Title: OPTIMIZING GLYCAN PROCESSING IN PLANTS

Abstract: The invention is directed to methods for optimizing glycan processing in organisms (and in particular, plants) so that a glycoprotein having complex type bi-antennary glycans and thus containing galactose residues on both arms and which are devoid of (or reduce in) xylose and fucosyl can be obtained. The invention is further directed to said glycoprotein obtained and host system comprising said protein.
Published:
— without international search report and to be republished upon receipt of that report

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

FIELD OF THE INVENTION

The invention is directed to methods for optimizing glycan processing of cell or an organism containing glycoproteins with N-glycans, in particular plants so that a glycoprotein having an N-glycan, high mannose type, hybrid or preferably complex type N-glycans, including but not limited to bi-antennary N-glycans, and containing a galactose residue on at least one arm of the N-glycan and which are devoid of (or reduced in) xylose and fucose residues can be obtained. The invention is further directed to said glycoprotein obtained and in particular a plant host system comprising said protein.

BACKGROUND OF THE INVENTION

N-linked glycans, specific oligosaccharide structures attached to asparagine residues of glycoproteins, can contribute significantly to the properties of the protein and, in turn, to the properties of the organism. Plant proteins can carry N-linked glycans but in marked contrast to mammals only few biological processes are known to which they contribute.

Biogenesis of N-linked glycans begins with the synthesis of a lipid linked oligosaccharide moiety (Glc3Man9GlcNAc2-) which is transferred en bloc to the nascent polypeptide chain in the endoplasmic reticulum (ER). Through a series of trimming reactions by exoglycosidases in the ER and cis-Golgi compartments, the so-called "high mannose" (Man9GlcNAc2 to Man5GlcNAc2) glycans are formed. Subsequently, the formation of complex type glycans starts with the transfer of the first GlcNAc onto Man5GlcNAc2 by GnTI and further trimming by mannosidase II (ManII) to form GlcNAcMan3GlcNAc2. Complex glycan biosynthesis continues while the glycoprotein is progressing through the secretory pathway with the transfer in the Golgi apparatus of the second GlcNAc residue by GnTII as well as other monosaccharide residues onto the GlcNAcMan3GlcNAc2 under the action of several other glycosyl transferases.

Plants and mammals differ with respect to the formation of complex glycans (see Figure 1, which compares the glycosylation pathway of glycoproteins in plants and mammals). In plants, complex glycans are characterized by the presence of β(1,2)-xylose residues linked to the Man-3 and/or an α(1,3)-fucose residue linked to GlcNAc-1, instead of an α(1,6)-fucose residue linked to the GlcNAc-1. Genes encoding the corresponding xylosyl (XylT) and fucosyl (FucT) transferases have been isolated [Strasser et al., "Molecular cloning and functional expression of beta1,2-xylosyltransferase cDNA from Arabidopsis thaliana," FEBS Lett. 472:105 (2000); Leiter et al., "Purification, cDNA cloning, and expression of GDP-L-Fuc:Asn-linked GlcNAc alpha 1,3-fucosyltransferase from mung beans," J. Biol. Chem. 274:21830 (1999)]. Plants do not possess β(1,4)-galactosyltransferases nor α(2,6)sialyltransferases and consequently plant glycans lack the β(1,4)-galactose and terminal α(2,6)NeuAc residues often found on mammalian glycans.
The final glycan structures are not only determined by the mere presence of enzymes involved in their biosynthesis and transport but to a large extent by the specific sequence of the various enzymatic reactions. The latter is controlled by discrete sequestering and relative position of these enzymes throughout the ER and Golgi, which is mediated by the interaction of determinants of the transferase and specific characteristics of the sub-Golgi compartment for which the transferase is destined. A number of studies using hybrid molecules have identified that the transmembrane domains of several glycosyltransferases, including that of β(1,4)-galactosyltransferases, play a central role in their sub-Golgi sorting [Grabenhorst et al., J. Biol. Chem 274:36107 (1999); Colley, K., Glycobiology 7:1 (1997); Munro, S., Trends Cell Biol. 8:11 (1998); Gleeson, P.A., Histochem. Cell Biol. 109:517 (1998)].

Although plants and mammals have diverged a relatively long time ago, N-linked glycosylation seems at least partly conserved. This is evidenced by the similar though not identical glycan structures and by the observation that a mammalian GlcNAcTI gene complements a Arabidopsis mutant that is deficient in GlcNAcTI activity, and vice versa. The differences in glycan structures can have important consequences. For example, xylose and α(1,3)-fucose epitopes are known to be highly immunogenic and possibly allergenic in some circumstances, which may pose a problem when plants are used for the production of therapeutic glycoproteins. Moreover, blood serum of many allergy patients contains IgE directed against these epitopes but also 50% of non-allergic blood donors contains in their sera antibodies specific for core-xylose whereas 25% have antibodies for core-alpha 1,3-fucose (Bardor et al., 2002, in press, Glycobiology ) (Advance Access published December 17, 2002) which make these individuals at risk to treatments with recombinant proteins produced in plants containing fucose and/or xylose. In addition, this carbohydrate directed IgE in sera might cause false positive reaction in in vitro tests using plant extracts since there is evidence that these carbohydrate specific IgE’s are not relevant for the allergenic reaction. In sum, a therapeutic failure with a glycoprotein produced in plants might be the result of accelerated clearance of the recombinant glycoprotein having xylose and/or fucose.

Accordingly, there is a need to better control glycosylation in plants, and particularly, glycosylation of glycoproteins intended for therapeutic use.

**DEFINITIONS**

To facilitate understanding of the invention, a number of terms as used in this specification are defined below.

The term "vector" refers to any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, retrovirus, virion, or similar genetic element, which is capable of replication when associated with the proper control elements and which can transfer gene sequences into cells and/or between cells. Thus, this term includes cloning and expression vehicles, as well as viral vectors.
The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence (or coding sequences) – such as the coding sequence(s) for the hybrid enzyme(s) described in more detail below - and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host cell or organism.

Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter, an operator (optional), and a ribosome binding site, often along with other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals. It is not intended that the present invention be limited to particular expression vectors or expression vectors with particular elements.

The term "transgenic" when used in reference to a cell refers to a cell which contains a transgene, or whose genome has been altered by the introduction of a transgene. The term "transgenic" when used in reference to a cell, tissue or to a plant refers to a cell, tissue or plant, respectively, which comprises a transgene, where one or more cells of the tissue contain a transgene (such as a gene encoding the hybrid enzyme(s) of the present invention), or a plant whose genome has been altered by the introduction of a transgene. Transgenic cells, tissues and plants may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into a target cell or integration of the transgene into a chromosome of a target cell by way of human intervention, such as by the methods described herein.

The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations. A transgene may be an "endogenous DNA sequence," or a "heterologous DNA sequence" (i.e., "foreign DNA"). The term "endogenous DNA sequence" refers to a nucleotide sequence which is naturally found in the cell into which it is introduced so long as it does not contain some modification (e.g., a point mutation, the presence of a selectable marker gene, or other like modifications) relative to the naturally-occurring sequence.

The term "heterologous DNA sequence" refers to a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Heterologous DNA also includes an endogenous DNA sequence which contains some modification. Generally, although not necessarily, heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed. Examples of heterologous DNA include reporter genes, transcriptional and translational regulatory sequences, selectable marker proteins (e.g., proteins which confer drug resistance), or other similar elements.

The term "foreign gene" refers to any nucleic acid (e.g., gene sequence) which is introduced into the genome of a cell by experimental manipulations and may include gene sequences found in that cell so long as the introduced gene contains some modification (e.g., a point mutation, the presence of a selectable marker gene, or other like modifications) relative to the naturally-occurring gene.
The term "fusion protein" refers to a protein wherein at least one part or portion is from a first protein and another part or portion is from a second protein. The term "hybrid enzyme" refers to a fusion protein which is a functional enzyme, wherein at least one part or portion is from a first species and another part or portion is from a second species. Preferred hybrid enzymes of the present invention are functional glycosyltransferases (or portions thereof) wherein at least one part or portion is from a plant and another part or portion is from a mammal (such as human).

The term "introduction into a cell" or "introduction into a host cell" in the context of nucleic acid (e.g., vectors) is intended to include what the art calls "transformation" or "transfection" or "transduction." Transformation of a cell may be stable or transient – and the present invention contemplates introduction of vectors under conditions where, on the one hand, there is stable expression, and on the other hand, where there is only transient expression. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA) which detects the presence of a polypeptide encoded by one or more of the transgenes. Alternatively, transient transformation may be detected by detecting the activity of the protein (e.g., antigen binding of an antibody) encoded by the transgene (e.g., the antibody gene). The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes. In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more transgenes into the genome of a cell. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell with nucleic acid sequences which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction (PCR) of genomic DNA of the cell to amplify transgene sequences. The term "stable transformant" refers to a cell which has stably integrated one or more transgenes into the genomic DNA. Thus, a stable transformant is distinguished from a transient transformant in that, whereas genomic DNA from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene.

The term "host cell" includes both mammalian (e.g. human B cell clones, Chinese hamster ovary cells, hepatocytes) and non-mammalian cells (e.g. insect cells, bacterial cells, plant cells). In one embodiment, the host cells are mammalian cells and the introduction of a vector expressing a hybrid protein of the present invention (e.g TmGnTII-GalT) inhibits (or at least reduces) fucosylation in said mammalian cells.

The term "nucleotide sequence of interest" refers to any nucleotide sequence, the manipulation of which may be deemed desirable for any reason (e.g., confer improved qualities, use for production of therapeutic proteins), by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (e.g., reporter genes, selection marker genes, oncogenes, antibody genes, drug resistance genes, growth factors, and other like genes), and non-coding regulatory sequences which do not encode an mRNA or protein product,
(e.g., promoter sequence, polyadenylation sequence, termination sequence, enhancer sequence, and other like sequences). The present invention contemplates host cells expressing a heterologous protein encoded by a nucleotide sequence of interest along with one or more hybrid enzymes.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated nucleic acid sequence" refers to a nucleic acid sequence that is identified and separated from one or more other components (e.g., separated from a cell containing the nucleic acid, or separated from at least one contaminant nucleic acid, or separated from one or more proteins, one or more lipids) with which it is ordinarily associated in its natural source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, an isolated nucleic acid sequence comprising SEQ ID NO:1 includes, by way of example, such nucleic acid sequences in cells which ordinarily contain SEQ ID NO:1 where the nucleic acid sequence is in a chromosomal or extrachromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid sequence may be present in single-stranded or double-stranded form. When an isolated nucleic acid sequence is to be utilized to express a protein, the nucleic acid sequence will contain at a minimum at least a portion of the sense or coding strand (i.e., the nucleic acid sequence may be single-stranded). Alternatively, it may contain both the sense and anti-sense strands (i.e., the nucleic acid sequence may be double-stranded).

As used herein, the term "purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated. An "isolated nucleic acid sequence" is therefore a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free, from other components with which they are naturally associated. The present invention contemplates both purified (including substantially purified) and unpurified hybrid enzyme(s) (which are described in more detail below).

As used herein, the terms "complementary" or "complementarity" are used in reference to nucleotide sequences related by the base-pairing rules. For example, the sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.
A "complement" of a nucleic acid sequence as used herein refers to a nucleotide sequence whose nucleic acids show total complementarity to the nucleic acids of the nucleic acid sequence. For example, the present invention contemplates the complements of SEQ ID NOS: 1, 3, 5, 9, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 40, 41 and 43.

The term "homology" when used in relation to nucleic acids refers to a degree of complementarity. There may be partial homology (i.e., partial identity) or complete homology (i.e., complete identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe (i.e., an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest) will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described infra.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize to the single-stranded nucleic acid sequence under conditions of low stringency as described infra.

The term "hybridization" as used herein includes "any process by which a strand of nucleic acid joins with a complementary strand through base pairing." [Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY]. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein, the term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: \( T_m = 81.5 + 0.41(\%\ G + C) \), when a nucleic acid is in aqueous solution.
at 1 M NaCl [see e.g., Anderson and Young, Quantitative Filter Hybridization, in: Nucleic Acid Hybridization (1985)]. Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T_m.

Low stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE (Saline, Sodium Phosphate, EDTA) (43.8 g/l NaCl, 6.9 g/l NaH2PO4•H2O and 1.85 g/l EDTA (Ethylenediaminetetraacetic Acid), pH adjusted to 7.4 with NaOH), 0.1% SDS (Sodium dodecyl sulfate), 5X Denhardt's reagent [50X Denhardt's contains the following per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Bovine Serum Albumin) (Fraction V; Sigma)] and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising between 0.2X and 2.0X SSPE, and 0.1% SDS at room temperature when a DNA probe of about 100 to about 1000 nucleotides in length is employed.

High stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE, 1% SDS, 5X Denhardt's reagent and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, and 0.1% SDS at 68°C when a probe of about 100 to about 1000 nucleotides in length is employed.

The term "equivalent" when made in reference to a hybridization condition as it relates to a hybridization condition of interest means that the hybridization condition and the hybridization condition of interest result in hybridization of nucleic acid sequences which have the same range of percent (%) homology. For example, if a hybridization condition of interest results in hybridization of a first nucleic acid sequence with other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence, then another hybridization condition is said to be equivalent to the hybridization condition of interest if this other hybridization condition also results in hybridization of the first nucleic acid sequence with the other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence.

When used in reference to nucleic acid hybridization the art knows well that numerous equivalent conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency hybridization different from, but equivalent to, the above-listed conditions.

The term "promoter," "promoter element," or "promoter sequence" as used herein, refers to a DNA sequence which when ligated to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into mRNA. A promoter is typically, though not necessarily, located 5' (i.e., upstream) of a nucleotide sequence of interest whose transcription into
mRNA it controls, and provides a site for specific binding by RNA polymerase and other transcription factors for initiation of transcription.

Promoters may be tissue specific or cell specific. The term "tissue specific" as it applies to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (e.g., petals) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (e.g., roots). Tissue specificity of a promoter may be evaluated by, for example, operably linking a reporter gene to the promoter sequence to generate a reporter construct, introducing the reporter construct into the genome of a plant such that the reporter construct is integrated into every tissue of the resulting transgenic plant, and detecting the expression of the reporter gene (e.g., detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic plant. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the promoter is specific for the tissues in which greater levels of expression are detected. The term "cell type specific" as applied to a promoter refers to a promoter which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a promoter also means a promoter capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue. Cell type specificity of a promoter may be assessed using methods well known in the art, e.g., immunohistochemical staining. Briefly, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is controlled by the promoter. A labeled (e.g., peroxidase conjugated) secondary antibody which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (e.g., with avidin/biotin) by microscopy.

Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (e.g., heat shock, chemicals, light, or similar stimuli). Typically, constitutive promoters are capable of directing expression of a transgene in substantially any cell and any tissue. In contrast, a "regulatable" promoter is one which is capable of directing a level of transcription of an operably linked nucleic acid sequence in the presence of a stimulus (e.g., heat shock, chemicals, light, or similar stimuli) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

The terms "infecting" and "infection" with a bacterium refer to co-incubation of a target biological sample, (e.g., cell, tissue, plant part) with the bacterium under conditions such that nucleic acid sequences contained within the bacterium are introduced into one or more cells of the target biological sample.
The term "Agrobacterium" refers to a soil-borne, Gram-negative, rod-shaped phytopathogenic bacterium which causes crown gall. The term "Agrobacterium" includes, but is not limited to, the strains Agrobacterium tumefaciens, (which typically causes crown gall in infected plants), and Agrobacterium rhizogenes (which causes hairy root disease in infected host plants). Infection of a plant cell with Agrobacterium generally results in the production of opines (e.g., nopaline, agropine, octopine) by the infected cell. Thus, Agrobacterium strains which cause production of nopaline (e.g., strain LBA4301, C58, A208) are referred to as "nopaline-type" Agrobacteria; Agrobacterium strains which cause production of octopine (e.g., strain LBA4404, Ach5, B6) are referred to as "octopine-type" Agrobacteria; and Agrobacterium strains which cause production of agropine (e.g., strain EHA105, EHA101, A281) are referred to as "agropine-type" Agrobacteria.

The terms "bombarding," "bombardment," and "biolistic bombardment" refer to the process of accelerating particles towards a target biological sample (e.g., cell, tissue, plant part — such as a leaf, or intact plant) to effect wounding of the cell membrane of a cell in the target biological sample and/or entry of the particles into the target biological sample. Methods for biolistic bombardment are known in the art (e.g., U.S. Patent Nos. 5,584,807 and 5,141,131, the contents of both are herein incorporated by reference), and are commercially available (e.g., the helium gas-driven microprojectile accelerator (PDS-1000/He) (BioRad).

The term "microwounding" when made in reference to plant tissue refers to the introduction of microscopic wounds in that tissue. Microwounding may be achieved by, for example, particle bombardment as described herein. The present invention specifically contemplates schemes for introducing nucleic acid which employ microwounding.

The term “organism” as used herein refers to all organisms and in particular organisms containing glycoproteins with n-linked glycans.

The term "plant" as used herein refers to a plurality of plant cells which are largely differentiated into a structure that is present at any stage of a plant's development. Such structures include, but are not limited to, a fruit, shoot, stem, root, leaf, seed, flower petal, or similar structure. The term "plant tissue" includes differentiated and undifferentiated tissues of plants including, but not limited to, roots, shoots, leaves, pollen, seeds, tumor tissue and various types of cells in culture (e.g., single cells, protoplasts, embryos, callus, protocorm-like bodies, and other types of cells). Plant tissue may be in planta, in organ culture, tissue culture, or cell culture. Similarly, "plant cells" may be cells in culture or may be part of a plant.

Glycosyltransferases are enzymes that catalyze the processing reactions that determine the structures of cellular oligosaccharides, including the oligosaccharides on glycoproteins. As used herein, “glycosyltransferase” is meant to include mannosidases, even though these enzymes trim glycans and do not “transfer” a monosaccharide. Glycosyltransferases share the feature of a type II membrane orientation. Each glycosyltransferase is comprised of an amino terminal cytoplasmic tail (shown for illustration purposes below as a made up of a string of amino acids arbitrarily labeled “X”)
without intending to suggest the actual size of the region), a signal anchor domain (shown below as made up of a string of amino acids labeled “H” for hydrophobic – without intending to suggest the actual size of the domain and without intending to suggest that the domain is only made up of hydrophobic amino acids) that spans the membrane (referred to herein as a “transmembrane domain”), followed by a luminal stem (shown below as made up of a string of amino acids arbitrarily labeled “S” – without intended to suggest the actual size of the region) or stalk region, and a carboxy-terminal catalytic domain (shown below as made up of a string of amino acids arbitrarily labeled “C” – without intending to suggest the actual size of the domain:

\[
\text{NH}_2-\text{XXXXXXXHHHHHHHHSSSSSSSSSSSCCCCCCCCC}
\]

Collectively, The Cytoplasmic Tail-Transmembrane-Stem Region or “CTS” (which has been underlined in the above schematic for clarity) can be used (or portions thereof) in embodiments contemplated by the present invention wherein the catalytic domain is exchanged or “swapped” with a corresponding catalytic domain from another molecule (or portions of such regions/domains) to create a hybrid protein.

For example, in a preferred embodiment, the present invention contemplates nucleic acid encoding a hybrid enzyme (as well as vectors containing such nucleic acid, host cells containing such vectors, and the hybrid enzyme itself), said hybrid enzyme comprising at least a portion of a CTS region (e.g., the cytoplasmic tail (“C”), the transmembrane domain (“T”), the cytoplasmic tail together with the transmembrane domain (“CT”), the transmembrane domain together with the stem (“TS”), or the complete CTS region] of a first glycosyltransferase (e.g. plant glycosyltransferase) and at least a portion of a catalytic region of a second glycosyltransferase (e.g. mammalian glycosyltransferase). To create such an embodiment, the coding sequence for the entire CTS region (or portion thereof) may be deleted from nucleic acid coding for the mammalian glycosyltransferase and replaced with the coding sequence for the entire CTS region (or portion thereof) of a plant glycosyltransferase. On the other hand, a different approach might be taken to create this embodiment; for example, the coding sequence for the entire catalytic domain (or portion thereof) may be deleted from the coding sequence for the plant glycosyltransferase and replaced with the coding sequence for the entire catalytic domain (or portion thereof) of the mammalian glycosyltransferase. In such a case, the resulting hybrid enzyme would have the amino-terminal cytoplasmic tail of the plant glycosyltransferase linked to the plant glycosyltransferase transmembrane domain linked to the stem region of the plant glycosyltransferase in the normal manner of the wild-type plant enzyme – but the stem region would be linked to the catalytic domain of the mammalian glycosyltransferase (or portion thereof).

It is not intended that the present invention be limited only to the two approaches outlined above. Other variations in the approach are contemplated. For example, to create nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a transmembrane region of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian
glycosyltransferase, one might use less than the entire coding sequence for the CTS region (e.g., only
the transmembrane domain of the plant glycosyltransferase, or the complete cytoplasmic tail together
with all or a portion of the transmembrane domain, or the complete cytoplasmic tail together with all
of the transmembrane domain together with a portion of the stem region). One might delete the
mammalian coding sequence for the entire cytoplasmic tail together with the coding sequence for the
transmembrane domain (or portion thereof) – followed by replacement with the corresponding
coding sequence for the cytoplasmic tail and transmembrane domain (or portion thereof) of the plant
glycosyltransferase. In such a case, the resulting hybrid enzyme would have the stem region of the
mammalian glycosyltransferase linked to the plant glycosyltransferase transmembrane domain (or
portion thereof) which in turn would be linked to the amino-terminal cytoplasmic tail of the plant
glycosyltransferase, with the stem region being linked to the catalytic domain of the mammalian
glycosyltransferase (i.e., two of the four regions/domains would be of plant origin and two would be
of mammalian origin).

In other embodiments, the present invention contemplates nucleic acid encoding a hybrid
enzyme (along with vectors, host cells containing the vectors, plants – or plant parts - containing the
host cells), said hybrid enzyme comprising at least a portion of an amino-terminal cytoplasmic tail of
a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian
glycosyltransferase. In this embodiment, the hybrid enzyme encoded by the nucleic acid might or
might not contain other plant sequences (e.g., the transmembrane domain or portion thereof, the stem
region or portion thereof). For example, to create such an embodiment, the coding sequence for the
entire cytoplasmic tail (or portion thereof) may be deleted from nucleic acid coding for the
mammalian glycosyltransferase and replaced with the coding sequence for the entire cytoplasmic
domain (or portion thereof) of a plant glycosyltransferase. In such a case, the resulting hybrid
enzyme would have the amino-terminal cytoplasmic tail (or portion thereof) of the plant
glycosyltransferase linked to the mammalian glycosyltransferase transmembrane domain, which in
turn is linked to stem region of the mammalian glycosyltransferase, the stem region being linked to
the catalytic domain of the mammalian glycosyltransferase. On the other hand, a different approach
might be taken to create this embodiment; for example, the coding sequence for the entire catalytic
domain (or portion thereof) may be deleted from the coding sequence for the plant
glycosyltransferase and replaced with the coding sequence for the entire catalytic domain (or portion
thereof) of the mammalian glycosyltransferase. In such a case, the resulting hybrid enzyme would
have the amino-terminal cytoplasmic tail of the plant glycosyltransferase linked to the plant
glycosyltransferase transmembrane domain linked to the stem region of the plant glycosyltransferase
in the normal manner of the wild-type plant enzyme – but the stem region would be linked to the
catalytic domain of the mammalian glycosyltransferase (or portion thereof).

In the above discussion, the use of the phrase “or portion thereof” was used to expressly
indicate that less than the entire region/domain might be employed in the particular case (e.g., a
fragment might be used). For example, the cytoplasmic tail of glycosyltransferases ranges from
approximately 5 to 50 amino acids in length, and more typically 15 to 30 amino acids, depending on the particular transferase. A “portion” of the cytoplasmic tail region is herein defined as no fewer than four amino acids and can be as large as up to the full length of the region/domain less one amino acid. It is desired that the portion function in a manner analogous to the full length region/domain — but need not function to the same degree. For example, to the extent the full-length cytoplasmic tail functions as a Golgi retention region or ER retention signal, it is desired that the portion employed in the above-named embodiments also function as a Golgi or ER retention region, albeit perhaps not as efficiently as the full-length region.

Similarly, the transmembrane domain is typically 15-25 amino acids in length and made up of primarily hydrophobic amino acids. A “portion” of the transmembrane domain is herein defined as no fewer than ten amino acids and can be as large as up to the full length of the region/domain (for the particular type of transferase) less one amino acid. It is desired that the portion function in a manner analogous to the full length region/domain — but need not function to the same degree. For example, to the extent the full-length transmembrane domain functions as the primary Golgi retention region or ER retention signal, it is desired that the portion employed in the above-named embodiments also function as a Golgi or ER retention region, albeit perhaps not as efficiently as the full-length region. The present invention specifically contemplates conservative substitutions to create variants of the wild-type transmembrane domain or portions thereof. For example, the present invention contemplates replacing one or more hydrophobic amino acids (shown as “H” in the schematic above) of the wild-type sequence with one or more different amino acids, preferably also hydrophobic amino acids.

A portion of the catalytic domain can be as large as the full length of the domain less on amino acid. Where the catalytic domain is from a beta1,4-galactosyltransferase, it is preferred that the portion include at a minimum residues 345-365 which are believed to be involved in the conformation conferring an oligosaccharide acceptor binding site (it is preferred that the portion include this region at a minimum and five to ten amino acids on either side to permit the proper conformation).

The present invention also includes synthetic CTS regions and portions thereof. A “portion” of a CTS region must include at least one (and may include more than one) entire domain (e.g., the entire transmembrane domain) but less than the entire CTS region.

Importantly, by using the term “CTS region” or “transmembrane domain” it is not intended that only wild type sequences be encompassed. Indeed, this invention is not limited to natural glycosyltransferases and enzymes involved in glycosylation, but also includes the use of synthetic enzymes exhibit the same or similar function. In one embodiment, wild type domains are changed (e.g. by deletion, insertion, replacement and the like).

Finally, by using the indicator “Tm” when referring to a particular hybrid (e.g., “TmXyl”), entire transmembrane/CTS domains (with or without changes to the wild-type sequence) as well as portions (with or without changes to the wild-type sequence) are intended to be encompassed.
SUMMARY OF THE INVENTION

The present invention contemplates nucleic acid (whether DNA or RNA) encoding hybrid enzymes (or "fusion proteins"), vectors containing such nucleic acid, host cells (including but not limited to cells in plant tissue and whole plants) containing such vectors an expressing the hybrid enzymes, and the isolated hybrid enzyme(s) themselves. In one embodiment, expression of said hybrid enzymes (or "fusion proteins") results in changes in glycosylation, such as, but not limited to, reduction of sugar moieties such as xylose, fucose, Lewis\textsuperscript{A/B/X} or other sugar structures that interfere with desired glycoform accumulation. In one embodiment, the present invention contemplates nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a CTS region (or portion thereof) of a glycosyltransferase (including but not limited to a plant glycosyltransferase) and a catalytic region (or portion thereof) of a non-plant glycosyltransferase (e.g., mammalian, fish, amphibian, fungal). It is preferred that, when expressed, the CTS region (or portion thereof) is linked (directly or indirectly) in operable combination to said catalytic region (or portion thereof).

The linking is preferably covalent and the combination is operable in that the catalytic region exhibits catalytic function (even if said catalytic function is reduced as compared to the wild-type enzyme). The linking can be direct in the sense that there are no intervening amino acids or other regions/domains. On the other hand, the linking can be indirect in that there are intervening amino acids (or other chemical groups) and/or other regions/domains between them. Of course, the nucleic acid used to make the nucleic acid encoding the above-described hybrid enzyme(s) can be obtained enzymatically from a physical sequence (e.g. genomic DNA, a cDNA, and the like) or alternatively, made synthetically using a reference sequence (e.g. electronic or hardcopy sequence) as a guide.

In a particular embodiment, the present invention contemplates nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region (e.g., at least a transmembrane region and optionally more of the CTS region) of a plant glycosyltransferase and a catalytic region (or portion thereof) of a non-plant (such as a mammalian) glycosyltransferase. Again, it is preferred that, when expressed, these regions are linked (directly or indirectly) in operable combination. In yet another embodiment, the present invention contemplates nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane domain (or portion thereof) of a plant glycosyltransferase and a catalytic region (or portion thereof) of a mammalian glycosyltransferase. Again, it is preferred that, when expressed, these regions are linked (directly or indirectly) in operable combination.

It is not intended that the present invention be limited to particular transferases. In one embodiment, the plant glycosyltransferase is a xylosyltransferase. In another embodiment, the plant glycosyltransferase is a N-acetylglicosaminyltransferase. In another embodiment, the plant glycosyltransferase is a fucosyltransferase. In a preferred embodiment, the mammalian glycosyltransferase is a human galactosyltransferase (such as the human beta 1,4-
galactosyltransferase encoded by SEQ ID NO:1 wherein the nucleotides encoding the transmembrane domain are deleted and replaced.

It is not intended that the present invention is limited to the use of a plant-derived glycosyltransferase CTS-domain and a human glycosyltransferase catalytic domain but also vice versa and the use of any CTS-domain of a glycosyltransferase in combination with the catalytic fragment of at least one other glycosyltransferase. Indeed, the present invention broadly contemplates, in one embodiment, nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a first glycosyltransferase and a catalytic region of a second glycosyltransferase. It is preferred that said first and second glycosyltransferases are from different species (and can be from a different genus or even from a different phylum). In one embodiment, said first glycosyltransferase comprises a plant glycosyltransferase. In another embodiment, said plant glycosyltransferase is a xylosyltransferase. In yet another embodiment, said plant glycosyltransferase is a fucosyltransferase. In a preferred embodiment said second glycosyltransferase comprises a mammalian glycosyltransferase. In a particularly preferred embodiment, said mammalian glycosyltransferase is a human galactosyltransferase.

It is not intended that the present invention be limited to circumstances where the first and second glycosyltransferases are plant and non-plant, respectively. In one embodiment, said first glycosyltransferase comprises a first mammalian glycosyltransferase and said second glycosyltransferase comprises a second mammalian glycosyltransferase. In a preferred embodiment, said first mammalian glycosyltransferase is a non-human glycosyltransferase and said second mammalian glycosyltransferase is a human glycosyltransferase.

It is not intended that the present invention be limited to the type of vector. In one embodiment, the present invention contemplates an expression vector, comprising the nucleic acid encoding the above-described hybrid enzyme.

It is also not intended that the present invention be limited to the type of host cells. A variety of prokaryotic and eukaryotic host cells are commercially available for expressing proteins. In one embodiment, the present invention contemplates a host cell containing the vector comprising the nucleic acid encoding the above-described hybrid enzyme (with or without other vectors or other nucleic acid encoding other hybrid enzymes or glycosyltransferases). In a preferred embodiment, the host cell is a plant cell. In a particularly preferred embodiment, the present invention contemplates a plant comprising such a host cell.

It is not intended that the present invention be limited by the method by which host cells are made to express the hybrid enzymes of the present invention. In one embodiment, the present invention contemplates a method, comprising: a) providing: i) a host cell (such as a plant cell, whether in culture or as part of plant tissue or even as part of an intact growing plant), and ii) an expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a CTS region of a plant glycosyltransferase (e.g. the transmembrane domain) and at least a portion of a catalytic region of a mammalian glycosyltransferase; and b)
introducing said expression vector into said plant cell under conditions such that said hybrid enzyme is expressed. Again, it is not intended that the present invention be limited to particular transferases. In one embodiment, the plant glycosyltransferase used in the above-described method is a xylosyltransferase. In another embodiment, the plant glycosyltransferase is an N-acetylgalactosaminyltransferase. In another embodiment, the plant glycosyltransferase is a fucosyltransferase. In a preferred embodiment, the mammalian glycosyltransferase used in the above-described method is a human galactosyltransferase (such as the human beta 1,4-galactosyltransferase encoded by SEQ ID NO:1 wherein the nucleotides encoding the transmembrane domain are deleted and replaced) (or simply where the nucleotides of SEQ ID NO:1 encoding the catalytic domain, or portion thereof, are taken and linked to nucleotides encoding the CTS region, or portion thereof, of a plant glycosyltransferase). It is not intended that the present invention be limited to a particular scheme for controlling glycosylation of a heterologous protein using the hybrid enzymes described above. In one embodiment, the present invention contemplates a method, comprising: a) providing: i) a host cell (such as a plant cell), ii) a first expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a CTS region (e.g., at least a transmembrane domain) of a first (such as a plant) glycosyltransferase and at least a portion of a catalytic region of a second (such as a mammalian) glycosyltransferase, and iii) a second expression vector comprising nucleic acid encoding a heterologous glycoprotein; (or portion thereof; and b) introducing said first and second expression vectors into said plant cell under conditions such that said hybrid enzyme and said heterologous protein are expressed. Alternatively, a single vector with nucleic acid encoding both the hybrid enzyme (or hybrid enzymes) and the heterologous glycoprotein might be used. Regardless of which method is used, the invention contemplates, in one embodiment, the additional step (c) of isolating the heterologous protein — as well as the isolated protein itself as a composition. On the other hand, the present invention also contemplates introducing different vectors into different plant cells (whether they are cells in culture, part of plant tissue, or even part of an intact growing plant). In one embodiment, the present invention contemplates a method, comprising: a) providing: i) a first plant comprising a first expression vector, said first vector comprising nucleic acid encoding a hybrid enzyme (or encoding two or more hybrid enzymes), said hybrid enzyme comprising at least a portion of a CTS region (e.g., the first approximately 40-60 amino acids of the N-terminus) of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian glycosyltransferase, and ii) a second plant comprising a second expression vector, said second vector comprising nucleic acid encoding a heterologous protein (or portion thereof); and crossing said first plant and said second plant to produce progeny expressing said hybrid enzyme and said heterologous protein. Of course, such progeny can be isolated, grown up, and analyzed for the presence of each (or both) of the proteins. Indeed, the heterologous protein can be used (typically first purified substantially free of plant cellular material) therapeutically (e.g., administered to a human or animal,
whether orally, by intravenous, transdermally or by some other route of administration) to treat or prevent disease.

It is not intended that the present invention be limited to a particular heterologous protein. In one embodiment, any peptide or protein that is not endogenous to the host cell (or organism) is contemplated. In one embodiment, the heterologous protein is an antibody or antibody fragment. In a particularly preferred embodiment, the antibody is a human antibody or “humanized” antibody expressed in a plant in high yield. “Humanized” antibodies are typically prepared from non-human antibodies (e.g. rodent antibodies) by taking the hypervariable regions (the so-called CDRs) of the non-human antibodies and “grafting” them on to human frameworks. The entire process can be synthetic (provided that the sequences are known) and frameworks can be selected from a database of common human frameworks. Many times, there is a loss of affinity in the process unless either the framework sequences are modified or the CDRs are modified. Indeed, increases in affinity can be revealed when the CDRs are systematically mutated (for example, by randomization procedures) and tested.

While the present invention is particularly useful in the context of heterologous proteins, in one embodiment, the hybrid enzymes of the present invention are used to change the glycosylation of endogenous proteins, i.e. proteins normally expressed by the host cell or organism.

The present invention specifically contemplates the plants themselves. In one embodiment, the present invention contemplates a plant, comprising first and second expression vectors, said first vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a CTS region (e.g. the cytoplasmic tail together with at least a portion of the transmembrane domain) of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian glycosyltransferase, said second expression vector, said second vector comprising nucleic acid encoding a heterologous protein (or portion thereof). In a preferred embodiment, by virtue of being expressed along with the hybrid enzyme (or hybrid enzymes) of the present invention, the heterologous protein displays reduced (10% to 99%) alpha 1,3-fucosylation (or even no fucosylation), as compared to when the heterologous protein is expressed in the plant in the absence of the hybrid enzyme (or enzymes). In a preferred embodiment, by virtue of being expressed along with the hybrid enzyme (or hybrid enzymes) of the present invention, the heterologous protein displays reduced (10% to 99%) xylosylation (or even no xylose), as compared to when the heterologous protein is expressed in the plant in the absence of the hybrid enzyme (or enzymes).

It is not intended that the present invention be limited to a particular theory by which reduced fucose and/or xylose is achieved. Very little is known about the sub-Golgi sorting mechanism in plants. The mammalian specific β(1,4)-galactosyltransferase (GaIT) has been used (see the
Examples below) as an excellent first marker to study this phenomenon since it generates glycan structures not normally found in plants. The glycan structures of plants that express galactosyltransferase have been compared with glycan structures from plants that express a chimeric galactosyltransferase of which the CTS domain is exchanged for that of a plant xylosyltransferase (or portion thereof). The change in observed glycan structures show that the galactosyltransferase is, as in mammals, confined to a specific sub-compartment of the plant Golgi. Without limiting the invention to any particular mechanism, the sorting mechanism of plants and mammals are apparently conserved even to the extent that glycosyltransferases unknown to plants are routed to specific analogous location in the Golgi. This location is later in the Golgi than where the endogenous xylosyl-, fucosyl- and GlcNAcTII (GnTII) transferases are located.

The finding that N-glycans in these plants that express relocalised variants of GaIT containing significantly less xylose and fucose is also of biotechnological relevance. For glycoproteins intended for therapeutic use in mammals, such as humans, the approach of certain embodiments of the present invention provides methods and compositions for controlling N-linked glycosylation of glycoproteins in plants so that glycoprotein essentially free of xylose and fucose and containing at least a bi-antennary N-glycans (but not limited to bi-antennary, also include tri-antennary, and the like) and (at least one) galactose residue on at least one of the arms of the N-glycan can be obtained. Hence, it is not intended that the present invention is limited to bi-antennary N-glycans but also includes bisected bi-antennary N-glycans, tri-antennary N-glycans, and the like.

Furthermore, the invention is not limited to complex-type N-glycans but also includes hybrid-type N-glycans and other type N-glycans. The present invention contemplates such resulting glycoproteins. In addition, the methods and compositions of the present invention may be applicable for plants and non-plant systems where besides xylose, fucose, Lewis\textsuperscript{A/BX} type N-glycan modifications (β1-3-GalT, α1-4-FucT, other) or other sugars, "interfere" with desired glycoform accumulation.

In one embodiment, the invention is directed to controlling N-linked glycosylation of plants by modulating the localization of enzymes involved in glycan biosynthesis in the Golgi apparatus. Specifically, embodiments of the invention are directed to a method of producing in a plant host system a glycoprotein having bi-antennary glycans and containing at least one galactose residues on at least one of the arms and which are devoid (or reduced in) of xylose and fucose, comprising:(a) preventing (or inhibiting) addition of xylose and fucose on the core of the glycan of said glycoprotein and (b) adding one or preferably two galactose residues to said arms.

Addition of xylose and fucose to said heterologous glycoprotein may be reduced or even prevented by introducing to said plant host system a nucleic acid encoding a hybrid enzyme comprising a CTS region (or portion thereof) of a protein, particularly an enzyme such as plant xylosyltransferase and catalytic region (or portion thereof) of a galactosyltransferase not normally found in a plant, or a modified galactosyltransferase where its transmembrane portion has been removed and endoplasmic reticulum retention signal have been inserted, wherein said protein or enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said
galactosyltransferase. It is preferred that the galactosyltransferase is a mammalian galactosyltransferase and in particular, a human galactosyltransferase. In a most specific embodiment, said galactosyltransferase is human β1,4 galactosyltransferase (GalT). In a preferred embodiment, said xylosyltransferase is a β1,2-xylosyltransferase. The exchange of the CTS region or CTS fragment of a mammalian glycosyltransferase (such as a galactosyltransferase) by one from the group of enzymes that act earlier in the Golgi apparatus than galactosyltransferase including but not limited to those from of XylT, FucT, GnTI, GnTII, GnTIII, GnTIV, GntV, GntVI, ManI, ManII and ManIII results in strongly reduced amounts of glycans that contain the undesired xylose and fucose residues (see Figure 2). In addition, galactosylation is improved and the diversity in glycans is reduced. While not limited to any particular mechanism, the increase in galactosylated glycans that carry neither xylose nor fucose is believed to be mainly attributed to the accumulation of GalGNMan5, GNMan5 or GalGNMan4. Also, galactosylation occurs on one glycan arm only. Apparently, the galactosylation earlier in the Golgi inhibits trimming of the said glycoforms by Mannosidase II (ManII) to GalGNMan3. Also addition of the second GlcNAc by GlcNAcTII (GnTII) is inhibited.

Therefore, in one embodiment, a further step is contemplated to obtain the desired glycoprotein that has both arms galactosylated and yet is essentially devoid of xylose and fucose. Thus, in one embodiment, the method of the invention as noted above further comprises adding galactose residues to the arms of said glycoprotein (see Figure 3). In one embodiment of the invention, galactose residues are added onto both arms by introducing to said plant host system (a) a nucleic acid sequence encoding a first hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of GnTI and the active domain (or portion thereof) of GnTII; (b) a nucleic acid sequence encoding the second hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane of GnTI and the active domain of ManII and (c) a nucleic acid sequence encoding a third hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of XylT and the active domain (or portion thereof) of human galactosyltransferase (TmXyl-GalT). In another embodiment of the invention, galactose residues are added onto both arms by introducing to said plant host system (a) a nucleic acid sequence encoding a first hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of ManI and the active domain (or portion thereof) of GnTI; (b) a nucleic acid sequence encoding the second hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of ManI and the active domain (or portion thereof) of GnTII; (c) a nucleic acid sequence encoding the third hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of ManI and the active domain (or portion thereof) of ManII, and (d) a nucleic acid sequence encoding a fourth hybrid enzyme comprising the CTS region (or fragment, such as one including the transmembrane domain) of XylT and the active domain (or portion thereof) of human galactosyltransferase (TmXyl-GalT).
It is not intended that the present invention be limited to particular combinations of hybrid enzymes or the number of such hybrid enzymes employed in a single cell, plant tissue or plant. In a preferred embodiment, the present invention contemplates host cells expressing TmXyl-GalT plus TmGnTI-GnTII plus TmGnTI-ManII. In one embodiment of the invention, galactose residues are added to said arms by introducing to said plant host system (a) a nucleic acid sequence encoding a first hybrid enzyme comprising a CTS region (or fragment thereof) of a protein, particularly an enzyme, including but not limited to N-acetylglucosaminyltransferase I (GnTI) and a catalytic region (or portion thereof) of a mannosidase II (ManII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said mannosidase II or modified mannosidase II where its transmembrane portion has been deleted and endoplasmic reticulum retention signal have been inserted and (b) a nucleic acid sequence encoding a second hybrid enzyme comprising a CTS region (or fragment, such as one including the transmembrane domain) of an enzyme including but not limited to N-acetylglucosaminyltransferase I (GnTI) and a catalytic region (or portion thereof) of a N-acetylglucosaminyl-transferase II (GnTII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said N acetylglucosaminyl-transferaseII (GnTII) or modified N-acetylglucosaminyltransferase II (GnTII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted. The sequences encoding N-acetylglucosaminyltransferases or mannosidase II or the said transmembrane fragments can originate from plants or from eukaryotic non-plant organisms (e.g., mammals).

In yet another preferred embodiment, the present invention contemplates a host cell expressing TmXyl-GalT plus TmManI-GnTII plus TmManI-ManII plus TmManI-GnTII. In another embodiment of the invention, galactose residues are added to said arms by introducing to said plant host system (a) a nucleic acid sequence encoding a first hybrid enzyme comprising a CTS region (or fragment, such as one including the transmembrane domain) of a protein, particularly an enzyme, including but not limited to Mannosidase I (ManI) and a catalytic region (or portion thereof) of a N acetylglucosaminyltransferase I (GnTI), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said N-acetylglucosaminyl-transferase I (GnTI) or modified N acetylglucosaminyltransferase I (GnTI) where its transmembrane portion has been deleted and endoplasmic reticulum retention signal have been inserted and (b) a nucleic acid sequence encoding a second hybrid enzyme comprising a CTS region (or fragment, such as one including the transmembrane domain) of an enzyme including but not limited to Mannosidase I (ManI) and a catalytic region (or portion thereof) of a Mannosidase II (ManII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said Mannosidase II (ManII) or modified Mannosidase II (ManII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted and (c) a nucleic acid sequence encoding a third hybrid enzyme comprising a CTS region (or fragment, such as one including the transmembrane domain) of an enzyme including but not limited to Mannosidase I (ManI) and a catalytic region (or portion thereof) of a N-acetylglucos-aminyltransferase II (GnTII), wherein said
enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said N-acetylglucosaminyltransferase II (GnTII) or modified N-acetylglucosaminyltransferase II (GnTIII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted. The sequences encoding N-acetylglucosaminyltransferases or mannosidases or the said transmembrane fragments can originate from plants or from eukaryotic non-plant organisms (e.g., mammals).

In still another preferred embodiment, the present invention contemplates host cells expressing TmXyl-GaIT plus ManIII. In another embodiment of the invention, galactose residues are added to said arms by introducing to said plant host system (a) a nucleic acid sequence encoding a Mannosidase III (ManIII, wildtype gene sequence but not limited to; also ManIII with endoplasmic reticulum retention signal; ManIII with transmembrane fragment of early (cis-) Golgi apparatus glycosyltransferase (GnTI, ManI, GnTIII). The sequences encoding Mannosidase III can originate form insects, preferably from Spodoptera frugiperda or Drosophila melanogaster (but not limited to), human or from other organisms.

In still another preferred embodiment, the present invention contemplates a host cell expressing TmXyl-GaIT plus ManIII plus TmGnTI-GnTII. In yet another preferred embodiment, the present invention contemplates a host cell expressing TmXyl-GaIT plus ManIII plus TmManI-GnTI plus TmManI-GnTII.

The method of the invention may optionally comprise, in one embodiment, introducing into said plant host system a mammalian N-acetylglucosaminyltransferase GnTIII, particularly a human GnTIII or hybrid protein comprising a catalytic portion of mammalian GnTIII and a transmembrane portion of a protein, said protein residing in the ER or earlier compartment of the Golgi apparatus of a eukaryotic cell. For example, in one embodiment, the hybrid enzyme TmXyl-GnTII is contemplated (along with nucleic acid coding for such a hybrid enzyme, vectors containing such nucleic acid, host cells containing such vectors, and plants – or plant parts – containing such host cells). In another embodiment, the hybrid enzyme TmFuc-GnTII is contemplated (along with nucleic acid coding for such a hybrid enzyme, vectors containing such nucleic acid, host cells containing such vectors, and plants – or plant parts – containing such host cells). The present invention specifically contemplates host cells expressing such hybrid enzymes (with or without additional hybrid enzymes or other glycosyltransferases).

The invention is further directed to said hybrid and modified enzymes, nucleic acid sequences encoding said hybrid enzymes, vectors comprising said nucleic acid sequences and methods for obtaining said hybrid enzymes. Furthermore, the invention is directed to a plant host system comprising a heterologous glycoprotein having preferably complex type bi-antennary glycans and containing at least one galactose residue on at least one of the arms and are devoid of xylose and fucose. A “heterologous glycoprotein” is a glycoprotein originating from a species other than the
plant host system. The glycoprotein may include but is not limited to antibodies, hormones, growth factors and growth factor receptors and antigens.

Indeed, the present invention is particularly useful for controlling the glycosylation of heterologous glycoproteins, such as antibodies or antibody fragments (single chain antibodies, Fab fragments, Fab₂ fragments, Fv fragments, and the like). To control the glycosylation of an antibody, the gene construct encoding a hybrid enzyme of the present invention (e.g., the TmXyl-GalT gene construct) can be introduced in transgenic plants expressing an antibody (e.g., monoclonal antibody) or antibody fragment. On the other hand, the gene(s) encoding the antibody (or antibody fragment) can be introduced by retransformation of plant expressing TmXyl-GalT gene construct. In still another embodiment, the binary vector harbouring the TmXyl-GalT expression cassette can be co-transformed to plants together with a plant binary vector harbouring the expression cassettes comprising both light and heavy chain sequences of a monoclonal antibody on a single T-DNA or with binary vectors harbouring the expression cassettes for light and heavy chain sequences both separately on independent T-DNA's but both encoding a monoclonal antibody. The present invention specifically contemplates, in one embodiment, crossing plants expressing antibodies with plant expressing the hybrid glycosyltransferase(s) of the present invention.

A “host system” may include but is not limited to any organism containing glycoproteins with N-glycans.

A “plant host system” may include but is not limited to a plant or portion thereof, which includes but is not limited to a plant cell, plant organ and/or plant tissue. The plant may be a monocotyledon (monocot) which is a flowering plant whose embryos have one cotyledon or seed leaf and includes but is not limited to lilies, grasses, corn (Zea mays), rice, grains including oats, wheat and barley, orchids, irises, onions and palms. Alternatively, the plant may be a dicotyledenon (dicot) which includes but is not limited to tobacco (Nicotiana), tomatoes, potatoes, legumes (e.g., alfalfa and soybeans), roses, daises, cacti, violets and duckweed. The plant may also be a moss which includes but is not limited to Physcomitrella patens.

The invention is further directed to a method for obtaining said plant host system. The method comprises crossing a plant expressing a heterologous glycoprotein with a plant comprising (a) a hybrid enzyme comprising a catalytic region (or portion thereof) of a galactosyltransferase not normally found in a plant and a CTS region (or fragment, such as one including the transmembrane domain) of a protein, wherein said protein acts earlier in the Golgi apparatus of a plant cell in said plant host system than said galactosyltransferase or a modified galactosyltransferase where its transmembrane portion has been deleted and endoplasmic reticulum retention signal has been inserted; (b) a hybrid enzyme comprising a CTS region (or portion thereof, such as one including the transmembrane domain) of a protein, particularly an enzyme, including but not limited to N-acetylglucosaminyltransferase I (GnTI) and a catalytic region (or portion thereof) of a mannosidase II (ManII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said mannosidase II or modified mannosidase II where its transmembrane
portion has been deleted and endoplasmic reticulum retention signal have been inserted and (c) a hybrid enzyme comprising at least a transmembrane region of an enzyme (such as the first 40-60 amino acids of the N-terminus) of a glycosyltransferase including but not limited to N-acetylgalactosaminyltransferase I (GnTI) and a catalytic region of a N-acetylgalactos-aminyltransferase I (GnTI), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said N-acetylgalactosaminyltransferase II (GnTII) and modified N-acetylgalactosaminyltransferase II (GnTII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted., harvesting progeny from said crossing and selecting a desired progeny plant expressing said heterologous glycoprotein.

The invention is further directed to said plant or portion thereof which would constitute a plant host system. Said plant host system may further comprise a mammalian GnTIII enzyme or hybrid protein comprising a catalytic portion of mammalian GnTIII and a transmembrane portion of a protein, said protein residing in the ER or earlier compartment of the Golgi apparatus of a eukaryotic cell.

Additionally, the invention also provides the use of a plant host system to produce a desired glycoprotein or functional fragment thereof. The invention additionally provides a method for obtaining a desired glycoprotein or functional fragment thereof comprising cultivating a plant according to the invention until said plant has reached a harvestable stage, for example when sufficient biomass has grown to allow profitable harvesting, followed by harvesting said plant with established techniques known in the art and fractionating said plant with established techniques known in the art to obtain fractionated plant material and at least partly isolating said glycoprotein from said fractionated plant material.

Alternatively, said plant host cell system comprising said heterologous glycoprotein may also be obtained by introducing into a plant host cell system or portion thereof (a) a nucleic acid sequence encoding a hybrid enzyme comprising a catalytic region of a galactosyltransferase not normally found in a plant and at least the transmembrane region (or more of the CTS) of a protein, wherein said protein acts earlier in the Golgi apparatus of a plant cell in said plant host system than said galactosyltransferase or a modified galactosyltransferase where its transmembrane portion has been deleted and endoplasmic reticulum retention signal have been inserted; (b) a nucleic acid sequence encoding a first hybrid enzyme comprising at least the transmembrane region (or more of the CTS if desired) of a protein, particularly an enzyme, including but not limited to N-acetylgalactosaminyltransferase I (GnTI) and a catalytic region of a mannosidase II (ManII), wherein said enzyme acts earlier in the Golgi apparatus of a plant cell in said plant host system than said mannosidase II, or modified mannosidase II where its transmembrane portion has been deleted and endoplasmic reticulum retention signal have been inserted and (c) a nucleic acid sequence encoding a second hybrid enzyme comprising at least a transmembrane region (more of the CTS if desired) of an enzyme including but not limited to N-acetylgalactosaminyl-transferase I (GnTI) and a catalytic region of a N-acetylgalactosaminyltransferase II (GnTII), wherein said enzyme acts earlier in the
Golgi apparatus of a plant cell in said plant host system than said N-acetylglucos-aminyltransferase-II (GnTII) or modified N-acetylglucosaminyltransferase II (GnTII) where its transmembrane portion has been deleted and an endoplasmic reticulum retention signal have been inserted and isolating a plant or portion thereof expressing said heterologous glycoprotein (or portion thereof). In one embodiment, one vector comprising all of the nucleic acid sequences is introduced into said plant host system. In another embodiment, each nucleic acid sequence is inserted into separate vectors and these vectors are introduced into said plant host system. In another embodiment combinations of two or more nucleic acid sequences are inserted into separate vectors which are then combined into said plant host system by retransformation or co-transformation or by crossing.

The invention also provides use of such a plant-derived glycoprotein or functional fragment thereof according to the invention for the production of a composition, particularly, pharmaceutical composition, for example for the treatment of a patient with an antibody, a hormone, a vaccine antigen, an enzyme, or the like. Such a pharmaceutical composition comprising a glycoprotein or functional fragment thereof is now also provided.

Finally, it is contemplated that the above-described approach may be useful in reducing the overall diversity in glycans in plants expressing one or more of the hybrid enzymes of the present invention (as compared to wild-type plants or plants simply transformed with only mammalian GalT).

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 compares the glycosylation pathway of glycoproteins in plants and in mammals.

Figure 2 shows the effect of exchanging the CTS fragment of galactosyltransferase with xylosyltransferase.

Figure 3 shows the further effect of relocating mannosidase II and GlcNAcTII.

Figure 4 top panel shows a T-DNA construct carrying the genes encoding glycan modifying enzymes to produce efficiently galactosylated glycans that are devoid of immunogenic xylose and fucose and the bottom panel shows a T-DNA construct carrying antibody light chain and heavy chain genes.

Figure 5 shows the nucleic acid sequence (SEQ ID NO:1) for a human galactosyltransferase (human B1,4-galactosyltransferase – GalT).

Figure 6 shows the nucleic acid sequence of Figure 5 along with the corresponding amino acid sequence (SEQ ID NO:2).

Figure 7 shows an illustrative mutated sequence (SEQ ID NO:59) derived the wild type amino acid sequence (SEQ ID NO:2) for a human galactosyltransferase, wherein a serine has been deleted from the cytoplasmic tail and a G-I-Y motif has been repeated. Of course, such changes are merely illustrative of the many possible changes within the scope of the present invention. For example, in one embodiment, the present invention contemplates mutated sequences wherein only deletions (one or more) are employed (e.g. deletions in the cytoplasmic tail domain or the stem
domain) – with no insertions or repeats. Similarly, in one embodiment, the present invention contemplates mutated sequences wherein only (one or more) insertions or replacements (e.g. in the transmembrane domain) are employed – with no deletions.

Figure 8 shows the nucleic acid sequence (SEQ ID NO:3) encoding a hybrid enzyme comprising human galactosyltransferase (human B1,4-galactosyltransferase – GaIT). The upper case letters are nucleotides of *Arabidopsis thaliana* mRNA for beta 1,2-xylosyltransferase (database entry: EMBL:ATH277603, the TmXyl-fragment used involves nucleotides 135-297 of this database sequence).

Figure 9 shows the nucleic acid sequence of Figure 8 along with the corresponding amino acid sequence (SEQ ID NO:4).

Figure 10 shows the amino acid sequence (SEQ ID NO:4) for the hybrid enzyme encoded by the nucleic acid shown in Figure 8.

Figure 11 shows the nucleic acid sequence (SEQ ID NO:5) for the human glycosyltransferase GnTIII (along with additional sequence encoding a myc-tag) (primary accession number Q09327 GnT3 HUMAN).

Figure 12 shows the nucleic acid sequence of Figure 11 along with the corresponding amino acid sequence (SEQ ID NO:6).

Figure 13 shows the amino acid sequence (SEQ ID NO:6) for a human GnTIII (along with additional amino acid sequence of the myc epitope tag SEQ ID NO:7).

Figure 14 shows the nucleic acid sequence (SEQ ID NO:9) encoding one embodiment of a hybrid enzyme of the present invention, said hybrid enzyme comprising the transmembrane domain of a plant xylosyltransferase (TmXyl-) and the catalytic domain (along with other regions) for human GnTIII (TmXyl-GnTIII) (along with additional sequence encoding a myc-tag).

Figure 15 shows the nucleic acid sequence of Figure 14 along with the corresponding amino acid sequence (SEQ ID NO:10).

Figure 16 shows the amino acid sequence (SEQ ID NO:10) for hybrid enzyme encoded by the nucleic acid of Figure 14 (along with additional sequence for the myc epitope tag SEQ ID NO:7).

Figure 17 shows the complete nucleic acid sequence (SEQ ID NO:27) for a cassette encoding the hybrid enzymes TmXyl-GaIT plus TmGnTI-GnTII plus TmGnTI-ManII).

Figure 18 shows the complete nucleic acid sequence (SEQ ID NO:28) for a cassette encoding the hybrid enzyme TmGnTI-ManII (with the RbcS1 promoter sequence SEQ ID NO:39 shown).

Figure 19 shows the nucleic acid sequence (SEQ ID NO:29) encoding the hybrid enzyme TmGnTI-ManII.

Figure 20 shows the nucleic acid sequence (SEQ ID NO:30) encoding the hybrid enzyme TmGnTI-GnTII.

Figure 21 shows the nucleic acid sequence (SEQ ID NO:31) encoding the hybrid enzyme TmGnTI-GnTII, wherein the transmembrane fragment used (designated TmGntI) has the nucleic acid sequence set forth in SEQ ID NO:32.
Figure 22A shows the nucleic acid sequence (SEQ ID NO:32) encoding one embodiment of a transmembrane domain fragment (TmGnTI). Figure 22B shows the nucleic acid sequence (SEQ ID NO:33) encoding another embodiment of a transmembrane domain fragment (TmManI).

Figure 23 shows the complete nucleic acid sequence (SEQ ID NO:34) for a triple cassette embodiment of the present invention.

Figure 24 shows the nucleic acid sequence (SEQ ID NO:35) for a hybrid gene expression cassette (TmManI-GnTI).

Figure 25 shows the nucleic acid sequence (SEQ ID NO:36) for the histone 3.1 promoter.

Figure 26 shows the nucleic acid sequence (SEQ ID NO:37) for the hybrid gene fusion (TmManI-GnTI).

Figure 27 shows the nucleic acid sequence (SEQ ID NO:38) for the hybrid gene fusion TmManI-ManII (with the RbcS1 promoter sequence SEQ ID NO:39 shown).

Figure 28 shows the nucleic acid sequence (SEQ ID NO:39) for the RbcS1 promoter.

Figure 29 shows the nucleic acid sequence (SEQ ID NO:40) for the hybrid gene TmManI-ManII wherein the nucleic acid sequence (SEQ ID NO:33) encoding the transmembrane fragment is shown.

Figure 30 shows the nucleic acid sequence (SEQ ID NO:41) for the hybrid gene TmManI-GnTII.

Figure 31 shows the nucleic acid sequence (SEQ ID NO:42) for the Lhca promoter.

Figure 32 shows the nucleic acid sequence (SEQ ID NO:43) for the hybrid gene TmManI-GnTII wherein the nucleic acid sequence (SEQ ID NO:33) encoding the transmembrane fragment is shown.

Figure 33 shows the nucleic acid sequence (SEQ ID NO:44) for the terminator sequence used (see below).

Figure 34 is a Western Blot which examines total protein glycosylation of plants of the present invention compared to control plants.

Figure 35 is a lectin blot with RCA on F1 progeny of crossed plants, said progeny made according to one embodiment of the present invention.

Figure 36 is a Western Blot. Panel A was assayed with anti-IgG antibody. Panel B was assayed with an anti-HRP antibody. Panel C was assayed with a specific anti-Xyl antibody fraction. Panel D was assayed with a specific anti-Fucose antibody fraction. Panel E was assayed with the lectin RCA.

Figure 37 shows the nucleic acid sequence (SEQ ID NO:49) of a hybrid gene wherein the aminoterminal CTS region of an insect Mannosidase III gene is replaced by a mouse signal peptide and a carboxyterminal endoplasmic reticulum retention signal (KDEL) was added.

Figure 38 shows the corresponding amino acid sequence (SEQ ID NO:50) for the nucleic acid sequence of Figure 37.
Figure 39 shows the nucleic acid sequence (SEQ ID NO:51) of a hybrid gene wherein the aminoterminal CTS region of a human beta-1,4-galactosyltransferase (GalT) gene is replaced by a mouse signal peptide and a carboxyterminal endoplasmic reticulum retention signal (KDEL) was added.

Figure 40 shows the corresponding amino acid sequence (SEQ ID NO:52) for the nucleic acid sequence of Figure 39.

Figure 41 shows the nucleic acid sequence (SEQ ID NO:53) of a hybrid gene wherein the aminoterminal CTS region of an Arabidopsis thaliana GnTII gene is replaced by a mouse signal peptide and a carboxyterminal endoplasmic reticulum retention signal (KDEL) was added.

Figure 42 shows the corresponding amino acid sequence (SEQ ID NO:54) for the nucleic acid sequence of Figure 41.

Figure 43 shows the nucleic acid sequence (SEQ ID NO:55) of a hybrid gene wherein the aminoterminal CTS region of an Arabidopsis thaliana GnTII gene is replaced by a mouse signal peptide and a carboxyterminal endoplasmic reticulum retention signal (KDEL) was added.

Figure 44 shows the corresponding amino acid sequence (SEQ ID NO:56) for the nucleic acid sequence of Figure 43.

Figure 45 shows the nucleic acid sequence (SEQ ID NO:57) of a hybrid gene wherein the aminoterminal CTS region of a human beta-1,4-galactosyltransferase (GalT) gene is replaced by the CTS region of the human gene for GnTII.

Figure 46 shows the corresponding amino acid sequence (SEQ ID NO:58) for the nucleic acid sequence of Figure 45.

Figure 47 is a schematic of how enzymes might be localized to the Golgi.

Figure 48 is a non-limiting speculative schematic of how the “swapping” of regions of transferases might cause relocalization.

**DETAILED DESCRIPTION OF THE INVENTION**

**Hybrid Enzymes**

The nucleic acid sequences encoding the various glycosylation enzymes such as mannosidases, GlcNAcTs, galactosyltransferases may be obtained using various recombinant DNA procedures known in the art, such as polymerase chain reaction (PCR) or screening of expression libraries to detect cloned DNA fragments with shared structural features. See, e.g., Innis *et al.*, 1990, *PCR: A Guide to Methods and Application*, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleic acid sequence-based amplification (NASBA) or long range PCR may be used.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene may be accomplished in a number of ways. For example, if an amount of a portion of a gene or its specific RNA, or a fragment thereof, is available and can be purified and
labeled, the generated DNA fragments may be screened by nucleic acid hybridization to the labeled probe [Benton and Davis, *Science* 196:180 (1977); Grunstein and Hogness, *Proc. Natl. Acad. Sci. U.S.A.* 72:3961 (1975)]. Alternatively, the presence of the gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper mRNAs, can be selected which produce a protein that, e.g., has similar or identical electrophoretic migration, isoelectric focusing behavior, proteolytic digestion maps, or antigenic properties as known for the protein of interest.

A nucleic acid sequence encoding a hybrid enzyme comprising a transmembrane portion of a first enzyme and a catalytic portion of a second enzyme may be obtained as follows. The sequence encoding the transmembrane portion is removed from the second enzyme, leaving a nucleic acid sequence comprising a nucleic acid sequence encoding the C-terminal portion of the second enzyme, which encompasses the catalytic site. The sequence encoding the transmembrane portion of the first enzyme is isolated or obtained via PCR and ligated to the sequence encoding a sequence comprising the C-terminal portion of the second enzyme.

**Modified Enzymes**

A nucleic acid sequence encoding a protein, particularly enzymes such as galactosyltransferases, mannosidases and N-acetylglucosamine transferases that are retained in the ER may be obtained by removing the sequence encoding the transmembrane fragment and substituting it for a methionine (initiation of translation) codon and by inserting between the last codon and the stop codon of galactosyltransferase the nucleic acid sequence encoding an ER retention signal such as the sequence encoding KDEL (amino acid residue sequence: lysine-aspartic acid-glutamic acid-leucine) [Rothman *Cell* 50:521 (1987)].

**Using Domains and Portions Thereof**

As noted above, the phrases "at least a portion of" or a "fragment of" refers to the minimal amino acid sequence necessary for a protein or a peptide to retain its natural or native function. For example, the function of an enzyme could refer to its enzymatic or catalytic role, its ability to anchor a protein in the Golgi apparatus, or as a signal peptide. Thus, the phrases “at least a portion of a transmembrane domain” or “a fragment of a transmembrane domain” each refer to the smallest amino acid segment of a larger transmembrane domain that still retains at least part of the native transmembrane functionality (for example, the function may be evident, albeit decreased). As another example, the phrases “at least a portion of a catalytic region” or “a fragment of a catalytic region” each refer to the smallest amino acid segment of a larger catalytic region that still retains at least part of the native catalytic functionality (again, even if somewhat decreased). As discussed herein, one skilled in the art will know the minimal amino acid segment that is necessary for a protein or a peptide to retain at least some of the functionality of the native protein or peptide.
The glycosyltransferase enzymes are typically grouped into families based on the type of sugar they transfer (galactosyltransferases, sialyltransferases, etc.). Based on amino-acid sequence similarity and the stereochemical course of the reaction, glycosyltransferases can be classified into at least 27 and perhaps as many as 47 different families [Campbell et al., Biochem. J. 326:929-939 (1997), Biochem. J. 329:719 (1998)]. The majority of glycosyltransferases cloned to date are type II transmembrane proteins (i.e., single transmembrane domain with the NH$_2$ terminus in the cytosol and the COOH terminus in the lumen of the Golgi apparatus). Regardless of how they are classified, all glycosyltransferases share some common structural features: a short NH$_2$-terminal cytoplasmic tail, a 16-20 amino acid signal-anchor or transmembrane domain, and an extended stem region which is followed by the large COOH-terminal catalytic domain. The cytoplasmic tail appears to be involved in the specific localization of some types of glycosyltransferases to the Golgi [Milland et al., J. Biol. Chem. 277:10374-10378]. The signal anchor domains can act as both uncleavable signal peptides and as membrane-spanning regions that orient the catalytic domains of the glycosyltransferases within the lumen of the Golgi apparatus.

In one embodiment of the present invention, a portion defined by the N-terminal 77 amino acids of Nicotiana benthamiana (tobacco) acetylglicosaminyltransferase I are contemplated for use in the hybrid enzyme(s), since this portion has been found to be sufficient to target to and to retain a reporter protein in the plant Golgi apparatus [Essl et al., FEBS Lett 453:169-173 (1999)]. Subcellular localization in tobacco of various fusion proteins between the putative cytoplasmic, transmembrane and stem domains revealed that the cytoplasmic-transmembrane domains alone were sufficient to sustain Golgi retention of β 1,2-xylosyltransferase without the contribution of any luminal sequences [Dirrberger et al., Plant Mol. Biol. 50:273-281 (2002)]. Thus, as noted above, certain embodiments of the present invention utilize portions of the CTS region which involve only the cytoplasmic-transmembrane domains (or portions thereof) without utilizing the stem region of the CTS region. However, while some types of glycosyltransferases rely primarily on their transmembrane domain for Golgi retention, other types require their transmembrane region and sequences flanking one or both sides of this region [Colley, Glycobiology 7:1-13 (1997)]. For example, the N-terminal peptide encompassing amino acids 1 to 32 appears to be the minimal targeting signal sufficient to localize β 1,6 N-acetylglicosaminyltransferase to the Golgi. This peptide makes up the cytoplasmic and transmembrane domains of this enzyme [Zerfaoui et al., Glycobiology 12:15-24].

A great deal of information is available on the amino acid sequences of the domains for specific glycosyltransferases. For example, the amino acid sequence of the mammalian galactosyltransferase provided in GenBank Accession No. AAM17731 has the "stem" and "catalytic" domains spanning residues 19 to 147 and residues 148 to 397, respectively [U.S. Patent No. 6,416,988, hereby incorporated by reference] – and the present invention, in certain embodiments, specifically contemplates such portions for use in the hybrid enzyme(s). The amino acid sequence of the rat liver sialyltransferase provided in GenBank Accession No. AAC91156 has a
9-amino acid NH$_2$-terminal cytoplasmic tail, a 17-amino acid signal-anchor domain, and a luminal domain that includes an exposed stem region followed by a 41 kDa catalytic domain [Hudgin et al., Can. J. Biochem. 49:829-837 (1971); U.S. Patent Nos. 5,032,519 and 5,776,772, hereby incorporated by reference]. Known human and mouse β 1,3-galactosyltransferases have a catalytic domain with eight conserved regions [Kolbinger et al., J. Biol. Chem. 273:433-440 (1998); Hennet et al., J. Biol. Chem. 273:58-65 (1998); U.S. Patent No. 5,955,282, hereby incorporated by reference]. For example, the amino acid sequence of mouse UDP-galactose: β-N-acetylgalactosamine β 1,3-galactosyltransferase-I provided in GenBank Accession No. NM020026 has the following catalytic regions: region 1 from residues 78-83; region 2 from residues 93-102; region 3 from residues 116-119; region 4 from residues 147-158; region 5 from residues 172-183; region 6 from residues 203-206; region 7 from amino acid residues 236-246; and region 8 from residues 264-275. [Hennet et al., supra.] – all of which are contemplated in certain embodiments of the present invention as useful portions in the context of the hybrid enzyme(s) discussed above.

While earlier comparisons amongst known cDNA clones of glycosyltransferases had revealed very little sequence homology between the enzymes [Paulson et al., J. Biol. Chem. 264:17615-17618 (1989)], more recent advances have made it possible to deduce conserved domain structures in glycosyltransferases of diverse specificity [Kapitonov et al., Glycobiology 9:961-978 (1999)]. For example, the nucleic acid and amino acid sequences of a number of glycosyltransferases have been identified using sequence data provided by the complete genomic sequences obtained for such diverse organisms as Homo sapiens (humans), Caenorhabditis elegans (soil nematode), Arabidopsis thaliana (thale cress, a mustard) and Oryza sativa (rice).

As a result of extensive studies, common amino acid sequences have been deduced for homologous binding sites of various families of glycosyltransferases. For example, sialyltransferases have sialyl motifs that appear to participate in the recognition of the donor substrate, CMP-sialic acid [Paulson et al., J. Biol. Chem., 264:17615-17618 (1989); Datta et al., J. Biol. Chem., 270:1497-1500 (1995); Katsutoshi, Trends Glycosci. Glycotech. 8:195-215 (1996)]. The hexapeptide RDKKND in Gal α1-3 galactosyltransferase and RDKKNE in GlcNAc β1-4 galactosyltransferase have been suggested as the binding site for UDP-Gal [Joziassse et al., J. Biol. Chem., 260:4941-4951 (1985), J. Biol. Chem., 264:14290-14297 (1989); Joziassse, Glycobiology, 2:271-277 (1992)].

A small, highly-conserved motif formed by two aspartic acid residues (DXD), which is frequently surrounded by a hydrophobic region, has been identified in a large number of different eukaryotic transferases, including α-1, 3-mannosyltransferase, β-1, 4-galactosyltransferases, α-1, 3-galactosyltransferases, glucuronyltransferases, fucosyltransferases, glycogenins and others [Wiggins et al., Proc. Natl. Acad. Sci. U.S.A. 95:7945-7950 (1998)]. Mutation studies indicate that this motif is necessary for enzymatic activity [Busch et al., J. Biol. Chem. 273:19566-19572 (1998); Wang et al., J. Biol. Chem. 277:18568-18573 (2002)]. Multiple peptide alignment showed several motifs corresponding to putative catalytic domains that are conserved throughout all members of the β 3-
galactosyltransferase family, namely, a type II transmembrane domain, a conserved DxD motif, an N-glycosylation site and five conserved cysteines [Gromova et al., Mol. Carcinog. 32:61-72 (2001)].

Through the use of BLAST searches and multiple alignments, the E-X_{\gamma}-E motif was found to be a highly conserved among the members of four families of retaining glycosyltransferases [Cid et al., J. Biol. Chem. 275:33614-33621 (2000)]. The O-linked acetylglucosaminyltransferases (GlcNAc) add a single β-N-acetylglucosamine moiety to specific serine or threonine hydroxyls. BLAST analyses, consensus secondary structure predictions and fold recognition studies indicate that a conserved motif in the second Rossmann domain points to the UDP-GlcNAc donor-binding site [Wrabl et al., J. Mol. Biol. 314:365-374 (2001)]. The β1, 3-glycosyltransferase enzymes identified to date share several conserved regions and conserved cysteine residues, all being located in the putative catalytic domain. Site-directed mutagenesis of the murine β3GalT-I gene (Accession No. AF029790) indicate that the conserved residues W101 and W162 are involved in the binding of the UDP-galactose donor, the residue W315 in the binding of the N-acetylglucosamine-β-p-nitrophenol acceptor, and the domain including E264 appears to participate in the binding of both substrates [Malissard et al., Eur. J. Biochem. 269:233-239 (2002)].

Expression of Proteins of Interest in Plant Host System

The nucleic acid encoding the hybrid or modified enzymes or other heterologous proteins, such as a heterologous glycoprotein may be inserted according to certain embodiments of the present invention into an appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence, or in the case of an RNA viral vector, the necessary elements for replication and translation, as well as selectable markers. These include but are not limited to a promoter region, a signal sequence, 5’ untranslated sequences, initiation codon (depending upon whether or not the structural gene comes equipped with one), and transcription and translation termination sequences. Methods for obtaining such vectors are known in the art (see WO 01/29242 for review).

Promoter sequences suitable for expression in plants are described in the art, e.g., WO 91/198696. These include non-constitutive promoters or constitutive promoters, such as, the nopaline synthetase and octopine synthetase promoters, cauliflower mosaic virus (CaMV) 19S and 35S promoters and the figwort mosaic virus (FMV) 35 promoter (see U.S. Pat. Nos. 5, 352,605 and 6,051,753, both of which are hereby incorporated by reference). Promoters used may also be tissue specific promoters targeted for example to the endosperm, aleurone layer, embryo, pericarp, stem, leaves, tubers, roots, and the like.

A signal sequence allows processing and translocation of a protein where appropriate. The signal can be derived from plants or could be non-plant signal sequences. The signal peptides direct the nascent polypeptide to the endoplasmic reticulum, where the polypeptide subsequently undergoes post-translational modification. Signal peptides can routinely be identified by those of skill in the art. They typically have a tripartite structure, with positively charged amino acids at the N-terminal end,
followed by a hydrophobic region and then the cleavage site within a region of reduced hydrophobicity.

The transcription termination is routinely at the opposite end from the transcription initiation regulatory region. It may be associated with the transcriptional initiation region or from a different gene and may be selected to enhance expression. An example is the NOS terminator from Agrobacterium Ti plasmid and the rice alpha-amylase terminator. Polyadenylation tails may also be added. Examples include but are not limited to Agrobacterium octopine synthetase signal, [Gielen et al., EMBO J. 3:835-846 (1984)] or nopaline synthase of the same species [Depicker et al., Mol. Appl. Genet. 1:561-573 (1982)].

Enhancers may be included to increase and/or maximize transcription of the heterologous protein. These include, but are not limited to peptide export signal sequence, codon usage, introns, polyadenylation, and transcription termination sites (see WO 01/29242). Markers include preferably prokaryote selectable markers. Such markers include resistance toward antibiotics such as ampicillin, tetracycline, kanamycin, and spectinomycin. Specific examples include but are not limited to streptomyces phosphotransferase (spt) gene coding for streptomycin resistance, neomycin phosphotransferase (nptII) gene encoding kanamycin or geneticin resistance, hygromycin phosphotransferase (hpt) gene encoding resistance to hygromycin.

The vectors constructed may be introduced into the plant host system using procedures known in the art (reviewed in WO 01/29242 and WO 01/31045). The vectors may be modified to intermediate plant transformation plasmids that contain a region of homology to an Agrobacterium tumefaciens vector, a T-DNA border region from A. tumefaciens. Alternatively, the vectors used in the methods of the present invention may be Agrobacterium vectors. Methods for introducing the vectors include but are not limited to microinjection, velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface and electroporation. The vector may be introduced into a plant cell, tissue or organ. In a specific embodiment, once the presence of a heterologous gene is ascertained, a plant may be regenerated using procedures known in the art. The presence of desired proteins may be screened using methods known in the art, preferably using screening assays where the biologically active site is detected in such a way as to produce a detectable signal. This signal may be produced directly or indirectly. Examples of such assays include ELISA or a radioimmunoassay.

**Transient Expression**

The present invention specifically contemplates both stable and transient expression of the above-described hybrid enzymes. Techniques for transforming a wide variety of higher plant species for transient expression of an expression cassette are well known [see, for example, Weising et al., Annu. Rev. Genet. 22:421-477(1988)]. Variables of different systems include type nucleic acid transferred (DNA, RNA, plasmid, viral), type of tissue transformed, means of introducing transgene(s), and conditions of transformation. For example, a nucleic acid construct may be
introduced directly into a plant cell using techniques ranging from electroporation, PEG poration, particle bombardment, silicon fiber delivery, microinjection of plant cell protoplasts or embryogenic callus or other plant tissue, or Agrobacterium-mediated transformation [Hiei et al., Plant J. 6:271-282 (1994)]. Because transformation efficiencies are variable, internal standards (e.g., 35S-Luc) are often used to standardize transformation efficiencies.

Expression constructs for transient assays include plasmids and viral vectors. A variety of plant viruses that can be employed as vectors are known in the art and include cauliflower mosaic virus (CaMV), geminivirus, brome mosaic virus, and tobacco mosaic virus.

Plant tissues suitable for transient expression include cultured cells, either intact or as protoplasts (in which the cell wall is removed), cultured tissue, cultured plants, and plant tissue such as leaves.

Some transient expression methods utilize gene transfer into plant cell protoplasts mediated by electroporation or polyethylene glycol (PEG). These methods require the preparation and culture of plant protoplasts, and involve creating pores in the protoplast through which nucleic acid is transferred into the interior of the protoplast.


Other methods allow visualization of transient expression of genes in situ, such as with onion epidermal peels, in which GFP expression in various cellular compartments was observed (Scott et al., (1999) Biotechniques 26(6): 1128-1132

Agrobacterium-mediated transformation is applicable to both dicots and monocots.

Optimized methods and vectors for Agrobacterium-mediated transformation of plants in the family Gramineae, such as rice and maize have been described (see, for example, Heath et al., (1997) Mol. Plant-Microbe Interact. 10:221-227; Hiei et al., (1994) Plant J. 6:271-282 and Ishida et al., (1996) Nat. Biotech. 14:745-750). The efficiency of maize transformation is affected by a variety of factors including the types and stages of tissue infected, the concentration of Agrobacterium, the tissue culture media, the Ti vectors and the maize genotype.

Another useful basic transformation protocol involves a combination of wounding by particle bombardment, followed by use of Agrobacterium for DNA delivery (see, for example, Bidney et al., (1992) Plant Mol. Biol. 18:301-313). Both intact meristem transformation and a split meristem transformation methods are also known (U. S. Patent 6,300,545, hereby incorporated by reference).

Additional methods utilizing Agrobacteria include agroinfection and agroinfiltration. By inserting a viral genome into the T-DNA, Agrobacterium can be used to mediate the viral infection of plants (see, for example, U. S. Patent 6,300,545, hereby incorporated by reference). Following transfer of the T-DNA to the plant cell, excision of the viral genome from the T-DNA (mobilization) is required for successful viral infection. This Agrobacterium-mediated method for introducing a virus into a plant host is known as agroinfection (see, for example, Grimsley, "Agroinfection" pp. 325-342, in Methods in Molecular Biology, vol 44: Agrobacterium Protocols, ed. Garland and Davey, Humana Press, Inc., Totowa, N.J.; and Grimsley (1990) Physiol. Plant. 79:147-153).

The development of plant virus gene vectors for expression of foreign genes in plants provides a means to provide high levels of gene expression within a short time.

Suitable viral replicons include double-stranded DNA from a virus having a double stranded DNA genome or replication intermediate. The excised viral DNA is capable of acting as a replicon or replication intermediate, either independently, or with factors supplied in trans. The viral DNA may or may not encode infectious viral particles and furthermore may contain insertions, deletions, substitutions, rearrangements or other modifications. The viral DNA may contain heterologous DNA, which is any non-viral DNA or DNA from a different virus. For example, the heterologous DNA may comprise an expression cassette for a protein or RNA of interest.

Super binary vectors carrying the vir genes of Agrobacterium strains A281 and A348 are useful for high efficiency transformation of monocots. However, even without the use of high
efficiency vectors, it has been demonstrated that T-DNA is transferred to maize at an efficiency that results in systemic infection by viruses introduced by agroinfection, although tumors are not formed (Grimsley et al., (1989) Mol. Gen. Genet. 217:309-316). This is because integration of the T-DNA containing the viral genome is not required for viral multiplication, since the excised viral genome acts as an independent replicon.

Another Agrobacterium-mediated transient expression assay is based on Agrobacterium-mediated transformation of tobacco leaves in planta (Yang et al., (2000) The Plant J 22(6): 543-551). The method utilizes infiltration of agrobacteria carrying plasmid constructs into tobacco leaves, and is referred to as agroinfiltration; it has been utilized to analyze in vivo expression of promoters and transcription factors in as little as 2-3 days. It also allows examination of effects of external stimuli such as pathogen infections and environmental stresses on promoter activity in situ.

Example 1

An Arabidopsis thaliana cDNA encoding β1,2-xylosyltransferase was isolated from a cDNA library by a previously described PCR based sibling selection procedure [Bakker et al., BBRC 261:829 (1999)]. Xylosyltransferase activity was confirmed by immunostaining of transfected CHO cells with a xylose specific antibody purified from rabbit-antihorseradish-peroxidase antiserum. A DNA fragment covering the N-terminal part of the xylosyltransferase was amplified using primers:

XylTpvuF:ATACTCGAGTAAACAATGAGTAAACCGGAATC (SEQ ID NO:45)
and XylTpvuR:TTCTCGATCGCCGATTGTTATTC (SEQ ID NO:46)

Xhol and HpaI restriction sites were introduced in front of the start codon and a PvuI was introduced at the reverse end. A C-terminal fragment from Human β1,4galactosyltransferase (acc.no. x55415, Aoki 1992) was amplified using primers GalTpvuF:GGCGCGCGATCGGGCGCTTCCTCC (SEQ ID NO:47) and GalTrev:AACCGATCCACGCTAGCTCGGTGTTCCCGAT (SEQ ID NO:48) thus introducing PvuI and BamHI sites. The Xhol/PvuI and PvuI/BamHI digested PCR fragments were ligated in Xhol/BamHI digested pBluescriptSK+ and sequenced. The resulting open reading frame encodes a fusion protein containing the first 54 amino acids of A. thaliana β1,2-xylosyltransferase fused with amino acid 69 to 398 of human β1,4galactosyltransferase and is designated as TmXylGalT. The fragment was cloned into a plant expression vector between the CaMV35S promoter and Nos terminator, using HpaI/BamHI. The clone was introduced into Nicotiana tabacum (samsun NN) as described for native human β1,4galactosyltransferase [Bakker et al., Proc. Nat. Acad. Sci. USA 98:2899 (2001)].

Protein extract of transgenic plants and Western Blots were made as described [Bakker et al., Proc. Nat. Acad. Sci. USA 98:2899 (2001)]. Based on reaction with the lectin RCA, a transgenic plant expressing TmXylGalT was selected for further glycan analysis by MALDI-TOF [Elbers et al., Plant Physiology 126:1314 (2001)] and compared with glycans isolated from plants expressing native β1,4galactosyltransferase and with glycans from wild-type plants. Relative peak areas of the
MALDI-TOF spectrum are given in Table 1. That is to say, Table 1 is a comparison of the results of mass spec (MALDI-TOF) analysis of N-glycans of endogenous glycoproteins of control tobacco ("Tobacco"), transgenic tobacco expressing human beta-1,4-galactosyltransferase ("GalT") and transgenic tobacco plants expressing the beta-1,4-galactosyltransferase gene of which the CTS region has been replaced with that of beta-1,2-xylosyltransferase ("TmXyl-GalT").

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th></th>
<th>Tobacco</th>
<th>GalT</th>
<th>TmXyl-GalT</th>
</tr>
</thead>
<tbody>
<tr>
<td>m/z</td>
<td>Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>933</td>
<td>M3</td>
<td>3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>1065</td>
<td>XM3</td>
<td>10</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>1079</td>
<td>FM3</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1095</td>
<td>M4</td>
<td></td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>1211</td>
<td>FXM3</td>
<td>41</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>1257</td>
<td>M5</td>
<td>4</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>1268</td>
<td>GNXM3</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1298</td>
<td>GalIGNM3</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>1298</td>
<td>GNM4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1414</td>
<td>GNFXM3</td>
<td>27</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>1419</td>
<td>M6</td>
<td>7</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>1460</td>
<td>GalIGNM4</td>
<td></td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>1460</td>
<td>GNM5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1485</td>
<td>GN2FM3</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1576</td>
<td>GalIGNFXM3</td>
<td>8</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>1576</td>
<td>GNFXM4</td>
<td></td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>1581</td>
<td>M7</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1606</td>
<td>GNFM5</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1606</td>
<td>GalIGNFM4</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>1617</td>
<td>GN2FXM3</td>
<td></td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>1622</td>
<td>GalIGNM5</td>
<td></td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>1622</td>
<td>GNM6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1743</td>
<td>M8</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1768</td>
<td>GalIGNFM5</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1768</td>
<td>GNFM6</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1779</td>
<td>GalIGN2FXM3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1905</td>
<td>M9</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1941</td>
<td>Gal2GN2FXM3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>100</td>
<td>100</td>
<td>101</td>
<td></td>
</tr>
</tbody>
</table>

These data show that:

1. In TmXylGalT plants, xylosylation and fucosylation of the glycans is dramatically reduced: 82% of the glycans do not carry xylene nor fucose as compared to 14% in wild-type plants.

2. Galactosylation has increased from 9% in GalT plants to 32% in TmXylGalT plants.
Example 2
A transgenic plant expressing said TmXyl-GalT gene (TmXyl-GalT-12 plant) was selected (above) based on lectin blotting using biotin-labelled RCA (Vector Laboratories, Burlingame, California). Comparison of protein extracts of MGR48 transgenic (control) plant, a selected transgenic plant expressing the unmodified human β1,4-galactosyltransferase gene and TmXyl-GalT-12 plant for the presence of xylose and fucose using anti-HRP (horseradish peroxidase) polyclonal antibody (known for high anti-xylose and anti-fucose reactivity) clearly showed reduced xylose and fucose (Figure 34: “Anti-HRP”). Western blotting using an anti-xylose fraction of the anti-HRP and an anti-fucose fraction (each of which can be prepared by affinity chromatography over the appropriate ligand) showed that especially xylose was reduced compared to control plants (Figure 34: anti-Fuc” and “anti-Xyl”).

Example 3
The TmXyl-GalT-12 plant was crossed with a transgenic plant expressing the monoclonal antibody MGR48 from a single T-DNA integration event (MGR48-31) and which was first made homozygous by selecting offspring plants not segregating for the kanamycin resistance marker and antibody production (MGR48-31-4). Pollen of MGR48-31-4 was used for pollination of emasculated TmXyl-GalT-12 plants. Vice versa, pollen of TmXyl-GalT-12 plant was used for fertilization on emasculated MGR48-31-4 plants. A number of F1 plants were analyzed for the presence of MGR48 by western blotting and for galactosylation of endogenous glycoproteins by lectin blotting using RCA (Figure 35). One plant expressing MGR48 and showing galactosylation of endogenous glycoproteins was selected for further analysis. This plant was identified as XGM8.

Seeds from TmXyl-GalT-12 (♀) x MGR48-31-4 (♂) were sown and F1 offspring plants (XGM) were analyzed for antibody production by Western blotting and for galactosylation by lectin blotting using biotinylated RCA120 (Vector Labs., Burlingame, California) using standard techniques as described before. All plants as expected expressed the monoclonal antibody MGR48 and the majority also had galactosylated glycans as depicted from lectin blotting using RCA120. A single plant expressing both antibody MGR48 and having galactosylated N-glycans was chosen for further analysis (XGM8) (TmXyl-GalT-12 X MGR48-31-4 offspring plant 8). The monoclonal recombinant MGR48 antibody was purified from this plant as described before and submitted to N-glycan analysis by MALDI-TOF.

Briefly, XGM8 plant was grown in greenhouse for antibody production under optimal conditions [Elbers et al., Plant Physiology 126:1314 (2001)]. Protein extract of leaves of transgenic XGM8 plant was made and monoclonal antibody was purified using protein G chromatography as described [Bakker et al, Proc. Nat. Acad. Sci. USA 98:2899 (2001)]. MALDI-TOF of N-glycans of purified monoclonal antibody was as described (Elbers et al., 2001, supra). The presence of galactose on glycans was established by enzyme sequencing using bovine testis β-galactosidase as described (Bakker et al., 2001, supra; Table 2). Table 2 (below) is a comparison of the results of
mass spec (MALDI-TOF) analysis of N-glycans of endogenous glycoproteins (“Xyl-GaIT Endo”) of a F1 hybrid of TmXyl-GaIT-12plant and plant producing rec-mAb (MGR48) and of N-glycans of rec-mAB purified by protein G chromatography from said F1 hybrid.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th></th>
<th>Xyl-GaIT Endo</th>
<th>Xyl-GaIT IgG</th>
</tr>
</thead>
<tbody>
<tr>
<td>m/z</td>
<td>Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>933</td>
<td>M3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>1065</td>
<td>XM3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1079</td>
<td>FM3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1095</td>
<td>M4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>1136</td>
<td>GNM3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1211</td>
<td>FXM3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>1241</td>
<td>FM4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1257</td>
<td>M5</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>1268</td>
<td>GNXM3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1282</td>
<td>GNFM3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1298</td>
<td>GalGNM3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1403</td>
<td>FM5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>1414</td>
<td>GNFXM3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1419</td>
<td>M6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>1430</td>
<td>GNXM4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1430</td>
<td>GalGNXM3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1444</td>
<td>GNFM4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1444</td>
<td>GalGNFM3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1460</td>
<td>GalGNM4</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>1460</td>
<td>GNM5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1471</td>
<td>GN2XM3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1485</td>
<td>GN2FM3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1501</td>
<td>GalGN2M3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1576</td>
<td>GalGNFXM3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1576</td>
<td>GNFXM4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1581</td>
<td>M7</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1593</td>
<td>GalGNXM4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1593</td>
<td>GNXM5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1606</td>
<td>GNFM5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1606</td>
<td>GalGNFM4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1617</td>
<td>GN2FXM3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1622</td>
<td>GalGNM5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>1622</td>
<td>GNM6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1647</td>
<td>GalGN2FM3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1663</td>
<td>Gal2GN2M3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1738</td>
<td>GNFXM5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1738</td>
<td>GalGNFXM4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1743</td>
<td>M8</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1754</td>
<td>GalGNXM5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1768</td>
<td>GalGNFM5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1768</td>
<td>GNFM6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
These data show that:

1. In the F1 hybrid, xylosylation and fucosylation of the glycans is dramatically reduced: 43% of the glycans of endogenous glycoproteins lack xylose and fucose as compared to only 14% in wild-type tobacco plants.

2. The glycans of purified mAb of this F1 hybrid have reduced xylose and fucose, 47% compared to 14% for wildtype tobacco. See also Figure 36, panels B-D.

3. Galactosylation of endogenous glycoproteins of F1 hybrid has increased from 9% in GalT plants to 37% in F1 TmXyl-GalT X MGR48 plant. See also Figure 35.

4. Purified rec-mAb from said F1 (see Figure 36, panel A) shows increased galactosylation; that is to say, 46% has galactose. See also Figure 36, panel E.

It should however be noted that the observed quantities (MALDI-TOF) do not necessarily reflect the molar rations of said glycoforms in vivo. Quantification based on MALDI-TOF can be under- or overestimated depending on the specific glycoform under study. Also, since there is no molecular weight difference between Gal and Man, some peaks can not be annotated unambiguously unless there are clear differences in relative height of specific molecules before and after galactosidase treatment.

**Example 4**

A more direct comparison of xylose, fucose and galactose content was done by examining the MGR48 IgG antibodies from hybridoma, transgenic tobacco and TmXyl-GalT transgenic tobacco. As mentioned above, the TmXyl-GalT-12 plant was crossed with tobacco plant expressing MGR48 IgG (MGR48 tobacco) resulting in an F1 hybrid harbouring MGR48 TmXyl-GalT. An F1 plant was chosen for extraction and purification of MGR48 IgG. Antibodies from said plants (tobacco and TmXyl-GalT) were isolated and purified using protein G chromatography (Elbers et al., 2001. *Plant Physiology* 126: 1314-1322). 300 nanograms amounts of each, hybridoma MGR48 and plant-derived recMGR48, were loaded on precast 12% SDS-PAGE gels (BioRad) and run. The contents of each lane were as follows: Lane 1, MGR48 from hybridoma; Lane 2, purified recMGR48 from normal transgenic tobacco plant; and Lane 3, purified recMGR48 from TmXyl-GalT transgenic plant. Following SDS-PAGE proteins were transferred to nitrocellulose using CAPS buffer. Blots were incubated with A, anti-mouse IgG; B, polyclonal rabbit anti-HRP (anti-xylose/alpha 1,3-fucose); C, anti-xylose; D, anti-(alpha 1,3-) fucose antibodies; and E, biotinylated RCA. Detection
was with LumiLight on Lumi Imager following incubation with HRP-labelled sheep anti-mouse (panel A) or goat-anti-rabbit (panels B-D) antibodies and HRP-labeled streptavidin (E).

Panel A shows that approximately similar amounts of the MGR48 IgG was loaded for all lanes (1-3). L refers to Light chain and H, heavy chain of MGR48 IgG.

Panel B shows that the heavy chain of MGR48 antibody in lane 2 (tobacco) strongly reacts with anti-HRP as expected, whereas the heavy chain of hybridoma derived MGR48 (lane 1) does not (as expected). Hybridoma derived antibodies do not carry xylose and alpha 1, 3-fucoside residues. Remarkably, MGR48 antibodies from TmXyl-GaIT tobacco plant also do not react, suggesting that the heavy chain of antibody from this plant have significantly reduced (perhaps by 90% or more) the amounts of xylose and fucose residues on the N-glycans. This is confirmed by experiments depicted in panels C (anti-xylose) and D (anti-fucose). Panel E shows that the heavy chain of MGR48 antibody of hybridoma (lane 1) has a galactosylated N-glycan, whereas tobacco-derived MGR48 (lane 2) has not, both as expected. Heavy chain of MGR48 from the TmXyl-GaIT plant (lane 3) also has galactosylated N-glycan due to the presence of the construct expressing the hybrid enzyme.

These data are in agreement with the data obtained from similar experiments using total protein extracts from similar plants (tobacco and TmXyl-GaIT-12 plant) as shown previously and confirm that the novel trait introduced in tobacco from expression of TmXyl-GaIT gene can be stably transmitted to offspring and a recombinant monoclonal antibody.

Example 5

Further characterization of the above-described F1 hybrid was performed by treatment with beta-galactosidase. Table 3 is a comparison of the results of mass spec (MALDI-TOF) analysis of N-glycans of rec-mAbs purified by protein G chromatography from an F1 hybrid of TmXyl-GaIT and MGR48 plant before and after treatment of the glycans with beta-galactosidase.

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Xyl-GaIT</th>
<th>Xyl-GaIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>m/z</td>
<td>Type</td>
<td>IgG</td>
</tr>
<tr>
<td>933</td>
<td>M3</td>
<td>4</td>
</tr>
<tr>
<td>1065</td>
<td>XM3</td>
<td>2</td>
</tr>
<tr>
<td>1079</td>
<td>FM3</td>
<td>3</td>
</tr>
<tr>
<td>1095</td>
<td>M4</td>
<td>5</td>
</tr>
<tr>
<td>1136</td>
<td>GNM3</td>
<td>2</td>
</tr>
<tr>
<td>1211</td>
<td>FXM3</td>
<td>3</td>
</tr>
<tr>
<td>1241</td>
<td>FM4</td>
<td>2</td>
</tr>
<tr>
<td>1257</td>
<td>M5</td>
<td>12</td>
</tr>
<tr>
<td>1288</td>
<td>GNXM3</td>
<td>2</td>
</tr>
<tr>
<td>1282</td>
<td>GNFM3</td>
<td>3</td>
</tr>
<tr>
<td>1298</td>
<td>GalGNM3</td>
<td>4</td>
</tr>
<tr>
<td>1403</td>
<td>FM5</td>
<td>3</td>
</tr>
<tr>
<td>1414</td>
<td>GNFXM3</td>
<td>4</td>
</tr>
<tr>
<td>1419</td>
<td>M6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1430</td>
<td>GNXM4</td>
<td>2</td>
</tr>
<tr>
<td>1430</td>
<td>GalGNXM3</td>
<td></td>
</tr>
<tr>
<td>1444</td>
<td>GNFM4</td>
<td>3</td>
</tr>
<tr>
<td>1444</td>
<td>GalGNFM3</td>
<td></td>
</tr>
<tr>
<td>1460</td>
<td>GalGNM4</td>
<td>10</td>
</tr>
<tr>
<td>1460</td>
<td>GNM5</td>
<td></td>
</tr>
<tr>
<td>1471</td>
<td>GN2XM3</td>
<td></td>
</tr>
<tr>
<td>1485</td>
<td>GN2FM3</td>
<td>1</td>
</tr>
<tr>
<td>1501</td>
<td>GalGN2M3</td>
<td>1</td>
</tr>
<tr>
<td>1576</td>
<td>GalGNFM3</td>
<td>3</td>
</tr>
<tr>
<td>1576</td>
<td>GNFXM4</td>
<td></td>
</tr>
<tr>
<td>1581</td>
<td>M7</td>
<td>2</td>
</tr>
<tr>
<td>1593</td>
<td>GalGNXM4</td>
<td>2</td>
</tr>
<tr>
<td>1593</td>
<td>GNXM5</td>
<td></td>
</tr>
<tr>
<td>1606</td>
<td>GNFM5</td>
<td>4</td>
</tr>
<tr>
<td>1606</td>
<td>GalGNFM4</td>
<td></td>
</tr>
<tr>
<td>1617</td>
<td>GN2FXM3</td>
<td>1</td>
</tr>
<tr>
<td>1622</td>
<td>GalGNM5</td>
<td>6</td>
</tr>
<tr>
<td>1622</td>
<td>GNM6</td>
<td></td>
</tr>
<tr>
<td>1647</td>
<td>GalGN2FM3</td>
<td>1</td>
</tr>
<tr>
<td>1663</td>
<td>Gal2GN2M3</td>
<td>1</td>
</tr>
<tr>
<td>1738</td>
<td>GNFXM5</td>
<td>2</td>
</tr>
<tr>
<td>1738</td>
<td>GalGNFXM4</td>
<td></td>
</tr>
<tr>
<td>1743</td>
<td>M8</td>
<td>2</td>
</tr>
<tr>
<td>1754</td>
<td>GalGNXM5</td>
<td>2</td>
</tr>
<tr>
<td>1768</td>
<td>GalGNFM5</td>
<td>3</td>
</tr>
<tr>
<td>1784</td>
<td>GNFM6</td>
<td></td>
</tr>
<tr>
<td>1784</td>
<td>GNM7</td>
<td>1</td>
</tr>
<tr>
<td>1784</td>
<td>GalGNM6</td>
<td></td>
</tr>
<tr>
<td>1809</td>
<td>Gal2GN2FM3</td>
<td>1</td>
</tr>
<tr>
<td>1900</td>
<td>GNFXM6</td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td>GalGNFXM5</td>
<td></td>
</tr>
<tr>
<td>1905</td>
<td>M9</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>102</td>
</tr>
</tbody>
</table>

These data show that:

1. Rec-mAbs from F1 hybrid contain galactose which can be deduced from the observed reduction of specific (galactose-containing) glycoforms after beta-galactosidase treatment and increase of glycoforms lacking galactose. Note the observed reduction of m/z 1622 from 6 to 1% and simultaneous increase of m/z 1460 from 10 to 14% which is the result of the removal of galactose from GalGNM5 to give rise to GNM5. The same is true for m/z 1768 (3 to 1% decrease) and corresponding m/z 1606 peak (4 to 6% increase). See also Figure 36, panel E.

2. Similarly a number of peaks that can be attributed to galactose containing glycans vanish upon treatment with galactosidase, especially m/z 1501, 1647 and 1663 confirming the presence of galactose.
Example 6

In another embodiment, the aminoterminal CTS region of an insect Mannosidase III gene (accession number: AF005034; mistakenly annotated as a Mannosidase II gene) is replaced by a mouse signal peptide coding sequence for import into the endoplasmic reticulum (see Figure 37). The signal peptide sequence encodes a fully active signal peptide normally present at the aminoterminus of IgG sequences and has been used successfully in plants and other organisms before. Furthermore a synthetic sequence coding for a so-called endoplasmic reticulum retention sequence (KDEL) is added to the carboxyterminus of the gene part encoding the catalytic fragment for ER retention. The hybrid Mannosidase III protein encoded by this gene sequence will hence accumulate preferentially in the endoplasmic reticulum.

Example 7

In another embodiment, the aminoterminal CTS region of the human beta-1,4-galactosyltransferase (GaIT) gene (accession A52551) is replaced by a mouse signal peptide coding sequence for import into the endoplasmic reticulum (see Figure 39). The signal peptide sequence encodes a fully active signal peptide normally present at the aminoterminus of IgG sequences and has been used successfully in plants and other organisms before. Furthermore a synthetic sequence coding for a so-called endoplasmic reticulum retention sequence (KDEL) is added to the carboxyterminus of the gene part encoding the catalytic fragment for ER retention. The hybrid beta-1,4-galactosyl-transferase protein encoded by this gene sequence will hence accumulate preferentially in the endoplasmic reticulum.

Example 8

In another embodiment, the aminoterminal CTS region of Arabidopsis thaliana GnTI (acc. AJ243198) is replaced by a mouse signal peptide coding sequence for import into the endoplasmic reticulum (see Figure 41). The signal peptide sequence encodes a fully active signal peptide normally present at the aminoterminus of IgG sequences and has been used successfully in plants and other organisms before. Furthermore a synthetic sequence coding for a so-called endoplasmic reticulum retention sequence (KDEL) is added to the carboxyterminus of the gene part encoding the catalytic fragment for ER retention. The hybrid GnTI protein encoded by this gene sequence will hence accumulate preferentially in the endoplasmic reticulum.

Example 9

In another embodiment, the aminoterminal CTS region of an Arabidopsis thaliana GnTII (acc. AJ249274) is replaced by a mouse signal peptide coding sequence for import into the endoplasmic reticulum (see Figure 43). The signal peptide sequence encodes a fully active signal peptide normally present at the aminoterminus of IgG sequences and has been used successfully in
plants and other organisms before. Furthermore a synthetic sequence coding for a so-called endoplasmic reticulum retention sequence (KDEL) is added to the carboxyterminus of the gene part encoding the catalytic fragment for ER retention. The hybrid GnTII protein encoded by this gene sequence will hence accumulate preferentially in the endoplasmic reticulum.

Example 10

In another embodiment, the aminoterminal CTS region of the human gene for beta-1,4-galactosyltransferase (GalT) gene is replaced by the CTS region of the human gene for GnTI (TmhuGnTI-GaLT) (see Figure 45).

It is understood that the present invention is not limited to any particular mechanism. Nor is it necessary to understand the mechanism in order to successfully use the various embodiments of the invention. Nonetheless, it is believed that there is a sequential distribution of Golgi enzymes (Figure 47) and that the swapping in of transmembrane domains of plant glycosyltransferases causes relocation (Figure 48).

It is understood that the present invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described herein, as these may vary. It is also to be understood that the terminology used herein is used for the purpose of describing particular embodiments only, and is not intend to limit the scope of the present invention. It must be noted that as used herein and in the appended claims, the singular forms “a”, “an”, and “the” include plural reference unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein disclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, 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. Such modifications are also intended to fall within the scope of the appended claims.

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.
WHAT IS CLAIMED IS:

1. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a plant glycosyltransferase and a catalytic region of a mammalian glycosyltransferase.

2. The nucleic acid of Claim 1, wherein said plant glycosyltransferase is a xylosyltransferase.

3. The nucleic acid of Claim 1, wherein said plant glycosyltransferase is a N-acetylglucosaminyltransferase.

4. The nucleic acid of Claim 1, wherein said plant glycosyltransferase is a fucosyltransferase.

5. The nucleic acid of Claim 1, wherein said mammalian glycosyltransferase is a human galactosyltransferase.

6. The nucleic acid of Claim 5, wherein said human galactosyltransferase is encoded by at least a portion of the nucleic acid sequence of SEQ ID NO:1.

7. An expression vector, comprising the nucleic acid of Claim 1.

8. A host cell transfected with the vector of Claim 7.

9. The host cell of Claim 8, wherein said host cell is a plant cell.

10. A cell suspension comprising the host cell of Claim 9.

11. The hybrid enzyme expressed by the plant cell of Claim 9.

12. The plant comprising the host cell of Claim 9.

13. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a first glycosyltransferase and a catalytic region of a second glycosyltransferase.

14. The nucleic acid of Claim 13, wherein said first glycosyltransferase comprises a plant glycosyltransferase.

15. The nucleic acid of Claim 14, wherein said plant glycosyltransferase is a xylosyltransferase.

16. The nucleic acid of Claim 14, wherein said plant glycosyltransferase is a fucosyltransferase.
17. The nucleic acid of Claim 13, wherein said second glycosyltransferase comprises a mammalian glycosyltransferase.

18. The nucleic acid of Claim 17, wherein said mammalian glycosyltransferase is a human galactosyltransferase.

19. The nucleic acid of Claim 13, wherein said first glycosyltransferase comprises a first mammalian glycosyltransferase and said second glycosyltransferase comprises a second mammalian glycosyltransferase.

20. The nucleic acid of Claim 19, wherein said first mammalian glycosyltransferase is a non-human glycosyltransferase.

21. The nucleic acid of Claim 19, wherein said second mammalian glycosyltransferase is a human glycosyltransferase.

22. A method, comprising:
   a. providing: i) a plant cell, and ii) an expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a plant glycosyltransferase and a catalytic region of a mammalian glycosyltransferase; and
   b. introducing said expression vector into said plant cell under conditions such that said hybrid enzyme is expressed.

23. The method of Claim 22, wherein said plant glycosyltransferase is a xylosyltransferase.

24. The method of Claim 23, wherein said plant glycosyltransferase is a N-acetylglucosaminyltransferase.

25. The method of Claim 23, wherein said plant glycosyltransferase is a fucosyltransferase.

26. The method of Claim 22, wherein said mammalian glycosyltransferase is a human galactosyltransferase.

27. The nucleic acid of Claim 26, wherein said human galactosyltransferase is encoded by at least a portion of the nucleic acid sequence of SEQ ID NO:1.

28. A method, comprising:
a. providing: i) a plant cell, ii) a first expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a plant glycosyltransferase and a catalytic region of a mammalian glycosyltransferase, and iii) a second expression vector comprising nucleic acid encoding a heterologous glycoprotein; and

b. introducing said first and second expression vectors into said plant cell under conditions such that said hybrid enzyme and said heterologous protein are expressed.

29. The method of Claim 28, wherein said heterologous protein is an antibody or antibody fragment.

30. A method, comprising:

a) providing: i) a first plant comprising a first expression vector, said first vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a transmembrane region of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian glycosyltransferase, and ii) a second plant comprising a second expression vector, said second vector comprising nucleic acid encoding a heterologous protein; and

b) crossing said first plant and said second plant to produce progeny expressing said hybrid enzyme and said heterologous protein.

31. A plant, comprising first and second expression vectors, said first vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a transmembrane region of a plant glycosyltransferase and at least a portion of a catalytic region of a mammalian glycosyltransferase, said second vector comprising nucleic acid encoding a heterologous protein.

32. The plant of Claim 31, wherein said heterologous protein displays reduced amounts of fucose as compared to when the heterologous protein is expressed in a plant in the absence of said hybrid enzyme.

33. The plant of Claim 31, wherein the heterologous protein displays reduced amounts of xylose as compared to when the heterologous protein is expressed in a plant in the absence of said hybrid enzyme.

34. The plant of Claim 31, wherein the heterologous protein displays both reduced fucose and xylose, as compared to when the heterologous protein is expressed in a plant in the absence of said hybrid enzyme.
35. The plant of Claim 31, wherein the heterologous protein displays complex type bi-antennary glycans and contains galactose residues on at least one of the arms.

36. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a modified mammalian glycosyltransferase, wherein a transmembrane portion has been deleted and endoplasmic reticulum retention signal have been inserted.

37. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a CTS region or portion thereof of a plant glycosyltransferase and a catalytic region of a mammalian glycosyltransferase, wherein said CTS region is from a N-acetylgalactosaminyltransferase I (GnTI) and said catalytic region is from a mannosidase II (ManII).

38. Nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a CTS region or portion thereof of a plant glycosyltransferase and a catalytic region of a mammalian glycosyltransferase, wherein said CTS region or portion thereof is from a N-acetylgalactosaminyltransferase I (GnTI) and said catalytic region is from a N-acetylgalactosaminyltransferase II (GnTII).

39. A plant host system, comprising (a) a nucleic acid sequence encoding a Mannosidase III glycosyltransferase; (b) a nucleic acid sequence encoding a hybrid enzyme, said hybrid enzyme comprising a CTS region or portion thereof of a plant glycosyltransferase and a catalytic domain of a mammalian glycosyltransferase.

40. A method, comprising (a) introducing into said plant host system a vector comprising (i) a nucleic acid sequence encoding a hybrid enzyme comprising a catalytic region of a galactosyltransferase not normally found in a plant and a transmembrane region of a protein, (ii) a nucleic acid sequence encoding a hybrid enzyme comprising a transmembrane region of a N-acetylgalactosaminyltransferase I (GnTI) and a catalytic region of a mannosidase II (ManII), (iii) a nucleic acid sequence encoding a hybrid enzyme comprising a transmembrane region of an N-acetylgalactosaminyltransferase I (GnTI) and a catalytic region of a N-acetylgalactosaminyltransferase II (GnTII); and (b) isolating a plant or portion thereof expressing said nucleic acid sequences.

41. A method, comprising:
   a. providing: i) a host cell, and ii) an expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a first glycosyltransferase and a catalytic region of a second glycosyltransferase; and
   b. introducing said expression vector into said host cell under conditions such that said hybrid enzyme is expressed.
42. The method of Claim 41, wherein said first glycosyltransferase comprises a plant glycosyltransferase.

43. The method of Claim 42, wherein said plant glycosyltransferase is a xylosyltransferase.

44. The method of Claim 42, wherein said plant glycosyltransferase is a N-acetylglucosaminyltransferase.

45. The method of Claim 42, wherein said plant glycosyltransferase is a fucosyltransferase.

46. The method of Claim 41, wherein said second glycosyltransferase comprises a mammalian glycosyltransferase.

47. The method of Claim 46, wherein said mammalian glycosyltransferase is a human galactosyltransferase.

48. A method, comprising:
   a. providing: i) a host cell, ii) a first expression vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising a transmembrane region of a first glycosyltransferase and a catalytic region of a second glycosyltransferase, and iii) a second expression vector comprising nucleic acid encoding a heterologous glycoprotein; and
   b. introducing said first and second expression vectors into said host cell under conditions such that said hybrid enzyme and said heterologous protein are expressed.

49. The method of Claim 48, wherein said heterologous protein is an antibody or antibody fragment.

50. The method of Claim 48, further comprising the step of c) isolating said heterologous protein.

51. The isolated heterologous protein produced according to the method of Claim 50.

52. A host cell, comprising first and second expression vectors, said first vector comprising nucleic acid encoding a hybrid enzyme, said hybrid enzyme comprising at least a portion of a transmembrane region of a first glycosyltransferase and at least a portion of a catalytic region of a second glycosyltransferase, said second vector comprising nucleic acid encoding a heterologous protein.

53. The heterologous protein isolated from the host cell of Claim 52.
FIG. 5
FIG. 6
FIG. 7

ATGAGTAAACGGAATCCGAAGATTTCTCAGAGATTTTCTTGTATATGTATTACCTCTCATCTCCCCATCATCATCTCCCCCCTTTTCTTCTCCACCTCAGCACACTCAGCCACAGCTCATCTACATCTACTCACTCAGCTCAGCTTCTTCTCTCATATATACCAACATCCGCGAATgggcatcetccg
gggacgtccgacacgggagccgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacgggacggg
FIG. 11
FIG. 12
FIG. 16
FIG. 17
FIG. 17 Cont.
FIG. 17 Cont.
FIG. 18
FIG. 18 Cont.
FIG. 21
FIG. 22A

FIG. 22B
AGCTTCCCGGTGTGACCCTGAATGATTGTAAGAAACAAGGGTGTAGGCGAACGGGCATTGTAAGGTAATCTCTG
ATCAGTATGGGAAATCATCGGTTCGCCGAAGATTGATCTTTGAAAGCATCACTGTTGATGGATGAACACTGTATG
GGATGGGTGGAGAGACTAAGGACATGGGGGACATCCTTTTCATGGAAGAAGCATCATTTCCTGTTTCTAAAT
GCATATCGTAAACATACAGACTCTTACGAGGCTGAAAACCGGAAAGTGTCCTGACTGTATTGCTGTAATTAGCAC
CGTCTGATGTAAGTCAAGAGAGAAGGCTTGAAGATTTGGTTGTCAGAGAGATGGGAATGTTGGGTATTCTTT
TAATAGAGAGTGTGTGGAGAGATATTCATCAAGGCAAGAAGCTCTTGTATGATTCAACTGGGATATA
ACGATGTGGGCAACGGTCTTTTGCTCCGTCTGGTTGCCCGGTGTACACATTCGGCGAGCTAGACTAGTCGGTAC
ACTTTGGAAATGTTGGTTCATCAAGAGTAGAGGAGATGAGGGTATGGCATGATAATGGGGTGCATAAACATAGA
AGTTAAGGAAAACAGATAAAAGTTGTGAACATAAAAGAAGGATGGGGAGTCCGGGTGTATAAGCATCAAAGCCTGGTTAT
AAAGCCGTTTCTGAAGGGGAGTTGGGCGATAGGAGGCCGACATTTATGGATTTGCTCAGACTATGT
ATCGTTACAGCAATGACAGTACCTCAGTGAAACGGATCCGCTAGAGTCCGCAAAAATACCCAGTCTCTCTCAC
AAATCTACTCTCCTCTATTTTTTCTCCAATTATGTCGAGTATTCCGATATAAGGGAATTAGGGTTTTTTATAGG
GTTCGCTCATGTGTGGAGCATATAAGAAACCTTATAGTATTTGTATTTGTAGAAATACCTTATCAATAAAAT
TTCTAATCTAAAACCAAAATCCCGCGAGAGACCTTAAATTTA

FIG. 23 Cont.
FIG. 25

GGATCCGATATAACAAAAATTTGAAATCGCAGATCGATCTCTTTTGAGATTCTATACCTAGAAAATGGAGAGCATT
TTCAAAATCTCTGTAAAAATTTCTGTTTCTTTCTTGAGCAAGAGAACCCGACTCAAATTTCCGTTAGTACTGAA
CCGGAAGTGTTGACTGTTGCAACCAATTTAATGTACCGTACGTAACCACCAAATCGAGATTGTATTTGATATCAATGGGCC
TTATCTGTGAGCCCATTAATTGTGACGCGCCTAAAAACTAAATCCGAAACGGTTTATTTCCAGCGATCCGCGACGGTT
TGATTTCAATAGCAATCAATTATGGACAGTGTTGATCCTCGTCAAAACCAGTAAGCAGATCTGACCGTT
GAATTGCTGAAAGAAACACTGTTGTGATATTGTAACCGTACGATAGAATTTGAGAAACACATGGATAATC
GATAAAAACGGTCGATCATATAATCCGCTTTTCCATCGTTGCTATAAAATTATATCAGAAATGGCTACACGCGT
GAAAGACTGAAATATTATTTCTTTCGAAATCGGAGCTCAAGGTTGAAAATTCGGAAGACAGCTAGAGAGTTTCTGATACCATGG
CCATGGCGAGAGGGACAGATCGTGGTGACAGCAAGCAAATGGAGGTACTGCAACCCTCTCTATTACTTTGAA
GCAGCCAAGGCTCTTGCTCTGCTCTCAGTTTGTTTGGTTTGCTTCTCTTTTCTGGACCGTCAAACCTCCT
GTCAGAGAGACCAGGTATGAAATTCTGAGCTGCAAGAAGAAATGACTGATTGAAAAATTGGGAAAGATGAT
ATATAAACAAAGGCTGTAACCTCTGGGAAAACGACTTGGGACGACGATGCTCTGCTGCTGCTGAGAGT
TATGGGCTGCAGTGCTGACAGACTCTTGGAAAAAGGACTGTTAACATCATCAACACTCCTCGTGCTTCA
AAATATCCTCTATTTTATATCACGAGATGGATCTGATCAAGGACAAGGACTCTACGATAGCTATAATCAA
CTATATAGCAAGCATTGGAGTTTGGAAACAGTGGTGAATGGAATAGCCTGGAGCCGACTGACCTGACTGAC
AGTCTACATAAGGGCAGCTGGAGCTGTTGTGTTACAAACAGAAATTTAGTGAGTGATATATACAGGAAGATG
ATAGGAAATTGCTCCAGACTCTCTTTGATTACCTGTGGAGCTGCTGCAAGGATAGCTATGCTAGGATAAACC
ATTATGCTGCTTCATCAATGAAATGACGACAGAGAGTTTTGTGCAATGATCTCTATTGCTGACTGATAGAAT
TCCTGGCTCTTTGCTGAGTCTCAAGAGATCGGACTTGGGAGGTTATCAACAAAGTGCCAAAGGCTTACCTG
GATGATTGCTGAGACTAAAGGAAACCATAAAAGGCCGCAATTCTCGACCCGGAAGTCTGAGAACATAACAATT
TGGAAACATGTGCTAGTTTGGGACGAGTTTCTACGTACCTGCTGAGGACCTAAAGCTAAACGATGACGCTTGG
AGTGAAAGACAAAGGACTGCGGATACCGGACAGGGAAACTATACCAAGTACTTTTCTGCTTGTAGGACAGAAC
CGAGCCAATACGTGCTTCGTCAGCCTCTTTAAGAGCTCAAACATAAGGATAGTGTCTGATCCGGTGTAAAAGAC
CAGTAGAGTTGGAACAGCATGGCAGGGTAATTTGGTATATATGGAAGAAATGGAAGGATGTGTGGCTGCGAACAC
GATAAAGGAGATGTGCTTCCGAATCCGACAACAAGACGTGATATCTCTGGTGCGCGAGATTCTGTAATGCGACG
CCGGATTCCGAAAATCTGATGCGGATCC

FIG. 26
FIG. 27 Cont.
FIG. 28

AGATCTAATCTAAACCAATTACGATACGCTTGGGTACACTTGAGTTTCATATATATCTTCTTT
ATATGCTATCTTTAAGGATCTGACAAAGATTATTTGTTGATGTTCTTGATGGGCTCAGAAGATTGTATAGATA
CACTCTAACTTTAGGAGATACCACACGAGTTATATTTCAGTAAAGACATCAAATTTAACGTTACACTCGTTA
TCTTTTCATTCAAAGGATGAGCCAGAATCTTTATAGAATGATTGCAATCAGAATATGTTCGCCAGATATGCTTT
GGTTGGCTTTAATATTCTACATATACACACACAGAAAATCGACCGATATTGTAACCTCTTTCTTCCAATAAGGAAAACACATAT
GCAAGATGCTTTTTTCCACATGCAATACATATAGGTATTTCAAAAATGGCTAAAGAAGTTGGAATAACAAATTTGAC
AACTATTTCCATTCTGTTATATAATTTTCACACACACAAAAAGCCCGTAATCAAGAGTCTGCCCATGTACGAAT
AACTTCTATTATTTGGTTATTTGGGCTAAGCCAGCTACAGTAGCGTGGGGGATCCACATATAGGAAGGTAAACAA
TAAGCAAGATAGCCCCCATAAACGTACACGCTCCTTACACGGGAGAGATAAGATATAAGACCCACCCCTGGCGACG
TGTCACATCGTCATGGTGTGTAAATGATAAGGATACATTTCTCTATGTTTGTGACATGATGCATGTAATGTCAAT
GAGCCACAGGATCACAATGGGCACAGGAAACGGTAAGAATGAGATAGATGATTGATTGTGTGGTTCCGTTAGTAGAAGAAACAC
ATTATAAAAGGTTGGTATCTAAATAGGAACATATAACACTCTATTTGGATTCAATAGGAATTCACACCTTCTATTCTTAGTATCTAG
AAACCATGG
FIG. 30 Cont.

GTAGAGGAGATGAGGTTGATGATCGATAATGGGGTCGTAACATAGAAGTTAAGGAACAGATAAAGTGTGAA
CATAAAAGAAGGATGGGAGTTGATGTGTAAGACATCAAGCGGGTTTACACGCAGTACGCAGTGCATCTC
CATGAAACGGATCCGCTAGAGTCCGCAAAAATCAACCAGTCTCTCTCTACAAAATCTATCTCTCTCTATTTTCTCCA
GAATAATGTGTGAGTGTAGTGCTCTAGATAAGGGAATTAGGGTTTCATGTCATGTGGTGACATATAAG
AAACCCTTAGTGATATTTGTATTTGTATTTGAAATAGCTACTCTACACAAAAATTTCTACATAAAAACCGCGAGAGACCTCTTAATTA
Lanes: 25μg total protein each
1, MGR48-plantibody control
2, TmXyl-GalT
3, GalT
4, SSNN control

FIG. 34
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anti-IgG</td>
<td>Anti-HRP (anti-Xyl/Fuc)</td>
<td>Anti-Xyl</td>
<td>Anti-Fuc</td>
<td>RCA *</td>
</tr>
<tr>
<td>1</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
</tr>
<tr>
<td>2</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

Lanes 1, 2 & 3: 300 ng MGR48 IgG
1, MGR48 hybridoma
2, MGR48 tobacco
3, MGR48 TmXylH-GaT

FIG. 36
FIG. 37 Cont.
FIG. 39

FIG. 40
FIG. 45

FIG. 46
Sequential distribution of Golgi enzymes

FIG. 47
SEQUENCE LISTING

<110> PLANT RESEARCH INTERNATIONAL BV
       BAKKER, Hendrikus A.C.
       FLORACK, Dionisius E.A.
       BOSCH, Hendrik J.
       ROUWENDAAL, Gerard J.A.

<120> Optimizing glycan processing in plants

<150> US-60/365,735

<160> 59

<170> PatentIn version 3.2

<210> 1
<211> DNA
<212> Homo sapiens
<400> 1
atgaggcttc gggagccgctc cctgagcggc aagcggccgca tgccaggcgct tggccctacag  60
cggaacctgc cctctgcctgt ggcgtctgcgt gctctgcacc ttggcccagct cctgttttaa  120
tacctgtagtg gccggccgccct gacccggctgt ccccaaccgg tgccagtgctc cacccggctg  180
caggccgctgc gggagccgctc cctgacctgc aggccgctgc cggggagggct cggaggccccg  240
gggcgcgcgc gcggccgcctc ttctgagcgcg tgcgcgcgcgc gcgggccccg gcgggctgc  300
ggagagccgct cccggccgcctc ctgctgagcgc gggaggccccg tggccgagctc tggagccccg  360
cacaccaccc caccctgaggc cacaggccacct cctggagccct cggcgggtcgt cgggggccccg  420
atgctgtatgg aagtttataat ggcctgctggc cctgctgagc ggcgagcgag ccaggggatc  480
gtgagagagtc gcgccccgctca cggccggtctt ccttccacctc ggtggcccttc gtcctctatc  540
atcattccat tccgcaaccc gcaggagccct cttccgctcct gctacatata ggtgctatccct  600
gtcctctagt gcgcagccgct gacatgggtc atctctatgt ttcaaccagc gcctgggcgtgc  660
tagttcatta gtcgctcatt ccctaatcag gccggtcctgc ggcctgggctt gatcggggctag  720
tacattccat ccctggctctg ctttccagtgc ctctgggggat gatcggcgcc ctctgggatlct  780
agggctcttt ggcagccgctc gcagttgaca atatatatat ataatgttgg agccaagctgg  840
tatatagtgc aatataataa tttgggggctg gcagggagag atagtagcttt ttttttttttt  900
ttagctttttt gaggagatgggc taagctgtcttt cctgctctgata cagatcctgt ctttaaaccct  960
taattgttttt ttttttttttt gaggagagat tggatatagcc tgcgggtttt ttttttttttt 1020
atccgccccat ccaacccgctg ggccttgtcag ttgctggagc ctctgggctc cctctgggctc 1080
ccacagaaaa aagacatgggt cctctgagtg tgaggctgcc tacaactctcc ttttttttttt 1140
gttatgggc taacaggtgcc ttgagagagtc gaagctgtgt gatagagagat gctgctgctg 1200

<210> 2
<211> PRT
<212> Homo sapiens
<400> 2
Met Arg Leu Arg Glu Pro Leu Leu Ser Gly Ser Ala Ala Ala Met Pro Gly
      1 5 10 15
Ala Ser Leu Gln Arg Ala Cys Arg Leu Leu Val Ala Val Cys Ala Leu
      20 25 30
His Leu Gly Val Thr Leu Val Tyr Tyr Leu Ala Gly Arg Asp Leu Ser
      35 40 45
Arg Leu Pro Gln Leu Val Gly Val Ser Thr Pro Leu Gln Gly Gly Ser
      50 55 60
Asn Ser Ala Ala Ala Ile Gly Gin Ser Ser Gly Glu Leu Arg Thr Gly
65    70    75    80
Gly Ala Arg Pro Pro Pro Leu Gly Ala Ser Ser Gln Pro Arg Pro
   85
Gly Gly Asp Ser Ser Pro Val Val Asp Ser Gly Pro Gly Pro Ala Ser
100   105  110
Asn Leu Thr Ser Val Pro Val Pro His Thr Thr Ala Leu Ser Leu Pro
115  120  125
 Ala Cys Pro Glu Glu Ser Pro Leu Leu Val Gly Pro Met Leu Ile Glu
130   135  140
 Phe Asn Met Pro Val Asp Leu Glu Leu Val Ala Lys Gln Asn Pro Asn
145  150  155  160
 Val Lys Met Gly Gly Arg Tyr Ala Pro Arg Asp Cys Val Ser Pro His
165  170
 Lys Val Ala Ile Ile Ile Pro Phe Arg Asn Arg Gln Glu His Leu Lys
175   180  185  190
 Tyr Trp Leu Tyr Tyr Leu His Pro Val Leu Gln Arg Gln Gln Leu Asp
195  200  205
 Tyr Gly Ile Tyr Val Ile Asn Gln Ala Gly Asp Thr Ile Phe Asn Arg
210  215  220
 Ala Lys Leu Leu Asn Val Gly Phe Gln Glu Ala Leu Lys Asp Tyr Asp
225  230  235  240
 Tyr Thr Cys Phe Val Phe Ser Asp Val Leu Ile Pro Met Asn Asp
245  250  255
 His Asn Ala Tyr Arg Cys Phe Ser Gln Pro Arg His Ile Ser Val Ala
260  265  270
 Met Asp Lys Phe Gly Phe Ser Leu Pro Tyr Val Gln Tyr Phe Gly Gly
275  280  285
 Val Ser Ala Leu Ser Lys Gln Gln Phe Leu Thr Ile Asn Gly Phe Pro
290  295  300
 Asn Asn Tyr Trp Gly Trp Gly Gly Glu Asp Asp Ile Phe Asn Arg
305  310  315  320
 Leu Val Phe Arg Gly Met Ser Ile Ser Arg Pro Asn Ala Val Val Gly
325  330  335
 Arg Cys Arg Met Ile Arg His Ser Arg Asp Lys Asn Gln Pro Asn
340  345  350
 Pro Gln Arg Phe Asp Arg Ile Ala His Thr Lys Glu Thr Met Leu Ser
355  360  365
 Asp Gly Leu Asn Ser Leu Thr Tyr Gln Val Leu Asp Val Gln Arg Tyr
370  375  380
 Pro Leu Tyr Thr Gln Ile Thr Val Asp Ile Gly Thr Pro Ser
385  390  395

<210> 3
<211> 1152
<212> DNA
<213> hybrid
<400> 3
atgagtaaac ggaatccgaa gattctgaaag atttttctgt atatgttact tctcaacctc 60
tctttcttca tcctctactc ctgcttttac cttctctcttc gatcctcag 120
cctcctcata tatccacggt tacagttgaat aaccaatcgg gatcggccgg 180
gagctgccgg ccggagaggg ccggccggccg ccctccctag gcgcctccct ccagccgcccgc 240
cgcggggtgct acttccagcct actcctgctg tgtgacccctg ccgccccgctag caaccttgacc 300
tcgttgcccag tggccccacat caccgcactcg cctctggcccg ctctgctgtaa ggatcccccgg 360
tcgtgctgct gcctggctgct gacagcgcgctg tcgctgctgca gctccttgccga 420
aagccagacc caaattgtgaa gattggccggc cgtctagccc ccagggactg cgctctcttc 480
caacaggtgg cacatctacat cttccttgccg aaccggcagg gagccctccaa gtactgctgtaa 540
tattatgcag acaccgtcctc gcagcgccaga cagctctgacct atgtgcataa tttatcacac 600
caggccggg acacatctatt cacagctctctg aagctctctca atgcttgcctt tcaagagcc 660
ttgagggact atgcaacctac cttgctctctg tttgagtcgcg tggacccctc tccaaatgaaa 720
gaccataatacg cgtgaggtt gttttccacag ccacggccaca ttcgcttggtc aatggataag 780
tttggattta ctctctcttt gattctccag tttgaggtgtc tcctgctctct aagttcacaaca 840
cagtttctctt ctctcataag atctccttactt aattatggg gctggggaggg agagatgatgt 900
acatgtttac agctattagtt ttttagaggg ctgctcttcat cttgcccaca ttcgctggtgc 960
gggagggttt gccttgcagcc caaattccag acaagacaaaa atgaaaccctaa tccctcagag 1020
ttggacccag tttgagccac acaagagccata atgctcctctg atggtttgaaag ccctcctcacc 1080
taccgggtgc tggatgtaccc gatcataaca tttgtatatccc aaatcagcg agatcagcgg 1140
acccggagct ag

<210> 4

<211> 383

<212> PRT

<213> hybrid

<400> 4
Met Ser Lys Arg Asn Pro Lys Ile Leu Lys Ile Phe Leu Tyr Met Leu
Leu Leu Asn Ser Leu Phe Leu Ile Tyr Phe Val Phe His Ser Ser
Ser Phe Ser Pro Glu Glu Ser Glu Pro Pro His Ile Tyr His Val Ser
Val Asn Glu Ser Ala Ile Gly Glu Ser Ser Gly Glu Leu Arg Thr
Gly Gly Ala Arg Pro Pro Pro Leu Gly Ala Ser Ser Glu Pro Arg
Gly Gly Asp Ser Ser Pro Val Val Asp Ser Gly Pro Gly Pro Ala
Ser Asn Leu Thr Ser Val Pro Val Pro His Thr Thr Ala Leu Ser Leu
Pro Ala Cys Pro Glu Glu Ser Pro Leu Leu Val Gly Pro Met Leu Ile
Glu Phe Asn Met Pro Val Asp Leu Glu Leu Val Ala Lys Glu Asn Pro
Asn Val Lys Met Gly Gly Arg Tyr Ala Pro Arg Asp Cys Val Ser Pro
His Lys Val Ala Ile Ile Ile Pro Phe Arg Asn Arg Glu Glu His Leu
Lys Tyr Trp Leu Tyr Val Leu Gly Arg Glu Arg Glu Leu
Asp Tyr Gly Ile Tyr Val Ile Asn Glu Ala Gly Asp Thr Ile Phe Asn
aaagggatct gaggaaggtc

<210> b
<211> 544
<212> PRT
<213> Homo sapiens

<400> b

| Met Val Met Arg Arg Tyr Lys Leu Phe Leu Met Phe Cys Met Ala Gly |
|-------------|--------|--------|--------|--------|--------|--------|--------|
| 1           | 5      | 10     | 15     |

| Leu Cys Leu Ile Ser Phe Leu His Phe Phe Lys Thr Leu Ser Tyr Val |
|---------------------|--------|--------|--------|--------|--------|--------|
| 20                   | 25     | 30     |

| Thr Phe Pro Arg Glu Leu Ala Ser Leu Ser Pro Asn Leu Val Ser Ser |
|---------------------|--------|--------|--------|--------|--------|--------|
| 35                   | 40     | 45     |

| Phe Phe Trp Asn Asn Ala Pro Val Thr Pro Gln Ala Ser Pro Glu Pro |
|---------------------|--------|--------|--------|--------|--------|--------|
| 50                   | 55     | 60     |

| Gly Gly Pro Asp Leu Leu Arg Thr Pro Leu Tyr Ser His Ser Pro Leu |
|---------------------|--------|--------|--------|--------|--------|--------|
| 65                   | 70     | 75     | 80     |

| Leu Gln Pro Leu Pro Pro Ser Lys Ala Ala Glu Glu Leu His Arg Val |
|---------------------|--------|--------|--------|--------|--------|--------|
| 85                   | 90     | 95     |

| Asp Leu Val Leu Pro Glu Asp Thr Thr Glu Tyr Phe Val Arg Arg Thr Lys |
|---------------------|--------|--------|--------|--------|--------|--------|
| 100                  | 105    | 110    |

| Ala Gly Gly Val Cys Phe Lys Pro Gly Thr Lys Met Leu Glu Arg Pro |
|---------------------|--------|--------|--------|--------|--------|--------|
| 115                  | 120    | 125    |

| Pro Pro Gly Arg Pro Glu Lys Pro Glu Gly Ala Asn Gly Ser Ser |
|---------------------|--------|--------|--------|--------|--------|--------|
| 130                  | 135    | 140    |

| Ala Arg Arg Pro Pro Arg Tyr Leu Leu Ser Ala Arg Glu Arg Thr Gly |
|---------------------|--------|--------|--------|--------|--------|--------|
| 145                  | 150    | 155    | 160    |

| Gly Arg Gly Ala Arg Arg Lys Trp Val Glu Cys Val Cys Leu Pro Gly |
|---------------------|--------|--------|--------|--------|--------|--------|
| 165                  | 170    | 175    |

| Trp His Gly Pro Ser Cys Gly Val Pro Thr Val Val Gln Tyr Ser Asn |
|---------------------|--------|--------|--------|--------|--------|--------|
| 180                  | 185    | 190    |

| Leu Pro Thr Lys Glu Arg Leu Val Pro Arg Glu Val Pro Arg Arg Val |
|---------------------|--------|--------|--------|--------|--------|--------|
| 195                  | 200    | 205    |

| Ile Asn Ala Ile Asn Val Asn His Glu Phe Asp Leu Leu Asp Val Arg |
|---------------------|--------|--------|--------|--------|--------|--------|
| 210                  | 215    | 220    |

| Phe His Glu Leu Gly Asp Val Val Asp Ala Phe Val Val Cys Gly Ser |
|---------------------|--------|--------|--------|--------|--------|--------|
| 225                  | 230    | 235    | 240    |

| Asn Phe Thr Ala Tyr Gly Glu Pro Arg Pro Leu Lys Phe Arg Glu Met |
|---------------------|--------|--------|--------|--------|--------|--------|
| 245                  | 250    | 255    |

| Leu Thr Asn Gly Thr Phe Glu Tyr Ile Arg His Lys Val Leu Tyr Val |
|---------------------|--------|--------|--------|--------|--------|--------|
| 260                  | 265    | 270    |

| Phe Leu Asp His Phe Pro Pro Gly Gly Arg Gln Asp Gly Trp Ile Ala |
|---------------------|--------|--------|--------|--------|--------|--------|
| 275                  | 280    | 285    |

| Asp Asp Tyr Leu Arg Thr Phe Leu Thr Gln Asp Gly Val Ser Arg Leu |
|---------------------|--------|--------|--------|--------|--------|--------|
| 290                  | 295    | 300    |

<p>| Arg Asn Leu Arg Pro Asp Asp Val Phe Ile Ile Asp Asp Ala Asp Glu |
|---------------------|--------|--------|--------|--------|--------|--------|
| 310                  | 315    | 320    | 325    | 330    | 335    | 340    |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>305</td>
<td>Ile</td>
<td>Pro</td>
<td>Ala</td>
<td>Arg</td>
</tr>
<tr>
<td>310</td>
<td>Asp</td>
<td>Gly</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>315</td>
<td>Phe</td>
<td>Leu</td>
<td>Lys</td>
<td>Leu</td>
</tr>
<tr>
<td>320</td>
<td>Tyr</td>
<td>Asp</td>
<td>Gly</td>
<td>Trp</td>
</tr>
<tr>
<td>325</td>
<td>Thr</td>
<td>Glu</td>
<td>Pro</td>
<td>Phe</td>
</tr>
<tr>
<td>330</td>
<td>Ala</td>
<td>Phe</td>
<td>His</td>
<td>Met</td>
</tr>
<tr>
<td>335</td>
<td>Arg</td>
<td>Lys</td>
<td>Ser</td>
<td>Leu</td>
</tr>
<tr>
<td>340</td>
<td>Tyr</td>
<td>Gly</td>
<td>Phe</td>
<td>Phe</td>
</tr>
<tr>
<td>345</td>
<td>Trp</td>
<td>Lys</td>
<td>Glu</td>
<td>Pro</td>
</tr>
<tr>
<td>350</td>
<td>Gly</td>
<td>Thr</td>
<td>Leu</td>
<td>Glu</td>
</tr>
<tr>
<td>355</td>
<td>Val</td>
<td>Val</td>
<td>Ser</td>
<td>Gly</td>
</tr>
<tr>
<td>360</td>
<td>Cys</td>
<td>Thr</td>
<td>Val</td>
<td>Asp</td>
</tr>
<tr>
<td>365</td>
<td>Met</td>
<td>Leu</td>
<td>Gln</td>
<td>Ala</td>
</tr>
<tr>
<td>370</td>
<td>Val</td>
<td>Tyr</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td>375</td>
<td>Asp</td>
<td>Gly</td>
<td>Ile</td>
<td>Arg</td>
</tr>
<tr>
<td>380</td>
<td>Leu</td>
<td>Arg</td>
<td>Arg</td>
<td>Arg</td>
</tr>
<tr>
<td>385</td>
<td>Gln</td>
<td>Tyr</td>
<td>Tyr</td>
<td>Thr</td>
</tr>
<tr>
<td>390</td>
<td>Met</td>
<td>Pro</td>
<td>Asn</td>
<td>Phe</td>
</tr>
<tr>
<td>395</td>
<td>Arg</td>
<td>Gln</td>
<td>Tyr</td>
<td>Glu</td>
</tr>
<tr>
<td>400</td>
<td>Asn</td>
<td>Arg</td>
<td>Thr</td>
<td>Gly</td>
</tr>
<tr>
<td>405</td>
<td>His</td>
<td>Ile</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td>410</td>
<td>Gln</td>
<td>Trp</td>
<td>Ser</td>
<td>Leu</td>
</tr>
<tr>
<td>415</td>
<td>Gly</td>
<td>Ser</td>
<td>Pro</td>
<td>Leu</td>
</tr>
<tr>
<td>420</td>
<td>His</td>
<td>Phe</td>
<td>Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>425</td>
<td>Trp</td>
<td>His</td>
<td>Cys</td>
<td>Ser</td>
</tr>
<tr>
<td>430</td>
<td>Trp</td>
<td>Cys</td>
<td>Phe</td>
<td>Thr</td>
</tr>
<tr>
<td>435</td>
<td>Pro</td>
<td>Glu</td>
<td>Gly</td>
<td>Ile</td>
</tr>
<tr>
<td>440</td>
<td>Tyr</td>
<td>Phe</td>
<td>Lys</td>
<td>Leu</td>
</tr>
<tr>
<td>445</td>
<td>Val</td>
<td>Ser</td>
<td>Ala</td>
<td>Gln</td>
</tr>
<tr>
<td>450</td>
<td>Asn</td>
<td>Gly</td>
<td>Asp</td>
<td>Phe</td>
</tr>
<tr>
<td>455</td>
<td>Pro</td>
<td>Arg</td>
<td>Trp</td>
<td>Gly</td>
</tr>
<tr>
<td>460</td>
<td>Asp</td>
<td>Tyr</td>
<td>Glu</td>
<td>Asp</td>
</tr>
<tr>
<td>465</td>
<td>Lys</td>
<td>Arg</td>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td>470</td>
<td>Asn</td>
<td>Tyr</td>
<td>Ile</td>
<td>Arg</td>
</tr>
<tr>
<td>475</td>
<td>Gly</td>
<td>Leu</td>
<td>Ile</td>
<td>Arg</td>
</tr>
<tr>
<td>480</td>
<td>Thr</td>
<td>Gly</td>
<td>Gly</td>
<td>Trp</td>
</tr>
<tr>
<td>485</td>
<td>Phe</td>
<td>Asp</td>
<td>Gly</td>
<td>Thr</td>
</tr>
<tr>
<td>490</td>
<td>Gln</td>
<td>Gln</td>
<td>Glu</td>
<td>Tyr</td>
</tr>
<tr>
<td>495</td>
<td>Pro</td>
<td>Pro</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>500</td>
<td>Pro</td>
<td>Ser</td>
<td>Glu</td>
<td>His</td>
</tr>
<tr>
<td>505</td>
<td>Met</td>
<td>Tyr</td>
<td>Ala</td>
<td>Pro</td>
</tr>
<tr>
<td>510</td>
<td>Lys</td>
<td>Tyr</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td>515</td>
<td>Lys</td>
<td>Asn</td>
<td>Tyr</td>
<td>Asp</td>
</tr>
<tr>
<td>520</td>
<td>Arg</td>
<td>Phe</td>
<td>His</td>
<td>Tyr</td>
</tr>
<tr>
<td>525</td>
<td>Leu</td>
<td>Leu</td>
<td>Asp</td>
<td>Pro</td>
</tr>
<tr>
<td>530</td>
<td>Tyr</td>
<td>Gln</td>
<td>Glu</td>
<td>Pro</td>
</tr>
<tr>
<td>535</td>
<td>Arg</td>
<td>Pro</td>
<td>Ala</td>
<td>Arg</td>
</tr>
<tr>
<td>540</td>
<td>Gly</td>
<td>Lys</td>
<td>Leu</td>
<td>Ile</td>
</tr>
</tbody>
</table>

<210> 7
<211> 10
<212> PRT
<213> Artificial Sequence

<220> Synthetic

<400> 7

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glu</td>
<td>Gln</td>
<td>Lys</td>
<td>Leu</td>
</tr>
<tr>
<td>5</td>
<td>Ile</td>
<td>Ser</td>
<td>Glu</td>
<td>Glu</td>
</tr>
<tr>
<td>10</td>
<td>Asp</td>
<td>Leu</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> 8
<211> 31
<212> PRT
<213> Homo sapiens

<400> 8

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gln</td>
<td>Glu</td>
<td>Pro</td>
<td>Arg</td>
</tr>
<tr>
<td>5</td>
<td>Ser</td>
<td>Thr</td>
<td>Ala</td>
<td>Ala</td>
</tr>
<tr>
<td>10</td>
<td>Gly</td>
<td>Gly</td>
<td>Trp</td>
<td>Arg</td>
</tr>
<tr>
<td>15</td>
<td>His</td>
<td>Arg</td>
<td>Gly</td>
<td>Pro</td>
</tr>
<tr>
<td>Met</td>
<td>Ser</td>
<td>Lys</td>
<td>Arg</td>
<td>Asn</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ser</td>
<td>Leu</td>
<td>Phe</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Phe</td>
<td>Ser</td>
<td>Pro</td>
<td>Glu</td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Asn</td>
<td>Gln</td>
<td>Ser</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Tyr</td>
<td>Ser</td>
<td>His</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Leu</td>
<td>His</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Tyr</td>
<td>Phe</td>
<td>Val</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Domain:** DNA

**Sequence:**

catgagttttgaa cggacataccga agatcctgga attttttcttg tatatatggac tttttaacttcc 40
ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta 50
ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta 60
ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta 70
ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta 80
ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta 90
ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta 100
ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta ttttttttta 110
Thr Lys Met Leu Gly Arg Pro Pro Pro Gly Arg Pro Gly Glu Lys Pro
115 120 125
Glu Gly Ala Asn Gly Ser Ser Ala Arg Arg Pro Arg Arg Tyr Leu Leu
130 135 140
Ser Ala Arg Glu Arg Thr Gly Gly Arg Gly Ala Arg Arg Lys Trp Val
145 150 155 160
Glu Cys Val Cys Leu Pro Gly Trp His Gly Pro Ser Cys Gly Val Pro
165 170 175
Thr Val Val Gln Tyr Ser Asn Leu Pro Thr Lys Glu Arg Leu Val Pro
180 185 190
Arg Glu Val Pro Arg Arg Val Ile Asn Ala Ile Asn Val Asn His Glu
195 200 205
Phe Asp Leu Leu Asp Val Arg Phe His Glu Leu Gly Asp Val Val Asp
210 215 220
Ala Phe Val Val Cys Glu Ser Asn Phe Thr Ala Tyr Gly Glu Pro Arg
225 230 235 240
Pro Leu Lys Phe Arg Glu Met Leu Thr Asn Gly Thr Phe Glu Tyr Ile
245 250 255
Arg His Lys Val Leu Tyr Val Phe Leu Asp His Phe Pro Pro Gly Gly
260 265 270
Arg Gln Asp Gly Trp Ile Ala Asp Asp Tyr Leu Arg Thr Phe Leu Thr
275 280 285
Gln Asp Gly Val Ser Arg Leu Arg Asn Leu Arg Pro Asp Asp Val Phe
290 295 300
Ile Ile Asp Asp Ala Asp Glu Ile Pro Ala Arg Asp Gly Val Leu Phe
305 310 315 320
Leu Lys Leu Tyr Asp Gly Trp Thr Glu Pro Phe Ala Phe His Met Arg
325 330 335
Lys Ser Leu Tyr Gly Phe Phe Trp Lys Gln Pro Gly Thr Leu Glu Val
340 345 350
Val Ser Gly Cys Thr Val Asp Met Leu Gln Ala Val Tyr Gly Leu Asp
355 360 365
Gly Ile Arg Leu Arg Arg Gln Tyr Tyr Thr Met Pro Asn Phe Arg
370 375 380
Gln Tyr Glu Asn Arg Thr Gly His Ile Leu Val Gln Trp Ser Leu Gly
385 390 395 400
Ser Pro Leu His Phe Ala Gly Trp His Cys Ser Trp Cys Phe Thr Pro
405 410 415
Glu Gly Ile Tyr Phe Lys Leu Val Ser Ala Gln Asn Gly Asp Phe Pro
420 425 430
Arg Trp Gly Asp Tyr Glu Asp Arg Asp Leu Asn Tyr Ile Arg Gly
435 440 445
Leu Ile Arg Thr Gly Gly Trp Phe Asp Gly Thr Gln Gln Glu Tyr Pro
450 455 460
Pro Ala Asp Pro Ser Gly His Met Tyr Ala Pro Lys Tyr Leu Leu Lys
465 470 475 480
Asn Tyr Asp Arg Phe His Tyr Leu Leu Asp Asn Pro Tyr Gln Glu Pro 485 490
Arg Ser Thr Ala Ala Gly Gly Trp Arg His Arg Gly Pro Glu Gly Arg 500 505 510
Pro Pro Ala Arg Gly Lys Leu Asp Glu Ala Glu Val Glu Gln Lys Leu 515 520 525
Ile Ser Glu Glu Asp Leu Asn 530 535

<210> 11
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 11
aatctttcga ccc 13

<210> 12
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 12
ccacccgtta acaatgaaga tgagacgta caag 34

<210> 13
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 13
gggcagagga gatgagacgc tacaagctc 29

<210> 14
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 14
ggatccaatg aagatgagac gctacaag 28

<210> 15
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 15
GGGCCGAAA GATCCATT CAGATCTCT TCTGAGATGA

<210> 16
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 16
CCCGATCTCT AATCAGATC CTCTCTGAG ATGAG

<210> 17
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 17
GGGTCTAGAT CCTAATTCG ATCCTCTCT GAGATGAG

<210> 18
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 18
CCACCCTTTA ACAATGAGT AACGGAATCC GAAGA

<210> 19
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 19
GGGCCATGGG TAAACCGGA ATCCGAGATC TGAG

<210> 20
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 20
CCCGATCCA TGAGTTACG GAATCCGAG ATTC

<210> 21
<211> 29
DNA
Artificial Sequence

Synthetic

21
ggcccccggg acgctagctc ggtgtcccg

22
cccggatcca cgctagctg gtcgc

23
gggtctagat ccacgctagc tgcgttcccc g

24
ccacccggtta acaatgagcc ttcgggagcc gctcctgag

25
26
27
28
29
30
31
32
33
34
35
<table>
<thead>
<tr>
<th>Line</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>210</td>
<td>ggcgcagccc gaggcgtcag cagcatcaat cttaaccaatt acagatcgtg tttggatcac</td>
</tr>
<tr>
<td>211</td>
<td>ggcgagattc gttcagaggt gtttctttt gctcattaata gaccagagct gtcgtacgct</td>
</tr>
<tr>
<td>212</td>
<td>atccttgatt gacatagggct gtttacttg</td>
</tr>
<tr>
<td>213</td>
<td>tgcctttacc acgagagatg agatataaag acctccctcgt ggttaataat tctccttttc ttcagtatatttttttgcggtggttttcggttcctttct gggccttggct ggttcggtgct</td>
</tr>
<tr>
<td>214</td>
<td>atccagatgg cggacgctgg ggtggcttgc</td>
</tr>
<tr>
<td>215</td>
<td>tgggcttgc ttttgcgggc tgtcttttgtc ggtgggtggc</td>
</tr>
<tr>
<td>216</td>
<td>gtttggag ggacagcttg</td>
</tr>
<tr>
<td>217</td>
<td>ggggataaca cttgcttgcc</td>
</tr>
<tr>
<td>218</td>
<td>cttcctttcc cacacactaat ctaatagcgtg</td>
</tr>
<tr>
<td>219</td>
<td>gtttggag ggacagcttg</td>
</tr>
<tr>
<td>220</td>
<td>ggggataaca cttgcttgcc</td>
</tr>
<tr>
<td>221</td>
<td>cttcctttcc cacacactaat ctaatagcgtg</td>
</tr>
</tbody>
</table>

**Note:** The text appears to be a sequence of nucleotides, possibly from DNA or RNA, but without context, it's difficult to provide a meaningful interpretation.
WO 03/078637

15/38

PCT/IB03/01626

ttgggtggag attatgctgg acaagccggt ttgcctgctat gacgccagcg tctctagggca 4140
aggtgtgatt gatacccgcg caatgaccgtt gttattttcct tctttcgcgg aacttaaactat 4200
tctcgaagc gccggctgct ccacactaa cccggaagaa ccctgcttct ttcctttactc 4250
cataaggctt cactatctta acccctaaaa ctcactttaa gcgaagaac gcgcgaacatc 4300
atctctgtct gtcttcaacat acggttccct tgtctcttta cgcacattagt taccattgca 4350
cctccacatct gtaaattcct aggttcttgc tttcactacaa actcctcactct aatgtgaaga 4400
gacagcagc aagtcgcttc ttaactctaa tagactgacct ttggacttcc tctttgatcc 4450
taaagaagga caagtaaatc gcaacagcatt gctaataagg gatcgatgat tcttcgcatac 4500
tggctaaagat ctgctcagctt caagagataa aacactttaa ccgtaactct gaaactacgg 4550
tatgagacatt ctggctctttg agcgagtttc agttcaggtc agtgctcaga 4600
agggcgtgcc gcagccttct tcgtcatgttc caatggctagg aagcagttgg ctgcaggtcc 4650
cagattgacac ctgctgctgc taacctgtac aaggttggag atcgctcgat 4700
taatgaagca ctgagaagct ctaaattcct gtttgctcttt actctcatca 4750
acgtgggttt ttggaatttt atcgtgagag atgatatatt cacagaactc acggcatact 4800
ttgagactct ccagactcttt tctgctagtt gttcgtttttt ctacttctgta 4850
tatctatca ctgctagtaa aacagcagcct tcctcttaaag ctaactattc 4900
atattgagaa ggcagccgct acggctgttt gttggtgattg aagcagcgat 4950
ttcccataatt cttgcatcct aatcagaggctc cttgctagtct gcagcgtcag 5000
aattgtgggtt atgccctaaa attttggctt tagagtccat caattgcatt 5050
ggcattttct acctcggctt gtttccttccttat ccatactag ccctttgctt 5100
gctcaagaa gacgctggtt accactgtatt cttgccttct tgcgcttctt 5150
tgccagatt cactttcact cttcctcagc cttcctttct actctctcatt ccgctccttt 5200
acactagtttg gatccagctt gctcaggttttag cttgctagtt gacccggtct 5250
atttaaagctt gaaatctgct gtgcagcttg ccgctgtttt ctcgctttttt 5300
ccgagagg ccagataaag ctctttgactat ttttatttattat ccgctcagtc 5350
tatactacatt ttttagtact gcgtgatgat ctgactgtaat catgcgtctt ggcctgcc 5400
ccgctgtttt gatccagttttt gactcattttct gctgattttt tgcgactccct aatacttttc 5450
cgctcgggtt ccgctaactg tatttccttg tgcgacttcg ctgtgcttttc 5500
tcctcttttt tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
<211> 1394
<212> DNA
<213> hybrid

<400> 31
ccatggcgag gatctcgtg t gacttgatg ttctttcctc t t cccgccagct t tcatgttca 160
cttcatcaca ttggagccct t t tccagacgc aactacagta tgcagatgc tttcaagctgg 120
tatatcagtc accgattgac gcctcctgaa atatgaagaa cccggcagac gaagacctg 140
ttcgagtctaa tggatctact cagacgggt taaaaaag aatgaaagaa ctctcaccag 300
tttgagacct tgccttcaga ttgcatagaa ggaacctatt ttgcgtcaga aatccgagtc 360
ttgctcccagg t t cctggtaaa ttatgcgtctt tttgcgctat aatgccggtc 420
atttttccct gctacagcgt gaaagttgg agaaggttaa aggttataag gagaacattg 480
tgtagttgag tcagttggag t actttggaag aagatgaaag tagttgagaag agttaaatg 540
ttggtcagag gaaacgcatc tttcgcctt c tttggcctca tatatactgt actagcttcc 600
cgggtgtgac ccagatgtat gtgaagaac a g ggtagtgat gggaaaggg cattgtgaa 660
gatattcggc tgcagatggc acatccggct ccggagatgt gatattcgtt aacatcacct 720
ggtgtggttg gatgaaacact gtattggagtg cggttgaaaga gactaaagga cattgagggc 780
atattttttt cattgaaaga gacatctttc ttggtttcaac tgcctacgtg aacatacaca 840
tttcatcaggg gctgaagggc gcaagttgcct ctagcttgct ttcgctcaat ttagcaccgt 900
tgctgggagc gtaagagacca gaaagttctg aagatgttt tgcagaaaaa ctagaaagat 960
<210> 32
<211> 32
<212> DNA
<213> Arabidopsis thaliana

<400> 32
ccatggcgag gatctcgtg t gacttgatg ttctttcctc t cccgccagct t tcatgttca 160
cttcatcaca ttggagccct t t tccagacgc aactacagta tgcagatgc tttcaagctgg 120
tatatcagtc accgattgac gcctcctgaa atatgaagaa cccggcagac gaagacctg 140
ttcgagtctaa tggatctact cagacgggt taaaaaag aatgaaagaa ctctcaccag 300
tttgagacct tgccttcaga ttgcatagaa ggaacctatt ttgcgtcaga aatccgagtc 420
atttttccct gctacagcgt gaaagttgg agaaggttaa aggttataag gagaacattg 480
tgtagttgag tcagttggag t actttggaag aagatgaaag tagttgagaag agttaaatg 540
ttggtcagag gaaacgcatc tttcgcctt c tttggcctca tatatactgt actagcttcc 600
cgggtgtgac ccagatgtat gtgaagaac a g ggtagtgat gggaaaggg cattgtgaa 660
gatattcggc tgcagatggc acatccggct ccggagatgt gatattcgtt aacatcacct 720
ggtgtggttg gatgaaacact gtattggagtg cggttgaaaga gactaaagga cattgagggc 780
atattttttt cattgaaaga gacatctttc ttggtttcaac tgcctacgtg aacatacaca 840
tttcatcaggg gctgaagggc gcaagttgcct ctagcttgct ttcgctcaat ttagcaccgt 900
tgctgggagc gtaagagacca gaaagttctg aagatgttt tgcagaaaaa ctagaaagat 960
<210> 33
<211> 33
<212> DNA
<213> Glycine max

<400> 33
ccatggcgag aggagcaga tcaagtggta tgcagacagac caaatggagg tactgcaacc 160
tttctcttaa cttcaacggtg ccacacgtcttg tgcagtccgct tttcatcgtttcttgtttg 120
ttatcagttc caccgatttg ctggctggttc cttcgtttcttg cggagattc gaaagctttg 140
atattttttt cattgaaaga gacatctttc ttggtttcaac tgcctacgtg aacatacaca 300
tttcatcaggg gctgaagggc gcaagttgcct ctagcttgct ttcgctcaat ttagcaccgt 420
atttttccct gctacagcgt gaaagttgg agaaggttaa aggttataag gagaacattg 480
tgtagttgag tcagttggag t actttggaag aagatgaaag tagttgagaag agttaaatg 540
ttggtcagag gaaacgcatc tttcgcctt c tttggcctca tatatactgt actagcttcc 600
cgggtgtgac ccagatgtat gtgaagaac a g ggtagtgat gggaaaggg cattgtgaa 660
gatattcggc tgcagatggc acatccggct ccggagatgt gatattcgtt aacatcacct 720
ggtgtggttg gatgaaacact gtattggagtg cggttgaaaga gactaaagga cattgagggc 780
atattttttt cattgaaaga gacatctttc ttggtttcaac tgcctacgtg aacatacaca 840
tttcatcaggg gctgaagggc gcaagttgcct ctagcttgct ttcgctcaat ttagcaccgt 900
tgctgggagc gtaagagacca gaaagttctg aagatgttt tgcagaaaaa ctagaaagat 960
<210> 34
<211> 34
<212> DNA
<213> hybrid

<400> 34
ggcgaggcacgcctt gacgtagcatc gacagtccccat attaaacaaatt ttcgatcgct cagatagtac 60
tttgtggagag ttcatactact tgaatatggg cagtattttc attaatctcg aaaaatcctg 120
gtttttcacct gacagagcgctt ttcagcttttc gtttaatgtagc umaacgcaaa cttttcctt 180
gttttaggagc tttgcaacaaat ttaatgttaca cttgacttaac gacaactggg attttgatt 240
WO 03/078637

19/38

tgtcggccgca gatagcagtc tattcttatg 4380
tatcttcga ccctgagtttagt tagtttgtttg 4440
tatcttttaattta gacctagcagac 5410
tgtcggcagt cagatgcagtc tattcttatg 4380
tatcttcgagc cctgacagttt gacctagcagac 5410
tatatcacaagc cctcaggtgcc gatagcagtc 4380
tatcttcgagc cctcaggtgcc gatagcagtc 4380
tatatcacaagc cctcaggtgcc gatagcagtc 4380
cttttcattg aagaagatca ttttctgttt cctaatgcct atctgaaatc acgacacctc tt460
acagggctga aacccgacaat gttgctctgc tgttttcgct ctaaatgagc accgctctgat 8520
gtaaactga gagaagagtt cttggtcag agagaaatgg aatagttgg 8580
tttcttttta atagagctt gttgggaat atatacagag aggggaagtt gttttttttt 8640
tttgtattt caaactgagga tataaagg ggggcaactcg tttttttcgc gtgattttcct 7800
cccgggttgca cattgcaaggg cctctccgct atctgaaatc acgacacctc 8000
ctacaactg gtaggtttgt atctgctttt atcgtgatcg aatagttgg 8420
ccaagacgcgg taacgcgggc ttggtggagtt ggggcaactcg tttttttcgc 8580
catatttttt cttttttttt cttttttttt cttttttttt 8740
catttatattattgtag cttttttttt 8900
acaagctatc ctagttttgc caaatacttc acctctttttc tctttttttt 9060
atatcttcgta caaataacat cttttttttt 9240
<400> 35
<413> DNA
<413> hybrid
<413> 35

<400> 35

ggcggcggctc gaggcgtacg cagatcgatc ataaacaaat ttgaatgcca gagatcctc 60
tcttttggaga ttcttaccct agaaatgta gacagtttttt aaatctcttg aaaaatcttg 120
gttcttcttt gacggaagaa gacgcagact caaatatttc gttggtactt gacgcagaat 180
ggttgaacctg taataatgtg gtagtacagc ccaactacag atttgttatt 240
catggtttgc ctttctggtg ccatattaat gatgtgcagg cctaaactta atccgaaacg 300
tttattttcc cgcagtcgga cgtgggttagt tgggggaatc tgggtgctagt 360
ggttagcctt gctcagatcc ctaaactcga tctgcagcct tgaattttggc caaagactc 420
catttggagt catattttt cgtacagttt aaacactcgg gacaaatctc tatactcatt 480
aaaaacccct gcatactgta aacgctcctt accattttttt cttataaatt atatacatca 540
gccgtagctg gtaggacact cacaatatta cttttttttt cttttttttt cttttttttt 600
atgttgtgact gcataagagt tttttctaa gatgcagcag taggtttttt 650
gcgcgcaag cagcttttct tttttttttt tttttttttt tttttttttt 720
ttcttgtctg ccgttttttt tttttttttt tttttttttt gttattttttt tattatcttt 780
ctcctcatca gcacacagtt gaaatttttt aagctgcaaa gatggactt gatttttttt 840
attttttttt gtaggtttttt aacaaataac ggtggcggcc tttttttttt 900
ccattggtga talattttttt tcggcagttt gagtttgctag ggcgagactc 960
tttgaaagac tgattataat cttttttttt atcataactct ccacagtcct 1020
tattattact ctgactttag gttcagtttt caagctattt acgctactaa 1080
aattaactata tagcagcatt ttttattttt gacatagttt cctttttttttt 1140
tgactcgtag tcaaagatt gcagcgtcact aacaagttgct acggtgcaag tggataagtt 1200
aacaacaaat ttgattgtatt attataatat aagtttattt gaaatatttt ccaaccttt 1260
ttgtactact gtaggtgtcc tttcagtttt aacactcatg atttgtcttt 1320
catctgtaga tgaataagtt cagaaatcgc tttttttttt cttattcag 1380
cagattttttt tctctcgtct ggttgggtact gctggtgactc 1440
ccaagtgggc cagtagtttt ttggtggtatt ggctggtgata aacaaattttt caaagctctc 1500
gcataattcat tcggcagcgg cttctgttac aggtttttttt tttttttttt 1560
ttgagcagtt tctctcgtc tttttttttt ctttataagttt attttttttt 1620
ggaagacaa gacgctactaa taccatagtt cgccacagctt tttttttttt 1680
taggtagacg acagcacagc attcagattt cttctggtct tttttttttt 1740
aggaggtgtg tctgcagtttt tttcagtttt cttttaagct cttttttttt 1800
ttttagatt aactaagatt gggtgtgctt gtagcctttt tttttttttt 1860
tgatggatgtg tgcgtttttt gtttctgtttt ggcacagtct gcagctattt 1920
tggagacagcg taaaatcggc gcgtgtgatc tctggtgactc 1980
tatcataact tttcttcagttt tttttttttt tttttttttt 2040
ggatagtttt tttttttttt gtttggtcag atataaaagtt cttctttttt 2100
gtattttttt ttggataattt cttttttttt aataatctttt ccacagtgct 2160
gcagagagc tctaaatttt 2180

<423> Arabidopsis thaliana
gaaatgtaaga agcttttcca aactcctgta aaataatcgg ttttccttct gacgagaag 120
agcagacctc caaatattcag gttagactcg acgccgaag tttgactggt gcaccaatt 160
taattggtac tacgtaaccc acctactcgg gcccttgatt ggcccttttc gcctgttggc 200
cctatacgg atgtgacgag ctaaactaaa ttgtgagcgt ttaaacacgc gtagccgcac 260
gttgtgatttt cacgcaatag caactaattacttgcaagctgtagcagttcccagctcaggtt 310
aagagtataaa aggctgagtc ggtaacactga aacatcgtt 150
agtgcattgaa gttcttactag tcgcacgtcct cacagtcttt ctgtgctagcattttctct 150
ccgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150
cgctctggt tctaacttactcg ctagtggaa ccgcacggtt tacgtagctcc aggcttctccctcttctcttc 150
ctctctactt ttcgctgagct ctagtcacgct gcagctgcgtc cagcacgct gcagctgctc 150

cggcgccctg ggggactctg cagatcttac ctaacccatt aagcaacgc ggcttacagc 60
tgtagctcgg gttgcagactc ttcattcact gtttgacttg atcgaaactc 100
acaaaattt ctggggactc atgcgaactc ttcctgctac aagatcagc gcgtttacct 120
atcactggct atcactggct atgcgaactc ttcctgctac aagatcagc gcgtttacct 120
acaaaattt ctggggactc atgcgaactc ttcctgctac aagatcagc gcgtttacct 120
acaaaattt ctggggactc atgcgaactc ttcctgctac aagatcagc gcgtttacct 120
acaaaattt ctggggactc atgcgaactc ttcctgctac aagatcagc gcgtttacct 120
acaaaattt ctggggactc atgcgaactc ttcctgctac aagatcagc gcgtttacct 120
<table>
<thead>
<tr>
<th>120</th>
<th>T</th>
<th>C</th>
<th>G</th>
<th>A</th>
<th>G</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>180</td>
</tr>
<tr>
<td>180</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>210</td>
</tr>
<tr>
<td>210</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td>240</td>
</tr>
<tr>
<td>240</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>270</td>
</tr>
<tr>
<td>270</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>300</td>
</tr>
<tr>
<td>300</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td>330</td>
</tr>
<tr>
<td>330</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>360</td>
</tr>
</tbody>
</table>

DNA sequence: TCGACTGAGTCAGTATCGAGT

RNA sequence: AGTCAGTACGACTGATCGACT

Protein sequence: ACTGAGTCAGTATCGAGT

RNA: AGTCAGTACGACTGATCGACT

Protein: ACTGAGTCAGTATCGAGT

DNA: TCGACTGAGTCAGTATCGAGT

Protein: ACTGAGTCAGTATCGAGT

RNA: AGTCAGTACGACTGATCGACT
<210> 42
<211> 541
<212> DNA
<213> Solanum tuberosum

<400> 42
agatcctcatt atacggttag aacgatatg attatatcttatgattttct gctttgctgtactctgattctccatcctttactttct 50
atttacatt gtttgggtctt ctatattatt taagatcccc attaagtctccataaagaattttaa taagatgttttagttgaa 120
atatcctag ctagttgttctcatttacatc aagattaattcatttatttattttatttttttatttttttttatttttttttatttttttttttattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Artificial Sequence

Synthetic

45
atactcgagt taacaatgag taaacggaat c

24

DNA

Artificial Sequence

Synthetic

46
ttctcgatcg cccgattggtt attc

24

47
gccgcccgca tggggcagtc ctcc

48

30

DNA

Artificial Sequence

Synthetic

48
aacggatcca cgctagctcg gtgtccccgat

49

3327

DNA

hybrid

49
atgggccatca agatggagac acatttcag gtctttgtat acatggtctgt gtgtttgtct

ggtgtcgcata tgaagcactt caaatctccta ctacctcaca ccgctaaag agcagacag 120
ccacacctccg ataatactgcc tcgattgaag gaaagcgaag cggacatcga caccgtggcg

33

ataatcaca cttttgtat tcgctgga gctgttgtgc ta caaagggat tttggcag

240

tctcgtgag ctaaatgtaq aagaattcata aacagacacta cagcgcctag actgaaagta

300

tctctgggcttc cttctcaca caaaggcccg gatgtgctga acaggtttgta agacattcct

360

gataatcgg aacaagacat tataaaacaa atagtcgaaca aactgacaa ctaacctccac aacgtgaga 420

tagbcccctca tttgaggcca gataatcgtt cttgaatcct gttggaagag gtgccacct 480

gtaggagaaaa acgtgtagca aaaaactattca aaaaatggac gttctggaat cacgagggg 540

ggtgtggagttg gaccgagcag aagttcgcag ctaatcatttg cgttaattgc ccaagtttatt 600

gagacatcag actgggtgaa aactaatcttc ggcgcactcc cgaagcagag atgtgctatt 660

gaccctttccg gccaggggcc aactgtgctcc ttacattcag accagacgga ctttgagaga 720

catttgacag aagataacctta tattgctggtg aaacagtggg tggctggaag cagagttgag 780

gatagtctct ggcgtgcggag ttggtctacat acgagacgct cccatgtcatt gcaccaactcag 840

cgcgttgatattatatc aaaaagcacgg tttggccgcgc acccttcatat ttgcttctagt 900

ttggacttca ggagattttc ccgggaatat tctggataca cagctgaga cgaagacatc 960

cagcgagcagc acttgccagcat caagcggaaag acatttgagat aggagtagcag cgtctacggg 1020
<table>
<thead>
<tr>
<th>Met</th>
<th>Gly</th>
<th>Ile</th>
<th>Lys</th>
<th>Met</th>
<th>Glu</th>
<th>Val</th>
<th>Phe</th>
<th>Val</th>
<th>Tyr</th>
<th>Met</th>
<th>Leu</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Trp</td>
<td>Leu</td>
<td>Ser</td>
<td>Gly</td>
<td>Val</td>
<td>Asp</td>
<td>Met</td>
<td>Lys</td>
<td>His</td>
<td>Phe</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Thr</td>
<td>Val</td>
<td>Lys</td>
<td>Ser</td>
<td>Arg</td>
<td>Asp</td>
<td>Glu</td>
<td>Pro</td>
<td>Thr</td>
<td>Pro</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Lys</td>
<td>Glu</td>
<td>Ser</td>
<td>Glu</td>
<td>Ala</td>
<td>Asp</td>
<td>Thr</td>
<td>Val</td>
<td>Ala</td>
<td>Ile</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Asp</td>
<td>Phe</td>
<td>Gln</td>
<td>Pro</td>
<td>Ser</td>
<td>Trp</td>
<td>Leu</td>
<td>Arg</td>
<td>Thr</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Phe</td>
<td>Glu</td>
<td>Asp</td>
<td>Arg</td>
<td>Tyr</td>
<td>Glu</td>
<td>Arg</td>
<td>Ile</td>
<td>His</td>
<td>Asn</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Leu</td>
<td>Lys</td>
<td>Val</td>
<td>Ile</td>
<td>Val</td>
<td>Pro</td>
<td>His</td>
<td>Ser</td>
<td>His</td>
<td>Asn</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

|<210> | 50 |
|<211> | 110A |
|<212> | PRT |
|<213> | hybrid |

<p>|&lt;400&gt; | 50 |</p>
<table>
<thead>
<tr>
<th>115</th>
<th>120</th>
<th>125</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td>Asn</td>
<td>Asn</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>Trp</td>
<td>Thr</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>Val</td>
<td>Lys</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>Ile</td>
<td>Thr</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>200</td>
<td>205</td>
</tr>
<tr>
<td>Tyr</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>215</td>
<td>220</td>
</tr>
<tr>
<td>Asn</td>
<td>Gly</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>His</td>
<td>Gly</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Thr</td>
<td>Ile</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>Arg</td>
<td>Gln</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>Pro</td>
<td>Ser</td>
<td>Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>290</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Ser</td>
<td>Thr</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>Lys</td>
<td>Ile</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Thr</td>
<td>Glu</td>
<td>His</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Asp</td>
<td>Arg</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Gly</td>
<td>Asp</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Val</td>
<td>Asn</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Phe</td>
<td>Asn</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Met</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>Val</td>
<td>Tyr</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>435</td>
<td>440</td>
<td>445</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyr</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>450</td>
<td>455</td>
<td>460</td>
</tr>
<tr>
<td>His</td>
<td>Gln</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Arg</td>
<td>Glu</td>
</tr>
</tbody>
</table>
Leu Glu Lys Ser Tyr Glu Gln Leu Ile Tyr Ala Arg Arg Asn Leu Gly 465
Leu Phe Gln His Asp Ala Ile Thr Gly Thr Ser Lys Ser Ser Ser Val 500
Met Gln Asp Tyr Gly Thr Lys Leu Phe Thr Ser Leu Tyr His Cys Ile 515
Arg Leu Gln Glu Ala Ala Leu Thr Thr Ile Met Leu Pro Asp Gln Ser 530
Leu His Ser Gln Ser Ile Ile Gln Ser Glu Val Glu Trp Glu Thr Tyr 545
Gly Lys Pro Pro Lys Lys Leu Gln Val Ser Phe Ile Asp Lys Lys Lys 565
Val Ile Leu Phe Asn Pro Leu Ala Glu Thr Arg Thr Glu Val Val Thr 580
Val Arg Ser Asn Thr Ser Asn Ile Arg Val Tyr Asp Thr His Lys Arg 595
Lys His Val Leu Tyr Gln Ile Met Pro Ser Ile Thr Ile Gln Asp Asn 610
Gly Lys Ser Ile Val Ser Asp Thr Thr Phe Asp Ile Met Phe Val Ala 625
Thr Ile Pro Pro Leu Thr Ser Ile Ser Tyr Lys Leu Gln Glu His Thr 645
Asn Thr Ser His His Cys Val Ile Phe Cys Asn Asn Cys Glu Glu Tyr 655
Gln Lys Ser Asn Val Phe Gln Ile Lys Lys Met Met Pro Gly Asp Ile 670
Gln Leu Glu Asn Ala Val Leu Lys Leu Leu Val Asn Arg Asn Thr Gly 695
Phe Leu Arg Gln Val Tyr Arg Lys Asp Ile Arg Lys Arg Thr Val Val 710
Asp Val Gln Phe Gly Ala Tyr Gln Ser Ala Gln Arg His Ser Gly Ala 725
Tyr Leu Phe Met Pro His Tyr Asp Ser Pro Glu Lys Asn Val Leu His 740
Pro Tyr Thr Asn Gln Asn Met Gln Asp Asp Asn Ile Ile Ile Val 755
Ser Gly Pro Ile Ser Thr Glu Ile Thr Thr Met Tyr Leu Pro Phe Leu 770
Val His Thr Ile Arg Ile Tyr Asn Val Pro Asp Pro Val Leu Ser Arg 785
Ala Ile Leu Leu Glu Thr Asp Val Asp Phe Glu Ala Pro Pro Lys Asn 805
Arg Glu Thr Glu Leu Phe Met Arg Leu Gln Thr Asp Ile Gln Asn Gly 820
Asp Ile Pro Glu Phe Tyr Thr Asp Gln Asn Gly Phe Gln Tyr Gln Lys
835 840 845

Arg Val Lys Val Asn Lys Leu Gly Ile Glu Ala Asn Tyr Tyr Pro Ile
850 855 860

Thr Thr Met Ala Cys Leu Gln Asp Glu Glu Thr Arg Leu Thr Leu Leu
865 870 875 880

Thr Asn His Ala Gln Gly Ala Ala Ala Tyr Glu Pro Gly Arg Leu Glu
885 890 895

Val Met Leu Asp Arg Arg Thr Leu Tyr Asp Asp Phe Arg Gly Ile Gly
900 905 910

Glu Gly Val Val Asp Asn Lys Pro Thr Thr Phe Gln Asn Trp Ile Leu
915 920 925

Ile Glu Ser Met Pro Gly Val Thr Arg Ala Lys Arg Asp Thr Ser Glu
930 935 940

Pro Gly Phe Lys Phe Val Asn Glu Arg Arg Phe Gly Pro Gly Glu Lys
945 950 955 960

Glu Ser Pro Tyr Gln Val Pro Ser Gln Thr Ala Asp Tyr Leu Ser Arg
970 975

Met Phe Asn Tyr Pro Val Asn Val Tyr Leu Val Asp Thr Ser Glu Val
980 985 990

Gly Glu Ile Glu Val Lys Pro Tyr Gln Ser Phe Leu Gln Ser Ser Phe Pro
995 1000 1005

Pro Gly Ile His Leu Val Thr Leu Arg Thr Ile Thr Asp Asp Val
1010 1015 1020

Leu Glu Leu Phe Pro Ser Asn Glu Ser Tyr Met Val Leu His Arg
1025 1030 1035

Pro Gly Tyr Ser Cys Ala Val Gly Glu Lys Pro Val Ala Lys Ser
1040 1045 1050

Pro Lys Phe Ser Ser Lys Thr Arg Phe Asn Gly Leu Asn Ile Gln
1055 1060 1065

Asn Ile Thr Ala Val Ser Leu Thr Gly Leu Lys Ser Leu Arg Pro
1070 1075 1080

Leu Thr Gly Leu Ser Asp Ile His Leu Asn Ala Met Glu Val Lys
1085 1090 1095

Thr Tyr Lys Ile Arg Phe Lys Asp Glu Leu
1100 1105

<210> 51
<211> 1068
<212> DNA
<213> hybrid

<400> 51
atgggcatac agatgggagac acattctcag gtctttgtat acatgtggtct tgtgtgtctct
60
ggtgctgaca tgcagtcttc cgggaggtct cggacggagag gggccccggcc gcggccctctt
120
cctaggccgct cctcccgccgg gccgcggggt tgtgacacca ggcctgagctt gattctgcc
180
cctgcccccg ccattcggctt ccgctgcccc acacgaccgc actgctgtctg
240
cGGcGcGctc cgtaggagctt ccggccctctt tgtgagagctgtt gtttatagctgttaacatg
300
cctgttgaccc tcggagctgctt ggacagacag aaccacaagat tgaagatgg cgccgcgtatt
360
gccccccccggt actgctgtct cctctcacaag atggcgcatac tcattccatt cccacccggt
420
<table>
<thead>
<tr>
<th>AA</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met</td>
<td>Gly</td>
<td>Ile</td>
<td>Lys</td>
<td>Met</td>
</tr>
<tr>
<td>Leu</td>
<td>Trp</td>
<td>Leu</td>
<td>Ser</td>
<td>Gly</td>
</tr>
<tr>
<td>Gly</td>
<td>Ala</td>
<td>Arg</td>
<td>Pro</td>
<td>Pro</td>
</tr>
<tr>
<td>Pro</td>
<td>Gly</td>
<td>Gly</td>
<td>Asp</td>
<td>Ser</td>
</tr>
<tr>
<td>Ser</td>
<td>Asn</td>
<td>Leu</td>
<td>Thr</td>
<td>Ser</td>
</tr>
<tr>
<td>Pro</td>
<td>Ala</td>
<td>Cys</td>
<td>Pro</td>
<td>Glu</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AA</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu</td>
<td>Phe</td>
<td>Asn</td>
<td>Met</td>
<td>Pro</td>
</tr>
<tr>
<td>Asn</td>
<td>Val</td>
<td>Lys</td>
<td>Met</td>
<td>Gly</td>
</tr>
<tr>
<td>His</td>
<td>Lys</td>
<td>Val</td>
<td>Ala</td>
<td>Ile</td>
</tr>
<tr>
<td>Lys</td>
<td>Tyr</td>
<td>Trp</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td>Asp</td>
<td>Tyr</td>
<td>Gly</td>
<td>Ile</td>
<td>Tyr</td>
</tr>
<tr>
<td>Arg</td>
<td>Ala</td>
<td>Lys</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td>Asp</td>
<td>Tyr</td>
<td>Thr</td>
<td>Cys</td>
<td>Phe</td>
</tr>
<tr>
<td>Asp</td>
<td>His</td>
<td>Asn</td>
<td>Ala</td>
<td>Tyr</td>
</tr>
<tr>
<td>Ala</td>
<td>Met</td>
<td>Asp</td>
<td>Lys</td>
<td>Phe</td>
</tr>
<tr>
<td>Gly</td>
<td>Val</td>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
</tr>
</tbody>
</table>

52
355
PRT
hybrid
52
32/38

Pro Asn Asn Tyr Trp Gly Trp Gly Gly Glu Asp Asp Asp Ile Phe Asn
260 265

Arg Leu Val Phe Arg Gly Met Ser Ile Ser Arg Pro Asn Ala Val Val
275 280 285

Gly Arg Cys Arg Met Ile Arg His Ser Arg Asp Lys Lys Asn Glu Pro
290 295

Asn Pro Gln Arg Phe Asp Arg Ile Ala His Thr Lys Glu Thr Met Leu
305 310 315 320

Ser Asp Gly Leu Asn Ser Leu Thr Tyr Gln Val Leu Asp Val Gln Arg
325 330 335

Tyr Pro Leu Tyr Thr Gln Ile Thr Val Asp Ile Gly Thr Pro Ser Lys
340 345 350

Asp Glu Leu
355

<210> 53
<211> 1119
<212> DNA
<213> hybrid

<400> 53
atggcagatca agatggagac acattctcag gttttttag tttttttag acatgtggct gtggggtctc 60
gttgtaaagcagat tcgctgtgcc ctttttag ttttttagc tggcagat gcttttagc ttttttagc 120
gactatccttg aagccttgg tcttctttag gaatttaaat cactttgg ttttttagc ttttttagc 180
tatccttttag ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc 240
ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc 300
gattttttgg ctttttttag gattttttgg ctttttttag gattttttgg ctttttttag gattttttgg 360
gcttttttag gattttttgg ctttttttag gattttttgg ctttttttag gattttttgg ctttttttag 420
gcttttttag gattttttgg ctttttttag gattttttgg ctttttttag gattttttgg ctttttttag 480
ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc 540
ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc 600

ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc ttttttagc 660

aaagccgccc aatccctttc agacgcgtgc tggatatgcag cagatgtaaa ccttctttgg 720
tttttttgg ctttttttag ctttttttag ctttttttag ctttttttag ctttttttag ctttttttag 780
gattttttgg ctttttttag ctttttttag ctttttttag ctttttttag ctttttttag ctttttttag 840
tttttttgg ctttttttag ctttttttag ctttttttag ctttttttag ctttttttag ctttttttag 900
gttttttgg ctttttttag ctttttttag ctttttttag ctttttttag ctttttttag ctttttttag 960

atccttttag ttttttttag cctttttttt gttttttttt gttttttttt gttttttttt gttttttttt 1020
gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt 1080
gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt 1140

<210> 54
<211> 372
<212> PRT
<213> hybrid

<400> 54
Met Gly Ile Lys Met Glu Thr His Ser Gln Val Phe Val Tyr Met Leu
1 5 10 15

Leu Trp Leu Ser Gly Val Asp Met Gly Gln Met Pro Val Ala Ala Val
20 25 30

Val Val Met Ala Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys
35 40 45

Ser Val Leu Thr Tyr Gln Thr Pro Ile Ala Ser Lys Tyr Pro Leu Phe
50 55 60
Ile Ser Gln Asp Gly Ser Asp Gln Ala Val Lys Ser Lys Ser Leu Ser
65  70  75  80
Tyr Asn Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val Val
85  90  95
Thr Glu Arg Pro Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His
100 105 110
Tyr Lys Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Lys Phe Ser Arg
115 120 125
Val Ile Ile Leu Glu Asp Met Glu Ile Ala Pro Asp Phe Phe Asp
130 135 140
Tyr Phe Glu Ala Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met
145 150 155 160
Ala Ala Ser Ser Trp Asn Asp Gly Gln Lys Gln Phe Val His Asp
165 170 175
Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met
180 185 190
Leu Lys Arg Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala
195 200 205
Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Glu
210 215 220
Phe Ile Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly
225 230 235 240
Ser Ser Leu Gly Gln Phe Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu
245 250 255
Asn Asp Val Thr Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr
260 265 270
Glu Gly Asn Tyr Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg
275 280 285
Pro Ile Gln Gly Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp
290 295 300
Asp Val Arg Ile Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala
305 310 315 320
Gly Glu Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr
325 330 335
Ala Tyr Lys Gly Val Val Val Phe Arg Ile Gln Thr Thr Arg Arg Val
340 345 350
Phe Leu Val Gly Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser
355 360 365
Lys Asp Glu Leu
370

<210>  55
<211> 1158
<212> DNA
<213> hybrid
<400>  55
atgggcatca agatggagc acatttcag gtctttgtat acatgttgct gtggttgc
Met Gly Ile Lys Met Glu Thr His Ser Gln Val Phe Val Tyr Met Leu
Leu Trp Leu Ser Gly Val Asp Met Ala Leu Arg Leu His Arg Arg Asn
His Phe Ser Pro Arg Asn Thr Asp Leu Phe Pro Asp Leu Ala Lys Asp
Arg Val Val Ile Val Leu Tyr Val His Asn Arg Ala Gln Tyr Phe Arg
Val Thr Val Glu Ser Leu Ser Lys Val Lys Gly Ile Ser Glu Thr Leu
Leu Ile Val Ser His Asp Gly Tyr Phe Glu Glu Met Asn Arg Ile Val
Glu Ser Ile Lys Phe Cys Gln Val Lys Gln Ile Phe Ser Pro Tyr Ser
Pro His Ile Tyr Arg Thr Ser Phe Pro Gly Val Thr Leu Asn Asp Cys
Lys Asn Lys Gly Asp Glu Ala Lys Gly His Cys Gly Gly Asn Pro Asp
Gln Tyr Gly Asn His Arg Ser Pro Lys Ile Val Ser Leu Lys His His
Trp Trp Trp Met Met Asn Thr Val Trp Asp Gly Leu Glu Glu Thr Lys
Gly His Gly His Ile Leu Phe Ile Glu Gly Asp His Phe Phe Phe
Pro Asn Ala Tyr Arg Asn Ile Gln Thr Leu Thr Arg Leu Lys Pro Ala
Lys Cys Pro Asp Cys Phe Ala Ala Asn Leu Ala Pro Ser Asp Val Lys
<table>
<thead>
<tr>
<th>210</th>
<th>215</th>
<th>220</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Arg</td>
<td>Gly</td>
</tr>
<tr>
<td>225</td>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>Val</td>
<td>Gly</td>
<td>Tyr</td>
</tr>
<tr>
<td>245</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Ala</td>
<td>Arg</td>
<td>Glu</td>
</tr>
<tr>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>Trp</td>
<td>Ala</td>
<td>Thr</td>
</tr>
<tr>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>Gly</td>
<td>Pro</td>
<td>Arg</td>
</tr>
<tr>
<td>290</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Gly</td>
</tr>
<tr>
<td>305</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>Glu</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Val</td>
<td>Arg</td>
<td>Val</td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Trp</td>
<td>Gly</td>
<td>Gly</td>
</tr>
<tr>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Ala</td>
<td>Thr</td>
<td>Met</td>
</tr>
<tr>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
</tbody>
</table>

Leu
385

<210> 57
<211> 1152
<212> DNA
<213> Homo sapiens

<table>
<thead>
<tr>
<th>&lt;100&gt; 57</th>
</tr>
</thead>
<tbody>
<tr>
<td>atgctgaaga</td>
</tr>
<tr>
<td>aatgcctctg</td>
</tr>
<tr>
<td>agcgtctccg</td>
</tr>
<tr>
<td>gagctcgggg</td>
</tr>
<tr>
<td>cccggtgggc</td>
</tr>
<tr>
<td>tcgcttgcag</td>
</tr>
<tr>
<td>ctgctttgta</td>
</tr>
<tr>
<td>aagcagcaac</td>
</tr>
<tr>
<td>caacagttag</td>
</tr>
<tr>
<td>tatttttgg</td>
</tr>
<tr>
<td>cagggcgag</td>
</tr>
<tr>
<td>tgtggact</td>
</tr>
<tr>
<td>gaccatataq</td>
</tr>
<tr>
<td>ttttgattca</td>
</tr>
<tr>
<td>cattttttta</td>
</tr>
<tr>
<td>ggaggaggtg</td>
</tr>
<tr>
<td>ttggacgaa</td>
</tr>
<tr>
<td>taccaggtgc</td>
</tr>
<tr>
<td>acgccctgc</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Met Leu Lys Lys Gln Ser Ala Gly Leu Val Leu Trp Gly Ala Ile Leu</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>Phe Val Ala Trp Asn Ala Leu Leu Leu Leu Phe Phe Thr Arg Pro</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>Ala Pro Gly Pro Pro Ser Val Ser Ala Leu Asp Gly Asp Pro Ala</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>Ser Leu Thr Arg Gly Val Asp Met Gln Ser Ser Gly Leu Arg Thr</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>Gly Gly Ala Arg Pro Pro Pro Leu Gly Ala Ser Ser Gln Pro Arg</td>
</tr>
<tr>
<td>65</td>
</tr>
<tr>
<td>Pro Gly Gly Asp Ser Ser Pro Val Val Asp Ser Gly Pro Gly Pro Ala</td>
</tr>
<tr>
<td>85</td>
</tr>
<tr>
<td>Ser Asn Leu Thr Ser Val Pro Val Pro His Thr Thr Ala Leu Ser Leu</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>Pro Ala Cys Pro Glu Glu Ser Pro Leu Leu Val Gly Pro Met Leu Ile</td>
</tr>
<tr>
<td>115</td>
</tr>
<tr>
<td>Glu Phe Asn Met Pro Val Asp Leu Glu Leu Val Ala Lys Gln Asn Pro</td>
</tr>
<tr>
<td>130</td>
</tr>
<tr>
<td>Asn Val Lys Met Gln Gly Arg Tyr Ala Pro Arg Asp Cys Val Ser Pro</td>
</tr>
<tr>
<td>145</td>
</tr>
<tr>
<td>His Lys Val Ala Ile Ile Ile Pro Phe Arg Asn Arg Gln Glu His Leu</td>
</tr>
<tr>
<td>165</td>
</tr>
<tr>
<td>Lys Tyr Trp Leu Tyr Tyr Leu His Pro Val Leu Gln Arg Gln Gin Leu</td>
</tr>
<tr>
<td>180</td>
</tr>
<tr>
<td>Asp Tyr Gly Ile Tyr Val Ile Asn Glu Ala Gly Asp Thr Ile Phe Asn</td>
</tr>
<tr>
<td>195</td>
</tr>
<tr>
<td>Arg Ala Lys Leu Leu Asn Val Gly Phe Gin Glu Ala Leu Lys Asp Tyr</td>
</tr>
<tr>
<td>210</td>
</tr>
<tr>
<td>Asp Tyr Thr Cys Phe Val Phe Ser Asp Val Asp Leu Ile Pro Met Asn</td>
</tr>
<tr>
<td>225</td>
</tr>
<tr>
<td>Asp His Asn Ala Tyr Arg Cys Phe Ser Gin Pro Arg Gin His Ile Ser Val</td>
</tr>
<tr>
<td>245</td>
</tr>
<tr>
<td>Ala Met Asp Lys Phe Gin Phe Gin Leu Gin Gin Phe Val Gin Tyr Phe Gin</td>
</tr>
<tr>
<td>260</td>
</tr>
<tr>
<td>Gly Val Ser Ala Leu Ser Lys Gin Gin Phe Leu Thr Ile Gin Gin Gin</td>
</tr>
<tr>
<td>275</td>
</tr>
<tr>
<td>Pro Asn Gin Asp Tyr Trp Gly Trp Gly Gin Asp Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>290</td>
</tr>
<tr>
<td>Arg Leu Val Phe Arg Gin Met Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>305</td>
</tr>
<tr>
<td>Gly Gin Cys Gin Met Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>325</td>
</tr>
<tr>
<td>Asn Pro Gin Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>340</td>
</tr>
</tbody>
</table>
Ser Asp Gly Leu Asn Ser Leu Thr Tyr Gln Val Leu Asp Val Gln Arg
355 360 365

Tyr Pro Leu Tyr Thr Gln Ile Thr Val Asp Ile Gly Thr Pro Ser
370 375 380

<210> 59
<211> 400
<212> PRT
<213> Homo sapiens

<400> 59

Met Arg Leu Arg Glu Pro Leu Leu Ser Gly Ala Ala Met Pro Gly Ala
1 5 10 15

Ser Leu Gln Arg Ala Cys Arg Leu Leu Val Ala Val Cys Ala Leu His
20 25 30

Leu Gly Val Thr Leu Val Tyr Tyr Leu Ala Gly Arg Asp Leu Ser Arg
35 40 45

Leu Pro Gln Leu Val Gly Val Ser Thr Pro Leu Gln Gly Gly Ser Asn
50 55 60

Ser Ala Ala Ala Ile Gly Gln Ser Ser Gly Leu Arg Thr Gly Gly
65 70 75 80

Ala Arg Pro Pro Pro Pro Leu Gly Ala Ser Ser Gln Pro Arg Pro Gly
85 90 95

Gly Asp Ser Ser Pro Val Val Asp Ser Gly Pro Gly Pro Ala Ser Asn
100 105 110

Leu Thr Ser Val Pro Val Pro His Thr Thr Ala Leu Ser Leu Pro Ala
115 120 125

Cys Pro Glu Glu Ser Pro Leu Leu Val Gly Pro Met Leu Ile Glu Phe
130 135 140

Asn Met Pro Val Asp Leu Glu Leu Val Ala Lys Gln Asn Pro Asn Val
145 150 155 160

Lys Met Gly Gly Arg Tyr Ala Pro Arg Asp Cys Val Ser Pro His Lys
165 170 175

Val Ala Ile Ile Ile Pro Phe Arg Asn Arg Gln Glu His Leu Lys Tyr
180 185 190

Trp Leu Tyr Tyr Leu His Pro Val Leu Gln Arg Gln Leu Asp Tyr
195 200 205

Gly Ile Tyr Gly Ile Tyr Val Ile Asn Gln Ala Gly Asp Thr Ile Phe
210 215 220

Asn Arg Ala Lys Leu Leu Asn Val Gly Phe Gln Glu Ala Leu Lys Asp
225 230 235 240

Tyr Asp Tyr Thr Cys Phe Val Phe Ser Asp Val Asp Leu Ile Pro Met
245 250 255

Asn Asp His Asn Ala Tyr Arg Cys Phe Ser Gln Pro Arg His Ile Ser
260 265 270

Val Ala Met Asp Lys Phe Gly Phe Ser Leu Pro Tyr Val Gln Tyr Phe
275 280 285
Gly Gly Val Ser Ala Leu Ser Lys Gln Gln Phe Leu Thr Ile Asn Gly 290 295 300
Phe Pro Asn Asn Tyr Trp Gly Trp Gly Gly Glu Asp Asp Asp Ile Phe 305 310 315 320
Asn Arg Leu Val Phe Arg Gly Met Ser Ile Ser Arg Pro Asn Ala Val 325 330 335
Val Gly Arg Cys Arg Met Ile Arg His Ser Asp Lys Lys Asn Glu 340 345 350
Pro Asn Pro Gln Arg Phe Asp Arg Ile Ala His Thr Lys Glu Thr Met 355 360 365
Leu Ser Asp Gly Leu Asn Ser Leu Thr Tyr Gln Val Leu Asp Val Gln 370 375 380
Arg Tyr Pro Leu Tyr Thr Gln Ile Thr Val Asp Ile Gly Thr Pro Ser 385 390 395 400