



- (51) **International Patent Classification:**
A61K 39/395 (2006.01) *A61K 47/48* (2006.01)
- (21) **International Application Number:**
PCT/US2012/025812
- (22) **International Filing Date:**
20 February 2012 (20.02.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/446,853 25 February 2011 (25.02.2011) US
- (71) **Applicant (for all designated States except US):** **MERCK SHARP & DOHME CORP.** [US/US]; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** **HAMILTON, Stephen** [GB/US]; 21 Lafayette Street, Suite 200, Lebanon, New Hampshire 03766 (US). **COOK, William, J.** [US/US]; 29 Carriage Lane, Hanover, New Hampshire 03755 (US). **GOMATHINAYAGAM, Sujatha** [IN/US]; 21 Lafayette Street, Suite 200, Lebanon, New Hampshire 03766 (US).
- (74) **Common Representative:** **MERCK SHARP & DOHME CORP.**; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US).
- (81) **Designated States (unless otherwise indicated, for every kind of national protection available):** AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States (unless otherwise indicated, for every kind of regional protection available):** ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** PRODUCTION OF N- AND O-SIALYLATED TNFR2-Fc FUSION PROTEIN IN YEAST

Comparison of Glycan Composition

Sample	Source	N-glycan Profile (% of total N-glycans)			Terminal O-glycan Profile (% of total O-glycans)			Total sialic acid (mol Sialic acid per mol protein)
		Neutral	Mono-sialylated	Bi-sialylated	Mannose	Neutral (exc. mannose)	Sialylated	
Form 1	YGLY10299	43.3	44.5	12.2	100	0	0	3
Form 2	YGLY11731	25.7	42.4	31.9	100	0	0	5
Form 3	YGLY12680	44.6	44.0	11.4	18	53	29	10
Control	Commercial ENBREL (liquid formulation, Lot A)	26.0	52.4	21.6	Not determined			27

FIG.32

(57) **Abstract:** Production of recombinant Tumor Necrosis Factor Receptor fused to the Fc region of an antibody (TNFR2-Fc fragment fusion protein) in a glycoengineered yeast strain that is capable of producing sialylated N-glycans and O-glycans is described. Compositions of TNFR2-Fc fragment fusion protein comprising dystroglycan type O-glycans and sialylated N- and O-glycans with only terminal N-acetylneuraminic acid (NANA) residues in an α 2,6-linkage are provided. In particular aspects, methods are provided for modulating the in vivo pharmacokinetics of the TNFR2-Fc fragment fusion protein by altering the O-glycan structure on the molecule.



PRODUCTION OF N- AND O-SIALYLATED TNFRII-FC FUSION PROTEIN IN YEAST

CROSS REFERENCE TO RELATED APPLICATIONS

5

BACKGROUND OF THE INVENTION

(1) Field of the Invention

10 The present invention relates to the production of recombinant soluble tumor necrosis factor receptor II (TNFRII) fused to the Fc region of an antibody (TNFRII-Fc fragment fusion protein) in a glycoengineered yeast strain that is capable of producing sialylated *N*-glycans and *O*-glycans. In particular aspects, the present invention further relates to compositions of TNFRII-Fc fragment fusion protein comprising dystroglycan type *O*-glycans and sialylated *N*-
15 and *O*-glycans with only terminal *N*-acetylneuraminic acid (NANA) residues in an α 2,6-linkage. In particular aspects, the present invention relates to methods for modulating the *in vivo* pharmacokinetics of the TNFRII-Fc fragment fusion protein by altering the sialylation state of the molecule.

20 (2) Background of the Invention

Tumor necrosis factor receptor II (TNFRII) is a type I membrane glycoprotein belonging to the tumor necrosis factor (TNF) receptor superfamily and has an important role in independent signaling in chronic inflammatory conditions. Several inflammatory diseases and cancers display an increased and/or unregulated level of soluble TNFRII or polymorphisms.
25 These observations have suggested that TNFRII might be an important target in treatments for these inflammatory diseases and cancers. Currently, TNFRII is used in therapies for treating rheumatoid arthritis. By binding TNF α , a cytokine, and blocking its interactions with receptors. Etanercept is a commercially available product marketed under the tradename ENBREL that is approved for treating moderate to severe rheumatoid arthritis; psoriatic arthritis; ankylosing
30 spondylitis; chronic, moderate to severe psoriasis; and moderate to severe active polyarticular juvenile idiopathic arthritis. Etanercept is produced in Chinese hamster ovary (CHO) cells as a fusion protein consisting of the soluble domain of the TNFRII fused to the Fc region of an antibody (TNFRII-Fc). Soluble TNFRII-Fc fusion proteins and methods for producing them have been disclosed in Scallon *et al.*, Cytokine 7: 759-770 (1995); Olsen & Stein, N. Engl. J. Med. 350: 2167-2179 (2004), Davis *et al.*, Biotechnol. Prog., 16: 736-743 (2000), US Patent No.
35 5,605,690, US Patent No. 7,476,722, and US Patent No. 7,157,557.

Soluble TNFRII-Fc contains several *N*-glycosylation sites and multiple *O*-glycosylation sites. The extent and type of glycosylation is important as it conveys many desirable properties to the glycoprotein, including but not limited regulation of serum half-life and regulation of biological activity. In general, TNFRII-Fc produced in mammalian cells such as CHO cells has a glycosylation pattern that is similar to but not identical to the glycosylation pattern that would be produced in human cells. (See Wilson *et al.*, Apollo Cytokine Research Pty., (2006); Jiang *et al.* Apollo Cytokine Research Pty.; Hossler *et al.*, Glycobiol. 19: 936-949 (2009)). In addition, sialic acid on glycoproteins obtained from human cells is primarily of the *N*-acetylneuraminic acid (NANA) type. In contrast, the sialic acid on glycoproteins obtained from non-human cells such as CHO cells can include mixtures of NANA and *N*-glycolylneuraminic acid (NGNA). The ratio of NANA to NGNA is variable and depends on culturing conditions and cell line (Raju *et al.*, Glycobiol. 10: 477-486 (2000); Baker *et al.*, Biotechnol. Bioeng. 73: 188-202 (2001)). High levels NGNA has been shown to elicit an immune response (Noguchi *et al.*, J. Biochem. 117: 59-62 (1995)) and can cause the rapid removal of glycoproteins from serum (Flesher *et al.*, Biotechnol. Bioeng. 46: 309-407 (1995)).

Commercially available soluble TNFRII-Fc has been shown to be a useful product for treating a variety of inflammatory conditions and cancers. However, in light of the difference in glycosylation pattern between TNFRII-Fc produced in human cells verses TNFRII-Fc produced in non-human mammalian cell lines and the general observation that varying the glycosylation profile of a therapeutic glycoprotein can affect the pharmacokinetics and/or pharmacodynamics of the therapeutic glycoprotein, there remains a need for providing TNFRII-Fc with other glycosylation patterns. For example, it would be desirable to provide a TNFRII-Fc wherein the sialic acid is of only the NANA type.

SUMMARY OF THE INVENTION

The present invention provides a soluble recombinant tumor necrosis factor receptor II (TNFRII) fused to the Fc region of an antibody (TNFRII-Fc fragment fusion protein) produced in a glycoengineered yeast strain. The soluble TNFRII-Fc fragment fusion protein has sialylated *N*-glycans and *O*-glycans comprising sialic acid of only the NANA type, which further aspects are linked to the *N*-glycan or *O*-glycan in an $\alpha 2,6$ or $\alpha 2,3$ linkage. By modulating the amount and sialylation of the *O*-glycan structure on the molecule, the present invention enables the *in vivo* half-life of the TNFRII-Fc to be regulated.

Therefore, the present invention provides a composition comprising or consisting essentially of a recombinant fragment of human tumor necrosis factor receptor fused to the constant region of an antibody (TNFRII-Fc) wherein the TNFRII-Fc has *N*-glycans and *O*-glycans and wherein the *O*-glycans are of the dystroglycan-type, and pharmaceutically acceptable salts thereof.

In further aspects of the invention, the *N*-glycans and *O*-glycans on the TNFRII-Fc are predominantly sialylated with α 2,6 sialic acid residues. In other aspects of the invention, the *N*-glycans and *O*-glycans on the TNFRII-Fc are predominantly sialylated with α 2,3 sialic acid residues. In further still aspects, the *N*-glycans on the TNFRII-Fc lack fucose residues. In further still aspects, the *N*-glycans and *O*-glycans on the TNFRII-Fc, which are sialylated, comprise *N*-acetylneuraminic acid (NANA) and no *N*-glycolylneuraminic acid (NGNA).

In further still aspects, a ratio of mole sialic acid to mole of the TNFRII-Fc is at least 10. In further still aspects, a ratio of mole sialic acid to mole of the TNFRII-Fc is about 10 to 21. In further still aspects, a ratio of mole sialic acid to mole of the TNFRII-Fc is greater than 21.

In particular aspects, at least 50%, 60%, 70%, 80%, 90%, or 100% of the *N*-glycans are sialylated. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly mono-, bi-, tri-, or tetra-sialylated *N*-glycans. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly mono-sialylated *N*-glycans. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly bi-sialylated *N*-glycans. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly tri-sialylated *N*-glycans. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly tetra-sialylated *N*-glycans.

In further still aspects, the *O*-glycans on the TNFRII-Fc comprise or consist of predominantly sialylated *O*-glycans. In further still aspects, greater than 10%, 20%, 30%, 40%, or 50% of the *O*-glycans on the TNFRII-Fc comprise or consist of sialylated *O*-glycans. In further still aspects, less than 10%, 20%, 40% or 50% of the *O*-glycans on the TNFRII-Fc terminate in mannose.

In further still aspects, the TNFRII domain of the TNFRII-Fc comprises or consists of an amino acid sequence with at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the amino acid sequence for the TNFRII domain set forth in SEQ ID NO:73 or 75. The receptor domain includes amino acids 1 to 235 of SEQ ID NO:73 or 75 and is encoded by nucleotides 1-705 of SEQ ID NO:72 or 74.

Further provided is a method for producing a recombinant human tumor necrosis factor fused to the constant region of an antibody (TNFRII-Fc) having sialylated *N*-glycans and *O*-glycans comprising or consisting of (a) providing a recombinant yeast host cell genetically engineered to produce glycoproteins having sialylated *N*-glycans and further comprising (i) a nucleic acid molecule encoding the TNFRII-Fc; (ii) a nucleic acid molecule encoding an α 1,2-mannosidase activity linked to a heterologous targeting or signaling peptide that targets the mannosidase activity to the secretory pathway; and (iii) a nucleic acid molecule encoding an *O*-linked mannose β 1,2-*N*-acetylglucosaminyltransferase 1 (POMGnT1); (b) culturing the host cell

under conditions suitable for producing the TNFRII-Fc; and (c) recovering the TNFRII-Fc from the culture fluid to produce the TNFRII-Fc having sialylated *N*-glycans and *O*-glycans.

In further aspects, the POMGnT1 is provided as a fusion protein comprising the receptor domain of the POMGnT1 fused to a heterologous cellular targeting or signaling (or leader) peptide that targets the POMGnT1 to the secretory pathway, e.g., the ER or Golgi apparatus. Particular heterologous targeting or signal peptides include but are not limited to the *Saccharomyces cerevisiae* *MNN2*, *MNN5* or *MNN6* targeting or signal peptide.

In further aspects of the method, the *N*-glycans and *O*-glycans on the TNFRII-Fc are predominantly sialylated with α 2,6 sialic acid residues. In further still aspects, the *N*-glycans on the TNFRII-Fc lack fucose residues. In further still aspects, the *N*-glycans and *O*-glycans on the TNFRII-Fc, which are sialylated, comprise *N*-acetylneuraminic acid (NANA) and no *N*-glycolylneuraminic acid (NGNA).

In further still aspects, a ratio of mole sialic acid to a mole of the TNFRII-Fc is at least 10. In further still aspects, a ratio of mole sialic acid to mole of the TNFRII-Fc is about 10 to 21. In further still aspects, a ratio of mole sialic acid to mole of the TNFRII-Fc is greater than 21.

In particular aspects, at least 50%, 60%, 70%, 80%, 90%, or 100% of the *N*-glycans are sialylated. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly mono-, bi-, tri-, or tetra-sialylated *N*-glycans. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly mono-sialylated *N*-glycans. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly bi-sialylated *N*-glycans. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly tri-sialylated *N*-glycans. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly tetra-sialylated *N*-glycans.

In further still aspects, the *O*-glycans on the TNFRII-Fc comprise or consist of predominantly sialylated *O*-glycans. In further still aspects, greater than 10%, 20%, 30%, 40%, or 50% of the *O*-glycans on the TNFRII-Fc comprise or consist of sialylated *O*-glycans. In further still aspects, less than 10%, 20%, 40% or 50% of the *O*-glycans on the TNFRII-Fc terminate in mannose.

In further still aspects, the TNFRII domain of the TNFRII-Fc comprises or consists of an amino acid sequence with at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the amino acid sequence for the TNFRII domain set forth in SEQ ID NO:73 or 75. The receptor domain includes amino acids 1 to 235 of SEQ ID NO:73 or 75 and is encoded by nucleotides 1-705 of SEQ ID NO:72 or 74.

In further aspects of the method, the TNFRII-Fc is recovered from the culture fluid in a process comprising a hydroxyapatite or aminophenyl borate chromatography step. In further aspects of the method, the TNFRII-Fc is recovered from the culture fluid in a process

comprising an affinity capture chromatography step and a hydroxyapatite or aminophenyl borate chromatography step. In further aspects of the method, the TNFRII-Fc is recovered from the culture fluid in a process comprising the steps of an affinity capture chromatography step, a hydrophobic interaction chromatography step, a hydroxyapatite or aminophenyl borate chromatography step, and a cation exchange chromatography step.

Further provided is a composition comprising or consisting essentially of a recombinant fragment of human tumor necrosis factor receptor fused to the constant region of an antibody (TNFRII-Fc) wherein the TNFRII-Fc has *N*-glycans and *O*-glycans and wherein the *O*-glycans are *O*-mannose reduced glycans, and pharmaceutically acceptable salts thereof. An *O*-mannose reduced glycan is an *O*-glycan in which the predominant *O*-glycan consists predominantly of a single mannose (mannose type) or mannoside type (two mannose residues). In further aspects of the composition, the TNFRII domain of the TNFRII-Fc comprises or consists of an amino acid sequence with at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the amino acid sequence for the TNFRII domain set forth in SEQ ID NO:73 or 75. The receptor domain includes amino acids 1 to 235 of SEQ ID NO:73 or 75 and is encoded by nucleotides 1-705 of SEQ ID NO:72 or 74.

Further provided is a method for producing a recombinant human tumor necrosis factor fused to the constant region of an antibody (TNFRII-Fc) having sialylated *N*-glycans and *O*-mannose reduced glycans comprising or consisting of (a) providing a recombinant lower eukaryote host cell genetically engineered to produce glycoproteins having sialylated *N*-glycans and further comprising (i) a nucleic acid molecule encoding the TNFRII-Fc; and (ii) a nucleic acid molecule encoding an α 1,2-mannosidase activity linked to a heterologous targeting or signaling peptide that targets the mannosidase activity to the secretory pathway; (b) culturing the host cell under conditions suitable for producing the TNFRII-Fc; and (c) recovering the TNFRII-Fc from the culture fluid to produce the TNFRII-Fc having sialylated *N*-glycans and *O*-mannose reduced glycans.

In further aspects of the method, the TNFRII domain of the TNFRII-Fc comprises or consists of an amino acid sequence with at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the amino acid sequence for the TNFRII domain set forth in SEQ ID NO:73 or 75. The receptor domain includes amino acids 1 to 235 of SEQ ID NO:73 or 75 and is encoded by nucleotides 1-705 of SEQ ID NO:72 or 74.

In further aspects of the method, the host cells are cultured in the presence of a *PMT* inhibitor which reduces the number of sites on the TNFRII-Fc that are *O*-glycosylated.

Further provided is a pharmaceutical composition comprising or consisting of the polypeptide of any one of aspects above and a pharmaceutically suitable carrier.

Further provided is the use of the above pharmaceutical composition in the manufacture of a medicament for inflammatory diseases and cancers that display an increased

and/or unregulated level of soluble TNFR II or polymorphisms or the use of the pharmaceutical composition of claim 25 in the manufacture of a medicament for treating rheumatoid arthritis.

DEFINITIONS

As used herein, the terms "*N*-glycan" and "glycoform" are used interchangeably and refer to an *N*-linked oligosaccharide, for example, 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 Man₃GlcNAc₂ ("Man" refers to mannose; "Glc" refers to glucose; and "NAc" refers to *N*-acetyl; GlcNAc refers to *N*-acetylglucosamine). Usually, *N*-glycan structures are presented with the non-reducing end to the left and the reducing end to the right. The reducing end of the *N*-glycan is the end that is attached to the Asn residue comprising the glycosylation site on the protein. *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 Man₃GlcNAc₂ ("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."

With respect to complex *N*-glycans, the terms "G-2", "G-1", "G0", "G1", "G2", "A1", and "A2" mean the following. "G-2" refers to an *N*-glycan structure that can be characterized as Man₃GlcNAc₂; the term "G-1" refers to an *N*-glycan structure that can be characterized as GlcNAcMan₃GlcNAc₂; the term "G0" refers to an *N*-glycan structure that can

be characterized as $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$; the term "G1" refers to an *N*-glycan structure that can be characterized as $\text{GalGlcNAc}_2\text{Man}_3\text{GlcNAc}_2$; the term "G2" refers to an *N*-glycan structure that can be characterized as $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$; the term "A1" refers to an *N*-glycan structure that can be characterized as $\text{NANA Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$; and, the term "A2" refers to an *N*-glycan structure that can be characterized as $\text{NANA}_2\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$. Unless otherwise indicated, the terms "G-2", "G-1", "G0", "G1", "G2", "A1", and "A2" refer to *N*-glycan species that lack fucose attached to the GlcNAc residue at the reducing end of the *N*-glycan. When the term includes an "F", the "F" indicates that the *N*-glycan species contains a fucose residue on the GlcNAc residue at the reducing end of the *N*-glycan. For example, G0F, G1F, G2F, A1F, and A2F all indicate that the *N*-glycan further includes a fucose residue attached to the GlcNAc residue at the reducing end of the *N*-glycan. Lower eukaryotes such as yeast and filamentous fungi do not normally produce *N*-glycans that produce fucose.

With respect to multiantennary *N*-glycans, the term "multiantennary *N*-glycan" refers to *N*-glycans that further comprise a GlcNAc residue on the mannose residue comprising the non-reducing end of the 1,6 arm or the 1,3 arm of the *N*-glycan or a GlcNAc residue on each of the mannose residues comprising the non-reducing end of the 1,6 arm and the 1,3 arm of the *N*-glycan. Thus, multiantennary *N*-glycans can be characterized by the formulas $\text{GlcNAc}_{(2-4)}\text{Man}_3\text{GlcNAc}_2$, $\text{Gal}_{(1-4)}\text{GlcNAc}_{(2-4)}\text{Man}_3\text{GlcNAc}_2$, or $\text{NANA}_{(1-4)}\text{Gal}_{(1-4)}\text{GlcNAc}_{(2-4)}\text{Man}_3\text{GlcNAc}_2$. The term "1-4" refers to 1, 2, 3, or 4 residues.

With respect to bisected *N*-glycans, the term "bisected *N*-glycan" refers to *N*-glycans in which a GlcNAc residue is linked to the mannose residue at the reducing end of the *N*-glycan. A bisected *N*-glycan can be characterized by the formula $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ wherein each mannose residue is linked at its non-reducing end to a GlcNAc residue. In contrast, when a multiantennary *N*-glycan is characterized as $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$, the formula indicates that two GlcNAc residues are linked to the mannose residue at the non-reducing end of one of the two arms of the *N*-glycans and one GlcNAc residue is linked to the mannose residue at the non-reducing end of the other arm of the *N*-glycan.

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.

When referring to "mole percent" of a glycan present in a preparation of a
5 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 PNGase and then quantified by a method that is not affected by glycoform composition, (for instance, labeling a PNGase released glycan pool with a fluorescent tag such as 2-aminobenzamide and then
10 separating by high performance liquid chromatography or capillary electrophoresis and then quantifying glycans by fluorescence intensity). For example, 50 mole percent NANA₂Gal₂GlcNAc₂Man₃GlcNAc₂ means that 50 percent of the released glycans are NANA₂Gal₂GlcNAc₂Man₃GlcNAc₂ 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%,
15 more preferably above 50%, 55%, 60%, 65% or 70% and most preferably above 75%, 80% 85%, 90% or 95%.

The term "operably linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in trans or at a distance to control the
20 gene of interest.

The term "expression control sequence" or "regulatory sequences" are used interchangeably and as used herein refer to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operably linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and
25 translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The
30 nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner
35 sequences.

The term "transfect", "transfection", "transfecting" and the like refer to the introduction of a heterologous nucleic acid into eukaryote cells, both higher and lower eukaryote

cells. Historically, the term "transformation" has been used to describe the introduction of a nucleic acid into a yeast or fungal cell; however, herein the term "transfection" is used to refer to the introduction of a nucleic acid into any eukaryote cell, including yeast and fungal cells.

The term "eukaryotic" refers to a nucleated cell or organism, and includes insect
5 cells, plant cells, mammalian cells, animal cells and lower eukaryotic cells.

The term "lower eukaryotic cells" includes yeast and filamentous fungi. Yeast and filamentous fungi include, but are not limited to *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*,
10 *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Physcomitrella patens* and *Neurospora crassa*. *Pichia* sp., any *Saccharomyces* sp., *Hansenula polymorpha*, any *Kluyveromyces* sp., *Candida albicans*, any *Aspergillus* sp., *Trichoderma reesei*,
15 *Chrysosporium lucknowense*, any *Fusarium* sp. and *Neurospora crassa*.

As used herein, the terms "antibody," "immunoglobulin," "immunoglobulins" and "immunoglobulin molecule" are used interchangeably. Each immunoglobulin molecule has a unique structure that allows it to bind its specific antigen, but all immunoglobulins have the same
20 overall structure as described herein. The basic immunoglobulin structural unit is known to comprise a tetramer of subunits. Each tetramer has two identical pairs of polypeptide chains, each pair having one "light" chain (about 25 kDa) and one "heavy" chain (about 50-70 kDa). The amino-terminal portion of each chain includes a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The carboxy-terminal portion of each
25 chain defines a constant region primarily responsible for effector function. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, and define the antibody's isotype as IgG, IgM, IgA, IgD, and IgE, respectively.

The term "Fc fragment" refers to the 'fragment crystallized' C-terminal region of the antibody containing the CH2 and CH3 domains.

As used herein, the term "consisting essentially of" will be understood to imply the inclusion of a stated integer or group of integers; while excluding modifications or other integers which would materially affect or alter the stated integer. With respect to species of *N*-glycans, the term "consisting essentially of" a stated *N*-glycan will be understood to include the *N*-glycan whether or not that *N*-glycan is fucosylated at the *N*-acetylglucosamine (GlcNAc)
35 which is directly linked to the asparagine residue of the glycoprotein.

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 or HPLC. 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 at 40 mole percent, species B at 35 mole percent and species C at 25 mole percent, the composition comprises predominantly species A, and species B would be the next most predominant species. Some host cells may produce compositions comprising neutral *N*-glycans and charged *N*-glycans such as mannosylphosphate or sialic acid. Therefore, a composition of glycoproteins can include a plurality of charged and uncharged or neutral *N*-glycans. In the present invention, it is within the context of the total plurality of *N*-glycans in the composition in which the predominant *N*-glycan determined. Thus, as used herein, “predominant *N*-glycan” means that of the total plurality of *N*-glycans in the composition, the predominant *N*-glycan is of a particular structure.

As used herein, the term “essentially free of” a particular sugar residue, such as fucose, or galactose and the like, is used to indicate that the glycoprotein composition is substantially devoid of *N*-glycans which contain such residues. Expressed in terms of purity, essentially free means that the amount of *N*-glycan structures containing such sugar residues does not exceed 10%, and preferably is below 5%, more preferably below 1%, most preferably below 0.5%, wherein the percentages are by weight or by mole percent. Thus, substantially all of the *N*-glycan structures in a glycoprotein composition according to the present invention are free of, for example, fucose, or galactose, or both.

As used herein, a glycoprotein composition “lacks” or “is lacking” a particular sugar residue, such as fucose or galactose, when no detectable amount of such sugar residue is present on the *N*-glycan structures at any time. For example, in preferred embodiments of the present invention, the glycoprotein compositions are produced by lower eukaryotic organisms, as defined above, including yeast (for example, *Pichia* sp.; *Saccharomyces* sp.; *Kluyveromyces* sp.; *Aspergillus* sp.), and will “lack fucose,” because the cells of these organisms do not have the enzymes needed to produce fucosylated *N*-glycan structures. Thus, the term “essentially free of fucose” encompasses the term “lacking fucose.” However, a composition may be “essentially free of fucose” even if the composition at one time contained fucosylated *N*-glycan structures or contains limited, but detectable amounts of fucosylated *N*-glycan structures as described above.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-G are flow-diagrams showing the construction of strains YGLY11731, YGLY10299, and YGLY13571, each strain capable of producing a TNFRII-Fc fragment fusion protein comprising sialylated *N*-glycans.

Figures 2A-B show the construction of **YGLY12680**, a strain capable of producing a TNFRII-Fc fragment fusion protein comprising sialylated *N*-glycans and *O*-glycans.

Figure 3 shows the construction of strain **YGLY14252**, a strain capable of producing a TNFRII-Fc fragment fusion protein comprising sialylated *N*-glycans and *O*-glycans.

Figure 4 shows the construction of strains **YGLY14954** and **YGLY14927**, each strain capable of producing a TNFRII-Fc fragment fusion protein comprising sialylated *N*-glycans and *O*-glycans.

Figure 5 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 6 shows a map of plasmid pGLY40. 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 7 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 8 shows a map of plasmid pGLY48. Plasmid pGLY48 is an integration vector that targets the *MNN4LI* 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 9 shows as map of plasmid pGLY45. Plasmid pGLY45 is an integration
 5 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'
 10 region of the *MNN4* gene (PpMNN4-3').

Figure 10 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
 15 (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.
 20

Figure 11 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)
 30 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.
 35

Figure 12 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 13 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 14 shows a map of plasmid pGLY3419 (pSH1110). Plasmid pGLY3419 (pSH1110) is an integration vector that contains an 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* *BMT1* gene (PBS1 5') and on the other side with the 3' nucleotide sequence of the *P. pastoris* *BMT1* gene (PBS1 3').

Figure 15 shows a map of plasmid pGLY3421 (pSH1106). Plasmid pGLY3421 (pSH1106) contains an 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* *BMT3* gene (PpPBS3 5') and on the other side with the 3' nucleotide sequence of the *P. pastoris* *BMT3* gene (PpPBS3 3').

Figure 16 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 α -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 17 shows a map of plasmid pGLY5048. Plasmid pGLY5048 is an integration vector that targets the *STE13* locus and contains expression cassettes encoding (1) the *T. reesei* α -1,2-mannosidase catalytic domain fused at the N-terminus to *S. cerevisiae* α MATpre signal peptide (α MATTrMan) to target the chimeric protein to the secretory pathway and secretion from the cell and (2) the *P. pastoris* *URA5* gene or transcription unit.

Figure 18 shows a map of plasmid pGLY5019. Plasmid pGLY5019 is an integration vector that targets the *DAP2* locus and contains an expression cassette comprising a nucleic acid molecule encoding the Nourseothricin resistance (NAT^R) ORF operably linked to the *Ashbya gossypii* *TEF1* promoter and *A. gossypii* *TEF1* termination sequences flanked one side with the 5' nucleotide sequence of the *P. pastoris* *DAP2* gene and on the other side with the 3' nucleotide sequence of the *P. pastoris* *DAP2* gene.

Figure 19 is a map of plasmid pGLY5045. Plasmid pGLY5045 is a roll-in integration vector that targets the *URA6* locus and contains an expression cassette encoding the TNFRII-Fc fragment fusion protein. The plasmid contains two expression cassettes, each comprising a nucleic acid molecule encoding the TNFRII-Fc fragment fusion protein fused at the 5' end to a nucleic acid molecule encoding the human serum albumin signal peptide, which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *AOX1* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* transcription termination sequence. The plasmid also includes a ZeocinR expression cassette comprising a nucleic acid molecule encoding the Sh ble ORF operably linked at the 5' end to the *S. cerevisiae* *TEF1* promoter and at the 3' end to the *S. cerevisiae* *CYC* termination sequence.

Figure 20 shows a plasmid map of pGLY6391. Plasmid pGLY6391 is a roll-in integration vector that targets the *THR1* locus and contains an expression cassette encoding the TNFRII-Fc fragment fusion protein. The plasmid contains two expression cassettes, each comprising a nucleic acid molecule encoding the TNFRII-Fc fragment fusion protein without the C-terminal lysine residue fused at the 5' end to a nucleic acid molecule encoding the human serum albumin signal peptide, which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *AOX1* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* transcription termination sequence. The plasmid also includes a ZeocinR expression cassette comprising a nucleic acid molecule encoding the Sh ble ORF operably linked at the 5' end to the *S. cerevisiae* *TEF1* promoter and at the 3' end to the *S. cerevisiae* *CYC* termination sequence.

Figure 21 shows a plasmid map of pGLY5085. Plasmid pGLY5085 is a KINKO plasmid for introducing a second set of the genes involved in producing sialylated *N*-glycans into *P. pastoris*. The plasmid is similar to plasmid YGLY2456 except that the *P. pastoris* *ARG1* gene

has been replaced with an expression cassette encoding hygromycin resistance (HygR) and the plasmid targets the *P. pastoris* *TRP5* locus. The six tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and ORF of the *TRP5* gene ending at the stop codon followed by a *P. pastoris* *ALG3* termination sequence and
 5 on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *TRP5* gene.

Figure 22 shows a plasmid map of pGLY5755. Plasmid pGLY5755 is a KINKO integration plasmid that encodes a chimeric mouse POMGnT I and targets the *HIS3* locus in *P. pastoris*. The expression cassette encoding the chimeric mouse POMGnT I comprises a nucleic
 10 acid molecule encoding the catalytic domain of the mouse POMGnT I ORF codon-optimized for effective expression in *P. pastoris* ligated in-frame with a nucleic acid molecule encoding *S. cerevisiae* *MNN2*-s signal peptide (53) operably linked at the 5' end to a nucleic acid molecule that has the inducible *P. pastoris* *AOX1* promoter sequence and at the 3' end to a nucleic acid molecule that has the *S. cerevisiae* *CYC* transcription termination sequence. For selecting
 15 transformants, the plasmid comprises an expression cassette encoding the *S. cerevisiae* *ARR3* ORF in which the nucleic acid molecule encoding the ORF is operably linked at the 5' end to a nucleic acid molecule having the *P. pastoris* *RPL10* promoter sequence and at the 3' end to a nucleic acid molecule having the *S. cerevisiae* *CYC* transcription termination sequence.

Figure 23 shows a plasmid map of pGLY5086. Plasmid pGLY5086 is a KINKO
 20 plasmid for introducing a second set of the genes involved in producing sialylated *N*-glycans into *P. pastoris*. The plasmid is similar to plasmid YGLY5085 except that the plasmid targets the *P. pastoris* *THR1* locus.

Figure 24 shows a plasmid map of pGLY5219. Plasmid pGLY5219 (**Figure 24**)
 is an integration plasmid that encodes a chimeric mouse POMGnT I and targets the *VPS10-1*
 25 locus in *P. pastoris*. The expression cassette encoding the chimeric mouse POMGnT I comprises a nucleic acid molecule encoding the catalytic domain of the mouse POMGnT I ORF ORF codon-optimized for effective expression in *P. pastoris* ligated in-frame with a nucleic acid molecule encoding *S. cerevisiae* *Mnn6*-s signal peptide (65) operably linked at the 5' end to a nucleic acid molecule that has the constitutive *P. pastoris* *GAPDH* promoter sequence (SEQ ID
 30 NO:5) and at the 3' end to a nucleic acid molecule that has the *S. cerevisiae* *CYC* transcription termination sequence. For selecting transformants, the plasmid comprises an expression cassette comprising the *URA5* gene flanked by *lacZ* repeats.

Figure 25 shows a map of pGLY5192. Plasmid pGLY5192 is an integration
 plasmid that targets the *VPS10-1* locus. The plasmid comprises an expression cassette
 35 comprising the *URA5* gene flanked by *lacZ* repeats flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *VPS10-1* gene and on the

other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *VPS10-1* gene.

Figure 26 shows a map of pGLY7087cv. Plasmid pGLY7087cv is a KINKO integration plasmid that encodes a chimeric mouse POMGnT I and targets the *HIS3* locus in *P. pastoris*. The expression cassette encoding the chimeric mouse POMGnT I comprises a nucleic acid molecule encoding the catalytic domain of the mouse POMGnT I ORF codon-optimized for effective expression in *P. pastoris* ligated in-frame with a nucleic acid molecule encoding *S. cerevisiae* Mnn5-s signal peptide (56) operably linked at the 5' end to a nucleic acid molecule that has the constitutive *P. pastoris GAPDH* promoter sequence and at the 3' end to a nucleic acid molecule that has the *S. cerevisiae CYC* transcription termination sequence. For selecting transformants, the plasmid comprises an expression cassette encoding the *S. cerevisiae ARR3* ORF in which the nucleic acid molecule encoding the ORF is operably linked at the 5' end to a nucleic acid molecule having the *P. pastoris RPL10* promoter sequence and at the 3' end to a nucleic acid molecule having the *S. cerevisiae CYC* transcription termination sequence.

Figure 27 shows the amino acid sequence of TNFRII-Fc (SEQ ID NO:75). Represented are the features: TNFRII ectodomain (in bold); IgG1 Fc domain (regular text); cysteine-rich subdomains of TNFRII domain (outlined by arrows); N-linked glycosylation sites ("N" residues encircled); and, optional C-terminal lysine (in brackets).

Figure 28 shows a comparison of mucin-type *O*-glycosylation and dystroglycan-type *O*-glycosylation.

Figure 29 shows a schematic representation of the *O*-glycosylation engineering strategy for TNFRII-Fc. Form 1: mannose-reduced *O*-glycans (strain YGLY10299); Form 2: mannose-reduced *O*-glycans and enhanced sialylation of *N*-glycans (strain YGLY11731); Form 3: sialylated *O*-glycans (strain YGLY12680). Forms 5A, 5B & 5C: sialylated *O*-glycans (strain YGLY14252). Form 7A: sialylated *O*-glycans (strain YGLY14954).

Figure 30 shows a schematic representation of a purification strategy for recovering TNFRII-Fc produced in recombinant strains.

Figure 31 shows a composite of gradient SDS-PAGE analyses of TNFRII-Fc purified using the method shown in **Figure 30**. Purified TNFRII-Fc samples were resolved on 4-20% Tris-HCl BIORAD gels loaded with 3 µg/mL of reduced (R) or non-reduced (NR) TNFRII-Fc. Form 1: mannose-reduced *O*-glycans (strain YGLY10299); Form 2: mannose-reduced *O*-glycans and enhanced sialylation of *N*-glycans (strain YGLY11731); Form 3: sialylated *O*-glycans (strain YGLY12680). The control was commercial ENBREL.

Figure 32 shows a table comparing the glycans composition of Form 1, Form 2, and Form 3 TNFRII-Fc. Form 1: mannose-reduced *O*-glycans (strain YGLY10299); Form 2: mannose-reduced *O*-glycans and enhanced sialylation of *N*-glycans (strain YGLY11731); Form 3: sialylated *O*-glycans (strain YGLY12680).

Figure 33 shows the results of *in vitro* TNFRII-Fc-induced cell killing of L929 cells. Experimental design: L929 cells seeded overnight in 96-well plate (1×10^4 /well); cells treated with human recombinant TNF α (0.25 ng/mL) +/- TNFRII-Fc and incubated for 24 hours; and cell viability measured by ATPlite (luminescence readout). Form 1: mannose-reduced *O*-glycans (strain YGLY10299); Form 2: mannose-reduced *O*-glycans and enhanced sialylation of *N*-glycans (strain YGLY11731); Form 3: sialylated *O*-glycans (strain YGLY12680). The control was commercial ENBREL.

Figure 34 shows the results of *in vitro* TNFRII-Fc-stimulated release of IL-6 in A549 cells. Experimental design: A549 cells seeded at 5×10^4 per well in a 96 well plate and allowed to recover overnight; TNFRII-Fc samples titrated in triplicate; cells stimulated with 3 ng/mL human recombinant TNF α overnight at 37°C; and IL6 production determined by AlphaLISA assay. Form 1: mannose-reduced *O*-glycans (strain YGLY10299); Form 2: mannose-reduced *O*-glycans and enhanced sialylation of *N*-glycans (strain YGLY11731); Form 3: sialylated *O*-glycans (strain YGLY12680). The control was commercial ENBREL.

Figure 35 shows the results of *in vivo* rat pharmacokinetic analysis of TNFRII-Fc. Sprague Dawley (SD) rats were dosed SC at 1mg/kg and serum samples collected at 4, 24, 48, 72, 96, 120, 144 and 168 hr. Serum TNFRII-Fc concentration was determined with a Gyro immunoassay using anti-TNFRII antibody for capture and labeled-anti-Fc antibody for detection. Form 1: mannose-reduced *O*-glycans (strain YGLY10299); Form 2: mannose-reduced *O*-glycans and enhanced sialylation of *N*-glycans (strain YGLY11731); Form 3: sialylated *O*-glycans (strain YGLY12680). The control was commercial ENBREL.

Figure 36 shows a schematic representation of a purification strategy for recovering TNFRII-Fc from strain YGLY14252. Form 5A, hydroxyl apatite (HA) unbound wash purified. Form 5C, HA bound TNFRII-Fc eluted and purified. Form B, a 1:1 mix of Form 5A and 5C. The control was commercial ENBREL.

Figure 37 shows a composite of gradient SDS-PAGE analyses of TNFRII-Fc purified using the method shown in **Figure 36**. Purified TNFRII-Fc samples were resolved on 4-20% Tris-HCl BIORAD gels loaded with 2.5 μ g/lane of non-reduced (NR) TNFRII-Fc. YGLY14252. The control was commercial ENBREL.

Figure 38 shows a table comparing the glycans composition of TNFRII-Fc in Form 5A, Form 5B, and Form 5C.

Figure 39 shows a table comparing the *in vitro* TNFRII-Fc-induced cell killing of L929 cells and the *in vitro* TNFRII-Fc fragment fusion protein-stimulated release of IL-6 in A549 cells of TNFRII-Fc Form 5A, Form 5B, and Form 5C. Assays were performed as in **Figures 33** and **34**. The control was commercial ENBREL.

Figure 40 shows the results of *in vivo* rat pharmacokinetic analysis of TNFRII-Fc fragment fusion protein. SD rats were dosed SC at 1mg/kg and serum samples collected at 4, 24,

48, 72, 96, 120, 144 and 168 hr. Serum TNFRII-Fc fragment fusion protein concentration was determined with a Gyro immunoassay using anti-TNFRII as capture and anti-Fc as detection. The control was commercial ENBREL.

Figure 41 shows the results of *in vivo* mouse pharmacokinetic analysis of TNFRII-Fc fragment fusion protein. Mice were dosed with TNFRII-Fc fragment fusion protein SC at varying doses (0.1, 1, 5, 10 and 20 mg/kg) and the serum harvested at 48 hours post-inoculation. Serum TNFRII-Fc fusion protein concentration was determined with a Gyro immunoassay using anti-TNFRII as capture and anti-Fc as detection. The control was commercial ENBREL.

Figure 42 shows the results of the *in vivo* mouse chronic rheumatoid arthritic model. Transgenic mice were separated into 7 groups consisting of 8 gender and age-matched mice each, which received intraperitoneally 10 μ l of test compounds per gram of body weight, twice weekly. The groups received different test materials and dose levels, as follows: Vehicle, *Pichia* TNFRII-Fc at 30, 10 and 3 mg/kg; commercial Enbrel at 30, 10 and 3 mg/kg. Treatment was initiated at the onset of arthritis (three weeks of age) and continued over 8 weeks; the study was concluded at 10 weeks of age.

Figure 43 shows a schematic representation of an alternative purification strategy for recovering TNFRII-Fc with enriched sialic acid content.

Figure 44 shows a composite of gradient SDS-PAGE analyses of TNFRII-Fc purified isolated from strain YGLY14954, using the method shown in **Figure 43**. Purified TNFRII-Fc samples were resolved on 4-20% Tris-HCl BIORAD gels loaded with 2.5 μ g/Lane of non-reduced TNFRII-Fc. The control was commercial ENBREL.

Figure 45 shows a table comparing the glycans composition of TNFRII-Fc in Form 7A and commercial ENBREL.

Figure 46 shows the results of *in vivo* rat pharmacokinetic analysis of TNFRII-Fc fragment fusion protein purified by the Prosep-PB strategy compared to commercial ENBREL. SD rats were dosed SC at 1 mg/kg and serum samples collected at 4, 24, 48, 72, 96, 120, 144 and 168 hours. Serum TNFRII-Fc fragment fusion protein concentration was determined with a Gyro immunoassay using anti-TNFRII as capture and anti-Fc as detection. The control was commercial ENBREL.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides compositions comprising a recombinant human tumor necrosis factor fused to the constant region of an antibody (TNFRII-Fc fragment fusion protein) wherein the recombinant TNFRII-Fc fragment fusion protein comprises sialylated, afucosylated *N*-glycans and *O*-glycans. The sialylated *O*-glycans are of the dystroglycan type and not the mucin type. The sialic acid residue comprising the *N*-glycans and *O*-glycans consist only

of *N*-acetylneuraminic acid (NANA) residues. In addition, the sialic acid residues are linked to the non-reducing end of the oligosaccharide comprising the *N*-glycan and *O*-glycans in an α 2,6 linkage. Further provided are host cells for making the a recombinant TNFRII-Fc fragment fusion protein.

5 *N*-linked and *O*-linked are two major types of glycosylation. *N*-linked glycosylation (*N*-glycosylation) is characterized by the β -glycosylamine linkage of *N*-acetylglucosamine (GlcNAc) to asparagine (Asn) (Spiro, *Glycobiol.* 12: 43R-56R (2002)). It has been well established that the consensus sequence motif Asn-X-Ser/Thr is essential in *N*-glycosylation (Blom *et al.*, *Proteomics* 4: 1633-1649 (2004)). The most abundant form of *O*-linked glycosylation (*O*-glycosylation) is of the mucin-type, which is characterized by α -*N*-acetylgalactosamine (GalNAc) attached to the hydroxyl group of serine/threonine (Ser/Thr) side chains by the enzyme UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase (Hang & Bertozzi, *Bioorg. Med. Chem.* 13: 5021-5034 (2005); Julenius *et al.*, *Glycobiol.* 15: 153-164 (2005)). Mucin-type *O*-glycans can further include galactose and sialic acid residues. Mucin-type *O*-glycosylation is commonly found in many secreted and membrane-bound mucins in mammal, although it also exists in other higher eukaryotes (Hanish, *Biol Chem.* 382: 143-149 (2001)). As the main component of mucus, a gel playing crucial role in defending epithelial surface against pathogens and environmental injury, mucins are in charge of organizing the framework and conferring the rheological property of mucus. Beyond the above properties exhibited by mucins, mucin-type *O*-glycosylation is also known to modulate various protein functions *in vivo* (Hang & Bertozzi, *Bioorg. Med. Chem.* 13: 5021-5034 (2005)). For instance, mucin-like glycans can serve as receptor-binding ligands during an inflammatory response (McEver & Cummings, *J. Chin. Invest.* 100: 485-491 (1997

Another form of *O*-glycosylation is that of the *O*-mannose-type glycosylation (T. Endo, *BBA* 1473: 237-246 (1999)). In mammalian organisms this form of glycosylation can be sub-divided into two forms. The first form is the addition of a single mannose to a serine or threonine residue of a protein. This is a rare occurrence and has been demonstrated on very few proteins, including IgG2 light chain (Martinez *et al.*, *J. Chromatogr. A.* 1156: 183-187 (2007)). A more common form of *O*-mannose-type glycosylation in mammalian systems is that of the dystroglycan-type, which is characterized by β -*N*-acetylglucosamine (GlcNAc) attached to a mannose residue attached to the hydroxyl group of serine/threonine side chains in an α 1 linkage by an *O*-linked mannose β 1,2-*N*-acetylglucosaminyltransferase 1 (POMGnT1) (T. Endo, *BBA* 1473: 237-246 (1999)). Dystroglycan-type *O*-glycans can further include galactose and sialic acid residues. Unlike *N*-glycosylation, the consensus motif has not been identified in the sequence context of mucin or dystroglycan *O*-glycosylation sites.

In fungi such as *Pichia pastoris*, *O*-glycosylation produces *O*-glycans that can include up to five or six mannose residues (*See* for example, Tanner & Lehle, *Biochim. Biophys.*

Acta 906: 81–89 (1987); Herscovics & Orlean, FASEB J. 7: 540–550 (1993); Trimble et al., GlycoBiol. 14: 265–274 (2004); Lommel & Strahl, Glycobiol. 19: 816–828 (2009). Wild-type *Pichia pastoris* as shown in **Figure 29** can produce *O*-mannose-type *O*-glycans consisting of up to six mannose residues in which the terminal mannose residue can be phosphorylated. By abrogating phosphomannosyltransferase activity and β -mannosyltransferase activity in the *Pichia pastoris*, which results in charge-free *O*-glycans without β -linked mannose residues, and cultivating the *Pichia pastoris* lacking phosphomannosyltransferase activity and β -mannosyltransferase activity in the presence of a protein *PMT* inhibitor, which reduces *O*-glycosylation site occupancy, and a secreted α 1,2-mannosidase, which reduces the chain length of the charge-free *O*-glycans, *O*-mannose reduced glycans (or mannose-reduced *O*-glycans) can be produced (See U.S. Published Application No. 20090170159 and U.S. Patent No.). The consensus motif has not been identified in the sequence context of fungal *O*-glycosylation sites.

Mucin-type *O*-glycosylation is primarily found on cell surface proteins and secreted proteins. Dystroglycan-type *O*-glycosylation is primarily associated with proteins comprising the extracellular matrix. Both mucin- and dystroglycan-type *O*-glycans may possess terminal sialic acid residues. As shown in **Figure 28**, the terminal sialic acid residues are in α 2,3 linkage with the preceding galactose residue. In some instances, as shown in **Figure 28**, mucin-type *O*-glycans can also possess a branched α 2,6 sialic acid residue. The sialic acid present on each type of structure on glycoproteins obtained from recombinant non-human cell lines can include mixtures of *N*-acetylneuraminic acid (NANA) and *N*-glycolylneuraminic acid (NGNA). However, in contrast to glycoproteins obtained from mammalian cells, the sialic acid present on each type of structure on glycoproteins obtained from human cells is primarily composed of NANA. Thus, glycoprotein compositions obtained from mammalian cell culture include sialylated *N*-glycans that have a structure primarily associated to glycoproteins produced in non-human mammalian cells. ENBREL (etanercept) is a commercially provided TNFRII-Fc fragment fusion protein composition that is produced in Chinese Hamster Ovary (CHO) cells. U.S. Patent No. 5,459,031 discloses that the level of NGNA in a glycoprotein produced by a mammalian recombinant host cell can be controlled by monitoring and adjusting the levels of CO₂ during production of the glycoprotein in the host cell. The method was shown to reduce but not eliminate the presence of NGNA in the glycoprotein. In contrast, the present invention provides methods for producing TNFRII-Fc fusion protein compositions wherein the NANA is the only sialic acid species on the glycoprotein.

The *N*-glycan and *O*-glycan profiles of the several compositions of TNFRII-Fc fragment fusion protein of the present invention are shown in **Figures 32 and 38**. **Figure 32** shows the glycosylation profiles for TNFRII-Fc fragment fusion protein produced in strain **YGLY12680**, a *Pichia pastoris* strain genetically engineered to produce sialylated *N*-glycans and *O*-glycans, compared to the profile of a TNFRII-Fc fragment fusion protein produced in strains

that lacks the ability to produce sialylated *O*-glycans. Strain **YGLY12680** is a genetically engineered strain that includes a chimeric POMGnT I comprising the catalytic domain of POMGnT I fused to a heterologous targeting or signaling peptide that targets the chimeric POMGnT to the endoplasmic reticulum (ER) or Golgi apparatus, which transfers a GlcNAc residue to the *O*-linked mannose residue of an *O*-glycan, and a duplication of the nucleic acid molecules encoding a chimeric α 2,6-sialyltransferase (α 2,6ST) comprising the catalytic domain of an α 2,6ST fused to a heterologous targeting or signaling peptide that targets the chimeric α 2,6ST to the ER or Golgi apparatus, and the enzymes involved in making the CMP-sialic acid substrate for the chimeric α 2,6ST. Because yeast do not include an endogenous sialic acid pathway, the sialylated *N*-glycans and *O*-glycans produced by the strain are only of the NANA type. Thus, the strains herein produce sialylated *N*-glycans and *O*-glycans that include only the NANA type, similar to the *N*-glycans and *O*-glycans produced in human cells. This is in contrast to mammalian cells that produce *N*-glycans and *O*-glycans in a mixture of NANA and NGNA types. In general, the mole of sialic acid per mole of protein produced in strain **YGLY12680** was about 10. Sialylated *N*-glycans were the predominant species in the strain of which the predominant subspecies was mono-sialylated. Neutral *O*-glycans were the predominant species in the strain and were of the dystroglycan type. Neutral *N*-glycans in either glycoform include galactose-, GlcNAc-, or mannose-terminated oligosaccharide chains.

Figure 38 shows the glycosylation profiles for TNFRII-Fc fragment fusion protein produced in strain **YGLY14252**. The TNFRII-Fc fragment fusion protein was fractionated into three fractions, and the glycosylation profiles determined for each fraction. The mole of sialic acid per mole of protein ranged from about 11 to 21 depending on the fraction. For Form 5A, the sialylated *N*-glycan and *O*-glycan glycoforms comprised the predominant species. As shown in **Figures 40-41**, Form 5A pharmacokinetics was similar to commercially available ENBREL where as the less sialylated forms (Form 5B and 5C) had reduced pharmacokinetics compared to ENBREL. The sialylated *N*-glycans and *O*-glycans produced by the strain are only of the NANA type. The TNFRII-Fc produced in the recombinant *Pichia pastoris* strains when compared to commercial Enbrel in the mouse chronic rheumatoid arthritic model demonstrated a dose dependent potency similar to commercial Enbrel (**Figure 42**).

Therefore, the present invention provides a composition comprising or consisting essentially of a recombinant fragment of human tumor necrosis factor receptor fused to the constant region of an antibody (TNFRII-Fc) wherein the TNFRII-Fc has *N*-glycans and *O*-glycans and wherein the *O*-glycans are of the dystroglycan- or *O*-man type, and pharmaceutically acceptable salts thereof.

In further aspects of the composition, the *N*-glycans and *O*-glycans on the TNFRII-Fc are predominantly sialylated with α 2,6 or α 2,3 sialic acid residues. In further still aspects of the composition, the *N*-glycans on the TNFRII-Fc lack fucose residues; however, in

particular aspects of the composition, one or more of the N-glycans on the TNFRII-Fc are fucosylated. In further still aspects, the N-glycans and O-glycans on the TNFRII-Fc, which are sialylated, comprise N-acetylneuraminic acid (NANA) and no N-glycolylneuraminic acid (NGNA).

5 In further still aspects of the composition, a ratio of mole sialic acid to mole of the TNFRII-Fc is at least 10. In further still aspects of the composition, a ratio of mole sialic acid to mole of the TNFRII-Fc is about 10 to 21. In further still aspects of the composition, a ratio of mole sialic acid to mole of the TNFRII-Fc is greater than 21.

10 In further aspects of the composition, at least 50%, 60%, 70%, 80%, 90%, or 100% of the N-glycans are sialylated. In further still aspects, the N-glycans on the TNFRII-Fc comprise or consist of predominantly mono-, bi-, tri-, or tetra-sialylated N-glycans. In further still aspects, the N-glycans on the TNFRII-Fc comprise or consist of predominantly mono-sialylated N-glycans. In further still aspects of the composition, the N-glycans on the TNFRII-Fc comprise or consist of predominantly bi-sialylated N-glycans. In further still aspects, the N-glycans on the TNFRII-Fc comprise or consist of predominantly tri-sialylated N-glycans. In further still aspects of the composition, the N-glycans on the TNFRII-Fc comprise or consist of predominantly tetra-sialylated N-glycans.

20 In further still aspects of the composition, the O-glycans on the TNFRII-Fc comprise or consist of predominantly sialylated O-glycans. In further still aspects, greater than 10%, 20%, 30%, 40%, or 50% of the O-glycans on the TNFRII-Fc comprise or consist of sialylated O-glycans. In further still aspects, less than 10%, 20%, 40% or 50% of the O-glycans on the TNFRII-Fc terminate in mannose.

25 In further still aspects of the composition, the TNFRII domain of the TNFRII-Fc comprises or consists of an amino acid sequence with at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the amino acid sequence for the TNFRII domain set forth in SEQ ID NO:73 or 75. The receptor domain includes amino acids 1 to 235 of SEQ ID NO:73 or 75 and is encoded by nucleotides 1-705 of SEQ ID NO:72 or 74. The receptor domain includes amino acids 1 to 235 of SEQ ID NO:73 or 75 and is encoded by nucleotides 1-705 of SEQ ID NO:72 or 74.

30 Further provided is a composition comprising or consisting essentially of a recombinant fragment of human tumor necrosis factor receptor fused to the constant region of an antibody (TNFRII-Fc) wherein the TNFRII-Fc has N-glycans and O-glycans and wherein the O-glycans are O-mannose reduced glycans, and pharmaceutically acceptable salts thereof. An O-mannose reduced glycan is an O-glycan in which the predominant O-glycan consists of a single mannose (mannose type) or mannanose type (two mannose residues). In further aspects of the composition, the TNFRII domain of the TNFRII-Fc comprises or consists of an amino acid sequence with at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the

amino acid sequence for the TNFRII domain set forth in SEQ ID NO:73 or 75. The receptor domain includes amino acids 1 to 235 of SEQ ID NO:73 or 75 and is encoded by nucleotides 1-705 of SEQ ID NO:72 or 74.

Lower eukaryotes such as yeast or filamentous fungi are often used for expression of recombinant glycoproteins because they can be economically cultured, give high yields, and when appropriately modified are capable of suitable glycosylation. Yeast in particular offers established genetics allowing for rapid transfections, tested protein localization strategies and facile gene knock-out techniques. Suitable vectors have expression control sequences, such as promoters, including 3-phosphoglycerate kinase or other glycolytic enzymes, and an origin of replication, termination sequences, and the like as desired. These glycoengineered host cells enable the production of the TNFRII-Fc comprising the compositions disclosed herein.

Therefore, further provided is a method for producing a recombinant human tumor necrosis factor fused to the constant region of an antibody (TNFRII-Fc) having sialylated *N*-glycans and *O*-glycans comprising or consisting of (a) providing a recombinant lower eukaryote host cell genetically engineered to produce glycoproteins having sialylated *N*-glycans and further comprising (i) a nucleic acid molecule encoding the TNFRII-Fc; (ii) a nucleic acid molecule encoding an α 1,2-mannosidase activity linked to a heterologous targeting or signaling peptide that targets the mannosidase activity to the secretory pathway; and (iii) a nucleic acid molecule encoding an *O*-linked mannanose β 1,2-*N*-acetylglucosaminyltransferase 1 (POMGnT1); (b) culturing the host cell under conditions suitable for producing the TNFRII-Fc; and (c) recovering the TNFRII-Fc from the culture fluid to produce the TNFRII-Fc having sialylated *N*-glycans and *O*-glycans.

In further aspects, the POMGnT1 is provided as a fusion protein comprising the catalytic domain of the POMGnT1 fused to a heterologous targeting or signaling peptide that targets the POMGnT1 to the secretory pathway, e.g., the ER or Golgi apparatus. Examples of heterologous targeting or signaling peptides include but are not limited to the *MNN2*, *MNN5* and *MNN6* targeting or signaling peptides.

In further aspects of the method, the *N*-glycans and *O*-glycans on the TNFRII-Fc are predominantly sialylated with α 2,6 or α 2,3 sialic acid residues. In further still aspects, the *N*-glycans on the TNFRII-Fc lack fucose residues. In further still aspects of the method, the *N*-glycans and *O*-glycans on the TNFRII-Fc, which are sialylated, comprise *N*-acetylneuraminic acid (NANA) and no *N*-glycolylneuraminic acid (NGNA).

In further still aspects of the method, a ratio of mole sialic acid to the mole of the TNFRII-Fc is at least 10. In further still aspects, a ratio of mole sialic acid to mole of the TNFRII-Fc is about 10 to 21. In further still aspects of the method, a ratio of mole sialic acid to mole of the TNFRII-Fc is greater than 21.

In further aspects of the method, at least 50%, 60%, 70%, 80%, 90%, or 100% of the *N*-glycans are sialylated. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly mono-, bi-, tri-, or tetra-sialylated *N*-glycans. In further still aspects of the method, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly mono-sialylated *N*-glycans. In further still aspects, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly bi-sialylated *N*-glycans. In further still aspects of the method, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly tri-sialylated *N*-glycans. In further still aspects of the method, the *N*-glycans on the TNFRII-Fc comprise or consist of predominantly tetra-sialylated *N*-glycans.

In further still aspects of the method, the *O*-glycans on the TNFRII-Fc comprise or consist of predominantly sialylated *O*-glycans. In further still aspects, greater than 10%, 20%, 30%, 40%, or 50% of the *O*-glycans on the TNFRII-Fc comprise or consist of sialylated *O*-glycans. In further still aspects of the method, less than 10%, 20%, 40% or 50% of the *O*-glycans on the TNFRII-Fc terminate in mannose.

In further still aspects of the method, the TNFRII domain of the TNFRII-Fc comprises or consists of an amino acid sequence with at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the amino acid sequence for the TNFRII domain set forth in SEQ ID NO:73 or 75. The receptor domain includes amino acids 1 to 235 of SEQ ID NO:73 or 75 and is encoded by nucleotides 1-705 of SEQ ID NO:72 or 74.

Further provided is a method for producing a recombinant human tumor necrosis factor fused to the constant region of an antibody (TNFRII-Fc) having sialylated *N*-glycans and *O*-mannose reduced glycans comprising or consisting of (a) providing a recombinant lower eukaryote host cell genetically engineered to produce glycoproteins having sialylated *N*-glycans and further comprising (i) a nucleic acid molecule encoding the TNFRII-Fc; and (ii) a nucleic acid molecule encoding an α 1,2-mannosidase activity linked to a heterologous targeting or signaling peptide that targets the mannosidase activity to the secretory pathway; (b) culturing the host cell under conditions suitable for producing the TNFRII-Fc; and (c) recovering the TNFRII-Fc from the culture fluid to produce the TNFRII-Fc having sialylated *N*-glycans and *O*-mannose reduced glycans.

In further aspects of the method, the TNFRII domain of the TNFRII-Fc comprises or consists of an amino acid sequence with at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the amino acid sequence for the TNFRII domain set forth in SEQ ID NO:73 or 75. The receptor domain includes amino acids 1 to 235 of SEQ ID NO:73 or 75 and is encoded by nucleotides 1-705 of SEQ ID NO:72 or 74.

In further aspects, the host cells are cultured in the presence of a *PMT* inhibitor which reduces the number of sites on the TNFRII-Fc that is *O*-glycosylated.

Host Cells

Useful lower eukaryote host cells for producing the TNFRII-Fc molecules disclosed herein are glycoengineered host cells that include but are not limited to *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* and *Neurospora crassa*. Various yeasts, such as *K. lactis*, *Pichia pastoris*, *Pichia methanolica*, and *Hansenula polymorpha* are particularly suitable for cell culture because they are able to grow to high cell densities and secrete large quantities of recombinant protein. Likewise, filamentous fungi, such as *Aspergillus niger*, *Fusarium sp.*, *Neurospora crassa* and others can be used to produce glycoproteins of the invention at an industrial scale. In the case of lower eukaryotes, cells are routinely grown from between about one and a half to three days.

The *Pichia pastoris* strains **YGLY11731**, **YGLY10299**, **YGLY13571**, **YGLY12680**, and **YGLY14252** shown in **Figures 1A-G, 2A-B, and 3** and their construction are described in Examples 1-3. Example 4 describes the construction of strains **YGLY14954** and **YGLY14927**, shown in **Figure 4**. These strains are similar to strain **YGLY14252** except that the chimeric POMGnT is fused to a different heterologous targeting or signaling peptide and it is inserted into a different locus in the *Pichia pastoris* genome. The methods for constructing the strains in Examples 1-4 can be used to construct other lower eukaryote host cells that express TNFRII-Fc fragment fusion protein with characteristics similar to the TNFRII-Fc fragment fusion protein described in Examples 1-4. In general, these lower eukaryote host cells can be achieved by eliminating selected endogenous glycosylation enzymes and/or supplying exogenous enzymes as described by Gerngross *et al.*, U.S. Patent No. 7,449,308, the disclosure of which is incorporated herein by reference. In particular aspects of the invention, the host cell is yeast, which in further aspects, a methylotrophic yeast such as *Pichia pastoris* or *Ogataea minuta* and mutants thereof. In general, the TNFRII-Fc fragment fusion protein produced in a lower eukaryote other than *Pichia pastoris* as exemplified in the examples or using variants or species of the enzymes and heterologous targeting or signaling peptides exemplified in the examples are expected to produce a TNFRII-Fc fragment fusion protein with general characteristics similar or the same as that for TNFRII-Fc fragment fusion protein produced as described in the examples. These general characteristics are that the O-glycans are of the dystroglycan type, the N-glycans are afucosylated, the N-glycans and O-glycans possess only NANA residues and no NGNA

residues, and provided the sialyltransferase is an $\alpha 2,6$ sialyltransferase, the sialic acid residues will be linked via an $\alpha 2,6$ linkage.

A general scheme for constructing a host cell that can produce the TNFRII-Fc fragment fusion protein disclosed herein can include the following. The host cell is selected that lacks in initiating 1,6-mannosyl transferase activity. Such host cells either naturally lack an endogenous initiating 1,6-mannosyl transferase activity or are genetically engineered to lack the initiating 1,6-mannosyl transferase activity. Then, the host cell further includes an $\alpha 1,2$ -mannosidase catalytic domain fused to a heterologous targeting or signal peptide not normally associated with the catalytic domain and selected to target the $\alpha 1,2$ -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 Man₅GlcNAc₂ glycoform, for example, a recombinant glycoprotein composition comprising predominantly a Man₅GlcNAc₂ glycoform. U.S. Patent No. 7,029,872, U.S. Patent No. 7,449,308, and U.S. Published Patent Application No. 2005/0170452, the disclosures of which are all incorporated herein by reference, disclose lower eukaryote host cells capable of producing a glycoprotein comprising a Man₅GlcNAc₂ glycoform.

The immediately preceding host cell further includes an *N*-acetylglucosaminyltransferase I (GlcNAc transferase I or GnT I) catalytic domain fused to a heterologous targeting or 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 GlcNAcMan₅GlcNAc₂ glycoform, for example a recombinant glycoprotein composition comprising predominantly a GlcNAcMan₅GlcNAc₂ glycoform. U.S. Patent No. 7,029,872, U.S. Patent No. 7,449,308, and U.S. Published Patent Application No. 2005/0170452, the disclosures of which are all incorporated herein by reference, disclose lower eukaryote host cells capable of producing a glycoprotein comprising a GlcNAcMan₅GlcNAc₂ glycoform.

The immediately preceding host cell further includes a mannosidase II catalytic domain fused to a heterologous targeting or 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 GlcNAcMan₃GlcNAc₂ glycoform, for example a recombinant glycoprotein composition comprising predominantly a GlcNAcMan₃GlcNAc₂ glycoform. U.S. Patent No. 7,029,872 and U.S. Patent No. 7,625,756, the disclosures of which are all incorporated herein by reference, disclose lower eukaryote host cells that express mannosidase II enzymes and are capable of producing glycoproteins having predominantly a GlcNAcMan₃GlcNAc₂ glycoform.

The immediately preceding host cell further includes *N*-acetylglucosaminyltransferase II (GlcNAc transferase II or GnT II) catalytic domain fused to a heterologous targeting or 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.

5 Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a GlcNAc₂Man₃GlcNAc₂ glycoform, for example a recombinant glycoprotein composition comprising predominantly a GlcNAc₂Man₃GlcNAc₂ glycoform. U.S. Patent Nos. 7,029,872 and 7,449,308 and U.S. Published Patent Application No. 2005/0170452, the disclosures of which are all incorporated
10 herein by reference, disclose lower eukaryote host cells capable of producing a glycoprotein comprising a GlcNAc₂Man₃GlcNAc₂ glycoform.

The immediately preceding host cell further includes a galactosyltransferase catalytic domain fused to a heterologous targeting or signal peptide not normally associated with the catalytic domain and selected to target galactosyltransferase activity to the ER or Golgi
15 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 GalGlcNAc₂Man₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ glycoform, or mixture thereof for example a recombinant glycoprotein composition comprising predominantly a GalGlcNAc₂Man₃GlcNAc₂ glycoform or Gal₂GlcNAc₂Man₃GlcNAc₂ glycoform or mixture
20 thereof. U.S. Patent No. 7,029,872 and U.S. Published Patent Application No. 2006/0040353, the disclosures of which are incorporated herein by reference, discloses lower eukaryote host cells capable of producing a glycoprotein comprising a Gal₂GlcNAc₂Man₃GlcNAc₂ glycoform.

The immediately preceding host cell further includes a sialyltransferase catalytic domain fused to a heterologous targeting or signal peptide not normally associated with the
25 catalytic domain and selected to target sialyltransferase activity to the ER or Golgi apparatus of the host cell. The sialyltransferase can be an α 2,6-sialyltransferase or an α 2,3-sialyltransferase. The type of sialyltransferase species will determine whether the sialic acid residue is attached in an α 2,6 linkage or an α 2,3 linkage. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising predominantly
30 a NANA₂Gal₂GlcNAc₂Man₃GlcNAc₂ glycoform or NANAGal₂GlcNAc₂Man₃GlcNAc₂ glycoform or mixture thereof. For lower eukaryote host cells such as yeast and filamentous fungi, the host cell further includes a means for providing CMP-sialic acid for transfer to the *N*-glycan. U.S. Published Patent Application No. 2005/0260729, the disclosure of which is incorporated herein by reference, discloses a method for genetically engineering lower eukaryotes
35 to have a CMP-sialic acid synthesis pathway and U.S. Published Patent Application No. 2006/0286637, the disclosure of which is incorporated herein by reference, discloses a method for genetically engineering lower eukaryotes to produce sialylated glycoproteins. To enhance the

amount of sialylation of the *N*-glycans and *O*-glycans, it can be advantageous to construct the host cell to include two or more copies of the CMP-sialic acid pathway and two or more copies of the sialyltransferase.

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. Patent No. 7,598,055 and U.S. Published Patent Application No. 2007/0037248, the disclosures of which are all incorporated herein by reference.

The above host cells are further genetically engineered to express a nucleic acid molecule encoding a protein *O*-mannose β -1,2-*N*-acetylglucosaminyltransferase I (POMGnT I) activity. In general, the POMGnT I catalytic domain is fused not normally associated with the catalytic domain and selected to target the fusion protein to a location in the ER or Golgi where it can then transfer a GlcNAc residue to *O*-linked mannose residues on the TNFRII-Fc fragment fusion protein as it traverses the secretory pathway. The human POMGnT and its expression in yeast have been disclosed in U.S. Patent No. 7,217,548.

The host cells are also genetically modified to control the chain length of the *O*-glycans on the TNFRII-Fc fragment fusion protein so as to provide single-mannose *O*-glycans. The single-mannose *O*-glycans serve as a substrate for the POMGnT I to transfer a GlcNAc residue thereto. Control can be accomplished by growing the cells in the presence of Pmtp inhibitors that inhibit *O*-mannosyltransferase (PMT) protein activity or an alpha-mannosidase as disclosed in U.S. Published Application No. 20090170159, the disclosure of which is incorporated herein by reference), or both. Thus, in one aspect, controlling *O*-glycosylation includes expressing one or more secreted α 1,2-mannosidase enzymes in the host cell to produce the recombinant protein having reduced *O*-linked glycosylation, also referred to herein as *O*-mannose reduced glycans. In particular embodiments, the α 1,2-mannosidase, which is capable of trimming multiple mannose residues from an *O*-linked glycan is produced by *Trichoderma* sp., *Saccharomyces* sp., or *Aspergillus* sp., *Coccidioides immitis*, *Coccidioides posadasii*, *Penicillium citrinum*, *Magnaporthe grisea*, *Aspergillus saitoi*, *Aspergillus oryzae*, or *Chaetomium globosum*. For example, α 1,2-mannosidases can be obtained from *Trichoderma reesei*, *Aspergillus niger*, or *Aspergillus oryzae*. *T. reesei* is also known as *Hypocrea jecorina*. As shown in the examples, a transformed yeast comprising an expression cassette, which expresses the *Trichoderma reesei* α 1,2-mannosidase catalytic domain fused to the *Saccharomyces cerevisiae* α MAT pre signal sequence, was used to produce the TNFRII-Fc fragment fusion protein in which the *O*-glycans are trimmed to a single mannose residue, which can serve as a substrate for POMGnT1.

The Pmtp inhibitor reduces *O*-glycosylation occupancy (lowers the number of serines and threonine residues with *O*-mannose glycans on the TNFRII-Fc fragment fusion protein) from about 80 *O*-glycans to about 20 *O*-glycans per protein molecule. In the presence of

the Pmtp inhibitor, the overall level of *O*-linked glycans on the TNFRII-Fc fragment fusion protein is significantly lowered. Thus, the Pmtp inhibitor and the secreted α 1,2-mannosidase results in a higher percentage of the *O*-glycans on the TNFRII-Fc fragment fusion protein being the desired sialylated *O*-glycan instead of the less desired *O*-linked mannobiose, mannotriose, and mannotetrose *O*-glycan structures or asialylated *O*-Man-GlcNAc or *O*-Man-GlcNAc-Gal. Thus, the control of *O*-glycosylation enables the overall levels of sialylated *O*-glycans to be increased while also reducing the level of asialylated or neutral charge *O*-glycans.

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.

Pichia pastoris host cells further include strains that have been genetically engineered to eliminate glycoproteins having phosphomannose residues. This can be achieved by deleting or disrupting one or both of the phosphomannosyltransferase genes *PNO1* and *MNN4B* (or *MNN4L1*) (See for example, U.S. Patent Nos. 7,198,921 and 7,259,007; the disclosures of which are all incorporated herein by reference), which in further aspects can also include deleting or disrupting the *MNN4A* (or *MNN4*) 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.

To reduce or eliminate the likelihood of *N*-glycans and *O*-glycans with β -linked mannose residues, which are resistant to α -mannosidases, the recombinant glycoengineered *Pichia pastoris* host cells are genetically engineered to eliminate glycoproteins having α -mannosidase-resistant *N*-glycans by deleting or disrupting one or more of the β -mannosyltransferase genes (e.g., *BMT1*, *BMT2*, *BMT3*, and *BMT4*) (See, U.S. Patent No. 7,465,577 and U.S. Patent No. 7,713,719). The deletion or disruption of *BMT2* and one or more of *BMT1*, *BMT3*, and *BMT4* also reduces or eliminates detectable cross reactivity to antibodies against host cell protein.

To reduce the risk of *N*-terminal clipping in *Pichia pastoris* host cells (LP diaminopeptidase activity), expression of the *STE13* and *DAP2* genes encoding the Ste13p and Dap2p proteases. Identification and deletion of the *STE13* or *DAP2* genes in *Pichia pastoris* has been described in Published PCT Application No. WO2007148345 and in Pabha *et al.*, Protein Express. Purif. 64: 155-161 (2009).

Proteins that are destined for the vacuole are sorted from proteins destined for the cell surface in the late Golgi compartment. The sorting process is similar to the mammalian lysosomal sorting system; however, unlike the mammalian lysosomal sorting system where the sorting signal is a carbohydrate moiety, in yeast the sorting signal is contained within the polypeptide chains themselves. The most thoroughly studied vacuolar protein in *S. cerevisiae* is carboxypeptidase Y (CPY encoded by *PRCI*), which has a sorting signal at the *N*-terminus of its prosegment that is QRPL. This sorting signal sequence is recognized by the CPY sorting receptor Vps10p/Pep1p, which binds and directs the CPY to the vacuole. Mutational analysis of the sorting signal sequence by Van Voors *et al.*, J. Biol. Chem. 271: 841-846 (1996) suggests that there may be cryptic sorting signals that if present in a recombinant protein such as TNFRII-Fc fragment fusion protein might direct the protein to the vacuole where it is degraded. To avoid potential sorting of the TNFRII-Fc fragment fusion protein to the vacuole, the *Pichia pastoris* host strain can further include a disruption or deletion of the expression of the *VPS10-1* gene. The *VPS10-1* gene in *Pichia pastoris* was identified and the gene deleted in the above glycoengineered *Pichia pastoris* to produce a *Pichia pastoris* strain that lacked CPY sorting mediated by the Vps10-1p.

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 Published International Application No. WO 2009105357 and WO2010019487 (the disclosures of which are incorporated herein by reference).

The host cell can be further genetically engineered to include a nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase but wherein the endogenous host cell genes encoding the proteins comprising the oligosaccharyltransferase (OTase) complex are expressed. This includes expression of the endogenous *STT3* gene, which in yeast is the *STT3* gene. In general, in the above methods and host cells, the single-subunit oligosaccharyltransferase is capable of functionally suppressing the lethal phenotype of a

mutation of at least one essential protein of the OTase complex. In further aspects, the essential protein of the OTase complex is encoded by the *STT3* locus, *WBP1* locus, *OST1* locus, *SWP1* locus, or *OST2* locus, or homologue thereof. In further aspects, the for example single-subunit oligosaccharyltransferase is the *Leishmania major* STT3D protein.

Promoters are DNA sequence elements for controlling gene expression. In particular, promoters specify transcription initiation sites and can include a TATA box and upstream promoter elements. The promoters selected are those which would be expected to be operable in the particular host system selected. For example, yeast promoters are used when a yeast such as *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Ogataea minuta*, or *Pichia pastoris* is the host cell whereas fungal promoters would be used in host cells such as *Aspergillus niger*, *Neurospora crassa*, or *Trichoderma reesei*. Examples of yeast promoters include but are not limited to the *GAPDH*, *AOX1*, *SEC4*, *HHI*, *PMA1*, *OCH1*, *GAL1*, *PGK*, *GAP*, *TPI*, *CYC1*, *ADH2*, *PHO5*, *CUP1*, *MFa1*, *FLD1*, *PMA1*, *PDI*, *TEF*, *RPL10*, and *GUT1* promoters. Romanos *et al.*, Yeast 8: 423-488 (1992) provide a review of yeast promoters and expression vectors. Hartner *et al.*, Nucl. Acid Res. 36: e76 (pub on-line 6 June 2008) describes a library of promoters for fine-tuned expression of heterologous proteins in *Pichia pastoris*.

The promoters that are operably linked to the nucleic acid molecules disclosed herein can be constitutive promoters or inducible promoters. An inducible promoter, for example the *AOX1* promoter, is a promoter that directs transcription at an increased or decreased rate upon binding of a transcription factor in response to an inducer. Transcription factors as used herein include any factor that can bind to a regulatory or control region of a promoter and thereby affect transcription. The RNA synthesis or the promoter binding ability of a transcription factor within the host cell can be controlled by exposing the host to an inducer or removing an inducer from the host cell medium. Accordingly, to regulate expression of an inducible promoter, an inducer is added or removed from the growth medium of the host cell. Such inducers can include sugars, phosphate, alcohol, metal ions, hormones, heat, cold and the like. For example, commonly used inducers in yeast are glucose, galactose, alcohol, and the like.

Transcription termination sequences that are selected are those that are operable in the particular host cell selected. For example, yeast transcription termination sequences are used in expression vectors when a yeast host cell such as *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, or *Pichia pastoris* is the host cell whereas fungal transcription termination sequences would be used in host cells such as *Aspergillus niger*, *Neurospora crassa*, or *Trichoderma reesei*. Transcription termination sequences include but are not limited to the *Saccharomyces cerevisiae* *CYC* transcription termination sequence (*ScCYC* TT), the *Pichia pastoris* *ALG3* transcription termination sequence (*ALG3* TT), the *Pichia pastoris* *ALG6* transcription termination sequence (*ALG6* TT), the *Pichia pastoris* *ALG12* transcription termination sequence (*ALG12* TT), the *Pichia pastoris* *AOX1* transcription termination sequence (*AOX1* TT), the *Pichia pastoris* *OCH1*

transcription termination sequence (*OCHI* TT) and *Pichia pastoris* *PMA1* transcription termination sequence (*PMA1* TT). Other transcription termination sequences can be found in the examples and in the art.

For genetically engineering yeast, selectable markers 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, nourseothricin, hygromycin, 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 (the disclosure of which is incorporated herein by reference) 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; the disclosures of which are all incorporated herein by reference). 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 disclosure of which is incorporated herein by reference), 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.

Therapeutic Administration of the TNFRII-Fc fragment fusion protein

The present invention provides methods of suppressing TNF-dependent inflammatory responses in humans comprising administering an effective amount of a composition comprising the TNFRII-Fc fragment fusion protein disclosed herein and a suitable diluent and carrier, for example, a pharmaceutical composition comprising a TNFRII-Fc fragment fusion protein in a pharmaceutically acceptable carrier.

For therapeutic use, a composition comprising the TNFRII-Fc fragment fusion protein is administered to a patient, preferably a human, for treatment of arthritis. Thus, for example, TNFRII-Fc fragment fusion protein compositions can be administered, for example, via

intra-articular, intraperitoneal or subcutaneous routes by bolus injection, continuous infusion, sustained release from implants, or other suitable techniques. Typically, a composition comprising the TNFRII-Fc fragment fusion protein will be administered in the form of a composition comprising purified protein in conjunction with physiologically acceptable carriers, excipients or diluents. Such carriers will be nontoxic to recipients at the dosages and concentrations employed. Ordinarily, the preparation of such compositions entails combining the TNFRII-Fc fragment fusion protein with buffers, antioxidants such as ascorbic acid, low molecular weight (less than about 10 residues) polypeptides, proteins, amino acids, carbohydrates including glucose, sucrose or dextrans, chelating agents such as EDTA, glutathione and other stabilizers and excipients. Neutral buffered saline or saline mixed with conspecific serum albumin are exemplary appropriate diluents. Preferably, product is formulated as a lyophilizate using appropriate excipient solutions (e.g., sucrose) as diluents. Appropriate dosages can be determined in trials. In accordance with appropriate industry standards, preservatives may also be added, such as benzyl alcohol. The amount and frequency of administration will depend, of course, on such factors as the nature and severity of the indication being treated, the desired response, the condition of the patient, and so forth.

TNFRII-Fc fragment fusion protein compositions are administered to a mammal, preferably a human, for the purpose treating TNF-dependent inflammatory diseases, such as arthritis. For example, the TNFRII-Fc fragment fusion protein inhibits TNF-dependent arthritic responses. Because of the primary roles IL-1 and IL-2 play in the production of TNF, combination therapy using TNFR in combination with IL-1R and/or IL-2R may be used in the treatment of TNF-associated clinical indications. In the treatment of humans, the TNFRII-Fc fragment fusion proteins disclosed herein are preferred. Either Type I IL-1R or Type II IL-1R, or a combination thereof, may be used in accordance with the present invention to treat TNF-dependent inflammatory diseases, such as arthritis. Other types of TNF binding proteins may be similarly used.

For treatment of arthritis, the TNFRII-Fc fragment fusion protein composition is administered in systemic amounts ranging from about 0.1 mg/kg/week to about 100 mg/kg/week. In further aspects, the TNFRII-Fc fragment fusion protein is administered in amounts ranging from about 0.5 mg/kg/week to about 50 mg/kg/week. For local intra-articular administration, dosages preferably range from about 0.01 mg/kg to about 1.0 mg/kg per injection.

Pharmaceutical compositions

The TNFRII-Fc fragment fusion proteins disclosed herein may be provided as a pharmaceutical composition when combined with a pharmaceutically acceptable carrier. Such compositions comprise a therapeutically-effective amount of the TNFRII-Fc fragment fusion protein and a pharmaceutically acceptable carrier. Such a composition may also be comprised of

(in addition to TNFRII-Fc fragment fusion protein and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art and generally regarded as safe by pharmaceutical and biological regulatory agencies. Compositions comprising the TNFRII-Fc fragment fusion protein can be administered, if desired, in the form of salts provided the salts are pharmaceutically acceptable. Salts may be prepared using standard procedures known to those skilled in the art of synthetic organic chemistry.

The term "pharmaceutically acceptable salts" refers to salts prepared from pharmaceutically acceptable non-toxic bases or acids including inorganic or organic bases and inorganic or organic acids. Salts derived from inorganic bases include aluminum, ammonium, calcium, copper, ferric, ferrous, lithium, magnesium, manganic salts, manganous, potassium, sodium, zinc, and the like. Particularly preferred are the ammonium, calcium, magnesium, potassium, and sodium salts. Salts derived from pharmaceutically acceptable organic non-toxic bases include salts of primary, secondary, and tertiary amines, substituted amines including naturally occurring substituted amines, cyclic amines, and basic ion exchange resins, such as arginine, betaine, caffeine, choline, N,N'-dibenzylethylenediamine, diethylamine, 2-diethylaminoethanol, 2-dimethylaminoethanol, ethanolamine, ethylenediamine, N-ethylmorpholine, N-ethylpiperidine, glucamine, glucosamine, histidine, hydrabamine, isopropylamine, lysine, methylglucamine, morpholine, piperazine, piperidine, polyamine resins, procaine, purines, theobromine, triethylamine, trimethylamine, tripropylamine, tromethamine, and the like. The term "pharmaceutically acceptable salt" further includes all acceptable salts such as acetate, lactobionate, benzenesulfonate, laurate, benzoate, malate, bicarbonate, maleate, bisulfate, mandelate, bitartrate, mesylate, borate, methylbromide, bromide, methylnitrate, calcium edetate, methylsulfate, camsylate, mucate, carbonate, napsylate, chloride, nitrate, clavulanate, N-methylglucamine, citrate, ammonium salt, dihydrochloride, oleate, edetate, oxalate, edisylate, pamoate (embonate), estolate, palmitate, esylate, pantothenate, fumarate, phosphate/diphosphate, gluceptate, polygalacturonate, gluconate, salicylate, glutamate, stearate, glycolylarsanilate, sulfate, hexylresorcinate, subacetate, hydrabamine, succinate, hydrobromide, tannate, hydrochloride, tartrate, hydroxynaphthoate, teoclate, iodide, tosylate, isothionate, triethiodide, lactate, pantoate, valerate, and the like which can be used as a dosage form for modifying the solubility or hydrolysis characteristics or can be used in sustained release or pro-drug formulations. It will be understood that, as used herein, references to the TNFRII-Fc fragment fusion protein disclosed herein are meant to also include the pharmaceutically acceptable salts.

As utilized herein, the term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s), approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopoeia or other generally recognized pharmacopoeia for use in animals and, more particularly, in humans. The term "carrier" refers to a diluent, adjuvant, excipient, or

vehicle with which the therapeutic is administered and includes, but is not limited to such sterile liquids as water and oils. The characteristics of the carrier will depend on the route of administration. The TNFRII-Fc fragment fusion protein disclosed herein may be in multimers (for example, heterodimers or homodimers) or complexes with itself or other peptides. As a result, pharmaceutical compositions of the invention may comprise one or more TNFRII-Fc fragment fusion protein molecules disclosed herein in such multimeric or complexed form.

As used herein, the term "therapeutically effective amount" means the total amount of each active component of the pharmaceutical composition or method that is sufficient to show a meaningful patient benefit, i.e., treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, the term refers to that ingredient alone. When applied to a combination, the term refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially, or simultaneously.

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

EXAMPLE 1

This example shows the construction of *Pichia pastoris* strains **YGLY10299**, **YGLY11731**, and **YGLY13571**, each strain a GS6.0 strain capable of producing TNFRII-Fc fragment fusion protein comprising sialylated *N*-glycans. **Figures 1A-G** provide a flow-diagram illustrating construction of the strains.

All yeast transformations were as follows. *P. pastoris* strains were grown in 50 mL YPD media (yeast extract (1%), peptone (2%), dextrose (2%)) overnight to an optical density ("OD") of between about 0.2 to 6. After incubation on ice for 30 minutes, cells were pelleted by centrifugation at 2500-3000 rpm for 5 minutes. Media was removed and the cells washed three times with ice cold sterile 1M sorbitol before resuspension in 0.5 ml ice cold sterile 1M sorbitol. Ten μ L DNA (5-20 μ g) and 100 μ L cell suspension was combined in an electroporation cuvette and incubated for 5 minutes on ice. Electroporation was in a Bio-Rad GenePulser Xcell following the preset *Pichia pastoris* protocol (2 kV, 25 μ F, 200 Ω), immediately followed by the addition of 1 mL YPDS recovery media (YPD media plus 1 M sorbitol). The transformed cells were allowed to recover for four hours to overnight at room temperature (26°C) before plating the cells on selective media.

The strain **YGLY9469** 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 5**) is an integration vector that targets the *URA5* locus. It contains a nucleic acid molecule comprising the *S. cerevisiae* invertase gene or transcription unit (*ScSUC2*; SEQ ID NO:17) 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:18) and on the other side by a nucleic acid molecule comprising the nucleotide sequence from the 3' region of the *P. pastoris URA5* gene (SEQ ID NO:19). 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 6**) is an integration vector that targets the *OCHI* locus and contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (SEQ ID NO:20) flanked by nucleic acid molecules comprising *lacZ* repeats (SEQ ID NO:21) which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *OCHI* gene (SEQ ID NO:22) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *OCHI* gene (SEQ ID NO:23). Plasmid pGLY40 was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY1-3** to produce a number of strains in which the *URA5* gene flanked by the *lacZ* repeats has been inserted into the *OCHI* 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 *OCHI* locus. This renders the strain auxotrophic for uracil. Strain **YGLY4-3** was selected.

Plasmid pGLY43a (**Figure 7**) 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 (*KIMNN2-2*, SEQ ID NO:24) 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: 25) 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:26). Plasmid pGLY43a was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY4-3** to produce a number of strains in which the *KIMNN2-2* 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 8**) 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:27) open reading frame (ORF) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *GAPDH* promoter (SEQ ID NO:5) and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* termination sequences (SEQ ID NO:3) 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:28) 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:29). 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 9**) is an integration vector that targets the *PNO1/MNN4* loci 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 *PNO1* gene (SEQ ID NO:30) 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:31). Plasmid pGLY45 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY12-3** to produce a number of strains in which the *URA5* gene flanked by the *lacZ* repeats has been inserted into the *PNO1/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 *MNN4A*) 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 pGLY1430 (**Figure 10**) 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:32) fused at the 5' end to a nucleic acid molecule encoding the *SEC12* leader 10 (SEQ ID NO:33), 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:34) and at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *OCH1* termination sequences (SEQ ID NO:35). The expression cassette encoding the FB8 comprises a nucleic acid molecule encoding the mouse mannosidase IA catalytic domain (SEQ ID NO:36) fused at the 5' end to a nucleic acid molecule encoding the *SEC12-m* leader 8 (SEQ ID NO:37), 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:38) followed by a *P. pastoris* *ALG3* termination sequence (SEQ ID NO:8) 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:39). Plasmid pGLY1430 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY16-3** 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 **YGLY2798** 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 **YGLY3794** was selected and is capable of making glycoproteins that have predominantly GlcNAcMan₅GlcNAc₂ terminated *N*-glycans.

Plasmid pGLY582 (**Figure 11**) 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:40) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* promoter (SEQ ID NO:1) and operably linked at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* transcription termination sequence (SEQ ID NO:41). 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:42) fused at the 5' end to a nucleic acid molecule encoding the *KRE2-s* leader 33 (SEQ ID NO:43), 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:44) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *OCH1* promoter (SEQ ID NO:45) and operably linked at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *ALG12* transcription termination sequence (SEQ ID NO:46). 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:47) 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:48). Plasmid pGLY582 was linearized and the linearized plasmid transformed into strain **YGLY3794** to produce a number of strains in which the four tandem expression cassette have been inserted into the *HIS1* locus by homologous recombination. Strain **YGLY3853** was selected and is auxotrophic for histidine and prototrophic for uridine.

Plasmid pGLY167b (**Figure 12**) 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:49) fused at the 5' end to a nucleic acid molecule encoding the *MNN2* leader 53 (SEQ ID NO:50), 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:51). 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:52) fused at the 5' end to a nucleic acid molecule encoding the *MNN2* leader 54 (SEQ ID NO:53), 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:54) 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:55). Plasmid pGLY167b was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY3853** to produce a number of strains (in which the three tandem expression cassettes have been inserted into the *ARG1* locus by double-crossover homologous recombination. The strain **YGLY4754** 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 **YGLY4799** was selected.

Plasmid pGLY3411 (**Figure 13**) 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:56) and on the other side with the 3' nucleotide sequence of the *P. pastoris* *BMT4* gene (SEQ ID NO:57). Plasmid pGLY3411 was linearized and the linearized plasmid transformed into **YGLY4799** 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 **YGLY6903** was selected from the strains produced and is prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan. The strain was then counterselected in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strain **YGLY7432** was selected.

Plasmid pGLY3419 (**Figure 14**) is an integration vector that contains an 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 BMT1* gene (SEQ ID NO:58) and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT1* gene (SEQ ID NO:59).

5 Plasmid pGLY3419 was linearized and the linearized plasmid transformed into strain **YGLY7432** to produce a number of strains in which the *URA5* expression cassette has been inserted into the *BMT1* locus by double-crossover homologous recombination. The strain **YGLY7651** was selected from the strains produced and are prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan. The strains were then counterselected in the
10 presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strain **YGLY7930** was selected.

Plasmid pGLY3421 (**Figure 15**) is an integration vector that contains an 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 BMT3* gene (SEQ ID NO:60) and on the
15 other side with the 3' nucleotide sequence of the *P. pastoris BMT3* gene (SEQ ID NO:61). Plasmid pGLY3419 was linearized and the linearized plasmid transformed into strain **YGLY7930** to produce a number of strains in which the *URA5* expression cassette has been inserted into the *BMT1* locus by double-crossover homologous recombination. The strain **YGLY7961** was selected from the strains produced and are prototrophic for uracil, adenine,
20 histidine, proline, arginine, and tryptophan.

Plasmid pGLY2456 (**Figure 16**) 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
25 transcription unit, (4) the human CMP-sialic acid synthase (hCSS), (5) the human *N*-acetylneuraminate-9-phosphate synthase (hSPS), (6) the mouse α -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
30 molecule encoding the mCMP Sia Transp ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:64), 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
35 encoding the hGNE ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:65), 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:66). The expression cassette encoding the human CMP-sialic acid synthase comprises a nucleic acid molecule encoding the hCSS ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:67), 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 human N-acetylneuraminase-9-phosphate synthase comprises a nucleic acid molecule encoding the hSIAP S ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:68), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* transcription termination sequence. The expression cassette encoding the chimeric mouse α -2,6-sialyltransferase comprises a nucleic acid molecule encoding the mST6 catalytic domain codon optimized for expression in *P. pastoris* (SEQ ID NO:69) 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:6) and at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *TEF* transcription termination sequence (SEQ ID NO:7). The six tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and ORF of the *TRP2* gene ending at the stop codon (SEQ ID NO:62) 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:63). Plasmid pGLY2456 was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY7961** 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 **YGLY8146** was selected from the strains produced. The strain was then counterscreened in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strain **YGLY9296** was selected.

Plasmid pGLY5048 (**Figure 17**) is an integration vector that targets the *STE13* locus and contains expression cassettes encoding (1) the *T. reesei* α -1,2-mannosidase catalytic domain fused at the *N*-terminus to *S. cerevisiae* α MATpre signal peptide (α MATTrMan) 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 α MATTrMan comprises a nucleic acid molecule encoding the *T. reesei* catalytic domain (SEQ ID NO:81) fused at the 5' end to a nucleic acid molecule encoding the *S. cerevisiae* α MATpre signal peptide (SEQ ID NO:80), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *AOX1* 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 of the *STE13* gene (SEQ ID NO:82) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *STE13* gene (SEQ ID NO:83). Plasmid pGLY5048 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY9296** to produce a number of strains. The strain **YGLY9469** was selected from the strains produced. This strain is capable of producing glycoproteins that have single-mannose *O*-glycosylation (See Published U.S. Application No. 20090170159).

Plasmid pGLY5019 (**Figure 18**) is an integration vector that targets the *DAP2* locus and contains an expression cassette comprising a nucleic acid molecule encoding the Nourseothricin resistance (NAT^R) expression cassette (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)). The NAT^R expression cassette (SEQ ID NO:13) is operably regulated to the *Ashbya gossypii* *TEF1* promoter and *A. gossypii* *TEF1* termination sequences flanked one side with the 5' nucleotide sequence of the *P. pastoris* *DAP2* gene (SEQ ID NO:84) and on the other side with the 3' nucleotide sequence of the *P. pastoris* *DAP2* gene (SEQ ID NO:85). Plasmid pGLY5019 was linearized and the linearized plasmid transformed into strain **YGLY9469** to produce a number of strains in which the NAT^R expression cassette has been inserted into the *DAP2* locus by double-crossover homologous recombination. The strains **YGLY9795** and **YGLY9797** were selected from the strains produced.

Strain **YGLY9795** was transformed with plasmids pGLY5045 to produce strain **YGLY10296**, and strain **YGLY9797** was transformed with plasmid pGLY5045 or pGLY6391 to produce strains **YGLY10299** and **YGLY12626**, respectively. Each strain can produce a TNFR^{II}-Fc fragment fusion protein.

Plasmid pGLY5045 (**Figure 19**) is a roll-in integration vector that targets the *URA6* locus and contains an expression cassette encoding the TNFR^{II}-Fc fragment fusion protein. The plasmid contains two expression cassettes, each comprising a nucleic acid molecule codon-optimized for expression in *P. pastoris* encoding the TNFR^{II}-Fc fragment fusion protein (SEQ ID NO:74; encoding SEQ ID NO:75) fused at the 5' end to a nucleic acid molecule encoding the human serum albumin signal peptide (SEQ ID NO:70; encoding SEQ ID NO:71), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *AOX1* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* transcription termination sequence. The plasmid also includes a Zeocin^R expression cassette comprising a nucleic acid molecule encoding the Sh ble ORF (SEQ ID NO:14) operably linked at the 5' end to the *S. cerevisiae* *TEF1* promoter (SEQ ID NO:16) and at the 3' end to the *S. cerevisiae* *CYC* termination sequence. The *P. pastoris* *URA6* gene is shown in SEQ ID NO:12.

Plasmid pGLY5045 was transformed into strains **YGLY9795** and **YGLY9797** to produce a number of strains of which strains **YGLY10296** and **YGLY10299** were selected.

Plasmid pGLY6391 (**Figure 20**) is a roll-in integration vector that targets the *THR1* locus and contains an expression cassette encoding the TNFRII-Fc fragment fusion protein. The plasmid contains two expression cassettes, each comprising a nucleic acid molecule codon-optimized for expression in *P. pastoris* encoding the TNFRII-Fc fragment fusion protein without the C-terminal lysine residue (SEQ ID NO:72; encoding SEQ ID NO:73) fused at the 5' end to a nucleic acid molecule encoding the human serum albumin signal peptide, which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris AOX1* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* transcription termination sequence. The plasmid also includes a Zeocin^R expression cassette comprising a nucleic acid molecule encoding the Sh ble ORF operably linked at the 5' end to the *S. cerevisiae TEF1* promoter and at the 3' end to the *S. cerevisiae CYC* termination sequence. The *P. pastoris THR1* gene is shown in SEQ ID NO:86. Plasmid pGLY6391 was transformed into strain **YGLY9797** to produce a number of strains of which strain **YGLY12626** was selected.

Plasmid pGLY5085 (**Figure 21**) is a KINKO plasmid for introducing a second set of the genes involved in producing sialylated *N*-glycans into *P. pastoris*. The plasmid is similar to plasmid YGLY2456 except that the *P. pastoris ARG1* gene has been replaced with an expression cassette encoding hygromycin resistance (HygR) and the plasmid targets the *P. pastoris TRP5* locus. The HYG^R resistance cassette is SEQ ID NO:79. The HYG^R expression cassette (SEQ ID NO:79) is operably regulated to the *Ashbya gossypii TEF1* promoter and *A. gossypii TEF1* termination sequences (See Goldstein *et al.*, Yeast 15: 1541 (1999)). The six tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and ORF of the *TRP5* gene ending at the stop codon (SEQ ID NO:93) 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 *TRP5* gene (SEQ ID NO:94). Plasmid pGLY5085 was transformed into strain **YGLY10296** to produce a number of strains of which strain **YGLY11731** was selected. Plasmid pGLY5085 was also transformed into strain **YGLY12626** to produce a number of strains of which strain **YGLY13430** was selected. **YGLY13430** was then counterselected in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine of which strain **YGLY13571** was selected.

Thus, shown are the construction of *Pichia pastoris* strains **YGLY10299**, **YGLY11731**, and **YGLY13571**, each strain a GS6.0 strain capable of producing TNFRII-Fc fragment fusion protein comprising sialylated *N*-glycans.

EXAMPLE 2

This example shows the construction of *Pichia pastoris* strains **YGLY12680**, a GS6.0 strain capable of producing TNFRII-Fc fragment fusion protein with sialylated *N*-glycans and *O*-glycans. **Figures 2A-2B** provide a flow-diagram illustrating construction of the strain. Strain **YGLY10299** was transformed as follows to produce strain **YGLY12680**.

Plasmid pGLY5755 (**Figure 22**) is a KINKO integration plasmid that encodes a chimeric mouse POMGnT I and targets the *HIS3* locus in *P. pastoris*. The expression cassette encoding the chimeric mouse POMGnT I comprises a nucleic acid molecule encoding the catalytic domain of the mouse POMGnT I ORF codon-optimized for effective expression in *P. pastoris* (SEQ ID NO:76) ligated in-frame with a nucleic acid molecule encoding *S. cerevisiae* *MNN2-s* signal peptide (53: SEQ ID NO:50) operably linked at the 5' end to a nucleic acid molecule that has the inducible *P. pastoris* *AOX1* promoter sequence (SEQ ID NO:2) and at the 3' end to a nucleic acid molecule that has the *S. cerevisiae* *CYC* transcription termination sequence (SEQ ID NO:3). For selecting transformants, the plasmid comprises an expression cassette encoding the *S. cerevisiae* *ARR3* ORF in which the nucleic acid molecule encoding the ORF (SEQ ID NO:11) is operably linked at the 5' end to a nucleic acid molecule having the *P. pastoris* *RPL10* promoter sequence (SEQ ID NO:4) and at the 3' end to a nucleic acid molecule having the *S. cerevisiae* *CYC* transcription termination sequence (SEQ ID NO:3). The expression cassettes are in tandem and are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and ORF of the *HIS3* gene ending at the stop codon (SEQ ID NO:87) 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 *HIS3* gene (SEQ ID NO:88). Plasmid pGLY5755 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY10299** to produce a number of strains in which the expression cassettes have been inserted into the *HIS3* locus immediately following the *HIS3* ORF by double-crossover homologous recombination. The strain **YGLY11566** was selected from the strains produced.

Plasmid pGLY5086 (**Figure 23**) is a KINKO plasmid for introducing a second set of the genes involved in producing sialylated *N*-glycans into *P. pastoris*. The plasmid is similar to plasmid YGLY5086 except that the plasmid targets the *P. pastoris* *THR1* locus. The expression cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and ORF of the *THR1* gene ending at the stop codon (SEQ ID NO:89) 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 *THR1* gene (SEQ ID NO:90). Plasmid pGLY5086 was transformed into strain **YGLY11566** to produce a number of strains of which strain **YGLY12680** was selected.

EXAMPLE 3

This example shows the construction of *Pichia pastoris* strain **YGLY14252**, a GS6.0 strain capable of producing TNFRII-Fc fragment fusion protein with sialylated *N*-glycans and *O*-glycans. **Figure 3** provides a flow diagram illustrating construction of the strain. Strain **YGLY13571** was transformed as follows to produce strain **YGLY14252**.

Plasmid pGLY5219 (**Figure 24**) is an integration plasmid that encodes a chimeric mouse POMGnT I and targets the *VPS10-1* locus in *P. pastoris*. The expression cassette encoding the chimeric mouse POMGnT I comprises a nucleic acid molecule encoding the catalytic domain of the mouse POMGnT I ORF codon-optimized for effective expression in *P. pastoris* (SEQ ID NO:76) ligated in-frame with a nucleic acid molecule encoding *S. cerevisiae* Mnn6-s signal peptide (65: SEQ ID NO:77;) operably linked at the 5' end to a nucleic acid molecule that has the inducible *P. pastoris GAPDH* promoter sequence (SEQ ID NO:5) and at the 3' end to a nucleic acid molecule that has the *S. cerevisiae CYC* transcription termination sequence (SEQ ID NO:3). For selecting transformants, the plasmid comprises an expression cassette comprising the *URA5* gene flanked by *lacZ* repeats as described previously. The expression cassettes are in tandem and are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *VPS10-1* gene (SEQ ID NO:91) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *VPS10-1* gene (SEQ ID NO:92). Plasmid pGLY5219 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY13571** to produce a number of strains in which the expression cassettes have been inserted into the *VPS10-1* locus. The strain **YGLY14252** was selected from the strains produced.

EXAMPLE 4

This example shows the construction of *Pichia pastoris* strains **YGLY14954** and **YGLY14297**, each a GS6.0 strain capable of producing TNFRII-Fc fragment fusion protein with sialylated *N*-glycans and *O*-glycans. **Figure 4** provides a flow diagram illustrating construction of the strains. Strain **YGLY13571** was transformed as follows to produce strains **YGLY14954** and **YGLY14927**.

Plasmid pGLY5192 (**Figure 25**) is an integration plasmid that targets the *VPS10-1* locus. The plasmid comprises an expression cassette comprising the *URA5* gene flanked by *lacZ* repeats as described previously. The expression cassette is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *VPS10-1* gene (SEQ ID NO:91) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *VPS10-1* gene (SEQ ID NO:92). Plasmid pGLY5192 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY13571** to produce a number of

strains in which the expression cassette has been inserted into the *VPS10-1* locus. The strain **YGLY13663** was selected from the strains produced.

Plasmid pGLY7087 (**Figure 26**) is a KINKO integration plasmid that encodes a chimeric mouse POMGnT I and targets the *HIS3* locus in *P. pastoris*. The expression cassette encoding the chimeric mouse POMGnT I comprises a nucleic acid molecule encoding the catalytic domain of the mouse POMGnT I ORF codon-optimized for effective expression in *P. pastoris* (SEQ ID NO:76) ligated in-frame with a nucleic acid molecule encoding *S. cerevisiae* Mnn5-s signal peptide (56: SEQ ID NO:78) operably linked at the 5' end to a nucleic acid molecule that has the inducible *P. pastoris GAPDH* promoter sequence (SEQ ID NO:5) and at the 3' end to a nucleic acid molecule that has the *S. cerevisiae CYC* transcription termination sequence (SEQ ID NO:3). For selecting transformants, the plasmid comprises an expression cassette encoding the *S. cerevisiae ARR3* ORF in which the nucleic acid molecule encoding the ORF (SEQ ID NO:11) is operably linked at the 5' end to a nucleic acid molecule having the *P. pastoris RPL10* promoter sequence (SEQ ID NO:4) and at the 3' end to a nucleic acid molecule having the *S. cerevisiae CYC* transcription termination sequence (SEQ ID NO:3). The expression cassettes are in tandem and are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and ORF of the *HIS3* gene ending at the stop codon (SEQ ID NO:87) 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 *HIS3* gene (SEQ ID NO:88). Plasmid pGLY7087 was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY13663** to produce a number of strains in which the expression cassettes have been inserted into the *HIS3* locus immediately following the *HIS3* ORF by double-crossover homologous recombination. The strains **YGLY14954** and **YGLY14927** were selected from the strains produced.

EXAMPLE 5

Purification strategy for YGLY10299 (produces Form 1 TNFRII-Fc fragment fusion protein), YGLY11731 (Form 2 TNFRII-Fc fragment fusion protein), and YGLY12680 (Form 3 TNFRII-Fc fragment fusion protein) as shown in **Figure 30**.

Form 1 is TNFRII-Fc fragment fusion protein in which the extent of *O*-glycosylation is reduced and the length of the *O*-glycans is about one mannose residue. Form 2 is TNFRII-Fc fragment fusion protein in which the extent of *O*-glycosylation is reduced and the length of the *O*-glycans is about one mannose residue as for Form 1 but wherein the amount of sialylated *N*-glycans on the glycoprotein is enhanced. Form 3 is a TNFRII-Fc fragment fusion protein that is similar to Form 2 but further having sialylated *O*-glycans.

YGLY10299, YGLY11731, and YGLY12680 were grown as follows. The primary culture was prepared by inoculating two 2.8 L baffled Fernbach flasks containing 500

mL of BSGY media with a 2 mL Research Cell Bank of the relevant strain. After 48 hours of incubation, the cells were transferred to inoculate the fermentor. The fermentation batch media contained: 40 g glycerol (Sigma Aldrich, St.Louis, MO), 18.2 g sorbitol (Acros Organics, Geel, Belgium), 2.3 g mono-basic potassium phosphate, (Fisher Scientific, Fair Lawn, NJ) 11.9 g di-
5 basic potassium phosphate (EMD, Gibbstown, NJ), 10 g Yeast Extract (Sensient, Milwaukee, WI), 20 g Hy-Soy (Sheffield Bioscience, Norwich, NY), 13.4 g YNB (BD, Franklin Lakes, NJ), and 4×10^{-3} g biotin (Sigma-Aldrich, St.Louis, MO) per liter of medium.

Fermentations were conducted in 3L & 15L dished-bottom glass autoclavable and 40 L SIP bioreactors (1.5L, 8L & 16 L starting volume respectively) (Applikon, Foster City, CA).

10 The fermenters were run in a simple fed-batch mode with the following conditions: temperature of $24 \pm 1^\circ\text{C}$; pH of 6.5 ± 0.2 maintained by the addition of 30% NH_4OH ; airflow of approximately 0.7 ± 0.1 vvm; dissolved oxygen of 20% of saturation was maintained by cascading feedback control of the agitation rate (from 350 to 1200 rpm) followed by supplementation of pure oxygen to the sparged air stream up to 0.1 vvm. After the depletion of the initial charge of glycerol as
15 seen by a sharp increase in dissolved oxygen concentration, a 50% (w/w) glycerol solution containing PTM2 Salts and Biotin was fed at an exponential rate of 5.33 g/L/h increasing at 0.08 1/h for 8 hours to achieve a target cell density of 200 ± 20 g/L (wet cell weight). After a 30 minute Transition period, a 100% methanol solution containing PTM2 Salts and Biotin was initiated. The methanol was fed at an exponential feeding rate of 1.33 g/L/h increasing at 0.01
20 1/h for 36 hours. At the end of the fermentation, the supernatant was obtained by centrifugation at $13,000 \times g$ for 30 minutes and subsequently purified via affinity chromatography.

The purification of TNFRII-Fc fragment fusion protein obtained from the three strains as shown in **Figure 30** was as follows. The TNFRII-Fc fragment fusion protein was captured by affinity chromatography from the culture medium (supernatant medium) of *P. pastoris* using
25 MABSELECT from GE Healthcare (PolyA-agarose media; Cat. # 17-5199-03). The cell free supernatant medium was loaded on to MABSELECT column pre-equilibrated with 3 column volume of 20 mM Tris-HCl pH7.0. The column was washed with 2 column volumes of 20mM Tris-HCl pH 7.0 and 5 column volume of 20mM Tris-HCl , 1 M NaCl pH 7.0 to remove the host cell protein contaminants. The TNFRII-Fc fragment fusion protein was eluted with 7 column
30 volumes of 50mM sodium citrate pH 3.0. The eluted fusion protein was neutralized immediately with 1 M Tris-HCl pH 8.0.

Macro-prep Ceramic Hydroxyapatite type I $40 \mu\text{m}$ Chromatography (Bio-Rad Laboratories, Cat #157-0040) was used as the first intermediate purification step to remove aggregated forms of TNFRII-Fc fragment fusion protein. The Hydroxyapatite column was
35 equilibrated with 3 column volumes of 5 mM Sodium phosphate pH6.5 and the mabselect pool containing TNFRII-Fc fragment fusion protein that was buffer exchanged into the equilibration buffer was applied on to the column. After loading, the column was washed with 3 column

volumes of the equilibration buffer and elution was performed by developing a gradient over 20 column volumes ranging from 0 to 1000 mM sodium chloride. The TNFRII-Fc fragment fusion protein that elutes around 550 -650 mM sodium chloride was pooled together.

Hydrophobic Interaction Chromatography (HIC) step was employed as the second intermediate purification step to separate the scrambled or misfolded TNFRII-Fc fragment fusion protein. The Hydroxyapatite pool sample of TNFRII-Fc fragment fusion protein was adjusted to 1 M Ammonium sulfate concentration and loaded on to the Phenyl SEPHAROSE 6 FF (low sub) (GE Healthcare Cat# 17-0965-05) column that was pre-equilibrated with 20 mM Sodium phosphate, 1M Ammonium sulfate pH 7.0. After loading, the column was washed with 3 column volumes of the equilibration buffer and elution was performed by developing a gradient over 30 column volumes ranging from 1 M to 0 M ammonium sulfate in 20 mM sodium phosphate pH 7.0. The unscrambled TNFRII-Fc fragment fusion protein that elutes out as a second peak from the HIC column was collected.

Cation Exchange Chromatography (CEX) was employed as the polishing step to clean up the endotoxins and formulate TNFRII-Fc fragment fusion protein into the formulation buffer containing, 25 mM sodium phosphate, 25 mM sodium chloride, 25 mM L-arginine hydrochloride, 1% sucrose pH 6.5 ± 0.2 . The HIC peak 2 TNFRII-Fc fragment fusion protein pool that was dialyzed in 25 mM sodium phosphate pH 5.0 was loaded on to the SP SEPHAROSE FF (GE Healthcare Cat# 17-0729-01) column that was pre-equilibrated with 25mM sodium phosphate pH 5.0. After loading, the column was washed with 10 column volumes of 25 mM sodium phosphate pH 5.0 containing 10mM CHAPS, 10 mM EDTA followed by 10 column volumes wash with 25 mM Sodium phosphate pH 7.0. TNFRII-Fc fragment fusion protein was eluted as a single step elution with the formulation buffer. The peak region containing the TNFRII-Fc fragment fusion protein was pooled and sterile filtered using 0.2 μ m PES (PolyEtherSulfone) membrane filter and stored @4°C until PK/PD studies.

EXAMPLE 6

The Glycan composition of TNFRII-Fc fragment fusion protein produced in YGLY10299 (produces Form1), YGLY11731 (produces Form 2), and YGLY12680 (produces Form 3) was performed as follows.

O-Glycan Analysis by HPAEC-PAD

Analysis of *O*-glycans on the TNFRII-Fc fragment fusion protein can use the following protocol.

Yeast strains are grown in shakeflasks containing 100mL of BMGY for 48 hours, centrifuged, and the cell pellet and washed 1X with BMMY, and then resuspended in 50 mL BMMY and grown an additional 48 hours prior to harvest by centrifugation. Secreted TNFRII-Fc fragment fusion protein is purified from cleared supernatants using protein A chromatography

(Li *et al.* Nat. Biotechnol. 24(2):210-5 (2006)), and the *O*-glycans released from and separated from protein by alkaline elimination (β -elimination) (Harvey, Mass Spectrometry Reviews 18: 349- 451 (1999), Stadheim *et al.*, Nat. Protoc. 3:1026-31 (2006)). This process also reduces the newly formed reducing terminus of the released *O*-glycan (either oligomannose or mannose) to mannitol. The mannitol group thus serves as a unique indicator of each *O*- glycan.

About 0.5 nmole or more of protein, contained within a volume of 100 μ L PBS buffer, is used for β -elimination. The protein sample is treated with 25 μ L alkaline borohydride reagent and incubated at 50°C for 16 hours. About 20 μ L arabinol internal standard is added, followed by 10 μ L glacial acetic acid. The sample is then centrifuged through a Millipore filter containing both SEPABEADS and AG 50W-X8 resin and washed with water. The samples, including wash, are transferred to plastic autosampler vials and evaporated to dryness in a centrifugal evaporator. 150 μ L 1% AcOH/MeOH is added to the samples and the samples evaporated to dryness in a centrifugal evaporator. This last step is repeated five more times. 200 μ L of water is added and 100 μ L of the sample is analyzed by high pH anion-exchange chromatography coupled with pulsed electrochemical detection-HPLC (HPAEC-PAD) according to the manufacturer (Dionex, Sunnyvale, CA).

N-glycan analysis

To quantify the relative amount of each glycoform, the N-glycosidase F released glycans were labeled with 2-aminobenzidine (2-AB) and analyzed by HPLC as described in Choi *et al.*, Proc. Natl. Acad. Sci. USA 100: 5022-5027 (2003) and Hamilton *et al.*, Science 313: 1441-1443 (2006).

Total Sialic Acid Determination

The following assay detects total sialic acid content on glycoproteins as a ratio of moles sialic acid/mole protein. Sialic acid was released from glycoprotein samples by acid hydrolysis and analysed by HPAEC-PAD using the following method: About 10-15 μ g of protein sample were buffer-exchanged into phosphate buffered saline. Four hundred μ L of 0.1M hydrochloric acid was added, and the sample heated at 80°C for 1 hour. After drying in a SpeedVac (Savant), the samples were reconstituted with 500 μ L of water. One hundred μ L was then subjected to HPAEC-PAD analysis.

Purified TNFRII-Fc fragment fusion protein was electrophoresed on Tris-buffered 4-20% gradient SDS-polyacrylamide gels obtained from BioRad Laboratories (Hercules, CA). About 3 μ g of protein prepared in either reducing or non-reducing loading buffer was applied to a lane. A control consisted of commercially-available ENBREL. **Figure 31** shows that all three forms of TNFRII-Fc fragment fusion protein appeared to be similar in size to commercial ENBREL.

The Glycan compositions of the three forms of TNFRII-Fc fragment fusion protein were determined and the results presented in **Figure 32**. The figure shows that the glycan composition of the TNFRII-Fc fragment fusion protein was distinguishable from the glycan composition of ENBREL.

EXAMPLE 7

TNFRII-Fc fragment fusion protein produced in YGLY10299 (produces Form1), YGLY11731 (produces Form 2), and YGLY12680 (produces Form 3) was analyzed to assess and compare the bioactivity of the forms of TNFRII-Fc fragment fusion protein. The assays that used were (1) an *in vitro* assay to measure the effect sialylation of TNFRII-Fc fragment fusion protein has on its ability to inhibit TNF α -induced cell killing of L929 cells, (2) an *in vitro* assay to measure the effect sialylation of TNFRII-Fc fragment fusion protein has on its ability to inhibit TNF α -stimulated release of IL-6 in A549 cells, and (3) an *in vivo* assay in rat to measure the effect sialylation of TNFRII-fc fusion protein has on pharmacokinetics.

The three forms were compared to commercial ENBREL for ability to inhibit TNF α -induced cell killing of L929 cells. L929 cells were seeded overnight in 96-well plates at about 10,000 cells/well in Eagle's Minimum Essential Medium (ATCC Cat No. 30-2003) supplemented with 10% Fetal Bovine Serum at 37°C and 5% CO₂. Cells were then treated with human recombinant TNF α at 25 ng/mL with or without TNFRII-Fc fragment fusion protein or commercial ENBREL and then incubated for 24 hours under the same conditions. Then cell viability was measured by ATPlite (luminescence readout from Perkin-Elmer, Waltham, MA, see also U.S. Patent No. 6,503,723). The results are shown in **Figure 33** and show that the three forms of TNFRII-Fc fragment fusion protein were comparable to commercial ENBREL in inhibiting cell killing.

The three forms were compared to commercial ENBREL for ability to inhibit TNF α -stimulated release of IL-6 in A549 cells. A549 cells were seeded overnight in 96-well plates at about 50,000 cells/well in F-12K Medium (ATCC Cat No.30-2009)medium supplemented with_ 10% Fetal Bovine Serum at 37°C and 5% CO₂. Cells were then treated in triplicate with one of the three forms of TNFRII-Fc fragment fusion protein or commercial ENBREL and then stimulated with 3 ng/mL human recombinant TNF α and then incubated overnight under the same conditions. Then IL6 production was determined by AlphaLISA assay (Perkin-Elmer, Waltham, MA). The results are shown in **Figure 34** and show that the three forms of TNFRII-Fc fragment fusion protein were comparable to commercial ENBREL in inhibiting TNF α -stimulated release of IL-6.

The *in vivo* pharmacokinetics for each of the three forms was compared to that of commercial ENBREL. Spargue Dawley (SD) rats were dosed subcutaneously (SC) at 1 mg/kg with one of the three forms or commercial ENBREL and serum samples collected at 4, 24, 48,

72, 96, 120, 144, and 168 hour time points following administration. Serum concentration of the TNFRII-Fc fragment fusion protein or commercial ENBREL was determined with a Gyro immunoassay (Gyros US Inc., Monmouth Junction, NJ) using anti-TNFRII antibody as the capture antibody and labeled anti-Fc antibody for detection. The results are shown in **Figure 35** and show that Forms 1 and 2 of the TNFRII-Fc fragment fusion protein exhibited about 155-900 fold lower exposure than commercial ENBREL following SC administration and Form 3 TNFRII-Fc fragment fusion protein exhibited about 9-10 fold lower exposure than commercial ENBREL following SC administration. The results show that there is an apparent correlation between the extent of sialylation and increased *in vivo* pharmacokinetics.

Although this example demonstrates that the *O*-sialylated form of TNFRII-Fc (Form 3) has more activity *in vivo* compared to the *O*-mannose reduced glycan forms (Forms 1 and 2), all three forms demonstrated similar activity in *in vitro* assays. As such, it is foreseeable that one skilled in the art could increase the bioavailability and/or half-life of the *O*-mannose reduced glycan forms, to provide a therapeutic molecule with similar *in vivo* characteristics to the *O*-sialylated form or commercial ENBREL. One such strategy would be to increase the bioavailability of the molecule by formulation buffer optimization. An alternative strategy would be to increase the half-life of the molecule by conjugation to a carrier molecule to increase its physical size, for example, covalent linkage to polyethylene glycol.

EXAMPLE 8

Purification strategy for TNFRII-Fc fragment fusion protein produced in strain YGLY14252 as shown in **Figure 36**. The purification strategy enabled isolation of three forms of TNFRII-Fc fragment fusion protein: Form 5A, which has high relative total sialic acid (TSA) content; Form 5B, which has medium TSA content; and, Form 5C, which has low TSA content.

YGLY14252 was grown as described in Example 5 above. The purification of Forms 5A, 5B, and 5C of TNFRII-Fc fragment fusion protein obtained from YGLY14252 as shown in **Figure 36** was as follows.

Briefly, the same strategy as described in Example 5 was used with the following changes in the first intermediate purification step using Macro-Prep Ceramic Hydroxyapatite type I 40 µm resin. This step was not only used to remove the aggregated forms of TNFRII-Fc fragment fusion protein, but also to separate highly sialylated *N*- and *O*-Glycan containing fractions of TNFRII-Fc fragment fusion protein.

The Hydroxyapatite column was equilibrated with 3 column volumes of 5 mM sodium phosphate pH 6.5 and the mabselect pool containing TNFRII-Fc fragment fusion protein that was buffer exchanged into the equilibration buffer was applied on to the column. After loading, the column was washed with 3 column volumes of the equilibration buffer. The TNFRII-Fc fragment fusion protein that was present in the flowthrough and wash-unbound were

collected together as one pool and used for generating Form 5A which contains highly sialylated *N*- and *O*-glycans. Elution was performed by developing a gradient over 20 column volume ranging from 0 to 1000 mM Sodium chloride. TNFRII-Fc fragment fusion protein that elutes around 550 -650 mM Sodium chloride was pooled together and used for Form 5C generation.

5 The final formulated TNFRII-Fc fragment fusion protein of Forms 5A and 5C were mixed 1:1 protein ratio to generate Form 5B. All the three Forms 5A, 5B and 5C final formulated samples were stored @4°C until PK/PD studies.

EXAMPLE 9

10 The three forms of TNFRII-Fc fragment fusion protein obtained as shown in **Figure 36** were analyzed to assess and compare the bioactivity of the 5A, 5B, and 5C forms of TNFRII-Fc fragment fusion protein. The assays that used were (1) an *in vitro* assay to measure the effect sialylation of TNFRII-Fc fragment fusion protein has on its ability to inhibit TNF α -induced cell killing of L929 cells, (2) an *in vitro* assay to measure the effect sialylation of
15 TNFRII-Fc fragment fusion protein has on its ability to inhibit TNF α -stimulated release of IL-6 in A549 cells, and (3) an *in vivo* assay in rat and mouse to measure the effect sialylation of TNFRII-fc fusion protein has on pharmacokinetics.

 Purified 5A, 5B, and 5C forms of TNFRII-Fc fragment fusion protein were electrophoresed on Tris-buffered 4-20% gradient SDS-polyacrylamide gels obtained from
20 BioRad Laboratories (Hercules, CA). About 3 μ g of non-reduced protein was applied to a lane. A control consisted of commercially-available ENBREL. **Figure 37** shows that the Form 5A of TNFRII-Fc fragment fusion protein appeared to be similar in size to commercial ENBREL.

 The glycan compositions of the three forms of TNFRII-Fc fragment fusion protein were determined as in Example 6 and the results presented in **Figure 38**. The figure shows that
25 the glycan composition of each of the three fractions of TNFRII-Fc fragment fusion protein was distinguishable from the glycan composition of ENBREL.

Figure 39 shows the results of an *in vitro* assay to measure the effect sialylation of TNFRII-Fc fragment fusion protein has on its ability to inhibit TNF α -induced cell killing of L929 cells or inhibit TNF α -stimulated release of IL-6 in A549 cells. No significant difference
30 was observed between Merck TNFRII-Fc samples and commercial ENBREL.

 TNFRII-Fc fragment fusion protein Form 5A had a similar PK profile to commercial ENBREL following SC administration in both rat and mouse models (**Figure 40** and **Figure 41**, respectively). In contrast, TNFRII-Fc fragment fusion protein Forms 5B and 5C, each possessing a lower TSA content to Form 5A, had markedly lower *in vivo* PK when compared to both
35 commercial ENBREL and Form 5A (**Figure 40** and **Figure 41**). The results show that there is a direct correlation between the extent of sialylation and increased *in vivo* pharmacokinetics.

EXAMPLE 10

Pichia TNFRII-Fc was tested together with ENBREL for efficacy in a chronic mouse model of rheumatoid arthritis. The Tg197 genetically engineered mice overexpress a human TNF transgene and develop progressive arthritis (Keffer *et al.*, EMBO J. (13): 4025-4031 (1991)). The primary intent of the study was to verify whether the ability of *Pichia* TNFRII-Fc to neutralize TNF bioactivity translates into an ability to block the chronic effects of overexpressed TNF; the secondary purpose of the study was to compare the chronic effects of *Pichia* TNFRII-Fc to those of ENBREL. Transgenic mice were separated into 7 groups consisting of 8 gender and age-matched mice each, which received intraperitoneally 10µl of test compounds per gram of body weight, twice weekly. The groups received different test materials and dose levels, as follows: Vehicle, *Pichia* TNFRII-Fc at 30, 10 and 3 mg/kg; commercial ENBREL at 30, 10 and 3 mg/kg. Treatment was initiated at the onset of arthritis (three weeks of age) and continued over 8 weeks; the study was concluded at 10 weeks of age.

The assessment indicates (**Figure 42**) that *Pichia* TNFRII-Fc has in vivo potency and target efficacy. Its effectiveness shows a dose effect relationship, with higher doses increasing the anti-arthritic effect. The effects that *Pichia* TNFRII-Fc and commercial Enbrel have on the arthritic scores are similar at 30, 10 and 3 mg/kg dose levels.

EXAMPLE 11

An alternative purification strategy for enrichment of highly sialylated glycoforms of TNFRII-Fc was developed using phenyl borate chromatography instead of hydroxyapatite chromatography as shown by the scheme in **Figure 43**. This strategy was similar to the strategy as described in EXAMPLE 8 above except with the following changes in the first intermediate purification step in which PROSEP-PB chromatography media (non-compressible media comprising *m*-aminophenylborate ligands attached to glass beads; Millipore Corp. Cat#113247327) was used instead of Macro-Prep Ceramic Hydroxyapatite type I 40 µm resin to enrich for highly sialylated *N* and *O*-linked glycan containing fractions of TNFRII-Fc fragment fusion protein.

The PROSEP-PB column was equilibrated with 3 column volumes of 50 mM HEPES (N'-2-hydroxyethylpiperazine-N'-2 ethanesulphonic acid) pH 8.0 and the mabselect pool containing TNFRII-Fc fragment fusion protein that was previously buffer exchanged into the equilibration buffer was applied on to the column. After loading, the column was washed with 3 column volumes of the equilibration buffer. Elution was performed by developing a linear gradient over 30 column volumes ranging from 0 to 125 mM sorbitol in 50 mM HEPES pH8.0. Highly sialylated forms of TNFRII-Fc fragment fusion protein that elutes earlier in the gradient ranging between 7 mM to 20 mM sorbitol were collected and further processed through second intermediate step purification utilizing Hydrophobic Interaction Chromatography.

Figure 44 demonstrates that the protein quality of the material isolated (Form 7A) using this purification strategy was of similar quality to that of the commercial ENBREL control. Characterization of the glycan quality of Form 7A material (Figure 45) indicates that the TSA content compared to the commercial Enbrel lot used is similar to that highlighted in **Figure 37**, when comparing Form 5A to a different lot of commercial ENBREL. The *in vivo* comparison of the material purified using the Prosep-PB purification strategy in a rat pharmacokinetic study (**Figure 46**) indicates that the Form 7A material was comparable to commercial ENBREL.

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.

TABLE OF SEQUENCES		
SEQ ID NO:	Description Pp= <i>Pichia pastoris</i> Sc= <i>Saccharomyces cerevisiae</i>	Sequence
1	Sequence of the PpPMA1 promoter:	AAATGCGTACCTCTTCTACGAGATTCAAGCGAATGAG AATAATGTAATATGCAAGATCAGAAAGAATGAAAGG AGTTGAAAAAAAAAAACCGTTGCGTTTTGACCTTGAAT GGGGTGGAGGTTTCCATTCAAAGTAAAGCCTGTGTCT TGGTATTTTCGGCGGCACAAGAAATCGTAATTTTCATC TTCTAAACGATGAAGATCGCAGCCCAACCTGTATGTA GTTAACCGGTCGGAATTATAAGAAAGATTTTCGATCA ACAAACCCTAGCAAATAGAAAGCAGGGTTACAACCTT AAACCGAAGTCACAAACGATAAACCACTCAGCTCCCA CCCAAATTCATTCCCCTAGCAGAAAGGAATTATTTA ATCCCTCAGGAAACCTCGATGATTCTCCCGTTCTTCCA TGGGCGGGTATCGCAAATGAGGAATTTTCAAATTT CTCTATTGTCAAGACTGTTTATTATCTAAGAAATAGCC

		CAATCCGAAGCTCAGTTTTGAAAAAATCACTTCCGCG TTTCTTTTTTACAGCCCGATGAATATCCAAATTTGGAA TATGGATTACTCTATCGGGACTGCAGATAATATGACA ACAACGCAGATTACATTTTAGGTAAGGCATAAACACC AGCCAGAAATGAAACGCCCACTAGCCATGGTCGAATA GTCCAATGAATTCAGATAGCTATGGTCTAAAAGCTGA TGTTTTTTATTGGGTAATGGCGAAGAGTCCAGTACGAC TTCCAGCAGAGCTGAGATGGCCATTTTTGGGGGTATT AGTAACTTTTTGAGCTCTTTTCACTTCGATGAAGTGTC CCATTCGGGATATAATCGGATCGCGTCGTTTTCTCGAA AATACAGCTTAGCGTCGTCCGCTTGTTGTAAAAGCAG CACCACATTCCTAATCTCTTATATAAAACAAAACAACCC AAATTATCAGTGCTGTTTTCCCAACAGATATAAGTTTC TTTTCTCTCCGCTTTTTGATTTTTTATCTTTTCCTTA AAAACCTCTTTACCTTAAAGGGCGGCC
2	Pp AOX1 promoter	AACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTG CCATCCGACATCCACAGGTCCATTCTCACACATAAGT GCCAAACGCAACAGGAGGGGATACACTAGCAGCAGA CCGTTGCAAACGCAGGACCTCCACTCCTCTTCTCCTCA ACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATT GGGCTTGATTGGAGCTCGCTCATTCCAATTCCTTCTAT TAGGCTACTAACACCATGACTTTATTAGCCTGTCTATC CTGGCCCCCTGGCGAGGTTTCATGTTTGTTTATTTCCG AATGCAACAAGCTCCGCATTACACCCGAACATCACTC CAGATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTT CATGTTCCCCAAATGGCCCAAACTGACAGTTTAAAC GCTGTCTTGGAACCTAATATGACAAAAGCGTGATCTC ATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTA ACGGCCAGTTGGTCAAAAAGAACTTCCAAAAGTCGG CATACCGTTTGTCTTGTTTGGTATTGATTGACGAATGC TCAAAAATAATCTCATTAAATGCTTAGCGCAGTCTCTCT ATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGC AAATGGGGAAACACCCGCTTTTTGGATGATTATGCAT TGTCTCCACATTGTATGCTTCCAAGATTCTGGTGGGAA TACTGCTGATAGCCTAACGTTTCATGATCAAAATTTAAC TGTTCTAACCCCTACTTGACAGCAATATATAAACAGA AGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATC

		ATTATTAGCTTACTTTTCATAATTGCGACTGGTTCCAAT TGACAAGCTTTTGGATTTTAACGACTTTTAACGACAACT TGAGAAGATCAAAAAACAATAATTATTCGAAACG
3	ScCYC TT	ACAGGCCCTTTTCCTTTGTCGATATCATGTAATTAGT TATGTCACGCTTACATTCACGCCCTCCTCCACATCCG CTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGT CTAGGTCCCTATTTATTTTTTTTAATAGTTATGTTAGTA TTAAGAACGTTATTTATATTTCAAATTTTCTTTTTTTT CTGTACAAACGCGTGTACGCATGTAACATTATACTGA AAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGC TTAATTTGCAAGCTGCCGGCTCTTAAG
4	PpRPL10 promoter	GTTCTTCGCTTGGTCTTGTATCTCCTTACACTGTATCTT CCCATTTGCGTTTAGGTGGTTATCAAAAATAAAAGG AAAAATTTAGATGTTTATCTCTAAGGTTTTTCTTTTT ACAGTATAACACGTGATGCGTCACGTGGTACTAGATT ACGTAAGTTATTTTGGTCCGGTGGGTAAGTGGGTAAG AATAGAAAGCATGAAGGTTTACAAAACGCAGTCACG AATTATTGCTACTTCGAGCTTGGAACCAACCCCAAAGA TTATATTGTACTGATGCACTACCTTCTCGATTTTGCTCC TCCAAGAACCTACGAAAAACATTTCTTGAGCCTTTTCA ACCTAGACTACACATCAAGTTATTTAAGGTATGTTCCG TTAACATGTAAGAAAAGGAGAGGATAGATCGTTTATG GGGTACGTCGCCTGATTCAAGCGTGACCATTGGAAGA ATAGGCCTTCGAAAGCTGAATAAAGCAAATGTCAGTT GCGATTGGTATGCTGACAAATTAGCATAAAAAGCAAT AGACTTTCTAACCACTGTTTTTTTCTTTTACTTTATT TATATTTTGCCACCGTACTAACAAGTTCAGACAAA
5	PpGAPDH promoter	TTTTTGTAAGAAATGTCTTGGTGTCTCGTCCAATCAGG TAGCCATCTCTGAAATATCTGGCTCCGTTGCAACTCCG AACGACCTGCTGGCAACGTAAAATTCTCCGGGGTAAA ACTTAAATGTGGAGTAATGGAACCAGAAACGTCTCTT CCCTTCTCTCTCCTTCCACCGCCCGTTACCGTCCCTAG GAAATTTTACTCTGCTGGAGAGCTTCTTCTACGGCCCC CTTGCAAGCAATGCTCTTCCCAGCATTACGTTGCGGGTA AAACGGAGGTCGTGTACCCGACCTAGCAGCCCAGGGA TGGAAGAGTCCCGGCCGTCGCTGGCAATAATAGCGGG CGGACGCATGTCATGAGATTATTGGAACCAACAGAA

		TCGAATATAAAAGGCGAACACCTTTCCCAATTTTGGTT TCTCCTGACCCAAAGACTTTAAATTTAATTTATTTGTC CCTATTTCAATCAATTGAACAACCTATCAAAACACA
6	PpTEF1 promoter	TTAAGGTTTGGAAACAACACTAAACTACCTTGCGGTAC TACCATTGACACTACACATCCTTAATTCCAATCCTGTC TGGCCTCCTTCACCTTTTAACCATCTTGCCCATTTCCAA CTCGTGTCAGATTGCGTATCAAGTGAAAAAAAAAAAAA TTTTAAATCTTTAACCCAATCAGGTAATAACTGTGCGC TCTTTTATCTGCCGCACTGCATGAGGTGTCCCCTTAGT GGGAAAGAGTACTGAGCCAACCCTGGAGGACAGCAA GGGAAAAATACCTACAACCTTGCTTCATAATGGTCGTA AAAACAATCCTTGTCGGATATAAGTGTTGTAGACTGT CCCTTATCCTCTGCGATGTTCTTCCTCTCAAAGTTTGC GATTCTCTCTATCAGAATTGCCATCAAGAGACTCAGG ACTAATTTGCGAGTCCCACACGCACTCGTACATGATTG GCTGAAATTTCCCTAAAGAATTTCTTTTTTCACGAAAAT TTTTTTTTTACACAAGATTTTCAGCAGATATAAAATGG AGAGCAGGACCTCCGCTGTGACTCTTCTTTTTTTTCTTT TATTCTCACTACATACATTTTAGTTATTCGCCAAC
7	PpTEF1 TT	ATTGCTTGAAGCTTTAATTTATTTTATTAACATAATAA TAATACAAGCATGATATATTTGTATTTTGTTTCGTTAAC ATTGATGTTTTCTTCATTTACTGTTATTGTTTGTAACCT TGATCGATTTATCTTTTCTACTTTACTGTAATATGGCTG GCGGGTGAGCCTTGAACCTCCCTGTATTACTTTACCTTG CTATTACTTAATCTATTGACTAGCAGCGACCTCTTCAA CCGAAGGGCAAGTACACAGCAAGTTCATGTCTCCGTA AGTGTCATCAACCCTGGAAACAGTGGGCCATGTC
8	PpALG3 TT	ATTTACAATTAGTAATATTAAGGTGGTAAAAACATTC GTAGAATTGAAATGAATTAATATAGTATGACAATGGT TCATGTCTATAAATCTCCGGCTTCGGTACCTTCTCCCC AATTGAATACATTGTCAAATGAATGGTTGAACTATT AGGTTGCGCCAGTTTCGTTATTAAGAAAACCTGTTAAAAT CAAATTCCATATCATCGGTTCCAGTGGGAGGACCAGT TCCATCGCCAAAATCCTGTAAGAATCCATTGTCAGAA CCTGTAAAGTCAGTTTGAGATGAAATTTTCCGGTCTT TGTTGACTTGGAAGCTTCGTTAAGGTTAGGTGAAACA GTTTGATCAACCAGCGGCTCCCGTTTTTCGTCGCTTAGT

		AG
9	PpTRP1 5' region and ORF	GCGGAAACGGCAGTAAACAATGGAGCTTCATTAGTGG GTGTTATTATGGTCCCTGGCCGGGAACGAACGGTGAA ACAAGAGGTTGCGAGGGAAATTTTCGCAGATGGTGC GG GAAAAGAGAATTTCAAAGGGCTCAAAATACTTGGATT CCAGACA ACTGAGGAAAGAGTGGGACGACTGTCCTCT GGAAGACTGGTTT GAGTACAACGTGAAAGAAATAAAC AGCAGTGGTCCATTTTTAGTTGGAGTTTTTCGTAATCA AAGTATAGATGAAATCCAGCAAGCTATCCACACTCAT GGTTTGGATTTCGTCCA ACTACATGGGTCTGAGGATTT TGATTCGTATATACGCAATATCCCAGTTCCTGTGATTA CCAGATACACAGATAATGCCGTCGATGGTCTTACCGG AGAAGACCTCGCTATAAATAGGGCCCTGGTGCTACTG GACAGCGAGCAAGGAGGTGAAGGAAAAACCATCGAT TGGGCTCGTGCACAAAAATTTGGAGAACGTAGAGGAA AATATTTACTAGCCGGAGGTTTGACACCTGATAATGTT GCTCATGCTCGATCTCATACTGGCTGTATTGGTGTTGA CGTCTCTGGTGGGGTAGAAACAAATGCCTCAAAAGAT ATGGACAAGATCACACAATTTATCAGAAACGCTACAT AA
10	PpTRP1 3' region	AAGTCAATTAAATACACGCTTGAAAGGACATTACATA GCTTTCGATTTAAGCAGAACCAGAAATGTAGAACCAC TTGTCAATAGATTGGTCAATCTTAGCAGGAGCGGCTG GGCTAGCAGTTGGAACAGCAGAGGTTGCTGAAGGTGA GAAGGATGGAGTGGATTGCAAAGTGGTGTTGGTTAAG TCAATCTCACCAGGGCTGGTTTTGCCAAAAATCAACTT CTCCCAGGCTTCACGGCATTCTTGAATGACCTCTTCTG CATACTTCTTGTTCTTGCATTACCCAGAGAAAGCAAAC TGGTTCTCAGGTTTTCCATCAGGGATCTTGTAATTCT GAACCATTCGTTGGTAGCTCTCAACAAGCCCGGCATG TGCTTTTCAACATCCTCGATGTCATTGAGCTTAGGAGC CAATGGGTCGTTGATGTCGATGACGATGACCTTCCAG TCAGTCTCTCCCTCATCCAACAAAGCCATAACACCGA GGACCTTGACTTGCTTGACCTGTCCAGTGTAACCTACG GCTTCACCAATTTTCGCAAACGTCCAATGGATCATTGTC ACCCTTGGCCTTGGTCTCTGGATGAGTGACGTTAGGGT CTTCCCATGTCTGAGGGAAGGCACCGTAGTTGTGAAT

		GTATCCGTGGTGAGGGAAACAGTTACGAACGAAACGA AGTTTTCCCTTCTTTGTGTCCTGAAGAATTGGGTTTCAG TTTCTCCTCCTTGGAATCTCCAACCTTGGCGTTGGTCC AACGGGGGACTTCAACAACCATGTTGAGAACCTTCTT GGATTCGTCAGCATAAAGTGGGATGTCGTGGAAAGGA GATACGACTT
11	ScARR3 ORF	ATGTCAGAAGATCAAAAAAGTGAAAATTCCGTACCTT CTAAGGTTAATATGGTGAATCGCACCGATATACTGAC TACGATCAAGTCATTGTCATGGCTTGACTTGATGTTGC CATTTACTATAATTCTCTCCATAATCATTGCAGTAATA ATTTCTGTCTATGTGCCTTCTTCCCGTCACACTTTTGAC GCTGAAGGTCATCCCAATCTAATGGGAGTGTCCATTC CTTTGACTGTTGGTATGATTGTAATGATGATTCCCCCG ATCTGCAAAGTTTCCTGGGAGTCTATTCACAAGTACTT CTACAGGAGCTATATAAGGAAGCAACTAGCCCTCTCG TTATTTTTGAATTGGGTCATCGGTCCTTTGTTGATGAC AGCATTGGCGTGGATGGCGCTATTCGATTATAAGGAA TACCGTCAAGGCATTATTATGATCGGAGTAGCTAGAT GCATTGCCATGGTGCTAATTTGGAATCAGATTGCTGG AGGAGACAATGATCTCTGCGTCGTGCTTGTTATTACAA ACTCGCTTTTACAGATGGTATTATATGCACCATTGCAG ATATTTTACTGTTATGTTATTTCTCATGACCACCTGAA TACTTCAAATAGGGTATTATTCGAAGAGGTTGCAAAG TCTGTCGGAGTTTTTCTCGGCATACCACTGGGAATTGG CATTATCATACGTTTGGGAAGTCTTACCATAGCTGGTA AAAGTAATTATGAAAAATACATTTTGAGATTTATTTCT CCATGGGCAATGATCGGATTTTATTACACTTTATTTGT TATTTTATTAGTAGAGGTTATCAATTTATCCACGAAA TTGGTTCTGCAATATTGTGCTTTGTCCCATTTGGTGCTTT ACTTCTTTATTGCATGGTTTTTGACCTTCGCATTAATG AGGTACTTATCAATATCTAGGAGTGATACACAAAGAG AATGTAGCTGTGACCAAGAAGTACTTTTAAAGAGGGT CTGGGGAAGAAAGTCTTGTGAAGCTAGCTTTTCTATTA CGATGACGCAATGTTTCACTATGGCTTCAAATAATTTT GAACTATCCCTGGCAATTGCTATTTCTTATATGGTAA CAATAGCAAGCAAGCAATAGCTGCAACATTTGGGCCG TTGCTAGAAGTTCCAATTTTATTGATTTTGGCAATAGT

		CGCGAGAATCCTTAAACCATATTATATATGGAACAAT AGAAATTAA
12	PpURA6 region	CAAATGCAAGAGGACATTAGAAATGTGTTTGGTAAGA ACATGAAGCCGGAGGCATACAAACGATTCACAGATTT GAAGGAGGAAAACAACTGCATCCACCGGAAGTGCC AGCAGCCGTGTATGCCAACCTTGCTCTCAAAGGCATT CCTACGGATCTGAGTGGGAAATATCTGAGATTCACAG ACCCACTATTGGAACAGTACCAAACCTAGTTTGGCCG ATCCATGATTATGTAATGCATATAGTTTTTGTTCGATGC TCACCCGTTTCGAGTCTGTCTCGTATCGTCTTACGTAT AAGTTCAAGCATGTTTACCAGGTCTGTTAGAACTCCT TTGTGAGGGCAGGACCTATTCGTCTCGGTCCCGTTGTT TCTAAGAGACTGTACAGCCAAGCGCAGAATGGTGGCA TTAACCATAAGAGGATTCTGATCGGACTTGGTCTATTG GCTATTGGAACCAACCTTTACGGGACAACCAACCTA CCAAGACTCCTATTGCATTTGTGGAACCAGCCACGGA AAGAGCGTTTAAGGACGGAGACGTCTCTGTGATTTTT GTTCTCGGAGGTCCAGGAGCTGGAAAAGGTACCCAAT GTGCCAAACTAGTGAGTAATTACGGATTTGTTACCTG TCAGCTGGAGACTTGTTACGTGCAGAACAGAAGAGGG AGGGGTCTAAGTATGGAGAGATGATTTCCCAGTATAT CAGAGATGGACTGATAGTACCTCAAGAGGTCACCATT GCGCTCTTGGAGCAGGCCATGAAGGAAAACCTTCGAGA AAGGGAAGACACGGTTCTTGATTGATGGATTCCCTCG TAAGATGGACCAGGCCAAAACCTTTTGAGGAAAAAGTC GCAAAGTCCAAGGTGACACTTTTCTTTGATTGTCCCGA ATCAGTGCTCCTTGAGAGATTACTTAAAAGAGGACAG ACAAGCGGAAGAGAGGATGATAATGCGGAGAGTATC AAAAAAAGATTCAAAACATTCGTGGAACTTCGATGC CTGTGGTGGACTATTTTCGGGAAGCAAGGACGCGTTTT GAAGGTATCTTGTGACCACCCTGTGGATCAAGTGTATT CACAGGTTGTGTGCGGTGCTAAAAGAGAAGGGGATCTT TGCCGATAACGAGACGGAGAATAAATAA
13	NatR expression cassette (CDS 385-954, represented in	TGTTTAGCTTGCCTCGTCCCCGCCGGGTACCCGGCCA GCGACATGGAGGCCCAGAATACCCTCCTTGACAGTCT TGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGC CCGTACATTTAGCCCATACATCCCCATGTATAATCATT

	bold)	TGCATCCATACATTTTGTATGGCCGCACGGCGCGAAGC AAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGG GAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCA CGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGAT TTGCCACTGAGGTTCTTCTTTTCATATACTTCCTTTTAAA ATCTTGCTAGGATACAGTTCTCACATCACATCCGAACA TAAACAACCATGGGTACCACTCTTGACGACACGGCT TACCGGTACCGCACCAGTGTCCCGGGGGACGCCGA GGCCATCGAGGCACTGGATGGGTCCTTCACCACCG ACACCGTCTTCCGCGTCACCGCCACCGGGGACGGC TTCACCCTGCGGGAGGTGCCGGTGGACCCGCCCT GACCAAGGTGTTCCCCGACGACGAATCGGACGACG AATCGGACGACGGGGAGGACGGCGACCCGGACTC CCGGACGTTCGTGCGGTACGGGGACGACGGCGACC TGGCGGGCTTCGTGGTCTGTCTCGTACTCCGGCTGG AACCGCCGGCTGACCGTCGAGGACATCGAGGTCTGC CCCGGAGCACCGGGGGCACGGGGTCCGGCGCGCG TTGATGGGGCTCGCGACGGAGTTCGCCCCGCGAGCG GGGCGCCGGGCACCTCTGGCTGGAGGTCACCAACG TCAACGCACCGGCGATCCACGCGTACCGGCGGATG GGGTTACCCCTCTGCGGCCTGGACACCGCCCTGTA CGACGGCACCGCCTCGGACGGCGAGCAGGCGCTCT ACATGAGCATGCCCTGCCCCCTAATCAGTACTGACAA TAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTAT AGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAATG TTAGCGTGATTTATATTTTTTTTCGCTCGACATCATCT GCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATC ATGCGTCAATCGTATGTGAATGCTGGTCGCTATACTGC TGTCGATTCGATACTAACGCCGCCATCCAGTGTCGAA AAC
14	Sequence of the Sh ble ORF (Zeocin resistance marker):	ATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACC CGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGA CCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGAC TTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTTAT CAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACC CTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGT ACGCCGAGTGGTCGGAGGTCGTGTCCACGAACCTCCG

		GGACGCCTCCGGGCGGGCCATGACCGAGATCGGCGAG CAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGG CCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGA CTGA
15	PpAOX1 TT	TCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATG CAGGCTTCATTTTGATACTTTTTTATTGTAACTATAT AGTATAGGATTTTTTTTGTCAATTTGTTTCTTCTCGTAC GAGCTTGCTCCTGATCAGCCTATCTCGCAGCTGATGAA TATCTTGTGGTAGGGGTTTGGGAAAATCATTGAGTTT GATGTTTTTCTTGGTATTTCCCACTCCTCTTCAGAGTAC AGAAGATTAAGTGAGACGTTTCGTTTGTGCA
16	ScTEF1 promoter	GATCCCCACACACCATAGCTTCAAAATGTTTCTACTC CTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATC GCCGTACCACTTCAAAACACCCAAGCACAGCATACTA AATTTCCCCTCTTTCTTCTCTAGGGTGTCGTTAATTAC CCGTACTAAAGGTTTGGAAAAGAAAAAGAGACCGC CTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAAT TTTTATCACGTTTCTTTTTCTTGAAAATTTTTTTTTTG ATTTTTTCTCTTTCGATGACCTCCCATTGATATTTAAG TTAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCA TTTTTCTTGTTCTATTACAACTTTTTTTACTTCTTGCTC ATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTAA ATTACAAA
17	<i>S. cerevisiae</i> invertase gene (ScSUC2) ORF underlined	AGGCTCGCAACAACCTATAATTGAGTTAAGTGCCTTT CCAAGCTAAAAAGTTTGAGGTTATAGGGGCTTAGCAT CCACACGTCACAATCTCGGGTATCGAGTATAGTATGT AGAATTACGGCAGGAGGTTTCCCAATGAACAAAGGAC AGGGGCACGGTGAGCTGTCGAAGGTATCCATTTTATC ATGTTTCGTTTGTACAAGCACGACATACTAAGACATTT ACCGTATGGGAGTTGTTGTCCTAGCGTAGTTCTCGCTC CCCCAGCAAAGCTCAAAAAAGTACGTCATTTAGAATA GTTTGTGAGCAAATTACCAGTCGGTATGCTACGTTAG AAAGGCCACAGTATTCTTCTACCAAAGGCGTGCCTTT GTTGAACTCGATCCATTATGAGGGCTTCCATTATTCCC CGCATTTTTATTACTCTGAACAGGAATAAAAAGAAAA AACCAGTTTAGGAAATTATCCGGGGGCGAAGAAATA CGCGTAGCGTTAATCGACCCACGTCCAGGGTTTTTCC

ATGGAGGTTTCTGGAAAACTGACGAGGAATGTGATT
 ATAAATCCCTTTATGTGATGTCTAAGACTTTTAAGGTA
 CGCCCGATGTTTGCCTATTACCATCATAGAGACGTTTC
 TTTTCGAGGAATGCTTAAACGACTTTGTTTGACAAAAA
 TGTTCCTAAGGGCTCTATAGTAAACCATTTGGAAGA
 AAGATTTGACGACTTTTTTTTTTTTGGATTTGATCCTAT
 AATCCTTCCTCCTGAAAAGAAACATATAAATAGATAT
 GTATTATTCTTCAAAACATTCTCTTGTTCTTGTCCTTT
 TTTTACCATATATCTTACTTTTTTTTTTCTCTCAGAGA
 AACAAGCAAAACAAAAAGCTTTTCTTTTCACTAACGT
 ATATGATGCTTTTGCAAGCTTTCCTTTTCCTTTTGGCTG
GTTTTGCAGCCAAAATATCTGCATCAATGACAAACGA
AACTAGCGATAGACCTTTGGTCCACTTCACACCCAAC
AAGGGCTGGATGAATGACCCAAATGGGTTGTGGTACG
ATGAAAAAGATGCCAAATGGCATCTGTACTTTCAATA
CAACCCAAATGACACCGTATGGGGTACGCCATTGTTT
TGGGGCCATGCTACTTCCGATGATTTGACTAATTGGGA
AGATCAACCCATTGCTATCGCTCCCAAGCGTAACGAT
TCAGGTGCTTTCTCTGGCTCCATGGTGGTTGATTACAA
CAACACGAGTGGGTTTTTCAATGATACTATTGATCCAA
GACAAAGATGCGTTGCGATTTGGACTTATAACACTCC
TGAAAGTGAAGAGCAATACATTAGCTATTCTCTTGAT
GGTGGTTACACTTTTACTGAATACCAAAAGAACCCTG
TTTAGCTGCCAACTCCACTCAATTCAGAGATCCAAAG
GTGTTCTGGTATGAACCTTCTCAAAAATGGATTATGAC
GGCTGCCAAATCACAAGACTACAAAATTGAAATTTAC
TCCTCTGATGACTTGAAGTCCTGGAAGCTAGAATCTGC
ATTTGCCAATGAAGGTTTCTTAGGCTACCAATACGAAT
GTCCAGGTTTGATTGAAGTCCCAACTGAGCAAGATCC
TTCCAAATCTTATTGGGTCATGTTTATTTCTATCAACC
CAGGTGCACCTGCTGGCGGTTCCCTTCAACCAATATTTT
GTTGGATCCTTCAATGGTACTCATTTTGAAGCGTTTGA
CAATCAATCTAGAGTGGTAGATTTTGGTAAGGACTAC
TATGCCTTGCAAACCTTCTTCAACACTGACCCAACCTA
CGGTTCAGCATTAGGTATTGCCTGGGCTTCAAACCTGG
GAGTACAGTGCCTTTGTCCCAACTAACCCATGGAGAT
CATCCATGTCTTTGGTCCGCAAGTTTTCTTTGAACACT

		<u>GAATATCAAGCTAATCCAGAGACTGAATTGATCAATT</u> <u>TGAAAGCCGAACCAATATTGAACATTAGTAATGCTGG</u> <u>TCCCTGGTCTCGTTTTGCTACTAACACAACCTCTAACTA</u> <u>AGGCCAATTCTTACAATGTCGATTTGAGCAACTCGACT</u> <u>GGTACCCTAGAGTTTGAGTTGGTTTACGCTGTTAACAC</u> <u>CACACAAACCATATCCAAATCCGTCTTTGCCGACTTAT</u> <u>CACTTTGGTTCAAGGGTTTAGAAGATCCTGAAGAATA</u> <u>TTTGAGAATGGGTTTTGAAGTCAGTGCTTCTTCCTTCT</u> <u>TTTTGGACCGTGGTAACTCTAAGGTCAAGTTTGTCAAG</u> <u>GAGAACCCATATTTACAAACAGAATGTCTGTCAACA</u> <u>ACCAACCATTCAAGTCTGAGAACGACCTAAGTTACTA</u> <u>TAAAGTGTACGGCCTACTGGATCAAAACATCTTGGAA</u> <u>TTGTACTTCAACGATGGAGATGTGGTTTCTACAAATAC</u> <u>CTACTTCATGACCACCGGTAACGCTCTAGGATCTGTGA</u> <u>ACATGACCACTGGTGTGCGATAATTTGTTCTACATTGAC</u> <u>AAGTTCCAAGTAAGGGAAGTAAAATAGAGGTTATAA</u> <u>AACTTATTGTCTTTTTTATTTTTTTCAAAGCCATTCTA</u> <u>AAGGGCTTTAGCTAACGAGTGACGAATGTAAAACCTT</u> <u>ATGATTTCAAAGAATACCTCCAAACCATTGAAAATGT</u> <u>ATTTTTATTTTTATTTTCTCCCGACCCAGTTACCTGGA</u> <u>ATTTGTTCTTTATGTACTTTATATAAGTATAATTCTCTT</u> <u>AAAAATTTTTACTACTTTGCAATAGACATCATTTTTTC</u> <u>ACGTAATAAACCCACAATCGTAATGTAGTTGCCTTAC</u> <u>ACTACTAGGATGGACCTTTTTGCCTTTATCTGTTTTGTT</u> <u>ACTGACACAATGAAACCGGGTAAAGTATTAGTTATGT</u> <u>GAAAATTTAAAAGCATTAAAGTAGAAGTATACCATATT</u> <u>GTAAAAAAGCGTTGTCTTCTACGTAAAAGTGT</u> <u>TCTCAAAAAGAAGTAGTGAGGGAAATGGATACCAAGC</u> <u>TATCTGTAACAGGAGCTAAAAAATCTCAGGGAAAAGC</u> <u>TTCTGGTTTGGGAAACGGTCGAC</u>
18	Sequence of the 5'-Region used for knock out of PpURA5:	ATCGGCCTTTGTTGATGCAAGTTTTACGTGGATCATGG ACTAAGGAGTTTTATTTGGACCAAGTTCATCGTCCTAG ACATTACGGAAAGGGTTCTGCTCCTCTTTTTGGAACT TTTTGGAACCTCTGAGTATGACAGCTTGGTGGATTGTA CCCATGGTATGGCTTCCTGTGAATTTCTATTTTTTCTAC ATTGGATTCACCAATCAAAACAAATTAGTCGCCATGG CTTTTTGGCTTTTGGGTCTATTTGTTTGGACCTTCTTGG

		AATATGCTTTGCATAGATTTTTGTTCCACTTGGACTAC TATCTTCCAGAGAATCAAATTGCATTTACCATTCATTT CTTATTGCATGGGATACACCACTATTTACCAATGGATA AATACAGATTGGTGATGCCACCTACACTTTTCATTGTA CTTTGCTACCCAATCAAGACGCTCGTCTTTTCTGTTCT ACCATATTACATGGCTTGTCTGGATTTGCAGGTGGAT TCCTGGGCTATATCATGTATGATGTCACTCATTACGTT CTGCATCACTCCAAGCTGCCTCGTTATTTCCAAGAGTT GAAGAAATATCATTTGGAACATCACTACAAGAATTAC GAGTTAGGCTTTGGTGTCACCTCCAAATTCTGGGACAA AGTCTTTGGGACTTATCTGGGTCCAGACGATGTGTATC AAAAGACAAATTAGAGTATTTATAAAGTTATGTAAGC AAATAGGGGCTAATAGGGAAAGAAAAATTTTGGTTCT TTATCAGAGCTGGCTCGCGCGCAGTGTTTTTCGTGCTC CTTTGTAATAGTCATTTTGGACTACTGTTTCAGATTGAA ATCACATTGAAGATGTCACTCGAGGGGTACCAAAAAA GGTTTTTGGATGCTGCAGTGGCTTCGC
19	Sequence of the 3'-Region used for knock out of PpURA5:	GGTCTTTTCAACAAAGCTCCATTAGTGAGTCAGCTGGC TGAATCTTATGCACAGGCCATCATTAAACAGCAACCTG GAGATAGACGTTGTATTTGGACCAGCTTATAAAGGTA TTCCTTTGGCTGCTATTACCGTGTTGAAGTTGTACGAG CTCGGCGGCAAAAAATACGAAAATGTCGGATATGCGT TCAATAGAAAAGAAAAGAAAGACCACGGAGAAGGTG GAAGCATCGTTGGAGAAAGTCTAAAGAATAAAAGAGT ACTGATTATCGATGATGTGATGACTGCAGGTACTGCT ATCAACGAAGCATTTGCTATAATTGGAGCTGAAGGTG GGAGAGTTGAAGGTAGTATTATTGCCCTAGATAGAAT GGAGACTACAGGAGATGACTCAAATACCAGTGCTACC CAGGCTGTTAGTCAGAGATATGGTACCCCTGTCTTGA GTATAGTGACATTGGACCATATTGTGGCCCATTTGGGC GAAACTTTCACAGCAGACGAGAAATCTCAAATGGAAA CGTATAGAAAAAAGTATTTGCCCAAATAAGTATGAAT CTGCTTCGAATGAATGAATTAATCCAATTATCTTCTCA CCATTATTTTCTTCTGTTTCGGAGCTTTGGGCACGGCG GCGGGTGGTGCGGGCTCAGGTTCCTTTCATAAACAG ATTTAGTACTTGGATGCTTAATAGTGAATGGCGAATGC

		AAAGGAACAATTTTCGTTTCATCTTTAACCCTTTCACTCG GGGTACACGTTCTGGAATGTACCCGCCCTGTTGCAACT CAGGTGGACCGGGCAATTCTTGAACCTTTCTGTAACGTT GTTGGATGTTCAACCAGAAATTGTCCTACCAACTGTAT TAGTTTCCTTTTGGTCTTATATTGTTTCATCGAGATACTT CCCACTCTCCTTGATAGCCACTCTCACTCTTCCTGGAT TACCAAAATCTTGAGGATGAGTCTTTTCAGGCTCCAG GATGCAAGGTATATCCAAGTACCTGCAAGCATCTAAT ATTGTCCTTGCCAGGGGGTTCTCCACACCATACTCCTT TTGGCGCATGC
20	Sequence of the PpURA5 auxotrophic marker:	TCTAGAGGGACTTATCTGGGTCCAGACGATGTGTATC AAAAGACAAATTAGAGTATTTATAAAGTTATGTAAGC AAATAGGGGCTAATAGGGGAAAGAAAAATTTTGGTTCT TTATCAGAGCTGGCTCGCGCGCAGTGTTTTTCGTGCTC CTTTGTAATAGTCATTTTTTGACTIONGTTTCAGATTGAA ATCACATTGAAGATGTCACTGGAGGGGTACCAAAAAA GGTTTTTGGATGCTGCAGTGGCTTCGCAGGCCTTGAAG TTTGGAACCTTTCACCTTGAAAAGTGGAAGACAGTCTC CATACTTCTTTAACATGGGTCTTTTCAACAAAGCTCCA TTAGTGAGTCAGCTGGCTGAATCTTATGCTCAGGCCAT CATTAACAGCAACCTGGAGATAGACGTTGTATTTGGA CCAGCTTATAAAGGTATTCCTTTGGCTGCTATTACCGT GTTGAAGTTGTACGAGCTGGGCGGCAAAAAATACGAA AATGTCGGATATGCGTTCAATAGAAAAGAAAAGAAAG ACCACGGAGAAGGTGGAAGCATCGTTGGAGAAAGTCT AAAGAATAAAAGAGTACTGATTATCGATGATGTGATG ACTGCAGGTACTGCTATCAACGAAGCATTGCTATAA TTGGAGCTGAAGGTGGGAGAGTTGAAGGTTGTATTAT TGCCCTAGATAGAATGGAGACTACAGGAGATGACTCA AATACCAGTGCTACCCAGGCTGTTAGTCAGAGATATG GTACCCCTGTCTTGAGTATAGTGACATTGGACCATATT GTGGCCCATTTGGGCGAACTTTCACAGCAGACGAGA AATCTCAAATGGAAACGTATAGAAAAAAGTATTTGCC CAAATAAGTATGAATCTGCTTCGAATGAATGAATTAA TCCAATTATCTTCTCACCATTATTTTCTTCTGTTTCGGA GCTTTGGGCACGGCGGCGGATCC

21	Sequence of the part of the Ec lacZ gene that was used to construct the PpURA5 blaster (recyclable auxotrophic marker)	CCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTG GCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAG GTAAACAGTTGATTGAACTGCCTGAACTACCGCAGCC GGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTA GTGCAACCGAACGCGACCGCATGGTCAGAAGCCGGGC ACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAA CCTCAGTGTGACGCTCCCCGCCGCTCCACGCCATCC CGCATCTGACCACCAGCGAAATGGATTTTTGCATCGA GCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCA GGCTTTCTTTCACAGATGTGGATTGGCGATAAAAAAC AACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGC ACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACC CGCATTGACCCTAACGCCTGGGTCTGAACGCTGGAAGG CGGCGGGCCATTACCAGGCCGAAGCAGCGTTGTTGCA GTGCACGGCAGATACACTTGCTGATGCGGTGCTGATT ACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCT TATTTATCAGCCGGAAAACCTACCGGATTGATGGTAG TGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCG AGCGATACACCGCATCCGGCGCGGATTGGCCTGAACT GCCAG
22	Sequence of the 5'-Region used for knock out of PpOCH1:	AAAACCTTTTTTCCTATTCAAACACAAGGCATTGCTTC AACACGTGTGCGTATCCTTAACACAGATACTCCATACT TCTAATAATGTGATAGACGAATACAAAGATGTTCACT CTGTGTTGTGTCTACAAGCATTTCTTATTCTGATTGGG GATATTCTAGTTACAGCACTAAACAACCTGGCGATACA AACTTAAATTAATAATCCGAATCTAGAAAATGAACT TTTGGATGGTCCGCCTGTTGGTTGGATAAATCAATACC GATTAAATGGATTCTATTCCAATGAGAGAGTAATCCA AGACACTCTGATGTCAATAATCATTGCTTGCAACAAC AAACCCGTCATCTAATCAAAGGGTTTGATGAGGCTTA CCTTCAATTGCAGATAAACTCATTGCTGTCCACTGCTG TATTATGTGAGAATATGGGTGATGAATCTGGTCTTCTC CACTCAGCTAACATGGCTGTTTGGGCAAAGGTGGTAC AATTATACGGAGATCAGGCAATAGTGAAATTGTTGAA TATGGCTACTGGACGATGCTTCAAGGATGTACGTCTA GTAGGAGCCGTGGGAAGATTGCTGGCAGAACCAGTTG GCACGTCGCAACAATCCCCAAGAAATGAAATAAGTGA

	AAACGTAACGTCAAAGACAGCAATGGAGTCAATATTG ATAACACCACTGGCAGAGCGGTTCGTACGTCGTTTTG GAGCCGATATGAGGCTCAGCGTGCTAACAGCACGATT GACAAGAAGACTCTCGAGTGACAGTAGGTTGAGTAAA GTATTCGCTTAGATTCCCAACCTTCGTTTTATTCTTTG TAGACAAAGAAGCTGCATGCGAACATAGGGACAACCTT TTATAAATCCAATTGTCAAACCAACGTAAAACCCTCT GGCACCATTTC AACATATATTTGTGAAGCAGTACGC AATATCGATAAATACTCACCGTTGTTTGTAACAGCCCC AACTTGCATACGCCTTCTAATGACCTCAAATGGATAA GCCGCAGCTTGTGCTAACATACCAGCAGCACCGCCCCG CGGTCAGCTGCGCCACACATATAAAGGCAATCTACG ATCATGGGAGGAATTAGTTTTGACCGTCAGGTCTTCA AGAGTTTTGAACTCTTCTTCTTGAAGTGTGTAACCTTT TAAATGACGGGATCTAAATACGTCATGGATGAGATCA TGTGTGTAAAACTGACTCCAGCATATGGAATCATTC CAAAGATTGTAGGAGCGAACCCACGATAAAAGTTTCC CAACCTTGCCAAAGTGTCTAATGCTGTGACTTGAAATC TGGGTTCCCTCGTTGAAGACCCTGCGTACTATGCCCAA AACTTTCCTCCACGAGCCCTATTAACCTTCTCTATGAGT TTCAAATGCCAAACGGACACGGATTAGGTCCAATGGG TAAGTGAAAAACACAGAGCAAACCCACGCTAATGAG CCGGCCAGTAACCGTCTTGGAGCTGTTTCATAAGAGT CATTAGGGATCAATAACGTTCTAATCTGTTCATAACAT ACAAATTTTATGGCTGCATAGGGAAAAATTCTCAACA GGGTAGCCGAATGACCCTGATATAGACCTGCGACACC ATCATACCCATAGATCTGCCTGACAGCCTTAAAGAGC CCGCTAAAAGACCCGGAACCCGAGAGAACTCTGGAT TAGCAGTCTGAAAAAGAATCTTCACTCTGTCTAGTGG AGCAATTAATGTCTTAGCGGCACTTCCTGCTACTCCGC CAGCTACTCCTGAATAGATCACATACTGCAAAGACTG CTTGTCGATGACCTTGGGGTTATTTAGCTTCAAGGGCA ATTTTTGGGACATTTTGGACACAGGAGACTCAGAAAC AGACACAGAGCGTTCTGAGTCCTGGTGCTCCTGACGT AGGCCTAGAACAGGAATTATTGGCTTTATTTGTTTGTC CATTTCATAGGCTTGGGGTAATAGATAGATGACAGAG AAATAGAGAAGACCTAATATTTTTTGTTCATGGCAAAT
--	---

		CGCGGGTTCGCGGTCGGGTCACACACGGAGAAGTAAT GAGAAGAGCTGGTAATCTGGGGTAAAAGGGTTCAAAA GAAGGTCGCCTGGTAGGGATGCAATACAAGGTTGTCT TGGAGTTTACATTGACCAGATGATTTGGCTTTTTCTCT GTTCAATTCACATTTTTTCAGCGAGAATCGGATTGACGG AGAAATGGCGGGGTGTGGGGTGGATAGATGGCAGAA ATGCTCGCAATCACCGCGAAAGAAAGACTTTATGGAA TAGAACTACTGGGTGGTGTAAAGGATTACATAGCTAGT CCAATGGAGTCCGTTGGAAAGGTAAGAAGAAGCTAAA ACCGGCTAAGTAACTAGGGAAGAATGATCAGACTTTG ATTTGATGAGGTCTGAAAATACTCTGCTGCTTTTTTCAG TTGCTTTTTCCCTGCAACCTATCATTTTCCTTTTCATAA GCCTGCCTTTTTCTGTTTTCACTTATATGAGTTCCGCCG AGACTTCCCCAAATTCTCTCCTGGAACATTCTCTATCG CTCTCCTTCCAAGTTGCGCCCCCTGGCACTGCCTAGTA ATATTACCACGCGACTTATATTCAGTTCCACAATTTCC AGTGTTTCGTAGCAAATATCATCAGCCATGGCGAAGGC AGATGGCAGTTTGCTCTACTATAATCCTCACAATCCAC CCAGAAGGTATTACTTCTACATGGCTATATTCGCCGTT TCTGTCATTTGCGTTTTGTACGGACCCTCACAACAATT ATCATCTCCAAAAATAGACTATGATCCATTGACGCTCC GATCACTTGATTTGAAGACTTTGGAAGCTCCTTCACAG TTGAGTCCAGGCACCGTAGAAGATAATCTTCG
23	Sequence of the 3'-Region used for knock out of PpOCH1	AAAGCTAGAGTAAAATAGATATAGCGAGATTAGAGA ATGAATACCTTCTTCTAAGCGATCGTCCGTCATCATAG AATATCATGGACTGTATAGTTTTTTTTTTGTACATATA ATGATTAAACGGTCATCCAACATCTCGTTGACAGATCT CTCAGTACGCGAAATCCCTGACTATCAAAGCAAGAAC CGATGAAGAAAAAAACAACAGTAACCCAAACACCAC AACAAACACTTTATCTTCTCCCCCCCCAACACCAATCAT CAAAGAGATGTCGGAACCAACACCAAGAAGCAAAA ACTAACCCCATATAAAAACATCCTGGTAGATAATGCT GGTAACCCGCTCTCCTTCCATATTCTGGGCTACTTCAC GAAGTCTGACCGGTCTCAGTTGATCAACATGATCCTC GAAATGGGTGGCAAGATCGTTCCAGACCTGCCTCCTC TGGTAGATGGAGTGTTGTTTTTGACAGGGGATTACAA

		GTCTATTGATGAAGATACCCTAAAGCAACTGGGGGAC GTTCCAATATACAGAGACTCCTTCATCTACCAGTGTTT TGTGCACAAGACATCTCTTCCCATTGACACTTTCCGAA TTGACAAGAACGTCGACTTGGCTCAAGATTTGATCAA TAGGGCCCTTCAAGAGTCTGTGGATCATGTCACTTCTG CCAGCACAGCTGCAGCTGCTGCTGTTGTTGTCGCTACC AACGGCCTGTCTTCTAAACCAGACGCTCGTACTAGCA AAATACAGTTCACTCCCGAAGAAGATCGTTTTATTCTT GACTTTGTTAGGAGAAATCCTAAACGAAGAAACACAC ATCAACTGTACACTGAGCTCGCTCAGCACATGAAAAA CCATACGAATCATTCTATCCGCCACAGATTTTCGTCGTA ATCTTTCCGCTCAACTTGATTGGGTTTATGATATCGAT CCATTGACCAACCAACCTCGAAAAGATGAAAACGGGA ACTACATCAAGGTACAAGGCCTTCCA
24	<i>K. lactis</i> UDP- GlcNAc transporter gene (KIMNN2-2) ORF underlined	AAACGTAACGCCTGGCACTCTATTTTCTCAAACCTTCTG GGACGGAAGAGCTAAATATTGTGTTGCTTGAACAAAC CCAAAAAACAATAATGAACAACTAAACTACA CCTAAATAAACCGTGTGTAAACGTAAGTACCATATTA CTAGAAAAGATCACAAGTGTATCACACATGTGCATCT CATATTACATCTTTTATCCAATCCATTCTCTATCCCG TCTGTTCTGTCAGATTCTTTTCCATAAAAAGAAGAA GACCCCGAATCTCACCGGTACAATGCAAACTGCTGA AAAAAAAAGAAAGTTCACTGGATACGGGAACAGTGC CAGTAGGCTTCACCACATGGACAAAACAATTGACGAT AAAATAAGCAGGTGAGCTTCTTTTCAAGTCACGATC CCTTTATGTCTCAGAAACAATATATACAAGCTAAACC CTTTTGAACCAGTTCTCTCTTCATAGTTATGTTACAT AAATTGCGGGAACAAGACTCCGCTGGCTGTCAGGTAC ACGTTGTAACGTTTTTCGTCGCCCAATTATTAGCACA CATTGGCAAAAAGAAAAACTGCTCGTTTTCTCTACAG GTAAATTACAATTTTTTTCAGTAATTTTCGCTGAAAAA TTTAAAGGGCAGGAAAAAAGACGATCTCGACTTTGC ATAGATGCAAGAACTGTGGTCAAACTTGAATAGTA ATTTTGCTGTGCGTGAATAATAATATATATATAT ATATATATATATTTGTGTATTTTGTATATGTAATTGTGC ACGTCTTGGCTATTGGATATAAGATTTTCGCGGGTTGA TGACATAGAGCGTGTACTACTGTAATAGTTGTATATTC

		AAAAGCTGCTGCGTGGAGAAAGACTAAAATAGATAA AAAGCACACATTTTGACTTCGGTACCGTCAACTTAGTG GGACAGTCTTTTATATTTGGTGTAAGCTCATTCTGGT ACTATTCGAAACAGAACAGTGTTTTCTGTATTACCGTC CAATCGTTTGTCA <u>ATGAGTTTTGTATTGATTTTGTCGTT</u> <u>AGTGTTTCGGAGGATGTTGTTCCAATGTGATTAGTTTCG</u> <u>AGCACATGGTGCAAGGCAGCAATATAAATTTGGGAAA</u> <u>TATTGTTACATTCACTCAATTCGTGTCTGTGACGCTAA</u> <u>TTCAGTTGCCCAATGCTTTGGACTTCTCTCACTTTCCGT</u> <u>TTAGGTTGCGACCTAGACACATTCCTCTTAAGATCCAT</u> <u>ATGTTAGCTGTGTTTTTGTCTTTACCAGTTCAGTCGCC</u> <u>AATAACAGTGTGTTTAAATTTGACATTTCCGTTCCGAT</u> <u>TCATATTATCATTAGATTTTCAGGTACCACTTTGACGA</u> <u>TGATAATAGGTTGGGCTGTTTGTAATAAGAGGTACTCC</u> <u>AACTTCAGGTGCAATCTGCCATCATTATGACGCTTGG</u> <u>TGCGATTGTCGCATCATTATACCGTGACAAAGAATTTT</u> <u>CAATGGACAGTTTAAAGTTGAATACGGATTCAGTGGG</u> <u>TATGACCCAAAAATCTATGTTTGGTATCTTTGTTGTGC</u> <u>TAGTGGCCACTGCCTTGATGTCATTGTTGTCGTTGCTC</u> <u>AACGAATGGACGTATAACAAGTACGGGAAACATTGGA</u> <u>AAGAACTTTGTTCTATTTCGCATTTCTTGGCTCTACCG</u> <u>TTGTTTATGTTGGGGTACACAAGGCTCAGAGACGAAT</u> <u>TCAGAGACCTCTTAATTTCTCAGACTCAATGGATATT</u> <u>CCTATTGTAAATTACCAATTGCTACGAACTTTTCAT</u> <u>GCTAATAGCAAATAACGTGACCCAGTTCATTTGTATC</u> <u>AAAGGTGTAAACATGCTAGCTAGTAACACGGATGCTT</u> <u>TGACACTTTCTGTCGTGCTTCTAGTGCGTAAATTTGTT</u> <u>AGTCTTTTACTCAGTGTCTACATCTACAAGAACGTCCT</u> <u>ATCCGTGACTGCATACCTAGGGACCATCACCGTGTTCC</u> <u>TGGGAGCTGGTTTGTATTCAATATGGTTTCGGTCAAACT</u> <u>GCACTGCCTCGCTGA</u> AAACAATCCACGTCTGTATGATA CTCGTTTCAGAAATTTTTTTGATTTTCTGCCGGATATGGT TTCTCATCTTTACAATCGCATTCTTAATTATACCAGAA CGTAATTCAATGATCCCAGTGACTCGTAACCTCTTATAT GTCAATTTAAGC
25	Sequence of the 5'-Region used	GGCCGAGCGGGCCTAGATTTTCACTACAAATTTCAA ACTACGCGGATTTATTGTCTCAGAGAGCAATTTGGCAT

	for knock out of PpBMT2	TTCTGAGCGTAGCAGGAGGCTTCATAAGATTGTATAG GACCGTACCAACAAATTGCCGAGGCACAACACGGTAT GCTGTGCACTTATGTGGCTACTTCCCTACAACGGAATG AAACCTTCCTCTTTCCGCTTAAACGAGAAAGTGTGTGCG CAATTGAATGCAGGTGCCTGTGCGCCTTGGTGTATTGT TTTTGAGGGCCCAATTTATCAGGCGCCTTTTTTCTTGG TTGTTTTCCCTTAGCCTCAAGCAAGGTTGGTCTATTTTC ATCTCCGCTTCTATACCGTGCCTGATACTGTTGGATGA GAACACGACTCAACTTCCTGCTGCTCTGTATTGCCAGT GTTTTGTCTGTGATTTGGATCGGAGTCCTCCTTACTTG GAATGATAATAATCTTGGCGGAATCTCCCTAAACGGA GGCAAGGATTCTGCCTATGATGATCTGCTATCATTGGG AAGCTTCAACGACATGGAGGTCGACTCCTATGTCACC AACATCTACGACAATGCTCCAGTGCTAGGATGTACGG ATTTGTCTTATCATGGATTGTTGAAAGTCACCCCAAAG CATGACTTAGCTTGCATTTGGAGTTCATAAGAGCTCA GATTTTGGACATTGACGTTTACTCCGCCATAAAAGACT TAGAAGATAAAGCCTTGACTGTAAAACAAAAGGTTGA AAAACACTGGTTTACGTTTTATGGTAGTTCAGTCTTTC TGCCCGAACACGATGTGCATTACCTGGTTAGACGAGT CATCTTTTCGGCTGAAGGAAAGGCGAACTCTCCAGTA ACATC
26	Sequence of the 3'-Region used for knock out of PpBMT2	CCATATGATGGGTGTTTGCTCACTCGTATGGATCAAAA TTCCATGGTTTTCTTCTGTACAACCTTGTAACCTTATTTGG ACTTTTCTAACGGTTTTTCTGGTGATTTGAGAAGTCCT TATTTTGGTGTTTCGCAGCTTATCCGTGATTGAACCATC AGAAATACTGCAGCTCGTTATCTAGTTTCAGAATGTGT TGTAAGAATACAATCAATTCTGAGTCTAGTTTGGGTGGG TCTTGGCGACGGGACCGTTATATGCATCTATGCAGTGT TAAGGTACATAGAATGAAAATGTAGGGGTTAATCGAA AGCATCGTTAATTTCAGTAGAACGTAGTTCTATTCCCT ACCCAAATAATTTGCCAAGAATGCTTCGTATCCACAT ACGCAGTGGACGTAGCAAATTTCACTTTGGACTGTGA CCTCAAGTCGTTATCTTCTACTTGGACATTGATGGTCA TTACGTAATCCACAAAGAATTGGATAGCCTCTCGTTTT ATCTAGTGCACAGCCTAATAGCACTTAAGTAAGAGCA

		ATGGACAAATTTGCATAGACATTGAGCTAGATACGTA ACTCAGATCTTGTTCACTCATGGTGTACTCGAAGTACT GCTGGAACCGTTACCTCTTATCATTTTCGCTACTGGCTC GTGAAACTACTGGATGAAAAAAAAAAAAAGAGCTGAA AGCGAGATCATCCCATTTTGTTCATCATACAAATTCACG CTTGCAGTTTTGCTTCGTTAACAAGACAAGATGTCTTT ATCAAAGACCCGTTTTTTCTTCTTGAAGAATACTTCCC TGTGAGCACATGCAAACCATATTTATCTCAGATTTCA CTCAACTTGGGTGCTTCCAAGAGAAGTAAAATTCTTCC CACTGCATCAACTTCCAAGAAACCCGTAGACCAGTTT CTCTTCAGCCAAAAGAAGTTGCTCGCCGATCACCGCG GTAACAGAGGAGTCAGAAGGTTTCACACCCTTCCATC CCGATTTCAAAGTCAAAGTGCTGCGTTGAACCAAGGT TTTCAGGTTGCCAAAGCCCAGTCTGCAAAAAGTAGTT CCAAATGGCCTATTAATTCCCATAAAAAGTGTTGGCTAC GTATGTATCGGTACCTCCATTCTGGTATTTGCTATTGT TGTCGTTGGTGGGTTGACTAGACTGACCGAATCCGGT CTTTCCATAACGGAGTGGAACCTATCACTGGTTCCGGT TCCCCCACTGACTGAGGAAGACTGGAAGTTGGAATTT GAAAAATACAAACAAAGCCCTGAGTTTCAGGAAGTAA ATTCTCACATAACATTGGAAGAGTTCAAGTTTATATTT TCCATGGAATGGGGACATAGATTGTTGGGAAGGGTCA TCGGCCTGTCGTTTGTCTTCCCACGTTTTACTTCATTG CCCGTCGAAAGTGTTCCAAAGATGTTGCATTGAAACT GCTTGCAATATGCTCTATGATAGGATTCCAAGGTTTCA TCGGCTGGTGGATGGTGTATTCCGGATTGGACAAACA GCAATTGGCTGAACGTAACCTCAAACCAACTGTGTCT CCATATCGCTTAACTACCCATCTTGGAAGTGCATTTGT TATTTACTGTTACATGATTTACACAGGGCTTCAAGTTT TGAAGAACTATAAGATCATGAAACAGCCTGAAGCGTA TGTTCAAATTTTCAAGCAAATTGCGTCTCCAAAATTGA AACTTTCAAGAGACTCTCTTCAGTTCTATTAGGCCTG GTG
27	DNA encodes MmSLC35A3 UDP-GlcNAc transporter	ATGTCTGCCAACCTAAAATATCTTTCCTTGGAATTTT GGTGTTTCAGACTACCAGTCTGGTTCTAACGATGCGGT ATTCTAGGACTTTAAAAGAGGAGGGGCCTCGTTATCT GTCTTCTACAGCAGTGGTTGTGGCTGAATTTTGAAGA

		TAATGGCCTGCATCTTTTTAGTCTACAAAGACAGTAAG TG TAGTGTGAGAGCACTGAATAGAGTACTGCATGATG AAATTCTTAATAAGCCCATGGAAACCCTGAAGCTCGC TATCCCGTCAGGGATATATACTCTTCAGAACAACTTAC TCTATGTGGCACTGTCAAACCTAGATGCAGCCACTTAC CAGGTTACATATCAGTTGAAAATACTTACAACAGCAT TATTTTCTGTGTCTATGCTTGGTAAAAAATTAGGTGTG TACCAGTGGCTCTCCCTAGTAATTCTGATGGCAGGAGT TGCTTTTGTACAGTGGCCTTCAGATTCTCAAGAGCTGA ACTCTAAGGACCTTTCAACAGGCTCACAGTTTGTAGG CCTCATGGCAGTTCTCACAGCCTGTTTTTCAAGTGGCT TTGCTGGAGTTTATTTTGAGAAAATCTTAAAAGAAAC AAAACAGTCAGTATGGATAAGGAACATTCAACTGGT TTCTTTGGAAGTATATTTGGATTAATGGGTGTATACGT TTATGATGGAGAATTGGTCTCAAAGAATGGATTTTTTC AGGGATATAATCAACTGACGTGGATAGTTGTTGCTCT GCAGGCACTTGGAGGCCTTGTAATAGCTGCTGTCATC AAATATGCAGATAACATTTTAAAAGGATTTGCGACCT CCTTATCCATAATATTGTCAACAATAATATCTTATTTT TGGTTGCAAGATTTTGTGCCAACCAGTGTCTTTTTCCT TGGAGCCATCCTTGTAATAGCAGCTACTTTCTTGATG GTTACGATCCCAAACCTGCAGGAAATCCCACTAAAGC ATAG
28	Sequence of the 5'-Region used for knock out of PpMNN4L1	GATCTGGCCATTGTGAAACTTGACACTAAAGACAAAA CTCTTAGAGTTTCCAATCACTTAGGAGACGATGTTTCC TACAACGAGTACGATCCCTCATTGATCATGAGCAATTT GTATGTGAAAAAAGTCATCGACCTTGACACCTTGGAT AAAAGGGCTGGAGGAGGTGGAACCACCTGTGCAGGC GGTCTGAAAGTGTTCAAGTACGGATCTACTACCAAAT ATACATCTGGTAACCTGAACGGCGTCAGGTTAGTATA CTGGAACGAAGGAAAGTTGCAAAGCTCCAAATTTGTG GTTCGATCCTCTAATTACTCTCAAAGCTTGGAGGAA ACAGCAACGCCGAATCAATTGACAACAATGGTGTGGG TTTTGCCTCAGCTGGAGACTCAGGCGCATGGATTCTTT CCAAGCTACAAGATGTTAGGGAGTACCAGTCATTCAC TGAAAAGCTAGGTGAAGCTACGATGAGCATTTTCGAT TTCCACGGTCTTAAACAGGAGACTTCTACTACAGGGC

		TTGGGGTAGTTGGTATGATTTCATTCTTACGACGGTGAG TTCAAACAGTTTGGTTTGTTCCTCCAATGACATCTAT TCTACAAAGACTTCAACGAGTGACCAATGTAGAATGG TGTGTAGCGGGTTGCGAAGATGGGGATGTGGACACTG AAGGAGAACACGAATTGAGTGATTTGGAACAACCTGCA TATGCATAGTGATTCCGACTAGTCAGGCAAGAGAGAG CCCTCAAATTTACCTCTCTGCCCCCTCCTCACTCCTTTTG GTACGCATAATTGCAGTATAAAGAACTTGCTGCCAGC CAGTAATCTTATTTTCATACGCAGTTCTATATAGCACAT AATCTTGCTTGTATGTATGAAATTTACCGCGTTTTAGT TGAAATTGTTTATGTTGTGTGCCTTGCATGAAATCTCT CGTTAGCCCTATCCTTACATTTAACTGGTCTCAAAACC TCTACCAATTCCATTGCTGTACAACAATATGAGGCGG CATTACTGTAGGGTTGGAAAAAAATTGTCATTCCAGC TAGAGATCACACGACTTCATCACGCTTATTGCTCCTCA TTGCTAAATCATTTACTCTTGACTTCGACCCAGAAAAG TTCGCC
29	Sequence of the 3'-Region used for knock out of PpMNN4L1	GCATGTCAAACCTTGAACACAACGACTAGATAGTTGTT TTTTCTATATAAAACGAAACGTTATCATCTTTAATAAT CATTGAGGTTTACCCTTATAGTTCCGTATTTTCGTTTCC AACTTAGTAATCTTTTGGAAATATCATCAAAGCTGGT GCCAATCTTCTTGTGTTGAAGTTTCAAACCTGCTCCACCA AGCTACTTAGAGACTGTTCTAGGTCTGAAGCAACTTC GAACACAGAGACAGCTGCCGCCGATTGTTCTTTTTTTGT GTTTTTCTTCTGGAAGAGGGGCATCATCTTGTATGTCC AATGCCCGTATCCTTTCTGAGTTGTCCGACACATTGTC CTTCGAAGAGTTTCCTGACATTGGGCTTCTTCTATCCG TGTATTAATTTTGGGTAAAGTTCCTCGTTTGCATAGCA GTGGATACCTCGATTTTTTTGGCTCCTATTTACCTGAC ATAATATTCTACTATAATCCAACCTTGGACGCGTCATCT ATGATAACTAGGCTCTCCTTTGTTCAAAGGGGACGTCT TCATAATCCACTGGCACGAAGTAAGTCTGCAACGAGG CGGCTTTTGCAACAGAACGATAGTGTCGTTTCGTA GGACTATGCTAAACAAAAGGATCTGTCAAACATTTCA ACCGTGTTTCAAGGCACTCTTTACGAATTATCGACCAA GACCTTCCTAGACGAACATTTCAACATATCCAGGCTA CTGCTTCAAGGTGGTGCAAATGATAAAGGTATAGATA

		TTAGATGTGTTTGGGACCTAAAACAGTTCTTGCCTGAA GATTCCCTTGAGCAACAGGCTTCAATAGCCAAGTTAG AGAAGCAGTACCAAATCGGTAACAAAAGGGGGAAGC ATATAAAACCTTTACTATTGCGACAAAATCCATCCTTG AAAGTAAAGCTGTTTGTTCATGTAAAGCATACGAAA CGAAGGAGGTAGATCCTAAGATGGTTAGAGAACTTAA CGGGACATACTCCAGCTGCATCCCATATTACGATCGCT GGAAGACTTTTTTCATGTACGTATCGCCACCAACCTT TCAAAGCAAGCTAGGTATGATTTTGACAGTTCTCACA ATCCATTGGTTTTTCATGCAACTTGAAAAAACCCACTC AACTTCATGGGGATCCATACAATGTAAATCATTACG AGAGGGCGAGGTTGAAAAGTTTCCATTGCAATCACGT CGCATCATGGCTACTGAAAGGCCTTAAC
30	Sequence of the 5'-Region used for knock out of PpPNO1 and PpMNN4	TCATTCTATATGTTCAAGAAAAGGGTAGTGAAAGGAA AGAAAAGGCATATAGGCGAGGGAGAGTTAGCTAGCA TACAAGATAATGAAGGATCAATAGCGGTAGTTAAAGT GCACAAGAAAAGAGCACCTGTTGAGGCTGATGATAAA GCTCCAATTACATTGCCACAGAGAAACACAGTAACAG AAATAGGAGGGGATGCACCACGAGAAGAGCATTTCAG TGAACAACCTTTGCCAAATTCATAACCCCAAGCGCTAA TAAGCCAATGTCAAAGTCGGCTACTAACATTAATAGT ACAACAACCTATCGATTTTCAACCAGATGTTTGCAAGG ACTACAAACAGACAGGTTACTGCGGATATGGTGACAC TTGTAAGTTTTTGCACCTGAGGGATGATTTCAAACAGG GATGGAAATTAGATAGGGAGTGGGAAAATGTCCAAA AGAAGAAGCATAATACTCTCAAAGGGGTTAAGGAGAT CCAAATGTTTAATGAAGATGAGCTCAAAGATATCCCG TTTAAATGCATTATATGCAAAGGAGATTACAAATCAC CCGTGAAAACCTTCTTGCAATCATTATTTTTGCGAACAA TGTTTCCTGCAACGGTCAAGAAGAAAACCAAATTGTA TTATATGTGGCAGAGACACTTAGGAGTTGCTTTACCA GCAAAGAAGTTGTCCCAATTTCTGGCTAAGATACATA ATAATGAAAGTAATAAAGTTTAGTAATTGCATTGCGTT GACTATTGATTGCATTGATGTCGTGTGATACTTTCACC GAAAAAAAACACGAAGCGCAATAGGAGCGGTTGCAT ATTAGTCCCCAAAGCTATTTAATTGTGCCTGAAACTGT TTTTTAAGCTCATCAAGCATAATTGTATGCATTGCGAC

		GTAACCAACGTTTAGGCGCAGTTTAATCATAGCCCAC TGCTAAGCC
31	Sequence of the 3'-Region used for knock out of PpPNO1 and PpMNN4	CGGAGGAATGCAAATAATAATCTCCTTAATTACCCAC TGATAAGCTCAAGAGACGCGGTTTGAAAACGATATAA TGAATCATTGGATTTTATAATAAACCCCTGACAGTTTT TCCACTGTATTGTTTTAACACTCATTGGAAGCTGTATT GATTCTAAGAAGCTAGAAATCAATACGGCCATACAAA AGATGACATTGAATAAGCACCGGCTTTTTTGATTAGC ATATACCTTAAAGCATGCATTCATGGCTACATAGTTGT TAAAGGGCTTCTTCCATTATCAGTATAATGAATTACAT AATCATGCACTTATATTTGCCCATCTCTGTTCTCTCACT CTTGCCCTGGGTATATTCTATGAAATTGCGTATAGCGTG TCTCCAGTTGAACCCCAAGCTTGGCGAGTTTGAAGAG AATGCTAACCTTGCGTATTCCTTGCTTCAGGAAACATT CAAGGAGAAACAGGTCAAGAAGCCAAACATTTTGATC CTTCCCGAGTTAGCATTGACTGGCTACAATTTTCAAAG CCAGCAGCGGATAGAGCCTTTTTTGGAGGAAACAACC AAGGGAGCTAGTACCCAATGGGCTCAAAAAGTATCCA AGACGTGGGATTGCTTTACTTTAATAGGATACCCAGA AAAAAGTTTAGAGAGCCCTCCCCGTATTTACAACAGT GCGGTACTTGATCGCCTCAGGGAAAAGTAATGAACA ACTACAGAAAGTCCTTCTTGATGAAGCTGATGAACA TTGGGGATGTTTCGGAATCTTCTGATGGGTTTCAAACAG TAGATTTATTAATTGAAGGAAAGACTGTAAAGACATC ATTTGGAATTTGCATGGATTTGAATCCTTATAAATTTG AAGCTCCATTACAGACTTCGAGTTCAGTGGCCATTGC TTGAAAACCGGTACAAGACTCATTTTGTGCCCAATGG CCTGGTTGTCCCCTCTATCGCCTTCCATTAAAAAGGAT CTTAGTGATATAGAGAAAAGCAGACTTCAAAAGTTCT ACCTTGAAAAAATAGATACCCCGGAATTTGACGTTAA TTACGAATTGAAAAAAGATGAAGTATTGCCACCCGT ATGAATGAAACGTTGGAAACAATTGACTTTGAGCCTT CAAAACCGGACTACTCTAATATAAATTATTGGATACT AAGGTTTTTCCCTTTCTGACTCATGTCTATAAACGAG ATGTGCTCAAAGAGAATGCAGTTGCAGTCTTATGCAA CCGAGTTGGCATTGAGAGTGATGTCTTGTACGGAGGA

		TCAACCACGATTCTAAACTTCAATGGTAAGTTAGCATC GACACAAGAGGAGCTGGAGTTGTACGGGCAGACTAAT AGTCTCAACCCCAGTGTGGAAGTATTGGGGGCCCTTG GCATGGGTCAACAGGGAATTCTAGTACGAGACATTGA ATTAACATAATATACAATATACAATAAACACAAATAA AGAATACAAGCCTGACAAAAATTACAAATTATTGCC TAGACTTGTCGTTATCAGCAGCGACCTTTTTCCAATGC TCAATTTACGATATGCCTTTTCTAGCTCTGCTTAAAG CTTCTCATTGGAATTGGCTAACTCGTTGACTGCTTGGT CAGTGATGAGTTTCTCCAAGGTCCATTTCTCGATGTTG TTGTTTTCGTTTTCTTTAATCTCTTGATATAATCAACA GCCTTCTTTAATATCTGAGCCTTGTTGAGTCCCCTGT TGGCAACAGAGCGGCCAGTTCCTTTATTCCGTGGTTTA TATTTTCTCTTCTACGCCTTTCTACTTCTTTGTGATTCT CTTTACGCATCTTATGCCATTCTTCAGAACCAGTGGCT GGCTTAACCGAATAGCCAGAGCCTGAAGAAGCCGCAC TAGAAGAAGCAGTGGCATTGTTGACTATGG
32	DNA encodes human GnTI catalytic domain (NA) Codon- optimized	TCAGTCAGTGCTCTTGATGGTGACCCAGCAAGTTTGAC CAGAGAAGTGATTAGATTGGCCCAAGACGCAGAGGTG GAGTTGGAGAGACAACGTGGACTGCTGCAGCAAATCG GAGATGCATTGTCTAGTCAAAGAGGTAGGGTGCCTAC CGCAGCTCCTCCAGCACAGCCTAGAGTGCATGTGACC CCTGCACCAGCTGTGATTCTATCTTGGTCATCGCCTG TGACAGATCTACTGTTAGAAGATGTCTGGACAAGCTG TTGCATTACAGACCATCTGCTGAGTTGTTCCCTATCAT CGTTAGTCAAGACTGTGGTCACGAGGAGACTGCCCAA GCCATCGCCTCCTACGGATCTGCTGTCACTCACATCAG ACAGCCTGACCTGTCATCTATTGCTGTGCCACCAGACC ACAGAAAGTTCCAAGGTTACTACAAGATCGCTAGACA CTACAGATGGGCATTGGGTCAAGTCTTCAGACAGTTT AGATTCCCTGCTGCTGTGGTGGTGGAGGATGACTTGG AGGTGGCTCCTGACTTCTTTGAGTACTTTAGAGCAACC TATCCATTGCTGAAGGCAGACCCATCCCTGTGGTGTGT CTCTGCCTGGAATGACAACGGTAAGGAGCAAATGGTG GACGCTTCTAGGCCTGAGCTGTTGTACAGAACCGACT TCTTTCCTGGTCTGGGATGGTTGCTGTTGGCTGAGTTG TGGGCTGAGTTGGAGCCTAAGTGGCCAAAGGCATTCT

		GGGACGACTGGATGAGAAGACCTGAGCAAAGACAGG GTAGAGCCTGTATCAGACCTGAGATCTCAAGAACCAT GACCTTTGGTAGAAAGGGAGTGTCTCACGGTCAATTC TTTGACCAACACTTGAAGTTTATCAAGCTGAACCAGC AATTTGTGCACTTCACCCAACTGGACCTGTCTTACTTG CAGAGAGAGGCCTATGACAGAGATTTCTAGCTAGAG TCTACGGAGCTCCTCAACTGCAAGTGGAGAAAGTGAG GACCAATGACAGAAAGGAGTTGGGAGAGGTGAGAGT GCAGTACACTGGTAGGGACTCCTTTAAGGCTTTCGCTA AGGCTCTGGGTGTCATGGATGACCTTAAGTCTGGAGT TCCTAGAGCTGGTTACAGAGGTATTGTCACCTTTCAAT TCAGAGGTAGAAGAGTCCACTTGGCTCCTCCACCTAC TTGGGAGGGTTATGATCCTTCTTGGAATTAG
33	DNA encodes Pp SEC12 (10) The last 9 nucleotides are the linker containing the AscI restriction site used for fusion to proteins of interest.	ATGCCCAGAAAAATATTTAACTACTTCATTTTGACTGT ATTCATGGCAATTCTTGCTATTGTTTTACAATGGTCTA TAGAGAATGGACATGGGCGCGCC
34	Sequence of the PpSEC4 promoter	GAAGTAAAGTTGGCGAACTTTGGGAACCTTTGGTTA AACTTTGTAAATTTTGTGCTACCCATTAGGCAGAAT CTGCATCTTGGGAGGGGGATGTGGTGGCGTTCTGAGA TGTACGCGAAGAATGAAGAGCCAGTGGTAACAACAG GCCTAGAGAGATACGGGCATAATGGGTATAACCTACA AGTTAAGAATGTAGCAGCCCTGGAAACCAGATTGAAA CGAAAAACGAAATCATTTAACTGTAGGATGTTTTGG CTCATTGTCTGGAAGGCTGGCTGTTTATTGCCCTGTTC TTTGCATGGGAATAAGCTATTATATCCCTCACATAATC CCAGAAAATAGATTGAAGCAACGCGAAATCCTTACGT ATCGAAGTAGCCTTCTTACACATTCACGTTGTACGGAT AAGAAAACACTCAAACGAACAATC

35	Sequence of the PpOCH1 terminator	AATAGATATAGCGAGATTAGAGAATGAATACCTTCTT CTAAGCGATCGTCCGTCATCATAGAATATCATGGACT GTATAGTTTTTTTTTTGTACATATAATGATTAAACGGT CATCCAACATCTCGTTGACAGATCTCTCAGTACGCGA AATCCCTGACTATCAAAGCAAGAACCGATGAAGAAAA AAACAACAGTAACCCAAACACCACAACAAACACTTTA TCTTCTCCCCCCCCAACACCAATCATCAAAGAGATGTCG GAACACAAACACCAAGAAGCAAAAATAACCCCATATA TAAAAACATCCTGGTAGATAATGCTGGTAACCCGCTC TCCTTCCATATTCTGGGCTACTTCACGAAGTCTGACCG GTCTCAGTTGATCAACATGATCCTCGAAATGG
36	DNA encodes Mm ManI catalytic domain (FB)	GAGCCCGCTGACGCCACCATCCGTGAGAAGAGGGCAA AGATCAAAGAGATGATGACCCATGCTTGAATAATTA TAAACGCTATGCGTGGGGCTTGAACGAACTGAAACCT ATATCAAAAGAAGGCCATTCAAGCAGTTTGTGGCA ACATCAAAGGAGCTACAATAGTAGATGCCCTGGATAC CCTTTTCATTATGGGCATGAAGACTGAATTTCAAGAA GCTAAATCGTGGATTAAAAAATATTTAGATTTTAATGT GAATGCTGAAGTTTCTGTTTTTGAAGTCAACATACGCT TCGTCGGTGGACTGCTGTCAGCCTACTATTTGTCCGGA GAGGAGATATTTGAAAGAAAGCAGTGGAACCTGGGG TAAAATTGCTACCTGCATTTTCACTCCCTCTGGAATA CCTTGGGCATTGCTGAATATGAAAAGTGGGATCGGGC GGAACCTGGCCCTGGGCCTCTGGAGGCAGCAGTATCCT GGCCGAATTTGGAACCTCTGCATTTAGAGTTTATGCACT TGTCCCACTTATCAGGAGACCCAGTCTTTGCCGAAAA GGTTATGAAAATTCGAACAGTGTTGAACAACTGGAC AAACCAGAAGGCCTTTATCCTAACTATCTGAACCCCA GTAGTGGACAGTGGGGTCAACATCATGTGTCGGTTGG AGGACTTGGAGACAGCTTTTATGAATATTTGCTTAAGG CGTGGTTAATGTCTGACAAGACAGATCTCGAAGCCAA GAAGATGTATTTTGATGCTGTTTCAGGCCATCGAGACTC ACTTGATCCGCAAGTCAAGTGGGGGACTAACGTACAT CGCAGAGTGGAAGGGGGGGCCTCCTGGAACACAAGAT GGGCCACCTGACGTGCTTTGCAGGAGGCATGTTTGCA CTTGGGGCAGATGGAGCTCCGGAAGCCCGGGCCCAAC ACTACCTTGAACCTCGGAGCTGAAATTGCCCGCACTTGT

		CATGAATCTTATAATCGTACATATGTGAAGTTGGGAC CGGAAGCGTTTCGATTTGATGGCGGTGTGGAAGCTAT TGCCACGAGGCAAAATGAAAAGTATTACATCTTACGG CCGGAGGTCATCGAGACATACATGTACATGTGGCGAC TGA CTCACGACCCCAAGTACAGGACCTGGGCCTGGGA AGCCGTGGAGGCTCTAGAAAGTCACTGCAGAGTGAAC GGAGGCTACTCAGGCTTACGGGATGTTTACATTGCCC GTGAGAGTTATGACGATGTCCAGCAAAGTTTCTTCCTG GCAGAGACACTGAAGTATTTGTA CTTGATATTTTCCGA TGATGACCTTCTTCCACTAGAACACTGGATCTTCAACA CCGAGGCTCATCCTTTCCCTATACTCCGTGAACAGAAG AAGGAAATTGATGGCAAAGAGAAATGA
37	DNA encodes ScSEC12 (8) The last 9 nucleotides are the linker containing the AscI restriction site used for fusion to proteins of interest	ATGAACACTATCCACATAATAAAATTACCGCTTA ACT ACGCCA ACTACACCTCAATGAAACAAAAAATCTCTAA ATTTTTCACCAACTTCATCCTTATTGTGCTGCTTTCTTA CATTTTACAGTTCTCCTATAAGCACAAATTTGCATTCCA TGCTTTTCAATTACGCGAAGGACAATTTTCTAACGAAA AGAGACACCATCTCTTCGCCCTACGTAGTTGATGAAG ACTTACATCAAACA ACTTTGTTTGGCAACCACGGTAC AAAAACATCTGTACCTAGCGTAGATTCCATAAAAGTG CATGGCGTGGGGCGCGCC
38	Sequence of the 5'-region that was used to knock into the PpADE1 locus	GAGTCGGCCAAGAGATGATAACTGTTACTAAGCTTCT CCGTAATTAGTGGTATTTTGTA ACTTTTACCAATAATC GTTTATGAATACGGATATTTTTCGACCTTATCCAGTGC CAAATCACGTA ACTTAATCATGGTTTAAATACTCCACT TGAACGATTCA TTATTCAGAAAAAAGTCAGGTTGGCA GAAACACTTGGGCGCTTTGAAGAGTATAAGAGTATTA AGCATTA AACATCTGA ACTTTACCGCCCCAATATACT ACTCTAGGAA ACTCGAAAAATTCCTTTCCATGTGTCAT CGCTTCCAACACACTTTGCTGTATCCTTCCAAGTATG T CCATTGTGAACACTGATCTGGACGGAATCCTACCTTTA ATCGCCAAAGGAAAGGTTAGAGACATTTATGCAGTCG ATGAGAACA ACTTGCTGTTTCGTCGCAACTGACCGTAT CTCCGCTTACGATGTGATTATGACAAACGGTATTCTCTG ATAAGGGAAAGATTTTGACTCAGCTCTCAGTTTCTGG

		TTTGATTTTTTGGCACCCCTACATAAAGAATCATTTGGT TGCTTCTAATGACAAGGAAGTCTTTGCTTTACTACCAT CAAAACTGTCTGAAGAAAAaTACAAATCTCAATTAGA GGGACGATCCTTGATAGTAAAAAAGCACAGACTGATA CCTTTGGAAGCCATTGTCAGAGGTTACATCACTGGAA GTGCATGGAAAGAGTACAAGAACTCAAAAACCTGTCCA TGGAGTCAAGGTTGAAAACGAGAACCTTCAAGAGAGC GACGCCTTTCCAACCTCCGATTTTTCACACCTTCAACGAA AGCTGAACAGGGTGAACACGATGAAAACATCTCTATT GAACAAGCTGCTGAGATTGTAGGTAAAGACATTTGTG AGAAGGTCGCTGTCAAGGCGGTCGAGTTGTATTCTGC TGCAAAAAACCTCGCCCTTTTGAAGGGGATCATTATT GCTGATACGAAATTCGAATTTGGACTGGACGAAAACA ATGAATTGGTACTAGTAGATGAAGTTTTAACTCCAGAT TCTTCTAGATTTTGGAAATCAAAGACTTACCAAGTGG GTAAATCGCAAGAGAGTTACGATAAGCAGTTTCTCAG AGATTGGTTGACGGCCAACGGATTGAATGGCAAAGAG GGCGTAGCCATGGATGCAGAAATTGCTATCAAGAGTA AAGAAAAGTATATTGAAGCTTATGAAGCAATTACTGG CAAGAAATGGGCTTGA
39	Sequence of the 3'-region that was used to knock into the PpADE1 locus	ATGATTAGTACCCTCCTCGCCTTTTTTCAGACATCTGAA ATTTCCCTTATTCTTCCAATTCCATATAAAATCCTATTT AGGTAATTAGTAAACAATGATCATAAAGTGAAATCAT TCAAGTAACCATTCCGTTTATCGTTGATTTAAAATCAA TAACGAATGAATGTCGGTCTGAGTAGTCAATTTGTTGC CTTGGAGCTCATTGGCAGGGGGTCTTTTGGCTCAGTAT GGAAGGTTGAAAGGAAAACAGATGGAAAGTGGTTTCG TCAGAAAAGAGGTATCCTACATGAAGATGAATGCCAA AGAGATATCTCAAGTGATAGCTGAGTTCAGAATTCTT AGTGAGTTAAGCCATCCCAACATTGTGAAGTACCTTC ATCACGAACATATTTCTGAGAATAAACTGTCAATTT ATACATGGAATACTGTGATGGTGGAGATCTCTCCAAG CTGATTCGAACACATAGAAGGAACAAAGAGTACATTT CAGAAGAAAAAATATGGAGTATTTTACGCAGGTTTT ATTAGCATTGTATCGTTGTCATTATGGAAGTATTTC CGGCTTCAAAGGAGTTTGAATCGCTCAATAAAGGTAA

		TAGACGAACCCAGAATCCTTCGTGGGTAGACTCGACA AGAGTTATTATTCACAGGGATATAAAACCCGACAACA TCTTTCTGATGAACAATTCAAACCTTGTCAAACCTGGGA GATTTTGGATTAGCAAAAATTCTGGACCAAGAAAACG ATTTTGCCAAAACATACGTCGGTACGCCGTATTACATG TCTCCTGAAGTGCTGTTGGACCAACCCTACTCACCATT ATGTGATATATGGTCTCTTGGGTGCGTCATGTATGAGC TATGTGCATTGAGGCCTCCTT
40	DNA encodes ScGAL10	ATGACAGCTCAGTTACAAAGTGAAAGTACTTCTAAAA TTGTTTTGGTTACAGGTGGTGCTGGATACATTGGTTCA CACTGTGGTAGAGCTAATTGAGAATGGATATGACT GTGTTGTTGCTGATAACCTGTCGAATTCAACTTATGAT TCTGTAGCCAGGTTAGAGGTCTTGACCAAGCATCACA TTCCCTTCTATGAGGTTGATTTGTGTGACCGAAAAGGT CTGGAAAAGGTTTTCAAAGAATATAAAATTGATTTCGG TAATTCACTTTGCTGGTTTAAAGGCTGTAGGTGAATCT ACACAAATCCCGCTGAGATACTATCACAATAACATTT TGGGAACTGTCGTTTTATTAGAGTTAATGCAACAATAC AACGTTTCCAAATTTGTTTTTTCATCTTCTGCTACTGTC TATGGTGATGCTACGAGATTCCCAAATATGATTCCCTAT CCCAGAAGAATGTCCCTTAGGGCCTACTAATCCGTAT GGTCATACGAAATACGCCATTGAGAATATCTTGAATG ATCTTTACAATAGCGACAAAAAAGTTGGAAGTTTGC TATCTTGCGTTATTTTAACCCAATTGGCGCACATCCCT CTGGATTAATCGGAGAAGATCCGCTAGGTATACCAA CAATTTGTTGCCATATATGGCTCAAGTAGCTGTTGGTA GCGCGAGAAGCTTTACATCTTCGGAGACGATTATGA TTCCAGAGATGGTACCCCGATCAGGGATTATATCCAC GTAGTTGATCTAGCAAAAGGTCATATTGCAGCCCTGC AATACCTAGAGGCCTACAATGAAAATGAAGGTTTGTG TCGTGAGTGGAACCTTGGGTTCGGTAAAGGTTCTACA GTTTTTGAAGTTTATCATGCATTCTGCAAAGCTTCTGG TATTGATCTTCCATACAAAGTTACGGGCAGAAGAGCA GGTGATGTTTTGAACCTTGACGGCTAAACCAGATAGGG CCAAACGCGAACTGAAATGGCAGACCGAGTTGCAGGT TGAAGACTCCTGCAAGGATTTATGGAAATGGACTACT GAGAATCCTTTTGGTTACCAGTTAAGGGGTGTCGAGG

		CCAGATTTTCCGCTGAAGATATGCGTTATGACGCAAG ATTTGTGACTATTGGTGCCGGCACCAGATTTCAAGCCA CGTTTGCCAATTTGGGCGCCAGCATTGTTGACCTGAAA GTGAACGGACAATCAGTTGTTCTTGGCTATGAAAATG AGGAAGGGTATTTGAATCCTGATAGTGCTTATATAGG CGCCACGATCGGCAGGTATGCTAATCGTATTTCTGAAG GGTAAGTTTAGTTTATGCAACAAAGACTATCAGTTAA CCGTTAATAACGGCGTTAATGCGAATCATAGTAGTAT CGGTTCTTTCCACAGAAAAAGATTTTTGGGACCCATCA TTCAAAATCCTTCAAAGGATGTTTTTACCGCCGAGTAC ATGCTGATAGATAATGAGAAGGACACCGAATTTCCAG GTGATCTATTGGTAACCATAACAGTATACTGTGAACGTT GCCCAAAAAAGTTTGGAAATGGTATATAAAGGTAAAT TGACTGCTGGTGAAGCGACGCCAATAAATTTAACAAA TCATAGTTATTTCAATCTGAACAAGCCATATGGAGAC ACTATTGAGGGTACGGAGATTATGGTGCGTTCAAAAA AATCTGTTGATGTCGACAAAAACATGATTCCTACGGG TAATATCGTCGATAGAGAAATTGCTACCTTTAACTCTA CAAAGCCAACGGTCTTAGGCCCAAAAAATCCCCAGTT TGATTGTTGTTTTGTGGTGGATGAAAATGCTAAGCCAA GTCAAATCAATACTCTAAACAATGAATTGACGCTTATT GTCAAGGCTTTTCATCCCGATTCCAATATTACATTAGA AGTTTTAAGTACAGAGCCAACCTTATCAATTTTATACCG GTGATTTCTTGTCTGCTGGTTACGAAGCAAGACAAGG TTTTGCAATTGAGCCTGGTAGATACATTGATGCTATCA ATCAAGAGAACTGGAAAGATTGTGTAACCTTGAAAAA CGGTGAAACTTACGGGTCCAAGATTGTCTACAGATTTT CCTGA
41	Sequence of the PpPMA1 terminator	TAAGCTTCACGATTTGTGTTCCAGTTTATCCCCCTTT ATATACCGTTAACCCTTTCCCTGTTGAGCTGACTGTTG TTGTATTACCGCAATTTTCCAAGTTTGCCATGCTTTTC GTGTTATTTGACCGATGTCTTTTTTCCCAAATCAAAC ATATTTGTTACCATTTAAACCAAGTTATCTTTTGTATT AAGAGTCTAAGTTTGTTCACGAGGCTTCATGTGAGAGT GATAACCATCCAGACTATGATTCTTGTTTTTTATTGGG TTTGTGTTGTGTGATACATCTGAGTTGTGATTCGTAAAG TATGTCAGTCTATCTAGATTTTAAATAGTTAATTGGTA

		ATCAATGACTTGTTTGTTTTAACTTTTAAATTGTGGGT CGTATCCACGCGTTTAGTATAGCTGTTTCATGGCTGTTA GAGGAGGGCGATGTTTATATACAGAGGACAAGAATGA GGAGGCGGCGTGTATTTTAAATGGAGACGCGACTC CTGTACACCTTATCGGTTGG
42	hGalT codon optimized (XB)	GGTAGAGATTTGTCTAGATTGCCACAGTTGGTTGGTGT TTCCACTCCATTGCAAGGAGGTTCTAACTCTGCTGCTG CTATTGGTCAATCTTCCGGTGAGTTGAGAACTGGTGG AGCTAGACCACCTCCACCATTGGGAGCTTCCTCTCAAC CAAGACCAGGTGGTGATTCTTCTCCAGTTGTTGACTCT GGTCCAGGTCCAGCTTCTAACTTGACTTCCGTTCCAGT TCCACACACTACTGCTTTGTCCTTGCCAGCTTGTCCAG AAGAATCCCCATTGTTGGTTGGTCCAATGTTGATCGAG TTCAACATGCCAGTTGACTTGGAGTTGGTTGCTAAGCA GAACCCAAACGTTAAGATGGGTGGTAGATACGCTCCA AGAGACTGTGTTTCCCCACACAAAGTTGCTATCATCAT CCCATTCAGAAACAGACAGGAGCACTTGAAGTACTGG TTGTACTACTTGCACCCAGTTTTGCAAAGACAGCAGTT GGACTACGGTATCTACGTTATCAACCAGGCTGGTGAC ACTATTTTCAACAGAGCTAAGTTGTTGAATGTTGGTTT CCAGGAGGCTTTGAAGGATTACGACTACACTTGTTTC GTTTTCTCCGACGTTGACTTGATTCCAATGAACGACCA CAACGCTTACAGATGTTTCTCCCAGCCAAGACACATTT CTGTTGCTATGGACAAGTTCGGTTTTCTCCTTGCCATAC GTTCAATACTTCGGTGGTGTTTCCGCTTTGTCCAAGCA GCAGTTCTTGACTATCAACGGTTTCCCAAACAATTACT GGGGATGGGGTGGTGAAGATGACGACATCTTTAACAG ATTGGTTTTTCAGAGGAATGTCCATCTCTAGACCAAAC GCTGTTGTTGGTAGATGTAGAATGATCAGACACTCCA GAGACAAGAAGAACGAGCCAAACCCACAAAGATTCCG ACAGAATCGCTCACACTAAGGAAACTATGTTGTCCGA CGGATTGAACTCCTTGACTTACCAGGTTTTGGACGTTT AGAGATACCCATTGTACACTCAGATCACTGTTGACAT CGGTACTCCATCCTAG
43	DNA encodes ScMnt1 (Kre2) (33)	ATGGCCCTCTTCTCAGTAAGAGACTGTTGAGATTAC CGTCATTGCAGGTGCGGTTATTGTTCTCCTCCTAACAT TGAATTCCAACAGTAGAACTCAGCAATATATTCCGAG

		TTCCATCTCCGCTGCATTTGATTTTACCTCAGGATCTA TATCCCCTGAACAACAAGTCATCGGGCGCGCC
44	DNA encodes DmUGT	ATGAATAGCATACACATGAACGCCAATACGCTGAAGT ACATCAGCCTGCTGACGCTGACCCTGCAGAATGCCAT CCTGGGCCTCAGCATGCGCTACGCCCCGACCCGGCCA GGCGACATCTTCCTCAGCTCCACGGCCGTACTCATGGC AGAGTTCGCCAAACTGATCACGTGCCTGTTCTGCTGT TCAACGAGGAGGGCAAGGATGCCCAGAAGTTTGTACG CTCGCTGCACAAGACCATCATTGCGAATCCCATGGAC ACGCTGAAGGTGTGCGTCCCCTCGCTGGTCTATATCGT TCAAAACAATCTGCTGTACGTCTCTGCCTCCCATTTGG ATGCGGCCACCTACCAGGTGACGTACCAGCTGAAGAT TCTCACCACGGCCATGTTTCGCGGTTGTCATTCTGCGCC GCAAGCTGCTGAACACGCAGTGGGGTGCCTGCTGCT CCTGGTGATGGGCATCGTCCTGGTGCAAGTTGGCCCAA ACGGAGGGTCCGACGAGTGGCTCAGCCGGTGGTGCCG CAGCTGCAGCCACGGCCGCCTCCTCTGGCGGTGCTCC CGAGCAGAACAGGATGCTCGGACTGTGGGCCGCACTG GGCGCCTGCTTCCTCTCCGGATTCGCGGGCATCTACTT TGAGAAGATCCTCAAGGGTGCCGAGATCTCCGTGTGG ATGCGGAATGTGCAGTTGAGTCTGCTCAGCATTCCCTT CGGCCTGCTCACCTGTTTCGTTAACGACGGCAGTAGG ATCTTCGACCAGGGATTCTTCAAGGGCTACGATCTGTT TGTCTGGTACCTGGTCTGCTGCAGGCCGGCGGTGGA TTGATCGTTGCCGTGGTGGTCAAGTACGCGGATAACA TTCTCAAGGGCTTCGCCACCTCGCTGGCCATCATCATC TCGTGCGTGGCCTCCATATACATCTTCGACTTCAATCT CACGCTGCAGTTCAGCTTCGGAGCTGGCCTGGTCATC GCCTCCATATTTCTCTACGGCTACGATCCGGCCAGGTC GGCGCCGAAGCCAACCTATGCATGGTCCTGGCGGCGAT GAGGAGAAGCTGCTGCCGCGCGTCTAG
45	Sequence of the PpOCH1 promoter	TGGACACAGGAGACTCAGAAACAGACACAGAGCGTT CTGAGTCCTGGTGCTCCTGACGTAGGCCTAGAACAGG AATTATTGGCTTTATTTGTTTGTCCATTTCATAGGCTTG GGGTAATAGATAGATGACAGAGAAATAGAGAAGACC TAATATTTTTTGTTCATGGCAAATCGCGGGTTCGCGGT CGGGTCACACACGGAGAAGTAATGAGAAGAGCTGGT

		AATCTGGGGTAAAAGGGTTCAAAAGAAGGTCGCCTGG TAGGGATGCAATACAAGGTTGTCTTGGAGTTTACATTG ACCAGATGATTTGGCTTTTTCTCTGTTCAATTCACATTT TTCAGCGAGAATCGGATTGACGGAGAAATGGCGGGGT GTGGGGTGGATAGATGGCAGAAATGCTCGCAATCACC GCGAAAGAAAGACTTTATGGAATAGAACTACTGGGTG GTGTAAGGATTACATAGCTAGTCCAATGGAGTCCGTT GGAAAGGTAAGAAGAAGCTAAAACCGGCTAAGTAAC TAGGGAAGAATGATCAGACTTTGATTTGATGAGGTCT GAAAATACTCTGCTGCTTTTTTCAGTTGCTTTTTCCCTGC AACCTATCATTTTCCTTTTCATAAGCCTGCCTTTTCTGT TTTCACTTATATGAGTTCCGCCGAGACTTCCCCAAATT CTCTCCTGGAACATTCTCTATCGCTCTCCTTCCAAGTT GCGCCCCCTGGCACTGCCTAGTAATATTACCACGCGA CTTATATTCAGTTCCACAATTTCCAGTGTTCTGATGCAA ATATCATCAGCC
46	Sequence of the PpALG12 terminator	AATATATACCTCATTTGTTCAATTTGGTGTAAGAGTG TGCGGATAGACTTCTTGTAATCAGGAAAGCTACAA TTCCAATTGCTGCAAAAAATACCAATGCCCATAAACC AGTATGAGCGGTGCCTTCGACGGATTGCTTACTTTCCG ACCCTTTGTCGTTTGATTCTTCTGCCTTTGGTGAGTCA GTTTGTTTCGACTTTATATCTGACTCATCAACTTCCTTT ACGGTTGCGTTTTTAATCATAATTTTAGCCGTTGGCTT ATTATCCCTTGAGTTGGTAGGAGTTTGTATGATGCTG
47	Sequence of the 5'-Region used for knock out of PpHIS1	TAACTGGCCCTTTGACGTTTCTGACAATAGTTCTAGAG GAGTCGTCCAAAACTCAACTCTGACTTGGGTGACAC CACCACGGGATCCGGTTCTTCCGAGGACCTTGATGAC CTTGGCTAATGTAAGTGGAGTTTTAGTATCCATTTTAA GATGTGTGTTTCTGTAGGTTCTGGGTGGAAAAAATT TTAGACACCAGAAGAGAGGAGTGAAGTGGTTTGCGTG GGTTTAGACTGTGTAAGGCACTACTCTGTCGAAGTTTT AGATAGGGGTTACCCGCTCCGATGCATGGGAAGCGAT TAGCCCGGCTGTTGCCCGTTTGGTTTTTGAAGGGTAAT TTTCAATATCTCTGTTTGAGTCATCAATTTTCATATTCA AAGATTCAAAAACAAAATCTGGTCCAAGGAGCGCATT

		TAGGATTATGGAGTTGGCGAATCACTTGAACGATAGA CTATTATTTGC
48	Sequence of the 3'-Region used for knock out of PpHIS1	GTGACATTCTTGTCTTTGAGATCAGTAATTGTAGAGCA TAGATAGAATAATATTCAAGACCAACGGCTTCTCTTC GGAAGCTCCAAGTAGCTTATAGTGATGAGTACCGGCA TATATTTATAGGCTTAAAATTTTCGAGGGTTCACTATAT TCGTTTAGTGGAAGAGTTCCTTTCACTCTTGTATCT ATATTGTCAGCGTGGACTGTTTATAACTGTACCAACTT AGTTTCTTTCAACTCCAGGTAAAGAGACATAAATGTCC TTTGATGCTGACAATAATCAGTGGAATTCAAGGAAGG ACAATCCCGACCTCAATCTGTTCAATTAATGAAGAGTTC GAATCGTCCTTAAATCAAGCGCTAGACTCAATTGTCA ATGAGAACCCTTTCTTTGACCAAGAACTATAAATAG ATCGAATGACAAAGTTGGAAATGAGTCCATTAGCTTA CATGATATTGAGCAGGCAGACCAAAATAAACCGTCCT TTGAGAGCGATATTGATGGTTCGGCGCCGTTGATAAG AGACGACAAATTGCCAAAGAAACAAAGCTGGGGGCT GAGCAATTTTTTTTCAAGAAGAAATAGCATATGTTTAC CACTACATGAAAATGATTCAAGTGTTGTTAAGACCGA AAGATCTATTGCAGTGGGAACACCCCATCTTCAATAC TGCTTCAATGGAATCTCCAATGCCAAGTACAATGCATT TACCTTTTTCCCAGTCATCCTATACGAGCAATTCAAAT TTTTTTTCAATTTATACTTTACTTTAGTGGCTCTCTCTC AAGCGATACCGCAACTTCGCATTGGATATCTTCTTCG TATGTCGTCCCACTTTTGTGTTGTAATCATAGTGACCAT GTCAAAAGAGGGCGATGGATGATATTCAACGCCGAAGA AGGGATAGAGAACAGAACAAATGAACCATATGAGGTTTC TGTCCAGCCCATCACCAGTTTTGTCCAAAACTTAAAA TGTGGTCACTTGGTTCGATTGCATAAGGGAATGAGAG TGCCCGCAGATATGGTTCTTGTCCAGTCAAGCGAATCC ACCGGAGAGTCATTTATCAAGACAGATCAGCTGGATG GTGAGACTGATTGGAAGCTTCGGATTGTTTCTCCAGTT ACACAATCGTTACCAATGACTGAACTTCAAAATGTGCG CCATCACTGCAAGCGCACCTCAAAATCAATTCACTC CTTCTTGGAAAGATTGACCTACAATGGGCAATCATATG GTCTTACGATAGACAACACAATGTGGTGTAATACTGT

		ATTAGCTTCTGGTTCAGCAATTGGTTGTATAATTTACA CAGGTAAAGATACTCGACAATCGATGAACACAACCTCA GCCCAAACCTGAAAACGGGCTTGTTAGAACTGGAAATC AATAGTTTGTCCAAGATCTTATGTGTTTGTGTGTTTGC ATTATCTGTCATCTTAGTGCTATTCCAAGGAATAGCTG ATGATTGGTACGTCGATATCATGCGGTTTCTCATTCTA TTCTCCACTATTATCCCAGTGTCTCTGAGAGTTAACCT TGATCTTGGAAAGTCAGTCCATGCTCATCAAATAGAA ACTGATAGCTCAATACCTGAAACCGTTGTTAGAACTA GTACAATACCGGAAGACCTGGGAAGAATTGAATACCT ATTAAGTGACAAAACCTGGAACCTTACTCAAAATGAT ATGGAAATGAAAAAACTACACCTAGGAACAGTCTCTT ATGCTGGTGATACCATGGATATTATTTCTGATCATGTT AAAGGTCTTAATAACGCTAAAACATCGAGGAAAGATC TTGGTATGAGAATAAGAGATTTGGTTACAACCTCTGGC CATCTG
49	DNA encodes <i>Drosophila</i> <i>melanogaster</i> ManII codon- optimized (KD)	AGAGACGATCCAATTAGACCTCCATTGAAGGTTGCTA GATCCCCAAGACCAGGTCAATGTCAAGATGTTGTTCA GGACGTCCCAAACGTTGATGTCCAGATGTTGGAGTTG TACGATAGAATGTCCTTCAAGGACATTGATGGTGGTG TTTGGAAAGCAGGGTTGGAACATTAAGTACGATCCATT GAAGTACAACGCTCATCACAAGTTGAAGGTCTTCGTT GTCCACACTCCCACAACGATCCTGGTTGGATTCAGA CCTTCGAGGAATACTACCAGCACGACACCAAGCACAT CTTGTCCAACGCTTTGAGACATTTGCACGACAACCCA GAGATGAAGTTCATCTGGGCTGAAATCTCCTACTTCGC TAGATTCTACCACGATTTGGGTGAGAACAAGAAGTTG CAGATGAAGTCCATCGTCAAGAACGGTCAGTTGGAAT TCGTCACTGGTGGATGGGTCATGCCAGACGAGGCTAA CTCCCACTGGAGAAACGTTTTGTTGCAGTTGACCGAA GGTCAAACCTTGGTTGAAGCAATTCATGAACGTCACTC CAACTGCTTCCTGGGCTATCGATCCATTCGGACACTCT CCAACATATGCCATACATTTTGCAGAAGTCTGGTTTCAA GAATATGTTGATCCAGAGAACCCACTACTCCGTTAAG AAGGAGTTGGCTCAACAGAGACAGTTGGAGTTCTTGT GGAGACAGATCTGGGACAACAAAGGTGACACTGCTTT

	GTTCACCCACATGATGCCATTCTACTCTTACGACATTC CTCATACCTGTGGTCCAGATCCAAAGGTTTGTGTGTCAG TTCGATTTCAAAGAATGGGTTCCTTCGGTTTGTCTTG TCCATGGAAGGTTCCACCTAGAACTATCTCTGATCAA AATGTTGCTGCTAGATCCGATTTGTTGGTTGATCAGTG GAAGAAGAAGGCTGAGTTGTACAGAACCAACGTCTTG TTGATTCCATTGGGTGACGACTTCAGATTCAAGCAGA ACACCGAGTGGGATGTTTCAGAGAGTCAACTACGAAAG ATTGTTCGAACACATCAACTCTCAGGCTCACTTCAATG TCCAGGCTCAGTTCGGTACTTTGCAGGAATACTTCGAT GCTGTTCAACCAGGCTGAAAGAGCTGGACAAGCTGAGT TCCCAACCTTGTCTGGTGACTTCTTCACTTACGCTGAT AGATCTGATAACTACTGGTCTGGTTACTACACTTCCAG ACCATACCATAAGAGAATGGACAGAGTCTTGATGCAC TACGTTAGAGCTGCTGAAATGTTGTCCGCTTGGCACTC CTGGGACGGTATGGCTAGAATCGAGGAAAGATTGGAG CAGGCTAGAAGAGAGTTGTCCTTGTTCAGCACCACG ACGGTATTACTGGTACTGCTAAAACTCACGTTGTGCTC GACTACGAGCAAAGAATGCAGGAAGCTTTGAAAGCTT GTCAAATGGTCATGCAACAGTCTGTCTACAGATTGTTG ACTAAGCCATCCATCTACTCTCCAGACTTCTCCTTCTC CTACTTCACTTTGGACGACTCCAGATGGCCAGGTTCTG GTGTTGAGGACTCTAGAACTACCATCATCTTGGGTGA GGATATCTTGCCATCCAAGCATGTTGTCATGCACAAC ACCTTGCCACACTGGAGAGAGCAGTTGGTTGACTTCT ACGTCTCCTCTCCATTCGTTTCTGTTACCGACTTGGCT AACAATCCAGTTGAGGCTCAGGTTTCTCCAGTTTGGTC TTGGCACCACGACACTTTGACTAAGACTATCCACCCA CAAGGTTCCACCACCAAGTACAGAATCATCTTCAAGG CTAGAGTTCCACCAATGGGTTTGGCTACCTACGTTTGTG ACCATCTCCGATTCCAAGCCAGAGCACACCTCCTACG CTTCCAATTTGTTGCTTAGAAAGAACCCAACCTTCTTG CCATTGGGTCAATACCCAGAGGATGTCAAGTTCGGTG ATCCAAGAGAGATCTCCTTGAGAGTTGGTAACGGTCC AACCTTGGCTTTCTCTGAGCAGGGTTTGTGGAAGTCCA TTCAGTTGACTCAGGATTCTCCACATGTTCCAGTTCAC TTCAAGTTCTTGAAGTACGGTGTTAGATCTCATGGTGA
--	--

		TAGATCTGGTGCTTACTTGTTCTTGCCAAATGGTCCAG CTTCTCCAGTCGAGTTGGGTCAGCCAGTTGTCTTGGTC ACTAAGGGTAAATTGGAGTCTTCCGTTTCTGTTGGTTT GCCATCTGTCGTTCAACCAGACCATCATGAGAGGTGGT GCTCCAGAGATTAGAAATTTGGTCGATATTGGTTCTTT GGACAACACTGAGATCGTCATGAGATTGGAGACTCAT ATCGACTCTGGTGATATCTTCTACACTGATTTGAATGG ATTGCAATTCATCAAGAGGAGAAGATTGGACAAGTTG CCATTGCAGGCTAACTACTACCCAATTCCATCTGGTAT GTTCATTGAGGATGCTAATAACCAGATTGACTTTGTTGA CCGGTCAACCATTGGGTGGATCTTCTTTGGCTTCTGGT GAGTTGGAGATTATGCAAGATAGAAGATTGGCTTCTG ATGATGAAAGAGGTTTGGGTCAGGGTGTTTTGGACAA CAAGCCAGTTTTGCATATTTACAGATTGGTCTTGGAGA AGGTAAACAACGTGTGTCAGACCATCTAAGTTGCATCC AGCTGGTTACTTGACTTCTGCTGCTCACAAAGCTTCTC AGTCTTTGTTGGATCCATTGGACAAGTTCATCTTCGCT GAAAATGAGTGGATCGGTGCTCAGGGTCAATTCCGGTG GTGATCATCCATCTGCTAGAGAGGATTTGGATGTCTCT GTCATGAGAAGATTGACCAAGTCTTCTGCTAAAACCC AGAGAGTTGGTTACGTTTTGCACAGAACCAATTTGAT GCAATGTGGTACTCCAGAGGAGCATACTCAGAAGTTG GATGTCTGTCACTTGTTGCCAAATGTTGCTAGATGTGA GAGAACTACCTTGACTTTCTTGCAAGATTTGGAGCACT TGGATGGTATGGTTGCTCCAGAAGTTTGTCCAATGGA AACCGCTGCTTACGTCTCTTCTCACTCTTCTTGA
50	DNA encodes ScMNN2-s leader (53)	ATGCTGCTTACCAAAAGGTTTTCAAAGCTGTTCAAGCT GACGTTTCATAGTTTTGATATTGTGCGGGCTGTTTCGTCA TTACAAACAAATACATGGATGAGAACACGTCG
51	Sequence of the PpHIS1 auxotrophic marker	CAAGTTGCGTCCGGTATACGTAACGTCTCACGATGAT CAAAGATAATACTTAATCTTCATGGTCTACTGAATAAC TCATTTAAACAATTGACTAATTGTACATTATATTGAAC TTATGCATCCTATTAACGTAATCTTCTGGCTTCTCTCTC AGACTCCATCAGACACAGAATATCGTTCTCTCTAACTG GTCCTTTGACGTTTCTGACAATAGTTCTAGAGGAGTCG TCCAAAACTCAACTCTGACTTGGGTGACACCACCAC GGGATCCGGTTCTTCCGAGGACCTTGATGACCTTGGCT

	AATGTA ACTGGAGT TTTAGTATCCATTTTAAGATGTGT GTTTCTGTAGGTTCTGGGTTGGAAAAAATTTTAGACA CCAGAAGAGAGGAGTGA ACTGGTTTGCGTGGGTTTAG ACTGTGTAAGGCACTACTCTGTCGAAGTTT TAGATAG GGGTTACCCGCTCCGATGCATGGGAAGCGATTAGCCC GGCTGTTGCCCGTTTGGTTTTTGAAGGGTAATTTTCAA TATCTCTGTTTGAGTCATCAATTT CATATTCAAAGATT CAAAAACAAAATCTGGTCCAAGGAGCGCATTTAGGAT TATGGAGTTGGCGAATCACTTGAACGATAGACTATTA TTTGCTGTTCCCTAAAGAGGGCAGATTGTATGAGAAAT GCGTTGAATTACTTAGGGGATCAGATATTCAGTTTCGA AGATCCAGTAGATTGGATATAGCTTTGTGCACTAACCT GCCCCTGGCATTGGTTTTCTTCCAGCTGCTGACATTC CCACGTTTGTAGGAGAGGGTAAATGTGATTTGGGTAT AACTGGTATTGACCAGGTT CAGGAAAGTGACGTAGAT GTCATACCTTTATTAGACTTGAATTT CGGTAAGTGCAA GTTGCAGATTCAAGTTCCCGAGAATGGTGACTTGAAA GAACCTAAACAGCTAATTGGTAAAGAAATTGTTTCCT CCTTTACTAGCTTAACCACCAGGTACTTTGAACA ACTG GAAGGAGTTAAGCCTGGTGAGCCACTAAAGACAAAA ATCAAATATGTTGGAGGGTCTGTTGAGGCCTCTTGTGC CCTAGGAGTTGCCGATGCTATTGTGGATCTTGT TGAGA GTGGAGAAACCATGAAAGCGGCAGGGCTGATCGATAT TGAAACTGTTCTTTCTACTTCCGCTTACCTGATCTCTTC GAAGCATCCTCAACACCCAGAACTGATGGATACTATC AAGGAGAGAATTGAAGGTGTACTGACTGCTCAGAAGT ATGTCTTGTGTAATTACAACGCACCTAGAGGTAACCTT CCTCAGCTGCTAAAACTGACTCCAGGCAAGAGAGCTG CTACCGTTTCTCCATTAGATGAAGAAGATTGGGTGGG AGTGTCCCTCGATGGTAGAGAAGAAAGATGTTGGAAGA ATCATGGACGAATTAAAGAAACAAGGTGCCAGTGACA TTCTTGTCTTTGAGATCAGTAATTGTAGAGCATAGATA GAATAATATTCAAGACCAACGGCTTCTCTTCGGAAGC TCCAAGTAGCTTATAGTGATGAGTACCGGCATATATTT ATAGGCTTAAAATTT CGAGGGTTCACTATATTCGTTTA GTGGGAAGAGTTCCTTTCACTCTTGTTATCTATATTGT CAGCGTGGACTGTTTATAACTGTACCAACTTAGTTTCT
--	---

		TTCAACTCCAGGTTAAGAGACATAAAATGTCCTTTGATG C
52	DNA encodes Rat GnT II (TC) Codon- optimized	TCCTTGGTTTACCAATTGAACTTCGACCAGATGTTGAG AAACGTTGACAAGGACGGTACTTGGTCTCCTGGTGAG TTGGTTTTGGTTGTTTCAGGTTTACAACAGACCAGAGTA CTTGAGATTGTTGATCGACTCCTTGAGAAAGGCTCAA GGTATCAGAGAGGTTTTTGGTTATCTTCTCCCACGATTT CTGGTCTGCTGAGATCAACTCCTTGATCTCCTCCGTTG ACTTCTGTCCAGTTTTTGCAGGTTTTCTTCCCATTCTCCA TCCAATTGTACCCATCTGAGTTCCCAGGTTCTGATCCA AGAGACTGTCCAAGAGACTTGAAGAAGAACGCTGCTT TGAAGTTGGGTTGTATCAACGCTGAATACCCAGATTCT TTCGGTCACTACAGAGAGGCTAAGTTCTCCCAAATA AGCATCATTGGTGGTGGAAGTTGCACTTTGTTTGGGAG AGAGTTAAGGTTTTTGCAGGACTACACTGGATTGATCTT GTTCTTGGAGGAGGATCATTACTTGGCTCCAGACTTCT ACCACGTTTTTCAAGAAGATGTGGAAGTTGAAGCAACA AGAGTGTCCAGGTTGTGACGTTTTTGTCTTGGGAACTT ACACTACTATCAGATCCTTCTACGGTATCGCTGACAAG GTTGACGTTAAGACTTGGAAGTCCACTGAACACAACA TGGGATTGGCTTTGACTAGAGATGCTTACCAGAAGTT GATCGAGTGTACTGACACTTTCTGTACTTACGACGACT ACAACTGGGACTGGACTTTGCAGTACTTGACTTTGGCT TGTTTGCCAAAAGTTTTGGAAGGTTTTTGGTTCCACAGGC TCCAAGAATTTTCCACGCTGGTGACTGTGGAATGCAC CACAAGAAAACCTTGTAGACCATCCACTCAGTCCGCTC AAATTGAGTCCTTGTGTAACAACAACAAGCAGTACTT GTTCCCAGAGACTTTGGTTATCGGAGAGAAGTTTCCA ATGGCTGCTATTTCCCCACCAAGAAAGAATGGTGGAT GGGGTGATATTAGAGACCACGAGTTGTGTAAATCCTA CAGAAGATTGCAGTAG
53	DNA encodes ScMNN2 leader (54) The last 9 nucleotides are the linker	ATGCTGCTTACCAAAAGGTTTTCAAAGCTGTTCAAGCT GACGTTTCATAGTTTTGATATTGTGCGGGCTGTTTCGTCA TTACAAACAAATACATGGATGAGAACACGTCGGTCAA GGAGTACAAGGAGTACTTAGACAGATATGTCCAGAGT TACTCCAATAAGTATTCATCTTCCTCAGACGCCGCCAG CGCTGACGATTCAACCCCATGAGGGACAATGATGAG

	containing the AscI restriction site	GCAGGCAATGAAAAGTTGAAAAGCTTCTACAACAACG TTTTCAACTTTCTAATGGTTGATTCGCCCCGGGCGCGCC
54	Sequence of the 5'-Region used for knock out of PpARG1	GATCTGGCCTTCCCTGAATTTTTACGTCCAGCTATACG ATCCGTTGTGACTGTATTTCTGAAATGAAGTTTCAAC CTAAAGTTTTGGTTGTACTTGCTCCACCTACCACGGAA ACTAATATCGAAACCAATGAAAAAGTAGAACTGGAAT CGTCAATCGAAATTCGCAACCAAGTGGAACCCAAAGA CTTGAATCTTTCTAAAGTCTATTCTAGTGACACTAATG GCAACAGAAGATTTGAGCTGACTTTTCAAATGAATCT CAATAATGCAATATCAACATCAGACAATCAATGGGCT TTGTCTAGTGACACAGGATCAATTATAGTAGTGTCTTC TGCAGGAAGAATAACTTCCCCGATCCTAGAAGTCGGG GCATCCGTCTGTGTCTTAAGATCGTACAACGAACACCT TTTGGCAATAACTTGTGAAGGAACATGCTTTTCATGGA ATTTAAAGAAGCAAGAATGTGTTCTAAACAGCATTTC ATTAGCACCTATAGTCAATTCACACATGCTAGTTAAG AAAGTTGGAGATGCAAGGAACTATTCTATTGTATCTG CCGAAGGAGACAACAATCCGTTACCCCAGATTCTAGA CTGCGAACTTTCCAAAAATGGCGCTCCAATTGTGGCTC TTAGCACGAAAGACATCTACTCTTATTCAAAGAAAAT GAAATGCTGGATCCATTTGATTGATTGAAATACTTTG AATTGTTGGGTGCTGACAATGCACTGTTTGAGTGTGTG GAAGCGCTAGAAGGTCCAATTGGAATGCTAATTCATA GATTGGTAGATGAGTTCTTCCATGAAAACACTGCCGG TAAAAAACTCAAACCTTTACAACAAGCGAGTACTGGAG GACCTTTCAAATTCACCTGAAGAACTAGGTGAAAATG CGTCTCAATTAAGAGAGAACTTGACAACTCTATGG TGATGAGGTTGAGGCTTCTTGACCTCTTCTCTCTATCT GCGTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGTTG AGCCAGACCGCGCTAAACGCATACCAATTGCCAAATC AGGCAATTGTGAGACAGTGGTAAAAAAGATGCCTGCA AAGTTAGATTCACACAGTAAGAGAGATCCTACTCATA AATGAGGCGCTTATTTAGTAGCTAGTGATAGCCACTG CGGTTCTGCTTTATGCTATTTGTTGTATGCCTTACTATC TTTGTGTTGGCTCCTTTTTCTTGACGTTTCCGTTGGAGG GACTCCCTATTCTGAGTCATGAGCCGCACAGATTATCG

		CCCAAAATTGACAAAATCTTCTGGCGAAAAAAGTATA AAAGGAGAAAAAAGCTCACCCCTTTTCCAGCGTAGAAA GTATATATCAGTCATTGAAGAC
55	Sequence of the 3'-Region used for knock out of PpARG1	GGGACTTTAACTCAAGTAAAAGGATAGTTGTACAATT ATATATACGAAGAATAAATCATTACAAAAAGTATTCG TTTCTTTGATTCTTAACAGGATTCATTTTCTGGGTGTCA TCAGGTACAGCGCTGAATATCTTGAAGTTAACATCGA GCTCATCATCGACGTTTCATCACACTAGCCACGTTTCCG CAACGGTAGCAATAATTAGGAGCGGACCACACAGTGA CGACATCTTTCTCTTTGAAATGGTATCTGAAGCCTTCC ATGACCAATTGATGGGCTCTAGCGATGAGTTGCAAGT TATTAATGTGGTTGAACTCACGTGCTACTCGAGCACCG AATAACCAGCCAGCTCCACGAGGAGAAACAGCCCAA CTGTCGACTTCATCTGGGTCAGACCAAACCAAGTCAC AAAATCCTCCTTCATGAGGGACCTCTTGCGCTCGGCTG AGAACTCTGATTTGATCTAACATGCGAATATCGGGAG AGAGACCACCATGGATACATAATATTTTACCATCAAT GATGGCACTAAGGGTTAAAAAGTCGAACACCTGGCAA CAGTACTTCCAGACAGTGGTGGAAACCATATTTATTGA GACATTCCTCATAAAATCCATAAACCTGAGTGATCTGT CTGGATTCATGATTTCCCCTTACCAATGTGATATGTTG AGGAACTTAATTTTAAATCATGAGTAACGTGAAC GTCTCCAACGAGAAATAGCCTCTATCCACATAGTCTCC TAGGAAGATATAGTTCTGTTTTATTCCATTAGAGGAGG ATCCGGGAAACCCACCACTAATCTTGAAAAGTTCCAG TAGATCGTGAAATTGGCCGTGAATATCTCCGCATACT GTCACTGGACTCTGCACTGGCTGTATATTGGATTCTTC CATCAGCAAATCCTTCACCCGTTTCGCAAAGATGCTTCA TATCATTTTCACTTAAAGCCTTGCAGCTTTTGA CTCTT CAAACCACTGATCTGGTCCTCTTTCTGGCATGATTAAG GTCTATAATATTTCTGAGCTGAGATGTAAAAAAAAT AATAAAAATGGGGAGTGAAAAAGTGTGTAGCTTTTAG GAGTTTGGGATTGATACCCCAAATGATCTTTATGAG AATTAAAAGGTAGATACGCTTTTAATAAGAACACCTA TCTATAGTACTTTGTGGTCTTGAGTAATTGAGATGTTC

		AGCTTCTGAGGTTTGCCGTTATTCTGGGATAGTAGTGC GCGACCAAACAACCCGCCAGGCAAAGTGTGTTGTGCT CGAAGACGATTGCCAGAAGAGTAAGTCCGTCCTGCCT CAGATGTTACACACTTTCTTCCCTAGACAGTCGATGCA TCATCGGATTTAAACCTGAAACTTTGATGCCATGATAC GCCTAGTCACGTCGACTGAGATTTTAGATAAGCCCCG ATCCCTTTAGTACATTCCTGTTATCCATGGATGGAATG GCCTGATA
56	Sequence of the 5'-Region used for knock out of PpBMT4	AAGCTTGTTACCGTTGGGACTTTTCCGTGGACAATGT TGACTACTCCAGGAGGGATTCCAGCTTTCTCTACTAGC TCAGCAATAATCAATGCAGCCCCAGGCGCCCGTTCTG ATGGCTTGATGACCGTTGTATTGCCTGTCACTATAGCC AGGGGTAGGGTCCATAAAGGAATCATAGCAGGGAAA TTAAAAGGGCATATTGATGCAATCACTCCCAATGGCT CTCTTGCCATTGAAGTCTCCATATCAGCACTAACTTCC AAGAAGGACCCCTTCAAGTCTGACGTGATAGAGCACG CTTGCTCTGCCACCTGTAGTCCTCTCAAACGTCACCT TGTGCATCAGCAAAGACTTTACCTTGCTCCAATACTAT GACGGAGGCAATTCTGTCAAATTTCTCTCTCAGCAATT CAACCAACTTGAAAGCAAATTGCTGTCTCTTGATGAT GGAGACTTTTTTCCAAGATTGAAATGCAATGTGGGAC GACTCAATTGCTTCTTCCAGCTCCTCTTCGGTTGATTG AGGAACTTTTGAAACCACAAAATTGGTTCGTTGGGTCA TGTACATCAAACCATTCTGTAGATTTAGATTTCGACGAA AGCGTTGTTGATGAAGGAAAAGGTTGGATACGGTTTG TCGGTCTCTTTGGTATGGCCGGTGGGGTATGCAATTGC AGTAGAAGATAATTGGACAGCCATTGTTGAAGGTAGA GAAAAGGTCAGGGAAGTTGGGGGTTATTTATACCATT TTACCCACAAATAACAACTGAAAAGTACCCATTCCA TAGTGAGAGGTAACCGACGGAAAAAGACGGGCCCAT GTTCTGGGACCAATAGAACTGTGTAATCCATTGGGAC TAATCAACAGACGATTGGCAATATAATGAAATAGTTC GTTGAAAAGCCACGTCAGCTGTCTTTTCATTAAGTTG GTCGGACACAACATTTTCTACTGTTGTATCTGTCTAC TTTGCTTATCATCTGCCACAGGGCAAGTGGATTTCCTT CTCGCGCGGCTGGGTGAAAACGGTTAACGTGAA

57	Sequence of the 3'-Region used for knock out of PpBMT4	GCCTTGGGGGACTTCAAGTCTTTGCTAGAACTAGAT GAGGTCAGGCCCTCTTATGGTTGTGTCCCAATTGGGCA ATTTCACTCACCTAAAAAGCATGACAATTATTTAGCG AAATAGGTAGTATATTTTCCCTCATCTCCCAAGCAGTT TCGTTTTTGCATCCATATCTCTCAAATGAGCAGCTACG ACTCATTAGAACCAGAGTCAAGTAGGGGTGAGCTCAG TCATCAGCCTTCGTTTCTAAAACGATTGAGTTCTTTTG TTGCTACAGGAAGCGCCCTAGGGAACTTTCGCACTTT GGAAATAGATTTTGATGACCAAGAGCGGGAGTTGATA TTAGAGAGGCTGTCCAAAGTACATGGGATCAGGCCGG CCAAATTGATTGGTGTGACTAAACCATTGTGTACTTGG ACACTCTATTACAAAAGCGAAGATGATTTGAAGTATT ACAAGTCCCGAAGTGTTAGAGGATTCTATCGAGCCCA GAATGAAATCATCAACCGTTATCAGCAGATTGATAAA CTCTTGGAAGCGGTATCCCATTTTCATTATTGAAGAA CTACGATAATGAAGATGTGAGAGACGGCGACCCTCTG AACGTAGACGAAGAAACAAATCTACTTTTGGGGTACA ATAGAGAAAGTGAATCAAGGGAGGTATTTGTGGCCAT AATACTCAACTCTATCATTAATG
58	Sequence of the 5'-Region used for knock out of PpBMT1	CATATGGTGAGAGCCGTTCTGCACA ACTAGATGTTTTC GAGCTTCGCATTGTTTCCTGCAGCTCGACTATTGAATT AAGATTTCCGGATATCTCCAATCTCACAAAACTTATG TTGACCACGTGCTTTCCTGAGGCGAGGTGTTTTATATG CAAGCTGCCAAAAATGGAAAACGAATGGCCATTTTTTC GCCCAGGCAAATTATTCGATTACTGCTGTCATAAAGA CAGTGTTGCAAGGCTCACATTTTTTTTTTAGGATCCGAG ATAAAGTGAATACAGGACAGCTTATCTCTATATCTTGT ACCATTCGTGAATCTTAAGAGTTCGGTTAGGGGGACT CTAGTTGAGGGTTGGCACTCACGTATGGCTGGGCGCA GAAATAAAATTCAGGCGCAGCAGCACTTATCGATG
59	Sequence of the 3'-Region used for knock out of PpBMT1	GAATTCACAGTTATAAATAAAAAACAAAACTCAAAAA GTTTGGGCTCCACAAAATAACTTAATTTAAATTTTGT CTAATAAATGAATGTAATTCCAAGATTATGTGATGCA AGCACAGTATGCTTCAGCCCTATGCAGCTACTAATGTC AATCTCGCCTGCGAGCGGGCCTAGATTTTCACTACAA ATTTCAAACTACGCGGATTTATTGTCTCAGAGAGCA ATTTGGCATTCTGAGCGTAGCAGGAGGCTTCATAAG

		ATTGTATAGGACCGTACCAACAAATTGCCGAGGCACA ACACGGTATGCTGTGCACTTATGTGGCTACTTCCCTAC AACGGAATGAAACCTTCCTCTTTCCGCTTAAACGAGA AAGTGTGTCGCAATTGAATGCAGGTGCCTGTGCGCCT TGGTGTATTGTTTTTGAGGGCCCAATTTATCAGGCGCC TTTTTCTTGTTGTTTTCCCTTAGCCTCAAGCAAGGTT GGTCTATTTTCATCTCCGCTTCTATACCGTGCCTGATAC TGTTGGATGAGAACACGACTCAACTTCCTGCTGCTCTG TATTGCCAGTGTTTTGTCTGTGATTTGGATCGGAGTCC TCCTTACTTGGAATGATAATAATCTTGGCGGAATCTCC CTAAACGGAGGCAAGGATTCTGCCTATGATGATCTGC TATCATTGGGAAGCTT
60	Sequence of the 5'-Region used for knock out of PpBMT3	GATATCTCCCTGGGGACAATATGTGTTGCAACTGTTTCG TTGTTGGTGCCCCAGTCCCCCAACCGGTACTAATCGGT CTATGTTCCCGTAACTCATATTCGGTTAGAACTAGAAC AATAAGTGCATCATTGTTCAACATTGTGGTTCAATTGT CGAACATTGCTGGTGCTTATATCTACAGGGAAGACGA TAAGCCTTTGTACAAGAGAGGTAACAGACAGTTAATT GGTATTTCTTTGGGAGTCGTTGCCCTCTACGTTGTCTC CAAGACATACTACATTCTGAGAAACAGATGGAAGACT CAAAAATGGGAGAAGCTTAGTGAAGAAGAGAAAGTT GCCTACTTGGACAGAGCTGAGAAGGAGAACCTGGGTT CTAAGAGGCTGGACTTTTTGTTTCGAGAGTTAACTGC ATAATTTTTTCTAAGTAAATTTTCATAGTTATGAAATTT CTGCAGCTTAGTGTTTACTGCATCGTTTACTGCATCAC CCTGTAAATAATGTGAGCTTTTTTCCTTCCATTGCTTG GTATCTTCCTTGCTGCTGTTT
61	Sequence of the 3'-Region used for knock out of PpBMT3	ACAAAACAGTCATGTACAGAACTAACGCCTTTAAGAT GCAGACCACTGAAAAGAATTGGGTCCCATTTTTCTTG AAAGACGACCAGGAATCTGTCCATTTTGTTTACTCGTT CAATCCTCTGAGAGTACTCAACTGCAGTCTTGATAAC GGTGCATGTGATGTTCTATTTGAGTTACCACATGATTT TGGCATGTCTTCCGAGCTACGTGGTGCCACTCCTATGC TCAATCTTCCTCAGGCAATCCCGATGGCAGACGACAA AGAAATTTGGGTTTCATTCCCAAGAACGAGAATATCA GATTGCGGGTGTCTGAAACAATGTACAGGCCAATGT TAATGCTTTTTGTTAGAGAAGGAACAACTTTTTTGCT

		GAGC
62	PpTRP2: 5' and ORF	ACTGGGCCTTTAGAGGGTGCTGAAGTTGACCCCTTGG TGCTTCTGGAAAAAGAACTGAAGGGCACCAGACAAGC GCAACTTCCTGGTATTCCTCGTCTAAGTGGTGGTGCCA TAGGATACATCTCGTACGATTGTATTAAGTACTTTGAA CCAAAAAGTGAAGAGAAAAGTGAAGATGTTTTGCAAC TTCCGGAAGCAGCTTTGATGTTGTTTCGACACGATCGTG GCTTTTGACAATGTTTATCAAAGATTCCAGGTAATTGG AAACGTTTCTCTATCCGTTGATGACTCGGACGAAGCTA TTCTTGAGAAATATTATAAGACAAGAGAAGAAGTGGA AAAGATCAGTAAAGTGGTATTTGACAATAAAAGTGT CCTACTATGAACAGAAAGATATTATTCAAGGCCAAA CGTTCACCTCTAATATTGGTCAGGAAGGGTATGAAAA CCATGTTTCGCAAGCTGAAAGAACATATTCTGAAAGGA GACATCTTCCAAGCTGTTCCCTCTCAAAGGGTAGCCA GGCCGACCTCATTGCACCCTTTCAACATCTATCGTCAT TTGAGAACTGTCAATCCTTCTCCATACATGTTCTATAT TGACTATCTAGACTTCCAAGTTGTTGGTGCTTCACCTG AATTACTAGTTAAATCCGACAACAACAACAAAATCAT CACACATCCTATTGCTGGAAGTCTTCCCAGAGGTAAA ACTATCGAAGAGGACGACAATTATGCTAAGCAATTGA AGTCGTCTTTGAAAGACAGGGCCGAGCACGTCATGCT GGTAGATTTGGCCAGAAATGATATTAACCGTGTGTGT GAGCCCACCAGTACCACGGTTGATCGTTTATTGACTGT GGAGAGATTTTCTCATGTGATGCATCTTGTGTCAGAAG TCAGTGGAACATTGAGACCAAACAAGACTCGCTTCGA TGCTTTCAGATCCATTTTCCCAGCAGGTACCGTCTCCG GTGCTCCGAAGGTAAGAGCAATGCAACTCATAGGAGA ATTGGAAGGAGAAAAGAGAGGTGTTTATGCGGGGGCC GTAGGACACTGGTCGTACGATGGAAAATCGATGGACA CATGTATTGCCTTAAGAACAATGGTCGTCAAGGACGG TGTCGCTTACCTTCAAGCCGGAGGTGGAATTGTCTACG ATTCTGACCCCTATGACGAGTACATCGAAACCATGAA CAAAATGAGATCCAACAATAACACCATCTTGAGGGCT GAGAAAATCTGGACCGATAGGTTGGCCAGAGACGAG AATCAAAGTGAATCCGAAGAAAACGATCAATGA
63	PpTRP2 3'	ACGGAGGACGTAAGTAGGAATTTATGTAATCATGCCA

	region	ATACATCTTTAGATTTCTTCCTCTTCTTTTTTAACGAAAG ACCTCCAGTTTTGCACTCTCGACTCTCTAGTATCTTCC CATTTCTGTTGCTGCAACCTCTTGCCTTCTGTTTCCTTC AATTGTTCTTCTTTCTTCTGTTGCACTTGGCCTTCTTCC TCCATCTTTCGTTTTTTTTCAAGCCTTTTCAGCAGTTCT TCTTCCAAGAGCAGTTCTTTGATTTTCTCTCTCCAATCC ACCAAAAACTGGATGAATTCAACCGGGCATCATCAA TGTTCCACTTTCTTTCTTATCAATAATCTACGTGCTT CGGCATACGAGGAATCCAGTTGCTCCCTAATCGAGTC ATCCACAAGGTTAGCATGGGCCTTTTTCAGGGTGTCA AAAGCATCTGGAGCTCGTTTATTCGGAGTCTTGTCTGG ATGGATCAGCAAAGACTTTTTGCGGAAAGTCTTTCTTA TATCTTCCGGAGAACAACCTGGTTTCAAATCCAAGAT GGCATAGCTGTCCAATTTGAAAGTGGAAGAATCCTG CCAATTTCTTCTCTCGTGTGCTCGTTCTCCTCCTTT TGCAACAGGTCCACTTCATCTGGCATTTTTCTTTATGT TAACTTTAATTATTATTAATTATAAAGTTGATTATCGT TATCAAAATAATCATATTCGAGAAATAATCCGTCCAT GCAATATATAAATAAGAATTCATAATAATGTAATGAT AACAGTACCTCTGATGACCTTTGATGAACCGCAATTTT CTTTCCAATGACAAGACATCCCTATAATACAATTATAC AGTTTATATATCACAAATAATCACCTTTTTATAAGAAA ACCGTCCTCTCCGTAACAGAACTTATTATCCGCACGTT ATGGTTAACACACTACTAATACCGATATAGTGTATGA AGTCGCTACGAGATAGCCATCCAGGAACTTACCAAT TCATCAGCACTTTCATGATCCGATTGTTGGCTTTATTC TTTGCAGAGACAGATACTTGCCAATGAAATAACTGATC CCACAGATGAGAATCCGGTGCTCGT
64	Mouse CMP- sialic acid transporter (MmCST) Codon optimized	ATGGCTCCAGCTAGAGAAAACGTTTCCTTGTTCCTCAA GTTGTACTGTTTGGCTGTTATGACTTTGGTTGCTGCTG CTTACACTGTTGCTTTGAGATACACTAGAACTACTGCT GAGGAGTTGTACTTCTCCACTACTGCTGTTTGTATCAC TGAGGTTATCAAGTTGTTGATCTCCGTTGGTTTGTG CTAAGGAGACTGGTTCTTTGGGAAGATTCAAGGCTTC CTTGTCCGAAAACGTTTTGGGTTCCCCAAAGGAGTTG GCTAAGTTGTCTGTTCCATCCTTGGTTTACGCTGTTCA GAACAACATGGCTTTCTTGGCTTTGTCTAACTTGGACG

		CTGCTGTTTACCAAGTTACTTACCAGTTGAAGATCCCA TGTACTGCTTTGTGTACTGTTTTGATGTTGAACAGAAC ATTGTCCAAGTTGCAGTGGATCTCCGTTTTTCATGTTGT GTGGTGGTGTACTTTGGTTCAGTGGAAGCCAGCTCA AGCTTCCAAAGTTGTTGTTGCTCAGAACCCATTGTTGG GTTTCGGTGCTATTGCTATCGCTGTTTTGTGTTCCGGTT TCGCTGGTGTTTACTTCGAGAAGGTTTTGAAGTCCTCC GACACTTCTTTGTGGGTTAGAAACATCCAGATGTACTT GTCCGGTATCGTTGTTACTTTGGCTGGTACTTACTTGT CTGACGGTGCTGAGATTCAAGAGAAGGGATTCTTCTA CGGTTACACTTACTATGTTTGGTTCGTTATCTTCTTGGC TTCCGTTGGTGGTTTGTACACTTCCGTTGTTGTTAAGT ACACTGACAACATCATGAAGGGATTCTCTGCTGCTGC TGCTATTGTTTTGTCCACTATCGCTTCCGTTTTGTTGTT CGGATTGCAGATCACATTGTCCTTTGCTTTGGGAGCTT TGTTGGTTTGTGTTTCCATCTACTTGTACGGATTGCCA AGACAAGACACTACTTCCATTCAGCAAGAGGCTACTT CCAAGGAGAGAATCATCGGTGTTTAGTAG
65	Human UDP- GlcNAc 2- epimerase/N- acetylmanno- samine kinase (HsGNE) codon opitimized	ATGGAAAAGAACGGTAACAACAGAAAGTTGAGAGTTT GTGTTGCTACTTGTAAACAGAGCTGACTACTCCAAGTTG GCTCCAATCATGTTCCGGTATCAAGACTGAGCCAGAGT TCTTCGAGTTGGACGTTGTTGTTTTGGGTTCCCACTTG ATTGATGACTACGGTAACACTTACAGAATGATCGAGC AGGACGACTTCGACATCAACACTAGATTGCACACTAT TGTTAGAGGAGAGGACGAAGCTGCTATGGTTGAATCT GTTGGATTGGCTTTGGTTAAGTTGCCAGACGTTTTGAA CAGATTGAAGCCAGACATCATGATTGTTACGGTGAC AGATTGATGCTTTGGCTTTGGCTACTTCCGCTGCTTT GATGAACATTAGAATCTTGCACATCGAGGGTGGTGAA GTTTCTGGTACTATCGACGACTCCATCAGACACGCTAT CACTAAGTTGGCTCACTACCATGTTTGTGTACTAGAT CCGCTGAGCAACACTTGATTTCCATGTGTGAGGACCA CGACAGAATTTTGTGGCTGGTTGTCCATCTTACGACA AGTTGTTGTCCGCTAAGAACAAGGACTACATGTCCAT CATCAGAATGTGGTTGGGTGACGACGTTAAGTCTAAG GACTACATCGTTGCTTTGCAGCACCCAGTTACTACTGA CATCAAGCACTCCATCAAGATGTTTCGAGTTGACTTTGG

	ACGCTTTGATCTCCTTCAACAAGAGAACTTTGGTTTTG TTCCCAAACATTGACGCTGGTTCCAAAGAGATGGTTA GAGTTATGAGAAAGAAGGGTATCGAACACCACCCAA ACTTCAGAGCTGTAAAGCACGTTCCATTCGACCAATTC ATCCAGTTGGTTGCTCATGCTGGTTGTATGATCGGTAA CTCCTCCTGTGGTGTAGAGAAGTTGGTGCTTTCGGTA CTCCAGTTATCAACTTGGGTACTAGACAGATCGGTAG AGAGACTGGAGAAAACGTTTTTGCATGTTAGAGATGCT GACACTCAGGACAAGATTTTGCAGGCTTTCGACTTGC AATTCGGAAAGCAGTACCCATGTTCCAAAATCTACGG TGACGGTAACGCTGTTCCAAGAATCTTGAAGTTTTTGA AGTCCATCGACTTGCAAGAGCCATTGCAGAAGAAGTT CTGTTTCCCACCAGTTAAGGAGAACATCTCCCAGGAC ATTGACCACATCTTGGAGACATTGTCCGCTTTGGCTGT TGATTTGGGTGGAACCTAAGTTGAGAGTTGCTATCGTTT CCATGAAGGGAGAGATCGTTAAGAAGTACACTCAGTT CAACCCAAAGACTTACGAGGAGAGAATCAACTTGATC TTGCAGATGTGTGTTGAAGCTGCTGCTGAGGCTGTAA GTTGAACTGTAGAATCTTGGGTGTTGGTATCTCTACTG GTGGTAGAGTTAATCCAAGAGAGGGTATCGTTTTGCA CTCCACTAAGTTGATTCAGGAGTGGAACCTCCGTTGATT TGAGAACTCCATTGTCCGACACATTGCACTTGCCAGTT TGGGTTGACAACGACGGTAATTGTGCTGCTTTGGCTG AGAGAAAGTTCGGTCAAGGAAAGGGATTGGAGAACTT CGTTACTTTGATCACTGGTACTGGTATTGGTGGTGGTA TCATTCACCAGCACGAGTTGATTCACGGTTCTTCCTTC TGTGCTGCTGAATTGGGACACTTGGTTGTTTCTTTGGA CGGTCCAGACTGTTCTTGTGGTTCCCACGGTTGTATTG AAGCTTACGCATCAGGAATGGCATTGCAGAGAGAGGC TAAGAAGTTGCACGACGAGGACTTGTGTGTTGGTTGAG GGAATGTCTGTTCCAAAGGACGAGGCTGTTGGTGCTT TGCATTTGATCCAGGCTGCTAAGTTGGGTAATGCTAA GGCTCAGTCCATCTTGAGAACTGCTGGTACTGCTTTGG GATTGGGTGTTGTTAATATCTTGCACACTATGAACCCA TCCTTGGTTATCTTGTCCGGTGTTTGGCTTCTCACTAC ATCCACATCGTTAAGGACGTTATCAGACAGCAAGCTT TGTCTCCGTTCAAGACGTTGATGTTGTTGTTCCGAC
--	---

		TTGGTTGACCCAGCTTTGTTGGGTGCTGCTTCCATGGT TTTGGACTACACTACTAGAAGAATCTACTAATAG
66	Sequence of the PpARG1 auxotrophic marker	CAGTTGAGCCAGACCGCGCTAAACGCATACCAATTGC CAAATCAGGCAATTGTGAGACAGTGGTAAAAAAGATG CCTGCAAAGTTAGATTCACACAGTAAGAGAGATCCTA CTCATAAATGAGGCGCTTATTTAGTAGCTAGTGATAG CCACTGCGGTTCTGCTTTATGCTATTTGTTGTATGCCTT ACTATCTTTGTTTGGCTCCTTTTTCTTGACGTTTTCCGT TGGAGGGACTCCCTATTCTGAGTCATGAGCCGCACAG ATTATCGCCCCAAAATTGACAAAATCTTCTGGCGAAAA AAGTATAAAAGGAGAAAAAAGCTCACCCTTTTCCAGC GTAGAAAGTATATATCAGTCATTGAAGACTATTATTTA AATAACACAATGTCTAAAGGAAAAGTTTGTTTGGCCT ACTCCGGTGGTTTGGATACCTCCATCATCCTAGCTTGG TTGTTGGAGCAGGGATACGAAGTCGTTGCCTTTTTAGC CAACATTGGTCAAGAGGAAGACTTTGAGGCTGCTAGA GAGAAAGCTCTGAAGATCGGTGCTACCAAGTTTATCG TCAGTGACGTTAGGAAGGAATTTGTTGAGGAAGTTTT GTTCCCAGCAGTCCAAGTTAACGCTATCTACGAGAAC GTCTACTTACTGGGTACCTCTTTGGCCAGACCAGTCAT TGCCAAGGCCCAAATAGAGGTTGCTGAACAAGAAGGT TGTTTTGCTGTTGCCACGGTTGTACCGGAAAGGGTAA CGATCAGGTTAGATTTGAGCTTTCCTTTTATGCTCTGA AGCCTGACGTTGTCTGTATCGCCCCATGGAGAGACCC AGAATTCTTCGAAAGATTCGCTGGTAGAAATGACTTG CTGAATTACGCTGCTGAGAAGGATATTCCAGTTGCTC AGACTAAAGCCAAGCCATGGTCTACTGATGAGAACAT GGCTCACATCTCCTTCGAGGCTGGTATTCTAGAAGATC CAAACACTACTCCTCCAAAGGACATGTGGAAGCTCAC TGTTGACCCAGAAGATGCACCAGACAAGCCAGAGTTC TTTGACGTCCACTTTGAGAAGGGTAAGCCAGTTAAAT TAGTTCTCGAGAACAACAACTGAGGTCACCGATCCGGT TGAGATCTTTTTGACTGCTAACGCCATTGCTAGAAGAA ACGGTGTGGTAGAATTGACATTGTCGAGAACAGATT CATCGGAATCAAGTCCAGAGGTTGTTATGAACTCCA GGTTTGACTCTACTGAGAACCACTCACATCGACTTGG AAGGTCTTACCGTTGACCGTGAAGTTAGATCGATCAG

		AGACACTTTTGTACCCCAACCTACTCTAAGTTGTTAT ACAACGGGTTGTACTTTACCCCAGAAGGTGAGTACGT CAGAACTATGATTCAGCCTTCTCAAAACACCGTCAAC GGTGTGTTAGAGCCAAGGCCTACAAAGGTAATGTGT ATAACCTAGGAAGATACTCTGAAACCGAGAAATTGTA CGATGCTACCGAATCTTCCATGGATGAGTTGACCGGA TTCCACCCTCAAGAAGCTGGAGGATTTATCACAACAC AAGCCATCAGAATCAAGAAGTACGGAGAAAGTGTCA GAGAGAAGGGAAAGTTTTTGGGACTTTAACTCAAGTA AAAGGATAGTTGTACAATTATATATACGAAGAATAAA TCATTACAAAAAGTATTCGTTTCTTTGATTCTTAACAG GATTCATTTTCTGGGTGTCATCAGGTACAGCGCTGAAT ATCTTGAAGTTAACATCGAGCTCATCATCGACGTTTAT CACACTAGCCACGTTTCCGCAACGGTAGCAATAATTA GGAGCGGACCACACAGTGACGACATC
67	Human CMP-sialic acid synthase (HsCSS) codon optimized	ATGGACTCTGTTGAAAAGGGTGCTGCTACTTCTGTTTC CAACCCAAGAGGTAGACCATCCAGAGGTAGACCTCCT AAGTTGCAGAGAAACTCCAGAGGTGGTCAAGGTAGAG GTGTTGAAAAGCCACCACACTTGGCTGCTTTGATCTTG GCTAGAGGAGGTTCTAAGGGTATCCCATTTGAAGAACA TCAAGCACTTGGCTGGTGTTCATTGATTGGATGGGTT TTGAGAGCTGCTTTGGACTCTGGTGCTTTCCAATCTGT TTGGGTTTCCACTGACCACGACGAGATTGAGAACGTT GCTAAGCAATTCGGTGCTCAGGTTACAGAAGATCCT CTGAGGTTTCCAAGGACTCTTCTACTTCCTTGGACGCT ATCATCGAGTTCTTGAACCTACCACAACGAGGTTGACA TCGTTGGTAACATCCAAGCTACTTCCCCATGTTTGCAC CCAAGTGAAGTTGCAAAAAGTTGCTGAGATGATCAGAG AAGAGGGTTACGACTCCGTTTTCTCCGTTGTTAGAAGG CACCAGTTCAGATGGTCCGAGATTCAGAAGGGTGTTA GAGAGGTTACAGAGCCATTGAACTTGAACCCAGCTAA AAGACCAAGAAGGCAGGATTGGGACGGTGAATTGTAC GAAAACGGTTCCTTCTACTTCGCTAAGAGACACTTGAT CGAGATGGGATACTTGCAAGGTGGAAAGATGGCTTAC TACGAGATGAGAGCTGAACACTCCGTTGACATCGACG TTGATATCGACTGGCCAATTGCTGAGCAGAGAGTTTT GAGATACGGTACTTCGGAAAGGAGAAGTTGAAGGAG

		ATCAAGTTGTTGGTTTGTAAACATCGACGGTTGTTTGAC TAACGGTCACATCTACGTTTCTGGTGACCAGAAGGAG ATTATCTCCTACGACGTTAAGGACGCTATTGGTATCTC CTTGTTGAAGAAGTCCGGTATCGAAGTTAGATTGATCT CCGAGAGAGCTTGTTCCAAGCAAACATTGTCCTCTTTG AAGTTGGACTGTAAGATGGAGGTTTCCGTTTCTGACA AGTTGGCTGTTGTTGACGAATGGAGAAAGGAGATGGG TTTGTGTTGGAAGGAAGTTGCTTACTTGGGTAACGAA GTTTCTGACGAGGAGTGTTTGAAGAGAGTTGGTTTGTG TGGTGCTCCAGCTGATGCTTGTTCCACTGCTCAAAAGG CTGTTGGTTACATCTGTAAGTGTAACGGTGGTAGAGGT GCTATTAGAGAGTTGCTGAGCACATCTGTTTGTGAT GGAGAAAGTTAATAACTCCTGTCAGAAGTAGTAG
68	Human N- acetylneuramin- ate-9-phosphate synthase (HsSPS) codon optimized	ATGCCATTGGAATTGGAGTTGTGTCCTGGTAGATGGGT TGGTGGTCAACACCCATGTTTCATCATCGCTGAGATCG GTCAAAACCACCAAGGAGACTTGGACGTTGCTAAGAG AATGATCAGAATGGCTAAGGAATGTGGTGCTGACTGT GCTAAGTTCCAGAAGTCCGAGTTGGAGTTCAAGTTCA ACAGAAAGGCTTTGGAAAGACCATACTTCCAAGCA CTCTTGGGGAAAGACTTACGGAGAACACAAGAGACAC TTGGAGTTCTCTCACGACCAATACAGAGAGTTGCAGA GATACGCTGAGGAAGTTGGTATCTTCTTCACTGCTTCT GGAATGGACGAAATGGCTGTTGAGTTCTTGCACGAGT TGAACGTTCCATTCTTCAAAGTTGGTTCCGGTGACACT AACAACTTCCCATACTTGGAAAAGACTGCTAAGAAAG GTAGACCAATGGTTATCTCCTCTGGAATGCAGTCTATG GACACTATGAAGCAGGTTTACCAGATCGTTAAGCCAT TGAACCCAACTTTTGTCTTGCAGTGACTTCCGCT TACCCATTGCAACCAGAGGACGTTAATTTGAGAGTTA TCTCCGAGTACCAGAAGTTGTTCCCAGACATCCCAATT GGTTACTCTGGTCACGAGACTGGTATTGCTATTTCCGT TGCTGCTGTTGCTTTGGGTGCTAAGGTTTTGGAGAGAC ACATCACTTTGGACAAGACTTGGAAAGGGTTCTGATCA CTCTGCTTCTTTGGAACCTGGTGAGTTGGCTGAAGTTG TTAGATCAGTTAGATTGGTTGAGAGAGCTTTGGGTCC CCAATAAGCAATTGTTGCCATGTGAGATGGCTTGTA ACGAGAAGTTGGGAAAGTCCGTTGTTGCTAAGGTAA

		GATCCCAGAGGGTACTATCTTGACTATGGACATGTTG ACTGTTAAAGTTGGAGAGCCAAAGGGTTACCCACCAG AGGACATCTTTAACTTGGTTGGTAAAAAGGTTTTGGTT ACTGTTGAGGAGGACGACACTATTATGGAGGAGTTGG TTGACAACCACGGAAAGAAGATCAAGTCCTAG
69	Mouse alpha- 2,6-sialyl transferase catalytic domain (MmmST6) codon optimized	GTTTTTCAAATGCCAAAGTCCCAGGAGAAAGTTGCTG TTGGTCCAGCTCCACAAGCTGTTTTCTCCAACCTCCAAG CAAGATCCAAAGGAGGGTGTTCAAATCTTGTCCTACC CAAGAGTTACTGCTAAGGTTAAGCCACAACCATCCTT GCAAGTTTGGGACAAGGACTCCACTTACTCCAAGTTG AACCCAAGATTGTTGAAGATTTGGAGAACTACTTGA ACATGAACAAGTACAAGGTTTCCTACAAGGGTCCAGG TCCAGGTGTTAAGTTCTCCGTTGAGGCTTTGAGATGTC ACTTGAGAGACCACGTTAACGTTTCCATGATCGAGGC TACTGACTTCCCATTCAACACTACTGAATGGGAGGGA TACTTGCCAAAGGAGAACTTCAGAACTAAGGCTGGTC CATGGCATAAGTGTGCTGTTGTTTCTTCTGCTGGTTCC TTGAAGAACTCCCAGTTGGGTAGAGAAATTGACAACC ACGACGCTGTTTTGAGATTCAACGGTGCTCCAACCTGA CAACTTCCAGCAGGATGTTGGTACTAAGACTACTATC AGATTGGTTAACTCCCAATTGGTTACTACTGAGAAGA GATTCTTGAAGGACTCCTTGTACACTGAGGGAATCTTG ATTTTGTGGGACCCATCTGTTTACCACGCTGACATTCC ACAATGGTATCAGAAGCCAGACTACAACCTTCTTCGAG ACTTACAAGTCCTACAGAAGATTGCACCCATCCCAGC CATTCTACATCTTGAAGCCACAAATGCCATGGGAATT GTGGGACATCATCCAGGAAATTTCCCCAGACTTGATC CAACCAAACCCACCATCTTCTGGAATGTTGGGTATCAT CATCATGATGACTTTGTGTGACCAGGTTGACATCTACG AGTTCTTGCCATCCAAGAGAAAGACTGATGTTTGTAC TACCACCAGAAGTTCTTCGACTCCGCTTGACTATGGG AGCTTACCACCCATTGTTGTTTCGAGAAGAACATGGTT AAGCACTTGAACGAAGGTACTGACGAGGACATCTACT TGTTTCGAAAGGCTACTTTGTCCGGTTTCAGAAACAA CAGATGTTAG
70	HSA signal peptide DNA	ATGAAGTGGGTTACCTTTATCTCTTTGTTGTTTCTTTT TCTTCTGCTTACTCT

71	HSA signal peptide	MKWVTFISLLFLFSSAYS
72	TNFRII-Fc fragment fusion protein (C-terminal K-less) 1-705 encodes TNFRII (underlined)	<u>CTGCCAGCTCAAGTTGCTTTTACTCCATACGCTCCAGA</u> <u>ACCAGGTTCTACTTGTAGATTGAGAGAGTACTACGAC</u> <u>CAAAGTCTCAGATGTGTTGTTCCAAGTGTCTCCAGG</u> <u>TCAACACGCTAAGGTTTTCTGTACTAAGACTTCCGACA</u> <u>CTGTTTGTGACTCTTGTGAGGACTCCACTTACACTCAA</u> <u>TTGTGGAAGTGGGTTCCAGAATGTTTGTCTGTGGTTC</u> <u>CAGATGTTCTTCCGACCAAGTTGAGACTCAGGCTTGTA</u> <u>CTAGAGAGCAGAACAGAATCTGTACTTGTAGACCTGG</u> <u>TTGGTACTGTGCTTTGTCCAAGCAAGAGGGTTGTAGAT</u> <u>TGTGTGCTCCATTGAGAAAGTGTAGACCAGGTTTCGG</u> <u>TGTTGCTAGACCAGGTACAGAACTTCCGACGTTGTTT</u> <u>GTAAGCCATGTGCTCCAGGAACCTTTCTCCAACACTACT</u> <u>TCCTCCACTGACATCTGTAGACCACACCAAATCTGTAA</u> <u>CGTTGTTGCTATCCCAGGTAACGCTTCTATGGACGCTG</u> <u>TTTGTACTTCTACTTCCCCAACTAGATCCATGGCTCCA</u> <u>GGTGCTGTTTCATTTGCCACAGCCAGTTTCCACTAGATC</u> <u>CCAACACACTCAACCAACTCCAGAACCATCTACTGCT</u> <u>CCATCCACTTCCTTTTTTGTGTTGCCAATGGGACCATCTCC</u> <u>ACCTGCTGAAGGTTCTACTGGTGACGAGCCAAAGTCC</u> <u>TGTGACAAGACACATACTTGTCCACCATGTCCAGCTCC</u> <u>AGAATTGTTGGGTGGTCCATCCGTTTTCTTGTTCAC</u> <u>CAAAGCCAAAGGACACTTTGATGATCTCCAGAACTCC</u> <u>AGAGGTTACATGTGTTGTTGTTGACGTTTCTCACGAGG</u> <u>ACCCAGAGGTTAAGTTCAACTGGTACGTTGACGGTGT</u> <u>TGAAGTTCACAACGCTAAGACTAAGCCAAGAGAAGA</u> <u>GCAGTACAACCTCACTTACAGAGTTGTTTCCGTTTTGA</u> <u>CTGTTTTGCACCAGGATTGGTTGAACGGTAAAGAATA</u> <u>CAAGTGTAAGGTTTCCAACAAGGCTTTGCCAGCTCCA</u> <u>ATCGAAAAGACAATCTCCAAGGCTAAGGGTCAACCAA</u> <u>GAGAGCCACAGGTTTACACTTTGCCACCATCCAGAGA</u> <u>AGAGATGACTAAGAACCAGGTTTCCTTGACTTGTTTG</u> <u>GTTAAAGGATTCTACCCATCCGACATTGCTGTTGAATG</u> <u>GGAATCTAACGGTCAACCAGAGAACAACTACAAGACT</u> <u>ACTCCACCAGTTTTTGATTCTGACGGTTCCTTCTTCTT</u> <u>GTAAGTCCAAGTTGACTGTTGACAAGTCCAGATGGCAA</u>

		CAGGGTAACGTTTTCTCCTGTTCCGTTATGCATGAGGC TTTGCACAACCACTACACTCAAAGTCCTTGTCTTTGT CCCCAGGTTAG
73	TNFRII-Fc fragment fusion protein (C- terminal K-less) 1-235 receptor domain (underlined)	<u>LPAQVAFTPYAPEPGSTCRLREYYDQTAQMCCSKCSPGQ</u> <u>HAKVFCTKTSDTVCDSCEDSTYTQLWNWVPECLSCGSR</u> <u>CSSDQVETQACTREQNRICTCRPGWYCALSKEGCRLC</u> <u>APLRKCRPGFGVARPGTETSDVVCKPCAPGTFSENNTSST</u> <u>DICRPHQICNVVAIPGNASMDAVCTSTSPTRSMAPGAVH</u> <u>LPQPVSTRSQHTQPTPEPSTAPSTSFLPMGPSPPAEGSTG</u> <u>DEPKSCDKTHTCPPCPAPPELLGGPSVFLFPPKPKDTLMIS</u> <u>RTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKP</u> <u>REEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALP</u> <u>APIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLV</u> <u>KGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYS</u> <u>KLTVDKSRWQQGNVFCSSVMHEALHNHYTQKSLSLSPG</u>
74	TNFRII-Fc fragment fusion protein (with C- terminal K) 1-705 encode TNFRII (underlined)	<u>CTGCCAGCTCAAGTTGCTTTTACTCCATACGCTCCAGA</u> <u>ACCAGGTTCTACTTGTAGATTGAGAGAGTACTACGAC</u> <u>CAAAGTCTCAGATGTGTTGTTCCAAGTGTCTCCAGG</u> <u>TCAACACGCTAAGGTTTTCTGTACTAAGACTTCCGACA</u> <u>CTGTTTGTGACTCTTGTGAGGACTCCACTTACACTCAA</u> <u>TTGTGGAAGTGGGTTCCAGAATGTTTGTCTGTGGTTC</u> <u>CAGATGTTCTTCCGACCAAGTTGAGACTCAGGCTTGTA</u> <u>CTAGAGAGCAGAACAGAATCTGTACTTGTAGACCTGG</u> <u>TTGGTACTGTGCTTTGTCCAAGCAAGAGGGTTGTAGAT</u> <u>TGTGTGCTCCATTGAGAAAGTGTAGACCAGGTTTCGG</u> <u>TGTTGCTAGACCAGGTACAGAACTTCCGACGTTGTTT</u> <u>GTAAGCCATGTGCTCCAGGAACCTTTCTCCAACACTACT</u> <u>TCCTCCACTGACATCTGTAGACCACACCAAATCTGTAA</u> <u>CGTTGTTGCTATCCCAGGTAACGCTTCTATGGACGCTG</u> <u>TTTGTACTTCTACTTCCCCAACTAGATCCATGGCTCCA</u> <u>GGTGCTGTTTCAATTTGCCACAGCCAGTTTCCACTAGATC</u> <u>CCAACACACTCAACCAACTCCAGAACCATCTACTGCT</u> <u>CCATCCACTTCCTTTTTGTTGCCAATGGGACCATCTCC</u> <u>ACCTGCTGAAGGTTCTACTGGTGACGAGCCAAAGTCC</u> <u>TGTGACAAGACACATACTTGTCCACCATGTCCAGCTCC</u> <u>AGAATTGTTGGGTGGTCCATCCGTTTTCTTGTTCCAC</u> <u>CAAAGCCAAAGGACACTTTGATGATCTCCAGAACTCC</u>

		AGAGGTTACATGTGTTGTTGTTGACGTTTCTCACGAGG ACCCAGAGGTTAAGTTCAACTGGTACGTTGACGGTGT TGAAGTTCACAACGCTAAGACTAAGCCAAGAGAAGA GCAGTACAACCTCCACTTACAGAGTTGTTTCCGTTTTGA CTGTTTTGCACCAGGATTGGTTGAACGGTAAAGAATA CAAGTGTAAGGTTTCCAACAAGGCTTTGCCAGCTCCA ATCGAAAAGACAATCTCCAAGGCTAAGGGTCAACCAA GAGAGCCACAGGTTTACACTTTGCCACCATCCAGAGA AGAGATGACTAAGAACCAGGTTTCCTTGACTTGTTTG GTTAAAGGATTCTACCCATCCGACATTGCTGTTGAATG GGAATCTAACGGTCAACCAGAGAACTACAAGACT ACTCCACCAGTTTTGGATTCTGACGGTTCCTTCTTCTT GTACTCCAAGTTGACTGTTGACAAGTCCAGATGGCAA CAGGGTAACGTTTTCTCCTGTTCCGTTATGCATGAGGC TTTGCACAACCACTACACTCAAAAGTCCTTGTCTTTGT CCCCAGGTAAGTAG
75	TNFRII-Fc fragment fusion protein (with C- terminal K) 1-235 receptor domain (underlined)	<u>LPAQVAFTPYAPEPGSTCRLREYYDQTAQMCCSKCSPGQ</u> <u>HAKVFCTKTSDTVCDSCEDSTYTQLWNWVPECLSCGSR</u> <u>CSSDQVETQACTREQNRICTCRPGWYCALSKQEGCRLC</u> <u>APLRKCRPGFGVARPGTETSDVVCKPCAPGTFSNTTSSST</u> <u>DICRPHQICNVVAIPGNASMDAVCTSTSPTRSMAPGAVH</u> <u>LPQPVSTRSQHTQPTPEPSTAPSTSFLPMGPSPPAEGSTG</u> <u>DEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMIS</u> RTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKP REEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALP APIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLV KGFYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYS KLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG K
76	Mouse POMGnT I	CGCGCCATTTCTGAAGCTAACGAGGACCCTGAACCAG AACAAGATTACGACGAGGCTTTGGGAAGATTGGAATC CCCAAGAAGAAGAGGATCCTCCCCTAGAAGAGTTTTG GACGTTGAGGTTTACTCTTCCAGATCCAAGGTTTACGT TGCTGTTGACGGTACTACTGTTTTGGAGGACGAGGCT AGAGAACAAGGTAGAGGTATCCACGTTATCGTTTTGA ACCAGGCTACTGGTCATGTTATGGCTAAGAGAGTTTTTC GACACTTACTCTCCACACGAAGATGAGGCTATGGTTTT

	GTTCTTGAACATGGTTGCTCCAGGTAGAGTTTTGATTT GTACTGTTAAGGACGAGGGATCCTTCCATTTGAAGGA CACTGCTAAGGCTTTGTTGAGATCCTTGGGTTCTCAAG CTGGTCCAGCTTTGGGATGGAGAGATACTTGGGCTTTC GTTGGTAGAAAGGGTGGTCCAGTTTTGGGTGAAAAGC ACTCTAAGTCCCCAGCTTTGTCCTCTTGGGGTGACCCA GTTTTGTTGAAAAGTACGTTCCATTGTCCTCTGCTGA AGAGGCTGAATGTCACTGGGCTGACACTGAGTTGAAC AGAAGAAGAAGAAGATTCTGTTCCAAGGTTGAGGGTT ACGGTTCTGTTTGTTCCTGTAAGGACCCAAGTCCAATT GAATTCTCCCCAGACCCATTGCCAGATAACAAGGTTTT GAACGTTCCAGTTGCTGTTATCGCTGGTAACAGACCA AACTACTTGTACAGAATGTTGAGATCTTTGTTGTCCGC TCAGGGAGTTTCTCCACAGATGATCACTGTTTTCATCG ACGGTTACTACGAAGAACCAATGGACGTTGTTGCTTT GTTTCGGATTGAGAGGTATTCAGCACACTCCAATCTCC ATCAAGAACGCTAGAGTTTCCCAACACTACAAGGCTT CCTTGACTGCTACTTTCAACTTGTTCCCAGAGGCTAAG TTCGCTGTTGTTTTGGAAGAGGACTTGGACATTGCTGT TGATTTCTTCTCCTTCTTGTTCCCAATCCATCCACTTGTT GGAAGAGGATGACTCCTTGTACTGTATCTCTGCTTGGA ACGACCAAGGTTACGAACACACTGCTGAGGATCCAGC TTTGTTGTACAGAGTTGAGACTATGCCAGGATTGGGAT GGGTTTTGAGAAAGTCCTTGTACAAAGAGGAGTTGGA GCCAAAGTGGCCAACTCCAGAAAAGTTGTGGGATTGG GACATGTGGATGAGAATGCCAGAGCAGAGAAGAGGT AGAGAGTGTATCATCCCAGACGTTTCCAGATCTTACC ACTTCGGTATTGTTGGATTGAACATGAACGGTTACTTC CACGAGGCTTACTTCAAGAAGCACAAGTTCAACACTG TTCCAGGTGTTTCAAGTTGAGAAACGTTGACTCCTTGAAG AAAGAGGCTTACGAGGTTGAGATCCACAGATTGTTGT CTGAGGCTGAGGTTTTGGATCACTCCAAGGATCCATG TGAGGACTCATTCTTGCCAGATACTGAGGGTCATACTT ACGTTGCTTTCATCAGAATGGAAACTGACGACGACTT TGCTACTTGGACTCAGTTGGCTAAGTGTTTGCACATTT GGGACTTGGATGTTAGAGGTAACCACAGAGGATTGTG GAGATTGTTTCAAGAAAGAAGCACTTCTTGGTTGTT
--	--

		GGTGTTCAGCTTCTCCATACTCCGTTAAGAAGCCACC ATCCGTTACTCCAATTTTCTTGGAGCCACCACCAAAGG AAGAAGGTGCTCCTGGAGCTGCTGAACAAACTTAGTA GTAA
77	DNA encodes Mnn6-s leader (65)	ATGCACGTACTGCTGAGCAAAAAAATAGCACGCTTTC TGTTGATTTCGTTTGTTCGTGCTTGCGCTAATGGTG ACAATAAATCATCCAGGGCGCGCC
78	DNA encodes Mnn5-s leader (56)	ATGCTGATTAGGTTAAAGAAGAGAAAAATCCTGCAGG TCATCGTGAGCGCAGTAGTGCTAATTTTATTTTTTTGT TCTGTGCATAATGATGTGTCTTCTAGTTGGGGGCGCGC C
79	HYG ^R resistance cassette	GATCTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCG GCCAGCGACATGGAGGCCAGAATACCCTCCTTGACA GTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTG TCGCCCCGTACATTTAGCCCATACATCCCCATGTATAAT CATTTGCATCCATACATTTTGATGGCCGCACGGCGCGA AGCAAAAATTACGGCTCCTCGCTGCGGACCTGCGAGC AGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCC CCACGCCGCGCCCCCTGTAGAGAAATATAAAAGGTTAG GATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTT AAAATCTTGCTAGGATACAGTTCTCACATCACATCCG AACATAAACAACCATGGGTAAAAAGCCTGAACCTACC GCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCTG ACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGA AGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGT GGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTT TCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCG GCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGG AATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGT GCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCG AACTGCCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCAT GGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGC GGGTTTCGGCCCATTCGGACCGCAAGGAATCGGTCAAT ACACTACATGGCGTGATTTTCATATGCGCGATTGCTGAT CCCCATGTGTATCACTGGCAAACCTGTGATGGACGACA CCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCT GATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCAC

		CTCGTGCACGCGGATTTCGGGCTCCAACAATGTCCTGAC GGACAATGGCCGCATAACAGCGGTCATTGACTGGAGC GAGGCGATGTTCTGGGGATTCCCAATACGAGGTCGCCA ACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAG CAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGC TTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCG CATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACG GCAATTCGATGATGCAGCTTGGGCGCAGGGTCGATG CGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGG CGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGA CCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGA CCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATAA TCAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAA CTTGTCATTTGTATAGTTTTTTTATATTGTAGTTGTTCT ATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCG CCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCG CAGAAAGTAATATCATGCGTCAATCGTATGTGAATGC TGGTCGCTATACTGCTGTCGATTTCGATACTAACGCCGC CATCCAGTGTCGAAAACGAGCT
80	DNA encodes <i>S. cerevisiae</i> Mating Factor pre signal sequence	ATG AGA TTC CCA TCC ATC TTC ACT GCT GTT TTG TTC GCT GCT TCT TCT GCT TTG GCT
81	DNA encodes Tr ManI catalytic domain	CGCGCCGGATCTCCAACCCTACGAGGGCGGCAGCAG TCAAGGCCGCATTCCAGACGTCGTGGAACGCTTACCA CCATTTTGCCTTTCCCCATGACGACCTCCACCCGGTCA GCAACAGCTTTGATGATGAGAGAAACGGCTGGGGCTC GTCGGCAATCGATGGCTTGGACACGGCTATCCTCATG GGGGATGCCGACATTGTGAACACGATCCTTCAGTATG TACCGCAGATCAACTTCACCACGACTGCGGTTGCCAA CCAAGGCATCTCCGTGTTTCGAGACCAACATTCGGTAC CTCGGTGGCCTGCTTTCTGCCTATGACCTGTTGCGAGG TCCTTTCAGCTCCTTGGCGACAAACCAGACCCTGGTAA ACAGCCTTCTGAGGCAGGCTCAAACACTGGCCAACGG CCTCAAGGTTGCGTTCACCACTCCCAGCGGTGTCCCGG ACCCTACCGTCTTCTTCAACCCTACTGTCCGGAGAAGT

		GGTGCATCTAGCAACAACGTCGCTGAAATTGGAAGCC TGGTGCTCGAGTGGACACGGTTGAGCGACCTGACGGG AAACCCGCAGTATGCCCAGCTTGCGCAGAAGGGGCGAG TCGTATCTCCTGAATCCAAAGGGAAGCCCGGAGGCAT GGCCTGGCCTGATTGGAACGTTTGTGTCAGCACGAGCAA CGGTACCTTTCAGGATAGCAGCGGCAGCTGGTCCGGC CTCATGGACAGCTTCTACGAGTACCTGATCAAGATGT ACCTGTACGACCCGGTTGCGTTTGCACACTACAAGGA TCGCTGGGTCCCTTGCTGCCGACTCGACCATTGCGCATC TCGCCTCTCACCCGTCGACGCGCAAGGACTTGACCTTT TTGTCTTCGTACAACGGACAGTCTACGTCGCCAAACTC AGGACATTTGGCCAGTTTTGCCGGTGGCAACTTCATCT TGGGAGGCATTCTCCTGAACGAGCAAAAAGTACATTGA CTTTGGAATCAAGCTTGCCAGCTCGTACTTTGCCACGT ACAACCAGACGGCTTCTGGAATCGGCCCCGAAGGCTT CGCGTGGGTGGACAGCGTGACGGGCGCCGGCGGCTCG CCGCCCTCGTCCCAGTCCGGGTCTACTCGTCGGCAGG ATTCTGGGTGACGGCACCGTATTACATCCTGCGGCCG GAGACGCTGGAGAGCTTGTACTACGCATACCGCGTCA CGGGCGACTCCAAGTGGCAGGACCTGGCGTGGGAAGC G TTCAGTGCCATTGAGGACGCATGCCGCGCCGGCAGC GCGTACTCGTCCATCAACGACGTGACGCAGGCCAACG GCGGGGGTGCCTCTGACGATATGGAGAGCTTCTGGTT TGCCGAGGCGCTCAAGTATGCGTACCTGATCTTTGCG GAGGAGTCGGATGTGCAGGTGCAGGCCAACGGCGGG AACAAATTTGTCTTTAACACGGAGGCGCACCCCTTTA GCATCCGTTTCATCATCACGACGGGGCGGCCACCTTGC TTAA
82	Sequence of the 5'-Region used for knock out of PpSTE13	TTGGGGGCCTCCAGGACTTGCTGAAATTTGCTGACTCA TCTTCGCCATCCAAGGATAATGAGTTAGCTAATGTGA CAGTTAATGAGTCGTCTTGACTAACGGGGAACATTTC ATTATTTATATCCAGAGTCAATTTGATAGCAGAGTTTG TGGTTGAAATACCTATGATTCGGGAGACTTTGTTGTAA CGACCATTATCCACAGTTTGGACCGTGAAAATGTCAT CGAAGAGAGCAGACGACATATTATCTATTGTGGTAAG TGATAGTTGGAAGTCCGACTAAGGCATGAAAATGAGA AGACTGAAAATTTAAAGTTTTTGAACACTAATCGG

		GTAATAACTTGGAATTACGTTTACGTGCCTTTAGCTC TTGTCCTTACCCCTGATAATCTATCCATTTCCCGAGAG ACAATGACATCTCGGACAGCTGAGAACCCGTTTCGATA TAGAGCTTCAAGAGAATCTAAGTCCACGTTCTTCCAAT TCGTCCATATTGGAAAACATTAATGAGTATGCTAGAA GACATCGCAATGATTCGCTTTCCCAAGAATGTGATAA TGAAGATGAGAACGAAAATCTCAATTATACTGATAAC TTGGCCAAGTTTTCAAAGTCTGGAGTATCAAGAAAGA GCTGTATGCTAATATTTGGTATTTGCTTTGTTATCTGG CTGTTTCTCTTTGCCTTGTATGCGAGGGACAATCGATT TTCCAATTTGAACGAGTACGTTCCAGATTCAAACAG
83	Sequence of the 3'-Region used for knock out of PpSTE13	CTACTGGGAACCACGAGACATCACTGCAGTAGTTTCC AAGTGGATTTTCAGATCACTCATTTGTGAATCCTGACAA AACTGCGATATGGGGGTGGTCTTACGGTGGGTTCCT ACGCTTAAGACATTGGAATATGATTCTGGAGAGGTTTT CAAATATGGTATGGCTGTTGCTCCAGTAACTAATTGGC TTTTGTATGACTCCATCTACACTGAAAGATACATGAAC CTTCCAAAGGACAATGTTGAAGGCTACAGTGAACACA GCGTCATTAAGAAGGTTTCCAATTTTAAGAATGTAAA CCGATTCTTGGTTTGTACGGGACTACTGATGATAACG TGCATTTTCAGAACACACTAACCTTACTGGACCAGTTC AATATTAATGGTGTGTGAATTACGATCTTCAGGTGTA TCCCGACAGTGAACATAGCATTGCCCATCACAACGCA AATAAAGTGATCTACGAGAGGTTATTCAAGTGGTTAG AGCGGGCATTTAACGATAGATTTTTGTAAACATTCCGTA CTTCATGCCATACTATATATCCTGCAAGGTTTCCCTTT CAGACACAATAATTGCTTTGCAATTTTACATAACCACCA ATTGGCAAAAATAATCTCTTCAGTAAGTTGAATGCTTT TCAAGCCAGCACCGTGAGAAATTGCTACAGCGCGCAT TCTAACATCACTTTAAAATTCCCTCGCCGGTGCTCACT GGAGTTTCCAACCCTTAGCTTATCAAAATCGGGTGAT AACTCTGAGTTTTTTTTTTTCACTTCTATTCTAAACCTT CGCCCAATGCTACCACCTCCAATCAACATCCCGAAAT GGATAGAAGAGAATGGACATCTCTTGCAACCTCCGGT TAATAATTACTGTCTCCACAGAGGAGGATTTACGGTA ATGATTGTAGGTGGGCCTAATG
84	Sequence of the	CACCTGGGCCTGTTGCTGCTGGTACTGCTGTTGGAAC

	5'-Region used for knock out of PpDAP2	GTTGGTATTGTTGCTGATCTAAGGCCGCCTGTTCCACA CCGTGTGTATCGAATGCTTGGGCAAAATCATCGCCTG CCGGAGGCCCCACTACCGCTTGTTCCCTCCTGCTCTTGT TTGTTTTGCTCATTGATGATATCGGCGTCAATGAATTG ATCCTCAATCGTGTGGTGGTGGTGTCTGATTCCCTCTT CTTTCTTGAGTGCCTTATCCATATTCCTATCTTAGTGTA CCAATAATTTTGTTAAACACACGCTGTTGTTTATGAAA AGTCGTCAAAAGGTTAAAAATTCTACTTGGTGTGTGTC AGAGAAAGTAGTGCAGACCCCCAGTTTGTTGACTAGT TGAGAAGGCGGCTCACTATTGCGCGAATAGCATGAGA AATTTGCAAACATCTGGCAAAGTGGTCAATACCTGCC AACCTGCCAATCTTCGCGACGGAGGCTGTAAAGCGGG TTGGGTTCCTCAAAGTGAATGGATATTACGGGCAGGAA AAACAGCCCCTTCACACTAGTCTTTGCTACTGACATC TTCCCTCTCATGTATCCCGAACACAAGTATCGGGAGTA TCAACGGAGGGTGCCCTTATGGCAGTACTCCCTGTTG GTGATTGTACTGCTATACGGGTCTCATTGCTTATCAG CACCATCAACTTGATACACTATAACCACAAAAATTAT CATGCACACCCAGTCAATAGTGGTATCGTTCTTAATGA GTTTGCTGATGACGATTCAATTCTCTTTGAATGGCACTC TGAACCTGGAGAACTGGAGAAATGGTACCTTTTCCCC TAAATTTCAATCCATTCACTGGACCGAAATAGGTCAG GAAGATGACCAGGGATATTACATTCTCTCTTCCAATTC CTCTTACATAGTAAAGTCTTTATCCGACCCAGACTTTG AATCTGTTCTATTCAACGAGTCTACAATCACTTACAAC G
85	Sequence of the 3'-Region used for knock out of PpDAP2	GGCAGCAAAGCCTTACGTTGATGAGAATAGACTGGCC ATTTGGGGTTGGTCTTATGGAGGTTACATGACGCTAAA GGTTTTAGAACAGGATAAAGGTGAAACATTCAAATAT GGAATGTCTGTTGCCCTGTGACGAATTGGAAATTCTA TGATTCTATCTACACAGAAAGATACATGCACACTCCTC AGGACAATCCAACTATTATAATTCGTCAATCCATGA GATTGATAATTTGAAGGGAGTGAAGAGGTTCTTGCTA ATGCACGGAACCTGGTGACGACAATGTTCACTTCCAAA ATACACTCAAAGTTCTAGATTTATTTGATTTACATGGT CTTGAAAACATATGATATCCACGTGTTCCCTGATAGTGA TCACAGTATTAGATATCACAACGGTAATGTTATAGTGT

		ATGATAAGCTATTCCATTGGATTAGGCGTGCATTCAA GGCTGGCAAATAAATAGGTGCAAAAATATTATTAGAC TTTTTTTTCGTTTCGCAAGTTATTACTGTGTACCATACC GATCCAATCCGTATTGTAATTCATGTTCTAGATCCAAA ATTTGGGACTCTAATTCATGAGGTCTAGGAAGATGAT CATCTCTATAGTTTTTCAGCGGGGGGCTCGATTTGCGGT TGGTCAAAGCTAACATCAAAATGTTTGTGAGGTTTCAG TGAATGGTAACTGCTGCTCTTGAATTGGTCGTCTGACA AATTCTCTAAGTGATAGCACTTCATCTACAATCATTG CTTTCATCGTTTCTATATCGTCCACGACCTCAAACGAGA AATCGAATTTGGAAGAACAGACGGGCTCATCGTTAGG ATCATGCCAAACCTTGAGATATGGATGCTCTAAAGCC TCAGTAACTGTAATTCTGTGAGTGGGATCTACCGTGA GCATTCGATCCAGTAAGTCTATCGCTTCAGGGTTGGCA CCGGGAAATAACTGGCTGAATGGGATCTTGGGCATGA ATGGCAGGGAGCGAACATAATCCTGGGCACGCTCTGA TCTGATAGACTGAAGTGTCTCTTCCGAAACAGTACCC AGCGTACTCAAAATCAAGTTCAATTGATCCACATAGT CTCTTCCTCTAAAAATGGGTTCGGCCACCTA
86	Sequence of the <i>PpTHR1</i> in loci	GGCCAGCCCATCACCATGAATGCTTAAACGCCAACT CCTTCCATCTCATTTCGTACCAGATTATGACTCTTAG GCGGGGAGAATCCCGTCCAGCATAGCGAACATTTCTT TTTTTTTTTTTTTTCGTTTCGCATCTCTCTATCGCATTCA GAAAAAATAACATATAATTCTTCCAGTTTCCGTCATTC ATTACGTTTAAAACTACGAAAGTTTATGCTCTCTTTTG TTTTGTTCCTAGATTTCGAAATATTTTCTTTATTGAGT TTAATTTGTGTGGCAGACAATGGTTAGATCTTTCACCA TCAAAGTGCCTGCTTCCTCAGCAAATATAGGACCGGG GTTTGACGTTCTGGGAATTGGTCTCAACCTTTACTTGG AACTACAAGTCACCATTGATCCCAAAATTGATACCTC AAGCGATCCAGAAAATGTGTTATTGTCGTATGAAGGT GAGGGGGCTGATGAGGTGTCATTGAAAAGTGACGAAA ACTTGATTACGCGCACAGCTCTCTATGTTCTACGTTGT GACGACGTCAGGACTTTCCTTAAGGGAACCAAGATTC ACGTCATTAACCCTATTCCTCTAGGAAGAGGCTTGGG ATCTTCGGGTGCTGCAGTTGTCGCCGGTGCAATTGCTCG GAAATTCATCGGACAGCTTGGATACTCCAAACAACG

		TTTACTGGATTACTGTTTGATGATAGAACGTCATCCAG ATAACATCACCGCAGCTATGGTGGGTGGTTTCGTTGG ATCTTATCTTAGAGATCTTTCACCAGAAGACACCCAG AGAAAAGAGATTCCATTAGCAGAAGTCCTGCCAGAAC CTCAAGGTGGTATTAACACCGGTCTCAACCCACCAGT GCCTCCAAAAAACATTGGGCACCACATCAAATACGGC TGGGCAAAAGAGATCAAATGTATTGCCATTATTCCAG ACTTTGAAGTATCAACCGCTTCATCTAGAGGCGTTCTT CCAACCACTTACGAGAGACATGACATTATTTTCAACCT GCAAAGGATAGCCGTTCTTACCACTGCCCTGACACAA TCTCCACCAGATCCAAGCTTGATATACCCAGCTATGCA GGACAGGATTCACCAACCTTACAGGAAAACCTTGATC CACGGACTGACTGAAATACTGTCTTCATTCACCCCAG AATTACACAAAGGTTTGTGTTGGGAATCTGTCTTCCGGT GCTGGGCCCACAATATTAGCCCTCGCAACTGAAAAC TCGATCAGATTGCTAAGGACATCATTGCCAGATTGCT GTCGAAGACATCACCTGTAGTTGGAACTCTTGACCC CAGCTCTTGAAGGTTCTGTTGTTGAGGAGCTTGCTTAA TAGAAATTAGAACATCCTCTTTAGATTATGATAATACG TTTTTAACCTTTTCCCCTAACTGTAGTGATGGTATCTGA CCCTCTTAGACCTTAGGTTGGACCTTCTCGAATTTCCT GCCTCTATCAAAAATCCGACCCTCGACATCGTTTACGT ACTTTGCAACCAATTAAGTACCGGCAGACGTTCA GTGATCATGGCTCTCTATACAAATACCCTGATAACGTT TGCATTCTGACAGTCGGAGGATGTACGTGCTTATTTT CTTGCTAGTCCCAAATGTTTTGAGATTGCTCCAATCGT TTTTTCAACAATACTAACTGCCAACAAATAGATCTTTT ATTCAACGGAAATGGGGAACAATTCAACGTGGGTGAC TTTTTGGAGACTACATCTCCCTATATGTGGGCAAATCT GGGTATAGCAAGTTGCATTGGATTCTCGGTCAATTGGTG CTGCATGGGGAATTTTCATAACAGGTTCTTCGATCATC GGTGCAGGTGTCAAAGCTCCCAGAATCACAACAAAA ATTTAATCTCCATCATTTTCTGTGAGGTGGTGGCTATT TATGGGCTTATTATGGCC
87	Sequence of <i>PpHIS3</i> 5' integration	CCTGTGAGTCTGGCTCAATCACTTTTCAAAGATAAGG ACTATTCTGCAGAACATGCAGCCCAGGCAACATCATC CCAGTTCATCTCTGTGAACACAGGAATAGGATTCCTG

	fragment	GACCATATGTTACACGCACTTGCTAAGCACGGCGGCT GGTCTGTCATTATCGAATGTGTAGGTGATTTGCACATT GATGACCATCATTTCAGCAGAAGATACTGGAATCGCAT TGGGGATGGCATTCAAAGAAGCCTTGGGCCATGTTTCG TGGTATCAAAAGATTTCGGGTCCGGATTGCTCCACTA GACGAAGCTCTCAGTCGGGCTGTTATTGATATGTCTAA CAGGCCCTATGCTGTTGTGATCTGGGTTTGAAAAGA GAGAAGATTGGAGACCTATCGTGTGAGATGATTCCCC ATGTTTTGGAAAGTTTTGCCCAAGGAGCCCATGTAAC CATGCACGTAGATTGTTTGCGAGGTTTCAACGACCATC ATCGTGCCGAGAGTGCATTCAAAGCTTTGGCTATAGC TATCAAAGAGGCCATTTCAAGCAACGGCACGGACGAC ATTCCAAGTACGAAGGGTGTCTTTTCTGA
88	Sequence of <i>PpHIS3</i> 3' integration fragment	GTCTGGAAGGTGTCTACATCTGTGAAATCCGTATTTAT TTAAGTAAAACAATCAGTAATATAAGATCTTAGTTGG TTTACCACATAGTCGGTACCGGTCGTGTGAACAATAG TTCAATGCCTCCGATTGTGCCTTATTGTTGTGGTCTGC ATTTTCGCGGGCGAAATTTCTACTTCAGATCGGGGCTGA GATGACCTTAGTACTCACATCAACCAGCTCGTTGAAA GTTCCACATGACCACTCAATGTTTAATAGCTTGGCAC CCATGAGGTTGAAGAACTACTTAAGGTGTTTTGTGC CTCAGTAGTGCTGTTAGCGGCGACATCTGTGGTGTAT TTTTCCACTTTGGAGGTCAGATCATAATCCCCATACCG GAACGCACTGTGACCTTAAGTACTCCTCCCGCAAACG ATACTTGGCAGTTTCAACAGTTCTTCAACGGCTATTTA GACGCCCTGTTAGAGAATAACCTGTCGTATCCGATAC CAGAAAGGTGGAATCATGAAGTTACAAATGTAAGATT CTTCAATCGCATAGGTGAATTGCTCTCGGAGAGTAGG CTACAGGAGCTGATTCATTTTAGTCCTGAGTTCATAGA GGATACCAGTGACAAATTCGACAATATTGTTGAACAA ATTCCAGCAAAATGGCCTTACGAAAACATGTACAGAG GAGATGGATACGTTATTGTTGGTGGTGGCAGACACAC CTTTTTGGCACTGCTGAATATCAACGCTTTGAGAAGAG CAGGCAATAAACTGCCAGTTGAGGTCGTGTTGCCAAC TTACGACGACTATGAGGAAGATTTCTGTGAAAATCAT TTTCCACTTTTGAATGCAAGATGCGTAATCTTAGAAGA ACGATTTGGTGACCAAGTTTATCCCCGGTTACAAC TAG

		GAGGCTACCAGTTTAAAATATTTGCGATAGCAGCAAG TTCATTCAAAAACGTCTTTTTGTTAGATTGAGATAATA TACCCTTGCGAAAGATGGATAAGATATTCTCAAGCGA ACTATACAAGAATAAGACAATGATTACTTGGCCAGAC T
89	Sequence of <i>PpTHR1</i> 5' integration fragment	CGAGTCGGCCAGCCCATCACCATGAATGCTTAAAACG CCAACTCCTTCCATCTCATTTTCGTACCAGATTATGAC TCTTAGGCGGGGAGAATCCCGTCCAGCATAGCGAACA TTTCTTTTTTTTTTTTTTTTTTCGTTTCGCATCTCTCTATCG CATTCAGAAAAAATAACATATAATTCTTCCAGTTTCCG TCATTCAATTACGTTTAAAACACTACGAAAGTTTATAGCTCT CTTTTGTTTTTGTTCCTAGATTCGAAATATTTTCTTTA TTGAGTTTAATTTGTGTGGCAGACAATGGTTAGATCTT TCACCATCAAAGTGCCTGCTTCCTCAGCAAATATAGG ACCGGGGTTTGACGTTCTGGGAATTGGTCTCAACCTTT ACTTGGAACATAAGTCACCATTGATCCCAAATGA TACCTCAAGCGATCCAGAAAATGTGTTATTGTCGTATG AAGGTGAGGGGGCTGATGAGGTGTCATTGAAAAGTGA CGAAAACCTTGATTACGCGCACAGCTCTCTATGTTCTAC GTTGTGACGACGTCAGGACTTCCCTAAGGGAACCAA GATTCACGTCATTAACCCTATTCCTCTAGGAAGAGGCT TGGGATCTTCGGGTGCTGCAGTTGTC
90	Sequence of <i>PpTHR1</i> 3' integration fragment	TAGAAATTAGAACATCCTCTTTAGATTATGATAATACG TTTTTAACTTTTCCCCTAACTGTAGTGATGGTATCTGA CCCTCTTAGACCTTAGGTTGGACCTTCTCGAATTTCTT GCCTCTATCAAAAATCCGACCCTCGACATCGTTTACGT ACTTTGCAACCAATTAAGTAGTACCGGCAGACGTTCA GTGATCATGGCTCTCTATACAAATACCCTGATAACGTT TGCATTCCTGACAGTCGGAGGATGTACGTGCTTATTTT CTTGCTAGTCCCAAATGTTTTGAGATTGCTCCAATCGT TTTTTCAACAATACTAACTGCCAACAATAGATCTTTT ATTCAACGGAAATGGGGAACAATTCAACGTGGGTGAC TTTTTGGAGACTACATCTCCCTATATGTGGGCAAATCT GGGTATAGCAAGTTGCATTGGATTCTCGGTCATTGGTG CTGCATGGGGAATTTTCATAACAGGTTCTTCGATCATC GGTGCAGGTGTCAAAGCTCCCAGAATCACAACAAAAA ATTTAATCTCCATCATTTTCTGTGAGGTGGTGGCTATT

		TATGGGCTTATTATGGCCATTGT
91	Sequence of the 5'-Region used for knock out of <i>PpVPS10-1</i>	AAGTGGGCCAGATTATATAAATATGGATCAACATGAA GCCTTGAAAGATTTCAAGGACAGGCTTAGGAATTACG AAAAAGTTTACGAGACTATTGACGACCAGGAGGAAGA GGAGAACGAACGGTACAATATTCAGTATCTGAAGATA ATCAACGCAGGAAAGAAGATAGTCAGTTATAACATAA ATGGGTATTTATCGTCCCACACCGTTTTTTATCTCCTG AATTTCAATCTTGCAGAACGTCAAATATGGTTGACGA CGAATGGAGAGACAGAGTATAACCTTCAAATAGGAT TGGAGGTGATTCCAAATTAAGCAATGAGGGATGGAAA TTTGCCAAAGCATTGCCCAAGTTTATAGCACAGAAAA GAAAAGAGTTTCAACTTAGACAGTTGACCAAACACTA TATCGAGACTCAAACGCCCATTTGAAGACGTACCGTTG GAGGAGCACACCAAGCCAGTCAAATATTCTGATCTGC ATTTCCATGTTTGGTCATCGGCTTTAAAGAGATCTACT CAATCAACAACATTTTTTCCATCGGAAAATTACTCTCT GAAGCAATTCAGAACGTTGAATGATCTCTGTTGCGGA TCACTGGATGGTTTGACTGAACAAGAGTTCAAAGTA AATACAAAGAAGAATACCAGAATTCTCAGACTGATAA ACTGAGTTTCAGTTTCCCTGGTATCGGTGGGGAGTCTT ATTTGGACGTGATCAACCGTTTGAGACCACTAATAGTT GAACTAGAAAGGTTGCCAGAACATGTCCTGGTCATTA CCCACCGGGTCATAGTAAGGATTTTACTAGGATATTTT ATGAATTTGGATAGAAATCTGTTGACAGATTTGGAAA TTTTGCATGGGTATGTTTATTGTATTGAGCCGAAACCT TATGGTTTAGACTTAAAGATCTGGCAGTATGATGAGG CGGACAACGAGTTTAATGAAGTTGATAAGCTGGAATT CATGAAAAGAAGAAGAAAATCGATCAACGTCAACAC GACAGATTTCAGAATGCAGTTAAACAAAGAGTTGCAA CAGGACGCTCTCAATAATAGTCCTGGTAATAATAGTC CGGGCGTATCATCTCTATCTTCATACTCGTCGTCCTCT TCCCTTTCCGCTGACGGGAGCGAGGGAGAAACATTAA TACCACAAGTATCCCAGGCGGAGAGCTACAACCTTTGA ATTTAACTCTCTTTCATCATCAGTTTCATCGTTGAAAA GGACGACATCTTCTTCCCAACATTTGAGCTCCAATCCT AGTTGTCTGAGCATGCATAATGCCTCATTGGACGAGA ATGACGACGAACATTTAATAGACCCGGCTTCTACAGA

		CGACAAGCTAAACATGGTATTACAGGACAAAACGCTA ATTAAAAAGCTCAAAAGTTTACTACTTGACGAGGCCG AAGGCTAGACAATCCACAGTTAATTTTGATACTGTACT TTATAACGAGTAACATACATATCTTATGTAATCATCTA TGTCACGTCACGTGCGCGCGACATTATTCCGAGAACTT GCGCCCTGCTAGCTCCACTGTCAGAGTGATAACTTCCC CAAAATAGGATCCAACGTGTTTCCAATTGCTTTTGAAA TGTGGATTGAAAGAAACCTCATAGCGTAA
92	Sequence of the 3'-Region used for knock out of <i>PpVPS10-1</i>	GACGACGAGGAGAATATCAATTTTGATTCCCGGTAGA TAGCTCACCCACGGTCACACACACAAACACACATACA CATTAACACACAGAGTTATTAGTTAACAGAGAAAAC CTAACAAAGTATTTATTTTCGTTACGTAATCCGACTTT TCTTTTACCGTTTTCTATTGCTCCTCTCATTTGCCCT AAAAGTTGCTCCTCATTACTAAAATCACCACACCATG CTCGAATATGATGTTACTAAATGCAAATTGTAGTCGTG CCTCTTGTGGTAATACTATAGGGAATATCTCTCGATTA CTCGATTCTGGTTAATTTTTCTTTTTTATAGGGGAAG TTTTTTTTCTTCCCCTTTCTCTCCAGTTTATTTATTAC TAAGAAAATCCAACAGATACCAACCACCCAAAAAGAT CCTAACAGCCTGTTTTTGAGGAGTTTTTCAGCAGCTA AGCTTCATCAGTTTTTAACTTAATTTATTGCCCTTC ACTTTGTTTCTTGTGGCTTTTAAGGCTCTCCGGAACAG CGGTTTCAAATCAAATCTCAGTTATTTGTTTGCTCCG CTTTGTCAGTTCAAAGATCATGGTTTCCGAAAACAAG AATCAATCTTCGATTTTGATGGACAACCTCCAAGAAGC TCTCTCCGAAGCCCATTTTGAATAACAAGAATGAACC GTTTGGCATCGGCGTCGATGGACTTCAACATCCTCAAC CGACTTTATGCCGCACAGAATCGGAACTCTTGTTCAAC TTGAGCCAAGTCAATAAATCCCAAATAACTTTGGACG GTGCAGTTACTCCACCTGCTGATGGTAATGGGAATGA AGCAAAAAGAGCAAATCTCATCTCTTTTGATGTTCCAT CGTCTCAAGTGAAACATAGAGGGTCTATTAGTGCAAG GCCCTCGGCAGTGAATGTGTCCCAAATTACCGGGGCC CTTCTCAATCCGGATCTTCTAGAAATCCCTACGATCA AACACAGTCACCTCCACCTAGCACTTACGCCTCCAGG CAGAACTCCACCCATGGAAATAATATCGATAGCTTGC AATATTTGGCAACAAGAGATCTTAGTGCTTTAAGGCT

		GGAAAGAGATGCTTCCGCACGAGAAGCTACCTCTTCT GCAGTGTCCACTCCTGTTTCAGTTCGATGTACCCAAACA ACATCATCTCCTTCATTTAGAACAAGACCCGACAAGG CCCATCC
93	Sequence of <i>PpTRP5</i> 5' integration fragment	ACGACGGCCAAATTCATGATACACACTCTGTTTCAGCT GGTTTGGACTACCCTGGAGTTGGTCCTGAATTGGCTGC CTGGAAAGCAAATGGTAGAGCCCAATTTTCCGCTGTA ACTGATGCCCAAGCATTAGAGGGATTCAAAATCCTGT CTCAATTGGAAGGGATCATTCCAGCACTAGAGTCTAG TCATGCAATCTACGGCGCATTGCAAATTGCAAAGACT ATGTCTTCGGACCAGTCCTTAGTTATTAATGTATCTGG AAGGGGTGATAAGGACGTCCAGAGTGTAGCTGAGATT TTACCTAAATTGGGACCTCAAATTGGATGGGATTTGC GTTTCAGCGAAGACATTACTAAAGAGTGA
94	Sequence of <i>PpTRP5</i> 3' integration fragment	TCGATAGCACAATATTCAACTTGACTGGGTGTTAAGA ACTAAGAGCTCTGGGAACTTTGTATTTATTACTACCA ACACAGTCAAATTATTGGATGTGTTTTTTTTTCCAGTA CATTTCACTGAGCAGTTTGTATACTCGGTCTTTAATC TCCATATACATGCAGATTGTAATACAGATCTGAACAG TTTGATTCTGATTGATCTTGCCACCAATATTCTATTTT GTATCAAGTAACAGAGTCAATGATCATTGGTAACGTA ACGGTTTTTCGTGTATAGTAGTTAGAGCCCATCTTGTA CCTCATTTCCCTCCCATATTAAAGTATCAGTGATTGCT GGAACGATTAATAAGAAAAAATATCTGCACAT ACTCATCAGTCTGTAAATCTAAGTCAAACTGCTGTAT CCAATAGAAATCGGGATATACCTGGATGTTTTTCCAC ATAAACAAACGGGAGTTCAGCTTACTTATGGTGTTGA TGCAATTCAGTATGATCCTACCAATAAACGAACTT TGGGATTTTGGCTGTTTGAGGGATCAAAGCTGCACC TTTACAAGATTGACGGATCGACCATTAGACCAAAGCA AATGGCCACCAA

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described

5 herein will become apparent to those skilled in the art from the foregoing description and the

accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Patents, patent applications, Genbank Accession Numbers and publications are cited throughout this application, the disclosures of which, particularly, including all disclosed
5 chemical structures and antibody amino acid sequences therein, are incorporated herein by reference. Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. All references cited herein are incorporated by
10 reference to the same extent as if each individual publication, patent application, or patent, was specifically and individually indicated to be incorporated by reference.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. Various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

15

WHAT IS CLAIMED:

1. A composition comprising a fragment of recombinant human tumor necrosis factor receptor fused to the constant region of an antibody (TNFRII-Fc) wherein the TNFRII-Fc has *N*-glycans and *O*-glycans and wherein the *O*-glycans are of the dystroglycan- or *O*-mannose reduced glycans, and pharmaceutically acceptable salts thereof.
2. The composition of claim 1, wherein the *N*-glycans and *O*-glycans on the TNFRII-Fc are predominantly sialylated with α 2,6 or α 2,3 sialic acid residues.
3. The composition of claim 1, wherein the *N*-glycans on the TNFRII-Fc lack fucose residues.
4. The composition of claim 1, wherein the *N*-glycans and *O*-glycans on the TNFRII-Fc which are sialylated comprise *N*-acetylneuraminic acid (NANA) and no *N*-glycolylneuraminic acid (NGNA).
5. The composition of claim 1, wherein a ratio of mole sialic acid to mole of the TNFRII-Fc is at least 10.
6. The composition of claim 5, wherein a ratio of mole sialic acid to mole of the TNFRII-Fc is about 10 to 21.
7. The composition of claim 5, wherein a ratio of mole sialic acid to mole of the TNFRII-Fc is greater than 21.
8. The composition of claim 1, wherein the *N*-glycans on the TNFRII-Fc are predominantly mono-, bi-, tri-, or tetra-sialylated *N*-glycans.
9. The composition of claim 1, wherein the *O*-glycans on the TNFRII-Fc are predominantly sialylated *O*-glycans.
10. The composition of claim 1, wherein greater than 40% of the *O*-glycans on the TNFRII-Fc are sialylated *O*-glycans.
11. The composition of claim 1, wherein about 20% of the *O*-glycans on the TNFRII-Fc are of the mannose type or a combination of mannose and mannanose types.

12. The composition of claim 1, wherein less than 50% of *O*-glycans on the TNFRII-Fc possess terminal mannose.

5 13. The composition of claim 1, wherein the TNRFII domain of the TNFRII-Fc has an amino acid sequence with at least 90% identity to the amino acid sequence set forth in SEQ ID NO:73 or 75.

10 14. A method for producing a recombinant human tumor necrosis factor receptor fused to the constant region of an antibody (TNFRII-Fc) having sialylated *N*-glycans and *O*-glycans comprising;

(a) providing a recombinant yeast host cell genetically engineered to produce glycoproteins having sialylated *N*-glycans and further comprising

15 (i) a nucleic acid molecule encoding the TNFRII-Fc;

(ii) a nucleic acids molecule encoding an α 1,2-mannosidase activity linked to a heterologous targeting or signaling peptide that targets the mannosidase activity to the secretory pathway; and

(iii) a nucleic acid molecule encoding an *O*-linked mannose β 1,2-*N*-acetylglucosaminyltransferase I (POMGnT I);

20 (b) culturing the host cell under conditions suitable for producing the TNFRII-Fc; and

(c) recovering the TNFRII-Fc from the culture fluid to produce the TNFRII-Fc having sialylated *N*-glycans and *O*-glycans.

25 15. The method of claim 14, wherein the *N*-glycans and *O*-glycans on the TNFRII-Fc are predominantly sialylated with α 2,6 or α 2,3 sialic acid residues.

16. The method of claim 14, wherein the *N*-glycans on the TNFRII-Fc lack fucose residues.

30 17. The method of claim 14, wherein the *N*-glycans and *O*-glycans on the TNFRII-Fc which are sialylated comprise *N*-acetylneuraminic acid (NANA) and no *N*-glycolylneuraminic acid (NGNA).

35 18. The method of claim 14, wherein a ratio of mole sialic acid to mole of the TNFRII-Fc is at least 10.

19. The method of claim 18, wherein a ratio of mole sialic acid to mole of the TNFRII-Fc is about 10 to 21.

20. The method of claim 18, wherein a ratio of mole sialic acid to mole of the TNFRII-Fc is greater than 21.

21. The method of claim 14, wherein the *N*-glycans on the TNFRII-Fc are predominantly mono-, bi-, tri-, or tetra-sialylated *N*-glycans.

22. The method of claim 14, wherein the *O*-glycans on the TNFRII-Fc are predominantly sialylated *O*-glycans.

23. The method of claim 14, wherein greater than 40% of the *O*-glycans on the TNFRII-Fc are sialylated *O*-glycans.

24. The method of claim 14, wherein less than 50% of *O*-glycans on the TNFRII-Fc possess terminal mannose.

25. The method of claim 14, wherein about 20% of the *O*-glycans on the TNFRII-Fc are of the mannose type or a combination of mannose and mannanose types.

26. The method of claim 14, wherein the TNFRII domain of the TNFRII-Fc has an amino acid sequence with 90% identity to the amino acid sequence set forth in SEQ ID NO:73 or 75.

27. The method of claim 14, wherein the TNFRII-Fc is recovered from the culture fluid in a process comprising a hydroxyapatite or aminophenyl borate chromatography step.

28. A pharmaceutical composition comprising the polypeptide of any one of claims 1 to 13 and a pharmaceutically suitable carrier.

29. Use of the pharmaceutical composition of claim 27 in the manufacture of a medicament for inflammatory diseases and cancers that display an increased and/or unregulated level of soluble TNFRII or polymorphisms.

30. Use of the pharmaceutical composition of claim 27 in the manufacture of a medicament for treating rheumatoid arthritis.

1/53

Construction of YGLY11731, YGLY10299, and YGLY13571

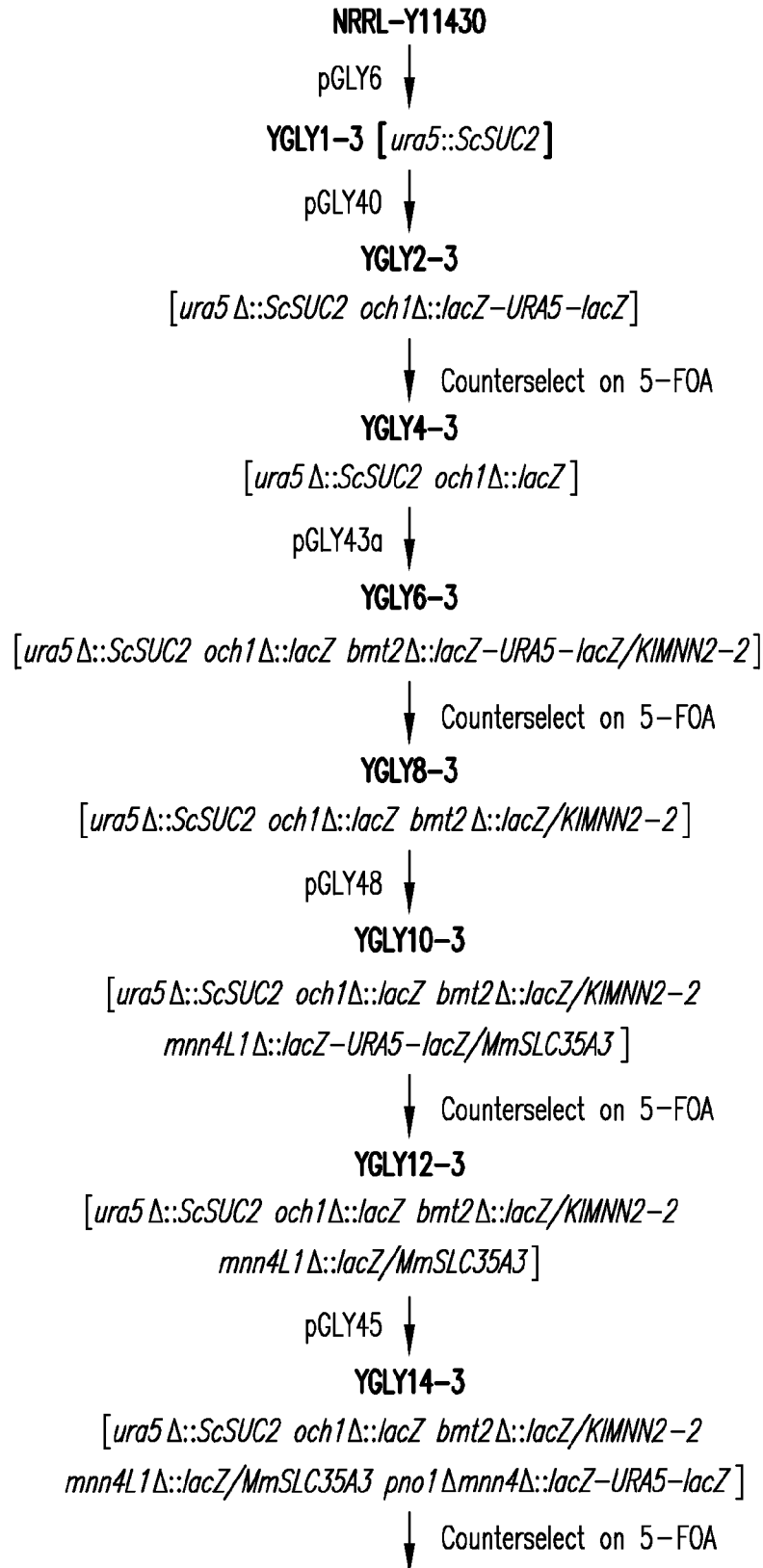


FIG. 1A

2/53

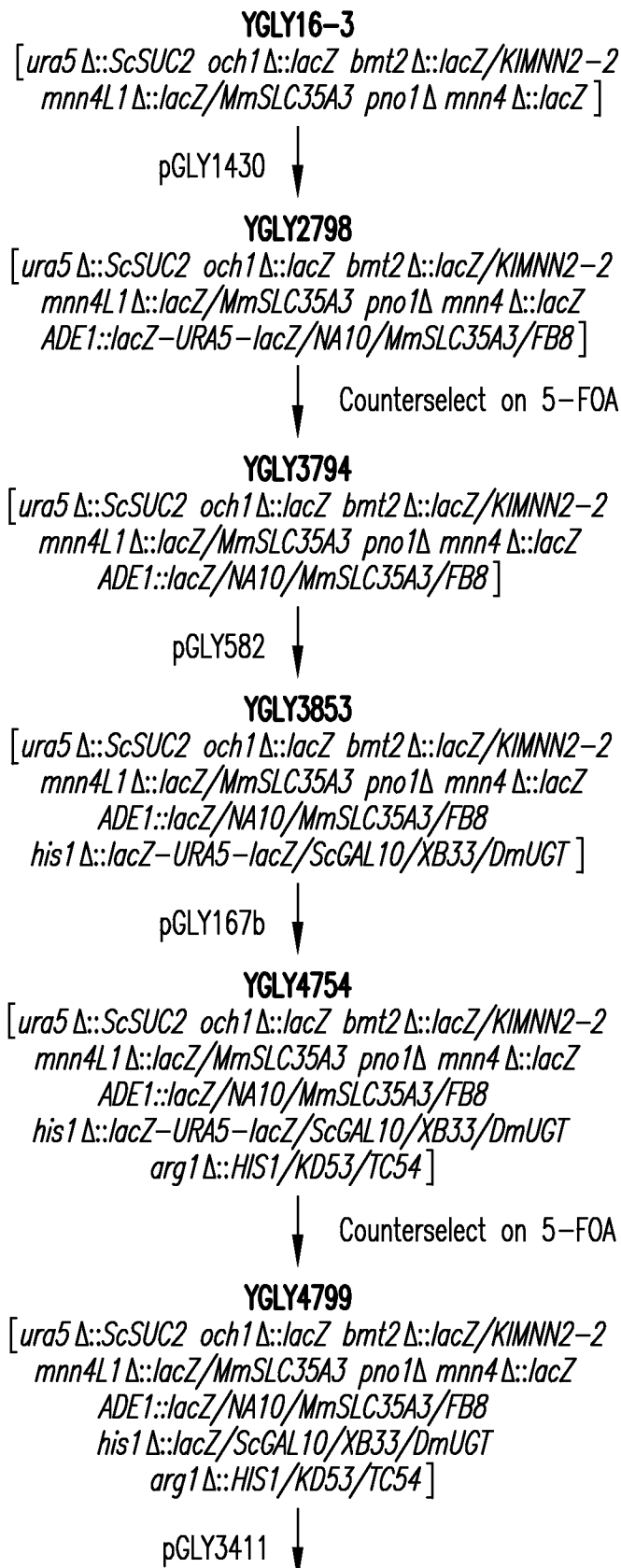


FIG. 1 B

3/53

YGLY6903

[*ura5*Δ::*ScSUC2 och1*Δ::*lacZ bmt2*Δ::*lacZ/KIMNN2-2*
*mnn4L1*Δ::*lacZ/MmSLC35A3 pno1*Δ *mnn4*Δ::*lacZ*
ADE1::*lacZ/NA10/MmSLC35A3/FB8*
*his1*Δ::*lacZ-URA5 lacZ/ScGAL10/XB33/DmUGT*
*arg1*Δ::*HIS1/KD53/TC54*
*bmt4*Δ::*lacZ-URA5-lacZ*]



Counterselect on 5-FOA

YGLY7432

[*ura5*Δ::*ScSUC2 och1*Δ::*lacZ bmt2*Δ::*lacZ/KIMNN2-2*
*mnn4L1*Δ::*lacZ/MmSLC35A3 pno1*Δ *mnn4*Δ::*lacZ*
ADE1::*lacZ/NA10/MmSLC35A3/FB8*
*his1*Δ::*lacZ/ScGAL10/XB33/DmUGT*
*arg1*Δ::*HIS1/KD53/TC54*
*bmt4*Δ::*lacZ*]

pGLY3419

**YGLY7651**

[*ura5*Δ::*ScSUC2 och1*Δ::*lacZ bmt2*Δ::*lacZ/KIMNN2-2*
*mnn4L1*Δ::*lacZ/MmSLC35A3 pno1*Δ *mnn4*Δ::*lacZ*
ADE1::*lacZ/NA10/MmSLC35A3/FB8*
*his1*Δ::*lacZ/ScGAL10/XB33/DmUGT*
*arg1*Δ::*HIS1/KD53/TC54*
*bmt4*Δ::*lacZ bmt1*Δ::*lacZ-URA5-lacZ*]



Counterselect on 5-FOA

YGLY7930

[*ura5*Δ::*ScSUC2 och1*Δ::*lacZ bmt2*Δ::*lacZ/KIMNN2-2*
*mnn4L1*Δ::*lacZ/MmSLC35A3 pno1*Δ *mnn4*Δ::*lacZ*
ADE1::*lacZ/NA10/MmSLC35A3/FB8*
*his1*Δ::*lacZ/ScGAL10/XB33/DmUGT*
*arg1*Δ::*HIS1/KD53/TC54*
*bmt4*Δ::*lacZ bmt1*Δ::*lacZ*]

pGLY3421

**YGLY7961**

[*ura5*Δ::*ScSUC2 och1*Δ::*lacZ bmt2*Δ::*lacZ/KIMNN2-2*
*mnn4L1*Δ::*lacZ/MmSLC35A3 pno1*Δ *mnn4*Δ::*lacZ*
ADE1::*lacZ/NA10/MmSLC35A3/FB8*
*his1*Δ::*lacZ/ScGAL10/XB33/DmUGT*
*arg1*Δ::*HIS1/KD53/TC54*
*bmt4*Δ::*lacZ bmt1*Δ::*lacZ bmt3*Δ::*lacZ-URA5-lacZ*]

pGLY2456

**FIG. 1C**

4/53

YGLY8146

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ-URA5-lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33]



Counterselect on 5-FOA

YGLY9296

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33]

pGLY5048

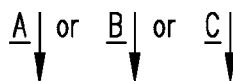
**YGLY9469**

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1]

pGLY5019

**YGLY9795 & YGLY9797**

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2Δ::Nat^R]

**FIG. 1D**

5/53

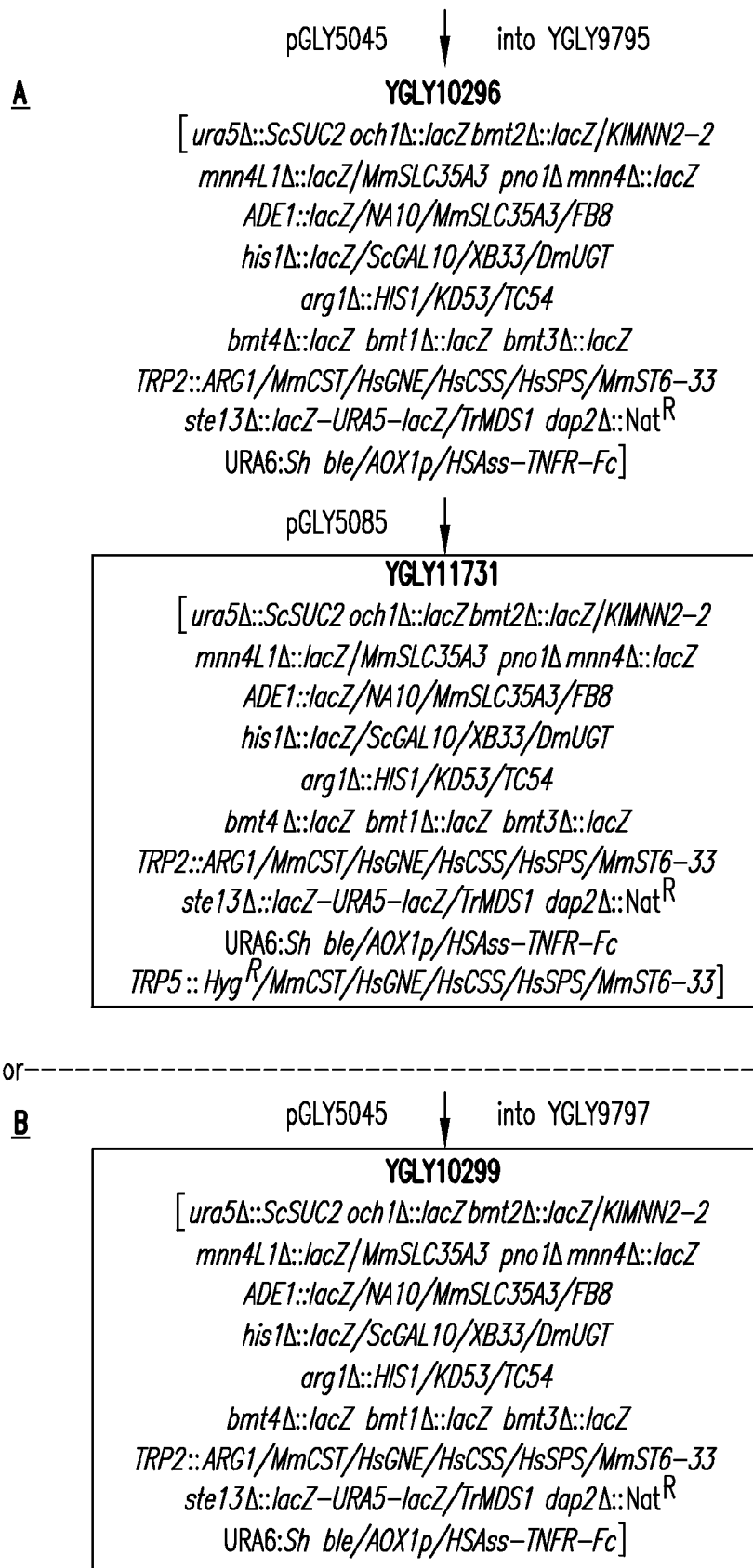


FIG.1E

6/53

C

pGLY6391 ↓ into YGLY9797

YGLY12626

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2 Δ::Nat^R
THR1:Sh ble/AOX1p/HSAss-TNFR-Fc]

pGLY5085 ↓

YGLY13430

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2 Δ::Nat^R
THR1:Sh ble/AOX1p/HSAss-TNFR-Fc
TRP5::Hyg^R/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33]

↓ Counterselect on 5-FOA

YGLY13571

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ/TrMDS1 dap2 Δ::Nat^R
THR1:Sh ble/AOX1p/HSAss-TNFR-Fc
TRP5::Hyg^R/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33]

FIG. 1F

7/53

Glossary:

ScSUC2	<i>S. cerevisiae</i> Invertase
OCH1	Alpha-1, 6-mannosyltransferase
K1MNN2-2:	<i>K. lactis</i> UDP-GlcNAc transporter
BMT1:	Beta-mannose-transfer (beta-mannose elimination)
BMT2:	Beta-mannose-transfer (beta-mannose elimination)
BMT3:	Beta-mannose-transfer (beta-mannose elimination)
BMT4:	Beta-mannose-transfer (beta-mannose elimination)
MNN4L1:	MNN4-like 1 (charge elimination)
MmSLC35A3	Mouse homologue of UDP-GlcNAc transporter
PN01:	Phosphomannosylation of N-linked oligosaccharides (charge elimination)
MNN4:	Mannosyltransferase (charge elimination)
ScGAL10	UDP-glucose 4-epimerase
XB33	Truncated HsGalT1 fused to ScKRE2 leader
DmUGT	UDP-Galactose transporter
KD53	Truncated DmMNSII fused to ScMNN2 leader
TC54	Truncated RnGNTII fused to ScMNN2 leader
NA10	Truncated HsGNTI fused to PpSEC12 leader
FB8:	Truncated MmMNS1A fused to ScSEC12 leader
MmCST	Mouse CMP-sialic acid transporter
HsGNE	Human UDP-GlcNAc 2-epimerase/N-acetylmannosamine kinase
HsCSS	Human CMP-sialic acid synthase
HsSPS	Human N-acetylneuraminase-9-phosphate synthase
MmST6-33	Truncated Mouse alpha-2,6-sialyl transferase fused to ScKRE2 leader
TrMDS1:	Secreted <i>T. reesei</i> MNS1
STE13	Golgi dipeptidyl aminopeptidase
DAP2	Vacuolar dipeptidyl aminopeptidase
NatR	Nourseothricin resistance marker
Sh ble:	Zeocin resistance marker
ARSR	Arsenite resistance marker
HYGR	Hygromycin resistance marker

FIG.1 G

8/53

Construction of YGLY12680

YGLY10299

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2 Δ::Nat^R
THR1:Sh ble/AOX1p/HSAss-TNFR-Fc]

pGLY5755



YGLY11566

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2 Δ::Nat^R
URA6:Sh ble/AOX1p/HSAss-TNFR-Fc]
HIS3:ARS^R/MmPOMGNT1-53]

pGLY5086



YGLY12680

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2 Δ::Nat^R
URA6:Sh ble/AOX1p/HSAss-TNFR-Fc
HIS3:ARS^R/MmPOMGNT1-53
THR1::HYG^R/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33]

FIG.2A

9/53

Construction of YGLY12680

Glossary:

VPS10

Sorting receptor for vacuolar hydrolases

ScARR3

Arsenite resistance marker

MmPomGNT1-53

Truncated Mouse PomGNT1 fused to ScMNN2-s leader

FIG.2B

10/53

Construction of YGLY14252**YGLY13571**

[ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ/TrMDS1 dap2 Δ::Nat^R
THR1:Sh ble/AOX1p/HSAss-TNFR-Fc
TRP5::Hyg^R/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33]
 pGLY5219 ↓

YGLY14252

[ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ/TrMDS1 dap2 Δ::Nat^R
THR1:Sh ble/AOX1p/HSAss-TNFR-Fc
TRP5::Hyg^R/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
vps10::lacZ-URA5-lacZ/MmPomGNT1-65]

MmPomGNT1-65**Truncated Mouse PomGNT1 fused to ScMNN6 leader****FIG.3**

11/53

Construction of YGLY14954 and YGLY14927**YGLY13571**

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ/TrMDS1 dap2 Δ::Nat^R
THR1:Sh ble/AOX1p/HSAss-TNFR-Fc
TRP5::Hyg^RMmCST/HsGNE/HsCSS/HsSPS/MmST6-33]

pGLY5192

**YGLY13663**

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ/TrMDS1 dap2 Δ::Nat^R
THR1:Sh ble/AOX1p/HSAss-TNFR-Fc
TRP5::Hyg^RMmCST/HsGNE/HsCSS/HsSPS/MmST6-33
vps10::lacZ-URA5-lacZ]

pGLY7087

**YGLY14954 & YGLY14927**

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1 Δ::lacZ/ScGAL10/XB33/DmUGT
arg1 Δ::HIS1/KD53/TC54
bmt4 Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ/TrMDS1 dap2 Δ::Nat^R
THR1:Sh ble/AOX1p/HSAss-TNFR-Fc
TRP5::Hyg^RMmCST/HsGNE/HsCSS/HsSPS/MmST6-33
vps10::lacZ-URA5-lacZ
HIS3::ScARR3/GAPp/MmPomGNT1-56]

MmPomGNT1-65

Truncated Mouse PomGNT1 fused to ScMNN6 leader

FIG.4

12/53

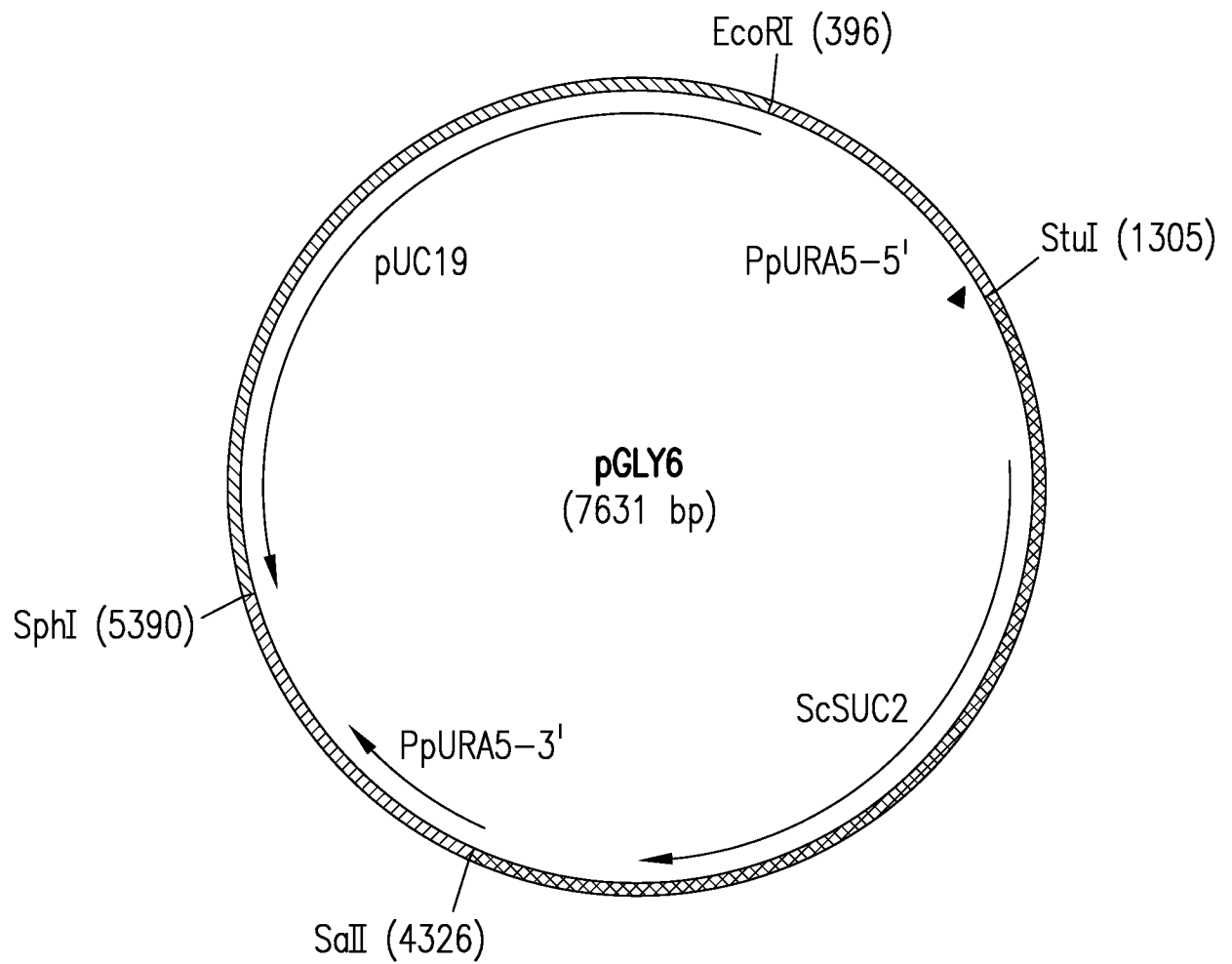


FIG.5

13/53

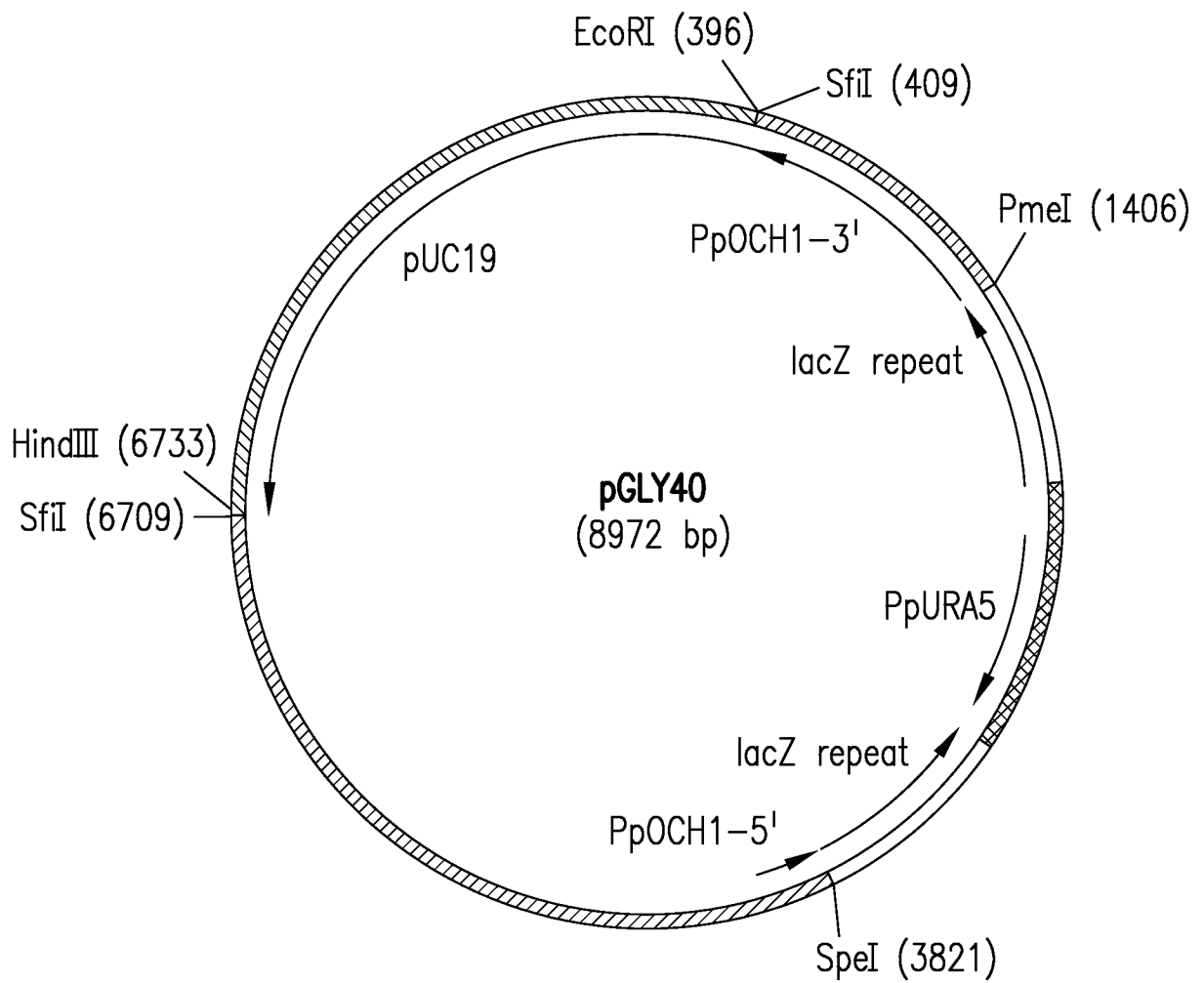


FIG. 6

14/53

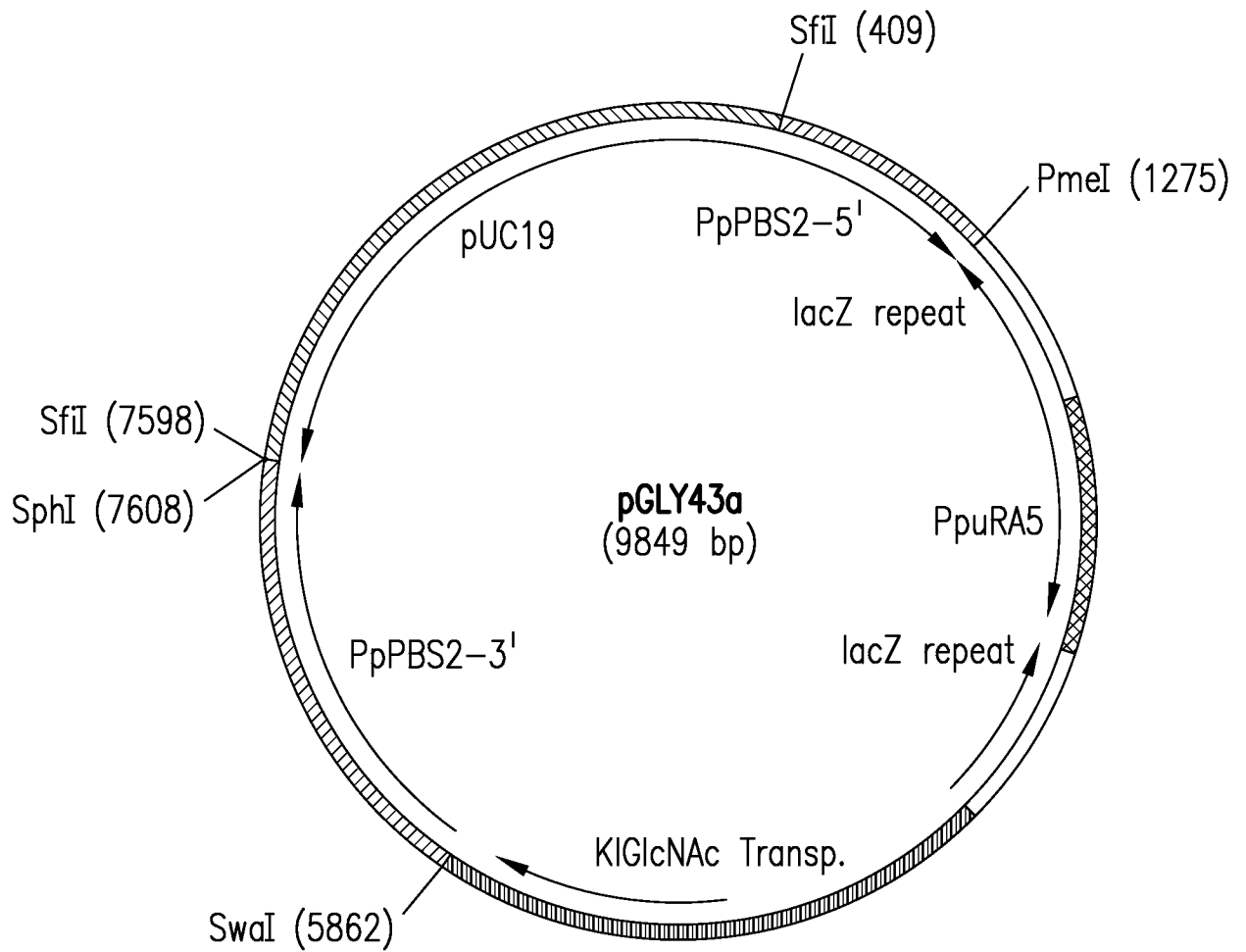


FIG. 7

15/53

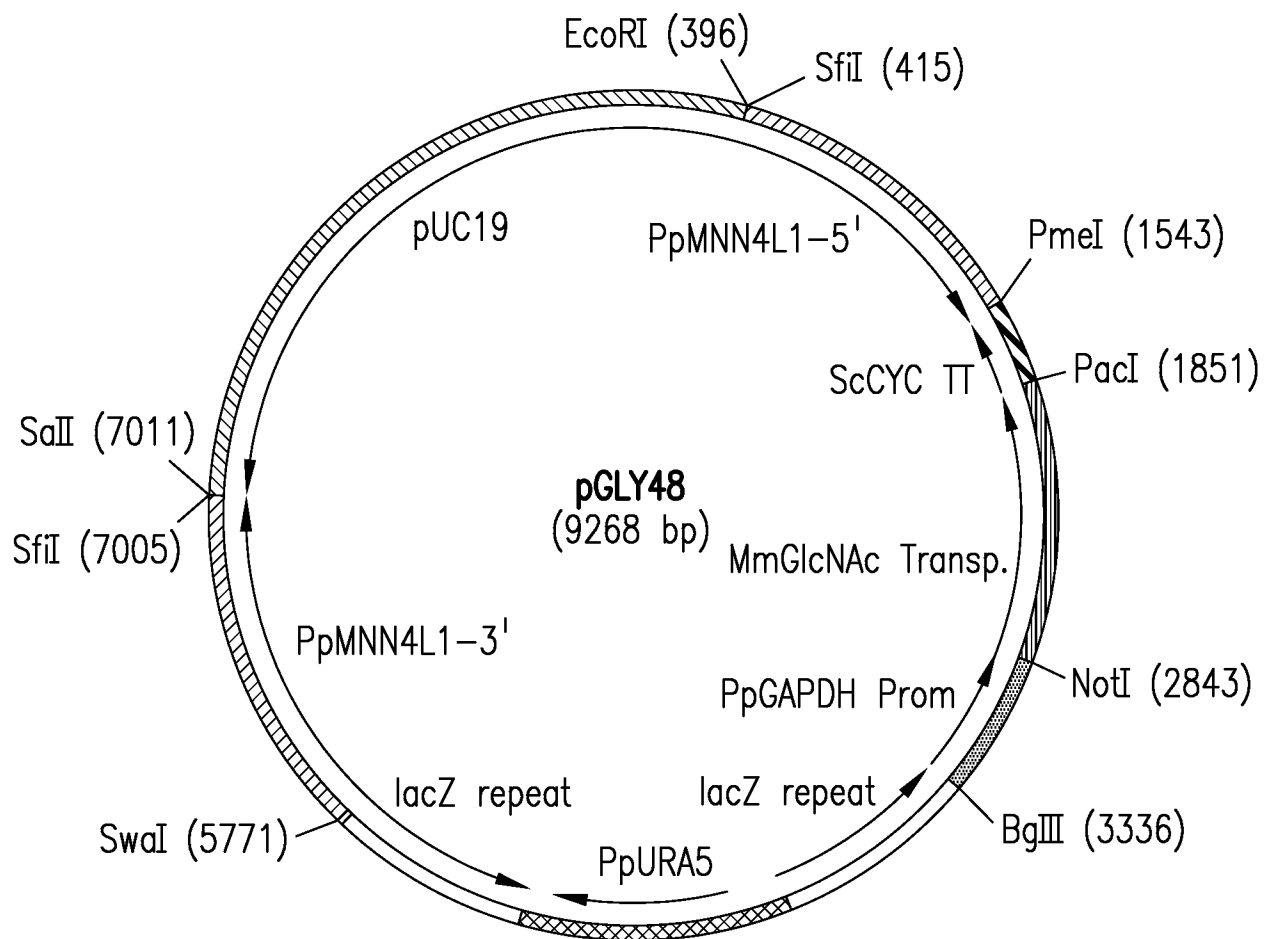


FIG.8

16/53

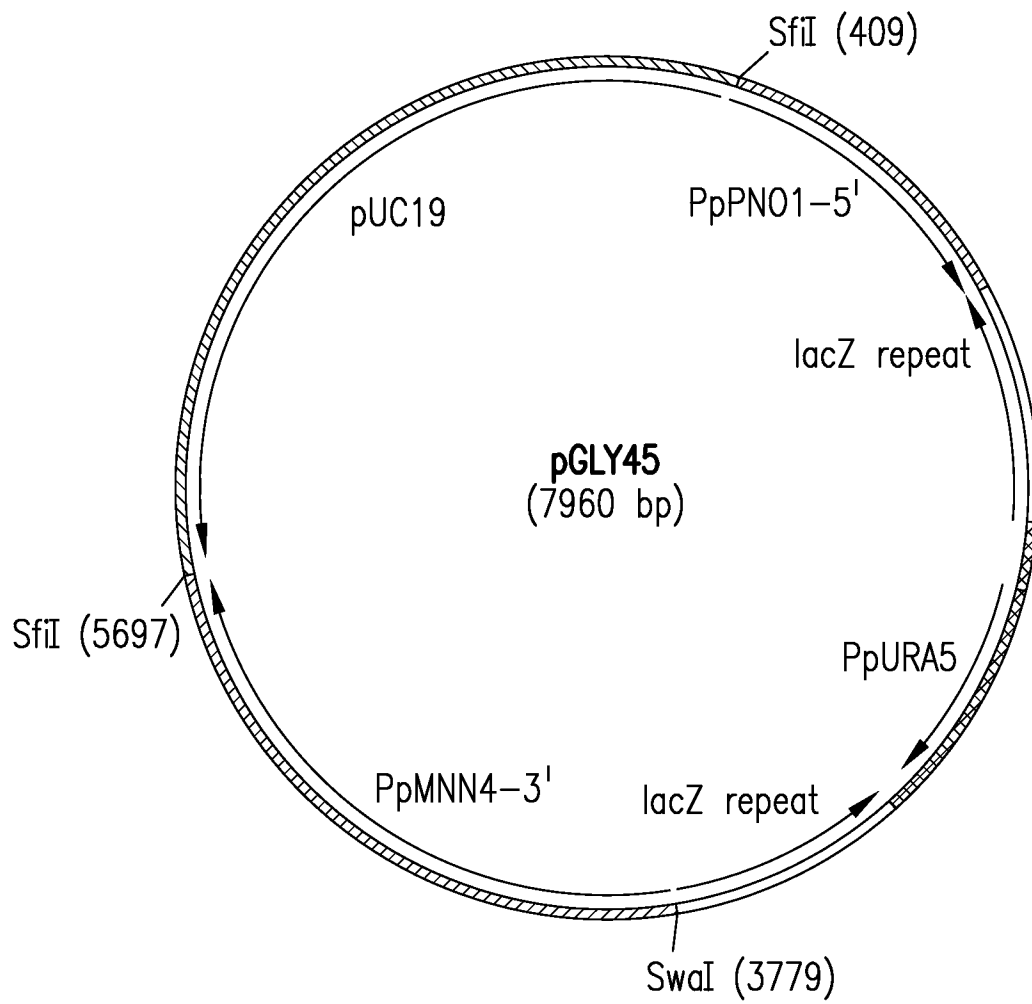


FIG.9

17/53

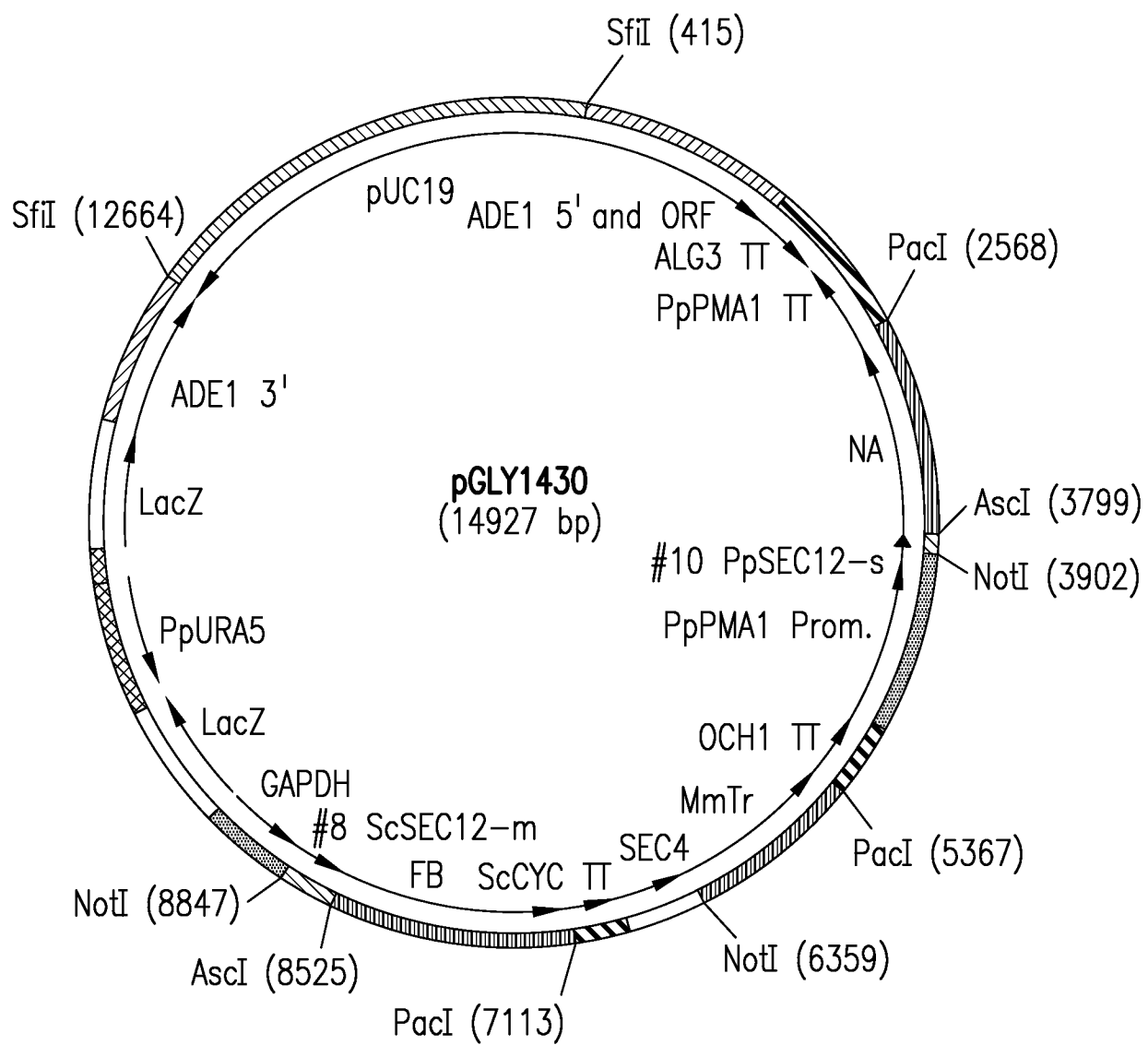


FIG.10

18/53

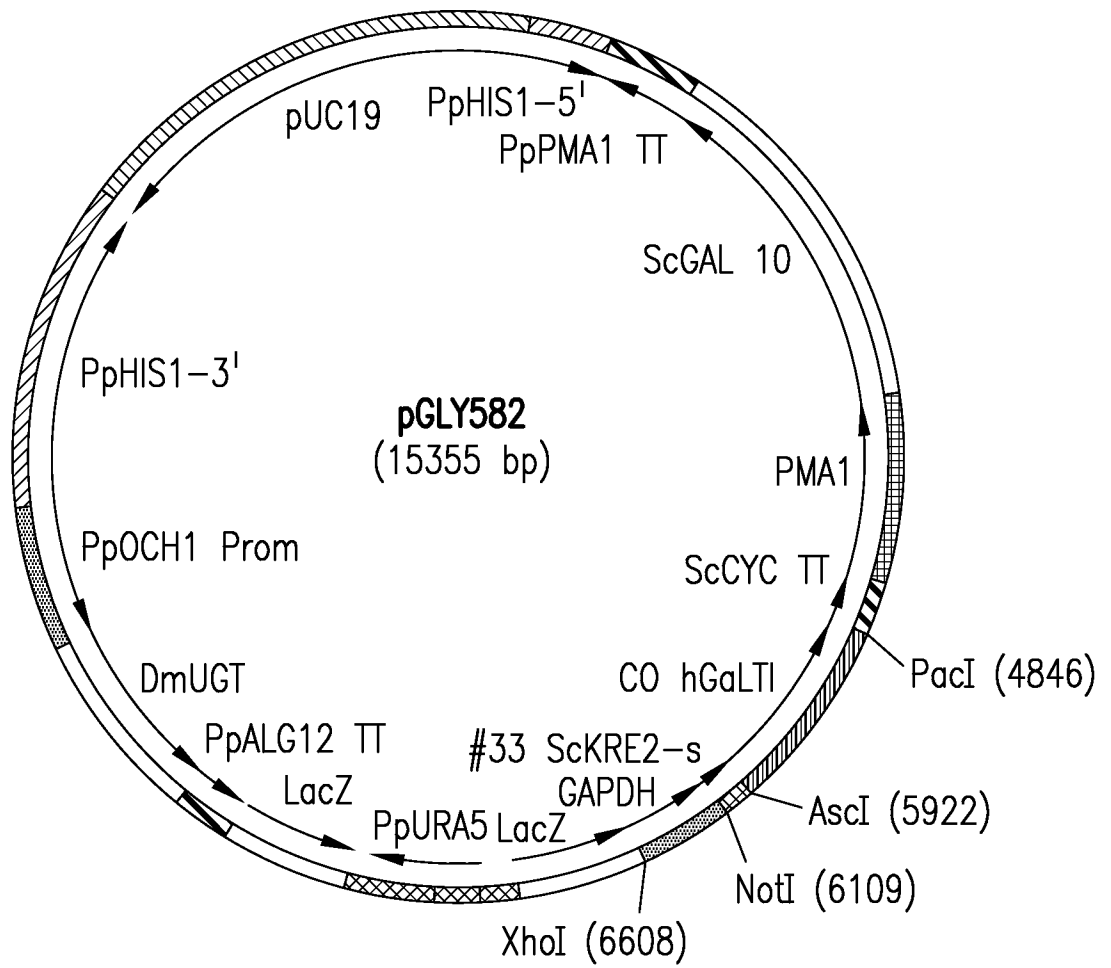


FIG.11

19/53

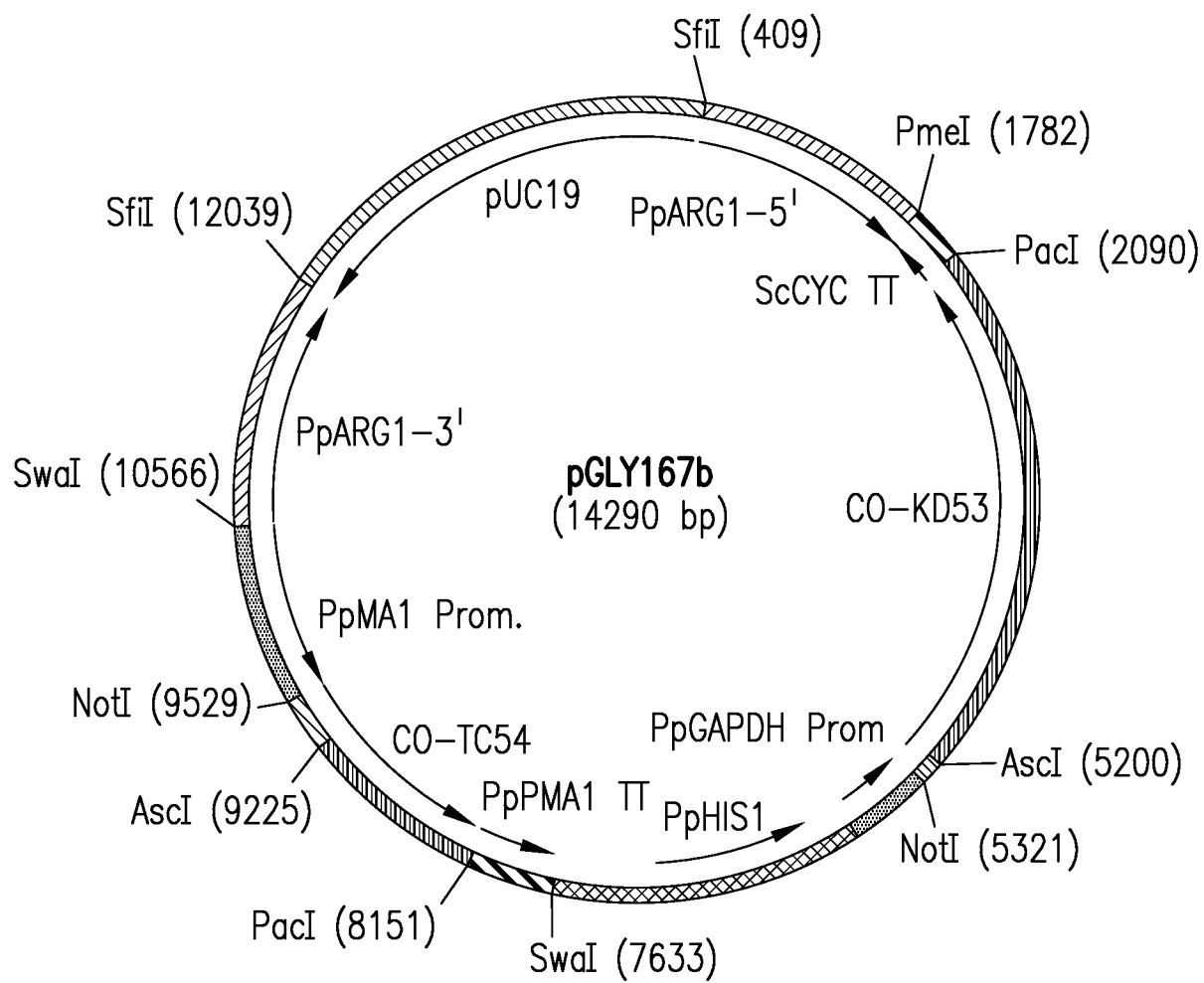


FIG. 12

20/53

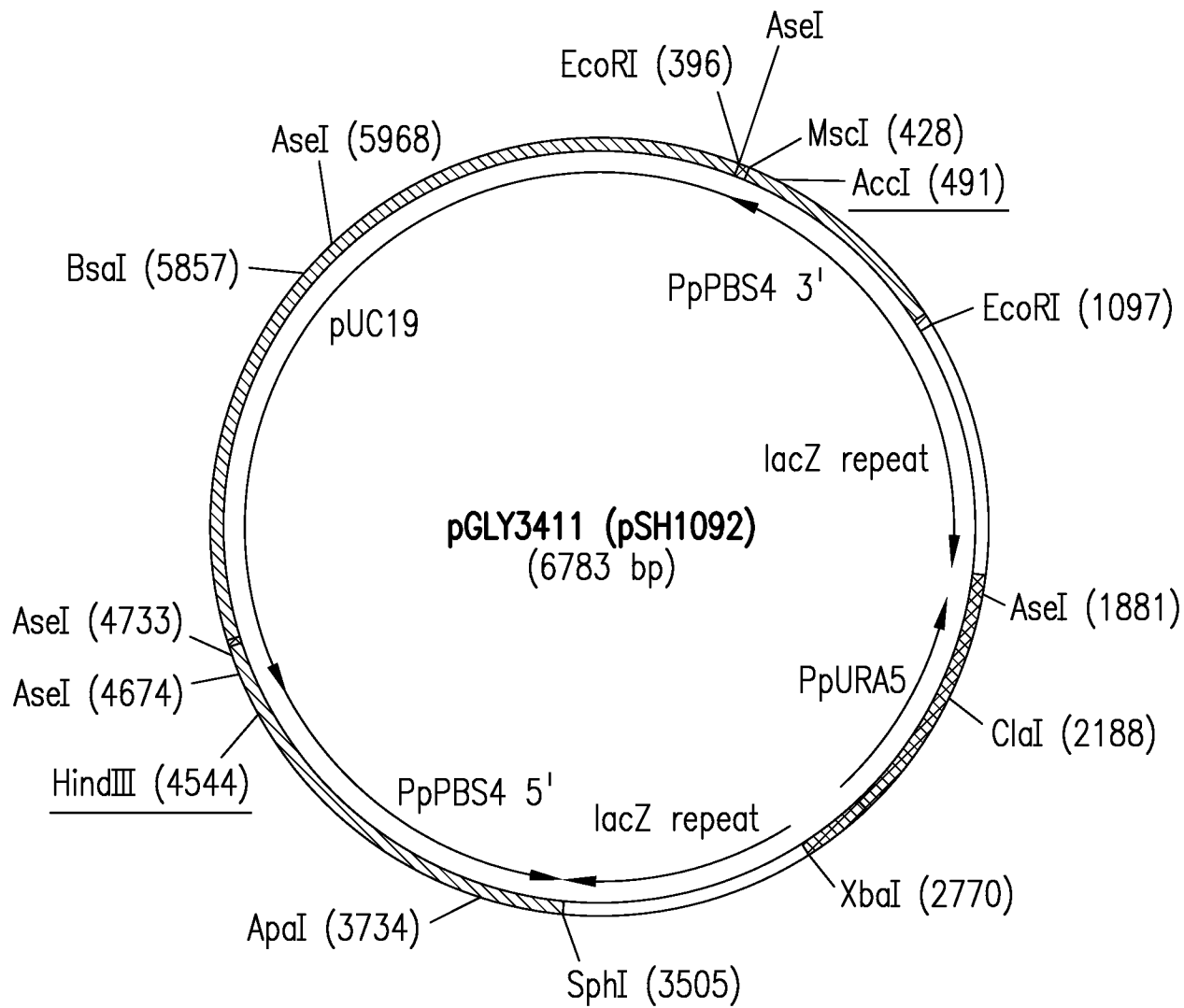


FIG.13

21/43

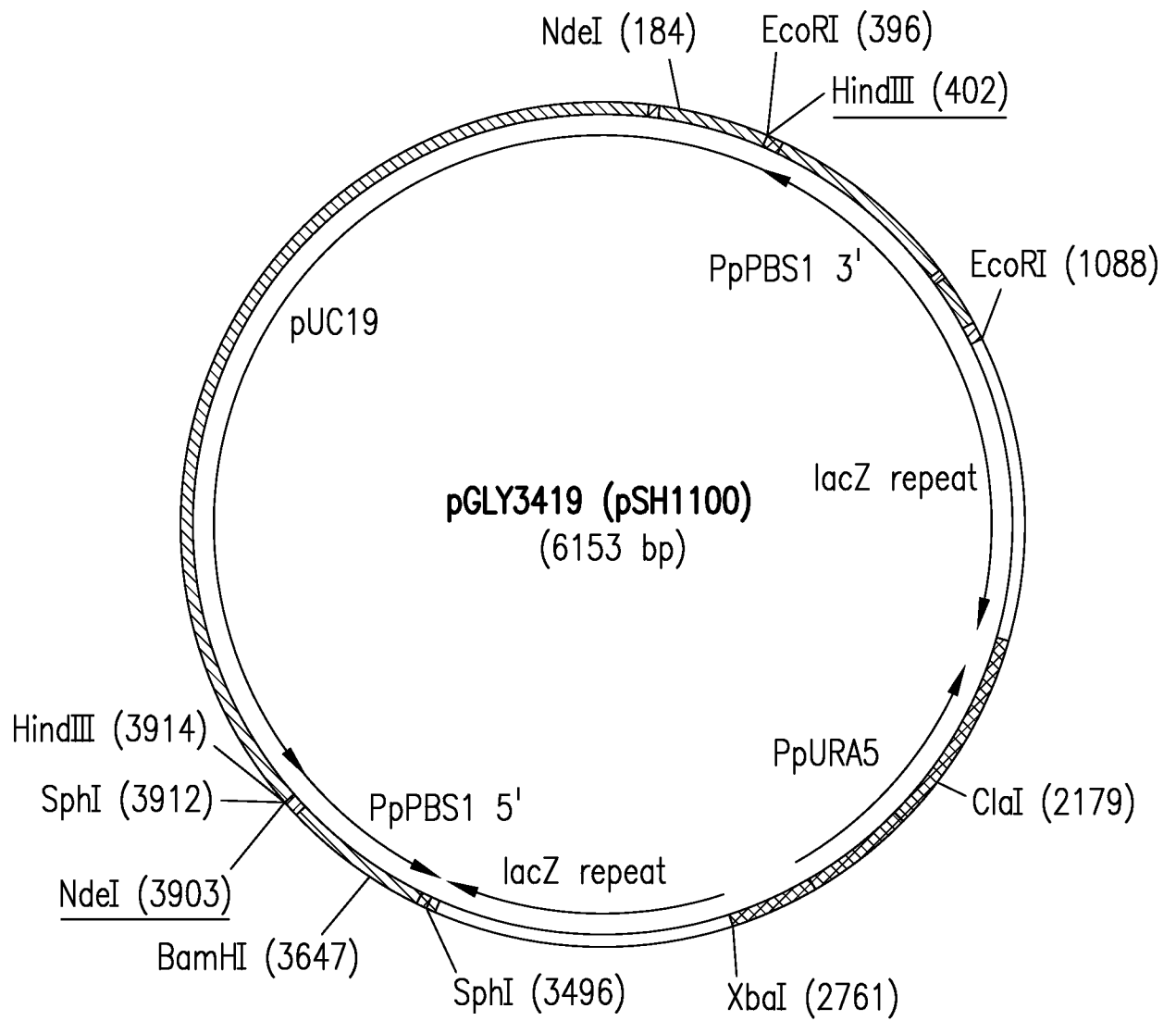


FIG. 14

22/53

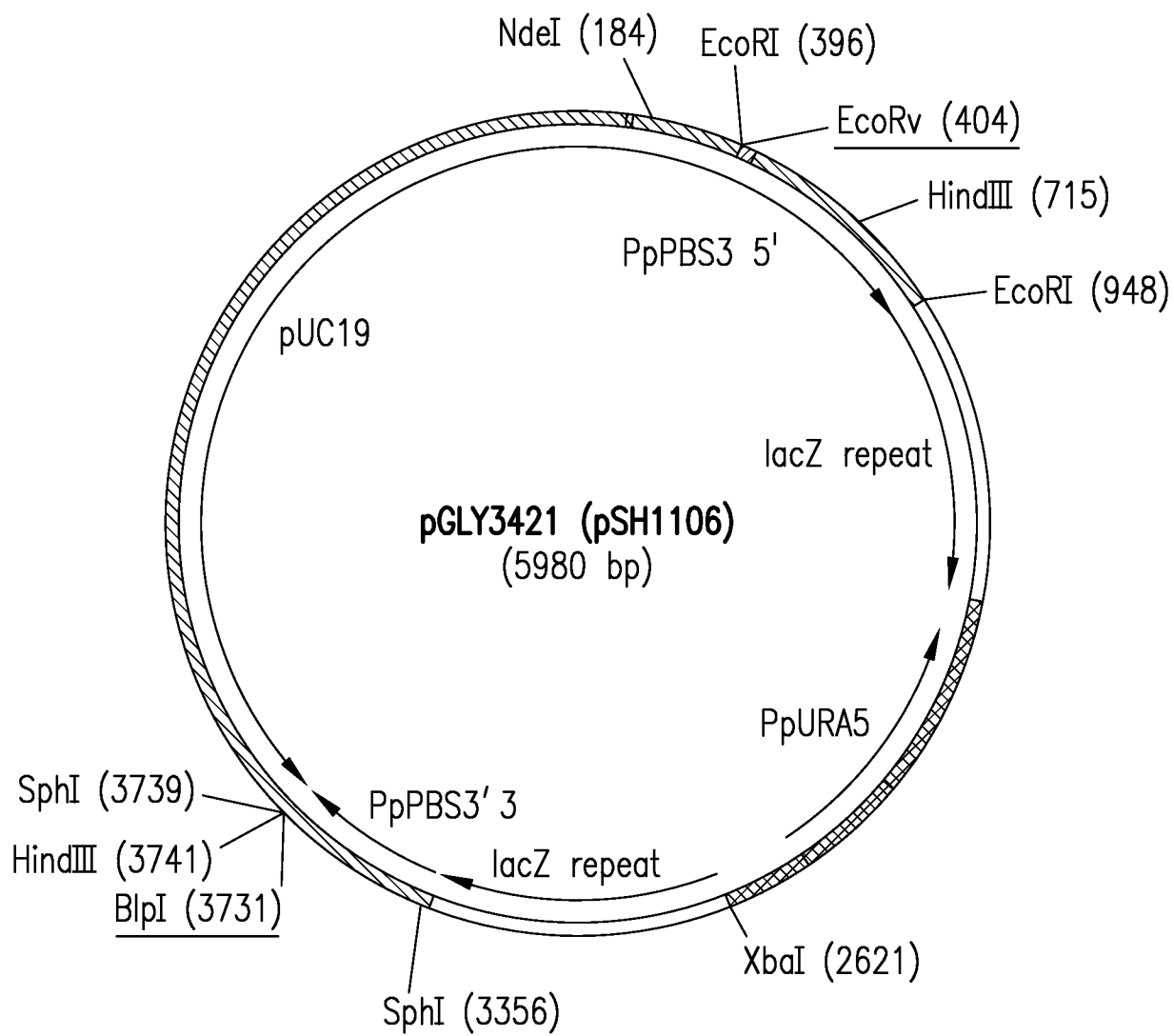


FIG. 15

23/53

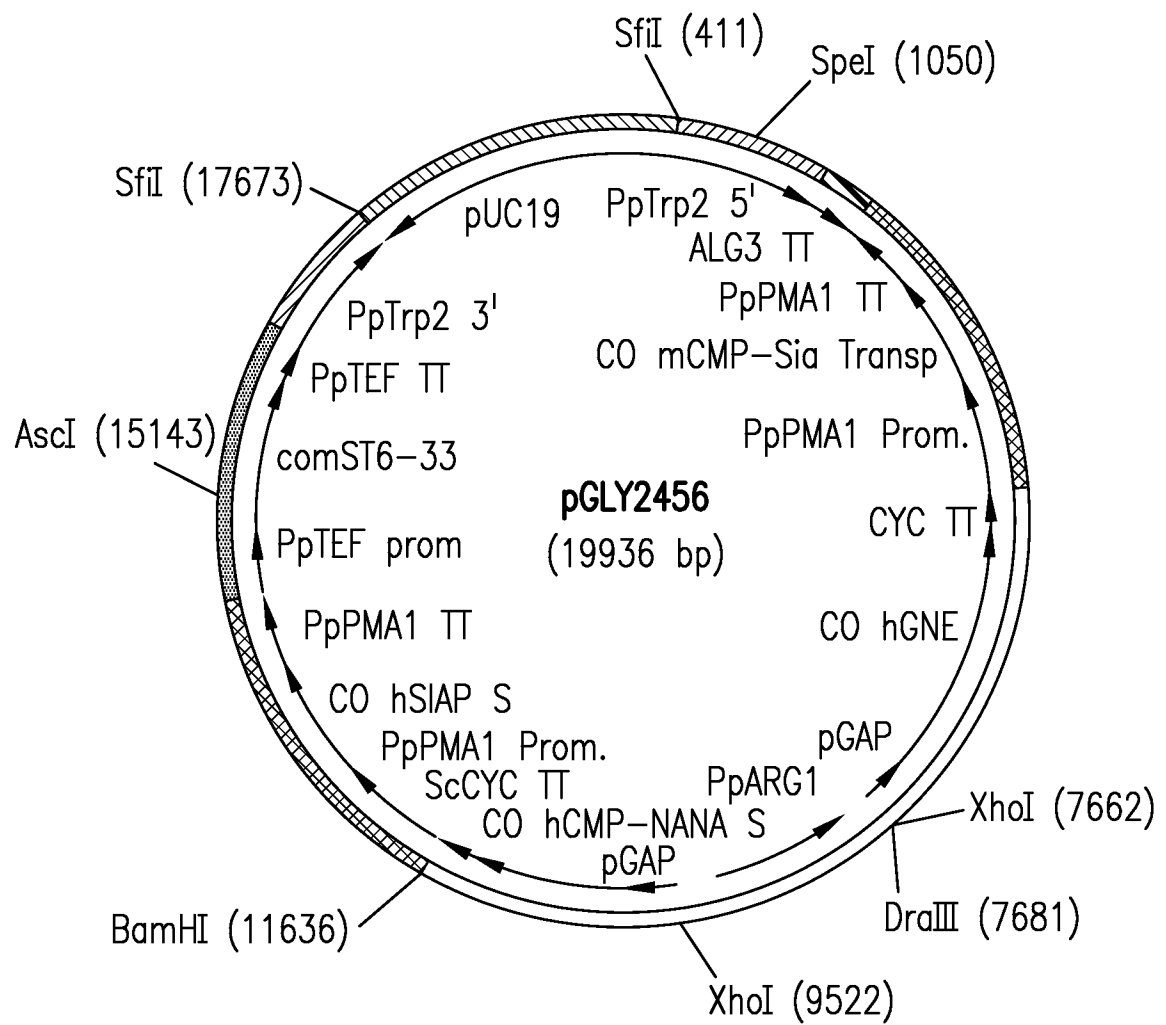


FIG.16

24/53

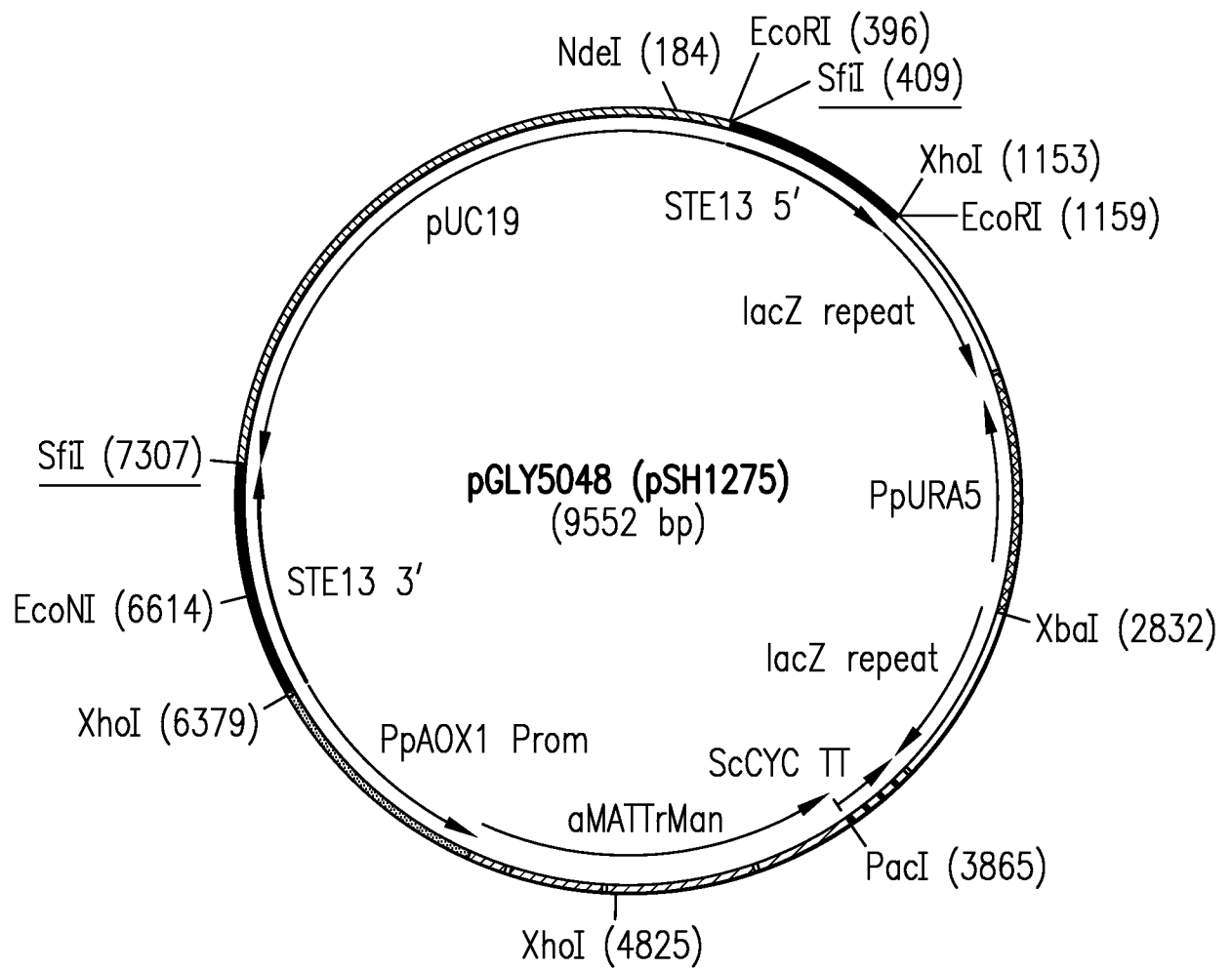


FIG. 17

25/53

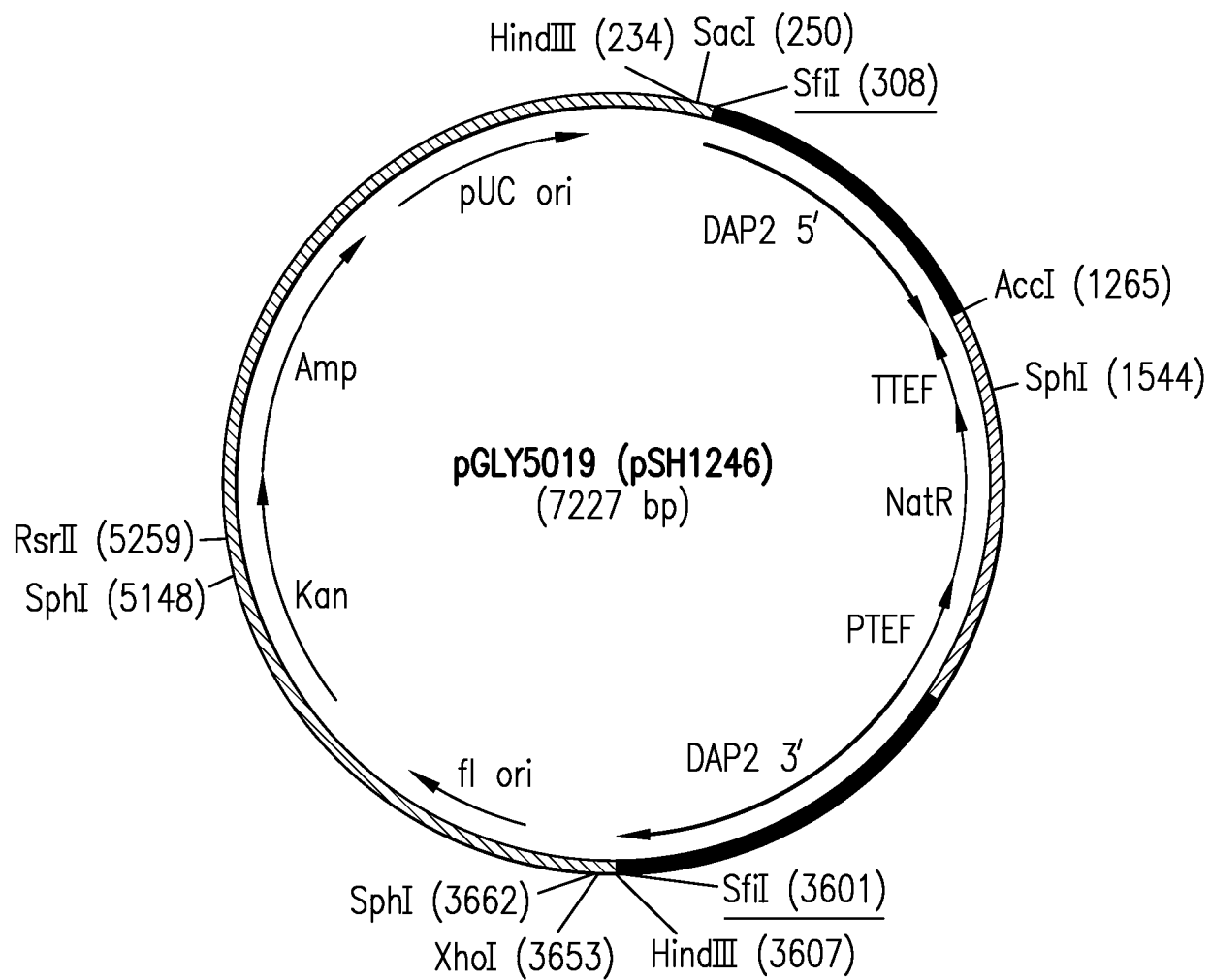


FIG.18

26/53

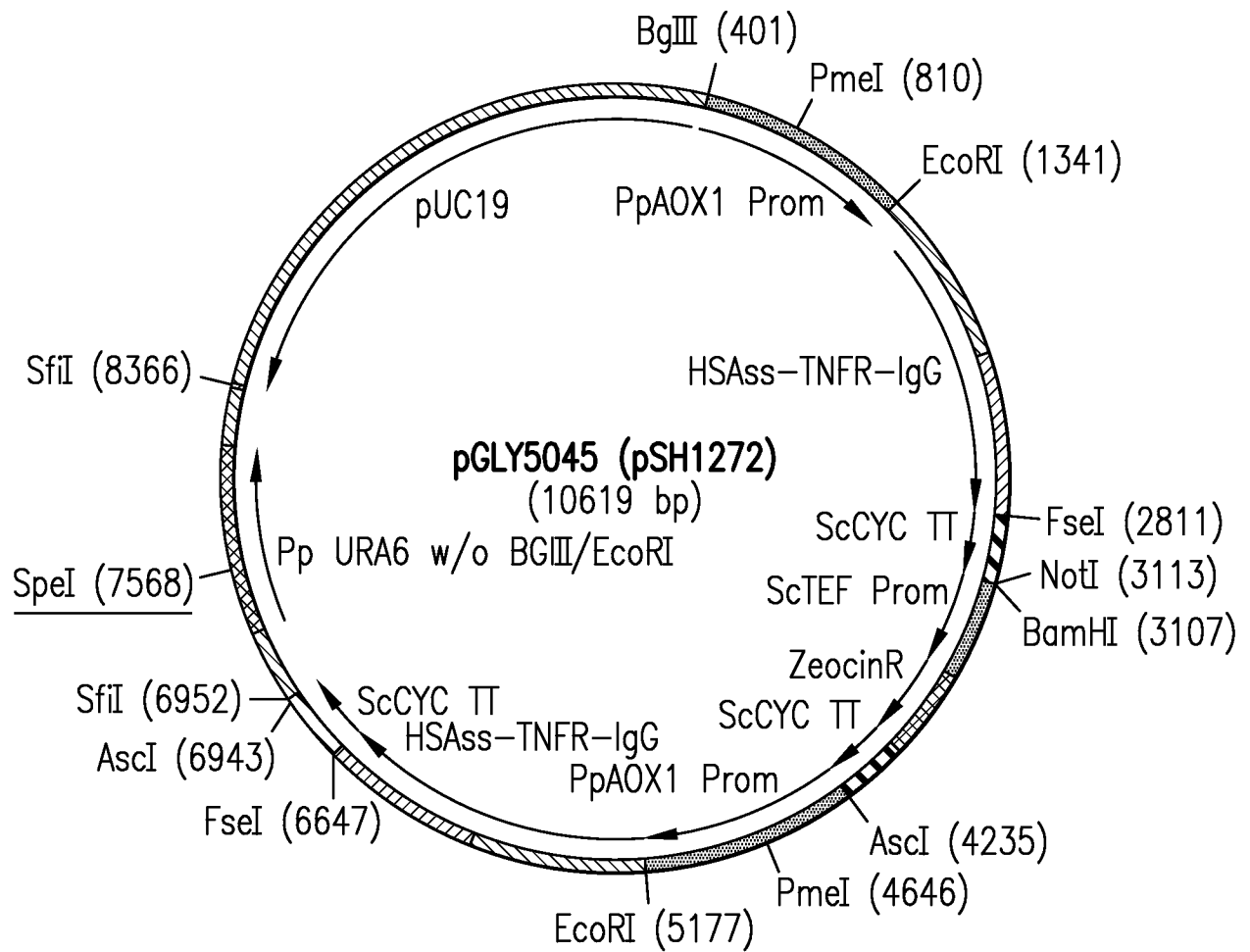


FIG.19

27/53

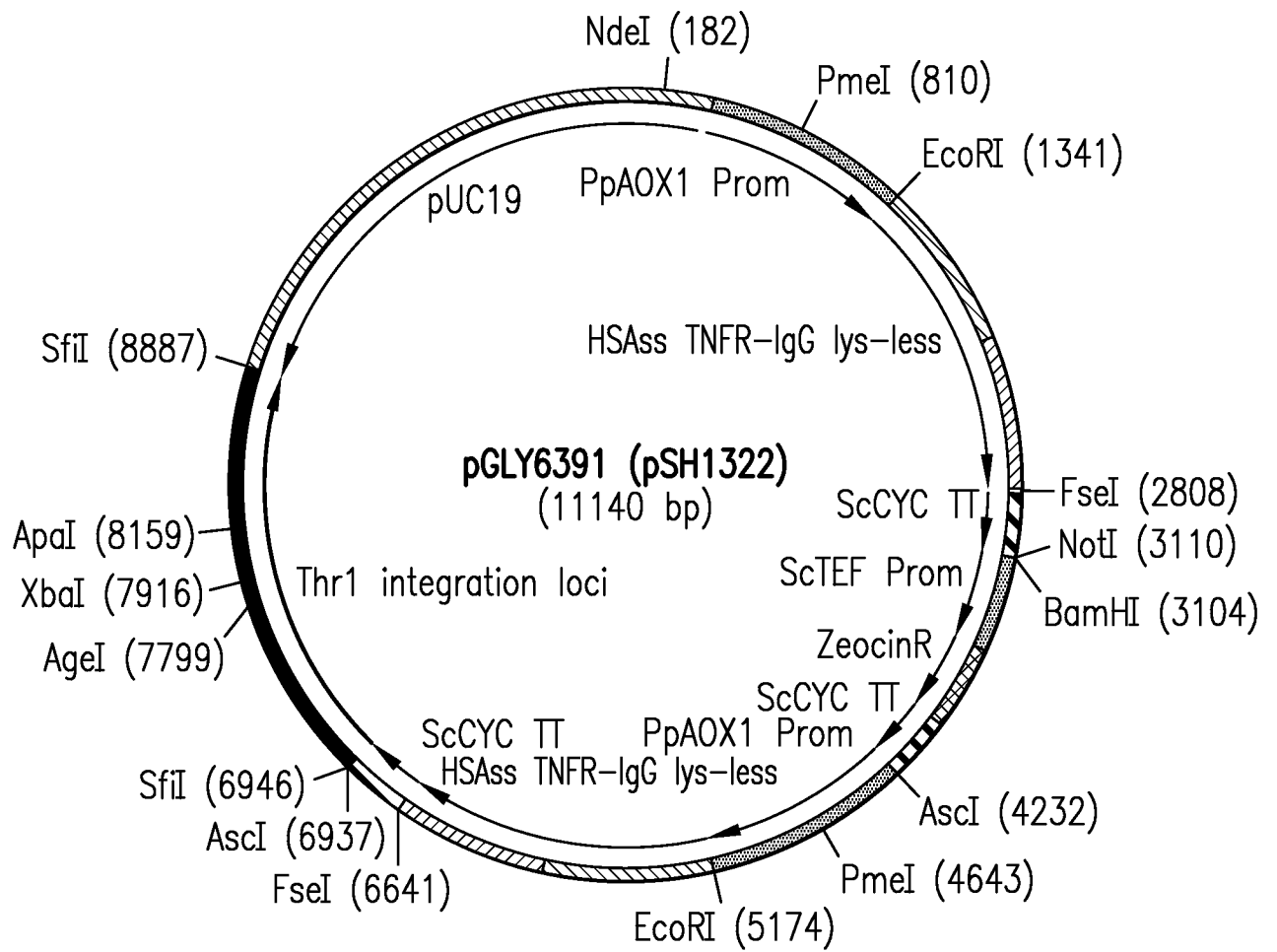


FIG.20

28/53

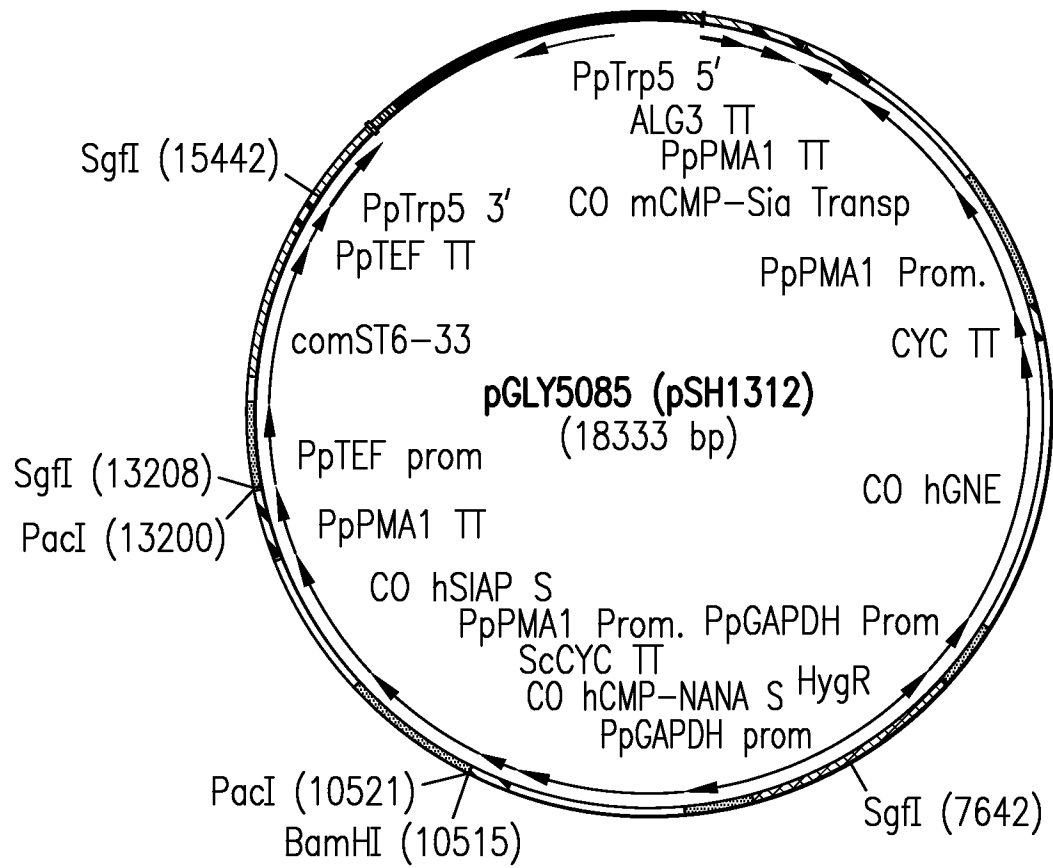


FIG.21

29/53

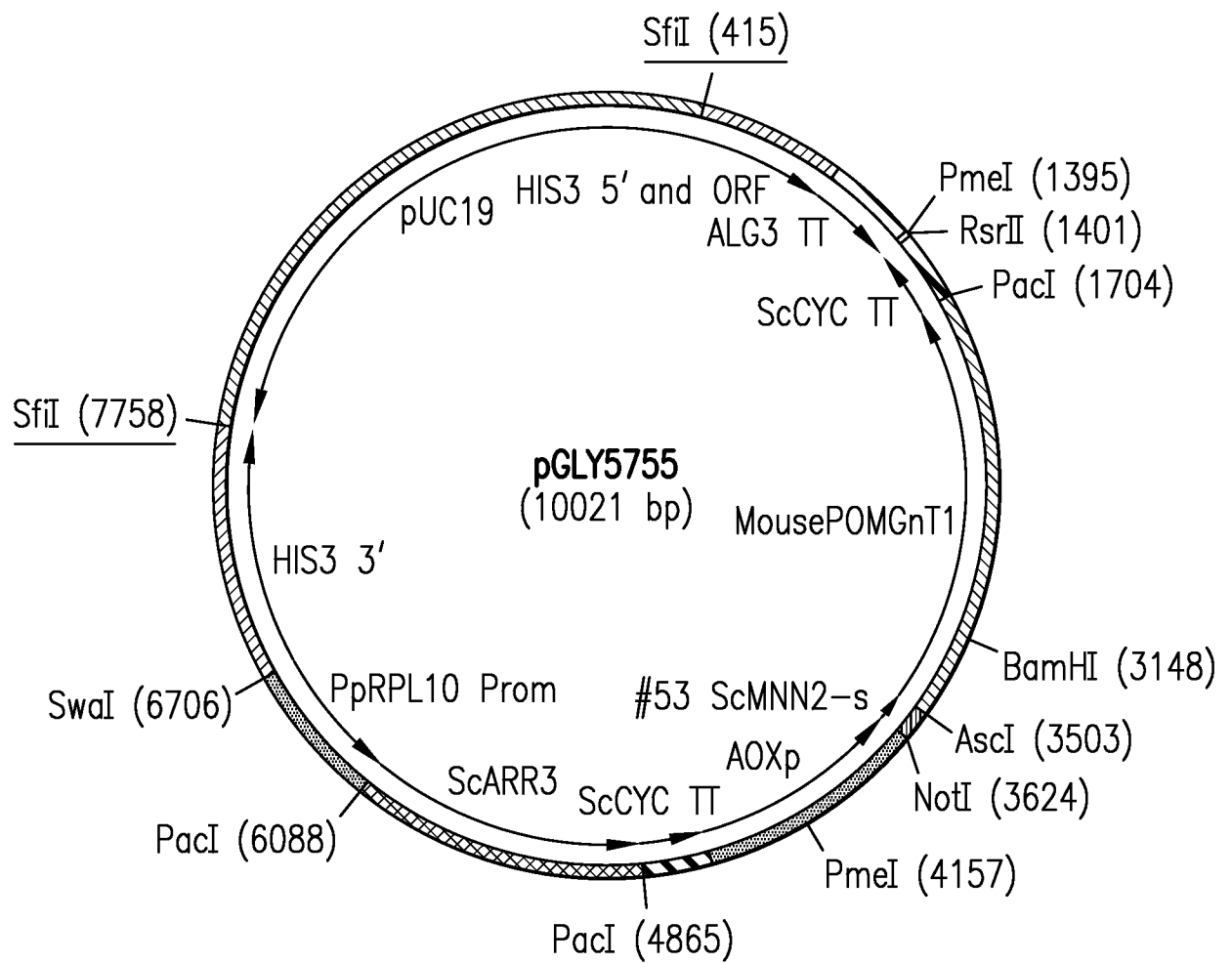


FIG.22

30/53

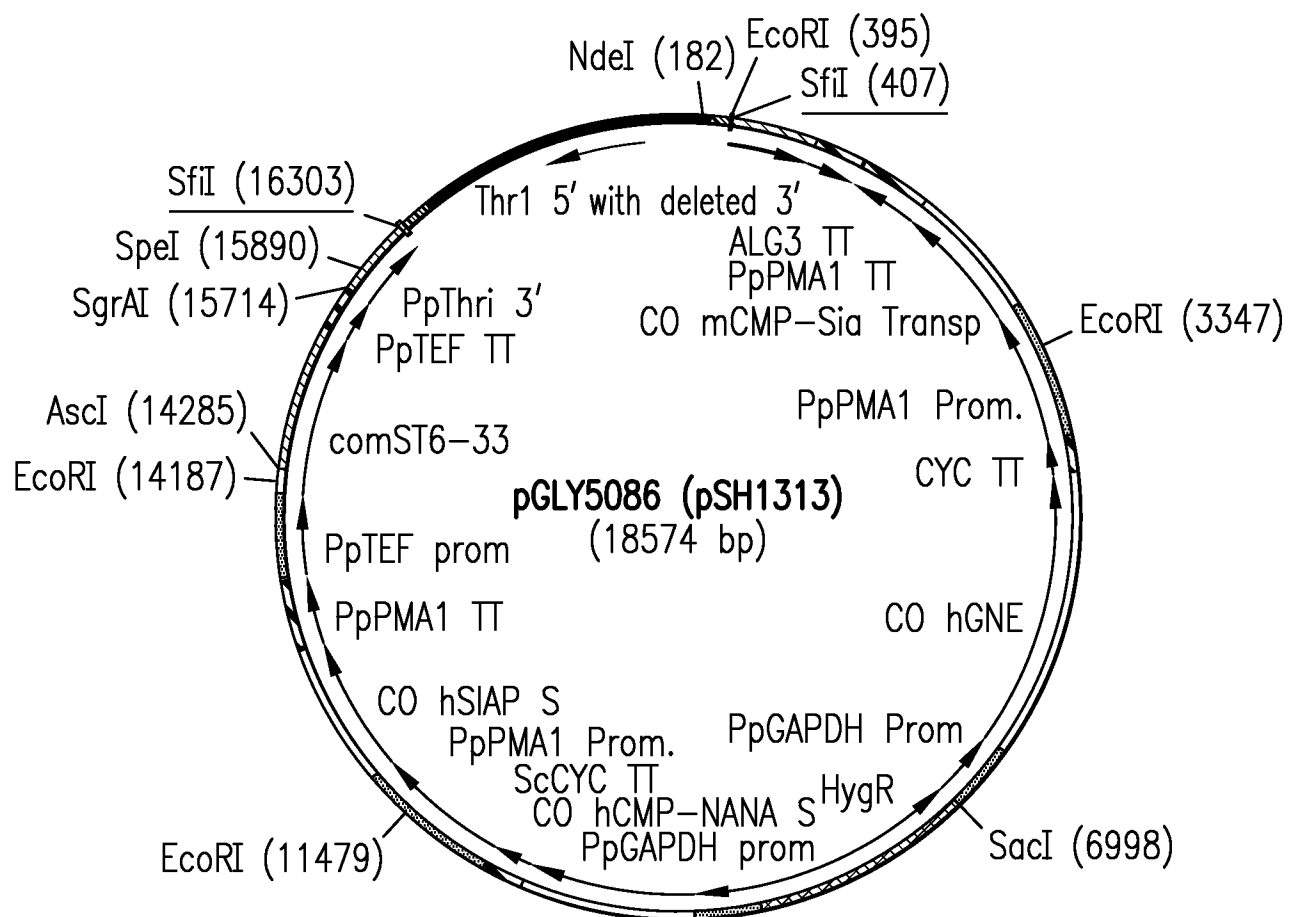


FIG.23

31/53

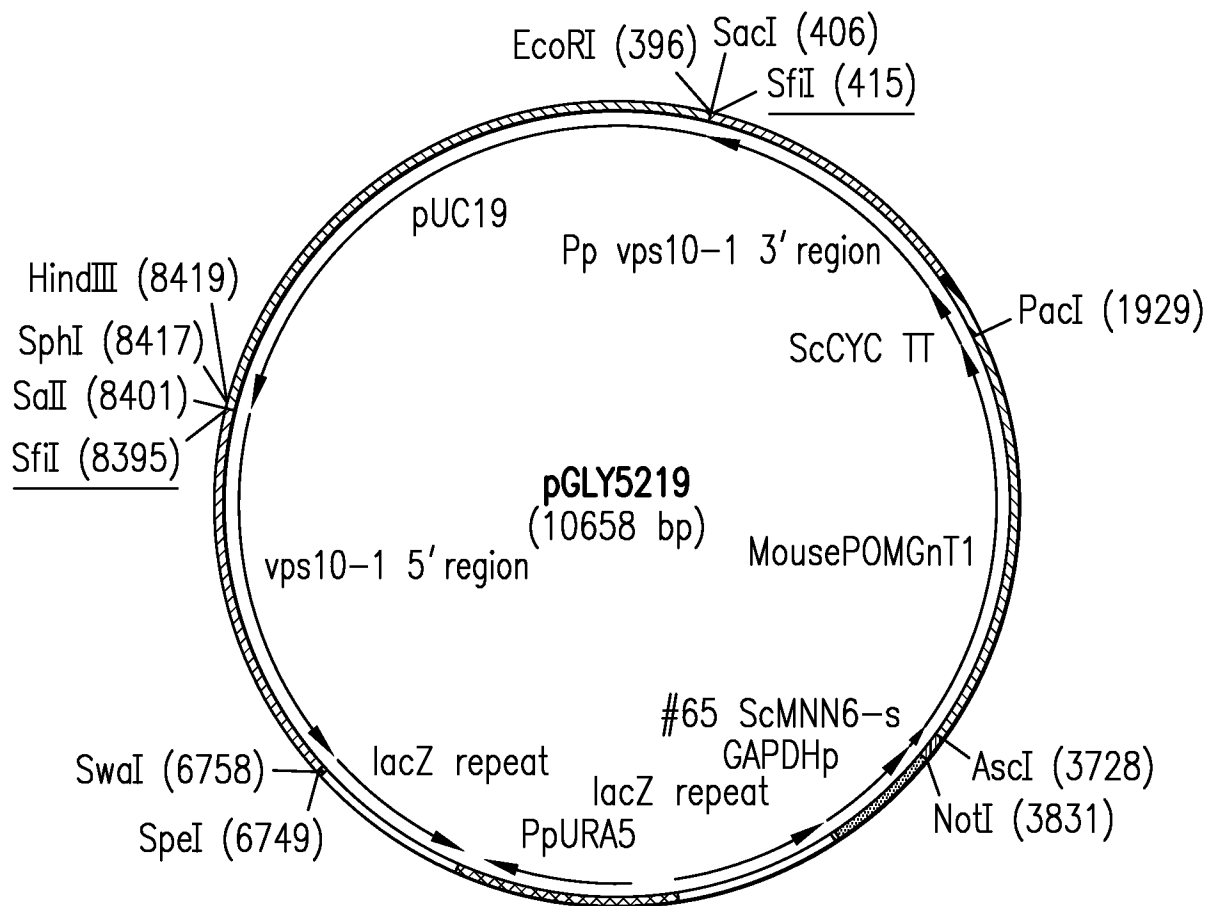


FIG.24

32/53

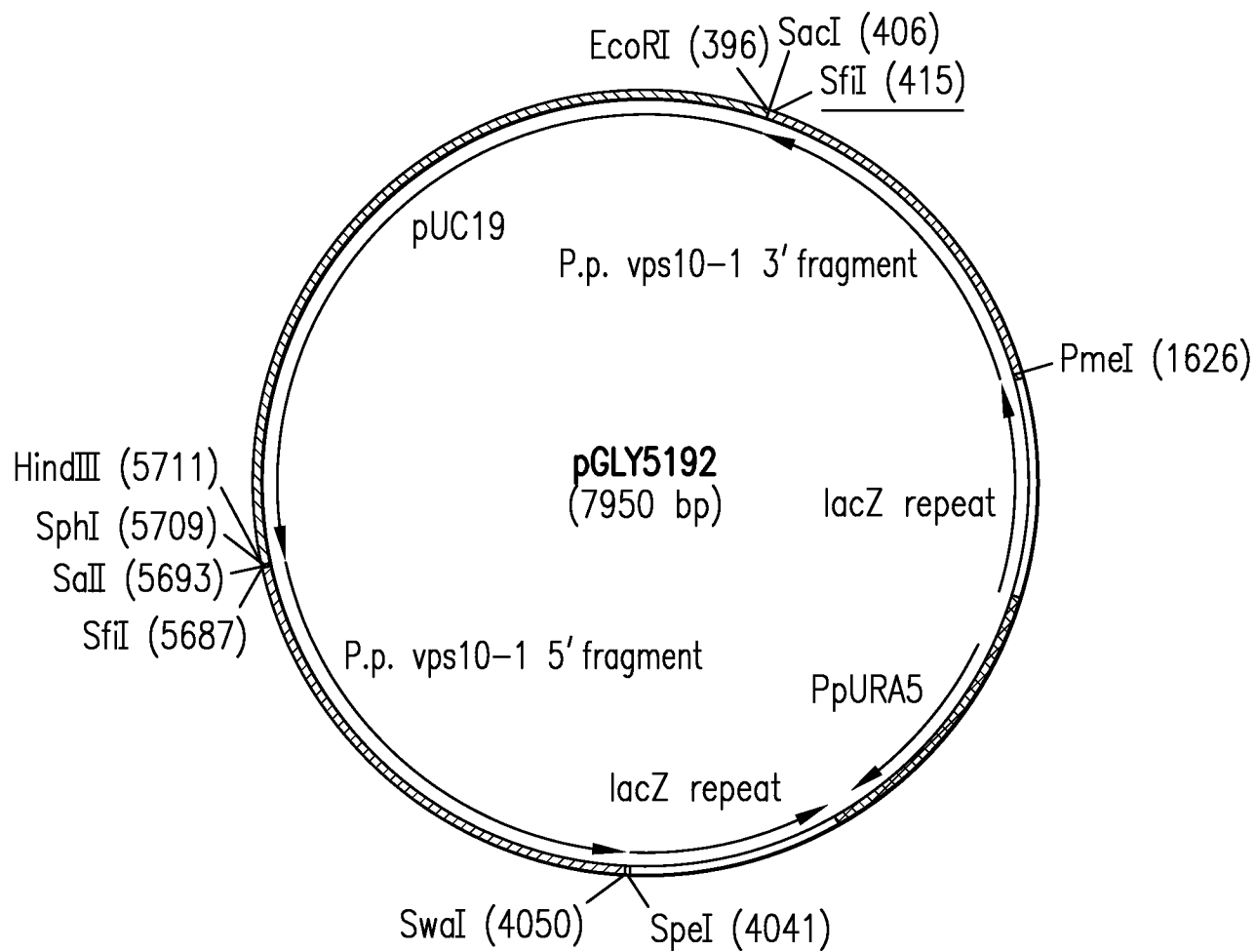


FIG.25

33/53

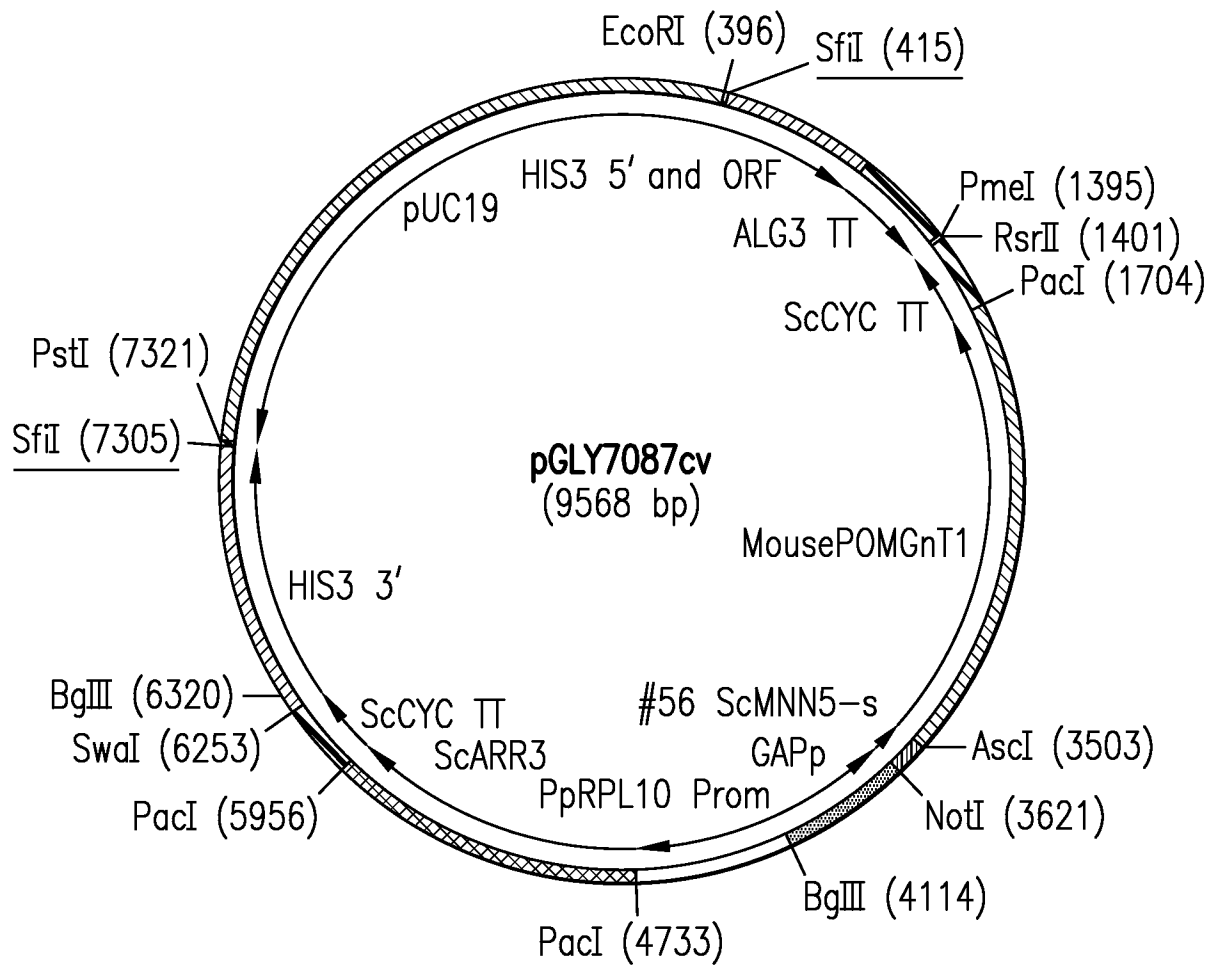


FIG.26

34/53

TNFR–Fc is a highly complex molecule

```

1  LPAQVAFTPYAPEGSTCRLREYYDQTAQMCCSKCSPGQHAKVFCTKTSD
51  TVQDSCEDSTYTQLWNWVPECLSCGSRCSDDQVETQACTREQNRIQTCRP
101 GWYCALSKQEGCRLCAPLRKCRPGFGVARPGTETSDVVCKPCAPGTFSNT
151 TSSTDICRPHQICNVVAIPGNASMDAVCTSTSPTRSMAPGAVHLPQPVSTR
202 SQHTQPTPEPSTAPSTFLLPMGPSPPAEGSTGDEPKSCDKTHTCPPCPAP
253 ELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVE
304 VHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEK
355 TISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNG
406 QPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNH
456 YTQKSLSLSPG(K)

```

• Key:

- TNFRII ectodomain (in bold)
- IgG1 Fc domain (regular text)
- Cysteine-rich subdomains of TNFRII domain (outlined by arrows)
- N-linked glycosylation sites ("N" residues encircled)
- Optional C-terminal lysine (in brackets)

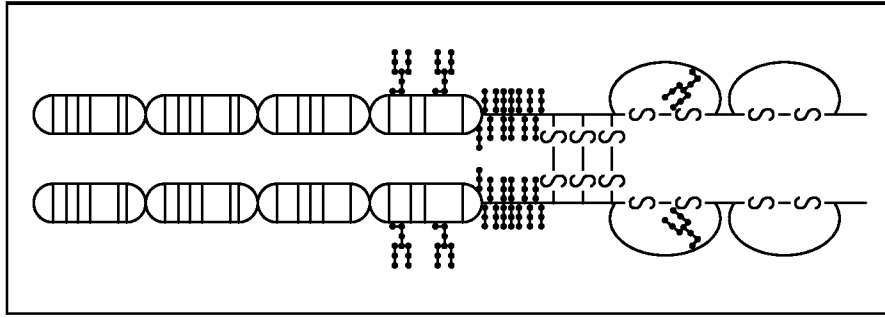


FIG. 27

35/53

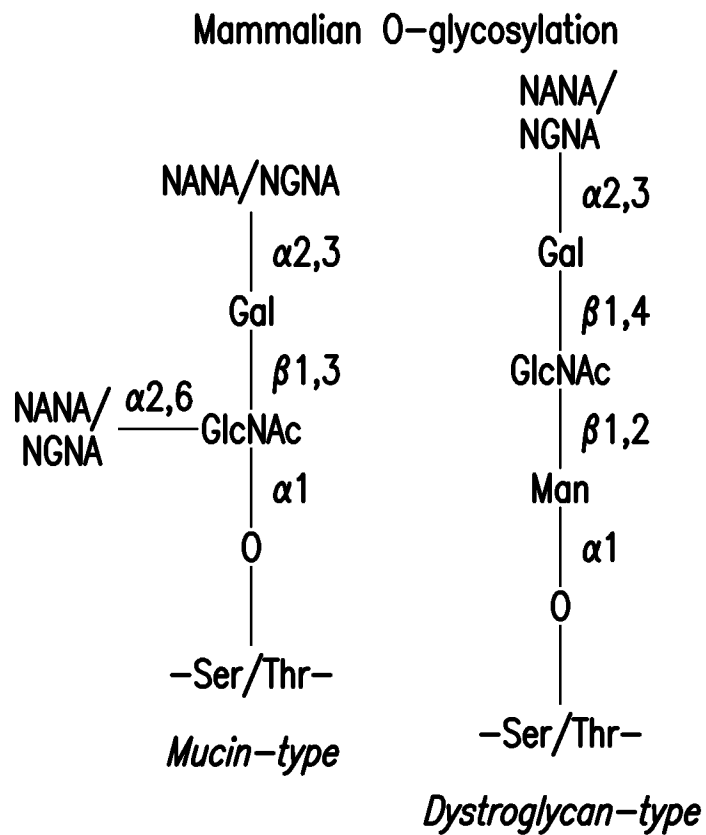


FIG.28

O-glycosylation engineering strategy for TNFRII-Fc

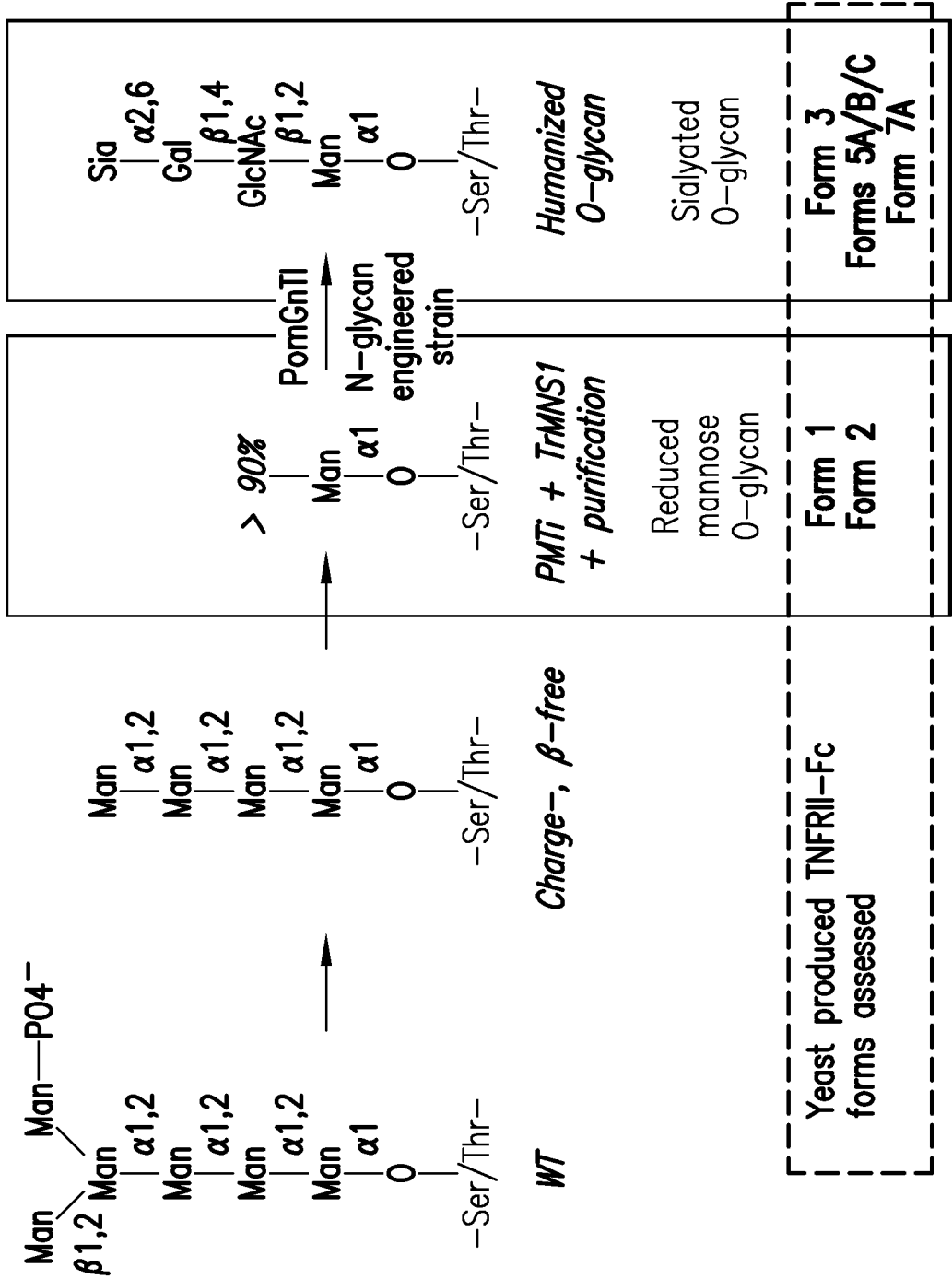


FIG.29

37/53

Example of Purification strategy

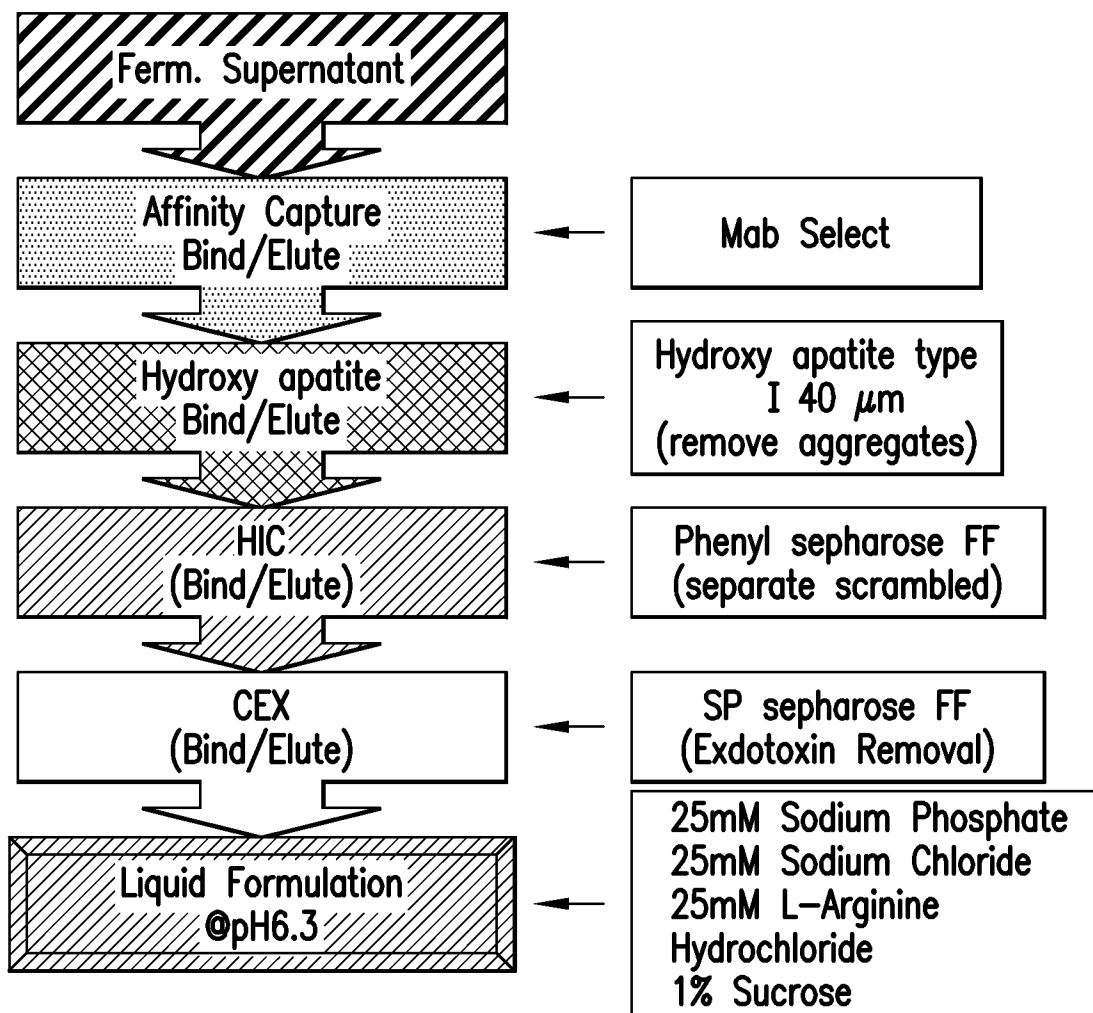


FIG.30

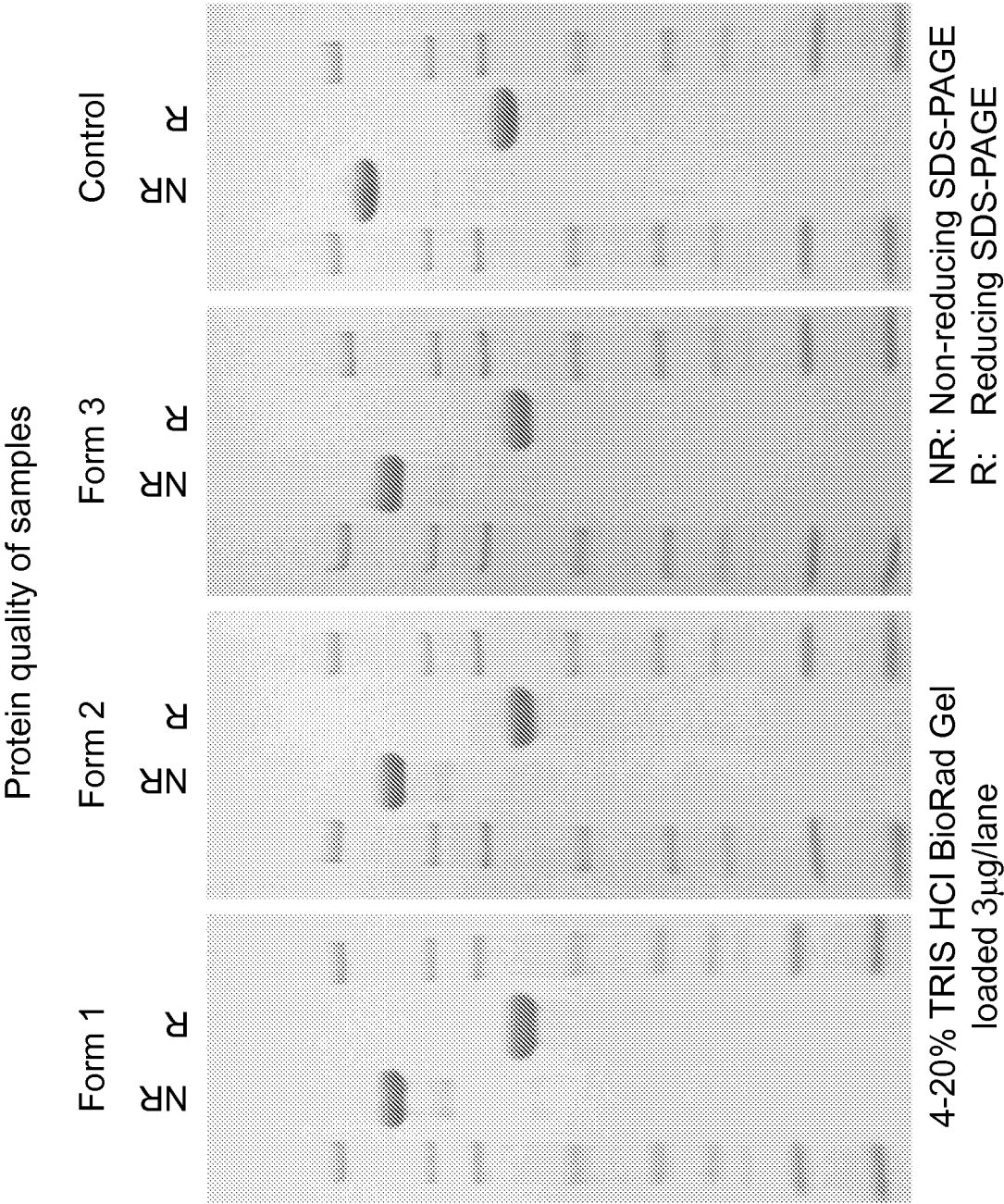


FIG.31

Comparison of Glycan Composition

Sample	Source	N-glycan Profile (% of total N-glycans)			Terminal O-glycan Profile (% of total O-glycans)			Total sialic acid (mol Sialic acid per mol protein)
		Neutral	Mono-sialylated	Bi-sialylated	Mannose	Neutral (exc. mannose)	Sialylated	
Form 1	YGLY10299	43.3	44.5	12.2	100	0	0	3
Form 2	YGLY11731	25.7	42.4	31.9	100	0	0	5
Form 3	YGLY12680	44.6	44.0	11.4	18	53	29	10
Control	Commercial ENBREL (liquid formulation, Lot A)	26.0	52.4	21.6	Not determined			27

FIG.32

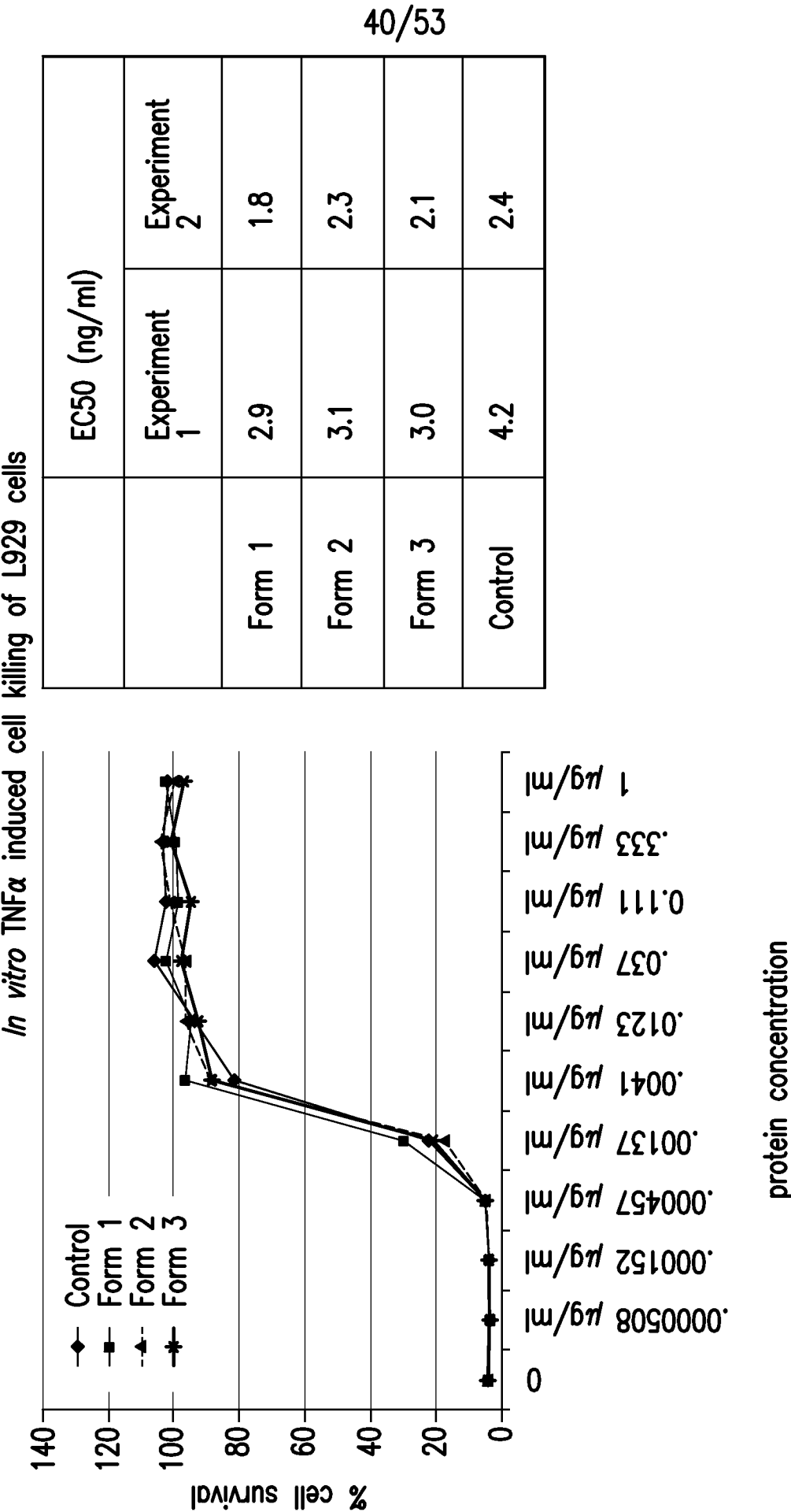
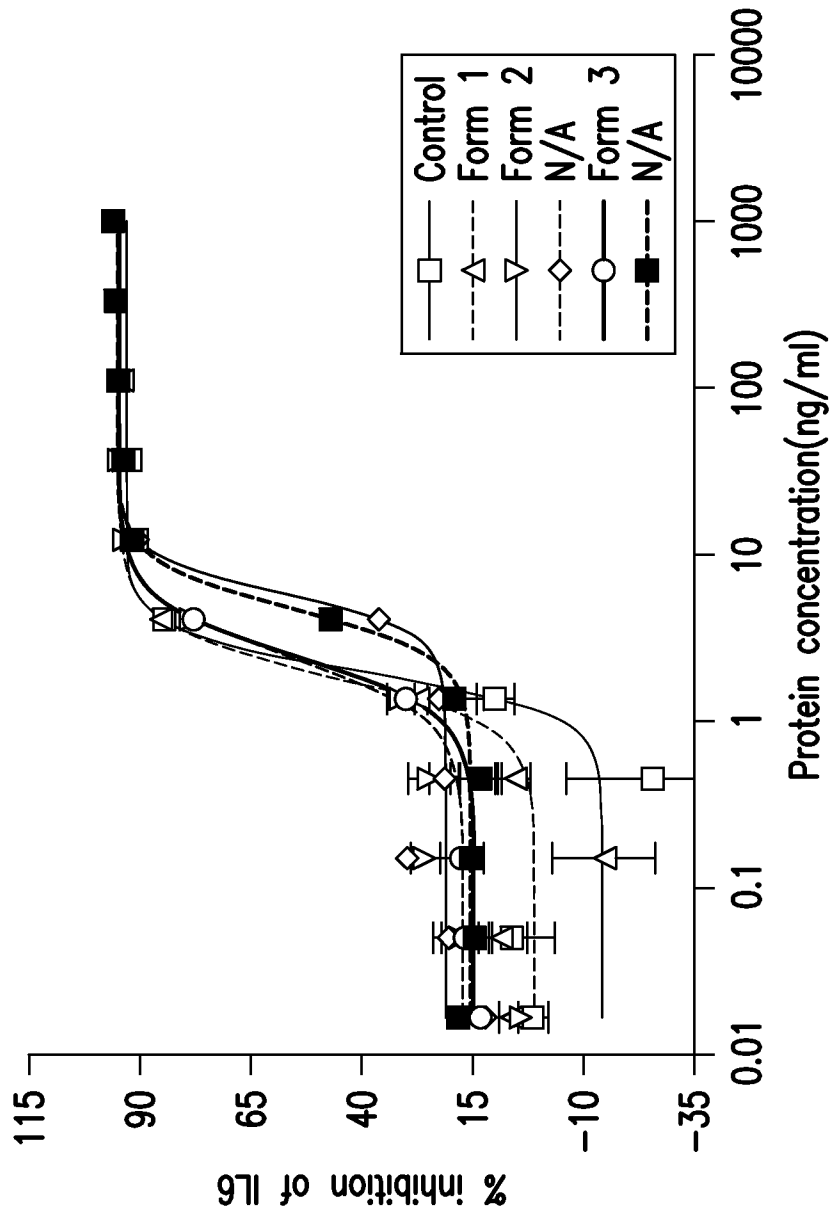


FIG.33

In vitro TNF α -stimulated release of IL-6 in
A549 cells

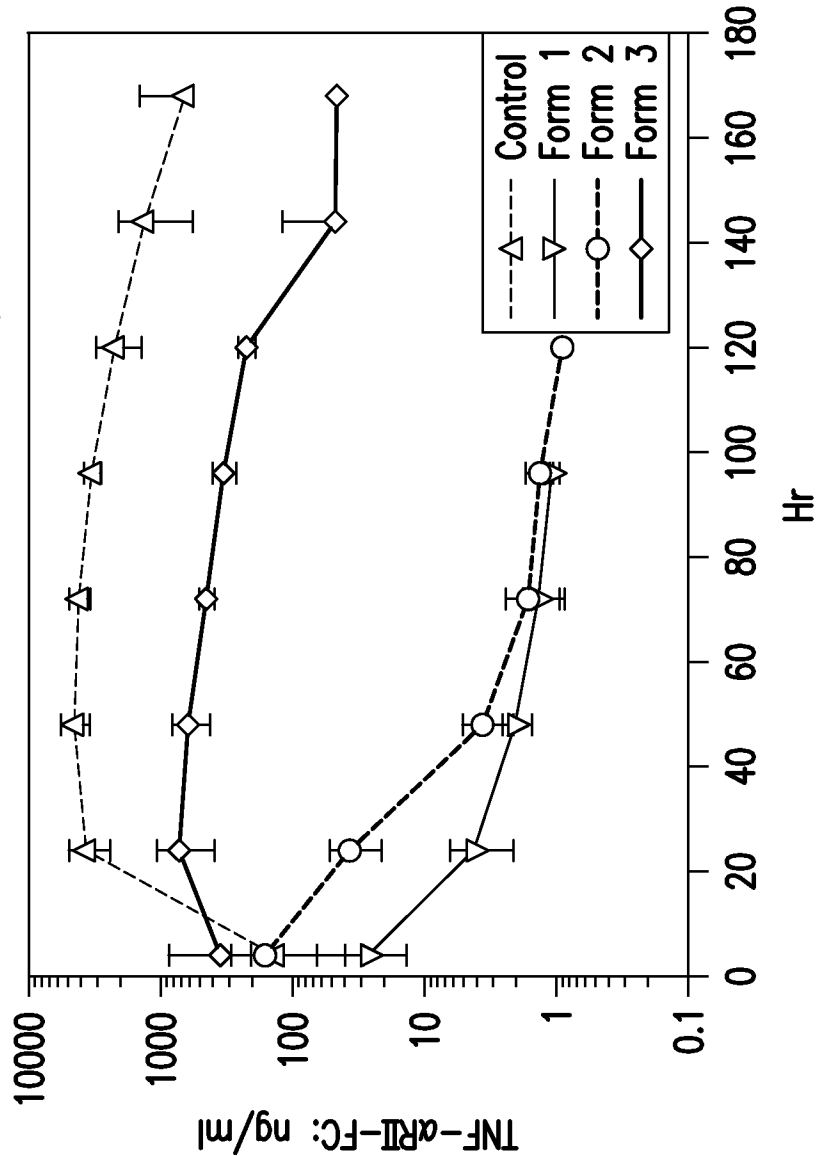


41/53

	EC50 (ng/ml)	
	Experiment 1	Experiment 2
Form 1	1.9	5.6
Form 2	2.5	6.9
Form 3	2.5	6.4
Control	2.0	5.6

FIG.34

In vivo rat pharmacokinetic analysis



PK parameters:

	C _{max} (ng/ml)	AUC _{last} (hr*ng/ml)
Form 1	26±13	461±243
Form 2	162±131	2879±1910
Form 3	773±333	59672±16489
Control	4574±944	447477±10882

FIG.35

Example of Purification strategy

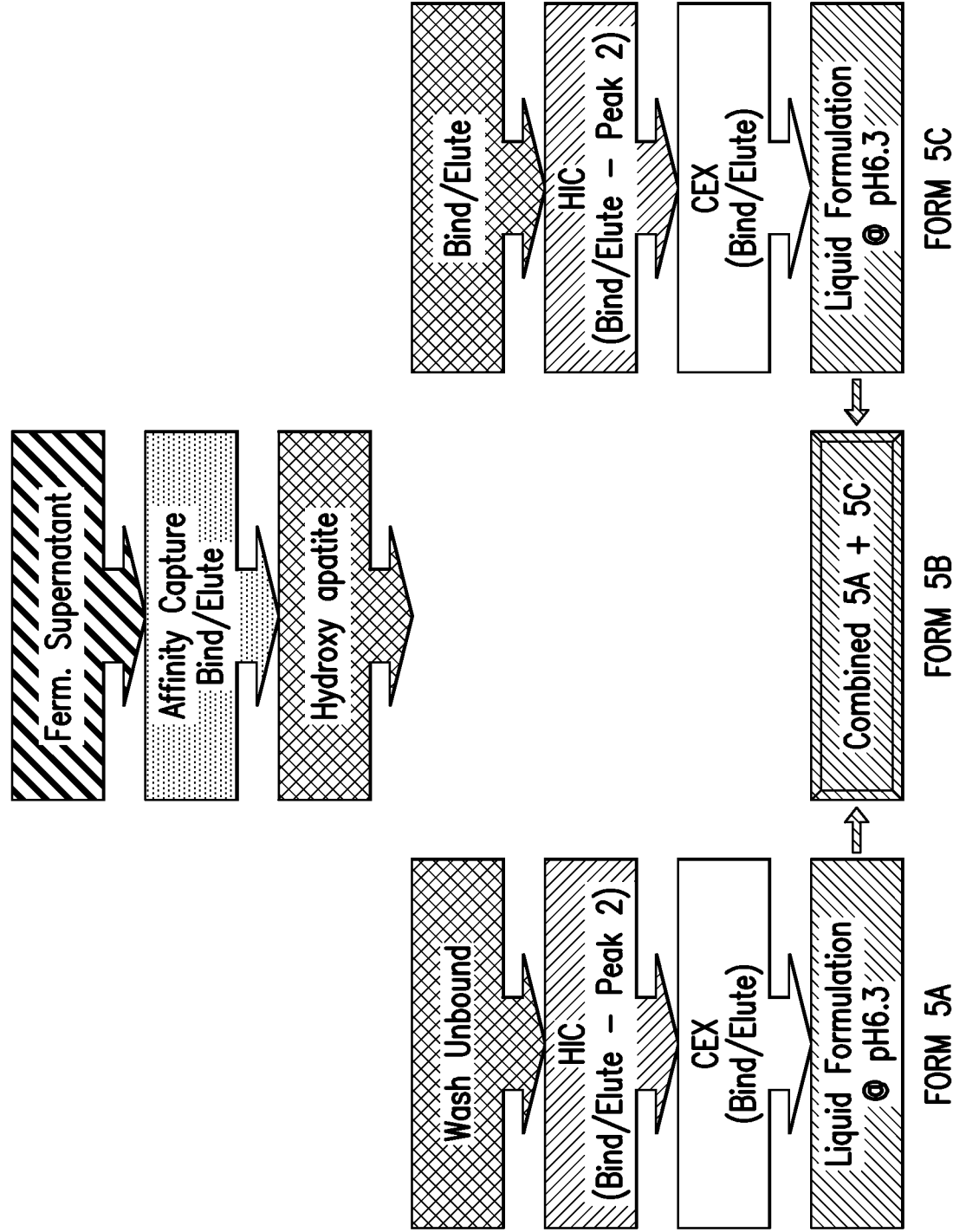


FIG.36

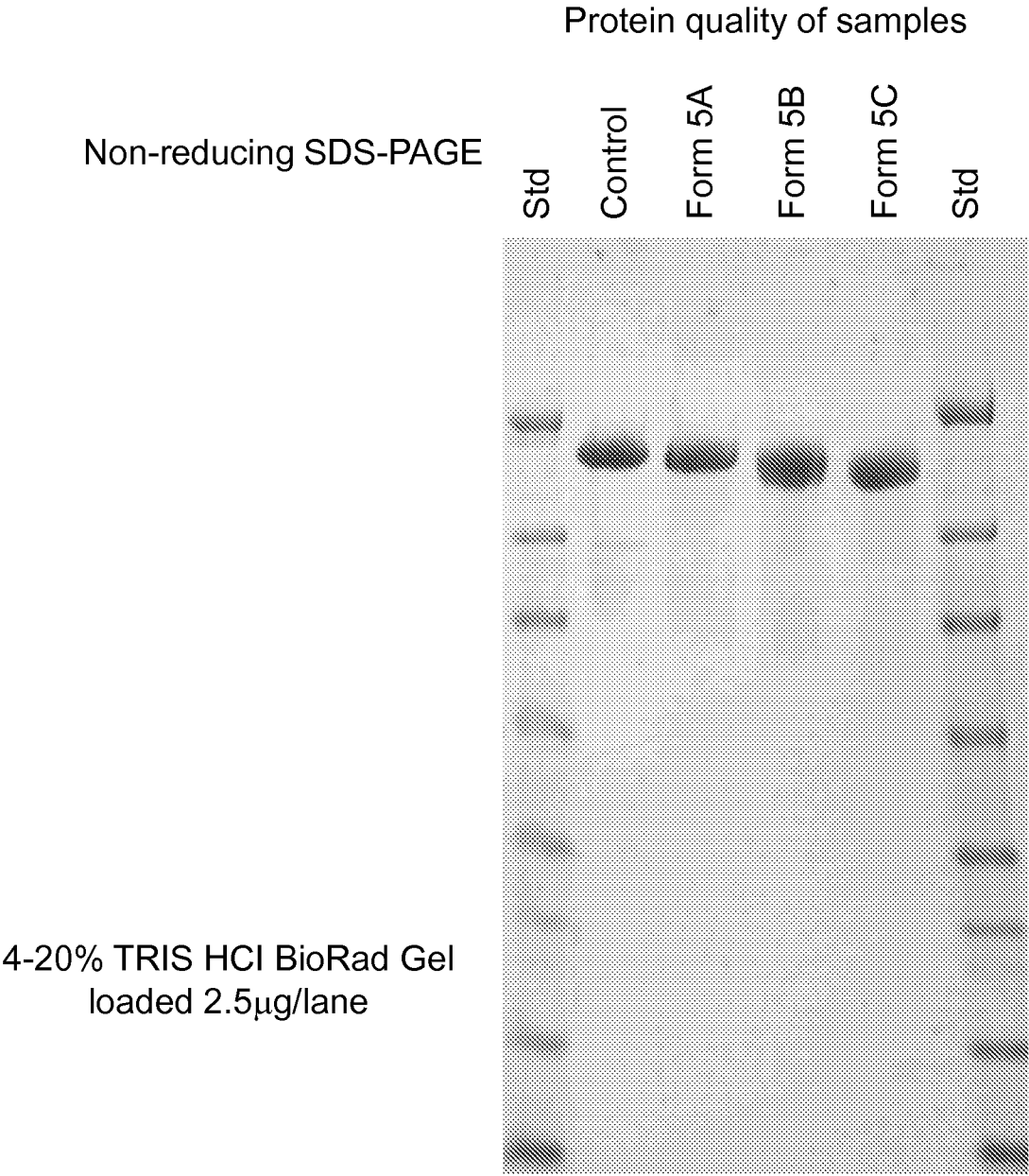


FIG.37

45/53

Comparison of Glycan Composition

Sample	Source	N-glycan Profile (% of total N-glycans)			Terminal O-glycan Profile (% of total O-glycans)			Total sialic acid (mol Sia per mol protein)
		Neutral	Mono-sialylated	Bi-sialylated	Mannose	Neutral (exc. mannose)	Sialylated	
Form 5A	YGLY14252 (HA wash/unbound)	8.9	28.5	62.6	26	7	67	21
Form 5B	YGLY14252 (1:1 mix of 5A & 5C)	22.4	37.7	39.9	23	24	53	16
Form 5C	YGLY14252 (HA bind/elute)	36.6	45	18.4	20	40	40	11
Control	Commercial ENBREL (liquid formulation, Lot B)	34.3	44.6	21.1	Not determined			25

FIG.38

46/53

In vitro TNF α induced cell killing of L929 cells and
TNF α -stimulated release of IL-6 in A549 cells

	EC ₅₀ (ng/ml)	
	L929 cytotoxicity assay	A549 TNF- induced IL-6 assay
Form 1	3.1	2.5
Form 2	2.7	2.7
Form 3	2.7	2.1
Control	2.6	1.9

FIG.39

In vivo rat pharmacokinetic analysis

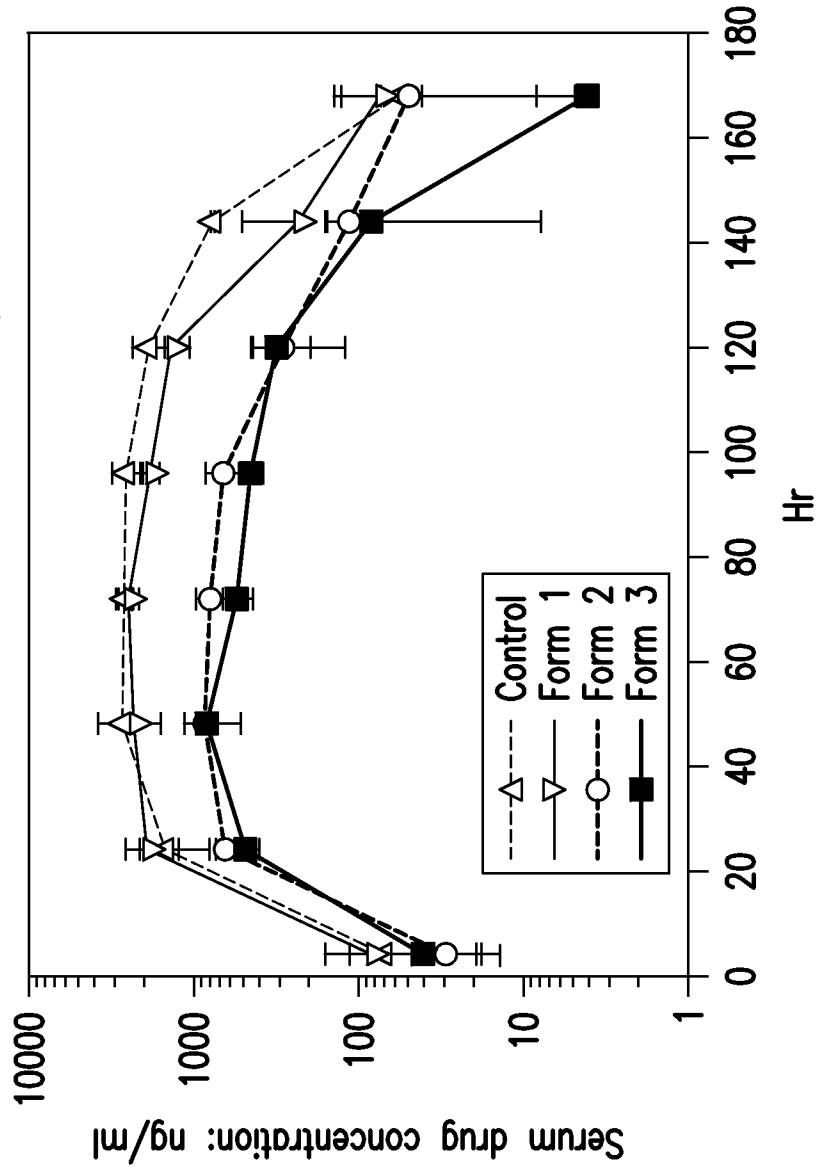


FIG.40

PK parameters:

	C _{max} (ng/ml)	AUC _{last} (hr*ng/ml)
Form 5A	2542±314	241793±30828
Form 5B	922±87	78275±13081
Form 5C	709±333	61204±9166
Control	2982±365	275043±35174

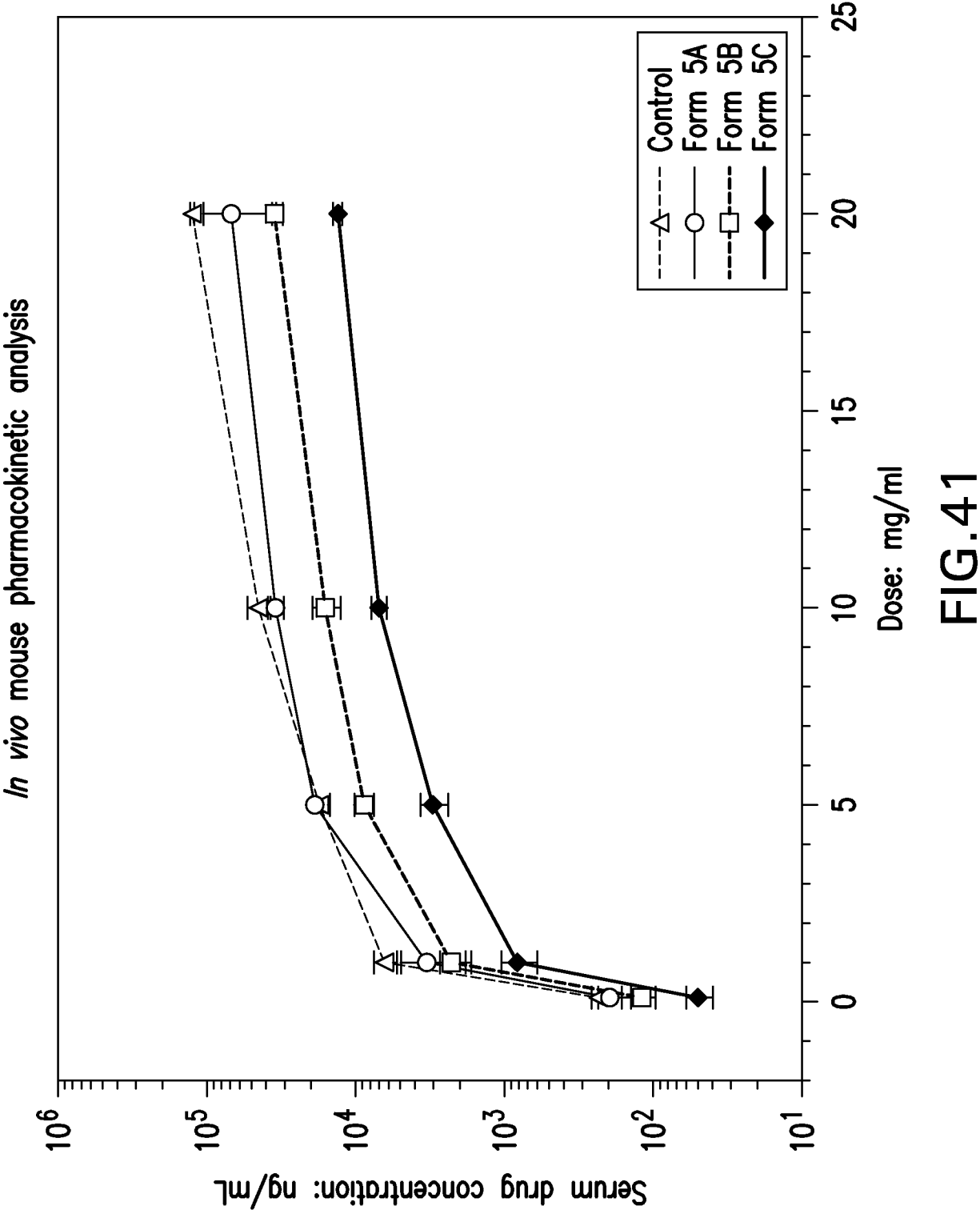


FIG.41

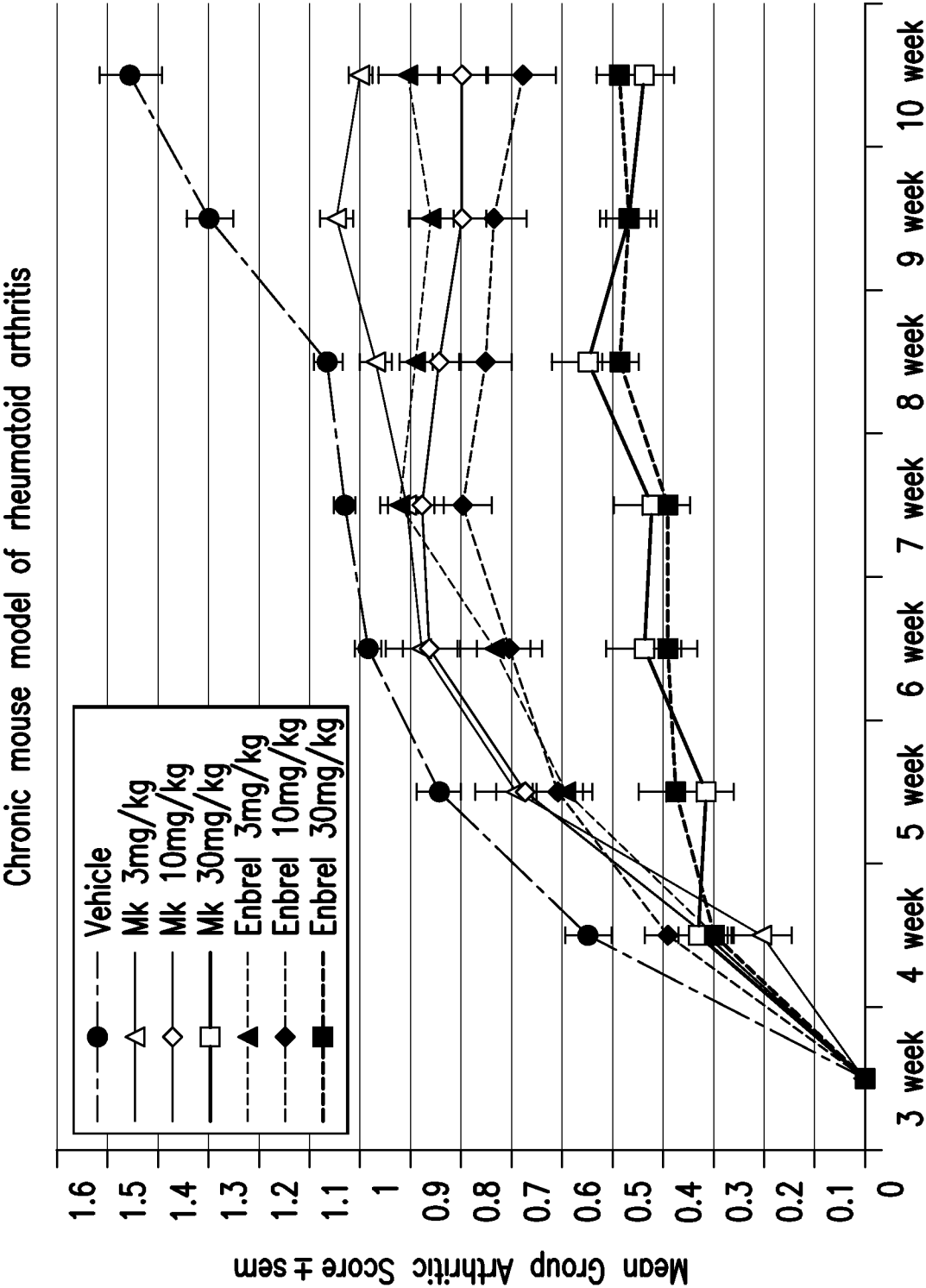


FIG.42

50/53

Example of Purification strategy

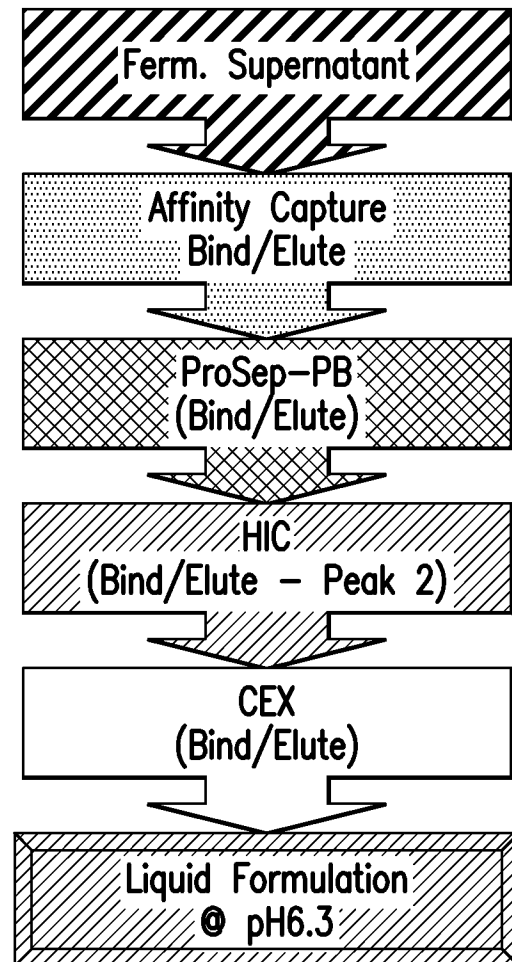


FIG.43

51/53

Protein quality of samples

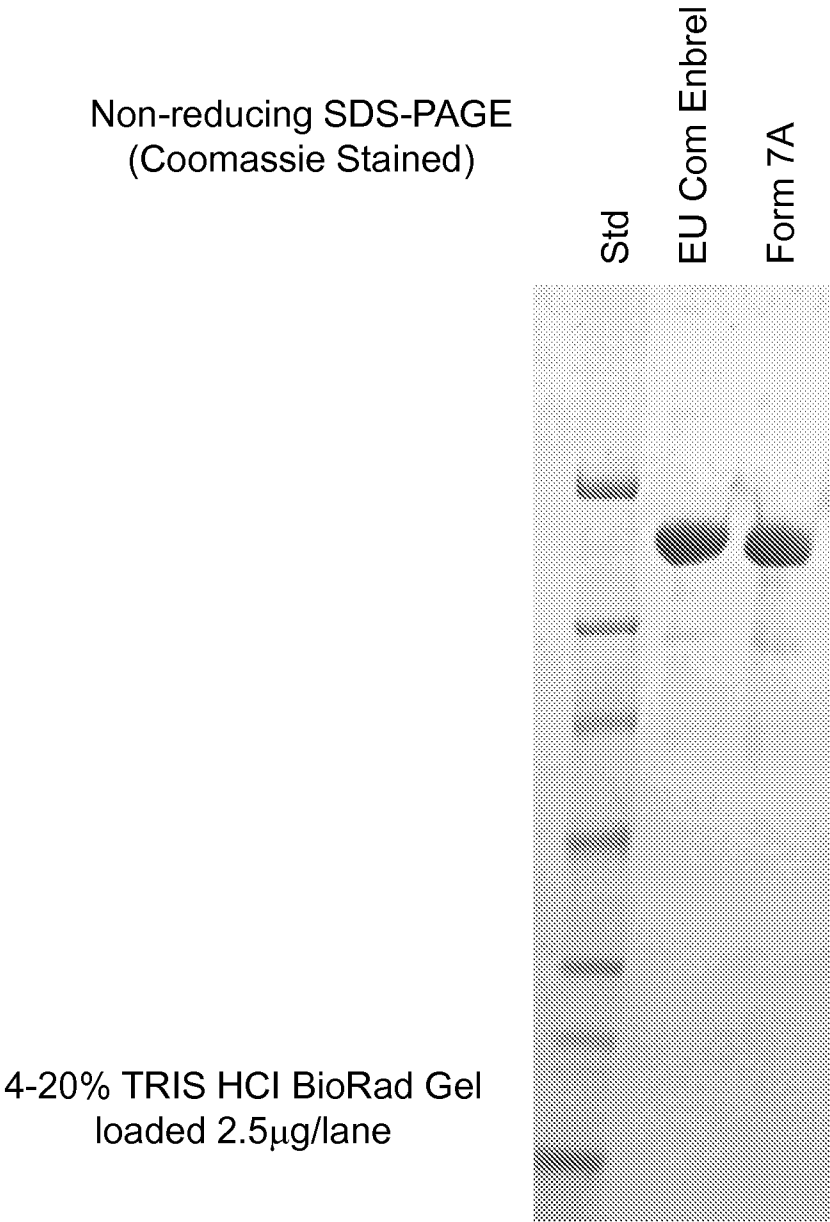


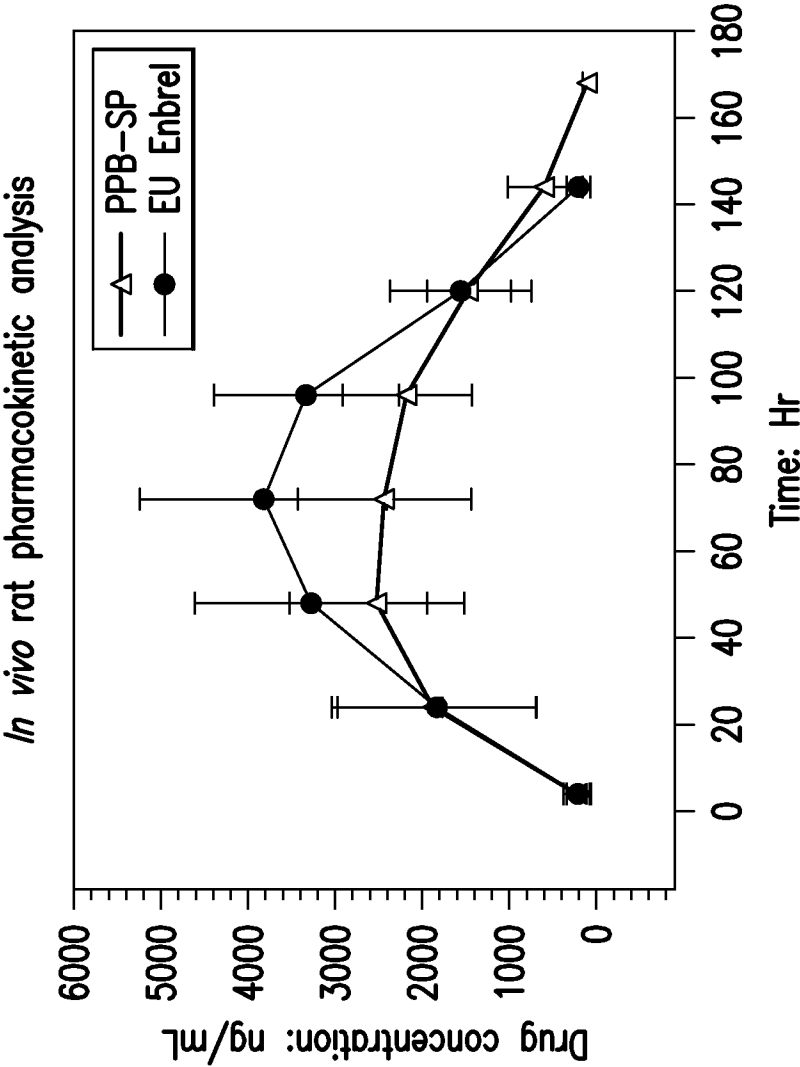
FIG.44

52/53

Comparison of Glycan Composition

Sample	Source	N-glycan Profile (% of total N-glycans)			Terminal O-glycan Profile (% of total O-glycans)			Total sialic acid (mol Sia per mol protein)
		Neutral	Mono-sialylated	Bi-sialylated	Mannose	Neutral (exc. mannose)	Sialylated	
Form 7A	YGLY14954 (PPB/SP pool)	15.3	31.3	53.4	22	10	68	18
Control	Commercial ENBREL (liquid formulation, Lot C)	31.3	52.3	16.4	Not determined			23

FIG. 45



	Process method	TSA (mol /mol)	Cmax (ng/mL)	AUClast (hr*ng/mL)	AUC0-INF (hr*ng/mL)	CL/F (mL/hr)
Commercial ENBREL (liquid formulation, LotC)	HA	23	3871±1413	328629±123096	338719±122991	3.48±1.89
Prosep	Prosep	18	2715±989	262552±96403	270892±90499	4.13±1.71

FIG.46