTEGPTSGSAGGAAAAATAASSGGA PEQNRMLGLWAALGACFLSGFAGIYFEKILKGAEISVWMRNVQLSL LSIPFGLLTCFVNDGSRIFDQGFFKGYDLFVWYLVLLQAGGGLIVAV VVKYADNILKGFATSLAIIISCVASIYIFDFNLTLQFSFGAGLVIASIFLY GYDPARSAPKPTMHGPGGDEEKLLPRV
21	DNA encodes ScGAL10	ATGACAGCTCAGTTACAAAGTGAAAGTACTTCTAAAATTGTTTTG GTTACAGGTGGTGCTGGATACATTGGTTCACACACTGTGGTAGA GCTAATTGAGAATGGATATGACTGTGTTGTTGCTGATAACCTGTC GAATTCAACTTATGATTCTGTAGCCAGGTTAGAGGTCTTGACCAA GCATCACATTCCCTTCTATGAGGTTGATTTGTGTGACCGAAAAGG TCTGGAAAAGGTTTTCAAAGAATATAAAATTGATTCGGTAATTCA CTTTGCTGGTTTAAAGGCTGTAGGTGAATCTACACAAATCCCGCT GAGATACTATCACAATAACATTTTGGGAACTGTCGTTTTATTAGA GTTAATGCAACAATACAACGTTTCCAAATTTGTTTTTTCATCTTCT GCTACTGTCTATGGTGATGCTACGAGATTCCCAAATATGATTCCT

		ATCCCAGAAGAATGTCCCTTAGGGCCTACTAATCCGTATGGTCAT ACGAAATACGCCATTGAGAATATCTTGAATGATCTTTACAATAGC GACAAAAAAAGTTGGAAGTTTGCTATCTTGCGTTATTTTAACCCA ATTGGCGCACATCCCTCTGGATTAATCGGAGAAGATCCGCTAGG TATACCAAACAATTTGTTGCCATATATGGCTCAAGTAGCTGTTGG TAGGCGCGAGAAGCTTTACATCTTCGGAGACGATTATGATTCCA GAGATGGTACCCCGATCAGGGATTATATCCACGTAGTTGATCTA GCAAAAGGTCATATTGCAGCCCTGCAATACCTAGAGGCCTACAA TGAAAATGAAGGTTTGTGTCGTGAGTGGAACTTGGGTTCCGGTA AAGGTTCTACAGTTTTTGAAGTTTATCATGCATTCTGCAAAGCTT CTGGTATTGATCTTCCATACAAAGTTACGGGCAGAAGAGCAGGT GATGTTTTGAACTTGACGGCTAAACCAGATAGGGCCAAACGCGA ACTGAAATGGCAGACCGAGTTGCAGGTTGAAGACTCCTGCAAGG ATTTATGGAAATGGACTACTGAGAATCCTTTTGGTTACCAGTTAA GGGGTGTCGAGGCCAGATTTTCCGCTGAAGATATGCGTTATGAC GCAAGATTTGTGACTATTGGTGCCGGCACCAGATTTCAAGCCAC GTTTGCCAATTTGGGCGCCAGCATTGTTGACCTGAAAGTGAACG GACAATCAGTTGTTCTTGGCTATGAAAATGAGGAAGGGTATTTG AATCCTGATAGTGCTTATATAGGCGCCACGATCGGCAGGTATGC TAATCGTATTTCGAAGGGTAAGTTTAGTTTATGCAACAAAGACTA TCAGTTAACCGTTAATAACGGCGTTAATGCGAATCATAGTAGTAT CGGTTCTTTCCACAGAAAAAGATTTTTGGGACCCATCATTCAAAA TCCTTCAAAGGATGTTTTTACCGCCGAGTACATGCTGATAGATAA TGAGAAGGACACCGAATTTCCAGGTGATCTATTGGTAACCATAC AGTATACTGTGAACGTTGCCCAAAAAAGTTTGGAAATGGTATAT AAAGGTAAATTGACTGCTGGTGAAGCGACGCCAATAAATTTAAC AAATCATAGTTATTTCAATCTGAACAAGCCATATGGAGACACTAT TGAGGGTACGGAGATTATGGTGCGTTCAAAAAAATCTGTTGATG TCGACAAAAACATGATTCCTACGGGTAATATCGTCGATAGAGAA ATTGCTACCTTTAACTCTACAAAGCCAACGGTCTTAGGCCCCAAA AATCCCCAGTTTGATTGTTGTTTTGTGGTGGATGAAAATGCTAAG CCAAGTCAAATCAATACTCTAAACAATGAATTGACGCTTATTGTC AAGGCTTTTCATCCCGATTCCAATATTACATTAGAAGTTTTAAGT ACAGAGCCAACTTATCAATTTTATACCGGTGATTTCTTGTCTGCT GGTTACGAAGCAAGACAAGGTTTTGCAATTGAGCCTGGTAGATA CATTGATGCTATCAATCAAGAGAACTGGAAAGATTGTGTAACCT TGAAAAACGGTGAAACTTACGGGTCCAAGATTGTCTACAGATTT TCCTGA
22	ScGal10	MTAQLQSESTSKIVLVTGGAGYIGSHTVVELIENGYDCVVADNLSNS TYDSVARLEVLTKHHIPFYEVDLCDRKGLEKVFKEYKIDSVIHFAGL KAVGESTQIPLRYYHNNILGTVVLLELMQQYNVSKFVFSSSATVYG DATRFPNMIPIPEECPLGPTNPYGHTKYAIENILNDLYNSDKKSWKFA ILRYFNPIGAHPSGLIGEDPLGIPNNLLPYMAQVAVGRREKLYIFGDD YDSRDGTPIRDYIHVVDLAKGHIAALQYLEAYNENEGLCREWNLGS GKGSTVFEVYHAFCKASGIDLPYKVTGRRAGDVLNLTAKPDRAKRE LKWQTELQVEDSCKDLWKWTTENPFGYQLRGVEARFSAEDMRYD ARFVTIGAGTRFQATFANLG ASIVDLKVNGQSVVLGYENEEGYLNPDSAYIGATIGRYANRISKGKF SLCNKDYQLTVNNGVNANHSSIGSFHRKRFLGPIIQNPSKDVFTAEY

		MLIDNEKDTEFPGDLLVTIQYTVNVAQKSLEMVYKGKLTAGEATPI NLTNHSYFNLNKPYGDTIEGTEIMVRSKKSVDVDKNMIPTGNIVDRE IATFNSTKPTVLGPKNPQFDCCFVVDENAKPSQINTLNNELTLIVKAF HPDSNITLEVLSTEPTYQFYTGDFLSAGYEARQGFAIEPGRYIDAINQ ENWKDCVTLKNGETYGSKIVYRFS
23	hGalT codon optimized (XB)	GGTAGAGATTTGTCTAGATTGCCACAGTTGGTTGGTGTTTCCACT CCATTGCAAGGAGGTTCTAACTCTGCTGCTGCTATTGGTCAATCT TCCGGTGAGTTGAGAACTGGTGGAGCTAGACCACCTCCACCATT GGGAGCTTCCTCTCAACCAAGACCAGGTGGTGATTCTTCTCCAGT TGTTGACTCTGGTCCAGGTCCAGCTTCTAACTTGACTTCCGTTCC AGTTCCACACACTACTGCTTTGTCCTTGCCAGCTTGTCCAGAAGA ATCCCCATTGTTGGTTGGTCCAATGTTGATCGAGTTCAACATGCC AGTTGACTTGGAGTTGGTTGCTAAGCAGAACCCAAACGTTAAGA TGGGTGGTAGATACGCTCCAAGAGACTGTGTTTCCCCACACAAA GTTGCTATCATCATCCCATTCAGAAACAGACAGGAGCACTTGAA GTACTGGTTGTACTACTTGCACCCAGTTTTGCAAAGACAGCAGTT GGACTACGGTATCTACGTTATCAACCAGGCTGGTGACACTATTTT CAACAGAGCTAAGTTGTTGAATGTTGGTTTCCAGGAGGCTTTGAA GGATTACGACTACACTTGTTTCGTTTTCTCCGACGTTGACTTGATT CCAATGAACGACCACAACGCTTACAGATGTTTCTCCCAGCCAAG ACACATTTCTGTTGCTATGGACAAGTTCGGTTTCTCCTTGCCATA CGTTCAATACTTCGGTGGTGTTTCCGCTTTGTCCAAGCAGCAGTT CTTGACTATCAACGGTTTCCCAAACAATTACTGGGGATGGGGTG GTGAAGATGACGACATCTTTAACAGATTGGTTTTCAGAGGAATG TCCATCTCTAGACCAAACGCTGTTGTTGGTAGATGTAGAATGATC AGACACTCCAGAGACAAGAAGAACGAGCCAAACCCACAAAGAT TCGACAGAATCGCTCACACTAAGGAAACTATGTTGTCCGACGGA TTGAACTCCTTGACTTACCAGGTTTTGGACGTTCAGAGATACCCA TTGTACACTCAGATCACTGTTGACATCGGTACTCCATCCTAG
24	hGalT I catalytic doman (XB)	GRDLSRLPQLVGVSTPLQGGSNSAAAIGQSSGELRTGGARPPPPLGA SSQPRPGGDSSPVVDSGPGPASNLTSVPVPHTTALSLPACPEESPLLV GPMLIEFNMPVDLELVAKQNPNVKMGGRYAPRDCVSPHKVAIIIPFR NRQEHLKYWLYYLHPVLQRQQLDYGIYVINQAGDTIFNRAKLLNVG FQEALKDYDYTCFVFSDVDLIPMNDHNAYRCFSQPRHISVAMDKFG FSLPYVQYFGGVSALSKQQFLTINGFPNNYWGWGGEDDDIFNRLVF RGMSISRPNAVVGRCRMIRHSRDKKNEPNPQRFDRIAHTKETMLSD GLNSLTYQVLDVQRYPLYTQITVDIGTPS
25	DNA encodes human GnTI catalytic doman (NA) Codon-optimized	TCAGTCAGTGCTCTTGATGGTGACCCAGCAAGTTTGACCAGAGA AGTGATTAGATTGGCCCAAGACGCAGAGGTGGAGTTGGAGAGAC AACGTGGACTGCTGCAGCAAATCGGAGATGCATTGTCTAGTCAA AGAGGTAGGGTGCCTACCGCAGCTCCTCCAGCACAGCCTAGAGT GCATGTGACCCCTGCACCAGCTGTGATTCCTATCTTGGTCATCGC CTGTGACAGATCTACTGTTAGAAGATGTCTGGACAAGCTGTTGCA TTACAGACCATCTGCTGAGTTGTTCCCTATCATCGTTAGTCAAGA CTGTGGTCACGAGGAGACTGCCCAAGCCATCGCCTCCTACGGAT CTGCTGTCACTCACATCAGACAGCCTGACCTGTCATCTATTGCTG TGCCACCAGACCACAGAAAGTTCCAAGGTTACTACAAGATCGCT AGACACTACAGATGGGCATTGGGTCAAGTCTTCAGACAGTTTAG ATTCCCTGCTGCTGTGGTGGTGGAGGATGACTTGGAGGTGGCTCC

		TGACTTCTTTGAGTACTTTAGAGCAACCTATCCATTGCTGAAGGC AGACCCATCCCTGTGGTGTGTCTCTGCCTGGAATGACAACGGTA AGGAGCAAATGGTGGACGCTTCTAGGCCTGAGCTGTTGTACAGA ACCGACTTCTTTCCTGGTCTGGGATGGTTGCTGTTGGCTGAGTTG TGGGCTGAGTTGGAGCCTAAGTGGCCAAAGGCATTCTGGGACGA CTGGATGAGAAGACCTGAGCAAAGACAGGGTAGAGCCTGTATCA GACCTGAGATCTCAAGAACCATGACCTTTGGTAGAAAGGGAGTG TCTCACGGTCAATTCTTTGACCAACACTTGAAGTTTATCAAGCTG AACCAGCAATTTGTGCACTTCACCCAACTGGACCTGTCTTACTTG CAGAGAGAGGCCTATGACAGAGATTTCCTAGCTAGAGTCTACGG AGCTCCTCAACTGCAAGTGGAGAAAGTGAGGACCAATGACAGAA AGGAGTTGGGAGAGGTGAGAGTGCAGTACACTGGTAGGGACTCC TTTAAGGCTTTCGCTAAGGCTCTGGGTGTCATGGATGACCTTAAG TCTGGAGTTCCTAGAGCTGGTTACAGAGGTATTGTCACCTTTCAA TTCAGAGGTAGAAGAGTCCACTTGGCTCCTCCACCTACTTGGGA GGGTTATGATCCTTCTTGGAATTAG
26	Human GnT I catalytic doman (NA)	SVSALDGDPASLTREVIRLAQDAEVELERQRGLLQQIGDALSSQRGR VPTAAPPAQPRVHVTPAPAVIPILVIACDRSTVRRCLDKLLHYRPSAE LFPIIVSQDCGHEETAQAIASYGSAVTHIRQPDLSSIAVPPDHRKFQG YYKIARHYRWALGQVFRQFRFPAAVVVEDDLEVAPDFFEYFRATYP LLKADPSLWCVSAWNDNGKEQMVDASRPELLYRTDFFPGLGWLLL AELWAELEPKWPKAFWDDWMRRPEQRQGRACIRPEISRTMTFGRK GVSHGQFFDQHLKFIKLNQQFVHFTQLDLSYLQREAYDRDFLARVY GAPQLQVEKVRTNDRKELGEVRVQYTGRDSFKAFAKALGVMDDL KSGVPRAGYRGIVTFQFRGRRVHLAPPPTWEGYDPSWN
27	DNA encodes Mm ManI catalytic doman (FB)	GAGCCCGCTGACGCCACCATCCGTGAGAAGAGGGCAAAGATCA AAGAGATGATGACCCATGCTTGGAATAATTATAAACGCTATGCG TGGGGCTTGAACGAACTGAAACCTATATCAAAAGAAGGCCATTC AAGCAGTTTGTTTGGCAACATCAAAGGAGCTACAATAGTAGATG CCCTGGATACCCTTTTCATTATGGGCATGAAGACTGAATTTCAAG AAGCTAAATCGTGGATTAAAAAATATTTAGATTTTAATGTGAATG CTGAAGTTTCTGTTTTTGAAGTCAACATACGCTTCGTCGGTGGAC TGCTGTCAGCCTACTATTTGTCCGGAGAGGAGATATTTCGAAAGA AAGCAGTGGAACTTGGGGTAAAATTGCTACCTGCATTTCATACTC CCTCTGGAATACCTTGGGCATTGCTGAATATGAAAAGTGGGATC GGGCGGAACTGGCCCTGGGCCTCTGGAGGCAGCAGTATCCTGGC CGAATTTGGAACTCTGCATTTAGAGTTTATGCACTTGTCCCACTT ATCAGGAGACCCAGTCTTTGCCGAAAAGGTTATGAAAATTCGAA CAGTGTTGAACAAACTGGACAAACCAGAAGGCCTTTATCCTAAC TATCTGAACCCCAGTAGTGGACAGTGGGGTCAACATCATGTGTC GGTTGGAGGACTTGGAGACAGCTTTTATGAATATTTGCTTAAGGC GTGGTTAATGTCTGACAAGACAGATCTCGAAGCCAAGAAGATGT ATTTTGATGCTGTTCAGGCCATCGAGACTCACTTGATCCGCAAGT CAAGTGGGGGACTAACGTACATCGCAGAGTGGAAGGGGGGCCTC CTGGAACACAAGATGGGCCACCTGACGTGCTTTGCAGGAGGCAT GTTTGCACTTGGGGCAGATGGAGCTCCGGAAGCCCGGGCCCAAC ACTACCTTGAACTCGGAGCTGAAATTGCCCGCACTTGTCATGAAT CTTATAATCGTACATATGTGAAGTTGGGACCGGAAGCGTTTCGAT TTGATGGCGGTGTGGAAGCTATTGCCACGAGGCAAAATGAAAAG

		TATTACATCTTACGGCCCGAGGTCATCGAGACATACATGTACATG TGGCGACTGACTCACGACCCCAAGTACAGGACCTGGGCCTGGGA AGCCGTGGAGGCTCTAGAAAGTCACTGCAGAGTGAACGGAGGCT ACTCAGGCTTACGGGATGTTTACATTGCCCGTGAGAGTTATGACG ATGTCCAGCAAAGTTTCTTCCTGGCAGAGACACTGAAGTATTTGT ACTTGATATTTTCCGATGATGACCTTCTTCCACTAGAACACTGGA TCTTCAACACCGAGGCTCATCCTTTCCCTATACTCCGTGAACAGA AGAAGGAAATTGATGGCAAAGAGAAATGA
28	Mm ManI catalytic doman (FB)	EPADATIREKRAKIKEMMTHAWNNYKRYAWGLNELKPISKEGHSSS LFGNIKGATIVDALDTLFIMGMKTEFQEAKSWIKKYLDFNVNAEVS VFEVNIRFVGGLLSAYYLSGEEIFRKKAVELGVKLLPAFHTPSGIPWA LLNMKSGIGRNWPWASGGSSILAEFGTLHLEFMHLSHLSGDPVFAE KVMKIRTVLNKLDKPEGLYPNYLNPSSGQWGQHHVSVGGLGDSFY EYLLKAWLMSDKTDLEAKKMYFDAVQAIETHLIRKSSGGLTYIAEW KGGLLEHKMGHLTCFAGGMFALGADGAPEARAQHYLELGAEIART CHESYNRTYVKLGPEAFRFDGGVEAIATRQNEKYYILRPEVIETYMY MWRLTHDPKYRTWAWEAVEALESHCRVNGGYSGLRDVYIARESY DDVQQSFFLAETLKYLYLIFSDDDLLPLEHWIFNTEAHPFPILREQKK EIDGKEK
29	DNA encodes Tr ManI catalytic doman	CGCGCCGGATCTCCCAACCCTACGAGGGCGGCAGCAGTCAAGGC CGCATTCCAGACGTCGTGGAACGCTTACCACCATTTTGCCTTTCC CCATGACGACCTCCACCCGGTCAGCAACAGCTTTGATGATGAGA GAAACGGCTGGGGCTCGTCGGCAATCGATGGCTTGGACACGGCT ATCCTCATGGGGGATGCCGACATTGTGAACACGATCCTTCAGTAT GTACCGCAGATCAACTTCACCACGACTGCGGTTGCCAACCAAGG CATCTCCGTGTTCGAGACCAACATTCGGTACCTCGGTGGCCTGCT TTCTGCCTATGACCTGTTGCGAGGTCCTTTCAGCTCCTTGGCGAC AAACCAGACCCTGGTAAACAGCCTTCTGAGGCAGGCTCAAACAC TGGCCAACGGCCTCAAGGTTGCGTTCACCACTCCCAGCGGTGTCC CGGACCCTACCGTCTTCTTCAACCCTACTGTCCGGAGAAGTGGTG CATCTAGCAACAACGTCGCTGAAATTGGAAGCCTGGTGCTCGAG TGGACACGGTTGAGCGACCTGACGGGAAACCCGCAGTATGCCCA GCTTGCGCAGAAGGGCGAGTCGTATCTCCTGAATCCAAAGGGAA GCCCGGAGGCATGGCCTGGCCTGATTGGAACGTTTGTCAGCACG AGCAACGGTACCTTTCAGGATAGCAGCGGCAGCTGGTCCGGCCT CATGGACAGCTTCTACGAGTACCTGATCAAGATGTACCTGTACG ACCCGGTTGCGTTTGCACACTACAAGGATCGCTGGGTCCTTGCTG CCGACTCGACCATTGCGCATCTCGCCTCTCACCCGTCGACGCGCA AGGACTTGACCTTTTTGTCTTCGTACAACGGACAGTCTACGTCGC CAAACTCAGGACATTTGGCCAGTTTTGCCGGTGGCAACTTCATCT TGGGAGGCATTCTCCTGAACGAGCAAAAGTACATTGACTTTGGA ATCAAGCTTGCCAGCTCGTACTTTGCCACGTACAACCAGACGGCT TCTGGAATCGGCCCCGAAGGCTTCGCGTGGGTGGACAGCGTGAC GGGCGCCGGCGGCTCGCCGCCCTCGTCCCAGTCCGGGTTCTACTC GTCGGCAGGATTCTGGGTGACGGCACCGTATTACATCCTGCGGC CGGAGACGCTGGAGAGCTTGTACTACGCATACCGCGTCACGGGC GACTCCAAGTGGCAGGACCTGGCGTGGGAAGCGTTCAGTGCCAT TGAGGACGCATGCCGCGCCGGCAGCGCGTACTCGTCCATCAACG ACGTGACGCAGGCCAACGGCGGGGGTGCCTCTGACGATATGGAG

		AGCTTCTGGTTTGCCGAGGCGCTCAAGTATGCGTACCTGATCTTT GCGGAGGAGTCGGATGTGCAGGTGCAGGCCAACGGCGGGAACA AATTTGTCTTTAACACGGAGGCGCACCCCTTTAGCATCCGTTCAT CATCACGACGGGGCGGCCACCTTGCTTAA
30	Tr Man I catalytic doman	RAGSPNPTRAAAVKAAFQTSWNAYHHFAFPHDDLHPVSNSFDDER NGWGSSAIDGLDTAILMGDADIVNTILQYVPQINFTTTAVANQGISV FETNIRYLGGLLSAYDLLRGPFSSLATNQTLVNSLLRQAQTLANGLK VAFTTPSGVPDPTVFFNPTVRRSGASSNNVAEIGSLVLEWTRLSDLT GNPQYAQLAQKGESYLLNPKGSPEAWPGLIGTFVSTSNGTFQDSSGS WSGLMDSFYEYLIKMYLYDPVAFAHYKDRWVLAADSTIAHLASHP STRKDLTFLSSYNGQSTSPNSGHLASFAGGNFILGGILLNEQKYIDFGI KLASSYFATYNQTASGIGPEGFAWVDSVTGAGGSPPSSQSGFYSSAG FWVTAPYYILRPETLESLYYAYRVTGDSKWQDLAWEAFSAIEDACR AGSAYSSINDVTQANGGGASDDMESFWFAEALKYAYLIFAEESDVQ VQANGGNKFVFNTEAHPFSIRSSSRRGGHLA
31	DNA encodes Rat GnT II (TC) Codon-optimized	TCCTTGGTTTACCAATTGAACTTCGACCAGATGTTGAGAAACGTT GACAAGGACGGTACTTGGTCTCCTGGTGAGTTGGTTTTGGTTGTT CAGGTTCACAACAGACCAGAGTACTTGAGATTGTTGATCGACTC CTTGAGAAAGGCTCAAGGTATCAGAGAGGTTTTGGTTATCTTCTC CCACGATTTCTGGTCTGCTGAGATCAACTCCTTGATCTCCTCCGT TGACTTCTGTCCAGTTTTGCAGGTTTTCTTCCCATTCTCCATCCAA TTGTACCCATCTGAGTTCCCAGGTTCTGATCCAAGAGACTGTCCA AGAGACTTGAAGAAGAACGCTGCTTTGAAGTTGGGTTGTATCAA CGCTGAATACCCAGATTCTTTCGGTCACTACAGAGAGGCTAAGTT CTCCCAAACTAAGCATCATTGGTGGTGGAAGTTGCACTTTGTTTG GGAGAGAGTTAAGGTTTTGCAGGACTACACTGGATTGATCTTGTT CTTGGAGGAGGATCATTACTTGGCTCCAGACTTCTACCACGTTTT CAAGAAGATGTGGAAGTTGAAGCAACAAGAGTGTCCAGGTTGTG ACGTTTTGTCCTTGGGAACTTACACTACTATCAGATCCTTCTACG GTATCGCTGACAAGGTTGACGTTAAGACTTGGAAGTCCACTGAA CACAACATGGGATTGGCTTTGACTAGAGATGCTTACCAGAAGTT GATCGAGTGTACTGACACTTTCTGTACTTACGACGACTACAACTG GGACTGGACTTTGCAGTACTTGACTTTGGCTTGTTTGCCAAAAGT TTGGAAGGTTTTGGTTCCACAGGCTCCAAGAATTTTCCACGCTGG TGACTGTGGAATGCACCACAAGAAAACTTGTAGACCATCCACTC AGTCCGCTCAAATTGAGTCCTTGTTGAACAACAACAAGCAGTAC TTGTTCCCAGAGACTTTGGTTATCGGAGAGAAGTTTCCAATGGCT GCTATTTCCCCACCAAGAAAGAATGGTGGATGGGGTGATATTAG AGACCACGAGTTGTGTAAATCCTACAGAAGATTGCAGTAG
32	Rat GnTII (TC)	SLVYQLNFDQMLRNVDKDGTWSPGELVLVVQVHNRPEYLRLLIDSL RKAQGIREVLVIFSHDFWSAEINSLISSVDFCPVLQVFFPFSIQLYPSEF PGSDPRDCPRDLKKNAALKLGCINAEYPDSFGHYREAKFSQTKHHW WWKLHFVWERVKVLQDYTGLILFLEEDHYLAPDFYHVFKKMWKL KQQECPGCDVLSLGTYTTIRSFYGIADKVDVKTWKSTEHNMGLALT RDAYQKLIECTDTFCTYDDYNWDWTLQYLTLACLPKVWKVLVPQA PRIFHAGDCGMHHKKTCRPSTQSAQIESLLNNNKQYLFPETLVIGEK FPMAAISPPRKNGGWGDIRDHELCKSYRRLQ
33	DNA encodes	AGAGACGATCCAATTAGACCTCCATTGAAGGTTGCTAGATCCCC AAGACCAGGTCAATGTCAAGATGTTGTTCAGGACGTCCCAAACG

	Drosophila melanogaster ManII codon-optimized (KD)	TTGATGTCCAGATGTTGGAGTTGTACGATAGAATGTCCTTCAAGG ACATTGATGGTGGTGTTTGGAAGCAGGGTTGGAACATTAAGTAC GATCCATTGAAGTACAACGCTCATCACAAGTTGAAGGTCTTCGTT GTCCCACACTCCCACAACGATCCTGGTTGGATTCAGACCTTCGAG GAATACTACCAGCACGACACCAAGCACATCTTGTCCAACGCTTT GAGACATTTGCACGACAACCCAGAGATGAAGTTCATCTGGGCTG AAATCTCCTACTTCGCTAGATTCTACCACGATTTGGGTGAGAACA AGAAGTTGCAGATGAAGTCCATCGTCAAGAACGGTCAGTTGGAA TTCGTCACTGGTGGATGGGTCATGCCAGACGAGGCTAACTCCCA CTGGAGAAACGTTTTGTTGCAGTTGACCGAAGGTCAAACTTGGTT GAAGCAATTCATGAACGTCACTCCAACTGCTTCCTGGGCTATCGA TCCATTCGGACACTCTCCAACTATGCCATACATTTTGCAGAAGTC TGGTTTCAAGAATATGTTGATCCAGAGAACCCACTACTCCGTTAA GAAGGAGTTGGCTCAACAGAGACAGTTGGAGTTCTTGTGGAGAC AGATCTGGGACAACAAAGGTGACACTGCTTTGTTCACCCACATG ATGCCATTCTACTCTTACGACATTCCTCATACCTGTGGTCCAGAT CCAAAGGTTTGTTGTCAGTTCGATTTCAAAAGAATGGGTTCCTTC GGTTTGTCTTGTCCATGGAAGGTTCCACCTAGAACTATCTCTGAT CAAAATGTTGCTGCTAGATCCGATTTGTTGGTTGATCAGTGGAAG AAGAAGGCTGAGTTGTACAGAACCAACGTCTTGTTGATTCCATTG GGTGACGACTTCAGATTCAAGCAGAACACCGAGTGGGATGTTCA GAGAGTCAACTACGAAAGATTGTTCGAACACATCAACTCTCAGG CTCACTTCAATGTCCAGGCTCAGTTCGGTACTTTGCAGGAATACT TCGATGCTGTTCACCAGGCTGAAAGAGCTGGACAAGCTGAGTTC CCAACCTTGTCTGGTGACTTCTTCACTTACGCTGATAGATCTGAT AACTACTGGTCTGGTTACTACACTTCCAGACCATACCATAAGAG AATGGACAGAGTCTTGATGCACTACGTTAGAGCTGCTGAAATGT TGTCCGCTTGGCACTCCTGGGACGGTATGGCTAGAATCGAGGAA AGATTGGAGCAGGCTAGAAGAGAGTTGTCCTTGTTCCAGCACCA CGACGGTATTACTGGTACTGCTAAAACTCACGTTGTCGTCGACTA CGAGCAAAGAATGCAGGAAGCTTTGAAAGCTTGTCAAATGGTCA TGCAACAGTCTGTCTACAGATTGTTGACTAAGCCATCCATCTACT CTCCAGACTTCTCCTTCTCCTACTTCACTTTGGACGACTCCAGAT GGCCAGGTTCTGGTGTTGAGGACTCTAGAACTACCATCATCTTGG GTGAGGATATCTTGCCATCCAAGCATGTTGTCATGCACAACACCT TGCCACACTGGAGAGAGCAGTTGGTTGACTTCTACGTCTCCTCTC CATTCGTTTCTGTTACCGACTTGGCTAACAATCCAGTTGAGGCTC AGGTTTCTCCAGTTTGGTCTTGGCACCACGACACTTTGACTAAGA CTATCCACCCACAAGGTTCCACCACCAAGTACAGAATCATCTTCA AGGCTAGAGTTCCACCAATGGGTTTGGCTACCTACGTTTTGACCA TCTCCGATTCCAAGCCAGAGCACACCTCCTACGCTTCCAATTTGT TGCTTAGAAAGAACCCAACTTCCTTGCCATTGGGTCAATACCCAG AGGATGTCAAGTTCGGTGATCCAAGAGAGATCTCCTTGAGAGTT GGTAACGGTCCAACCTTGGCTTTCTCTGAGCAGGGTTTGTTGAAG TCCATTCAGTTGACTCAGGATTCTCCACATGTTCCAGTTCACTTC AAGTTCTTGAAGTACGGTGTTAGATCTCATGGTGATAGATCTGGT GCTTACTTGTTCTTGCCAAATGGTCCAGCTTCTCCAGTCGAGTTG GGTCAGCCAGTTGTCTTGGTCACTAAGGGTAAATTGGAGTCTTCC GTTTCTGTTGGTTTGCCATCTGTCGTTCACCAGACCATCATGAGA

		GGTGGTGCTCCAGAGATTAGAAATTTGGTCGATATTGGTTCTTTG GACAACACTGAGATCGTCATGAGATTGGAGACTCATATCGACTC TGGTGATATCTTCTACACTGATTTGAATGGATTGCAATTCATCAA GAGGAGAAGATTGGACAAGTTGCCATTGCAGGCTAACTACTACC CAATTCCATCTGGTATGTTCATTGAGGATGCTAATACCAGATTGA CTTTGTTGACCGGTCAACCATTGGGTGGATCTTCTTTGGCTTCTG GTGAGTTGGAGATTATGCAAGATAGAAGATTGGCTTCTGATGAT GAAAGAGGTTTGGGTCAGGGTGTTTTGGACAACAAGCCAGTTTT GCATATTTACAGATTGGTCTTGGAGAAGGTTAACAACTGTGTCAG ACCATCTAAGTTGCATCCAGCTGGTTACTTGACTTCTGCTGCTCA CAAAGCTTCTCAGTCTTTGTTGGATCCATTGGACAAGTTCATCTT CGCTGAAAATGAGTGGATCGGTGCTCAGGGTCAATTCGGTGGTG ATCATCCATCTGCTAGAGAGGATTTGGATGTCTCTGTCATGAGAA GATTGACCAAGTCTTCTGCTAAAACCCAGAGAGTTGGTTACGTTT TGCACAGAACCAATTTGATGCAATGTGGTACTCCAGAGGAGCAT ACTCAGAAGTTGGATGTCTGTCACTTGTTGCCAAATGTTGCTAGA TGTGAGAGAACTACCTTGACTTTCTTGCAGAATTTGGAGCACTTG GATGGTATGGTTGCTCCAGAAGTTTGTCCAATGGAAACCGCTGCT TACGTCTCTTCTCACTCTTCTTGA
34	*Drosophila melanogaster* ManII catalytic doman (KD)	RDDPIRPPLKVARSPRPGQCQDVVQDVPNVDVQMLELYDRMSFKDI DGGVWKQGWNIKYDPLKYNAHHKLKVFVVPHSHNDPGWIQTFEE YYQHDTKHILSNALRHLHDNPEMKFIWAEISYFARFYHDLGENKKL QMKSIVKNGQLEFVTGGWVMPDEANSHWRNVLLQLTEGQTWLKQ FMNVTPTASWAIDPFGHSPTMPYILQKSGFKNMLIQRTHYSVKKELA QQRQLEFLWRQIWDNKGDTALFTHMMPFYSYDIPHTCGPDPKVCC QFDFKRMGSFGLSCPWKVPPRTISDQNVAARSDLLVDQWKKKAEL YRTNVLLIPLGDDFRFKQNTEWDVQRVNYERLFEHINSQAHFNVQA QFGTLQEYFDAVHQAERAGQAEFPTLSGDFFTYADRSDNYWSGYY TSRPYHKRMDRVLMHYVRAAEMLSAWHSWDGMARIEERLEQARR ELSLFQHHDGITGTAKTHVVVDYEQRMQEALKACQMVMQQSVYR LLTKPSIYSPDFSFSYFTLDDSRWPGSGVEDSRTTIILGEDILPSKHVV MHNTLPHWREQLVDFYVSSPFVSVTDLANNPVEAQVSPVWSWHH DTLTKTIHPQGSTTKYRIIFKARVPPMGLATYVLTISDSKPEHTSYAS NLLLRKNPTSLPLGQYPEDVKFGDPREISLRVGNGPTLAFSEQGLLKS IQLTQDSPHVPVHFKFLKYGVRSHGDRSGAYLFLPNGPASPVELGQP VVLVTKGKLESSVSVGLPSVVHQTIMRGGAPEIRNLVDIGSLDNTEIV MRLETHIDSGDIFYTDLNGLQFIKRRRLDKLPLQANYYPIPSGMFIED ANTRLTLLTGQPLGGSSLASGELEIMQDRRLASDDERGLGQGVLDN KPVLHIYRLVLEKVNNCVRPSKLHPAGYLTSAAHKASQSLLDPLDKF IFAENEWIGAQGQFGGDHPSAREDLDVSVMRRLTKSSAKTQRVGYV LHRTNLMQCGTPEEHTQKLDVCHLLPNVARCERTTLTFLQNLEHLD GMVAPEVCPMETAAYVSSHSS
35	Mouse CMP-sialic acid transporter (MmCST) Codon optimized	ATGGCTCCAGCTAGAGAAAACGTTTCCTTGTTCTTCAAGTTGTAC TGTTTGGCTGTTATGACTTTGGTTGCTGCTGCTTACACTGTTGCTT TGAGATACACTAGAACTACTGCTGAGGAGTTGTACTTCTCCACTA CTGCTGTTTGTATCACTGAGGTTATCAAGTTGTTGATCTCCGTTG GTTTGTTGGCTAAGGAGACTGGTTCTTTGGGAAGATTCAAGGCTT CCTTGTCCGAAAACGTTTTGGGTTCCCCAAAGGAGTTGGCTAAGT TGTCTGTTCCATCCTTGGTTTACGCTGTTCAGAACAACATGGCTT

		TCTTGGCTTTGTCTAACTTGGACGCTGCTGTTTACCAAGTTACTTA CCAGTTGAAGATCCCATGTACTGCTTTGTGTACTGTTTTGATGTT GAACAGAACATTGTCCAAGTTGCAGTGGATCTCCGTTTTCATGTT GTGTGGTGGTGTTACTTTGGTTCAGTGGAAGCCAGCTCAAGCTTC CAAAGTTGTTGTTGCTCAGAACCCATTGTTGGGTTTCGGTGCTAT TGCTATCGCTGTTTTGTGTTCCGGTTTCGCTGGTGTTTACTTCGAG AAGGTTTTGAAGTCCTCCGACACTTCTTTGTGGGTTAGAAACATC CAGATGTACTTGTCCGGTATCGTTGTTACTTTGGCTGGTACTTAC TTGTCTGACGGTGCTGAGATTCAAGAGAAGGGATTCTTCTACGGT TACACTTACTATGTTTGGTTCGTTATCTTCTTGGCTTCCGTTGGTG GTTTGTACACTTCCGTTGTTGTTAAGTACACTGACAACATCATGA AGGGATTCTCTGCTGCTGCTGCTATTGTTTTGTCCACTATCGCTTC CGTTTTGTTGTTCGGATTGCAGATCACATTGTCCTTTGCTTTGGGA GCTTTGTTGGTTTGTGTTTCCATCTACTTGTACGGATTGCCAAGA CAAGACACTACTTCCATTCAGCAAGAGGCTACTTCCAAGGAGAG AATCATCGGTGTTTAGTAG
36	Mouse CMP-sialic acid transporter (MmCST)	MAPARENVSLFFKLYCLAVMTLVAAAYTVALRYTRTTAEELYFSTT AVCITEVIKLLISVGLLAKETGSLGRFKASLSENVLGSPKELAKLSVPS LVYAVQNNMAFLALSNLDAAVYQVTYQLKIPCTALCTVLMLNRTL SKLQWISVFMLCGGVTLVQWKPAQASKVVVAQNPLLGFGAIAIAVL CSGFAGVYFEKVLKSSDTSLWVRNIQMYLSGIVVTLAGTYLSDGAEI QEKGFFYGYTYYVWFVIFLASVGGLYTSVVVKYTDNIMKGFSAAA AIVLSTIASVLLFGLQITLSFALGALLVCVSIYLYGLPRQDTTSIQQEA TSKERIIGV
37	Human UDP-GlcNAc 2-epimerase/ N-acetylmannosamine kinase (HsGNE) codon opitimized	ATGGAAAAGAACGGTAACAACAGAAAGTTGAGAGTTTGTGTTGC TACTTGTAACAGAGCTGACTACTCCAAGTTGGCTCCAATCATGTT CGGTATCAAGACTGAGCCAGAGTTCTTCGAGTTGGACGTTGTTGT TTTGGGTTCCCACTTGATTGATGACTACGGTAACACTTACAGAAT GATCGAGCAGGACGACTTCGACATCAACACTAGATTGCACACTA TTGTTAGAGGAGAGGACGAAGCTGCTATGGTTGAATCTGTTGGA TTGGCTTTGGTTAAGTTGCCAGACGTTTTGAACAGATTGAAGCCA GACATCATGATTGTTCACGGTGACAGATTCGATGCTTTGGCTTTG GCTACTTCCGCTGCTTTGATGAACATTAGAATCTTGCACATCGAG GGTGGTGAAGTTTCTGGTACTATCGACGACTCCATCAGACACGCT ATCACTAAGTTGGCTCACTACCATGTTTGTTGTACTAGATCCGCT GAGCAACACTTGATTTCCATGTGTGAGGACCACGACAGAATTTT GTTGGCTGGTTGTCCATCTTACGACAAGTTGTTGTCCGCTAAGAA CAAGGACTACATGTCCATCATCAGAATGTGGTTGGGTGACGACG TTAAGTCTAAGGACTACATCGTTGCTTTGCAGCACCCAGTTACTA CTGACATCAAGCACTCCATCAAGATGTTCGAGTTGACTTTGGACG CTTTGATCTCCTTCAACAAGAGAACTTTGGTTTTGTTCCCAAACA TTGACGCTGGTTCCAAAGAGATGGTTAGAGTTATGAGAAAGAAG GGTATCGAACACCACCCAAACTTCAGAGCTGTTAAGCACGTTCC ATTCGACCAATTCATCCAGTTGGTTGCTCATGCTGGTTGTATGAT CGGTAACTCCTCCTGTGGTGTTAGAGAAGTTGGTGCTTTCGGTAC TCCAGTTATCAACTTGGGTACTAGACAGATCGGTAGAGAGACTG GAGAAAACGTTTTGCATGTTAGAGATGCTGACACTCAGGACAAG ATTTTGCAGGCTTTGCACTTGCAATTCGGAAAGCAGTACCCATGT TCCAAAATCTACGGTGACGGTAACGCTGTTCCAAGAATCTTGAA

		GTTTTTGAAGTCCATCGACTTGCAAGAGCCATTGCAGAAGAAGTT CTGTTTCCCACCAGTTAAGGAGAACATCTCCCAGGACATTGACC ACATCTTGGAGACATTGTCCGCTTTGGCTGTTGATTTGGGTGGAA CTAACTTGAGAGTTGCTATCGTTTCCATGAAGGGAGAGATCGTTA AGAAGTACACTCAGTTCAACCCAAAGACTTACGAGGAGAGAATC AACTTGATCTTGCAGATGTGTGTTGAAGCTGCTGCTGAGGCTGTT AAGTTGAACTGTAGAATCTTGGGTGTTGGTATCTCTACTGGTGGT AGAGTTAATCCAAGAGAGGGTATCGTTTTGCACTCCACTAAGTTG ATTCAGGAGTGGAACTCCGTTGATTTGAGAACTCCATTGTCCGAC ACATTGCACTTGCCAGTTTGGGTTGACAACGACGGTAATTGTGCT GCTTTGGCTGAGAGAAAGTTCGGTCAAGGAAAGGGATTGGAGAA CTTCGTTACTTTGATCACTGGTACTGGTATTGGTGGTGGTATCATT CACCAGCACGAGTTGATTCACGGTTCTTCCTTCTGTGCTGCTGAA TTGGGACACTTGGTTGTTTCTTTGGACGGTCCAGACTGTTCTTGT GGTTCCCACGGTTGTATTGAAGCTTACGCATCAGGAATGGCATTG CAGAGAGAGGCTAAGAAGTTGCACGACGAGGACTTGTTGTTGGT TGAGGGAATGTCTGTTCCAAAGGACGAGGCTGTTGGTGCTTTGC ATTTGATCCAGGCTGCTAAGTTGGGTAATGCTAAGGCTCAGTCCA TCTTGAGAACTGCTGGTACTGCTTTGGGATTGGGTGTTGTTAATA TCTTGCACACTATGAACCCATCCTTGGTTATCTTGTCCGGTGTTTT GGCTTCTCACTACATCCACATCGTTAAGGACGTTATCAGACAGCA AGCTTTGTCCTCCGTTCAAGACGTTGATGTTGTTGTTTCCGACTTG GTTGACCCAGCTTTGTTGGGTGCTGCTTCCATGGTTTTGGACTAC ACTACTAGAAGAATCTACTAATAG
38	Human UDP-GlcNAc 2-epimerase/ N-acetylmannosamine kinase (HsGNE)	MEKNGNNRKLRVCVATCNRADYSKLAPIMFGIKTEPEFFELDVVVL GSHLIDDYGNTYRMIEQDDFDINTRLHTIVRGEDEAAMVESVGLAL VKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIEGGEVS GTIDDSIRHAITKLAHYHVCCTRSAEQHLISMCEDHDRILLAGCPSYD KLLSAKNKDYMSIIRMWLGDDVKSKDYIVALQHPVTTDIKHSIKMF ELTLDALISFNKRTLVLFPNIDAGSKEMVRVMRKKGIEHHPNFRAVK HVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQIGRET GENVLHVRDADTQDKILQALHLQFGKQYPCSKIYGDGNAVPRILKF LKSIDLQEPLQKKFCFPPVKENISQDIDHILETLSALAVDLGGTNLRV AIVSMKGEIVKKYTQFNPKTYEERINLILQMCVEAAAEAVKLNCRIL GVGISTGGRVNPREGIVLHSTKLIQEWNSVDLRTPLSDTLHLPVWVD NDGNCAALAERKFGQGKGLENFVTLITGTGIGGGIIHQHELIHGSSFC AAELGHLVVSLDGPDCSCGSHGCIEAYASGMALQREAKKLHDEDLL LVEGMSVPKDEAVGALHLIQAAKLGNAKAQSILRTAGTALGLGVV NILHTMNPSLVILSGVLASHYIHIVKDVIRQQALSSVQDVDVVVSDL VDPALLGAASMVLDYTTRRIY
39	Human CMP-sialic acid synthase (HsCSS) codon optimized	ATGGACTCTGTTGAAAAGGGTGCTGCTACTTCTGTTTCCAACCCA AGAGGTAGACCATCCAGAGGTAGACCTCCTAAGTTGCAGAGAAA CTCCAGAGGTGGTCAAGGTAGAGGTGTTGAAAAGCCACCACACT TGGCTGCTTTGATCTTGGCTAGAGGAGGTTCTAAGGGTATCCCAT TGAAGAACATCAAGCACTTGGCTGGTGTTCCATTGATTGGATGG GTTTTGAGAGCTGCTTTGGACTCTGGTGCTTTCCAATCTGTTTGG GTTTCCACTGACCACGACGAGATTGAGAACGTTGCTAAGCAATT CGGTGCTCAGGTTCACAGAAGATCCTCTGAGGTTTCCAAGGACT CTTCTACTTCCTTGGACGCTATCATCGAGTTCTTGAACTACCACA

		ACGAGGTTGACATCGTTGGTAACATCCAAGCTACTTCCCCATGTT TGCACCCAACTGACTTGCAAAAAGTTGCTGAGATGATCAGAGAA GAGGGTTACGACTCCGTTTTCTCCGTTGTTAGAAGGCACCAGTTC AGATGGTCCGAGATTCAGAAGGGTGTTAGAGAGGTTACAGAGCC ATTGAACTTGAACCCAGCTAAAAGACCAAGAAGGCAGGATTGGG ACGGTGAATTGTACGAAAACGGTTCCTTCTACTTCGCTAAGAGA CACTTGATCGAGATGGGATACTTGCAAGGTGGAAAGATGGCTTA CTACGAGATGAGAGCTGAACACTCCGTTGACATCGACGTTGATA TCGACTGGCCAATTGCTGAGCAGAGAGTTTTGAGATACGGTTACT TCGGAAAGGAGAAGTTGAAGGAGATCAAGTTGTTGGTTTGTAAC ATCGACGGTTGTTTGACTAACGGTCACATCTACGTTTCTGGTGAC CAGAAGGAGATTATCTCCTACGACGTTAAGGACGCTATTGGTAT CTCCTTGTTGAAGAAGTCCGGTATCGAAGTTAGATTGATCTCCGA GAGAGCTTGTTCCAAGCAAACATTGTCCTCTTTGAAGTTGGACTG TAAGATGGAGGTTTCCGTTTCTGACAAGTTGGCTGTTGTTGACGA ATGGAGAAAGGAGATGGGTTTGTGTTGGAAGGAAGTTGCTTACT TGGGTAACGAAGTTTCTGACGAGGAGTGTTTGAAGAGAGTTGGT TTGTCTGGTGCTCCAGCTGATGCTTGTTCCACTGCTCAAAAGGCT GTTGGTTACATCTGTAAGTGTAACGGTGGTAGAGGTGCTATTAGA GAGTTCGCTGAGCACATCTGTTTGTTGATGGAGAAAGTTAATAAC TCCTGTCAGAAGTAGTAG
40	Human CMP-sialic acid synthase (HsCSS)	MDSVEKGAATSVSNPRGRPSRGRPPKLQRNSRGGQGRGVEKPPHL AALILARGGSKGIPLKNIKHLAGVPLIGWVLRAALDSGAFQSVWVST DHDEIENVAKQFGAQVHRRSSEVSKDSSTSLDAIIEFLNYHNEVDIV GNIQATSPCLHPTDLQKVAEMIREEGYDSVFSVVRRHQFRWSEIQKG VREVTEPLNLNPAKRPRRQDWDGELYENGSFYFAKRHLIEMGYLQG GKMAYYEMRAEHSVDIDVDIDWPIAEQRVLRYGYFGKEKLKEIKLL VCNIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKKSGIEVRLISER ACSKQTLSSLKLDCKMEVSVSDKLAVVDEWRKEMGLCWKEVAYL GNEVSDEECLKRVGLSGAPADACSTAQKAVGYICKCNGGRGAIREF AEHICLLMEKVNNSCQK
41	Human N-acetylneuraminate-9-phosphate synthase (HsSPS) codon optimized	ATGCCATTGGAATTGGAGTTGTGTCCTGGTAGATGGGTTGGTGGT CAACACCCATGTTTCATCATCGCTGAGATCGGTCAAAACCACCA AGGAGACTTGGACGTTGCTAAGAGAATGATCAGAATGGCTAAGG AATGTGGTGCTGACTGTGCTAAGTTCCAGAAGTCCGAGTTGGAG TTCAAGTTCAACAGAAAGGCTTTGGAAAGACCATACACTTCCAA GCACTCTTGGGGAAAGACTTACGGAGAACACAAGAGACACTTGG AGTTCTCTCACGACCAATACAGAGAGTTGCAGAGATACGCTGAG GAAGTTGGTATCTTCTTCACTGCTTCTGGAATGGACGAAATGGCT GTTGAGTTCTTGCACGAGTTGAACGTTCCATTCTTCAAAGTTGGT TCCGGTGACACTAACAACTTCCCATACTTGGAAAAGACTGCTAA GAAAGGTAGACCAATGGTTATCTCCTCTGGAATGCAGTCTATGG ACACTATGAAGCAGGTTTACCAGATCGTTAAGCCATTGAACCCA AACTTTTGTTTCTTGCAGTGTACTTCCGCTTACCCATTGCAACCA GAGGACGTTAATTTGAGAGTTATCTCCGAGTACCAGAAGTTGTTC CCAGACATCCCAATTGGTTACTCTGGTCACGAGACTGGTATTGCT ATTTCCGTTGCTGCTGTTGCTTTGGGTGCTAAGGTTTTGGAGAGA CACATCACTTTGGACAAGACTTGGAAGGGTTCTGATCACTCTGCT TCTTTGGAACCTGGTGAGTTGGCTGAACTTGTTAGATCAGTTAGA

		TTGGTTGAGAGAGCTTTGGGTTCCCCAACTAAGCAATTGTTGCCA TGTGAGATGGCTTGTAACGAGAAGTTGGGAAAGTCCGTTGTTGC TAAGGTTAAGATCCCAGAGGGTACTATCTTGACTATGGACATGTT GACTGTTAAAGTTGGAGAGCCAAAGGGTTACCCACCAGAGGACA TCTTTAACTTGGTTGGTAAAAAGGTTTTGGTTACTGTTGAGGAGG ACGACACTATTATGGAGGAGTTGGTTGACAACCACGGAAAGAAG ATCAAGTCCTAG
42	Human N-acetylneuraminate-9-phosphate synthase (HsSPS)	MPLELELCPGRWVGGQHPCFIIAEIGQNHQGDLDVAKRMIRMAKEC GADCAKFQKSELEFKFNRKALERPYTSKHSWGKTYGEHKRHLEFSH DQYRELQRYAEEVGIFFTASGMDEMAVEFLHELNVPFFKVGSGDTN NFPYLEKTAKKGRPMVISSGMQSMDTMKQVYQIVKPLNPNFCFLQC TSAYPLQPEDVNLRVISEYQKLFPDIPIGYSGHETGIAISVAAVALGA KVLERHITLDKTWKGSDHSASLEPGELAELVRSVRLVERALGSPTKQ LLPCEMACNEKLGKSVVAKVKIPEGTILTMDMLTVKVGEPKGYPPE DIFNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS
43	Mouse alpha-2,6-sialyl transferase catalytic domain (MmmST6) codon optimized	GTTTTTCAAATGCCAAAGTCCCAGGAGAAAGTTGCTGTTGGTCCA GCTCCACAAGCTGTTTTCTCCAACTCCAAGCAAGATCCAAAGGA GGGTGTTCAAATCTTGTCCTACCCAAGAGTTACTGCTAAGGTTAA GCCACAACCATCCTTGCAAGTTTGGGACAAGGACTCCACTTACTC CAAGTTGAACCCAAGATTGTTGAAGATTTGGAGAAACTACTTGA ACATGAACAAGTACAAGGTTTCCTACAAGGGTCCAGGTCCAGGT GTTAAGTTCTCCGTTGAGGCTTTGAGATGTCACTTGAGAGACCAC GTTAACGTTTCCATGATCGAGGCTACTGACTTCCCATTCAACACT ACTGAATGGGAGGGATACTTGCCAAAGGAGAACTTCAGAACTAA GGCTGGTCCATGGCATAAGTGTGCTGTTGTTTCTTCTGCTGGTTC CTTGAAGAACTCCCAGTTGGGTAGAGAAATTGACAACCACGACG CTGTTTTGAGATTCAACGGTGCTCCAACTGACAACTTCCAGCAGG ATGTTGGTACTAAGACTACTATCAGATTGGTTAACTCCCAATTGG TTACTACTGAGAAGAGATTCTTGAAGGACTCCTTGTACACTGAGG GAATCTTGATTTTGTGGGACCCATCTGTTTACCACGCTGACATTC CACAATGGTATCAGAAGCCAGACTACAACTTCTTCGAGACTTAC AAGTCCTACAGAAGATTGCACCCATCCCAGCCATTCTACATCTTG AAGCCACAAATGCCATGGGAATTGTGGGACATCATCCAGGAAAT TTCCCCAGACTTGATCCAACCAAACCCACCATCTTCTGGAATGTT GGGTATCATCATCATGATGACTTTGTGTGACCAGGTTGACATCTA CGAGTTCTTGCCATCCAAGAGAAAGACTGATGTTTGTTACTACCA CCAGAAGTTCTTCGACTCCGCTTGTACTATGGGAGCTTACCACCC ATTGTTGTTCGAGAAGAACATGGTTAAGCACTTGAACGAAGGTA CTGACGAGGACATCTACTTGTTCGGAAAGGCTACTTTGTCCGGTT TCAGAAACAACAGATGTTAG
44	Mouse alpha-2,6-sialyl transferase catalytic domain (MmmST6)	VFQMPKSQEKVAVGPAPQAVFSNSKQDPKEGVQILSYPRVTAKVKP QPSLQVWDKDSTYSKLNPRLLKIWRNYLNMNKYKVSYKGPGPGVK FSVEALRCHLRDHVNVSMIEATDFPFNTTEWEGYLPKENFRTKAGP WHKCAVVSSAGSLKNSQLGREIDNHDAVLRFNGAPTDNFQQDVGT KTTIRLVNSQLVTTEKRFLKDSLYTEGILILWDPSVYHADIPQWYQK PDYNFFETYKS YRRLHPSQPFYILKPQMPWELWDIIQEISPDLIQPNPPSSGMLGIIIMM TLCDQVDIYEFLPSKRKTDVCYYHQKFFDSACTMGAYHPLLFEKNM VKHLNEGTDEDIYLFGKATLSGFRNNRC

45	Sequence of the PpPMA1 promoter:	AAATGCGTACCTCTTCTACGAGATTCAAGCGAATGAGAATAATG TAATATGCAAGATCAGAAAGAATGAAAGGAGTTGAAAAAAAAA ACCGTTGCGTTTTGACCTTGAATGGGGTGGAGGTTTCCATTCAAA GTAAAGCCTGTGTCTTGGTATTTTCGGCGGCACAAGAAATCGTAA TTTTCATCTTCTAAACGATGAAGATCGCAGCCCAACCTGTATGTA GTTAACCGGTCGGAATTATAAGAAAGATTTTCGATCAACAAACC CTAGCAAATAGAAAGCAGGGTTACAACTTTAAACCGAAGTCACA AACGATAAACCACTCAGCTCCCACCCAAATTCATTCCCACTAGC AGAAAGGAATTATTTAATCCCTCAGGAAACCTCGATGATTCTCCC GTTCTTCCATGGGCGGGTATCGCAAAATGAGGAATTTTTCAAATT TCTCTATTGTCAAGACTGTTTATTATCTAAGAAATAGCCCAATCC GAAGCTCAGTTTTGAAAAAATCACTTCCGCGTTTCTTTTTTACAG CCCGATGAATATCCAAATTTGGAATATGGATTACTCTATCGGGAC TGCAGATAATATGACAACAACGCAGATTACATTTTAGGTAAGGC ATAAACACCAGCCAGAAATGAAACGCCCACTAGCCATGGTCGAA TAGTCCAATGAATTCAGATAGCTATGGTCTAAAAGCTGATGTTTT TTATTGGGTAATGGCGAAGAGTCCAGTACGACTTCCAGCAGAGC TGAGATGGCCATTTTTGGGGGTATTAGTAACTTTTTGAGCTCTTTT CACTTCGATGAAGTGTCCCATTCGGGATATAATCGGATCGCGTCG TTTTCTCGAAAATACAGCTTAGCGTCGTCCGCTTGTTGTAAAAGC AGCACCACATTCCTAATCTCTTATATAAACAAAACAACCCAAATT ATCAGTGCTGTTTTCCCACCAGATATAAGTTTCTTTTCTCTTCCGC TTTTTGATTTTTTATCTCTTTCCTTTAAAAACTTCTTTACCTTAAA GGGCGGCC
46	Sequence of the PpPMA1 terminator:	TAAGCTTCACGATTTGTGTTCCAGTTTATCCCCCCTTTATATACCG TTAACCCTTTCCCTGTTGAGCTGACTGTTGTTGTATTACCGCAATT TTTCCAAGTTTGCCATGCTTTTCGTGTTATTTGACCGATGTCTTTT TTCCCAAATCAAACTATATTTGTTACCATTTAAACCAAGTTATCT TTTGTATTAAGAGTCTAAGTTTGTTCCCAGGCTTCATGTGAGAGT GATAACCATCCAGACTATGATTCTTGTTTTTTATTGGGTTTGTTTG TGTGATACATCTGAGTTGTGATTCGTAAAGTATGTCAGTCTATCT AGATTTTTAATAGTTAATTGGTAATCAATGACTTGTTTGTTTTAAC TTTTAAATTGTGGGTCGTATCCACGCGTTTAGTATAGCTGTTCAT GGCTGTTAGAGGAGGGCGATGTTTATATACAGAGGACAAGAATG AGGAGGCGGCGTGTATTTTTAAAATGGAGACGCGACTCCTGTAC ACCTTATCGGTTGG
47	Sequence of the PpOCH1 promoter:	TGGACACAGGAGACTCAGAAACAGACACAGAGCGTTCTGAGTCC TGGTGCTCCTGACGTAGGCCTAGAACAGGAATTATTGGCTTTATT TGTTTGTCCATTTCATAGGCTTGGGGTAATAGATAGATGACAGAG AAATAGAGAAGACCTAATATTTTTTGTTCATGGCAAATCGCGGGT TCGCGGTCGGGTCACACACGGAGAAGTAATGAGAAGAGCTGGTA ATCTGGGGTAAAAGGGTTCAAAAGAAGGTCGCCTGGTAGGGATG CAATACAAGGTTGTCTTGGAGTTTACATTGACCAGATGATTTGGC TTTTTCTCTGTTCAATTCACATTTTTCAGCGAGAATCGGATTGAC GGAGAAATGGCGGGGTGTGGGGTGGATAGATGGCAGAAATGCT CGCAATCACCGCGAAAGAAAGACTTTATGGAATAGAACTACTGG GTGGTGTAAGGATTACATAGCTAGTCCAATGGAGTCCGTTGGAA AGGTAAGAAGAAGCTAAAACCGGCTAAGTAACTAGGGAAGAAT GATCAGACTTTGATTTGATGAGGTCTGAAAATACTCTGCTGCTTT

		TTCAGTTGCTTTTTCCCTGCAACCTATCATTTTCCTTTTCATAAGC CTGCCTTTTCTGTTTTCACTTATATGAGTTCCGCCGAGACTTCCCC AAATTCTCTCCTGGAACATTCTCTATCGCTCTCCTTCCAAGTTGC GCCCCCTGGCACTGCCTAGTAATATTACCACGCGACTTATATTCA GTTCCACAATTTCCAGTGTTCGTAGCAAATATCATCAGCC
48	Sequence of the PpALG12 terminator:	AATATATACCTCATTTGTTCAATTTGGTGTAAAGAGTGTGGCGGA TAGACTTCTTGTAAATCAGGAAAGCTACAATTCCAATTGCTGCAA AAAATACCAATGCCCATAAACCAGTATGAGCGGTGCCTTCGACG GATTGCTTACTTTCCGACCCTTTGTCGTTTGATTCTTCTGCCTTTG GTGAGTCAGTTTGTTTCGACTTTATATCTGACTCATCAACTTCCTT TACGGTTGCGTTTTTAATCATAATTTTAGCCGTTGGCTTATTATCC CTTGAGTTGGTAGGAGTTTTGATGATGCTG
49	Sequence of the PpSEC4 promoter:	GAAGTAAAGTTGGCGAAACTTTGGGAACCTTTGGTTAAAACTTT GTAATTTTTGTCGCTACCCATTAGGCAGAATCTGCATCTTGGGAG GGGGATGTGGTGGCGTTCTGAGATGTACGCGAAGAATGAAGAGC CAGTGGTAACAACAGGCCTAGAGAGATACGGGCATAATGGGTAT AACCTACAAGTTAAGAATGTAGCAGCCCTGGAAACCAGATTGAA ACGAAAAACGAAATCATTTAAACTGTAGGATGTTTTGGCTCATTG TCTGGAAGGCTGGCTGTTTATTGCCCTGTTCTTTGCATGGGAATA AGCTATTATATCCCTCACATAATCCCAGAAAATAGATTGAAGCA ACGCGAAATCCTTACGTATCGAAGTAGCCTTCTTACACATTCACG TTGTACGGATAAGAAAACTACTCAAACGAACAATC
50	Sequence of the PpOCH1 terminator:	AATAGATATAGCGAGATTAGAGAATGAATACCTTCTTCTAAGCG ATCGTCCGTCATCATAGAATATCATGGACTGTATAGTTTTTTTTTT GTACATATAATGATTAAACGGTCATCCAACATCTCGTTGACAGAT CTCTCAGTACGCGAAATCCCTGACTATCAAAGCAAGAACCGATG AAGAAAAAAACAACAGTAACCCAAACACCACAACAAACACTTT ATCTTCTCCCCCCCAACACCAATCATCAAAGAGATGTCGGAACA CAAACACCAAGAAGCAAAAACTAACCCCATATAAAAACATCCTG GTAGATAATGCTGGTAACCCGCTCTCCTTCCATATTCTGGGCTAC TTCACGAAGTCTGACCGGTCTCAGTTGATCAACATGATCCTCGAA ATGG
51	Sequence of the PpTEF1 promoter	TTAAGGTTTGGAACAACACTAAACTACCTTGCGGTACTACCATTG ACACTACACATCCTTAATTCCAATCCTGTCTGGCCTCCTTCACCTT TTAACCATCTTGCCCATTCCAACTCGTGTCAGATTGCGTATCAAG TGAAAAAAAAAAAAATTTTAAATCTTTAACCCAATCAGGTAATAA CTGTCGCCTCTTTTATCTGCCGCACTGCATGAGGTGTCCCCTTAG TGGGAAAGAGTACTGAGCCAACCCTGGAGGACAGCAAGGGAAA AATACCTACAACTTGCTTCATAATGGTCGTAAAAACAATCCTTGT CGGATATAAGTGTTGTAGACTGTCCCTTATCCTCTGCGATGTTCT TCCTCTCAAAGTTTGCGATTTCTCTCTATCAGAATTGCCATCAAG AGACTCAGGACTAATTTCGCAGTCCCACACGCACTCGTACATGA TTGGCTGAAATTTCCCTAAAGAATTTCTTTTTCACGAAAATTTTTT TTTTACACAAGATTTTCAGCAGATATAAAATGGAGAGCAGGACC TCCGCTGTGACTCTTCTTTTTTTTCTTTTATTCTCACTACATACATT TTAGTTATTCGCCAAC

52	Sequence of the PpTEF1 terminator:	ATTGCTTGAAGCTTTAATTTATTTTATTAACATAATAATAATACA AGCATGATATATTTGTATTTTGTTCGTTAACATTGATGTTTTCTTC ATTTACTGTTATTGTTTGTAACTTTGATCGATTTATCTTTTCTACTT TACTGTAATATGGCTGGCGGGTGAGCCTTGAACTCCCTGTATTAC TTTACCTTGCTATTACTTAATCTATTGACTAGCAGCGACCTCTTCA ACCGAAGGGCAAGTACACAGCAAGTTCATGTCTCCGTAAGTGTC ATCAACCCTGGAAACAGTGGGCCATGTC
53	Sequence of the PpGAPDH promoter:	TTTTTGTAGAAATGTCTTGGTGTCCTCGTCCAATCAGGTAGCCAT CTCTGAAATATCTGGCTCCGTTGCAACTCCGAACGACCTGCTGGC AACGTAAAATTCTCCGGGGTAAAACTTAAATGTGGAGTAATGGA ACCAGAAACGTCTCTTCCCTTCTCTCTCCTTCCACCGCCCGTTAC CGTCCCTAGGAAATTTTACTCTGCTGGAGAGCTTCTTCTACGGCC CCCTTGCAGCAATGCTCTTCCCAGCATTACGTTGCGGGTAAAACG GAGGTCGTGTACCCGACCTAGCAGCCCAGGGATGGAAAAGTCCC GGCCGTCGCTGGCAATAATAGCGGGCGGACGCATGTCATGAGAT TATTGGAAACCACCAGAATCGAATATAAAAGGCGAACACCTTTC CCAATTTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTATT TGTCCCTATTTCAATCAATTGAACAACTATCAAAACACA
54	Sequence of the PpALG3 terminator:	ATTTACAATTAGTAATATTAAGGTGGTAAAAACATTCGTAGAATT GAAATGAATTAATATAGTATGACAATGGTTCATGTCTATAAATCT CCGGCTTCGGTACCTTCTCCCCAATTGAATACATTGTCAAAATGA ATGGTTGAACTATTAGGTTCGCCAGTTTCGTTATTAAGAAAACTG TTAAAATCAAATTCCATATCATCGGTTCCAGTGGGAGGACCAGTT CCATCGCCAAAATCCTGTAAGAATCCATTGTCAGAACCTGTAAA GTCAGTTTGAGATGAAATTTTTCCGGTCTTTGTTGACTTGGAAGC TTCGTTAAGGTTAGGTGAAACAGTTTGATCAACCAGCGGCTCCC GTTTTCGTCGCTTAGTAG
55	Sequence of the PpAOX1 promoter and integration locus:	AACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCCG ACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAG GAGGGGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCC ACTCCTCTTCTCCTCAACACCCACTTTTGCCATCGAAAAACCAGC CCAGTTATTGGGCTTGATTGGAGCTCGCTCATTCCAATTCCTTCT ATTAGGCTACTAACACCATGACTTTATTAGCCTGTCTATCCTGGC CCCCCTGGCGAGGTTCATGTTTGTTTATTTCCGAATGCAACAAGC TCCGCATTACACCCGAACATCACTCCAGATGAGGGCTTTCTGAGT GTGGGGTCAAATAGTTTCATGTTCCCCAAATGGCCCAAAACTGA CAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAGCGTGAT CTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACGGC CAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGT CTTGTTTGGTATTGATTGACGAATGCTCAAAAATAATCTCATTAA TGCTTAGCGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCTG TGCCGAAACGCAAATGGGGAAACACCCGCTTTTTGGATGATTAT GCATTGTCTCCACATTGTATGCTTCCAAGATTCTGGTGGGAATAC TGCTGATAGCCTAACGTTCATGATCAAAATTTAACTGTTCTAACC CCTACTTGACAGCAATATATAAACAGAAGGAAGCTGCCCTGTCT

		TAAACCTTTTTTTTTATCATCATTATTAGCTTACTTTCATAATTGC GACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTTTTAACG ACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACG
56	Sequence of the ScCYC1 terminator:	ACAGGCCCCTTTTCCTTTGTCGATATCATGTAATTAGTTATGTCA CGCTTACATTCACGCCCTCCTCCCACATCCGCTCTAACCGAAAAG GAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTT TAATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTT TCTTTTTTTTCTGTACAAACGCGTGTACGCATGTAACATTATACT GAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAA TTTGCAAGCTGCCGGCTCTTAAG
57	Sequence of the ScTEF1 promoter:	GATCCCCCACACACCATAGCTTCAAAATGTTTCTACTCCTTTTTT ACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCGTACCACTTC AAAACACCCAAGCACAGCATACTAAATTTCCCCTCTTTCTTCCTC TAGGGTGTCGTTAATTACCCGTACTAAAGGTTTGGAAAAGAAAA AAGAGACCGCCTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAA AATTTTTATCACGTTTCTTTTTCTTGAAAATTTTTTTTTTTGATTTT TTTCTCTTTCGATGACCTCCCATTGATATTTAAGTTAATAAACGG TCTTCAATTTCTCAAGTTTCAGTTTCATTTTTCTTGTTCTATTACA ACTTTTTTTACTTCTTGCTCATTAGAAAGAAAGCATAGCAATCTA ATCTAAGTTTTAATTACAAA
58	Sequence of the Sh ble ORF (Zeocin resistance marker):	ATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGA CGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCT CCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGAC GACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCC GGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGC TGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGGAC GCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGG GCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACT TCGTGGCCGAGGAGCAGGACTGA
59	Sequence of the 5'-Region used for knock out of PpURA5:	ATCGGCCTTTGTTGATGCAAGTTTTACGTGGATCATGGACTAAGG AGTTTTATTTGGACCAAGTTCATCGTCCTAGACATTACGGAAAGG GTTCTGCTCCTCTTTTTGGAAACTTTTTGGAACCTCTGAGTATGAC AGCTTGGTGGATTGTACCCATGGTATGGCTTCCTGTGAATTTCTA TTTTTTCTACATTGGATTCACCAATCAAAACAAATTAGTCGCCAT GGCTTTTTGGCTTTTGGGTCTATTTGTTTGGACCTTCTTGGAATAT GCTTTGCATAGATTTTTGTTCCACTTGGACTACTATCTTCCAGAG AATCAAATTGCATTTACCATTCATTTCTTATTGCATGGGATACAC CACTATTTACCAATGGATAAATACAGATTGGTGATGCCACCTAC ACTTTTCATTGTACTTTGCTACCCAATCAAGACGCTCGTCTTTTCT GTTCTACCATATTACATGGCTTGTTCTGGATTTGCAGGTGGATTC CTGGGCTATATCATGTATGATGTCACTCATTACGTTCTGCATCAC TCCAAGCTGCCTCGTTATTTCCAAGAGTTGAAGAAATATCATTTG GAACATCACTACAAGAATTACGAGTTAGGCTTTGGTGTCACTTCC AAATTCTGGGACAAAGTCTTTGGGACTTATCTGGGTCCAGACGAT GTGTATCAAAAGACAAATTAGAGTATTTATAAAGTTATGTAAGC AAATAGGGGCTAATAGGGAAAGAAAAATTTTGGTTCTTTATCAG AGCTGGCTCGCGCGCAGTGTTTTTCGTGCTCCTTTGTAATAGTCA

		TTTTTGACTACTGTTCAGATTGAAATCACATTGAAGATGTCACTC GAGGGGTACCAAAAAAGGTTTTTGGATGCTGCAGTGGCTTCGC
60	Sequence of the 3'-Region used for knock out of PpURA5:	GGTCTTTTCAACAAAGCTCCATTAGTGAGTCAGCTGGCTGAATCT TATGCACAGGCCATCATTAACAGCAACCTGGAGATAGACGTTGT ATTTGGACCAGCTTATAAAGGTATTCCTTTGGCTGCTATTACCGT GTTGAAGTTGTACGAGCTCGGCGGCAAAAAATACGAAAATGTCG GATATGCGTTCAATAGAAAAGAAAAGAAAGACCACGGAGAAGG TGGAAGCATCGTTGGAGAAAGTCTAAAGAATAAAAGAGTACTGA TTATCGATGATGTGATGACTGCAGGTACTGCTATCAACGAAGCAT TTGCTATAATTGGAGCTGAAGGTGGGAGAGTTGAAGGTAGTATT ATTGCCCTAGATAGAATGGAGACTACAGGAGATGACTCAAATAC CAGTGCTACCCAGGCTGTTAGTCAGAGATATGGTACCCCTGTCTT GAGTATAGTGACATTGGACCATATTGTGGCCCATTTGGGCGAAA CTTTCACAGCAGACGAGAAATCTCAAATGGAAACGTATAGAAAA AAGTATTTGCCCAAATAAGTATGAATCTGCTTCGAATGAATGAAT TAATCCAATTATCTTCTCACCATTATTTTCTTCTGTTTCGGAGCTT TGGGCACGGCGGCGGGTGGTGCGGGCTCAGGTTCCCTTTCATAA ACAGATTTAGTACTTGGATGCTTAATAGTGAATGGCGAATGCAA AGGAACAATTTCGTTCATCTTTAACCCTTTCACTCGGGGTACACG TTCTGGAATGTACCCGCCCTGTTGCAACTCAGGTGGACCGGGCA ATTCTTGAACTTTCTGTAACGTTGTTGGATGTTCAACCAGAAATT GTCCTACCAACTGTATTAGTTTCCTTTTGGTCTTATATTGTTCATC GAGATACTTCCCACTCTCCTTGATAGCCACTCTCACTCTTCCTGG ATTACCAAAATCTTGAGGATGAGTCTTTTCAGGCTCCAGGATGCA AGGTATATCCAAGTACCTGCAAGCATCTAATATTGTCTTTGCCAG GGGGTTCTCCACACCATACTCCTTTTGGCGCATGC
61	Sequence of the PpURA5 auxotrophic marker:	TCTAGAGGGACTTATCTGGGTCCAGACGATGTGTATCAAAAGAC AAATTAGAGTATTTATAAAGTTATGTAAGCAAATAGGGGCTAAT AGGGAAAGAAAAATTTTGGTTCTTTATCAGAGCTGGCTCGCGCG CAGTGTTTTTCGTGCTCCTTTGTAATAGTCATTTTTGACTACTGTT CAGATTGAAATCACATTGAAGATGTCACTGGAGGGGTACCAAAA AAGGTTTTTGGATGCTGCAGTGGCTTCGCAGGCCTTGAAGTTTGG AACTTTCACCTTGAAAAGTGGAAGACAGTCTCCATACTTCTTTAA CATGGGTCTTTTCAACAAAGCTCCATTAGTGAGTCAGCTGGCTGA ATCTTATGCTCAGGCCATCATTAACAGCAACCTGGAGATAGACG TTGTATTTGGACCAGCTTATAAAGGTATTCCTTTGGCTGCTATTA CCGTGTTGAAGTTGTACGAGCTGGGCGGCAAAAAATACGAAAAT GTCGGATATGCGTTCAATAGAAAAGAAAAGAAAGACCACGGAG AAGGTGGAAGCATCGTTGGAGAAAGTCTAAAGAATAAAAGAGT ACTGATTATCGATGATGTGATGACTGCAGGTACTGCTATCAACGA AGCATTTGCTATAATTGGAGCTGAAGGTGGGAGAGTTGAAGGTT GTATTATTGCCCTAGATAGAATGGAGACTACAGGAGATGACTCA AATACCAGTGCTACCCAGGCTGTTAGTCAGAGATATGGTACCCC TGTCTTGAGTATAGTGACATTGGACCATATTGTGGCCCATTTGGG CGAAACTTTCACAGCAGACGAGAAATCTCAAATGGAAACGTATA

		GAAAAAAGTATTTGCCCAAATAAGTATGAATCTGCTTCGAATGA ATGAATTAATCCAATTATCTTCTCACCATTATTTTCTTCTGTTTCG GAGCTTTGGGCACGGCGGCGGATCC
62	Sequence of the part of the Ec lacZ gene that was used to construct the PpURA5 blaster (recyclable auxotrophi c marker)	CCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCG GTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATT GAACTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTG GCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGTCAG AAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAA AACCTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCAT CTGACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAA GCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGTG GATTGGCGATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGT TCACCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCG ACCCGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGC GGGCCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGCAG ATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGC AGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGG ATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTG GCGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCA G
63	PpURA5 amino acid sequence	MSLEGYQKRFLDAAVASQALKFGTFTLKSGRQSPYFFNMGLFNKAP LVSQLAESYAQAIINSNLEIDVVFGPAYKGIPLAAITVLKLYELGGKK YENVGYAFNRKEKKDHGEGGSIVGESLKNKRVLIIDDVMTAGTAIN EAFAIIGAEGGRVEGCIIALDRMETTGDDSNTSATQAVSQRYGTPVL SIVTLDHIVAHLGETFTADEKSQMETYRKKYLPKZ
64	Sequence of the 5'-Region used for knock out of PpOCH1:	AAAACCTTTTTTCCTATTCAAACACAAGGCATTGCTTCAACACGT GTGCGTATCCTTAACACAGATACTCCATACTTCTAATAATGTGAT AGACGAATACAAAGATGTTCACTCTGTGTTGTGTCTACAAGCATT TCTTATTCTGATTGGGGATATTCTAGTTACAGCACTAAACAACTG GCGATACAAACTTAAATTAAATAATCCGAATCTAGAAAATGAAC TTTTGGATGGTCCGCCTGTTGGTTGGATAAATCAATACCGATTAA ATGGATTCTATTCCAATGAGAGAGTAATCCAAGACACTCTGATGT CAATAATCATTTGCTTGCAACAACAAACCCGTCATCTAATCAAA GGGTTTGATGAGGCTTACCTTCAATTGCAGATAAACTCATTGCTG TCCACTGCTGTATTATGTGAGAATATGGGTGATGAATCTGGTCTT CTCCACTCAGCTAACATGGCTGTTTGGGCAAAGGTGGTACAATT ATACGGAGATCAGGCAATAGTGAAATTGTTGAATATGGCTACTG GACGATGCTTCAAGGATGTACGTCTAGTAGGAGCCGTGGGAAGA TTGCTGGCAGAACCAGTTGGCACGTCGCAACAATCCCCAAGAAA TGAAATAAGTGAAAACGTAACGTCAAAGACAGCAATGGAGTCA ATATTGATAACACCACTGGCAGAGCGGTTCGTACGTCGTTTTGGA GCCGATATGAGGCTCAGCGTGCTAACAGCACGATTGACAAGAAG ACTCTCGAGTGACAGTAGGTTGAGTAAAGTATTCGCTTAGATTCC CAACCTTCGTTTTATTCTTTCGTAGACAAAGAAGCTGCATGCGAA CATAGGGACAACTTTTATAAATCCAATTGTCAAACCAACGTAAA ACCCTCTGGCACCATTTTCAACATATATTTGTGAAGCAGTACGCA ATATCGATAAATACTCACCGTTGTTTGTAACAGCCCCAACTTGCA

		TACGCCTTCTAATGACCTCAAATGGATAAGCCGCAGCTTGTGCTA ACATACCAGCAGCACCGCCCGCGGTCAGCTGCGCCCACACATAT AAAGGCAATCTACGATCATGGGAGGAATTAGTTTTGACCGTCAG GTCTTCAAGAGTTTTGAACTCTTCTTCTTGAACTGTGTAACCTTTT AAATGACGGGATCTAAATACGTCATGGATGAGATCATGTGTGTA AAAACTGACTCCAGCATATGGAATCATTCCAAAGATTGTAGGAG CGAACCCACGATAAAAGTTTCCCAACCTTGCCAAAGTGTCTAAT GCTGTGACTTGAAATCTGGGTTCCTCGTTGAAGACCCTGCGTACT ATGCCCAAAAACTTTCCTCCACGAGCCCTATTAACTTCTCTATGA GTTTCAAATGCCAAACGGACACGGATTAGGTCCAATGGGTAAGT GAAAAACACAGAGCAAACCCCAGCTAATGAGCCGGCCAGTAAC CGTCTTGGAGCTGTTTCATAAGAGTCATTAGGGATCAATAACGTT CTAATCTGTTCATAACATACAAATTTTATGGCTGCATAGGGAAAA ATTCTCAACAGGGTAGCCGAATGACCCTGATATAGACCTGCGAC ACCATCATACCCATAGATCTGCCTGACAGCCTTAAAGAGCCCGC TAAAAGACCCGGAAAACCGAGAGAACTCTGGATTAGCAGTCTGA AAAAGAATCTTCACTCTGTCTAGTGGAGCAATTAATGTCTTAGCG GCACTTCCTGCTACTCCGCCAGCTACTCCTGAATAGATCACATAC TGCAAAGACTGCTTGTCGATGACCTTGGGGTTATTTAGCTTCAAG GGCAATTTTTGGGACATTTTGGACACAGGAGACTCAGAAACAGA CACAGAGCGTTCTGAGTCCTGGTGCTCCTGACGTAGGCCTAGAA CAGGAATTATTGGCTTTATTTGTTTGTCCATTTCATAGGCTTGGG GTAATAGATAGATGACAGAGAAATAGAGAAGACCTAATATTTTT TGTTCATGGCAAATCGCGGGTTCGCGGTCGGGTCACACACGGAG AAGTAATGAGAAGAGCTGGTAATCTGGGGTAAAAGGGTTCAAAA GAAGGTCGCCTGGTAGGGATGCAATACAAGGTTGTCTTGGAGTT TACATTGACCAGATGATTTGGCTTTTTCTCTGTTCAATTCACATTT TTCAGCGAGAATCGGATTGACGGAGAAATGGCGGGGTGTGGGGT GGATAGATGGCAGAAATGCTCGCAATCACCGCGAAAGAAAGACT TTATGGAATAGAACTACTGGGTGGTGTAAGGATTACATAGCTAG TCCAATGGAGTCCGTTGGAAAGGTAAGAAGAAGCTAAAACCGGC TAAGTAACTAGGGAAGAATGATCAGACTTTGATTTGATGAGGTC TGAAAATACTCTGCTGCTTTTTCAGTTGCTTTTTCCCTGCAACCTA TCATTTTCCTTTTCATAAGCCTGCCTTTTCTGTTTTCACTTATATG AGTTCCGCCGAGACTTCCCCAAATTCTCTCCTGGAACATTCTCTA TCGCTCTCCTTCCAAGTTGCGCCCCCTGGCACTGCCTAGTAATAT TACCACGCGACTTATATTCAGTTCCACAATTTCCAGTGTTCGTAG CAAATATCATCAGCCATGGCGAAGGCAGATGGCAGTTTGCTCTA CTATAATCCTCACAATCCACCCAGAAGGTATTACTTCTACATGGC TATATTCGCCGTTTCTGTCATTTGCGTTTTGTACGGACCCTCACAA CAATTATCATCTCCAAAAATAGACTATGATCCATTGACGCTCCGA TCACTTGATTTGAAGACTTTGGAAGCTCCTTCACAGTTGAGTCCA GGCACCGTAGAAGATAATCTTCG
65	Sequence of the 3'-Region used for knock out	AAAGCTAGAGTAAAATAGATATAGCGAGATTAGAGAATGAATAC CTTCTTCTAAGCGATCGTCCGTCATCATAGAATATCATGGACTGT ATAGTTTTTTTTTTGTACATATAATGATTAAACGGTCATCCAACA TCTCGTTGACAGATCTCTCAGTACGCGAAATCCCTGACTATCAAA GCAAGAACCGATGAAGAAAAAAACAACAGTAACCCAAACACCA

	of PpOCH1:	CAACAAACACTTTATCTTCTCCCCCCCAACACCAATCATCAAAGA GATGTCGGAACCAAACACCAAGAAGCAAAAACTAACCCCATATA AAAACATCCTGGTAGATAATGCTGGTAACCCGCTCTCCTTCCATA TTCTGGGCTACTTCACGAAGTCTGACCGGTCTCAGTTGATCAACA TGATCCTCGAAATGGGTGGCAAGATCGTTCCAGACCTGCCTCCTC TGGTAGATGGAGTGTTGTTTTTGACAGGGGATTACAAGTCTATTG ATGAAGATACCCTAAAGCAACTGGGGGACGTTCCAATATACAGA GACTCCTTCATCTACCAGTGTTTTGTGCACAAGACATCTCTTCCC ATTGACACTTTCCGAATTGACAAGAACGTCGACTTGGCTCAAGA TTTGATCAATAGGGCCCTTCAAGAGTCTGTGGATCATGTCACTTC TGCCAGCACAGCTGCAGCTGCTGCTGTTGTTGTCGCTACCAACGG CCTGTCTTCTAAACCAGACGCTCGTACTAGCAAAATACAGTTCAC TCCCGAAGAAGATCGTTTTATTCTTGACTTTGTTAGGAGAAATCC TAAACGAAGAAACACACATCAACTGTACACTGAGCTCGCTCAGC ACATGAAAAACCATACGAATCATTCTATCCGCCACAGATTTCGTC GTAATCTTTCCGCTCAACTTGATTGGGTTTATGATATCGATCCAT TGACCAACCAACCTCGAAAAGATGAAAACGGGAACTACATCAA GGTACAAGGCCTTCCA
66	Sequence of the 5'-Region used for knock out of PpBMT2:	GGCCGAGCGGGCCTAGATTTTCACTACAAATTTCAAAACTACGC GGATTTATTGTCTCAGAGAGCAATTTGGCATTTCTGAGCGTAGCA GGAGGCTTCATAAGATTGTATAGGACCGTACCAACAAATTGCCG AGGCACAACACGGTATGCTGTGCACTTATGTGGCTACTTCCCTAC AACGGAATGAAACCTTCCTCTTTCCGCTTAAACGAGAAAGTGTG TCGCAATTGAATGCAGGTGCCTGTGCGCCTTGGTGTATTGTTTTT GAGGGCCCAATTTATCAGGCGCCTTTTTTCTTGGTTGTTTTCCCTT AGCCTCAAGCAAGGTTGGTCTATTTCATCTCCGCTTCTATACCGT GCCTGATACTGTTGGATGAGAACACGACTCAACTTCCTGCTGCTC TGTATTGCCAGTGTTTTGTCTGTGATTTGGATCGGAGTCCTCCTTA CTTGGAATGATAATAATCTTGGCGGAATCTCCCTAAACGGAGGC AAGGATTCTGCCTATGATGATCTGCTATCATTGGGAAGCTTCAAC GACATGGAGGTCGACTCCTATGTCACCAACATCTACGACAATGC TCCAGTGCTAGGATGTACGGATTTGTCTTATCATGGATTGTTGAA AGTCACCCCAAAGCATGACTTAGCTTGCGATTTGGAGTTCATAA GAGCTCAGATTTTGGACATTGACGTTTACTCCGCCATAAAAGACT TAGAAGATAAAGCCTTGACTGTAAAACAAAAGGTTGAAAAACAC TGGTTTACGTTTTATGGTAGTTCAGTCTTTCTGCCCGAACACGAT GTGCATTACCTGGTTAGACGAGTCATCTTTTCGGCTGAAGGAAAG GCGAACTCTCCAGTAACATC
67	Sequence of the 3'-Region used for knock out of PpBMT2:	CCATATGATGGGTGTTTGCTCACTCGTATGGATCAAAATTCCATG GTTTCTTCTGTACAACTTGTACACTTATTTGGACTTTTCTAACGGT TTTTCTGGTGATTTGAGAAGTCCTTATTTTGGTGTTCGCAGCTTAT CCGTGATTGAACCATCAGAAATACTGCAGCTCGTTATCTAGTTTC AGAATGTGTTGTAGAATACAATCAATTCTGAGTCTAGTTTGGGTG GGTCTTGGCGACGGGACCGTTATATGCATCTATGCAGTGTTAAGG TACATAGAATGAAAATGTAGGGGTTAATCGAAAGCATCGTTAAT TTCAGTAGAACGTAGTTCTATTCCCTACCCAAATAATTTGCCAAG AATGCTTCGTATCCACATACGCAGTGGACGTAGCAAATTTCACTT TGGACTGTGACCTCAAGTCGTTATCTTCTACTTGGACATTGATGG

		TCATTACGTAATCCACAAAGAATTGGATAGCCTCTCGTTTTATCT AGTGCACAGCCTAATAGCACTTAAGTAAGAGCAATGGACAAATT TGCATAGACATTGAGCTAGATACGTAACTCAGATCTTGTTCACTC ATGGTGTACTCGAAGTACTGCTGGAACCGTTACCTCTTATCATTT CGCTACTGGCTCGTGAAACTACTGGATGAAAAAAAAAAAAGAGC TGAAAGCGAGATCATCCCATTTTGTCATCATACAAATTCACGCTT GCAGTTTTGCTTCGTTAACAAGACAAGATGTCTTTATCAAAGACC CGTTTTTTCTTCTTGAAGAATACTTCCCTGTTGAGCACATGCAAA CCATATTTATCTCAGATTTCACTCAACTTGGGTGCTTCCAAGAGA AGTAAAATTCTTCCCACTGCATCAACTTCCAAGAAACCCGTAGA CCAGTTTCTCTTCAGCCAAAAGAAGTTGCTCGCCGATCACCGCGG TAACAGAGGAGTCAGAAGGTTTCACACCCTTCCATCCCGATTTCA AAGTCAAAGTGCTGCGTTGAACCAAGGTTTTCAGGTTGCCAAAG CCCAGTCTGCAAAAACTAGTTCCAAATGGCCTATTAATTCCCATA AAAGTGTTGGCTACGTATGTATCGGTACCTCCATTCTGGTATTTG CTATTGTTGTCGTTGGTGGGTTGACTAGACTGACCGAATCCGGTC TTTCCATAACGGAGTGGAAACCTATCACTGGTTCGGTTCCCCCAC TGACTGAGGAAGACTGGAAGTTGGAATTTGAAAAATACAAACAA AGCCCTGAGTTTCAGGAACTAAATTCTCACATAACATTGGAAGA GTTCAAGTTTATATTTTCCATGGAATGGGGACATAGATTGTTGGG AAGGGTCATCGGCCTGTCGTTTGTTCTTCCCACGTTTTACTTCATT GCCCGTCGAAAGTGTTCCAAAGATGTTGCATTGAAACTGCTTGC AATATGCTCTATGATAGGATTCCAAGGTTTCATCGGCTGGTGGAT GGTGTATTCCGGATTGGACAAACAGCAATTGGCTGAACGTAACT CCAAACCAACTGTGTCTCCATATCGCTTAACTACCCATCTTGGAA CTGCATTTGTTATTTACTGTTACATGATTTACACAGGGCTTCAAG TTTTGAAGAACTATAAGATCATGAAACAGCCTGAAGCGTATGTT CAAATTTTCAAGCAAATTGCGTCTCCAAAATTGAAAACTTTCAAG AGACTCTCTTCAGTTCTATTAGGCCTGGTG
68	Sequence of the 5'-Region used for knock out of BMT1	CATATGGTGAGAGCCGTTCTGCACAACTAGATGTTTTCGAGCTTC GCATTGTTTCCTGCAGCTCGACTATTGAATTAAGATTTCCGGATA TCTCCAATCTCACAAAAACTTATGTTGACCACGTGCTTTCCTGAG GCGAGGTGTTTTATATGCAAGCTGCCAAAAATGGAAAACGAATG GCCATTTTTCGCCCAGGCAAATTATTCGATTACTGCTGTCATAAA GACAGTGTTGCAAGGCTCACATTTTTTTTTAGGATCCGAGATAAA GTGAATACAGGACAGCTTATCTCTATATCTTGTACCATTCGTGAA TCTTAAGAGTTCGGTTAGGGGGACTCTAGTTGAGGGTTGGCACTC ACGTATGGCTGGGCGCAGAAATAAAATTCAGGCGCAGCAGCACT TATCGATG
69	Sequence of the 3'-Region used for knock out of BMT1	GAATTCACAGTTATAAATAAAAACAAAAACTCAAAAAGTTTGGG CTCCACAAAATAACTTAATTTAAATTTTTGTCTAATAAATGAATG TAATTCCAAGATTATGTGATGCAAGCACAGTATGCTTCAGCCCTA TGCAGCTACTAATGTCAATCTCGCCTGCGAGCGGGCCTAGATTTT CACTACAAATTTCAAAACTACGCGGATTTATTGTCTCAGAGAGC AATTTGGCATTTCTGAGCGTAGCAGGAGGCTTCATAAGATTGTAT AGGACCGTACCAACAAATTGCCGAGGCACAACACGGTATGCTGT GCACTTATGTGGCTACTTCCCTACAACGGAATGAAACCTTCCTCT TTCCGCTTAAACGAGAAAGTGTGTCGCAATTGAATGCAGGTGCC TGTGCGCCTTGGTGTATTGTTTTTGAGGGCCCAATTTATCAGGCG

		CCTTTTTTCTTGGTTGTTTTCCCTTAGCCTCAAGCAAGGTTGGTCT ATTTCATCTCCGCTTCTATACCGTGCCTGATACTGTTGGATGAGA ACACGACTCAACTTCCTGCTGCTCTGTATTGCCAGTGTTTTGTCT GTGATTTGGATCGGAGTCCTCCTTACTTGGAATGATAATAATCTT GGCGGAATCTCCCTAAACGGAGGCAAGGATTCTGCCTATGATGA TCTGCTATCATTGGGAAGCTT
70	Sequence of the 5'-Region used for knock out of BMT3	GATATCTCCCTGGGGACAATATGTGTTGCAACTGTTCGTTGTTGG TGCCCCAGTCCCCCAACCGGTACTAATCGGTCTATGTTCCCGTAA CTCATATTCGGTTAGAACTAGAACAATAAGTGCATCATTGTTCAA CATTGTGGTTCAATTGTCGAACATTGCTGGTGCTTATATCTACAG GGAAGACGATAAGCCTTTGTACAAGAGAGGTAACAGACAGTTAA TTGGTATTTCTTTGGGAGTCGTTGCCCTCTACGTTGTCTCCAAGA CATACTACATTCTGAGAAACAGATGGAAGACTCAAAAATGGGAG AAGCTTAGTGAAGAAGAGAAAGTTGCCTACTTGGACAGAGCTGA GAAGGAGAACCTGGGTTCTAAGAGGCTGGACTTTTTGTTCGAGA GTTAAACTGCATAATTTTTTCTAAGTAAATTTCATAGTTATGAAA TTTCTGCAGCTTAGTGTTTACTGCATCGTTTACTGCATCACCCTGT AAATAATGTGAGCTTTTTTCCTTCCATTGCTTGGTATCTTCCTTGC TGCTGTTT
71	Sequence of the 3'-Region used for knock out of BMT3	ACAAAACAGTCATGTACAGAACTAACGCCTTTAAGATGCAGACC ACTGAAAAGAATTGGGTCCCATTTTTCTTGAAAGACGACCAGGA ATCTGTCCATTTTGTTTACTCGTTCAATCCTCTGAGAGTACTCAAC TGCAGTCTTGATAACGGTGCATGTGATGTTCTATTTGAGTTACCA CATGATTTTGGCATGTCTTCCGAGCTACGTGGTGCCACTCCTATG CTCAATCTTCCTCAGGCAATCCCGATGGCAGACGACAAAGAAAT TTGGGTTTCATTCCCAAGAACGAGAATATCAGATTGCGGGTGTTC TGAAACAATGTACAGGCCAATGTTAATGCTTTTTGTTAGAGAAG GAACAAACTTTTTTGCTGAGC
72	Sequence of the 5'-Region used for knock out of BMT4	AAGCTTGTTCACCGTTGGGACTTTTCCGTGGACAATGTTGACTAC TCCAGGAGGGATTCCAGCTTTCTCTACTAGCTCAGCAATAATCAA TGCAGCCCCAGGCGCCCGTTCTGATGGCTTGATGACCGTTGTATT GCCTGTCACTATAGCCAGGGGTAGGGTCCATAAAGGAATCATAG CAGGGAAATTAAAAGGGCATATTGATGCAATCACTCCCAATGGC TCTCTTGCCATTGAAGTCTCCATATCAGCACTAACTTCCAAGAAG GACCCCTTCAAGTCTGACGTGATAGAGCACGCTTGCTCTGCCACC TGTAGTCCTCTCAAAACGTCACCTTGTGCATCAGCAAAGACTTTA CCTTGCTCCAATACTATGACGGAGGCAATTCTGTCAAAATTCTCT CTCAGCAATTCAACCAACTTGAAAGCAAATTGCTGTCTCTTGATG ATGGAGACTTTTTTCCAAGATTGAAATGCAATGTGGGACGACTC AATTGCTTCTTCCAGCTCCTCTTCGGTTGATTGAGGAACTTTTGA AACCACAAAATTGGTCGTTGGGTCATGTACATCAAACCATTCTGT AGATTTAGATTCGACGAAAGCGTTGTTGATGAAGGAAAAGGTTG GATACGGTTTGTCGGTCTCTTTGGTATGGCCGGTGGGGTATGCAA TTGCAGTAGAAGATAATTGGACAGCCATTGTTGAAGGTAGAGAA AAGGTCAGGGAACTTGGGGGTTATTTATACCATTTTACCCCACAA ATAACAACTGAAAAGTACCCATTCCATAGTGAGAGGTAACCGAC GGAAAAAGACGGGCCCATGTTCTGGGACCAATAGAACTGTGTAA TCCATTGGGACTAATCAACAGACGATTGGCAATATAATGAAATA GTTCGTTGAAAAGCCACGTCAGCTGTCTTTTCATTAACTTTGGTC

		GGACACAACATTTTCTACTGTTGTATCTGTCCTACTTTGCTTATCA TCTGCCACAGGGCAAGTGGATTTCCTTCTCGCGCGGCTGGGTGA AAACGGTTAACGTGAA
73	Sequence of the 3'-Region used for knock out of BMT4	GCCTTGGGGGACTTCAAGTCTTTGCTAGAAACTAGATGAGGTCA GGCCCTCTTATGGTTGTGTCCCAATTGGGCAATTTCACTCACCTA AAAAGCATGACAATTATTTAGCGAAATAGGTAGTATATTTTCCCT CATCTCCCAAGCAGTTTCGTTTTTGCATCCATATCTCTCAAATGA GCAGCTACGACTCATTAGAACCAGAGTCAAGTAGGGGTGAGCTC AGTCATCAGCCTTCGTTTCTAAAACGATTGAGTTCTTTTGTTGCT ACAGGAAGCGCCCTAGGGAACTTTCGCACTTTGGAAATAGATTT TGATGACCAAGAGCGGGAGTTGATATTAGAGAGGCTGTCCAAAG TACATGGGATCAGGCCGGCCAAATTGATTGGTGTGACTAAACCA TTGTGTACTTGGACACTCTATTACAAAAGCGAAGATGATTTGAAG TATTACAAGTCCCGAAGTGTTAGAGGATTCTATCGAGCCCAGAA TGAAATCATCAACCGTTATCAGCAGATTGATAAACTCTTGGAAA GCGGTATCCCATTTTCATTATTGAAGAACTACGATAATGAAGATG TGAGAGACGGCGACCCTCTGAACGTAGACGAAGAAACAAATCTA CTTTTGGGGTACAATAGAGAAAGTGAATCAAGGGAGGTATTTGT GGCCATAATACTCAACTCTATCATTAATG
74	Sequence of the 5'-Region used for knock out of PpPNO1 and PpMNN4:	TCATTCTATATGTTCAAGAAAAGGGTAGTGAAAGGAAAGAAAAG GCATATAGGCGAGGGAGAGTTAGCTAGCATACAAGATAATGAAG GATCAATAGCGGTAGTTAAAGTGCACAAGAAAAGAGCACCTGTT GAGGCTGATGATAAAGCTCCAATTACATTGCCACAGAGAAACAC AGTAACAGAAATAGGAGGGGATGCACCACGAGAAGAGCATTCA GTGAACAACTTTGCCAAATTCATAACCCCAAGCGCTAATAAGCC AATGTCAAAGTCGGCTACTAACATTAATAGTACAACAACTATCG ATTTTCAACCAGATGTTTGCAAGGACTACAAACAGACAGGTTAC TGCGGATATGGTGACACTTGTAAGTTTTTGCACCTGAGGGATGAT TTCAAACAGGGATGGAAATTAGATAGGGAGTGGGAAAATGTCCA AAAGAAGAAGCATAATACTCTCAAAGGGGTTAAGGAGATCCAA ATGTTTAATGAAGATGAGCTCAAAGATATCCCGTTTAAATGCATT ATATGCAAAGGAGATTACAAATCACCCGTGAAAACTTCTTGCAA TCATTATTTTTGCGAACAATGTTTCCTGCAACGGTCAAGAAGAAA ACCAAATTGTATTATATGTGGCAGAGACACTTTAGGAGTTGCTTT ACCAGCAAAGAAGTTGTCCCAATTTCTGGCTAAGATACATAATA ATGAAAGTAATAAAGTTTAGTAATTGCATTGCGTTGACTATTGAT TGCATTGATGTCGTGTGATACTTTCACCGAAAAAAAACACGAAG CGCAATAGGAGCGGTTGCATATTAGTCCCCAAAGCTATTTAATTG TGCCTGAAACTGTTTTTTAAGCTCATCAAGCATAATTGTATGCAT TGCGACGTAACCAACGTTTAGGCGCAGTTTAATCATAGCCCACT GCTAAGCC
75	Sequence of the 3'-Region used for knock out of PpPNO1 and	CGGAGGAATGCAAATAATAATCTCCTTAATTACCCACTGATAAG CTCAAGAGACGCGGTTTGAAAACGATATAATGAATCATTTGGAT TTTATAATAAACCCTGACAGTTTTTCCACTGTATTGTTTTAACACT CATTGGAAGCTGTATTGATTCTAAGAAGCTAGAAATCAATACGG CCATACAAAAGATGACATTGAATAAGCACCGGCTTTTTTGATTAG CATATACCTTAAAGCATGCATTCATGGCTACATAGTTGTTAAAGG GCTTCTTCCATTATCAGTATAATGAATTACATAATCATGCACTTA

	PpMNN4:	TATTTGCCCATCTCTGTTCTCTCACTCTTGCCTGGGTATATTCTAT GAAATTGCGTATAGCGTGTCTCCAGTTGAACCCCAAGCTTGGCG AGTTTGAAGAGAATGCTAACCTTGCGTATTCCTTGCTTCAGGAAA CATTCAAGGAGAAACAGGTCAAGAAGCCAAACATTTTGATCCTT CCCGAGTTAGCATTGACTGGCTACAATTTTCAAAGCCAGCAGCG GATAGAGCCTTTTTTGGAGGAAACAACCAAGGGAGCTAGTACCC AATGGGCTCAAAAAGTATCCAAGACGTGGGATTGCTTTACTTTA ATAGGATACCCAGAAAAAAGTTTAGAGAGCCCTCCCCGTATTTA CAACAGTGCGGTACTTGTATCGCCTCAGGGAAAAGTAATGAACA ACTACAGAAAGTCCTTCTTGTATGAAGCTGATGAACATTGGGGA TGTTCGGAATCTTCTGATGGGTTTCAAACAGTAGATTTATTAATT GAAGGAAAGACTGTAAAGACATCATTTGGAATTTGCATGGATTT GAATCCTTATAAATTTGAAGCTCCATTCACAGACTTCGAGTTCAG TGGCCATTGCTTGAAAACCGGTACAAGACTCATTTTGTGCCCAAT GGCCTGGTTGTCCCCTCTATCGCCTTCCATTAAAAAGGATCTTAG TGATATAGAGAAAAGCAGACTTCAAAAGTTCTACCTTGAAAAAA TAGATACCCCGGAATTTGACGTTAATTACGAATTGAAAAAAGAT GAAGTATTGCCCACCCGTATGAATGAAACGTTGGAAACAATTGA CTTTGAGCCTTCAAAACCGGACTACTCTAATATAAATTATTGGAT ACTAAGGTTTTTTCCCTTTCTGACTCATGTCTATAAACGAGATGT GCTCAAAGAGAATGCAGTTGCAGTCTTATGCAACCGAGTTGGCA TTGAGAGTGATGTCTTGTACGGAGGATCAACCACGATTCTAAACT TCAATGGTAAGTTAGCATCGACACAAGAGGAGCTGGAGTTGTAC GGGCAGACTAATAGTCTCAACCCCAGTGTGGAAGTATTGGGGGC CCTTGGCATGGGTCAACAGGGAATTCTAGTACGAGACATTGAAT TAACATAATATACAATATACAATAAACACAAATAAAGAATACAA GCCTGACAAAAATTCACAAATTATTGCCTAGACTTGTCGTTATCA GCAGCGACCTTTTTCCAATGCTCAATTTCACGATATGCCTTTTCT AGCTCTGCTTTAAGCTTCTCATTGGAATTGGCTAACTCGTTGACT GCTTGGTCAGTGATGAGTTTCTCCAAGGTCCATTTCTCGATGTTG TTGTTTTCGTTTTCCTTTAATCTCTTGATATAATCAACAGCCTTCT TTAATATCTGAGCCTTGTTCGAGTCCCCTGTTGGCAACAGAGCGG CCAGTTCCTTTATTCCGTGGTTTATATTTTCTCTTCTACGCCTTTCT ACTTCTTTGTGATTCTCTTTACGCATCTTATGCCATTCTTCAGAAC CAGTGGCTGGCTTAACCGAATAGCCAGAGCCTGAAGAAGCCGCA CTAGAAGAAGCAGTGGCATTGTTGACTATGG
76	Sequence of the 5'-Region used for knock out of PpMNN4L 1:	GATCTGGCCATTGTGAAACTTGACACTAAAGACAAAACTCTTAG AGTTTCCAATCACTTAGGAGACGATGTTTCCTACAACGAGTACG ATCCCTCATTGATCATGAGCAATTTGTATGTGAAAAAAGTCATCG ACCTTGACACCTTGGATAAAAGGGCTGGAGGAGGTGGAACCACC TGTGCAGGCGGTCTGAAAGTGTTCAAGTACGGATCTACTACCAA ATATACATCTGGTAACCTGAACGGCGTCAGGTTAGTATACTGGA ACGAAGGAAAGTTGCAAAGCTCCAAATTTGTGGTTCGATCCTCT AATTACTCTCAAAAGCTTGGAGGAAACAGCAACGCCGAATCAAT TGACAACAATGGTGTGGGTTTTGCCTCAGCTGGAGACTCAGGCG CATGGATTCTTTCCAAGCTACAAGATGTTAGGGAGTACCAGTCAT TCACTGAAAAGCTAGGTGAAGCTACGATGAGCATTTTCGATTTCC ACGGTCTTAAACAGGAGACTTCTACTACAGGGCTTGGGGTAGTT GGTATGATTCATTCTTACGACGGTGAGTTCAAACAGTTTGGTTTG

		TTCACTCCAATGACATCTATTCTACAAAGACTTCAACGAGTGACC AATGTAGAATGGTGTGTAGCGGGTTGCGAAGATGGGGATGTGGA CACTGAAGGAGAACACGAATTGAGTGATTTGGAACAACTGCATA TGCATAGTGATTCCGACTAGTCAGGCAAGAGAGAGCCCTCAAAT TTACCTCTCTGCCCCTCCTCACTCCTTTTGGTACGCATAATTGCAG TATAAAGAACTTGCTGCCAGCCAGTAATCTTATTTCATACGCAGT TCTATATAGCACATAATCTTGCTTGTATGTATGAAATTTACCGCG TTTTAGTTGAAATTGTTTATGTTGTGTGCCTTGCATGAAATCTCTC GTTAGCCCTATCCTTACATTTAACTGGTCTCAAAACCTCTACCAA TTCCATTGCTGTACAACAATATGAGGCGGCATTACTGTAGGGTTG GAAAAAAATTGTCATTCCAGCTAGAGATCACACGACTTCATCAC GCTTATTGCTCCTCATTGCTAAATCATTTACTCTTGACTTCGACCC AGAAAAGTTCGCC
77	Sequence of the 3'-Region used for knock out of PpMNN4L 1:	GCATGTCAAACTTGAACACAACGACTAGATAGTTGTTTTTTCTAT ATAAAACGAAACGTTATCATCTTTAATAATCATTGAGGTTTACCC TTATAGTTCCGTATTTTCGTTTCCAAACTTAGTAATCTTTTGGAAA TATCATCAAAGCTGGTGCCAATCTTCTTGTTTGAAGTTTCAAACT GCTCCACCAAGCTACTTAGAGACTGTTCTAGGTCTGAAGCAACTT CGAACACAGAGACAGCTGCCGCCGATTGTTCTTTTTTGTGTTTTT CTTCTGGAAGAGGGGCATCATCTTGTATGTCCAATGCCCGTATCC TTTCTGAGTTGTCCGACACATTGTCCTTCGAAGAGTTTCCTGACA TTGGGCTTCTTCTATCCGTGTATTAATTTTGGGTTAAGTTCCTCGT TTGCATAGCAGTGGATACCTCGATTTTTTTGGCTCCTATTTACCTG ACATAATATTCTACTATAATCCAACTTGGACGCGTCATCTATGAT AACTAGGCTCTCCTTTGTTCAAAGGGGACGTCTTCATAATCCACT GGCACGAAGTAAGTCTGCAACGAGGCGGCTTTTGCAACAGAACG ATAGTGTCGTTTCGTACTTGGACTATGCTAAACAAAAGGATCTGT CAAACATTTCAACCGTGTTTCAAGGCACTCTTTACGAATTATCGA CCAAGACCTTCCTAGACGAACATTTCAACATATCCAGGCTACTGC TTCAAGGTGGTGCAAATGATAAAGGTATAGATATTAGATGTGTTT GGGACCTAAAACAGTTCTTGCCTGAAGATTCCCTTGAGCAACAG GCTTCAATAGCCAAGTTAGAGAAGCAGTACCAAATCGGTAACAA AAGGGGGAAGCATATAAAACCTTTACTATTGCGACAAAATCCAT CCTTGAAAGTAAAGCTGTTTGTTCAATGTAAAGCATACGAAACG AAGGAGGTAGATCCTAAGATGGTTAGAGAACTTAACGGGACATA CTCCAGCTGCATCCCATATTACGATCGCTGGAAGACTTTTTTCAT GTACGTATCGCCCACCAACCTTTCAAAGCAAGCTAGGTATGATTT TGACAGTTCTCACAATCCATTGGTTTTCATGCAACTTGAAAAAAC CCAACTCAAACTTCATGGGGATCCATACAATGTAAATCATTACG AGAGGGCGAGGTTGAAAAGTTTCCATTGCAATCACGTCGCATCA TGGCTACTGAAAGGCCTTAAC
78	Sequence of the PpTRP2 gene integration locus:	TAATGGCCAAACGGTTTCTCAATTACTATATACTACTAACCATTT ACCTGTAGCGTATTTCTTTTCCCTCTTCGCGAAAGCTCAAGGGCA TCTTCTTGACTCATGAAAAATATCTGGATTTCTTCTGACAGATCA TCACCCTTGAGCCCAACTCTCTAGCCTATGAGTGTAAGTGATAGT CATCTTGCAACAGATTATTTTGGAACGCAACTAACAAAGCAGAT ACACCCTTCAGCAGAATCCTTTCTGGATATTGTGAAGAATGATCG CCAAAGTCACAGTCCTGAGACAGTTCCTAATCTTTACCCCATTTA CAAGTTCATCCAATCAGACTTCTTAACGCCTCATCTGGCTTATAT

		CAAGCTTACCAACAGTTCAGAAACTCCCAGTCCAAGTTTCTTGCT TGAAAGTGCGAAGAATGGTGACACCGTTGACAGGTACACCTTTA TGGGACATTCCCCCAGAAAAATAATCAAGACTGGGCCTTTAGAG GGTGCTGAAGTTGACCCCTTGGTGCTTCTGGAAAAAGAACTGAA GGGCACCAGACAAGCGCAACTTCCTGGTATTCCTCGTCTAAGTG GTGGTGCCATAGGATACATCTCGTACGATTGTATTAAGTACTTTG AACCAAAAACTGAAAGAAAACTGAAAGATGTTTTGCAACTTCCG GAAGCAGCTTTGATGTTGTTCGACACGATCGTGGCTTTTGACAAT GTTTATCAAAGATTCCAGGTAATTGGAAACGTTTCTCTATCCGTT GATGACTCGGACGAAGCTATTCTTGAGAAATATTATAAGACAAG AGAAGAAGTGGAAAAGATCAGTAAAGTGGTATTTGACAATAAA ACTGTTCCCTACTATGAACAGAAAGATATTATTCAAGGCCAAAC GTTCACCTCTAATATTGGTCAGGAAGGGTATGAAAACCATGTTCG CAAGCTGAAAGAACATATTCTGAAAGGAGACATCTTCCAAGCTG TTCCCTCTCAAAGGGTAGCCAGGCCGACCTCATTGCACCCTTTCA ACATCTATCGTCATTTGAGAACTGTCAATCCTTCTCCATACATGT TCTATATTGACTATCTAGACTTCCAAGTTGTTGGTGCTTCACCTG AATTACTAGTTAAATCCGACAACAACAACAAAATCATCACACAT CCTATTGCTGGAACTCTTCCCAGAGGTAAAACTATCGAAGAGGA CGACAATTATGCTAAGCAATTGAAGTCGTCTTTGAAAGACAGGG CCGAGCACGTCATGCTGGTAGATTTGGCCAGAAATGATATTAAC CGTGTGTGTGAGCCCACCAGTACCACGGTTGATCGTTTATTGACT GTGGAGAGATTTTCTCATGTGATGCATCTTGTGTCAGAAGTCAGT GGAACATTGAGACCAAACAAGACTCGCTTCGATGCTTTCAGATC CATTTTCCCAGCAGGAACCGTCTCCGGTGCTCCGAAGGTAAGAG CAATGCAACTCATAGGAGAATTGGAAGGAGAAAAGAGAGGTGT TTATGCGGGGGCCGTAGGACACTGGTCGTACGATGGAAAATCGA TGGACACATGTATTGCCTTAAGAACAATGGTCGTCAAGGACGGT GTCGCTTACCTTCAAGCCGGAGGTGGAATTGTCTACGATTCTGAC CCCTATGACGAGTACATCGAAACCATGAACAAAATGAGATCCAA CAATAACACCATCTTGGAGGCTGAGAAAATCTGGACCGATAGGT TGGCCAGAGACGAGAATCAAAGTGAATCCGAAGAAAACGATCA ATGAACGGAGGACGTAAGTAGGAATTTATGGTTTGGCCAT
79	Sequence of the 5'-Region used for knock out of PpARG1:	GATCTGGCCTTCCCTGAATTTTTACGTCCAGCTATACGATCCGTT GTGACTGTATTTCCTGAAATGAAGTTTCAACCTAAAGTTTTGGTT GTACTTGCTCCACCTACCACGGAAACTAATATCGAAACCAATGA AAAAGTAGAACTGGAATCGTCAATCGAAATTCGCAACCAAGTGG AACCCAAAGACTTGAATCTTTCTAAAGTCTATTCTAGTGACACTA ATGGCAACAGAAGATTTGAGCTGACTTTTCAAATGAATCTCAAT AATGCAATATCAACATCAGACAATCAATGGGCTTTGTCTAGTGA CACAGGATCAATTATAGTAGTGTCTTCTGCAGGAAGAATAACTTC CCCGATCCTAGAAGTCGGGGCATCCGTCTGTGTCTTAAGATCGTA CAACGAACACCTTTTGGCAATAACTTGTGAAGGAACATGCTTTTC ATGGAATTTAAAGAAGCAAGAATGTGTTCTAAACAGCATTTCAT TAGCACCTATAGTCAATTCACACATGCTAGTTAAGAAAGTTGGA GATGCAAGGAACTATTCTATTGTATCTGCCGAAGGAGACAACAA TCCGTTACCCCAGATTCTAGACTGCGAACTTTCCAAAAATGGCGC TCCAATTGTGGCTCTTAGCACGAAAGACATCTACTCTTATTCAAA GAAAATGAAATGCTGGATCCATTTGATTGATTCGAAATACTTTGA

		ATTGTTGGGTGCTGACAATGCACTGTTTGAGTGTGTGGAAGCGCT AGAAGGTCCAATTGGAATGCTAATTCATAGATTGGTAGATGAGT TCTTCCATGAAAACACTGCCGGTAAAAAACTCAAACTTTACAAC AAGCGAGTACTGGAGGACCTTTCAAATTCACTTGAAGAACTAGG TGAAAATGCGTCTCAATTAAGAGAGAAACTTGACAAACTCTATG GTGATGAGGTTGAGGCTTCTTGACCTCTTCTCTCTATCTGCGTTTC TTTTTTTTTTTTTTTTTTTTTTTTTTTCAGTTGAGCCAGACCGCGCT AAACGCATACCAATTGCCAAATCAGGCAATTGTGAGACAGTGGT AAAAAAGATGCCTGCAAAGTTAGATTCACACAGTAAGAGAGATC CTACTCATAAATGAGGCGCTTATTTAGTAGCTAGTGATAGCCACT GCGGTTCTGCTTTATGCTATTTGTTGTATGCCTTACTATCTTTGTT TGGCTCCTTTTTCTTGACGTTTTCCGTTGGAGGGACTCCCTATTCT GAGTCATGAGCCGCACAGATTATCGCCCAAAATTGACAAAATCT TCTGGCGAAAAAAGTATAAAAGGAGAAAAAAGCTCACCCTTTTC CAGCGTAGAAAGTATATATCAGTCATTGAAGAC
80	Sequence of the 3'-Region used for knock out of PpARG1:	GGGACTTTAACTCAAGTAAAAGGATAGTTGTACAATTATATATA CGAAGAATAAATCATTACAAAAAGTATTCGTTTCTTTGATTCTTA ACAGGATTCATTTTCTGGGTGTCATCAGGTACAGCGCTGAATATC TTGAAGTTAACATCGAGCTCATCATCGACGTTCATCACACTAGCC ACGTTTCCGCAACGGTAGCAATAATTAGGAGCGGACCACACAGT GACGACATCTTTCTCTTTGAAATGGTATCTGAAGCCTTCCATGAC CAATTGATGGGCTCTAGCGATGAGTTGCAAGTTATTAATGTGGTT GAACTCACGTGCTACTCGAGCACCGAATAACCAGCCAGCTCCAC GAGGAGAAACAGCCCAACTGTCGACTTCATCTGGGTCAGACCAA ACCAAGTCACAAAATCCTCCTTCATGAGGGACCTCTTGCGCTCGG CTGAGAACTCTGATTTGATCTAACATGCGAATATCGGGAGAGAG ACCACCATGGATACATAATATTTTACCATCAATGATGGCACTAA GGGTTAAAAAGTCGAACACCTGGCAACAGTACTTCCAGACAGTG GTGGAACCATATTTATTGAGACATTCCTCATAAAATCCATAAACC TGAGTGATCTGTCTGGATTCATGATTTCCCCTTACCAATGTGATA TGTTGAGGAAACTTAATTTTTAAAATCATGAGTAACGTGAACGTC TCCAACGAGAAATAGCCTCTATCCACATAGTCTCCTAGGAAGAT ATAGTTCTGTTTTATTCCATTAGAGGAGGATCCGGGAAACCCACC ACTAATCTTGAAAAGTTCCAGTAGATCGTGAAATTGGCCGTGAA TATCTCCGCATACTGTCACTGGACTCTGCACTGGCTGTATATTGG ATTCCTCCATCAGCAAATCCTTCACCCGTTCGCAAAGATGCTTCA TATCATTTTCACTTAAAGCCTTGCAGCTTTTGACTTCTTCAAACCA CTGATCTGGTCCTCTTTCTGGCATGATTAAGGTCTATAATATTTCT GAGCTGAGATGTAAAAAAAAATAATAAAAATGGGGAGTGAAAA AGTGTGTAGCTTTTAGGAGTTTGGGATTGATACCCCAAAATGATC TTTATGAGAATTAAAAGGTAGATACGCTTTTAATAAGAACACCT ATCTATAGTACTTTGTGGTCTTGAGTAATTGAGATGTTCAGCTTC TGAGGTTTGCCGTTATTCTGGGATAGTAGTGCGCGACCAAACAA CCCGCCAGGCAAAGTGTGTTGTGCTCGAAGACGATTGCCAGAAG AGTAAGTCCGTCCTGCCTCAGATGTTACACACTTTCTTCCCTAGA CAGTCGATGCATCATCGGATTTAAACCTGAAACTTTGATGCCATG ATACGCCTAGTCACGTCGACTGAGATTTTAGATAAGCCCCGATCC

		CTTTAGTACATTCCTGTTATCCATGGATGGAATGGCCTGATA
81	Sequence of the PpARG1 auxotrophic marker:	CAGTTGAGCCAGACCGCGCTAAACGCATACCAATTGCCAAATCA GGCAATTGTGAGACAGTGGTAAAAAAGATGCCTGCAAAGTTAGA TTCACACAGTAAGAGAGATCCTACTCATAAATGAGGCGCTTATTT AGTAGCTAGTGATAGCCACTGCGGTTCTGCTTTATGCTATTTGTT GTATGCCTTACTATCTTTGTTTGGCTCCTTTTTCTTGACGTTTTCC GTTGGAGGGACTCCCTATTCTGAGTCATGAGCCGCACAGATTATC GCCCAAAATTGACAAAATCTTCTGGCGAAAAAAGTATAAAAGGA GAAAAAAGCTCACCCTTTTCCAGCGTAGAAAGTATATATCAGTC ATTGAAGACTATTATTTAAATAACACAATGTCTAAAGGAAAAGT TTGTTTGGCCTACTCCGGTGGTTTGGATACCTCCATCATCCTAGC TTGGTTGTTGGAGCAGGGATACGAAGTCGTTGCCTTTTTAGCCAA CATTGGTCAAGAGGAAGACTTTGAGGCTGCTAGAGAGAAAGCTC TGAAGATCGGTGCTACCAAGTTTATCGTCAGTGACGTTAGGAAG GAATTTGTTGAGGAAGTTTTGTTCCCAGCAGTCCAAGTTAACGCT ATCTACGAGAACGTCTACTTACTGGGTACCTCTTTGGCCAGACCA GTCATTGCCAAGGCCCAAATAGAGGTTGCTGAACAAGAAGGTTG TTTTGCTGTTGCCCACGGTTGTACCGGAAAGGGTAACGATCAGGT TAGATTTGAGCTTTCCTTTTATGCTCTGAAGCCTGACGTTGTCTGT ATCGCCCCATGGAGAGACCCAGAATTCTTCGAAAGATTCGCTGG TAGAAATGACTTGCTGAATTACGCTGCTGAGAAGGATATTCCAG TTGCTCAGACTAAAGCCAAGCCATGGTCTACTGATGAGAACATG GCTCACATCTCCTTCGAGGCTGGTATTCTAGAAGATCCAAACACT ACTCCTCCAAAGGACATGTGGAAGCTCACTGTTGACCCAGAAGA TGCACCAGACAAGCCAGAGTTCTTTGACGTCCACTTTGAGAAGG GTAAGCCAGTTAAATTAGTTCTCGAGAACAAAACTGAGGTCACC GATCCGGTTGAGATCTTTTTGACTGCTAACGCCATTGCTAGAAGA AACGGTGTTGGTAGAATTGACATTGTCGAGAACAGATTCATCGG AATCAAGTCCAGAGGTTGTTATGAAACTCCAGGTTTGACTCTACT GAGAACCACTCACATCGACTTGGAAGGTCTTACCGTTGACCGTG AAGTTAGATCGATCAGAGACACTTTTGTTACCCCAACCTACTCTA AGTTGTTATACAACGGGTTGTACTTTACCCCAGAAGGTGAGTACG TCAGAACTATGATTCAGCCTTCTCAAAACACCGTCAACGGTGTTG TTAGAGCCAAGGCCTACAAAGGTAATGTGTATAACCTAGGAAGA TACTCTGAAACCGAGAAATTGTACGATGCTACCGAATCTTCCATG GATGAGTTGACCGGATTCCACCCTCAAGAAGCTGGAGGATTTAT CACAACACAAGCCATCAGAATCAAGAAGTACGGAGAAAGTGTC AGAGAGAAGGGAAAGTTTTTGGGACTTTAACTCAAGTAAAAGGA TAGTTGTACAATTATATATACGAAGAATAAATCATTACAAAAAG TATTCGTTTCTTTGATTCTTAACAGGATTCATTTTCTGGGTGTCAT CAGGTACAGCGCTGAATATCTTGAAGTTAACATCGAGCTCATCA TCGACGTTCATCACACTAGCCACGTTTCCGCAACGGTAGCAATA ATTAGGAGCGGACCACACAGTGACGACATC
82	Sequence of the 5'-region that was used to knock into	GAGTCGGCCAAGAGATGATAACTGTTACTAAGCTTCTCCGTAATT AGTGGTATTTTGTAACTTTTACCAATAATCGTTTATGAATACGGA TATTTTTCGACCTTATCCAGTGCCAAATCACGTAACTTAATCATG GTTTAAATACTCCACTTGAACGATTCATTATTCAGAAAAAAGTCA GGTTGGCAGAAACACTTGGGCGCTTTGAAGAGTATAAGAGTATT

	the PpADE1 locus:	AAGCATTAAACATCTGAACTTTCACCGCCCCAATATACTACTCTA GGAAACTCGAAAAATTCCTTTCCATGTGTCATCGCTTCCAACACA CTTTGCTGTATCCTTCCAAGT**ATG**TCCATTGTGAACACTGATCTG GACGGAATCCTACCTTTAATCGCCAAAGGAAAGGTTAGAGACAT TTATGCAGTCGATGAGAACAACTTGCTGTTCGTCGCAACTGACCG TATCTCCGCTTACGATGTGATTATGACAAACGGTATTCCTGATAA GGGAAAGATTTTGACTCAGCTCTCAGTTTTCTGGTTTGATTTTTTG GCACCCTACATAAAGAATCATTTGGTTGCTTCTAATGACAAGGA AGTCTTTGCTTTACTACCATCAAAACTGTCTGAAGAAAAaTACAA ATCTCAATTAGAGGGACGATCCTTGATAGTAAAAAAGCACAGAC TGATACCTTTGGAAGCCATTGTCAGAGGTTACATCACTGGAAGTG CATGGAAAGAGTACAAGAACTCAAAAACTGTCCATGGAGTCAAG GTTGAAAACGAGAACCTTCAAGAGAGCGACGCCTTTCCAACTCC GATTTTCACACCTTCAACGAAAGCTGAACAGGGTGAACACGATG AAAACATCTCTATTGAACAAGCTGCTGAGATTGTAGGTAAAGAC ATTTGTGAGAAGGTCGCTGTCAAGGCGGTCGAGTTGTATTCTGCT GCAAAAAACCTCGCCCTTTTGAAGGGGATCATTATTGCTGATAC GAAATTCGAATTTGGACTGGACGAAAACAATGAATTGGTACTAG TAGATGAAGTTTTAACTCCAGATTCTTCTAGATTTTGGAATCAAA AGACTTACCAAGTGGGTAAATCGCAAGAGAGTTACGATAAGCAG TTTCTCAGAGATTGGTTGACGGCCAACGGATTGAATGGCAAAGA GGGCGTAGCCATGGATGCAGAAATTGCTATCAAGAGTAAAGAAA AGTATATTGAAGCTTATGAAGCAATTACTGGCAAGAAATGGGCT TGA
83	Sequence of the 3'-region that was used to knock into the PpADE1 locus:	ATGATTAGTACCCTCCTCGCCTTTTTCAGACATCTGAAATTTCCCT TATTCTTCCAATTCCATATAAAATCCTATTTAGGTAATTAGTAAA CAATGATCATAAAGTGAAATCATTCAAGTAACCATTCCGTTTATC GTTGATTTAAAATCAATAACGAATGAATGTCGGTCTGAGTAGTC AATTTGTTGCCTTGGAGCTCATTGGCAGGGGGTCTTTTGGCTCAG TATGGAAGGTTGAAAGGAAAACAGATGGAAAGTGGTTCGTCAGA AAAGAGGTATCCTACATGAAGATGAATGCCAAAGAGATATCTCA AGTGATAGCTGAGTTCAGAATTCTTAGTGAGTTAAGCCATCCCAA CATTGTGAAGTACCTTCATCACGAACATATTTCTGAGAATAAAAC TGTCAATTTATACATGGAATACTGTGATGGTGGAGATCTCTCCAA GCTGATTCGAACACATAGAAGGAACAAAGAGTACATTTCAGAAG AAAAAATATGGAGTATTTTTACGCAGGTTTTATTAGCATTGTATC GTTGTCATTATGGAACTGATTTCACGGCTTCAAAGGAGTTTGAAT CGCTCAATAAAGGTAATAGACGAACCCAGAATCCTTCGTGGGTA GACTCGACAAGAGTTATTATTCACAGGGATATAAAACCCGACAA CATCTTTCTGATGAACAATTCAAACCTTGTCAAACTGGGAGATTT TGGATTAGCAAAAATTCTGGACCAAGAAAACGATTTTGCCAAAA CATACGTCGGTACGCCGTATTACATGTCTCCTGAAGTGCTGTTGG ACCAACCCTACTCACCATTATGTGATATATGGTCTCTTGGGTGCG TCATGTATGAGCTATGTGCATTGAGGCCTCCTT
84	MET16 5'	GGGTGGGCCTGGTAATGTTCACTCCTAGGAACTACTAGAAAAAC TGTGCTAAACGGATTACGTAATTATTATACAAATTCTCTATGGTC TATGGTACATATGGGCTGGTTCAATAATGAATCTATGAAGAATTT GTGCCCATGGGGACCGTTTCTATAAACGTTCTCTTCTTTATGTTTT

		CCACCTGCTCTTTGAGTTCCGGAAATTCGTTGACAATCTTTTGTC CCAATGTCGATTGGGCGTATTTAAAGCCCAGCTGTTTTCCTCTGA GAAATTGATTCAACTTCCTCACCACCTCCACAAACTCACGCGTGT ATATATCAGGGTTTCTACCGTCTTCGATATAATTGACTACGTCCA CGGGGATGGGAATGTTCAAATCTGTGTTGTGGAGCTTTTGCAAGT GCTCTACAACCTTGTTAATGTTGTTGGAAAGACCCAATTGACTTT CCGCTGTACCGGCGTAATCGTGCACCTGAACACCCAAATGGATG AGGGTTTCGATGAGTTGACTTAGTTCATTTTCAACTTGATCTAAT GTTGTCGCAGGTGCACTCATACTTGTCATGGAGAATGAAAGTAA GTTGATAGAGAGCAGACTTCGAGGATGGGATGAACTTGATTAGG TAATCTTTGACAATGTCTTAGAGGTAGGCAGAGGATGCTGGAAA AAAAAAATTGAAAACGCCCAAGCTTCCAGCTTTGCAAGGAAAGA AGAAAAGGGAGTTGCCAGCACGAAATCGGCTTCCTCCGAAAGGT TCACAATTGCAGAATTGTCACCATTCAAATGCCTTTACCCTTCAT CTGTGGTACCTCAGGCTAAGAACGGGTCACGTGATATTTCGACA CTCATCGCCACAATATGTACTAGCAAGAACTTTTCAGATTTAGTA ATCCGTTCGAAACGGG
85	MET16 3'	CTAGATTTGCACAATATTTGAAAGCTCAGCAAAACATATGAATA TAATTTTTTTTTTCTCTACACTATTTATCCTGTAAGTTTCTGTTTCC CCATGTAGGATCTTTTTCTCCTTCTCTGTCTCCCATTTTTTTTGTTC CCTGTAGTCTTGCCTTGCCTGAGATGCGAGCTCGTCCGCCCATCC AGTCGTGTGAAGGGCCTAGCTTTTCAAAAAGAAAATACCTCCCG CTAAAGGAGGCGTTGCCCCTTCTATCAGTAGTGTCGTAACCAATT TTCACAAACAATAAAAAAAGGACACCAACAACGAAATCAACTAT TTACACACATCCAGATCCGTCCCCCTCCCCATCCAAGAGTTAAAG ACAAATATGGCTGTTAATAATCCGTCTGAATTTAGAAAGAAGTT GGTCGTAGTAGGAGATGGTGCTTGCGGTAAAACTTGTCTATTGAT GGTGTTTGCCGAGGGCGAGTTCCCTCCATCTTATGTTCCAACTGT TTTTGAGAACTATGCCACCCCAGTAGAGGTTGACAACAGAATAG TACAACTCACTCTATGGGATACTGCCGGACAGGAAGATTATGAT AGACTGAGACCTCTTTCCTATCCCGATGCCAATGTGGTCTTGATT TGTTTTGCTATTGACATTCCTGACACCTTAGATAACGTTCAAGAG AAGTGGATTAGTGAGGTGTTGCATTTCTGTCCTGGAGTCCCTATC ATTTTAGTTGGTTGTAAACTTGACTTGAGAAACGATCCAGAGGTT ATCCGTGAATTACAAGCTGTTGGAAAGCAACCAGTCTCCACCAG TGAGGGTCAGGCCGTTGC
86	Sequence of the PpMET16 auxotrophic marker:	CAACTTCCTCACCACCTCCACAAACTCACGCGTGTATATATCAGG GTTTCTACCGTCTTCGATATAATTGACTACGTCCACGGGGATGGG AATGTTCAAATCTGTGTTGTGGAGCTTTTGCAAGTGCTCTACAAC CTTGTTAATGTTGTTGGAAAGACCCAATTGACTTTCCGCTGTACC GGCGTAATCGTGCACCTGAACACCCAAATGGATGAGGGTTTCGA TGAGTTGACTTAGTTCATTTTCAACTTGATCTAATGTTGTCGCAG GTGCACTCATACTTGTCATGGAGAATGAAAGTAAGTTGATAGAG AGCAGACTTCGAGGATGGGATGAACTTGATTAGGTAATCTTTGA CAATGTCTTAGAGGTAGGCAGAGGATGCTGGAAAAAAAAAATTG AAAACGCCCAAGCTTCCAGCTTTGCAAGGAAAGAAGAAAAGGG AGTTGCCAGCACGAAATCGGCTTCCTCCGAAAGGTTCACAATTG CAGAATTGTCACCATTCAAATGCCTTTACCCTTCATCTGTGGTAC CTCAGGCTAAGAACGGGTCACGTGATATTTCGACACTCATCGCC

		ACAATATGTACTAGCAAGAACTTTTCAGATTTAGTAATCCGTTCG AAACGGGAAAAAATGTTTTTACCCTTCTATCAACTGCTAATCTTT CTAGGTTTATACTGCCAGCAGCCCGTTCCAGATACCAACATGCCA TTCACTATAGGCCAGTCAAAAACCAGTTTGAACCTCTCCAAGGTC CAAGTGGACCACCTTAACCTTTCTCTTCAGAATCTCAGTCCAGAA GAAATCATACAATGGTCTATCATTACCTTCCCACACCTGTATCAA ACTACGGCATTCGGATTGACTGGGTTGTGTATAACTGACATGGTT CACAAAATAACAGCCAAAAGAGGCAAAAAGCATGCTATTGACTT GATTTTCATAGACACCTTACATCATTTTCCACAGACTTTAGATCT CGTTGAACGAGTCAAAGATAAATACCACTGCAATGTTCATGTCTT CAAACCACAGAATGCCACTACTGAGCTCGAGTTTGGGGCGCAAT ATGGCGAAAACTTATGGGAAACAGATGATAACAAGTATGACTAC CTCGTAAAAGTTGAACCCTCACAACGTGCCTACCATGCATTAGA CGTCTGCGCCGTCTTCACAGGAAGAAGACGGTCTCAAGGTGGTA AAAGGGGAGAATTGCCCGTGATTGAAATTGATGAAATTTCTCAG GTGGTCAAGATTAATCCGTTAGCATCCTGGGGGTTTGAACAAGTT CAAAACTATATCCAAGCTAATAGCGTTCCATACAACGAATTGCT GGATTTGGGATACAAGTCAGTTGGAGATTACCATTCCACACAAC CCACTAAAAATGGTGAAGATGAAAGAGCAGGCAGGTGGAGAGG TAAACAAAAGAGTGAGTGTGGTATCCACGAAGCTTCTAGATTTG CACAATATTTGAAAGCTCAGCAAAACATATGAATATAATTTTTTT TTTCTCTACACTATTTATCCTGTAAGTTTCTGTTTCCCCATGTAGG ATCTTTTTCTCCTTCTCTGTCTCCCATTTTTTTTGTTCCCTGTAGTC TTGCCTTGCCTGAGATGCGAGCTCGTCCGCCCATCCAGTCGTGTG AAGGGCCTAGCTTTTCAAAAAGAAAATACCTCCCGCTAAAGGAG GCGTTGCCCCTTCTATCAGTAGTGTCGTAACCAATTTTCACAAAC AATAAAAAAAAGGACACCAACAACGAAATCAACTATTTACACACA TCCAGATCCGTCCC
87	Sequence of the 5'-Region used for knock out of PpHIS1:	TAACTGGCCCTTTGACGTTTCTGACAATAGTTCTAGAGGAGTCGT CCAAAAACTCAACTCTGACTTGGGTGACACCACCACGGGATCCG GTTCTTCCGAGGACCTTGATGACCTTGGCTAATGTAACTGGAGTT TTAGTATCCATTTTAAGATGTGTGTTTCTGTAGGTTCTGGGTTGG AAAAAAATTTTAGACACCAGAAGAGAGGAGTGAACTGGTTTGCG TGGGTTTAGACTGTGTAAGGCACTACTCTGTCGAAGTTTTAGATA GGGGTTACCCGCTCCGATGCATGGGAAGCGATTAGCCCGGCTGT TGCCCGTTTGGTTTTTGAAGGGTAATTTTCAATATCTCTGTTTGAG TCATCAATTTCATATTCAAAGATTCAAAAACAAAATCTGGTCCAA GGAGCGCATTTAGGATTATGGAGTTGGCGAATCACTTGAACGAT AGACTATTATTTGC
88	Sequence of the 3'-Region used for knock out of PpHIS1:	GTGACATTCTTGTCTTTGAGATCAGTAATTGTAGAGCATAGATAG AATAATATTCAAGACCAACGGCTTCTCTTCGGAAGCTCCAAGTA GCTTATAGTGATGAGTACCGGCATATATTTATAGGCTTAAAATTT CGAGGGTTCACTATATTCGTTTAGTGGGAAGAGTTCCTTTCACTC TTGTTATCTATATTGTCAGCGTGGACTGTTTATAACTGTACCAAC TTAGTTTCTTTCAACTCCAGGTTAAGAGACATAAATGTCCTTTGA TGCTGACAATAATCAGTGGAATTCAAGGAAGGACAATCCCGACC TCAATCTGTTCATTAATGAAGAGTTCGAATCGTCCTTAAATCAAG

		CGCTAGACTCAATTGTCAATGAGAACCCTTTCTTTGACCAAGAAA CTATAAATAGATCGAATGACAAAGTTGGAAATGAGTCCATTAGC TTACATGATATTGAGCAGGCAGACCAAAATAAACCGTCCTTTGA GAGCGATATTGATGGTTCGGCGCCGTTGATAAGAGACGACAAAT TGCCAAAGAAACAAAGCTGGGGGCTGAGCAATTTTTTTTCAAGA AGAAATAGCATATGTTTACCACTACATGAAAATGATTCAAGTGTT GTTAAGACCGAAAGATCTATTGCAGTGGGAACACCCCATCTTCA ATACTGCTTCAATGGAATCTCCAATGCCAAGTACAATGCATTTAC CTTTTTCCCAGTCATCCTATACGAGCAATTCAAATTTTTTTTCAAT TTATACTTTACTTTAGTGGCTCTCTCTCAAGCGATACCGCAACTT CGCATTGGATATCTTTCTTCGTATGTCGTCCCACTTTTGTTTGTAC TCATAGTGACCATGTCAAAAGAGGCGATGGATGATATTCAACGC CGAAGAAGGGATAGAGAACAGAACAATGAACCATATGAGGTTC TGTCCAGCCCATCACCAGTTTTGTCCAAAAACTTAAAATGTGGTC ACTTGGTTCGATTGCATAAGGGAATGAGAGTGCCCGCAGATATG GTTCTTGTCCAGTCAAGCGAATCCACCGGAGAGTCATTTATCAAG ACAGATCAGCTGGATGGTGAGACTGATTGGAAGCTTCGGATTGT TTCTCCAGTTACACAATCGTTACCAATGACTGAACTTCAAAATGT CGCCATCACTGCAAGCGCACCCTCAAAATCAATTCACTCCTTTCT TGGAAGATTGACCTACAATGGGCAATCATATGGTCTTACGATAG ACAACACAATGTGGTGTAATACTGTATTAGCTTCTGGTTCAGCAA TTGGTTGTATAATTTACACAGGTAAAGATACTCGACAATCGATGA ACACAACTCAGCCCAAACTGAAAACGGGCTTGTTAGAACTGGAA ATCAATAGTTTGTCCAAGATCTTATGTGTTTGTGTGTTTGCATTAT CTGTCATCTTAGTGCTATTCCAAGGAATAGCTGATGATTGGTACG TCGATATCATGCGGTTTCTCATTCTATTCTCCACTATTATCCCAGT GTCTCTGAGAGTTAACCTTGATCTTGGAAAGTCAGTCCATGCTCA TCAAATAGAAACTGATAGCTCAATACCTGAAACCGTTGTTAGAA CTAGTACAATACCGGAAGACCTGGGAAGAATTGAATACCTATTA AGTGACAAAACTGGAACTCTTACTCAAAATGATATGGAAATGAA AAAACTACACCTAGGAACAGTCTCTTATGCTGGTGATACCATGG ATATTATTTCTGATCATGTTAAAGGTCTTAATAACGCTAAAACAT CGAGGAAAGATCTTGGTATGAGAATAAGAGATTTGGTTACAACT CTGGCCATCTG
89	Sequence of the PpHIS1 auxotrophic marker:	CAAGTTGCGTCCGGTATACGTAACGTCTCACGATGATCAAAGAT AATACTTAATCTTCATGGTCTACTGAATAACTCATTTAAACAATT GACTAATTGTACATTATATTGAACTTATGCATCCTATTAACGTAA TCTTCTGGCTTCTCTCTCAGACTCCATCAGACACAGAATATCGTT CTCTCTAACTGGTCCTTTGACGTTTCTGACAATAGTTCTAGAGGA GTCGTCCAAAAACTCAACTCTGACTTGGGTGACACCACCACGGG ATCCGGTTCTTCCGAGGACCTTGATGACCTTGGCTAATGTAACTG GAGTTTTAGTATCCATTTTAAGATGTGTGTTTCTGTAGGTTCTGG GTTGGAAAAAAATTTTAGACACCAGAAGAGAGGAGTGAACTGGT TTGCGTGGGTTTAGACTGTGTAAGGCACTACTCTGTCGAAGTTTT AGATAGGGGTTACCCGCTCCGATGCATGGGAAGCGATTAGCCCG GCTGTTGCCCGTTTGGTTTTTGAAGGGTAATTTTCAATATCTCTGT TTGAGTCATCAATTTCATATTCAAAGATTCAAAAACAAAATCTGG TCCAAGGAGCGCATTTAGGATTATGGAGTTGGCGAATCACTTGA

		ACGATAGACTATTATTTGCTGTTCCTAAAGAGGGCAGATTGTATG AGAAATGCGTTGAATTACTTAGGGGATCAGATATTCAGTTTCGA AGATCCAGTAGATTGGATATAGCTTTGTGCACTAACCTGCCCCTG GCATTGGTTTTCCTTCCAGCTGCTGACATTCCCACGTTTGTAGGA GAGGGTAAATGTGATTTGGGTATAACTGGTATTGACCAGGTTCA GGAAAGTGACGTAGATGTCATACCTTTATTAGACTTGAATTTCGG TAAGTGCAAGTTGCAGATTCAAGTTCCCGAGAATGGTGACTTGA AAGAACCTAAACAGCTAATTGGTAAAGAAATTGTTTCCTCCTTTA CTAGCTTAACCACCAGGTACTTTGAACAACTGGAAGGAGTTAAG CCTGGTGAGCCACTAAAGACAAAAATCAAATATGTTGGAGGGTC TGTTGAGGCCTCTTGTGCCCTAGGAGTTGCCGATGCTATTGTGGA TCTTGTTGAGAGTGGAGAAACCATGAAAGCGGCAGGGCTGATCG ATATTGAAACTGTTCTTTCTACTTCCGCTTACCTGATCTCTTCGAA GCATCCTCAACACCCAGAACTGATGGATACTATCAAGGAGAGAA TTGAAGGTGTACTGACTGCTCAGAAGTATGTCTTGTGTAATTACA ACGCACCTAGAGGTAACCTTCCTCAGCTGCTAAAACTGACTCCA GGCAAGAGAGCTGCTACCGTTTCTCCATTAGATGAAGAAGATTG GGTGGGAGTGTCCTCGATGGTAGAGAAGAAAGATGTTGGAAGAA TCATGGACGAATTAAAGAAACAAGGTGCCAGTGACATTCTTGTC TTTGAGATCAGTAATTGTAGAGCATAGATAGAATAATATTCAAG ACCAACGGCTTCTCTTCGGAAGCTCCAAGTAGCTTATAGTGATGA GTACCGGCATATATTTATAGGCTTAAAATTTCGAGGGTTCACTAT ATTCGTTTAGTGGGAAGAGTTCCTTTCACTCTTGTTATCTATATTG TCAGCGTGGACTGTTTATAACTGTACCAACTTAGTTTCTTTCAAC TCCAGGTTAAGAGACATAAATGTCCTTTGATGC
90	Sequence of the 5'-region that was used to knock into the PpPRO1 locus:	GAAGGGCCATCGAATTGTCATCGTCTCCTCAGGTGCCATCGCTGT GGGCATGAAGAGAGTCAACATGAAGCGGAAACCAAAAAAGTTA CAGCAAGTGCAGGCATTGGCTGCTATAGGACAAGGCCGTTTGAT AGGACTTTGGGACGACCTTTTCCGTCAGTTGAATCAGCCTATTGC GCAGATTTTACTGACTAGAACGGATTTGGTCGATTACACCCAGTT TAAGAACGCTGAAAATACATTGGAACAGCTTATTAAAATGGGTA TTATTCCTATTGTCAATGAGAATGACACCCTATCCATTCAAGAAA TCAAATTTGGTGACAATGACACCTTATCCGCCATAACAGCTGGTA TGTGTCATGCAGACTACCTGTTTTTGGTGACTGATGTGGACTGTC TTTACACGGATAACCCTCGTACGAATCCGGACGCTGAGCCAATC GTGTTAGTTAGAAATATGAGGAATCTAAACGTCAATACCGAAAG TGGAGGTTCCGCCGTAGGAACAGGAGGAATGACAACTAAATTGA TCGCAGCTGATTTGGGTGTATCTGCAGGTGTTACAACGATTATTT GCAAAAGTGAACATCCCGAGCAGATTTTGGACATTGTAGAGTAC AGTATCCGTGCTGATAGAGTCGAAAATGAGGCTAAATATCTGGT CATCAACGAAGAGGAAACTGTGGAACAATTTCAAGAGATCAATC GGTCAGAACTGAGGGAGTTGAACAAGCTGGACATTCCTTTGCAT ACACGTTTCGTTGGCCACAGTTTTAATGCTGTTAATAACAAAGAG TTTTGGTTACTCCATGGACTAAAGGCCAACGGAGCCATTATCATT GATCCAGGTTGTTATAAGGCTATCACTAGAAAAAACAAAGCTGG TATTCTTCCAGCTGGAATTATTTCCGTAGAGGGTAATTTCCATGA ATACGAGTGTGTTGATGTTAAGGTAGGACTAAGAGATCCAGATG ACCCACATTCACTAGACCCCAATGAAGAACTTTACGTCGTTGGCC GTGCCCGTTGTAATTACCCCAGCAATCAAATCAACAAAATTAAG

		GGTCTACAAAGCTCGCAGATCGAGCAGGTTCTAGGTTACGCTGA CGGTGAGTATGTTGTTCACAGGGACAACTTGGCTTTCCCAGTATT TGCCGATCCAGAACTGTTGGATGTTGTTGAGAGTACCCTGTCTGA ACAGGAGAGAGAATCCAAACCAAATAAATAG
91	Sequence of the 3'-region that was used to knock into the PpPRO1 locus:	AATTTCACATATGCTGCTTGATTATGTAATTATACCTTGCGTTCG ATGGCATCGATTTCCTCTTCTGTCAATCGCGCATCGCATTAAAAG TATACTTTTTTTTTTTTCCTATAGTACTATTCGCCTTATTATAAAC TTTGCTAGTATGAGTTCTACCCCCAAGAAAGAGCCTGATTTGACT CCTAAGAAGAGTCAGCCTCCAAAGAATAGTCTCGGTGGGGGTAA AGGCTTTAGTGAGGAGGGTTTCTCCCAAGGGGACTTCAGCGCTA AGCATATACTAAATCGTCGCCCTAACACCGAAGGCTCTTCTGTGG CTTCGAACGTCATCAGTTCGTCATCATTGCAAAGGTTACCATCCT CTGGATCTGGAAGCGTTGCTGTGGGAAGTGTGTTGGGATCTTCGC CATTAACTCTTTCTGGAGGGTTCCACGGGCTTGATCCAACCAAGA ATAAAATAGACGTTCCAAAGTCGAAACAGTCAAGGAGACAAAGT GTTCTTTCTGACATGATTTCCACTTCTCATGCAGCTAGAAATGAT CACTCAGAGCAGCAGTTACAAACTGGACAACAATCAGAACAAA AAGAAGAAGATGGTAGTCGATCTTCTTTTTCTGTTTCTTCCCCCG CAAGAGATATCCGGCACCCAGATGTACTGAAAACTGTCGAGAAA CATCTTGCCAATGACAGCGAGATCGACTCATCTTTACAACTTCAA GGTGGAGATGTCACTAGAGGCATTTATCAATGGGTAACTGGAGA AAGTAGTCAAAAAGATAACCCGCCTTTGAAACGAGCAAATAGTT TTAATGATTTTTCTTCTGTGCATGGTGACGAGGTAGGCAAGGCAG ATGCTGACCACGATCGTGAAAGCGTATTCGACGAGGATGATATC TCCATTGATGATATCAAAGTTCCGGGAGGGATGCGTCGAAGTTTT TTATTACAAAAGCATAGAGACCAACAACTTTCTGGACTGAATAA AACGGCTCACCAACCAAAACAACTTACTAAACCTAATTTCTTCAC GAACAACTTTATAGAGTTTTTGGCATTGTATGGGCATTTTGCAGG TGAAGATTTGGAGGAAGACGAAGATGAAGATTTAGACAGTGGTT CCGAATCAGTCGCAGTCAGTGATAGTGAGGGAGAATTCAGTGAG GCTGACAACAATTTGTTGTATGATGAAGAGTCTCTCCTATTAGCA CCTAGTACCTCCAACTATGCGAGATCAAGAATAGGAAGTATTCG TACTCCTACTTATGGATCTTTCAGTTCAAATGTTGGTTCTTCGTCT ATTCATCAGCAGTTAATGAAAAGTCAAATCCCGAAGCTGAAGAA ACGTGGACAGCACAAGCATAAAACACAATCAAAAATACGCTCG AAGAAGCAAACTACCACCGTAAAAGCAGTGTTGCTGCTATTAAA
92	Truncated hEPO DNA (codon optimized)	GCTCCACCAAGATTGATTTGTGACTCCAGAGTTTTGGAGAGATAC TTGTTGGAGGCTAAAGAGGCTGAGAACATCACTACTGGTTGTGC TGAACACTGTTCCTTGAACGAGAACATCACAGTTCCAGACACTA AGGTTAACTTCTACGCTTGGAAGAGAATGGAAGTTGGACAACAG GCTGTTGAAGTTTGGCAAGGATTGGCTTTGTTGTCCGAGGCTGTT TTGAGAGGTCAAGCTTTGTTGGTTAACTCCTCCCAACCATGGGAA CCATTGCAATTGCACGTTGACAAGGCTGTTTCTGGATTGAGATCC TTGACTACTTTGTTGAGAGCTTTGGGTGCTCAGAAAGAGGCTATT TCTCCACCAGATGCTGCTTCAGCTGCTCCATTGAGAACTATCACT GCTGACACTTTCAGAAAGTTGTTCAGAGTTTACTCCAACTTCTTG AGAGGAAAGTTGAAGTTGTACACTGGTGAAGCTTGTAGAACTGG TGACTAGTAA
93	Truncated	APPRLICDSR VLERYLLEAK EAENITTGCA EHCSLNENIT VPDTKVNFYA

	hEPO protein	WKRMEVGQQA VEVWQGLALL SEAVLRGQAL LVNSSQPWEP LQLHVDKAVS GLRSLTTLLR ALGAQKEAIS PPDAASAAPL RTITADTFRK LFRVYSNFLR GKLKLYTGEA CRTGD
94	Chicken lysosome signal DNA (CLSP)	ATGCTGGGTAAGAACGACCCAATGTGTCTTGTTTTGGTCTTGTTG GGATTGACTGCTTTGTTGGGTATCTGTCAAGGT
95	Chicken lysosome signal peptide (CLSP)	MLGKNDPMCLVLVLLGLTALLGICQG
96	Sequence of the PpAde2 gene without its promoter but including its termination sequences	ATGGATTCTCAGGTAATAGGTATTCTAGGAGGAGGCCAGCTAGG CCGAATGATTGTTGAGGCCGCTAGCAGGCTCAATATCAAGACCG TGATTCTTGATGATGGTTTTTCACCTGCTAAGCACATTAATGCTG CGCAAGACCACATCGACGGATCATTCAAAGATGAGGAGGCTATC GCCAAGTTAGCTGCCAAATGTGATGTTCTCACTGTAGAGATTGAG CATGTCAACACAGATGCTCTAAAGAGAGTTCAAGACAGAACTGG AATCAAGATATATCCTTTACCAGAGACAATCGAACTAATCAAGG ATAAGTACTTGCAAAAGGAACATTTGATCAAGCACAACATTTCG GTGACAAAGTCTCAGGGTATAGAATCTAATGAAAAGGCGCTGCT TTTGTTTGGAGAAGAGAATGGATTTCCATATCTGTTGAAGTCCCG GACTATGGCTTATGATGGAAGAGGCAATTTTGTAGTGGAGTCTA AAGAGGACATCAGTAAGGCATTAGAATTCTTGAAAGATCGTCCA TTGTATGCCGAGAAGTTTGCTCCTTTTGTTAAAGAATTAGCGGTA ATGGTTGTGAGATCACTGGAAGGCGAAGTATTCTCCTACCCAAC CGTAGAAACTGTGCACAAGGACAATATCTGTCATATTGTGTATGC TCCGGCCAGAGTTAATGACACCATCCAAAAGAAAGCTCAAATAT TAGCTGAAAACACTGTGAAGACTTTCCCAGGCGCTGGAATCTTC GGAGTTGAGATGTTCCTATTGTCTGATGGAGAACTTCTTGTAAAT GAGATTGCTCCAAGGCCCCACAATTCTGGTCACTATACAATCGAT GCATGTGTAACATCTCAGTTCGAAGCACATGTAAGAGCCATAAC TGGTCTGCCAATGCCACTAGATTTCACCAAACTATCTACTTCCAA CACCAACGCTATTATGCTCAATGTTTTGGGTGCTGAAAAATCTCA CGGGGAATTAGAGTTTTGTAGAAGAGCCTTAGAAACACCCGGTG CTTCTGTATATCTGTACGGAAAGACCACCCGATTGGCTCGTAAGA TGGGTCATATCAACATAATAGGATCTTCCATGTTGGAAGCAGAA CAAAAGTTAGAGTACATTCTAGAAGAATCAACCCACTTACCATC CAGTACTGTATCAGCTGACACTAAACCGTTGGTTGGAGTTATCAT GGGTTCAGACTCTGATCTACCTGTGATTTCGAAAGGTTGCGATAT TTTAAAACAGTTTGGTGTTCCATTCGAAGTTACTATTGTCTCTGCT CATAGAACACCACAGAGAATGACCAGATATGCCTTTGAAGCCGC TAGTAGAGGTATCAAGGCTATCATTGCAGGTGCTGGTGGTGCTG CTCATCTTCCAGGAATGGTTGCTGCCATGACTCCGTTGCCAGTCA TTGGTGTTCCTGTCAAGGGCTCTACGTTGGATGGTGTAGACTCGC TACACTCGATTGTCCAAATGCCTAGAGGTGTTCCTGTGGCTACGG TTGCTATCAACAACGCCACCAATGCCGCTCTGTTGGCCATCAGGA TTTTAGGTACAATTGACCACAAATGGCAAAAGGAAATGTCCAAG

		TATATGAATGCAATGGAGACCGAAGTGTTGGGGAAGGCATCCAA CTTGGAATCTGAAGGGTATGAATCCTATTTGAAGAATCGTCTT*TG* *AATTTAGTATTGTTTTTTAATAGATGTATATATAATAGTACACGTAACTT* *ATCTATTCCATTCATAATTTTATTTTAAAGGTTCGGTAGAAATTTGTCCT* *CCAAAAAGTTGGTTAGAGCCTGGCAGTTTTGATAGGCATTATTATAGA* *TTGGGTAATATTTACCCTGCACCTGGAGGAACTTTGCAAAGAGCCTCA* *TGTGC*
97	PpADE2	MDSQVIGILGGGQLGRMIVEAASRLNIKTVILDDGFSPAKHINAAQD HIDGSFKDEEAIAKLAAKCDVLTVEIEHVNTDALKRVQDRTGIKIYP LPETIELIKDKYLQKEHLIKHNISVTKSQGIESNEKALLLFGEENGFPY LLKSRTMAYDGRGNFVVESKEDISKALEFLKDRPLYAEKFAPFVKEL AVMVVRSLEGEVFSYPTVETVHKDNICHIVYAPARVNDTIQKKAQIL AENTVKTFPGAGIFGVEMFLLSDGELLVNEIAPRPHNSGHYTIDACV TSQFEAHVRAITGLPMPLDFTKLSTSNTNAIMLNVLGAEKSHGELEF CRRALETPGASVYLYGKTTRLARKMGHINIIGSSMLEAEQKLEYILE ESTHLPSSTVSADTKPLVGVIMGSDSDLPVISKGCDILKQFGVPFEVTI VSAHRTPQRMTRYAFEAASRGIKAIIAGAGGAAHLPGMVAAMTPLP VIGVPVKGSTLDGVDSLHSIVQMPRGVPVATVAINNATNAALLAIRI LGTIDHKWQKEMSKYMNAMETEVLGKASNLESEGYESYLKNRL
98	Pp TRP2: 5' and ORF	ACTGGGCCTTTAGAGGGTGCTGAAGTTGACCCCTTGGTGCTTCTG GAAAAAGAACTGAAGGGCACCAGACAAGCGCAACTTCCTGGTAT TCCTCGTCTAAGTGGTGGTGCCATAGGATACATCTCGTACGATTG TATTAAGTACTTTGAACCAAAAACTGAAAGAAAACTGAAAGATG TTTTGCAACTTCCGGAAGCAGCTTTGATGTTGTTCGACACGATCG TGGCTTTTGACAATGTTTATCAAAGATTCCAGGTAATTGGAAACG TTTCTCTATCCGTTGATGACTCGGACGAAGCTATTCTTGAGAAAT ATTATAAGACAAGAGAAGAAGTGGAAAAGATCAGTAAAGTGGT ATTTGACAATAAAACTGTTCCCTACTATGAACAGAAAGATATTAT TCAAGGCCAAACGTTCACCTCTAATATTGGTCAGGAAGGGTATG AAAACCATGTTCGCAAGCTGAAAGAACATATTCTGAAAGGAGAC ATCTTCCAAGCTGTTCCCTCTCAAAGGGTAGCCAGGCCGACCTCA TTGCACCCTTTCAACATCTATCGTCATTTGAGAACTGTCAATCCT TCTCCATACATGTTCTATATTGACTATCTAGACTTCCAAGTTGTTG GTGCTTCACCTGAATTACTAGTTAAATCCGACAACAACAACAAA ATCATCACACATCCTATTGCTGGAACTCTTCCCAGAGGTAAAACT ATCGAAGAGGACGACAATTATGCTAAGCAATTGAAGTCGTCTTT GAAAGACAGGGCCGAGCACGTCATGCTGGTAGATTTGGCCAGAA ATGATATTAACCGTGTGTGTGAGCCCACCAGTACCACGGTTGATC GTTTATTGACTGTGGAGAGATTTTCTCATGTGATGCATCTTGTGT CAGAAGTCAGTGGAACATTGAGACCAAACAAGACTCGCTTCGAT GCTTTCAGATCCATTTTCCCAGCAGGTACCGTCTCCGGTGCTCCG AAGGTAAGAGCAATGCAACTCATAGGAGAATTGGAAGGAGAAA AGAGAGGTGTTTATGCGGGGGCCGTAGGACACTGGTCGTACGAT GGAAAATCGATGGACACATGTATTGCCTTAAGAACAATGGTCGT CAAGGACGGTGTCGCTTACCTTCAAGCCGGAGGTGGAATTGTCT ACGATTCTGACCCCTATGACGAGTACATCGAAACCATGAACAAA ATGAGATCCAACAATAACACCATCTTGGAGGCTGAGAAAATCTG GACCGATAGGTTGGCCAGAGACGAG AATCAAAGTGAATCCGAAGAAAACGATCAATGA

99	PpTRP2 3' region	ACGGAGGACGTAAGTAGGAATTTATGTAATCATGCCAATACATC TTTAGATTTCTTCCTCTTCTTTTTAACGAAAGACCTCCAGTTTTGC ACTCTCGACTCTCTAGTATCTTCCCATTTCTGTTGCTGCAACCTCT TGCCTTCTGTTTCCTTCAATTGTTCTTCTTTCTTCTGTTGCACTTGG CCTTCTTCCTCCATCTTTCGTTTTTTTTCAAGCCTTTTCAGCAGTTC TTCTTCCAAGAGCAGTTCTTTGATTTTCTCTCTCCAATCCACCAAA AAACTGGATGAATTCAACCGGGCATCATCAATGTTCCACTTTCTT TCTCTTATCAATAATCTACGTGCTTCGGCATACGAGGAATCCAGT TGCTCCCTAATCGAGTCATCCACAAGGTTAGCATGGGCCTTTTTC AGGGTGTCAAAAGCATCTGGAGCTCGTTTATTCGGAGTCTTGTCT GGATGGATCAGCAAAGACTTTTTGCGGAAAGTCTTTCTTATATCT TCCGGAGAACAACCTGGTTTCAAATCCAAGATGGCATAGCTGTC CAATTTGAAAGTGGAAAGAATCCTGCCAATTTCCTTCTCTCGTGT CAGCTCGTTCTCCTCCTTTTGCAACAGGTCCACTTCATCTGGCATT TTTCTTTATGTTAACTTTAATTATTATTAATTATAAAGTTGATTAT CGTTATCAAAATAATCATATTCGAGAAATAATCCGTCCATGCAAT ATATAAATAAGAATTCATAATAATGTAATGATAACAGTACCTCT GATGACCTTTGATGAACCGCAATTTTCTTTCCAATGACAAGACAT CCCTATAATACAATTATACAGTTTATATATCACAAATAATCACCT TTTTATAAGAAAACCGTCCTCTCCGTAACAGAACTTATTATCCGC ACGTTATGGTTAACACACTACTAATACCGATATAGTGTATGAAGT CGCTACGAGATAGCCATCCAGGAAACTTACCAATTCATCAGCAC TTTCATGATCCGATTGTTGGCTTTATTCTTTGCGAGACAGATACTT GCCAATGAAATAACTGATCCCACAGATGAGAATCCGGTGCTCGT
100	Pp ADE2 5' region	CTTAAAATCATCTGCCTCACCCCACCGACCAATGGGAATTCTAGA AACAATTTCATTGCTCTTCTTCTCGTTACCATAAGAATCGGCTGT CATGTTTGACTTAACGAACCCTGGAACAAGGGAATTCACGGTAA TACCTTTTGGAGCAAGTTCAACCGATAGAGCCTTCATTAATGAGT TGATTGCACCTTTGGTGGTCGCATATACCGATTGATTCGGGTAGG TCACTTCGAAACTGTACAGGGAGGCAGTAAAGATGATCCTACCC TTAATCTGGTTCTTAATAAAGTGTTTAGTGACTAGCTGTGTCAAT CTAAATGGAAAATCGACATTTACCTTTTGGATAGCCGCGTAATCT TTCTCCGTAAAACTTGTAAACTCAGATTTAATGGCAATGGCAGCG TTGTTGATTAAAATGTCAATCTTTCCAGTGGAACTCTTCTCCACC GCAGGACTCGTTACGGTCTCTTCCAGCTTTGCAAGATCGGCATCC ACTAGATCCAACTCAATTGTATGTATGGAGGCACCATCGGCATTT GACATTCTCACCTCTTCAATGAAAGCCGTTGGGTCTGTAGAAGGT CTATGGATAAGAATAAGTTCTGCACCTGCTTCATAAAGTCCTCGA ACTATTCCTTGGCCTAATCCGCTGGTACCACCGGTGATCAAGGCG ACCTTACCATTCAAAGAAAACAAATCAGCGGACATTAGCGACTT GAATAGGGAATGGGTTAGACAAATGAAAGCCGACGAGCCAGCA CTTTATAGTAAGTGCAGGTGAGTCAATAAGAATAAATGTATGGC TTGCTGTCCCTATCGCGTAAGAAGCTTACTAAGATCGCCTAAATT GAAAAGTTGAACAAATCAGTTCTAGCTGGCCTCCATCAGCATTTC GTTCTCCTCTGATCATCTTTGCCAATCGCTAGCATGCCCTCAGCG TGCAAGGAAAAGCACGCTTCTTTCTTATCGACGTATTTTCAACTA TGGCAGAGCCAGGTTAGCAAGTC
101	Pp ADE2 3' region	ATTTAGTATTGTTTTTTAATAGATGTATATATAATAGTACACGTA ACTTATCTATTCCATTCATAATTTTATTTTAAAGGTTCGGTAGAA

		ATTTGTCCTCCAAAAAGTTGGTTAGAGCCTGGCAGTTTTGATAGG CATTATTATAGATTGGGTAATATTTACCCTGCACCTGGAGGAACT TTGCAAAGAGCCTCATGTGCTCTAAAAGGATGTCAGAATTCCAA CATTTCAAAATTATATCTGCATGCGTCTGTAATACTGGAACTGTT ATTTTTCTGGTCAGGATTTCACCGCTCTTGTCGTCATGTTTCTCGT CGTCTGAAAGTAAACTGACTTTCCTCTTTCCATAAACACAAAAAT CGATTGCAACTTGGTTATTCTTGAGATTGAAATTTGCTGTGTCTTC AGTGCTTAGCTGAATATCAACAAACTTACTTAGTACTAATAACGA AGCACTATGGTAAGTGGCATAACATAGTGGTATTGAAGCGAACA GTGGATATTGAACCCAAGCATTGGCAACATCTGGCTCTGTTGATA CTGATCCGGATCGTTTGGCACCAATTCCTGAAACGGCGTAGTGCC ACCAAGGTTTCGATTTGAGAACAGGTTCATCATCAGAGTCAACC ACCCCAATGTCAATGGCAGGCTCCAACGAAGTAGGTCCAACAAC AACAGGAAGTATTTGACCTTGAAGATCTGTTCCTTTATGATCCAC CACACCTTGCCCCAATTCCAATAACTTTACCAGTCCCGATGCAGA CATGATAACTGGTACTAATGATCTCCATTGATTTTCGTCGGCACT ACGTAAAGCCTCCAAAAATGAATTCAGAATATCTTCTGAAACTA GATTCTGCTTCTGTGATTCAAGCATTGCTTTATGTAGACATCTCTT GAATAAAAGCAATTCTCCACATATTGGTGTGTGTAAGATAGATCT GGAAAGATGTATCTGGAATAGTCCAGTCAACGTTGTGCAATTGA TTAGCATTACCTTACTGTGAACATCTCTATCTACAACAACAGACT CAATTCGATAGACGTTCCGGGAAAGTTTTTCAAGCGCATTCAGTT TGCTGTTGAACAAAGTGACTTTGCTTTCCAATGTGCAAATACCCC TGTATATCAAGTCCATCACATCACTCAAGACCTTGGTGGAAAAG AATGAAACAGCTGGAGCATAATTTTCGAATGAATTAGGTAAGGT CACTTCATCCTTATCTGTTGTAATGCTATAATCAATAGCGGAACT AACATCTTCCCATGTAACAGGTTTCTTGATCTCTGAATCTGAATC TTTATTTGAAAAAGAATTGAAAAAAGACTCATCACTCATTGGGA ATTCAAGGTCATTAGGGTATTCCATTGTTAGTTCTGGTCTAGGTT TAAAGGGATCACCTTCGTTAAGACGATGGAAAATAGCTAATCTG TACAATAACCAGATACTTCTAACGAAGCTCTCTCTATCCATCAGT TGACGTGTTGAGGATATCTGAACTAGCTCTTTCCACTGCGAATCA GGCATGCTCGTATAGCTGGCAAGCATGTTATTCAGCTTTACCAAG TTAGAAGCCCTTTGGAAACCATCTATAGATTCCCGAAAAAACTT ATACCCACTGAGGGTTTCACTGAGCATAGTCAGTGACATCAAAG AGCATTTCAAATCCATCTCA
102	NATR ORF	ATGGGTACCACTCTTGACGACACGGCTTACCGGTACCGCACCAG TGTCCCGGGGGACGCCGAGGCCATCGAGGCACTGGATGGGTCCT TCACCACCGACACCGTCTTCCGCGTCACCGCCACCGGGGACGGC TTCACCCTGCGGGAGGTGCCGGTGGACCCGCCCCTGACCAAGGT GTTCCCCGACGACGAATCGGACGACGAATCGGACGACGGGGAG GACGGCGACCCGGACTCCCGGACGTTCGTCGCGTACGGGGACGA CGGCGACCTGGCGGGCTTCGTGGTCGTCTCGTACTCCGGCTGGA ACCGCCGGCTGACCGTCGAGGACATCGAGGTCGCCCCGGAGCAC CGGGGGCACGGGGTCGGGCGCGCGTTGATGGGGCTCGCGACGG AGTTCGCCCGCGAGCGGGGCGCCGGGCACCTCTGGCTGGAGGTC ACCAACGTCAACGCACCGGCGATCCACGCGTACCGGCGGATGGG GTTCACCCTCTGCGGCCTGGACACCGCCCTGTACGACGGCACCG CCTCGGACGGCGAGCAGGCGCTCTACATGAGCATGCCCTGCCCC

		TAATCAGTACTG
103	HygR ORF	ATGGGTAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTT TCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTC GGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGC GTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACA AAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGA TTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACC TATTGCATCTCCCGCCGTGCACAGGGTGTCACGTTGCAAGACCTG CCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGC CATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGT TCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGG CGTGATTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGG CAAACTGTGATGGACGACACCGTCAGTGCGTCCGTCGCGCAGGC TCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCC GGCACCTCGTGCACGCGGATTTCGGCTCCAACAATGTCCTGACG GACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAGGCGAT GTTCGGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAG GCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGC GGAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTAT ATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGAC GGCAATTTCGATGATGCAGCTTGGGCGCAGGGTCGATGCGACGC AATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCG CCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTA CTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCCGAGGGC AAAGGAATAG
104	*PpPEP4* region (including upstream knock-out fragment, promoter, open reading frame, and downstrea m knock-out fragment)	ATTTGAGTCACCTGCTTTAGGGCTGGAAGATATTTGGTTACTAGA TTTTAGTACAAACTCTTGCTTTGTCAATGACATTAAAATAGGCAA GAATCGCAAAACTCAAATATTTCATGGAGATGAGATATGCTTGTT CAAAGATGCCCAGAAAAAAGAGCAACTCGTTTATAGGGTTCATA TTGATGATGGAACAGGCCTTTTCCAGGGAGGTGAAAGAACCCAA GCCAATTCTGATGACATTCTGGATATTGATGAGGTTGATGAAAA GTTAAGAGAACTATTGACAAGAGCCTCAAGGAAACGGCATATCA CCCCTGCATTGGAAACTCCTGATAAACGTGTAAAAAGAGCTTAT TTGAACAGTATTACTGATAACTCTTGATGGACCTTAAAGATGTAT AATAGTAGACAGAATTCATAATGGTGAGATTAGGTAATCGTCCG GAATAGGAATAGTGGTTTGGGGCGATTAATCGCACCTGCCTTAT ATGGTAAGTACCTTGACCGATAAGGTGGCAACTATTTAGAACAA AGCAAGCCACCTTTCTTTATCTGTAACTCTGTCGAAGCAAGCATC TTTACTAGAGAACATCTAAACCATTTTACATTCTAGAGTTCCATT TCTCAATTACTGATAATCAATTTAAAGATGATATTTGACGGTACT ACGATGTCAATTGCCATTGGTTTGCTCTCTACTCTAGGTATTGGT GCTGAAGCCAAAGTTCATTCTGCTAAGATACACAAGCATCCAGT CTCAGAAACTTTAAAAGAGGCCAATTTTGGGCAGTATGTCTCTGC TCTGGAACATAAATATGTTTCTCTGTTCAACGAACAAAATGCTTT GTCCAAGTCGAATTTTATGTCTCAGCAAGATGGTTTTGCCGTTGA AGCTTCGCATGATGCTCCACTTACAAACTATCTTAACGCTCAGTA TTTTACTGAGGTATCATTAGGTACCCCTCCACAATCGTTCAAGGT GATTCTTGACACAGGATCCTCCAATTTATGGGTTCCTAGCAAAGA TTGTGGATCATTAGCTTGCTTCTTGCATGCTAAGTATGACCATGA

		TGAGTCTTCTACTTATAAGAAGAATGGTAGTAGCTTTGAAATTAG GTATGGATCCGGTTCCATGGAAGGGTATGTTTCTCAGGATGTGTT GCAAATTGGGGATTTGACCATTCCCAAAGTTGATTTTGCTGAGGC CACATCGGAGCCGGGGTTGGCCTTCGCTTTTGGCAAATTTGACGG AATTTTGGGGCTTGCTTATGATTCAATATCAGTAAATAAGATTGT TCCTCCAATTTACAAGGCTTTGGAATTAGATCTCCTTGACGAACC AAAATTTGCCTTCTACTTGGGGGATACGGACAAAGATGAATCCG ATGGCGGTTTGGCCACATTTGGTGGTGTGGACAAATCTAAGTATG AAGGAAAGATCACCTGGTTGCCTGTCAGAAGAAAGGCTTACTGG GAGGTCTCTTTTGATGGTGTAGGTTTGGGATCCGAATATGCTGAA TTGCAAAAAACTGGTGCAGCCATCGACACTGGAACCTCATTGAT TGCTTTGCCCAGTGGCCTAGCTGAAATTCTCAATGCAGAAATTGG TGCTACCAAGGGTTGGTCTGGTCAATACGCTGTGGACTGTGACA CTAGAGACTCTTTGCCAGACTTAACTTTAACCTTCGCCGGTTACA ACTTTACCATTACTCCATATGACTATACTTTGGAGGTTTCTGGGT CATGTATTAGTGCTTTCACCCCCATGGACTTTCCTGAACCAATAG GTCCTTTGGCAATCATTGGTGACTCGTTCTTGAGAAAATATTACT CAGTTTATGACCTAGGCAAAGATGCAGTAGGTTTAGCCAAGTCT ATTTAGGCAAGAATAAAAGTTGCTCAGCTGAACTTATTTGGTTAC TTATCAGGTAGTGAAGATGTAGAGAATATATGTTTAGGTATTTTT TTTTAGTTTTTCTCCTATAACTCATCTTCAGTACGTGATTGCTTGT CAGCTACCTTGACAGGGGCGCATAAGTGATATCGTGTACTGCTC AATCAAGATTTGCCTGCTCCATTGATAAGGGTATAAGAGACCCA CCTGCTCCTCTTTAAAATTCTCTCTTAACTGTTGTGAAAATCATCT TCGAAGCAAATTCGAGTTTAAATCTATGCGGTTGGTAACTAAAG GTATGTCATGGTGGTATATAGTTTTTCATTTTACCTTTTACTAATC AGTTTTACAGAAGAGGAACGTCTTTCTCAAGATCGAAATAGGAC TAAATACTGGAGACGATGGGGTCCTTATTTGGGTGAAAGGCAGT GGGCTACAGTAAGGGAAGACTATTCCGATGATGGAGATGCTTGG TCTGCTTTTCCTTTTGAGCAATCTCATTTGAGAACTTATCGCTGGG GAGAGGATGGACTAGCTGGAGTCTCAGACAATCATCAACTAATT TGTTTCTCAATGGCACTGTGGAATGAGAATGATGATATTTTGAAG GAGCGATTATTTGGGGTCACTGGAGAGGCTGCAAATCATGGAGA GGATGTTAAGGAGCTTTATTATTATCTTGATAATACACCTTCTCA CTCTTATATGAAATACCTTTACAAATATCCACAATCGAAATTTCC TTACGAAGAATTGATTTCAGAGAACCGTAAACGTTCCAGATTAG AAAGAGAGTACGAGATTACTGACTCTGAAGTACTGAAGGATAAC AGATATTTTGATGTGATCTTTGAAATGGCAAAGGACGATGAAGA TGAGAATGAACTTTACTTTAGAATTACCGCTTACAACCGAGGTCC CACCCCTGCCCCTTTACATGTCGCTCCACAGGTAACCTTTAGAAA TACCTGGTCCTGGGGTATAGATGAGGAAAAGGATCACGACAAAC CTATAGCTTGCAAGGAATACCAAGACAACAACTATTCTATTCGG TTAGATAGTT
105	*Ashbya gossypii TEF1* promoter	GATCTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCGGCCAGCG ACATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGC AGCTCAGGGGCATGATGTGACTGTCGCCCGTACATTTAGCCCAT ACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCG CACGGCGCGAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCG AGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCAC

		GCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACT GAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATAC AGTTCTCACATCACATCCGAACATAAACAACC
106	*Ashbya gossypii TEF1* termination sequence	TAATCAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGT CATTTGTATAGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAA TGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCA GATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCG TATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATACTAACG CCGCCATCCAGTGTCGAAAAC

While the present invention is described herein with reference to illustrated embodiments, it should be understood that the invention is not limited hereto. Those having ordinary skill in the art and access to the teachings herein will recognize additional modifications and embodiments within the scope thereof. Therefore, the present invention is limited only by the claims attached herein.

WHAT IS CLAIMED:

1. A method for producing a recombinant glycoprotein in *Pichia pastoris* that lacks detectable cross binding activity with antibodies made against host cell antigens, comprising:

(a) providing a recombinant *Pichia pastoris* host cell which does not display β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan and does not display at least one activity selected from β-mannosyltransferase 1 activity and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan and which includes a nucleic acid molecule encoding the recombinant glycoprotein;

(b) growing the host cell in a medium under conditions effective for expressing the recombinant glycoprotein; and

(c) recovering the recombinant glycoprotein from the medium to produce the recombinant glycoprotein that lacks detectable cross binding activity with antibodies made against host cell antigens.

2. The method of claim 1, wherein the host cell does not display β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan.

3. The method of claim 1, wherein the host cell further does not display β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan.

4. The method of claim 1, wherein the detectable cross binding activity with antibodies made against host cell antigens is determined in a sandwich ELISA.

5. The method of claim 1, wherein the detectable cross binding activity with antibodies made against host cell antigens is determined in a Western blot.

6. The method of claim 1, wherein the recombinant glycoprotein is a therapeutic glycoprotein.

7. The method of claim 5, wherein the therapeutic glycoprotein is selected from the group consisting erythropoietin (EPO); cytokines such as interferon α, interferon β, interferon γ, and interferon ω; and granulocyte-colony stimulating factor (GCSF); GM-CSF; coagulation factors such as factor VIII, factor IX, and human protein C; antithrombin III; thrombin,; soluble IgE receptor α-chain; immunoglobulins such as IgG, IgG fragments, IgG fusions, and IgM; immunoadhesions and other Fc fusion proteins such as soluble TNF receptor-Fc fusion proteins;

RAGE-Fc fusion proteins; interleukins; urokinase; chymase; and urea trypsin inhibitor; IGF-binding protein; epidermal growth factor; growth hormone-releasing factor; annexin V fusion protein; angiostatin; vascular endothelial growth factor-2; myeloid progenitor inhibitory factor-1; osteoprotegerin; α-1-antitrypsin; α-feto proteins; DNase II; kringle 3 of human plasminogen; glucocerebrosidase; TNF binding protein 1; follicle stimulating hormone; cytotoxic T lymphocyte associated antigen 4 – Ig; transmembrane activator and calcium modulator and cyclophilin ligand; glucagon like protein 1; and IL-2 receptor agonist.

8. The method of claim 1, wherein the host cell is genetically engineered to produce glycoproteins that have human-like *N*-glycans.

9. The method of claim 1, wherein the host cell is genetically engineered to produce glycoproteins that have predominantly an *N*-glycan selected from $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GalGlcNAcMan_5GlcNAc_2$, $NANAGalGlcNAcMan_5GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$, $GlcNAc_{(1-4)}Man_3GlcNAc_2$, $Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$, and $NANA_{(1-4)}Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$.

10. A composition comprising one or more recombinant glycoproteins obtained by the method of claim 1.

11. A method for producing a mature human erythropoietin in *Pichia pastoris* comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens, comprising:

(a) providing a recombinant *Pichia pastoris* host cell genetically engineered to produce sialic acid-terminated biantennary *N*-glycans and does not display a β-mannosyltransferase 2 activity with respect to an *N*-glycan or *O*-glycan, and does not display at least one activity selected from a β-mannosyltransferase 1 activity and a β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan and which includes two or more nucleic acid molecules, each encoding a fusion protein comprising a mature human erythropoietin fused to a signal peptide that targets the ER and which is removed when the fusion protein is in the ER;

(b) growing the host cell in a medium under conditions effective for expressing and processing the first and second fusion proteins; and

(c) recovering the mature human erythropoietin from the medium to produce the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens.

12. The method of claim 11, wherein the host cell does not display β-mannosyltransferase 2 activity, β-mannosyltransferase 1 activity, and β-mannosyltransferase 3 activity with respect to an *N*-glycan or *O*-glycan.

13. The method of claim 12, wherein the host cell further does not display β-mannosyltransferase 4 activity with respect to an *N*-glycan or *O*-glycan.

14. The method of claim 11, wherein the signal peptide is a *S. cerevisiae* αMATpre signal peptide or a chicken lysozyme signal peptide.

15. The method of claim 11, wherein at least one nucleic acid molecule encodes a fusion protein wherein the erythropoietin is fused to the *S. cerevisiae* αMATpre signal peptide and at least one nucleic acid molecule encodes a fusion protein wherein the erythropoietin is fused to the *S. cerevisiae* αMATpre signal peptide a chicken lysozyme signal peptide.

16. The method of claim 11, wherein the codons of the nucleic acid sequence of the nucleic acid molecule encoding the erythropoietin is optimized for expression in *Pichia pastoris*.

17. The method of claim 11, wherein the detectable cross binding activity with antibodies made against host cell antigens is determined in a sandwich ELISA.

18. The method of claim 11, wherein the detectable cross binding activity with antibodies made against host cell antigens is determined in a Western blot.

19. The method of claim 11, wherein recovering the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens from the medium includes a cation exchange chromatography step.

20. The method of claim 11, wherein recovering the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens from the medium includes a hydroxyapatite chromatography step.

21. The method of claim 11, wherein recovering the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable

cross binding activity with antibodies made against host cell antigens from the medium includes an anion exchange chromatography step.

22. A composition comprising a mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens obtained from the method of claim 18 and a pharmaceutically acceptable salt.

23. The composition of claim 22, wherein the mature human erythropoietin comprising predominantly sialic acid-terminated biantennary *N*-glycans and having no detectable cross binding activity with antibodies made against host cell antigens is conjugated to a hydrophilic polymer.

24. The composition of claim 23, wherein the hydrophilic polymer is a polyethylene glycol polymer.