Recombinant Production of Steviol Glycosides

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Application Serial No. 61/521,084, filed August 8, 2011; U.S. Application Serial No. 61/521,203, filed August 8, 2011; U.S. Application Serial No. 61/521,051, filed August 8, 2011; U.S. Application Serial No. 61/523,487, filed August 15, 2011; U.S. Application Serial No. 61/567,929, filed December 7, 2011; and U.S. Application Serial No. 61/603,639, filed February 27, 2012, all of which are incorporated by reference herein in their entirety.

TECHNICAL FIELD

This disclosure relates to the recombinant production of steviol glycosides. In particular, this disclosure relates to the production of steviol glycosides such as rebaudioside D by recombinant hosts such as recombinant microorganisms, plants, or plant cells. This disclosure also provides compositions containing steviol glycosides. The disclosure also relates to tools and methods for producing terpenoids by modulating the biosynthesis of terpenoid precursors of the squalene pathway.

BACKGROUND

Sweeteners are well known as ingredients used most commonly in the food, beverage, or confectionary industries. The sweetener can either be incorporated into a final food product during production or for stand-alone use, when appropriately diluted, as a tabletop sweetener or an at-home replacement for sugars in baking. Sweeteners include natural sweeteners such as sucrose, high fructose corn syrup, molasses, maple syrup, and honey and artificial sweeteners such as aspartame, saccharine and sucralose. Stevia extract is a natural sweetener that can be isolated and extracted from a perennial shrub, *Stevia rebaudiana*. Stevia is commonly grown in South America and Asia for commercial production of stevia extract. Stevia extract, purified to various degrees, is used commercially as a high intensity sweetener in foods and in blends or alone as a tabletop sweetener.

Extracts of the Stevia plant contain rebaudiosides and other steviol glycosides that contribute to the sweet flavor, although the amount of each glycoside often varies among different production batches. Existing commercial products are predominantly rebaudioside A with lesser amounts of other glycosides such as rebaudioside C, D, and F. Stevia extracts may also contain contaminants such as plant-derived compounds that contribute to off-flavors. These off-flavors can be more or less problematic depending on the food system or application of choice. Potential contaminants include pigments, lipids, proteins, phenolics, saccharides, spathulenol and other sesquiterpenes, labdane diterpenes, monoterpenes, decanoic acid, 8,11,14-eicosatrienoic acid, 2-methyloctadecane, pentacosane, octacosane, tetracosane, octadecanol, stigmasterol, β-sitosterol, α- and β-amyrin, lupeol, β-amryin acetate, pentacyclic triterpenes, centauredin, quercitin, epi-alpha-cadinol, carophyllenes and derivatives, beta-pinene, beta-sitosterol, and gibberellin.

SUMMARY

Provided herein is a recombinant host, such as a microorganism, plant, or plant cell, comprising one or more biosynthesis genes whose expression results in production of steviol glycosides such as rebaudioside A, rebaudioside C, rebaudioside D, rebaudioside E, rebaudioside F, or dulcoside A. In particular, EUGT11, a uridine 5'-diphospho (UDP) glycosyl transferase described herein, can be used alone or in combination with one or more other UDP glycosyl transferases such as UGT74G1, UGT76G1, UGT85C2, and UGT91D2e, to allow the production and accumulation of rebaudioside D in recombinant hosts or using *in vitro* systems. As described herein, EUGT11 has a strong 1,2-19-O-glucose glycosylation activity, which is an important step for rebaudioside D production.

Typically, stevioside and rebaudioside A are the primary compounds in commercially-produced stevia extracts. Stevioside is reported to have a more bitter and less sweet taste than rebaudioside A. The composition of stevia extract can vary from lot to lot depending on the soil and climate in which the plants are grown. Depending upon the sourced plant, the climate conditions, and the extraction process, the amount of

rebaudioside A in commercial preparations is reported to vary from 20 to 97% of the total steviol glycoside content. Other steviol glycosides are present in varying amounts in stevia extracts. For example, Rebaudioside B is typically present at less than 1-2%, whereas Rebaudioside C can be present at levels as high as 7-15%. Rebaudioside D is typically present in levels of 2% or less, and Rebaudioside F is typically present in compositions at 3.5% or less of the total steviol glycosides. The amount of the minor steviol glycosides affects the flavor profile of a Stevia extract. In addition, Rebaudioside D and other higher glycosylated steviol glycosides are thought to be higher quality sweeteners than Rebaudioside A. As such, the recombinant hosts and methods described herein are particularly useful for producing steviol glycoside compositions having an increased amount of Rebaudioside D for use, for example, as a non-caloric sweetener with functional and sensory properties superior to those of many high-potency sweeteners.

In one aspect, this document features a recombinant host that includes a recombinant gene encoding a polypeptide having at least 80% identity to the amino acid sequence set forth in SEQ ID NO:152.

This document also features a recombinant host that includes a recombinant gene encoding a polypeptide having the ability to transfer a second sugar moiety to the C-2' of a 19-O-glucose of rubusoside. This document also features a recombinant host that includes a recombinant gene encoding a polypeptide having the ability to transfer a second sugar moiety to the C-2' of a 19-O-glucose of stevioside.

In another aspect, this document features a recombinant host that includes a recombinant gene encoding a polypeptide having the ability to transfer a second sugar moiety to the C-2' of the 19-O-glucose of rubusoside and to the C-2' of the 13-O-glucose of rubusoside.

This document also features a recombinant host that includes a recombinant gene encoding a polypeptide having the ability to transfer a second sugar moiety to the C-2' of a 19-O-glucose of rebaudioside A to produce rebaudioside D, wherein the catalysis rate of the polypeptide is at least 20 times faster (e.g., 25 or 30 times faster) than a 91D2e

polypeptide having the amino acid sequence set forth in SEQ ID NO: 5 when the reactions are performed under corresponding conditions.

In any of the recombinant hosts described herein, the polypeptide can have at least 85% sequence identity (e.g., 90%, 95%, 98%, or 99% sequence identity) to the amino acid sequence set forth in SEQ ID NO:152. The polypeptide can have the amino acid sequence set forth in SEQ ID NO: 152.

Any of the hosts described herein further can include a recombinant gene encoding a UGT85C polypeptide having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:3. The UGT85C polypeptide can include one or more amino acid substitutions at residues 9, 10, 13, 15, 21, 27, 60, 65, 71, 87, 91, 220, 243, 270, 289, 298, 334, 336, 350, 368, 389, 394, 397, 418, 420, 440, 441, 444, and 471 of SEQ ID NO:3.

Any of the hosts described herein further can include a recombinant gene encoding a UGT76G polypeptide having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:7. The UGT76G polypeptide can have one or more amino acid substitutions at residues 29, 74, 87, 91, 116, 123, 125, 126, 130, 145, 192, 193, 194, 196, 198, 199, 200, 203, 204, 205, 206, 207, 208, 266, 273, 274, 284, 285, 291, 330, 331, and 346 of SEQ ID NO:7.

Any of the hosts described herein further can include a gene (e.g., a recombinant gene) encoding a UGT74G1 polypeptide.

Any of the hosts described herein further can include a gene (e.g., a recombinant gene) encoding a functional UGT91D2 polypeptide. The UGT91D2 polypeptide can have at least 80% sequence identity to the amino acid sequence set forth in SEQ ID NO:5. The UGT91D2 polypeptide can have a mutation at position 206, 207, or 343 of SEQ ID NO:5. The UGT91D2 polypeptide also can have a mutation at positions 211 and 286 of SEQ ID NO:5 (e.g., L211M and V286A, referred to as UGT91D2e-b). The UGT91D2 polypeptide can have the amino acid sequence set forth in SEQ ID NOs: 5, 10, 12, 76, 78, or 95.

Any of the hosts described herein further can include one or more of

(i) a gene encoding a geranylgeranyl diphosphate synthase;

(ii) a gene encoding a bifunctional copalyl diphosphate synthase and kaurene synthase, or a gene encoding a copalyl diphosphate synthase and a gene encoding a kaurene synthase;

(iii) a gene encoding a kaurene oxidase; and

(iv) a gene encoding a steviol synthetase. Each of the genes of (i), (ii), (iii), and (iv) can be a recombinant gene.

Any of the hosts described herein further can include one or more of

(v) a gene encoding a truncated HMG-CoA;

(vi) a gene encoding a CPR;

(vii) a gene encoding a rhamnose synthetase;

(viii) a gene encoding a UDP-glucose dehydrogenase; and

(ix) a gene encoding a UDP-glucuronic acid decarboxylase. At least one of the genes of (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), or (ix) can be a recombinant gene.

The geranylgeranyl diphosphate synthase can have greater than 90 % sequence identity to one of the amino acid sequences set forth in SEQ ID NOs: 121-128. The copalyl diphosphate synthase can have greater than 90 % sequence identity to the amino acid sequence set forth in SEQ ID NOs: 129-131. The kaurene synthase can have greater than 90 % sequence identity to one of the amino acid sequences set forth in SEQ ID NOs: 132-135. The kaurene oxidase can have greater than 90 % sequence identity to one of the amino acid sequences set forth in SEQ ID NOs: 138-141. The steviol synthetase can have greater than 90 % sequence identity to one of the amino acid sequences set forth in SEQ ID NOs: 142-146.

Any of the recombinant hosts can produce at least one steviol glycoside when cultured under conditions in which each of the genes is expressed. The steviol glycoside can be selected from the group consisting of rubusoside, rebaudioside A, rebaudioside B, rebaudioside C, rebaudioside D, rebaudioside E, rebaudioside F, dulcoside A, stevioside, steviol-19-*O*-Glucoside, steviol-13-*O*-glucoside, steviol-1,2-bioside, steviol-1,3-bioside, 1,3-stevioside, as well as other rhamnosylated or xylosylated intermediates. The steviol glycoside (e.g., rebaudioside D) can accumulate to at least 1 mg/liter (e.g., at least 10 mg/liter, 20 mg/liter, 100 mg/liter, 200 mg/liter, 300 mg/liter, 400 mg/liter, 500 mg/liter,

600 mg/liter, or 700 mg/liter, or greater) of culture medium when cultured under said conditions.

This document also features a method of producing a steviol glycoside. The method includes growing any of the hosts described herein in a culture medium, under conditions in which the genes are expressed; and recovering the steviol glycoside produced by the host. The growing step can include inducing expression of one or more of the genes. The steviol glycoside can be a 13-*O*-1,2-diglycosylated and/or a 19-*O*-1,2-diglycosylated steviol glycoside (e.g., stevioside, steviol 1,2 bioside, rebaudioside D, or rebaudioside E). For example, the steviol glycoside can be rebaudioside D or rebaudioside E. Other examples of steviol glycosides can include rebaudioside A, rebaudioside B, rebaudioside C, rebaudioside F, and dulcoside A.

This document also features a recombinant host. The host includes (i) a gene encoding a UGT74G1; (ii) a gene encoding a UGT85C2; (iii) a gene encoding a UGT76G1; (iv) a gene encoding a glycosyltransferase having the ability to transfer a second sugar moiety to the C-2' of a 19-O-glucose of rubusoside or stevioside; and (v) optionally a gene encoding a UGT91D2e, wherein at least one of the genes is a recombinant gene. In some embodiments, each of the genes is a recombinant gene. The host can produce at least one steviol glycoside (e.g., rebaudioside D) when cultured under conditions in which each of the genes (e.g., recombinant genes) is expressed. The host further can include (a) a gene encoding a bifunctional copalyl diphosphate synthase and kaurene synthase, or a gene encoding a copalyl diphosphate synthase and a gene encoding a kaurene synthase; (b) a gene encoding a kaurene oxidase; (c) a gene encoding a steviol synthetase; (d) a gene encoding a geranylgeranyl diphosphate synthase.

This document also features a steviol glycoside composition produced by any of the hosts described herein. The composition has reduced levels of stevia plant-derived contaminants relative to a stevia extract.

In another aspect, this document features a steviol glycoside composition produced by any of the hosts described herein. The composition has a steviol glycoside composition enriched for rebaudioside D relative to the steviol glycoside composition of a wild-type Stevia plant.

In yet another aspect, this document features a method of producing a steviol glycoside composition. The method includes growing a host described herein in a culture medium, under conditions in which each of the genes is expressed; and recovering the steviol glycoside composition produced by the host (e.g., a microorganism). The composition is enriched for rebaudioside A, rebaudioside B, rebaudioside C, rebaudioside D, rebaudioside E, rebaudioside F or dulcoside A relative to the steviol glycoside composition of a wild-type Stevia plant. The steviol glycoside composition produced by the host (e.g., microorganism) can have a reduced level of stevia plant-derived contaminants relative to a stevia extract.

This document also features a method for transferring a second sugar moiety to the C-2' of a 19-*O*-glucose or the C-2' of a 13-*O*-glucose in a steviol glycoside. The method includes contacting the steviol glycoside with a EUGT11 polypeptide described herein or UGT91D2 polypeptide described herein (e.g., UGT91D2e-b) and a UDP-sugar under suitable reaction conditions for the transfer of the second sugar moiety to the steviol glycoside. The steviol glycoside can be rubusoside, wherein the second sugar moiety is glucose, and stevioside is produced upon transfer of the second glucose moiety. The steviol glycoside can be stevioside, wherein the second sugar moiety is glucose, and Rebaudioside E is produced upon transfer of the second glucose moiety. The steviol glycoside can be Rebaudioside A, and Rebaudioside D is produced upon transfer of the second glucose moiety.

In another embodiment of an improved downstream steviol glycoside pathway as disclosed herein, materials and methods are provided for the recombinant production of sucrose synthase, and to materials and methods for increasing production of UDP-glucose in a host, specifically for increasing the availability of UDP-glucose *in vivo*, with the purpose of promoting glycosylation reactions in the cells, and methods for reducing UDP concentrations in the cells are provided.

The document also provides a recombinant host comprising one or more exogenous nucleic acids encoding a sucrose transporter and a sucrose synthase, wherein expression of the one or more exogenous nucleic acids with a glucosyltransferase results in increased levels of UDP-glucose in the host. Optionally, the one or more exogenous

nucleic acids comprise a SUS1 sequence. Optionally, the SUS1 sequence is from *Coffea arabic*a, or encodes a functional homolog of the sucrose synthase encoded by the *Coffea arabica* SUS1 sequence, but equally an *Arabidopsis thaliana* or *Stevia rebaudiana* SUS may be used as described herein. In the recombinant host of the invention, the one or more exogenous nucleic acids may comprise a sequence encoding a polypeptide having the sequence set forth in SEQ ID NO:180, or an amino acid sequence at least 90 percent identical thereto, and optionally the one or more exogenous nucleic acids comprise a SUC1 sequence. In one embodiment, the SUC1 sequence is from *Arabidopsis thaliana*, or the SUC1 sequence encodes a functional homolog of the sucrose transporter encoded by the *Arabidopsis thaliana* SUC1 sequence. In the recombinant host, the one or more exogenous nucleic acids may comprise a sequence encoding a polypeptide having the sequence set forth in SEQ ID NO:179, or an amino acid sequence at least 90 percent identical thereto. The recombinant host has reduced ability to degrade external sucrose, as compared to a corresponding host that lacks the one or more exogenous nucleic acids.

The recombinant host may be a microorganism, such as a Saccharomycete, for example *Saccharomyces cerevisiae*. Alternatively, the microorganism is *Escherichia coli*. In an alternative embodiment, the recominbant host is a plant or plant cell.

The invention also provides a method for increasing the level of UDP-glucose and reducing the level of UDP in a cell, the method comprising expressing in the cell a recombinant sucrose synthase sequence and a recombinant sucrose transporter sequence, in a medium comprising sucrose, wherein the cell is deficient in sucrose degradation.

The invention additionally provides a method for promoting a glycosylation reaction in a cell, comprising expressing in the cell a recombinant sucrose synthase sequence and a recombinant sucrose transporter sequence, in a medium comprising sucrose, wherein the expressing results in a decreased level of UDP in the cell and an increased level of UDP-glucose in the cell, such that glycosylation in the cell is increased.

In either method for increasing the level of UDP-glucose or promoting glycosylation, the cell may produce vanillin glucoside, resulting in increased production of vanillin glucoside by the cell, or may produce steviol glucoside, resulting in increased production of steviol glucoside by the cell. Optionally, the SUS1 sequence is a *A.*

thaliana, S. rebaudiana, or *Coffea arabica* SUS1 sequence (see e.g., FIG. 17, SEQ ID NOs. 175-177), or is a sequence that encodes a functional homolog of the sucrose synthase encoded by the *A. thaliana, S. rebaudiana,* or *Coffea arabica* SUS1 sequence. The recombinant sucrose synthase sequence optionally comprises a nucleic acid encoding a polypeptide having the sequence set forth in SEQ ID NO:180, or an amino acid sequence at least 90% identical thereto, wherein optionally the recombinant sucrose transporter sequence is a SUC1 sequence, or wherein optionally the SUC1 sequence is an *Arabidopsis thaliana* SUC1 sequence, or is a sequence that encodes a functional homolog of the sucrose transporter encoded by the *Arabidopsis thaliana* SUC1 sequence, or wherein optionally the recombinant sucrose transporter sequence comprises a nucleic acid encoding a polypeptide having the sequence set forth in SEQ ID NO:179, or an amino acid sequence at least 90 % identical thereto. In either method, the host is a microorganism, for example a Saccharomycete, optionally such as *Saccharomyces cerevisiae*. Or the host may be *Escherichia coli*. Or the host may be a plant cell.

Also provided herein is a recombinant host, such as a microorganism, comprising one or more biosynthesis genes whose expression results in production of diterpenoids. Such genes include a gene encoding an *ent*-copalyl diphosphate synthase (CDPS) (EC 5.5.1.13), a gene encoding an *ent*-kaurene synthase, a gene encoding an *ent*-kaurene oxidase; or a gene encoding a steviol synthetase. At least one of the genes is a recombinant gene. The host can also be a plant cell. Expression of these gene(s) in a Stevia plant can result in increased steviol glycoside levels in the plant. In some embodiments the recombinant host further comprises a plurality of copies of a recombinant gene encoding a CDPS polypeptide (EC 5.5.1.13) lacking a chloroplast transit peptide sequence. The CDPS polypeptide can have at least 90%, 95%, 99 %, or 100% identity to the truncated CDPS amino acid sequence set forth in FIG. 14. The host can further comprise a plurality of copies of a recombinant gene encoding a KAH polypeptide, e.g., a KAH polypeptide that has at least 90%, 95%, 99 %, or 100% identity to the KAH amino acid sequence set forth in FIG. 12. The host can further comprise one or more of: (i) a gene encoding a geranylgeranyl diphosphate synthase; (ii) a gene encoding a *ent*-kaurene oxidase; and (iii) a gene encoding a *ent*-kaurene synthase. The

host can further comprise one or more of (iv) a gene encoding a truncated HMG-CoA; (v) a gene encoding a CPR; (vi) a gene encoding a rhamnose synthetase; (vii) a gene encoding a UDP-glucose dehydrogenase; and (viii) a gene encoding a UDP-glucuronic acid decarboxylase. Two or more exogenous CPRs can be present, for example. The expression of one or more of such genes can be inducible. At least one of genes (i), (ii), (iii), (iv), (v), (vi), (vii), or (viii) can be a recombinant gene, and in some cases each of the genes of (i), (ii), (iii), (iv), (v), (vi), (vii), and (viii) is a recombinant gene. The geranylgeranyl diphosphate synthase can have greater than 90 % sequence identity to the amino acid sequence set forth in SEQ ID NO:127; the kaurene oxidase can have greater than 90 % sequence identity to the amino acid sequence set forth in SEQ ID NO:138; a CPR can have greater than 90 % sequence identity to the amino acid sequence set forth in SEQ ID NO:168; a CPR can have greater than 90 % sequence identity to the amino acid sequence set forth in SEQ ID NO:170, and a kaurene synthase can have greater than 90 % sequence identity to the amino acid sequence set forth in SEQ ID NO: 156.

In one aspect, this document features an isolated nucleic acid encoding a polypeptide having the amino acid sequence set forth in SEQ ID NO:5, wherein the polypeptide contains substitutions position 211and 286 of SEQ ID NO:5. For example, the polypeptide can include a methionine at position 211 and an alanine at position 286.

In one aspect, this document features an isolated nucleic acid encoding a polypeptide having at least 80% identity (e.g., at least 85%, 90%, 95%, or 99% identity) to the amino acid sequence set forth in FIG. 12C (SEQ ID NO:164). The polypeptide can have the amino acid sequence set forth in FIG. 12C.

In another aspect, this document features a nucleic acid construct that included a regulatory region operably linked to a nucleic acid encoding a polypeptide having at least 80% identity(e.g., at least 85%, 90%, 95%, or 99% identity) to the amino acid sequence set forth in FIG. 12C (SEQ ID NO:164). The polypeptide can have the amino acid sequence set forth in FIG. 12C.

This document also features a recombinant host that includes a recombinant gene (e.g., a plurality of copies of a recombinant gene) encoding a KAH polypeptide having at least 80% identity (e.g., at least 85%, 90%, 95%, or 99% identity) to the amino acid

sequence set forth in FIG. 12C. The polypeptide can have the amino acid sequence set forth in FIG. 12C. The host can be a microorganism such as a saccharomycete (e.g., *Saccharomyces cerevisiae*) or *Escherichia coli*. The host can be a plant or plant cell (e.g., a *Stevia*, *Physcomitrella*, or tobacco plant or plant cell). The *Stevia* plant or plant cell is a *Stevia rebaudiana* plant or plant cell. The recombinant host can produce steviol when cultured under conditions in which each of the genes is expressed. The recombinant host can further comprise a gene encoding a UGT74G1 polypeptide; a gene encoding a UGT85C2 polypeptide; a gene encoding a UGT76G1 polypeptide; a gene encoding a UGT91D2 polypeptide; and/or a gene encoding a EUGT11 polypeptide. Such a host can produce at least one steviol glycoside when cultured under conditions in which each of the genes is expressed. The steviol glycoside can be steviol-13-O-glucoside, steviol-19-O-glucoside, rubusoside, rebaudioside A, rebaudioside B, rebaudioside C, rebaudioside D, rebaudioside E, rebaudioside F, and/or dulcoside A. The recombinant host can further comprise one or more of: a gene encoding a deoxyxylulose 5-phosphate synthase (DXS); a gene encoding a D-1-deoxyxylulose 5-phosphate reductoisomerase (DXR); a gene encoding a 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (CMS); a gene encoding a 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK); a gene encoding a 4-diphosphocytidyl-2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS); a gene encoding a 1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate synthase (HDS); and a gene encoding a 1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate reductase (HDR). The recombinant host can further comprise one or more of: a gene encoding an acetoacetyl-CoA thiolase; a gene encoding a truncated HMG-CoA reductase; a gene encoding a mevalonate kinase; a gene encoding a phosphomevalonate kinase; and a gene encoding a mevalonate pyrophosphate decarboxylase. In another aspect, this document features a recombinant host that further comprises a gene encoding an *ent*-kaurene synthase (EC 4.2.3.19) and/or a gene encoding a gibberellin 20-oxidase (EC 1.14.11.12). Such a host produces gibberellin GA3 when cultured under conditions in which each of the genes is expressed.

This document also features an isolated nucleic acid encoding a CPR polypeptide having at least 80% sequence identity (e.g., at least 85%, 90%, 95%, or 99% sequence

identity) to the *S. rebaudiana* CPR amino acid sequence set forth in FIG. 13. In some embodiments, the polypeptide has the *S. rebaudiana* CPR amino acid sequence set forth in FIG.13 (SEQ ID NOs: 169 and 170).

In any of the hosts described herein, expression of one or more of the genes can be inducible.

In any of the hosts described herein, one or more genes encoding endogenous phosphatases can be deleted or disrupted such that endogenous phosphatase activity is reduced. For example, the yeast gene *DPP1* and/or *LPP1* can be disrupted or deleted such that the degradation of farnesyl pyrophosphate (FPP) to farnesol is reduced and the degradation of geranylgeranylpyrophosphate (GGPP)) to geranylgeraniol (GGOH) is reduced.

In another aspect, as described herein, *ERG9* can be modified as defined below, resulting in the decreased production of squalene synthase (SQS) and an accumulation of terpenoid precursors. The precursors may or may not be secreted into the culture medium and can in turn be used as substrates to enzymes capable of metabolizing the terpenoid precursors into desired terpenoids.

Thus, in a main aspect the present invention relates to a cell comprising a nucleic acid sequence, said nucleic acid comprising

i) a promoter sequence operably linked to

ii) a heterologous insert sequence operably linked to

iii) an open reading frame operably linked to

iv) a transcription termination signal,

wherein the heterologous insert sequence has the general formula (I):

$$-X_1-X_2-X_3-X_4-X_5-$$

wherein X_2 comprises at least 4 consecutive nucleotides being complementary to, and forming a hairpin secondary structure element with at least 4 consecutive nucleotides of X_4, and

wherein X_3 is optional and if present comprises unpaired nucleotidesinvolved in forming a hairpin loop between X2 and X4, and

wherein X_1 and X_5 individually and optionally comprises one or more nucleotides, and

wherein the open reading frame upon expression encodes a polypeptide sequence having at least 70% identity to a squalene synthase (EC 2.5.1.21) or a biologically active fragment thereof, said fragment having at least 70% sequence identity to said squalene synthase in a range of overlap of at least 100 amino acids.

The cell of the present invention is useful in enhancing yield of industrially interesting terpenoids. Accordingly, in another aspect the present invention relates to a method for producing a terpenoid compound synthesized through the squalene pathway, in a cell culture, said method comprising the steps of

(a) providing the cell as defined herein above,

(b) culturing the cell of (a).

(c) recovering the terpenoid product compound.

By providing the cell comprising the genetically modified construct defined herein above, the accumulation of terpenoid precursors is enhanced (see e.g., FIG. 20).

Thus, in another aspect, the invention relates to a method for producing a terpenoid derived from a terpenoid precursor selected from the group consisting of Farnesyl-pyrophosphate (FPP), Isopentenyl-pyrophosphate (IPP), Dimethylallyl-pyrophosphate (DMAPP), Geranyl-pyrophosphate (GPP) and/or Geranylgeranyl-pyrophosphate (GGPP), said method comprising:

(a) contacting said precursor with an enzyme of the squalene synthase pathway,

(b) recovering the terpenoid product.

The present invention may operate by at least partly, sterically hindering binding of the ribosome to the RNA thus reducing the translation of squalene synthase. Accordingly, in one aspect the present invention relates to a method for reducing the translation rate of a functional squalene synthase (EC 2.5.1.21) said method comprising:

(a) providing the cell defined herein above,

(b) culturing the cell of (a).

Similarly, the invention in another aspect relates to a method for decreasing turnover of farnesyl-pp to squalene, said method comprising:

(a) providing the cell defined herein above,

(b) culturing the cell of (a).

As depicted in FIG. 20, the knocking down of the ERG9 results in build-up of precursors to squalene synthase. Thus in one aspect, the present invention relates to a method for enhancing accumulation of a compound selected from the group consisting of Farnesyl-pyrophosphate, Isopentenyl-pyrophosphate, Dimethylallyl-pyrophosphate, Geranyl-pyrophosphate and Geranylgeranyl-pyrophosphate, said method comprising the steps of:

(a) providing the cell defined herein above, and

(b) culturing the cell of (a).

In one embodiment the invention relates to the production of Geranylgeranyl Pyrophosphate (GGPP) as well as other terpenoids, which can be prepared from Geranylgeranyl Pyrophosphate (GGPP).

In this embodiment of the invention the above described decrease of production of squalene synthase (SQS) may be combined with an increase in activity of Geranylgeranyl Pyrophosphate Synthase (GGPPS), which converts FPP to Geranylgeranyl Pyrophosphate (GGPP), leading to increased production of GGPP.

Thus, in one embodiment the invention relates to a microbial cell comprising a nucleic acid sequence, said nucleic acid comprising

i) a promoter sequence operably linked to

ii) a heterologous insert sequence operably linked to

iii) an open reading frame operably linked to

iv) a transcription termination signal,

wherein the heterologous insert sequence and the open reading frame are as defined herein above,

wherein said microbial cell furthermore comprises a heterologous nucleic acid encoding GGPPS operably linked to a nucleic acid sequence directing expression of GGPPS in said cell.

In addition, the document relates to a method for producing steviol or a steviol glycoside, wherein the method comprises use of any one of the above-mentioned microbial cells.

Any of the hosts described herein can be a microorganism (e.g., a Saccharomycete such as *Saccharomyces cerevisiae*, or *Escherichia coli*), or a plant or plant cell (e.g., a *Stevia* such as a *Stevia rebaudiana*, *Physcomitrella*, or tobacco plant or plant cell).

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention belongs. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting. Other features and advantages of the invention will be apparent from the following detailed description. Applicants reserve the right to alternatively claim any disclosed invention using the transitional phrase "comprising," "consisting essentially of," or "consisting of," according to standard practice in patent law.

DESCRIPTION OF DRAWINGS

FIG. 1 is the chemical structure of various steviol glycosides.

FIGS. 2A- D show representative pathways for the biosynthesis of steviol glycosides from steviol.

FIG. 3 is a schematic representation of 19-O-1,2-diglycosylation reactions by EUGT11 and UGT91D2e. The numbers are the average signal intensity for the substrates

or the products of the reaction, from liquid chromatography-mass spectrometry (LC-MS) chromatograms

FIG. 4 contains LC-MS chromatograms showing the production of rebaudioside D (RebD) from Rebaudioside A (RebA) using *in vitro* transcribed and translated UGT91D2e (SEQ ID NO:5) (left panel) or EUGT11 (SEQ ID NO:152) (right panel). The LC-MS was set to detect certain masses corresponding to steviol + 5 glucoses (such as RebD), steviol + 4 glucoses (such as RebA) etc. Each ´lane´ is scaled according to the highest peak.

FIG. 5 contains LC-MS chromatograms showing the conversion of rubusoside to stevioside and compounds ´2´ and ´3´ (RebE) by UGT91D2e (left panel) and EUGT11 (right panel).

FIG. 6 is an alignment of the amino acid sequence of EUGT11 (SEQ ID NO:152, top line) with the amino acid sequence of UGT91D2e (SEQ ID NO:5, bottom line).

FIG. 7 contains the amino acid sequence of EUGT11 (SEQ ID NO:152), the nucleotide sequence (SEQ ID NO:153) encoding EUGT11, and the nucleotide sequence encoding EUGT11 that has been codon optimized for expression in yeast (SEQ ID NO: 154).

FIG. 8 is an alignment of the secondary structure predictions of UGT91D2e with UGT85H2 and UGT71G1. Secondary structure predictions were made by subjecting the amino acid sequences of the three UGTs to NetSurfP ver. 1.1 - Protein Surface Accessibility and Secondary Structure Predictions, at the world wide web at cbs.dtu.dk/services/NetSurfP/. This predicted the presence and location of alpha helices, beta sheets and coils in the proteins. These were subsequently labeled as shown for UGT91D2e. For example, the first N-terminal beta-sheet was labeled Nβ1. The y-axis represents the certainty of the prediction, the higher the more confident and the x-axis represents amino acid position. Although the primary sequence identity between these UGTs is very low, the secondary structures show a very high degree of conservation.

FIG. 9 is an alignment of the amino acid sequences of UGT91D1 and UGT91D2e (SEQ ID NO: 5).

FIG. 10 is a bar graph of the activity of double amino acid substitution mutants of UGT91D2e. The filled bars represent stevioside production and the open bars represent 1,2-bioside production.

FIG. 11 is a schematic representation of UDP-glucose regeneration for the biosynthesis of steviol glycosides. SUS = sucrose synthase; Steviol = steviol or steviol glycoside substrate; UGT = UDP glycosyl transferase.

FIG. 12A is the nucleotide sequence encoding the *Stevia rebaudiana* KAH (SEQ ID NO:163), designated SrKAHe1 herein.

FIG. 12B is the nucleotide sequence encoding the *Stevia rebaudiana* KAHe1 that has been codon-optimized for expression in yeast (SEQ ID NO:165).

FIG. 12C is the amino acid sequence of the *Stevia rebaudiana* KAHe1 (SEQ ID NO:164).

FIG. 13A contains the amino acid sequences of CPR polypeptides from *S. cerevisiae* (encoded by NCP1 gene) (SEQ ID NO:166), *A. thaliana* (encoded by ATR1 and encoded by ATR2) ((SEQ ID NOs: 148 and 168), and *S. rebaudiana* (encoded by CPR7 and encoded by CPR8) (SEQ ID NOs: 169 and 170).

FIG. 13B contains ATR1 nucleotide sequence (Accession No. CAA23011) that has been codon optimized for expression in yeast (SEQ ID NO:171); ATR2 nucleotide sequence that has been codon optimized for expression in yeast (SEQ ID NO:172); the *Stevia rebaudiana* CPR7 nucleotide sequence (SEQ ID NO:173); and the *Stevia rebaudiana* CPR8 nucleotide sequence (SEQ ID NO:174).

FIG. 14A contains the nucleotide sequence (SEQ ID NO:157) encoding a CDPS polypeptide (SEQ ID NO:158) from *Zea mays*. The sequence that is in bold and underlined can be deleted to remove the sequence encoding the chloroplast transit sequence.

FIG. 14B contains the amino acid sequence of the CDPS polypeptide (SEQ ID NO:158) from *Zea mays*. The sequence that is in bold and underlined can be deleted to remove the chloroplast transit sequence.

FIG. 15A contains a codon-optimized nucleotide sequence (SEQ ID NO:161) encoding a bifunctional CDPS-KS polypeptide (SEQ ID NO:162) from *Gibberella*

fujikuroi. FIG. 15B contains the amino acid sequence of the bifunctional CDPS-KS polypeptide (SEQ ID NO:162) from *Gibberella fujikuroi*.

FIG. 16 is a graph of the growth of two strains of *S. cerevisiae,* enhanced EFSC1972 (designated T2) and enhanced EFSC1972 with further overexpression of the Arabidopsis thaliana kaurene synthase (KS-5) (designated T7, squares). Numbers on the y-axis are OD600 values of the cell culture, while numbers on the x-axis represent hours of growth in synthetic complete based medium at 30°C.

FIG. 17 contains the nucleic acid sequences encoding the *A. thaliana*, *S. rebaudiana* (from contig10573 selection_ORF S11E, with the mutation that changes S11 to glutamate (E) in bold, lowercase letters), and coffee (*Coffea arabica*) sucrose synthases, SEQ ID NOs:175, 176, and 177, respectively.

FIG. 18 is a bar graph of rebD production in permeabilized *S. cerevisiae*, which had been transformed with EUGT11 or an empty plasmid ("Empty"). Cells were grown to exponential growth phase, washed in PBS buffer and subsequently treated with Triton X-100 (0.3% or 0.5% in PBS) 30°C, 30 min. After permeabilization cells were washed in PBS and resuspended in reaction mix containing 100 μM RebA and 300 μM UDP-glucose. Reactions proceeded for 20h, 30°C.

FIG. 19A is the amino acid sequence of the *A. thaliana* UDP-glycosyltransferase UGT72E2 (SEQ ID NO:178).

FIG. 19B is the amino acid sequence of the sucrose transporter SUC1 from *A. thaliana* (SEQ ID NO:179).

FIG. 19C is the amino acid sequence of the sucrose synthase from coffee (SEQ ID NO:180).

FIG. 20 is a schematic of the isoprenoid pathway in yeast, showing the position of ERG9.

FIG. 21 contains the nucleotide sequence of the *Saccharomyces cerevisiae* Cyc1 promoter (SEQ ID NO:185) and *Saccharomyces cerevisiae* Kex2 promoter (SEQ ID NO:186).

FIG. 22 is a schematic of the PCR product containing two regions, HR1 and HR2, which are homologous to parts of the genome sequence within the ERG9 promoter or 5'

end of the ERG9 open reading frame (ORF), respectively. Also, on the PCR product is an antibiotic marker, NatR, which can be embedded between two Lox sites (L) for subsequent excision with Cre recombinase. The PCR product further can include a promoter, such as either the wild type ScKex2, wild type ScCyc1, and the promoter further can include a heterologous insert such as a hairpin (SEQ ID NO: 181-184) at its 3'-end (See FIG. 23).

FIG. 23 is a schematic of promoter and ORF with a hairpin stemloop immediately upstream of the translation startsite (arrow) and an alignment of a portion of the wild-type *S. cerevisiae* Cyc1 promoter sequence and initial ATG of the ERG9 OPR without a heterologous insert (SEQ ID NO:187) and with four different heterologous inserts (SEQ ID NOs. 188-191). 75% refers to construct comprising the ScCyc1 promoter followed by SEQ ID NO: 184 (SEQ ID NO:191); 50% refers to construct comprising the ScCyc1 promoter followed by SEQ ID NO: 183 (SEQ ID NO:190); 20% refers to construct comprising the ScCyc1 promoter followed by SEQ ID NO: 182 (SEQ ID NO:189); 5% refers to construct comprising the ScCyc1 promoter followed by SEQ ID NO: 181 (SEQ ID NO:188).

FIG. 24 is a bar graph showing amorphadiene produced in yeast strains with different promoter constructs inserted in front of the ERG9 gene of the host genome. CTRL-ADS refers to control strain with no modification; ERG9-CYC1-100% refers to construct comprising the ScCyc1 promoter and no insert; ERG9-CYC1-50% refers to construct comprising the ScCyc1 promoter followed by SEQ ID NO: 183 (SEQ ID NO:190); ERG9-CYC1-20% refers to construct comprising the ScCyc1 promoter followed by SEQ ID NO: 182 (SEQ ID NO:189); ERG9-CYC1-5% refers to construct comprising the ScCyc1 promoter followed by SEQ ID NO: 181 (SEQ ID NO:188); ERG9-KEX2-100% refers to construct comprising the ScKex2 promoter.

FIG. 25 contains the amino acid sequence of squalene synthase polypeptides from *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Candida glabrata*, *Ashbya gossypii*, *Cyberlindnera jadinii*, *Candida albicans*, *Saccharomyces cerevisiae*, *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus* (SEQ ID NOs:192-202),

and the amino acid sequence of a geranylgeranyl diphosphate synthase (GGPPS) from *Aspergilus nidulans* and *S. cerevisiae* (SEQ ID NOs. 203 and 167).

FIG. 26 is a bar graph of geranylgeraniol (GGOH) accumulation in the ERG9-CYC1-5% strain and ERG9-KEX2 strain after 72 hours.

FIG. 27 is a representative chromatograph showing the conversion of rubusoside to xylosylated intermediates for RebF production by UGT91D2e and EUGT11.

Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

This document is based on the discovery that recombinant hosts such as plant cells, plants, or microorganisms can be developed that express polypeptides useful for the biosynthesis of steviol glycosides such as rebaudioside A, rebaudioside C, rebaudioside D, rebaudioside E, rebaudioside F, or dulcoside A. The recombinant hosts described herein are particularly useful for producing Rebaudioside D. Such hosts can express one or more Uridine 5'-diphospho (UDP) glycosyl transferases suitable for producing steviol glycosides. Expression of these biosynthetic polypeptides in various microbial chassis allows steviol glycosides to be produced in a consistent, reproducible manner from energy and carbon sources such as sugars, glycerol, CO_2, H_2, and sunlight. The proportion of each steviol glycoside produced by a recombinant host can be tailored by incorporating preselected biosynthetic enzymes into the hosts and expressing them at appropriate levels, to produce a sweetener composition with a consistent taste profile. Furthermore, the concentrations of steviol glycosides produced by recombinant hosts are expected to be higher than the levels of steviol glycosides produced in the *Stevia* plant, which improves the efficiency of the downstream purification. Such sweetener compositions contain little or no plant based contaminants, relative to the amount of contaminants present in *Stevia* extracts.

At least one of the genes is a recombinant gene, the particular recombinant gene(s) depending on the species or strain selected for use. Additional genes or biosynthetic modules can be included in order to increase steviol glycoside yield, improve efficiency with which energy and carbon sources are converted to steviol and its

glycosides, and/or to enhance productivity from the cell culture or plant. Such additional biosynthetic modules include genes involved in the synthesis of the terpenoid precursors, isopentenyl diphosphate and dimethylallyl diphosphate. Additional biosynthetic modules include terpene synthase and terpene cyclase genes, such as genes encoding geranylgeranyl diphosphate synthase and copalyl diphosphate synthase; these genes may be endogenous genes or recombinant genes.

I. Steviol and Steviol Glycoside Biosynthesis Polypeptides

A. Steviol Biosynthesis Polypeptides

Chemical structures for several of the compounds found in Stevia extracts are shown in FIG. 1, including the diterpene steviol and various steviol glycosides. CAS numbers are shown in Table A below. See also, Steviol Glycosides Chemical and Technical Assessment 69th JECFA, prepared by Harriet Wallin, Food Agric. Org. (2007).

Table A.

COMPOUND	CAS #
Steviol	471-80-7
Rebaudioside A	58543-16-1
Steviolbioside	41093-60-1
Stevioside	57817-89-7
Rebaudioside B	58543-17-2
Rebaudioside C	63550-99-2
Rebaudioside D	63279-13-0
Rebaudioside E	63279-14-1
Rebaudioside F	438045-89-7
Rubusoside	63849-39-4
Dulcoside A	64432-06-0

It has been discovered that expression of certain genes in a host such as a microorganism confers the ability to synthesize steviol glycosides upon that host. As discussed in more detail below, one or more of such genes may be present naturally in a host. Typically, however, one or more of such genes are recombinant genes that have been transformed into a host that does not naturally possess them.

The biochemical pathway to produce steviol involves formation of geranylgeranyl diphosphate, cyclization to (-) copalyl diphosphate, followed by oxidation and hydroxylation to form steviol. Thus, conversion of geranylgeranyl diphosphate to steviol in a recombinant microorganism involves the expression of a gene encoding a kaurene synthase (KS), a gene encoding a kaurene oxidase (KO), and a gene encoding a steviol synthetase (KAH). Steviol synthetase also is known as kaurenoic acid 13-hydroxylase.

Suitable KS polypeptides are known. For example, suitable KS enzymes include those made by *Stevia rebaudiana*, *Zea mays*, *Populus trichocarpa*, and *Arabidopsis thaliana*. See, Table 1 and SEQ ID NOs: 132-135 and 156. Nucleotide sequences encoding these polypeptides are set forth in SEQ ID NOs: 40-47 and 155. The nucleotide sequences set forth in SEQ ID NOs:40-43 were modified for expression in yeast while the nucleotide sequences set forth in SEQ ID NOs: 44-47 are from the source organisms from which the KS polypeptides were identified.

Table 1. KS Clones

Enzyme Source Organism	gi Number	Accession Number	Construct Name	Length (nts)	SEQ ID (DNA)	SEQ ID (protein)
Stevia rebaudiana	4959241	AAD34295	MM-12	2355	40	132
Stevia rebaudiana	4959239	AAD34294	MM-13	2355	41	133
Zea mays	162458963	NP_001105097	MM-14	1773	42	134
Populus trichocarpa	224098838	XP_002311286	MM-15	2232	43	135
Arabidopsis thaliana	3056724	AF034774	EV-70	2358	155	156

Suitable KO polypeptides are known. For example, suitable KO enzymes include those made by *Stevia rebaudiana*, *Arabidopsis thaliana*, *Gibberella fujikoroi* and *Trametes versicolor*. See, Table 2 and SEQ ID NOs: 138-141. Nucleotide sequences encoding these polypeptides are set forth in in SEQ ID NOs: 52-59. The nucleotide sequences set forth in SEQ ID NOs: 52-55 were modified for expression in yeast. The nucleotide sequences set forth in SEQ ID NOs: 56-59 are from the source organisms from which the KO polypeptides were identified.

Table 2. KO Clones

Enzyme Source Organism	gi Number	Accession Number	Construct Name	Length (nts)	SEQ ID (DNA)	SEQ ID (protein)
Stevia rebaudiana	76446107	ABA42921	MM-18	1542	52	138
Arabidopsis thaliana	3342249	AAC39505	MM-19	1530	53	139
Gibberella fujikoroi	4127832	CAA76703	MM-20	1578	54	140
Trametes versicolor	14278967	BAB59027	MM-21	1500	55	141

Suitable KAH polypeptides are known. For example, suitable KAH enzymes include those made by *Stevia rebaudiana*, *Arabidopsis thaliana*, *Vitis vinifera* and *Medicago trunculata*. See, e.g., Table 3, SEQ ID NOs: 142-146; U.S. Patent Publication No. 2008-0271205; U.S. Patent Publication No. 2008-0064063 and Genbank Accession No. gi 189098312. The steviol synthetase from *Arabidopsis thaliana* is classified as a CYP714A2. Nucleotide sequences encoding these KAH enzymes are set forth in SEQ ID NOs: 60-69. The nucleotide sequences set forth in SEQ ID NOs: 60-64 were modified for expression in yeast while the nucleotide sequences from the source organisms from which the polypeptides were identified are set forth in SEQ ID NOs: 65-69.

Table 3. KAH Clones

Enzyme Source Organism	gi Number	Accession Number	Plasmid Name	Construct Name	Length (nts)	SEQ ID (DNA)	SEQ ID (protein)
Stevia rebaudiana		--*	pMUS35	MM-22	1578	60	142
Stevia rebaudiana	189418962	ACD93722	pMUS36	MM-23	1431	61	143
Arabidopsis thaliana	15238644	NP_197872	pMUS37	MM-24	1578	62	144
Vitis vinifera	225458454	XP_002282091	pMUS38	MM-25	1590	63	145
Medicago trunculata	84514135	ABC59076	pMUS39	MM-26	1440	64	146

* = Sequence is shown in U.S. Patent Publication No. 2008-0064063.

In addition, a KAH polypeptide from *Stevia rebaudiana* that was identified herein is particularly useful in a recombinant host. The nucleotide sequence (SEQ ID NO:163) encoding the *S. rebaudiana* KAH (SrKAHe1) (SEQ ID NO:164) is set forth in FIG. 12A. A nucleotide sequence encoding the *S. rebaudiana* KAH that has been codon-optimized for expression in yeast (SEQ ID NO:165) is set forth in FIG. 12B. The amino acid sequence of the *S. rebaudiana* KAH is set forth in FIG. 12C. The *S. rebaudiana* KAH shows significantly higher steviol synthase activity as compared to the *Arabidopsis thaliana ent*-kaurenoic acid hydroxylase described by Yamaguchi *et al.* (U.S. Patent Publication No. 2008/0271205 A1) when expressed in *S. cerevisiae*. The *S. rebaudiana* KAH polypeptide set forth in FIG. 12C has less than 20% identity to the KAH from U.S. Patent Publication No. 2008/0271205, and less than 35% identity to the KAH from U.S. Patent Publication No. 2008/0064063.

In some embodiments, a recombinant microorganism contains a recombinant gene encoding a KO and/or a KAH polypeptide. Such microorganisms also typically contain a recombinant gene encoding a cytochrome P450 reductase (CPR) polypeptide, since certain combinations of KO and/or KAH polypeptides require expression of an exogenous CPR polypeptide. In particular, the activity of a KO and/or a KAH polypeptide of plant origin can be significantly increased by the inclusion of a recombinant gene encoding an exogenous CPR polypeptide. Suitable CPR polypeptides are known. For example, suitable CPR enzymes include those made by *Stevia rebaudiana* and *Arabidopsis thaliana*. See, e.g., Table 4 and SEQ ID NOs: 147 and 148. Nucleotide sequences encoding these polypeptides are set forth in SEQ ID NOs: 70, 71, 73, and 74. The nucleotide sequences set forth in SEQ ID NOs: 70-72 were modified for expression in yeast. The nucleotide sequences from the source organisms from which the polypeptides were identified are set forth in SEQ ID NOs:73-75.

Table 4. CPR Clones

Enzyme Source Organism	gi Number	Accession Number	Plasmid Name	Construct Name	Length (nts)	SEQ ID (DNA)	SEQ ID (protein)
Stevia rebaudiana	93211213	ABB88839	pMUS40	MM-27	2133	70	147
Arabidopsis	15233853	NP_194183	pMUS41	MM-28	2079	71	148

thaliana							
Giberella fujikuroi	32562989	CAE09055	pMUS42	MM-29	2142	72	149

For example, the steviol synthase encoded by SrKAHe1 is activated by the *S. cerevisiae* CPR encoded by gene NCP1 (YHR042W). Even better activation of the steviol synthase encoded by SrKAHe1 is observed when the *Arabidopsis thaliana* CPR encoded by the gene *ATR2* or the *S. rebaudiana* CPR encoded by the gene *CPR8* are co-expressed. FIG. 13A contains the amino acid sequence of the *S. cerevisiae*, *A. thaliana* (from ATR1 and ATR2 genes) and *S. rebaudiana* CPR polypeptides (from *CPR7* and *CPR8* genes) (SEQ ID NOs: 166-170). FIG. 13 B contains the nucleotide sequence encoding the *A. thaliana* and *S. rebaudiana* CPR polypeptides (SEQ ID NOs:171-174).

For example, the yeast gene *DPP1* and/or the yeast gene *LPP1* can be disrupted or deleted such that the degradation of farnesyl pyrophosphate (FPP) to farnesol is reduced and the degradation of geranylgeranylpyrophosphate (GGPP)) to geranylgeraniol (GGOH) is reduced. Alternatively, the promoter or enhancer elements of an endogenous gene encoding a phosphatase can be altered such that the expression of their encoded proteins is altered. Homologous recombination can be used to disrupt an endogenous gene. For example, a "gene replacement" vector can be constructed in such a way to include a selectable marker gene. The selectable marker gene can be operably linked, at both 5' and 3' end, to portions of the gene of sufficient length to mediate homologous recombination. The selectable marker can be one of any number of genes that complement host cell auxotrophy, provide antibiotic resistance, or result in a color change. Linearized DNA fragments of the gene replacement vector then are introduced into the cells using methods well known in the art (see below). Integration of the linear fragments into the genome and the disruption of the gene can be determined based on the selection marker and can be verified by, for example, Southern blot analysis. Subsequent to its use in selection, a selectable marker can be removed from the genome of the host cell by, e.g., Cre-loxP systems (see, e.g., Gossen et al. (2002) *Ann. Rev. Genetics* 36:153-173 and U.S. Application Publication No. 20060014264). Alternatively, a gene replacement vector can be constructed in such a way as to include a portion of the gene to

be disrupted, where the portion is devoid of any endogenous gene promoter sequence and encodes none, or an inactive fragment of, the coding sequence of the gene. An "inactive fragment" is a fragment of the gene that encodes a protein having, e.g., less than about 10% (e.g., less than about 9%, less than about 8%, less than about 7%, less than about 6%, less than about 5%, less than about 4%, less than about 3%, less than about 2%, less than about 1%, or 0%) of the activity of the protein produced from the full-length coding sequence of the gene. Such a portion of the gene is inserted in a vector in such a way that no known promoter sequence is operably linked to the gene sequence, but that a stop codon and a transcription termination sequence are operably linked to the portion of the gene sequence. This vector can be subsequently linearized in the portion of the gene sequence and transformed into a cell. By way of single homologous recombination, this linearized vector is then integrated in the endogenous counterpart of the gene.

Expression in a recombinant microorganism of these genes results in the conversion of geranylgeranyl diphosphate to steviol.

B. Steviol Glycoside Biosynthesis Polypeptides

A recombinant host described herein can convert steviol to a steviol glycoside. Such a host (e.g., microorganism) contains genes encoding one or more UDP Glycosyl Transferases, also known as UGTs. UGTs transfer a monosaccharide unit from an activated nucleotide sugar to an acceptor moiety, in this case, an –OH or –COOH moiety on steviol or steviol derivative. UGTs have been classified into families and subfamilies based on sequence homology. Li et al. J. Biol. Chem. 276:4338-4343 (2001).

B. 1 Rubusoside Biosynthesis Polypeptides

The biosynthesis of rubusoside involves glycosylation of the 13-OH and the 19-COOH of steviol. See FIG. 2A. Conversion of steviol to rubusoside in a recombinant host such as a microorganism can be accomplished by the expression of gene(s) encoding UGTs 85C2 and 74G1, which transfer a glucose unit to the 13-OH or the 19-COOH, respectively, of steviol.

A suitable UGT85C2 functions as a uridine 5'-diphospho glucosyl:steviol 13-OH transferase, and a uridine 5'-diphospho glucosyl:steviol-19-*O*-glucoside 13-OH transferase. Functional UGT85C2 polypeptides also may catalyze glucosyl transferase reactions that utilize steviol glycoside substrates other than steviol and steviol-19-*O*-glucoside.

A suitable UGT74G1 polypeptide functions as a uridine 5'-diphospho glucosyl: steviol 19-COOH transferase and a uridine 5'-diphospho glucosyl: steviol-13-*O*-glucoside 19-COOH transferase. Functional UGT74G1 polypeptides also may catalyze glycosyl transferase reactions that utilize steviol glycoside substrates other than steviol and steviol-13-*O*-glucoside, or that transfer sugar moieties from donors other than uridine diphosphate glucose.

A recombinant microorganism expressing a functional UGT74G1 and a functional UGT85C2 can make rubusoside and both steviol monosides (i.e., steviol 13-*O*-monoglucoside and steviol 19-*O*-monoglucoside) when steviol is used as a feedstock in the medium. One or more of such genes may be present naturally in the host. Typically, however, such genes are recombinant genes that have been transformed into a host (e.g., microorganism) that does not naturally possess them.

As used herein, the term recombinant host is intended to refer to a host, the genome of which has been augmented by at least one incorporated DNA sequence. Such DNA sequences include but are not limited to genes that are not naturally present, DNA sequences that are not normally transcribed into RNA or translated into a protein ("expressed"), and other genes or DNA sequences which one desires to introduce into the non-recombinant host. It will be appreciated that typically the genome of a recombinant host described herein is augmented through the stable introduction of one or more recombinant genes. Generally, the introduced DNA is not originally resident in the host that is the recipient of the DNA, but it is within the scope of the invention to isolate a DNA segment from a given host, and to subsequently introduce one or more additional copies of that DNA into the same host, e.g., to enhance production of the product of a gene or alter the expression pattern of a gene. In some instances, the introduced DNA will modify or even replace an endogenous gene or DNA sequence by, e.g., homologous

recombination or site-directed mutagenesis. Suitable recombinant hosts include microorganisms, plant cells, and plants.

The term "recombinant gene" refers to a gene or DNA sequence that is introduced into a recipient host, regardless of whether the same or a similar gene or DNA sequence may already be present in such a host. "Introduced," or "augmented" in this context, is known in the art to mean introduced or augmented by the hand of man. Thus, a recombinant gene may be a DNA sequence from another species, or may be a DNA sequence that originated from or is present in the same species, but has been incorporated into a host by recombinant methods to form a recombinant host. It will be appreciated that a recombinant gene that is introduced into a host can be identical to a DNA sequence that is normally present in the host being transformed, and is introduced to provide one or more additional copies of the DNA to thereby permit overexpression or modified expression of the gene product of that DNA.

Suitable UGT74G1 and UGT85C2 polypeptides include those made by *Stevia rebaudiana*. Genes encoding functional UGT74G1 and UGT85C2 polypeptides from *Stevia* are reported in Richman, et al. Plant J. 41: 56-67 (2005). Amino acid sequences of *S. rebaudiana* UGT74G1 and UGT85C2 polypeptides are set forth in SEQ ID NOs: 1 and 3, respectively. Nucleotide sequences encoding UGT74G1 and UGT85C2 that have been optimized for expression in yeast are set forth in SEQ ID NOs: 2 and 4, respectively. DNA 2.0 codon-optimized sequence for UGTs 85C2, 91D2e, 74G1 and 76G1 are set forth in SEQ ID NOs: 82, 84, 83, and 85, respectively. See also the UGT85C2 and UGT74G1 variants described below in the "Functional Homolog" section. For example, an UGT85C2 polypeptide containing substitutions at positions 65, 71, 270, 289, and 389 can be used (e.g., A65S, E71Q, T270M, Q289H, and A389V).

In some embodiments, the recombinant host is a microorganism. The recombinant microorganism can be grown on media containing steviol in order to produce rubusoside. In other embodiments, however, the recombinant microorganism expresses one or more recombinant genes involved in steviol biosynthesis, e.g., a CDPS gene, a KS gene, a KO gene and/or a KAH gene. Suitable CDPS polypeptides are known. For example, suitable CDPS enzymes include those made by *Stevia rebaudiana*,

Streptomyces clavuligerus, *Bradyrhizobium japonicum*, *Zea mays*, and *Arabidopsis*. See, e.g., Table 5 and SEQ ID NOs: 129-131, 158, and 160. Nucleotide sequences encoding these polypeptides are set forth in SEQ ID NOs: 34-39, 157, and 159. The nucleotide sequences set forth in SEQ ID NOs: 34-36 were modified for expression in yeast. The nucleotide sequences from the source organisms from which the polypeptides were identified are set forth in SEQ ID NOs:37-39.

In some embodiments, CDPS polypeptides that lack a chloroplast transit peptide at the amino terminus of the unmodified polypeptide can be used. For example, the first 150 nucleotides from the 5' end of the *Zea mays* CDPS coding sequence shown in FIG. 14 (SEQ ID NO:157) can be removed. Doing so removes the amino terminal 50 residues of the amino acid sequence shown in FIG. 14 (SEQ ID NO:158), which encode a chloroplast transit peptide. The truncated CDPS gene can be fitted with a new ATG translation start site and operably linked to a promoter, typically a constitutive or highly expressing promoter. When a plurality of copies of the truncated coding sequence are introduced into a microorganism, expression of the CDPS polypeptide from the promoter results in an increased carbon flux towards *ent*-kaurene biosynthesis.

Table 5. CDPS Clones

Enzyme Source Organism	gi Number	Accession Number	Plasmid Name	Construct Name	Length (nts)	SEQ ID: (DNA)	SEQ ID (protein)
Stevia rebaudiana	2642661	AAB87091	pMUS22	MM-9	2364	34	129
Streptomyces clavuligerus	197705855	EDY51667	pMUS23	MM-10	1584	35	130
Bradyrhizobium japonicum	529968	AAC28895.1	pMUS24	MM-11	1551	36	131
Zea mays	50082774	AY562490		EV65	2484	157	158
Arabidopsis thaliana	18412041	NM_116512		EV64	2409	159	160

CDPS-KS bifunctional proteins (SEQ ID NOs: 136 and 137) also can be used. Nucleotide sequences encoding the CDPS-KS bifunctional enzymes shown in Table 6 were modified for expression in yeast (see SEQ ID NOs: 48 and 49). The nucleotide sequences from the source organisms from which the polypeptides were originally

identified are set forth in SEQ ID NOs: 50 and 51. A bifunctional enzyme from *Gibberella fujikuroi* (SEQ ID NO:162) also can be used. A nucleotide sequence encoding the *Gibberella fujikuroi* bifunctional CDPS-KS enzyme was modified for expression in yeast (see FIG. 15A, SEQ ID NO:161).

Table 6. CDPS-KS Clones

Enzyme Source Organism	gi Number	Accession Number	Construct Name	Length (nts)	SEQ ID (DNA)	SEQ ID (protein)
Phomopsis amygdali	186704306	BAG30962	MM-16	2952	48	136
Physcomitrella patens	146325986	BAF61135	MM-17	2646	49	137
Gibberella fujikuroi	62900107	Q9UVY5.1		2859	161	162

Thus, a microorganism containing a CDPS gene, a KS gene, a KO gene and a KAH gene in addition to a UGT74G1 and a UGT85C2 gene is capable of producing both steviol monosides and rubusoside without the necessity for using steviol as a feedstock.

In some embodiments, the recombinant microorganism further expresses a recombinant gene encoding a geranylgeranyl diphosphate synthase (GGPPS). Suitable GGPPS polypeptides are known. For example, suitable GGPPS enzymes include those made by *Stevia rebaudiana*, *Gibberella fujikuroi*, *Mus musculus*, *Thalassiosira pseudonana*, *Streptomyces clavuligerus*, *Sulfulobus acidocaldarius*, *Synechococcus* sp. and *Arabidopsis thaliana*. See, Table 7 and SEQ ID NOs: 121-128. Nucleotide sequences encoding these polypeptides are set forth in SEQ ID NOs:18-33. The nucleotide sequences set forth in SEQ ID NOs: 18-25 were modified for expression in yeast while the nucleotide sequences from the source organisms from which the polypeptides were identified are set forth in SEQ ID NOs: 26-33.

Table 7. GGPPS Clones

Enzyme Source Organism	gi Number	Accession Number	Plasmid Name	Construct Name	Length (nts)	SEQ ID (DNA)	SEQ ID (protein)
Stevia rebaudiana	90289577	ABD92926	pMUS14	MM-1	1086	18	121

Gibberella fujikuroi	3549881	CAA75568	pMUS15	MM-2	1029	19	122
Mus musculus	47124116	AAH69913	pMUS16	MM-3	903	20	123
Thalassiosira pseudonana	223997332	XP_002288339	pMUS17	MM-4	1020	21	124
Streptomyces clavuligerus	254389342	ZP_05004570	pMUS18	MM-5	1068	22	125
Sulfulobus acidocaldarius	506371	BAA43200	pMUS19	MM-6	993	23	126
Synechococcus sp.	86553638	ABC98596	pMUS20	MM-7	894	24	127
Arabidopsis thaliana	15234534	NP_195399	pMUS21	MM-8	1113	25	128

In some embodiments, the recombinant microorganism further can express recombinant genes involved in diterpene biosynthesis or production of terpenoid precursors, e.g., genes in the methylerythritol 4-phosphate (MEP) pathway or genes in the mevalonate (MEV) pathway discussed below, have reduced phosphatase activity, and/or express a sucrose synthase (SUS) as discussed herein.

B. 2 Rebaudioside A, Rebaudioside D, and Rebaudioside E Biosynthesis Polypeptides

The biosynthesis of rebaudioside A involves glucosylation of the aglycone steviol. Specifically, rebaudioside A can be formed by glucosylation of the 13-OH of steviol which forms the 13-*O*-steviolmonoside, glucosylation of the C-2' of the 13-O-glucose of steviolmonoside which forms steviol-1,2-bioside, glucosylation of the C-19 carboxyl of steviol-1,2-bioside which forms stevioside, and glucosylation of the C-3' of the C-13-O-glucose of stevioside. The order in which each glucosylation reaction occurs can vary. See FIG. 2A.

The biosynthesis of rebaudioside E and/or rebaudioside D involves glucosylation of the aglycone steviol. Specifically, rebaudioside E can be formed by glucosylation of the 13-OH of steviol which forms steviol-13-*O*-glucoside, glucosylation of the C-2' of the 13-O-glucose of steviol-13-*O*-glucoside which forms the steviol-1,2-bioside, glucosylation of the C-19 carboxyl of the 1,2-bioside to form 1,2-stevioside, and glucosylation of the C-2' of the 19-O-glucose of the 1,2-stevioside to form rebaudioside

E. Rebaudioside D can be formed by glucosylation of the C-3' of the C-13-O-glucose of rebaudioside E. The order in which each glycosylation reaction occurs can vary. For example, the glucosylation of the C-2' of the 19-O-glucose may be the last step in the pathway, wherein Rebaudioside A is an intermediate in the pathway. See FIG. 2C.

It has been discovered that conversion of steviol to rebaudioside A, rebaudioside D, and/or rebaudioside E in a recombinant host can be accomplished by expressing the following functional UGTs: EUGT11, 74G1, 85C2, and 76G1, and optionally 91D2. Thus, a recombinant microorganism expressing combinations of these four or five UGTs can make rebaudioside A and rebaudioside D when steviol is used as a feedstock. Typically, one or more of these genes are recombinant genes that have been transformed into a microorganism that does not naturally possess them. It has also been discovered that UGTs designated herein as SM12UGT can be substituted for UGT91D2.

In some embodiments, less than five (e.g., one, two, three, or four) UGTs are expressed in a host. For example, a recombinant microorganism expressing a functional EUGT11 can make rebaudioside D when rebaudioside A is used as a feedstock. A recombinant microorganism expressing two functional UGTs, EUGT11 and 76G1, and optionally a functional 91D12, can make rebaudioside D when rubusoside or 1,2-stevioside is used as a feedstock. As another alternative, a recombinant microorganism expressing three functional UGTs, EUGT11, 74G1, 76G1, and optionally 91D2, can make rebaudioside D when fed the monoside, steviol-13-*O*-glucoside, in the medium. Similarly, conversion of steviol-19-*O*-glucoside to rebaudioside D in a recombinant microorganism can be accomplished by the expression of genes encoding UGTs EUGT11, 85C2, 76G1, and optionally 91D2, when fed steviol-19-*O*-glucoside. Typically, one or more of these genes are recombinant genes that have been transformed into a host that does not naturally possess them.

Suitable UGT74G1 and UGT85C2 polypeptides include those discussed above. A suitable UGT76G1 adds a glucose moiety to the C-3'of the C-13-*O*-glucose of the acceptor molecule, a steviol 1,2 glycoside. Thus, UGT76G1 functions, for example, as a uridine 5'-diphospho glucosyl: steviol 13-*O*-1,2 glucoside C-3' glucosyl transferase and a uridine 5'-diphospho glucosyl: steviol-19-*O*-glucose, 13-O-1,2 bioside C-3' glucosyl

transferase. Functional UGT76G1 polypeptides may also catalyze glucosyl transferase reactions that utilize steviol glycoside substrates that contain sugars other than glucose, e.g., steviol rhamnosides and steviol xylosides. See, FIGS 2A, 2B, 2C and 2D. Suitable UGT76G1 polypeptides include those made by *S. rebaudiana* and reported in Richman, et al. Plant J. 41: 56-67 (2005). The amino acid sequence of a *S. rebaudiana* UGT76G1 polypeptide is set forth in SEQ ID NO:7. The nucleotide sequence encoding the UGT76G1 polypeptide of SEQ ID NO:7 has been optimized for expression in yeast and is set forth in SEQ ID NO:8. See also the UGT76G1 variants set forth in the "Functional Homolog" section.

A suitable EUGT11 or UGT91D2 polypeptide functions as a uridine 5'-diphospho glucosyl: steviol-13-*O*-glucoside transferase (also referred to as a steviol-13-monoglucoside 1,2-glucosylase), transferring a glucose moiety to the C-2' of the 13-O-glucose of the acceptor molecule, steviol-13-*O*-glucoside.

A suitable EUGT11 or UGT91D2 polypeptide also functions as a uridine 5'-diphospho glucosyl: rubusoside transferase transferring a glucose moiety to the C-2' of the 13-*O*-glucose of the acceptor molecule, rubusoside, to produce stevioside. EUGT11 polypeptides also can transfer a glucose moiety to the C-2' of the 19-*O*-glucose of the acceptor molecule, rubusoside, to produce a 19-*O*-1,2-diglycosylated rubusoside (compound 2 in FIG. 3).

Functional EUGT11 or UGT91D2 polypeptides also can catalyze reactions that utilize steviol glycoside substrates other than steviol-13-*O*-glucoside and rubusoside. For example, a functional EUGT11 polypeptide may utilize stevioside as a substrate, transferring a glucose moiety to the C-2' of the 19-*O*-glucose residue to produce Rebaudioside E (see compound 3 in FIG. 3). Functional EUGT11 and UGT91D2 polypeptides may also utilize Rebaudioside A as a substrate, transferring a glucose moiety to the C-2' of the 19-*O*-glucose residue of Rebaudioside A to produce Rebaudioside D. As set forth in the Examples, EUGT11 can convert Rebaudioside A to Rebaudioside D at a rate that is least 20 times faster (e.g., as least 25 times or at least 30 times faster) than the corresponding rate of UGT91D2e (SEQ ID NO: 5) when the reactions are performed under similar conditions, i.e., similar time, temperature, purity,

and substrate concentration. As such, EUGT11 produces greater amounts of RebD than UGT91D2e when incubated under similar conditions.

In addition, a functional EUGT11 exhibits significant C-2' 19-O-diglycosylation activity with rubusoside or stevioside as substrates, whereas UGT91D2e has no detectable diglycosylation activity with these substrates. Thus, a functional EUGT11 can be distinguished from UGT91D2e by the differences in steviol glycoside substrate-specificity. FIG. 3 provides a schematic overview of the 19-*O*-1,2 diglycosylation reactions that are performed by EUGT11 and UGT91D2e.

A functional EUGT11 or UGT91D2 polypeptide typically does not transfer a glucose moiety to steviol compounds having a 1,3-bound glucose at the C-13 position, i.e., transfer of a glucose moiety to steviol 1,3-bioside and 1,3-stevioside does not occur.

Functional EUGT11 and UGT91D2 polypeptides can transfer sugar moieties from donors other than uridine diphosphate glucose. For example, a functional EUGT11 or UGT91D2 polypeptide can act as a uridine 5'-diphospho D-xylosyl: steviol-13-*O*-glucoside transferase, transferring a xylose moiety to the C-2' of the 13-O-glucose of the acceptor molecule, steviol-13-*O*-glucoside. As another example, a functional EUGT11 or UGT91D2 polypeptide can act as a uridine 5'-diphospho L-rhamnosyl: steviol-13-*O*-glucoside transferase, transferring a rhamnose moiety to the C-2' of the 13-O-glucose of the acceptor molecule, steviol-13-*O*-glucoside

Suitable EUGT11 polypeptides are described herein and can include the EUGT11 polypeptide from *Oryza sativa* (GenBank Accession No. AC133334). For example, an EUGT11 polypeptide can have an amino acid sequence with at least 70% sequence identity (e.g., at least 75, 80, 85, 90, 95, 96, 97, 98, or 99% sequence identity) to the amino acid sequence set forth in SEQ ID NO: 152 (see FIG. 7). The nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 152 is set forth in SEQ ID NO: 153. SEQ ID NO: 154 is a nucleotide sequence encoding the polypeptide of SEQ ID NO: 152 that has been codon optimized for expression in yeast.

Suitable functional UGT91D2 polypeptides include those disclosed herein, e.g., the polypeptides designated UGT91D2e and UGT91D2m. The amino acid sequence of an exemplary UGT91D2e polypeptide from *Stevia rebaudiana* is set forth in SEQ ID NO:

5. SEQ ID NO:6 is a nucleotide sequence encoding the polypeptide of SEQ ID NO:5 that has been codon optimized for expression in yeast. The *S. rebaudiana* nucleotide sequence encoding the polypeptide of SEQ ID NO:5 is set forth in SEQ ID NO:9. The amino acid sequences of exemplary UGT91D2m polypeptides from *S. rebaudiana* are set forth in SEQ ID NOs: 10 and 12, and are encoded by the nucleic acid sequences set forth in SEQ ID NOs: 11 and 13, respectively. In addition, UGT91D2 variants containing a substitution at amino acid residues 206, 207, and 343 of SEQ ID NO: 5 can be used. For example, the amino acid sequence set forth in SEQ ID NO:95 and having the following mutations with respect to wild-type UGT92D2e (SEQ ID NO:5) G206R, Y207C, and W343R can be used. In addition, a UGT91D2 variant containing substitutions at amino acid residues 211 and 286 can be used. For example, a UGT91D2 variant can include a substitution of a methionine for leucine at position 211 and a substitution of an alanine for valine at position 286 of SEQ ID NO:5 (UGT91D2e-b).

As indicated above, UGTs designated herein as SM12UGT can be substituted for UGT91D2. Suitable functional SM12UGT polypeptides include those made by *Ipomoea purpurea* (Japanese morning glory) and described in Morita et al. Plant J. 42, 353-363 (2005). The amino acid sequence encoding the *I. purpurea* IP3GGT polypeptide is set forth in SEQ ID NO:76. SEQ ID NO:77 is a nucleotide sequence encoding the polypeptide of SEQ ID NO:76 that has been codon optimized for expression in yeast. Another suitable SM12UGT polypeptide is a Bp94B1 polypeptide having an R25S mutation. See Osmani et al. Plant Phys. 148: 1295-1308 (2008) and Sawada et al. J. Biol. Chem. 280:899-906 (2005). The amino acid sequence of the *Bellis perennis* (red daisy) UGT94B1 polypeptide is set forth in SEQ ID NO:78. SEQ ID NO:79 is the nucleotide sequence encoding the polypeptide of SEQ ID NO:78 that has been codon optimized for expression in yeast.

In some embodiments, the recombinant microorganism is grown on media containing steviol-13-*O*-glucoside or steviol-19-*O*-glucoside in order to produce rebaudioside A and/or rebaudioside D. In such embodiments, the microorganism contains and expresses genes encoding a functional EUGT11, a functional UGT74G1, a functional UGT85C2, a functional UGT76G1, and an optional functional UGT91D2, and

is capable of accumulating rebaudioside A and rebaudioside D when steviol, one or both of the steviolmonosides, or rubusoside is used as feedstock.

In other embodiments, the recombinant microorganism is grown on media containing rubusoside in order to produce rebaudioside A and/or rebaudioside D. In such embodiments, the microorganism contains and expresses genes encoding a functional EUGT11, a functional UGT76G1, and an optional functional UGT91D2, and is capable of producing rebaudioside A and/or rebaudioside D when rubusoside is used as feedstock.

In other embodiments the recombinant microorganism expresses one or more genes involved in steviol biosynthesis, e.g., a CDPS gene, a KS gene, a KO gene and/or a KAH gene. Thus, for example, a microorganism containing a CDPS gene, a KS gene, a KO gene and a KAH gene, in addition to a EUGT11, a UGT74G1, a UGT85C2, a UGT76G1, and optionally a functional UGT91D2 (e.g., UGT91D2e), is capable of producing rebaudioside A, rebaudioside D, and/or rebaudioside E without the necessity for including steviol in the culture media.

In some embodiments, the recombinant host further contains and expresses a recombinant GGPPS gene in order to provide increased levels of the diterpene precursor geranylgeranyl diphosphate, for increased flux through the steviol biosynthetic pathway. In some embodiments, the recombinant host further contains a construct to silence the expression of non-steviol pathways consuming geranylgeranyl diphosphate, ent-Kaurenoic acid or farnesyl pyrophosphate, thereby providing increased flux through the steviol and steviol glycosides biosynthetic pathways. For example, flux to sterol production pathways such as ergosterol may be reduced by downregulation of the *ERG9* gene. See, the ERG9 section below and Examples 24-25. In cells that produce gibberellins, gibberellin synthesis may be downregulated to increase flux of ent-kaurenoic acid to steviol. In carotenoid-producing organisms, flux to steviol may be increased by downregulation of one or more carotenoid biosynthetic genes. In some embodiments, the recombinant microorganism further can express recombinant genes involved in diterpene biosynthesis or production of terpenoid precursors, e.g., genes in the MEP or MEV) pathways discussed below, have reduced phosphatase activity, and/or express a SUS as discussed herein.

One with skill in the art will recognize that by modulating relative expression levels of different UGT genes, a recombinant host can be tailored to specifically produce steviol glycoside products in a desired proportion. Transcriptional regulation of steviol biosynthesis genes and steviol glycoside biosynthesis genes can be achieved by a combination of transcriptional activation and repression using techniques known to those in the art. For *in vitro* reactions, one with skill in the art will recognize that addition of different levels of UGT enzymes in combination or under conditions which impact the relative activities of the different UGTS in combination will direct synthesis towards a desired proportion of each steviol glycoside. One with skill in the art will recognize that a higher proportion of rebaudioside D or E or more efficient conversion to rebaudioside D or E can be obtained with a diglycosylation enzyme that has a higher activity for the 19-*O*-glucoside reaction as compared to the 13-*O*-glucoside reaction (substrates rebaudioside A and stevioside).

In some embodiments, a recombinant host such as a microorganism produces rebaudioside D-enriched steviol glycoside compositions that have greater than at least 3% rebaudioside D by weight total steviol glycosides, e.g., at least 4% rebaudioside D at least 5% rebaudioside D, 10-20% rebaudioside D, 20-30% rebaudioside D, 30-40% rebaudioside D, 40-50% rebaudioside D, 50-60% rebaudioside D, 60-70% rebaudioside D, 70-80% rebaudioside D. In some embodiments, a recombinant host such as a microorganism produces steviol glycoside compositions that have at least 90% rebaudioside D, e.g., 90-99% rebaudioside D. Other steviol glycosides present may include those depicted in Figure 2 C such as steviol monosides, steviol glucobiosides, rebaudioside A, rebaudioside E, and stevioside. In some embodiments, the rebaudioside D-enriched composition produced by the host (e.g., microorganism) can be further purified and the rebaudioside D or rebaudioside E so purified can then be mixed with other steviol glycosides, flavors, or sweeteners to obtain a desired flavor system or sweetening composition. For instance, a rebaudioside D-enriched composition produced by a recombinant host can be combined with a rebaudioside A, C, or F-enriched composition produced by a different recombinant host, with rebaudioside A, F, or C purified from a Stevia extract, or with rebaudioside A, F, or C produced *in vitro*.

In some embodiments, rebaudioside A, rebaudioside D, rebaudioside B, steviol monoglucosides, steviol-1,2-bioside, rubusoside, stevioside, or rebaudioside E can be produced using *in vitro* methods while supplying the appropriate UDP-sugar and/or a cell-free system for regeneration of UDP-sugars. See, for example, Jewett MC, et al. Molecular Systems Biology, Vol. 4, article 220 (2008); Masada S et al. FEBS Letters, Vol. 581, 2562-2566 (2007). In some embodiments, sucrose and a sucrose synthase may be provided in the reaction vessel in order to regenerate UDP-glucose from the UDP generated during glycosylation reactions. See Figure 11. The sucrose synthase can be from any suitable organism. For example, a sucrose synthase coding sequence from *Arabidopsis thaliana*, *Stevia rebaudiana*, or *Coffea arabica* can be cloned into an expression plasmid under control of a suitable promoter, and expressed in a host such as a microorganism or a plant.

Conversions requiring multiple reactions may be carried out together, or stepwise. For example, rebaudioside D may be produced from rebaudioside A that is commercially available as an enriched extract or produced via biosynthesis, with the addition of stoichiometric or excess amounts of UDP-glucose and EUGT11. As an alternative, rebaudioside D may be produced from steviol glycoside extracts that are enriched for stevioside and rebaudioside A, using EUGT11 and a suitable UGT76G1 enzyme. In some embodiments, phosphatases are used to remove secondary products and improve the reaction yields. UGTs and other enzymes for *in vitro* reactions may be provided in soluble forms or in immobilized forms.

In some embodiments, rebaudioside A, rebaudioside D, or rebaudioside E can be produced using whole cells that are fed raw materials that contain precursor molecules such as steviol and/or steviol glycosides, including mixtures of steviol glycosides derived from plant extracts. The raw materials may be fed during cell growth or after cell growth. The whole cells may be in suspension or immobilized. The whole cells may be entrapped in beads, for example calcium or sodium alginate beads. The whole cells may be linked to a hollow fiber tube reactor system. The whole cells may be concentrated and entrapped within a membrane reactor system. The whole cells may be in fermentation broth or in a reaction buffer. In some embodiments, a permeabilizing agent is utilized for efficient

transfer of substrate into the cells. In some embodiments, the cells are permeabilized with a solvent such as toluene, or with a detergent such as Triton-X or Tween. In some embodiments, the cells are permeabilized with a surfactant, for example a cationic surfactant such as cetyltrimethylammonium bromide (CTAB). In some embodiments, the cells are permeabilized with periodic mechanical shock such as electroporation or a slight osmotic shock. The cells can contain one recombinant UGT or multiple recombinant UGTs. For example, the cells can contain UGT 76G1 and EUGT11 such that mixtures of stevioside and RebA are efficiently converted to RebD. In some embodiments, the whole cells are the host cells described in section III A. In some embodiments, the whole cells are a Gram-negative bacterium such as *E. coli*. In some embodiments, the whole cell is a Gram-positive bacterium such as *Bacillus*. In some embodiments, the whole cell is a fungal species such as *Aspergillus*, or a yeast such as *Saccharomyces*. In some embodiments, the term "whole cell biocatalysis" is used to refer to the process in which the whole cells are grown as described above (e.g., in a medium and optionally permeabilized) and a substrate such as rebA or stevioside is provided and converted to the end product using the enzymes from the cells. The cells may or may not be viable, and may or may not be growing during the bioconversion reactions. In contrast, in fermentation, the cells are cultured in a growth medium and fed a carbon and energy source such as glucose and the end product is produced with viable cells.

B. 3 Dulcoside A and Rebaudioside C Biosynthesis Polypeptides

The biosynthesis of rebaudioside C and/or dulcoside A involves glucosylation and rhamnosylation of the aglycone steviol. Specifically, dulcoside A can be formed by glucosylation of the 13-OH of steviol which forms steviol-13-*O*-glucoside, rhamnosylation of the C-2' of the 13-O-glucose of steviol-13-*O*-glucoside which forms the 1,2 rhamnobioside, and glucosylation of the C-19 carboxyl of the 1,2 rhamnobioside. Rebaudioside C can be formed by glucosylation of the C-3' of the C-13-O-glucose of dulcoside A. The order in which each glycosylation reaction occurs can vary. See FIG. 2B.

It has been discovered that conversion of steviol to dulcoside A in a recombinant host can be accomplished by the expression of gene(s) encoding the following functional UGTs: 85C2, EUGT11 and/or 91D2e, and 74G1. Thus, a recombinant microorganism expressing these three or four UGTs and a rhamnose synthetase can make dulcoside A when fed steviol in the medium. Alternatively, a recombinant microorganism expressing two UGTs, EUGT11 and 74G1, and rhamnose synthetase can make dulcoside A when fed the monoside, steviol-13-*O*-glucoside or steviol-19-*O*-glucoside, in the medium. Similarly, conversion of steviol to rebaudioside C in a recombinant microorganism can be accomplished by the expression of gene(s) encoding UGTs 85C2, EUGT11, 74G1, 76G1, optionally 91D2e, and rhamnose synthetase when fed steviol, by the expression of genes encoding UGTs EUGT11 and/or 91D2e, 74G1, and 76G1, and rhamnose synthetase when fed steviol-13-*O*-glucoside, by the expression of genes encoding UGTs 85C2, EUGT11 and/or 91D2e, 76G1, and rhamnose synthetase when fed steviol-19-*O*-glucoside, or by the expression of genes encoding UGTs EUGT11 and/or 91D2e, 76G1, and rhamnose synthetase when fed rubusoside. Typically, one or more of these genes are recombinant genes that have been transformed into a microorganism that does not naturally possess them.

Suitable EUGT11, UGT91D2, UGT74G1, UGT76G1 and UGT85C2 polypeptides include the functional UGT polypeptides discussed herein. Rhamnose synthetase provides increased amounts of the UDP-rhamnose donor for rhamnosylation of the steviol compound acceptor. Suitable rhamnose synthetases include those made by *Arabidopsis thaliana*, such as the product of the *A. thaliana* RHM2 gene.

In some embodiments, a UGT79B3 polypeptide is substituted for a UGT91D2 polypeptide. Suitable UGT79B3 polypeptides include those made by *Arabidopsis thaliana*, which are capable of rhamnosylation of steviol 13-O-monoside *in vitro*. *A. thaliana* UGT79B3 can rhamnosylate glucosylated compounds to form 1,2-rhamnosides. The amino acid sequence of an *Arabidopsis* thaliana UGT79B3 is set forth in SEQ ID NO:150. The nucleotide sequence encoding the amino acid sequence of SEQ ID NO:150 is set forth in SEQ ID NO:151.

In some embodiments rebaudioside C can be produced using *in vitro* methods while supplying the appropriate UDP-sugar and/or a cell-free system for regeneration of UDP-sugars. See, for example, "An integrated cell-free metabolic platform for protein production and synthetic biology" by Jewett MC, Calhoun KA, Voloshin A, Wuu JJ and Swartz JR in Molecular Systems Biology, 4, article 220 (2008); Masada S et al. FEBS Letters, Vol. 581, 2562-2566 (2007). In some embodiments, sucrose and a sucrose synthase may be provided in the reaction vessel in order to regenerate UDP-glucose from UDP during the glycosylation reactions. See FIG. 11. The sucrose synthase can be from any suitable organism. For example, a sucrose synthase coding sequence from *Arabidopsis thaliana*, *Stevia rebaudiana*, or *Coffea arabica* can be cloned into an expression plasmid under control of a suitable promoter, and expressed in a host (e.g., a microorganism or a plant). In some embodiments a RHM2 enzyme (Rhamnose synthase) may also be provided, with NADPH, to generate UDP-rhamnose from UDP-glucose.

Reactions may be carried out together, or stepwise. For instance, rebaudioside C may be produced from rubusoside with the addition of stoichiometric amounts of UDP-rhamnose and EUGT11, followed by addition of UGT76G1 and an excess or stoichiometric supply of UDP-glucose. In some embodiments, phosphatases are used to remove secondary products and improve the reaction yields. UGTs and other enzymes for *in vitro* reactions may be provided in soluble forms or immobilized forms. In some embodiments, rebaudioside C, Dulcoside A, or other steviol rhamnosides can be produced using whole cells as discussed above. The cells can contain one recombinant UGT or multiple recombinant UGTs. For example, the cells can contain UGT 76G1 and EUGT11 such that mixtures of stevioside and RebA are efficiently converted to RebD. In some embodiments, the whole cells are the host cells described in section III A.

In other embodiments, the recombinant host expresses one or more genes involved in steviol biosynthesis, e.g., a CDPS gene, a KS gene, a KO gene and/or a KAH gene. Thus, for example, a microorganism containing a CDPS gene, a KS gene, a KO gene and a KAH gene, in addition to a UGT85C2, a UGT74G1, a EUGT11 gene, optionally a UGT91D2e gene, and a UGT76G1 gene, is capable of producing rebaudioside C without the necessity for including steviol in the culture media. In

addition, the recombinant host typically expresses an endogenous or a recombinant gene encoding a rhamnose synthetase. Such a gene is useful in order to provide increased amounts of the UDP-rhamnose donor for rhamnosylation of the steviol compound acceptor. Suitable rhamnose synthetases include those made by *Arabidopsis thaliana*, such as the product of the *A. thaliana* RHM2 gene.

One with skill in the art will recognize that by modulating relative expression levels of different UGT genes as well as modulating the availability of UDP-rhamnose, a recombinant host can be tailored to specifically produce steviol and steviol glycoside products in a desired proportion. Transcriptional regulation of steviol biosynthesis genes, and steviol glycoside biosynthesis genes can be achieved by a combination of transcriptional activation and repression using techniques known to those in the art. For *in vitro* reactions, one with skill in the art will recognize that addition of different levels of UGT enzymes in combination or under conditions which impact the relative activities of the different UGTS in combination will direct synthesis towards a desired proportion of each steviol glycoside.

In some embodiments, the recombinant host further contains and expresses a recombinant GGPPS gene in order to provide increased levels of the diterpene precursor geranylgeranyl diphosphate, for increased flux through the rebaudioside A biosynthetic pathway. In some embodiments, the recombinant host further contains a construct to silence or reduce the expression of non-steviol pathways consuming geranylgeranyl diphosphate, ent-Kaurenoic acid or farnesyl pyrophosphate, thereby providing increased flux through the steviol and steviol glycosides biosynthetic pathways. For example, flux to sterol production pathways such as ergosterol may be reduced by downregulation of the ERG9 gene. See, the ERG9 section below and Examples 24-25. In cells that produce gibberellins, gibberellin synthesis may be downregulated to increase flux of ent-kaurenoic acid to steviol. In carotenoid-producing organisms, flux to steviol may be increased by downregulation of one or more carotenoid biosynthetic genes.

In some embodiments, the recombinant host further contains and expresses recombinant genes involved in diterpene biosynthesis or production of terpenoid

precursors, e.g., genes in the MEP or MEV pathway, have reduced phosphatase activity, and/or express a SUS as discussed herein.

In some embodiments, a recombinant host such as a microorganism produces steviol glycoside compositions that have greater than at least 15% rebaudioside C of the total steviol glycosides, e.g., at least 20% rebaudioside C, 30-40% rebaudioside C, 40-50% rebaudioside C, 50-60% rebaudioside C, 60-70% rebaudioside C, 70-80% rebaudioside C, 80-90% rebaudioside C. In some embodiments, a recombinant host such as a microorganism produces steviol glycoside compositions that have at least 90% rebaudioside C, e.g., 90-99% rebaudioside C. Other steviol glycosides present may include those depicted in Figures 2 A and B such as steviol monosides, steviol glucobiosides, steviol rhamnobiosides, rebaudioside A, and Dulcoside A. In some embodiments, the rebaudioside C-enriched composition produced by the host can be further purified and the rebaudioside C or Dulcoside A so purified may then be mixed with other steviol glycosides, flavors, or sweeteners to obtain a desired flavor system or sweetening composition. For instance, a rebaudioside C-enriched composition produced by a recombinant microorganism can be combined with a rebaudioside A, F, or D-enriched composition produced by a different recombinant microorganism, with rebaudioside A, F, or D purified from a Stevia extract, or with rebaudioside A, F, or D produced *in vitro*.

B. 4 Rebaudioside F Biosynthesis Polypeptides

The biosynthesis of rebaudioside F involves glucosylation and xylosylation of the aglycone steviol. Specifically, rebaudioside F can be formed by glucosylation of the 13-OH of steviol which forms steviol-13-*O*-glucoside, xylosylation of the C-2' of the 13-O-glucose of steviol-13-*O*-glucoside which forms steviol-1,2-xylobioside, glucosylation of the C-19 carboxyl of the 1,2-xylobioside to form 1,2-stevioxyloside, and glucosylation of the C-3' of the C-13-O-glucose of 1,2-stevioxyloside to form rebaudioside F. The order in which each glycosylation reaction occurs can vary. See FIG. 2D.

It has been discovered that conversion of steviol to rebaudioside F in a recombinant host can be accomplished by the expression of genes encoding the following

functional UGTs: 85C2, EUGT11 and/or 91D2e, 74G1, and 76G1, along with endogenous or recombinantly expressed UDP-glucose dehydrogenase and UDP-glucuronic acid decarboxylase. Thus, a recombinant microorganism expressing these four or five UGTs along with endogenous or recombinant UDP-glucose dehydrogenase and UDP-glucuronic acid decarboxylase can make rebaudioside F when fed steviol in the medium. Alternatively, a recombinant microorganism expressing two functional UGTs, EUGT11 or 91D2e, and 76G1, can make rebaudioside F when fed rubusoside in the medium. As another alternative, a recombinant microorganism expressing a functional UGT 76G1 can make rebaudioside F when fed 1,2 steviorhamnoside. As another alternative, a recombinant microorganism expressing 74G1, EUGT11 and/or 91D2e, 76G1, and can make rebaudioside F when fed the monoside, steviol-13-*O*-glucoside, in the medium. Similarly, conversion of steviol-19-*O*-glucoside to rebaudioside F in a recombinant microorganism can be accomplished by the expression of genes encoding UGTs 85C2, EUGT11 and/or 91D2e, and 76G1, when fed steviol-19-*O*-glucoside. Typically, one or more of these genes are recombinant genes that have been transformed into a host that does not naturally possess them.

Suitable EUGT11, UGT91D2, UGT74G1, UGT76G1 and UGT85C2 polypeptides include the functional UGT polypeptides discussed herein. In some embodiments, a UGT79B3 polypeptide is substituted for a UGT91, as discussed above. UDP-glucose dehydrogenase and UDP-glucuronic acid decarboxylase provide increased amounts of the UDP-xylose donor for xylosylation of the steviol compound acceptor. Suitable UDP-glucose dehydrogenases and UDP-glucuronic acid decarboxylases include those made by *Arabidopsis thaliana* or *Cryptococcus neoformans*. For example, suitable UDP-glucose dehydrogenase and UDP-glucuronic acid decarboxylases polypeptides can be encoded by the *A. thaliana* UGD1 gene and UXS3 gene, respectively. See, Oka and Jigami, FEBS J. 273:2645-2657 (2006).

In some embodiments rebaudioside F can be produced using *in vitro* methods while supplying the appropriate UDP-sugar and/or a cell-free system for regeneration of UDP-sugars. See, for example, Jewett MC, et al. Molecular Systems Biology, Vol. 4, article 220 (2008); Masada S et al. FEBS Letters, Vol. 581, 2562-2566 (2007). In some

embodiments, sucrose and a sucrose synthase are provided in the reaction vessel in order to regenerate UDP-glucose from UDP during the glycosylation reactions. See FIG. 11. The sucrose synthase can be from any suitable organism. For example, a sucrose synthase coding sequence from *Arabidopsis thaliana*, *Stevia rebaudiana*, or *Coffea arabica* can be cloned into an expression plasmid under control of a suitable promoter, and expressed in a host, e.g., a microorganism or a plant. In some embodiments, UDP-xylose can be produced from UDP-glucose by supplying suitable enzymes, for example, the *Arabidopsis thaliana* UGD1 (UDP-glucose dehydrogenase) and UXS3 (UDP-glucuronic acid decarboxylase) enzymes along with NAD+ cofactor.

Reactions may be carried out together, or stepwise. For instance, rebaudioside F may be produced from rubusoside with the addition of stoichiometric amounts of UDP-xylose and EUGT11, followed by addition of UGT76G1 and an excess or stoichiometric supply of UDP-glucose. In some embodiments, phosphatases are used to remove secondary products and improve the reaction yields. UGTs and other enzymes for *in vitro* reactions may be provided in soluble forms or immobilized forms. In some embodiments, rebaudioside F or other steviol xylosides can be produced using whole cells as discussed above. For example, the cells may contain UGT 76G1 and EUGT11 such that mixtures of stevioside and RebA are efficiently converted to RebD. In some embodiments, the whole cells are the host cells described in section III A.

In other embodiments, the recombinant host expresses one or more genes involved in steviol biosynthesis, e.g., a CDPS gene, a KS gene, a KO gene and/or a KAH gene. Thus, for example, a microorganism containing a CDPS gene, a KS gene, a KO gene and a KAH gene, in addition to a EUGT11, UGT85C2, a UGT74G1, an optional UGT91D2 gene, and a UGT76G1 gene, is capable of producing rebaudioside F without the necessity for including steviol in the culture media. In addition, the recombinant host typically expresses an endogenous or a recombinant gene encoding a UDP-glucose dehydrogenase and a UDP-glucuronic acid decarboxylase. Such genes are useful in order to provide increased amounts of the UDP-xylose donor for xylosylation of the steviol compound acceptor. Suitable UDP-glucose dehydrogenases and UDP-glucuronic acid decarboxylases include those made by *Arabidopsis thaliana* or *Cryptococcus*

neoformans. For example, suitable UDP-glucose dehydrogenase and UDP-glucuronic acid decarboxylases polypeptides can be encoded by the *A. thaliana* UGD1 gene and UXS3 gene, respectively. See, Oka and Jigami, FEBS J. 273:2645-2657 (2006).

One with skill in the art will recognize that by modulating relative expression levels of different UGT genes as well as modulating the availability of UDP-xylose, a recombinant microorganism can be tailored to specifically produce steviol and steviol glycoside products in a desired proportion. Transcriptional regulation of steviol biosynthesis genes can be achieved by a combination of transcriptional activation and repression using techniques known to those in the art. For *in vitro* reactions, one with skill in the art will recognize that addition of different levels of UGT enzymes in combination or under conditions which impact the relative activities of the different UGTS in combination will direct synthesis towards a desired proportion of each steviol glycosides.

In some embodiments, the recombinant host further contains and expresses a recombinant GGPPS gene in order to provide increased levels of the diterpene precursor geranylgeranyl diphosphate, for increased flux through the steviol biosynthetic pathway. In some embodiments, the recombinant host further contains a construct to silence the expression of non-steviol pathways consuming geranylgeranyl diphosphate, ent-Kaurenoic acid or farnesyl pyrophosphate, thereby providing increased flux through the steviol and steviol glycosides biosynthetic pathways. For example, flux to sterol production pathways such as ergosterol may be reduced by downregulation of the ERG9 gene. See, the ERG9 section below and Examples 24-25. In cells that produce gibberellins, gibberellin synthesis may be downregulated to increase flux of ent-kaurenoic acid to steviol. In carotenoid-producing organisms, flux to steviol may be increased by downregulation of one or more carotenoid biosynthetic genes. In some embodiments, the recombinant host further contains and expresses recombinant genes involved in diterpene biosynthesis, e.g., genes in the MEP pathway discussed below.

In some embodiments, a recombinant host such as a microorganism produces rebaudioside F-enriched steviol glycoside compositions that have greater than at least 4% rebaudioside F by weight total steviol glycosides, e.g., at least 5% rebaudioside F, at

least 6% of rebaudioside F, 10-20% rebaudioside F, 20-30% rebaudioside F, 30-40% rebaudioside F, 40-50% rebaudioside F, 50-60% rebaudioside F, 60-70% rebaudioside F, 70-80% rebaudioside F. In some embodiments, a recombinant host such as a microorganism produces steviol glycoside compositions that have at least 90% rebaudioside F, e.g., 90-99% rebaudioside F. Other steviol glycosides present may include those depicted in FIGs 2A and D such as steviol monosides, steviol glucobiosides, steviol xylobiosides, rebaudioside A, stevioxyloside, rubusoside and stevioside. In some embodiments, the rebaudioside F-enriched composition produced by the host can be mixed with other steviol glycosides, flavors, or sweeteners to obtain a desired flavor system or sweetening composition. For instance, a rebaudioside F-enriched composition produced by a recombinant microorganism can be combined with a rebaudioside A, C, or D-enriched composition produced by a different recombinant microorganism, with rebaudioside A, C, or D purified from a Stevia extract, or with rebaudioside A, C, or D produced *in vitro*.

C. Other Polypeptides

Genes for additional polypeptides whose expression facilitates more efficient or larger scale production of steviol or a steviol glycoside can also be introduced into a recombinant host. For example, a recombinant microorganism, plant, or plant cell can also contain one or more genes encoding a geranylgeranyl diphosphate synthase (GGPPS, also referred to as GGDPS). As another example, the recombinant host can contain one or more genes encoding a rhamnose synthetase, or one or more genes encoding a UDP-glucose dehydrogenase and/or a UDP-glucuronic acid decarboxylase. As another example, a recombinant host can also contain one or more genes encoding a cytochrome P450 reductase (CPR). Expression of a recombinant CPR facilitates the cycling of NADP+ to regenerate NADPH, which is utilized as a cofactor for terpenoid biosynthesis. Other methods can be used to regenerate NADHP levels as well. In circumstances where NADPH becomes limiting; strains can be further modified to include exogenous transhydrogenase genes. See, e.g., Sauer et al., J. Biol. Chem. 279: 6613–6619 (2004).

Other methods are known to those with skill in the art to reduce or otherwise modify the ratio of NADH/NADPH such that the desired cofactor level is increased.

As another example, the recombinant host can contain one or more genes encoding one or more enzymes in the MEP pathway or the mevalonate pathway. Such genes are useful because they can increase the flux of carbon into the diterpene biosynthesis pathway, producing geranylgeranyl diphosphate from isopentenyl diphosphate and dimethylallyl diphosphate generated by the pathway. The geranylgeranyl diphosphate so produced can be directed towards steviol and steviol glycoside biosynthesis due to expression of steviol biosynthesis polypeptides and steviol glycoside biosynthesis polypeptides.

As another example the recombinant host can contain one or more genes encoding a sucrose synthase, and additionally can contain sucrose uptake genes if desired. The sucrose synthase reaction can be used to increase the UDP-glucose pool in a fermentation host, or in a whole cell bioconversion process. This regenerates UDP-glucose from UDP produced during glycosylation and sucrose, allowing for efficient glycosylation. In some organisms, disruption of the endogenous invertase is advantageous to prevent degradation of sucrose. For example, the *S. cerevisiae* SUC2 invertase may be disrupted. The sucrose synthase (SUS) can be from any suitable organism. For example, a sucrose synthase coding sequence from, without limitation, *Arabidopsis thaliana*, *Stevia rebaudiana*, or *Coffea arabica* can be cloned into an expression plasmid under control of a suitable promoter, and expressed in a host (e.g., a microorganism or a plant). The sucrose synthase can be expressed in such a strain in combination with a sucrose transporter (e.g., the *A. thaliana* SUC1 transporter or a functional homolog thereof) and one or more UGTs (e.g., one or more of UGT85C2, UGT74G1, UGT76G1, and UGT91D2e, EUGT11 or functional homologs thereof). Culturing the host in a medium that contains sucrose can promote production of UDP-glucose, as well as one or more glucosides (e.g., steviol glycosides).

In addition, a recombinant host can have reduced phosphatase activity as discussed herein.

C. 1 MEP Biosynthesis Polypeptides

In some embodiments, a recombinant host contains one or more genes encoding enzymes involved in the methylerythritol 4-phosphate (MEP) pathway for isoprenoid biosynthesis. Enzymes in the MEP pathway include deoxyxylulose 5-phosphate synthase (DXS), D-1-deoxyxylulose 5-phosphate reductoisomerase (DXR), 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (CMS), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), 4-diphosphocytidyl-2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS), 1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate synthase (HDS) and 1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate reductase (HDR). One or more DXS genes, DXR genes, CMS genes, CMK genes, MCS genes, HDS genes and/or HDR genes can be incorporated into a recombinant microorganism. See, Rodríguez-Concepción and Boronat, Plant Phys. 130: 1079-1089 (2002).

Suitable genes encoding DXS, DXR, CMS, CMK, MCS, HDS and/or HDR polypeptides include those made by *E. coli*, *Arabidopsis thaliana* and *Synechococcus leopoliensis*. Nucleotide sequences encoding DXR polypeptides are described, for example, in U.S. Patent No. 7,335,815.

C. 2 Mevalonate Biosynthesis Polypeptides

In some embodiments, a recombinant host contains one or more genes encoding enzymes involved in the mevalonate pathway for isoprenoid biosynthesis. Genes suitable for transformation into a host encode enzymes in the mevalonate pathway such as a truncated 3-hydroxy-3-methyl-glutaryl (HMG)-CoA reductase (tHMG), and/or a gene encoding a mevalonate kinase (MK), and/or a gene encoding a phosphomevalonate kinase (PMK), and/or a gene encoding a mevalonate pyrophosphate decarboxylase (MPPD). Thus, one or more HMG-CoA reductase genes, MK genes, PMK genes, and/or MPPD genes can be incorporated into a recombinant host such as a microorganism.

Suitable genes encoding mevalonate pathway polypeptides are known. For example, suitable polypeptides include those made by *E. coli*, *Paracoccus denitrificans*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Kitasatospora griseola*, *Homo sapiens*, *Drosophila melanogaster*, *Gallus gallus*, *Streptomyces* sp. KO-3988, *Nicotiana attenuata*,

Kitasatospora griseola, *Hevea brasiliensis*, *Enterococcus faecium* and *Haematococcus pluvialis*. See, e.g., Table 8 and U.S. Patent Nos. 7,183,089, 5,460,949, and 5,306,862.

Table 8. Sources of HMG CoA Reductases and other Mevalonate Genes

Accession#	Organism	Enzyme	Size (nt)	Gene name	SEQ ID (codon optimized)	SEQ ID (protein)
XM_001467423	*Leishmania infantum*	Acetyl-CoA C-acetyltransferase	1323	MEV-4	103	104
YML075C	*Saccharomyces cerevisiae*	Truncated HMG (tHMG1)	1584	tHMG1	105	106
EU263989	*Ganoderma lucidum*	3-HMG-CoA reductase	3681	MEV-11	107	108
BC153262	*Bos taurus*	3-HMG-CoA reductase	2667	MEV-12	109	110
AAD47596	*Artemisia annua*	3-HMG-CoA reductase	1704	MEV-13	111	112
AAB62280	*Trypanosoma cruzi*	3-HMG-CoA reductase	1308	MEV-14	113	114
CAG41604	*Staph aureus*	3-HMG-CoA reductase	1281	MEV-15	115	116
DNA2.0 sequence	*Archaeoglobus fulgidus*	3-HMG-CoA reductase	1311	HMG reductase	117	118
DNA2.0 sequence	*Pseudomonas mevalonii*	3-HMG-CoA reductase	1287	HMG reductase	119	120

C.3 Sucrose Synthase Polypeptides

Sucrose synthase (SUS) can be used as a tool for generating UDP-sugar. SUS (EC 2.4.1.13) catalyzes the formation of UDP-glucose and fructose from sucrose and UDP (FIG. 11). UDP generated by the reaction of UGTs thus can be converted into UDP-glucose in the presence of sucrose. *See*, e.g., Chen et al. (2001) *J. Am. Chem. Soc.* 123:8866-8867; Shao et al. (2003) *Appl. Env. Microbiol.* 69:5238-5242; Masada et al. (2007) *FEBS Lett.* 581:2562-2566; and Son et al. (2009) *J. Microbiol. Biotechnol.* 19:709-712.

Sucrose synthases can be used to generate UDP-glucose and remove UDP, facilitating efficient glycosylation of compounds in various systems. For example, yeast deficient in the ability to utilize sucrose can be made to grow on sucrose by introducing a

sucrose transporter and a SUS. For example, *Saccharomyces cerevisiae* does not have an efficient sucrose uptake system, and relies on extracellular SUC2 to utilize sucrose. The combination of disrupting the endogenous *S. cerevisiae* SUC2 invertase and expressing recombinant SUS resulted in a yeast strain that was able to metabolize intracellular but not extracellular sucrose (Riesmeier et al. ((1992) *EMBO J.* 11:4705-4713). The strain was used to isolate sucrose transporters by transformation with a cDNA expression library and selection of transformants that had gained the ability to take up sucrose.

As described herein, the combined expression of recombinant sucrose synthase and a sucrose transporter *in vivo* can lead to increased UDP-glucose availability and removal of unwanted UDP. For example, functional expression of a recombinant sucrose synthase, a sucrose transporter, and a glycosyltransferase, in combination with knockout of the natural sucrose degradation system (SUC2 in the case of *S. cerevisiae*) can be used to generate a cell that is capable of producing increased amounts of glycosylated compounds such as steviol glycosides. This higher glycosylation capability is due to at least (a) a higher capacity for producing UDP-glucose in a more energy efficient manner, and (b) removal of UDP from growth medium, as UDP can inhibit glycosylation reactions.

The sucrose synthase can be from any suitable organism. For example, a sucrose synthase coding sequence from, without limitation, *Arabidopsis thaliana*, *Stevia rebaudiana*, or *Coffea arabica* (see, e.g., FIGs. 19A-19C, SEQ ID NOs:178, 179, and 180) can be cloned into an expression plasmid under control of a suitable promoter, and expressed in a host (e.g., a microorganism or a plant). As described in the Examples herein, a SUS coding sequence may be expressed in a *SUC2* (sucrose hydrolyzing enzyme) deficient *S. cerevisiae* strain, so as to avoid degradation of extracellular sucrose by the yeast. The sucrose synthase can be expressed in such a strain in combination with a sucrose transporter (e.g., the *A. thaliana* SUC1 transporter or a functional homolog thereof) and one or more UGTs (e.g., one or more of UGT85C2, UGT74G1, UGT76G1, EUGT11, and UGT91D2e, or functional homologs thereof). Culturing the host in a medium that contains sucrose can promote production of UDP-glucose, as well as one or more glucosides (e.g., steviol glucoside). It is to be noted that in some cases, a sucrose

synthase and a sucrose transporter can be expressed along with a UGT in a host cell that also is recombinant for production of a particular compound (e.g., steviol).

C. 4 Modulation of ERG9 activity

It is an object of the disclosure to produce terpenoids based on the concept of increasing the accumulation of terpenoid precursors of the squalene pathway. Non-limiting examples of terpenoids include Hemiterpenoids, 1 isoprene unit (5 carbons); Monoterpenoids, 2 isoprene units (10C); Sesquiterpenoids, 3 isoprene units (15C); Diterpenoids, 4 isoprene units (20C) (e.g. ginkgolides); Triterpenoids, 6 isoprene units (30C); Tetraterpenoids, 8 isoprene units (40C) (e.g. carotenoids); and polyterpenoid with a larger number of isoprene units.

Hemiterpenoids include isoprene, prenol and isovaleric acid. Monoterpenoids include Geranyl pyrophosphate, Eucalyptol, Limonene and Pinene. Sesquiterpenoids include Farnesyl pyrophosphate, Artemisinin and Bisabolol. Diterpenoids include Geranylgeranyl pyrophosphate, steviol, Retinol, Retinal, Phytol, Taxol, Forskolin and Aphidicolin. Triterpenoids include Squalene and Lanosterol. Tetraterpenoids include Lycopene and Carotene.

Terpenes are hydrocarbons resulting from the combination of several isoprene units. Terpenoids can be thought of as terpene derivatives. The term terpene is sometimes used broadly to include the terpenoids. Just like terpenes, the terpenoids can be classified according to the number of isoprene units used. The present invention is focussed on terpenoids and in particular terpenoids derived through the squalene pathway from the precursors Farnesyl-pyrophosphate (FPP), Isopentenyl-pyrophosphate (IPP), Dimethylallyl-pyrophosphate (DMAPP), Geranyl-pyrophosphate (GPP) and/or Geranylgeranyl-pyrophosphate (GGPP).

By terpenoids is understood terpenoids of the Hemiterpenoid class such as but not limited to isoprene, prenol and isovaleric acid; terpenoids of the Monoterpenoid class such as but not limited to geranyl pyrophosphate, eucalyptol, limonene and pinene; terpenoids of the Sesquiterpenoids class such as but not limited to farnesyl pyro-phosphate, artemisinin and bisabolol; terpenoids of the diterpenoid class such as but not

limited to geranylgeranyl pyrophosphate, steviol, retinol, retinal, phytol, taxol , forskolin and aphidicolin; terpenoids of the Triterpenoid class such as but not limited to lanosterol; terpenoids of the Tetraterpenoid class such as but not limited to lycopene and carotene.

In one embodiment the invention relates to production of terpenoids, which are biosynthesized from Geranylgeranyl-pyrophosphate (GGPP). In particular such terpenoids may be steviol.

In one embodiment the invention relates to production of terpenoids, which are biosynthesized from Geranylgeranyl-pyrophosphate (GGPP). In particular such terpenoids may be steviol.

The cell

The present invention relates to a cell, such as any of the hosts described in section III, modified to comprise the construct depicted in FIG. 22. Accordingly, in a main aspect, the present invention relates to a cell comprising a nucleic acid, said nucleic acid comprising

i) a promoter sequence operably linked to
ii) a heterologous insert sequence operably linked to
iii) an open reading frame operably linked to
iv) a transcription termination signal,

wherein the heterologous insert sequence has the general formula (I):

$$-X_1-X_2-X_3-X_4-X_5-$$

wherein X_2 comprises at least 4 consecutive nucleotides being complementary to, and forms a hairpin secondary structure element with at least 4 consecutive nucleotides of X_4, and

wherein X_3 is optional and if present comprises nucleotides involved in forming a hairpin loop between X_2 and X_4, and

wherein X_1 and X_5 individually and optionally comprise one or more nucleotides, and

wherein the open reading frame upon expression encodes a squalene synthase (EC 2.5.1.21), e.g., a polypeptide sequence having at least 70% identity to a squalene synthase

(EC 2.5.1.21) or a biologically active fragment thereof, said fragment having at least 70% sequence identity to said squalene synthase in a range of overlap of at least 100 amino acids.

In addition to above mentioned nucleic acid comprising a heterologous insert sequence, the cell may also comprise one or more additional heterologous nucleic acid sequences (e.g., nucleic acids encoding any of the steviol and steviol glycoside biosynthesis polypeptides of section I). In one preferred embodiment the cell comprises a heterologous nucleic acid encoding GGPPS operably linked to a nucleic acid sequence directing expression of GGPPS in said cell.

Heterologous insert sequence

The heterologous insert sequence can adapt the secondary structure element of a hairpin with a hairpin loop. The hairpin part comprises sections X_2 and X_4 which are complementary and hybridize to one another. Sections X_2 and X_4 flank section X_3, which comprises nucleotides that form a loop – the hairpin loop. The term complementary is understood by the person skilled in the art as meaning two sequences compared to each other, nucleotide by nucleotide counting from the 5' end to the 3' end, or vice versa.

The heterologous insert sequence is long enough to allow a hairpin to be completed, but short enough to allow limited translation of an ORF that is present in-frame and immediately 3' to the heterologous insert sequence. Thus, in one embodiment, the heterologous insert sequence comprises 10-50 nucleotides, preferably 10-30 nucleotides, more preferably 15-25 nucleotides, more preferably 17-22 nucleotides, more preferably 18-21 nucleotides, more preferably 18-20 nucleotides, more preferably 19 nucleotides.

X_2 and X_4 may individually consist of any suitable number of nucleotides, so long as a consecutive sequence of at least 4 nucleotides of X_2 is complementary to a consecutive sequence of at least 4 nucleotides of X_4. In a preferred embodiment X_2 and X_4 consist of the same number of nucleotides.

X_2 may for example consist of in the range of 4 to 25, such as in the range of 4 to 20, for example of in the range of 4 to 15, such as in the range of 6 to 12, for example in

the range of 8 to 12, such as in the range of 9 to 11 nucleotides.

X_4 may for example consist of in the range of 4 to 25, such as in the range of 4 to 20, for example of in the range of 4 to 15, such as in the range of 6 to 12, for example in the range of 8 to 12, such as in the range of 9 to 11 nucleotides.

In one preferred embodiment, X_2 consists of a nucleotide sequence, which is complementary to the nucleotide sequence of X_4, i.e. it is preferred that all nucleotides of X_2 are complementary to the nucleotide sequence of X_4.

In one preferred embodiment X_4 consists of a nucleotide sequence, which is complementary to the nucleotide sequence of X_2, i.e. it is preferred that all nucleotides of X_4 are complementary to the nucleotide sequence of X_2. Very preferably, X_2 and X_4 consists of the same number of nucleotides, wherein X_2 is complementary to X_4 over the entire length of X_2 and X_4.

X_3 may be absent, i.e., X_3 may consist of zero nucleotides. It is also possible that X_3 consists of in the range of 1 to 5, such as in the range of 1 to 3 nucleotides.

X_1 may be absent, i.e., X_1 may consist of zero nucleotides. It is also possible that X_1 consists of in the range of 1 to 25, such as in the range of 1 to 20, for example in the range of 1 to 15, such as in the range of 1 to 10, for example in the range of 1 to 5, such as in the range of 1 to 3 nucleotides.

X_5 may be absent, i.e., X_5 may consist of zero nucleotides. It is also possible that X_5 may consist of in the range 1 to 5, such as in the range of 1 to 3 nucleotides.

The sequence may be any suitable sequence fulfilling the requirements defined herein above. Thus, the heterologous insert sequence may comprise a sequence selected from the group consisting of SEQ ID NO: 181, SEQ ID NO:182, SEQ ID NO:183, and SEQ ID NO:184. In a preferred embodiment the insert sequence is selected from the group consisting of SEQ ID NO: 181, SEQ ID NO:182, SEQ ID NO:183, and SEQ ID NO:184.

Squalene synthase

Squalene synthase (SQS) is the first committed enzyme of the biosynthesis pathway that leads to the production of sterols. It catalyzes the synthesis of squalene from farnesyl pyrophosphate via the intermediate presqualene pyrophosphate. This enzyme is a

critical branch point enzyme in the biosynthesis of terpenoids/isoprenoids and is thought to regulate the flux of isoprene intermediates through the sterol pathway. The enzyme is sometimes referred to as farnesyl-diphosphate farnesyltransferase (FDFT1).

The mechanism of SQS is to convert two units of farnesyl pyrophosphate into squalene.

SQS is considered to be an enzyme of eukaryotes or advanced organisms, although at least one prokaryote has been shown to possess a functionally similar enzyme.

In terms of structure and mechanics, squalene synthase most closely resembles phytoene syntase, which serves a similar role in many plants in the elaboration of phytoene, a precursor of many carotenoid compounds.

A high level of sequence identity indicates likelihood that the first sequence is derived from the second sequence. Amino acid sequence identity requires identical amino acid sequences between two aligned sequences. Thus, a candidate sequence sharing 70% amino acid identity with a reference sequence requires that, following alignment, 70% of the amino acids in the candidate sequence are identical to the corresponding amino acids in the reference sequence. Identity may be determined by aid of computer analysis, such as, without limitations, the ClustalW computer alignment program as described in section D. Using this program with its default settings, the mature (bioactive) part of a query and a reference polypeptide are aligned. The number of fully conserved residues are counted and divided by the length of the reference polypeptide. The ClustalW algorithm may similarly be used to align nucleotide sequences. Sequence identities may be calculated in a similar way as indicated for amino acid sequences.

In one important embodiment, the cell of the present invention comprises a nucleic acid sequence coding, as defined herein, upon expression for a squalene synthase wherein the squalene synthase is at least 75%, such as at least 76%, such as at least 77%, such as at least 78%, such as at least 79%, such as at least 80%, such as at least 81%, such as at least 82%, such as at least 83%, such as at least 84%, such as at least 85%, such as at least 86%, such as at least 87%, such as at least 88%, such as at least 89%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%,

such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as at least 99.5%, such as at least 99.6%, such as at least 99.7%, such as at least 99.8%, such as at least 99.9%, such as 100% identical to a squalene synthase wherein the squalene synthase is selected from the group consisting of SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, and SEQ ID NO:202.

Promoter

A promoter is a region of DNA that facilitates the transcription of a particular gene. Promoters are located near the genes they regulate, on the same strand and typically upstream (towards the 5' region of the sense strand). In order for the transcription to take place, the enzyme that synthesizes RNA, known as RNA polymerase, must attach to the DNA near a gene. Promoters contain specific DNA sequences and response elements which provide a secure initial binding site for RNA polymerase and for proteins called transcription factors that recruit RNA polymerase. These transcription factors have specific activator or repressor sequences of corresponding nucleotides that attach to specific promoters and regulate gene expressions.

In bacteria, the promoter is recognized by RNA polymerase and an associated sigma factor, which in turn are often brought to the promoter DNA by an activator protein binding to its own DNA binding site nearby. In eukaryotes the process is more complicated, and at least seven different factors are necessary for the binding of an RNA polymerase II to the promoter. Promoters represent critical elements that can work in concert with other regulatory regions (enhancers, silencers, boundary elements/insulators) to direct the level of transcription of a given gene.

As promoters are normally immediately adjacent to the open reading frame (ORF) in question, positions in the promoter are designated relative to the transcriptional start site, where transcription of RNA begins for a particular gene (i.e., positions upstream are negative numbers counting back from -1, for example -100 is a position 100 base pairs upstream).

Promoter elements

- Core promoter - the minimal portion of the promoter required to properly initiate transcription
  - Transcription Start Site (TSS)
  - Approximately -35 bp upstream and/or downstream of the start site
  - A binding site for RNA polymerase
    - RNA polymerase I: transcribes genes encoding ribosomal RNA
    - RNA polymerase II: transcribes genes encoding messenger RNA and certain small nuclear RNAs
    - RNA polymerase III: transcribes genes encoding tRNAs and other small RNAs
  - General transcription factor binding sites
- Proximal promoter - the proximal sequence upstream of the gene that tends to contain primary regulatory elements
  - Approximately -250 bp upstream of the start site
  - Specific transcription factor binding sites
- Distal promoter - the distal sequence upstream of the gene that may contain additional regulatory elements, often with a weaker influence than the proximal promoter
  - Anything further upstream (but not an enhancer or other regulatory region whose influence is positional/orientation independent)
  - Specific transcription factor binding sites

Prokaryotic promoters

In prokaryotes, the promoter consists of two short sequences at -10 and -35 positions upstream from the transcription start site. Sigma factors not only help in enhancing RNAP binding to the promoter but also help RNAP target specific genes to transcribe.

The sequence at -10 is called the Pribnow box, or the -10 element, and usually consists of the six nucleotides TATAAT. The Pribnow box is essential to start transcription in prokaryotes.

The other sequence at -35 (the -35 element) usually consists of the seven

nucleotides TTGACAT. Its presence allows a very high transcription rate.

Both of the above consensus sequences, while conserved on average, are not found intact in most promoters. On average only 3 of the 6 base pairs in each consensus sequence is found in any given promoter. No promoter has been identified to date that has intact consensus sequences at both the -10 and -35; artificial promoters with complete conservation of the -10/-35 hexamers has been found to promote RNA chain initiation at very high efficiencies.

Some promoters contain a UP element (consensus sequence 5'-AAAWWTWTTTTNNNAAANNN-3'; W = A or T; N = any base) centered at -50; the presence of the -35 element appears to be unimportant for transcription from the UP element-containing promoters.

Eukaryotic promoters

Eukaryotic promoters are typically located upstream of the gene (ORF) and can have regulatory elements several kilobases (kb) away from the transcriptional start site. In eukaryotes, the transcriptional complex can cause the DNA to fold back on itself, which allows for placement of regulatory sequences far from the actual site of transcription. Many eukaryotic promoters, contain a TATA box (sequence TATAAA), which in turn binds a TATA binding protein which assists in the formation of the RNA polymerase transcriptional complex. The TATA box typically lies very close to the transcriptional start site (often within 50 bases).

The cell of the present invention comprises a nucleic acid sequence which comprises a promoter sequence. The promoter sequence is not limiting for the invention and can be any promoter suitable for the host cell of choice.

In one embodiment of the present invention the promoter is a constitutive or inducible promoter.

In a further embodiment of the invention, the promoter is selected from the group consisting of an endogenous promoter, PGK-1, GPD1, PGK1, ADH1, ADH2, PYK1, TPI1, PDC1, TEF1, TEF2, FBA1, GAL1-10, CUP1, MET2, MET14, MET25, CYC1, GAL1-S, GAL1-L, TEF1, ADH1, CAG, CMV, human UbiC, RSV, EF-1alpha, SV40, Mt1, Tet-On, Tet-Off, Mo-MLV-LTR, Mx1, progesterone, RU486 and Rapamycin-

inducible promoter.

Post-transcriptional regulation

Post-transcriptional regulation is the control of gene expression at the RNA level, therefore between the transcription and the translation of the gene.

The first instance of regulation is at transcription (transcriptional regulation) where due to the chromatin arrangement and due to the activity of transcription factors, genes are differentially transcribed.

After being produced, the stability and distribution of the different transcripts is regulated (post-transcriptional regulation) by means of RNA binding protein (RBP) that control the various steps and rates of the transcripts: events such as alternative splicing, nuclear degradation (exosome), processing, nuclear export (three alternative pathways), sequestration in DCP2-bodies for storage or degradation, and ultimately translation. These proteins achieve these events thanks to a RNA recognition motif (RRM) that binds a specific sequence or secondary structure of the transcripts, typically at the 5' and 3' UTR of the transcript.

Modulating the capping, splicing, addition of a Poly(A) tail, the sequence-specific nuclear export rates and in several contexts sequestration of the RNA transcript occurs in eukaryotes but not in prokaryotes. This modulation is a result of a protein or transcript which in turn is regulated and may have an affinity for certain sequences.

Capping

Capping changes the five prime end of the mRNA to a three prime end by 5'-5' linkage, which protects the mRNA from 5' exonuclease, which degrades foreign RNA. The cap also helps in ribosomal binding.

Splicing

Splicing removes the introns, noncoding regions that are transcribed into RNA, in order to make the mRNA able to create proteins. Cells do this by spliceosomes binding on either side of an intron, looping the intron into a circle and then cleaving it off. The two ends of the exons are then joined together.

Polyadenylation

Polyadenylation is the addition of a poly(A) tail to the 3' end, i.e. the poly(A) tail

consists of multiple adenosine monophosphates. The poly-A sequence acts as a buffer to the 3' exonuclease and thus increases half-life of mRNA. In addition, a long poly(A) tail can increase translation. Thus the poly-A tail may be used to further modulate translation of the construct of the present invention, in order to arrive at the optimal translation rate.

In eukaryotes, polyadenylation is part of the process that produces mature messenger RNA (mRNA) for translation.

The poly(A) tail is also important for the nuclear export, translation, and stability of mRNA.

In one embodiment the nucleic acid sequence of the cell of the present invention, as defined herein above, further comprises a polyadenyl/polyadenylation sequence, preferably the 5' end of said polyadenyl l/polyadenylation sequence is operably linked to the 3' end of the open reading frame, such as to the open reading frame encoding squalene synthase.

RNA editing

RNA editing is a process which results in sequence variation in the RNA molecule, and is catalyzed by enzymes. These enzymes include the Adenosine Deaminase Acting on RNA (ADAR) enzymes, which convert specific adenosine residues to inosine in an mRNA molecule by hydrolytic deamination. Three ADAR enzymes have been cloned, ADAR1, ADAR2 and ADAR3, although only the first two subtypes have been shown to have RNA editing activity. Many mRNAs are vulnerable to the effects of RNA editing, including the glutamate receptor subunits GluR2, GluR3, GluR4, GluR5 and GluR6 (which are components of the AMPA and kainate receptors), the serotonin2C receptor, the GABA-alpha3 receptor subunit, the tryptophan hydroxlase enzyme TPH2, the hepatitis delta virus and more than 16% of microRNAs. In addition to ADAR enzymes, CDAR enzymes exist and these convert cytosines in specific RNA molecules, to uracil. These enzymes are termed 'APOBEC' and have genetic loci at 22q13, a region close to the chromosomal deletion which occurs in velocardiofacial syndrome (22q11) and which is linked to psychosis. RNA editing is extensively studied in relation to infectious diseases, because the editing process alters viral function.

Post-transcriptional regulatory elements

Use of a post-transcriptional regulatory elements (PRE) is often necessary to obtain vectors with sufficient performance for certain applications. Schambach et al in Gene Ther. (2006) 13(7):641-5 reports that introduction of a post-transcriptional regulatory element (PRE) of woodchuck hepatitis virus (WHV) into the 3' untranslated region of retroviral and lentiviral gene transfer vectors enhances both titer and transgene expression. The enhancing activity of the PRE depends on the precise configuration of its sequence and the context of the vector and cell into which it is introduced.

Thus use of a PRE such as a woodchuck hepatitis virus post-transcriptional regulatory elements (WPRE) may be useful in the preparation of the cell of the present invention when using a gene therapeutic approach.

Accordingly, in one embodiment the nucleic acid sequence of the cell defined herein further comprises a post-transcriptional regulatory element.

In a further embodiment, the post-transcriptional regulatory element is a Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE).

Terminal repeats

To insert genetic sequences into host DNA, viruses often use sequences of DNA that repeats up to thousands of times, so called repeats, or terminal repeats including long terminal repeats (LTR) and inverted terminal repeats (ITR), wherein said repeat sequences may be both 5' and 3' terminal repeats. ITRs aid in concatamer formation in the nucleus after the single-stranded vector DNA is converted by host cell DNA polymerase complexes into double-stranded DNA. ITR sequences may be derived from viral vectors, such as AAV, e.g. AAV2.

In one embodiment, the nucleic acid sequence or the vector of the cell defined herein comprises a 5' terminal repeat and a 3' terminal repeat.

In one embodiment said 5' and 3' terminal repeats are selected from Inverted Terminal Repeats [ITR] and Long Terminal Repeats [LTR].

In one embodiment of said 5' and 3' terminal repeats are AAV Inverted Terminal Repeats [ITR].

Geranylgeranyl Pyrophosphate Synthase

The microbial cells of the present invention may in preferred embodiments

contain a heterologous nucleic acid sequence encoding Geranylgeranyl Pyrophosphate Synthase (GGPPS). See, e.g., Table 7. GGPPS is an enzyme, which catalyzes the chemical reaction that turns one farnesyl pyrophosphate (FPP) molecule into one Geranylgeranyl Pyrophosphate (GGPP) molecule. Genes encoding GGPPS may for example be found in organisms that contain the mevalonate pathway.

The GGPPS to be used with the present invention may be any useful enzyme, which is capable of catalysing conversion of a farnesyl pyrophosphate (FPP) molecule into a Geranylgeranyl Pyrophosphate (GGPP) molecule. In particular, the GGPPS to be used with the present invention may be any enzyme capable of catalysing the following reaction:

(2E,6E)-farnesyl diphosphate + isopentenyl diphosphate -> diphosphate + geranylgeranyl diphosphate.

It is preferred that the GGPPS used with the present invention is an enzyme categorised under EC 2.5.1.29.

The GGPPS may be GGPPS from a variety of sources, such as from bacteria, fungi or mammals. The GGPPS may be any kind of GGPPS, for example GGPPS-1, GGPPS-2, GGPPS-3 or GGPPS-4. The GGPPS may be wild type GGPPS or a functional homologue thereof.

For example, the GGPPS may be GGPPS-1 of *S. acidicaldarius* (SEQ ID NO: 126), GGPPS-2 of *A.nidulans* (SEQ ID NO: 203), GGPPS-3 of *S. cerevisiae* (SEQ ID NO: 167) or GGPPS-4 of *M. musculus* (SEQ ID NO:123) or a functional homologue of any of the aforementioned.

The heterologous nucleic acid encoding said GGPPS may be any nucleic acid sequence encoding said GGPPS. Thus, in embodiments of the invention where GGPPS is a wild type protein, the nucleic acid sequence may for example be a wild type cDNA sequence encoding said protein. However, it is frequently the case that the heterologous nucleic acid is nucleic acid sequence encoding any particular GGPPS, where said nucleic acid has been codon optimised for the particular microbial cell. Thus, by way of example, if the microbial cell is S. cerevisiae, then the nucleic acid encoding GGPPS has preferably been codon optimised for optimal expression in S. cerevisiae.

Functional homologues of GGPPS are preferably protein having above-mentioned activity and sharing at least 70% amino acid identity with the sequence of a reference GGPPS. Methods for determining sequence identity are described herein above in the section "Squalene synthase" and in section D.

In one embodiment, the cell, such as the microbial of the present invention comprises a nucleic acid sequence coding a GGPPS or a functional homologue thereof, where said functional homologue is at least 75%, such as at least 76%, such as at least 77%, such as at least 78%, such as at least 79%, such as at least 80%, such as at least 81%, such as at least 82%, such as at least 83%, such as at least 84%, such as at least 85%, such as at least 86%, such as at least 87%, such as at least 88%, such as at least 89%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as at least 99.5%, such as at least 99.6%, such as at least 99.7%, such as at least 99.8%, such as at least 99.9%, such as 100% identical to a GGPPS selected from the group consisting of SEQ ID NO:123, SEQ ID NO:126, SEQ ID NO:167 and SEQ ID NO:203.

Said heterologous nucleic acid sequence encoding a GGPPS is in general operably linked to a nucleic acid sequence directing expression of GGPPS in the microbial cell. The nucleic acid sequence directing expression of GGPPS in the microbial cell may be a promoter sequence, and preferably said promoter sequence is selected according the particular microbial cell. The promoter may for example be any of the promoters described herein above in the section "Promoter"

Vectors

A vector is a DNA molecule used as a vehicle to transfer foreign genetic material into another cell. The major types of vectors are plasmids, viruses, cosmids, and artificial chromosomes. Common to all engineered vectors is an origin of replication, a multicloning site, and a selectable marker.

The vector itself is generally a DNA sequence that consists of an insert (transgene) and a larger sequence that serves as the "backbone" of the vector. The purpose of a vector which transfers genetic information to another cell is typically to

isolate, multiply, or express the insert in the target cell. Vectors called expression vectors (expression constructs) specifically are for the expression of the transgene in the target cell, and generally have a promoter sequence that drives expression of the transgene. Simpler vectors called transcription vectors are only capable of being transcribed but not translated: they can be replicated in a target cell but not expressed, unlike expression vectors. Transcription vectors are used to amplify their insert.

Insertion of a vector into the target cell is usually called transformation for bacterial cells, transfection for eukaryotic cells, although insertion of a viral vector is often called transduction.

Plasmids

Plasmid vectors are double-stranded generally circular DNA sequences that are capable of automatically replicating in a host cell. Plasmid vectors minimalistically consist of an origin of replication that allows for semi-independent replication of the plasmid in the host and also the transgene insert. Modern plasmids generally have many more features, notably including a "multiple cloning site" which includes nucleotide overhangs for insertion of an insert, and multiple restriction enzyme consensus sites to either side of the insert. In the case of plasmids utilized as transcription vectors, incubating bacteria with plasmids generates hundreds or thousands of copies of the vector within the bacteria in hours, and the vectors can be extracted from the bacteria, and the multiple cloning site can be cut by restriction enzymes to excise the hundredfold or thousandfold amplified insert. These plasmid transcription vectors characteristically lack crucial sequences that code for polyadenylation sequences and translation termination sequences in translated mRNAs, making protein expression from transcription vectors impossible. plasmids may be conjugative/transmissible and non-conjugative:

- conjugative: mediate DNA transfer through conjugation and therefore spread rapidly among the bacterial cells of a population; e.g., F plasmid, many R and some col plasmids.
- nonconjugative- do not mediate DNA through conjugation, e.g., many R and col plasmids.

Viral vectors

Viral vectors are generally genetically-engineered viruses carrying modified viral DNA or RNA that has been rendered noninfectious, but still contain viral promoters and also the transgene, thus allowing for translation of the transgene through a viral promoter. However, because viral vectors frequently are lacking infectious sequences, they require helper viruses or packaging lines for large-scale transfection. Viral vectors are often designed for permanent incorporation of the insert into the host genome, and thus leave distinct genetic markers in the host genome after incorporating the transgene. For example, retroviruses leave a characteristic retroviral integration pattern after insertion that is detectable and indicates that the viral vector has incorporated into the host genome.

In one embodiment the invention concerns a viral vector capable of transfecting a host cell, such as a cell that can be cultured, e.g. a yeast cell or any other suitable eukaryotic cell. The vector is then capable of transfecting said cell with a nucleic acid that includes the heterologous insert sequence as described herein.

The viral vector can be any suitable viral vector such as a viral vector selected from the group consisting of vectors derived from the Retroviridae family including lentivirus, HIV, SIV, FIV, EAIV, CIV.

The viral vector may also be selected from the group consisting of alphavirus, adenovirus, adeno associated virus, baculovirus, HSV, coronavirus, Bovine papilloma virus, Mo-MLV and adeno associated virus.

In embodiments of the invention wherein the microbial cell comprises a heterologous nucleic acid encoding GGPPS, then said heterologous nucleic acid may be positioned on the vector also containing the nucleic acid encoding squalene synthase, or the heterologous nucleic acid encoding GGPPS may be positioned on a different vector. Said heterologous nucleic acid encoding GGPPS may be contained in any of the vectors described herein above.

In embodiments of the invention wherein the microbial cell comprises a heterologous nucleic acid encoding HMCR, then said heterologous nucleic acid may be positioned on the vector also containing the nucleic acid encoding squalene synthase, or the heterologous nucleic acid encoding HMCR may be positioned on a different vector.

Said heterologous nucleic acid encoding HMCR may be contained in any of the vectors described herein above. It is also contained within the invention that the heterologous nucleic acid encoding GGPPS and the heterologous nucleci acid encoding HMCR may be positioned on the same or on individual vectors.

Transcription

Transcription is a necessary component in all vectors: the premise of a vector is to multiply the insert (although expression vectors later also drive the translation of the multiplied insert). Thus, even stable expression is determined by stable transcription, which generally depends on promoters in the vector. However, expression vectors have a variety of expression patterns: constitutive (consistent expression) or inducible (expression only under certain conditions or chemicals). This expression is based on different promoter activities, not post-transcriptional activities. Thus, these two different types of expression vectors depend on different types of promoters.

Viral promoters are often used for constitutive expression in plasmids and in viral vectors because they normally reliably force constant transcription in many cell lines and types.

Inducible expression depends on promoters that respond to the induction conditions: for example, the murine mammary tumor virus promoter only initiates transcription after dexamethasone application and the Drosophilia heat shock promoter only initiates after high temperatures. transcription is the synthesis of mRNA. Genetic information is copied from DNA to RNA

Expression

Expression vectors require sequences that encode for e.g. polyadenylation tail (see herein above): Creates a polyadenylation tail at the end of the transcribed pre-mRNA that protects the mRNA from exonucleases and ensures transcriptional and translational termination: stabilizes mRNA production.

Minimal UTR length: UTRs contain specific characteristics that may impede transcription or translation, and thus the shortest UTRs or none at all are encoded for in optimal expression vectors.

Kozak sequence: Vectors should encode for a Kozak sequence in the mRNA, which assembles the ribosome for translation of the mRNA.

Above conditions are necessary for expression vectors in eukaryotes but not in prokaryotes.

Modern vectors may encompass additional features besides the transgene insert and a backbone such as a promoter (discussed above), genetic markers to e.g. allow for confirmation that the vector has integrated with the host genomic DNA, antibiotic resistance genes for antibiotic selection, and affinity tags for purification.

In one embodiment the cell of the present invention comprises a nucleic acid sequence integrated in a vector such as an expression vector.

In one embodiment the the vector is selected from the group consisting of plasmid vectors, cosmids, artificial chromosomes and viral vectors.

The plasmid vector should be able to be maintained and replicated in bacteria, fungi and yeast.

The present invention also concerns cells comprising plasmid and cosmid vectors as well as artificial chromosome vectors.

The important factor is that the vector is functional and that the vector comprises at least the nucleic acid sequence comprising the heterologous insert sequence as described herein.

In one embodiment the vector is functional in fungi and in mammalian cells.

In one embodiment the invention concerns a cell transformed or transduced with the vector as defined herein above.

Methods for producing terpenoids

As mentioned herein above, the cell of the present invention (e.g., recombinant host cells) is useful in enhancing yield of industrially relevant terpenoids.

The cell of the invention may therefore be used in various set-ups in order to increase accumulation of terpenoid precursors and thus to increase yield of terpenoid products resulting from enzymatic conversion of said (upstream) terpenoid precursors.

Accordingly, in one aspect the present invention relates to a method for producing a terpenoid compound synthesized through the squalene pathway, in a cell culture, said method comprising the steps of

(a) providing the cell as defined herein above,

(b) culturing the cell of (a).

(c) recovering the terpenoid product compound.

By providing the cell comprising the genetically modified construct defined herein above, the accumulation of terpenoid precursors is enhanced (FIG. 20).

Thus, in another aspect, the invention relates to a method for producing a terpenoid derived from a terpenoid precursor selected from the group consisting of Farnesyl-pyrophosphate (FPP), Isopentenyl-pyrophosphate (IPP), Dimethylallyl-pyrophosphate (DMAPP), Geranyl-pyrophosphate (GPP) and/or Geranylgeranyl-pyrophosphate (GGPP), said method comprising:

(a) contacting said precursor with an enzyme of the squalene synthase pathway,

(b) recovering the terpenoid product.

In one embodiment the terpenoid (product) of the method of the present invention as defined herein above, is selected from the group consisting of hemiterpenoids, monoterpenes, sesquiterpenoids, diterpenoids, sesterpenes, triterpenoids, tetraterpenoids and polyterpenoids.

In a further embodiment the terpenoid is selected from the group consisting of farnesyl phosphate, farnesol, geranylgeranyl, geranylgeraniol, isoprene, prenol, isovaleric acid, geranyl pyrophosphate, eucalyptol, limonene, pinene, farnesyl pyrophosphate, artemisinin, bisabolol, geranylgeranyl pyrophosphate, retinol, retinal, phytol, taxol , forskolin, aphidicolin, lanosterol, lycopene and carotene.

The terpenoid prioduct can be used as starting point in an additional refining process. Thus, in one embodiment said method further comprises dephosphorylating the farnesyl phosphate to produce farnesol.

The enzyme or enzymes used in the process of preparing the target product terpenoid compound is preferably an enzyme “located downstream” of the terpenoid

precursors Farnesyl-pyrophosphate, Isopentenyl-pyrophosphate, Dimethylallyl-pyrophosphate, Geranyl-pyrophosphate and Geranylgeranyl-pyrophosphate such as an enzyme located downstream of the terpenoid precursors Farnesyl-pyrophosphate, Isopentenyl-pyrophosphate, Dimethylallyl-pyrophosphate, Geranyl-pyrophosphate and Geranylgeranyl-pyrophosphate depicted in the squalene pathway of FIG. 20. The enzyme used in the process of preparing the target product terpenoid, based on the accumulation of precursors achieved through the present invention, may thus be selected from the group consisting of Dimethylallyltransferase (EC 2.5.1.1), Isoprene synthase (EC 4.2.3.27) and Geranyltranstransferase (EC 2.5.1.10).

The present invention may operate by at least partly, sterically hindering binding of the ribosome to the RNA thus reducing the translation of squalene synthase. Accordingly, in one aspect the present invention relates to a method for reducing the translation rate of a functional squalene synthase (EC 2.5.1.21) said method comprising:

(a) providing the cell defined herein above,

(b) culturing the cell of (a).

Similarly, the invention in another aspect relates to a method for decreasing turnover of farnesyl-pp to squalene, said method comprising:

(a) providing the cell defined herein above,

(b) culturing the cell of (a).

As depicted in FIG. 20, the knocking down of the ERG9 results in build-up of precursors to squalene synthase. Thus in one aspect, the present invention relates to a method for enhancing accumulation of a compound selected from the group consisting of Farnesyl-pyrophosphate, Isopentenyl-pyrophosphate, Dimethylallyl-pyrophosphate, Geranyl-pyrophosphate and Geranylgeranyl-pyrophosphate, said method comprising the steps of:

(a) providing the cell defined herein above, and

(b) culturing the cell of (a).

In one embodiment the method of the invention as define herein above further comprises recovering the Farnesyl-pyrophosphate, Isopentenyl-pyrophosphate, Dimethylallyl-pyrophosphate, Geranyl-pyrophosphate or Geranylgeranyl-pyrophosphate

compound. The recovered compound may be used in further processes for producing the desired terpenoid product compound. The further process may take place in the same cell culture as the process performed and defined herein above, such as the accumulation of the terpenoid precursors by the cell of the present invention. Alternatively, the recovered precursors may be added to another cell culture, or a cell free system, to produce the desired products.

As the precursors are intermediates, however mainly stable intermediates, a certain endogenous production of terpenoid products may occur based on the terpenoid precursor substrates. Also, the cells of the invention may have additional genetic modifications such that they are capable of performing both the accumulation of the terpenoid precursors (construct of the cell of the invention) and whole or substantially the whole subsequent biosynthesis process to the desired terpenoid product.

Thus, in one embodiment the method of the invention further comprises recovering a compound synthesized through the squalene pathway, said compound being derived from said Farnesyl-pyrophosphate, Isopentenyl-pyrophosphate, Dimethylallyl-pyrophosphate, Geranyl-pyrophosphate and/or Geranylgeranyl-pyrophosphate. Occasionally it may be advantageous to include a squalene synthase inhibitor when culturing the cell of the present invention. Chemical inhibition of squalene synthase, e.g. by lapaquistat, is known in the art and is under investigation e.g. as a method of lowering cholesterol levels in the prevention of cardiovascular disease. It has also been suggested that variants in this enzyme may be part of a genetic association with hypercholesterolemia. Other squalene synthase inhibitors include Zaragozic acid and RPR 107393.

Thus, in one embodiment the culturing step of the method(s) defined herein above is performed in the presence of a squalene synthase inhibitor.

The cell of the invention may furthermore be genetically modified to further enhance production of certain key terpenoid precursors. In one embodiment the cell is additionally genetically modified to enhance activity of and/or overexpress one or more enzymes selected from the group consisting of Phosphomevalonate kinase (EC 2.7.4.2), Diphosphomevalonate decarboxylase (EC 4.1.1.33), 4-hydroxy-3-methylbut-2-en-1-yl

diphosphate synthase (EC 1.17.7.1), 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (EC 1.17.1.2), Isopentenyl-diphosphate Delta-isomerase 1 (EC 5.3.3.2), Short-chain Z-isoprenyl diphosphate synthase (EC 2.5.1.68), Dimethylallyltransferase (EC 2.5.1.1), Geranyltranstransferase (EC 2.5.1.10) and Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29).

As described herein above in one embodiment of the invention the microbial cell comprises both a nucleic acid encoding a sqalene synthase as described herein above as well as a heterologous nucleic acid encoding a GGPPS. Such microbial cells are particularly useful for the preparation of GGPP as well as terpenoids, wherein GGPP is an intermediate in their biosynthesis.

Accordingly, in one aspect the invention relates to a method for preparing GGPP, wherein the method comprises the steps of

a. providing a microbial cell comprising a nucleic acid sequence, said nucleic acid comprising

i) a promoter sequence operably linked to

ii) a heterologous insert sequence operably linked to

iii) an open reading frame operably linked to

iv)a transcription termination signal,

wherein the heterologous insert sequence and the open reading frame are as defined herein above,

wherein said microbial cell furthermore comprises a heterologous nucleic acid encoding GGPPS operably linked to a nucleic acid sequence directing expression of GGPPS in said cell;

b. Cultivating the microbial cell of a.;

c. Recovering the GGPP.

In another aspect the invention relates to a method for preparing a terpenoid of which GGPP is an intermediate in the biosynthesis pathway, wherein the method comprises the steps of

a. providing a microbial cell, wherein said microbial cell comprises a nucleic acid sequence, said nucleic acid comprising

i) a promoter sequence operably linked to

ii) a heterologous insert sequence operably linked to

iii) an open reading frame operably linked to

iv)a transcription termination signal,

wherein the heterologous insert sequence and the open reading frame are as defined herein above,

wherein said microbial cell furthermore comprises a heterologous nucleic acid encoding GGPPS operably linked to a nucleic acid sequence directing expression of GGPPS in said cell;

b. Cultivating the microbial cell of a; and

c. Recovering the terpenoid,

wherein said terpenoid may be any terpenoid described herein above in the section "Terpenoids" having GGPP as intermediate in its biosynthesises; and said microbial cell may be any of the microbial cells described herein above in the section "The cell"; and said promoter may be any promoter, such as any of the promoters described herein above in the section "Promoter"; and said heterologous insert sequence may be any of the heterologous insert sequences described herein above in the section "Heterologous insert sequence"; and said open reading frame encodes a squalene synthase, which may be any of the squalene synthases described herein above in the section "Squalene synthase"; and said GGPPS may be any of the GGPPS described herein above in the section "Geranylgeranyl Pyrophosphate Synthase".

In this embodiment said microbial cell may also optionally contain one or more additional heterologous nucleic acids encoding one or more enzymes involved in the biosynthesis pathway of said terpenoid.

In one particular aspect the invention relates to a method for preparing steviol, wherein the method comprises the steps of

a. providing a microbial cell, wherein said microbial cell comprises a nucleic acid sequence, said nucleic acid comprising

i) a promoter sequence operably linked to

ii) a heterologous insert sequence operably linked to

iii) an open reading frame operably linked to

iv)a transcription termination signal,

wherein the heterologous insert sequence and the open reading frame are as defined herein above,

wherein said microbial cell furthermore comprises a heterologous nucleic acid encoding GGPPS operably linked to a nucleic acid sequence directing expression of GGPPS in said cell;

b. Cultivating the microbial cell of a.;

c. Recovering steviol,

wherein said microbial cell may be any of the microbial cells described herein above in the section "The cell"; and said promoter may be any promoter, such as any of the promoters described herein above in the section "Promoter"; and said heterologous insert sequence may be any of the heterologous insert sequences described herein above in the section "Heterologous insert sequence"; and said open reading frame encodes a squalene synthase, which may be any of the squalene synthases described herein above in the section "Squalene synthase"; and said GGPPS may be any of the GGPPS described herein above in the section "Geranylgeranyl Pyrophosphate Synthase".

In this embodiment said microbial cell may also optionally contain one or more additional heterologous nucleic acids encoding one or more enzymes involved in the biosynthesis pathway of steviol.

In another aspect the invention relates to a method for preparing a terpenoid of which GGPP is an intermediate in the biosynthesis pathway, wherein the method comprises the steps of

a. providing a microbial cell, wherein said microbial cell comprises a nucleic acid sequence, said nucleic acid comprising

i) a promoter sequence operably linked to

ii) a heterologous insert sequence operably linked to

iii) an open reading frame operably linked to

iv)a transcription termination signal,

wherein the heterologous insert sequence and the open reading frame are as defined herein above,

wherein said microbial cell furthermore comprises a heterologous nucleic acid encoding GGPPS operably linked to a nucleic acid sequence directing expression of GGPPS in said cell

and wherein said microbial cell furthermore comprises a heterologous nucleic acid encoding HMCR operably linked to a nucleic acid sequence directing expression of HMCR in said cell;

b. Cultivating the microbial cell of a.;

c. Recovering the terpenoid,

wherein said terpenoid may be any terpenoid described herein above in the section "Terpenoids" having GGPP as intermediate in its biosynthesises; and said microbial cell may be any of the microbial cells described herein above in the section "The cell"; and said promoter may be any promoter, such as any of the promoters described herein above in the section "Promoter"; and said heterologous insert sequence may be any of the heterologous insert sequences described herein above in the section "Heterologous insert sequence"; and said open reading frame encodes a squalene synthase, which may be any of the squalene synthases described herein above in the section "Squalene synthase"; and said GGPPS may be any of the GGPPS described herein above in the section "Geranylgeranyl Pyrophosphate Synthase"; and said HMCR may be any of the HMCR described herein above in the section "HMCR".

In this embodiment said microbial cell may also optionally contain one or more additional heterologous nucleic acids encoding one or more enzymes involved in the biosynthesis pathway of said terpenoid.

In one particular aspect the invention relates to a method for preparing steviol, wherein the method comprises the steps of

a. providing a microbial cell, wherein said microbial cell comprises a nucleic acid sequence, said nucleic acid comprising

i) a promoter sequence operably linked to

ii) a heterologous insert sequence operably linked to

iii) an open reading frame operably linked to

iv)a transcription termination signal,

wherein the heterologous insert sequence and the open reading frame are as defined herein above,

wherein said microbial cell furthermore comprises a heterologous nucleic acid encoding GGPPS operably linked to a nucleic acid sequence directing expression of GGPPS in said cell;

b. Cultivating the microbial cell of a.;

c. Recovering steviol,

wherein said microbial cell may be any of the microbial cells described herein above in the section "The cell"; and said promoter may be any promoter, such as any of the promoters described herein above in the section "Promoter"; and said heterologous insert sequence may be any of the heterologous insert sequences described herein above in the section "Heterologous insert sequence"; and said open reading frame encodes a squalene synthase, which may be any of the squalene synthases described herein above in the section "Squalene synthase"; and said GGPPS may be any of the GGPPS described herein above in the section "Geranylgeranyl Pyrophosphate Synthase" and said HMCR may be any of the HMCR described herein above in the section "HMCR".

In this embodiment said microbial cell may also optionally contain one or more additional heterologous nucleic acids encoding one or more enzymes involved in the biosynthesis pathway of steviol.

In one embodiment the cell is additionally genetically modified to enhance activity of and/or overexpress one or more enzymes selected from the group consisting of acetoacetyl CoA thiolose, HMG-CoA reductase or the catalytic domain thereof, HMG-CoA synthase, mevalonate kinase, phosphomevalonate kinase, phosphomevalonate decarboxylase, isopentenyl pyrophosphate isomerase, farnesyl pyrophosphate synthase, D-1-deoxyxylulose 5-phosphate synthase, and 1-deoxy-D-xylulose 5-phosphate reductoisomerase and farnesyl pyrophosphate synthase.

In one embodiment of the method of the present invention, the cell comprises a mutation in the ERG9 open reading frame.

In another embodiment of the method of the present invention the cell comprises an ERG9[Delta]::HIS3 deletion/insertion allele.

In yet another embodiment the step of recovering the compound in the method of the present invention further comprises purification of said compound from the cell culture media.

D. Functional Homologs

Functional homologs of the polypeptides described above are also suitable for use in producing steviol or steviol glycosides in a recombinant host. A functional homolog is a polypeptide that has sequence similarity to a reference polypeptide, and that carries out one or more of the biochemical or physiological function(s) of the reference polypeptide. A functional homolog and the reference polypeptide may be natural occurring polypeptides, and the sequence similarity may be due to convergent or divergent evolutionary events. As such, functional homologs are sometimes designated in the literature as homologs, or orthologs, or paralogs. Variants of a naturally occurring functional homolog, such as polypeptides encoded by mutants of a wild type coding sequence, may themselves be functional homologs. Functional homologs can also be created via site-directed mutagenesis of the coding sequence for a polypeptide, or by combining domains from the coding sequences for different naturally-occurring polypeptides ("domain swapping"). Techniques for modifying genes encoding functional UGT polypeptides described herein are known and include, *inter alia*, directed evolution techniques, site-directed mutagenesis techniques and random mutagenesis techniques, and can be useful to increase specific activity of a polypeptide, alter substrate specificity, alter expression levels, alter subcellular location, or modify polypeptide:polypeptide interactions in a desired manner. Such modified polypeptides are considered functional homologs. The term "functional homolog" is sometimes applied to the nucleic acid that encodes a functionally homologous polypeptide.

Functional homologs can be identified by analysis of nucleotide and polypeptide sequence alignments. For example, performing a query on a database of nucleotide or polypeptide sequences can identify homologs of steviol or steviol glycoside biosynthesis

polypeptides. Sequence analysis can involve BLAST, Reciprocal BLAST, or PSI-BLAST analysis of nonredundant databases using a GGPPS, a CDPS, a KS, a KO or a KAH amino acid sequence as the reference sequence. Amino acid sequence is, in some instances, deduced from the nucleotide sequence. Those polypeptides in the database that have greater than 40% sequence identity are candidates for further evaluation for suitability as a steviol or steviol glycoside biosynthesis polypeptide. Amino acid sequence similarity allows for conservative amino acid substitutions, such as substitution of one hydrophobic residue for another or substitution of one polar residue for another. If desired, manual inspection of such candidates can be carried out in order to narrow the number of candidates to be further evaluated. Manual inspection can be performed by selecting those candidates that appear to have domains present in steviol biosynthesis polypeptides, *e.g.*, conserved functional domains.

Conserved regions can be identified by locating a region within the primary amino acid sequence of a steviol or a steviol glycoside biosynthesis polypeptide that is a repeated sequence, forms some secondary structure (*e.g.*, helices and beta sheets), establishes positively or negatively charged domains, or represents a protein motif or domain. See, *e.g.*, the Pfam web site describing consensus sequences for a variety of protein motifs and domains on the World Wide Web at sanger.ac.uk/Software/Pfam/ and pfam.janelia.org/. The information included at the Pfam database is described in Sonnhammer *et al*., Nucl. Acids Res., 26:320-322 (1998); Sonnhammer *et al*., Proteins, 28:405-420 (1997); and Bateman *et al*., Nucl. Acids Res., 27:260-262 (1999). Conserved regions also can be determined by aligning sequences of the same or related polypeptides from closely related species. Closely related species preferably are from the same family. In some embodiments, alignment of sequences from two different species is adequate.

Typically, polypeptides that exhibit at least about 40% amino acid sequence identity are useful to identify conserved regions. Conserved regions of related polypeptides exhibit at least 45% amino acid sequence identity (*e.g.*, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% amino acid sequence identity). In some embodiments, a conserved region exhibits at least 92%, 94%, 96%, 98%, or 99% amino acid sequence identity.

For example, polypeptides suitable for producing steviol glycosides in a recombinant host include functional homologs of EUGT11, UGT91D2e, UGT91D2m, UGT85C, and UGT76G. Such homologs have greater than 90% (e.g., at least 95% or 99%) sequence identity to the amino acid sequence of EUGT11 (SEQ ID NO: 152), UGT91D2e (SEQ ID NO:5), UGT91D2m (SEQ ID NO:10), UGT85C (SEQ ID NO:3), or UGT76G (SEQ ID NO:7). Variants of EUGT11, UGT91D2, UGT85C, and UGT76G polypeptides typically have 10 or fewer amino acid substitutions within the primary amino acid sequence, *e.g.*, 7 or fewer amino acid substitutions, 5 or conservative amino acid substitutions, or between 1 and 5 substitutions. However, in some embodiments, variants of EUGT11, UGT91D2, UGT85C, and UGT76G polypeptides can have 10 or more amino acid substitutions (e.g., 10, 15, 20, 25, 30, 35, 10-20, 10-35, 20-30, or 25-35 amino acid substitutions). The substitutions may be conservative, or in some embodiments, non-conservative. Non-limiting examples of non-conservative changes in UGT91D2e polypeptides include glycine to arginine and tryptophan to arginine. Non-limiting examples of non-conservative substitutions in UGT76G polypeptides include valine to glutamic acid, glycine to glutamic acid, glutamine to alanine, and serine to proline. Non-limiting examples of changes to UGT85C polypeptides include histidine to aspartic acid, proline to serine, lysine to threonine, and threonine to arginine.

In some embodiments, a useful UGT91D2 homolog can have amino acid substitutions (e.g., conservative amino acid substitutions) in regions of the polypeptide that are outside of predicted loops, e.g., residues 20-26, 39-43, 88-95, 121-124, 142-158, 185-198, and 203-214 are predicted loops in the N-terminal domain and residues 381-386 are predicted loops in the C-terminal domain of SEQ ID NO:5. For example, a useful UGT91D2 homolog can include at least one amino acid substitution at residues 1-19, 27-38, 44-87, 96-120, 125-141, 159-184, 199-202, 215-380, or 387-473 of SEQ ID NO:5. In some embodiments, a UGT91D2 homolog can have an amino acid substitution at one or more residues selected from the group consisting of residues 30, 93, 99, 122, 140, 142, 148, 153, 156, 195, 196, 199, 206, 207, 211, 221, 286, 343, 427, and 438 of SEQ ID NO:5. For example, a UGT91D2 functional homolog can have an amino acid substitution at one or more of residues 206, 207, and 343, such as an arginine at residue

206, a cysteine at residue 207, and an arginine at residue 343 of SEQ ID NO:5. See, SEQ ID NO:95. Other functional homologs of UGT91D2 can have one or more of the following: a tyrosine or phenylalanine at residue 30, a proline or glutamine at residue 93, a serine or valine at residue 99, a tyrosine or a phenylalanine at residue 122, a histidine or tyrosine at residue 140, a serine or cysteine at residue 142, an alanine or threonine at residue 148, a methionine at residue 152, an alanine at residue 153, an alanine or serine at residue 156, a glycine at residue 162, a leucine or methionine at residue 195, a glutamic acid at residue 196, a lysine or glutamic acid at residue 199, a leucine or methionine at residue 211, a leucine at residue 213, a serine or phenylalanine at residue 221, a valine or isoleucine at residue 253, a valine or alanine at residue 286, a lysine or asparagine at residue 427, an alanine at residue 438, and either an alanine or threonine at residue 462 of SEQ ID NO:5. In another embodiment, a UGT91D2 functional homolog contains a methionine at residue 211 and an alanine at residue 286.

In some embodiments, a useful UGT85C homolog can have one or more amino acid substitutions at residues 9, 10, 13, 15, 21, 27, 60, 65, 71, 87, 91, 220, 243, 270, 289, 298, 334, 336, 350, 368, 389, 394, 397, 418, 420, 440, 441, 444, and 471 of SEQ ID NO:3. Non-limiting examples of useful UGT85C homologs include polypeptides having substitutions (with respect to SEQ ID NO:3) at residue 65 (e.g., a serine at residue 65), at residue 65 in combination with residue 15 (a leucine at residue 15), 270 (e.g., a methionine, arginine, or alanine at residue 270), 418 (e.g., a valine at residue 418), 440 (e.g., an aspartic acid at residue at residue 440), or 441 (e.g., an asparagine at residue 441); residues 13 (e.g., a phenylalanine at residue 13), 15, 60 (e.g., an aspartic acid at residue 60), 270, 289 (e.g., a histidine at residue 289), and 418; substitutions at residues 13, 60, and 270; substitutions at residues 60 and 87 (e.g., a phenylalanine at residue 87); substitutions at residues 65, 71 (e.g., a glutamine at residue 71), 220 (e.g., a threonine at residue 220), 243 (e.g., a tryptophan at residue 243), and 270; substitutions at residues 65, 71, 220, 243, 270, and 441; substitutions at residues 65, 71, 220, 389 (e.g., a valine at residue 389), and 394 (e.g., a valine at residue 394); substitutions at residues 65, 71, 270, and 289; substitutions at residues 220, 243, 270, and 334 (e.g., a serine at residue 334); or substitutions at residues 270 and 289. The following amino acid mutations did not result

in a loss of activity in 85C2 polypeptides: V13F, F15L, H60D, A65S, E71Q, I87F, K220T, R243W, T270M, T270R, Q289H, L334S, A389V, I394V, P397S, E418V, G440D, and H441N. Additional mutations that were seen in active clones include K9E, K10R, Q21H, M27V, L91P, Y298C, K350T, H368R, G420R, L431P, R444G, and M471T. In some embodiments, an UGT85C2 contains substitutions at positions 65 (e.g., a serine), 71 (a glutamine), 270 (a methionine), 289 (a histidine), and 389 (a valine).

The amino acid sequence of *Stevia rebaudiana* UGTs 74G1,76G1 and 91D2e with N-terminal, in-frame fusions of the first 158 amino acids of human MDM2 protein, and *Stevia rebaudiana* UGT85C2 with an N-terminal in-frame fusion of 4 repeats of the synthetic PMI peptide (4 X TSFAEYWNLLSP, SEQ ID NO:86) are set forth in SEQ ID NOs: 90, 88, 94, and 92, respectively; see SEQ ID NOs: 89, 92, 93, and 95 for the nucleotide sequences encoding the fusion proteins.

In some embodiments, a useful UGT76G homolog can have one or more amino acid substitutions at residues 29, 74, 87, 91, 116, 123, 125, 126, 130, 145, 192, 193, 194, 196, 198, 199, 200, 203, 204, 205, 206, 207, 208, 266, 273, 274, 284, 285, 291, 330, 331, and 346 of SEQ ID NO:7. Non-limiting examples of useful UGT76G homologs include polypeptides having substitutions (with respect to SEQ ID NO:7) at residues 74 , 87, 91, 116, 123, 125, 126, 130, 145, 192, 193, 194, 196, 198, 199, 200, 203, 204, 205, 206, 207, 208, and 291; residues 74 , 87, 91, 116, 123, 125, 126, 130, 145, 192, 193, 194, 196, 198, 199, 200, 203, 204, 205, 206, 207, 208, 266, 273, 274, 284, 285, and 291; or residues 74 , 87, 91, 116, 123, 125, 126, 130, 145, 192, 193, 194, 196, 198, 199, 200, 203, 204, 205, 206, 207, 208, 266, 273, 274, 284, 285, 291, 330, 331, and 346. See, Table 9.

Table 9

Clone	Mutations
76G_G7	M29I, V74E, V87G, L91P, G116E, A123T, Q125A, I126L, T130A, V145M, C192S, S193A, F194Y, M196N, K198Q, K199I, Y200L, Y203I, F204L, E205G, N206K, I207M, T208I, P266Q, S273P, R274S, G284T, T285S, 287-3 bp deletion, L330V, G331A, L346I
76G_H12	M29I, V74E, V87G, L91P, G116E, A123T, Q125A, I126L, T130A, V145M, C192S, S193A, F194Y, M196N, K198Q, K199I, Y200L,

	Y203I, F204L, E205G, N206K, I207M, T208I, P266Q, S273P, R274S, G284T, T285S, 287-3 bp deletion
76G_C4	M29I, V74E, V87G, L91P, G116E, A123T, Q125A, I126L, T130A, V145M, C192S, S193A, F194Y, M196N, K198Q, K199I, Y200L, Y203I, F204L, E205G, N206K, I207M, T208I

Methods to modify the substrate specificity of, for example, EUGT11 or UGT91D2e, are known to those skilled in the art, and include without limitation site-directed/rational mutagenesis approaches, random directed evolution approaches and combinations in which random mutagenesis/saturation techniques are performed near the active site of the enzyme. For example see Sarah A. Osmani, et al. Phytochemistry 70 (2009) 325–347.

A candidate sequence typically has a length that is from 80 percent to 200 percent of the length of the reference sequence, *e.g.*, 82, 85, 87, 89, 90, 93, 95, 97, 99, 100, 105, 110, 115, 120, 130, 140, 150, 160, 170, 180, 190, or 200 percent of the length of the reference sequence. A functional homolog polypeptide typically has a length that is from 95 percent to 105 percent of the length of the reference sequence, e.g., 90, 93, 95, 97, 99, 100, 105, 110, 115, or 120 percent of the length of the reference sequence, or any range between. A percent identity for any candidate nucleic acid or polypeptide relative to a reference nucleic acid or polypeptide can be determined as follows. A reference sequence (*e.g.*, a nucleic acid sequence or an amino acid sequence) is aligned to one or more candidate sequences using the computer program ClustalW (version 1.83, default parameters), which allows alignments of nucleic acid or polypeptide sequences to be carried out across their entire length (global alignment). Chenna *et al.*, *Nucleic Acids Res.*, 31(13):3497-500 (2003).

ClustalW calculates the best match between a reference and one or more candidate sequences, and aligns them so that identities, similarities and differences can be determined. Gaps of one or more residues can be inserted into a reference sequence, a candidate sequence, or both, to maximize sequence alignments. For fast pairwise alignment of nucleic acid sequences, the following default parameters are used: word

size: 2; window size: 4; scoring method: percentage; number of top diagonals: 4; and gap penalty: 5. For multiple alignment of nucleic acid sequences, the following parameters are used: gap opening penalty: 10.0; gap extension penalty: 5.0; and weight transitions: yes. For fast pairwise alignment of protein sequences, the following parameters are used: word size: 1; window size: 5; scoring method: percentage; number of top diagonals: 5; gap penalty: 3. For multiple alignment of protein sequences, the following parameters are used: weight matrix: blosum; gap opening penalty: 10.0; gap extension penalty: 0.05; hydrophilic gaps: on; hydrophilic residues: Gly, Pro, Ser, Asn, Asp, Gln, Glu, Arg, and Lys; residue-specific gap penalties: on. The ClustalW output is a sequence alignment that reflects the relationship between sequences. ClustalW can be run, for example, at the Baylor College of Medicine Search Launcher site on the World Wide Web (searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) and at the European Bioinformatics Institute site on the World Wide Web (ebi.ac.uk/clustalw).

To determine percent identity of a candidate nucleic acid or amino acid sequence to a reference sequence, the sequences are aligned using ClustalW, the number of identical matches in the alignment is divided by the length of the reference sequence, and the result is multiplied by 100. It is noted that the percent identity value can be rounded to the nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 are rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 are rounded up to 78.2.

It will be appreciated that functional UGTs can include additional amino acids that are not involved in glucosylation or other enzymatic activities carried out by the enzyme, and thus such a polypeptide can be longer than would otherwise be the case. For example, a EUGT11 polypeptide can include a purification tag (e.g., HIS tag or GST tag), a chloroplast transit peptide, a mitochondrial transit peptide, an amyloplast peptide, signal peptide, or a secretion tag added to the amino or carboxy terminus. In some embodiments, a EUGT11 polypeptide includes an amino acid sequence that functions as a reporter, *e.g.*, a green fluorescent protein or yellow fluorescent protein.

II. Steviol and Steviol Glycoside Biosynthesis Nucleic Acids

A recombinant gene encoding a polypeptide described herein comprises the

coding sequence for that polypeptide, operably linked in sense orientation to one or more regulatory regions suitable for expressing the polypeptide. Because many microorganisms are capable of expressing multiple gene products from a polycistronic mRNA, multiple polypeptides can be expressed under the control of a single regulatory region for those microorganisms, if desired. A coding sequence and a regulatory region are considered to be operably linked when the regulatory region and coding sequence are positioned so that the regulatory region is effective for regulating transcription or translation of the sequence. Typically, the translation initiation site of the translational reading frame of the coding sequence is positioned between one and about fifty nucleotides downstream of the regulatory region for a monocistronic gene.

In many cases, the coding sequence for a polypeptide described herein is identified in a species other than the recombinant host, *i.e.*, is a heterologous nucleic acid. Thus, if the recombinant host is a microorganism, the coding sequence can be from other prokaryotic or eukaryotic microorganisms, from plants or from animals. In some case, however, the coding sequence is a sequence that is native to the host and is being reintroduced into that organism. A native sequence can often be distinguished from the naturally occurring sequence by the presence of non-natural sequences linked to the exogenous nucleic acid, *e.g.*, non-native regulatory sequences flanking a native sequence in a recombinant nucleic acid construct. In addition, stably transformed exogenous nucleic acids typically are integrated at positions other than the position where the native sequence is found.

"Regulatory region" refers to a nucleic acid having nucleotide sequences that influence transcription or translation initiation and rate, and stability and/or mobility of a transcription or translation product. Regulatory regions include, without limitation, promoter sequences, enhancer sequences, response elements, protein recognition sites, inducible elements, protein binding sequences, 5′ and 3′ untranslated regions (UTRs), transcriptional start sites, termination sequences, polyadenylation sequences, introns, and combinations thereof. A regulatory region typically comprises at least a core (basal) promoter. A regulatory region also may include at least one control element, such as an enhancer sequence, an upstream element or an upstream activation region (UAR). A

regulatory region is operably linked to a coding sequence by positioning the regulatory region and the coding sequence so that the regulatory region is effective for regulating transcription or translation of the sequence. For example, to operably link a coding sequence and a promoter sequence, the translation initiation site of the translational reading frame of the coding sequence is typically positioned between one and about fifty nucleotides downstream of the promoter. A regulatory region can, however, be positioned as much as about 5,000 nucleotides upstream of the translation initiation site, or about 2,000 nucleotides upstream of the transcription start site.

The choice of regulatory regions to be included depends upon several factors, including, but not limited to, efficiency, selectability, inducibility, desired expression level, and preferential expression during certain culture stages. It is a routine matter for one of skill in the art to modulate the expression of a coding sequence by appropriately selecting and positioning regulatory regions relative to the coding sequence. It will be understood that more than one regulatory region may be present, *e.g.*, introns, enhancers, upstream activation regions, transcription terminators, and inducible elements.

One or more genes can be combined in a recombinant nucleic acid construct in "modules" useful for a discrete aspect of steviol and/or steviol glycoside production. Combining a plurality of genes in a module, particularly a polycistronic module, facilitates the use of the module in a variety of species. For example, a steviol biosynthesis gene cluster, or a UGT gene cluster, can be combined in a polycistronic module such that, after insertion of a suitable regulatory region, the module can be introduced into a wide variety of species. As another example, a UGT gene cluster can be combined such that each UGT coding sequence is operably linked to a separate regulatory region, to form a UGT module. Such a module can be used in those species for which monocistronic expression is necessary or desirable. In addition to genes useful for steviol or steviol glycoside production, a recombinant construct typically also contains an origin of replication, and one or more selectable markers for maintenance of the construct in appropriate species.

It will be appreciated that because of the degeneracy of the genetic code, a number of nucleic acids can encode a particular polypeptide; *i.e.*, for many amino acids,

there is more than one nucleotide triplet that serves as the codon for the amino acid. Thus, codons in the coding sequence for a given polypeptide can be modified such that optimal expression in a particular host is obtained, using appropriate codon bias tables for that host (e.g., microorganism). SEQ ID NOs:18-25, 34-36, 40-43, 48-49, 52-55, 60-64, 70-72, and 154 set forth nucleotide sequences encoding certain enzymes for steviol and steviol glycoside biosynthesis, modified for increased expression in yeast. As isolated nucleic acids, these modified sequences can exist as purified molecules and can be incorporated into a vector or a virus for use in constructing modules for recombinant nucleic acid constructs.

In some cases, it is desirable to inhibit one or more functions of an endogenous polypeptide in order to divert metabolic intermediates towards steviol or steviol glycoside biosynthesis. For example, it may be desirable to downregulate synthesis of sterols in a yeast strain in order to further increase steviol or steviol glycoside production, e.g., by downregulating squalene epoxidase. As another example, it may be desirable to inhibit degradative functions of certain endogenous gene products, e.g., glycohydrolases that remove glucose moieties from secondary metabolites or phosphatases as discussed herein. As another example, expression of membrane transporters involved in transport of steviol glycosides can be inhibited, such that secretion of glycosylated steviosides is inhibited. Such regulation can be beneficial in that secretion of steviol glycosides can be inhibited for a desired period of time during culture of the microorganism, thereby increasing the yield of glycoside product(s) at harvest. In such cases, a nucleic acid that inhibits expression of the polypeptide or gene product may be included in a recombinant construct that is transformed into the strain. Alternatively, mutagenesis can be used to generate mutants in genes for which it is desired to inhibit function.

III. Hosts

A. Microorganisms

A number of prokaryotes and eukaryotes are suitable for use in constructing the recombinant microorganisms described herein, e.g., gram-negative bacteria, yeast and fungi. A species and strain selected for use as a steviol or steviol glycoside production

strain is first analyzed to determine which production genes are endogenous to the strain and which genes are not present. Genes for which an endogenous counterpart is not present in the strain are assembled in one or more recombinant constructs, which are then transformed into the strain in order to supply the missing function(s).

Exemplary prokaryotic and eukaryotic species are described in more detail below. However, it will be appreciated that other species may be suitable. For example, suitable species may be in a genus selected from the group consisting of *Agaricus*, *Aspergillus*, *Bacillus*, *Candida*, *Corynebacterium*, *Escherichia*, *Fusarium/Gibberella*, *Kluyveromyces*, *Laetiporus*, *Lentinus*, *Phaffia*, *Phanerochaete*, *Pichia*, *Physcomitrella*, *Rhodoturula*, *Saccharomyces*, *Schizosaccharomyces*, *Sphaceloma*, *Xanthophyllomyces* and *Yarrowia*. Exemplary species from such genera include *Lentinus tigrinus*, *Laetiporus sulphureus*, *Phanerochaete chrysosporium*, *Pichia pastoris*, *Physcomitrella patens*, *Rhodoturula glutinis* 32, *Rhodoturula mucilaginosa*, *Phaffia rhodozyma* UBV-AX, *Xanthophyllomyces dendrorhous*, *Fusarium fujikuroi/Gibberella fujikuroi*, *Candida utilis* and *Yarrowia lipolytica*. In some embodiments, a microorganism can be an Ascomycete such as *Gibberella fujikuroi*, *Kluyveromyces lactis*, *Schizosaccharomyces pombe*, *Aspergillus niger*, or *Saccharomyces cerevisiae*. In some embodiments, a microorganism can be a prokaryote such as *Escherichia coli*, *Rhodobacter sphaeroides*, or *Rhodobacter capsulatus*. It will be appreciated that certain microorganisms can be used to screen and test genes of interest in a high throughput manner, while other microorganisms with desired productivity or growth characteristics can be used for large-scale production of steviol glycosides.

Saccharomyces cerevisiae

Saccharomyces cerevisiae is a widely used chassis organism in synthetic biology, and can be used as the recombinant microorganism platform. There are libraries of mutants, plasmids, detailed computer models of metabolism and other information available for *S. cerevisiae*, allowing for rational design of various modules to enhance product yield. Methods are known for making recombinant microorganisms.

A steviol biosynthesis gene cluster can be expressed in yeast using any of a number of known promoters. Strains that overproduce terpenes are known and can be

used to increase the amount of geranylgeranyl diphosphate available for steviol and steviol glycoside production.

Aspergillus spp.

Aspergillus species such as *A. oryzae*, *A. niger* and *A. sojae* are widely used microorganisms in food production, and can also be used as the recombinant microorganism platform. Nucleotide sequences are available for genomes of *A. nidulans*, *A. fumigatus*, *A. oryzae*, *A. clavatus*, *A. flavus*, *A. niger*, and *A. terreus*, allowing rational design and modification of endogenous pathways to enhance flux and increase product yield. Metabolic models have been developed for *Aspergillus*, as well as transcriptomic studies and proteomics studies. *A. niger* is cultured for the industrial production of a number of food ingredients such as citric acid and gluconic acid, and thus species such as *A. niger* are generally suitable for the production of food ingredients such as steviol and steviol glycosides.

Escherichia coli

Escherichia coli, another widely used platform organism in synthetic biology, can also be used as the recombinant microorganism platform. Similar to *Saccharomyces*, there are libraries of mutants, plasmids, detailed computer models of metabolism and other information available for *E. coli*, allowing for rational design of various modules to enhance product yield. Methods similar to those described above for *Saccharomyces* can be used to make recombinant *E. coli* microorganisms.

Agaricus, *Gibberella*, and *Phanerochaete* spp.

Agaricus, *Gibberella*, and *Phanerochaete* spp. can be useful because they are known to produce large amounts of gibberellin in culture. Thus, the terpene precursors for producing large amounts of steviol and steviol glycosides are already produced by endogenous genes. Thus, modules containing recombinant genes for steviol or steviol glycoside biosynthesis polypeptides can be introduced into species from such genera without the necessity of introducing mevalonate or MEP pathway genes.

Arxula adeninivorans (Blastobotrys adeninivorans)

Arxula adeninivorans is a dimorphic yeast (it grows as a budding yeast like the baker's yeast up to a temperature of 42 °C, above this threshold it grows in a filamentous

form) with unusual biochemical characteristics. It can grow on a wide range of substrates and can assimilate nitrate. It has successfully been applied to the generation of strains that can produce natural plastics or the development of a biosensor for estrogens in environmental samples.

Yarrowia lipolytica

Yarrowia lipolytica is a dimorphic yeast (see *Arxula adeninivorans*) that can grow on a wide range of substrates. It has a high potential for industrial applications but there are no recombinant products commercially available yet.

Rhodobacter spp.

Rhodobacter can be use as the recombinant microorganism platform. Similar to *E. coli*, there are libraries of mutants available as well as suitable plasmid vectors, allowing for rational design of various modules to enhance product yield. Isoprenoid pathways have been engineered in membraneous bacterial species of *Rhodobacter* for increased production of carotenoid and CoQ10. See, U.S. Patent Publication Nos. 20050003474 and 20040078846. Methods similar to those described above for *E. coli* can be used to make recombinant *Rhodobacter* microorganisms.

Candida boidinii

Candida boidinii is a methylotrophic yeast (it can grow on methanol). Like other methylotrophic species such as *Hansenula polymorpha* and *Pichia pastoris*, it provides an excellent platform for the production of heterologous proteins. Yields in a multigram range of a secreted foreign protein have been reported. A computational method, IPRO, recently predicted mutations that experimentally switched the cofactor specificity of *Candida boidinii* xylose reductase from NADPH to NADH.

Hansenula polymorpha (Pichia angusta)

Hansenula polymorpha is another methylotrophic yeast (see *Candida boidinii*). It can furthermore grow on a wide range of other substrates; it is thermo-tolerant and can assimilate nitrate (see also *Kluyveromyces lactis*). It has been applied to the production of hepatitis B vaccines, insulin and interferon alpha-2a for the treatment of hepatitis C, furthermore to a range of technical enzymes.

Kluyveromyces lactis

Kluyveromyces lactis is a yeast regularly applied to the production of kefir. It can grow on several sugars, most importantly on lactose which is present in milk and whey. It has successfully been applied among others to the production of chymosin (an enzyme that is usually present in the stomach of calves) for the production of cheese. Production takes place in fermenters on a 40,000 L scale.

Pichia pastoris

Pichia pastoris is a methylotrophic yeast (see *Candida boidinii* and *Hansenula polymorpha*). It provides an efficient platform for the production of foreign proteins. Platform elements are available as a kit and it is worldwide used in academia for the production of proteins. Strains have been engineered that can produce complex human N-glycan (yeast glycans are similar but not identical to those found in humans).

Physcomitrella spp.

Physcomitrella mosses, when grown in suspension culture, have characteristics similar to yeast or other fungal cultures. This genera is becoming an important type of cell for production of plant secondary metabolites, which can be difficult to produce in other types of cells.

B. Plant Cells or Plants

In some embodiments, the nucleic acids and polypeptides described herein are introduced into plants or plant cells to increase overall steviol glycoside production or enrich for the production of specific steviol glycosides in proportion to others. Thus, a host can be a plant or a plant cell that includes at least one recombinant gene described herein. A plant or plant cell can be transformed by having a recombinant gene integrated into its genome, *i.e.*, can be stably transformed. Stably transformed cells typically retain the introduced nucleic acid with each cell division. A plant or plant cell can also be transiently transformed such that the recombinant gene is not integrated into its genome. Transiently transformed cells typically lose all or some portion of the introduced nucleic acid with each cell division such that the introduced nucleic acid cannot be detected in daughter cells after a sufficient number of cell divisions. Both transiently transformed

and stably transformed transgenic plants and plant cells can be useful in the methods described herein.

Transgenic plant cells used in methods described herein can constitute part or all of a whole plant. Such plants can be grown in a manner suitable for the species under consideration, either in a growth chamber, a greenhouse, or in a field. Transgenic plants can be bred as desired for a particular purpose, *e.g.*, to introduce a recombinant nucleic acid into other lines, to transfer a recombinant nucleic acid to other species, or for further selection of other desirable traits. Alternatively, transgenic plants can be propagated vegetatively for those species amenable to such techniques. As used herein, a transgenic plant also refers to progeny of an initial transgenic plant provided the progeny inherits the transgene. Seeds produced by a transgenic plant can be grown and then selfed (or outcrossed and selfed) to obtain seeds homozygous for the nucleic acid construct.

Transgenic plants can be grown in suspension culture, or tissue or organ culture. For the purposes of this invention, solid and/or liquid tissue culture techniques can be used. When using solid medium, transgenic plant cells can be placed directly onto the medium or can be placed onto a filter that is then placed in contact with the medium. When using liquid medium, transgenic plant cells can be placed onto a flotation device, *e.g.*, a porous membrane that contacts the liquid medium.

When transiently transformed plant cells are used, a reporter sequence encoding a reporter polypeptide having a reporter activity can be included in the transformation procedure and an assay for reporter activity or expression can be performed at a suitable time after transformation. A suitable time for conducting the assay typically is about 1-21 days after transformation, *e.g.*, about 1-14 days, about 1-7 days, or about 1-3 days. The use of transient assays is particularly convenient for rapid analysis in different species, or to confirm expression of a heterologous polypeptide whose expression has not previously been confirmed in particular recipient cells.

Techniques for introducing nucleic acids into monocotyledonous and dicotyledonous plants are known in the art, and include, without limitation, *Agrobacterium*-mediated transformation, viral vector-mediated transformation, electroporation and particle gun transformation, U.S. Patent Nos 5,538,880; 5,204,253;

6,329,571; and 6,013,863. If a cell or cultured tissue is used as the recipient tissue for transformation, plants can be regenerated from transformed cultures if desired, by techniques known to those skilled in the art.

A population of transgenic plants can be screened and/or selected for those members of the population that have a trait or phenotype conferred by expression of the transgene. For example, a population of progeny of a single transformation event can be screened for those plants having a desired level of expression of a steviol or steviol glycoside biosynthesis polypeptide or nucleic acid. Physical and biochemical methods can be used to identify expression levels. These include Southern analysis or PCR amplification for detection of a polynucleotide; Northern blots, S1 RNase protection, primer-extension, or RT-PCR amplification for detecting RNA transcripts; enzymatic assays for detecting enzyme or ribozyme activity of polypeptides and polynucleotides; and protein gel electrophoresis, Western blots, immunoprecipitation, and enzyme-linked immunoassays to detect polypeptides. Other techniques such as *in situ* hybridization, enzyme staining, and immunostaining also can be used to detect the presence or expression of polypeptides and/or nucleic acids. Methods for performing all of the referenced techniques are known. As an alternative, a population of plants comprising independent transformation events can be screened for those plants having a desired trait, such as production of a steviol glycoside or modulated biosynthesis of a steviol glycoside. Selection and/or screening can be carried out over one or more generations, and/or in more than one geographic location. In some cases, transgenic plants can be grown and selected under conditions which induce a desired phenotype or are otherwise necessary to produce a desired phenotype in a transgenic plant. In addition, selection and/or screening can be applied during a particular developmental stage in which the phenotype is expected to be exhibited by the plant. Selection and/or screening can be carried out to choose those transgenic plants having a statistically significant difference in a steviol glycoside level relative to a control plant that lacks the transgene.

The nucleic acids, recombinant genes, and constructs described herein can be used to transform a number of monocotyledonous and dicotyledonous plants and plant cell systems. Non-limiting examples of suitable monocots include, for example, cereal

crops such as rice, rye, sorghum, millet, wheat, maize, and barley. The plant may be a non-cereal monocot such as asparagus, banana, or onion. The plant also may be a dicot such as stevia (*Stevia rebaudiana*), soybean, cotton, sunflower, pea, geranium, spinach, or tobacco. In some cases, the plant may contain the precursor pathways for phenyl phosphate production such as the mevalonate pathway, typically found in the cytoplasm and mitochondria. The non-mevalonate pathway is more often found in plant plastids [Dubey, et al., 2003 J. Biosci. **28** 637–646]. One with skill in the art may target expression of steviol glycoside biosynthesis polypeptides to the appropriate organelle through the use of leader sequences, such that steviol glycoside biosynthesis occurs in the desired location of the plant cell. One with skill in the art will use appropriate promoters to direct synthesis, e.g., to the leaf of a plant, if so desired. Expression may also occur in tissue cultures such as callus culture or hairy root culture, if so desired.

In one embodiment, one or more nucleic acid or polypeptides described herein are introduced into Stevia (e.g., *Stevia rebaudiana*) such that overall steviol glycoside biosynthesis is increased or that the overall steviol glycoside composition is selectively enriched for one or more specific steviol glycosides (e.g., rebaudioside D). For example, one or more recombinant genes can be introduced into *Stevia* such that a EUGT11 enzyme (e.g., SEQ ID NO: 152 or a functional homolog thereof) is expressed alone or in combination with one or more of: a UGT91D enzyme such as UGT91D2e (e.g., SEQ ID NO:5 or a functional homolog thereof), UGT91D2m (e.g., SEQ ID NO:10); a UGT85C enzyme such as a variant described in the "Functional Homolog" section, a UGT76G1 enzyme such as a variant described in the "Functional Homolog" section, or a UGT74G1 enzyme. Nucleic acid constructs typically include a suitable promoter (e.g., 35S, e35S, or ssRUBISCO promoters) operably linked to a nucleic acid encoding the UGT polypeptide. Nucleic acids can be introduced into Stevia by *Agrobacterium*-mediated transformation; electroporation-mediated gene transfer to protoplasts; or by particle bombardment. See, e.g., Singh, et al., Compendium of Transgenic Crop Plants: Transgenic Sugar, Tuber and Fiber, Edited by Chittaranjan Kole and Timothy C. Hall, Blackwell Publishing Ltd. (2008), pp. 97-115. For particle bombardment of stevia leaf

derived callus, the parameters can be as follows: 6 cm distance, 1100 psi He pressure, gold particles, and one bombardment.

Stevia plants can be regenerated by somatic embryogenesis as described by Singh et al., 2008, *supra*. In particular, leaf segments (approximately 1-2cm long) can be removed from 5 to 6-week-old *in vitro* raised plants and incubated (adaxial side down) on MS medium supplemented with B5 vitamins, 30 g sucrose and 3 g Gelrite. 2,4-dichlorophenoxyacetic acid (2,4-D) can be used in combination with 6-benzyl adenine (BA), kinetin (KN), or zeatin. Proembryogenic masses appear after 8 weeks of subculture. Within 2-3 weeks of subcultures, somatic embryos will appear on the surface of cultures. Embryos can be matured in medium containing BA in combination with 2,4-D, a-naphthaleneacetic acid (NAA), or indolbutyric acid (IBA). Mature somatic embryos that germinate and form plantlets can be excised from calli. After plantlets reach 3-4 weeks, the plantlets can be transferred to pots with vermiculite and grown for 6-8 weeks in growth chambers for acclimatization and transferred to greenhouses.

In one embodiment, steviol glycosides are produced in rice. Rice and maize are readily transformable using techniques such as *Agrobacterium*-mediated transformation. Binary vector systems are commonly utilized for *Agrobacterium* exogenous gene introduction to monocots. See, for example, U.S. Patent Nos. 6,215,051 and 6,329,571. In a binary vector system, one vector contains the T-DNA region, which includes a gene of interest (e.g., a UGT described herein) and the other vector is a disarmed Ti plasmid containing the *vir* region. Co-integrated vectors and mobilizable vectors also can be used. The types and pretreatment of tissues to be transformed, the strain of *Agrobacterium* used, the duration of the inoculation, the prevention of overgrowth and necrosis by the *Agrobacterium*, can be readily adjusted by one of skill in the art. Immature embryo cells of rice can be prepared for transformation with *Agrobacterium* using binary vectors. The culture medium used is supplemented with phenolic compounds. Alternatively, the transformation can be done *in planta* using vacuum infiltration. See, for example, WO 2000037663, WO 2000063400, and WO 2001012828.

IV. Methods of Producing Steviol Glycosides

Recombinant hosts described herein can be used in methods to produce steviol or steviol glycosides. For example, if the recombinant host is a microorganism, the method can include growing the recombinant microorganism in a culture medium under conditions in which steviol and/or steviol glycoside biosynthesis genes are expressed. The recombinant microorganism may be grown in a fed batch or continuous process. Typically, the recombinant microorganism is grown in a fermentor at a defined temperature(s) for a desired period of time. Depending on the particular microorganism used in the method, other recombinant genes such as isopentenyl biosynthesis genes and terpene synthase and cyclase genes may also be present and expressed. Levels of substrates and intermediates, e.g., isopentenyl diphosphate, dimethylallyl diphosphate, geranylgeranyl diphosphate, kaurene and kaurenoic acid, can be determined by extracting samples from culture media for analysis according to published methods.

After the recombinant microorganism has been grown in culture for the desired period of time, steviol and/or one or more steviol glycosides can then be recovered from the culture using various techniques known in the art. In some embodiments, a permeabilizing agent can be added to aid the feedstock entering into the host and product getting out. If the recombinant host is a plant or plant cells, steviol or steviol glycosides can be extracted from the plant tissue using various techniques known in the art. For example, a crude lysate of the cultured microorganism or plant tissue can be centrifuged to obtain a supernatant. The resulting supernatant can then be applied to a chromatography column, e.g., a C-18 column, and washed with water to remove hydrophilic compounds, followed by elution of the compound(s) of interest with a solvent such as methanol. The compound(s) can then be further purified by preparative HPLC. See also WO 2009/140394.

The amount of steviol glycoside (e.g., rebaudioside D) produced can be from about 1 mg/L to about 1500 mg/L, e.g., about 1 to about 10 mg/L, about 3 to about 10 mg/L, about 5 to about 20 mg/L, about 10 to about 50 mg/L, about 10 to about 100 mg/L, about 25 to about 500 mg/L, about 100 to about 1,500 mg/L, or about 200 to about 1,000 mg/L. In general, longer culture times will lead to greater amounts of product. Thus, the

recombinant microorganism can be cultured for from 1 day to 7 days, from 1 day to 5 days, from 3 days to 5 days, about 3 days, about 4 days, or about 5 days.

It will be appreciated that the various genes and modules discussed herein can be present in two or more recombinant microorganisms rather than a single microorganism. When a plurality of recombinant microorganisms is used, they can be grown in a mixed culture to produce steviol and/or steviol glycosides. For example, a first microorganism can comprise one or more biosynthesis genes for producing steviol while a second microorganism comprises steviol glycoside biosynthesis genes. Alternatively, the two or more microorganisms each can be grown in a separate culture medium and the product of the first culture medium, e.g., steviol, can be introduced into second culture medium to be converted into a subsequent intermediate, or into an end product such as rebaudioside A. The product produced by the second, or final microorganism is then recovered. It will also be appreciated that in some embodiments, a recombinant microorganism is grown using nutrient sources other than a culture medium and utilizing a system other than a fermentor.

Steviol glycosides do not necessarily have equivalent performance in different food systems. It is therefore desirable to have the ability to direct the synthesis to steviol glycoside compositions of choice. Recombinant hosts described herein can produce compositions that are selectively enriched for specific steviol glycosides (e.g., rebaudioside D) and have a consistent taste profile. Thus, the recombinant microorganisms, plants, and plant cells described herein can facilitate the production of compositions that are tailored to meet the sweetening profile desired for a given food product and that have a proportion of each steviol glycoside that is consistent from batch to batch. Microorganisms described herein do not produce the undesired plant byproducts found in Stevia extracts. Thus, steviol glycoside compositions produced by the recombinant microorganisms described herein are distinguishable from compositions derived from Stevia plants.

V. Food Products

The steviol glycosides obtained by the methods disclosed herein can be used to make food products, dietary supplements and sweetener compositions. For example, substantially pure steviol or steviol glycoside such as rebaudioside A or rebaudioside D can be included in food products such as ice cream, carbonated beverages, fruit juices, yogurts, baked goods, chewing gums, hard and soft candies, and sauces. Substantially pure steviol or steviol glycoside can also be included in non-food products such as pharmaceutical products, medicinal products, dietary supplements and nutritional supplements. Substantially pure steviol or steviol glycosides may also be included in animal feed products for both the agriculture industry and the companion animal industry. Alternatively, a mixture of steviol and/or steviol glycosides can be made by culturing recombinant microorganisms separately or growing different plants/plant cells, each producing a specific steviol or steviol glycoside, recovering the steviol or steviol glycoside in substantially pure form from each microorganism or plant/plant cells and then combining the compounds to obtain a mixture containing each compound in the desired proportion. The recombinant microorganisms, plants, and plant cells described herein permit more precise and consistent mixtures to be obtained compared to current Stevia products. In another alternative, a substantially pure steviol or steviol glycoside can be incorporated into a food product along with other sweeteners, e.g. saccharin, dextrose, sucrose, fructose, erythritol, aspartame, sucralose, monatin, or acesulfame potassium. The weight ratio of steviol or steviol glycoside relative to other sweeteners can be varied as desired to achieve a satisfactory taste in the final food product. See, e.g., U.S. Patent Publication No. 2007/0128311. In some embodiments, the steviol or steviol glycoside may be provided with a flavor (e.g., citrus) as a flavor modulator. For example, Rebaudioside C can be used as a sweetness enhancer or sweetness modulator, in particular for carbohydrate based sweeteners, such that the amount of sugar can be reduced in the food product.

Compositions produced by a recombinant microorganism, plant, or plant cell described herein can be incorporated into food products. For example, a steviol glycoside composition produced by a recombinant microorganism, plant, or plant cell can

be incorporated into a food product in an amount ranging from about 20 mg steviol glycoside/kg food product to about 1800 mg steviol glycoside/kg food product on a dry weight basis, depending on the type of steviol glycoside and food product. For example, a steviol glycoside composition produced by a recombinant microorganism, plant, or plant cell can be incorporated into a dessert, cold confectionary (e.g., ice cream), dairy product (e.g., yogurt), or beverage (e.g., a carbonated beverage) such that the food product has a maximum of 500 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism, plant, or plant cell can be incorporated into a baked good (e.g., a biscuit) such that the food product has a maximum of 300 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism, plant, or plant cell can be incorporated into a sauce (e.g., chocolate syrup) or vegetable product (e.g., pickles) such that the food product has a maximum of 1000 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism, plant, or plant cell can be incorporated into a bread such that the food product has a maximum of 160 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism, plant, or plant cell can be incorporated into a hard or soft candy such that the food product has a maximum of 1600 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism, plant, or plant cell can be incorporated into a processed fruit product (e.g., fruit juices, fruit filling, jams, and jellies) such that the food product has a maximum of 1000 mg steviol glycoside/kg food on a dry weight basis.

For example, such a steviol glycoside composition can have from 90-99% rebaudioside A and an undetectable amount of stevia plant-derived contaminants, and be incorporated into a food product at from 25-1600 mg/kg, e.g., 100-500 mg/kg, 25-100 mg/kg, 250-1000 mg/kg, 50-500 mg/kg or 500-1000 mg/kg on a dry weight basis.

Such a steviol glycoside composition can be a rebaudioside B-enriched composition having greater than 3% rebaudioside B and be incorporated into the food product such that the amount of rebaudioside B in the product is from 25-1600 mg/kg,

e.g., 100-500 mg/kg, 25-100 mg/kg, 250-1000 mg/kg, 50-500 mg/kg or 500-1000 mg/kg on a dry weight basis. Typically, the rebaudioside B-enriched composition has an undetectable amount of stevia plant-derived contaminants.

Such a steviol glycoside composition can be a rebaudioside C-enriched composition having greater than 15% rebaudioside C and be incorporated into the food product such that the amount of rebaudioside C in the product is from 20-600 mg/kg, e.g., 100-600 mg/kg, 20-100 mg/kg, 20-95 mg/kg, 20-250 mg/kg, 50-75 mg/kg or 50-95 mg/kg on a dry weight basis. Typically, the rebaudioside C-enriched composition has an undetectable amount of stevia plant-derived contaminants.

Such a steviol glycoside composition can be a rebaudioside D-enriched composition having greater than 3% rebaudioside D and be incorporated into the food product such that the amount of rebaudioside D in the product is from 25-1600 mg/kg, e.g., 100-500 mg/kg, 25-100 mg/kg, 250-1000 mg/kg, 50-500 mg/kg or 500-1000 mg/kg on a dry weight basis. Typically, the rebaudioside D-enriched composition has an undetectable amount of stevia plant-derived contaminants.

Such a steviol glycoside composition can be a rebaudioside E-enriched composition having greater than 3% rebaudioside E and be incorporated into the food product such that the amount of rebaudioside E in the product is from 25-1600 mg/kg, e.g., 100-500 mg/kg, 25-100 mg/kg, 250-1000 mg/kg, 50-500 mg/kg or 500-1000 mg/kg on a dry weight basis. Typically, the rebaudioside E-enriched composition has an undetectable amount of stevia plant-derived contaminants.

Such a steviol glycoside composition can be a rebaudioside F-enriched composition having greater than 4% rebaudioside F and be incorporated into the food product such that the amount of rebaudioside F in the product is from 25-1000 mg/kg, e.g., 100-600 mg/kg, 25-100 mg/kg, 25-95 mg/kg, 50-75 mg/kg or 50-95 mg/kg on a dry weight basis. Typically, the rebaudioside F-enriched composition has an undetectable amount of stevia plant-derived contaminants.

Such a steviol glycoside composition can be a dulcoside A-enriched composition having greater than 4% dulcoside A and be incorporated into the food product such that the amount of dulcoside A in the product is from 25-1000 mg/kg, e.g., 100-600 mg/kg,

25-100 mg/kg, 25-95 mg/kg, 50-75 mg/kg or 50-95 mg/kg on a dry weight basis. Typically, the dulcoside A-enriched composition has an undetectable amount of stevia plant-derived contaminants.

Such a steviol glycoside composition can be a composition enriched for rubusoside xylosylated on either of the two positions--the 13-O-glucose or the 19-O-glucose. Such a composition can have greater than 4% of the xylosylated rubusoside compound, and can be incorporated into the food product such that the amount of xylosylated rubusoside compound in the product is from 25-1000 mg/kg, e.g., 100-600 mg/kg, 25-100 mg/kg, 25-95 mg/kg, 50-75 mg/kg or 50-95 mg/kg on a dry weight basis. Typically, the xylosylated rubusoside enriched composition has an undetectable amount of stevia plant-derived contaminants.

Such a steviol glycoside composition can be a composition enriched for compounds rhamnosylated on either of the two positions--the 13-*O*-glucose or the 19-*O*-glucose, or compounds containing one rhamnose and multiple glucoses (e.g., steviol 13-*O*-1,3-diglycoside-1,2-rhamnoside). Such a composition can have greater than 4% of the rhamnosylated compound, and can be incorporated into the food product such that the amount of rhamnosylated compound in the product is from 25-1000 mg/kg, e.g., 100-600 mg/kg, 25-100 mg/kg, 25-95 mg/kg, 50-75 mg/kg or 50-95 mg/kg on a dry weight basis. Typically, the composition enriched for rhamnosylated compounds has as an undetectable amount of stevia plant-derived contaminants.

In some embodiments, a substantially pure steviol or steviol glycoside is incorporated into a tabletop sweetener or "cup-for-cup" product. Such products typically are diluted to the appropriate sweetness level with one or more bulking agents, e.g., maltodextrins, known to those skilled in the art. Steviol glycoside compositions enriched for rebaudioside A, rebaudioside C, rebaudioside D, rebaudioside E, rebaudioside F, dulcoside A, or rhamnosylated or xylosylated compounds, can be package in a sachet, for example, at from 10,000 to 30,000 mg steviol glycoside/kg product on a dry weight basis, for tabletop use.

In some embodiments, this disclosure relates to the following items:

1. A recombinant host cell comprising a nucleic acid sequence, said nucleic acid comprising a heterologous insert sequence operably linked to an open reading frame, wherein the heterologous insert sequence has the general formula (I):

$-X_1-X_2-X_3-X_4-X_5-$

wherein X_2 comprises at least 4 consecutive nucleotides complementary to at least 4 consecutive nucleotides of X_4,

wherein X_3 comprises zero nucleotides, or one or more nucleotides forming a hairpin loop,

wherein X_1 and X_5 each individually consists of zero nucleotides, or one or more nucleotides,

wherein the open reading frame encodes a squalene synthase (EC 2.5.1.21).

2. The recombinant cell according to item 1, said nucleic acid comprises in 5' to 3' order, a promoter sequence operably linkted to a heterologous insert sequence operably linked to an open reading frame, wherein the heterologous insert sequence and the open reading frame are as defined in item 1.

3. A cell comprising a nucleic acid sequence, said nucleic acid comprising

i) a promoter sequence operably linked to

ii) a heterologous insert sequence operably linked to

iii) an open reading frame operably linked to

iv) a transcription termination signal,

wherein the heterologous insert sequence has the general formula (I):

$-X_1-X_2-X_3-X_4-X_5-$

wherein X_2 comprises at least 4 consecutive nucleic acids being complementary to, and forms a hairpin secondary structure element with at least 4 consecutive nucleic acids of X_4, and

wherein X_3 comprises unpaired nucleic acids thus forming a hairpin loop between X_2 and X_4, and

wherein X_1 and X_5 individually and optionally comprises one or more nucleic acids, and

wherein the open reading frame upon expression encodes a polypeptide sequence having at least 70% identity to a squalene synthase (EC 2.5.1.21) or a biologically active fragment thereof, said fragment having at least 70% sequence identity to said squalene synthase in a range of overlap of at least 100 amino acids.

4. The cell according to any one of items 1 to 3, wherein the heterologous insert sequence comprises 10-50 nucleotides, preferably 10-30 nucleotides, more preferably 15-25 nucleotides, more preferably 17-22 nucleotides, more preferably 18-21 nucleotides, more preferably 18-20 nucleotides, more preferably 19 nucleotides.

5. The cell according to any one of items 1 to 4, wherein X_2 and X_4 consists of the same number of nucleotides.

6. The cell according to any one of items 1 to 5, wherein all X_2 consists of in the range of 4 to 25, such as in the range of 4 to 20, for example of in the range of 4 to 15, such as in the range of 6 to 12, for example in the range of 8 to 12, such as in the range of 9 to 11 nucleotides.

7. The cell according to any one of items 1 to 6, wherein all X_4 consists of in the range of 4 to 25, such as in the range of 4 to 20, for example of in the range of 4 to 15, such as in the range of 6 to 12, for example in the range of 8 to 12, such as in the range of 9 to 11 nucleotides.

8. The cell according to any one of items 1 to 7, wherein X_2 consists of a nucleotide sequence, which is complementary to the nucleotide sequence of X_4.

9. The cell according to any one of items 1 to 8, wherein X_4 consists of a nucleotide sequence, which is complementary to the nucleotide sequence of X_2.

10. The cell according to any one of items 1 to 9, wherein X_3 is absent, i.e. X_3 consists of zero nucleotides.

11. The cell according to any one of items 1 to 9, wherein X_3 consists of in the range of 1 to 5, such as in the range of 1 to 3 nucleotides.

12. The cell according to any one of items 1 to 11, wherein X_1 is absent, i.e. X_1 consists of zero nucleotides.

13. The cell according to any one of items 1 to 11, wherein X_1 consists of in the range of 1 to 25, such as in the range of 1 to 20, for example in the range of 1 to 15, such

as in the range of 1 to 10, for example in the range of 1 to 5, such as in the range of 1 to 3 nucleotides.

14. The cell according to any one items 1 to 13, wherein X_5 is absent, i.e. X_5 consists of zero nucleotides.

15. The cell according to any one of items 1 to 11, wherein X_5 consists of in the range 1 to 5, such as in the range of 1 to 3 nucleotides.

16. The cell according to any one of items 1 to 15, wherein the heterologous insert sequence comprises a sequence selected from the group consisting of SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 183 and SEQ ID NO: 184.

17. The cell according to any one of items 1 to 16, wherein the heterologous insert sequence is selected from the group consisting of SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 183 and SEQ ID NO: 184.

18. The cell according to any one of items 1 to 17, wherein the squalene synthase is at least 75%, such as at least 80%, such as at least 85%, such as at least 87%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% identical to a squalene synthase selected from the group consisting of SEQ ID NO: 192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, and SEQ ID NO:202.

19. The cell according to any one of items 1 to 18, wherein said promoter is a constitutive or inducible promoter.

20. The cell according to any one of items 1 to 19, wherein said promoter is selected from the group consisting of an endogenous promoter, GPD1, PGK1, ADH1, ADH2, PYK1, TPI1, PDC1, TEF1, TEF2, FBA1, GAL1-10, CUP1, MET2, MET14, MET25, CYC1, GAL1-S, GAL1-L, TEF1, ADH1, CAG, CMV, human UbiC, RSV, EF-1alpha, SV40, Mt1, Tet-On, Tet-Off, Mo-MLV-LTR, Mx1, progesterone, RU486 and Rapamycin-inducible promoter.

21. The cell according to any one of items 1 to 20, wherein the nucleic acid sequence further comprises a polyadenyl sequence.

22. The cell according to item 21, wherein the 5' end of said polyadenyl sequence is operably linked to the 3' end of the nucleic acid of item 1.

23. The cell according to any one of items 1 to 22, wherein the nucleic acid sequence further comprises a post-transcriptional regulatory element.

24. The cell of item 23, wherein said post-transcriptional regulatory element is a Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE).

25. The cell of any of items 1 to 24 , wherein the nucleic acid comprises a 5' terminal repeat and a 3' terminal repeat.

26. The cell of item 25, wherein the 5' and 3' terminal repeats are selected from Inverted Terminal Repeats [ITR] and Long Terminal Repeats [LTR].

27. The cell according to any one of items 1 to 26, wherein the nucleic acid sequence is integrated in a vector.

28. The cell of item 27, wherein the vector is an expression vector.

29. The cell of item 27, wherein the vector is selected from the group consisting of plasmid vectors, cosmids, artificial chromosomes and viral vectors.

30. The cell of item 29 wherein the plasmid vector can be maintained and replicated in bacteria, fungi and yeast.

31. The cell of item 29, wherein the viral vector is selected from the group consisting of vectors derived from the Retroviridae family including lentivirus, HIV, SIV, FIV, EAIV, CIV.

32. The cell of item 31, wherein the viral vector is selected from the group consisting of alphavirus, adenovirus, adeno associated virus, baculovirus, HSV, coronavirus, Bovine papilloma virus, Mo-MLV and adeno associated virus.

33. The cell of any of items 27 to 32, wherein said vector is functional in mammalian cells.

34. The cell of any of the preceding items wherein the cell is transformed or transduced with the vector of any of the items 27 to 33.

35. The cell of any of items 1 to 34, wherein said cell is a eukaryotic cell.

36. The cell of any of items 1 to 34, wherein said cell is a prokaryotic cell.

37. The cell of item 35, wherein said cell is selected from the group consisting of fungi cells such as yeast and aspergillus; microalgae such as Chlorella and Prototheca; plant cells; and mammalian cells, such as human, feline, porcine, simian, canine, murine, rat, mouse and rabbit cells.

38.The cell of item 37, wherein the yeast is selected from the group consisting of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Candida glabrata*, *Ashbya gossypii*, *Cyberlindnera jadinii*, and *Candida albicans*.

39. The cell of item 37, wherein the cell is selected from the group consisting of CHO, CHO-K1, HEI193T, HEK293, COS, PC12, HiB5, RN33b, BHK cells.

40. The cell of item 36, wherein said cell is *E. coli*, *Corynebacterium*, *Bacillus*, *Pseudomonas* or *Streptomyces*.

41. The cell of any of items 35 to 40, wherein said prokaryotic cell, or said fungi cell, has been genetically modified to express at least a portion of the enzymes of the mevalonate independent pathway.

42. The cell according to any one of items 1 to 41, wherein the cell furthermore comprises a heterologous nucleic acid encoding GGPPS operably linked to a nucleic acid sequence directing expression of GGPPS in said cell.

43. The cell according to item 42, wherein said GGPPS is selected from the groups consisting of SEQ ID NO: 126, SEQ ID NO:123, SEQ ID NO:203, SEQ ID NO:167, and functional homologues thereof sharing at least 75% sequence identity with any of the aforementioned.

43. A method for producing a terpenoid compound synthesized through the squalene pathway, in a cell culture, said method comprising the steps of

(a) providing the cell of of any one of items 1 to 42,

(b) culturing the cell of (a).

(c) recovering the terpenoid product compound.

44. A method for producing a terpenoid derived from a terpenoid precursor selected from the group consisting of Farnesyl-pyrophosphate (FPP), Isopentenyl-pyrophosphate (IPP), Dimethylallyl-pyrophosphate (DMAPP), Geranyl-pyrophosphate

(GPP) and/or Geranylgeranyl-pyrophosphate (GGPP),

said method comprising:

(a) contacting said precursor with an enzyme of the squalene synthase pathway,

(b) recovering the terpenoid product.

45. The method of any of items 44 and 45, wherein the terpenoid product is selected from the group consisting of hemiterpenoids, monoterpenes, sesquiterpenoids, diterpenoids, sesterpenes, triterpenoids, tetraterpenoids and polyterpenoids.

46. The method of item 44, wherein the terpenoid is selected from the group consisting of farnesyl phosphate, farnesol, geranylgeranyl, geranylgeraniol, isoprene, prenol, isovaleric acid, geranyl pyrophosphate, eucalyptol, limonene, pinene, farnesyl pyrophosphate, artemisinin, bisabolol, geranylgeranyl pyrophosphate, retinol, retinal, phytol, taxol , forskolin, aphidicolin, lanosterol, lycopene and carotene.

47. The method of item 46, wherein said method further comprises dephosphorylating the farnesyl phosphate to produce farnesol.

48. The method of item 44, wherein the enzyme of the squalene synthase pathway is selected from the group consisting of Dimethylallyltransferase (EC 2.5.1.1), Isoprene synthase (EC 4.2.3.27) and Geranyltranstransferase (EC 2.5.1.10).

49. A method for reducing the translation rate of a functional squalene synthase (EC 2.5.1.21) said method comprising:

(a) providing the cell of any one of items 1 to 42,

(b) culturing the cell of (a).

50. A method for decreasing turnover of farnesyl-pp to squalene, said method comprising:

(d) providing the cell of any one of items 1 to 42,

(e) culturing the cell of (a).

51. A method for enhancing accumulation of a compound selected from the group consisting of Farnesyl-pyrophosphate, Isopentenyl-pyrophosphate, Dimethylallyl-pyrophosphate, Geranyl-pyrophosphate and Geranylgeranyl-pyrophosphate, said method comprising the steps of:

(a) providing the cell of any one of items 1 to 42, and

(b) culturing the cell of (a).

52. The method of item 51, further comprising recovering the Farnesyl-pyrophosphate, Isopentenyl-pyrophosphate, Dimethylallyl-pyrophosphate, Geranyl-pyrophosphate or Geranylgeranyl-pyrophosphate compound.

53. The method of any one of items 51 and 52, further comprising recovering a compound synthesized through the squalene pathway, said compound being derived from said Farnesyl-pyrophosphate, Isopentenyl-pyrophosphate, Dimethylallyl-pyrophosphate, Geranyl-pyrophosphate and/or Geranylgeranyl-pyrophosphate.

54. The method of any of items 43 to 53, wherein the step of culturing the cell is performed in the presence of a squalene synthase inhibitor.

55. The method of any of items 43 to 54, wherein the cell additionally is genetically modified to enhance activity of and/or overexpress one or more enzymes selected from the group consisting of Phosphomevalonate kinase (EC 2.7.4.2), Diphosphomevalonate decarboxylase (EC 4.1.1.33), 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (EC 1.17.7.1), 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (EC 1.17.1.2), Isopentenyl-diphosphate Delta-isomerase 1 (EC 5.3.3.2), Short-chain Z-isoprenyl diphosphate synthase (EC 2.5.1.68), Dimethylallyltransferase (EC 2.5.1.1), Geranyltranstransferase (EC 2.5.1.10) and Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29).

56. The method of any of items 43 to 55, wherein the cell additionally is genetically modified to enhance activity of and/or overexpress one or more enzymes selected from the group consisting of acetoacetyl CoA thiolose, HMG-CoA reductase or the catalytic domain thereof, HMG-CoA synthase, mevalonate kinase, phosphomevalonate kinase, phosphomevalonate decarboxylase, isopentenyl pyrophosphate isomerase, farnesyl pyrophosphate synthase, D-1-deoxyxylulose 5-phosphate synthase, and 1-deoxy-D-xylulose 5-phosphate reductoisomerase and farnesyl pyrophosphate synthase.

57. The method of any of items 43 to 56, wherein the cell comprises a mutation in the ERG9 open reading frame.

58. The method of any of items 43 to 57, wherein the cell comprises an ERG9[Delta]::HIS3 deletion/insertion allele.

59. The method of any of items 43 to 58, wherein the step of recovering the compound comprises purification of said compound from the cell culture media.

VI. Examples

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims. In the examples described herein, the following LC-MS methodology was used to analyze steviol glycosides and steviol pathway intermediates unless otherwise indicated.

1) Analyses of steviol glycosides

LC-MS analyses were performed using an Agilent 1200 Series HPLC system (Agilent Technologies, Wilmington, DE, USA) fitted with a Phenomenex® kinetex C18 column (150 x 2.1 mm, 2.6 µm particles, 100 Å pore size) connected to a TSQ Quantum Access (ThermoFisher Scientific) triple quadropole mass spectrometer with a heated electrospray ion (HESI) source. Elution was carried out using a mobile phase of eluent B (MeCN with 0.1% Formic acid) and eluent A (water with 0.1% Formic acid) by increasing the gradient from 10->40 % B from min 0.0 to 1.0, increasing 40 -> 50% B in min 1.0 to 6.5, 50->100% B from min 6.5 to 7.0 and finally washing and re-equilibration. The flow rate was 0.4ml/min and the column temperature 30°C. The steviol glycosides were detected using SIM (Single Ion Monitoring) in positive mode with the following m/z-traces.

Description	Exact Mass	m/z trace	compound (typical t_R in min)
Steviol + 1 Glucose	$[M+H]^+$ 481.2796 $[M+Na]^+$ 503.2615	481.2± 0.5 503.1± 0.5	19-SMG (6.1), 13-SMG (6.4)
Steviol + 2 Glucose	$[M+Na]^+$ 665.3149	665± 0.5	Rubusoside (4.7) Steviol-1,2-bioside (5.2) Steviol-1,3-bioside (5.8)
Steviol + 3 Glucose	$[M+Na]^+$ 827.3677	827.4 ± 0.5	1,2-Stevioside (4.0) 1,3-Stevioside (4.4) Rebaudioside B (5.0)
Steviol + 4 Glucose	$[M+Na]^+$ 989.4200	989.4 ± 0.5	Rebaudioside A (3.9)

Steviol + 5 Glucose	$[M+Na]^+$ 1151.4728	1151.4 ± 0.5	Rebaudioside D (3.3)

The level of steviol glycosides were quantified by comparing with calibration curves obtained with authentic standards from LGC Standards. For example, standard solutions of 0.5 to 100 μM Rebaudioside A were typically utilized to construct a calibration curve.

2) Analyses of Steviol and *ent*-kaurenoic acid

LC-MS analyses of steviol and *ent*-kaurenoic acid were performed on the system described above. For the separation, a Thermo Science Hypersil Gold (C-18, 3μm, 100x2.1mm) column was used and a 20mM ammonium acetate aqueous solution was used as eluent A and acetonitrile as eluent B. The gradient conditions were: 20->55% B in min 0.0 to 1.0, 55->100 in min 1.0-7.0 and finally washing and re-equilibration. The flow rate was 0.5 mL/min and the column temperature 30°C. Steviol and *ent*-kaurenoic acid were detected using SIM (Single Ion Monitoring) in negative mode with the following m/z-traces.

Description	Exact Mass	m/z trace	typical t_R in min
Steviol	$[M-H]^-$ 317.2122	317.4 ± 0.5	3.3
Ent-kaurenoic acid	$[M-H]^-$ 301.2173	301.4 ± 0.5	5.5

3) HPLC quantification of UDP-Glucose

For the quantification of UDP-glucose, an Agilent 1200 Series HPLC system was used, with a Waters XBridge BEH amide (2.5μm, 3.0x50mm) column. Eluent A was a 10 mM ammonium acetate aqueous solution (pH 9.0) and Eluent B acetonitrile. The gradient conditions were: 95% B holding from min 0.0-0.5, decreasing from 95-50% B in min 0.5-4.5, holding 50% B from min 4.5-6.8 and finally re-equilibrating to 95% B. The flow rate was 0.9 mL/min and the column temperature 20°C. UDP-glucose was detected by UV_{262nm} absorbance.

The amount of UDP-glucose was quantified by comparing with a calibration curve obtained with a commercially available standard (e.g., from Sigma Aldrich).

Example 1 –Identification of EUGT11

Fifteen genes were tested for RebA 1,2-glycosylation activity. See Table 10.

Table 10

Name	Source	GenBank Accession No.
EUGT2	Oryza sativa UGT91 homolog	AP003270
EUGT3	Oryza sativa UGT91 homolog	AP005171
EUGT4	Oryza sativa UGT91 homolog	AP005643
EUGT6	Oryza sativa UGT91 homolog	AP005259
EUGT7	Oryza sativa UGT91 homolog	AP005171
EUGT8	Oryza sativa UGT91 homolog	XM_470006
EUGT9	Oryza sativa UGT91 homolog	AP005643
EUGT10	Oryza sativa UGT91 homolog	AC133334
EUGT11	Oryza sativa UGT91 homolog	AC133334
EUGT12	Oryza sativa UGT91 homolog	AC133334
EUGT15	Petunia x hybrid UGT79 homolog	Z25802
EUGT16	Arabidopsis thaliana UGT79 homolog	AC004786
EUGT17	Dianthus caryophyllus UGT79 homolog	AB294391
EUGT18	Ipomoea nil UGT79 homolog	AB192314
EUGT19	Oryza sativa UGT79 homolog	NM_001074394

In vitro transcription and translation of these genes was performed, and the resulting UGTs incubated with RebA and UDP-glucose. Following incubation, the reactions were analyzed by LC-MS. The reaction mixture containing EUGT11 (Rice, AC133334, SEQ ID NO:152) was shown to convert significant quantities of RebA to RebD. See LC-MS chromatograms in FIG. 4. As shown in the left panel of FIG. 4, UGT91D2e produced a trace amount of RebD when RebA was used as the feedstock. As shown in the right panel of FIG. 4, EUGT11 produced a significant amount of RebD when RebA was used as the feedstock. Preliminary quantification of the amount of RebD that was produced indicated that EUGT11 was approximately 30 times more efficient than UGT91D2e at converting RebA to RebD.

To further characterize EUGT11 and for quantitative comparison to UGT91D2e, the nucleotide sequence encoding EUGT11 (SEQ ID NO: 153, non-codon optimized, FIG. 7) was cloned into two *E. coli* expression vectors, one containing an *N*-terminal

HIS-tag and one containing an *N*-terminal GST-tag. EUGT11 was expressed using both systems and purified. When the purified enzymes were incubated with UDP-glucose and RebA, RebD was produced.

Example 2 –Identification of EUGT11 Reactions

EUGT11 was produced by *in vitro* transcription and translation, and incubated with various substrates in the RebD pathway. Similar experiments were carried out using *in vitro* transcribed and translated UGT91D2e. FIG. 3 shows a schematic overview of 19-O-1,2-diglycosylation reactions performed by EUGT11 and UGT91D2e. Compounds 1-3 were identified solely by mass and expected retention time. The numbers shown in FIG. 3 are the average peak height of the indicated steviol glycoside obtained from a LC-MS chromatogram, and, although not quantitative, can be used to compare the activity of the two enzymes. EUGT11 and UGT91D2e were not able to use steviol as a substrate. Both enzymes were able to convert steviol 19-*O*-monoglucoside (SMG) to compound 1, with EUGT11 being about ten times more efficient than UGT91D2e at converting 19-SMG to compound 1.

Both enzymes were able to convert rubusoside to stevioside with comparable activity but only EUGT11 was able to convert rubusoside to compound 2 and compound 3 (RebE). See FIG. 5. The left panel of FIG. 5 contains LC-MS chromatograms of the conversion of rubusoside to stevioside. The right panel of FIG. 5 contains chromatograms of the conversion of rubusoside to stevioside, to compound 2, and to compound 3 (RebE). Conversion of rubusoside to compound 3 requires two consecutive 1, 2-O-glycosylations at the 19- and 13-positions of steviol. UGT91D2e was able to produce a trace amount of compound 3 (RebE) in one experiment whereas EUGT11 produced a significant amount of compound 3.

Both enzymes were able to convert RebA to RebD. However, EUGT11 was approximately 30 times better at converting RebA to RebD. Overall, it appears that EUGT11 produces more product than UGT91D2e in all reactions (with similar time, concentrations, temperature, and purity of enzyme) except the conversion of rubusoside to stevioside.

Example 3 –Expression of EUGT11 in Yeast

The nucleotide sequence encoding EUGT11 was codon-optimized (SEQ ID NO:154) and transformed into yeast along with nucleic acids encoding all four UGTs (UGT91D2e, UGT74G1, UGT76G1, and UGT85C2). The resulting yeast strain was grown in medium containing steviol and steviol glycosides that accumulated were analyzed by LC-MS. EUGT11 was required for the production of RebD. In other experiments, RebD production has been observed with UGT91D2e, UGT74G1, UGT76G1, and UGT85C2.

Example 4 –UGT Activity on 19-O-1,2-diglycosylated steviol glycosides

The 19-O-1,2-diglycosylated steviol glycosides produced by EUGT11 need further glycosylation to be converted to RebD. The following experiments were performed to determine if other UGTs could use these intermediates as substrates.

In one experiment, compound 1 was produced *in vitro* from 19-SMG by either EUGT11 or UGT91D2e in the presence of UDP-glucose. After boiling the sample, UGT85C2 and UDP-glucose were added. The sample was analyzed by LC-MS and compound 2 was detected. This experiment indicated that UGT85C2 can use compound 1 as a substrate.

In another experiment, compound 2 was incubated with UGT91D2e and UDP-glucose. The reaction was analyzed by LC-MS. UGT91D2e was not able to convert compound 2 to compound 3 (RebE). Incubation of compound 2 with EUGT11 and UDP-glucose results in the production of compound 3. UGT76G1 was able to use RebE as a substrate to produce RebD.

This shows that the 19-O-1,2-diglycosylation of the steviol glycosides is able to take place at any time during production of RebD as the downstream enzymes are able to metabolize the 19-O-1,2-diglycosylated intermediates.

Example 5 – Comparison of EUGT11 and UGT91D2e Sequence

The amino acid sequence of EUGT11 (SEQ ID NO:152, FIG. 7) and the amino acid sequence of UGT91D2e (SEQ ID NO:5) were aligned using the FASTA algorithm (Pearson and Lipman, Proc. Natl. Acad. Sci., 85:2444-2448 (1998)). See FIG. 6. EUGT11 and UGT91D2e are 42.7% identical over 457 amino acids.

Example 6 –Modification of 19-1,2-diglycosylating activity of UGT91D2e

Crystal structures are available for a number of UGTs. Generally, the N-terminal half of a UGT is primarily involved with substrate binding whereas the C-terminal half is involved in binding the UDP-sugar donor.

Modeling the secondary structure of UGT91D2e onto the secondary structure of the UGTs that have been crystalized revealed a conserved pattern of secondary structure, despite a highly diverged primary sequence as shown in FIG. 8. The crystal structures of UGT71G1 and UGT85H2 (see, for example H. Shao *et al*, The Plant Cell November 2005 vol. 17 no 11 3141-3154 and L. Li *et al*., J Mol Biol. 2007 370(5):951-63) have been reported. Known loops, alpha-helices and beta-sheets are indicated on UGT91D2e in FIG. 8. Although the homology at the primary structure level of these UGTs is fairly low, the secondary structure appears to be conserved, allowing predictions regarding the locations of amino acids involved in substrate binding on UGT91D2e based on the location of such amino acids in UGT85H2 and UGT71G1.

Regions commonly involved in substrate binding were superimposed on UGT91D2e and largely shown to coincide with the 22 amino acid differences from UGT91D1 (GenBank Accession No. Protein Accession number AAR06918, GI:37993665). UGT91D1 is highly expressed in *Stevi*a and thought to be a functional UGT. However, its substrate is not a steviol glycoside. This suggests that UGT91D1 has a different substrate, which may be defined by the 22 amino acids with which it differs from UGT91D2e. FIG. 9 is an alignment of the amino acid sequences of UGT91D1 and UGT91D2e. The boxes represent areas that are reported to be involved in substrate binding. The amino acids highlighted in dark grey show the 22 amino acid differences between UGT91D1 and UGT91D2e. Stars denote amino acids that have been shown to

be involved in substrate binding in UGTs that have had their crystal structure resolved (more stars under one particular amino acids means substrate binding has been shown with more than one structure-resolved UGTs). There is a strong correlation between the 22 amino acid differences between the two UGT91s, the regions known to be involved in substrate binding, and the actual amino acids involved in substrate binding in the crystal structure-resolved UGTs. This suggests that the 22 amino acid differences between the two UGT91s are involved in substrate binding.

All 22 altered 91D2es were expressed in a XJb Autolysis *E. coli* strain from a pGEX-4T1 vector. In order to assess the activity of the enzymes, two substrate feeding experiments were performed - *in vivo* and *in vitro*. Most mutants had lower activity than wild type, however, 5 mutants showed increased activity. This was reproduced by *in vitro* transcription and translation (IVT) and showed that C583A, C631A and T857C have approximately 3-fold higher stevioside-forming activity than the wild-type UGT91D2e, whereas C662t and A1313C had approximately twice the stevioside-forming activity (nucleotide numbering). These changes result in amino acid mutations corresponding to L195M, L211M, V286A; and S221F and E438A, respectively. The increased activity differed depending on substrate, with C583A and C631A showing almost a 10-fold increase using 13-SMG as substrate and about a 3-fold increase using rubusoside as substrate, whereas T857C showed a 3-fold increase when using either 13-SMG or rubusoside as substrate.

To investigate if these mutations were additive, a range of double mutants were made and analyzed for activity (FIG. 10). In this particular experiment, a higher wild type level of activity was observed than the previous four experiments; however, the relative activities of the mutations remain the same. As rubusoside accumulates in many of the *S. cerevisiae* strains expressing the 4 UGTs (UGT74G1, UGT85C2, UGT76G1, and UGT91D2e), the stevioside-forming activity may be more important for increasing steviol glycoside production. As such, the double mutant C631A/T857C (nucleotide numbering) may be useful. This mutant has been named UGT91D2e-b, which contains the amino acid modifications L211M and V286A. The experiments have been reproduced *in vitro* using *S. cerevisiae*-expressed UGT91D2e-mutants.

To improve 19-1,2-diglycosylating activity of UGT91D2e, a directed saturated mutagenic screen of UGT91D2e of the 22 amino acid differences between UGT91D2e and UGT91D1 was performed. GeneArt's® (Life Technologies, Carlsbad, CA) site-saturation mutagenesis was used to obtain a library containing each of the mutations. The library was cloned into the BamHI and NotI sites of pGEX4T1 bacterial expression plasmid expressing the mutated versions of 91D2e as GST fusion proteins, resulting in a new library (Lib#116). Lib#116 was transformed into XJbAutolysis *E. coli* strain (ZymoResearch, Orange, CA) to produce approximately 1600 clones containing the 418 expected mutations (i.e., 22 positions with 19 different amino acids at each position). Other plasmids expressing GST-tagged versions of 91D2e (EPCS1314), 91D2e-b (EPSC1888) or EUGT11 (EPSC1744) as well as the empty pGEX4T1 (PSB12) were transformed as well.

Screening by LC-MS

To analyze the approximately 1600 mutant clones of UGT91D2e, the *E. coli* transformants were grown overnight at 30°C in 1 ml of NZCYM containing ampicillin (100 mg/l) and chloramphenicol (33 mg/l), in 96-well format. The next day, 150 µl of each culture was inoculated into 3 ml NZCYM containing ampicillin (100 mg/l), chloramphenicol (33 mg/l), arabinose 3 mM, IPTG 0.1 mM and ethanol 2% v/v, in 24-well format, and incubated at 20°C and 200 rpm for ~20h. The following day, cells were spun down and pellets were resuspended in 100 µl of lysis buffer containing 10mM Tris-HCl pH 8, 5 mM $MgCl_2$, 1 mM $CaCl_2$ and complete mini protease inhibitor EDTA-free (3 tablets/100 ml) (Hoffmann-La Roche, Basel, Switzerland) and frozen -80°C for at least 15 minutes to promote cell lysis. Pellets were thawed at room temperature and 50 µl of DNase mix (1 µl of 1.4 mg/ml DNase in H2O (~80000u/ml), 1.2 µl of $MgCl_2$ 500 mM and 47.8 µl of 4x PBS buffer solution) was added to each well. Plates were shaken at 500 rpm for 5 min at room temperature to allow degradation of genomic DNA. Plates were spun down at 4000 rpm for 30 min at 4°C and six µl of the lysates were used in UGT *in vitro* reactions as described for GST-91D2e-b, using rubusoside or rebaudioside A as substrates. In each case, the resulting compounds, stevioside or rebaudioside D (rebD), were measured by LC-MS. Results were analyzed in comparison with the stevioside or

rebD produced by the lysates expressing the corresponding controls (91D2e, 91D2e-b, EUGT11 and the empty plasmid). Clones showing activity similar to or higher than the ones expressing 91D2e-b were selected as primary hits.

Half of the 1600 clones and the corresponding controls were assayed for their capacity to glycosylate rubusoside and rebaudiosideA. Stevioside and RebD were quantified by LC-MS. Under the conditions used, lysates from clones expressing the native UGT91D2e show activity just around background with both substrates (approximately 0.5 μM stevioside and 1 μM RebD), while clones expressing UGT91D2e-b show consistently improved product formation (>10 μM Stevioside; > 1.5 μM RebD). Clones expressing EUGT11 consistently display a higher level of activity, especially using RebA as substrate. Cutoff for considering clones as *primary hits* in the screening was generally set at 1.5 μM for both products, but in some cases was adjusted for each independent assay.

Example 7 –EUGT11 Homologs

A Blastp search of the NCBI nr database using the EUGT11 protein sequence revealed approximately 79 potential UGT homologs from 14 plant species (one of which is the Stevia UGT91D1, approximately 67% identical to EUGT11 in conserved UGT regions but less than 45% overall). Homologs with greater than 90% identity in conserved regions were identified from corn, soybean, *Arabidopsis*, grape, and *Sorghum*. The overall homology of the full-length EUGT11 homologs, at the amino acid level, was only 28-68%.

RNA was extracted from plant material by the method described by Iandolino *et al.* (Iandolino et al., *Plant Mol Biol Reporter* 22, 269-278, 2004), the RNeasy Plant mini Kit (Qiagen) according to the manufacturer´s instructions, or using the Fast RNA Pro Green Kit (MP Biomedicals) according to the manufacturer´s instructions. cDNA was produced by AffinityScript QPCR cDNA Synthesis Kit (Agilent) according to the manufacturer´s instructions. Genomic DNA was extracted using the FastDNA kit (MP biomedicals) according to the manufacturer´s instructions. PCR was performed on cDNA

using either the Dream Taq polymerase (Fermentas) or the Phusion polymerase (New England Biolabs) and a series of primers designed to amplify the homologs.

PCR-reactions were analyzed by electrophoresis in SyberSafe-containing agarose-TAE gels. DNA was visualized by UV-irradiation in a trans illuminator. Bands of the correct size were cut out, purified through spin columns according to the manufacturer´s specifications, and cloned into TOPO-Zero blunt (for Phusion polymerase-generated products) or TOPO-TA (for Dream Taq-generated products). The TOPO-vectors containing the PCR-products were transformed into *E. coli* DH5Bα and plated on LB-agar plates containing the appropriate selective antibiotics. DNA was extracted from surviving colonies and sequenced. The genes with the correct sequence were cut out by restriction digest with SbfI and AscI, cloned into similarly digested IVT8 vector and transformed into *E. coli*. PCRs were performed on all cloned genes to amplify the gene and flanking regions required for *in vitro* transcription and translation. Proteins were produced from the PCR products by *in vitro* transcription and translation using the Promega L5540, TNT T7 Quick for PCR DNA Kit according to the manufacturer´s instructions. Production of protein was evaluated by incorporation of ^{35}S-methionine followed by separation by SDS-PAGE and visualization on a Typhoon phosphor-imager.

Activity assays were set up totaling 20% (by volume) of each *in vitro* reaction, 0.1 mM rubusoside or RebA, 5% DMSO, 100 mM Tris-HCl pH 7.0, 0.01 units Fast alkaline phosphatase (Fermentas), and 0.3 mM UDP-glucose (final concentrations). Following incubation at 30°C for one hour, the samples were analyzed by LC-MS for production of stevioside and RebD as described above. The UGT91D2e and UGT91D2e-b (double mutant described in Example 6) were used as positive controls, along with EUGT11. Under the initial assay conditions, clone P 64B (see Table 11) produced a trace amount of product using rubusoside and RebA. Table 11 lists the percent identity at the amino acid level compared to EUGT11 for the whole length of the UGTs, which ranges from 28-58%. High amounts of homology (96-100%) were observed over shorter stretches of sequences, which may indicate highly conserved domains of plant UGTs.

Table 11.

List of cloned EUGT11 homologs and their amino acid percent identity to EUGT11.

UGT	Accession	% identity to EUGT11
P44G	XP_002297733.1	32.16
P54A	XP_002532392.1	34.20
P51H	XP_002325254.1	32.53
P55D	XP_002533517.1	31.90
P5F	AAM12787.1	31.73
P48G	XP_002318358.1	33.20
P52F	XP_002334090.1	32.80
P48F	XP_002318358.1	33.00
T4B-b	NP_565540.4	31.19
P56C	XP_002533518.1	32.60
T67H	XP_002270294.1	34.06
T65E	CAN80742.1	34.98
T74G	XP_002270331.1	35.48
T65D	CAN80742.1	34.98
T69F1	XP_003635103.1	34.69
P6B	Q66PF2.1	33.20
P6D variant	Q66PF2.1	33.60
P64B	ACE87855.1	34.64
T3F	AT5G65550	34.94
P53H	XP_002527371.1	33.40
P53F	XP_002527371.1	33.40
P46H	XP_002303861.1	32.40
2-b	NP_199780.1	35.79
T70F	XP_002275802.1	36.67
T72A	XP_002275850.1	36.42
T71G	XP_002275824.1	37.25
P49G	XP_002320817.1	35.15
P57H	XP_002511902.1	36.23
45 Pop	XP_002302598.1	34.21
P50G	XP_002323718.1	32.86
P50H	XP_002323718.1	32.66
T73G	XP_002281094.1	32.05
63	XP_002458816.1	37.25
P78B	NP_001147674.1	35.33
62	XP_002458815.1	34.06
P9F	BAJ84800.1	37.92
T7H	NP_001240857.1	31.30
16-1	BAJ93155.1	58.03
T16H	BAJ93155.1	58.03
31TA	BAD35324.1	51.81

P41G	NP_001174664.1	35.40
P37G	NP_001051010.1	56.71
P60aH	XP_002466606.1	57.35
12	BAJ89368.1	44.35
P12A	BAJ89368.1	44.35
P12H	BAJ89368.1	44.35
P10B	BAJ86656.1	45.16
P58aF	XP_002463702.1	43.71
P59aG	XP_002463705.1	43.51
P76H	NP_001140711.1	28.81

Example 8 - Cell-free Biocatalytic Production of Reb-D

The cell-free approach is an *in vitro* system where RebA, stevioside or a steviol glycoside mixture is enzymatically converted to RebD. The system requires stoichiometric amounts of UDP-glucose and therefore UDP-glucose regeneration from UDP and sucrose using sucrose synthase can be used. Additionally, sucrose synthase removes UDP produced during the reaction, which improves conversion to glycosylated products by alleviated product inhibition observed for glycosylation reactions. See, WO 2011/153378.

Enzyme expression and purification

UGT91D2e-b (described in Example 6) and EUGT11 are key enzymes that catalyze the glycosylation of RebA yielding RebD. These UGTs were expressed in bacteria (*E. coli*) but one of ordinary skill in the art will appreciate that such proteins also can be prepared using different methods and hosts (e.g., other bacteria such as *Bacillus sp.*, yeast such as *Pichia sp.* or *Saccharomyces sp.*, other fungi (e.g., *Aspergillus*), or other organisms). For example, the proteins can be produced by *in vitro* transcription and translation or by protein synthesis.

The UGT91D2e-b and EUGT11 genes were cloned in pET30a or pGEX4T1 plasmids. Resulting vectors were transformed into an XJb (DE3) Autolysis *E. coli* strain (ZymoResearch, Orange, CA). Initially, *E. coli* transformants were grown overnight at 30°C in NZCYM medium, followed by induction with 3 mM arabinose and 0.1 mM IPTG, and further incubation overnight at 30°C. The corresponding fusion proteins were purified by affinity chromatography using included 6HIS- or GST- tags and standard

methods. One skilled in the art will appreciate that other protein purification methods such as gel filtration or other chromatography techniques also can be used, along with precipitation/crystallization or fractionation with e.g., ammonium sulfate. While EUGT11 was expressed well using the initial conditions, UGT91D2e-b required several modifications to the base protocol to increase protein solubility, including lowering the temperature of the overnight expression from 30°C to 20°C and adding 2% ethanol to the expression medium. Generally 2-4 mg/L of soluble GST-EUGT11 and 400-800 μg/l of GST-UGT91D2e-b were purified with this method.

Stability of EUGT11

Reactions were conducted to explore the stability of EUGT11 under various RebA to RebD reaction conditions. Omitting the substrate from the reaction mixture, EUGT11 was pre-incubated for various periods of time before substrate was added. Following a pre-incubation of the enzyme in 100 mM Tris-HCl buffer, substrate (100 μM RebA) and other reaction components (300 μM UDP-glucose, and 10 U/mL Alkaline Phosphatase (Fermentas/Thermo Fisher, Waltham, MA)) were added (0, 1, 4 or 24 hours after the incubation was started). The reaction was then allowed to proceed for 20h, after which the reactions were stopped and RebD product-formation measured. Experiments were repeated at different temperatures: 30°C, 32.7°C, 35.8°C and 37°C.

The activity of EUGT11 was reduced rapidly when the enzyme was pre-incubated at 37°C, reaching approximately half activity after 1 hour, and having almost no activity after 4 hours. At 30°C, the activity was not significantly reduced after 4 hours and after 24 hours, approximately one-third of the activity remained. This suggests that EUGT11 is heat-labile.

To assess the thermal stability of EUGT11 and to compare it with the other UGTs in the steviol glycosylation pathway, denaturation temperatures of the proteins were determined using differential scanning calorimetry (DSC). Use of DSC thermograms to estimate denaturation temperatures, T_D, is described, for example, by E. Freire in *Methods in Molecular Biology* 1995, Vol. 40 191-218. DSC was performed using 6HIS-purified EUGT11, yielding an apparent T_D of 39°C; while when GST-purified 91D2e-b was used, the measured T_D was 79°C. For reference, the measured T_D when using 6HIS-purified

UGT74G1, UGT76G1 and UGT85C2 was 86°C in all cases. One of skill in the art will recognize that enzyme immobilization or addition of thermal protectants can be added to the reactions to improve stability of the protein. Non-limiting examples of thermal protectants include trehalose, glycerol, ammonium sulphate, betaine, trimethylamine oxide, and proteins.

Enzyme kinetics

A series of experiments were performed to determine kinetic parameters of EUGT11 and 91D2e-b. For both enzymes, 100 μM RebA, 300 μM UDP-glucose, and 10 U/mL Alkaline Phosphatase (Fermentas/Thermo Fisher, Waltham, MA) were used in the reactions. For EUGT11, the reactions were performed at 37°C using 100 mM Tris-HCl, pH 7, and 2% enzyme. For 91D2e-b, the reactions were performed at 30°C using 20 mM Hepes-NaOH, pH 7.6, 20% (by volume) enzyme. The initial velocities (V_0) were calculated in the linear range of a product versus time plot.

To first investigate the linearity intervals, initial time-courses were done for each enzyme. EUGT11 was assayed at 37˚C for 48h at initial concentrations of 100 μM RebA and 300 μM UDP-glucose. UGT91D2e-b was assayed at 37˚C for 24h at initial concentrations of 200 μM RebA and 600 μM UDP-glucose. Based on these range-finding studies, it was determined that the initial 10 minutes in the case of EUGT11, and the initial 20 minutes in UGT91D2e-b would be in the linear range with respect to product formation, and therefore initial velocities of each reaction were calculated in those intervals. In the case of EUGT11, RebA concentrations assayed were 30 μM, 50 μM, 100 μM, 200 μM, 300 μM and 500 μM. Concentration of UDP-glucose was always three times the concentration of RebA and incubation was performed at 37˚C. By plotting the calculated V_0 as a function of the substrate concentrations, Michaelis-Menten curves were generated. By plotting the reciprocal of V_0 and the reciprocal of [S], a Lineweaver-Burk graphic was obtained, with $y = 339.85x + 1.8644$; $R^2 = 0.9759$.

V_{max} and K_M parameters were determined from the curve-fit Lineweaver-Burk data, calculated from the x- and y intercepts ($x=0$, $y=1/V_{max}$) and ($y=0$, $x=-1/K_M$). Additionally, the same parameters also were calculated by a non-linear least squares regression, using the SOLVER function in Excel. The results obtained with both methods

for EUGT11 and RebA are presented in Table 12, along with all the kinetic parameters of this example. Results from both Nonlinear Least Square Fit method and Lineweaver-Burk plot are presented in Table 12. K_{cat} is calculated based on V_{max} divided by the approximate amount of protein in the assay.

Table 12

Comparison of kinetic parameters for EUGT11 and UGT 91D2e-b, with RebA or UDP-glucose as substrate.

	Nonlinear Least Square Fit				Lineweaver-Burk plot			
	Reb A		UDP-glucose		Reb A		UDP-glucose	
	EUGT11	91D2e-b	EUGT11	91D2e-b	EUGT11	91D2e-b	EUGT11	91D2e-b
V_{max} ($\mu M.min^{-1}$)	0.52	0.34	0.79	0.19	0.54	0.44	0.78	0.18
K_{cat} (min^{-1})	8.11	0.32	12.32	0.2	8.42	0.41	12.1	0.19
K_M (μM)	162.5	1150	130	45.1	182.3	1580	118	41.9
K_{cat}/K_M ($min^{-1}.\mu M^{-1}$)	0.05	0.000275	0.095	0.00454	0.046	0.000258	0.102	0.00463

In order to investigate the influence of UDP-glucose concentration in the glycosylation reaction, as well as the affinity of EUGT11 for UDP-glucose, similar kinetics analysis were performed. EUGT11 was incubated with increasing amounts of UDP-glucose (20 μM, 50 μM, 100 μM, and 200 μM), maintaining an excess of RebA (500 μM). The kinetic parameters were calculated as described above, and shown in Table 12.

In the case of UGT91D2e-b, RebA concentrations assayed were 50 μM, 100 μM, 200 μM, 300 μM, 400 μM and 500 μM. Concentration of UDP-glucose was always three times the concentration of RebA and incubation was performed at 30˚C, in the reaction conditions described above for UGT91D2e-b. The kinetic parameters were calculated as previously described; and the resulting kinetic parameters are shown in Table 12. Additionally, kinetic parameters of UGT91D2e-b towards UDP-glucose were determined. UGT91D2e-b was incubated with increasing amounts of UDP-glucose (30 μM, 50 μM , 100 μM, and 200 μM), maintaining an excess of RebA (1500 μM). incubation was performed at 30˚C, in optimal conditions for UGT91D2e-b. The kinetic parameters were calculated as previously described and results are presented in Table 12.

By comparison of the kinetics parameters for EUGT11 and 91D2e-b, it was concluded that 91D2e-b has a lower K_{cat} and has lower affinity for RebA (higher K_M) although the K_M for UDP-glucose of 91D2e-b is lower than EUGT11. UGT91D2e-b has a lower K_{cat}/K_M which is a measure of catalytic efficiency, combining information on rate of catalysis with a particular substrate (Kcat) and the strength of enzyme-substrate binding (K_M).

Determining the limiting factor in reactions

Under the conditions described above for EUGT11, approximately 25% of the RebA administered was converted to RebD. The limiting factor in these conditions could be either the enzyme, UDP-glucose or RebA. Experiments were set up to distinguish between these possibilities. A standard assay was allowed to run its course during 4 hours. This was followed by addition of either extra RebA substrate, extra enzyme, extra UDP-glucose or extra enzyme and UDP-glucose. Addition of extra enzyme resulted in a relative increase of the conversion of around 50%, adding extra RebA or UDP-glucose

alone did not increase the conversion significantly, but the simultaneous addition of enzyme and UDP-glucose increased the conversion approximately 2-fold.

Experiments were conducted to examine the limit to this benefit of adding bolus amounts of UDP-glucose and fresh enzyme in the conversion of RebA to RebD reaction. Additional enzyme or enzyme and UDP-glucose were added after 1, 6, 24 and 28 hours. In the case of the addition of both extra EUGT11 and UDP-glucose, a conversion of more than 70% was achieved. No other components had a significant effect on the conversion. This indicates that EUGT11 is a primary limiting factor for the reaction but UDP-glucose also is limiting. As UDP-glucose is present at 3-fold higher concentration than RebA, this indicates that UDP-glucose may be somewhat unstable in the reaction mixture, at least in the presence of EUGT11. Alternatively, as explained below, EUGT11 may be metabolizing the UDP-glucose.

Inhibition studies

Experiments were conducted to determine if factors such as sucrose, fructose, UDP, product (RebD) and impurities in the less pure Stevia extracts raw materials inhibited the extent of the conversion of steviol glycoside substrates to RebD. In a standard reaction mixture, excess of the potential inhibitors (sucrose, fructose, UDP, RebD, or a commercial blend of steviol glycosides (*Steviva*, Steviva Brands, Inc., Portland, OR)) were added. Following incubation, RebD-production was quantified. Addition of 500 μg/ml of the commercial *Steviva* mix (approximately 60% 1,2-stevioside, 30% RebA, 5% Rubusoside, 2% 1,2-bioside, less than 1% of RebD, RebC and others, as evaluated by LC-MS) was not found to be inhibitory, but rather increased the overall RebD production (to around 60 μM from around 30 μM without any addition) well beyond the RebD originally added with the blend (around 5 μM). From the molecules tested, only UDP was shown to have an inhibitory effect on RebD-production at the concentration used (500 μM), as measured by LC-MS. The RebD that was produced was less than 7 μM. This inhibition can be alleviated in the *in vivo* or *in vitro* reactions for RebD production, by including an UDP recycling system to UDP-glucose, either by yeast or by an added SUS (sucrose synthase enzyme) in conjunction with sucrose. Moreover, when working with lower amounts of UDP-glucose (300 μM), the

addition of alkaline phosphatase to remove UDP-G does not increase the amount of RebD produced in the *in vitro* glycosylations substantially, suggesting that the UDP produced may not be inhibitory at these concentrations.

RebA vs crude steviol glycoside mix

In some experiments, a crude steviol glycoside mix was used as a source of RebA instead of purified RebA. As such a crude steviol glycoside mix contains a high percentage of stevioside along with RebA, UGT76G1 was included in the reactions. *In vitro* reactions were performed as described above using 0.5 g/l of the *Steviva*® mix as substrate and enzyme (UGT76G1 and/or EUGT11) and incubated at 30°C. The presence of steviol glycosides was analyzed by LC-MS.

When only UGT76G1 was added to the reactions, stevioside was converted to RebA quite efficiently. An unknown penta-glycoside (with a retention time peak at 4.02 min) also was detected. When only EUTG11 was added to the reaction, large amounts of RebE, RebA, RebD and an unknown steviol-pentaglycoside (with a retention time peak at 3.15 min) were found. When both EUGT11 and UGT76G1 were added to the reactions, the stevioside peak was reduced, and almost entirely converted to RebA and RebD. There were trace amounts of the unknown steviol-pentaglycoside (peak at 4.02 min). No RebE was detected nor was the second unknown steviol-pentaglycoside (peak at 3.15 min). This result indicated that the use of stevia extracts as a substrate to produce RebD *in vitro* is possible when EUGT11 and UGT76G1 are used in combination.

Non-specific UDP-glucose metabolism

To determine if EUGT11 can metabolize UDP-glucose independently of the conversion of RebA to RebD, GST-purified EUGT11 was incubated in the presence or absence of RebA substrate, and UDP-glucose usage was measured as UDP-release, using the TR-FRET Transcreener® kit (BellBrook Labs). The Transcreener® kit is based on a tracer molecule bound to an antibody. The tracer molecule is displaced by UDP or ADP in a highly sensitive and quantitative manner. The FP kit includes an Alexa633 tracer bound to an antibody. The tracer is displaced by UDP/ADP. The displaced tracer freely rotates leading to a decrease in fluorescence polarization. Therefore, UDP production is proportional to a decrease in polarization. The FI kit includes a quenched Alexa594

Tracer bound to an antibody, which is conjugated to an IRDye® QC-1 quencher. The tracer is displaced by UDP/ADP, whereby the displaced tracer is un-quenched, leading to a positive increase in fluorescence intensity. Therefore, UDP production is proportional to an increase in fluorescence. A TR-FRET kit includes a HiLyte647 Tracer bound to an Antibody-Tb conjugate. Excitation of the terbium complex in the UV range (ca. 330 nm) results in energy transfer to the tracer and emission at a higher wavelength (665nm) after a time delay. The tracer is displaced by UDP/ADP causing a decrease in TR-FRET.

It was observed that UDP-glucose measured was the same independent of the presence of RebA substrate. UDP release was not detectable in the absence of enzyme. This indicates a non-specific degradation of UDP-glucose by EUGT11. Nevertheless, RebD was still produced when RebA was added, suggesting that EUGT11 would preferentially catalyze RebA glycosylation over the non-specific UDP-glucose degradation.

Experiments were set up to find out the destiny of the glucose molecule in the absence of RebA or other obvious glycosylation substrates. One common factor in all previous reactions was the presence of Tris buffer and/or trace amounts of glutathione, which both contain potential glycosylation sites. The effect of these molecules on the non-specific UDP-glucose consumption was assayed using GST-purified EUGT11 (with glutathione) and HIS-purified enzyme (without glutathione) in *in vitro* reactions, in the presence or absence of RebA. UDP-glucose usage was measured as UDP-release, using the TR-FRET Transcreener® kit. UDP release occurred in all cases and was independent of the presence of RebA. UDP release was slower when the HIS-purified enzyme was used, but the overall catalytic activity of the enzyme in conversion of RebA to RebD was also lower, suggesting a lower amount of active soluble enzyme present in the assay. Therefore, it appears that the UDP-glucose metabolism by EUGT11 is independent of the presence of substrate and independent of the presence of glutathione in the reaction, under the conditions tested.

To test the effect of Tris on the metabolism of UDP-glucose by EUGT11, GST-EUGT11 was purified using a Tris- or a PBS-based buffer for the elution, obtaining similar amounts of protein in both cases. Tris- and PBS-purified enzymes were used in *in*

vitro reactions using Tris and HEPES as buffers respectively, in the presence or absence of RebA in a similar manner as above. In both conditions, the UDP release was the same in the reactions whether RebA was added or not, indicating that the metabolism of UDP-glucose by EUGT11 is independent of both the presence of RebA and Tris in the reaction. This suggests that the UDP-release detected may somehow be an artifact caused by a property of EUGT11 or, alternatively, EUGT11 may be hydrolyzing UDP-glucose. EUGT11 is still efficient at converting RebA to RebD preferentially and the loss of UDP-G can be compensated by addition of the sucrose synthase recycling system described below.

RebA solubility

The solubility of RebA determines the concentration that can be used both for the whole-cell approach and for the cell-free approach. Several different solutions of RebA in water were made and left at room temperature for several days. After 24 hours of storage, RebA precipitated at concentrations of 50 mM or higher. Twenty-five mM RebA started to precipitate after 4-5 days, while 10 mM or lower concentrations remained in solution, even when stored at 4°C.

RebD solubility

The solubility of RebD was assessed by making several different solutions of RebD in water were made and incubated at 30°C for 72 hours. RebD was found to be soluble in water initially in concentrations of 1 mM or lower while concentrations of 0.5 mM or less were found to be stable for longer periods of time. One with skill in the art will recognize that the solubility can be influenced by any number of conditions such as pH, temperature, or different matrices.

Sucrose synthase

Sucrose synthase (SUS) has been used to regenerate UDP-glucose from UDP and sucrose (FIG. 11) for other small molecule glycosylations (Masada Sayaka *et al. FEBS Letters* 581 (2007) 2562–2566.). Three SUS1 genes from *A. thaliana*, *S. rebaudiana* and coffee (*Coffea arabica*) were cloned into pGEX4T1 *E. coli* expression vectors (see FIG. 17 for the sequences). Using methods similar to those described for EUGT11, around 0.8 mg/l of GST-AtSUS1 (*A. thaliana* SUS1) was purified. Initial expression of CaSUS1

(*Coffea arabica* SUS1) and SrSUS1 (*S. rebaudiana* SUS1) followed by GST-purification did not produce significant amounts of protein although, when analyzed by western blot, the presence of GST-SrSUS1 was verified. When GST-SrSUS1 was expressed at 20°C in the presence of 2% ethanol, approximately 50 μg/l of enzyme was produced.

Experiments were performed to evaluate the UDP-glucose regenerating activity of the purified GST-AtSUS1 and GST-SrSUS1. *In vitro* assays were conducted in 100 mM Tris-HCl pH=7.5 and 1 mM UDP (final concentration). Either ~2.4 μg of purified GST-AtSUS1, ~0.15 μg of GST-SrSUS1, or ~1.5 μg commercial BSA (New England Biolabs, Ipswich, MA) were also added. Reactions were done in presence or absence of ~200 mM sucrose and incubated at 37°C for 24h. Product UDP-glucose was measured by HPLC as described in the analytical section. AtSUS1 produced ~0.8 mM UDP-glucose when sucrose was present. No UDP-glucose was observed when SrSUS1 or the negative control (BSA) was used. The lack of activity observed for SrSUS1 could be explained by the poor quality and concentration of the purified enzyme. UDP-glucose production by AtSUS1 was sucrose dependent and, therefore, it was concluded that AtSUS1 can be used in a coupled reaction to regenerate the UDP-glucose used by EUGT1 or other UGTs for small molecule glycosylation (FIG. 11, above).

SUS catalyzes the formation of UDP-glucose and fructose from sucrose and from UDP as depicted in FIG. 11. This UDP-glucose then can be used by EUGT11 for glycosylation of RebA to produce RebD. *In vitro* assays as described above were performed, adding ~200 mM sucrose, 1 mM UDP, 100 μM RebA, ~1.6 μg purified GST-AtSUS1 and ~0.8 μg GST-EUGT11. Formation of product, RebD, was evaluated by LC-MS. When AtSUS, EUGT11, sucrose and UDP were mixed with RebA, 81±5 μM of RebD was formed. The reaction was dependent on the presence of AtSUS, EUGT11 and sucrose. The conversion rate was similar to what has been observed previously using UDP-glucose provided extraneously. This shows that AtSUS can be used to regenerate UDP-glucose for RebD-formation by EUGT11.

Example 9: Whole-cell Biocatalytic Production of RebD

In this example, several parameters were studied that are factors for using whole cell biocatalytic systems in the production of RebD from RebA or other steviol glycosides. The ability of raw materials to cross the cell membrane and availability of UDP-glucose are two such factors. Permeabilizing agents were studied as well as different cell types to ascertain which systems may be the most beneficial for RebD production.

Permeabilizing agents

Several different permeabilization agents have previously been shown to allow intracellular enzymatic conversion of various compounds that are normally not able to cross a cell membrane (Chow and Palecek, *Biotechnol Prog*. 2004 Mar-Apr;20(2):449-56). In several cases, the approaches resemble a partial lysis of the cells and, in yeast, often rely on the removal of the cell membrane by a detergent and the encapsulation of the enzymes inside of the remaining cell wall, which is permeable to smaller molecules. Common to these methods is the exposure to the permeabilizing agent followed by a centrifugation step to pellet cells before the addition of the substrate. See , for example, Flores *et al.*, *Enzyme Microb. Technol.*, 16, pp. 340-346 (1994); Presecki & Vasic-Racki, *Biotechnology Letters*, 27, pp. 1835-1839 (2005); . Yu *et al.*, *J Ind Microbiol Biotechnol*, 34, 151-156 (2007); Chow and Palecek, *Cells. Biotechol. Prog.*, 20, pp. 449-456 (2004); Fernandez *et al.*, *Journal of Bacteriology*, 152, pp. 1255-1264 (1982); Kondo *et al.*, *Enzyme and microbial technology*, 27, pp. 806-811 (2000); Abraham and Bhat, *J Ind Microbiol Biotechnol*, 35, pp. 799-804 (2008); Liu *et al.*,: *Journal of bioscience and bioengineering*, 89, pp. 554-558 (2000); and Gietz and Schiestl, *Nature Protocols*, 2, pp. 31-34 (2007) regarding permeabilization of yeast. See, Naglak and Wang, *Biotechnology and Bioengineering*, 39, pp. 732-740 (1991); Alakomi *et al.*, *Applied and environmental Microbiology*, 66, pp. 2001-2005 (2000); and Fowler and Zabin, *Journal of bacteriology*, 92, pp. 353-357 (1966) regarding permeabilization of bacteria. As described in this example, it was determined if cells could remain viable and therefore could retain *de novo* UDP-glucose biosynthesis.

Experiments were done to establish conditions for permeabilization in *E. coli* and in yeast. Growing cells (*S. cerevisiae* or *E. coli*) were treated with different

concentrations/combinations of permeabilization agents: toluene, chloroform and ethanol for permeabilization of *S. cerevisiae*, and guanidine, lactic acid, DMSO and/or Triton X-100 for permeabilization of *E. coli*. Tolerance of both model organisms to high concentrations of RebA and other potential substrates also was evaluated. The permeabilization was measured by the amount of RebD produced from a EUGT11-expressing organism after incubation in a RebA containing medium (feeding experiment). Enzyme activity was monitored before and after exposure to the permeabilizing agents by lysing the cells and analyzing the activity of the released UGTs in an *in vitro* assay.

In yeast, none of the permeabilization conditions tested resulted in an increase on RebD above the detected background (i.e., contaminating RebD levels present in the RebA stock used for feeding). This indicates that, under the tested conditions, yeast cells remain impermeable to RebA and/or the reduced cell viability caused by the solvents results in a decrease of EUGT11 activity as well.

In *E.coli*, none of the conditions tested resulted in permeabilization of the cells and subsequent production of RebD above background levels. Detectable levels of RebD were measured when lysates from strains expressing EUGT11 were used in the *in vitro* reactions (data not shown), indicating that EUGT11 enzyme is present and active even after all permeabilization treatments (though the level of activity varies). The permeabilization treatments had little or no effect on cell viability, except treating cultures with 0.2 M guanidine and 0.5% TritonX-100, which severely decreased viability.

S. cerevisiae also was subjected to permeabilization assays not allowing further growth of the cells using Triton X-100, N-lauryl sarcosine (LS), or Lithium acetate + polyethylene glycol (LiAc+PEG). That is, under these conditions, permeabilization renders the cells unviable by removing the cell membrane altogether while retaining the cell-wall as a barrier to keep enzymes and gDNA inside. In such methods, UDP-glucose can be supplemental or recycled as described above. The advantage of permeabilization versus the purely *in vitro* approach is that individual enzymes do not need to be separately produced and isolated.

N-Lauryl sarcosine treatment resulted in inactivation of EUGT11 and only a minor increase in RebD was detected when LiAc/PEG was applied (data not shown). Treatment with Triton X-100 0.3% or 0.5%, however, increased the amount of RebD above background levels (see Figure 18) while sustaining the activity of EUGT11. For Triton X-100 assays, overnight cultures were washed three times in PBS buffer. Cells corresponding to 6 OD_{600} units were resuspended in PBS containing 0.3% or 0.5% Triton X-100 respectively. Treated cells were vortexed and incubated 30 minutes at 30˚C. After treatment, cells were washed in PBS buffer. Cells corresponding to 5 OD_{600} units were used in an *in vitro* assay, as described for GST-EUGT11 and 0.6 OD_{600} units were resuspended in reaction buffer and incubated overnight at 30°C as described for the LS treated samples. Untreated samples were used as controls.

Lysates from transformants expressing EUGT11 were able to convert some RebA into RebD (8 to 50 μM were measured in the reactions) when cells were untreated or after treatment with LiAC/PEG or Triton X100. However, no RebD was measured in lysates of cell pellets treated with LS. Permeabilized but non-lysed cells were able to produce some RebD (1.4 to 1.5 μM measured) when treated with 0.3% or 0.5% Triton X100 (Figure 18) while no RebD was found on the samples treated with LS or LiAC/PEG. These results show that RebD can be produced from RebA biocatalytically using whole cells and using Triton X100 as the permeabilizing agent.

Example 10 – Assessment of codon optimized UGT sequences

Optimal coding sequences for UGT 91d2e, 74G1, 76G1, and 85C2 were designed and synthesized for yeast expression using two methodologies, supplied by GeneArt (Regensburg, Germany) (SEQ ID NOs: 6, 2, 8, and 4, respectively) or DNA 2.0 (Menlo Park, CA) (SEQ ID NOs: 84, 83, 85, and 82, respectively). The amino acid sequences of UGT 91d2e, 74G1, 76G1, and 85C2 (SEQ ID NOs: 5, 1, 7, and 3, respectively) were not changed.

The wild-type, DNA 2.0, and GeneArt sequences were assayed for *in vitro* activity to compare reactivity on substrates in the steviol glycosides pathway. UGTs were inserted in high copy (2μ) vectors and expressed from a strong constitutive

promoter (GPD1) (vectors P423-GPD, P424-GPD, P425-GPD, and P426-GPD). The plasmids were transformed individually into the universal Watchmaker strain, EFSC301 (described in Example 3 of WO2011/153378) and assays were carried out using cell lysates prepared from equal amount of cells (8 OD units). For the enzymatic reactions, 6 μL of each cell lysate were incubated in a 30 μL reaction with 0.25 mM steviol (final concentration) to test UGT74G1 and UGT85C2 clones, and with 0.25 mM 13-SMG (13SMG) (final concentration) to test 76G1 and 91D2e UGTs. Assays were carried out for 24 hours at 30˚C. Prior to LC-MS analysis, one volume of 100% DMSO was added to each reaction, samples were centrifuged at 16000g, and the supernatants analysed.

The lysates expressing the GeneArt-optimized genes provided higher levels of UGT activity under the conditions tested. Expressed as a percentage of the wild-type enzyme, the GeneArt lysates showed equivalent activity to the wild-type for UGT74G1, 170% activity for UGT76G1, 340% activity for UGT85C2 and 130% activity for UGT91D2e. Using UGT85C2 may improve the overall flux and productivity of cells for production of Reb-A and Reb-D when expressed in *S. cerevisiae*.

Further experiments were conducted to determine if the codon-optimized UGT85C2 could reduce 19-SMG accumulation and increase rubusoside and higher glycosylated steviol glycosides production. The production of 19-SMG and rubusoside were analysed in a steviol-feeding experiment of *S. cerevisiae* strain BY4741 expressing the wild type UGT74G1 as well as the codon-optimized UGT85C2 from high copy (2μ) vectors under strong constitutive promoter (GPD1) (vectors P426-GPD and P423-GPD, respectively). Whole culture samples (without cell removal) were taken and boiled in an equal volume of DMSO for total glycosides levels. Intracellular concentrations reported were obtained by pelleting cells, and resuspending in 50% DMSO to the volume of the original culture sample taken, followed by boiling. The "total" glycosides level and the normalized intracellular level then were measured using LC-MS. Using wild type UGT74G1 and wild type UGT85C2, approximately 13.5 μM rubusoside was produced in total with a maximum normalized intracellular concentration of about 7 μM. In contrast, when wild type UGT74G1 and codon-optimized UGT85C2 were used, a maximum of 26 μM rubusoside was produced, or approximately double of what was produced using the

wild type UGT85C2. Additionally, the maximum normalized intracellular concentration of rubusoside was 13 μM, again an approximate doubling of what was produced using wild type UGT85C2. Intracellular concentration of 19-SMG was significantly reduced from a maximum of 35 μM using the wild type UGT85C2 to 19 μM using the codon-optimized UGT85C2. Consequently, about 10 μM less total 19-SMG was measured for the codon-optimized UGT85C2. This shows that more 19-SMG is converted into rubusoside and confirms that the wild type UGT85C2 is a bottleneck.

During diversity screening, another homolog of UGT85C2 was discovered during *Stevia rebaudiana* cDNA cloning. The homolog has the following combination of conserved amino acid polymorphisms (with respect to the amino acid numbering of the wild-type *S. rebaudiana* UGT85C coding sequence set forth in Accession No. AY345978.1): A65S, E71Q, T270M, Q289H, and A389V. This clone, termed UGT85C2 D37, was expressed through coupled *in vitro* transcription-translation of PCR products (TNT®T7 Quick for PCR DNA kit, Promega). The expression product was assayed for glycosylation activity using steviol (0.5 mM) as the sugar acceptor, as described in WO/2011/153378 with the exception that assays were allowed to incubate for 24 hours. As compared to the wildtype UGT85C2 control assay, the D37 enzyme appears to have approximately 30% higher glycosylation activity.

Example 11 - Identification of a novel *S. rebaudiana* KAH

A partial sequence (GenBank Accession No. BG521726) was identified in the *Stevia rebaudiana* EST data base that had some homology to a *Stevia* KAH. The partial sequence was blasted against raw *Stevia rebaudiana* pyrosequencing reads using CLC main workbench software. Reads that partially overlapped with the ends of the partial sequence were identified and used to increase the length of the partial sequence. This was done several times until the sequence encompassed both the start- and the stop codons. The complete sequence was analyzed for frameshift mutations and nucleotide substitutions that may have been introduced by blasting the complete sequence against the raw pyrosequencing reads. The resulting sequence was designated SrKAHe1. See FIG. 12.

Activity of the KAH encoded by SrKAHe1 was assessed *in vivo* in *S. cerevisiae* background strain CEN.PK 111-61A, which expresses genes encoding enzymes constituting the entire biosynthetic pathway from the yeast secondary metabolites isopentenyl pyrophosphate (IPP) and farnesyl pyrophosphate (FPP) to steviol-19-O-monoside, except the steviol synthase enzyme that converts *ent*-kaurenoic acid to steviol.

Briefly, the *S. cerevisiae* strain CEN.PK 111-61A was modified to express an *Aspergillus nidulans* GGPPS, a 150 nt truncated *Zea mays* CDPS (with a new start codon, see below), a *S. rebaudiana* KS, a *S. rebaudiana* KO and the *S. rebaudiana* UGT74G1 from chromosomally integrated gene copies, with TPI1 and GPD1 yeast promoters driving transcription. The CEN.PK 111-61A yeast strain that expresses all of these genes was designated EFSC2386. Thus, strain EFSC2386 contained the following integrated genes: *Aspergillus nidulans* Geranyl geranyl pyrophosphate synthase (GGPPS); *Zea mays* ent-Copalyl diphosphate synthase (CDPS); *Stevia rebaudiana ent*-Kaurene synthase (KS); *Stevia rebaudiana ent*-kaurene oxidase (KO); and *Stevia rebaudiana* UGT74G1; in combination with the pathway from IPP and FPP to steviol-19-O-monoside, without a steviol synthase (KAH).

Expression of different steviol synthases (from episomal expression plasmids) was tested in strain EFSC2386 in combination with the expression of various CPRs (from episomal expression plasmids), and production of steviol-19-O-monoside was detected by LC-MS analysis of culture sample extracts. The nucleic acids encoding the CPRs were inserted in the multi cloning site of the p426 GPD basic plasmid while the nucleic acids encoding the steviol synthases were inserted in the multi cloning site the p415 TEF basic plasmid (p4XX basic plasmid series by Mumberg *et al.*, *Gene* 156 (1995), 119-122). Production of steviol-19-O-monoside occurs when a functional steviol synthase enzyme is present.

The KAHs that were expressed from episomal expression plasmids in strain EFSC2386 were "indKAH" (Kumar *et al*, Accession no. DQ398871; Reeja *et al.*, Accession No. EU722415); "KAH1" (*S. rebaudiana* steviol synthase from Brandle *et al.*, U.S. Patent Publication No. 2008/0064063 A1); "KAH3" (*A. thaliana* steviol synthase from Yamaguchi *et al.*, U.S. Patent Publication No. 2008/0271205 A1); "SrKAHe1" (*S.*

rebaudiana steviol synthase cloned from *S. rebaudiana* cDNA as described above); and "DNA2.0.SrKAHe1" (codon optimized sequence (DNA2.0) encoding *S. rebaudiana* steviol synthase, see FIG. 12B).

The CPRs that were expressed from episomal expression plasmids in strain EFSC2386 were "CPR1" (*S. rebaudiana* NADPH dependent cytochrome P450 reductase (Kumar *et al.*, Accession no. DQ269454); "ATR1" (*A. thaliana* CPR, Accession No. CAA23011, see also FIG. 13); "ATR2" (*A. thaliana* CPR, Accession No. CAA46815, see also FIG. 13); "CPR7" (*S. rebaudiana* CPR, see FIG. 13, CPR7 is similar to "CPR1"); "CPR8" (*S. rebaudiana* CPR, similar to *Artemisia annua* CPR, see FIG. 13); and "CPR4" (*S. cerevisiae* NCP1 (Accession No. YHR042W, see also FIG. 13).

Table 13 provides the levels of steviol-19-O-monoside (μM) in strain EFSC2386 with the various combination of steviol synthases and CPRs.

Table 13

19-SMG Production

Strain	19-SMG production (μM)
"indKAH" CPR1	0.000
"indKAH" ATR1	0.000
"indKAH" ATR2	0.000
"indKAH" CPR7	0.000
"indKAH" CPR8	0.000
"indKAH" CPR4	0.000
"KAH1" CPR1	0.000
"KAH1" ATR1	0.000
"KAH1" ATR2	0.000
"KAH1" CPR7	0.000
"KAH1" CPR8	0.000
"KAH1" CPR4	0.000
"KAH3" CPR1	5.300
"KAH3" ATR1	5.921
"KAH3" ATR2	0.000
"KAH3" CPR7	5.693
"KAH3" CPR8	0.000

"KAH3" CPR4	0.000
"SrKAHe1" CPR1	20.129
"SrKAHe1" ATR1	15.613
"SrKAHe1" ATR2	40.407
"SrKAHe1" CPR7	33.724
"SrKAHe1" CPR8	41.695
"SrKAHe1" CPR4	28.949
"DNA2.0.SrKAHe1" CPR1	26.065
"DNA2.0.SrKAHe1" ATR1	26.974
"DNA2.0.SrKAHe1" ATR2	54.354
"DNA2.0.SrKAHe1" CPR7	30.797
"DNA2.0.SrKAHe1" CPR8	50.956
"DNA2.0.SrKAHe1" CPR4	30.368

Only KAH3 and the steviol synthase encoded by SrKAHe1 had activity when expressed in *S. cerevisiae*. The DNA 2.0. codon optimized SrKAHe1 sequence encoding steviol synthase resulted in a level of steviol-19-O-monoside accumulation that was approximately one order of magnitude higher as compared with a codon optimized KAH3 when each were co-expressed with optimal CPRs. In the experiments presented in this example, the combination of KAH1 and ATR2 CPR did not result in the production of steviol-19-O-monoside.

Example 12 - Pairings of CPRs and KO

The CEN.PK *S. Cerevisiae* EFSC2386 strain and the CPRs referred to in this Example are described in the Example 11 ("Identification of *S. rebaudiana* KAH"). EFSC2386 contained the following integrated genes: *Aspergillus nidulans* Geranyl geranyl pyrophosphate synthase (GGPPS); *Zea mays* ent-copalyl diphosphate synthase (CDPS); *Stevia rebaudiana* ent-kaurene synthase (KS); and *Stevia rebaudiana* ent-kaurene oxidase (KO). This strain produces *ent*-kaurenoic acid that was detected by LC-MS analysis.

A collection of cytochrome P450 reductases (CPRs) were expressed and tested in strain EFSC2386; "CPR1" (*S. rebaudiana* NADPH dependent cytochrome P450 reductase, Kumar *et al.*, Accession no. DQ269454); "ATR1" (*A. thaliana* CPR, Accession

No. CAA23011); "ATR2" (*A. thaliana* CPR, Accession No. CAA46815); "CPR7" (*S. rebaudiana* CPR, CPR7 is similar to "CPR1"); "CPR8" (*S. rebaudiana* CPR, similar to *Artemisia annua* CPR; and "CPR4" (*S. cerevisiae* NCP1, Accession No. YHR042W).

Overexpression of the *S. cerevisiae* endogenous native CPR (referred to as CPR4 in Table 14), and especially overexpression of one of the *A. thaliana* CPRs namely ATR2, gives good activation of the *Stevia rebaudiana* kaurene oxidase (the latter called KO1 in Table 14) and results in increased accumulation of *ent*-kaurenoic acid. See Table 14, which presents the area under curve (AUC) of the ent-kaurenoic acid peak in the LC-MS chromatograms. KO1 is an *ent*-kaurenoic acid producing yeast control strain without additional overexpression of CPRs.

Table 14 - Effect of Different Cytochrome P450 Reductase Enzymes with KO-1

Cytochrome P450 Reductase	Ent-Kaurenoic Acid (AUC)
CPR-1	14113
ATR-1	13558
ATR-2	29412
CPR-7	18918
CPR-8	12590
CPR-4	25103
Control	16593

Example 13 - Evaluating KS-5 and KS-1 in steviol pathways

The yeast strain EFSC1972 is a CEN.PK 111-61A *S. cerevisiae* strain that has the biosynthetic pathway from IPP/FPP to rubusoside expressed by integrated gene copies encoding the *Aspergillus nidulans* GGPPS (internal name GGPPS-10), the *Stevia rebaudiana* KS (KS1, SEQ ID NO:133), the *Arabidopsis thaliana* KAH (KAH-3, SEQ ID NO:144), the *Stevia rebaudiana* KO (KO1, SEQ ID NO:138), the *Stevia rebaudiana* CPR (CPR-1, SEQ ID NO:147), the full length *Zea mays* CDPS (CDPS-5, SEQ ID NO:158), the *Stevia rebaudiana* UGT74G1 (SEQ ID NO:1) and *Stevia rebaudiana* UGT85C2 (SEQ ID NO:3). Furthermore EFSC1972 has down regulation of the ERG9 gene expression by displacement of the endogenous promoter with the cupper inducible promoter CUP1.

When EFSC1972 is transformed with a CEN/ARS-based plasmid that expresses the *Stevia rebaudiana* SrKAHe1 from a TEF1 promoter, and simultaneously transformed with 2μ-based plasmids that express the *Synechococcus sp* GGPPS (GGPPS-7) and a truncated version of the *Zea mays* CDPS (truncated CDPS-5) from a GPD promoter, the result is growth-impaired *S. cerevisiae* producer of rubusoside (and 19-SMG). This strain is referred to as the "enhanced EFSC1972" in the following text. To determine whether the slow growth rate is caused by accumulation of the toxic pathway intermediate ent-copalyl diphosphate, a collection of kaurene synthase (KS) genes was expressed in the "enhanced EFSC1972" strain then growth and steviol glycoside production was assessed.

Expression of the *A. thaliana* KS (KS5) results in improved growth and steviol glycoside production of the "enhanced EFSC1972" strain. See FIG. 16. The same positive effect on growth cannot be achieved by further overexpression of the *Stevia rebaudiana* kaurene synthase (KS-1) in the enhanced EFSC1972 (data not shown).

Example 14 -- Yeast Strain EFSC1859

Saccharomyces cerevisiae strain EFSC1859 contains GGPPS-10, CDPS-5, KS-1, KO-1, KAH-3, CPR-1 and UGT74G1 coding sequences integrated into the genome and expressed from the strong constitutive GPD1 and TPI promoters. See Table 15. In addition, the endogenous promoter for the yeast ERG9 gene was replaced with the copper inducible promoter CUP1 for downregulation of the ERG9 squalene synthase. In standard yeast growth medium, the *ERG9* gene is transcribed at very low levels, since the concentration of copper in such medium is low. The decrease in ergosterol production in this strain results in increased amounts of isoprene units available for isoprenoid biosynthesis. In addition, strain EFSC1859 also expresses UGT85C2 from a 2 micron multicopy vector using a GPD1 promoter. EFSC1859 produces rubusoside and steviol 19-O-glycoside.

Zea mays CDPS DNA, with and without the chloroplast signal peptide, was expressed from a 2 micron multicopy plasmid using the GPD promoter. The nucleotide sequence and amino acid sequence of the *Zea mays* CDPS are forth in FIG. 14. The

chloroplast signal peptide is encoded by nucleotides 1-150 and corresponds to residues 1 to 50 of the amino acid sequence.

Table 15

Gene Source	Enzyme	Designation	gi Number	Accession No.
Aspergillus nidulans	GGPP synthase	GGPPS-10 (C301)	29468175	AF479566
Zea mays	CDP synthase	CDPS-5 (EV65)	50082774	AY562490
Stevia rebaudiana	Kaurene synthase	KS-1	4959241	AAD34295
Stevia rebaudiana	KO	KO-1	76446107	ABA42921
Arabidopsis thaliana	KAH	KAH-3	15238644	NP_197872
Stevia rebaudiana	UGT74G1			
Stevia rebaudiana	UGT85C2			
Stevia rebaudiana	CPR	CPR-1	93211213	ABB88839

EFSC1859 + maize full-length CDPS plasmid, and EFSC + maize truncated CDPS plasmid were grown in selective yeast medium with 4% glucose. Rubusoside and 19-SMG production were measured by LC-MS to estimate the production level. The removal of the plastid leader sequence did not appear to increase steviol glycoside production as compared to the wild-type sequence, and demonstrates that the CDPS transit peptide can be removed without causing a loss of steviol glycoside biosynthesis.

Example 15 -- Yeast Strain EFSC1923

Saccharomyces cerevisiae strain CEN.PK 111-61A was modified to produce steviol glycosides by introduction of steviol glycoside pathway enzymes from various organisms. The modified strain was designated EFSC1923.

Strain EFSC1923 contains an *Aspergillus nidulans* GGPP synthase gene expression cassette in the *S. cerevisiae PRP5-YBR238C* intergenic region, a *Zea mays* full-length CDPS and *Stevia rebaudiana* CPR gene expression cassette in the *MPT5-*

YGL176C intergenic region, a *Stevia rebaudiana* kaurene synthase and CDPS-1 gene expression cassette in the *ECM3-YOR093C* intergenic region, an *Arabidopsis thaliana* KAH and *Stevia rebaudiana* KO gene expression cassette in the *KIN1-INO2* intergenic region, a *Stevia rebaudiana* UGT74G1 gene expression cassette in the MGA1-YGR250C intergenic region and a *Stevia rebaudiana* UGT85C2 gene expression cassette integrated by displacing the *TRP1* gene ORF. See Table 15. In addition, the endogenous promoter for the yeast ERG9 gene was replaced with the copper inducible promoter CUP1.

Strain EFSC1923 produced approximately 5μM of the steviol glycoside, steviol 19-*O*-monoside, on selective yeast medium with 4% glucose.

Example 16 – Expression of a Truncated Maize CDPS in Yeast Strain EFSC1923

The 150 nucleotides at the 5' end of the *Zea mays* CDP synthase coding sequence in Table 15 (SEQ ID NO:157, see FIG. 14) was deleted, the remainder of the coding sequence was provided with a new translation start ATG, and the truncated sequence was operably linked to the GPD1 promoter in the multicopy plasmid p423GPD in *Saccharomyces cerevisiae* EFSC1923. Plasmid p423GPD is described in Mumberg, D et al, Gene, **156** : 119-122 (1995). EFSC1923 and EFSC1923 plus p423GPD-Z.m.tCDPS were grown in for 96 hours in selective yeast medium containing 4% glucose. The amount of steviol 19-*O*-monoside produced by EFSC1923 + p423GPD-Z.m.tCDPS (the truncated *Zea mays* CDPS) under these conditions was approximately 2.5 fold more than that produced by EFSC1923 without the plasmid.

The *Arabidopsis thaliana* KAH coding sequence from Table 15 was inserted in a multicopy plasmid designated p426GPD, under the control of the GPD1 promoter. Plasmid p426GPD is described in Mumberg, D et al, Gene, **156** : 119-122 (1995). No significant difference was observed between the amount of steviol 19-*O*-monoside produced by EFSC1923 + p426GPD-A.t.KAH, and EFSC1923 lacking the plasmid.

EFSC1923 was transformed with both p423GPD-Z.m.tCDPS and p426 p426GPD-A.t.KAH. Surprisingly, the amount of steviol 19-*O*-monoside produced under these conditions by EFSC1923 harboring both plasmids (i.e., the truncated *Zea mays*

CDPS and *Arabidopsis* KAH) was more than 6 fold greater than the amount produced by EFSC1923 alone.

A bifunctional CDPS-KS from *Gibberella fujikuroi* (NCBI Accession no: Q9UVY5.1, FIG. 15) was cloned and compared to the truncated CDPS-5. The bifunctional *Gibberella* CDPS-KS was cloned into a 2μ plasmid with a GPD promoter and transformed with a plasmid expressing the *Arabidopsis thaliana* KAH-3 from a 2 μ based-plasmid from a GPD promoter into EFSC1923. In shake flask studies, this bi-functional CDPS-KS was about 5.8 times more active in producing steviol 19-O-monoside than strain EFSC1923 with the KAH-3 alone. However, it was found to be less optimal than the KAH-3 and truncated CDPS combination under the conditions tested. Therefore, further strains were constructed with KS-5 and truncated CDPS.

Example 17 - Toxicity of Intermediates

The effect on *S. cerevisiae* vitality of geranyl geranyl pyrophosphate (GGPP), *ent*-copalyl diphosphate (CDP) or *ent*-kaurene production was investigated by expression of *Synechococcus sp* GGPPS alone (GGPP production), the GGPPS and the 50 amino acid N-terminally truncated *Zea mays* CDPS (see Example 16) together (CDP production), or the GGPPS, truncated CDPS and the *Arabidopsis thaliana* kaurene synthase (KS5) together (*ent* kaurene production) in the laboratory *S. cerevisiae* strain CEN.PK background. Genes were expressed from 2μ plasmids with GPD promoters driving transcription of truncated CDPS and KS5, while transcription of the GGPPS was driven by the ADH1 promoter. The growth of *S. cerevisiae* CEN.PK transformed with various combinations of these plasmids (GGPPS alone; GGPPS + truncated CDPS; or GGPPS + truncated CDPS + KS5) or plasmids without gene insertions was observed. GGPP production, and especially CDP production, was toxic to *S. cerevisiae* when produced as end products. Interestingly, *ent*-kaurene appeared to not be toxic to yeast in the amounts produced in this experiment.

Example 18 - Disruption of endogenous phosphatase activity

The yeast genes *DPP1* and *LPP1* encode phosphatases that can degrade FPP and GGPP to farnesol and geranylgeraniol, respectively. The gene-encoding *DPP1* was deleted in strain EFSC1923 (described in Example 15) to determine if there was an effect on steviol glycoside production. When this *dpp1* mutant strain was further transformed with a plasmid expressing the *Z. mays* CDPS lacking the chloroplast transit sequence (Example 16), both small and large transformants emerged. Strains of the "large colony" type produced ~40% more 19-SMG as compared to "small colony" type and the non-*DPP1* deleted strain, under the conditions tested. These results indicate that deletion of *DPP1* can have a positive effect on steviol glycoside production and that the degradation of prenyl pyrophosphates in yeast therefore could influence steviol glycoside production negatively.

Example 19 - Construction of a genetically stable yeast reporter strain producing vanillin glucoside from glucose with a disrupted SUC2 gene

A yeast strain producing vanillin glucoside from glucose was created basically as described in Brochado et al. ((2010) *Microbial Cell Factories* 9:84-98) (strain VG4), but with additional integration into the ECM3 inter-locus region in the yeast genome of an expression cassette with *E. coli* EntD PPTase controlled by the yeast TPI1 promoter (as described in Hansen et al. (2009) *Appl. Environ. Microbiol.* 75(9):2765–2774), disruption of SUC2 by replacing coding sequence with a MET15 expression cassette, and disruption of LEU2 by replacing coding sequence with a Tn5ble expression cassette conferring resistance to phleomycin. The resulting yeast strain was called V28. This strain also encodes a recombinant *A. thaliana* UDP-glycosyltransferase (UGT72E2, GenBank Accession No. Q9LVR1) having the amino acid sequence set forth in FIG. 19 (SEQ ID NO:178).

Example 20 - Expression of sucrose transporter and sucrose synthase in yeast already biosynthesizing vanillin glucoside

A sucrose transporter SUC1 from *Arabidopsis thaliana* was isolated by PCR amplification from cDNA prepared from *A. thaliana*, using proof-reading PCR

polymerase. The resulting PCR fragment was transferred by restriction digestion with *Spe*I and *Eco*RI and inserted into the corresponding in the low copy number yeast expression vector p416-TEF (a CEN-ARS based vector), from which the gene can be expressed from the strong TEF promoter. The resulting plasmid was named pVAN192. The sequence of the encoded sucrose transporter is set forth in FIG. 19B (GenBank Accession No. AEE35247, SEQ ID NO:179).

A sucrose synthase SUS1 from *Coffea arabica* (Accession No. CAJ32596) from was isolated by PCR amplification from cDNA prepared from *C. arabica*, using proof-reading PCR polymerase. The PCR fragment was transferred by restriction digestion with *Spe*I and *Sal*I and inserted into the corresponding position in the high copy number yeast expression vector p425-GPD (a 2μm based vector), from which the gene can be expressed from the strong GPD promoter. The resulting plasmid was named pMUS55. The sequence of the encoded sucrose synthase is set forth in FIG. 19C (GenBank Accession No. CAJ32596; SEQ ID NO:180).

pVAN192 and pMUS55 were introduced into the yeast strain V28 by genetic transformation, using a lithium acetate transformation protocol, creating the yeast strain V28::pVAN192::pMUS55. A control strain was made by transforming V28 with the empty plasmids P146-TEF and P425-GPD.

These two yeast strains were grown in 200 ml cultures in 500 ml Erlenmeyer shake flasks using SC (synthetic complete) growth medium without aromatic amino acids supplemented with 2% glucose and 2% sucrose and adjusted to pH 5.0. Cultures were incubated at moderate revolution (150 rpm), at 30°C for 72 hours. Samples were taken at 72 hours, and the content of vanillin glucoside determined. As can be seen from the table below, VG production in the control strain (containing empty plasmids p416-TEF and p425-GPD) was 330 mg/L VG, while the yeast strain V28::pVAN192::pMUS55 expressing sucrose synthase and sucrose transporter produced 445 mg/l VG, corresponding to a 34.8% increase in VG production.

Strain	Vanillin glucoside (g/L after 72 h)
V28 (p416-TEF + P425-GPD)	330
V28::pVAN192::pMUS55	445

This indicates that co-expression of a sucrose synthase and a sucrose transporter together with a glucosyltransferase increased the ability to glycosylate a small molecule aglycon, and concentration of the glycosylated aglycon was significantly increased. In this case, a significant improvement in vanillin glucosylation was achieved, resulting in a significant increase in titer of the end product, vanillin-O-β-glucoside.

Example 21 - Improved Steviol Glycoside Producing Strains

Strain construction of Saccharomyces cerevisiae EFSC2763

EFSC2763 yeast strain is derived from a wild type *Saccharomyces cerevisiae* strain containing three auxotrophic modifications, namely the deletions of *URA3, LEU2* and *HIS3*. The genetics of the strain have been stabilized and can be used as a regular diploid or haploid yeast strain. EFSC2763 has been converted to a steviol glycoside producing yeast by genomic-integration of four DNA constructs. Each construct contains multiple genes that were introduced into the yeast genome by homologous recombination. Furthermore, construct one and two were assembled by homologous recombination.

The first construct contains eight genes and is inserted in the *DPP1* locus and disrupts and partially deletes *DPP1* (see Example 18). The DNA inserted contains: the *A. gossypii* TEF promoter expressing the NatMX gene (selectable marker) followed by the TEF terminator from *A. gossypii;* Gene Art codon optimized *S. rebaudiana UGT85C2* (see Example 10) expressed from the native yeast GPD1 promoter and followed by the native yeast CYC1 terminator; *S. rebaudiana CPR-8* (see FIG. 13) expressed using the TPI1 promoter followed by the native yeast TDH1 terminator*; A. thaliana* Kaurene synthase (KS-5, see Example 13, SEQ ID NO:156) expressed from the PDC1 promoter and followed by the native yeast FBA1 terminator*; Synechococcus sp.* GGPPS (GGPPS-7) expressed using the TEF2 promoter and followed by the native yeast PFI1 terminator*;*

DNA2.0 codon-optimized *S. rebaudiana* KAHe1 (see Example 11, SEQ ID NO:165), expressed from the TEF1 promoter and followed by the ENO2 terminator; *S. rebaudiana* KO-1 expressed using the FBA1 promoter and followed by the native yeast TDH2 terminator; and *Zea mays* truncated CDPS (see Example 14) expressed using the PGK1 promoter and followed by the native yeast ADH2 terminator.

The second construct was inserted at the *YPRCΔ15* locus and contains the native yeast TEF promoter from *A. gossypii* in front expressing the KanMX gene (selectable marker) followed by the TEF terminator from *A. gossypii,* the Gene Art codon optimized *A. thaliana ATR2* (see FIG. 13B) expressed from the PGK1 promoter followed by the yeast ADH2 terminator, *S. rebaudiana* UGT74G1 expressed from the TPI1 promoter followed by the yeast TDH1 terminator, Gene Art codon-optimized *S. rebaudiana* UGT76G1 expressed from the TEF1 promoter followed by the yeast ENO2 terminator, and GeneArt codon-optimized *S. rebaudiana* UGT91D2e-b (see Example 6) expressed from the GPD1 promoter and followed by the yeast CYC1 terminator.

The first and the second construct were combined in the same spore clone by mating and dissection. This yeast strain was subsequently transformed with construct three and four in two successive events.

Construct three was integrated between genes *PRP5* and *YBR238C* and contained the TEF promoter from *A. gossypii* in expressing the *K. lactis LEU2* gene followed by the TEF terminator from *A. gossypii,* the GPD1 promoter expressing the DNA2.0-optimized *S. rebaudiana* KAHe1 followed by the CYC1 terminator, and the TPI1 promoter expressing the *Zea mays* truncated CDPS. Construct four was integrated in the genome between genes *ECM3* and *YOR093C* with an expression cassette containing the TEF promoter from *A. gossypii* expressing the *K. pneumoniae* hph gene followed by the TEF terminator from *A. gossypii, Synechococcus sp.* GGPPS expressed from the GPD1 promoter followed by the CYC1 terminator, and the TPI1 promoter expressing the *A. thaliana* Kaurene synthase. The four utilized genetic markers were subsequently removed.

As analyzed by LC-MS following the DMSO-extraction of total steviol glycosides from cells and broth, EFSC2772 produces between 40-50 μM or 2-3

μM/OD600 Rebaudioside A, after growth for four days in 3 ml SC (Synthetic Complete) media at 30°C with 320 RPM shaking in deep-well plates.

Strain construction of Saccharomyces cerevisiae EFSC2772

EFSC2772 is very similar to strain 2763 with the exception that the genetic markers were not removed, and the strain was made prototrophic by introduction of the two plasmids p413TEF (public domain CEN/ARS shuttle plasmid with *HIS3* marker) and p416-TEF (public domain CEN/ARS shuttle plasmid with *URA3* marker) by transformation, and designated *EFSC2772*.

As analyzed by LC-MS following the DMSO-extraction of total steviol glycosides from cells and broth, EFSC2772 produces similar levels of Rebaudioside A as 2763, after growth in deep-well plates. Higher optical densities and higher titers were obtained through aerobic fed-batch growth in 2L (working volume) fermentors which included a ~16 hour growth phase in the base medium (Synthetic Complete media) followed by ~100 hours of feeding with glucose utilized as the carbon and energy source combined with trace metals, vitamins, salts, and Yeast Nitrogen Base (YNB) and/or amino acid supplementation. The pH was kept near pH 5 and the temperature setpoint was 30ºC. As evidenced by LC-MS, combined cellular and extracellular product concentrations were between 920-1660 mg/L of Reb-A and approximately 300-320 mg/L of Reb-D in the two different experiments, approximately 700 mg/L of Reb-A was detected in the broth when the higher titer results were obtained. Additionally a large peak was seen for Reb-B, and one skilled in the art will recognize that additional copies of UGT74G1 or upregulation of UGT74G1 will further increase the conversion of RebB to RebA.

Strain EFSC2743 was made in a similar manner as above, but without the two plasmids conferring prototrophy and with the addition of a p416 (CEN/ARS)-based plasmid expressing EUGT11 from the TEF promoter. This strain was grown in a fed-batch fermentation as above. This strain produced a total amount of RebD of 920 mg/L and furthermore approximately a 9:1 ratio of RebD to RebA was seen. Approximately 360 mg/L of RebD was found in the broth.

Example 22 - UDP-glucose capacity

In Example 21, it was shown that yeast can fully glycosylate over 1 mM steviol e.g., to RebD, RebB, and RebA. Similarly, *Saccharomyces* strains are able to glycosylate as much as 60 mM of other small molecule products (data not shown). However, the glycosylation limit of the yeast native UDP-glucose regenerating system is unknown, or the rate at which it replenishes the UDP-glucose pool needed for cell wall synthesis. Therefore, experiments were designed to investigate if an increase in UDP-glucose production would increase the glycosylation rate in yeast. A *suc2* deletion mutant was transformed with plasmids harboring the *A. thaliana suc1* gene encoding a sucrose transporter, UGT74G1 and *A. thaliana* SUS. UGT74G1 can rapidly glycosylate steviol to steviol 19-O-monoglucoside (19-SMG). Transformants were pre-grown overnight in 13-ml culture tubes containing 4 ml of SC medium lacking leucine, histidine and uracil. The next day, cells corresponding to 2 OD_{600} units were spun down and resuspended in 2 ml of fresh media was containing 2% sucrose and or 100 μM steviol. Cultures were shaken at 30°C for 3 days in culture tubes. After 1h, 3h, 6h, 21h and 46h, aliquots were taken. Aliquots of 100 μl of culture were spun down and an equal volume of DMSO was added. Samples were vortexed, heated at 80°C for 15 minutes, centrifuged, and the 19-SMG content analyzed by LC-MS. No difference in the rate of glycosylation of steviol was observed between wild-type and SUS1-augmented strains at the time points tested. This suggests that glycosylation of steviol by UGT74G1 proceeds at a slower rate than UDP-glucose is regenerated by the yeast and that extra UDP-glucose may not be needed to achieve high titers of small molecule glycosylation *in vivo*. Nevertheless, the use of a SUS to recycle UDP-glucose *in vitro* is shown in Example 8 and therefore its use in an *in vivo* system is expected to increase the rate of production of steviol glycosides, if UDP-glucose should become limiting.

Example 23 - Reb-C and Reb-F production *in vivo* from glucose

Production of RebC from steviol

Previous experiments (Publication No. WO/2011/153378) have shown that recombinantly expressed *Arabidopsis thaliana* RHM2 (rhamnose synthetase, locus tag AT1G53500) is able to convert UDP-glucose to UDP-rhamnose. This UDP-rhamnose can be used to produce steviol-13-*O*-glucopyranosyl-1,2-rhamnoside, when incubated with UGT91D2e and steviol-13-*O*-monoglucoside *in vitro*.

Further experiments were conducted to confirm production of RebC from steviol by expressing all 4 UGTs and the RHM2 in yeast *in vivo*, followed by steviol feeding. EFSC301 strain (MAT alpha, *lys2*ADE8 h*is3ura3leu2trp1*) was transformed with the following plasmids expressing wild type gene sequences: p424GPD expressing wild type UGT74G1 (Accession no: AY345982); p423GPD expressing wild type 85C2 (Accession no.: AY345978.1); and a p426GPD derived-plasmid expressing wildtype UGT76G1 (Accession no: AY345974) and UGT91D2e under GPD promoters. Plasmid p425GPD expressing either RHM2 or an empty p425GPD control plasmid was cotransformed with the UGTs. Transformants were pre-grown overnight in 13 mL culture tubes containing 2-3 ml of SC medium lacking leucine, histidine, tryptophan and uracil. The next day, after growth had reached 0.4 OD_{600} units, cells were spun down, resuspended in fresh medium containing 25 μM steviol and shaken at 30°C for 3 days in culture tubes. An aliquot of 100 μL of culture was spun down. An equal volume of DMSO was added to the supernatant of this sample while 200 μL of 50% DMSO was added to the pellet. Samples were vortexed, heated at 80°C for 15 minutes, centrifuged, and the steviol glycoside content analyzed by LC-MS. RebC was detected in growth media and cellular extracts only when the RHM2 gene was coexpressed with the UGTs. Quantification showed that approximately equal amounts of RebA and RebC were produced. This shows that RHM2 is able to produce significant quantities of UDP-rhamnose *in vivo* and that UGT91D2e is capable of efficient rhamnosylation *in vivo*. Two other compounds were observed via LC-MS with retention times of 5.64 and 5.33 minutes and m/z ratios corresponding to steviol with 1 glucose- and 1 rhamnose (steviol-1,2 rhamnobioside), and 2 glucoses- and 1 rhamnose (Dulcoside A), respectively. This suggests that the remaining UGTs in the steviol glycoside pathway are capable of accepting rhamnosylated intermediates, i.e, the rhamnosylation step does not need to occur last.

In addition, a series of sequential *in vitro* experiments were conducted to determine whether any dead-end reactions occur in the rebaudioside C pathway. See FIG.2B. For example, the rhamnosylation activity of UGT91D2e on rubusoside and subsequent conversion of the product to RebC by UGT76G1 was demonstrated using *in vitro* reactions. In this experiment, UGT91D2e and RHM2 recombinantly expressed in *E. coli* and purified were incubated overnight with rubusoside, NADPH, NAD^+ and UDP-glucose. The reaction mixture was subsequently boiled to denature the enzymes. An aliquot of the reaction was added to an enzyme preparation of UGT76G1 with UDP-glucose. The rubusoside was converted in the presence of UGT91D2e and RHM2 to a compound with m/z corresponding to steviol with 2-glucoses and 1-rhamnose. Subsequently, this compound was converted in the presence of UGT76G1 to RebC, which indicates that the intermediate is Dulcoside A. This experiment therefore demonstrates that UGT91D2e is able to rhamnosylate rubusoside and that UGT76G1 is able to convert the product to RebC.

Similarly, it was shown through *in vitro* reactions that rhamnosylation of 13-SMG by UGT91D2e (forming a steviol compound with one glucose and one rhamnose) and subsequent formation of a compound with 2 glucoses and 1 rhamnose by UGT76G1. This compound has a unique retention time (4.56 min) and is thought to be steviol 13-*O*-1,3-diglycoside-1,2-rhamnoside. This compound also was observed when steviol was fed to yeast expressing the four UGTs and RHM2.

From the current data, it is shown that UGT91D2e is able to rhamnosylate 13-SMG and rubusoside. It is also shown that UGT74G1 and UGT76G1 are able to metabolize the rhamnosylated compound produced by UGT91D2e from 13-SMG. When these compounds are incubated with the remaining UGT (UGT74G1 or UGT76G1 depending on which UGT was used for the previous step), RebC is formed. This indicates that the order of glycosylation is of little importance as UGT74G1 and UGT76G1 are able to glycosylate rhamnosylated substrates.

Production of RebC from glucose

Plasmids expressing RHM2 and UGTs 76G1 and 91D2e were transformed into a stable rubusoside producer, the EFSC1923 strain (see Example 15). This yeast is a *Saccharomyces cerevisiae* CEN.PK 111-61A derivative with the UGTs 85C2 (Accession no.: AY345978.1) and 74G1 (Accession no: AY345982) integrated into the genome as well as auxotrophic modifications. In strain EFSC1923 (see Example 15), expression of squalene synthase, which is encoded by ERG9, was downregulated by displacement of the endogenous promoter with the CUP1 copper-inducible promoter. Strain EFSC1923 also contains an *Aspergillus nidula*ns GGPP synthase (GGPPS-10) expression cassette in the *S. cerevisiae* PRP5-YBR238C intergenic region, a *Zea mays* full-length CDPS (CDPS-5) and *Stevia rebaudiana* CPR (CPR-1) gene expression cassette in the *MPT5-YGL176C* intergenic region, a *Stevia rebaudiana* Kaurene synthase and CDPS (KS-1/CDPS-1) gene expression cassette in the *ECM3-YOR093C* intergenic region, an *Arabidopsis thaliana* KAH (KAH-3) and *Stevia rebaudiana* KO (KO-1) gene expression cassette in the *KIN1-INO2* intergenic region, a *Stevia rebaudiana* UGT74G1 gene expression cassette in the MGA1-YGR250C intergenic region and a *Stevia rebaudiana* UGT85C2 gene expression cassette integrated by displacing the *TRP1* gene ORF20. Inserted steviol pathway genes are described in Table 11 of published PCT WO/2011/153378.

EFSC1923 strain was transformed with a p423GPD-derived plasmid expressing wildtype UGT74G1 and UGT85C2 sequences using GPD promoters and a p426GPD-derived plasmid expressing wildtype UGT76G1 (Accession no: AY345974) and UGT91D2e (see SEQ ID NO:5) under the control of GPD promoters. Plasmid p425GPD expressing *Arabidopsis thaliana* RHM2 (enzyme locus tag AT1G53500) or an empty p425GPD control plasmid was co-transformed. Transformants were pre-grown overnight in 13-ml culture tubes containing 2-3 ml of SC medium lacking leucine, histidine and uracil. The next day, when the culture reached an OD_{600} of 0.4 units it was centrifuged, resuspended in fresh medium, and shaken at 30°C for 3 days in culture tubes. One hundred μL of culture were spun down; to this an equivalent volume of DMSO was added to the supernatant while 200 μL of 50% DMSO was added to the pellet. Samples

were vortexed, heated at 80°C for 15 minutes, spun down and the steviol glycoside content analyzed by LC-MS.

Analyses of the medium and normalized intracellular content of this strain showed production of RebC. Approximately 8 μM RebC and 4 μM RebA was produced as determined by LC-MS. Furthermore, the intermediates produced following steviol feeding were not detected in this experiment. Accumulation of RebC was strictly dependent on expression of RHM2. This example demonstrates *de novo* biosynthesis of RebC from glucose.

Production of additional Steviol glycosides from steviol and glucose

Using the same GPD-based plasmids described above, the stable steviol-producing strain EFSC1923 containing UGT74G1 and UGT85C2 was transformed with the UGTs required to produce RebB (UGT76G1 and UGT91D2e/EUGT11), RebE (UGT91D2e/EUGT11) and dulcoside A (RHM2, UGT91D2e/EUGT11). Wildtype EUGT11 (NCBI: NP_001051007), which was found to have higher diglycosylation activity, was cloned into p424GPD for this experiment. Transformants were pre-grown overnight in 13-ml culture tubes containing 2-3 ml of SC medium lacking leucine, histidine, tryptophan and uracil. The next day, after growth had reached 0.4 OD_{600} units, cells were spun down, resuspended in fresh medium containing 25 μM steviol (except for glucose experiments) and shaken at 30°C for 3 days in culture tubes. An aliquot of 100 μL of culture was spun down. An equal volume of DMSO was added to the supernatant of this sample while 200 μL of 50% DMSO was added to the pellet. Samples were vortexed, heated at 80°C for 15 minutes, centrifuged, and the steviol glycoside content analyzed by LC-MS. LC-MS analyses confirmed *in vivo* production of RebB, RebE, and Dulcoside A in *S. cerevisiae* from glucose or steviol. See, e.g., FIGs. 2A and 2B. A higher concentration of steviol-glycosides was observed following steviol-feeding (as judged by chromatograms).

Characterization of RebF pathway intermediates using EUGT11.

The xylosylating properties of UGT91D2e and EUGT11 were compared *in vitro*. By using UDP-xylose as the sugar-donor, UGT91D2e was previously shown to xylosylate steviol-13-O-monoglucoside forming a key intermediate in RebF biosynthesis

(Publication No. WO/2011/153378). Similar *in vitro* experiments using EUGT11 and UGT91D2e have shown that these UGTs are capable of xylosylating rubusoside. When UGT91D2e is used, the LC-MS analysis shows a new peak with an m/z ratio corresponding to steviol with 2 glucose molecules and 1 xylose. See, FIG. 26. Because of the shift in the retention time this peak is thought to correspond to rubusoside xylosylated on the 13-O-glucose. When EUGT11 is used, the LC-MS analysis shows two new, similar sized peaks at retention time 3.99 and 4.39 minutes with m/z ratios corresponding to steviol with 2 glucoses and 1 xylose. These products most likely correspond to rubusoside xylosylated on either of the two positions--the 13-O-glucose or the 19-O-glucose.

Production of RebF from glucose

In vivo production of RebF requires cloning of UGD1 (UDP-glucose dehydrogenase) and USX3 (UDP-glucoronic acid decarboxylase) from *Arabidopsis* for production of UDP-xylose. UGD1 and UXS3 were inserted in a high copy (2μ) vector, derived from P425-GPD, containing two expression cassettes, and expressed from strong constitutive promoters (*TPI1* and *GPD1*, respectively). The plasmid was transformed into the RebA producer strain EFSC2763 (described in Example 21) and cultivated during 3 days in selection medium (SC -leu). The LC-MS results clearly show the appearance of a new peak at retention time 4.13 minutes with m/z ratios corresponding to steviol with 3 glucoses and 1 xylose and identified as RebF (based on a commercial RebF standard), as well as other new peaks with m/z ratios corresponding to steviol with 2 glucoses and 1 xylose (as above), indicating that UGT91D2e was capable of carrying out xylosylation *in vivo*. These peaks were not seen in the negative controls.

Example 24 - Effect of squalene synthase (ERG9) down regulation using a heterologous insert

In yeast such as *Saccharomyces cerevisiae*, the mevalonate pathway produces a number of isoprenoid phosphate intermediates in the biosynthetic pathway to squalene (See FIG. 20). The squalene synthase in yeast is *ERG9*. See GenBank Accession No.

P29704.2 for the *Saccharomyces cerevisiae* squalene synthase; P36596 for the *Schizosaccharomyces pombe* squalene synthase; Q9Y753 for the *Yarrowia lipolytica* squalene synthase; Q9HGZ6 for the *Candida glabrata* squalene synthase; Q752X9 for the *Ashbya gossypii* squalene synthase; O74165 for the *Cyberlindnera jadinii* squalene synthase; P78589 for the *Candida albicans* squalene synthase; P38604 for the *Saccharomyces cerevisiae* lanosterol synthase; P37268 for the *Homo sapiens* squalene synthase; P53798 for the *Mus musculus* squalene synthase; and Q02769 for the *Rattus norvegicus* squalene synthase. See FIG. 25 (SEQ ID NOs:192-202).

Introduction of stemloop structure in 5'UTR of ERG9 gene

The wild-type *ERG9* promoter region was replaced with the CYC1 promoter sequence and a 5'UTR sequence by homologous recombination. The 5'UTR region contains a sequence that can form a stemloop structure. See SEQ ID NOs. 181-183. SEQ ID NO:184 is another sequence that also can be used.

SEQ ID NO: 181 (heterologous insert 1): TGAATTCGTTAACGAATTC

SEQ ID NO: 182 (heterologous insert 2): TGAATTCGTTAACGAACTC

SEQ ID NO: 183 (heterologous insert 3): TGAATTCGTTAACGAAGTC

SEQ ID NO: 184 (heterologous insert 4): TGAATTCGTTAACGAAATT

Without being bound to a particular mechanism, the stemloop may partially block the 5'-3' directed ribosomal scanning for the AUG and reduce the translation of the transcript. Stemloops with different degree of basepairing were tested to find stemloops that reduced the *ERG9* transcript translation sufficiently to boost FPP levels without affecting the growth of the yeast strain.

DNA fragments encompassing an *ERG9* promoter upstream sequence (for homologous recombination), an expression cassette for the gene (NatR) that confers resistance to Nourseothricin, a CYC1 promoter (SEQ ID NO: 185, FIG. 21), a 5' UTR sequence with a stemloop structure, and an *ERG9* ORF sequence (for homologous recombination) were generated by PCR. DNA fragments that contained either the CYC1

promoter or the KEX2 promoter (SEQ ID NO: 186) but no stemloops were also generated as controls. The flanking ERG9 sequences for recombination as well as the stemloop structure were introduced via the PCR oligos. An overview of the construct for homologous recombinantion is shown in FIG. 22. The DNA fragments were transformed into an *S.cerevisiae* host strain that subsequently was selected on nourseothricin containing growth plates. Clones with successful exchange of the native ERG9 promoter with the CYC1 promoter and stemloop-containing 5'UTR sequence were identified. Overview and sequence of the stem-loop region is provided in FIG. 23. The sequence identified as 5% corresponds with the heterologous insert having SEQ ID NO:181; the sequence identified as 20% corresponds with the heterologous insert having SEQ ID NO:182; and the sequence identified as 50% corresponds with the heterologous insert having SEQ ID NO:183.

Assessment of FPP accumulation (boosting effect)

The Amorpha-4,11-diene Synthase (ADS) gene catalyzes the chemical reaction that turns one FPP molecule into Amorpha-4,11-diene in the plant *Artemisia annua*. The gene is functional and efficient in *S.cerevisiae* and can be used to indirectly assess the accumulation of FPP in the strains with the stemloop structure introduced in the heterologous 5'UTR of the ERG9 gene. An *S.cerevisiae* codon optimized nucleic acid encoding ADS (GenBank Accession No. AAF61439) was cloned on a multicopy plasmid (2μ) under the control of the PGK1 promoter and transformed in the wild type and engineered *S. cerevisiae* strains. Amorpha-4,11-diene production was measured and compared to the standard compound caryophyllene, as described by (Ro *et al*. 2006. *Nature* 440(7086):940–943; Paradise *et al. Biotechnol Bioeng.* 2008 Jun 1;100(2):371-8; Newman *et al*. *Biotechnol Bioeng* 95(4):684–691).

Chemicals

Dodecane and caryophyllene were purchased from Sigma-Aldrich (St.Louis, MO). Complete Supplement Mixtures for formulation of Synthetic Complete (SC) media were purchased from Formedium (UK). All others chemical were purchased from Sigma-Aldrich.

Yeast cultivation

Engineered yeast strains were grown in SC 2% glucose with uracil dropped out. Cultures were grown at 30°C overnight and then used to inoculate main cultures in 250 mL shake flasks containing 25 mL SC medium, and grown to an optical density of 0.1 at 600nm. The main cultures were grown for 72h at 30°C. Because amorphadiene at very low concentrations is volatile from aqueous cultures, 2.5 mL dodecane was added to each culture flask in order to trap and retain the amorphadiene produced. 10 µl of the dodecane layer was sampled and diluted 100 fold in ethyl acetate for quantification by GC-MS

GC-MS analysis of amorphadiene

GC-MS was used to measure amorphadiene production from yeast cultures. Samples were analysed using the method as follow: The GC oven temperature program used 80°C for 2 min, followed by a ramping of 30°C/min to 160°C, then 3°C/min up to 170°C, and finally 30°C/min up to 300°C with a 2 min final hold. Injector and MS quadrupole detector temperatures were 250°C and 150°C, respectively. 1µL was injected in split less mode. The MS was operated in full scan mode. Amorphadiene concentration was calculated in (-)-tran-caryophyllene equivalents using a caryophyllene standard curve using the total ions.

The analysis of the different strains, including the different promoter constructs, showed an increased production of amorphadiene ((2.5 x) when using the heterologous insert having the nucleotide sequence set forth in SEQ ID NO: 181 compared to either no insert or the inserts having the nucleotide sequences set forth in SEQ ID NO:182 and 183. See FIG. 24. The heterologous insert set forth in SEQ ID NO:181 has the most stable secondary structure. For comparison the wild type yeast, with unmodified ERG9, was also analyzed (FIG. 24: CTRL-ADS) and this strain showed even lower production of amorphadiene. Conversely, the construct that comprised the very weak promoter ScKex2 showed an even higher level of amorphadiene (6 x).

Example 25 - Analysis of the effect of Squalene Synthase (ERG9) down regulation and GGPPS overexpression on GGPP production.

Assessment of GGPP accumulation

S.cerevisiae contains a GGPPS (BTS1). In addition to BTS1 there are several heterologous GGPPS enzymes that are functional and efficient in *S. cerevisiae*. When a functional GGPPS is overexpressed in *S.cerevisiae*, it leads to accumulation of GGPP, which may be converted to geranylgeraniol (GGOH) by the *S. cerevisiae* enzymes DPP1 and LPP1. The GGOH is partly exported to the yeast culture medium. GGOH can be measured by GC-MS and its accumulation can indirectly be used to assess the potential pool of GGPP that is available for enzymes that use GGPP as substrate.

Four different GGPPSs (GGPPS-1 (*S.acidicaldarius*, see Table 7), GGPPS-2 (*A. nidulans*, FIG. 25, SEQ ID NO:203), GGPPS-3 (S.cerevisiae, BTS1, FIG. 25, SEQ ID NO:167), and GGPPS-4 (*M. musculus,* see Table 7)) were assessed. The nucleotide sequences encoding GGPPS-1, GGPPS-2, and GGPPS-4 were *S. cerevisiae* codon optimized. All nucleic acids encoding the GGPPS polypeptides were cloned on a multitcopy plasmid (2μ) under the control of the PGK1 promoter and transformed in two different ERG9 down regulated strains: KEX2-ERG9 and CYC1(5%)-ERG9 (see Example 24).

Engineered yeast strains were grown in SC 2% glucose with uracil dropped out. Complete Supplement Mixtures for formulation of Synthetic Complete (SC) media were purchased from Formedium (UK). All others chemical were purchased from Sigma-Aldrich (St.Louis, MO). All optical density measurements were done at OD 600nm. Cultures were grown at 30°C overnight and then used to inoculate 250 ml unbaffled culture flasks containing 25ml SC medium at an OD600 of 0.1. The main cultures were grown for 72h at 30°C.

To measure GGOH accumulation, yeast cells (pellet) and yeast culture medium (supernatant) were extracted separately and then combined before analysis by GC-MS. The supernatant was extracted with Hexane in a 1:1 ratio. The pellet was first subjected to a saponification in solution containing 20% KOH and 50% Ethanol and the lysed cells were finally extracted with Hexane in a 1:1 ratio. The GC oven temperature program used was 80°C for 2 min, followed by a ramp to 160°C at 30°C/min, then to 170°C at 3°C/min and finally to 300°C at 30°C/min with a 2 min hold. Injector and MS quadrupole

detector temperatures were 250°C and 150°C, respectively. 2ul was injected in split less mode. The MS was operated in full scan mode.

When the GGPPS were overexpressed in the CYC1(5%)-ERG9 strain or KEX2-ERG9 strain, there was a significant increase in GGOH (GGPP) production observed with all four GGPPS polypeptides compared to the control where no GGPPS was expressed. Notably, the CYC1(5%)-ERG9 strain showed a 2-4 fold higher GGOH (GGPP) accumulation than the KEX2-ERG9 strain. The results are shown in FIG. 26.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

Claims

1. A method for increasing a level of one or more UDP-sugars comprising bioconversion of UDP using one or more polypeptides and one or more substrates, wherein the bioconversion occurs by contacting UDP with the one or more polypeptides in the presence of the one or more substrates, wherein the bioconversion comprises *in vitro* enzymatic or whole cell bioconversion.

2. The method of claim 1, wherein the one or more UDP-sugars comprises UDP-glucose, UDP-rhamnose, or UDP-xylose.

3. The method of claim 1, wherein the one or more substrates comprises sucrose, NADPH, and NAD^+.

4. The method of claim 1, wherein the one or more polypeptides is a sucrose synthase polypeptide, a rhamnose synthase (RHM2) polypeptide, a UDP-glucose dehydrogenase (UGD1) polypeptide, or a UDP-glucuronic acid decarboxylase (UXS3) polypeptide.

5. The method of claim 1, further comprising using the sucrose synthase polypeptide to regenerate UDP-glucose from UDP and sucrose.

6. The method of claim 5, wherein the sucrose synthase polypeptide is encoded by a *Coffea Arabica*, *Arabidopsis thaliana*, or *Stevia rebaudiana* SUS1 gene or a functional homolog thereof.

7. The method of claim 5, wherein the sucrose synthase polypeptide is encoded by a nucleotide sequence having at least 90% identity to the nucleotide sequence set forth in any one of SEQ ID NOs:175-177 or has an amino acid sequence of SEQ ID NO:180.

8. The method of claim 1, further comprising using the rhamnose synthase (RHM2) polypeptide to generate UDP-rhamnose from UDP-glucose and NADPH.

9. The method of claim 8, wherein the RHM2 polypeptide comprises an *Arabidopsis thaliana* RHM2 polypeptide.

10. The method of claim 1, further comprising using the UGD1 polypeptide and the UXS3 polypeptide to generate UDP-xylose from UDP-glucose and NAD^+.

11. The method of claim 10, wherein the UGD1 polypeptide comprises an *Arabidopsis thaliana* UGD1 polypeptide or a *Cryptococcus neoformans* UGD1 polypeptide, and the UXS3 polypeptide comprises an *Arabidopsis thaliana* UXS3 polypeptide or a *Cryptococcus neoformans* UXS3 polypeptide.

12. The method of claim 1, wherein the whole cell bioconversion comprises feeding whole cell raw materials comprising UDP and the one or more substrates.

13. The method of claim 12, wherein the cell has a reduced ability to degrade external sucrose.

14. The method of claim 12, wherein the cell comprises one or more recombinant genes encoding a sucrose synthase polypeptide, a rhamnose synthase (RHM2) polypeptide, a UDP-glucose dehydrogenase (UGD1) polypeptide, or a UDP-glucuronic decarboxylase (UXS3) polypeptide.

15. The method of claim 14, wherein the cell further comprises a recombinant gene encoding a sucrose transporter polypeptide.

16. The method of claim 15, wherein the sucrose transporter polypeptide comprises a sucrose transporter polypeptide having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:179.

17. The method of claim 12, wherein the cell further comprises an endogenous sucrose invertase, wherein the endogenous sucrose invertase is disrupted.

18. The method of claim 12, further comprising feeding raw materials during or after cell growth.

19. The method of claim 12, wherein the cell is in suspension or immobilized.

20. The method of claim 12, wherein cell is entrapped in a bead.

21. The method of claim 20, wherein the bead is a calcium or sodium alginate bead.

22. The method of claim 12, wherein the cell is linked to a hollow fiber tube reactor system.

23. The method of claim 12, wherein the cell is concentrated and entrapped within a membrane reactor system.

24. The method of claim 12, wherein the cell is in fermentation broth or in a reaction buffer.

25. The method of claim 12, further comprising using a permeabilizing agent.

26. The method of claim 25, wherein the permeabilizing agent is a solvent, a detergent, a surfactant or a mechanical shock.

27. The method of claim 26, wherein the permeabilizing agent is toluene, Triton-X, Tween cetyltrimethylammonium bromide (CTAB), electroporation, or an osmotic shock.

28. The method of claim 1, wherein the sucrose synthase is produced in a permeabilized recombinant host cell which has been transformed with a gene encoding a polypeptide having at least 80% identity to the amino acid sequence forth in any of SEQ ID NOs: 175-177 or has an amino acid sequence of SEQ ID NO:180.

29. The method of claim 4, wherein the rhamnose synthase (RHM2) polypeptide is produced in a permeabilized recombinant host cell which has been transformed with a gene encoding the *Arabidopsis thaliana* RHM2 polypeptide.

30. The method of claim 4, wherein the UDP-glucose dehydrogenase (UGD1) polypeptide is produced in a permeabilized recombinant host cell which has been transformed with a gene encoding an *Arabidopsis thaliana* UGD1 polypeptide or a *Cryptococcus neoformans* UGD1 polypeptide.

31. The method of claim 4, wherein the UDP-glucuronic decarboxylase (UXS3) polypeptide is produced in a permeabilized recombinant host cell which has been transformed with a gene encoding an *Arabidopsis thaliana* UXS3 polypeptide or a *Cryptococcus neoformans* UXS3 polypeptide.

32. The method of claim 12, wherein the cell is a yeast cell, a bacterial cell, a fungal cell, or a plant cell.

33. The method of claim 32, wherein the yeast cell is a *Saccharomycete*.

34. The method of claim 32, wherein the yeast cell is a cell from *Saccharomyces cerevisiae* species.

35. The method of claim 32, wherein the bacterial cell is from *Escherichia coli* species or *Bacillus* genus.

36. The method of claim 32, wherein the fungal cell is from *Aspergillus* genus.

SEQUENCE LISTING

<110> Evolva SA

<120> Recombinant Production of Steviol Glycosides

<130> 30998-0010WO1

<150> 61/521,084
<151> 2011-08-08

<150> 61/521,203
<151> 2011-08-08

<150> 61/521,051
<151> 2011-08-08

<150> 61/523,487
<151> 2011-08-15

<150> 61/567,929
<151> 2011-12-07

<150> 61/603,639
<151> 2012-02-27

<160> 203

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 460
<212> PRT
<213> Stevia rebaudiana

<400> 1

```
Met Ala Glu Gln Gln Lys Ile Lys Lys Ser Pro His Val Leu Leu Ile
1               5                  10                  15

Pro Phe Pro Leu Gln Gly His Ile Asn Pro Phe Ile Gln Phe Gly Lys
            20                  25                  30

Arg Leu Ile Ser Lys Gly Val Lys Thr Thr Leu Val Thr Thr Ile His
        35                  40                  45

Thr Leu Asn Ser Thr Leu Asn His Ser Asn Thr Thr Thr Thr Ser Ile
    50                  55                  60

Glu Ile Gln Ala Ile Ser Asp Gly Cys Asp Glu Gly Gly Phe Met Ser
65                  70                  75                  80

Ala Gly Glu Ser Tyr Leu Glu Thr Phe Lys Gln Val Gly Ser Lys Ser
                85                  90                  95
```

```
Leu Ala Asp Leu Ile Lys Lys Leu Gln Ser Glu Gly Thr Thr Ile Asp
            100                 105                 110

Ala Ile Ile Tyr Asp Ser Met Thr Glu Trp Val Leu Asp Val Ala Ile
        115                 120                 125

Glu Phe Gly Ile Asp Gly Gly Ser Phe Phe Thr Gln Ala Cys Val Val
    130                 135                 140

Asn Ser Leu Tyr Tyr His Val His Lys Gly Leu Ile Ser Leu Pro Leu
145                 150                 155                 160

Gly Glu Thr Val Ser Val Pro Gly Phe Pro Val Leu Gln Arg Trp Glu
                165                 170                 175

Thr Pro Leu Ile Leu Gln Asn His Glu Gln Ile Gln Ser Pro Trp Ser
            180                 185                 190

Gln Met Leu Phe Gly Gln Phe Ala Asn Ile Asp Gln Ala Arg Trp Val
        195                 200                 205

Phe Thr Asn Ser Phe Tyr Lys Leu Glu Glu Glu Val Ile Glu Trp Thr
    210                 215                 220

Arg Lys Ile Trp Asn Leu Lys Val Ile Gly Pro Thr Leu Pro Ser Met
225                 230                 235                 240

Tyr Leu Asp Lys Arg Leu Asp Asp Asp Lys Asp Asn Gly Phe Asn Leu
                245                 250                 255

Tyr Lys Ala Asn His His Glu Cys Met Asn Trp Leu Asp Asp Lys Pro
            260                 265                 270

Lys Glu Ser Val Val Tyr Val Ala Phe Gly Ser Leu Val Lys His Gly
        275                 280                 285

Pro Glu Gln Val Glu Glu Ile Thr Arg Ala Leu Ile Asp Ser Asp Val
    290                 295                 300

Asn Phe Leu Trp Val Ile Lys His Lys Glu Glu Gly Lys Leu Pro Glu
305                 310                 315                 320

Asn Leu Ser Glu Val Ile Lys Thr Gly Lys Gly Leu Ile Val Ala Trp
                325                 330                 335

Cys Lys Gln Leu Asp Val Leu Ala His Glu Ser Val Gly Cys Phe Val
            340                 345                 350
```

Thr His Cys Gly Phe Asn Ser Thr Leu Glu Ala Ile Ser Leu Gly Val
355 360 365

Pro Val Val Ala Met Pro Gln Phe Ser Asp Gln Thr Thr Asn Ala Lys
370 375 380

Leu Leu Asp Glu Ile Leu Gly Val Gly Val Arg Val Lys Ala Asp Glu
385 390 395 400

Asn Gly Ile Val Arg Arg Gly Asn Leu Ala Ser Cys Ile Lys Met Ile
405 410 415

Met Glu Glu Glu Arg Gly Val Ile Ile Arg Lys Asn Ala Val Lys Trp
420 425 430

Lys Asp Leu Ala Lys Val Ala Val His Glu Gly Gly Ser Ser Asp Asn
435 440 445

Asp Ile Val Glu Phe Val Ser Glu Leu Ile Lys Ala
450 455 460

<210> 2
<211> 1383
<212> DNA
<213> Stevia rebaudiana

<400> 2
atggcagagc aacaaaagat caaaaagtca cctcacgtct tacttattcc atttcctctg 60

caaggacata tcaacccatt catacaattt gggaaaagat tgattagtaa gggtgtaaag 120

acaacactgg taaccactat ccacactttg aattctactc tgaaccactc aaatactact 180

actacaagta tagaaattca agctatatca gacggatgcg atgagggtgg ctttatgtct 240

gccggtgaat cttacttgga aacattcaag caagtgggat ccaagtctct ggccgatcta 300

atcaaaaagt tacagagtga aggcaccaca attgacgcca taatctacga ttctatgaca 360

gagtgggttt tagacgttgc tatcgaattt ggtattgatg gaggttcctt tttcacacaa 420

gcatgtgttg tgaattctct atactaccat gtgcataaag ggttaatctc tttaccattg 480

ggtgaaactg tttcagttcc aggttttcca gtgttacaac gttgggaaac cccattgatc 540

ttacaaaatc atgaacaaat acaatcacct tggtcccaga tgttgtttgg tcaattcgct 600

aacatcgatc aagcaagatg ggtctttact aattcattct ataagttaga ggaagaggta 660

attgaatgga ctaggaagat ctggaatttg aaagtcattg gtccaacatt gccatcaatg 720

tatttggaca aaagacttga tgatgataaa gataatggtt tcaatttgta caaggctaat 780

catcacgaat gtatgaattg gctggatgac aaaccaaagg aatcagttgt atatgttgct 840

ttcggctctc ttgttaaaca tggtccagaa caagttgagg agattacaag agcacttata 900

gactctgacg taaacttttt gtgggtcatt aagcacaaag aggaggggaa actgccagaa 960

aacctttctg aagtgataaa gaccggaaaa ggtctaatcg ttgcttggtg taaacaattg 1020

gatgttttag ctcatgaatc tgtaggctgt tttgtaacac attgcggatt caactctaca 1080

ctagaagcca tttccttagg cgtacctgtc gttgcaatgc ctcagttctc cgatcagaca 1140

accaacgcta aacttttgga cgaaatacta ggggtgggtg tcagagttaa agcagacgag 1200

aatggtatcg tcagaagagg gaacctagct tcatgtatca aaatgatcat ggaagaggaa 1260

agaggagtta tcataaggaa aaacgcagtt aagtggaagg atcttgcaaa ggttgccgtc 1320

catgaaggcg gctcttcaga taatgatatt gttgaatttg tgtccgaact aatcaaagcc 1380

taa 1383

<210> 3
<211> 481
<212> PRT
<213> Stevia rebaudiana

<400> 3

Met Asp Ala Met Ala Thr Thr Glu Lys Lys Pro His Val Ile Phe Ile
1 5 10 15

Pro Phe Pro Ala Gln Ser His Ile Lys Ala Met Leu Lys Leu Ala Gln
20 25 30

Leu Leu His His Lys Gly Leu Gln Ile Thr Phe Val Asn Thr Asp Phe
35 40 45

Ile His Asn Gln Phe Leu Glu Ser Ser Gly Pro His Cys Leu Asp Gly
50 55 60

Ala Pro Gly Phe Arg Phe Glu Thr Ile Pro Asp Gly Val Ser His Ser
65 70 75 80

Pro Glu Ala Ser Ile Pro Ile Arg Glu Ser Leu Leu Arg Ser Ile Glu
85 90 95

Thr Asn Phe Leu Asp Arg Phe Ile Asp Leu Val Thr Lys Leu Pro Asp
100 105 110

Pro Pro Thr Cys Ile Ile Ser Asp Gly Phe Leu Ser Val Phe Thr Ile
115 120 125

Asp Ala Ala Lys Lys Leu Gly Ile Pro Val Met Met Tyr Trp Thr Leu
130 135 140

```
Ala Ala Cys Gly Phe Met Gly Phe Tyr His Ile His Ser Leu Ile Glu
145                 150                 155                 160

Lys Gly Phe Ala Pro Leu Lys Asp Ala Ser Tyr Leu Thr Asn Gly Tyr
                165                 170                 175

Leu Asp Thr Val Ile Asp Trp Val Pro Gly Met Glu Gly Ile Arg Leu
            180                 185                 190

Lys Asp Phe Pro Leu Asp Trp Ser Thr Asp Leu Asn Asp Lys Val Leu
        195                 200                 205

Met Phe Thr Thr Glu Ala Pro Gln Arg Ser His Lys Val Ser His His
    210                 215                 220

Ile Phe His Thr Phe Asp Glu Leu Glu Pro Ser Ile Ile Lys Thr Leu
225                 230                 235                 240

Ser Leu Arg Tyr Asn His Ile Tyr Thr Ile Gly Pro Leu Gln Leu Leu
                245                 250                 255

Leu Asp Gln Ile Pro Glu Glu Lys Lys Gln Thr Gly Ile Thr Ser Leu
            260                 265                 270

His Gly Tyr Ser Leu Val Lys Glu Glu Pro Glu Cys Phe Gln Trp Leu
        275                 280                 285

Gln Ser Lys Glu Pro Asn Ser Val Val Tyr Val Asn Phe Gly Ser Thr
    290                 295                 300

Thr Val Met Ser Leu Glu Asp Met Thr Glu Phe Gly Trp Gly Leu Ala
305                 310                 315                 320

Asn Ser Asn His Tyr Phe Leu Trp Ile Ile Arg Ser Asn Leu Val Ile
                325                 330                 335

Gly Glu Asn Ala Val Leu Pro Pro Glu Leu Glu Glu His Ile Lys Lys
            340                 345                 350

Arg Gly Phe Ile Ala Ser Trp Cys Ser Gln Glu Lys Val Leu Lys His
        355                 360                 365

Pro Ser Val Gly Gly Phe Leu Thr His Cys Gly Trp Gly Ser Thr Ile
    370                 375                 380

Glu Ser Leu Ser Ala Gly Val Pro Met Ile Cys Trp Pro Tyr Ser Trp
385                 390                 395                 400
```

Asp Gln Leu Thr Asn Cys Arg Tyr Ile Cys Lys Glu Trp Glu Val Gly
405 410 415

Leu Glu Met Gly Thr Lys Val Lys Arg Asp Glu Val Lys Arg Leu Val
420 425 430

Gln Glu Leu Met Gly Glu Gly Gly His Lys Met Arg Asn Lys Ala Lys
435 440 445

Asp Trp Lys Glu Lys Ala Arg Ile Ala Ile Ala Pro Asn Gly Ser Ser
450 455 460

Ser Leu Asn Ile Asp Lys Met Val Lys Glu Ile Thr Val Leu Ala Arg
465 470 475 480

Asn

<210> 4
<211> 1446
<212> DNA
<213> Stevia rebaudiana

<400> 4
atggatgcaa tggcaactac tgagaaaaag cctcatgtga tcttcattcc atttcctgca 60

caatctcaca taaaggcaat gctaaagtta gcacaactat tacaccataa gggattacag 120

ataactttcg tgaataccga cttcatccat aatcaatttc tggaatctag tggccctcat 180

tgtttggacg gagccccagg gtttagattc gaaacaattc ctgacggtgt ttcacattcc 240

ccagaggcct ccatcccaat aagagagagt ttactgaggt caatagaaac caacttttg 300

gatcgtttca ttgacttggt cacaaaactt ccagacccac caacttgcat aatctctgat 360

ggctttctgt cagtgtttac tatcgacgct gccaaaaagt tgggtatccc agttatgatg 420

tactggactc ttgctgcatg cggtttcatg ggtttctatc acatccattc tcttatcgaa 480

aagggttttg ctccactgaa agatgcatca tacttaacca acggctacct ggatactgtt 540

attgactggg taccaggtat ggaaggtata agacttaaag attttccttt ggattggtct 600

acagacctta atgataaagt attgatgttt actacagaag ctccacaaag atctcataag 660

gtttcacatc atatctttca cacctttgat gaattggaac catcaatcat caaaaccttg 720

tctctaagat acaatcatat ctacactatt ggtccattac aattacttct agatcaaatt 780

cctgaagaga aaaagcaaac tggtattaca tccttacacg gctactcttt agtgaaagag 840

gaaccagaat gttttcaatg gctacaaagt aaagagccta attctgtggt ctacgtcaac 900

ttcggaagta caacagtcat gtccttggaa gatatgactg aatttggttg gggccttgct 960

aattcaaatc attactttct atggattatc aggtccaatt tggtaatagg ggaaaacgcc 1020

gtattacctc cagaattgga ggaacacatc aaaaagagag gtttcattgc ttcctggtgt 1080

tctcaggaaa aggtattgaa acatccttct gttggtggtt tccttactca ttgcggttgg 1140

ggctctacaa tcgaatcact aagtgcagga gttccaatga tttgttggcc atattcatgg 1200

gaccaactta caaattgtag gtatatctgt aaagagtggg aagttggatt agaaatggga 1260

acaaaggtta aacgtgatga agtgaaaaga ttggttcagg agttgatggg ggaaggtggc 1320

cacaagatga gaaacaaggc caaagattgg aaggaaaaag ccagaattgc tattgctcct 1380

aacgggtcat cctctctaaa cattgataag atggtcaaag agattacagt cttagccaga 1440

aactaa 1446

<210> 5
<211> 473
<212> PRT
<213> Stevia rebaudiana

<400> 5

```
Met Ala Thr Ser Asp Ser Ile Val Asp Asp Arg Lys Gln Leu His Val
1               5                  10                  15

Ala Thr Phe Pro Trp Leu Ala Phe Gly His Ile Leu Pro Tyr Leu Gln
            20                  25                  30

Leu Ser Lys Leu Ile Ala Glu Lys Gly His Lys Val Ser Phe Leu Ser
        35                  40                  45

Thr Thr Arg Asn Ile Gln Arg Leu Ser Ser His Ile Ser Pro Leu Ile
    50                  55                  60

Asn Val Val Gln Leu Thr Leu Pro Arg Val Gln Glu Leu Pro Glu Asp
65                  70                  75                  80

Ala Glu Ala Thr Thr Asp Val His Pro Glu Asp Ile Pro Tyr Leu Lys
                85                  90                  95

Lys Ala Ser Asp Gly Leu Gln Pro Glu Val Thr Arg Phe Leu Glu Gln
            100                 105                 110

His Ser Pro Asp Trp Ile Ile Tyr Asp Tyr Thr His Tyr Trp Leu Pro
        115                 120                 125

Ser Ile Ala Ala Ser Leu Gly Ile Ser Arg Ala His Phe Ser Val Thr
    130                 135                 140

Thr Pro Trp Ala Ile Ala Tyr Met Gly Pro Ser Ala Asp Ala Met Ile
145                 150                 155                 160
```

```
Asn Gly Ser Asp Gly Arg Thr Thr Val Glu Asp Leu Thr Thr Pro Pro
                165                 170                 175

Lys Trp Phe Pro Phe Pro Thr Lys Val Cys Trp Arg Lys His Asp Leu
            180                 185                 190

Ala Arg Leu Val Pro Tyr Lys Ala Pro Gly Ile Ser Asp Gly Tyr Arg
        195                 200                 205

Met Gly Leu Val Leu Lys Gly Ser Asp Cys Leu Leu Ser Lys Cys Tyr
    210                 215                 220

His Glu Phe Gly Thr Gln Trp Leu Pro Leu Leu Glu Thr Leu His Gln
225                 230                 235                 240

Val Pro Val Val Pro Val Gly Leu Leu Pro Pro Glu Ile Pro Gly Asp
                245                 250                 255

Glu Lys Asp Glu Thr Trp Val Ser Ile Lys Lys Trp Leu Asp Gly Lys
            260                 265                 270

Gln Lys Gly Ser Val Val Tyr Val Ala Leu Gly Ser Glu Val Leu Val
        275                 280                 285

Ser Gln Thr Glu Val Val Glu Leu Ala Leu Gly Leu Glu Leu Ser Gly
    290                 295                 300

Leu Pro Phe Val Trp Ala Tyr Arg Lys Pro Lys Gly Pro Ala Lys Ser
305                 310                 315                 320

Asp Ser Val Glu Leu Pro Asp Gly Phe Val Glu Arg Thr Arg Asp Arg
                325                 330                 335

Gly Leu Val Trp Thr Ser Trp Ala Pro Gln Leu Arg Ile Leu Ser His
            340                 345                 350

Glu Ser Val Cys Gly Phe Leu Thr His Cys Gly Ser Gly Ser Ile Val
        355                 360                 365

Glu Gly Leu Met Phe Gly His Pro Leu Ile Met Leu Pro Ile Phe Gly
    370                 375                 380

Asp Gln Pro Leu Asn Ala Arg Leu Leu Glu Asp Lys Gln Val Gly Ile
385                 390                 395                 400

Glu Ile Pro Arg Asn Glu Glu Asp Gly Cys Leu Thr Lys Glu Ser Val
```

405 410 415

Ala Arg Ser Leu Arg Ser Val Val Val Glu Lys Glu Gly Glu Ile Tyr
420 425 430

Lys Ala Asn Ala Arg Glu Leu Ser Lys Ile Tyr Asn Asp Thr Lys Val
435 440 445

Glu Lys Glu Tyr Val Ser Gln Phe Val Asp Tyr Leu Glu Lys Asn Ala
450 455 460

Arg Ala Val Ala Ile Asp His Glu Ser
465 470

<210> 6
<211> 1422
<212> DNA
<213> Stevia rebaudiana

<400> 6
atggctacat ctgattctat tgttgatgac aggaagcagt tgcatgtggc tactttccct 60

tggcttgctt tcggtcatat actgccttac ctacaactat caaaactgat agctgaaaaa 120

ggacataaag tgtcattcct ttcaacaact agaaacattc aaagattatc ttcccacata 180

tcaccattga ttaacgtcgt tcaattgaca cttccaagag tacaggaatt accagaagat 240

gctgaagcta caacagatgt gcatcctgaa gatatccctt acttgaaaaa ggcatccgat 300

ggattacagc ctgaggtcac tagattcctt gagcaacaca gtccagattg gatcatatac 360

gactacactc actattggtt gccttcaatt gcagcatcac taggcatttc tagggcacat 420

ttcagtgtaa ccacaccttg ggccattgct tacatgggtc catccgctga tgctatgatt 480

aacggcagtg atggtagaac taccgttgaa gatttgacaa ccccaccaaa gtggtttcca 540

tttccaacta aagtctgttg gagaaaacac gacttagcaa gactggttcc atacaaggca 600

ccaggaatct cagacggcta tagaatgggt ttagtcctta aagggtctga ctgcctattg 660

tctaagtgtt accatgagtt tgggacacaa tggctaccac ttttggaaac attacaccaa 720

gttcctgtcg taccagttgg tctattacct ccagaaatcc ctggtgatga gaaggacgag 780

acttgggttt caatcaaaaa gtggttagac gggaagcaaa aaggctcagt ggtatatgtg 840

gcactgggtt ccgaagtttt agtatctcaa acagaagttg tggaacttgc cttaggtttg 900

gaactatctg gattgccatt tgtctgggcc tacagaaaac caaaaggccc tgcaaagtcc 960

gattcagttg aattgccaga cggctttgtc gagagaacta gagatagagg gttggtatgg 1020

acttcatggg ctccacaatt gagaatcctg agtcacgaat ctgtgtgcgg tttcctaaca 1080

cattgtggtt ctggttctat agttgaagga ctgatgtttg gtcatccact tatcatgttg 1140

ccaatctttg gtgaccagcc tttgaatgca cgtctgttag aagataaaca agttggaatt 1200

gaaatcccac gtaatgagga agatggatgt ttaaccaagg agtctgtggc cagatcatta 1260

cgttccgttg tcgttgaaaa ggaaggcgaa atctacaagg ccaatgcccg tgaactttca 1320

aagatctaca atgacacaaa agtagagaag gaatatgttt ctcaatttgt agattaccta 1380

gagaaaaacg ctagagccgt agctattgat catgaatcct aa 1422

<210> 7
<211> 458
<212> PRT
<213> Stevia rebaudiana

<400> 7

```
Met Glu Asn Lys Thr Glu Thr Thr Val Arg Arg Arg Arg Arg Ile Ile
1               5                  10                  15

Leu Phe Pro Val Pro Phe Gln Gly His Ile Asn Pro Ile Leu Gln Leu
            20                  25                  30

Ala Asn Val Leu Tyr Ser Lys Gly Phe Ser Ile Thr Ile Phe His Thr
        35                  40                  45

Asn Phe Asn Lys Pro Lys Thr Ser Asn Tyr Pro His Phe Thr Phe Arg
    50                  55                  60

Phe Ile Leu Asp Asn Asp Pro Gln Asp Glu Arg Ile Ser Asn Leu Pro
65                  70                  75                  80

Thr His Gly Pro Leu Ala Gly Met Arg Ile Pro Ile Ile Asn Glu His
                85                  90                  95

Gly Ala Asp Glu Leu Arg Arg Glu Leu Glu Leu Leu Met Leu Ala Ser
            100                 105                 110

Glu Glu Asp Glu Glu Val Ser Cys Leu Ile Thr Asp Ala Leu Trp Tyr
        115                 120                 125

Phe Ala Gln Ser Val Ala Asp Ser Leu Asn Leu Arg Arg Leu Val Leu
    130                 135                 140

Met Thr Ser Ser Leu Phe Asn Phe His Ala His Val Ser Leu Pro Gln
145                 150                 155                 160

Phe Asp Glu Leu Gly Tyr Leu Asp Pro Asp Asp Lys Thr Arg Leu Glu
                165                 170                 175

Glu Gln Ala Ser Gly Phe Pro Met Leu Lys Val Lys Asp Ile Lys Ser
            180                 185                 190
```

```
Ala Tyr Ser Asn Trp Gln Ile Leu Lys Glu Ile Leu Gly Lys Met Ile
        195                 200                 205

Lys Gln Thr Lys Ala Ser Ser Gly Val Ile Trp Asn Ser Phe Lys Glu
    210                 215                 220

Leu Glu Glu Ser Glu Leu Glu Thr Val Ile Arg Glu Ile Pro Ala Pro
225                 230                 235                 240

Ser Phe Leu Ile Pro Leu Pro Lys His Leu Thr Ala Ser Ser Ser Ser
                245                 250                 255

Leu Leu Asp His Asp Arg Thr Val Phe Gln Trp Leu Asp Gln Gln Pro
            260                 265                 270

Pro Ser Ser Val Leu Tyr Val Ser Phe Gly Ser Thr Ser Glu Val Asp
        275                 280                 285

Glu Lys Asp Phe Leu Glu Ile Ala Arg Gly Leu Val Asp Ser Lys Gln
    290                 295                 300

Ser Phe Leu Trp Val Val Arg Pro Gly Phe Val Lys Gly Ser Thr Trp
305                 310                 315                 320

Val Glu Pro Leu Pro Asp Gly Phe Leu Gly Glu Arg Gly Arg Ile Val
                325                 330                 335

Lys Trp Val Pro Gln Gln Glu Val Leu Ala His Gly Ala Ile Gly Ala
            340                 345                 350

Phe Trp Thr His Ser Gly Trp Asn Ser Thr Leu Glu Ser Val Cys Glu
        355                 360                 365

Gly Val Pro Met Ile Phe Ser Asp Phe Gly Leu Asp Gln Pro Leu Asn
    370                 375                 380

Ala Arg Tyr Met Ser Asp Val Leu Lys Val Gly Val Tyr Leu Glu Asn
385                 390                 395                 400

Gly Trp Glu Arg Gly Glu Ile Ala Asn Ala Ile Arg Arg Val Met Val
                405                 410                 415

Asp Glu Glu Gly Glu Tyr Ile Arg Gln Asn Ala Arg Val Leu Lys Gln
            420                 425                 430

Lys Ala Asp Val Ser Leu Met Lys Gly Gly Ser Ser Tyr Glu Ser Leu
```

435 440 445

Glu Ser Leu Val Ser Tyr Ile Ser Ser Leu
450 455

<210> 8
<211> 1377
<212> DNA
<213> Stevia rebaudiana

<400> 8
atggaaaaca agaccgaaac aacagttaga cgtaggcgta gaatcattct gtttccagta 60

cctttttcaag ggcacatcaa tccaatacta caactagcca acgttttgta ctctaaaggt 120

ttttctatta caatctttca caccaatttc aacaaaccaa aaacatccaa ttacccacat 180

ttcacattca gattcatact tgataatgat ccacaagatg aacgtatttc aaacttacct 240

acccacggtc ctttagctgg aatgagaatt ccaatcatca atgaacatgg tgccgatgag 300

cttagaagag aattagagtt acttatgttg gcatccgaag aggacgagga agtctcttgt 360

ctgattactg acgctctatg gtactttgcc caatctgtgg ctgatagttt gaatttgagg 420

agattggtac taatgacatc cagtctgttt aactttcacg ctcatgttag tttaccacaa 480

tttgacgaat tgggatactt ggaccctgat gacaagacta ggttagagga acaggcctct 540

ggttttccta tgttgaaagt caaagatatc aagtctgcct attctaattg gcaaatcttg 600

aaagagatct taggaaagat gatcaaacag acaaaggctt catctggagt gatttggaac 660

agtttcaaag agttagaaga gtctgaattg gagactgtaa tcagagaaat tccagcacct 720

tcattcctga taccattacc aaaacatttg actgcttcct cttcctcttt gttggatcat 780

gacagaacag tttttcaatg gttggaccaa caaccaccta gttctgtttt gtacgtgtca 840

tttggtagta cttctgaagt cgatgaaaag gacttccttg aaatcgcaag aggcttagtc 900

gatagtaagc agtcattcct ttgggtcgtg cgtccaggtt tcgtgaaagg ctcaacatgg 960

gtcgaaccac ttccagatgg ttttctaggc gaaagaggta gaatagtcaa atgggttcct 1020

caacaggaag ttttagctca tggcgctatt ggggcattct ggactcattc cggatggaat 1080

tcaactttag aatcagtatg cgaaggggta cctatgatct tttcagattt tggtcttgat 1140

caaccactga acgcaagata catgtctgat gttttgaaag tgggtgtata tctagaaaat 1200

ggctgggaaa ggggtgaaat agctaatgca ataagacgtg ttatggttga tgaagagggg 1260

gagtatatca gacaaaacgc aagagtgctg aagcaaaagg ccgacgtttc tctaatgaag 1320

ggaggctctt catacgaatc cttagaatct cttgtttcct acatttcatc actgtaa 1377

<210> 9
<211> 1422
<212> DNA

<213> Stevia rebaudiana

<400> 9
atggctacca gtgactccat agttgacgac cgtaagcagc ttcatgttgc gacgttccca 60

tggcttgctt tcggtcacat cctcccttac cttcagcttt cgaaattgat agctgaaaag 120

ggtcacaaag tctcgtttct ttctaccacc agaaacattc aacgtctctc ttctcatatc 180

tcgccactca taaatgttgt tcaactcaca cttccacgtg tccaagagct gccggaggat 240

gcagaggcga ccactgacgt ccaccctgaa gatattccat atctcaagaa ggcttctgat 300

ggtcttcaac cggaggtcac ccggtttcta gaacaacact ctccggactg gattatttat 360

gattatactc actactggtt gccatccatc gcggctagcc tcggtatctc acgagcccac 420

ttctccgtca ccactccatg ggccattgct tatatgggac cctcagctga cgccatgata 480

aatggttcag atggtcgaac cacggttgag gatctcacga caccgcccaa gtggtttccc 540

tttccgacca aagtatgctg gcggaagcat gatcttgccc gactggtgcc ttacaaagct 600

ccggggatat ctgatggata ccgtatgggg ctggttctta agggatctga ttgtttgctt 660

tccaaatgtt accatgagtt tggaactcaa tggctacctc ttttggagac actacaccaa 720

gtaccggtgg ttccggtggg attactgcca ccggaaatac ccggagacga gaaagatgaa 780

acatgggtgt caatcaagaa atggctcgat ggtaaacaaa aaggcagtgt ggtgtacgtt 840

gcattaggaa gcgaggtttt ggtgagccaa accgaggttg ttgagttagc attgggtctc 900

gagctttctg ggttgccatt tgtttgggct tatagaaaac caaaaggtcc cgcgaagtca 960

gactcggtgg agttgccaga cgggttcgtg gaacgaactc gtgaccgtgg gttggtctgg 1020

acgagttggg cacctcagtt acgaatactg agccatgagt cggtttgtgg tttcttgact 1080

cattgtggtt ctggatcaat tgtggaaggg ctaatgtttg gtcaccctct aatcatgcta 1140

ccgatttttg gggaccaacc tctgaatgct cgattactgg aggacaaaca ggtgggaatc 1200

gagataccaa gaaatgagga agatggttgc ttgaccaagg agtcggttgc tagatcactg 1260

aggtccgttg ttgtggaaaa agaaggggag atctacaagg cgaacgcgag ggagctgagt 1320

aaaatctata acgacactaa ggttgaaaaa gaatatgtaa gccaattcgt agactatttg 1380

gaaaagaatg cgcgtgcggt tgccatcgat catgagagtt aa 1422

<210> 10
<211> 473
<212> PRT
<213> Stevia rebaudiana

<400> 10
Met Ala Thr Ser Asp Ser Ile Val Asp Asp Arg Lys Gln Leu His Val
1 5 10 15

Ala Thr Phe Pro Trp Leu Ala Phe Gly His Ile Leu Pro Tyr Leu Gln

```
            20                  25                  30

Leu Ser Lys Leu Ile Ala Glu Lys Gly His Lys Val Ser Phe Leu Ser
        35                  40                  45

Thr Thr Arg Asn Ile Gln Arg Leu Ser Ser His Ile Ser Pro Leu Ile
    50                  55                  60

Asn Val Val Gln Leu Thr Leu Pro Arg Val Gln Glu Leu Pro Glu Asp
65                  70                  75                  80

Ala Glu Ala Thr Thr Asp Val His Pro Glu Asp Ile Pro Tyr Leu Lys
                85                  90                  95

Lys Ala Ser Asp Gly Leu Gln Pro Glu Val Thr Arg Phe Leu Glu Gln
            100                 105                 110

His Ser Pro Asp Trp Ile Ile Tyr Asp Tyr Thr His Tyr Trp Leu Pro
        115                 120                 125

Ser Ile Ala Ala Ser Leu Gly Ile Ser Arg Ala His Phe Ser Val Thr
    130                 135                 140

Thr Pro Trp Ala Ile Ala Tyr Met Gly Pro Ser Ala Asp Ala Met Ile
145                 150                 155                 160

Asn Gly Ser Asp Gly Arg Thr Thr Val Glu Asp Leu Thr Thr Pro Pro
                165                 170                 175

Lys Trp Phe Pro Phe Pro Thr Lys Val Cys Trp Arg Lys His Asp Leu
            180                 185                 190

Ala Arg Leu Val Pro Tyr Lys Ala Pro Gly Ile Ser Asp Gly Tyr Arg
        195                 200                 205

Met Gly Leu Val Leu Lys Gly Ser Asp Cys Leu Leu Ser Lys Cys Tyr
    210                 215                 220

His Glu Phe Gly Thr Gln Trp Leu Pro Leu Leu Glu Thr Leu His Gln
225                 230                 235                 240

Val Pro Val Val Pro Val Gly Leu Leu Pro Pro Glu Val Pro Gly Asp
                245                 250                 255

Glu Lys Asp Glu Thr Trp Val Ser Ile Lys Lys Trp Leu Asp Gly Lys
            260                 265                 270
```

Gln Lys Gly Ser Val Val Tyr Val Ala Leu Gly Ser Glu Val Leu Val
275 280 285

Ser Gln Thr Glu Val Val Glu Leu Ala Leu Gly Leu Glu Leu Ser Gly
290 295 300

Leu Pro Phe Val Trp Ala Tyr Arg Lys Pro Lys Gly Pro Ala Lys Ser
305 310 315 320

Asp Ser Val Glu Leu Pro Asp Gly Phe Val Glu Arg Thr Arg Asp Arg
325 330 335

Gly Leu Val Trp Thr Ser Trp Ala Pro Gln Leu Arg Ile Leu Ser His
340 345 350

Glu Ser Val Cys Gly Phe Leu Thr His Cys Gly Ser Gly Ser Ile Val
355 360 365

Glu Gly Leu Met Phe Gly His Pro Leu Ile Met Leu Pro Ile Phe Gly
370 375 380

Asp Gln Pro Leu Asn Ala Arg Leu Leu Glu Asp Lys Gln Val Gly Ile
385 390 395 400

Glu Ile Pro Arg Asn Glu Glu Asp Gly Cys Leu Thr Lys Glu Ser Val
405 410 415

Ala Arg Ser Leu Arg Ser Val Val Val Glu Lys Glu Gly Glu Ile Tyr
420 425 430

Lys Ala Asn Ala Arg Glu Leu Ser Lys Ile Tyr Asn Asp Thr Lys Val
435 440 445

Glu Lys Glu Tyr Val Ser Gln Phe Val Asp Tyr Leu Glu Lys Asn Thr
450 455 460

Arg Ala Val Ala Ile Asp His Glu Ser
465 470

<210> 11
<211> 1422
<212> DNA
<213> Stevia rebaudiana

<400> 11
atggctacca gtgactccat agttgacgac cgtaagcagc ttcatgttgc gacgttccca 60

tggcttgctt tcggtcacat cctcccttac cttcagcttt cgaaattgat agctgaaaag 120

ggtcacaaag tctcgtttct ttctaccacc agaaacattc aacgtctctc ttctcatatc 180

tcgccactca taaatgttgt tcaactcaca cttccacgtg tccaagagct gccggaggat 240

gcagaggcga ccactgacgt ccaccctgaa gatattccat atctcaagaa ggcttctgat 300

ggtcttcaac cggaggtcac ccggtttcta gaacaacact ctccggactg gattatttat 360

gattatactc actactggtt gccatccatc gcggctagcc tcggtatctc acgagcccac 420

ttctccgtca ccactccatg ggccattgct tatatgggac cctcagctga cgccatgata 480

aatggttcag atggtcgaac cacggttgag gatctcacga caccgcccaa gtggtttccc 540

tttccgacca aagtatgctg gcggaagcat gatcttgccc gactggtgcc ttacaaagct 600

ccggggatat ctgatggata ccgtatgggg ctggttctta agggatctga ttgtttgctt 660

tccaaatgtt accatgagtt tggaactcaa tggctacctc ttttggagac actacaccaa 720

gtaccggtgg ttccggtggg attactgcca ccggaagtac ccggagacga gaaagatgaa 780

acatgggtgt caatcaagaa atggctcgat ggtaaacaaa aaggcagtgt ggtgtacgtt 840

gcattaggaa gcgaggtttt ggtgagccaa accgaggttg ttgagttagc attgggtctc 900

gagctttctg ggttgccatt tgtttgggct tatagaaaac caaaaggtcc cgcgaagtca 960

gactcggtgg agttgccaga cgggttcgtg gaacgaactc gtgaccgtgg gttggtctgg 1020

acgagttggg cacctcagtt acgaatactg agccatgagt cggtttgtgg tttcttgact 1080

cattgtggtt ctggatcaat tgtggaaggg ctaatgtttg gtcaccctct aatcatgcta 1140

ccgatttttg gggaccaacc tctgaatgct cgattactgg aggacaaaca ggtgggaatc 1200

gagataccaa gaaatgagga agatggttgc ttgaccaagg agtcggttgc tagatcactg 1260

aggtccgttg ttgtggaaaa agaaggggag atctacaagg cgaacgcgag ggagctgagt 1320

aaaatctata acgacactaa ggttgaaaaa gaatatgtaa gccaattcgt agactatttg 1380

gaaaagaata cgcgtgcggt tgccatcgat catgagagtt aa 1422

<210> 12
<211> 473
<212> PRT
<213> Stevia rebaudiana

<400> 12
Met Ala Thr Ser Asp Ser Ile Val Asp Asp Arg Lys Gln Leu His Val
1 5 10 15

Ala Thr Phe Pro Trp Leu Ala Phe Gly His Ile Leu Pro Phe Leu Gln
20 25 30

Leu Ser Lys Leu Ile Ala Glu Lys Gly His Lys Val Ser Phe Leu Ser
35 40 45

Thr Thr Arg Asn Ile Gln Arg Leu Ser Ser His Ile Ser Pro Leu Ile

50 55 60

Asn Val Val Gln Leu Thr Leu Pro Arg Val Gln Glu Leu Pro Glu Asp
65 70 75 80

Ala Glu Ala Thr Thr Asp Val His Pro Glu Asp Ile Gln Tyr Leu Lys
85 90 95

Lys Ala Val Asp Gly Leu Gln Pro Glu Val Thr Arg Phe Leu Glu Gln
100 105 110

His Ser Pro Asp Trp Ile Ile Tyr Asp Phe Thr His Tyr Trp Leu Pro
115 120 125

Ser Ile Ala Ala Ser Leu Gly Ile Ser Arg Ala Tyr Phe Cys Val Ile
130 135 140

Thr Pro Trp Thr Ile Ala Tyr Leu Ala Pro Ser Ser Asp Ala Met Ile
145 150 155 160

Asn Asp Ser Asp Gly Arg Thr Thr Val Glu Asp Leu Thr Thr Pro Pro
165 170 175

Lys Trp Phe Pro Phe Pro Thr Lys Val Cys Trp Arg Lys His Asp Leu
180 185 190

Ala Arg Met Glu Pro Tyr Glu Ala Pro Gly Ile Ser Asp Gly Tyr Arg
195 200 205

Met Gly Met Val Phe Lys Gly Ser Asp Cys Leu Leu Phe Lys Cys Tyr
210 215 220

His Glu Phe Gly Thr Gln Trp Leu Pro Leu Leu Glu Thr Leu His Gln
225 230 235 240

Val Pro Val Val Pro Val Gly Leu Leu Pro Pro Glu Ile Pro Gly Asp
245 250 255

Glu Lys Asp Glu Thr Trp Val Ser Ile Lys Lys Trp Leu Asp Gly Lys
260 265 270

Gln Lys Gly Ser Val Val Tyr Val Ala Leu Gly Ser Glu Ala Leu Val
275 280 285

Ser Gln Thr Glu Val Val Glu Leu Ala Leu Gly Leu Glu Leu Ser Gly
290 295 300

Leu Pro Phe Val Trp Ala Tyr Arg Lys Pro Lys Gly Pro Ala Lys Ser
305 310 315 320

Asp Ser Val Glu Leu Pro Asp Gly Phe Val Glu Arg Thr Arg Asp Arg
325 330 335

Gly Leu Val Trp Thr Ser Trp Ala Pro Gln Leu Arg Ile Leu Ser His
340 345 350

Glu Ser Val Cys Gly Phe Leu Thr His Cys Gly Ser Gly Ser Ile Val
355 360 365

Glu Gly Leu Met Phe Gly His Pro Leu Ile Met Leu Pro Leu Phe Gly
370 375 380

Asp Gln Pro Leu Asn Ala Arg Leu Leu Glu Asp Lys Gln Val Gly Ile
385 390 395 400

Glu Ile Pro Arg Asn Glu Glu Asp Gly Cys Leu Thr Lys Glu Ser Val
405 410 415

Ala Arg Ser Leu Arg Ser Val Val Val Glu Asn Glu Gly Glu Ile Tyr
420 425 430

Lys Ala Asn Ala Arg Glu Leu Ser Lys Ile Tyr Asn Asp Thr Lys Val
435 440 445

Glu Lys Glu Tyr Val Ser Gln Phe Val Asp Tyr Leu Glu Lys Asn Ala
450 455 460

Arg Ala Val Ala Ile Asp His Glu Ser
465 470

<210> 13
<211> 1422
<212> DNA
<213> Stevia rebaudiana

<400> 13
atggctacca gtgactccat agttgacgac cgtaagcagc ttcatgttgc gacgttccca 60

tggcttgctt tcggtcacat cctccctttc cttcagcttt cgaaattgat agctgaaaag 120

ggtcacaaag tctcgtttct ttctaccacc agaaacattc aacgtctctc ttctcatatc 180

tcgccactca taaatgttgt tcaactcaca cttccacgtg tccaagagct gccggaggat 240

gcagaggcga ccactgacgt ccaccctgaa gatattcaat atctcaagaa ggctgttgat 300

ggtcttcaac cggaggtcac ccggtttcta gaacaacact ctccggactg gattatttat 360

gattttactc actactggtt gccatccatc gcggctagcc tcggtatctc acgagcctac 420

ttctgcgtca tcactccatg gaccattgct tatttggcac cctcatctga cgccatgata 480

aatgattcag atggtcgaac cacggttgag gatctcacga caccgcccaa gtggtttccc 540

tttccgacca aagtatgctg gcggaagcat gatcttgccc gaatggagcc ttacgaagct 600

ccagggatat ctgatggata ccgtatgggg atggttttta agggatctga ttgtttgctt 660

ttcaaatgtt accatgagtt tggaactcaa tggctacctc ttttggagac actacaccaa 720

gtaccggtgg ttccggtggg attactgccg ccggaaatac ccggagacga gaaagatgaa 780

acatgggtgt caatcaagaa atggctcgat ggtaaacaaa aaggcagtgt ggtgtacgtt 840

gcattaggaa gcgaggcttt ggtgagccaa accgaggttg ttgagttagc attgggtctc 900

gagctttctg ggttgccatt tgtttgggct tatagaaaac caaaaggtcc cgcgaagtca 960

gactcggtgg agttgccaga cgggttcgtg gaacgaactc gtgaccgtgg gttggtctgg 1020

acgagttggg cacctcagtt acgaatactg agccatgagt cggtttgtgg tttcttgact 1080

cattgtggtt ctggatcaat tgtggaaggg ctaatgtttg gtcaccctct aatcatgcta 1140

ccgctttttg gggaccaacc tctgaatgct cgattactgg aggacaaaca ggtgggaatc 1200

gagataccaa gaaatgagga agatggttgc ttgaccaagg agtcggttgc tagatcactg 1260

aggtccgttg ttgtggaaaa cgaaggggag atctacaagg cgaacgcgag ggagctgagt 1320

aaaatctata acgacactaa ggtggaaaaa gaatatgtaa gccaattcgt agactatttg 1380

gaaaagaatg cgcgtgcggt tgccatcgat catgagagtt aa 1422

<210> 14
<211> 485
<212> PRT
<213> Stevia rebaudiana

<400> 14

```
Met Tyr Asn Val Thr Tyr His Gln Asn Ser Lys Ala Met Ala Thr Ser
1               5                   10                  15

Asp Ser Ile Val Asp Asp Arg Lys Gln Leu His Val Ala Thr Phe Pro
            20                  25                  30

Trp Leu Ala Phe Gly His Ile Leu Pro Phe Leu Gln Leu Ser Lys Leu
        35                  40                  45

Ile Ala Glu Lys Gly His Lys Val Ser Phe Leu Ser Thr Thr Arg Asn
    50                  55                  60

Ile Gln Arg Leu Ser Ser His Ile Ser Pro Leu Ile Asn Val Val Gln
65                  70                  75                  80

Leu Thr Leu Pro Arg Val Gln Glu Leu Pro Glu Asp Ala Glu Ala Thr
```

```
                85                  90                  95

Thr Asp Val His Pro Glu Asp Ile Gln Tyr Leu Lys Lys Ala Val Asp
            100                 105                 110

Gly Leu Gln Pro Glu Val Thr Arg Phe Leu Glu Gln His Ser Pro Asp
        115                 120                 125

Trp Ile Ile Tyr Asp Phe Thr His Tyr Trp Leu Pro Ser Ile Ala Ala
    130                 135                 140

Ser Leu Gly Ile Ser Arg Ala Tyr Phe Cys Val Ile Thr Pro Trp Thr
145                 150                 155                 160

Ile Ala Tyr Leu Ala Pro Ser Ser Asp Ala Met Ile Asn Asp Ser Asp
                165                 170                 175

Gly Arg Thr Thr Val Glu Asp Leu Thr Thr Pro Pro Lys Trp Phe Pro
            180                 185                 190

Phe Pro Thr Lys Val Cys Trp Arg Lys His Asp Leu Ala Arg Met Glu
        195                 200                 205

Pro Tyr Glu Ala Pro Gly Ile Ser Asp Gly Tyr Arg Met Gly Met Val
    210                 215                 220

Phe Lys Gly Ser Asp Cys Leu Leu Phe Lys Cys Tyr His Glu Phe Gly
225                 230                 235                 240

Thr Gln Trp Leu Pro Leu Leu Glu Thr Leu His Gln Val Pro Val Val
                245                 250                 255

Pro Val Gly Leu Leu Pro Pro Glu Ile Pro Gly Asp Glu Lys Asp Glu
            260                 265                 270

Thr Trp Val Ser Ile Lys Lys Trp Leu Asp Gly Lys Gln Lys Gly Ser
        275                 280                 285

Val Val Tyr Val Ala Leu Gly Ser Glu Ala Leu Val Ser Gln Thr Glu
    290                 295                 300

Val Val Glu Leu Ala Leu Gly Leu Glu Leu Ser Gly Leu Pro Phe Val
305                 310                 315                 320

Trp Ala Tyr Arg Lys Pro Lys Gly Pro Ala Lys Ser Asp Ser Val Glu
                325                 330                 335
```

Leu Pro Asp Gly Phe Val Glu Arg Thr Arg Asp Arg Gly Leu Val Trp
340 345 350

Thr Ser Trp Ala Pro Gln Leu Arg Ile Leu Ser His Glu Ser Val Cys
355 360 365

Gly Phe Leu Thr His Cys Gly Ser Gly Ser Ile Val Glu Gly Leu Met
370 375 380

Phe Gly His Pro Leu Ile Met Leu Pro Ile Phe Cys Asp Gln Pro Leu
385 390 395 400

Asn Ala Arg Leu Leu Glu Asp Lys Gln Val Gly Ile Glu Ile Pro Arg
405 410 415

Asn Glu Glu Asp Gly Cys Leu Thr Lys Glu Ser Val Ala Arg Ser Leu
420 425 430

Arg Ser Val Val Val Glu Asn Glu Gly Glu Ile Tyr Lys Ala Asn Ala
435 440 445

Arg Ala Leu Ser Lys Ile Tyr Asn Asp Thr Lys Val Glu Lys Glu Tyr
450 455 460

Val Ser Gln Phe Val Asp Tyr Leu Glu Lys Asn Ala Arg Ala Val Ala
465 470 475 480

Ile Asp His Glu Ser
485

<210> 15
<211> 1458
<212> DNA
<213> Stevia rebaudiana

<400> 15
atgtacaacg ttacttatca tcaaaattca aaagcaatgg ctaccagtga ctccatagtt 60

gacgaccgta agcagcttca tgttgcgacg ttcccatggc ttgctttcgg tcacatcctc 120

cctttccttc agctttcgaa attgatagct gaaaagggtc acaaagtctc gtttctttct 180

accaccagaa acattcaacg tctctcttct catatctcgc cactcataaa tgttgttcaa 240

ctcacacttc cacgtgtcca agagctgccg gaggatgcag aggcgaccac tgacgtccac 300

cctgaagata ttcaatatct caagaaggct gttgatggtc ttcaaccgga ggtcacccgg 360

tttctagaac aacactctcc ggactggatt atttatgatt ttactcacta ctggttgcca 420

tccatcgcgg ctagcctcgg tatctcacga gcctacttct gcgtcatcac tccatggacc 480

attgcttatt tggcaccctc atctgacgcc atgataaatg attcagatgg tcgaaccacg 540

gttgaggatc tcacgacacc gcccaagtgg tttccctttc cgaccaaagt atgctggcgg 600

aagcatgatc ttgcccgaat ggagccttac gaagctccgg ggatatctga tggataccgt 660

atggggatgg tttttaaggg atctgattgt ttgcttttca aatgttacca tgagtttgga 720

actcaatggc tacctctttt ggagacacta caccaagtac cggtggttcc ggtgggatta 780

ctgccgccgg aaatacccgg agacgagaaa gatgaaacat gggtgtcaat caagaaatgg 840

ctcgatggta aacaaaaagg cagtgtggtg tacgttgcat taggaagcga ggctttggtg 900

agccaaaccg aggttgttga gttagcattg ggtctcgagc tttctgggtt gccatttgtt 960

tgggcttata gaaaaccaaa aggtcccgcg aagtcagact cggtggagtt gccagacggg 1020

ttcgtggaac gaactcgtga ccgtgggttg gtctggacga gttgggcacc tcagttacga 1080

atactgagcc acgagtcagt ttgtggtttc ttgactcatt gtggttctgg atcaattgtg 1140

gaagggctaa tgtttggtca ccctctaatc atgctaccga ttttttgtga ccaacctctg 1200

aatgctcgat tactggagga caaacaggtg ggaatcgaga taccaagaaa tgaggaagat 1260

ggttgcttga ccaaggagtc ggttgctaga tcactgaggt ccgttgttgt ggaaaacgaa 1320

ggggagatct acaaggcgaa cgcgagggcg ctgagtaaaa tctataacga cactaaggtg 1380

gaaaaagaat atgtaagcca attcgtagac tatttggaaa agaatgcgcg tgcggttgcc 1440

atcgatcatg agagttaa 1458

<210> 16
<211> 473
<212> PRT
<213> Stevia rebaudiana

<400> 16

Met Ala Thr Ser Asp Ser Ile Val Asp Asp Arg Lys Gln Leu His Val
1 5 10 15

Ala Thr Phe Pro Trp Leu Ala Phe Gly His Ile Leu Pro Phe Leu Gln
20 25 30

Leu Ser Lys Leu Ile Ala Glu Lys Gly His Lys Val Ser Phe Leu Ser
35 40 45

Thr Thr Arg Asn Ile Gln Arg Leu Ser Ser His Ile Ser Pro Leu Ile
50 55 60

Asn Val Val Gln Leu Thr Leu Pro Arg Val Gln Glu Leu Pro Glu Asp
65 70 75 80

Ala Glu Ala Thr Thr Asp Val His Pro Glu Asp Ile Gln Tyr Leu Lys
85 90 95

```
Lys Ala Val Asp Gly Leu Gln Pro Glu Val Thr Arg Phe Leu Glu Gln
            100                 105                 110

His Ser Pro Asp Trp Ile Ile Tyr Asp Phe Thr His Tyr Trp Leu Pro
        115                 120                 125

Ser Ile Ala Ala Ser Leu Gly Ile Ser Arg Ala Tyr Phe Cys Val Ile
    130                 135                 140

Thr Pro Trp Thr Ile Ala Tyr Leu Ala Pro Ser Ser Asp Ala Met Ile
145                 150                 155                 160

Asn Asp Ser Asp Gly Arg Thr Thr Val Glu Asp Leu Thr Thr Pro Pro
                165                 170                 175

Lys Trp Phe Pro Phe Pro Thr Lys Val Cys Trp Arg Lys His Asp Leu
            180                 185                 190

Ala Arg Met Glu Pro Tyr Glu Ala Pro Gly Ile Ser Asp Gly Tyr Arg
        195                 200                 205

Met Gly Met Val Phe Lys Gly Ser Asp Cys Leu Leu Phe Lys Cys Tyr
    210                 215                 220

His Glu Phe Gly Thr Gln Trp Leu Pro Leu Leu Glu Thr Leu His Gln
225                 230                 235                 240

Val Pro Val Val Pro Val Gly Leu Leu Pro Pro Glu Ile Pro Gly Asp
                245                 250                 255

Glu Lys Asp Glu Thr Trp Val Ser Ile Lys Lys Trp Leu Asp Gly Lys
            260                 265                 270

Gln Lys Gly Ser Val Val Tyr Val Ala Leu Gly Ser Glu Ala Leu Val
        275                 280                 285

Ser Gln Thr Glu Val Val Glu Leu Ala Leu Gly Leu Glu Leu Ser Gly
    290                 295                 300

Leu Pro Phe Val Trp Ala Tyr Arg Lys Pro Lys Gly Pro Ala Lys Ser
305                 310                 315                 320

Asp Ser Val Glu Leu Pro Asp Gly Phe Val Glu Arg Thr Arg Asp Arg
                325                 330                 335

Gly Leu Val Trp Thr Ser Trp Ala Pro Gln Leu Arg Ile Leu Ser His
            340                 345                 350
```

Glu Ser Val Cys Gly Phe Leu Thr His Cys Gly Ser Gly Ser Ile Val
355 360 365

Glu Gly Leu Met Phe Gly His Pro Leu Ile Met Leu Pro Ile Phe Cys
370 375 380

Asp Gln Pro Leu Asn Ala Arg Leu Leu Glu Asp Lys Gln Val Gly Ile
385 390 395 400

Glu Ile Pro Arg Asn Glu Glu Asp Gly Cys Leu Thr Lys Glu Ser Val
405 410 415

Ala Arg Ser Leu Arg Ser Val Val Val Glu Asn Glu Gly Glu Ile Tyr
420 425 430

Lys Ala Asn Ala Arg Ala Leu Ser Lys Ile Tyr Asn Asp Thr Lys Val
435 440 445

Glu Lys Glu Tyr Val Ser Gln Phe Val Asp Tyr Leu Glu Lys Asn Ala
450 455 460

Arg Ala Val Ala Ile Asp His Glu Ser
465 470

<210> 17
<211> 1422
<212> DNA
<213> Stevia rebaudiana

<400> 17
atggctacca gtgactccat agttgacgac cgtaagcagc ttcatgttgc gacgttccca 60

tggcttgctt tcggtcacat cctccctttc cttcagcttt cgaaattgat agctgaaaag 120

ggtcacaaag tctcgtttct ttctaccacc agaaacattc aacgtctctc ttctcatatc 180

tcgccactca taaatgttgt tcaactcaca cttccacgtg tccaagagct gccggaggat 240

gcagaggcga ccactgacgt ccaccctgaa gatattcaat atctcaagaa ggctgttgat 300

ggtcttcaac cggaggtcac ccggtttcta gaacaacact ctccggactg gattatttat 360

gattttactc actactggtt gccatccatc gcggctagcc tcggtatctc acgagcctac 420

ttctgcgtca tcactccatg gaccattgct tatttggcac cctcatctga cgccatgata 480

aatgattcag atggtcgaac cacggttgag gatctcacga caccgcccaa gtggtttccc 540

tttccgacca aagtatgctg gcggaagcat gatcttgccc gaatggagcc ttacgaagct 600

ccggggatat ctgatggata ccgtatgggg atggttttta agggatctga ttgtttgctt 660

ttcaaatgtt accatgagtt tggaactcaa tggctacctc ttttggagac actacaccaa 720

gtaccggtgg ttccggtggg attactgccg ccggaaatac ccggagacga gaaagatgaa 780

acatgggtgt caatcaagaa atggctcgat ggtaaacaaa aaggcagtgt ggtgtacgtt 840

gcattaggaa gcgaggcttt ggtgagccaa accgaggttg ttgagttagc attgggtctc 900

gagctttctg ggttgccatt tgtttgggct tatagaaaac caaaaggtcc cgcgaagtca 960

gactcggtgg agttgccaga cgggttcgtg gaacgaactc gtgaccgtgg gttggtctgg 1020

acgagttggg cacctcagtt acgaatactg agccacgagt cagtttgtgg tttcttgact 1080

cattgtggtt ctggatcaat tgtggaaggg ctaatgtttg gtcaccctct aatcatgcta 1140

ccgatttttt gtgaccaacc tctgaatgct cgattactgg aggacaaaca ggtgggaatc 1200

gagataccaa gaaatgagga agatggttgc ttgaccaagg agtcggttgc tagatcactg 1260

aggtccgttg ttgtggaaaa cgaaggggag atctacaagg cgaacgcgag ggcgctgagt 1320

aaaatctata acgacactaa ggtggaaaaa gaatatgtaa gccaattcgt agactatttg 1380

gaaaagaatg cgcgtgcggt tgccatcgat catgagagtt aa 1422

<210> 18
<211> 1086
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 18
atggctttgg taaacccaac cgctcttttc tatggtacct ctatcagaac aagacctaca 60

aacttactaa atccaactca aaagctaaga ccagtttcat catcttcctt accttctttc 120

tcatcagtta gtgcgattct tactgaaaaa catcaatcta atccttctga gaacaacaat 180

ttgcaaactc atctagaaac tcctttcaac tttgatagtt atatgttgga aaaagtcaac 240

atggttaacg aggcgcttga tgcatctgtc ccactaaaag acccaatcaa aatccatgaa 300

tccatgagat actctttatt ggcaggcggt aagagaatca gaccaatgat gtgtattgca 360

gcctgcgaaa tagtcggagg taatatcctt aacgccatgc cagccgcatg tgccgtggaa 420

atgattcata ctatgtcttt ggtgcatgac gatcttccat gtatggataa tgatgacttc 480

agaagaggta aacctatttc acacaaggtc tacggggagg aaatggcagt attgaccggc 540

gatgctttac taagtttatc tttcgaacat atagctactg ctacaaaggg tgtatcaaag 600

gatagaatcg tcagagctat aggggagttg gcccgttcag ttggctccga aggtttagtg 660

gctggacaag ttgtagatat cttgtcagag ggtgctgatg ttggattaga tcacctagaa 720

tacattcaca tccacaaaac agcaatgttg cttgagtcct cagtagttat tggcgctatc 780

atgggaggag gatctgatca gcagatcgaa aagttgagaa aattcgctag atctattggt 840

ctactattcc aagttgtgga tgacattttg gatgttacaa aatctaccga agagttgggg 900

aaaacagctg gtaaggattt gttgacagat aagacaactt acccaaagtt gttaggtata 960

gaaaagtcca gagaatttgc cgaaaaactt aacaaggaag cacaagagca attaagtggc 1020

tttgatagac gtaaggcagc tcctttgatc gcgttagcca actacaatgc gtaccgtcaa 1080

aattga 1086

<210> 19
<211> 1029
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 19
atggctgagc aacaaatatc taacttgctg tctatgtttg atgcttcaca tgctagtcag 60

aaattagaaa ttactgtcca aatgatggac acataccatt acagagaaac gcctccagat 120

tcctcatctt ctgaaggcgg ttcattgtct agatacgacg agagaagagt ctctttgcct 180

ctcagtcata atgctgcctc tccagatatt gtatcacaac tatgtttttc cactgcaatg 240

tcttcagagt tgaatcacag atggaaatct caaagattaa aggtggccga ttctccttac 300

aactatatcc taacattacc atcaaaagga attagaggtg cctttatcga ttccctgaac 360

gtatggttgg aggttccaga ggatgaaaca tcagtcatca aggaagttat tggtatgctc 420

cacaactctt cattaatcat tgatgacttc caagataatt ctccacttag aagaggaaag 480

ccatctaccc atacagtctt cggccctgcc caggctatca atactgctac ttacgttata 540

gttaaagcaa tcgaaaagat acaagacata gtgggacacg atgcattggc agatgttacg 600

ggtactatta caactatttt ccaaggtcag gccatggact tgtggtggac agcaaatgca 660

atcgttccat caatacagga atacttactt atggtaaacg ataaaaccgg tgctctcttt 720

agactgagtt tggagttgtt agctctgaat tccgaagcca gtatttctga ctctgcttta 780

gaaagtttat ctagtgctgt ttccttgcta ggtcaatact tccaaatcag agacgactat 840

atgaacttga tcgataacaa gtatacagat cagaaaggct tctgcgaaga tcttgatgaa 900

ggcaagtact cactaacact tattcatgcc ctccaaactg attcatccga tctactgacc 960

aacatccttt caatgagaag agtgcaagga aagttaacgg cacaaaagag atgttggttc 1020

tggaaatga 1029

<210> 20
<211> 903
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 20
atggaaaaga ctaaggagaa agcagaacgt atcttgctgg agccatacag atacttatta 60

caactaccag gaaagcaagt ccgttctaaa ctatcacaag cgttcaatca ctggttaaaa 120

gttcctgaag ataagttaca aatcattatt gaagtcacag aaatgctaca caatgcttct 180

ttactgatcg atgatataga ggattcttcc aaactgagaa gaggttttcc tgtcgctcat 240

tccatatacg gggtaccaag tgtaatcaac tcagctaatt acgtctactt cttgggattg 300

gaaaaagtat tgacattaga tcatccagac gctgtaaagc tattcaccag acaacttctt 360

gaattgcatc aaggtcaagg tttggatatc tattggagag acacttatac ttgcccaaca 420

gaagaggagt acaaagcaat ggttctacaa aagactggcg gtttgttcgg acttgccgtt 480

ggtctgatgc aacttttctc tgattacaag gaggacttaa agcctctgtt ggataccttg 540

ggcttgtttt tccagattag agatgactac gctaacttac attcaaagga atattcagaa 600

aacaaatcat tctgtgaaga tttgactgaa gggaagttta gttttccaac aatccacgcc 660

atttggtcaa gaccagaatc tactcaagtg caaaacattc tgcgtcagag aacagagaat 720

attgacatca aaaagtattg tgttcagtac ttggaagatg ttggttcttt tgcttacaca 780

agacatacac ttagagaatt agaggcaaaa gcatacaagc aaatagaagc ctgtggaggc 840

aatccttctc tagtggcatt ggttaaacat ttgtccaaaa tgttcaccga ggaaaacaag 900

taa 903

<210> 21
<211> 1020
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 21
atggcaagat tctattttct taacgcacta ttgatggtta tctcattaca atcaactaca 60

gccttcactc cagctaaact tgcttatcca acaacaacaa cagctctaaa tgtcgcctcc 120

gccgaaactt ctttcagtct agatgaatac ttggcctcta agataggacc tatagagtct 180

gccttggaag catcagtcaa atccagaatt ccacagaccg ataagatctg cgaatctatg 240

gcctactctt tgatggcagg aggcaagaga attagaccag tgttgtgtat cgctgcatgt 300

gagatgttcg gtggatccca agatgtcgct atgcctactg ctgtggcatt agaaatgata 360

cacacaatgt ctttgattca tgatgatttg ccatccatgg ataacgatga cttgagaaga 420

ggtaaaccaa caaaccatgt cgttttcggc gaagatgtag ctattcttgc aggtgactct 480

ttattgtcaa cttccttcga gcacgtcgct agagaaacaa aaggagtgtc agcagaaaag 540

atcgtggatg ttatcgctag attaggcaaa tctgttggtg ccgagggcct tgctggcggt 600

caagttatgg acttagaatg tgaagctaaa ccaggtacca cattagacga cttgaaatgg 660

attcatatcc ataaaaccgc tacattgtta caagttgctg tagcttctgg tgcagttcta 720

ggtggtgcaa ctcctgaaga ggttgctgca tgcgagttgt ttgctatgaa tataggtctt 780

gcctttcaag ttgccgacga tatccttgat gtaaccgctt catcagaaga tttgggtaaa 840

actgcaggca aagatgaagc tactgataag acaacttacc caaagttatt aggattagaa 900

gagagtaagg catacgcaag acaactaatc gatgaagcca aggaaagttt ggctcctttt 960

ggagatagag ctgccccttt attggccatt gcagatttca ttattgatag aaagaattga 1020

<210> 22
<211> 1068
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 22
atgcacttag caccacgtag agtccctaga ggtagaagat caccacctga cagagttcct 60

gaaagacaag gtgccttggg tagaagacgt ggagctggct ctactggctg tgcccgtgct 120

gctgctggtg ttcaccgtag aagaggagga ggcgaggctg atccatcagc tgctgtgcat 180

agaggctggc aagccggtgg tggcaccggt ttgcctgatg aggtggtgtc taccgcagcc 240

gccttagaaa tgtttcatgc ttttgcttta atccatgatg atatcatgga tgatagtgca 300

actagaagag gctccccaac tgttcacaga gccctagctg atcgtttagg cgctgctctg 360

gacccagatc aggccggtca actaggagtt tctactgcta tcttggttgg agatctggct 420

ttgacatggt ccgatgaatt gttatacgct ccattgactc cacatagact ggcagcagta 480

ctaccattgg taacagctat gagagctgaa accgttcatg gccaatatct tgatataact 540

agtgctagaa gacctgggac cgatacttct cttgcattga gaatagccag atataagaca 600

gcagcttaca caatggaacg tccactgcac attggtgcag ccctggctgg ggcaagacca 660

gaactattag cagggctttc agcatacgcc ttgccagctg gagaagcctt ccaattggca 720

gatgacctgc taggcgtctt cggtgatcca agacgtacag ggaaacctga cctagatgat 780

cttagaggtg gaaagcatac tgtcttagtc gccttggcaa gagaacatgc cactccagaa 840

cagagacaca cattggatac attattgggt acaccaggtc ttgatagaca aggcgcttca 900

agactaagat gcgtattggt agcaactggt gcaagagccg aagccgaaag acttattaca 960

gagagaagag atcaagcatt aactgcattg aacgcattaa cactgccacc tcctttagct 1020

gaggcattag caagattgac attagggtct acagctcatc ctgcctaa 1068

<210> 23
<211> 993
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 23
atgtcatatt tcgataacta cttcaatgag atagttaatt ccgtgaacga catcattaag 60

tcttacatct ctggcgacgt accaaaacta tacgaagcct cctaccattt gtttacatca 120

ggaggaaaga gactaagacc attgatcctt acaatttctt ctgatctttt cggtggacag 180

agagaaagag catactatgc tggcgcagca atcgaagttt tgcacacatt cactttggtt 240

cacgatgata tcatggatca agataacatt cgtagaggtc ttcctactgt acatgtcaag 300

tatggcctac ctttggccat tttagctggt gacttattgc atgcaaaagc ctttcaattg 360

ttgactcagg cattgagagg tctaccatct gaaactatca tcaaggcgtt tgatatcttt 420

acaagatcta tcattatcat atcagaaggt caagctgtcg atatggaatt cgaagataga 480

attgatatca aggaacaaga gtatttggat atgatatctc gtaaaaccgc tgccttattc 540

tcagcttctt cttccattgg ggcgttgata gctggagcta atgataacga tgtgagatta 600

atgtccgatt tcggtacaaa tcttgggatc gcatttcaaa ttgtagatga tatacttggt 660

ttaacagctg atgaaaaaga gctaggaaaa cctgttttca gtgatatcag agaaggtaaa 720

aagaccatat tagtcattaa gactttagaa ttgtgtaagg aagacgagaa aaagattgtg 780

ttaaaagcgc taggcaacaa gtcagcatca aaggaagagt tgatgagttc tgctgacata 840

atcaaaaagt actcattgga ttacgcctac aacttagctg agaaatacta caaaaacgcc 900

atcgattctc taaatcaagt ttcaagtaaa agtgatattc cagggaaggc attgaaatat 960

cttgctgaat tcaccatcag aagacgtaag taa 993

<210> 24
<211> 894
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 24
atggtcgcac aaactttcaa cctggatacc tacttatccc aaagacaaca acaagttgaa 60

gaggccctaa gtgctgctct tgtgccagct tatcctgaga gaatatacga agctatgaga 120

tactccctcc tggcaggtgg caaaagatta agacctatct tatgtttagc tgcttgcgaa 180

ttggcaggtg gttctgttga acaagccatg ccaactgcgt gtgcacttga aatgatccat 240

acaatgtcac taattcatga tgacctgcca gccatggata acgatgattt cagaagagga 300

aagccaacta atcacaaggt gttcggggaa gatatagcca tcttagcggg tgatgcgctt 360

ttagcttacg cttttgaaca tattgcttct caaacaagag gagtaccacc tcaattggtg 420

ctacaagtta ttgctagaat cggacacgcc gttgctgcaa caggcctcgt tggaggccaa 480

gtcgtagacc ttgaatctga aggtaaagct atttccttag aaacattgga gtatattcac 540

tcacataaga ctggagcctt gctggaagca tcagttgtct caggcggtat tctcgcaggg 600

gcagatgaag agcttttggc cagattgtct cattacgcta gagatatagg cttggctttt 660

caaatcgtcg atgatatcct ggatgttact gctacatctg aacagttggg gaaaaccgct 720

ggtaaagacc aggcagccgc aaaggcaact tatccaagtc tattgggttt agaagcctct 780

agacagaaag cggaagagtt gattcaatct gctaaggaag ccttaagacc ttacggttca 840

caagcagagc cactcctagc gctggcagac ttcatcacac gtcgtcagca ttaa 894

<210> 25
<211> 1116
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 25
atggcctctg ttactttggg ttcctggatc gtcgtccacc accataacca tcaccatcca 60

tcatctatcc taactaaatc tcgttcaaga tcctgtccta ttacactaac caaaccaatc 120

tcttttcgtt caaagagaac agtttcctct agtagttcta tcgtgtcctc tagtgtcgtc 180

actaaggaag acaatctgag acagtctgaa ccttcttcct ttgatttcat gtcatatatc 240

attactaagg cagaactagt gaataaggct cttgattcag cagttccatt aagagagcca 300

ttgaaaatcc atgaagcaat gagatactct cttctagctg gcgggaagag agtcagacct 360

gtactctgca tagcagcgtg cgaattagtt ggtggcgagg aatcaaccgc tatgcctgcc 420

gcttgtgctg tagaaatgat tcatacaatg tcactgatac acgatgattt gccatgtatg 480

gataacgatg atctgagaag gggtaagcca actaaccata aggttttcgg cgaagatgtt 540

gccgtcttag ctggtgatgc tttgttatct ttcgcgttcg aacatttggc atccgcaaca 600

tcaagtgatg ttgtgtcacc agtaagagta gttagagcag ttggagaact ggctaaagct 660

attggaactg agggtttagt tgcaggtcaa gtcgtcgata tctcttccga aggtcttgat 720

ttgaatgatg taggtcttga acatctcgaa ttcatccatc ttcacaagac agctgcactt 780

ttagaagcca gtgcggttct cggcgcaatt gttggcggag ggagtgatga cgaaattgag 840

agattgagga agtttgctag atgtatagga ttactgttcc aagtagtaga cgatatacta 900

gatgtgacaa agtcttccaa agagttggga aaaacagctg gtaaagattt gattgccgac 960

aaattgacct accctaagat tatggggcta gaaaaatcaa gagaatttgc cgagaaactc 1020

aatagagagg cgcgtgatca actgttgggt ttcgattctg ataaagttgc accactctta 1080

gccttagcca actacatcgc ttacagacaa aactaa 1116

<210> 26
<211> 1086
<212> DNA
<213> Stevia rebaudiana

<400> 26
atggctcttg taaatcccac agctttgttc tatggaacct ccataagaac cagacccaca 60

aacttgctca acccgaccca aaaacttcga cccgtttcct cgtcttcttt gccttccttc 120

tcttcagttt ctgcaatctt gacggaaaaa caccaatcaa acccatcaga aaacaataac 180

ttgcaaaccc atctcgaaac accattcaat ttcgactctt acatgctgga gaaagtaaac 240

atggtgaatg aagctctgga cgcctcggtt ccactcaaag acccgataaa gatccatgaa 300

tccatgcggt actcccttct agctggcggg aaacgcatcc gaccgatgat gtgcatcgcc 360

gcttgcgaaa tagtcggagg caacatatta aacgccatgc cagctgcatg cgcggtcgag 420

atgattcaca ccatgtcact agttcatgac gaccttccat gcatggataa cgacgacttc 480

cgacgtggaa aaccaataag ccacaaggtg tacggtgaag aaatggcggt tctaaccggg 540

gacgcgttac tctcattatc cttcgaacat atcgcgaccg cgacaaaagg cgtatccaaa 600

gacaggatcg tccgagccat tggtgaactc gcaaggtccg ttggctcgga gggtttggtc 660

gccggtcagg tggttgatat tttatccgaa ggggctgatg ttgggttaga ccacttggag 720

tatattcata tacacaagac tgcaatgttg cttgagagct cggtcgtgat cggcgcgatc 780

atgggcggtg ggtctgacca acagatcgaa aagttgcgaa agtttgcgag atcgattggt 840

ttgttgtttc aggtggtaga tgatattctt gatgtcacaa agtcgactga ggaattgggg 900

aaaacggcgg gaaaagattt gctgacggac aagacaacgt atccgaagtt gttggggatc 960

gaaaaatcga gagaatttgc ggagaaatta aacaaggaag cgcaagaaca attgtcgggg 1020

tttgatcgcc gcaaggcggc tccgttaatt gcccttgcta attacaatgc ttataggcaa 1080

aactga 1086

<210> 27
<211> 1029
<212> DNA
<213> Gibberella fujikuroi

<400> 27
atggctgaac aacagatctc caaccttctt tcaatgtttg atgcttctca cgcaagccag 60

aagttggaga ttacggttca gatgatggat acctaccatt acagagaaac tcctccagac 120

tcttcctctt cagaaggcgg ttccttatct cgctatgatg agcgacgggt ctcccttccg 180

ctctctcaca atgcagcctc cccagacata gtctcccagt tatgcttctc aacagctatg 240

agctcggagc tcaatcacag gtggaagtca cagcgcctca aggttgctga ctctccctac 300

aactacatcc tgactcttcc atctaaaggt attcgtgggg ctttcattga ctcactgaat 360

gtctggctcg aggtccccga agacgagacc tcggtgatca aagaggtgat tggcatgctc 420

cacaactcgt ctctcataat cgatgacttc caagacaact ccccacttcg gcggggcaag 480

ccatctacac atactgtctt cggtccagca caagcaatca acacagcaac atatgtcatc 540

gtcaaggcca tcgagaaaat acaggatatc gtcggtcacg atgcattggc agatgtaact 600

ggcactataa ccacaatctt ccagggtcag gcaatggatc tgtggtggac tgctaatgcc 660

attgttccgt ctatccaaga atatctcctg atggtcaatg acaagactgg tgccctgttc 720

aggttatcgc ttgaactact ggcgctgaac tctgaagcat ccatcagtga cagcgcgctt 780

gaatctctca gcagcgctgt ctcactgctc gggcagtatt tccagataag agatgattac 840

atgaatctca ttgacaacaa gtatactgat cagaaaggat tttgcgagga tctggacgag 900

gggaaatact cgttgactct aatccatgct ctgcagaccg actccagcga ccttctcacc 960

aacatcttat cgatgagaag agtccaagga aaacttacgg cgcagaaaag atgctggttt 1020

tggaagtga 1029

<210> 28
<211> 903
<212> DNA
<213> Mus musculus

<400> 28
atggagaaaa ctaaagagaa agctgagagg attcttctag agccctatag gtacttactt 60

cagttaccag gtaaacaggt gagaagcaaa ctttcacagg catttaatca ctggctgaaa 120

gttccagaag acaagctaca gattatcatt gaagtgactg aaatgttgca taatgccagt 180

ttactcattg atgatattga agacagttca aagctccgac gtggtttccc agtggctcac 240

agcatctatg gtgtcccatc tgtcattaat tctgccaatt acgtctactt ccttggactg 300

gaaaaagtct taacccttga tcacccggat gcggtgaagc ttttcacacg ccagcttctg 360

gaacttcatc agggacaagg cctcgatatt tactggaggg acacctacac ttgtccaact 420

gaagaagaat ataaagccat ggtgttgcag aagacaggtg gtttgtttgg attagcagta 480

ggtcttatgc agctgttctc tgattacaaa gaagatctaa agccactgct tgacacactt 540

gggctctttt tccagattag agatgattat gccaatctac actccaaaga atacagtgaa 600

aacaaaagtt tctgtgaaga cttgacagaa gggaagttct cattccccac tatccatgcc 660

atttggtcaa ggccagaaag cacccaggta cagaacatcc tgcgccagag aacagagaat 720

atagatatta aaaagtattg tgtgcagtac ctggaggatg taggttcttt tgcatacact 780

cgacacactc ttagagagct tgaagctaaa gcctacaaac aaattgaggc ctgtggtggg 840

aacccttcac tagtggcttt agtcaagcac ttaagtaaga tgttcacaga agaaaataaa 900

taa 903

<210> 29
<211> 1020
<212> DNA
<213> Thalassiosira pseudonana

<400> 29
atggctcgtt tctacttcct gaacgctctc ctcatggtga tttctttaca aagcaccacg 60

gcattcaccc cggcaaaact cgcctaccca acaaccacca ctgcattaaa cgttgcctct 120

gccgaaacat catttagcct cgatgaatac ctagcctcca aaatcggacc cattgaatca 180

gctctcgagg catctgtcaa atctcgcatt cctcaaactg acaagatatg cgagtctatg 240

gcatactcac tcatggctgg aggaaagcgt atccgtcccg ttttgtgcat tgctgcttgt 300

gaaatgtttg ggggaagtca agatgtggct atgccgacgg ctgtggcttt ggagatgatt 360

catactatga gtcttattca tgacgatttg ccttcaatgg acaacgatga tctccgacga 420

ggaaagccaa ctaatcatgt tgtctttgga gaggatgttg ctattcttgc tggggattct 480

cttctcagta cgtcttttga acatgttgcc cgtgaaacca aaggagtgtc agctgaaaag 540

attgtagatg ttatcgctcg cctcgggaag tctgtgggtg cagagggtct tgctggtgga 600

caggttatgg atcttgagtg tgaggcgaag ccaggaacta ccctcgacga tctcaagtgg 660

attcacattc acaaaactgc cactcttctt caagtggcag tggcatcagg tgctgttctt 720

ggaggggcca caccagagga ggttgctgct tgtgaactgt tcgcaatgaa tattggactt 780

gccttccagg tcgctgatga tattttggac gtgacggcat cgagtgagga tcttggcaaa 840

actgctggaa aggatgaagc cacagataag acaacttatc ctaagctttt gggattggag 900

gagagtaagg catacgctcg acaactcata gacgaagcaa aggaatcttt ggctcctttc 960

ggtgatcgtg ctgctccatt gttggcaatt gccgacttta tcattgatcg aaagaactag 1020

<210> 30
<211> 1068
<212> DNA
<213> Streptomyces clavuligerus

<400> 30
atgcacctgg ctccccgccg agtaccgcgc ggccgtcgaa gcccacctga ccgcgttcct 60

gaacgccaag gagcgctcgg ccgccgccgg ggggccggtt ccacaggatg tgcccgcgct 120

gctgcgggag ttcatcggcg ccgggggggg ggggaagcgg atccgtccgc tgctgtgcat 180

cgcggctggc aggccggcgg cggaacagga ctgccggacg aggtggtgtc cacagcggcg 240

gcgctggaga tgttccacgc gttcgcgctg atccacgacg acatcatgga tgactccgcg 300

accaggcgcg gcagcccgac ggtgcaccgg gcactcgccg accggctcgg cgccgctctc 360

gaccccgacc aagccggaca actgggggtg agcacggcga tcctcgtcgg ggacctcgcc 420

ctgacctggt cggacgaact gctgtacgct cccctgaccc cccaccggct ggccgcggta 480

ctgcccctgg tcacggccat gcgcgcggaa acggtccacg gccagtacct ggacatcacc 540

tccgcccgcc ggcccggcac ggacacctca ctggcgctgc gaatcgcgcg ctacaaaacc 600

gctgcttaca ccatggaacg ccccctgcac atcggagcag cgctcgccgg cgcacgaccg 660

gaactcctgg cagggctcag cgcctacgcg ctgccggcgg gcgaggcatt ccagctcgcc 720

gacgacctcc tgggagtgtt cggcgatcca cggagaaccg gcaaacccga cctcgacgac 780

ctccgcggcg gcaagcacac cgtcctcgtg gccctcgccc gggaacacgc cacacctgaa 840

cagcggcaca ccctggacac cctgctcggc acaccaggcc tcgaccggca gggcgcgtcc 900

cggctgcgct gcgtcctcgt cgccaccggg gcccgggcgg aagccgaacg cctgatcacc 960

gaacggcgcg accaggccct caccgcgctc aacgccctga cactgccccc accgctcgcc 1020

gaggcactcg cccgcctcac cctcgggagt accgcacacc cggcctga 1068

<210> 31
<211> 993
<212> DNA
<213> Sulfulobus acidicaldarius

<400> 31
atgagttact ttgacaacta ttttaatgag attgttaatt ctgtaaacga cattattaag 60

agctatatat ctggagatgt tcctaaacta tatgaagcct catatcattt gtttacatct 120

ggaggtaaga ggttaagacc attaatctta actatatcat cagatttatt cggaggacag 180

agagaaagag cttattatgc aggtgcagct attgaagttc ttcatacttt tacgcttgtg 240

catgatgata ttatggatca agataatatc agaagagggt tacccacagt ccacgtgaaa 300

tacggcttac ccttagcaat attagctggg gatttactac atgcaaaggc ttttcagctc 360

ttaacccagg ctcttagagg tttgccaagt gaaaccataa ttaaggcttt cgatattttc 420

actcgttcaa taataattat atccgaagga caggcagtag atatggaatt tgaggacaga 480

attgatataa aggagcagga ataccttgac atgatctcac gtaagacagc tgcattattc 540

tcggcatcct caagtatagg cgcacttatt gctggtgcta atgataatga tgtaagactg 600

atgtctgatt tcggtacgaa tctaggtatt gcatttcaga ttgttgacga tatcttaggt 660

ctaacagcag acgaaaagga acttggaaag cctgttttta gtgatattag ggagggtaaa 720

aagactatac ttgtaataaa aacactggag ctttgtaaag aggacgagaa gaagattgtc 780

ctaaaggcgt taggtaataa gtcagcctca aaagaagaat taatgagctc agcagatata 840

attaagaaat actctttaga ttatgcatac aatttagcag agaaatatta taaaaatgct 900

atagactctt taaatcaagt ctcctctaag agtgatatac ctggaaaggc tttaaaatat 960

ctagctgaat ttacgataag aaggagaaaa taa 993

<210> 32
<211> 894
<212> DNA
<213> Synechococcus sp.

<400> 32
ttggttgccc aaaccttcaa cctggacacc tacttgagcc aacgccagca acaggtggaa 60

gaggcgcttt ctgcggcatt ggttcccgcc tatccggagc gcatttacga ggcgatgcgc 120

tacagcctgc tggcgggggg gaaacgcctg aggccgatcc tctgtctggc ggcctgtgag 180

ttggccggcg gctctgtgga gcaggccatg cccaccgcct gcgccctgga gatgatccac 240

accatgtcgc tgatccacga cgatctgccg gcgatggaca acgacgattt tcgccgcggc 300

aagcccacca atcacaaggt attcggcgag gatatcgcca ttttggcagg agatgccctg 360

ttggcctatg cctttgagca tatcgccagc caaacgcggg gggtgccgcc gcagttggtg 420

ctgcaagtca ttgcccgcat tggccatgct gtggcggcaa ccggcttggt agggggccag 480

gtggtggatc tggagtccga aggcaaagcc atttccctag aaactttgga gtacatccac 540

agtcacaaga cgggtgctct gctggaggcc tcggtggttt cgggagggat cctggcaggg 600

gccgatgagg agctgctggc gcggctgagc cactacgctc gggacatcgg cctggctttt 660

cagatcgtgg acgacatttt ggatgttact gccaccagcg agcaactggg caaaacggca 720

ggcaaggatc aagctgccgc caaagccacc taccccagct tgttgggcct agaggcttcc 780

cggcagaaag ctgaggaact gatccaatcg gccaaggagg cgttgcgccc ctacggatcc 840

caggccgagc ccctgttggc tctggccgat ttcatcaccc gccgccagca ttga 894

<210> 33
<211> 1116
<212> DNA
<213> Arabidopsis thaliana

<400> 33
atggcttcag tgactctagg ttcatggatt gttgttcacc accacaatca tcatcatcca 60

tcttcaatcc ttaccaaatc cagatccaga tcttgtccta taactcttac taaacccatc 120

tcctttcgat caaaacgcac cgtttcatca tcttcttcaa tcgtttcttc ttccgttgtt 180

acaaaagaag acaatctacg ccaatctgaa ccatcctctt tcgatttcat gtcgtacatc 240

atcaccaaag ccgaattagt caacaaagct ttagattcag ctgttcctct ccgtgagcca 300

ctcaagatcc acgaagcgat gcgttactct cttctcgccg gtggcaaaag agttagacca 360

gttctctgca tcgctgcttg tgaactcgtc ggaggtgaag aatcaaccgc tatgccagca 420

gcttgcgccg tcgagatgat tcacaccatg tcgttgatcc acgacgatct cccttgtatg 480

gataacgacg atctccgccg tggaaaaccg accaaccaca aagtgtttgg tgaagacgtc 540

gctgttttag ccggagacgc gcttctctct ttcgctttcg agcatttagc ttcggcgacg 600

agttctgatg ttgtttctcc ggtgagagtg gttcgagccg ttggagaatt ggctaaagcg 660

ataggaacag aagggttagt ggcgggtcaa gtcgtggata ttagtagtga agggttagat 720

ttaaacgacg tcggtttaga gcatttggag tttatccatt tgcataaaac ggcggcgttg 780

cttgaagctt ctgctgtttt gggagctatt gttggtggag gaagtgatga tgagattgag 840

aggttaagaa agtttgcgag atgtattggt ttgttgtttc aggtggttga tgatatcttg 900

gatgtgacga aatcgtcgaa agagttaggg aaaactgctg ggaaagattt gattgctgat 960

aagttgacgt atcctaagat tatgggtttg gagaaatcga gagagtttgc tgagaaattg 1020

aatagagagg ctcgtgatca gcttttaggg tttgattctg ataaggttgc tcctttgttg 1080

gctttggcta attacattgc ctatagacag aactga 1116

<210> 34
<211> 2364
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 34
atgaaaaccg ggtttatctc accagcaaca gtatttcatc acagaatctc accagcgacc 60

actttcagac atcacttatc acctgctact acaaactcta caggcattgt cgccttaaga 120

gacatcaact tcagatgtaa agcagtttct aaagagtact ctgatctgtt gcagaaagat 180

gaggcttctt tcacaaaatg ggacgatgac aaggtgaaag atcatcttga taccaacaaa 240

aacttatacc caaatgatga gattaaggaa tttgttgaat cagtaaaggc tatgttcggt 300

agtatgaatg acggggagat aaacgtctct gcatacgata ctgcatgggt tgctttggtt 360

caagatgtcg atggatcagg tagtcctcag ttcccttctt ctttagaatg gattgccaac 420

aatcaattgt cagatggatc atggggagat catttgctgt tctcagctca cgatagaatc 480

atcaacacat tagcatgcgt tattgcactt acaagttgga atgttcatcc ttctaagtgt 540

gaaaaaggtt tgaattttct gagagaaaac atttgcaaat tagaagatga aaacgcagaa 600

catatgccaa ttggttttga agtaacattc ccatcactaa ttgatatcgc gaaaaagttg 660

aacattgaag tacctgagga tactccagca cttaaagaga tctacgcacg tagagatatc 720

aagttaacta agatcccaat ggaagttctt cacaaggtac ctactacttt gttacattct 780

ttggaaggaa tgcctgattt ggagtgggaa aaactgttaa agctacaatg taaagatggt 840

agtttcttgt tttccccatc tagtaccgca ttcgccctaa tgcaaacaaa agatgagaaa 900

tgcttacagt atctaacaaa tatcgtcact aagttcaacg gtggcgtgcc taatgtgtac 960

ccagtcgatt tgtttgaaca tatttgggtt gttgatagac tgcagagatt ggggattgcc 1020

agatacttca aatcagagat aaaagattgt gtagagtata tcaataagta ctggaccaaa 1080

aatggaattt gttgggctag aaatactcac gttcaagata tcgatgatac agccatggga 1140

ttcagagtgt tgagagcgca cggttatgac gtcactccag atgtttttag acaatttgaa 1200

aaagatggta aattcgtttg ctttgcaggg caatcaacac aagccgtgac aggaatgttt 1260

aacgtttaca gagcctctca aatgttgttc ccaggggaga gaattttgga agatgccaaa 1320

aagttctctt acaattactt aaaggaaaag caaagtacca acgaattgct ggataaatgg 1380

ataatcgcta aagatctacc tggtgaagtt ggttatgctc tggatatccc atggtatgct 1440

tccttaccaa gattggaaac tcgttattac cttgaacaat acggcggtga agatgatgtc 1500

tggataggca agacattata cagaatgggt tacgtgtcca ataacacata tctagaaatg 1560

gcaaagctgg attacaataa ctatgttgca gtccttcaat tagaatggta cacaatacaa 1620

caatggtacg tcgatattgg tatagagaag ttcgaatctg acaacatcaa gtcagtcctg 1680

gtttcttact acttggctgc ggcttcaata ttcgaacctg agagatctaa ggagagaatc 1740

gcttgggcaa agacaacaat cttagtcgat aagatcacat caattttcga ttcctctcag 1800

tcaagtaagg aagatattac tgcctttatt gacaagtttc gtaacaagtc ctcctctaaa 1860

aagcactcta tcaacggtga accatggcat gaagttatgg tagctttgaa aaagacctta 1920

cacggctttg ctctggatgc tcttatgact cattctcaag atatacatcc acagttacat 1980

caagcctggg aaatgtggtt gactaaacta caagacggcg tagatgttac tgctgagcta 2040

atggtccaaa tgatcaacat gactgctggc agatgggtat caaaggaatt acttactcat 2100

ccacaatatc aaagattgtc tactgtgaca aattctgtgt gtcacgatat taccaaactt 2160

cacaatttca aggagaattc caccacagtg gattcaaagg ttcaggaact agtccagttg 2220

gtttttagtg acacaccaga tgatttggat caagatatga aacaaacatt cctgacagtg 2280

atgaagacat tctactacaa ggcgtggtgt gatccaaaca ctataaacga tcatatatct 2340

aaagttttcg aaatcgtaat ttga 2364

<210> 35
<211> 1584

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 35
atgcctgatg cacacgatgc tccacctcca caaataagac agagaacact agtagatgag 60

gctacccaac tgctaactga gtccgcagaa gatgcatggg gtgaagtcag tgtgtcagaa 120

tacgaaacag caaggctagt tgcccatgct acatggttag gtggacacgc cacaagagtg 180

gccttccttc tggagagaca acacgaagac gggtcatggg gtccaccagg tggatatagg 240

ttagtcccta cattatctgc tgttcacgca ttattgacat gtcttgcctc tcctgctcag 300

gatcatggcg ttccacatga tagactttta agagctgttg acgcaggctt gactgccttg 360

agaagattgg ggacatctga ctccccacct gatactatag cagttgagct ggttatccca 420

tctttgctag agggcattca acacttactg gaccctgctc atcctcatag tagaccagcc 480

ttctctcaac atagaggctc tcttgtttgt cctggtggac tagatgggag aactctagga 540

gctttgagat cacacgccgc agcaggtaca ccagtaccag gaaaagtctg gcacgcttcc 600

gagactttgg gcttgagtac cgaagctgct tctcacttgc aaccagccca aggtataatc 660

ggtggctctg ctgctgccac agcaacatgg ctaaccaggg ttgcaccatc tcaacagtca 720

gattctgcca gaagatacct tgaggaatta caacacagat actctggccc agttccttcc 780

attaccccta tcacatactt cgaaagagca tggttattga acaattttgc agcagccggt 840

gttccttgtg aggctccagc tgctttgttg gattccttag aagcagcact tacaccacaa 900

ggtgctcctg ctggagcagg attgcctcca gatgctgatg atacagccgc tgtgttgctt 960

gcattggcaa cacatgggag aggtagaaga ccagaagtac tgatggatta caggactgac 1020

gggtatttcc aatgctttat tggggaaagg actccatcaa tttcaacaaa cgctcacgta 1080

ttggaaacat tagggcatca tgtggcccaa catccacaag atagagccag atacggatca 1140

gccatggata ccgcatcagc ttggctgctg gcagctcaaa agcaagatgg ctcttggtta 1200

gataaatggc atgcctcacc atactacgct actgtttgtt gcacacaagc cctagccgct 1260

catgcaagtc ctgcaactgc accagctaga cagagagctg tcagatgggt tttagccaca 1320

caaagatccg atggcggttg gggtctatgg cattcaactg ttgaagagac tgcttatgcc 1380

ttacagatct tggccccacc ttctggtggt ggcaatatcc cagtccaaca agcacttact 1440

agaggcagag caagattgtg tggagccttg ccactgactc ctttatggca tgataaggat 1500

ttgtatactc cagtaagagt agtcagagct gccagagctg ctgctctgta cactaccaga 1560

gatctattgt taccaccatt gtaa 1584

<210> 36
<211> 1551
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 36
atgaacgccc tatccgaaca cattttgtct gaattgagaa gattattgtc tgaaatgagt 60

gatggcggat ctgttggtcc atctgtgtat gatacggccc aggccctaag attccacggt 120

aacgtaacag gtagacaaga tgcatatgct tggttgatcg cccagcaaca agcagatgga 180

ggttggggct ctgccgactt tccactcttt agacatgctc caacatgggc tgcacttctc 240

gcattacaaa gagctgatcc acttcctggc gcagcagacg cagttcagac cgcaacaaga 300

ttcttgcaaa gacaaccaga tccatacgct catgccgttc ctgaggatgc ccctattggt 360

gctgaactga tcttgcctca gttttgtgga gaggctgctt ggttgttggg aggtgtggcc 420

ttccctagac acccagccct attaccatta agacaggctt gtttagtcaa actgggtgca 480

gtcgccatgt tgccttcagg acacccattg ctccactcct gggaggcatg gggtacttct 540

ccaacaacag cctgtccaga cgatgatggt tctataggta tctcaccagc agctacagcc 600

gcctggagag cccaggctgt gaccagaggc tcaactcctc aagtgggcag agctgacgca 660

tacttacaaa tggcttcaag agcaacgaga tcaggcatag aaggagtctt ccctaatgtt 720

tggcctataa acgtattcga accatgctgg tcactgtaca ctctccatct tgccggtctg 780

ttcgcccatc cagcactggc tgaggctgta agagttatcg ttgctcaact tgaagcaaga 840

ttgggagtgc atggcctcgg accagcttta cattttgctg ccgacgctga tgatactgca 900

gttgccttat gcgttctgca tttggctggc agagatcctg cagttgacgc attgagacat 960

tttgaaattg gtgagctctt tgttacattc ccaggagaga gaaatgctag tgtctctacg 1020

aacattcacg ctcttcatgc tttgagattg ttaggtaaac cagctgccgg agcaagtgca 1080

tacgtcgaag caaatagaaa tccacatggt ttgtgggaca acgaaaaatg gcacgtttca 1140

tggctttatc caactgcaca cgccgttgca gctctagctc aaggcaagcc tcaatggaga 1200

gatgaaagag cactagccgc tctactacaa gctcaaagag atgatggtgg ttggggagct 1260

ggtagaggat ccactttcga ggaaaccgcc tacgctcttt tcgctttaca cgttatggac 1320

ggatctgagg aagccacagg cagaagaaga atcgctcaag tcgtcgcaag agccttagaa 1380

tggatgctag ctagacatgc cgcacatgga ttaccacaaa caccactctg gattggtaag 1440

gaattgtact gtcctactag agtcgtaaga gtagctgagc tagctggcct gtggttagca 1500

ttaagatggg gtagaagagt attagctgaa ggtgctggtg ctgcacctta a 1551

<210> 37
<211> 2364
<212> DNA
<213> Stevia rebaudiana

<400> 37
atgaagaccg gcttcatctc tcccgccacc gtcttccacc accgtatttc tccggcaacc 60

accttccgcc accacctttc tccggcgacc accaactcca ctggaattgt agctcttaga 120

gacatcaact tccggtgtaa agcggtatcc aaagagtact ctgatttact acaaaaagat 180

gaggcttcat ttaccaagtg ggacgatgac aaagtgaagg accatttgga cacaaataag 240

aatttgtatc caaacgatga gatcaaggag tttgttgaga gcgtgaaagc aatgtttggt 300

tctatgaatg acggagaaat aaatgtgtca gcgtatgata cggcttgggt tgcactcgtg 360

caagatgttg atggaagtgg ttcccctcaa tttccatcaa gtttggagtg gatcgcgaac 420

aatcaactct cagatgggtc ttggggcgat catttgttat tttcggctca tgataggatc 480

attaacacgt tggcatgtgt tatagcgctt acttcttgga acgtccatcc aagtaaatgt 540

gaaaaaggac tgaattttct tagagaaaac atatgtaaac tcgaagacga gaacgcggaa 600

catatgccaa ttggttttga agtcacgttc ccgtcgctaa tagatatcgc aaagaagcta 660

aatattgaag ttcctgagga tactcctgcc ttaaaagaaa tttatgcaag aagagacata 720

aaactcacaa agataccaat ggaagtattg cacaaagtgc ccacaacttt acttcatagt 780

ttggaaggaa tgccagattt ggaatgggaa aaacttctga aattgcaatg caaagatgga 840

tcatttctgt tttctccatc atctactgct tttgcactca tgcaaacaaa agatgaaaag 900

tgtcttcagt atttgacaaa tattgttacc aaattcaatg gtggagttcc gaatgtgtac 960

ccggtggatc tattcgaaca tatttgggta gttgatcgac ttcaacgact tgggattgct 1020

cgttatttca aatcagagat caaagattgc gttgaatata ttaacaagta ttggacaaag 1080

aatgggattt gttgggcaag aaacacgcac gtacaagata ttgatgatac cgcaatggga 1140

tttagggttt taagagcaca tggttatgat gttactccag atgtatttcg acaatttgag 1200

aaggatggta aattcgtatg tttcgctgga cagtcaacac aagccgtcac cggaatgttc 1260

aatgtgtata gagcgtcaca aatgctcttt cccggagaaa gaattcttga agatgcaaag 1320

aaattttcat ataattattt gaaagaaaaa caatcgacaa atgagcttct tgataaatgg 1380

atcatcgcca aagacttacc tggagaggtt ggatatgcgc tagacatacc atggtatgca 1440

agcttaccgc gactcgagac aagatattac ttagagcaat acgggggcga ggatgatgtt 1500

tggattggaa aaactctata caggatggga tatgtgagca ataatacgta ccttgaaatg 1560

gccaaattgg actacaataa ctatgtggcc gtgcttcaac tcgaatggta cactatccag 1620

caatggtatg ttgatatcgg tatcgaaaag tttgaaagtg acaatatcaa aagcgtatta 1680

gtgtcgtatt acttggctgc agccagcata ttcgagccgg aaaggtccaa ggaacgaatc 1740

gcgtgggcta aaaccaccat attagttgac aagatcacct caatttttga ttcatcacaa 1800

tcctcaaaag aggacataac agcctttata gacaaattta ggaacaaatc gtcttctaag 1860

aagcattcaa taaatggaga accatggcac gaggtgatgg ttgcactgaa aaagacccta 1920

cacggcttcg ctttggatgc actcatgact catagtcaag acatccaccc gcaactccat 1980

caagcttggg agatgtggtt gacgaaattg caagatggag tagatgtgac agcggaatta 2040

atggtacaaa tgataaatat gacagctggt cgttgggtat ccaaagaact tttaactcat 2100

cctcaatacc aacgcctctc aaccgtcaca aatagtgtgt gtcacgatat aactaagctc 2160

cataacttca aggagaattc cacgacggta gactcgaaag ttcaagaact agtgcaactt 2220

gtgtttagcg acacgcccga tgatcttgat caggatatga aacagacgtt tctaaccgtc 2280

atgaaaacct tctactacaa ggcgtggtgt gatccgaaca cgataaatga ccatatctcc 2340

aaggtgttcg agattgtaat atga 2364

<210> 38
<211> 1584
<212> DNA
<213> Streptomyces clavuligerus

<400> 38
ttgcccgacg cgcatgatgc ccctccgcct cagatacgac agcggaccct tgtcgatgag 60

gcgacgcaac tcctcacgga gtcggccgag gacgcctggg gtgaggtgtc cgtgtccgag 120

tacgaaacgg cgcggctggt ggcccacgcc acctggctcg gcggtcacgc cacacgggtg 180

gccttcctgc tggagcggca gcatgaggac ggctcgtggg gcccgcccgg cgggtaccgt 240

ctcgtaccca cgctgagtgc cgtacacgcc ctgctcacct gtctggcgtc tcccgcgcag 300

gaccacggag tgcctcatga ccggctcctg cgcgcagttg acgcgggcct gacggcactg 360

cgtcgtcttg ggacgagcga cagcccgccg gacaccattg cggtcgaact ggtcataccc 420

tcgctccttg agggcatcca gcacctcctg gacccggcgc acccgcattc ccgacccgct 480

ttttcgcaac accgcggcag cctcgtctgc cccgggggcc tcgacggccg cacgctgggg 540

gccttgcgct cccacgccgc agccggcaca cctgtcccgg gcaaggtgtg gcacgcctcg 600

gaaaccttgg ggctatcgac cgaggcagcc tcccaccttc aacccgccca gggcatcatc 660

ggtggctccg ccgccgcgac agcaacatgg ctcaccaggg tcgccccgtc gcaacagagc 720

gacagcgcac ggcgctacct ggaagaactc cagcaccgat acagcggccc ggtgccctcc 780

atcaccccga tcacctattt cgaacgggcc tggctgctca acaacttcgc tgccgcgggg 840

gttccatgcg aggctccggc agcccttctc gacagcctgg aggcagcgct cacaccacag 900

ggcgctccag cgggtgcggg actgccgccg gacgcggatg acaccgccgc cgttctgctg 960

gcgcttgcca cgcacggccg cgggcgccgt cccgaggtcc tcatggacta ccgcacggac 1020

ggctacttcc agtgcttcat cggcgaacgc acccccttcca tcagcaccaa tgcccatgtc 1080

ctggagacgc tcggtcacca cgtcgcccaa caccctcagg acagggcccg atacggctca 1140

gccatggaca ccgcatcagc gtggctcctc gcggctcaga agcaggatgg cagctggctc 1200

gacaagtggc acgcctcccc ctactacgcc accgtctgct gcacccaggc actggcagcc 1260

cacgcttccc ctgccaccgc ccccgcacgg cagcgtgctg tgcggtgggt gctggcaaca 1320

caacgctcgg acggcggctg gggcctgtgg cactccacgg tcgaggagac cgcctacgcc 1380

ctgcagatcc tcgccccacc ttccggcggc gggaacatcc ccgtgcaaca ggcgctcacc 1440

agggggcgcg cccgcctctg cggcgctttg ccgctgactc ccctatggca tgacaaggac 1500

ctgtacacgc cggtacgtgt cgtccgcgcc gcccgtgccg ccgccctgta caccacccgt 1560

gacctgcttc tgccgcccct gtga 1584

<210> 39
<211> 1551
<212> DNA
<213> Bradyrhizobium japonicum

<400> 39
gtgaacgcgc tgtccgaaca tatcctttcc gaattgcgcc gcctgctgag cgaaatgagc 60

gatggcggca gcgtcggtcc gtccgtctac gacacggcgc aggcgctgcg cttccacggc 120

aacgtcaccg gtcggcagga cgcatacgcg tggctcatcg cgcagcaaca ggccgacggc 180

ggatggggaa gcgcggactt cccgctgttc cgccatgcgc ccacgtgggc ggcgttactg 240

gcattgcagc gtgccgatcc tcttcccgga gctgcggacg cagtccagac tgcaacgagg 300

ttcctccagc gccagcccga tccctacgca catgcggtgc cagaagacgc gccgatcggc 360

gcggagctga tcctgccgca gttttgcggt gaggccgcat ggttgctggg tggcgtagcg 420

tttccgcgcc atcctgcgct gttgccattg cggcaagcgt gcctggtcaa gctgggggcg 480

gtggcgatgt tgccgagcgg ccatccgttg ctacactcct gggaagcctg ggggacgtcg 540

ccgaccaccg catgcccgga tgacgacggc agcatcggca tcagtccggc ggccaccgcc 600

gcgtggcgtg cccaggccgt gacacggggg agcacgccgc aggtcgggcg cgccgatgcg 660

tatctgcaga tggcatcgcg ggcgacgcgc agcggcatcg aaggtgtctt tcccaacgtc 720

tggccgatca atgtgttcga gccatgctgg tcgctgtaca ccctgcatct ggccgggctt 780

ttcgcgcatc ccgcgctcgc ggaggcggtt cgcgtgatcg tcgcgcagct cgaggcccgt 840

ctgggcgtgc acggtctggg cccggccttg cacttcgcgg ctgatgcgga cgacaccgcc 900

gttgcgttgt gcgtcctgca ccttgcaggc cgtgacccgg cggtcgatgc gttgcgccat 960

ttcgaaatcg gcgagctgtt cgtcaccttc cccggcgaac gcaatgcctc ggtgtcgacc 1020

aacattcatg ccctgcatgc gttgcgactg ttgggaaagc ccgccgcggg cgccagcgcg 1080

tacgtcgagg ccaatcgcaa cccgcacggt ctatgggaca acgaaaaatg gcacgtttcg 1140

tggctgtatc ccaccgcgca tgcggtcgct gcgctggcgc aaggcaagcc ccagtggcga 1200

gatgagcgcg cgctggcggc gctgctgcag gcgcagcgcg acgacggtgg ctggggcgcg 1260

ggtcgcgggt ccacgttcga ggaaaccgcc tatgcgctgt ttgcgttgca cgtgatggat 1320

gggagcgaag aggcgacagg gcgccggcgc atcgcgcagg tggtggcgcg tgcgctggag 1380

tggatgctcg cccgccatgc ggcgcatgga ttgccgcaga cgccgctgtg gatcggcaag 1440

gaactgtatt gccccactcg ggtcgtgcgc gtggccgaac tcgccgggtt gtggctggcg 1500

cttcgttggg ggcggcgcgt cctggccgag ggggcaggag cggcgccatg a 1551

<210> 40
<211> 2355
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 40
atgaatttga gtttgtgtat agcatctcca ctattgacca aatctaatag accagctgct 60

ttatcagcaa ttcatacagc tagtacatcc catggtggcc aaaccaaccc tacgaatctg 120

ataatcgata cgaccaagga gagaatacaa aaacaattca aaaatgttga aatttcagtt 180

tcttcttatg atactgcgtg ggttgccatg gttccatcac ctaattctcc aaagtctcca 240

tgtttcccag aatgtttgaa ttggctgatt aacaaccagt tgaatgatgg atcttggggt 300

ttagtcaatc acacgcacaa tcacaaccat ccacttttga aagattcttt atcctcaact 360

ttggcttgca tcgtggccct aaagagatgg aacgtaggtg aggatcagat taacaagggg 420

cttagtttca ttgaatctaa cttggcttcc gcgactgaaa aatctcaacc atctccaata 480

ggattcgata tcatctttcc aggtctgtta gagtacgcca aaaatctaga tatcaactta 540

ctgtctaagc aaactgattt ctcactaatg ttacacaaga gagaattaga acaaaagaga 600

tgtcattcaa acgaaatgga tggttaccta gcttatatct ctgaaggtct tggtaatctt 660

tacgattgga atatggtgaa aaagtaccag atgaaaaatg gctcagtttt caattcccct 720

tctgcaactg cggcagcatt cattaaccat caaaatccag gatgcctgaa ctatttgaat 780

tcactactag acaaattcgg caacgcagtt ccaactgtat accctcacga tttgtttatc 840

agattgagta tggtggatac aattgaaaga cttggtatat cccaccactt tagagtcgag 900

atcaaaaatg ttttggatga gacataccgt tgttgggtgg agagagatga acaaatcttt 960

atggatgttg tgacgtgcgc gttggccttt agattgttgc gtattaacgg ttacgaagtt 1020

agtccagatc cacttgccga aattacaaac gaattagctt taaaggatga atacgccgct 1080

cttgaaacat atcatgcgtc acatatcctt taccaagagg acttatcatc tggaaaacaa 1140

attcttaaat ctgctgattt cctgaaggaa atcatatcca ctgatagtaa tagactgtcc 1200

aaactgatcc ataaagaggt tgaaaatgca cttaagttcc ctattaacac cggcttagaa 1260

cgtattaaca caagacgtaa catccagctt tacaacgtag acaatactag aatcttgaaa 1320

accacttacc attcttccaa catatcaaac actgattacc taagattagc tgttgaagat 1380

ttctacacat gtcagtctat ctatagagaa gagctgaaag gattagagag atgggtcgtt 1440

gagaataagc tagatcaatt gaaatttgcc agacaaaaga cagcttattg ttacttctca 1500

gttgccgcca ctttatcaag tccagaattg tcagatgcac gtatttcttg ggctaaaaac 1560

ggaattttga caactgttgt tgatgatttc tttgatattg gcgggacaat cgacgaattg 1620

acaaacctga ttcaatgcgt tgaaaagtgg aatgtcgatg tcgataaaga ctgttgctca 1680

gaacatgtta gaatactgtt cttggctctg aaagatgcta tctgttggat cggggatgag 1740

gctttcaaat ggcaagctag agatgtgacg tctcacgtca ttcaaacctg gctagaactg 1800

atgaactcta tgttgagaga agcaatttgg actagagatg catacgttcc tacattaaac 1860

gagtatatgg aaaacgctta tgtctccttt gctttgggtc ctatcgttaa gcctgccata 1920

tactttgtag gaccaaagct atccgaggaa atcgtcgaat catcagaata ccataacttg 1980

ttcaagttaa tgtccacaca aggcagatta cttaatgata ttcattcttt caaaagagag 2040

tttaaggaag gaaagttaaa tgctgttgct ctgcatcttt ctaatggcga aagtggtaaa 2100

gtcgaagagg aagtagttga ggaaatgatg atgatgatca aaaacaagag aaaggagttg 2160

atgaaactaa tcttcgaaga gaacggttca attgttccta gagcatgtaa ggatgcattt 2220

tggaacatgt gtcatgtgct aaacttttтc tacgcaaacg acgatggttt tactgggaac 2280

acaatactag atacagtaaa agacatcata tacaaccctt tggtcttagt aaacgaaaac 2340

gaggagcaaa gataa 2355

<210> 41
<211> 2355
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 41
atgaatctgt ccctttgtat agctagtcca ctgttgacaa aatcttctag accaactgct 60

ctttctgcaa ttcatactgc cagtactagt catggaggtc aaacaaaccc aacaaatttg 120

ataatcgata ctactaagga gagaatccaa aagctattca aaaatgttga aatctcagta 180

tcatcttatg acaccgcatg ggttgcaatg gtgccatcac ctaattcccc aaaaagtcca 240

tgttttccag agtgcttgaa ttggttaatc aataatcagt taaacgatgg ttcttggggt 300
ttagtcaacc acactcataa ccacaatcat ccattattga aggactcttt atcatcaaca 360
ttagcctgta ttgttgcatt gaaaagatgg aatgtaggtg aagatcaaat caacaagggt 420
ttatcattca tagaatccaa tctagcttct gctaccgaca aatcacaacc atctccaatc 480
gggttcgaca taatcttccc tggtttgctg gagtatgcca aaaaccttga tatcaactta 540
ctgtctaaac aaacagattt ctctttgatg ctacacaaaa gagagttaga gcagaaaaga 600
tgccattcta acgaaattga cgggtactta gcatatatct cagaaggttt gggtaatttg 660
tatgactgga acatggtcaa aaagtatcag atgaaaaatg gatccgtatt caattctcct 720
tctgcaactg ccgcagcatt cattaatcat caaaaccctg ggtgtcttaa ctacttgaac 780
tcactattag ataagtttgg aaatgcagtt ccaacagtct atcctttgga cttgtacatc 840
agattatcta tggttgacac tatagagaga ttaggtattt ctcatcattt cagagttgag 900
atcaaaaatg ttttggacga gacatacaga tgttgggtcg aaagagatga gcaaatcttt 960
atggatgtcg tgacctgcgc tctggctttt agattgctaa ggatacacgg atacaaagta 1020
tctcctgatc aactggctga gattacaaac gaactggctt tcaaagacga atacgccgca 1080
ttagaaacat accatgcatc ccaaatactt taccaggaag acctaagttc aggaaaacaa 1140
atcttgaagt ctgcagattt cctgaaaggc attctgtcta cagatagtaa taggttgtct 1200
aaattgatac acaaggaagt agaaaacgca ctaaagtttc ctattaacac tggtttagag 1260
agaatcaata ctaggagaaa cattcagctg tacaacgtag ataatacaag gattcttaag 1320
accacctacc atagttcaaa catttccaac acctattact taagattagc tgtcgaagac 1380
ttttacactt gtcaatcaat ctacagagag gagttaaagg gcctagaaag atgggtagtt 1440
caaaacaagt tggatcaact gaagtttgct agacagaaga cagcatactg ttatttctct 1500
gttgctgcta ccctttcatc cccagaattg tctgatgcca gaataagttg ggccaaaaat 1560
ggtattctta caactgtagt cgatgatttc tttgatattg gaggtactat tgatgaactg 1620
acaaatctta ttcaatgtgt tgaaaagtgg aacgtggatg tagataagga ttgctgcagt 1680
gaacatgtga gaatactttt cctggctcta aaagatgcaa tatgttggat tggcgacgag 1740
gccttcaagt ggcaagctag agatgttaca tctcatgtca tccaaacttg gcttgaactg 1800
atgaactcaa tgctaagaga agcaatctgg acaagagatg catacgttcc aacattgaac 1860
gaatacatgg aaaacgctta cgtctcattt gccttgggtc ctattgttaa gccagccata 1920
tactttgttg ggccaaagtt atccgaagag attgttgagt cttccgaata tcataaccta 1980
ttcaagttaa tgtcaacaca aggcagactt ctgaacgata tccactcctt caaaagagaa 2040
ttcaaggaag gtaagctaaa cgctgttgct ttgcacttgt ctaatggtga atctggcaaa 2100

gtggaagagg aagtcgttga ggaaatgatg atgatgatca aaaacaagag aaaggaattg 2160

atgaaattga ttttcgagga aaatggttca atcgtaccta gagcttgtaa agatgctttt 2220

tggaatatgt gccatgttct taacttcttt tacgctaatg atgatggctt cactggaaat 2280

acaatattgg atacagttaa agatatcatc tacaacccac ttgttttggt caatgagaac 2340

gaggaacaaa gataa 2355

<210> 42
<211> 1773
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 42
atggctatgc cagtgaagct aacacctgcg tcattatcct taaaagctgt gtgctgcaga 60

ttctcatccg gtggccatgc tttgagattc gggagtagtc tgccatgttg gagaaggacc 120

cctacccaaa gatctacttc ttcctctact actagaccag ctgccgaagt gtcatcaggt 180

aagagtaaac aacatgatca ggaagctagt gaagcgacta tcagacaaca attacaactt 240

gtggatgtcc tggagaatat gggaatatcc agacattttg ctgcagagat aaagtgcata 300

ctagacagaa cttacagatc ttggttacaa agacacgagg aaatcatgct ggacactatg 360

acatgtgcta tggcttttag aatcctaaga ttgaacggat acaacgtttc atcagatgaa 420

ctataccacg ttgtagaggc atctggtctg cataattctt tgggtgggta tcttaacgat 480

accagaacac tacttgaatt acacaaggct tcaacagtta gtatctctga ggatgaatct 540

atcttagatt caattggctc tagatccaga acattgctta gagaacaatt ggagtctggt 600

ggcgcactga gaaagccttc tttattcaaa gaggttgaac atgcactgga tggacctttt 660

tacaccacac ttgatagact tcatcatagg tggaatattg aaaacttcaa cattattgag 720

caacacatgt tggagactcc atacttatct aaccagcata catcaaggga tatcctagca 780

ttgtcaatta gagatttttc ctcctcacaa ttcacttatc aacaagagct acagcatctg 840

gagagttggg ttaaggaatg tagattagat caactacagt tcgcaagaca gaaattagcg 900

tacttttacc tatcagccgc aggcaccatg ttttctcctg agctttctga tgcgagaaca 960

ttatgggcca aaaacggggt gttgacaact attgttgatg atttctttga tgttgccggt 1020

tctaaagagg aattggaaaa cttagtcatg ctggtcgaaa tgtgggatga acatcacaaa 1080

gttgaattct attctgagca ggtcgaaatc atcttctctt ccatctacga ttctgtcaac 1140

caattgggtg agaaggcctc tttggttcaa gacagatcaa ttacaaaaca ccttgttgaa 1200

atatggttag acttgttaaa gtccatgatg acggaagttg aatggagact gtcaaaatac 1260

gtgcctacag aaaaggaata catgattaat gcctctctta tcttcggcct aggtccaatc 1320

gttttaccag ctttgtattt cgttggtcca aagatttcag aaagtatagt aaaggaccca 1380

gaatatgatg aattgttcaa actaatgtca acatgtggta gattgttgaa tgacgtgcaa 1440

acgttcgaaa gagaatacaa tgagggtaaa ctgaattctg tcagtctatt ggttcttcac 1500

ggaggcccaa tgtctatttc agacgcaaag aggaaattac aaaagcctat tgatacgtgt 1560

agaagagatc ttctttcttt ggtccttaga gaagagtctg tagtaccaag accatgtaag 1620

gaactattct ggaaaatgtg taaagtgtgc tatttctttt actcaacaac tgatgggttt 1680

tctagtcaag tcgaaagagc aaaagaggta gacgctgtca taaatgagcc actgaagttg 1740

caaggttctc atacactggt atctgatgtt taa 1773

<210> 43
<211> 2232
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 43
atgcagaact tccatggtac aaaggaaagg atcaaaaaga tgtttgacaa gattgaattg 60

tccgtttctt cttatgatac agcctgggtt gcaatggtcc catcccctga ttgcccagaa 120

acaccttgtt ttccagaatg tactaaatgg atcctagaaa atcagttggg tgatggtagt 180

tggtcacttc ctcatggcaa tccacttcta gttaaagatg cattatcttc cactcttgct 240

tgtattctgg ctcttaaaag atggggaatc ggtgaggaac agattaacaa aggactgaga 300

ttcatagaac tcaactctgc tagtgtaacc gataacgaac aacacaaacc aattggattt 360

gacattatct ttccaggtat gattgaatac gctatagact tagacctgaa tctaccacta 420

aaaccaactg acattaactc catgttgcat cgtagagccc ttgaattgac atcaggtgga 480

ggcaaaaatc tagaaggtag aagagcttac ttggcctacg tctctgaagg aatcggtaag 540

ctgcaagatt gggaaatggc tatgaaatac caacgtaaaa acggatctct gttcaatagt 600

ccatcaacaa ctgcagctgc attcatccat atacaagatg ctgaatgcct ccactatatt 660

cgttctcttc tccagaaatt tggaaacgca gtccctacaa tataccctct cgatatctat 720

gccagacttt caatggtaga tgccctggaa cgtcttggta ttgatagaca tttcagaaag 780

gagagaaagt tcgttctgga tgaaacatac agattttggt tgcaaggaga agaggagatt 840

ttctccgata acgcaacctg tgctttggcc ttcagaatat tgagacttaa tggttacgat 900

gtctctcttg aagatcactt ctctaactct ctgggcggtt acttaaagga ctcaggagca 960

gctttagaac tgtacagagc cctccaattg tcttacccag acgagtccct cctggaaaag 1020

caaaattcta gaacttctta cttcttaaaa caaggtttat ccaatgtctc cctctgtggt 1080

gacagattgc gtaaaaacat aattggagag gtgcatgatg ctttaaactt ttccgaccac 1140

gctaacttac aaagattagc tattcgtaga aggattaagc attacgctac tgacgataca 1200

aggattctaa aaacttccta cagatgctca acaatcggta accaagattt tctaaaactt 1260

gcagtggaag atttcaatat ctgtcaatca atacaaagag aggaattcaa gcatattgaa 1320

agatgggtcg ttgaaagacg tctagacaag ttaaagttcg ctagacaaaa agaggcctat 1380

tgctatttct cagccgcagc aacattgttt gcccctgaat tgtctgatgc tagaatgtct 1440

tgggccaaaa atggtgtatt gacaactgtg gttgatgatt tcttcgatgt cggaggctct 1500

gaagaggaat tagttaactt gatagaattg atcgagcgtt gggatgtgaa tggcagtgca 1560

gatttttgta gtgaggaagt tgagattatc tattctgcta tccactcaac tatctctgaa 1620

ataggtgata agtcatttgg ctggcaaggt agagatgtaa agtctcaagt tatcaagatc 1680

tggctggact tattgaaatc aatgttaact gaagctcaat ggtcttcaaa caagtctgtt 1740

cctaccctag atgagtatat gacaaccgcc catgtttcat tcgcacttgg tccaattgta 1800

cttccagcct tatacttcgt tggcccaaag ttgtcagaag aggttgcagg tcatcctgaa 1860

ctactaaacc tctacaaagt cacatctact tgtggcagac tactgaatga ttggagaagt 1920

tttaagagag aatccgagga aggtaagctc aacgctatta gtttatacat gatccactcc 1980

ggtggtgctt ctacagaaga ggaaacaatc gaacatttca aaggtttgat tgattctcag 2040

agaaggcaac tgttacaatt ggtgttgcaa gagaaggata gtatcatacc tagaccatgt 2100

aaagatctat tttggaatat gattaagtta ttacacactt tctacatgaa agatgatggc 2160

ttcacctcaa atgagatgag gaatgtagtt aaggcaatca ttaacgaacc aatctcactg 2220

gatgaattat ga 2232

<210> 44
<211> 2355
<212> DNA
<213> Stevia rebaudiana

<400> 44
atgaatcttt cactatgcat cgcgtcccct ttgttaacca aatcaaatcg acccgcggct 60

ctgtcagcta ttcatacagc atcaacttca catggtggac aaactaatcc cactaatctg 120

atcattgata caaccaaaga acggatccaa aaacagttta aaaatgtaga aatttctgtt 180

tcttcatatg acacagcatg ggtagccatg gtcccttctc caaactcacc caaatcgcct 240

tgtttccctg agtgtctcaa ttggttaatt aataatcagc ttaatgatgg ttcatggggt 300

cttgttaatc acactcataa tcataatcac ccgttgctta aagattctct atcttcaaca 360

ttagcatgta ttgttgcatt aaaaagatgg aatgttgggg aagatcaaat aaataaaggt 420

ctaagtttta ttgagtcaaa tcttgcttca gctactgaaa aaagtcaacc atctcccatt 480

ggttttgaca tcatatttcc tggtttgctt gagtatgcga aaaacttgga cataaacctc 540

ctttcaaaac aaacagattt tagtttgatg ctacataaga gggaattgga gcaaaaaaga 600

tgccattcaa atgagatgga tggatacttg gcgtatatct ctgaaggact cggtaattta 660

tatgattgga atatggtgaa gaaatatcag atgaaaaatg gttctgtttt caactcacca 720

tcagcaacag ctgctgcttt cattaatcat caaaatcctg gttgtcttaa ttatttaaat 780

tcacttttgg acaagtttgg taatgcagtc ccaacagttt atcctcatga tttatttatc 840

cgactttcta tggttgacac aattgaaaga ttaggaattt cacaccattt cagagtggaa 900

attaaaaatg ttttagatga aacatacaga tgttgggtgg aacgagatga gcaaatattc 960

atggatgttg taacatgtgc tttagccttt cggttattaa ggatcaatgg gtatgaagtt 1020

tccccagatc cattggctga aattactaat gaattagctt tgaaagacga atatgcagct 1080

cttgaaacat atcatgcgtc acatatatta taccaagagg atttatcttc tggaaaacaa 1140

atcttgaagt cagctgattt cctcaaagag ataatatcca ctgattcaaa caggctttct 1200

aaattaattc acaaagaggt ggaaaatgct cttaagttcc ctatcaatac cggtttagaa 1260

cgcataaaca ctagacgaaa tatacagctt tacaatgtag acaatacaag aattctgaaa 1320

actacatatc actcatcaaa tattagtaac actgattacc taaggttggc tgttgaagat 1380

ttctacacct gccaatctat ttatcgtgaa gaattaaaag gtcttgaaag gtgggtggta 1440

gagaataagt tggaccagct caagtttgct aggcaaaaga ccgcctactg ttatttctct 1500

gttgctgcaa cactttcgtc tcccgaatta tcagatgcgc gtatttcatg ggccaaaaat 1560

ggcatattaa ctacagtagt tgatgacttt tttgatatcg gtggtacaat cgatgaattg 1620

accaacctga ttcaatgtgt tgaaaaatgg aatgtagatg tcgacaagga ttgttgttca 1680

gagcatgttc ggattttatt tttagcatta aaagatgcaa tctgttggat tggagatgaa 1740

gcttttaaat ggcaagcgcg cgatgtaact agccatgtta ttcaaacttg gttggaacta 1800

atgaatagta tgttgagaga agctatatgg acaagagatg cttatgtgcc aacattaaat 1860

gaatatatgg aaaacgctta cgtgtcattt gcattaggcc cgattgtcaa gccggctatt 1920

tactttgtgg ggcccaaatt atcagaggag attgttgaaa gctctgaata tcataatcta 1980

tttaagctaa tgagcacgca gggtcgactt ctaaacgata tccatagctt caagagggaa 2040

tttaaggaag gcaaattaaa cgcggtagca ttgcatttga gtaacggaga aagtgggaaa 2100

gtggaagaag aggttgtgga ggagatgatg atgatgatta aaaacaagag gaaagaatta 2160

atgaaattaa tttttgaaga aaatggtagc attgttccta gagcttgtaa agatgcattt 2220

tggaacatgt gtcacgtgtt gaatttttt tacgcaaacg atgacgggtt tactggaaac 2280

acgattcttg atactgtgaa ggacatcatt tacaacccgt tggtgcttgt gaatgaaaat 2340

gaagaacaaa ggtaa 2355

<210> 45
<211> 2355
<212> DNA
<213> Stevia rebaudiana

<400> 45
atgaatcttt cactatgcat tgcgtcccct ttgttaacca aatcaagtcg acccacggct 60

ctgtcagcta ttcatacagc atcaacttca catggtggac aaactaatcc cactaatctg 120

atcattgata caaccaaaga acggatccaa aaactgttta aaaatgtaga aatttctgtt 180

tcttcatatg acacagcatg ggtagccatg gtcccttctc caaactcacc caaatcgcct 240

tgtttccctg agtgtctcaa ttggttaatt aataatcagc ttaatgatgg ttcatggggt 300

cttgttaatc acactcataa tcataatcac ccgttgctta aagattctct atcttcaaca 360

ttagcatgta ttgttgcatt aaaaagatgg aatgttgggg aagatcaaat aaataaaggt 420

ctaagtttta ttgagtcaaa tcttgcttca gcaactgaca aaagtcaacc atctcccatt 480

ggttttgata tcatatttcc tggtttgctt gagtatgcga aaaacttgga cataaacctc 540

ctttcaaaac aaacagattt tagtttgatg ctacataaga gggaattgga gcaaaaaaga 600

tgccattcaa atgagattga tggatacttg gcgtatatct ctgaaggact cggtaattta 660

tatgattgga atatggtgaa gaaatatcag atgaaaaatg gttctgtttt caactcacca 720

tcagcaacag cagctgcttt cattaatcat caaaatcccg gttgtcttaa ttatttaaat 780

tcacttttgg acaagtttgg taatgcagtc ccaacagttt atcctcttga tttatatatc 840

cggctttcta tggttgacac aattgaaaga ttaggaattt cacaccattt cagagtggaa 900

attaaaaatg ttttagatga aacatacaga tgttgggtgg aacgagatga gcaaatattc 960

atggatgttg taacatgtgc tttagccttt cggttattaa ggatccacgg gtataaagtc 1020

tccccagatc aattggctga aattactaat gaattagctt tcaaagacga atacgcagct 1080

cttgaaacat atcatgcatc acagatatta taccaagagg atttatcttc tggaaaacaa 1140

atcttgaagt cagctgattt cctcaaaggg atattatcca ctgattcaaa caggctttct 1200

aaattaattc acaaagaggt ggaaaatgct cttaagttcc ctatcaatac cggtttagaa 1260

cgcataaaca ctagacgaaa tatacagctt tacaatgtag acaatacaag aattctgaaa 1320

actacatatc actcatcaaa tattagtaac acttattacc taaggttggc tgttgaagat 1380

ttctacacct gccaatctat ttatcgtgaa gaattaaaag gtcttgaaag gtgggtggta 1440

cagaataagt tggaccagct caagtttgct aggcaaaaga ccgcctactg ttatttctct 1500

gttgctgcaa cactttcgtc tcccgaatta tcagatgcgc gtatttcatg ggccaaaaat 1560

ggcatattaa ctacagtagt tgatgacttt tttgatatcg gtggtacaat cgatgaattg 1620

accaacctga ttcaatgtgt tgaaaaatgg aatgtagatg tcgacaagga ttgttgttca 1680

gagcatgttc ggattttatt tttagcatta aaagatgcaa tctgttggat tggagatgaa 1740

gcttttaaat ggcaagcgcg cgatgtaact agccatgtta ttcaaacttg gttggaacta 1800

atgaatagta tgttgagaga agctatatgg acaagagatg cttatgtgcc aacattaaat 1860

gaatatatgg aaaacgctta cgtgtcattt gcattaggcc cgattgtcaa gccggctatt 1920

tactttgtgg ggcccaaatt atcagaggag attgttgaaa gctctgaata tcataatcta 1980

tttaagctaa tgagcacgca gggtcgactt ctaaacgata tccatagctt caagagggaa 2040

tttaaggaag gcaaattaaa cgcggtagca ttgcatttga gtaacggaga aagtgggaaa 2100

gtggaagaag aggttgtgga ggagatgatg atgatgatta aaaacaagag gaaagaatta 2160

atgaaattaa tttttgaaga aaatggtagc attgttccta gagcttgtaa agatgcattt 2220

tggaacatgt gtcacgtgtt gaatttttt tacgcaaacg atgacgggtt tactggaaac 2280

acgattcttg atactgtgaa ggacatcatt tacaacccgt tggtgcttgt gaatgaaaat 2340

gaagaacaaa ggtaa 2355

<210> 46
<211> 1773
<212> DNA
<213> Zea mays

<400> 46
atggccatgc cagtgaagct gactcctgcc tccctctcgc tgaaggcggt ctgctgccgc 60

ttcagctccg gagggcatgc gctgcgcttc ggctcgtcgc taccgtgctg gaggaggacg 120

ccgacgcaac ggagcacgtc gtcgtctacg acgcgccctg cggctgaggt tagctctggc 180

aaaagcaagc agcacgatca agaagcatcg gaggctacga taagacagca gctccagcta 240

gtcgatgtgc ttgagaacat ggggatttct cggcattttg ctgctgaaat caaatgcatc 300

cttgacagga catacagaag ttggttacag agacatgagg aaattatgct ggacacaatg 360

acctgtgcga tggcatttcg tattctaagg ttgaatggat acaatgtctc ttctgatgag 420

ttgtatcatg ttgttgaagc ttccggactc cataattcac ttggaggata tctcaatgat 480

acaagaacct tgttagaatt acacaaggcc tcgacagtta gtatctctga agatgagtct 540

atcctggata gcataggctc aaggtcacgt accttactga gggaacaact agagtctggt 600

ggtgctctac gaaaaccttc actctttaaa gaggtggaac atgctctgga cggtcccttc 660

tacaccacat tggaccgtct acaccatagg tggaacatcg aaaatttcaa tattatagag 720

cagcacatgc tagagacacc atacttgtca aatcaacata ccagtagaga tattctagcg 780

ttgagtatta gagacttcag ttcctctcag tttacttacc agcaagaact tcaacatctt 840

gaaagctggg tgaaagagtg caggttagac cagctacaat ttgcgcgaca gaagttggca 900

tacttctact tgtctgctgc tggcaccatg ttctctcctg agctgtctga tgctcgaact 960

ttgtgggcca aaaatggtgt gctcacaact attgttgacg acttctttga tgttgcggga 1020

tcaaaagaag aacttgaaaa ccttgtcatg ttggttgaga tgtgggacga gcatcacaaa 1080

gttgagttct actcagaaca agtagagatt atattttctt caatttatga ctcagttaac 1140

caacttggtg aaaaggcttc tttggtacaa gaccgcagta ttaccaaaca cctagtagaa 1200

atatggttgg atttgctaaa gtctatgatg acagaggtag agtggcgttt gagcaaatat 1260

gtgccaacag agaaggaata catgataaat gcatctttaa tatttggact aggccccatt 1320

gtattgccag cattatattt tgttgggcca aagatctcag agtctattgt taaagatcca 1380

gaatatgatg aattgttcaa actgatgagc acatgtggtc gcctcttgaa tgatgttcag 1440

acttttgaga gggagtacaa cgagggcaag ttgaatagtg tttctctcct cgttcttcat 1500

ggtggcccca tgtccatatc agacgccaaa aggaaattac agaagcccat agacacatgc 1560

agaagagacc tcctaagttt agttcttcgt gaagaaagtg ttgttcctag gccctgcaag 1620

gaattatttt ggaaaatgtg caaggtgtgc tacttcttct actcgacgac ggatgggttt 1680

agctcacaag tggagagggc taaagaagtg gatgcggtga tcaatgagcc actaaagcta 1740

caaggaagtc atacgctggt gtctgatgtg tga 1773

<210> 47
<211> 2232
<212> DNA
<213> Populus trichocarpa

<400> 47
atgcagaact ttcatggaac taaggaaagg atcaagaaga tgtttgataa gattgaattg 60

tcagtgtctt catatgacac tgcttgggtg gcaatggtcc catctccaga ttgtccggaa 120

actccttgtt ttccagagtg cacaaaatgg attttggaaa atcaacttgg tgatggctcc 180

tggagtcttc ctcatggcaa tccattatta gttaaggatg ctctttcatc tacattagcg 240

tgcatccttg cattgaagcg atggggtatc ggtgaagaac aaataaataa aggccttcga 300

tttattgagt tgaattccgc ttcagttacg gataacgagc aacataaacc aattggattt 360

gatataatat ttcctggcat gattgaatat gccatagatt tggatttgaa cctccctttg 420

aagccgacag atataaattc catgctccac aggagggctt tggagcttac aagtggcggt 480

ggcaagaact tggagggaag aagagcctac ttagcatatg tttcggaagg aattggaaaa 540

ttacaggatt gggaaatggc catgaaatat caaagaaaga atggatcact gttcaattca 600

ccatccacca cagcagctgc ctttattcat attcaagatg ctgagtgtct ccattatatt 660

cgttcactct tacagaagtt tgggaatgca gttccaacca tttatccttt ggatatatat 720

gctcgtcttt ctatggttga tgctcttgaa aggttgggaa tcgatcggca ttttaggaag 780

gaaagaaaat ttgttttgga cgaaacatac cgattttggt tgcaggggga ggaagagata 840

ttttctgata atgccacttg tgctttggca tttaggatat tacgtttgaa cggatatgat 900

gtctctctag aagatcattt ctctaattca ctgggaggat atttgaagga ttcgggagct 960

gccttagagt tgtacagagc tctgcagcta agttatccag atgaatcact tctggaaaaa 1020

caaaattctc ggacaagcta tttcctgaaa cagggattat ccaacgtttc actttgtgga 1080

gataggcttc gtaaaaatat tatcggagag gtgcatgatg ctctcaattt ttctgaccat 1140

gcaaatttgc aacgcttagc tatcagaaga agaattaaac attatgctac agatgatacg 1200

aggattttga aaacttcgta tcgttgttcg actattggta accaggattt tctcaaattg 1260

gctgtagaag acttcaatat ctgtcaatca atacagcgtg aagaatttaa acatatcgag 1320

aggtgggttg tagagaggag actggacaag ctaaagtttg ctaggcagaa ggaggcctac 1380

tgttacttct ctgctgcagc aactctcttc gctccagaac tatctgatgc acgcatgtca 1440

tgggcaaaaa atggtgtgct tactactgtt gttgatgact tctttgatgt tggtggttct 1500

gaagaagaac tggtaaacct tattgaattg attgagaggt gggatgtcaa tggcagtgct 1560

gatttttgtt ctgaggaagt tgagatcata tattcggcaa ttcacagcac tataagtgag 1620

ataggagaca aatctttcgg atggcaagga cgcgatgtga aaagtcaggt tatcaagatt 1680

tggttggatt tgctcaaatc catgttgaca gaagcacaat ggtcaagtaa caaatcagtg 1740

ccgacccttg atgaatatat gacaactgca catgtatcgt tcgctctagg gcctattgtt 1800

cttccagctc tgtattttgt ggggcctaag ctttcagagg aggttgctgg acatcctgaa 1860

ttgcttaatc tatacaaggt tacgagcact tgcgggcgtc tgctcaatga ctggagaagc 1920

tttaagagag aatctgaaga agggaaattg aatgccatct cattgtacat gattcacagc 1980

ggtggtgctt caactgaaga agagaccatc gaacatttta aaggattgat cgacagccag 2040

agaagacaat tgcttcaatt agttttgcag gaaaaggata gtataattcc tagaccctgc 2100

aaggatttgt tttggaacat gataaaatta ttgcacacgt tctacatgaa ggatgatgga 2160

ttcacttcaa acgagatgag aaatgttgtc aaggcaataa taaatgaacc catctctcta 2220

gatgaattat aa 2232

<210> 48
<211> 2952
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 48
atggaatttg atgaaccatt ggttgacgaa gcaagatctt tagtgcagcg tactttacaa 60

gattatgatg acagatacgg cttcggtact atgtcatgtg ctgcttatga tacagcctgg 120

gtgtctttag ttacaaaaac agtcgatggg agaaaacaat ggcttttccc agagtgtttt 180

gaatttctac tagaaacaca atctgatgcc ggaggatggg aaatcgggaa ttcagcacca 240

atcgacggta tattgaatac agctgcatcc ttacttgctc taaaacgtca cgttcaaact 300

gagcaaatca tccaacctca acatgaccat aaggatctag caggtagagc tgaacgtgcc 360

gctgcatctt tgagagcaca attggctgca ttggatgtgt ctacaactga acacgtcggt 420

tttgagataa ttgttcctgc aatgctagac ccattagaag ccgaagatcc atctctagtt 480

ttcgattttc cagctaggaa acctttgatg aagattcatg atgctaagat gagtagattc 540

aggccagaat acttgtatgg caaacaacca atgaccgcct tacattcatt agaggctttc 600

ataggcaaaa tcgacttcga taaggtaaga caccaccgta cccatgggtc tatgatgggt 660

tctccttcat ctaccgcagc ctacttaatg cacgcttcac aatgggatgg tgactcagag 720

gcttacctta gacacgtgat taaacacgca gcagggcagg gaactggtgc tgtaccatct 780

gctttcccat caacacattt tgagtcatct tggattctta ccacattgtt tagagctgga 840

ttttcagctt ctcatcttgc ctgtgatgag ttgaacaagt tggtcgagat acttgagggc 900

tcattcgaga aggaaggtgg ggcaatcggt tacgctccag ggtttcaagc agatgttgat 960

gatactgcta aaacaataag tacattagca gtccttggaa gagatgctac accaagacaa 1020

atgatcaagg tatttgaagc taatacacat tttagaacat accctggtga aagagatcct 1080

tctttgacag ctaattgtaa tgctctatca gccttactac accaaccaga tgcagcaatg 1140

tatggatctc aaattcaaaa gattaccaaa tttgtctgtg actattggtg gaagtctgat 1200

ggtaagatta aagataagtg gaacacttgc tacttgtacc catctgtctt attagttgag 1260

gttttggttg atcttgttag tttattggag cagggtaaat tgcctgatgt tttggatcaa 1320

gagcttcaat acagagtcgc catcacattg ttccaagcat gtttaaggcc attactagac 1380

caagatgccg aaggatcatg gaacaagtct atcgaagcca cagcctacgg catccttatc 1440

ctaactgaag ctaggagagt ttgtttcttc gacagattgt ctgagccatt gaatgaggca 1500

atccgtagag gtatcgcttt cgccgactct atgtctggaa ctgaagctca gttgaactac 1560

atttggatcg aaaaggttag ttacgcacct gcattattga ctaaatccta tttgttagca 1620

gcaagatggg ctgctaagtc tcctttaggc gcttccgtag gctcttcttt gtggactcca 1680

ccaagagaag gattggataa gcatgtcaga ttattccatc aagctgagtt attcagatcc 1740

cttccagaat gggaattaag agcctccatg attgaagcag ctttgttcac accacttcta 1800

agagcacata gactagacgt tttccctaga caagatgtag gtgaagacaa atatcttgat 1860

gtagttccat tcttttggac tgccgctaac aacagagata gaacttacgc ttccactcta 1920

ttcctttacg atatgtgttt tatcgcaatg ttaaacttcc agttagacga attcatggag 1980

gccacagccg gtatcttatt cagagatcat atggatgatt tgaggcaatt gattcatgat 2040

cttttggcag agaaaacttc cccaaagagt tctggtagaa gtagtcaggg cacaaaagat 2100

gctgactcag gtatagagga agacgtgtca atgtccgatt cagcttcaga ttcccaggat 2160

agaagtccag aatacgactt ggttttcagt gcattgagta cctttacaaa acatgtcttg 2220

caacacccat ctatacaaag tgcctctgta tgggatagaa aactacttgc tagagagatg 2280

aaggcttact tacttgctca tatccaacaa gcagaagatt caactccatt gtctgaattg 2340

aaagatgtgc ctcaaaagac tgatgtaaca agagtttcta catctactac taccttcttt 2400

aactgggtta gaacaacttc cgcagaccat atatcctgcc catactcctt ccactttgta 2460

gcatgccatc taggcgcagc attgtcacct aaagggtcta acggtgattg ctatccttca 2520

gctggtgaga agttcttggc agctgcagtc tgcagacatt tggccaccat gtgtagaatg 2580

tacaacgatc ttggatcagc tgaacgtgat tctgatgaag gtaatttgaa ctccttggac 2640

ttccctgaat tcgccgattc cgcaggaaac ggagggatag aaattcagaa ggccgctcta 2700

ttaaggttag ctgagtttga gagagattca tacttagagg ccttccgtcg tttacaagat 2760

gaatccaata gagttcacgg tccagccggt ggtgatgaag ccagattgtc cagaaggaga 2820

atggcaatcc ttgaattctt cgcccagcag gtagatttgt acggtcaagt atacgtcatt 2880

agggatattt ccgctcgtat tcctaaaaac gaggttgaga aaaagagaaa attggatgat 2940

gctttcaatt ga 2952

<210> 49
<211> 2646
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 49
atggcttcta gtacacttat ccaaaacaga tcatgtggcg tcacatcatc tatgtcaagt 60

tttcaaatct tcagaggtca accactaaga tttcctggca ctagaacccc agctgcagtt 120

caatgcttga aaaagaggag atgccttagg ccaaccgaat ccgtactaga atcatctcct 180

ggctctggtt catatagaat agtaactggc ccttctggaa ttaaccctag ttctaacggg 240

cacttgcaag agggttcctt gactcacagg ttaccaatac caatggaaaa atctatcgat 300

aacttccaat ctactctata tgtgtcagat atttggtctg aaacactaca gagaactgaa 360

tgtttgctac aagtaactga aaacgtccag atgaatgagt ggattgagga aattagaatg 420

tactttagaa atatgacttt aggtgaaatt tccatgtccc cttacgacac tgcttgggtg 480

gctagagttc cagcgttgga cggttctcat gggcctcaat tccacagatc tttgcaatgg 540
attatcgaca accaattacc agatggggac tggggcgaac cttctctttt cttgggttac 600
gatagagttt gtaatacttt agcctgtgtg attgcgttga aaacatgggg tgttggggca 660
caaaacgttg aaagaggaat tcagttccta caatctaaca tatacaagat ggaggaagat 720
gacgctaatc atatgccaat aggattcgaa atcgtattcc ctgctatgat ggaagatgcc 780
aaagcattag gtttggattt gccatacgat gctactattt tgcaacagat ttcagccgaa 840
agagagaaaa agatgaaaaa gatcccaatg gcaatggtgt acaaataccc aaccacttta 900
cttcactcct tagaaggctt gcatagagaa gttgattgga ataagttgtt acaattacaa 960
tctgaaaatg gtagttttct ttattcacct gcttcaaccg catgcgcctt aatgtacact 1020
aaggacgtta aatgttttga ttacttaaac cagttgttga tcaagttcga ccacgcatgc 1080
ccaaatgtat atccagtcga tctattcgaa agattatgga tggttgacag attgcagaga 1140
ttagggatct ccagatactt tgaaagagag attagagatt gtttacaata cgtctacaga 1200
tattggaaag attgtggaat cggatgggct tctaactctt ccgtacaaga tgttgatgat 1260
acagccatgg cgtttagact tttaaggact catggtttcg acgtaaagga agattgcttt 1320
agacagtttt tcaaggacgg agaattcttc tgcttcgcag gccaatcatc tcaagcagtt 1380
acaggcatgt ttaatctttc aagagccagt caaacattgt ttccaggaga atctttattg 1440
aaaaaggcta gaaccttctc tagaaacttc ttgagaacaa agcatgagaa caacgaatgt 1500
ttcgataaat ggatcattac taaagatttg gctggtgaag tcgagtataa cttgaccttc 1560
ccatggtatg cctctttgcc tagattagaa cataggacat acttagatca atatggaatc 1620
gatgatatct ggataggcaa atctttatac aaaatgcctg ctgttaccaa cgaagttttc 1680
ctaaagttgg caaaggcaga ctttaacatg tgtcaagctc tacacaaaaa ggaattggaa 1740
caagtgataa agtggaacgc gtcctgtcaa ttcagagatc ttgaattcgc cagacaaaaa 1800
tcagtagaat gctattttgc tggtgcagcc acaatgttcg aaccagaaat ggttcaagct 1860
agattagtct gggcaagatg ttgtgtattg acaactgtct tagacgatta ctttgaccac 1920
gggacacctg ttgaggaact tagagtgttt gttcaagctg tcagaacatg gaatccagag 1980
ttgatcaacg gtttgccaga gcaagctaaa atcttgttta tgggcttata caaaacagtt 2040
aacacaattg cagaggaagc attcatggca cagaaaagag acgtccatca tcatttgaaa 2100
cactattggg acaagttgat aacaagtgcc ctaaaggagg ccgaatgggc agagtcaggt 2160
tacgtcccaa catttgatga atacatggaa gtagctgaaa tttctgttgc tctagaacca 2220
attgtctgta gtaccttgtt ctttgcgggt catagactag atgaggatgt tctagatagt 2280
tacgattacc atctagttat gcatttggta aacagagtcg gtagaatctt gaatgatata 2340
caaggcatga agagggaggc ttcacaaggt aagatctcat cagttcaaat ctacatggag 2400

gaacatccat ctgttccatc tgaggccatg gcgatcgctc atcttcaaga gttagttgat 2460

aattcaatgc agcaattgac atacgaagtt cttaggttca ctgcggttcc aaaaagttgt 2520

aagagaatcc acttgaatat ggctaaaatc atgcatgcct tctacaagga tactgatgga 2580

ttctcatccc ttactgcaat gacaggattc gtcaaaaagg ttcttttcga acctgtgcct 2640

gagtaa 2646

<210> 50
<211> 2952
<212> DNA
<213> Phomopsis amygdali

<400> 50
atggagttcg atgaaccact tgtggacgag gcgaggtcct tggtccaaag aaccctgcaa 60

gattatgacg accgctatgg ctttggcact atgagctgtg cggcctatga cacagcatgg 120

gtatcgctgg tgactaaaac agtcgatggg cgtaaacaat ggttgttccc tgagtgcttc 180

gaatttctcc tagaaacgca gtccgatgct ggcggctggg aaatcggcaa cagcgcaccc 240

atcgatggga tccttaacac tgctgcttca ctgctggcat tgaagcgcca cgtccaaaca 300

gagcagatta ttcagccgca acacgaccat aaagacctgg ccgggcgtgc ggaaagagcg 360

gcggcgtctt tgcgagcaca gttggcggct ctggatgtgt cgacaacgga gcatgtgggc 420

ttcgaaatca tcgtcccggc catgctcgac cctctcgagg ccgaagaccc gtctttggtg 480

ttcgactttc cagcacgcaa accactgatg aagatccacg acgctaagat gtcgcgattc 540

cgaccagagt acctctacgg taaacagccg atgacggcat tgcattcgct cgaggccttt 600

atcgggaaaa tagacttcga caaagtacgg catcacagga cacacggttc gatgatgggg 660

tcgccctcgt cgacggctgc atacctgatg catgcttctc agtgggacgg cgactctgag 720

gcctatctac gccatgtcat caagcacgca gctggccagg gcaccggagc tgttccgagt 780

gcatttcctt cgacgcattt cgagtcttct tggattttga caacattgtt tcgagctggg 840

ttctcagcct ctcatctagc atgcgacgaa ttaaacaagc tggtagagat cctcgaaggc 900

tcatttgaga aagaaggggg agccatcggt tatgctcctg ggtttcaagc agatgtggat 960

gataccgcaa agaccatctc cactttggct gtgcttggga gagatgccac tccccggcaa 1020

atgatcaagg tttttgaagc caatacacac tttcggactt accctggtga aagagatcca 1080

agcttgactg ccaattgcaa cgcgctctcg gctcttcttc accagccaga cgcagcaatg 1140

tacggcagcc agatccagaa gatcacaaag tttgtttgtg actactggtg gaaaagtgac 1200

ggcaaaatca aggacaagtg gaatacctgc tacttgtatc catcggtcct cctcgtcgag 1260

gtgttagtag accttgtgtc cctgttggag caaggaaagc tacccgacgt gctggatcag 1320

gagctgcaat acagggtcgc cattacgtta ttccaggcct gcttgcgacc gctacttgat 1380

caagatgctg aaggttcatg gaacaaatcc attgaagcca cagcctacgg cattctaatc 1440

cttacggagg cgcggcgagt atgctttttt gaccgtctga gtgagcctct gaatgaggct 1500

attcgacgcg ggattgcgtt tgcagattcg atgagcggta ctgaagctca gctgaattat 1560

atatggatcg agaaagtgag ctacgcacct gctcttctga ccaaatcata cctcctcgca 1620

gctcggtggg cggcaaagtc cccgcttggc gcttccgttg gatccagcct ttggacgcct 1680

ccaagagaag gcttggataa gcacgtccgt ctattccacc aggcagagct cttcaggtcg 1740

ttgccggagt gggagctgcg cgcgtccatg atcgaggcag ccctgttcac tcctttgctg 1800

cgtgcgcata ggctggatgt atttccacgc caagacgtcg gcgaggacaa gtacctggac 1860

gttgtgccgt tcttctggac ggccgccaat aaccgcgatc gcacgtacgc atccactctg 1920

tttctgtatg acatgtgctt tatcgccatg cttaacttcc agctggatga gttcatggag 1980

gctacagcgg gaatcctctt ccgggaccat atggatgatt tgcgccaact catccacgac 2040

ctgcttgccg aaaagacgag ccccaagtca tcgggcagaa gtagccaagg aaccaaagac 2100

gcggactcgg gcatcgaaga agacgtttct atgagcgact cagcgtcaga ctcccaggac 2160

cgcagccctg aatacgacct ggtcttctct gcgctctcta ccttcaccaa acatgtcctg 2220

cagcaccctt caatccagtc agccagtgtc tgggatagga aactactcgc tcgcgagatg 2280

aaagcatacc tcctagctca tattcaacag gctgaggaca gcacgccctt gagtgagctc 2340

aaggacgtcc ctcaaaaaac tgacgtgaca cgcgtctcaa cgtccacaac gactttcttc 2400

aactgggtac gcacaacatc cgcagaccac atatcctgcc catattcatt ccatttcgtg 2460

gcgtgtcacc tcggcgccgc gctgagcccc aagggcagca acggcgactg ttacccgtca 2520

gccggtgaaa agttcctcgc ggccgccgta tgccgccatt tggccacgat gtgccgcatg 2580

tacaatgact tgggatcggc ggagcgcgac agtgacgagg gaaatttgaa ttcactcgac 2640

tttcccgagt tcgccgactc agcggggaat ggtgggattg agatccagaa agctgccttg 2700

ctcaggctgg ccgagttcga acgcgactcg tatctcgagg ctttccggcg acttcaggat 2760

gaaagcaacc gcgttcacgg accggctggt ggggatgaag ccagactcag caggcggcgc 2820

atggccatcc ttgagttctt tgcccagcag gtggacttgt atggccaggt ctacgttatt 2880

cgcgatatct cggccaggat tccaaagaac gaggttgaga agaaaaggaa actagatgat 2940

gctttcaatt ag 2952

<210> 51
<211> 2646
<212> DNA
<213> Physcomitrella patens

<400> 51
atggcttcca gcaccttgat acagaatcgc tcttgtggcg ttacgtcaag catgtcttcc 60

tttcagattt ttcgagggca acctctacgt tttccaggca ctagaactcc tgctgcagtt 120

caatgcctaa agaagcgtcg atgtttgcga cctactgaat cagtcctcga gagctctcct 180

ggtagcggat cttacaggat tgtaactgga ccctccggca tcaatccttc ttcaaacggc 240

cacttgcaag aggggtccct tactcacaga cttccgatac ccatggaaaa atccattgat 300

aacttccagt ctactttgta cgtatcagac atatggtcag aaaccttgca aagaacggaa 360

tgtttgttgc aggtgactga gaatgtacag atgaacgagt ggattgagga aatcagaatg 420

tacttccgaa atatgacact gggggaaata tccatgtctc catacgacac agcttgggta 480

gcgcgagtgc cagcgctgga tggctcacat ggccctcagt tccatcggtc tttgcagtgg 540

attattgata atcagctccc ggatggcgat tggggtgaac cgtctctttt ccttggatac 600

gatcgcgttt gcaacactct cgcctgtgta attgccctga aaacttgggg tgttggggct 660

cagaacgtag agcgtggaat ccagtttctg caatctaaca tctacaaaat ggaggaagat 720

gacgccaatc atatgccgat tggatttgag attgtcttcc cagcgatgat ggaagatgcc 780

aaggcactgg gactggattt accatacgat gccactatct tgcaacaaat ctcggctgaa 840

agagagaaga aaatgaaaaa gattcctatg gcgatggtgt acaagtaccc cactactttg 900

ctgcattctc tggaaggcct gcaccgggaa gtggactgga acaagctcct ccagctacag 960

tccgagaatg gctcctttct gtattcaccc gcatccactg catgcgcact tatgtacaca 1020

aaagatgtga agtgcttcga ctacttgaac cagctcctca tcaagttcga ccacgcttgt 1080

ccaaacgtgt accccgttga tctcttcgag cgtttgtgga tggtagaccg cctacaaagg 1140

ctgggaatat cccgctactt cgagcgagaa atcagagact gtctacaata tgtataccga 1200

tactggaagg attgtggtat tggctgggca agcaattcgt ccgtgcagga cgtggacgac 1260

acggccatgg ccttccgcct tctccgcaca cacggattcg acgtcaagga ggactgcttc 1320

agacagtttt tcaaagatgg tgagttcttc tgcttcgccg gccagtccag ccaagccgtc 1380

acgggaatgt tcaacctcag cagagcatcg caaacgctct tcccagggga atcactccta 1440

aaaaaggcca gaacctttc cagaaacttt ttgagaacca agcatgaaaa caatgaatgc 1500

ttcgacaagt ggataatcac gaaggatcta gcgggcgagg tggaatacaa tctcacattc 1560

ccctggtatg ctagccttcc tcgtcttgag catcgcacct acttggacca atatgggatt 1620

gatgatatct ggattggcaa gtcgctctac aaaatgccgg ccgtcaccaa cgaagtgttt 1680

ctcaaattgg ccaaagccga cttcaacatg tgccaagctc ttcacaagaa ggaactcgag 1740

caggtcatca aatggaatgc cagctgccaa tttagagacc tcgagtttgc tagacagaaa 1800

tccgtggagt gctacttcgc aggcgctgca accatgtttg agcccgaaat ggtgcaggcg 1860

aggctcgttt gggcacgctg ttgcgtgctc accaccgttc tagacgatta cttcgatcac 1920

ggtacacctg tggaagagct tcgggttttt gtgcaggccg taaggacttg gaatcccgag 1980

ctcatcaacg gactacctga gcaagccaag attctcttta tgggactgta caagactgtg 2040

aacactatcg ccgaggaggc attcatggca cagaaacgag acgtacatca tcatctcaag 2100

cattactggg acaaattgat cacttcagct ttgaaagaag ccgaatgggc agagtccggc 2160

tacgtcccca ccttcgacga gtatatggaa gtcgctgaaa tctccgtcgc actagagccc 2220

attgtatgta gcactctctt cttcgccggc cataggctcg atgaggatgt gcttgacagt 2280

tatgactacc atcttgtcat gcatctcgtc aaccgcgtag gtcgcatcct caacgacatc 2340

caaggaatga agagggaagc cagccaaggg aagatatcga gcgtgcagat ctacatggag 2400

gagcatccaa gtgtgccttc agaggccatg gccatcgctc atctgcagga attggtcgac 2460

aactccatgc aacagctgac atacgaagtg ctgcgcttca ctgcagtccc gaagtcctgt 2520

aagagaatcc atttaaacat ggcgaagatc atgcacgctt tctacaagga cactgatggg 2580

ttttcgtcac tgacagccat gacagggttt gtgaagaagg tgctcttcga gccagtacct 2640

gaatag 2646

<210> 52
<211> 1542
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 52
atggatgctg tgacgggttt gttaactgtc ccagcaaccg ctataactat tggtggaact 60

gctgtagcat tggcggtagc gctaatcttt tggtacctga aatcctacac atcagctaga 120

agatcccaat caaatcatct tccaagagtg cctgaagtcc caggtgttcc attgttagga 180

aatctgttac aattgaagga gaaaaagcca tacatgactt ttacgagatg ggcagcgaca 240

tatggaccta tctatagtat caaaactggg gctacaagta tggttgtggt atcatctaat 300

gagatagcca aggaggcatt ggtgaccaga ttccaatcca tatctacaag gaacttatct 360

aaagccctga aagtacttac agcagataag acaatggtcg caatgtcaga ttatgatgat 420

tatcataaaa cagttaagag acacatactg accgccgtct tgggtcctaa tgcacagaaa 480

aagcatagaa ttcacagaga tatcatgatg gataacatat ctactcaact tcatgaattc 540

gtgaaaaaca acccagaaca ggaagaggta gaccttagaa aaatctttca atctgagtta 600

ttcggcttag ctatgagaca agccttagga aaggatgttg aaagtttgta cgttgaagac 660

ctgaaaatca ctatgaatag agacgaaatc tttcaagtcc ttgttgttga tccaatgatg 720

ggagcaatcg atgttgattg gagagacttc tttccatacc taaagtgggt cccaaacaaa 780

aagttcgaaa atactattca acaaatgtac atcagaagag aagctgttat gaaatcttta 840

atcaaagagc acaaaaagag aatagcgtca ggcgaaaagc taaatagtta tatcgattac 900

cttttatctg aagctcaaac tttaaccgat cagcaactat tgatgtcctt gtgggaacca 960

atcattgaat cttcagatac aacaatggtc acaacagaat gggcaatgta cgaattagct 1020

aaaaacccta aattgcaaga taggttgtac agagacatta agtccgtctg tggatctgaa 1080

aagataaccg aagagcatct atcacagctg ccttacatta cagctatttt ccacgaaaca 1140

ctgagaagac actcaccagt tcctatcatt cctctaagac atgtacatga agataccgtt 1200

ctaggcggct accatgttcc tgctggcaca gaacttgccg ttaacatcta cggttgcaac 1260

atggacaaaa acgtttggga aaatccagag gaatggaacc cagaaagatt catgaaagag 1320

aatgagacaa ttgattttca aaagacgatg gccttcggtg gtggtaagag agtttgtgct 1380

ggttccttgc aagccctttt aactgcatct attgggattg ggagaatggt tcaagagttc 1440

gaatggaaac tgaaggatat gactcaagag gaagtgaaca cgataggcct aactacacaa 1500

atgttaagac cattgagagc tattatcaaa cctaggatct aa 1542

<210> 53
<211> 1530
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 53
atggcatttt tctctatgat ttcaattttg ttgggatttg ttatttcttc tttcatcttc 60

atctttttct tcaaaaagtt acttagtttt agtaggaaaa acatgtcaga agtttctact 120

ttgccaagtg ttccagtagt gcctggtttt ccagttattg ggaatttgtt gcaactaaag 180

gagaaaaagc ctcataaaac tttcactaga tggtcagaga tatatggacc tatctactct 240

ataaagatgg gttcttcatc tcttattgta ttgaacagta cagaaactgc taaggaagca 300

atggtcacta gattttcatc aatatctacc agaaaattgt caaacgccct aacagttcta 360

acctgcgata agtctatggt cgccacttct gattatgatg acttccacaa attagttaag 420

agatgtttgc taaatggact tcttggtgct aatgctcaaa agagaaaaag acactacaga 480

gatgctttga ttgaaaatgt gagttccaag ctacatgcac acgctagaga tcatccacaa 540

gagccagtta actttagagc aattttcgaa cacgaattgt ttggtgtagc attaaagcaa 600

gccttcggta aagacgtaga atccatatac gtcaaggagt taggcgtaac attatcaaaa 660

gatgaaatct ttaaggtgct tgtacatgat atgatggagg gtgcaattga tgtagattgg 720

agagatttct tcccatattt gaaatggatc cctaataagt cttttgaagc taggatacaa 780

caaaagcaca agagaagact agctgttatg aacgcactta tacaggacag attgaagcaa 840

aatgggtctg aatcagatga tgattgttac cttaacttct taatgtctga ggctaaaaca 900

ttgactaagg aacagatcgc aatccttgtc tgggaaacaa tcattgaaac agcagatact 960

accttagtca caactgaatg ggccatatac gagctagcca aacatccatc tgtgcaagat 1020

aggttgtgta aggagatcca gaacgtgtgt ggtggagaga aattcaagga agagcagttg 1080

tcacaagttc cttaccttaa cggcgttttc catgaaacct tgagaaaata ctcacctgca 1140

ccattagttc ctattagata cgcccacgaa gatacacaaa tcggtggcta ccatgttcca 1200

gctgggtccg aaattgctat aaacatctac gggtgcaaca tggacaaaaa gagatgggaa 1260

agaccagaag attggtggcc agaaagattc ttagatgatg gcaaatatga aacatctgat 1320

ttgcataaaa caatggcttt cggagctggc aaaagagtgt gtgccggtgc tctacaagcc 1380

tccctaatgg ctggtatcgc tattggtaga ttggtccaag agttcgaatg gaaacttaga 1440

gatggtgaag aggaaaatgt cgatacttat gggttaacat ctcaaaagtt atacccacta 1500

atggcaatca tcaatcctag aagatcctaa 1530

<210> 54
<211> 1578
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 54
atgagtaagt ctaatagtat gaattctaca tcacacgaaa ccctttttca acaattggtc 60

ttgggtttgg accgtatgcc attgatggat gttcactggt tgatctacgt tgctttcggc 120

gcatggttat gttcttatgt gatacatgtt ttatcatctt cctctacagt aaaagtgcca 180

gttgttggat acaggtctgt attcgaacct acatggttgc ttagacttag attcgtctgg 240

gaaggtggct ctatcatagg tcaagggtac aataagttta aagactctat tttccaagtt 300

aggaaattgg gaactgatat tgtcattata ccacctaact atattgatga agtgagaaaa 360

ttgtcacagg acaagactag atcagttgaa cctttcatta atgattttgc aggtcaatac 420

acaagaggca tggttttctt gcaatctgac ttacaaaacc gtgttataca acaaagacta 480

actccaaaat tggtttcctt gaccaaggtc atgaaggaag agttggatta tgctttaaca 540

aaagagatgc ctgatatgaa aaatgacgaa tgggtagaag tagatatcag tagtataatg 600

gtgagattga tttccaggat ctccgccaga gtctttctag ggcctgaaca ctgtcgtaac 660

caggaatggt tgactactac agcagaatat tcagaatcac ttttcattac agggtttatc 720

ttaagagttg tacctcatat cttaagacca ttcatcgccc ctctattacc ttcatacagg 780

actctactta gaaacgtttc aagtggtaga agagtcatcg gtgacatcat aagatctcag 840

caaggggatg gtaacgaaga tatactttcc tggatgagag atgctgccac aggagaggaa 900

aagcaaatcg ataacattgc tcagagaatg ttaattcttt ctttagcatc aatccacact 960

actgcgatga ccatgacaca tgccatgtac gatctatgtg cttgccctga gtacattgaa 1020

ccattaagag atgaagttaa atctgttgtt ggggcttctg gctgggacaa gacagcgtta 1080

aacagatttc ataagttgga ctccttccta aaagagtcac aaagattcaa cccagtattc 1140

ttattgacat tcaatagaat ctaccatcaa tctatgacct tatcagatgg cactaacatt 1200

ccatctggaa cacgtattgc tgttccatca cacgcaatgt tgcaagattc tgcacatgtc 1260

ccaggtccaa ccccacctac tgaatttgat ggattcagat atagtaagat acgttctgat 1320

agtaactacg cacaaaagta cctattctcc atgaccgatt cttcaaacat ggctttcgga 1380

tacggcaagt atgcttgtcc aggtagattt tacgcgtcta atgagatgaa actaacatta 1440

gccattttgt tgctacaatt tgagttcaaa ctaccagatg gtaaaggtcg tcctagaaat 1500

atcactatcg attctgatat gattccagac ccaagagcta gactttgcgt cagaaaaaga 1560

tcacttagag atgaatga 1578

<210> 55
<211> 1500
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 55
atggaagatc ctactgtctt atatgcttgt cttgccattg cagttgcaac tttcgttgtt 60

agatggtaca gagatccatt gagatccatc ccaacagttg gtggttccga tttgcctatt 120

ctatcttaca tcggcgcact aagatggaca agacgtggca gagagatact tcaagaggga 180

tatgatggct acagaggatc tacattcaaa atcgcgatgt tagaccgttg gatcgtgatc 240

gcaaatggtc ctaaactagc tgatgaagtc agacgtagac cagatgaaga gttaaacttt 300

atggacggat taggagcatt cgtccaaact aagtacacct taggtgaagc tattcataac 360

gatccatacc atgtcgatat cataagagaa aaactaacaa gaggccttcc agccgtgctt 420

cctgatgtca ttgaagagtt gacacttgcg gttagacagt acattccaac agaaggtgat 480

gaatgggtgt ccgtaaactg ttcaaaggcc gcaagagata ttgttgctag agcttctaat 540

agagtctttg taggtttgcc tgcttgcaga aaccaaggtt acttagattt ggcaatagac 600

tttacattgt ctgttgtcaa ggatagagcc atcatcaata tgtttccaga attgttgaag 660

ccaatagttg gcagagttgt aggtaacgcc accagaaatg ttcgtagagc tgttcctttt 720

gttgctccat tggtggagga aagacgtaga cttatggaag agtacggtga agactggtct 780

gaaaaaccta atgatatgtt acagtggata atggatgaag ctgcatccag agatagttca 840

gtgaaggcaa tcgcagagag attgttaatg gtgaacttcg cggctattca tacctcatca 900

aacactatca ctcatgcttt gtaccacctt gccgaaatgc ctgaaacttt gcaaccactt 960

agagaagaga tcgaaccatt agtcaaagag gagggctgga ccaaggctgc tatgggaaaa 1020

atgtggtggt tagattcatt tctaagagaa tctcaaagat acaatggcat taacatcgta 1080

tctttaacta gaatggctga caaagatatt acattgagtg atggcacatt tttgccaaaa 1140

ggtactctag tggccgttcc agcgtattct actcatagag atgatgctgt ctacgctgat 1200

gccttagtat tcgatccttt cagattctca cgtatgagag cgagagaagg tgaaggtaca 1260

aagcaccagt tcgttaatac ttcagtcgag tacgttccat ttggtcacgg aaagcatgct 1320

tgtccaggaa gattcttcgc cgcaaacgaa ttgaaagcaa tgttggctta cattgttcta 1380

aactatgatg taaagttgcc tggtgacggt aaacgtccat tgaacatgta ttggggtcca 1440

acagttttgc ctgcaccagc aggccaagta ttgttcagaa agagacaagt tagtctataa 1500

<210> 56
<211> 1542
<212> DNA
<213> Stevia rebaudiana

<400> 56
atggatgccg tcaccggttt gctgacagtt ccggcaaccg caataaccat cggcggtacg 60

gccgtcgcac tcgccgtcgc tctgatattc tggtacctca aaagctacac atctgcacgc 120

aggagccaat caaaccatct ccctcgggtt cccgaggtac ctggtgtgcc attattgggg 180

aatttattgc agttgaagga gaagaaacct tacatgactt ttacaagatg ggcggcaact 240

tatggtccga tttattcgat taaaaccgga gcaacttcta tggtggtcgt cagttcaaat 300

gaaattgcaa aggaggcatt ggttaccaga tttcaatcta tctcaaccag aaacctatca 360

aaggcattaa aggttctcac agcagataaa accatggtgg cgatgagtga ttatgatgat 420

tatcataaga ctgtcaaacg ccatatactg accgctgttt tgggaccaaa tgctcagaag 480

aaacaccgca tccataggga catcatgatg gataatatat caacccaact tcatgaattt 540

gttaaaaata atcctgaaca agaggaagtg gatctaagga aaatattcca atccgaactt 600

tttggattag ctatgagaca agcattggga aaggatgtgg agagcttata tgttgaggat 660

cttaaaatca ccatgaaccg agacgagata tttcaggtat tggttgttga cccgatgatg 720

ggtgcaattg acgtcgactg gagagatttc ttcccgtatc taaagtgggt cccgaataaa 780

aagtttgaaa acacgatcca acaaatgtat atccggagag aagctgtgat gaagtctctt 840

attaaagaac ataaaaaacg tattgcatcc ggagagaaat taaacagcta cattgattac 900

ttgctatcgg aagcacaaac gttaaccgat caacaactac ttatgtctct atgggaacct 960

attattgaat catcagacac cactatggtt acaactgaat gggctatgta tgaacttgca 1020

aaaaacccca aacttcagga tcgtttgtat cgggatatca aaagtgtttg cgggtcagag 1080

aagattacag aagaacactt gtctcaactg ccatacataa ctgccatttt tcatgaaacc 1140

ttgagaaggc atagtccagt tcctataatt ccattaagac acgtgcatga agacacagtg 1200

ttaggagggt accatgtgcc agctggaacc gagctagcgg taaacattta tggatgtaac 1260

atggataaga atgtgtggga gaatcctgaa gaatggaatc cagagagatt catgaaggaa 1320

aatgaaacga tagatttcca gaaaacaatg gcgtttggag gtggaaagcg cgtatgtgct 1380

ggttcgcttc aagcattgtt gactgcttcc attggaattg gaagaatggt gcaagagttt 1440

gagtggaaac tgaaagayat gacccaagaa gaagttaata cgattgggct tacgacccag 1500

atgcttcgtc cactgcgggc cataataaag cccaggatat ga 1542

<210> 57
<211> 1530
<212> DNA
<213> Arabidopsis thaliana

<400> 57
atggccttct tctccatgat ctccattctc cttggctttg ttatctcctc cttcatcttc 60

atcttcttct tcaagaaact tctctccttc tccagaaaga acatgtctga agtctccact 120

ctcccctctg ttccagtggt accagggttt cctgttattg ggaacttgct gcaactaaaa 180

gagaagaaac ctcacaagac tttcactaga tggtcagaga tttatggtcc tatttactct 240

ataaagatgg gttcttcttc tcttattgtc ctcaattcta ctgagactgc caaagaggcc 300

atggtgacgc ggttttcgtc tatctcaacg aggaagttgt caaatgcgtt gacagtcctt 360

acttgtgaca aatctatggt tgctactagt gattatgatg atttccacaa gttggtgaaa 420

cggtgtctct tgaacggtct tttgggtgct aatgcacaga aacgaaaaag acattacaga 480

gatgcactca ttgaaaatgt gtcttccaag ttgcatgccc atgctaggga ccatccacaa 540

gaacctgtaa acttcagagc tatatttgag catgagcttt tcggtgtagc attgaagcaa 600

gcttttggga aagatgtgga atccatttat gttaaagaac tcggtgtgac tttgtcgaaa 660

gacgagatct tcaaggtttt agtacatgac atgatggaag gtgcaattga tgttgattgg 720

agagacttct tcccatactt gaaatggatt ccaaataaaa gttttgaagc aagaatccag 780

caaaagcata aacgtagact cgcggtgatg aatgctctga ttcaagatcg actgaagcag 840

aatggttcag aatcggatga tgattgctat ctcaacttct tgatgtcgga agcgaaaaca 900

ctaaccaagg agcaaattgc tatcttggtt tgggagacga ttatcgagac agctgacact 960

actttggtta caactgaatg ggccatctat gagctcgcta agcatccaag tgtccaagat 1020

cgtctgtgta aagaaatcca aaatgtctgc ggaggagaaa agttcaaaga agagcaattg 1080

tctcaagttc cttatctcaa tggagtattc catgaaacgc ttaggaaata cagtcctgct 1140

cctctagtcc ccattcgcta tgcccacgaa gatacgcaaa tcggaggcta tcatgtccct 1200

gcaggaagtg agattgcaat aaacatctat ggatgcaaca tggataagaa gcgttgggag 1260

agaccagagg actggtggcc ggagcggttt cttgatgatg gcaaatacga aacgtcggat 1320

cttcacaaga caatggcgtt tggagcggga aagagggttt gtgctggtgc tcttcaagca 1380

tctctcatgg caggcattgc cattgggagg ttagtgcaag aattcgagtg gaagcttaga 1440

gacggtgaag aagagaatgt ggatacatat ggcttgacct ctcagaagct ttatcctctt 1500

atggctatta tcaatccaag gcgttcttaa 1530

<210> 58
<211> 1578
<212> DNA
<213> Gibberella fujikoroi

<400> 58
atgagtaagt ccaacagcat gaacagtacc agccatgaaa cgttattcca gcagctcgtc 60

ttaggtcttg acagaatgcc gctaatggac gttcactggc tgatctacgt ggcctttggc 120

gcttggttat gctcttatgt catccatgtc ctatcgtcct cttctacagt caaagtgccc 180

gtcgtaggct accgcagcgt ctttgagcct acatggcttc tccgtttgcg ctttgtttgg 240

gaaggggat ctatcatcgg ccaaggctac aacaaattta aagactctat cttccaggtg 300

cgaaagcttg gtaccgatat cgtcatcatc ccgccaaact acatcgatga ggtcagaaag 360

ctgtcccaag acaagactcg ctcggtcgag cccttcatca atgactttgc gggacagtat 420

acacggggca tggtctttct gcaaagtgat ttgcagaacc gtgtgattca gcagcggttg 480

acgccaaaac tcgtatcgtt gacaaaggta atgaaggagg agcttgacta tgccttgacc 540

aaagagatgc ctgacatgaa gaatgatgaa tgggttgaag tcgacatttc ttccatcatg 600

gtcaggctca tatcacgcat ctcagccaga gtgtttctcg gtccagagca ctgccgcaac 660

caagaatggt tgacgaccac tgcagagtac agcgagagcc tgttcataac tggctttatt 720

ctccgcgttg tcccccatat tctaagacca ttcatagccc cgctgctacc ctcctacaga 780

acactacttc gcaacgtctc gtcaggtcga agagttattg gagacatcat tcgctcccag 840

caaggtgatg gcaacgagga catcctgtca tggatgaggg atgctgcgac aggggaagaa 900

aagcaaattg acaacattgc ccagcggatg cttatcctga gtctcgcgtc tattcacact 960

acggcaatga cgatgacgca tgctatgtat gacttatgtg cttgccctga gtacatagag 1020

cctcttagag atgaggtcaa aagtgtcgtt ggcgctagtg gttgggacaa gacggcgttg 1080

aatcgattcc acaaactcga cagctttctc aaagagtcac aacgcttcaa ccccgtgttc 1140

ctcttaacgt tcaatcgcat ttatcaccaa tccatgacac tctcagatgg caccaacatc 1200

ccatcaggca ctcgcatcgc ggttccctct cacgcgatgc ttcaggactc agcgcatgtc 1260

ccaggcccga cgccaccaac cgagtttgat ggatttagat actcaaagat tcgctcagac 1320

tcaaactatg cacagaaata tctcttctcc atgactgatt ctagtaacat ggcgtttggg 1380

tatgggaaat acgcctgccc agggcggttc tatgcatcta atgagatgaa gctgactttg 1440

gcgatactcc ttttacaatt tgagttcaag ttgccagatg ggaaaggaag accacgaaat 1500

atcactattg atagtgacat gatacctgat ccgagagcta ggctgtgcgt taggaagcga 1560

tcactgagag atgaatga 1578

<210> 59
<211> 1500
<212> DNA
<213> Trametes versicolor

<400> 59
atggaggatc ccaccgtact ctacgcttgc ctcgccatcg ctgtcgctac tttcgttgtc 60

agatggtaca gagacccgct tcggtccatt cctacggttg ggggctctga ccttcccatc 120

ctctcataca tcggggcgct caggtggacc cgccgcggaa gagagatact gcaagaaggt 180

tatgatgggt atcgcggatc cacgttcaag atcgcgatgc tcgaccggtg gatcgtcatc 240

gccaacggcc caaagctcgc cgacgaggtg aggaggcgtc ctgacgaaga gctaaacttc 300

atggacggac tgggagcgtt cgtgcagacg aagtataccc ttggggaagc aatccacaat 360

gacccgtacc acgtggacat tattcgtgag aagctgacgc gaggcctccc ggcagtcctg 420

ccggacgtca tcgaggaact cacgctagcc gttcgccagt acatcccgac ggaaggagat 480

gaatgggtca gcgtgaactg ctccaaagca gcgcgggaca tcgtcgcccg ggcaagcaac 540

cgcgtctttg tcgggttgcc cgcttgccgc aaccagggtt atctcgacct cgccattgac 600

ttcaccctga gcgttgtcaa agacagggcg atcatcaata tgttcccgga gttgctgaaa 660

cctatcgtcg gacgcgtggt tggaaatgcc actaggaacg tgcgccgcgc ggtcccattc 720

gtagcgccgt tggtggagga acgtcgccgc ctcatggagg agtacggtga ggattggtcg 780

gagaaaccga acgacatgct ccagtggatc atggacgagg cagcctcgcg ggactcctcc 840

gtcaaagcga tcgctgagcg tcttctcatg gtcaactttg ccgcaattca cacgtcgtcg 900

aacaccatca cccacgctct ttaccacctc gccgagatgc cggagaccct acagccgctg 960

cgggaagaga tcgagccgct cgtcaaggaa gaaggctgga cgaaggccgc catgggcaag 1020

atgtggtggc tcgacagctt cctgcgggag tcacagcgct acaatggcat caacatcgtc 1080

tccctgacgc gcatggccga caaggacata acgctcagcg acggcacgtt cctcccgaag 1140

ggcacgctcg tcgcggtccc cgcgtactcg acgcaccgcg acgacgcggt gtacgcggac 1200

gcgctggtct tcgacccgtt ccgcttctcc cgcatgcgcg cccgcgaggg cgagggcacg 1260

aagcaccagt tcgtcaacac ctccgtggag tacgtgccct tcggccacgg gaagcacgcc 1320

tgccccgggc ggttcttcgc ggccaacgag ctgaaggcga tgctcgcgta catcgtgctc 1380

aactacgacg tgaagctgcc cggcgatggc aagcgccccc tgaacatgta ctggggcccg 1440

acggtcttgc ctgctccggc tgggcaggtg ctcttccgca agaggcaggt gtcgctgtag 1500

<210> 60
<211> 1578
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 60
atgggtttgt tcccattaga ggattcctac gcgctggtct ttgaaggact agcaataaca 60

ctggctttgt actatctact gtctttcatc tacaaaacat ctaaaaagac atgtacacct 120

cctaaagcat ctggtgaaat cattccaatt acaggaatca tattgaatct gctatctggc 180

tcaagtggtc tacctattat cttagcactt gcctctttag cagacagatg tggtcctatt 240

ttcaccatta ggctgggtat taggagagtg ctagtagtat caaattggga aatcgctaag 300

gagattttca ctacccacga tttgatagtt tctaatagac caaaatactt agccgctaag 360

attcttggtt tcaattatgt ttcattctct ttcgctccat acggcccata ttgggtcgga 420

atcagaaaga ttattgctac aaaactaatg tcttcttcca gacttcagaa gttgcaattt 480

gtaagagttt ttgaactaga aaactctatg aaatctatca gagaatcatg gaaggagaaa 540

aaggatgaag agggaaaggt attagttgag atgaaaaagt ggttctggga actgaatatg 600

aacatagtgt taaggacagt tgctggtaaa caatacactg gtacagttga tgatgccgat 660

gcaaagcgta tctccgagtt attcagagaa tggtttcact acactggcag atttgtcgtt 720

ggagacgctt ttccttttct aggttggttg gacctgggcg gatacaaaaa gacaatggaa 780

ttagttgcta gtagattgga ctcaatggtc agtaaatggt tagatgagca tcgtaaaaag 840

caagctaacg atgacaaaaa ggaggatatg gatttcatgg atatcatgat ctccatgaca 900

gaagcaaatt caccacttga aggatacggc actgatacta ttatcaagac cacatgtatg 960

actttgattg tttcaggagt tgatacaacc tcaatcgtac ttacttgggc cttatcactt 1020

ttgttaaaca acagagatac tttgaaaaag gcacaagagg aattagatat gtgcgtaggt 1080

aaaggaagac aagtcaacga gtctgatctt gttaacttga tatacttgga agcagtgctt 1140

aaagaggctt taagacttta cccagcagcg ttcttaggcg gaccaagagc attcttggaa 1200

gattgtactg ttgctggtta tagaattcca aagggcacct gcttgttgat taacatgtgg 1260

aaactgcata gagatccaaa catttggagt gatccttgcg aattcaagcc agaaagattt 1320

ttgacaccta atcaaaagga tgttgatgtg atcggtatgg atttcgaatt gataccattt 1380

ggtgccggca gaagatattg tccaggtact agattggctt tacagatgtt gcatatcgta 1440

ttagcgacat tgctgcaaaa cttcgaaatg tcaacaccaa acgatgcgcc agtcgatatg 1500

actgcttctg ttggcatgac aaatgccaaa gcatcacctt tagaagtctt gctatcacct 1560

cgtgttaaat ggtcctaa 1578

<210> 61
<211> 1431
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 61
atgatacaag ttttaactcc aattctactc ttcctcatct tcttcgtttt ctggaaagtc 60

tacaaacatc aaaagactaa aatcaatcta ccaccaggtt ccttcggctg gccatttttg 120

ggtgaaacct tagccttact tagagcaggc tgggattctg agccagaaag attcgtaaga 180

gagcgtatca aaaagcatgg atctccactt gttttcaaga catcactatt tggagacaga 240

ttcgctgttc tttgcggtcc agctggtaat aagtttttgt tctgcaacga aaacaaatta 300

gtggcatctt ggtggccagt ccctgtaagg aagttgttcg gtaaaagttt actcacaata 360

agaggagatg aagcaaaatg gatgagaaaa atgctattgt cttacttggg tccagatgca 420

tttgccacac attatgccgt tactatggat gttgtaacac gtagacatat tgatgtccat 480

tggaggggca aggaggaagt taatgtattt caaacagtta agttgtacgc attcgaatta 540

gcttgtagat tattcatgaa cctagatgac ccaaaccaca tcgcgaaact cggtagtctt 600

ttcaacattt tcctcaaagg gatcatcgag cttcctatag acgttcctgg aactagattt 660

tactccagta aaaaggccgc agctgccatt agaattgaat tgaaaaagct cattaaagct 720

agaaaactcg aattgaagga gggtaaggcg tcttcttcac aggacttgct ttctcatcta 780

ttaacatcac ctgatgagaa tgggatgttc ttgacagaag aggaaatagt cgataacatt 840

ctacttttgt tattcgctgg tcacgatacc tctgcactat caataacact tttgatgaaa 900

accttaggtg aacacagtga tgtgtacgac aaggttttga aggaacaatt agaaatttcc 960

aaaacaaagg aggcttggga atcactaaag tgggaagata tccagaagat gaagtactca 1020

tggtcagtaa tctgtgaagt catgagattg aatcctcctg tcatagggac atacagagag 1080

gcgttggttg atatcgacta tgctggttac actatcccaa aaggatggaa gttgcattgg 1140

tcagctgttt ctactcaaag agacgaagcc aatttcgaag atgtaactag attcgatcca 1200

tccagatttg aaggggcagg ccctactcca ttcacatttg tgcctttcgg tggaggtcct 1260

agaatgtgtt taggcaaaga gtttgccagg ttagaagtgt tagcatttct ccacaacatt 1320

gttaccaact ttaagtggga tcttctaatc cctgatgaga agatcgaata tgatccaatg 1380

gctactccag ctaagggctt gccaattaga cttcatccac accaagtcta a 1431

<210> 62
<211> 1578
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 62
atggagtctt tagtggttca tacagtaaat gctatctggt gtattgtaat cgtcgggatt 60

ttctcagttg gttatcacgt ttacggtaga gctgtggtcg aacaatggag aatgagaaga 120

tcactgaagc tacaaggtgt taaaggccca ccaccatcca tcttcaatgg taacgtctca 180

gaaatgcaac gtatccaatc cgaagctaaa cactgctctg gcgataacat tatctcacat 240

gattattctt cttcattatt cccacacttc gatcactgga gaaaacagta cggcagaatc 300

tacacatact ctactggatt aaagcaacac ttgtacatca atcatccaga aatggtgaag 360

gagctatctc agactaacac attgaacttg ggtagaatca cccatataac caaaagattg 420

aatcctatct taggtaacgg aatcataacc tctaatggtc ctcattgggc ccatcagcgt 480

agaattatcg cctacgagtt tactcatgat aagatcaagg gtatggttgg tttgatggtt 540

gagtctgcta tgcctatgtt gaataagtgg gaggagatgg taaagagagg cggagaaatg 600

ggatgcgaca taagagttga tgaggacttg aaagatgttt cagcagatgt gattgcaaaa 660

gcctgtttcg gatcctcatt ttctaaaggt aaggctattt tctctatgat aagagatttg 720

cttacagcta tcacaaagag aagtgttcta ttcagattca acggattcac tgatatggtc 780

tttgggagta aaaagcatgg tgacgttgat atagacgctt tagaaatgga attggaatca 840

tccatttggg aaactgtcaa ggaacgtgaa atagaatgta aagatactca caaaaaggat 900

ctgatgcaat tgattttgga aggggcaatg cgttcatgtg acggtaacct ttgggataaa 960

tcagcatata gaagatttgt tgtagataat tgtaaatcta tctacttcgc agggcatgat 1020

agtacagctg tctcagtgtc atggtgtttg atgttactgg ccctaaaccc atcatggcaa 1080

gttaagatcc gtgatgaaat tctgtcttct tgcaaaaatg gtattccaga tgccgaaagt 1140

atcccaaacc ttaaaacagt gactatggtt attcaagaga caatgagatt ataccctcca 1200

gcaccaatcg tcgggagaga agcctctaaa gatatcagat tgggcgatct agttgttcct 1260

aaaggcgtct gtatatggac actaatacca gctttacaca gagatcctga gatttgggga 1320

ccagatgcaa acgatttcaa accagaaaga ttttctgaag gaatttcaaa ggcttgtaag 1380

tatcctcaaa gttacattcc atttggtctg ggtcctagaa catgcgttgg taaaaacttt 1440

ggcatgatgg aagtaaaggt tcttgtttcc ctgattgtct ccaagttctc tttcactcta 1500

tctcctacct accaacatag tcctagtcac aaacttttag tagaaccaca acatggggtg 1560

gtaattagag tggtttaa 1578

<210> 63
<211> 1590
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 63
atgtacttcc tactacaata cctcaacatc acaaccgttg gtgtctttgc cacattgttt 60

ctctcttatt gtttacttct ctggagaagt agagcgggta acaaaaagat tgccccagaa 120

gctgccgctg catggcctat tatcggccac ctccacttac ttgcaggtgg atcccatcaa 180

ctaccacata ttacattggg taacatggca gataagtacg gtcctgtatt cacaatcaga 240

ataggcttgc atagagctgt agttgtctca tcttgggaaa tggcaaagga atgttcaaca 300

gctaatgatc aagtgtcttc ttcaagacct gaactattag cttctaagtt gttgggttat 360

aactacgcca tgtttggttt ttcaccatac ggttcatact ggagagaaat gagaaagatc 420

atctctctcg aattactatc taattccaga ttggaactat tgaaagatgt tagagcctca 480

gaagttgtca catctattaa ggaactatac aaattgtggg cggaaaagaa gaatgagtca 540

ggattggttt ctgtcgagat gaaacaatgg ttcggagatt tgactttaaa cgtgatcttg 600

agaatggtgg ctggtaaaag atacttctcc gcgagtgacg cttcagaaaa caaacaggcc 660

cagcgttgta gaagagtctt cagagaattc ttccatctct ccggcttgtt tgtggttgct 720

gatgctatac cttttcttgg atggctcgat tggggaagac acgagaagac cttgaaaaag 780

accgccatag aaatggattc catcgcccag gagtggcttg aggaacatag acgtagaaaa 840

gattctggag atgataattc tacccaagat ttcatggacg ttatgcaatc tgtgctagat 900

ggcaaaaatc taggcggata cgatgctgat acgattaaca aggctacatg cttaactctt 960

atatcaggtg gcagtgatac tactgtagtt tctttgacat gggctcttag tcttgtgtta 1020

aacaatagag atactttgaa aaaggcacag gaagagttag acatccaagt cggtaaggaa 1080

agattggtta acgagcaaga catcagtaag ttagtttact tgcaagcaat agtaaaagag 1140

acactcagac tttatccacc aggtcctttg ggtggtttga gacaattcac tgaagattgt 1200

acactaggtg gctatcacgt ttcaaaagga actagattaa tcatgaactt atccaagatt 1260

caaaaagatc cacgtatttg gtctgatcct actgaattcc aaccagagag attccttacg 1320

actcataaag atgtcgatcc acgtggtaaa cactttgaat tcattccatt cggtgcagga 1380

agacgtgcat gtcctggtat cacattcgga ttacaagtac tacatctaac attggcatct 1440

ttcttgcatg cgtttgaatt ttcaacacca tcaaatgagc aggttaacat gagagaatca 1500

ttaggtctta cgaatatgaa atctacccca ttagaagttt tgatttctcc aagactatcc 1560

cttaattgct tcaaccttat gaaaatttga 1590

<210> 64
<211> 1440
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 64
atggaaccta acttttactt gtcattacta ttgttgttcg tgaccttcat ttctttaagt 60

ctgtttttca tcttttacaa acaaaagtcc ccattgaatt tgccaccagg gaaaatgggt 120

taccctatca taggtgaaag tttagaattc ctatccacag gctggaaggg acatcctgaa 180

aagttcatat ttgatagaat gcgtaagtac agtagtgagt tattcaagac ttctattgta 240

ggcgaatcca cagttgtttg ctgtggggca gctagtaaca aattcctatt ctctaacgaa 300

aacaaactgg taactgcctg gtggccagat tctgttaaca aaatcttccc aacaacttca 360

ctggattcta atttgaagga ggaatctata aagatgagaa agttgctgcc acagttcttc 420

aaaccagaag cacttcaaag atacgtcggc gttatggatg taatcgcaca aagacatttt 480

gtcactcact gggacaacaa aaatgagatc acagtttatc cacttgctaa aagatacact 540

ttcttgcttg cgtgtagact gttcatgtct gttgaggatg aaaatcatgt ggcgaaattc 600

tcagacccat tccaactaat cgctgcaggc atcatttcac ttcctatcga tcttcctggt 660

actccattca acaaggccat aaaggcttca aatttcatta gaaaagagct gataaagatt 720

atcaaacaaa gacgtgttga tctggcagag ggtacagcat ctccaaccca ggatatcttg 780

tcacatatgc tattaacatc tgatgaaaac ggtaaatcta tgaacgagtt gaacattgcc 840

gacaagattc ttggactatt gataggaggc cacgatacag cttcagtagc ttgcacattt 900

ctagtgaagt acttaggaga attaccacat atctacgata aagtctacca agagcaaatg 960

gaaattgcca agtccaaacc tgctggggaa ttgttgaatt gggatgactt gaaaaagatg 1020

aagtattcat ggaatgtggc atgtgaggta atgagattgt caccaccttt acaaggtggt 1080

tttagagagg ctataactga ctttatgttt aacggtttct ctattccaaa agggtggaag 1140

ttatactggt ccgccaactc tacacacaaa aatgcagaat gtttcccaat gcctgagaaa 1200

ttcgatccta ccagatttga aggtaatggt ccagcgcctt atacatttgt accattcggt 1260

ggaggcccta gaatgtgtcc tggaaaggaa tacgctagat tagaaatctt ggttttcatg 1320

cataatctgg tcaaacgttt taagtgggaa aaggttattc cagacgaaaa gattattgtc 1380

gatccattcc caatcccagc taaagatctt ccaatccgtt tgtatcctca caaagcttaa 1440

<210> 65
<211> 1572
<212> DNA
<213> Stevia rebaudiana

<400> 65
atgggtctct tccctttgga agatagttac gcactcgtct ttgaaggttt agcaataact 60

actctagctc tctactactt attatccttc atctataaaa cctctaaaaa gacttgtact 120

ccacctaaag caagcggtga gcaccctata acaggccact taaaccttct tagtggttca 180

tccggtcttc cccatctagc cttagcatct ttggctgacc gatgtgggcc catattcacc 240

atccgacttg gcatacgtag agttttggtg gttagtaatt gggaaattgc taaggagatc 300

ttcactaccc atgatttgat tgtttcaaac cgtcccaaat acctcgctgc aaagattttg 360

ggattcaact atgtgtcctt ttcgtttgct ccatatggcc cctattgggt tggaatccgt 420

aagatcatcg ccacaaaact gatgtcaagt agcaggctcc agaagcttca gtttgtccga 480

gttttcgaac tagaaaactc catgaaaagc atacgcgagt cttggaaaga gaaaaaagac 540

gaagaaggta aagtgttggt ggagatgaaa aaatggtttt gggaattgaa tatgaatata 600

gttcttagaa ctgttgctgg taaacagtac actggaactg ttgatgatgc ggatgcgaag 660

aggattagtg aattgtttag agaatggttt cattacacag gaaggtttgt tgtgggagat 720

gcttttcctt ttcttgggtg gttggatttg ggtggatata agaagaccat ggaactagtg 780

gcttccagac tagattccat ggtctcaaaa tggttagacg agcatcgcaa aaagcaggct 840

aacgacgaca aaaaagagga catggatttc atggacatca tgatatcgat gactgaagcc 900

aattcccctt tggagggtta tggtacggat acaataatta aaaccacttg catgactctt 960

attgtcagtg gtgtagatac aacctccatc gtgctaactt gggcactctc gttactactg 1020

aacaaccgtg acactcttaa gaaagctcaa gaagagctag acatgtgtgt gggaaaaggt 1080

cgacaagtaa acgaatcaga tctagtaaac ctaatctacc ttgaagccgt attaaaagaa 1140

gcattgcgac tatacccagc agcattcctt ggaggtccta gagccttttt agaagactgc 1200

accgtggcag ggtaccgtat cccaaaaggc acatgtctac ttattaacat gtggaaactt 1260

catcgtgatc caaacatatg gtcagaccca tgtgagttta aaccagagag gttcttaacc 1320

ccaaaccaaa aggacgtaga tgttattgga atggattttg agttaatccc atttggtgcg 1380

ggaagaaggt attgtccagg gacacgtttg gcattacaaa tgttacacat agttctggcc 1440

actctactac aaaactttga gatgtcaact ccaaatgatg cacccgttga tatgaccgcg 1500

agtgttggaa tgacaaatgc gaaggcaagt ccacttgaag ttctactttc gccacgtgtt 1560

aagtggtcat ag 1572

<210> 66
<211> 1431
<212> DNA
<213> Stevia rebaudiana

<400> 66
atgattcaag ttctaacacc gatccttctc ttcctcattt tcttcgtttt ctggaaggtt 60

tacaagcacc agaaaaccaa aatcaatctt ccaccgggaa gcttcggatg gccatttctg 120

ggcgaaactc tggcactcct acgtgcaggt tgggactcag agccggagag atttgttcgt 180

gaacggatca agaaacacgg aagtcctcta gtgtttaaga cgtcgttgtt tggcgaccgt 240

tttgcggtgt tgtgtggacc tgccggaaac aagttcctgt tctgcaacga gaacaagctg 300

gtggcgtcgt ggtggccggt tccggtgagg aagcttttcg gcaagtctct gctcacgatt 360

cgtggtgatg aagctaagtg gatgaggaag atgttgttat cgtatctcgg tcctgatgct 420

ttcgcaactc attatgccgt caccatggac gtcgtcaccc gtcggcatat cgacgttcat 480

tggcgaggga aggaagaggt gaacgtattc caaaccgtta agttatatgc ctttgagctt 540

gcatgtcgtt tattcatgaa cctagacgac ccaaaccaca ttgcaaaact cggttccttg 600

ttcaacattt tcttgaaagg catcattgag cttccaatcg acgtcccagg gacacgattt 660

tatagctcca aaaaagcagc agcagctatc aggattgaac taaaaaaatt gattaaagca 720

agaaaactgg aactgaaaga agggaaggca tcatcttcac aagacctctt atcacatttg 780

cttacatctc cagatgaaaa tggtatgttt ctaaccgaag aagagattgt agacaacatc 840

ttgttactac tctttgcggg tcatgatacc tcggctcttt caatcacttt gctcatgaag 900

actcttggcg aacattctga tgtttatgac aaggtgttaa aagagcaact agagatatcg 960

aagacgaaag aagcatggga gtccctgaaa tgggaggaca tacaaaagat gaaatactcc 1020

tggagtgtta tatgtgaagt catgagacta aatccacctg ttataggaac ctatagagag 1080

gcccttgtgg atattgatta tgcgggttat accatcccca aaggatggaa gctgcactgg 1140

agtgctgtat cgacacaaag ggacgaggct aactttgaag acgtaacacg ttttgaccca 1200

tcacggtttg aaggcgcagg accgactcca ttcacctttg ttccgtttgg agggggcct 1260

agaatgtgtt tagggaaaga atttgctcga ttggaagtac ttgcgtttct tcacaatatt 1320

gtcaccaatt tcaaatggga cctgttgata cctgatgaga aaatagaata tgatcccatg 1380

gctaccccag caaaggggct tccaattcgt cttcatcccc atcaagtttg a 1431

<210> 67

<211> 1578
<212> DNA
<213> Arabidopsis thaliana

<400> 67
atggagagtt tggttgttca tacggtaaat gcaatttggt gcatagttat tgtcggaatc 60

ttcagcgtag gttatcatgt gtatggaaga gcggtggtgg agcagtggag gatgcggagg 120

agtttaaagt tgcaaggcgt gaagggtcct ccaccgtcga tctttaacgg caatgtgtcg 180

gagatgcaac ggattcagtc ggaggctaaa cactgttccg gcgataacat catttctcat 240

gactattctt cttctctatt tcctcatttc gatcactggc gaaaacaata cggaaggatt 300

tacacatact caacggggtt aaagcagcac ctttacataa accacccgga aatggtgaag 360

gagcttagcc aaaccaacac acttaacctt ggtagaatca ctcacatcac caaacgcctt 420

aaccccattc tcggcaatgg catcatcacc tctaatgggc ctcattgggc ccatcaacgt 480

cgtatcattg cctatgagtt tacccacgac aaaatcaagg gaatggttgg tttaatggtg 540

gaatctgcca tgccaatgtt gaacaaatgg gaagagatgg tgaaaagagg aggagaaatg 600

ggttgtgaca taagagtgga cgaagacctt aaggatgtct cagctgatgt catcgctaag 660

gcttgctttg ggagctcttt ttcaaaaggc aaagcaatat tctctatgat tagggatctt 720

ttaaccgcca ttactaaacg aagcgtcctc ttcagattca atggcttcac tgatatggtg 780

tttggaagta agaagcatgg tgatgtggat attgatgcgc ttgagatgga attagaatct 840

tctatatggg aaacggttaa ggagagggaa attgaatgta aggatactca caagaaggat 900

ctaatgcagt tgatactcga gggagcgatg cgaagctgcg atggtaactt gtgggacaag 960

tcagcctata gacggtttgt ggtggacaat tgcaagagca tctatttcgc cggacatgat 1020

tcaaccgcag tctcagtgtc ttggtgcctt atgctcctcg ctctcaatcc tagttggcag 1080

gttaaaattc gcgatgaaat cttgagttct tgcaagaatg gcattcccga cgcagaatca 1140

attcctaatc tcaaaacggt gacaatggta atacaagaaa caatgagact atacccacca 1200

gcaccaatcg tgggaagaga agcatccaaa gacataagac ttggagacct tgtggtgcca 1260

aaaggagtgt gcatttggac actcattcct gccttacacc gagaccccga gatctgggga 1320

ccagacgcaa acgacttcaa gccagagagg tttagtgagg gaatctctaa ggcttgcaaa 1380

taccctcagt catacatccc atttggcctt ggaccaagaa catgcgtagg caaaaacttt 1440

ggtatgatgg aagtgaaagt gcttgtttca cttattgtct caaagttcag ttttactctt 1500

tccccgactt atcagcactc tccaagccat aaactccttg tagagcctca acatggtgtt 1560

gtcattaggg ttgtttga 1578

<210> 68
<211> 1590
<212> DNA

<213> Vitis vinifera

<400> 68
atgtatttcc ttctccaata cctaaacatc accacggtcg gagtctttgc cacacttttc 60

ctttcctact gtctattatt atggaggtct agagctggta acaaaaaaat agcacctgaa 120

gctgctgctg catggcccat aatcggtcac ctacacctgt tagctggtgg ttctcatcag 180

cttccccaca taaccttggg aaacatggcc gacaaatatg gaccggtctt cacaattcgg 240

attgggttgc atcgagctgt ggtggtaagt tcttgggaga tggctaaaga atgctcgacc 300

gccaatgacc aggtttcatc ctcgcgtccc gaacttttag cctcaaaact tttgggctac 360

aactacgcca tgtttggttt ctctccatac ggttcttact ggcgtgaaat gcgcaagata 420

atcagcctag agctactctc taacagccgc ttagagctgc tgaaggacgt ccgagcttca 480

gaagtggtga catccataaa agagctatac aagctctggg cagagaaaaa aaatgaatcg 540

ggccttgtct cggtggagat gaagcagtgg tttggagact tgactctgaa cgtaattctt 600

aggatggtgg cagggaagcg ttatttcagt gcttcagatg caagtgaaaa taaacaggcg 660

cagaggtgcc ggagagtgtt cagggaattc tttcatttgt cagggctctt tgtggtggcg 720

gacgctattc catttcttgg atggctcgac tgggggagac atgagaaaac cctaaagaag 780

acagcaatag aaatggacag tattgctcaa gaatggttag aggagcaccg tcggaggaaa 840

gactccggtg atgataatag tacgcaagac ttcatggatg tgatgcagtc agttcttgat 900

ggcaaaaacc ttggtggtta cgacgctgat accatcaata aagccacatg cctgactcta 960

atctccggag gtagcgacac aactgttgtc tctctaacat gggcactctc tcttgtacta 1020

aacaaccgtg acaccttaaa aaaagctcaa gaagaattag acatccaagt tggtaaggaa 1080

agattagtga atgaacaaga tataagtaag ttggtctatc tccaagccat tgttaaagag 1140

acattacggt tatatccacc aggaccactt ggaggactac gccaatttac cgaggattgc 1200

accttgggtg gataccatgt ctctaaaggc acccgtttaa taatgaacct ttcgaagatc 1260

caaaaggatc caagaatttg gtcagatccg acagaattcc aaccagagag gtttctcacc 1320

acccataaag atgttgatcc tcggggaaaa cattttgagt ttataccatt tggagctggt 1380

cgaagagcat gtccaggaat aacttttggt cttcaagtat tacatttaac attggctagt 1440

ttcttacatg cgtttgaatt ttcaactcca tcaaatgaac aggtcaatat gcgcgagagc 1500

cttggactta caaatatgaa atctacccca cttgaagttc tcatttctcc acgcttatca 1560

ttgaattgtt ttaacctaat gaagatataa 1590

<210> 69
<211> 1440
<212> DNA
<213> Medicago trunculata

<400> 69
atggagccta atttctatct ctcccttctc cttctctttg tcactttcat atctctctct 60

ctttttttca tattctacaa acagaaatct ccattaaatt tgccacctgg taaaatgggt 120

tacccaatca taggtgaaag ccttgagttc ttatcaacag gatggaaagg acatcctgaa 180

aaattcattt tcgaccgtat gcgtaaatat tcctcagaac tctttaaaac atcaatcgta 240

ggagaatcta cggtggtttg ttgcggagca gcaagtaaca agtttttgtt ttcaaacgag 300

aataaacttg tgactgcatg gtggccagat agtgtaaaca aaatcttccc tactacttct 360

cttgactcta acttgaagga agaatccatc aagatgagaa aattgcttcc acaattcttt 420

aaacccgaag ctctacaacg ttatgttggt gtcatggatg ttattgctca aagacatttt 480

gttactcatt gggataataa aaatgaaatc accgtctacc ccttggccaa gaggtacacc 540

tttttgttag cttgtcggtt gttcatgagc gttgaagacg agaatcatgt agcaaaattt 600

agtgatccat ttcagttaat tgcggccgga atcatatctc taccaattga tttgccagga 660

acaccattca acaaagctat aaaggcctca aactttataa gaaaggagtt gattaagatc 720

ataaagcaaa ggagggtaga tttggcagaa gggacagcat caccaacaca agatatattg 780

tctcacatgt tgttgacaag tgatgaaaat ggaaagagta tgaatgaact taatattgct 840

gataagattc ttggcctttt gatcggagga catgacactg ctagcgtcgc atgcactttc 900

cttgtcaaat atctcggcga gttacctcac atttatgata aagtctatca agagcaaatg 960

gaaattgcaa aatcgaaacc agcaggagaa ttgttgaatt gggatgacct gaagaaaatg 1020

aaatactctt ggaacgtagc ttgtgaagta atgagacttt cccctccact ccaaggaggt 1080

ttcagggaag ccatcactga ctttatgttc aatggattct caattcctaa gggatggaag 1140

ctttattgga gtgcaaattc aacacataag aacgcagaat gttttcccat gccagagaaa 1200

tttgacccaa caagatttga aggaaatgga ccagctcctt atacttttgt tccctttggt 1260

ggaggaccaa ggatgtgtcc tggaaaagag tatgcaagat tagaaatact tgttttcatg 1320

cacaatttgg tgaaaaggtt taagtgggaa aaggtgattc cagatgagaa gattattgtt 1380

gatccattcc ccatccctgc aaaggatctt ccaattcgcc tttatccaca caaagcttaa 1440

<210> 70
<211> 2133
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 70
atgcaatcag attcagtcaa agtctctcca tttgatttgg tttccgctgc tatgaatggc 60

aaggcaatgg aaaagttgaa cgctagtgaa tctgaagatc caacaacatt gcctgcacta 120

aagatgctag ttgaaaatag agaattgttg acactgttca caacttcctt cgcagttctt 180
attgggtgtc ttgtatttct aatgtggaga cgttcatcct ctaaaaagct ggtacaagat 240
ccagttccac aagttatcgt tgtaaagaag aaagagaagg agtcagaggt tgatgacggg 300
aaaaagaaag tttctatttt ctacggcaca caaacaggaa ctgccgaagg ttttgctaaa 360
gcattagtcg aggaagcaaa agtgagatat gaaaagacct ctttcaaggt tatcgatcta 420
gatgactacg ctgcagatga tgatgaatat gaggaaaaac tgaaaaagga atccttagcc 480
ttcttcttct tggccacata cggtgatggt gaacctactg ataatgctgc taacttctac 540
aagtggttca cagaaggcga cgataaaggt gaatggctga aaaagttaca atacggagta 600
tttggtttag gtaacagaca atatgaacat ttcaacaaga tcgctattgt agttgatgat 660
aaacttactg aaatgggagc caaaagatta gtaccagtag gattagggga tgatgatcag 720
tgtatagaag atgacttcac cgcctggaag gaattggtat ggccagaatt ggatcaactt 780
ttaagggacg aagatgatac ttctgtgact accccataca ctgcagccgt attggagtac 840
agagtggttt accatgataa accagcagac tcatatgctg aagatcaaac ccatacaaac 900
ggtcatgttg ttcatgatgc acagcatcct tcaagatcta atgtggcttt caaaaaggaa 960
ctacacacct ctcaatcaga taggtcttgt actcacttag aattcgatat ttctcacaca 1020
ggactgtctt acgaaactgg cgatcacgtt ggcgtttatt ccgagaactt gtccgaagtt 1080
gtcgatgaag cactaaaact gttagggtta tcaccagaca catacttctc agtccatgct 1140
gataaggagg atgggacacc tatcggtggt gcttcactac caccaccttt tcctccttgc 1200
acattgagag acgctctaac cagatacgca gatgtcttat cctcacctaa aaaggtagct 1260
ttgctggcat tggctgctca tgctagtgat cctagtgaag ccgataggtt aaagttcctg 1320
gcttcaccag ccggaaaaga tgaatatgca caatggatcg tcgccaacca acgttctttg 1380
ctagaagtga tgcaaagttt tccatctgcc aagcctccat taggtgtgtt cttcgcagca 1440
gtagctccac gtttacaacc aagatactac tctatcagtt catctcctaa gatgtctcct 1500
aacagaatac atgttacatg tgctttggtg tacgagacta ctccagcagg cagaattcac 1560
agaggattgt gttcaacctg gatgaaaaat gctgtccctt taacagagtc acctgattgc 1620
tctcaagcat ccattttcgt tagaacatca aatttcagac ttccagtgga tccaaaagtt 1680
ccagtcatta tgataggacc aggcactggt cttgccccat tcaggggctt tcttcaagag 1740
agattggcct tgaaggaatc tggtacagaa ttgggttctt ctatcttttt ctttggttgc 1800
cgtaatagaa aagttgactt tatctacgag gacgagctta acaattttgt tgagacagga 1860
gcattgtcag aattgatcgt cgcattttca agagaaggga ctgccaaaga gtacgttcag 1920
cacaagatga gtcaaaaagc ctccgatata tggaaacttc taagtgaagg tgcctatctt 1980

tatgtctgtg gcgatgcaaa gggcatggcc aaggatgtcc atagaactct gcatacaatt 2040

gttcaggaac aagggagtct ggattcttcc aaggctgaat tgtacgtcaa aaacttacag 2100

atgtctggaa gatacttaag agatgtttgg taa 2133

<210> 71
<211> 2079
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 71
atgacttctg cactttatgc ctccgatctt ttcaaacaat tgaaaagtat catgggaacg 60

gattctttgt ccgatgatgt tgtattagtt attgctacaa cttctctggc actggttgct 120

ggtttcgttg tcttattgtg gaaaaagacc acggcagatc gttccggcga gctaaagcca 180

ctaatgatcc ctaagtctct gatggcgaaa gatgaggatg atgacttaga tctaggttct 240

ggaaaaacga gagtctctat cttcttcggc acacaaaccg gaacagccga aggattcgct 300

aaagcacttt cagaagagat caaagcaaga tacgaaaagg cggctgtaaa agtaatcgat 360

ttggatgatt acgctgccga tgatgaccaa tatgaggaaa agttgaaaaa ggaaacattg 420

gctttctttt gtgtagccac gtatggtgat ggtgaaccaa ccgataacgc cgcaagattc 480

tacaagtggt ttactgaaga gaacgaaaga gatatcaagt tgcagcaact tgcttacggc 540

gtttttgcct taggtaacag acaatacgag cactttaaca agataggtat tgtcttagat 600

gaagagttat gcaaaaaggg tgcgaagaga ttgattgaag tcggtttagg agatgatgat 660

caatctatcg aggatgactt taatgcatgg aaggaatctt tgtggtctga attagataag 720

ttacttaagg acgaagatga taaatccgtt gccactccat acacagccgt cattccagaa 780

tatagagtag ttactcatga tccaagattc acaacacaga aatcaatgga aagtaatgtg 840

gctaatggta atactaccat cgatattcat catccatgta gagtagacgt tgcagttcaa 900

aaggaattgc acactcatga atcagacaga tcttgcatac atcttgaatt tgatatatca 960

cgtactggta tcacttacga aacaggtgat cacgtgggtg tctacgctga aaaccatgtt 1020

gaaattgtag aggaagctgg aaagttgttg ggccatagtt tagatcttgt tttctcaatt 1080

catgccgata aagaggatgg ctcaccacta gaaagtgcag tgcctccacc atttccagga 1140

ccatgcaccc taggtaccgg tttagctcgt tacgcggatc tgttaaatcc tccacgtaaa 1200

tcagctctag tggccttggc tgcgtacgcc acagaacctt ctgaggcaga aaaactgaaa 1260

catctaactt caccagatgg taaggatgaa tactcacaat ggatagtagc tagtcaacgt 1320

tctttactag aagttatggc tgctttccca tccgctaaac ctcctttggg tgttttcttc 1380

gccgcaatag cgcctagact gcaaccaaga tactattcaa tttcatcctc acctagactg 1440

gcaccatcaa gagttcatgt cacatccgct ttagtgtacg gtccaactcc tactggtaga 1500

atccataagg gcgtttgttc aacatggatg aaaaacgcgg ttccagcaga gaagtctcac 1560

gaatgttctg gtgctccaat ctttatcaga gcctccaact tcaaactgcc ttccaatcct 1620

tctactccta ttgtcatggt cggtcctggt acaggtcttg ctccattcag aggtttctta 1680

caagagagaa tggccttaaa ggaggatggt gaagagttgg gatcttcttt gttgtttttc 1740

ggctgtagaa acagacaaat ggatttcatc tacgaagatg aactgaataa ctttgtagat 1800

caaggagtta tttcagagtt gataatggct ttttctagag aaggtgctca gaaggagtac 1860

gtccaacaca aaatgatgga aaaggccgca caagtttggg acttaatcaa agaggaaggc 1920

tatctatatg tctgtggtga tgcaaagggt atggcaagag atgttcacag aacacttcat 1980

actatagtcc aggaacagga aggcgttagt tcttctgaag cggaagcaat tgtgaaaaag 2040

ttacaaacag agggaagata cttgagagat gtgtggtaa 2079

<210> 72
<211> 2142
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 72
atggcagaat tagatacact tgatatagta gtattaggtg ttatcttttt gggtactgtg 60

gcatacttta ctaagggtaa attgtggggt gttaccaagg atccatacgc taacggattc 120

gctgcaggtg gtgcttccaa gcctggcaga actagaaaca tcgtcgaagc tatggaggaa 180

tcaggtaaaa actgtgttgt tttctacggc agtcaaacag gtacagcgga ggattacgca 240

tcaagacttg caaaggaagg aaagtccaga ttcggtttga acactatgat cgccgatcta 300

gaagattatg acttcgataa cttagacact gttccatctg ataacatcgt tatgtttgta 360

ttggctactt acggtgaagg cgaaccaaca gataacgccg tggatttcta tgagttcatt 420

actggcgaag atgcctcttt caatgagggc aacgatcctc cactaggtaa cttgaattac 480

gttgcgttcg gtctgggcaa caatacctac gaacactaca actcaatggt caggaacgtt 540

aacaaggctc tagaaaagtt aggagctcat agaattggag aagcaggtga gggtgacgac 600

ggagctggaa ctatggaaga ggacttttta gcttggaaag atccaatgtg ggaagccttg 660

gctaaaaaga tgggcttgga ggaaagagaa gctgtatatg aacctatttt cgctatcaat 720

gagagagatg atttgacccc tgaagcgaat gaggtatact tgggagaacc taataagcta 780

cacttggaag gtacagcgaa aggtccattc aactcccaca acccatatat cgcaccaatt 840

gcagaatcat acgaactttt ctcagctaag gatagaaatt gtctgcatat ggaaattgat 900

atttctggta gtaatctaaa gtatgaaaca ggcgaccata tcgcgatctg gcctaccaac 960

ccaggtgaag aggtcaacaa atttcttgac attctagatc tgtctggtaa gcaacattcc 1020

gtcgtaacag tgaaagcctt agaacctaca gccaaagttc cttttccaaa tccaactacc 1080

tacgatgcta tattgagata ccatctggaa atatgcgctc cagtttctag acagtttgtc 1140

tcaactttag cagcattcgc ccctaatgat gatatcaaag ctgagatgaa ccgtttggga 1200

tcagacaaag attacttcca cgaaaagaca ggaccacatt actacaatat cgctagattt 1260

ttggcctcag tctctaaagg tgaaaaatgg acaaagatac cattttctgc tttcatagaa 1320

ggccttacaa aactacaacc aagatactat tctatctctt cctctagttt agttcagcct 1380

aaaaagatta gtattactgc tgttgtcgaa tctcagcaaa ttccaggtag agatgaccca 1440

ttcagaggtg tagcgactaa ctacttgttc gctttgaagc agaaacaaaa cggtgatcca 1500

aatccagctc cttttggcca atcatacgag ttgacaggac caaggaataa gtatgatggt 1560

atacatgttc cagtccatgt aagacattct aactttaagc taccatctga tccaggcaaa 1620

cctattatca tgatcggtcc aggtaccggt gttgcccctt ttagaggctt cgtccaagag 1680

agggcaaaac aagccagaga tggtgtagaa gttggtaaaa cactgctgtt ctttggatgt 1740

agaaagagta cagaagattt catgtatcaa aaagagtggc aagagtacaa ggaagctctt 1800

ggcgacaaat tcgaaatgat tacagctttt tcaagagaag gatctaaaaa ggtttatgtt 1860

caacacagac tgaaggaaag atcaaaggaa gtttctgatc ttctatccca aaaagcatac 1920

ttctacgttt gcggagacgc cgcacatatg gcacgtgaag tgaacactgt gttagcacag 1980

atcatagcag aaggccgtgg tgtatcagaa gccaagggtg aggaaattgt caaaaacatg 2040

agatcagcaa atcaatacca agtgtgttct gatttcgtaa ctttacactg taaagagaca 2100

acatacgcga attcagaatt gcaagaggat gtctggagtt aa 2142

<210> 73
<211> 2133
<212> DNA
<213> Stevia rebaudiana

<400> 73
atgcaatcag attccgtaaa agtgtcgccg ttcgatctcg tatctgcagc tatgaacgga 60

aaagcaatgg agaaattgaa cgcatcggaa tcggaagatc cgacgacgct accggcgttg 120

aagatgctgg tggagaatcg cgagctgctg acactgttta cgacgtcgtt tgctgtattg 180

atcggatgtc tcgtgttttt gatgtggcgg agatcgtcct cgaagaaact ggttcaggat 240

ccggtgccgc aggtaatcgt tgttaagaag aaagagaagg agtctgaggt tgatgatggc 300

aagaagaaag tttcgatatt ctacggaact caaacaggaa ccgctgaagg ttttgccaag 360

gcacttgtag aggaagctaa agttcgatat gaaaagacat cctttaaagt tattgatctg 420

gatgattatg ctgctgatga cgatgagtat gaggagaagc ttaagaaaga atctttggcg 480

tttttctttt tggcaacgta tggagatggt gaaccaacag ataatgcagc caatttttac 540

aaatggttta cagagggaga tgacaaaggc gaatggctga agaaacttca atatggcgtg 600

tttggcctcg gtaacagaca atatgagcat ttcaataaga ttgcaatagt ggttgatgac 660

aaactcacag aaatgggcgc aaaacgcctt gttcctgtgg gtcttggaga tgacgatcaa 720

tgtatagaag atgactttac agcatggaaa gagttagtgt ggcccgagtt ggatcaattg 780

ttgcgtgatg aggatgacac gagtgttacg actccttaca ctgctgcggt tttggaatac 840

cgagttgtat atcatgataa acctgcagac tcgtatgcag aagatcaaac tcatacaaat 900

ggtcatgttg ttcatgatgc tcaacatcca tctagatcca atgtggcatt taaaaaggaa 960

ttgcacacct ctcaatctga ccggtcttgc actcatttgg aatttgatat ctctcacacc 1020

gggctatcat acgagacggg ggatcatgtt ggtgtctaca gtgagaatct aagtgaagtt 1080

gtagatgaag ctttaaaatt actcggtttg tcacccgaca cttatttctc agtccatgct 1140

gacaaggaag acggaacacc tattggcggc gcctccttgc cgccaccttt ccctccatgc 1200

actttaagag atgcattaac gcgctacgca gatgttttga gttctcctaa aaaggttgct 1260

ttgcttgctc tggctgctca tgcttctgat cctagcgaag ccgatcgatt aaaatttcta 1320

gcatctccgg ctggcaagga tgaatatgct caatggatag ttgcaaacca aagaagtctt 1380

cttgaagtta tgcagtcatt tccgtcagct aaaccgccac ttggtgtttt cttcgcagct 1440

gtcgccccac gtttacaacc tcgatattac tcgatttctt cttctccaaa gatgtcacca 1500

aacagaattc atgtgacttg tgcattagtt tatgagacaa cacctgcagg acgtattcac 1560

agaggattgt gttcaacatg gatgaagaat gctgtgcctt tgaccgaaag tccagattgt 1620

agtcaggcgt cgatttttgt tagaacgtct aacttccgac ttccggttga cccgaaagtc 1680

ccggtcatca tgatcggtcc cgggactggg ttagcccctt tcagaggttt tcttcaagaa 1740

cggttagctt tgaaggaatc tggaaccgaa ctcgggtcat ctattttctt tttcggatgc 1800

agaaaccgca aagtggattt tatatacgaa gacgaactaa acaactttgt ggagaccggt 1860

gctttatcgg agcttattgt tgcattctcc cgtgaaggaa ccgcaaagga gtatgtgcaa 1920

cataaaatga gccagaaggc ttcagatatc tggaagttgc tttcagaggg agcatattta 1980

tatgtatgtg gtgatgctaa aggcatggct aaagatgtac acagaaccct tcacacaatt 2040

gtacaagaac agggatctct agattcttcc aaggcagaat tgtatgtaaa gaacctacaa 2100

atgtcgggaa gatatcttcg tgatgtttgg taa 2133

<210> 74
<211> 2079

<212> DNA
<213> Arabidopsis thaliana

<400> 74
atgacttctg ctttgtatgc ttccgatttg tttaagcagc tcaagtcaat tatggggaca 60

gattcgttat ccgacgatgt tgtacttgtg attgcaacga cgtctttggc actagtagct 120

ggatttgtgg tgttgttatg gaagaaaacg acggcggatc ggagcgggga gctgaagcct 180

ttgatgatcc ctaagtctct tatggctaag gacgaggatg atgatttgga tttgggatcc 240

gggaagacta gagtctctat cttcttcggt acgcagactg gaacagctga gggatttgct 300

aaggcattat ccgaagaaat caaagcgaga tatgaaaaag cagcagtcaa agtcattgac 360

ttggatgact atgctgccga tgatgaccag tatgaagaga aattgaagaa ggaaactttg 420

gcatttttct gtgttgctac ttatggagat ggagagccta ctgacaatgc tgccagattt 480

tacaaatggt ttacggagga aaatgaacgg gatataaagc ttcaacaact agcatatggt 540

gtgtttgctc ttggtaatcg ccaatatgaa cattttaata agatcgggat agttcttgat 600

gaagagttat gtaagaaagg tgcaaagcgt cttattgaag tcggtctagg agatgatgat 660

cagagcattg aggatgattt taatgcctgg aaagaatcac tatggtctga gctagacaag 720

ctcctcaaag acgaggatga taaaagtgtg gcaactcctt atacagctgt tattcctgaa 780

taccgggtgg tgactcatga tcctcggttt acaactcaaa aatcaatgga atcaaatgtg 840

gccaatggaa atactactat tgacattcat catccctgca gagttgatgt tgctgtgcag 900

aaggagcttc acacacatga atctgatcgg tcttgcattc atctcgagtt cgacatatcc 960

aggacgggta ttacatatga aacaggtgac catgtaggtg tatatgctga aaatcatgtt 1020

gaaatagttg aagaagctgg aaaattgctt ggccactctt tagatttagt attttccata 1080

catgctgaca aggaagatgg ctccccattg gaaagcgcag tgccgcctcc tttccctggt 1140

ccatgcacac ttgggactgg tttggcaaga tacgcagacc ttttgaaccc tcctcgaaag 1200

tctgcgttag ttgccttggc ggcctatgcc actgaaccaa gtgaagccga gaaacttaag 1260

cacctgacat cacctgatgg aaaggatgag tactcacaat ggattgttgc aagtcagaga 1320

agtcttttag aggtgatggc tgcttttcca tctgcaaaac ccccactagg tgtatttttt 1380

gctgcaatag ctcctcgtct acaacctcgt tactactcca tctcatcctc gccaagattg 1440

gcgccaagta gagttcatgt tacatccgca ctagtatatg gtccaactcc tactggtaga 1500

atccacaagg gtgtgtgttc tacgtggatg aagaatgcag ttcctgcgga gaaaagtcat 1560

gaatgtagtg gagccccaat ctttattcga gcatctaatt tcaagttacc atccaaccct 1620

tcaactccaa tcgttatggt gggacctggg actgggctgg caccttttag aggttttctg 1680

caggaaagga tggcactaaa agaagatgga gaagaactag gttcatcttt gctcttcttt 1740

gggtgtagaa atcgacagat ggactttata tacgaggatg agctcaataa ttttgttgat 1800

caaggcgtaa tatctgagct catcatggca ttctcccgtg aaggagctca gaaggagtat 1860

gttcaacata agatgatgga gaaggcagca caagtttggg atctaataaa ggaagaagga 1920

tatctctatg tatgcggtga tgctaagggc atggcgaggg acgtccaccg aactctacac 1980

accattgttc aggagcagga aggtgtgagt tcgtcagagg cagaggctat agttaagaaa 2040

cttcaaaccg aaggaagata cctcagagat gtctggtga 2079

<210> 75
<211> 2142
<212> DNA
<213> Giberella fujikuroi

<400> 75
atggctgaac tcgacactct ggacatcgtc gtcctcggcg ttatcttcct cggaacggtt 60

gcatacttta caaagggcaa gctatggggt gttaccaagg atccctacgc gaatggcttc 120

gctgccggcg gcgcttctaa gccgggtcgc acgaggaaca tcgtcgaggc aatggaagaa 180

tccggcaaga actgtgttgt cttctatggt tctcagaccg gtactgctga agattatgct 240

tctcgcctcg ccaaggaggg taagagtcga ttcggactaa acaccatgat tgccgatctt 300

gaggactacg atttcgacaa cctggatacc gttcccagtg acaacattgt catgttcgtt 360

ctcgcaactt atggtgaagg tgagcctacc gataacgcgg tcgacttcta tgaattcatt 420

accggcgagg atgccagctt caatgagggc aatgatcctc cgctgggcaa cctcaactac 480

gttgctttcg gtctcggaaa caacacgtac gagcactaca actctatggt ccgcaatgtt 540

aacaaggctc tcgagaagct tggcgctcac cgcatcggtg aagctggtga gggtgatgat 600

ggtgctggta ccatggaaga ggacttcttg gcctggaagg atcccatgtg ggaagccctc 660

gctaagaaaa tgggactgga agagcgtgaa gcagtctacg agcctatttt tgccattaac 720

gaacgcgacg acctgactcc tgaagccaat gaagtgtatc tcggtgagcc caacaagctg 780

catctcgaag gcaccgccaa gggaccattc aactctcaca acccctacat tgcccctatc 840

gctgaatctt atgagttgtt ctccgccaag gacagaaact gcctccacat ggaaattgac 900

atcagcggtt ctaacctcaa gtacgaaact ggagaccata ttgctatctg gcctaccaac 960

cctggtgagg aggtcaacaa attcctggat attctcgacc tctctggaaa gcagcacagc 1020

gttgtcactg tcaaggctct cgagcctacc gccaaggttc ctttccccaa ccctacaacc 1080

tacgatgcca ttctgcgata ccacctcgag atctgcgctc ctgtttcacg tcaattcgtc 1140

tctactctcg ccgcatttgc tcccaacgat gatatcaagg ctgagatgaa ccgccttggc 1200

agcgataagg attatttcca cgagaagact ggcccgcatt actacaacat tgcccgtttc 1260

cttgccagcg tcagcaaggg cgagaagtgg accaaaatcc cgttctctgc cttcatcgag 1320

ggtctcacca agctccagcc ccgttactac tccatttctt cctcgtctct ggttcagccc 1380

aagaaaatct cgatcactgc cgtcgttgaa tcccagcaga ttcctggccg ggatgatcct 1440

ttccgtggtg ttgctacaaa ctatcttttt gccctaaagc aaaagcagaa cggtgacccc 1500

aaccctgcac cttttggtca gagctacgag cttacaggcc cccgcaataa gtatgatggc 1560

atccacgttc ctgtccatgt tcgtcactcc aacttcaagc tcccctcgga ccccggtaag 1620

cccatcatca tgattggtcc tggtactggt gtcgctccct tccgcggttt cgtgcaggag 1680

cgtgctaagc aagcccgtga tggtgttgag gttggaaaga cactcttgtt ctttggttgc 1740

cgaaagtcaa ccgaggattt catgtaccaa aaggagtggc aggaatacaa ggaggctctt 1800

ggcgataagt ttgaaatgat caccgccttt tctcgagagg gctccaagaa ggtttatgtt 1860

cagcaccgac ttaaggagcg atctaaggag gtcagcgatc tgctctccca gaaggcttat 1920

ttctatgtct gcggtgatgc agcccacatg gcccgcgagg tcaataccgt cttggcacaa 1980

atcattgccg agggacgtgg ggtgtctgag gccaagggcg aggagatcgt gaagaacatg 2040

agatcagcga accaatacca ggtatgtagt gactttgtta ctcttcactg caaagaaacc 2100

acatatgcta actcagaatt acaggaggat gtttggtcat ag 2142

<210> 76
<211> 459
<212> PRT
<213> Ipomoea purpurea

<400> 76
Met Gly Ser Gln Ala Thr Thr Tyr His Met Ala Met Tyr Pro Trp Phe
1 5 10 15

Gly Val Gly His Leu Thr Gly Phe Phe Arg Leu Ala Asn Lys Leu Ala
20 25 30

Gly Lys Gly His Arg Ile Ser Phe Leu Ile Pro Lys Asn Thr Gln Ser
35 40 45

Lys Leu Glu Ser Phe Asn Leu His Pro His Leu Ile Ser Phe Val Pro
50 55 60

Ile Val Val Pro Ser Ile Pro Gly Leu Pro Pro Gly Ala Glu Thr Thr
65 70 75 80

Ser Asp Val Pro Phe Pro Ser Thr His Leu Leu Met Glu Ala Met Asp
85 90 95

Lys Thr Gln Asn Asp Ile Glu Ile Ile Leu Lys Asp Leu Lys Val Asp
100 105 110

Val Val Phe Tyr Asp Phe Thr His Trp Leu Pro Ser Leu Ala Arg Lys

115 120 125

Ile Gly Ile Lys Ser Val Phe Tyr Ser Thr Ile Ser Pro Leu Met His
130 135 140

Gly Tyr Ala Leu Ser Pro Glu Arg Arg Val Val Gly Lys Gln Leu Thr
145 150 155 160

Glu Ala Asp Met Met Lys Ala Pro Ala Ser Phe Pro Asp Pro Ser Ile
165 170 175

Lys Leu His Ala His Glu Ala Arg Gly Phe Thr Ala Arg Thr Val Met
180 185 190

Lys Phe Gly Gly Asp Ile Thr Phe Phe Asp Arg Ile Phe Thr Ala Val
195 200 205

Ser Glu Ser Asp Gly Leu Ala Tyr Ser Thr Cys Arg Glu Ile Glu Gly
210 215 220

Gln Phe Cys Asp Tyr Ile Glu Thr Gln Phe Gln Lys Pro Val Leu Leu
225 230 235 240

Ala Gly Pro Ala Leu Pro Val Pro Ser Lys Ser Thr Met Glu Gln Lys
245 250 255

Trp Ser Asp Trp Leu Gly Lys Phe Lys Glu Gly Ser Val Ile Tyr Cys
260 265 270

Ala Phe Gly Ser Glu Cys Thr Leu Arg Lys Asp Lys Phe Gln Glu Leu
275 280 285

Leu Trp Gly Leu Glu Leu Thr Gly Met Pro Phe Phe Ala Ala Leu Lys
290 295 300

Pro Pro Phe Glu Thr Glu Ser Val Glu Ala Ala Ile Pro Glu Glu Leu
305 310 315 320

Lys Glu Lys Ile Gln Gly Arg Gly Ile Val His Gly Glu Trp Val Gln
325 330 335

Gln Gln Leu Phe Leu Gln His Pro Ser Val Gly Cys Phe Val Ser His
340 345 350

Cys Gly Trp Ala Ser Leu Ser Glu Ala Leu Val Asn Asp Cys Gln Ile
355 360 365

Val Leu Leu Pro Gln Val Gly Asp Gln Ile Ile Asn Ala Arg Ile Met
370 375 380

Ser Val Ser Leu Lys Val Gly Val Glu Val Glu Lys Gly Glu Glu Asp
385 390 395 400

Gly Val Phe Ser Arg Glu Ser Val Cys Lys Ala Val Lys Ala Val Met
405 410 415

Asp Glu Lys Ser Glu Ile Gly Arg Glu Val Arg Gly Asn His Asp Lys
420 425 430

Leu Arg Gly Phe Leu Met Asn Ala Asp Leu Asp Ser Lys Tyr Met Asp
435 440 445

Ser Phe Asn Gln Lys Leu Gln Asp Leu Leu Gly
450 455

<210> 77
<211> 1380
<212> DNA
<213> Ipomoea purpurea

<400> 77
atgggttctc aagctacaac ttaccatatg gccatgtatc catggtttgg ggttggacat 60

ttgactggtt tcttccgttt ggcaaacaaa ttagctggca aaggacatag aatctcattt 120

ctaattccta aaaacactca atctaagtta gaatctttca accttcatcc acacttaatc 180

tcttttgtgc ctatcgttgt cccaagtata ccaggcctgc cacctggtgc agagactaca 240

tcagatgttc ctttcccaag tacacatttg ctaatggaag caatggacaa gactcaaaac 300

gatatagaga ttatcctgaa ggatcttaaa gtagatgttg ttttctatga ttttactcac 360

tggttgcctt ctctggccag aaagattggc attaagagtg tcttttactc caccatttct 420

cctttaatgc atggatatgc tttatcacca gaaagacgtg tagttggtaa gcaattgaca 480

gaggcagata tgatgaaggc cccagcttct ttcccagacc catccattaa gctacatgca 540

catgaagcta ggggttttac agccagaacc gttatgaaat tcggtggtga catcaccttt 600

ttcgatagaa tattcacagc agtttccgaa agtgatggcc tggcctactc tacttgtaga 660

gagatcgagg gacaattctg tgattacatt gaaacacaat tccagaagcc agtcttgtta 720

gccggtccag ctttgccagt cccatccaaa tccactatgg aacaaaagtg gtcagattgg 780

ttggggaaat tcaaggaagg ctccgtcatc tactgtgctt tcgggtctga atgtacattg 840

agaaaggaca aatttcagga acttttatgg ggtttggaat tgacaggaat gcctttcttc 900

gctgctctga agccaccttt tgagactgag tctgttgagg ctgctatccc tgaggaacta 960

aaggaaaaga ttcagggaag aggtatagta catggagaat gggtacaaca acaattgttt 1020

cttcaacacc catctgtcgg gtgcttcgtt tctcactgcg gctgggcaag tttatctgaa 1080

gcccttgtta atgattgtca aatcgtgtta cttccacaag ttggcgatca gattatcaac 1140

gccagaataa tgtcagtatc acttaaagtg ggcgtggaag ttgaaaaggg tgaggaggac 1200

ggtgtctttt caagagaatc tgtgtgcaag gctgttaaag cagtaatgga tgaaaaatct 1260

gaaatcggta gagaagtcag aggtaatcat gataaactga ggggtttctt gatgaatgca 1320

gacttagatt caaagtacat ggattcattc aatcaaaagc tacaagattt gctaggttaa 1380

<210> 78
<211> 438
<212> PRT
<213> Bellis perennis

<400> 78

```
Met Asp Ser Lys Ile Asp Ser Lys Thr Phe Arg Val Val Met Leu Pro
1               5                   10                  15

Trp Leu Ala Tyr Ser His Ile Ser Ser Phe Leu Val Phe Ala Lys Arg
            20                  25                  30

Leu Thr Asn His Asn Phe His Ile Tyr Ile Cys Ser Ser Gln Thr Asn
        35                  40                  45

Met Gln Tyr Leu Lys Asn Asn Leu Thr Ser Gln Tyr Ser Lys Ser Ile
    50                  55                  60

Gln Leu Ile Glu Leu Asn Leu Pro Ser Ser Ser Glu Leu Pro Leu Gln
65                  70                  75                  80

Tyr His Thr Thr His Gly Leu Pro Pro His Leu Thr Lys Thr Leu Ser
                85                  90                  95

Asp Asp Tyr Gln Lys Ser Gly Pro Asp Phe Glu Thr Ile Leu Ile Lys
            100                 105                 110

Leu Asn Pro His Leu Val Ile Tyr Asp Phe Asn Gln Leu Trp Ala Pro
        115                 120                 125

Glu Val Ala Ser Thr Leu His Ile Pro Ser Ile Gln Leu Leu Ser Gly
    130                 135                 140

Cys Val Ala Leu Tyr Ala Leu Asp Ala His Leu Tyr Thr Lys Pro Leu
145                 150                 155                 160

Asp Glu Asn Leu Ala Lys Phe Pro Phe Pro Glu Ile Tyr Pro Lys Asn
                165                 170                 175
```

```
Arg Asp Ile Pro Lys Gly Gly Ser Lys Tyr Ile Glu Arg Phe Val Asp
            180                 185                 190

Cys Met Arg Arg Ser Cys Glu Ile Ile Leu Val Arg Ser Thr Met Glu
        195                 200                 205

Leu Glu Gly Lys Tyr Ile Asp Tyr Leu Ser Lys Thr Leu Gly Lys Lys
    210                 215                 220

Val Leu Pro Val Gly Pro Leu Val Gln Glu Ala Ser Leu Leu Gln Asp
225                 230                 235                 240

Asp His Ile Trp Ile Met Lys Trp Leu Asp Lys Lys Glu Glu Ser Ser
                245                 250                 255

Val Val Phe Val Cys Phe Gly Ser Glu Tyr Ile Leu Ser Asp Asn Glu
            260                 265                 270

Ile Glu Asp Ile Ala Tyr Gly Leu Glu Leu Ser Gln Val Ser Phe Val
        275                 280                 285

Trp Ala Ile Arg Ala Lys Thr Ser Ala Leu Asn Gly Phe Ile Asp Arg
    290                 295                 300

Val Gly Asp Lys Gly Leu Val Ile Asp Lys Trp Val Pro Gln Ala Asn
305                 310                 315                 320

Ile Leu Ser His Ser Ser Thr Gly Gly Phe Ile Ser His Cys Gly Trp
                325                 330                 335

Ser Ser Thr Met Glu Ser Ile Arg Tyr Gly Val Pro Ile Ile Ala Met
            340                 345                 350

Pro Met Gln Phe Asp Gln Pro Tyr Asn Ala Arg Leu Met Glu Thr Val
        355                 360                 365

Gly Ala Gly Ile Glu Val Gly Arg Asp Gly Glu Gly Arg Leu Lys Arg
    370                 375                 380

Glu Glu Ile Ala Ala Val Val Arg Lys Val Val Val Glu Asp Ser Gly
385                 390                 395                 400

Glu Ser Ile Arg Glu Lys Ala Lys Glu Leu Gly Glu Ile Met Lys Lys
                405                 410                 415

Asn Met Glu Ala Glu Val Asp Gly Ile Val Ile Glu Asn Leu Val Lys
            420                 425                 430
```

Leu Cys Glu Met Asn Asn
435

<210> 79
<211> 1317
<212> DNA
<213> Bellis perennis

<400> 79
atggattcta aaatcgattc aaagacattc agagtcgtta tgttgccttg gcttgcatac 60

tcacacattt catcattcct agtgtttgcc aagagactaa caaatcataa cttccacatc 120

tacatttgtt cctctcaaac aaatatgcaa tacctgaaaa acaacttgac gtctcagtat 180

tcaaaatcta tacaactgat tgagttgaat cttccatcta gttccgaatt gcctctgcag 240

tatcatacta ctcacggact accaccacac cttacgaaaa cattgtctga tgattatcaa 300

aagtccggac ctgactttga aaccattttg atcaaattga acccacatct ggtaatctac 360

gactttaatc aactttgggc tccagaggtt gctagtacac ttcatattcc atccatacag 420

ttactgtctg gttgcgtcgc cttatatgcc ttagacgccc atctgtacac aaagccacta 480

gacgaaaact tggctaagtt tcctttccca gaaatctatc ctaaaaacag agatattcct 540

aagggaggta gtaaatacat cgaaaggttc gtagactgta tgagaagatc ttgtgaaatc 600

atattagtca gaagtaccat ggaacttgaa ggaaaataca ttgattactt gtctaagaca 660

ttagggaaaa aggtgttgcc agtagggcct ctggtgcaag aggcttcttt gttgcaagat 720

gatcatatat ggattatgaa gtggttagac aaaaaggagg agtcatccgt cgtgtttgtt 780

tgttttggtt ctgagtacat cttatcagac aacgaaatag aagatattgc ttatggccta 840

gagttgtccc aagtaagttt cgtttgggca ataagagcta agacttctgc cttaaatggc 900

ttcattgata gagtgggtga taaaggctta gtcatcgata aatgggttcc acaggctaac 960

atcttatctc actcttctac tggtggattc attagtcatt gcggttggtc atcaacaatg 1020

gaatctatta gatatggggt tcctattatc gccatgccaa tgcaattcga tcaaccttac 1080

aatgctaggt tgatggaaac tgttggtgca ggtatcgaag ttggcagaga tggcgaaggt 1140

agattgaaaa gagaagagat tgctgccgtg gttagaaagg tcgttgttga agattctggg 1200

gaatccataa gggagaaggc aaaggaattg ggagaaatca tgaaaaaaaa catggaggcc 1260

gaagtagatg gtatagtgat tgaaaatcta gttaagctat gtgagatgaa caattaa 1317

<210> 80
<211> 2490
<212> DNA
<213> Zea mays

<400> 80

```
atggttttgt cttcttcttg tactacagta ccacacttat cttcattagc tgtcgtgcaa       60
cttggtcctt ggagcagtag gattaaaaag aaaaccgata ctgttgcagt accagccgct      120
gcaggaaggt ggagaagggc cttggctaga gcacagcaca catcagaatc cgcagctgtc      180
gcaaagggca gcagtttgac ccctatagtg agaactgacg ctgagtcaag gagaacaaga      240
tggccaaccg atgacgatga cgccgaacct ttagtggatg agatcagggc aatgcttact      300
tccatgtctg atggtgacat ttccgtgagc gcatacgata cagcctgggt cggattggtt      360
ccaagattag acggcggtga aggtcctcaa tttccagcag ctgtgagatg gataagaaat      420
aaccagttgc ctgacggaag ttggggcgat gccgcattat tctctgccta tgacaggctt      480
atcaataccc ttgcctgcgt tgtaactttg acaaggtggt ccctagaacc agagatgaga      540
ggtagaggac tatcttttt gggtaggaac atgtggaaat tagcaactga agatgaagag      600
tcaatgccta ttggcttcga attagcattt ccatctttga tagagcttgc taagagccta      660
ggtgtccatg acttccctta tgatcaccag gccctacaag gaatctactc ttcaagagag      720
atcaaaatga agaggattcc aaaagaagtg atgcataccg ttccaacatc aatattgcac      780
agtttggagg gtatgcctgg cctagattgg gctaaactac ttaaactaca gagcagcgac      840
ggaagttttt tgttctcacc agctgccact gcatatgctt taatgaatac cggagatgac      900
aggtgtttta gctacatcga tagaacagta aagaaattca acggcggcgt ccctaatgtt      960
tatccagtgg atctatttga acatatttgg gccgttgata gacttgaaag attaggaatc     1020
tccaggtact tccaaaagga gatcgaacaa tgcatggatt atgtaaacag gcattggact     1080
gaggacggta tttgttgggc aaggaactct gatgtcaaag aggtggacga cacagctatg     1140
gcctttagac ttcttaggtt gcacggctac agcgtcagtc ctgatgtgtt taaaaacttc     1200
gaaaaggacg gtgaattttt cgcatttgtc ggacagtcta atcaagctgt taccggtatg     1260
tacaacttaa acagagcaag ccagatatcc ttcccaggcg aggatgtgct tcatagagct     1320
ggtgccttct catatgagtt cttgaggaga aaagaagcag agggagcttt gagggacaag     1380
tggatcattt ctaaagatct acctggtgaa gttgtgtata ctttggattt tccatggtac     1440
ggcaacttac ctagagtcga ggccagagac tacctagagc aatacggagg tggtgatgac     1500
gtttggattg gcaagacatt gtataggatg ccacttgtaa acaatgatgt atatttggaa     1560
ttggcaagaa tggatttcaa ccactgccag gctttgcatc agttagagtg gcaaggacta     1620
aaaagatggt atactgaaaa taggttgatg gactttggtg tcgcccaaga agatgccctt     1680
agagcttatt ttcttgcagc cgcatctgtt tacgagcctt gtagagctgc cgagaggctt     1740
gcatgggcta gagccgcaat actagctaac gccgtgagca cccacttaag aaatagccca     1800
tcattcagag aaaggttaga gcattctctt aggtgtagac ctagtgaaga gacagatggc     1860
tcctggttta actcctcaag tggctctgat gcagttttag taaaggctgt cttaagactt     1920
```

actgattcat tagccaggga agcacagcca atccatggag gtgacccaga agatattata 1980

cacaagttgt taagatctgc ttgggccgag tgggttaggg aaaaggcaga cgctgccgat 2040

agcgtgtgca atggtagttc tgcagtagaa caagagggat caagaatggt ccatgataaa 2100

cagacctgtc tattattggc tagaatgatc gaaatttctg ccggtagggc agctggtgaa 2160

gcagccagtg aggacggcga tagaagaata attcaattaa caggctccat ctgcgacagt 2220

cttaagcaaa aaatgctagt ttcacaggac cctgaaaaaa atgaagagat gatgtctcac 2280

gtggatgacg aattgaagtt gaggattaga gagttcgttc aatatttgct tagactaggt 2340

gaaaaaaaga ctggatctag cgaaaccagg caaacatttt taagtatagt gaaatcatgt 2400

tactatgctg ctcattgccc acctcatgtc gttgatagac acattagtag agtgattttc 2460

gagccagtaa gtgccgcaaa gtaaccgcgg 2490

<210> 81
<211> 827
<212> PRT
<213> Zea mays

<400> 81

```
Met Val Leu Ser Ser Ser Cys Thr Thr Val Pro His Leu Ser Ser Leu
1               5                  10                  15

Ala Val Val Gln Leu Gly Pro Trp Ser Ser Arg Ile Lys Lys Lys Thr
            20                  25                  30

Asp Thr Val Ala Val Pro Ala Ala Ala Gly Arg Trp Arg Arg Ala Leu
        35                  40                  45

Ala Arg Ala Gln His Thr Ser Glu Ser Ala Ala Val Ala Lys Gly Ser
    50                  55                  60

Ser Leu Thr Pro Ile Val Arg Thr Asp Ala Glu Ser Arg Arg Thr Arg
65                  70                  75                  80

Trp Pro Thr Asp Asp Asp Asp Ala Glu Pro Leu Val Asp Glu Ile Arg
                85                  90                  95

Ala Met Leu Thr Ser Met Ser Asp Gly Asp Ile Ser Val Ser Ala Tyr
            100                 105                 110

Asp Thr Ala Trp Val Gly Leu Val Pro Arg Leu Asp Gly Gly Glu Gly
        115                 120                 125

Pro Gln Phe Pro Ala Ala Val Arg Trp Ile Arg Asn Asn Gln Leu Pro
    130                 135                 140
```

```
Asp Gly Ser Trp Gly Asp Ala Ala Leu Phe Ser Ala Tyr Asp Arg Leu
145                 150                 155                 160

Ile Asn Thr Leu Ala Cys Val Val Thr Leu Thr Arg Trp Ser Leu Glu
                165                 170                 175

Pro Glu Met Arg Gly Arg Gly Leu Ser Phe Leu Gly Arg Asn Met Trp
            180                 185                 190

Lys Leu Ala Thr Glu Asp Glu Glu Ser Met Pro Ile Gly Phe Glu Leu
        195                 200                 205

Ala Phe Pro Ser Leu Ile Glu Leu Ala Lys Ser Leu Gly Val His Asp
    210                 215                 220

Phe Pro Tyr Asp His Gln Ala Leu Gln Gly Ile Tyr Ser Ser Arg Glu
225                 230                 235                 240

Ile Lys Met Lys Arg Ile Pro Lys Glu Val Met His Thr Val Pro Thr
                245                 250                 255

Ser Ile Leu His Ser Leu Glu Gly Met Pro Gly Leu Asp Trp Ala Lys
            260                 265                 270

Leu Leu Lys Leu Gln Ser Ser Asp Gly Ser Phe Leu Phe Ser Pro Ala
        275                 280                 285

Ala Thr Ala Tyr Ala Leu Met Asn Thr Gly Asp Asp Arg Cys Phe Ser
    290                 295                 300

Tyr Ile Asp Arg Thr Val Lys Lys Phe Asn Gly Gly Val Pro Asn Val
305                 310                 315                 320

Tyr Pro Val Asp Leu Phe Glu His Ile Trp Ala Val Asp Arg Leu Glu
                325                 330                 335

Arg Leu Gly Ile Ser Arg Tyr Phe Gln Lys Glu Ile Glu Gln Cys Met
            340                 345                 350

Asp Tyr Val Asn Arg His Trp Thr Glu Asp Gly Ile Cys Trp Ala Arg
        355                 360                 365

Asn Ser Asp Val Lys Glu Val Asp Asp Thr Ala Met Ala Phe Arg Leu
    370                 375                 380

Leu Arg Leu His Gly Tyr Ser Val Ser Pro Asp Val Phe Lys Asn Phe
385                 390                 395                 400
```

Glu Lys Asp Gly Glu Phe Phe Ala Phe Val Gly Gln Ser Asn Gln Ala
405 410 415

Val Thr Gly Met Tyr Asn Leu Asn Arg Ala Ser Gln Ile Ser Phe Pro
420 425 430

Gly Glu Asp Val Leu His Arg Ala Gly Ala Phe Ser Tyr Glu Phe Leu
435 440 445

Arg Arg Lys Glu Ala Glu Gly Ala Leu Arg Asp Lys Trp Ile Ile Ser
450 455 460

Lys Asp Leu Pro Gly Glu Val Val Tyr Thr Leu Asp Phe Pro Trp Tyr
465 470 475 480

Gly Asn Leu Pro Arg Val Glu Ala Arg Asp Tyr Leu Glu Gln Tyr Gly
485 490 495

Gly Gly Asp Asp Val Trp Ile Gly Lys Thr Leu Tyr Arg Met Pro Leu
500 505 510

Val Asn Asn Asp Val Tyr Leu Glu Leu Ala Arg Met Asp Phe Asn His
515 520 525

Cys Gln Ala Leu His Gln Leu Glu Trp Gln Gly Leu Lys Arg Trp Tyr
530 535 540

Thr Glu Asn Arg Leu Met Asp Phe Gly Val Ala Gln Glu Asp Ala Leu
545 550 555 560

Arg Ala Tyr Phe Leu Ala Ala Ala Ser Val Tyr Glu Pro Cys Arg Ala
565 570 575

Ala Glu Arg Leu Ala Trp Ala Arg Ala Ala Ile Leu Ala Asn Ala Val
580 585 590

Ser Thr His Leu Arg Asn Ser Pro Ser Phe Arg Glu Arg Leu Glu His
595 600 605

Ser Leu Arg Cys Arg Pro Ser Glu Glu Thr Asp Gly Ser Trp Phe Asn
610 615 620

Ser Ser Ser Gly Ser Asp Ala Val Leu Val Lys Ala Val Leu Arg Leu
625 630 635 640

Thr Asp Ser Leu Ala Arg Glu Ala Gln Pro Ile His Gly Gly Asp Pro

645 650 655

Glu Asp Ile Ile His Lys Leu Leu Arg Ser Ala Trp Ala Glu Trp Val
660 665 670

Arg Glu Lys Ala Asp Ala Ala Asp Ser Val Cys Asn Gly Ser Ser Ala
675 680 685

Val Glu Gln Glu Gly Ser Arg Met Val His Asp Lys Gln Thr Cys Leu
690 695 700

Leu Leu Ala Arg Met Ile Glu Ile Ser Ala Gly Arg Ala Ala Gly Glu
705 710 715 720

Ala Ala Ser Glu Asp Gly Asp Arg Arg Ile Ile Gln Leu Thr Gly Ser
725 730 735

Ile Cys Asp Ser Leu Lys Gln Lys Met Leu Val Ser Gln Asp Pro Glu
740 745 750

Lys Asn Glu Glu Met Met Ser His Val Asp Asp Glu Leu Lys Leu Arg
755 760 765

Ile Arg Glu Phe Val Gln Tyr Leu Leu Arg Leu Gly Glu Lys Lys Thr
770 775 780

Gly Ser Ser Glu Thr Arg Gln Thr Phe Leu Ser Ile Val Lys Ser Cys
785 790 795 800

Tyr Tyr Ala Ala His Cys Pro Pro His Val Val Asp Arg His Ile Ser
805 810 815

Arg Val Ile Phe Glu Pro Val Ser Ala Ala Lys
820 825

<210> 82
<211> 1461
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 82
actagtaaaa tggatgcaat ggcaactact gagaaaaagc ctcatgtgat cttcattcca 60

tttcctgcac aatctcacat aaaggcaatg ctaaagttag cacaactatt acaccataag 120

ggattacaga taactttcgt gaataccgac ttcatccata atcaatttct ggaatctagt 180

ggccctcatt gtttggacgg agccccaggg tttagattcg aaacaattcc tgacggtgtt 240

tcacattccc cagaggcctc catcccaata agagagagtt tactgaggtc aatagaaacc 300

aacttttgg atcgtttcat tgacttggtc acaaaacttc cagacccacc aacttgcata 360

atctctgatg gctttctgtc agtgtttact atcgacgctg ccaaaaagtt gggtatccca 420

gttatgatgt actggactct tgctgcatgc ggtttcatgg gtttctatca catccattct 480

cttatcgaaa agggttttgc tccactgaaa gatgcatcat acttaaccaa cggctacctg 540

gatactgtta ttgactgggt accaggtatg gaaggtataa gacttaaaga ttttcctttg 600

gattggtcta cagaccttaa tgataaagta ttgatgttta ctacagaagc tccacaaaga 660

tctcataagg tttcacatca tatctttcac acctttgatg aattggaacc atcaatcatc 720

aaaaccttgt ctctaagata caatcatatc tacactattg gtccattaca attacttcta 780

gatcaaattc ctgaagagaa aaagcaaact ggtattacat ccttacacgg ctactcttta 840

gtgaaagagg aaccagaatg ttttcaatgg ctacaaagta aagagcctaa ttctgtggtc 900

tacgtcaact tcggaagtac aacagtcatg tccttggaag atatgactga atttggttgg 960

ggccttgcta attcaaatca ttactttcta tggattatca ggtccaattt ggtaataggg 1020

gaaaacgccg tattacctcc agaattggag gaacacatca aaaagagagg tttcattgct 1080

tcctggtgtt ctcaggaaaa ggtattgaaa catccttctg ttggtggttt ccttactcat 1140

tgcggttggg gctctacaat cgaatcacta agtgcaggag ttccaatgat ttgttggcca 1200

tattcatggg accaacttac aaattgtagg tatatctgta aagagtggga agttggatta 1260

gaaatgggaa caaaggttaa acgtgatgaa gtgaaaagat tggttcagga gttgatgggg 1320

gaaggtggcc acaagatgag aaacaaggcc aaagattgga aggaaaaagc cagaattgct 1380

attgctccta acgggtcatc ctctctaaac attgataaga tggtcaaaga gattacagtc 1440

ttagccagaa actaagtcga c 1461

<210> 83
<211> 1398
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 83
actagtaaaa tggcagagca acaaaagatc aaaaagtcac ctcacgtctt acttattcca 60

tttcctctgc aaggacatat caacccattc atacaatttg ggaaaagatt gattagtaag 120

ggtgtaaaga caacactggt aaccactatc cacactttga attctactct gaaccactca 180

aatactacta ctacaagtat agaaattcaa gctatatcag acggatgcga tgagggtggc 240

tttatgtctg ccggtgaatc ttacttggaa acattcaagc aagtgggatc caagtctctg 300

gccgatctaa tcaaaaagtt acagagtgaa ggcaccacaa ttgacgccat aatctacgat 360

tctatgacag agtgggtttt agacgttgct atcgaatttg gtattgatgg aggttccttt 420

ttcacacaag catgtgttgt gaattctcta tactaccatg tgcataaagg gttaatctct 480

ttaccattgg gtgaaactgt ttcagttcca ggttttccag tgttacaacg ttgggaaacc 540

ccattgatct tacaaaatca tgaacaaata caatcacctt ggtcccagat gttgtttggt 600

caattcgcta acatcgatca agcaagatgg gtctttacta attcattcta taagttagag 660

gaagaggtaa ttgaatggac taggaagatc tggaatttga aagtcattgg tccaacattg 720

ccatcaatgt atttggacaa aagacttgat gatgataaag ataatggttt caatttgtac 780

aaggctaatc atcacgaatg tatgaattgg ctggatgaca aaccaaagga atcagttgta 840

tatgttgctt tcggctctct tgttaaacat ggtccagaac aagttgagga gattacaaga 900

gcacttatag actctgacgt aaacttttg tgggtcatta agcacaaaga ggaggggaaa 960

ctgccagaaa acctttctga agtgataaag accggaaaag gtctaatcgt tgcttggtgt 1020

aaacaattgg atgttttagc tcatgaatct gtaggctgtt ttgtaacaca ttgcggattc 1080

aactctacac tagaagccat ttccttaggc gtacctgtcg ttgcaatgcc tcagttctcc 1140

gatcagacaa ccaacgctaa acttttggac gaaatactag gggtgggtgt cagagttaaa 1200

gcagacgaga atggtatcgt cagaagaggg aacctagctt catgtatcaa aatgatcatg 1260

gaagaggaaa gaggagttat cataaggaaa aacgcagtta agtggaagga tcttgcaaag 1320

gttgccgtcc atgaaggcgg ctcttcagat aatgatattg ttgaatttgt gtccgaacta 1380

atcaaagcct aagtcgac 1398

<210> 84
<211> 1437
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 84
actagtaaaa tggctacatc tgattctatt gttgatgaca ggaagcagtt gcatgtggct 60

actttccctt ggcttgcttt cggtcatata ctgccttacc tacaactatc aaaactgata 120

gctgaaaaag gacataaagt gtcattcctt tcaacaacta gaaacattca aagattatct 180

tcccacatat caccattgat taacgtcgtt caattgacac ttccaagagt acaggaatta 240

ccagaagatg ctgaagctac aacagatgtg catcctgaag atatccctta cttgaaaaag 300

gcatccgatg gattacagcc tgaggtcact agattccttg agcaacacag tccagattgg 360

atcatatacg actacactca ctattggttg ccttcaattg cagcatcact aggcatttct 420

agggcacatt tcagtgtaac cacaccttgg gccattgctt acatgggtcc atccgctgat 480

gctatgatta acggcagtga tggtagaact accgttgaag atttgacaac cccaccaaag 540

tggtttccat ttccaactaa agtctgttgg agaaaacacg acttagcaag actggttcca 600

tacaaggcac caggaatctc agacggctat agaatgggtt tagtccttaa agggtctgac 660

tgcctattgt ctaagtgtta ccatgagttt gggacacaat ggctaccact tttggaaaca 720

ttacaccaag ttcctgtcgt accagttggt ctattacctc cagaaatccc tggtgatgag 780

aaggacgaga cttgggtttc aatcaaaaag tggttagacg ggaagcaaaa aggctcagtg 840

gtatatgtgg cactgggttc cgaagtttta gtatctcaaa cagaagttgt ggaacttgcc 900

ttaggtttgg aactatctgg attgccattt gtctgggcct acagaaaacc aaaaggccct 960

gcaaagtccg attcagttga attgccagac ggctttgtcg agagaactag agatagaggg 1020

ttggtatgga cttcatgggc tccacaattg agaatcctga gtcacgaatc tgtgtgcggt 1080

ttcctaacac attgtggttc tggttctata gttgaaggac tgatgtttgg tcatccactt 1140

atcatgttgc caatctttgg tgaccagcct ttgaatgcac gtctgttaga agataaacaa 1200

gttggaattg aaatcccacg taatgaggaa gatggatgtt taaccaagga gtctgtggcc 1260

agatcattac gttccgttgt cgttgaaaag gaaggcgaaa tctacaaggc caatgcccgt 1320

gaactttcaa agatctacaa tgacacaaaa gtagagaagg aatatgtttc tcaatttgta 1380

gattacctag agaaaaacgc tagagccgta gctattgatc atgaatccta agtcgac 1437

<210> 85
<211> 1392
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 85
actagtaaaa tggaaaacaa gaccgaaaca acagttagac gtaggcgtag aatcattctg 60

tttccagtac cttttcaagg gcacatcaat ccaatactac aactagccaa cgttttgtac 120

tctaaaggtt tttctattac aatctttcac accaatttca acaaaccaaa aacatccaat 180

tacccacatt tcacattcag attcatactt gataatgatc cacaagatga acgtatttca 240

aacttaccta cccacggtcc tttagctgga atgagaattc caatcatcaa tgaacatggt 300

gccgatgagc ttagaagaga attagagtta cttatgttgg catccgaaga ggacgaggaa 360

gtctcttgtc tgattactga cgctctatgg tactttgccc aatctgtggc tgatagtttg 420

aatttgagga gattggtact aatgacatcc agtctgttta actttcacgc tcatgttagt 480

ttaccacaat ttgacgaatt gggatacttg gaccctgatg acaagactag gttagaggaa 540

caggcctctg gttttcctat gttgaaagtc aaagatatca agtctgccta ttctaattgg 600

caaatcttga aagagatctt aggaaagatg atcaaacaga caaaggcttc atctggagtg 660

atttggaaca gtttcaaaga gttagaagag tctgaattgg agactgtaat cagagaaatt 720

ccagcacctt cattcctgat accattacca aaacatttga ctgcttcctc ttcctctttg 780

ttggatcatg acagaacagt ttttcaatgg ttggaccaac aaccacctag ttctgttttg 840

tacgtgtcat ttggtagtac ttctgaagtc gatgaaaagg acttccttga aatcgcaaga 900

ggcttagtcg atagtaagca gtcattcctt tgggtcgtgc gtccaggttt cgtgaaaggc 960

tcaacatggg tcgaaccact tccagatggt tttctaggcg aaagaggtag aatagtcaaa 1020

tgggttcctc aacaggaagt ttagctcat ggcgctattg gggcattctg gactcattcc 1080

ggatggaatt caactttaga atcagtatgc gaaggggtac ctatgatctt ttcagatttt 1140

ggtcttgatc aaccactgaa cgcaagatac atgtctgatg ttttgaaagt gggtgtatat 1200

ctagaaaatg gctgggaaag gggtgaaata gctaatgcaa taagacgtgt tatggttgat 1260

gaagaggggg agtatatcag acaaaacgca agagtgctga agcaaaaggc cgacgtttct 1320

ctaatgaagg gaggctcttc atacgaatcc ttagaatctc ttgtttccta catttcatca 1380

ctgtaagtcg ac 1392

<210> 86
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 86
Thr Ser Phe Ala Glu Tyr Trp Asn Leu Leu Ser Pro
1 5 10

<210> 87
<211> 1602
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 87
atgaccagct ttgccgagta ttggaatctg ttaagtccca cttcttttgc agaatattgg 60

aaccttctat caccgacgag tttcgcggag tactggaatt tgttttctcc aacatcgttc 120

gctgaatact ggaacttact cagccctgct agtaaaatgg atgcaatggc aactactgag 180

aaaaagcctc atgtgatctt cattccattt cctgcacaat ctcacataaa ggcaatgcta 240

aagttagcac aactattaca ccataaggga ttacagataa ctttcgtgaa taccgacttc 300

atccataatc aatttctgga atctagtggc cctcattgtt tggacggagc cccagggttt 360

agattcgaaa caattcctga cggtgtttca cattccccag aggcctccat cccaataaga 420

gagagtttac tgaggtcaat agaaaccaac tttttggatc gtttcattga cttggtcaca 480

aaacttccag acccaccaac ttgcataatc tctgatggct ttctgtcagt gtttactatc 540

gacgctgcca aaaagttggg tatcccagtt atgatgtact ggactcttgc tgcatgcggt 600

ttcatgggtt tctatcacat ccattctctt atcgaaaagg gttttgctcc actgaaagat 660

gcatcatact taaccaacgg ctacctggat actgttattg actgggtacc aggtatggaa 720

ggtataagac ttaaagattt tcctttggat tggtctacag accttaatga taaagtattg 780

atgtttacta cagaagctcc acaaagatct cataaggttt cacatcatat ctttcacacc 840

tttgatgaat tggaaccatc aatcatcaaa accttgtctc taagatacaa tcatatctac 900

actattggtc cattacaatt acttctagat caaattcctg aagagaaaaa gcaaactggt 960

attacatcct tacacggcta ctctttagtg aaagaggaac cagaatgttt tcaatggcta 1020

caaagtaaag agcctaattc tgtggtctac gtcaacttcg gaagtacaac agtcatgtcc 1080

ttggaagata tgactgaatt tggttggggc cttgctaatt caaatcatta ctttctatgg 1140

attatcaggt ccaatttggt aataggggaa aacgccgtat tacctccaga attggaggaa 1200

cacatcaaaa agagaggttt cattgcttcc tggtgttctc aggaaaaggt attgaaacat 1260

ccttctgttg gtggtttcct tactcattgc ggttggggct ctacaatcga atcactaagt 1320

gcaggagttc caatgatttg ttggccatat tcatgggacc aacttacaaa ttgtaggtat 1380

atctgtaaag agtgggaagt tggattagaa atgggaacaa aggttaaacg tgatgaagtg 1440

aaaagattgg ttcaggagtt gatgggggaa ggtggccaca agatgagaaa caaggccaaa 1500

gattggaagg aaaaagccag aattgctatt gctcctaacg ggtcatcctc tctaaacatt 1560

gataagatgg tcaaagagat tacagtctta gccagaaact aa 1602

<210> 88
<211> 533
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 88
Met Thr Ser Phe Ala Glu Tyr Trp Asn Leu Leu Ser Pro Thr Ser Phe
1 5 10 15

```
Ala Glu Tyr Trp Asn Leu Leu Ser Pro Thr Ser Phe Ala Glu Tyr Trp
            20                  25                  30

Asn Leu Phe Ser Pro Thr Ser Phe Ala Glu Tyr Trp Asn Leu Leu Ser
        35                  40                  45

Pro Ala Ser Lys Met Asp Ala Met Ala Thr Thr Glu Lys Lys Pro His
    50                  55                  60

Val Ile Phe Ile Pro Phe Pro Ala Gln Ser His Ile Lys Ala Met Leu
65                  70                  75                  80

Lys Leu Ala Gln Leu Leu His His Lys Gly Leu Gln Ile Thr Phe Val
                85                  90                  95

Asn Thr Asp Phe Ile His Asn Gln Phe Leu Glu Ser Ser Gly Pro His
            100                 105                 110

Cys Leu Asp Gly Ala Pro Gly Phe Arg Phe Glu Thr Ile Pro Asp Gly
        115                 120                 125

Val Ser His Ser Pro Glu Ala Ser Ile Pro Ile Arg Glu Ser Leu Leu
    130                 135                 140

Arg Ser Ile Glu Thr Asn Phe Leu Asp Arg Phe Ile Asp Leu Val Thr
145                 150                 155                 160

Lys Leu Pro Asp Pro Pro Thr Cys Ile Ile Ser Asp Gly Phe Leu Ser
                165                 170                 175

Val Phe Thr Ile Asp Ala Ala Lys Lys Leu Gly Ile Pro Val Met Met
            180                 185                 190

Tyr Trp Thr Leu Ala Ala Cys Gly Phe Met Gly Phe Tyr His Ile His
        195                 200                 205

Ser Leu Ile Glu Lys Gly Phe Ala Pro Leu Lys Asp Ala Ser Tyr Leu
    210                 215                 220

Thr Asn Gly Tyr Leu Asp Thr Val Ile Asp Trp Val Pro Gly Met Glu
225                 230                 235                 240

Gly Ile Arg Leu Lys Asp Phe Pro Leu Asp Trp Ser Thr Asp Leu Asn
                245                 250                 255

Asp Lys Val Leu Met Phe Thr Thr Glu Ala Pro Gln Arg Ser His Lys
            260                 265                 270
```

```
Val Ser His His Ile Phe His Thr Phe Asp Glu Leu Glu Pro Ser Ile
        275                 280                 285

Ile Lys Thr Leu Ser Leu Arg Tyr Asn His Ile Tyr Thr Ile Gly Pro
    290                 295                 300

Leu Gln Leu Leu Leu Asp Gln Ile Pro Glu Glu Lys Lys Gln Thr Gly
305                 310                 315                 320

Ile Thr Ser Leu His Gly Tyr Ser Leu Val Lys Glu Glu Pro Glu Cys
                325                 330                 335

Phe Gln Trp Leu Gln Ser Lys Glu Pro Asn Ser Val Val Tyr Val Asn
            340                 345                 350

Phe Gly Ser Thr Thr Val Met Ser Leu Glu Asp Met Thr Glu Phe Gly
        355                 360                 365

Trp Gly Leu Ala Asn Ser Asn His Tyr Phe Leu Trp Ile Ile Arg Ser
    370                 375                 380

Asn Leu Val Ile Gly Glu Asn Ala Val Leu Pro Pro Glu Leu Glu Glu
385                 390                 395                 400

His Ile Lys Lys Arg Gly Phe Ile Ala Ser Trp Cys Ser Gln Glu Lys
                405                 410                 415

Val Leu Lys His Pro Ser Val Gly Gly Phe Leu Thr His Cys Gly Trp
            420                 425                 430

Gly Ser Thr Ile Glu Ser Leu Ser Ala Gly Val Pro Met Ile Cys Trp
        435                 440                 445

Pro Tyr Ser Trp Asp Gln Leu Thr Asn Cys Arg Tyr Ile Cys Lys Glu
    450                 455                 460

Trp Glu Val Gly Leu Glu Met Gly Thr Lys Val Lys Arg Asp Glu Val
465                 470                 475                 480

Lys Arg Leu Val Gln Glu Leu Met Gly Glu Gly Gly His Lys Met Arg
                485                 490                 495

Asn Lys Ala Lys Asp Trp Lys Glu Lys Ala Arg Ile Ala Ile Ala Pro
            500                 505                 510

Asn Gly Ser Ser Ser Leu Asn Ile Asp Lys Met Val Lys Glu Ile Thr
        515                 520                 525
```

Val Leu Ala Arg Asn
530

<210> 89
<211> 1893
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 89
atgtgcaata ccaacatgtc tgtacctact gatggtgctg taaccacctc acagattcca 60

gcttcggaac aagagaccct ggttagacca aagccattgc ttttgaagtt attaaagtct 120

gttggtgcac aaaaagacac ttatactatg aaagaggttc ttttttatct tggccagtat 180

attatgacta aacgattata tgatgagaag caacaacata ttgtatattg ttcaaatgat 240

cttctaggag atttgtttgg cgtgccaagc ttctctgtga aagagcacag gaaaatatat 300

accatgatct acaggaactt ggtagtagtc aatcagcagg aatcatcgga ctcaggtaca 360

tctgtgagtg agaacaggtg tcaccttgaa ggtgggagtg atcaaaagga ccttgtacaa 420

gagcttcagg aagagaaacc ttcatcttca catttggttt ctagaccatc taccggtggt 480

agcggatcct ctggaggcag tgctagtaaa atggcagagc aacaaaagat caaaaagtca 540

cctcacgtct tacttattcc atttcctctg caaggacata tcaacccatt catacaattt 600

gggaaaagat tgattagtaa gggtgtaaag acaacactgg taaccactat ccacactttg 660

aattctactc tgaaccactc aaatactact actacaagta tagaaattca agctatatca 720

gacggatgcg atgagggtgg ctttatgtct gccggtgaat cttacttgga aacattcaag 780

caagtgggat ccaagtctct ggccgatcta atcaaaaagt tacagagtga aggcaccaca 840

attgacgcca taatctacga ttctatgaca gagtgggttt tagacgttgc tatcgaattt 900

ggtattgatg gaggttcctt tttcacacaa gcatgtgttg tgaattctct atactaccat 960

gtgcataaag ggttaatctc tttaccattg ggtgaaactg tttcagttcc aggttttcca 1020

gtgttacaac gttgggaaac cccattgatc ttacaaaatc atgaacaaat acaatcacct 1080

tggtcccaga tgttgtttgg tcaattcgct aacatcgatc aagcaagatg ggtctttact 1140

aattcattct ataagttaga ggaagaggta attgaatgga ctaggaagat ctggaatttg 1200

aaagtcattg gtccaacatt gccatcaatg tatttggaca aaagacttga tgatgataaa 1260

gataatggtt tcaatttgta caaggctaat catcacgaat gtatgaattg gctggatgac 1320

aaaccaaagg aatcagttgt atatgttgct ttcggctctc ttgttaaaca tggtccagaa 1380

caagttgagg agattacaag agcacttata gactctgacg taaacttttt gtgggtcatt 1440

aagcacaaag aggaggggaa actgccagaa aacctttctg aagtgataaa gaccggaaaa 1500

ggtctaatcg ttgcttggtg taaacaattg gatgttttag ctcatgaatc tgtaggctgt 1560

tttgtaacac attgcggatt caactctaca ctagaagcca tttccttagg cgtacctgtc 1620

gttgcaatgc ctcagttctc cgatcagaca accaacgcta aacttttgga cgaaatacta 1680

ggggtgggtg tcagagttaa agcagacgag aatggtatcg tcagaagagg gaacctagct 1740

tcatgtatca aaatgatcat ggaagaggaa agaggagtta tcataaggaa aaacgcagtt 1800

aagtggaagg atcttgcaaa ggttgccgtc catgaaggcg gctcttcaga taatgatatt 1860

gttgaatttg tgtccgaact aatcaaagcc taa 1893

<210> 90
<211> 630
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 90

Met Cys Asn Thr Asn Met Ser Val Pro Thr Asp Gly Ala Val Thr Thr
1 5 10 15

Ser Gln Ile Pro Ala Ser Glu Gln Glu Thr Leu Val Arg Pro Lys Pro
20 25 30

Leu Leu Leu Lys Leu Leu Lys Ser Val Gly Ala Gln Lys Asp Thr Tyr
35 40 45

Thr Met Lys Glu Val Leu Phe Tyr Leu Gly Gln Tyr Ile Met Thr Lys
50 55 60

Arg Leu Tyr Asp Glu Lys Gln Gln His Ile Val Tyr Cys Ser Asn Asp
65 70 75 80

Leu Leu Gly Asp Leu Phe Gly Val Pro Ser Phe Ser Val Lys Glu His
85 90 95

Arg Lys Ile Tyr Thr Met Ile Tyr Arg Asn Leu Val Val Val Asn Gln
100 105 110

Gln Glu Ser Ser Asp Ser Gly Thr Ser Val Ser Glu Asn Arg Cys His
115 120 125

Leu Glu Gly Gly Ser Asp Gln Lys Asp Leu Val Gln Glu Leu Gln Glu
130 135 140

```
Glu Lys Pro Ser Ser Ser His Leu Val Ser Arg Pro Ser Thr Gly Gly
145                 150                 155                 160

Ser Gly Ser Ser Gly Gly Ser Ala Ser Lys Met Ala Glu Gln Gln Lys
                165                 170                 175

Ile Lys Lys Ser Pro His Val Leu Leu Ile Pro Phe Pro Leu Gln Gly
            180                 185                 190

His Ile Asn Pro Phe Ile Gln Phe Gly Lys Arg Leu Ile Ser Lys Gly
        195                 200                 205

Val Lys Thr Thr Leu Val Thr Thr Ile His Thr Leu Asn Ser Thr Leu
    210                 215                 220

Asn His Ser Asn Thr Thr Thr Thr Ser Ile Glu Ile Gln Ala Ile Ser
225                 230                 235                 240

Asp Gly Cys Asp Glu Gly Gly Phe Met Ser Ala Gly Glu Ser Tyr Leu
                245                 250                 255

Glu Thr Phe Lys Gln Val Gly Ser Lys Ser Leu Ala Asp Leu Ile Lys
            260                 265                 270

Lys Leu Gln Ser Glu Gly Thr Thr Ile Asp Ala Ile Ile Tyr Asp Ser
        275                 280                 285

Met Thr Glu Trp Val Leu Asp Val Ala Ile Glu Phe Gly Ile Asp Gly
    290                 295                 300

Gly Ser Phe Phe Thr Gln Ala Cys Val Val Asn Ser Leu Tyr Tyr His
305                 310                 315                 320

Val His Lys Gly Leu Ile Ser Leu Pro Leu Gly Glu Thr Val Ser Val
                325                 330                 335

Pro Gly Phe Pro Val Leu Gln Arg Trp Glu Thr Pro Leu Ile Leu Gln
            340                 345                 350

Asn His Glu Gln Ile Gln Ser Pro Trp Ser Gln Met Leu Phe Gly Gln
        355                 360                 365

Phe Ala Asn Ile Asp Gln Ala Arg Trp Val Phe Thr Asn Ser Phe Tyr
    370                 375                 380

Lys Leu Glu Glu Glu Val Ile Glu Trp Thr Arg Lys Ile Trp Asn Leu
385                 390                 395                 400
```

Lys Val Ile Gly Pro Thr Leu Pro Ser Met Tyr Leu Asp Lys Arg Leu
405 410 415

Asp Asp Asp Lys Asp Asn Gly Phe Asn Leu Tyr Lys Ala Asn His His
420 425 430

Glu Cys Met Asn Trp Leu Asp Asp Lys Pro Lys Glu Ser Val Val Tyr
435 440 445

Val Ala Phe Gly Ser Leu Val Lys His Gly Pro Glu Gln Val Glu Glu
450 455 460

Ile Thr Arg Ala Leu Ile Asp Ser Asp Val Asn Phe Leu Trp Val Ile
465 470 475 480

Lys His Lys Glu Glu Gly Lys Leu Pro Glu Asn Leu Ser Glu Val Ile
485 490 495

Lys Thr Gly Lys Gly Leu Ile Val Ala Trp Cys Lys Gln Leu Asp Val
500 505 510

Leu Ala His Glu Ser Val Gly Cys Phe Val Thr His Cys Gly Phe Asn
515 520 525

Ser Thr Leu Glu Ala Ile Ser Leu Gly Val Pro Val Val Ala Met Pro
530 535 540

Gln Phe Ser Asp Gln Thr Thr Asn Ala Lys Leu Leu Asp Glu Ile Leu
545 550 555 560

Gly Val Gly Val Arg Val Lys Ala Asp Glu Asn Gly Ile Val Arg Arg
565 570 575

Gly Asn Leu Ala Ser Cys Ile Lys Met Ile Met Glu Glu Glu Arg Gly
580 585 590

Val Ile Ile Arg Lys Asn Ala Val Lys Trp Lys Asp Leu Ala Lys Val
595 600 605

Ala Val His Glu Gly Gly Ser Ser Asp Asn Asp Ile Val Glu Phe Val
610 615 620

Ser Glu Leu Ile Lys Ala
625 630

<210> 91

<211> 1932
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 91
atgtgcaata ccaacatgtc tgtacctact gatggtgctg taaccacctc acagattcca 60

gcttcggaac aagagaccct ggttagacca aagccattgc ttttgaagtt attaaagtct 120

gttggtgcac aaaaagacac ttatactatg aaagaggttc ttttttatct tggccagtat 180

attatgacta aacgattata tgatgagaag caacaacata ttgtatattg ttcaaatgat 240

cttctaggag atttgtttgg cgtgccaagc ttctctgtga aagagcacag gaaaatatat 300

accatgatct acaggaactt ggtagtagtc aatcagcagg aatcatcgga ctcaggtaca 360

tctgtgagtg agaacaggtg tcaccttgaa ggtgggagtg atcaaaagga ccttgtacaa 420

gagcttcagg aagagaaacc ttcatcttca catttggttt ctagaccatc taccggtggt 480

agcggatcct ctggaggcag tgctagtaaa atggctacat ctgattctat tgttgatgac 540

aggaagcagt tgcatgtggc tactttccct tggcttgctt tcggtcatat actgccttac 600

ctacaactat caaaactgat agctgaaaaa ggacataaag tgtcattcct ttcaacaact 660

agaaacattc aaagattatc ttcccacata tcaccattga ttaacgtcgt tcaattgaca 720

cttccaagag tacaggaatt accagaagat gctgaagcta caacagatgt gcatcctgaa 780

gatatccctt acttgaaaaa ggcatccgat ggattacagc ctgaggtcac tagattcctt 840

gagcaacaca gtccagattg gatcatatac gactacactc actattggtt gccttcaatt 900

gcagcatcac taggcatttc tagggcacat ttcagtgtaa ccacaccttg ggccattgct 960

tacatgggtc catccgctga tgctatgatt aacggcagtg atggtagaac taccgttgaa 1020

gatttgacaa ccccaccaaa gtggtttcca tttccaacta aagtctgttg gagaaaacac 1080

gacttagcaa gactggttcc atacaaggca ccaggaatct cagacggcta tagaatgggt 1140

ttagtcctta aagggtctga ctgcctattg tctaagtgtt accatgagtt tgggacacaa 1200

tggctaccac ttttggaaac attacaccaa gttcctgtcg taccagttgg tctattacct 1260

ccagaaatcc ctggtgatga gaaggacgag acttgggttt caatcaaaaa gtggttagac 1320

gggaagcaaa aaggctcagt ggtatatgtg gcactgggtt ccgaagtttt agtatctcaa 1380

acagaagttg tggaacttgc cttaggtttg gaactatctg gattgccatt tgtctgggcc 1440

tacagaaaac caaaaggccc tgcaaagtcc gattcagttg aattgccaga cggctttgtc 1500

gagagaacta gagatagagg gttggtatgg acttcatggg ctccacaatt gagaatcctg 1560

agtcacgaat ctgtgtgcgg tttcctaaca cattgtggtt ctggttctat agttgaagga 1620

ctgatgtttg gtcatccact tatcatgttg ccaatctttg gtgaccagcc tttgaatgca 1680

cgtctgttag aagataaaca agttggaatt gaaatcccac gtaatgagga agatggatgt 1740

ttaaccaagg agtctgtggc cagatcatta cgttccgttg tcgttgaaaa ggaaggcgaa 1800

atctacaagg ccaatgcccg tgaactttca aagatctaca atgacacaaa agtagagaag 1860

gaatatgttt ctcaatttgt agattaccta gagaaaaacg ctagagccgt agctattgat 1920

catgaatcct aa 1932

<210> 92
<211> 643
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 92

Met Cys Asn Thr Asn Met Ser Val Pro Thr Asp Gly Ala Val Thr Thr
1 5 10 15

Ser Gln Ile Pro Ala Ser Glu Gln Glu Thr Leu Val Arg Pro Lys Pro
20 25 30

Leu Leu Leu Lys Leu Leu Lys Ser Val Gly Ala Gln Lys Asp Thr Tyr
35 40 45

Thr Met Lys Glu Val Leu Phe Tyr Leu Gly Gln Tyr Ile Met Thr Lys
50 55 60

Arg Leu Tyr Asp Glu Lys Gln Gln His Ile Val Tyr Cys Ser Asn Asp
65 70 75 80

Leu Leu Gly Asp Leu Phe Gly Val Pro Ser Phe Ser Val Lys Glu His
85 90 95

Arg Lys Ile Tyr Thr Met Ile Tyr Arg Asn Leu Val Val Val Asn Gln
100 105 110

Gln Glu Ser Ser Asp Ser Gly Thr Ser Val Ser Glu Asn Arg Cys His
115 120 125

Leu Glu Gly Gly Ser Asp Gln Lys Asp Leu Val Gln Glu Leu Gln Glu
130 135 140

Glu Lys Pro Ser Ser Ser His Leu Val Ser Arg Pro Ser Thr Gly Gly
145 150 155 160

```
Ser Gly Ser Ser Gly Gly Ser Ala Ser Lys Met Ala Thr Ser Asp Ser
                165                 170                 175

Ile Val Asp Asp Arg Lys Gln Leu His Val Ala Thr Phe Pro Trp Leu
            180                 185                 190

Ala Phe Gly His Ile Leu Pro Tyr Leu Gln Leu Ser Lys Leu Ile Ala
        195                 200                 205

Glu Lys Gly His Lys Val Ser Phe Leu Ser Thr Thr Arg Asn Ile Gln
    210                 215                 220

Arg Leu Ser Ser His Ile Ser Pro Leu Ile Asn Val Val Gln Leu Thr
225                 230                 235                 240

Leu Pro Arg Val Gln Glu Leu Pro Glu Asp Ala Glu Ala Thr Thr Asp
                245                 250                 255

Val His Pro Glu Asp Ile Pro Tyr Leu Lys Lys Ala Ser Asp Gly Leu
            260                 265                 270

Gln Pro Glu Val Thr Arg Phe Leu Glu Gln His Ser Pro Asp Trp Ile
        275                 280                 285

Ile Tyr Asp Tyr Thr His Tyr Trp Leu Pro Ser Ile Ala Ala Ser Leu
    290                 295                 300

Gly Ile Ser Arg Ala His Phe Ser Val Thr Thr Pro Trp Ala Ile Ala
305                 310                 315                 320

Tyr Met Gly Pro Ser Ala Asp Ala Met Ile Asn Gly Ser Asp Gly Arg
                325                 330                 335

Thr Thr Val Glu Asp Leu Thr Thr Pro Pro Lys Trp Phe Pro Phe Pro
            340                 345                 350

Thr Lys Val Cys Trp Arg Lys His Asp Leu Ala Arg Leu Val Pro Tyr
        355                 360                 365

Lys Ala Pro Gly Ile Ser Asp Gly Tyr Arg Met Gly Leu Val Leu Lys
    370                 375                 380

Gly Ser Asp Cys Leu Leu Ser Lys Cys Tyr His Glu Phe Gly Thr Gln
385                 390                 395                 400

Trp Leu Pro Leu Leu Glu Thr Leu His Gln Val Pro Val Val Pro Val
                405                 410                 415
```

```
Gly Leu Leu Pro Pro Glu Ile Pro Gly Asp Glu Lys Asp Glu Thr Trp
            420                 425                 430

Val Ser Ile Lys Lys Trp Leu Asp Gly Lys Gln Lys Gly Ser Val Val
        435                 440                 445

Tyr Val Ala Leu Gly Ser Glu Val Leu Val Ser Gln Thr Glu Val Val
    450                 455                 460

Glu Leu Ala Leu Gly Leu Glu Leu Ser Gly Leu Pro Phe Val Trp Ala
465                 470                 475                 480

Tyr Arg Lys Pro Lys Gly Pro Ala Lys Ser Asp Ser Val Glu Leu Pro
                485                 490                 495

Asp Gly Phe Val Glu Arg Thr Arg Asp Arg Gly Leu Val Trp Thr Ser
            500                 505                 510

Trp Ala Pro Gln Leu Arg Ile Leu Ser His Glu Ser Val Cys Gly Phe
        515                 520                 525

Leu Thr His Cys Gly Ser Gly Ser Ile Val Glu Gly Leu Met Phe Gly
    530                 535                 540

His Pro Leu Ile Met Leu Pro Ile Phe Gly Asp Gln Pro Leu Asn Ala
545                 550                 555                 560

Arg Leu Leu Glu Asp Lys Gln Val Gly Ile Glu Ile Pro Arg Asn Glu
                565                 570                 575

Glu Asp Gly Cys Leu Thr Lys Glu Ser Val Ala Arg Ser Leu Arg Ser
            580                 585                 590

Val Val Val Glu Lys Glu Gly Glu Ile Tyr Lys Ala Asn Ala Arg Glu
        595                 600                 605

Leu Ser Lys Ile Tyr Asn Asp Thr Lys Val Glu Lys Glu Tyr Val Ser
    610                 615                 620

Gln Phe Val Asp Tyr Leu Glu Lys Asn Ala Arg Ala Val Ala Ile Asp
625                 630                 635                 640

His Glu Ser
```

<210> 93
<211> 1887

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 93
atgtgcaata ccaacatgtc tgtacctact gatggtgctg taaccacctc acagattcca 60

gcttcggaac aagagaccct ggttagacca aagccattgc ttttgaagtt attaaagtct 120

gttggtgcac aaaaagacac ttatactatg aaagaggttc ttttttatct tggccagtat 180

attatgacta aacgattata tgatgagaag caacaacata ttgtatattg ttcaaatgat 240

cttctaggag atttgtttgg cgtgccaagc ttctctgtga aagagcacag gaaaatatat 300

accatgatct acaggaactt ggtagtagtc aatcagcagg aatcatcgga ctcaggtaca 360

tctgtgagtg agaacaggtg tcaccttgaa ggtgggagtg atcaaaagga ccttgtacaa 420

gagcttcagg aagagaaacc ttcatcttca catttggttt ctagaccatc taccggtggt 480

agcggatcct ctggaggcag tgctagtaaa atggaaaaca agaccgaaac aacagttaga 540

cgtaggcgta gaatcattct gtttccagta ccttttcaag ggcacatcaa tccaatacta 600

caactagcca acgttttgta ctctaaaggt ttttctatta caatctttca caccaatttc 660

aacaaaccaa aaacatccaa ttacccacat ttcacattca gattcatact tgataatgat 720

ccacaagatg aacgtatttc aaacttacct acccacggtc ctttagctgg aatgagaatt 780

ccaatcatca atgaacatgg tgccgatgag cttagaagag aattagagtt acttatgttg 840

gcatccgaag aggacgagga agtctcttgt ctgattactg acgctctatg gtactttgcc 900

caatctgtgg ctgatagttt gaatttgagg agattggtac taatgacatc cagtctgttt 960

aactttcacg ctcatgttag tttaccacaa tttgacgaat tgggatactt ggaccctgat 1020

gacaagacta ggttagagga acaggcctct ggttttccta tgttgaaagt caaagatatc 1080

aagtctgcct attctaattg gcaaatcttg aaagagatct taggaaagat gatcaaacag 1140

acaaaggctt catctggagt gatttggaac agtttcaaag agttagaaga gtctgaattg 1200

gagactgtaa tcagagaaat tccagcacct tcattcctga taccattacc aaaacatttg 1260

actgcttcct cttcctcttt gttggatcat gacagaacag tttttcaatg gttggaccaa 1320

caaccaccta gttctgtttt gtacgtgtca tttggtagta cttctgaagt cgatgaaaag 1380

gacttccttg aaatcgcaag aggcttagtc gatagtaagc agtcattcct ttgggtcgtg 1440

cgtccaggtt tcgtgaaagg ctcaacatgg gtcgaaccac ttccagatgg ttttctaggc 1500

gaaagaggta gaatagtcaa atgggttcct caacaggaag ttttagctca tggcgctatt 1560

ggggcattct ggactcattc cggatggaat tcaactttag aatcagtatg cgaaggggta 1620

cctatgatct tttcagattt tggtcttgat caaccactga acgcaagata catgtctgat 1680

gttttgaaag tgggtgtata tctagaaaat ggctgggaaa ggggtgaaat agctaatgca 1740

ataagacgtg ttatggttga tgaagagggg gagtatatca gacaaaacgc aagagtgctg 1800

aagcaaaagg ccgacgtttc tctaatgaag ggaggctctt catacgaatc cttagaatct 1860

cttgtttcct acatttcatc actgtaa 1887

<210> 94
<211> 628
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 94

```
Met Cys Asn Thr Asn Met Ser Val Pro Thr Asp Gly Ala Val Thr Thr
1               5                  10                  15

Ser Gln Ile Pro Ala Ser Glu Gln Glu Thr Leu Val Arg Pro Lys Pro
            20                  25                  30

Leu Leu Leu Lys Leu Leu Lys Ser Val Gly Ala Gln Lys Asp Thr Tyr
        35                  40                  45

Thr Met Lys Glu Val Leu Phe Tyr Leu Gly Gln Tyr Ile Met Thr Lys
    50                  55                  60

Arg Leu Tyr Asp Glu Lys Gln Gln His Ile Val Tyr Cys Ser Asn Asp
65                  70                  75                  80

Leu Leu Gly Asp Leu Phe Gly Val Pro Ser Phe Ser Val Lys Glu His
                85                  90                  95

Arg Lys Ile Tyr Thr Met Ile Tyr Arg Asn Leu Val Val Val Asn Gln
            100                 105                 110

Gln Glu Ser Ser Asp Ser Gly Thr Ser Val Ser Glu Asn Arg Cys His
        115                 120                 125

Leu Glu Gly Gly Ser Asp Gln Lys Asp Leu Val Gln Glu Leu Gln Glu
    130                 135                 140

Glu Lys Pro Ser Ser Ser His Leu Val Ser Arg Pro Ser Thr Gly Gly
145                 150                 155                 160

Ser Gly Ser Ser Gly Gly Ser Ala Ser Lys Met Glu Asn Lys Thr Glu
                165                 170                 175
```

```
Thr Thr Val Arg Arg Arg Arg Arg Ile Ile Leu Phe Pro Val Pro Phe
            180                 185                 190

Gln Gly His Ile Asn Pro Ile Leu Gln Leu Ala Asn Val Leu Tyr Ser
        195                 200                 205

Lys Gly Phe Ser Ile Thr Ile Phe His Thr Asn Phe Asn Lys Pro Lys
    210                 215                 220

Thr Ser Asn Tyr Pro His Phe Thr Phe Arg Phe Ile Leu Asp Asn Asp
225                 230                 235                 240

Pro Gln Asp Glu Arg Ile Ser Asn Leu Pro Thr His Gly Pro Leu Ala
                245                 250                 255

Gly Met Arg Ile Pro Ile Ile Asn Glu His Gly Ala Asp Glu Leu Arg
            260                 265                 270

Arg Glu Leu Glu Leu Leu Met Leu Ala Ser Glu Glu Asp Glu Glu Val
        275                 280                 285

Ser Cys Leu Ile Thr Asp Ala Leu Trp Tyr Phe Ala Gln Ser Val Ala
    290                 295                 300

Asp Ser Leu Asn Leu Arg Arg Leu Val Leu Met Thr Ser Ser Leu Phe
305                 310                 315                 320

Asn Phe His Ala His Val Ser Leu Pro Gln Phe Asp Glu Leu Gly Tyr
                325                 330                 335

Leu Asp Pro Asp Asp Lys Thr Arg Leu Glu Glu Gln Ala Ser Gly Phe
            340                 345                 350

Pro Met Leu Lys Val Lys Asp Ile Lys Ser Ala Tyr Ser Asn Trp Gln
        355                 360                 365

Ile Leu Lys Glu Ile Leu Gly Lys Met Ile Lys Gln Thr Lys Ala Ser
    370                 375                 380

Ser Gly Val Ile Trp Asn Ser Phe Lys Glu Leu Glu Glu Ser Glu Leu
385                 390                 395                 400

Glu Thr Val Ile Arg Glu Ile Pro Ala Pro Ser Phe Leu Ile Pro Leu
                405                 410                 415

Pro Lys His Leu Thr Ala Ser Ser Ser Ser Leu Leu Asp His Asp Arg
            420                 425                 430
```

Thr Val Phe Gln Trp Leu Asp Gln Gln Pro Pro Ser Ser Val Leu Tyr
435 440 445

Val Ser Phe Gly Ser Thr Ser Glu Val Asp Glu Lys Asp Phe Leu Glu
450 455 460

Ile Ala Arg Gly Leu Val Asp Ser Lys Gln Ser Phe Leu Trp Val Val
465 470 475 480

Arg Pro Gly Phe Val Lys Gly Ser Thr Trp Val Glu Pro Leu Pro Asp
485 490 495

Gly Phe Leu Gly Glu Arg Gly Arg Ile Val Lys Trp Val Pro Gln Gln
500 505 510

Glu Val Leu Ala His Gly Ala Ile Gly Ala Phe Trp Thr His Ser Gly
515 520 525

Trp Asn Ser Thr Leu Glu Ser Val Cys Glu Gly Val Pro Met Ile Phe
530 535 540

Ser Asp Phe Gly Leu Asp Gln Pro Leu Asn Ala Arg Tyr Met Ser Asp
545 550 555 560

Val Leu Lys Val Gly Val Tyr Leu Glu Asn Gly Trp Glu Arg Gly Glu
565 570 575

Ile Ala Asn Ala Ile Arg Arg Val Met Val Asp Glu Glu Gly Glu Tyr
580 585 590

Ile Arg Gln Asn Ala Arg Val Leu Lys Gln Lys Ala Asp Val Ser Leu
595 600 605

Met Lys Gly Gly Ser Ser Tyr Glu Ser Leu Glu Ser Leu Val Ser Tyr
610 615 620

Ile Ser Ser Leu
625

<210> 95
<211> 473
<212> PRT
<213> Stevia rebaudiana

<400> 95
Met Ala Thr Ser Asp Ser Ile Val Asp Asp Arg Lys Gln Leu His Val
1 5 10 15

```
Ala Thr Phe Pro Trp Leu Ala Phe Gly His Ile Leu Pro Tyr Leu Gln
            20                  25                  30

Leu Ser Lys Leu Ile Ala Glu Lys Gly His Lys Val Ser Phe Leu Ser
        35                  40                  45

Thr Thr Arg Asn Ile Gln Arg Leu Ser Ser His Ile Ser Pro Leu Ile
    50                  55                  60

Asn Val Val Gln Leu Thr Leu Pro Arg Val Gln Glu Leu Pro Glu Asp
65                  70                  75                  80

Ala Glu Ala Thr Thr Asp Val His Pro Glu Asp Ile Pro Tyr Leu Lys
                85                  90                  95

Lys Ala Ser Asp Gly Leu Gln Pro Glu Val Thr Arg Phe Leu Glu Gln
            100                 105                 110

His Ser Pro Asp Trp Ile Ile Tyr Asp Tyr Thr His Tyr Trp Leu Pro
        115                 120                 125

Ser Ile Ala Ala Ser Leu Gly Ile Ser Arg Ala His Phe Ser Val Thr
    130                 135                 140

Thr Pro Trp Ala Ile Ala Tyr Met Gly Pro Ser Ala Asp Ala Met Ile
145                 150                 155                 160

Asn Gly Ser Asp Gly Arg Thr Thr Val Glu Asp Leu Thr Thr Pro Pro
                165                 170                 175

Lys Trp Phe Pro Phe Pro Thr Lys Val Cys Trp Arg Lys His Asp Leu
            180                 185                 190

Ala Arg Leu Val Pro Tyr Lys Ala Pro Gly Ile Ser Asp Arg Cys Arg
        195                 200                 205

Met Gly Leu Val Leu Lys Gly Ser Asp Cys Leu Leu Ser Lys Cys Tyr
    210                 215                 220

His Glu Phe Gly Thr Gln Trp Leu Pro Leu Leu Glu Thr Leu His Gln
225                 230                 235                 240

Val Pro Val Val Pro Val Gly Leu Leu Pro Pro Glu Ile Pro Gly Asp
                245                 250                 255

Glu Lys Asp Glu Thr Trp Val Ser Ile Lys Lys Trp Leu Asp Gly Lys
            260                 265                 270
```

Gln Lys Gly Ser Val Val Tyr Val Ala Leu Gly Ser Glu Val Leu Val
275 280 285

Ser Gln Thr Glu Val Val Glu Leu Ala Leu Gly Leu Glu Leu Ser Gly
290 295 300

Leu Pro Phe Val Trp Ala Tyr Arg Lys Pro Lys Gly Pro Ala Lys Ser
305 310 315 320

Asp Ser Val Glu Leu Pro Asp Gly Phe Val Glu Arg Thr Arg Asp Arg
325 330 335

Gly Leu Val Trp Thr Ser Arg Ala Pro Gln Leu Arg Ile Leu Ser His
340 345 350

Glu Ser Val Cys Gly Phe Leu Thr His Cys Gly Ser Gly Ser Ile Val
355 360 365

Glu Gly Leu Met Phe Gly His Pro Leu Ile Met Leu Pro Ile Phe Gly
370 375 380

Asp Gln Pro Leu Asn Ala Arg Leu Leu Glu Asp Lys Gln Val Gly Ile
385 390 395 400

Glu Ile Pro Arg Asn Glu Glu Asp Gly Cys Leu Thr Lys Glu Ser Val
405 410 415

Ala Arg Ser Leu Arg Ser Val Val Val Glu Lys Glu Gly Glu Ile Tyr
420 425 430

Lys Ala Asn Ala Arg Glu Leu Ser Lys Ile Tyr Asn Asp Thr Lys Val
435 440 445

Glu Lys Glu Tyr Val Ser Gln Phe Val Asp Tyr Leu Glu Lys Asn Ala
450 455 460

Arg Ala Val Ala Ile Asp His Glu Ser
465 470

<210> 96
<211> 1422
<212> DNA
<213> Stevia rebaudiana

<400> 96
atggctacca gtgactccat agttgacgac cgtaagcagc ttcatgttgc gacgttccca 60

tggcttgctt tcggtcacat cctcccttac cttcagcttt cgaaattgat agctgaaaag 120

ggtcacaaag tctcgtttct ttctaccacc agaaacattc aacgtctctc ttctcatatc 180

tcgccactca taaatgttgt tcaactcaca cttccacgtg tccaagagct gccggaggat 240

gcagaggcga ccactgacgt ccaccctgaa gatattccat atctcaagaa ggcttctgat 300

ggtcttcaac cggaggtcac ccggtttcta gaacaacact ctccggactg gattatttat 360

gattatactc actactggtt gccatccatc gcggctagcc tcggtatctc acgagcccac 420

ttctccgtca ccactccatg ggccattgct tatatgggac cctcagctga cgccatgata 480

aatggttcag atggtcgaac cacggttgag gatctcacga caccgcccaa gtggtttccc 540

tttccgacca aagtatgctg gcggaagcat gatcttgccc gactggtgcc ttacaaagct 600

ccggggatat ctgatcgatg ccgtatgggg ctggttctta agggatctga ttgtttgctc 660

tccaaatgtt accatgagtt tggaactcaa tggctacctc ttttggagac actacaccaa 720

gtaccggtgg ttccggtggg attactgcca ccggaaatac ccggagacga gaaagatgaa 780

acatgggtgt caatcaagaa atggctcgat ggtaaacaaa aaggcagtgt ggtgtacgtt 840

gcattaggaa gcgaggtttt ggtgagccaa accgaggttg ttgagttagc attgggtctc 900

gagctttctg ggttgccatt tgtttgggct tatagaaaac caaaaggtcc cgcgaagtca 960

gactcggtgg agttgccaga cgggttcgtg gaacgaactc gtgaccgtgg gttggtctgg 1020

acgagtcggg cacctcagtt acgaatactg agccatgagt cggtttgtgg gttcttgacg 1080

cattgtggtt ctggatcaat tgtggaaggg ctaatgtttg gtcaccctct aatcatgcta 1140

ccgatttttg gggaccaacc tctgaatgct cgattactgg aggacaaaca ggtgggaatc 1200

gagataccaa gaaatgagga agatggttgc ttgaccaagg agtcggttgc tagatcactg 1260

aggtccgttg ttgtggaaaa agaaggggag atctacaagg cgaacgcgag ggagctgagt 1320

aaaatctata acgacactaa ggttgaaaaa gaatatgtaa gccaattcgt agactatttg 1380

gaaaagaatg cgcgtgcggt tgccatcgat catgagagtt aa 1422

<210> 97
<211> 1380
<212> DNA
<213> Stevia rebaudiana

<400> 97
atggaaaata aaacggagac caccgttcgc cggcgccgga gaataatatt attcccggta 60

ccatttcaag gccacattaa cccaattctt cagctagcca atgtgttgta ctctaaagga 120

ttcagtatca ccatctttca caccaacttc aacaaaccca aaacatctaa ttaccctcac 180

ttcactttca gattcatcct cgacaacgac ccacaagacg aacgcatttc caatctaccg 240

actcatggtc cgctcgctgg tatgcggatt ccgattatca acgaacacgg agctgacgaa 300

ttacgacgcg aactggaact gttgatgtta gcttctgaag aagatgaaga ggtatcgtgt 360

ttaatcacgg atgctctttg gtacttcgcg caatctgttg ctgacagtct taacctccga 420

cggcttgttt tgatgacaag cagcttgttt aattttcatg cacatgtttc acttcctcag 480

tttgatgagc ttggttacct cgatcctgat gacaaaaccc gtttggaaga acaagcgagt 540

gggtttccta tgctaaaagt gaaagacatc aagtctgcgt attcgaactg gcaaatactc 600

aaagagatat tagggaagat gataaaacaa acaaaagcat cttcaggagt catctggaac 660

tcatttaagg aactcgaaga gtctgagctc gaaactgtta tccgtgagat cccggctcca 720

agtttcttga taccactccc caagcatttg acagcctctt ccagcagctt actagaccac 780

gatcgaaccg tttttccatg gttagaccaa caaccgtcac gttcggtact gtatgttagt 840

tttggtagtg gtactgaagt actggatgag aaagatttct tggaaatagc tcgtgggttg 900

gttgatagca agcagtcgtt tttatgggtg gttcgacctg ggtttgtcaa gggttcgacg 960

tgggtcgaac cgttgccaga tgggttcttg ggtgaaagag gacgtattgt gaaatgggtt 1020

ccacagcaag aagtgctagc tcatggagca ataggcgcat tctggactca tagcggatgg 1080

aactctacgt tggaaagcgt ttgtgaaggt gttcctatga ttttctcgga ttttgggctc 1140

gatcaaccgt tgaatgctag atacatgagt gatgttttga aggtaggggt gtatttggaa 1200

aatgggtggg aaagaggaga gatagcaaat gcaataagaa gagttatggt ggatgaagaa 1260

ggagaataca ttagacagaa tgcaagagtt ttgaaacaaa aggcagatgt ttctttgatg 1320

aagggtggtt cgtcttacga atcattagag tctctagttt cttacatttc atcgttgtaa 1380

<210> 98
<211> 459
<212> PRT
<213> Stevia rebaudiana

<400> 98
Met Glu Asn Lys Thr Glu Thr Thr Val Arg Arg Arg Arg Arg Ile Ile
1 5 10 15

Leu Phe Pro Val Pro Phe Gln Gly His Ile Asn Pro Ile Leu Gln Leu
20 25 30

Ala Asn Val Leu Tyr Ser Lys Gly Phe Ser Ile Thr Ile Phe His Thr
35 40 45

Asn Phe Asn Lys Pro Lys Thr Ser Asn Tyr Pro His Phe Thr Phe Arg
50 55 60

Phe Ile Leu Asp Asn Asp Pro Gln Asp Glu Arg Ile Ser Asn Leu Pro
65 70 75 80

Thr His Gly Pro Leu Ala Gly Met Arg Ile Pro Ile Ile Asn Glu His

85 90 95

Gly Ala Asp Glu Leu Arg Arg Glu Leu Glu Leu Leu Met Leu Ala Ser
100 105 110

Glu Glu Asp Glu Glu Val Ser Cys Leu Ile Thr Asp Ala Leu Trp Tyr
115 120 125

Phe Ala Gln Ser Val Ala Asp Ser Leu Asn Leu Arg Arg Leu Val Leu
130 135 140

Met Thr Ser Ser Leu Phe Asn Phe His Ala His Val Ser Leu Pro Gln
145 150 155 160

Phe Asp Glu Leu Gly Tyr Leu Asp Pro Asp Asp Lys Thr Arg Leu Glu
165 170 175

Glu Gln Ala Ser Gly Phe Pro Met Leu Lys Val Lys Asp Ile Lys Ser
180 185 190

Ala Tyr Ser Asn Trp Gln Ile Leu Lys Glu Ile Leu Gly Lys Met Ile
195 200 205

Lys Gln Thr Lys Ala Ser Ser Gly Val Ile Trp Asn Ser Phe Lys Glu
210 215 220

Leu Glu Glu Ser Glu Leu Glu Thr Val Ile Arg Glu Ile Pro Ala Pro
225 230 235 240

Ser Phe Leu Ile Pro Leu Pro Lys His Leu Thr Ala Ser Ser Ser Ser
245 250 255

Leu Leu Asp His Asp Arg Thr Val Phe Pro Trp Leu Asp Gln Gln Pro
260 265 270

Ser Arg Ser Val Leu Tyr Val Ser Phe Gly Ser Gly Thr Glu Val Leu
275 280 285

Asp Glu Lys Asp Phe Leu Glu Ile Ala Arg Gly Leu Val Asp Ser Lys
290 295 300

Gln Ser Phe Leu Trp Val Val Arg Pro Gly Phe Val Lys Gly Ser Thr
305 310 315 320

Trp Val Glu Pro Leu Pro Asp Gly Phe Leu Gly Glu Arg Gly Arg Ile
325 330 335

Val Lys Trp Val Pro Gln Gln Glu Val Leu Ala His Gly Ala Ile Gly
340 345 350

Ala Phe Trp Thr His Ser Gly Trp Asn Ser Thr Leu Glu Ser Val Cys
355 360 365

Glu Gly Val Pro Met Ile Phe Ser Asp Phe Gly Leu Asp Gln Pro Leu
370 375 380

Asn Ala Arg Tyr Met Ser Asp Val Leu Lys Val Gly Val Tyr Leu Glu
385 390 395 400

Asn Gly Trp Glu Arg Gly Glu Ile Ala Asn Ala Ile Arg Arg Val Met
405 410 415

Val Asp Glu Glu Gly Glu Tyr Ile Arg Gln Asn Ala Arg Val Leu Lys
420 425 430

Gln Lys Ala Asp Val Ser Leu Met Lys Gly Gly Ser Ser Tyr Glu Ser
435 440 445

Leu Glu Ser Leu Val Ser Tyr Ile Ser Ser Leu
450 455

<210> 99
<211> 1380
<212> DNA
<213> Stevia rebaudiana

<220>
<221> modified_base
<222> (861)..(863)
<223> a, c, t, g, unknown or other

<400> 99
atggaaaata aaacggagac caccgttcgc cggcgccgga gaataatatt attcccggta 60

ccatttcaag gccacattaa cccaattctt cagctagcca atgtgttgta ctctaaagga 120

ttcagtatca ccatctttca caccaacttc aacaaaccca aaacatctaa ttaccctcac 180

ttcactttca gattcatcct cgacaacgac ccacaagacg aacgcatttc caatctaccg 240

actcatggtc cgctcgctgg tatgcggatt ccgattatca acgaacacgg agctgacgaa 300

ttacgacgcg aactggaact gttgatgtta gcttctgaag aagatgaaga ggtatcgtgt 360

ttaatcacgg atgctctttg gtacttcgcg caatctgttg ctgacagtct taacctccga 420

cggcttgttt tgatgacaag cagcttgttt aattttcatg cacatgtttc acttcctcag 480

tttgatgagc ttggttacct cgatcctgat gacaaaaccc gtttggaaga acaagcgagt 540

gggtttccta tgctaaaagt gaaagacatc aagtctgcgt attcgaactg gcaaatactc 600

aaagagatat tagggaagat gataaaacaa acaaaagcat cttcaggagt catctggaac 660

tcatttaagg aactcgaaga gtctgagctc gaaactgtta tccgtgagat cccggctcca 720

agtttcttga taccactccc caagcatttg acagcctctt ccagcagctt actagaccac 780

gatcgaaccg tttttcaatg gttagaccaa caaccgccaa gttcggtact gtatgttagt 840

tttggtagta ctagtgaagt nnnggatgag aaagatttct tggaaatagc tcgtgggttg 900

gttgatagca agcagtcgtt tttatgggtg gttcgacctg ggtttgtcaa gggttcgacg 960

tgggtcgaac cgttgccaga tgggttcgtg gccgaaagag ggcgtattgt gaaatgggtt 1020

ccgcaacagg aagtgatagc tcatggagca atcggtgcat tctggactca tagcggatgg 1080

aactctacat tggaaagcgt ttgtgaaggt gttcctatga ttttctcgga ttttgggctc 1140

gatcaaccgt tgaatgctag atacatgagt gatgttttga aggtaggggt gtatttggaa 1200

aatgggtggg aaagaggaga gatagcaaat gcaatacgaa gagttatggt ggatgaagaa 1260

ggagaataca ttagacagaa tgcaagagtt ttgaaacaaa aggcagatgt ttctttgatg 1320

aagggtggtt catcttacga atcattagag tctctagttt cttacatttc atcgttgtaa 1380

<210> 100
<211> 459
<212> PRT
<213> Stevia rebaudiana

<220>
<221> MOD_RES
<222> (288)..(288)
<223> Any amino acid

<400> 100
Met Glu Asn Lys Thr Glu Thr Thr Val Arg Arg Arg Arg Arg Ile Ile
1 5 10 15

Leu Phe Pro Val Pro Phe Gln Gly His Ile Asn Pro Ile Leu Gln Leu
20 25 30

Ala Asn Val Leu Tyr Ser Lys Gly Phe Ser Ile Thr Ile Phe His Thr
35 40 45

Asn Phe Asn Lys Pro Lys Thr Ser Asn Tyr Pro His Phe Thr Phe Arg
50 55 60

Phe Ile Leu Asp Asn Asp Pro Gln Asp Glu Arg Ile Ser Asn Leu Pro
65 70 75 80

Thr His Gly Pro Leu Ala Gly Met Arg Ile Pro Ile Ile Asn Glu His
85 90 95

```
Gly Ala Asp Glu Leu Arg Arg Glu Leu Glu Leu Leu Met Leu Ala Ser
            100                 105                 110

Glu Glu Asp Glu Glu Val Ser Cys Leu Ile Thr Asp Ala Leu Trp Tyr
        115                 120                 125

Phe Ala Gln Ser Val Ala Asp Ser Leu Asn Leu Arg Arg Leu Val Leu
    130                 135                 140

Met Thr Ser Ser Leu Phe Asn Phe His Ala His Val Ser Leu Pro Gln
145                 150                 155                 160

Phe Asp Glu Leu Gly Tyr Leu Asp Pro Asp Asp Lys Thr Arg Leu Glu
                165                 170                 175

Glu Gln Ala Ser Gly Phe Pro Met Leu Lys Val Lys Asp Ile Lys Ser
            180                 185                 190

Ala Tyr Ser Asn Trp Gln Ile Leu Lys Glu Ile Leu Gly Lys Met Ile
        195                 200                 205

Lys Gln Thr Lys Ala Ser Ser Gly Val Ile Trp Asn Ser Phe Lys Glu
    210                 215                 220

Leu Glu Glu Ser Glu Leu Glu Thr Val Ile Arg Glu Ile Pro Ala Pro
225                 230                 235                 240

Ser Phe Leu Ile Pro Leu Pro Lys His Leu Thr Ala Ser Ser Ser Ser
                245                 250                 255

Leu Leu Asp His Asp Arg Thr Val Phe Gln Trp Leu Asp Gln Gln Pro
            260                 265                 270

Pro Ser Ser Val Leu Tyr Val Ser Phe Gly Ser Thr Ser Glu Val Xaa
        275                 280                 285

Asp Glu Lys Asp Phe Leu Glu Ile Ala Arg Gly Leu Val Asp Ser Lys
    290                 295                 300

Gln Ser Phe Leu Trp Val Val Arg Pro Gly Phe Val Lys Gly Ser Thr
305                 310                 315                 320

Trp Val Glu Pro Leu Pro Asp Gly Phe Val Ala Glu Arg Gly Arg Ile
                325                 330                 335

Val Lys Trp Val Pro Gln Gln Glu Val Ile Ala His Gly Ala Ile Gly
            340                 345                 350
```

Ala Phe Trp Thr His Ser Gly Trp Asn Ser Thr Leu Glu Ser Val Cys
355 360 365

Glu Gly Val Pro Met Ile Phe Ser Asp Phe Gly Leu Asp Gln Pro Leu
370 375 380

Asn Ala Arg Tyr Met Ser Asp Val Leu Lys Val Gly Val Tyr Leu Glu
385 390 395 400

Asn Gly Trp Glu Arg Gly Glu Ile Ala Asn Ala Ile Arg Arg Val Met
405 410 415

Val Asp Glu Glu Gly Glu Tyr Ile Arg Gln Asn Ala Arg Val Leu Lys
420 425 430

Gln Lys Ala Asp Val Ser Leu Met Lys Gly Gly Ser Ser Tyr Glu Ser
435 440 445

Leu Glu Ser Leu Val Ser Tyr Ile Ser Ser Leu
450 455

<210> 101
<211> 1380
<212> DNA
<213> Stevia rebaudiana

<220>
<221> modified_base
<222> (861)..(863)
<223> a, c, t, g, unknown or other

<400> 101
atggaaaata aaacggagac caccgttcgc cggcgccgga gaataatatt attcccggta 60

ccatttcaag gccacattaa cccaattctt cagctagcca atgtgttgta ctctaaagga 120

ttcagtatca ccatctttca caccaacttc aacaaaccca aaacatctaa ttaccctcac 180

ttcactttca gattcatcct cgacaacgac ccacaagacg aacgcatttc caatctaccg 240

actcatggtc cgctcgctgg tatgcggatt ccgattatca acgaacacgg agctgacgaa 300

ttacgacgcg aactggaact gttgatgtta gcttctgaag aagatgaaga ggtatcgtgt 360

ttaatcacgg atgctctttg gtacttcgcg caatctgttg ctgacagtct taacctccga 420

cggcttgttt tgatgacaag cagcttgttt aattttcatg cacatgtttc acttcctcag 480

tttgatgagc ttggttacct cgatcctgat gacaaaaccc gtttggaaga acaagcgagt 540

gggtttccta tgctaaaagt gaaagacatc aagtctgcgt attcgaactg gcaaatactc 600

aaagagatat tagggaagat gataaaacaa acaaaagcat cttcaggagt catctggaac 660

tcatttaagg aactcgaaga gtctgagctc gaaactgtta tccgtgagat cccggctcca 720

agtttcttga taccactccc caagcatttg acagcctctt ccagcagctt actagaccac 780

gatcgaaccg tttttcaatg gttagaccaa caaccgccaa gttcggtact gtatgttagt 840

tttggtagta ctagtgaagt nnnggatgag aaagatttct tggaaatagc tcgtgggttg 900

gttgatagca agcagtcgtt tttatgggtg gttcgacctg ggtttgtcaa gggttcgacg 960

tgggtcgaac cgttgccaga tgggttcttg ggtgaaagag gacgtattgt gaaatgggtt 1020

ccacagcaag aagtgctagc tcatggagca ataggcgcat tctggactca tagcggatgg 1080

aactctacgt tggaaagcgt ttgtgaaggt gttcctatga ttttctcgga ttttgggctc 1140

gatcaaccgt tgaatgctag atacatgagt gatgttttga aggtaggggt gtatttggaa 1200

aatgggtggg aaagaggaga gatagcaaat gcaataagaa gagttatggt ggatgaagaa 1260

ggagaataca ttagacagaa tgcaagagtt ttgaaacaaa aggcagatgt ttctttgatg 1320

aagggtggtt cgtcttacga atcattagag tctctagttt cttacatttc atcgttgtaa 1380

<210> 102
<211> 459
<212> PRT
<213> Stevia rebaudiana

<220>
<221> MOD_RES
<222> (288)..(288)
<223> Any amino acid

<400> 102
Met Glu Asn Lys Thr Glu Thr Thr Val Arg Arg Arg Arg Arg Ile Ile
1 5 10 15

Leu Phe Pro Val Pro Phe Gln Gly His Ile Asn Pro Ile Leu Gln Leu
20 25 30

Ala Asn Val Leu Tyr Ser Lys Gly Phe Ser Ile Thr Ile Phe His Thr
35 40 45

Asn Phe Asn Lys Pro Lys Thr Ser Asn Tyr Pro His Phe Thr Phe Arg
50 55 60

Phe Ile Leu Asp Asn Asp Pro Gln Asp Glu Arg Ile Ser Asn Leu Pro
65 70 75 80

Thr His Gly Pro Leu Ala Gly Met Arg Ile Pro Ile Ile Asn Glu His
85 90 95

Gly Ala Asp Glu Leu Arg Arg Glu Leu Glu Leu Leu Met Leu Ala Ser

```
            100                 105                 110

Glu Glu Asp Glu Glu Val Ser Cys Leu Ile Thr Asp Ala Leu Trp Tyr
        115                 120                 125

Phe Ala Gln Ser Val Ala Asp Ser Leu Asn Leu Arg Arg Leu Val Leu
    130                 135                 140

Met Thr Ser Ser Leu Phe Asn Phe His Ala His Val Ser Leu Pro Gln
145                 150                 155                 160

Phe Asp Glu Leu Gly Tyr Leu Asp Pro Asp Asp Lys Thr Arg Leu Glu
                165                 170                 175

Glu Gln Ala Ser Gly Phe Pro Met Leu Lys Val Lys Asp Ile Lys Ser
            180                 185                 190

Ala Tyr Ser Asn Trp Gln Ile Leu Lys Glu Ile Leu Gly Lys Met Ile
        195                 200                 205

Lys Gln Thr Lys Ala Ser Ser Gly Val Ile Trp Asn Ser Phe Lys Glu
    210                 215                 220

Leu Glu Glu Ser Glu Leu Glu Thr Val Ile Arg Glu Ile Pro Ala Pro
225                 230                 235                 240

Ser Phe Leu Ile Pro Leu Pro Lys His Leu Thr Ala Ser Ser Ser Ser
                245                 250                 255

Leu Leu Asp His Asp Arg Thr Val Phe Gln Trp Leu Asp Gln Gln Pro
            260                 265                 270

Pro Ser Ser Val Leu Tyr Val Ser Phe Gly Ser Thr Ser Glu Val Xaa
        275                 280                 285

Asp Glu Lys Asp Phe Leu Glu Ile Ala Arg Gly Leu Val Asp Ser Lys
    290                 295                 300

Gln Ser Phe Leu Trp Val Val Arg Pro Gly Phe Val Lys Gly Ser Thr
305                 310                 315                 320

Trp Val Glu Pro Leu Pro Asp Gly Phe Leu Gly Glu Arg Gly Arg Ile
                325                 330                 335

Val Lys Trp Val Pro Gln Gln Glu Val Leu Ala His Gly Ala Ile Gly
            340                 345                 350
```

Ala Phe Trp Thr His Ser Gly Trp Asn Ser Thr Leu Glu Ser Val Cys
355 360 365

Glu Gly Val Pro Met Ile Phe Ser Asp Phe Gly Leu Asp Gln Pro Leu
370 375 380

Asn Ala Arg Tyr Met Ser Asp Val Leu Lys Val Gly Val Tyr Leu Glu
385 390 395 400

Asn Gly Trp Glu Arg Gly Glu Ile Ala Asn Ala Ile Arg Arg Val Met
405 410 415

Val Asp Glu Glu Gly Glu Tyr Ile Arg Gln Asn Ala Arg Val Leu Lys
420 425 430

Gln Lys Ala Asp Val Ser Leu Met Lys Gly Gly Ser Ser Tyr Glu Ser
435 440 445

Leu Glu Ser Leu Val Ser Tyr Ile Ser Ser Leu
450 455

<210> 103
<211> 1323
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 103
atgcattcta ccagacatat cttaagacaa agggccgtcc tagttacagg cgctagaaca 60

ccattcgtga aatcatttgg ggctcttatg aaagcagata ccttggaatt ggcatcagca 120

tcagtcgctg ggttgctgaa caagacctca ctggacccta gagatatcga tcatatcgtt 180

tggggtaatg ttgtacttca aggatcagct cataactgcg ccagagaaat agttatcgac 240

cttaacatgc ctaaaaagat catcggtaat ttgacatcta tggcctgtgc ttcaggctta 300

tcttctttgt cacaagcctg tatgctaata gagggtggtc atgccgatgt cgtcattgct 360

ggcggttctg attcagtctc caacactgaa gtgcctttgc caagatccgt cacttacggt 420

ctaatgatgg cccaaaggaa gggtgttatg ggcttcttta aggaagcagg atacaaccca 480

ttcaaatggt ttccaggcgg tattgcttta accgaacgta gtacaggaaa aactatgggt 540

tggcatggag acttaattgc tgagttaaac tctatatcta gagatgacca ggaagccctg 600

gctgtggctt ctcatgcaaa tgctgctaga gcagaaaaag ctgggtactt taaggaggaa 660

attgtacctg tgacaatcga caaaaagggc aaaaagactg aagtaacatg tgatgatgtt 720

atgcaaagag atacagaaaa gatgaaggcc aagatgccat cattgaagcc tgttttcaga 780

aaagagggag gtacaataac agcagccact tccagtactc tgactgatgg tggctctgca 840

atgttggtta tgtcagagga aaaggccaaa aagttgggtt atccaactga tgtctgcgtg 900

aagtcttggt atttcagtgg tatcgatcct tacccacaac ttttgttagc accagttcta 960

ggttggggtc cagctttgaa aaaggccgga ttaaccccta aagatatcga tttgtacgaa 1020

attcacgaag catttgctgc acaagttcta gccacaatta agtgtttgaa gtctcaggaa 1080

ttcttcgata ggtacgctaa cggtgcaaag ccagtattaa ctgaggatat tgatctttct 1140

aaactaaatg ttaatggcgg ttccttagca cttggccacc cattcgccgc tacaggaggt 1200

agaatcgtaa tctctctagc aaatgagttg agaagatccg gaaagagaca cgggctggtc 1260

agtatttgtg cagctggagg gttaggcgga gtagctatac ttgagcatac agcaagtaag 1320

taa 1323

<210> 104
<211> 440
<212> PRT
<213> Leishmania infantum

<400> 104
Met His Ser Thr Arg His Ile Leu Arg Gln Arg Ala Val Leu Val Thr
1 5 10 15

Gly Ala Arg Thr Pro Phe Val Lys Ser Phe Gly Ala Leu Met Lys Ala
20 25 30

Asp Thr Leu Glu Leu Ala Ser Ala Ser Val Ala Gly Leu Leu Asn Lys
35 40 45

Thr Ser Leu Asp Pro Arg Asp Ile Asp His Ile Val Trp Gly Asn Val
50 55 60

Val Leu Gln Gly Ser Ala His Asn Cys Ala Arg Glu Ile Val Ile Asp
65 70 75 80

Leu Asn Met Pro Lys Lys Ile Ile Gly Asn Leu Thr Ser Met Ala Cys
85 90 95

Ala Ser Gly Leu Ser Ser Leu Ser Gln Ala Cys Met Leu Ile Glu Gly
100 105 110

Gly His Ala Asp Val Val Ile Ala Gly Gly Ser Asp Ser Val Ser Asn
115 120 125

Thr Glu Val Pro Leu Pro Arg Ser Val Thr Tyr Gly Leu Met Met Ala
130 135 140

```
Gln Arg Lys Gly Val Met Gly Phe Phe Lys Glu Ala Gly Tyr Asn Pro
145                 150                 155                 160

Phe Lys Trp Phe Pro Gly Gly Ile Ala Leu Thr Glu Arg Ser Thr Gly
                165                 170                 175

Lys Thr Met Gly Trp His Gly Asp Leu Ile Ala Glu Leu Asn Ser Ile
            180                 185                 190

Ser Arg Asp Asp Gln Glu Ala Leu Ala Val Ala Ser His Ala Asn Ala
        195                 200                 205

Ala Arg Ala Glu Lys Ala Gly Tyr Phe Lys Glu Glu Ile Val Pro Val
    210                 215                 220

Thr Ile Asp Lys Lys Gly Lys Lys Thr Glu Val Thr Cys Asp Asp Val
225                 230                 235                 240

Met Gln Arg Asp Thr Glu Lys Met Lys Ala Lys Met Pro Ser Leu Lys
                245                 250                 255

Pro Val Phe Arg Lys Glu Gly Gly Thr Ile Thr Ala Ala Thr Ser Ser
            260                 265                 270

Thr Leu Thr Asp Gly Gly Ser Ala Met Leu Val Met Ser Glu Glu Lys
        275                 280                 285

Ala Lys Lys Leu Gly Tyr Pro Thr Asp Val Cys Val Lys Ser Trp Tyr
    290                 295                 300

Phe Ser Gly Ile Asp Pro Tyr Pro Gln Leu Leu Leu Ala Pro Val Leu
305                 310                 315                 320

Gly Trp Gly Pro Ala Leu Lys Lys Ala Gly Leu Thr Pro Lys Asp Ile
                325                 330                 335

Asp Leu Tyr Glu Ile His Glu Ala Phe Ala Ala Gln Val Leu Ala Thr
            340                 345                 350

Ile Lys Cys Leu Lys Ser Gln Glu Phe Phe Asp Arg Tyr Ala Asn Gly
        355                 360                 365

Ala Lys Pro Val Leu Thr Glu Asp Ile Asp Leu Ser Lys Leu Asn Val
    370                 375                 380

Asn Gly Gly Ser Leu Ala Leu Gly His Pro Phe Ala Ala Thr Gly Gly
385                 390                 395                 400
```

Arg Ile Val Ile Ser Leu Ala Asn Glu Leu Arg Arg Ser Gly Lys Arg
405 410 415

His Gly Leu Val Ser Ile Cys Ala Ala Gly Gly Leu Gly Gly Val Ala
420 425 430

Ile Leu Glu His Thr Ala Ser Lys
435 440

<210> 105
<211> 1584
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 105
atggcagctg accaattggt gaaaactgaa gtcaccaaga agtcttttac tgctcctgta 60

caaaaggctt ctacaccagt tttaaccaat aaaacagtca tttctggatc gaaagtcaaa 120

agtttatcat ctgcgcaatc gagctcatca ggaccttcat catctagtga ggaagatgat 180

tcccgcgata ttgaaagctt ggataagaaa atacgtcctt tagaagaatt agaagcatta 240

ttaagtagtg gaaatacaaa acaattgaag aacaaagagg tcgctgcctt ggttattcac 300

ggtaagttac ctttgtacgc tttggagaaa aaattaggtg atactacgag agcggttgcg 360

gtacgtagga aggctctttc aattttggca gaagctcctg tattagcatc tgatcgttta 420

ccatataaaa attatgacta cgaccgcgta tttggcgctt gttgtgaaaa tgttataggt 480

tacatgcctt tgcccgttgg tgttataggc cccttggtta tcgatggtac atcttatcat 540

ataccaatgg caactacaga gggttgtttg gtagcttctg ccatgcgtgg ctgtaaggca 600

atcaatgctg gcggtggtgc aacaactgtt ttaactaagg atggtatgac aagaggccca 660

gtagtccgtt tcccaacttt gaaaagatct ggtgcctgta agatatggtt agactcagaa 720

gagggacaaa acgcaattaa aaaagctttt aactctacat caagatttgc acgtctgcaa 780

catattcaaa cttgtctagc aggagattta ctcttcatga gatttagaac aactactggt 840

gacgcaatgg gtatgaatat gatttctaaa ggtgtcgaat actcattaaa gcaaatggta 900

gaagagtatg gctgggaaga tatggaggtt gtctccgttt ctggtaacta ctgtaccgac 960

aaaaaaccag ctgccatcaa ctggatcgaa ggtcgtggta agagtgtcgt cgcagaagct 1020

actattcctg gtgatgttgt cagaaaagtg ttaaaaagtg atgtttccgc attggttgag 1080

ttgaacattg ctaagaattt ggttggatct gcaatggctg ggtctgttgg tggatttaac 1140

gcacatgcag ctaatttagt gacagctgtt ttcttggcat taggacaaga tcctgcacaa 1200

aatgttgaaa gttccaactg tataacattg atgaaagaag tggacggtga tttgagaatt 1260

tccgtatcca tgccatccat cgaagtaggt accatcggtg gtggtactgt tctagaacca 1320

caaggtgcca tgttggactt attaggtgta agaggcccgc atgctaccgc tcctggtacc 1380

aacgcacgtc aattagcaag aatagttgcc tgtgccgtct tggcaggtga attatcctta 1440

tgtgctgccc tagcagccgg ccatttggtt caaagtcata tgacccacaa caggaaacct 1500

gctgaaccaa caaaacctaa caatttggac gccactgata taaatcgttt gaaagatggg 1560

tccgtcacct gcattaaatc ctaa 1584

<210> 106
<211> 527
<212> PRT
<213> Saccharomyces cerevisiae

<400> 106
Met Ala Ala Asp Gln Leu Val Lys Thr Glu Val Thr Lys Lys Ser Phe
1 5 10 15

Thr Ala Pro Val Gln Lys Ala Ser Thr Pro Val Leu Thr Asn Lys Thr
20 25 30

Val Ile Ser Gly Ser Lys Val Lys Ser Leu Ser Ser Ala Gln Ser Ser
35 40 45

Ser Ser Gly Pro Ser Ser Ser Ser Glu Glu Asp Asp Ser Arg Asp Ile
50 55 60

Glu Ser Leu Asp Lys Lys Ile Arg Pro Leu Glu Glu Leu Glu Ala Leu
65 70 75 80

Leu Ser Ser Gly Asn Thr Lys Gln Leu Lys Asn Lys Glu Val Ala Ala
85 90 95

Leu Val Ile His Gly Lys Leu Pro Leu Tyr Ala Leu Glu Lys Lys Leu
100 105 110

Gly Asp Thr Thr Arg Ala Val Ala Val Arg Arg Lys Ala Leu Ser Ile
115 120 125

Leu Ala Glu Ala Pro Val Leu Ala Ser Asp Arg Leu Pro Tyr Lys Asn
130 135 140

Tyr Asp Tyr Asp Arg Val Phe Gly Ala Cys Cys Glu Asn Val Ile Gly
145 150 155 160

Tyr Met Pro Leu Pro Val Gly Val Ile Gly Pro Leu Val Ile Asp Gly

165 170 175

Thr Ser Tyr His Ile Pro Met Ala Thr Thr Glu Gly Cys Leu Val Ala
180 185 190

Ser Ala Met Arg Gly Cys Lys Ala Ile Asn Ala Gly Gly Gly Ala Thr
195 200 205

Thr Val Leu Thr Lys Asp Gly Met Thr Arg Gly Pro Val Val Arg Phe
210 215 220

Pro Thr Leu Lys Arg Ser Gly Ala Cys Lys Ile Trp Leu Asp Ser Glu
225 230 235 240

Glu Gly Gln Asn Ala Ile Lys Lys Ala Phe Asn Ser Thr Ser Arg Phe
245 250 255

Ala Arg Leu Gln His Ile Gln Thr Cys Leu Ala Gly Asp Leu Leu Phe
260 265 270

Met Arg Phe Arg Thr Thr Thr Gly Asp Ala Met Gly Met Asn Met Ile
275 280 285

Ser Lys Gly Val Glu Tyr Ser Leu Lys Gln Met Val Glu Glu Tyr Gly
290 295 300

Trp Glu Asp Met Glu Val Val Ser Val Ser Gly Asn Tyr Cys Thr Asp
305 310 315 320

Lys Lys Pro Ala Ala Ile Asn Trp Ile Glu Gly Arg Gly Lys Ser Val
325 330 335

Val Ala Glu Ala Thr Ile Pro Gly Asp Val Val Arg Lys Val Leu Lys
340 345 350

Ser Asp Val Ser Ala Leu Val Glu Leu Asn Ile Ala Lys Asn Leu Val
355 360 365

Gly Ser Ala Met Ala Gly Ser Val Gly Gly Phe Asn Ala His Ala Ala
370 375 380

Asn Leu Val Thr Ala Val Phe Leu Ala Leu Gly Gln Asp Pro Ala Gln
385 390 395 400

Asn Val Glu Ser Ser Asn Cys Ile Thr Leu Met Lys Glu Val Asp Gly
405 410 415

Asp Leu Arg Ile Ser Val Ser Met Pro Ser Ile Glu Val Gly Thr Ile
420 425 430

Gly Gly Gly Thr Val Leu Glu Pro Gln Gly Ala Met Leu Asp Leu Leu
435 440 445

Gly Val Arg Gly Pro His Ala Thr Ala Pro Gly Thr Asn Ala Arg Gln
450 455 460

Leu Ala Arg Ile Val Ala Cys Ala Val Leu Ala Gly Glu Leu Ser Leu
465 470 475 480

Cys Ala Ala Leu Ala Ala Gly His Leu Val Gln Ser His Met Thr His
485 490 495

Asn Arg Lys Pro Ala Glu Pro Thr Lys Pro Asn Asn Leu Asp Ala Thr
500 505 510

Asp Ile Asn Arg Leu Lys Asp Gly Ser Val Thr Cys Ile Lys Ser
515 520 525

<210> 107
<211> 3681
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 107
atgagagctg tccttagatt gttatcaaca catactgttt tctctcctat tgaaacaatt 60

gtatctgttt tcgtgttagc tacattagct tacttccaca tcttgtccgg aatcaagcac 120

tcaagtttct ttgcatcttc tcatcctcct gctatcagac ctgcttttgc acatctgacc 180

aacggggaat gggttgccgt ctcccaacat gattggactg aagcatggaa gcatcctggc 240

ggttcacttg atgcattaga acttcaacaa gtagttttca ctttagatga caagactcaa 300

ccatctgctg tgctagatgc atccgcaatt agtcagcact tagtttccaa tgttcctgca 360

ttatctggaa aagcctactc ttcattgtgc caccatccaa atgtatcagg cacctcctgt 420

tttacatcag tttctggtcc aggagcttca ccaatcttga cactgagttt taagcctgga 480

actagagacg attggttagg atcattaagg aaggagaaaa ctatcacact agatggggtt 540

aagtacgacg ttggagccgg aaaaagacaa gagtcaatcg gcgatatgga atcatctaag 600

tgggttgctt atgcattatc agctttggta cttagatttt gggaattaac aaaggcagat 660

tccttagata tactagtggt tctaactggg tacatcctaa tgcacgtaac attcatgaga 720

ttgttcttgg catccagagc acttggcagt aacttttggt tatcagctgg catattctcc 780

tccgcaacaa tttctttcct attcacttta ccaatgtgta gatctatgga tattccactt 840

gatccaattg ccttgacaga agccctgcca ttcttggtgt gtaccgtagg ttttgacaaa 900

ccacttagat tggcaagagc tgtgatggct catcctaata tccttaaacc tcaagatgat 960

ggtaggatga aagctgccgg agatgtcatt cttgaggcac tggacagagt tggtaacatg 1020

atattgagag attacgcttt agagatcgca gttctattcg ttggcgttaa ctccagagtt 1080

ggcggtctta aggaattttg tgctgtagct gcagcattac ttgctatgga cagattaatg 1140

acattcacac tttatacagc agtgttaacc atcatggttg aggtaaggcg tatcaaaaag 1200

gtcagagata tgactaaggc tagatctaga agttcttcta ttaccgccgt tacagccaac 1260

ggcaccgcca taagaggcgt tttgagtaga aaatcttcaa aacaatctgt gacagaacca 1320

gagacaacta aaaacctaag acaaagagcc actgattcag ccatcggtgt taagggttca 1380

ttgctgaaag atggaggcag attgcaggaa gccgaggaga atccaatggc aagattaaag 1440

ctattgttaa tcgcttcctt cttaacacta cacatcttga acttttgtac tactttgact 1500

tcagccacag ctaacgcaag acatcaaaga catcctttta gaaccgttca agaggtagta 1560

ccaattccta gagttgacat tactacccca gccatagcca atatcttgtc tcatctagct 1620

gtggctcagg aacctatgtt cactgttgtt ggcagtgaac ctatcgaact tcttgttaaa 1680

gtcgctgctc cagtctacgt ccatgctcta ccattggccc ctgctttaag agcttcaaac 1740

actaatactg gagaagctat tgaaaacttt atgagttcat ggtctagtct ggtaggtgac 1800

ccagttgtta gtaagtggat cgtagcattg ctagctgtct ctgttgcatt gaatggatac 1860

ttgttaaagg gtatagccgc aggttccggg ttggctgcca tgagagctgt tagatctcaa 1920

ggtgttcgtt tcagatctag agctagaagt atcgtaaaga tatctgatga acctgagcca 1980

gagccagaac actctatcga cccagcacca gtagtgttct tcgcttccgc agcaccagct 2040

gtagaggccc ctgctccagc tcctgcacct gaaccagaac caccagtcaa cagaccacca 2100

ccattgacta ttttctcaag accactgaac ttagaaacag tggacaaaaa gttacaagat 2160

gctctgccaa taagatcccc accacctgtt gaaccaatca ctccagaatc tagagaagtg 2220

gaaccaaccc aagtagaagt aagatctcta gctgaatgtg tggatgtgtt cgagaatggg 2280

ccaagaccag tctcagtggc tttaaagact ctgaatgatg aggaagttat cctgctttgc 2340

caaacaggta agatagctcc atatgcattg gttaagatgt tggctgattt cgatagggcc 2400

gtacgtgtca gaagagcact tattagtaga gcttcacgta caaaaacttt agaaaactca 2460

ctggttccta tgaaagatta tgattacgcc agagtcatgg gtgcctgttg tgaaaacgtt 2520

atcggataca tgccattacc actagggatt gcaggtccat tgaagattga tggcttgatg 2580

tatcctatac caatggcaac cgcagaaggt accttggttg catctacttc taggggctgt 2640

aaggccttaa atgctggtgg aggggtcaca actgtcttga cagcagatgg catgacaaga 2700

gggccagcta tagactttcc ttccatcgtc agagctgcag aggctaaggc cttcattgaa 2760

tcagaagatg gatacgctac aatcagggag gctttcgagt ctacttctag atttgccaag 2820

ttgcaaaaga tcaagtgtgc actagctggt cgtactcttt ttgtcagatt tgctactaga 2880

acaggagatg ccatgggtat gaacatgatt tctaaggcta ccgaaaaggc acttgatgtc 2940

ctgagtcacg agttccctga aatggtcgtc cttgctttgt ctggtaacta ctgcacagac 3000

aaaaagcctg cagctatttc atggatcgaa ggtaggggaa aatctattgt agcagaagca 3060

gttattcctg gtaaggtcgt taagtcagtc ctgaaaacaa cagtcgagtc tctttgcaat 3120

gtcaacacta agaaaaacct gattggttca gccatggcag gttctgttgg tggtttcaac 3180

gctcatgccg ccaacatcct aacagctgtg ttcctagcca caggtcagga tcctgctcaa 3240

aatgtcgaat cttctaattg catgacttta atggaaccaa caaacggcgg tgaggatttg 3300

ctaatgacaa tttcaatgcc atgtatagag gtaggaaccg ttggtggagg gacaattctg 3360

gaaccacaag gtgcagtttt ggatttgttg ggcgttagag gggctcaccc tactaatcct 3420

ggtcaaaacg ctcaacagtt agccagaatt atcgcatcag ctgtaatggc aggcgaattg 3480

tctttgataa gtgccttagc cgcaggtcat ttggttagag ctcatcttgc ccacaatcgt 3540

tctcaattga atacaccaat gccatccaga ccacatactc ctggccctga ggatgtctca 3600

catgtgcagc agctacctac accatctgca tctgatgata aaggtgttac agctcaaggt 3660

tacgttgtcg aagcaaaata a 3681

<210> 108
<211> 1226
<212> PRT
<213> Ganoderma lucidum

<400> 108

```
Met Arg Ala Val Leu Arg Leu Leu Ser Thr His Thr Val Phe Ser Pro
1               5                   10                  15

Ile Glu Thr Ile Val Ser Val Phe Val Leu Ala Thr Leu Ala Tyr Phe
            20                  25                  30

His Ile Leu Ser Gly Ile Lys His Ser Ser Phe Phe Ala Ser Ser His
        35                  40                  45

Pro Pro Ala Ile Arg Pro Ala Phe Ala His Leu Thr Asn Gly Glu Trp
    50                  55                  60

Val Ala Val Ser Gln His Asp Trp Thr Glu Ala Trp Lys His Pro Gly
65                  70                  75                  80
```

```
Gly Ser Leu Asp Ala Leu Glu Leu Gln Gln Val Val Phe Thr Leu Asp
                85                  90                  95

Asp Lys Thr Gln Pro Ser Ala Val Leu Asp Ala Ser Ala Ile Ser Gln
            100                 105                 110

His Leu Val Ser Asn Val Pro Ala Leu Ser Gly Lys Ala Tyr Ser Ser
        115                 120                 125

Leu Cys His His Pro Asn Val Ser Gly Thr Ser Cys Phe Thr Ser Val
    130                 135                 140

Ser Gly Pro Gly Ala Ser Pro Ile Leu Thr Leu Ser Phe Lys Pro Gly
145                 150                 155                 160

Thr Arg Asp Asp Trp Leu Gly Ser Leu Arg Lys Glu Lys Thr Ile Thr
                165                 170                 175

Leu Asp Gly Val Lys Tyr Asp Val Gly Ala Gly Lys Arg Gln Glu Ser
            180                 185                 190

Ile Gly Asp Met Glu Ser Ser Lys Trp Val Ala Tyr Ala Leu Ser Ala
        195                 200                 205

Leu Val Leu Arg Phe Trp Glu Leu Thr Lys Ala Asp Ser Leu Asp Ile
    210                 215                 220

Leu Val Val Leu Thr Gly Tyr Ile Leu Met His Val Thr Phe Met Arg
225                 230                 235                 240

Leu Phe Leu Ala Ser Arg Ala Leu Gly Ser Asn Phe Trp Leu Ser Ala
                245                 250                 255

Gly Ile Phe Ser Ser Ala Thr Ile Ser Phe Leu Phe Thr Leu Pro Met
            260                 265                 270

Cys Arg Ser Met Asp Ile Pro Leu Asp Pro Ile Ala Leu Thr Glu Ala
        275                 280                 285

Leu Pro Phe Leu Val Cys Thr Val Gly Phe Asp Lys Pro Leu Arg Leu
    290                 295                 300

Ala Arg Ala Val Met Ala His Pro Asn Ile Leu Lys Pro Gln Asp Asp
305                 310                 315                 320

Gly Arg Met Lys Ala Ala Gly Asp Val Ile Leu Glu Ala Leu Asp Arg
                325                 330                 335
```

```
Val Gly Asn Met Ile Leu Arg Asp Tyr Ala Leu Glu Ile Ala Val Leu
            340                 345                 350

Phe Val Gly Val Asn Ser Arg Val Gly Gly Leu Lys Glu Phe Cys Ala
        355                 360                 365

Val Ala Ala Ala Leu Leu Ala Met Asp Arg Leu Met Thr Phe Thr Leu
    370                 375                 380

Tyr Thr Ala Val Leu Thr Ile Met Val Glu Val Arg Arg Ile Lys Lys
385                 390                 395                 400

Val Arg Asp Met Thr Lys Ala Arg Ser Arg Ser Ser Ser Ile Thr Ala
                405                 410                 415

Val Thr Ala Asn Gly Thr Ala Ile Arg Gly Val Leu Ser Arg Lys Ser
            420                 425                 430

Ser Lys Gln Ser Val Thr Glu Pro Glu Thr Thr Lys Asn Leu Arg Gln
        435                 440                 445

Arg Ala Thr Asp Ser Ala Ile Gly Val Lys Gly Ser Leu Leu Lys Asp
    450                 455                 460

Gly Gly Arg Leu Gln Glu Ala Glu Glu Asn Pro Met Ala Arg Leu Lys
465                 470                 475                 480

Leu Leu Leu Ile Ala Ser Phe Leu Thr Leu His Ile Leu Asn Phe Cys
                485                 490                 495

Thr Thr Leu Thr Ser Ala Thr Ala Asn Ala Arg His Gln Arg His Pro
            500                 505                 510

Phe Arg Thr Val Gln Glu Val Val Pro Ile Pro Arg Val Asp Ile Thr
        515                 520                 525

Thr Pro Ala Ile Ala Asn Ile Leu Ser His Leu Ala Val Ala Gln Glu
    530                 535                 540

Pro Met Phe Thr Val Val Gly Ser Glu Pro Ile Glu Leu Leu Val Lys
545                 550                 555                 560

Val Ala Ala Pro Val Tyr Val His Ala Leu Pro Leu Ala Pro Ala Leu
                565                 570                 575

Arg Ala Ser Asn Thr Asn Thr Gly Glu Ala Ile Glu Asn Phe Met Ser
            580                 585                 590
```

```
Ser Trp Ser Ser Leu Val Gly Asp Pro Val Val Ser Lys Trp Ile Val
        595                 600                 605

Ala Leu Leu Ala Val Ser Val Ala Leu Asn Gly Tyr Leu Leu Lys Gly
    610                 615                 620

Ile Ala Ala Gly Ser Gly Leu Ala Ala Met Arg Ala Val Arg Ser Gln
625                 630                 635                 640

Gly Val Arg Phe Arg Ser Arg Ala Arg Ser Ile Val Lys Ile Ser Asp
                645                 650                 655

Glu Pro Glu Pro Glu Pro Glu His Ser Ile Asp Pro Ala Pro Val Val
            660                 665                 670

Phe Phe Ala Ser Ala Ala Pro Ala Val Glu Ala Pro Ala Pro Ala Pro
        675                 680                 685

Ala Pro Glu Pro Glu Pro Pro Val Asn Arg Pro Pro Pro Leu Thr Ile
    690                 695                 700

Phe Ser Arg Pro Leu Asn Leu Glu Thr Val Asp Lys Lys Leu Gln Asp
705                 710                 715                 720

Ala Leu Pro Ile Arg Ser Pro Pro Pro Val Glu Pro Ile Thr Pro Glu
                725                 730                 735

Ser Arg Glu Val Glu Pro Thr Gln Val Glu Val Arg Ser Leu Ala Glu
            740                 745                 750

Cys Val Asp Val Phe Glu Asn Gly Pro Arg Pro Val Ser Val Ala Leu
        755                 760                 765

Lys Thr Leu Asn Asp Glu Glu Val Ile Leu Leu Cys Gln Thr Gly Lys
    770                 775                 780

Ile Ala Pro Tyr Ala Leu Val Lys Met Leu Ala Asp Phe Asp Arg Ala
785                 790                 795                 800

Val Arg Val Arg Arg Ala Leu Ile Ser Arg Ala Ser Arg Thr Lys Thr
                805                 810                 815

Leu Glu Asn Ser Leu Val Pro Met Lys Asp Tyr Asp Tyr Ala Arg Val
            820                 825                 830

Met Gly Ala Cys Cys Glu Asn Val Ile Gly Tyr Met Pro Leu Pro Leu
```

```
        835                 840                 845

Gly Ile Ala Gly Pro Leu Lys Ile Asp Gly Leu Met Tyr Pro Ile Pro
    850                 855                 860

Met Ala Thr Ala Glu Gly Thr Leu Val Ala Ser Thr Ser Arg Gly Cys
865                 870                 875                 880

Lys Ala Leu Asn Ala Gly Gly Gly Val Thr Thr Val Leu Thr Ala Asp
                885                 890                 895

Gly Met Thr Arg Gly Pro Ala Ile Asp Phe Pro Ser Ile Val Arg Ala
            900                 905                 910

Ala Glu Ala Lys Ala Phe Ile Glu Ser Glu Asp Gly Tyr Ala Thr Ile
        915                 920                 925

Arg Glu Ala Phe Glu Ser Thr Ser Arg Phe Ala Lys Leu Gln Lys Ile
    930                 935                 940

Lys Cys Ala Leu Ala Gly Arg Thr Leu Phe Val Arg Phe Ala Thr Arg
945                 950                 955                 960

Thr Gly Asp Ala Met Gly Met Asn Met Ile Ser Lys Ala Thr Glu Lys
                965                 970                 975

Ala Leu Asp Val Leu Ser His Glu Phe Pro Glu Met Val Val Leu Ala
            980                 985                 990

Leu Ser Gly Asn Tyr Cys Thr Asp  Lys Lys Pro Ala Ala  Ile Ser Trp
        995                 1000                1005

Ile Glu  Gly Arg Gly Lys Ser  Ile Val Ala Glu Ala  Val Ile Pro
    1010                1015                1020

Gly Lys  Val Val Lys Ser Val  Leu Lys Thr Thr Val  Glu Ser Leu
    1025                1030                1035

Cys Asn  Val Asn Thr Lys Lys  Asn Leu Ile Gly Ser  Ala Met Ala
    1040                1045                1050

Gly Ser  Val Gly Gly Phe Asn  Ala His Ala Ala Asn  Ile Leu Thr
    1055                1060                1065

Ala Val  Phe Leu Ala Thr Gly  Gln Asp Pro Ala Gln  Asn Val Glu
    1070                1075                1080
```

Ser Ser Asn Cys Met Thr Leu Met Glu Pro Thr Asn Gly Gly Glu
1085 1090 1095

Asp Leu Leu Met Thr Ile Ser Met Pro Cys Ile Glu Val Gly Thr
1100 1105 1110

Val Gly Gly Gly Thr Ile Leu Glu Pro Gln Gly Ala Val Leu Asp
1115 1120 1125

Leu Leu Gly Val Arg Gly Ala His Pro Thr Asn Pro Gly Gln Asn
1130 1135 1140

Ala Gln Gln Leu Ala Arg Ile Ile Ala Ser Ala Val Met Ala Gly
1145 1150 1155

Glu Leu Ser Leu Ile Ser Ala Leu Ala Ala Gly His Leu Val Arg
1160 1165 1170

Ala His Leu Ala His Asn Arg Ser Gln Leu Asn Thr Pro Met Pro
1175 1180 1185

Ser Arg Pro His Thr Pro Gly Pro Glu Asp Val Ser His Val Gln
1190 1195 1200

Gln Leu Pro Thr Pro Ser Ala Ser Asp Asp Lys Gly Val Thr Ala
1205 1210 1215

Gln Gly Tyr Val Val Glu Ala Lys
1220 1225

<210> 109
<211> 2667
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 109
atgttatcaa gattgttcag aatgcatggt ctatttgttg cttctcaccc ttgggaagta 60

atagttggta ctgtaacatt aacgatctgt atgatgtcta tgaacatgtt taccggaaac 120

aacaagattt gtggttggaa ttatgagtgt cctaagctgg aagaggatgt gttgagttca 180

gacatcatca tacttactat aacaagatgc attgcaatat tgtatatcta cttccaattt 240

caaaacctta gacaattggg tagtaaatac atcctaggca tcgccggatt gttcactatt 300

ttctctagtt ttgttttctc aaccgtcgtt attcactttt tggacaaaga gttaactggt 360

ttgaacgaag ctctaccatt cttcttgctg ctggtagatt tgtccagagc ttccgcttta 420

gctaaattcg ctctgtcctc taattctcaa gatgaagtta gagagaatat agcaagggga 480

atggccatac ttggacctac tttcacactt gatgcccttg tcgaatgttt ggttattggg 540

gttggcacaa tgtccggcgt tagacagtta gaaatcatgt gttgttttgg ctgtatgagt 600

gtcttggcta actactttgt ctttatgaca ttctttccag cttgcgtttc tttggtattg 660

gagctgtcaa gagaatcaag agaaggcaga ccaatatggc aactatcaca tttcgccaga 720

gtgttagaag aggaggaaaa caaacctaat cctgtcacac agagagtgaa aatgatcatg 780

tctttgggtt tagtcctagt gcatgctcat tctagatgga tcgcagatcc atcccctcag 840

aattctacag ctgataactc taaagttagt ttaggtttag atgaaaatgt aagtaagagg 900

attgaacctt ccgtgtcttt gtggcaattc tacttatcaa aaatgatttc catggatatt 960

gaacaagtga taacgttgtc tttggcttta ttgttagccg ttaagtacat tttctttgag 1020

caagccgaaa cggaatctac attatcactg aaaaacccaa ttacatcccc agtcgttacc 1080

cagaaaaaga taactgatga ttgctgtaga agagatccag tgttggtcag gaatgatcaa 1140

aagttccacg ccatggagga ggaaactagg aaaaacagag aaaggaaagt tgaagttatc 1200

aagcctctat tagcagaaaa tgacacttca catagggcca ctttcgttgt cggcaattca 1260

tctcttttag gtacgtcatt ggagctggaa acacaggaac cagaaatgga actaccagtt 1320

gaaccaagac caaatgagga atgtttgcaa atactagaga acgctgaaaa gggagccaag 1380

ttcctatctg atgccgagat tatccagctg gtcaatgcca agcacattcc tgcctacaag 1440

ttggaaaccc ttatggagac acatgagaga ggtgtgtcta ttaggagaca attactatct 1500

aaaaagttac ctgaaccaag ttccctacaa tacctgcctt atagagatta caattactcc 1560

ttggtaatgg gagcttgttg tgaaaatgtc attgggtaca tgccaattcc agtgggtgtc 1620

gccggtccac tatgtttgga cggtaaggaa tttcaagtac ctatggcaac gactgaaggc 1680

tgcttagttg catctacaaa cagaggttgt agagccattg gattaggtgg cggtgcttct 1740

tcaagagtct tggctgacgg tatgactaga ggtcctgttg tgagatttcc tagggcctgt 1800

gactctgcag aagttaaggc ttggttggaa actccagaag gtttcaccgt aatcaaagag 1860

gcctttgatt ccacatcaag ggtggccaga ttacaaaaac tacacatgtc tgtcgctggg 1920

agaaatctgt atatcagatt tcaatccaga tccggcgacg caatgggtat gaatatgatt 1980

tcaaaaggga cagaaaaggc tttgtcaaag ctgcaggagt atttcccaga gatgcaaatc 2040

ttggccgtat ctggcaacta ttgcacagac aaaaagcctg ccgccatcaa ctggattgaa 2100

ggaagaggca aatctgtggt ttgtgaagct gtaattccag ccaaagttgt tagagaagtg 2160

ttaaagacca caacagaagc tatgattgaa gtaaacataa acaaaaactt agtagggtct 2220

gccatggctg gttcaattgg aggatacaac gctcatgctg ccaatattgt aaccgctatc 2280

tacatcgcat gtggacaaga tgctgcccaa aatgtcggtt cctcaaattg catcacattg 2340

atggaagcat ctggccctac aaacgaggat ttgtatatca gttgcacaat gccatctata 2400

gaaataggga ctgtgggagg aggaactaac ttacttccac agcaagcctg cttacaaatg 2460

ctgggtgtac aaggagcctg tagagataat ccaggggaga acgctagaca acttgccaga 2520

attgtttgtg ggacagttat ggctggtgaa cttagtctaa tggcagcttt ggctgctggg 2580

cacctggtga gatctcatat gattcataat agaagtaaga ttaaccttca agatttgcaa 2640

ggtacgtgta cgaaaaaggc tgcctaa 2667

<210> 110
<211> 888
<212> PRT
<213> Bos taurus

<400> 110

```
Met Leu Ser Arg Leu Phe Arg Met His Gly Leu Phe Val Ala Ser His
1               5                  10                  15

Pro Trp Glu Val Ile Val Gly Thr Val Thr Leu Thr Ile Cys Met Met
            20                  25                  30

Ser Met Asn Met Phe Thr Gly Asn Asn Lys Ile Cys Gly Trp Asn Tyr
        35                  40                  45

Glu Cys Pro Lys Leu Glu Glu Asp Val Leu Ser Ser Asp Ile Ile Ile
    50                  55                  60

Leu Thr Ile Thr Arg Cys Ile Ala Ile Leu Tyr Ile Tyr Phe Gln Phe
65                  70                  75                  80

Gln Asn Leu Arg Gln Leu Gly Ser Lys Tyr Ile Leu Gly Ile Ala Gly
                85                  90                  95

Leu Phe Thr Ile Phe Ser Ser Phe Val Phe Ser Thr Val Val Ile His
            100                 105                 110

Phe Leu Asp Lys Glu Leu Thr Gly Leu Asn Glu Ala Leu Pro Phe Phe
        115                 120                 125

Leu Leu Leu Val Asp Leu Ser Arg Ala Ser Ala Leu Ala Lys Phe Ala
    130                 135                 140

Leu Ser Ser Asn Ser Gln Asp Glu Val Arg Glu Asn Ile Ala Arg Gly
145                 150                 155                 160

Met Ala Ile Leu Gly Pro Thr Phe Thr Leu Asp Ala Leu Val Glu Cys
                165                 170                 175
```

```
Leu Val Ile Gly Val Gly Thr Met Ser Gly Val Arg Gln Leu Glu Ile
            180                 185                 190

Met Cys Cys Phe Gly Cys Met Ser Val Leu Ala Asn Tyr Phe Val Phe
        195                 200                 205

Met Thr Phe Phe Pro Ala Cys Val Ser Leu Val Leu Glu Leu Ser Arg
    210                 215                 220

Glu Ser Arg Glu Gly Arg Pro Ile Trp Gln Leu Ser His Phe Ala Arg
225                 230                 235                 240

Val Leu Glu Glu Glu Glu Asn Lys Pro Asn Pro Val Thr Gln Arg Val
                245                 250                 255

Lys Met Ile Met Ser Leu Gly Leu Val Leu Val His Ala His Ser Arg
            260                 265                 270

Trp Ile Ala Asp Pro Ser Pro Gln Asn Ser Thr Ala Asp Asn Ser Lys
        275                 280                 285

Val Ser Leu Gly Leu Asp Glu Asn Val Ser Lys Arg Ile Glu Pro Ser
    290                 295                 300

Val Ser Leu Trp Gln Phe Tyr Leu Ser Lys Met Ile Ser Met Asp Ile
305                 310                 315                 320

Glu Gln Val Ile Thr Leu Ser Leu Ala Leu Leu Leu Ala Val Lys Tyr
                325                 330                 335

Ile Phe Phe Glu Gln Ala Glu Thr Glu Ser Thr Leu Ser Leu Lys Asn
            340                 345                 350

Pro Ile Thr Ser Pro Val Val Thr Gln Lys Lys Ile Thr Asp Asp Cys
        355                 360                 365

Cys Arg Arg Asp Pro Val Leu Val Arg Asn Asp Gln Lys Phe His Ala
    370                 375                 380

Met Glu Glu Glu Thr Arg Lys Asn Arg Glu Arg Lys Val Glu Val Ile
385                 390                 395                 400

Lys Pro Leu Leu Ala Glu Asn Asp Thr Ser His Arg Ala Thr Phe Val
                405                 410                 415

Val Gly Asn Ser Ser Leu Leu Gly Thr Ser Leu Glu Leu Glu Thr Gln
```

420 425 430

Glu Pro Glu Met Glu Leu Pro Val Glu Pro Arg Pro Asn Glu Glu Cys
435 440 445

Leu Gln Ile Leu Glu Asn Ala Glu Lys Gly Ala Lys Phe Leu Ser Asp
450 455 460

Ala Glu Ile Ile Gln Leu Val Asn Ala Lys His Ile Pro Ala Tyr Lys
465 470 475 480

Leu Glu Thr Leu Met Glu Thr His Glu Arg Gly Val Ser Ile Arg Arg
485 490 495

Gln Leu Leu Ser Lys Lys Leu Pro Glu Pro Ser Ser Leu Gln Tyr Leu
500 505 510

Pro Tyr Arg Asp Tyr Asn Tyr Ser Leu Val Met Gly Ala Cys Cys Glu
515 520 525

Asn Val Ile Gly Tyr Met Pro Ile Pro Val Gly Val Ala Gly Pro Leu
530 535 540

Cys Leu Asp Gly Lys Glu Phe Gln Val Pro Met Ala Thr Thr Glu Gly
545 550 555 560

Cys Leu Val Ala Ser Thr Asn Arg Gly Cys Arg Ala Ile Gly Leu Gly
565 570 575

Gly Gly Ala Ser Ser Arg Val Leu Ala Asp Gly Met Thr Arg Gly Pro
580 585 590

Val Val Arg Phe Pro Arg Ala Cys Asp Ser Ala Glu Val Lys Ala Trp
595 600 605

Leu Glu Thr Pro Glu Gly Phe Thr Val Ile Lys Glu Ala Phe Asp Ser
610 615 620

Thr Ser Arg Val Ala Arg Leu Gln Lys Leu His Met Ser Val Ala Gly
625 630 635 640

Arg Asn Leu Tyr Ile Arg Phe Gln Ser Arg Ser Gly Asp Ala Met Gly
645 650 655

Met Asn Met Ile Ser Lys Gly Thr Glu Lys Ala Leu Ser Lys Leu Gln
660 665 670

Glu Tyr Phe Pro Glu Met Gln Ile Leu Ala Val Ser Gly Asn Tyr Cys
675 680 685

Thr Asp Lys Lys Pro Ala Ala Ile Asn Trp Ile Glu Gly Arg Gly Lys
690 695 700

Ser Val Val Cys Glu Ala Val Ile Pro Ala Lys Val Val Arg Glu Val
705 710 715 720

Leu Lys Thr Thr Thr Glu Ala Met Ile Glu Val Asn Ile Asn Lys Asn
725 730 735

Leu Val Gly Ser Ala Met Ala Gly Ser Ile Gly Gly Tyr Asn Ala His
740 745 750

Ala Ala Asn Ile Val Thr Ala Ile Tyr Ile Ala Cys Gly Gln Asp Ala
755 760 765

Ala Gln Asn Val Gly Ser Ser Asn Cys Ile Thr Leu Met Glu Ala Ser
770 775 780

Gly Pro Thr Asn Glu Asp Leu Tyr Ile Ser Cys Thr Met Pro Ser Ile
785 790 795 800

Glu Ile Gly Thr Val Gly Gly Gly Thr Asn Leu Leu Pro Gln Gln Ala
805 810 815

Cys Leu Gln Met Leu Gly Val Gln Gly Ala Cys Arg Asp Asn Pro Gly
820 825 830

Glu Asn Ala Arg Gln Leu Ala Arg Ile Val Cys Gly Thr Val Met Ala
835 840 845

Gly Glu Leu Ser Leu Met Ala Ala Leu Ala Ala Gly His Leu Val Arg
850 855 860

Ser His Met Ile His Asn Arg Ser Lys Ile Asn Leu Gln Asp Leu Gln
865 870 875 880

Gly Thr Cys Thr Lys Lys Ala Ala
885

<210> 111
<211> 1704
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 111

atggatttga gaaggaaatt accacctaag cctccatctt caacaacaac aaaacagcca 60

agtcataggt cccattctcc tacgccaatt ccaaaggctt cagatgcatt gcctcttcca 120

ttgtacctga ccaatacgtt tttcttcact cttttctttt ccgtagcata ttacctgttg 180

cataggtgga gagacaagat tagatccgga acacctttac acgttgtgac actgactgaa 240

ctatccgcaa ttgtactgct gattgcttcc ttcatctatc ttttaggctt tttcggtatt 300

gattttgtgc aatctttcac atcaagagaa aatgagcaac taaacaacga tgatcacaac 360

gtcgtgtcaa caaacaatgt tttatctgat agaaggttag tttacgacta tggattcgat 420

gtgacaggag acaacgataa cgataatgat gacgatgtta ttgtgaaaag tgtcgtttct 480

ggggaagtta attcttatag tttggaggct tccctaggag attgttacag agccgcaaag 540

attagaaaga gagccgtcga gagaattgtc gggagagaag tattaggctt gggtttcgag 600

ggatttgatt atgaatctat cctggggcaa tgttgtgaaa tgcctatcgg gtacgtccaa 660

gtgccagtag gtgtcgctgg acctttattg ttaaatggtg gggaattcat ggttccaatg 720

gctacaactg aaggctgtct tgtagcttcc actaatagag gttgtaaagc catatgctta 780

tcaggtggtg ccactgccat attgctaaaa gatggtatga caagagcccc agtagtgaga 840

ttcgccacag ctgagagagc ttcacaacta aagttttact tggaagatgg tgtcaatttc 900

gatacattgt ctgttgtctt taacaaaagt tcaagatttg ccagattgca aaacatccaa 960

tgctcaattg ccggtaaaaa cttgtacatt aggtttactt gctccacagg cgacgccatg 1020

ggtatgaaca tggtttcaaa aggagtacaa aatgtattag acttttaca aaatgatttt 1080

cctgatatgg acgtaattgg gatctcttgg aagttctgct ctgacaaaaa gccaacagct 1140

gtcaactgga ttgagggcag aggaaagtct gtcgttttcc aggccgtaat taccaaaaag 1200

gtggttagaa agtctgcact gaaccctcaa acttgcacat gtagaacttt gacctgttta 1260

agaccattat tggttctgct acttctggtt ttgctagtgg acttaatgca tatgcttcat 1320

atcgtgtctg ccgtgttcat cgctaccggt caagatccag ctcagaatat cgaatctagt 1380

cactgtatca ctatgatgga ggctgtcaac aatggtaagg atttgcacgt taatgttacg 1440

atgccatcta tagaagttgg cacggtggga ggtggcactc agctagcctc tcaatcagcc 1500

tgtttgaact tgcttggtgt aaagggtgcc tgtatagaat ccccaggatc aaacgcccag 1560

ttgttagcta gaatcgttgc tggttctgtt ctggcaggcg aattaagttt gatgtcagct 1620

ataagtgctg ggcaactagt taaatctcat atgaaataca ataggtctag tagagatatg 1680

tcagcaatag cttctaaggt ctaa 1704

<210> 112

<211> 567
<212> PRT
<213> Artemisia annua

<400> 112

```
Met Asp Leu Arg Arg Lys Leu Pro Pro Lys Pro Pro Ser Ser Thr Thr
1               5                  10                  15

Thr Lys Gln Pro Ser His Arg Ser His Ser Pro Thr Pro Ile Pro Lys
            20                  25                  30

Ala Ser Asp Ala Leu Pro Leu Pro Leu Tyr Leu Thr Asn Thr Phe Phe
        35                  40                  45

Phe Thr Leu Phe Phe Ser Val Ala Tyr Tyr Leu Leu His Arg Trp Arg
    50                  55                  60

Asp Lys Ile Arg Ser Gly Thr Pro Leu His Val Val Thr Leu Thr Glu
65                  70                  75                  80

Leu Ser Ala Ile Val Leu Leu Ile Ala Ser Phe Ile Tyr Leu Leu Gly
                85                  90                  95

Phe Phe Gly Ile Asp Phe Val Gln Ser Phe Thr Ser Arg Glu Asn Glu
            100                 105                 110

Gln Leu Asn Asn Asp Asp His Asn Val Val Ser Thr Asn Asn Val Leu
        115                 120                 125

Ser Asp Arg Arg Leu Val Tyr Asp Tyr Gly Phe Asp Val Thr Gly Asp
    130                 135                 140

Asn Asp Asn Asp Asn Asp Asp Asp Val Ile Val Lys Ser Val Val Ser
145                 150                 155                 160

Gly Glu Val Asn Ser Tyr Ser Leu Glu Ala Ser Leu Gly Asp Cys Tyr
                165                 170                 175

Arg Ala Ala Lys Ile Arg Lys Arg Ala Val Glu Arg Ile Val Gly Arg
            180                 185                 190

Glu Val Leu Gly Leu Gly Phe Glu Gly Phe Asp Tyr Glu Ser Ile Leu
        195                 200                 205

Gly Gln Cys Cys Glu Met Pro Ile Gly Tyr Val Gln Val Pro Val Gly
    210                 215                 220

Val Ala Gly Pro Leu Leu Leu Asn Gly Gly Glu Phe Met Val Pro Met
225                 230                 235                 240
```

```
Ala Thr Thr Glu Gly Cys Leu Val Ala Ser Thr Asn Arg Gly Cys Lys
                245                 250                 255

Ala Ile Cys Leu Ser Gly Gly Ala Thr Ala Ile Leu Leu Lys Asp Gly
            260                 265                 270

Met Thr Arg Ala Pro Val Val Arg Phe Ala Thr Ala Glu Arg Ala Ser
        275                 280                 285

Gln Leu Lys Phe Tyr Leu Glu Asp Gly Val Asn Phe Asp Thr Leu Ser
    290                 295                 300

Val Val Phe Asn Lys Ser Ser Arg Phe Ala Arg Leu Gln Asn Ile Gln
305                 310                 315                 320

Cys Ser Ile Ala Gly Lys Asn Leu Tyr Ile Arg Phe Thr Cys Ser Thr
                325                 330                 335

Gly Asp Ala Met Gly Met Asn Met Val Ser Lys Gly Val Gln Asn Val
            340                 345                 350

Leu Asp Phe Leu Gln Asn Asp Phe Pro Asp Met Asp Val Ile Gly Ile
        355                 360                 365

Ser Trp Lys Phe Cys Ser Asp Lys Lys Pro Thr Ala Val Asn Trp Ile
    370                 375                 380

Glu Gly Arg Gly Lys Ser Val Val Phe Gln Ala Val Ile Thr Lys Lys
385                 390                 395                 400

Val Val Arg Lys Ser Ala Leu Asn Pro Gln Thr Cys Thr Cys Arg Thr
                405                 410                 415

Leu Thr Cys Leu Arg Pro Leu Leu Val Leu Leu Leu Leu Val Leu Leu
            420                 425                 430

Val Asp Leu Met His Met Leu His Ile Val Ser Ala Val Phe Ile Ala
        435                 440                 445

Thr Gly Gln Asp Pro Ala Gln Asn Ile Glu Ser Ser His Cys Ile Thr
    450                 455                 460

Met Met Glu Ala Val Asn Asn Gly Lys Asp Leu His Val Asn Val Thr
465                 470                 475                 480

Met Pro Ser Ile Glu Val Gly Thr Val Gly Gly Gly Thr Gln Leu Ala
```

485 490 495

Ser Gln Ser Ala Cys Leu Asn Leu Leu Gly Val Lys Gly Ala Cys Ile
500 505 510

Glu Ser Pro Gly Ser Asn Ala Gln Leu Leu Ala Arg Ile Val Ala Gly
515 520 525

Ser Val Leu Ala Gly Glu Leu Ser Leu Met Ser Ala Ile Ser Ala Gly
530 535 540

Gln Leu Val Lys Ser His Met Lys Tyr Asn Arg Ser Ser Arg Asp Met
545 550 555 560

Ser Ala Ile Ala Ser Lys Val
565

<210> 113
<211> 1308
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 113
atgtttagaa gagctatact gttaggatgc tctgctgcca agacaccatg gtctgagtgt 60

tctaacgctc aattagttga tgcagttaag tctagaaaga tctcattcta cggtcttgaa 120

caagccttgg aaccagatta tagaagggct atcgaagtaa ggagagaggt tgtctctgaa 180

atcgcctcac aacagccaga agcaaaaaag aagcaatccg cattgcacac aataccattt 240

gagaattatg attggaataa ggtcgttggc caaaactgtg aaaacattat tggatacgtc 300

ccaataccac tgggcgttgc tggccctatt ttgattgatg gtaaagagta cccaatacca 360

atggctacaa cagaaggcgc tttggtcgct agtactcata gaggtgctag agctattaca 420

agatccggag gttgtaagac attgttatta ggtgaaggta tgacaagagc accagtggtt 480

gaattgcctt cattagagga agctgggcgt ttgcacaagt actgtaatga gaacttctta 540

tctttaaagg aagcatttga atcaactacc caatatggaa aacttaattc tttaaagtgc 600

gtactagctg gtagaaaagc ataccttaga ttcagagcca ctacaggcga tgctatgggc 660

atgaacatga taacaaaggg tgtagacaaa gcactgtctg ttctacagca acatttccct 720

tcaatggaaa tcctagccct aagtggtaat tactgtaccg acaaaaagcc atctgctgta 780

aattggattg atggcagagg taaatcagtg gttgcagaag ccactttatt ggctgatgtt 840

gtcgaagata ctctgaaatg tacagtcgat tctttggtat ccttgaatat cgacaaaaac 900

cttgttgggt cagctatggc tggttctgtt ggaggtttta acgcccaggc tgcaaacgct 960

gtggcagcca ttttcattgc aaccggtcaa gatcctgctc aagtggtaga aagttcaatg 1020

tgtatcacta caatgtccaa ggtaggtaac gatctattga tctctgtgac catgccttct 1080

atcgaggtcg gggtcgtggg aggagggact ggtcttgctg cccaaagagg atgcttagag 1140

ttaatagggt gcggaggccc atctaaggag tctcctggta ctaatgccca acttctaagt 1200

agagttgttg cagctggcgt tttatcagcc gaactttcct tgatgtccgg actggcagca 1260

ggtcatctat tgtcagcaca tatgagattg aacagaaaga agaaataa 1308

<210> 114
<211> 435
<212> PRT
<213> Trypanosoma cruzi

<400> 114

Met Phe Arg Arg Ala Ile Leu Leu Gly Cys Ser Ala Ala Lys Thr Pro
1 5 10 15

Trp Ser Glu Cys Ser Asn Ala Gln Leu Val Asp Ala Val Lys Ser Arg
20 25 30

Lys Ile Ser Phe Tyr Gly Leu Glu Gln Ala Leu Glu Pro Asp Tyr Arg
35 40 45

Arg Ala Ile Glu Val Arg Arg Glu Val Val Ser Glu Ile Ala Ser Gln
50 55 60

Gln Pro Glu Ala Lys Lys Lys Gln Ser Ala Leu His Thr Ile Pro Phe
65 70 75 80

Glu Asn Tyr Asp Trp Asn Lys Val Val Gly Gln Asn Cys Glu Asn Ile
85 90 95

Ile Gly Tyr Val Pro Ile Pro Leu Gly Val Ala Gly Pro Ile Leu Ile
100 105 110

Asp Gly Lys Glu Tyr Pro Ile Pro Met Ala Thr Thr Glu Gly Ala Leu
115 120 125

Val Ala Ser Thr His Arg Gly Ala Arg Ala Ile Thr Arg Ser Gly Gly
130 135 140

Cys Lys Thr Leu Leu Leu Gly Glu Gly Met Thr Arg Ala Pro Val Val
145 150 155 160

Glu Leu Pro Ser Leu Glu Glu Ala Gly Arg Leu His Lys Tyr Cys Asn
165 170 175

```
Glu Asn Phe Leu Ser Leu Lys Glu Ala Phe Glu Ser Thr Thr Gln Tyr
            180                 185                 190

Gly Lys Leu Asn Ser Leu Lys Cys Val Leu Ala Gly Arg Lys Ala Tyr
        195                 200                 205

Leu Arg Phe Arg Ala Thr Thr Gly Asp Ala Met Gly Met Asn Met Ile
    210                 215                 220

Thr Lys Gly Val Asp Lys Ala Leu Ser Val Leu Gln Gln His Phe Pro
225                 230                 235                 240

Ser Met Glu Ile Leu Ala Leu Ser Gly Asn Tyr Cys Thr Asp Lys Lys
                245                 250                 255

Pro Ser Ala Val Asn Trp Ile Asp Gly Arg Gly Lys Ser Val Val Ala
            260                 265                 270

Glu Ala Thr Leu Leu Ala Asp Val Val Glu Asp Thr Leu Lys Cys Thr
        275                 280                 285

Val Asp Ser Leu Val Ser Leu Asn Ile Asp Lys Asn Leu Val Gly Ser
    290                 295                 300

Ala Met Ala Gly Ser Val Gly Gly Phe Asn Ala Gln Ala Ala Asn Ala
305                 310                 315                 320

Val Ala Ala Ile Phe Ile Ala Thr Gly Gln Asp Pro Ala Gln Val Val
                325                 330                 335

Glu Ser Ser Met Cys Ile Thr Thr Met Ser Lys Val Gly Asn Asp Leu
            340                 345                 350

Leu Ile Ser Val Thr Met Pro Ser Ile Glu Val Gly Val Val Gly Gly
        355                 360                 365

Gly Thr Gly Leu Ala Ala Gln Arg Gly Cys Leu Glu Leu Ile Gly Cys
    370                 375                 380

Gly Gly Pro Ser Lys Glu Ser Pro Gly Thr Asn Ala Gln Leu Leu Ser
385                 390                 395                 400

Arg Val Val Ala Ala Gly Val Leu Ser Ala Glu Leu Ser Leu Met Ser
                405                 410                 415

Gly Leu Ala Ala Gly His Leu Leu Ser Ala His Met Arg Leu Asn Arg
```

420 425 430

Lys Lys Lys
435

<210> 115
<211> 1281
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 115
atgcaatccc tggacaaaaa ctttagacac ttatcaagac aacagaagtt acaacagcta 60

gttgataaac aatggctatc agaggaacaa ttcaatattc tacttaacca cccacttatt 120

gatgaagagg tagcaaactc attgatagaa aatgtcatcg cacagggcgc actgcctgtt 180

ggtttactac caaatatcat cgttgatgac aaagcatacg tcgtgcctat gatggtggaa 240

gagccatctg ttgttgccgc tgcttcatac ggcgctaaat tggtgaacca aacaggtggt 300

ttcaaaaccg tgtcctcaga acgtatcatg ataggtcaaa tagtatttga tggagtcgat 360

gataccgaga aactgtctgc agatatcaag gctcttgaaa aacaaatcca tcagattgca 420

gatgaggctt acccttctat taaggccaga ggtggaggct atcaaaggat cgccatcgat 480

acattcccag aacaacagtt gctttcattg aaggttttcg ttgatactaa ggatgctatg 540

ggcgctaata tgttaaacac aatcctagaa gcaatcacag ccttttgaa aaacgaattc 600

ccacaatctg atatcttgat gtctatcctt tccaaccacg caacagccag tgttgtcaag 660

gtccagggtg aaatagacgt taaggatttg gcaagaggag aacgtactgg agaagaggtc 720

gctaagagaa tggaaagagc atctgtgtta gctcaagtgg acattcatag agcagcaaca 780

cacaataagg gtgttatgaa tggcattcat gctgtagtct tggctacagg taatgatact 840

agaggtgcag aagcctctgc tcacgcttac gcttccaaag acggtcaata tagagggata 900

gctacatgga gatacgatca agagagacaa aggttaatag gaactataga agttccaatg 960

actctggcca ttgttggtgg cggtaccaag gtactgccta ttgctaaggc ctctttagaa 1020

ctgttaaacg tagaaagtgc ccaagagttg ggacatgttg tcgctgccgt tggactagct 1080

caaaacttcg ctgcatgtag agctttggtt tccgaaggta ttcaacaagg gcatatgtct 1140

ttgcaataca agtctttagc catcgtagtc ggggctaagg gcgatgaaat tgctcaggta 1200

gccgaagcac taaagcaaga gccaagagca aacactcaag ttgcagagag aattttgcaa 1260

gatttgagaa gtcaacaata a 1281

<210> 116

<211> 426
<212> PRT
<213> Staphylococcus aureus

<400> 116

```
Met Gln Ser Leu Asp Lys Asn Phe Arg His Leu Ser Arg Gln Gln Lys
1               5                  10                  15

Leu Gln Gln Leu Val Asp Lys Gln Trp Leu Ser Glu Glu Gln Phe Asn
            20                  25                  30

Ile Leu Leu Asn His Pro Leu Ile Asp Glu Glu Val Ala Asn Ser Leu
        35                  40                  45

Ile Glu Asn Val Ile Ala Gln Gly Ala Leu Pro Val Gly Leu Leu Pro
    50                  55                  60

Asn Ile Ile Val Asp Asp Lys Ala Tyr Val Val Pro Met Met Val Glu
65                  70                  75                  80

Glu Pro Ser Val Val Ala Ala Ala Ser Tyr Gly Ala Lys Leu Val Asn
                85                  90                  95

Gln Thr Gly Gly Phe Lys Thr Val Ser Ser Glu Arg Ile Met Ile Gly
            100                 105                 110

Gln Ile Val Phe Asp Gly Val Asp Asp Thr Glu Lys Leu Ser Ala Asp
        115                 120                 125

Ile Lys Ala Leu Glu Lys Gln Ile His Gln Ile Ala Asp Glu Ala Tyr
    130                 135                 140

Pro Ser Ile Lys Ala Arg Gly Gly Gly Tyr Gln Arg Ile Ala Ile Asp
145                 150                 155                 160

Thr Phe Pro Glu Gln Gln Leu Leu Ser Leu Lys Val Phe Val Asp Thr
                165                 170                 175

Lys Asp Ala Met Gly Ala Asn Met Leu Asn Thr Ile Leu Glu Ala Ile
            180                 185                 190

Thr Ala Phe Leu Lys Asn Glu Phe Pro Gln Ser Asp Ile Leu Met Ser
        195                 200                 205

Ile Leu Ser Asn His Ala Thr Ala Ser Val Val Lys Val Gln Gly Glu
    210                 215                 220

Ile Asp Val Lys Asp Leu Ala Arg Gly Glu Arg Thr Gly Glu Glu Val
225                 230                 235                 240
```

Ala Lys Arg Met Glu Arg Ala Ser Val Leu Ala Gln Val Asp Ile His
245 250 255

Arg Ala Ala Thr His Asn Lys Gly Val Met Asn Gly Ile His Ala Val
260 265 270

Val Leu Ala Thr Gly Asn Asp Thr Arg Gly Ala Glu Ala Ser Ala His
275 280 285

Ala Tyr Ala Ser Lys Asp Gly Gln Tyr Arg Gly Ile Ala Thr Trp Arg
290 295 300

Tyr Asp Gln Glu Arg Gln Arg Leu Ile Gly Thr Ile Glu Val Pro Met
305 310 315 320

Thr Leu Ala Ile Val Gly Gly Gly Thr Lys Val Leu Pro Ile Ala Lys
325 330 335

Ala Ser Leu Glu Leu Leu Asn Val Glu Ser Ala Gln Glu Leu Gly His
340 345 350

Val Val Ala Ala Val Gly Leu Ala Gln Asn Phe Ala Ala Cys Arg Ala
355 360 365

Leu Val Ser Glu Gly Ile Gln Gln Gly His Met Ser Leu Gln Tyr Lys
370 375 380

Ser Leu Ala Ile Val Val Gly Ala Lys Gly Asp Glu Ile Ala Gln Val
385 390 395 400

Ala Glu Ala Leu Lys Gln Glu Pro Arg Ala Asn Thr Gln Val Ala Glu
405 410 415

Arg Ile Leu Gln Asp Leu Arg Ser Gln Gln
420 425

<210> 117
<211> 1311
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 117
atgcaggtct taagattgga taggagacat tacaaaagtg gcaagattag aagagcaatg 60

agttctagaa ttcctggttt ctacaaattg tcagtcgagg aaagactgaa aaaggttgct 120

gaatttgcag ggttatctga tgaggaagtg aaagctgttt tgtcacaagg tttacctttg 180

gacgtagctg atagaatgat cgaaaatgtg atcggtacat ttgaattacc acttggtata 240

gcaaccaatt tccttattga tggcaaggat tatctaatcc ctatggctat agaggaacca 300

tcagtagttg cagctgcttc taacgcagct agaatggcca gagagtctgg cgggtttaca 360

actgattaca cagggtccct gatgattggt caaattcaag tcacaaaact gttgaatcca 420

aatgcagcta agttcgaagt tctacgtcaa aaagacgaaa tcatagaaag agcaaatgag 480

tgtgatccaa tgttggtgaa tttgggcggt ggatgtaaag atatagaagc aagggtgatc 540

gatacaatca tgggtaagat gctaattgtt catctgatcg ttgatgttaa agacgctatg 600

ggtgcaaatg ctgtcaacac tatgtgtgaa aaagttgctc ctttcatcga acgtattact 660

gggggaaagg tctatcttag aatcatttcc aacttggctg catatagact tgctagagca 720

aaggccgttt ttgacaaaga cgttattggc ggagaggagg ttgtagaagg gatcatgctt 780

gcatacgcct tcgctgccgc tgacccattt cgttgcgcca cccacaataa gggtatcatg 840

aatggcatat cagccttaat gatcgctaca ggaaacgact ttagagccat tgaagcagga 900

gctcattcct atgctgcaat aggtggatac aaaccactaa ctacctacga agttgataga 960

aaaggtaatc tagtaggcac aattgaaata cctatggcag taggcgtgat tggtggtgca 1020

accaaagtca acccactagc caagatctct cttaagatac taggagtgaa cactgctgaa 1080

gagttagcca gagtcgcagc cgctctaggt ttggctcaaa actttgctgc cttaagagcc 1140

ttggccacag aaggtatcca aagaggtcac atggaattac atgccaggaa cttagcaatc 1200

atggctggag ctactggaga tgaggttgac agagttgtag agattatggt gagagatggc 1260

aaaatcagat tggactacgc taaggaagta ttggagagac tgcgttccta a 1311

<210> 118
<211> 436
<212> PRT
<213> Archaeoglobus fulgidus

<400> 118
Met Gln Val Leu Arg Leu Asp Arg Arg His Tyr Lys Ser Gly Lys Ile
1 5 10 15

Arg Arg Ala Met Ser Ser Arg Ile Pro Gly Phe Tyr Lys Leu Ser Val
20 25 30

Glu Glu Arg Leu Lys Lys Val Ala Glu Phe Ala Gly Leu Ser Asp Glu
35 40 45

Glu Val Lys Ala Val Leu Ser Gln Gly Leu Pro Leu Asp Val Ala Asp
50 55 60

```
Arg Met Ile Glu Asn Val Ile Gly Thr Phe Glu Leu Pro Leu Gly Ile
65                  70                  75                  80

Ala Thr Asn Phe Leu Ile Asp Gly Lys Asp Tyr Leu Ile Pro Met Ala
                85                  90                  95

Ile Glu Glu Pro Ser Val Val Ala Ala Ala Ser Asn Ala Ala Arg Met
            100                 105                 110

Ala Arg Glu Ser Gly Gly Phe Thr Thr Asp Tyr Thr Gly Ser Leu Met
        115                 120                 125

Ile Gly Gln Ile Gln Val Thr Lys Leu Leu Asn Pro Asn Ala Ala Lys
    130                 135                 140

Phe Glu Val Leu Arg Gln Lys Asp Glu Ile Ile Glu Arg Ala Asn Glu
145                 150                 155                 160

Cys Asp Pro Met Leu Val Asn Leu Gly Gly Gly Cys Lys Asp Ile Glu
                165                 170                 175

Ala Arg Val Ile Asp Thr Ile Met Gly Lys Met Leu Ile Val His Leu
            180                 185                 190

Ile Val Asp Val Lys Asp Ala Met Gly Ala Asn Ala Val Asn Thr Met
        195                 200                 205

Cys Glu Lys Val Ala Pro Phe Ile Glu Arg Ile Thr Gly Gly Lys Val
    210                 215                 220

Tyr Leu Arg Ile Ile Ser Asn Leu Ala Ala Tyr Arg Leu Ala Arg Ala
225                 230                 235                 240

Lys Ala Val Phe Asp Lys Asp Val Ile Gly Gly Glu Glu Val Val Glu
                245                 250                 255

Gly Ile Met Leu Ala Tyr Ala Phe Ala Ala Ala Asp Pro Phe Arg Cys
            260                 265                 270

Ala Thr His Asn Lys Gly Ile Met Asn Gly Ile Ser Ala Leu Met Ile
        275                 280                 285

Ala Thr Gly Asn Asp Phe Arg Ala Ile Glu Ala Gly Ala His Ser Tyr
    290                 295                 300

Ala Ala Ile Gly Gly Tyr Lys Pro Leu Thr Thr Tyr Glu Val Asp Arg
305                 310                 315                 320
```

Lys Gly Asn Leu Val Gly Thr Ile Glu Ile Pro Met Ala Val Gly Val
325 330 335

Ile Gly Gly Ala Thr Lys Val Asn Pro Leu Ala Lys Ile Ser Leu Lys
340 345 350

Ile Leu Gly Val Asn Thr Ala Glu Glu Leu Ala Arg Val Ala Ala Ala
355 360 365

Leu Gly Leu Ala Gln Asn Phe Ala Ala Leu Arg Ala Leu Ala Thr Glu
370 375 380

Gly Ile Gln Arg Gly His Met Glu Leu His Ala Arg Asn Leu Ala Ile
385 390 395 400

Met Ala Gly Ala Thr Gly Asp Glu Val Asp Arg Val Val Glu Ile Met
405 410 415

Val Arg Asp Gly Lys Ile Arg Leu Asp Tyr Ala Lys Glu Val Leu Glu
420 425 430

Arg Leu Arg Ser
435

<210> 119
<211> 1287
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 119
atgtccttag attcaagact gccagctttc agaaatctgt ctccagctgc aagactagat 60

cacattggcc aacttttggg actaagtcat gacgacgttt ccctttagc aaacgccggt 120

gctttaccaa tggatatcgc taatggtatg attgaaaatg taatcgggac ctttgaactg 180

ccatatgcag tggccagtaa ctttcagatc aatggccgtg acgtcttagt accattagtt 240

gtggaggaac ctagtatcgt tgctgcagcc tcttacatgg caaagttagc tagagccaat 300

ggtgggttca ctacatcttc atctgctcca ctaatgcatg cacaagtaca aattgtcggc 360

attcaggatc cactaaacgc aagattgtct ttactgcgta gaaaggatga gatcatagaa 420

ttagccaata ggaaggacca acttctgaat tcattgggcg gtggttgcag agacatagag 480

gtgcatacat ttgccgatac tccaagagga ccaatgcttg tagcacacct tattgtcgat 540

gtgcgtgatg ccatgggagc taatactgtt aacactatgg ctgaagcagt agcacctctg 600

atggaagcca taacaggtgg ccaggtaaga ttgagaatcc tttccaattt ggctgatctt 660

agattggcca gagcccaagt gagaatcact cctcagcaat tggaaactgc cgaattctca 720

ggtgaggcag taattgaggg tatcttggac gcatatgctt ttgccgctgt ggacccttac 780

agagccgcta cccacaacaa aggcataatg aacggtatcg atcctttgat cgtcgctaca 840

ggaaatgatt ggagagctgt tgaggcagga gctcatgcat acgcttgtag atccggacat 900

tacggttcat taacaacatg ggaaaaagat aacaatggac acttggtcgg gacattggaa 960

atgcctatgc cagttggttt agttgggggt gctacaaaaa cccatcctct tgctcaattg 1020

tctttgagga tacttggtgt caaaactgct caagcactag ccgaaattgc cgttgctgtt 1080

ggtttggcac aaaacttggg tgcaatgcgt gctttagcta cagaaggcat ccaaagagga 1140

catatggctc tacacgctag aaacattgca gttgttgcag gagccagagg tgatgaggtt 1200

gattgggtgg ctagacaact tgtcgaatat catgatgtca gagcagacag ggctgtggca 1260

ttactgaaac agaagagagg tcaataa 1287

<210> 120
<211> 428
<212> PRT
<213> Pseudomonas mevalonii

<400> 120
Met Ser Leu Asp Ser Arg Leu Pro Ala Phe Arg Asn Leu Ser Pro Ala
1 5 10 15

Ala Arg Leu Asp His Ile Gly Gln Leu Leu Gly Leu Ser His Asp Asp
20 25 30

Val Ser Leu Leu Ala Asn Ala Gly Ala Leu Pro Met Asp Ile Ala Asn
35 40 45

Gly Met Ile Glu Asn Val Ile Gly Thr Phe Glu Leu Pro Tyr Ala Val
50 55 60

Ala Ser Asn Phe Gln Ile Asn Gly Arg Asp Val Leu Val Pro Leu Val
65 70 75 80

Val Glu Glu Pro Ser Ile Val Ala Ala Ala Ser Tyr Met Ala Lys Leu
85 90 95

Ala Arg Ala Asn Gly Gly Phe Thr Thr Ser Ser Ser Ala Pro Leu Met
100 105 110

His Ala Gln Val Gln Ile Val Gly Ile Gln Asp Pro Leu Asn Ala Arg
115 120 125

```
Leu Ser Leu Leu Arg Arg Lys Asp Glu Ile Ile Glu Leu Ala Asn Arg
    130                 135                 140

Lys Asp Gln Leu Leu Asn Ser Leu Gly Gly Gly Cys Arg Asp Ile Glu
145                 150                 155                 160

Val His Thr Phe Ala Asp Thr Pro Arg Gly Pro Met Leu Val Ala His
                165                 170                 175

Leu Ile Val Asp Val Arg Asp Ala Met Gly Ala Asn Thr Val Asn Thr
            180                 185                 190

Met Ala Glu Ala Val Ala Pro Leu Met Glu Ala Ile Thr Gly Gly Gln
        195                 200                 205

Val Arg Leu Arg Ile Leu Ser Asn Leu Ala Asp Leu Arg Leu Ala Arg
    210                 215                 220

Ala Gln Val Arg Ile Thr Pro Gln Gln Leu Glu Thr Ala Glu Phe Ser
225                 230                 235                 240

Gly Glu Ala Val Ile Glu Gly Ile Leu Asp Ala Tyr Ala Phe Ala Ala
                245                 250                 255

Val Asp Pro Tyr Arg Ala Ala Thr His Asn Lys Gly Ile Met Asn Gly
            260                 265                 270

Ile Asp Pro Leu Ile Val Ala Thr Gly Asn Asp Trp Arg Ala Val Glu
        275                 280                 285

Ala Gly Ala His Ala Tyr Ala Cys Arg Ser Gly His Tyr Gly Ser Leu
    290                 295                 300

Thr Thr Trp Glu Lys Asp Asn Asn Gly His Leu Val Gly Thr Leu Glu
305                 310                 315                 320

Met Pro Met Pro Val Gly Leu Val Gly Gly Ala Thr Lys Thr His Pro
                325                 330                 335

Leu Ala Gln Leu Ser Leu Arg Ile Leu Gly Val Lys Thr Ala Gln Ala
            340                 345                 350

Leu Ala Glu Ile Ala Val Ala Val Gly Leu Ala Gln Asn Leu Gly Ala
        355                 360                 365

Met Arg Ala Leu Ala Thr Glu Gly Ile Gln Arg Gly His Met Ala Leu
    370                 375                 380
```

His Ala Arg Asn Ile Ala Val Val Ala Gly Ala Arg Gly Asp Glu Val
385 390 395 400

Asp Trp Val Ala Arg Gln Leu Val Glu Tyr His Asp Val Arg Ala Asp
405 410 415

Arg Ala Val Ala Leu Leu Lys Gln Lys Arg Gly Gln
420 425

<210> 121
<211> 361
<212> PRT
<213> Stevia rebaudiana

<400> 121
Met Ala Leu Val Asn Pro Thr Ala Leu Phe Tyr Gly Thr Ser Ile Arg
1 5 10 15

Thr Arg Pro Thr Asn Leu Leu Asn Pro Thr Gln Lys Leu Arg Pro Val
20 25 30

Ser Ser Ser Ser Leu Pro Ser Phe Ser Ser Val Ser Ala Ile Leu Thr
35 40 45

Glu Lys His Gln Ser Asn Pro Ser Glu Asn Asn Asn Leu Gln Thr His
50 55 60

Leu Glu Thr Pro Phe Asn Phe Asp Ser Tyr Met Leu Glu Lys Val Asn
65 70 75 80

Met Val Asn Glu Ala Leu Asp Ala Ser Val Pro Leu Lys Asp Pro Ile
85 90 95

Lys Ile His Glu Ser Met Arg Tyr Ser Leu Leu Ala Gly Gly Lys Arg
100 105 110

Ile Arg Pro Met Met Cys Ile Ala Ala Cys Glu Ile Val Gly Gly Asn
115 120 125

Ile Leu Asn Ala Met Pro Ala Ala Cys Ala Val Glu Met Ile His Thr
130 135 140

Met Ser Leu Val His Asp Asp Leu Pro Cys Met Asp Asn Asp Asp Phe
145 150 155 160

Arg Arg Gly Lys Pro Ile Ser His Lys Val Tyr Gly Glu Glu Met Ala
165 170 175

```
Val Leu Thr Gly Asp Ala Leu Leu Ser Leu Ser Phe Glu His Ile Ala
            180                 185                 190

Thr Ala Thr Lys Gly Val Ser Lys Asp Arg Ile Val Arg Ala Ile Gly
        195                 200                 205

Glu Leu Ala Arg Ser Val Gly Ser Glu Gly Leu Val Ala Gly Gln Val
    210                 215                 220

Val Asp Ile Leu Ser Glu Gly Ala Asp Val Gly Leu Asp His Leu Glu
225                 230                 235                 240

Tyr Ile His Ile His Lys Thr Ala Met Leu Leu Glu Ser Ser Val Val
                245                 250                 255

Ile Gly Ala Ile Met Gly Gly Gly Ser Asp Gln Gln Ile Glu Lys Leu
            260                 265                 270

Arg Lys Phe Ala Arg Ser Ile Gly Leu Leu Phe Gln Val Val Asp Asp
        275                 280                 285

Ile Leu Asp Val Thr Lys Ser Thr Glu Glu Leu Gly Lys Thr Ala Gly
    290                 295                 300

Lys Asp Leu Leu Thr Asp Lys Thr Thr Tyr Pro Lys Leu Leu Gly Ile
305                 310                 315                 320

Glu Lys Ser Arg Glu Phe Ala Glu Lys Leu Asn Lys Glu Ala Gln Glu
                325                 330                 335

Gln Leu Ser Gly Phe Asp Arg Arg Lys Ala Ala Pro Leu Ile Ala Leu
            340                 345                 350

Ala Asn Tyr Asn Ala Tyr Arg Gln Asn
        355                 360


<210> 122
<211> 342
<212> PRT
<213> Gibberella fujikuroi

<400> 122
Met Ala Glu Gln Gln Ile Ser Asn Leu Leu Ser Met Phe Asp Ala Ser
1               5                   10                  15

His Ala Ser Gln Lys Leu Glu Ile Thr Val Gln Met Met Asp Thr Tyr
            20                  25                  30
```

```
His Tyr Arg Glu Thr Pro Pro Asp Ser Ser Ser Ser Glu Gly Gly Ser
        35                  40                  45

Leu Ser Arg Tyr Asp Glu Arg Arg Val Ser Leu Pro Leu Ser His Asn
    50                  55                  60

Ala Ala Ser Pro Asp Ile Val Ser Gln Leu Cys Phe Ser Thr Ala Met
65                  70                  75                  80

Ser Ser Glu Leu Asn His Arg Trp Lys Ser Gln Arg Leu Lys Val Ala
                85                  90                  95

Asp Ser Pro Tyr Asn Tyr Ile Leu Thr Leu Pro Ser Lys Gly Ile Arg
            100                 105                 110

Gly Ala Phe Ile Asp Ser Leu Asn Val Trp Leu Glu Val Pro Glu Asp
        115                 120                 125

Glu Thr Ser Val Ile Lys Glu Val Ile Gly Met Leu His Asn Ser Ser
    130                 135                 140

Leu Ile Ile Asp Asp Phe Gln Asp Asn Ser Pro Leu Arg Arg Gly Lys
145                 150                 155                 160

Pro Ser Thr His Thr Val Phe Gly Pro Ala Gln Ala Ile Asn Thr Ala
                165                 170                 175

Thr Tyr Val Ile Val Lys Ala Ile Glu Lys Ile Gln Asp Ile Val Gly
            180                 185                 190

His Asp Ala Leu Ala Asp Val Thr Gly Thr Ile Thr Thr Ile Phe Gln
        195                 200                 205

Gly Gln Ala Met Asp Leu Trp Trp Thr Ala Asn Ala Ile Val Pro Ser
    210                 215                 220

Ile Gln Glu Tyr Leu Leu Met Val Asn Asp Lys Thr Gly Ala Leu Phe
225                 230                 235                 240

Arg Leu Ser Leu Glu Leu Leu Ala Leu Asn Ser Glu Ala Ser Ile Ser
                245                 250                 255

Asp Ser Ala Leu Glu Ser Leu Ser Ser Ala Val Ser Leu Leu Gly Gln
            260                 265                 270

Tyr Phe Gln Ile Arg Asp Asp Tyr Met Asn Leu Ile Asp Asn Lys Tyr
        275                 280                 285
```

Thr Asp Gln Lys Gly Phe Cys Glu Asp Leu Asp Glu Gly Lys Tyr Ser
290 295 300

Leu Thr Leu Ile His Ala Leu Gln Thr Asp Ser Ser Asp Leu Leu Thr
305 310 315 320

Asn Ile Leu Ser Met Arg Arg Val Gln Gly Lys Leu Thr Ala Gln Lys
325 330 335

Arg Cys Trp Phe Trp Lys
340

<210> 123
<211> 300
<212> PRT
<213> Mus musculus

<400> 123
Met Glu Lys Thr Lys Glu Lys Ala Glu Arg Ile Leu Leu Glu Pro Tyr
1 5 10 15

Arg Tyr Leu Leu Gln Leu Pro Gly Lys Gln Val Arg Ser Lys Leu Ser
20 25 30

Gln Ala Phe Asn His Trp Leu Lys Val Pro Glu Asp Lys Leu Gln Ile
35 40 45

Ile Ile Glu Val Thr Glu Met Leu His Asn Ala Ser Leu Leu Ile Asp
50 55 60

Asp Ile Glu Asp Ser Ser Lys Leu Arg Arg Gly Phe Pro Val Ala His
65 70 75 80

Ser Ile Tyr Gly Val Pro Ser Val Ile Asn Ser Ala Asn Tyr Val Tyr
85 90 95

Phe Leu Gly Leu Glu Lys Val Leu Thr Leu Asp His Pro Asp Ala Val
100 105 110

Lys Leu Phe Thr Arg Gln Leu Leu Glu Leu His Gln Gly Gln Gly Leu
115 120 125

Asp Ile Tyr Trp Arg Asp Thr Tyr Thr Cys Pro Thr Glu Glu Glu Tyr
130 135 140

Lys Ala Met Val Leu Gln Lys Thr Gly Gly Leu Phe Gly Leu Ala Val
145 150 155 160

Gly Leu Met Gln Leu Phe Ser Asp Tyr Lys Glu Asp Leu Lys Pro Leu
165 170 175

Leu Asp Thr Leu Gly Leu Phe Phe Gln Ile Arg Asp Asp Tyr Ala Asn
180 185 190

Leu His Ser Lys Glu Tyr Ser Glu Asn Lys Ser Phe Cys Glu Asp Leu
195 200 205

Thr Glu Gly Lys Phe Ser Phe Pro Thr Ile His Ala Ile Trp Ser Arg
210 215 220

Pro Glu Ser Thr Gln Val Gln Asn Ile Leu Arg Gln Arg Thr Glu Asn
225 230 235 240

Ile Asp Ile Lys Lys Tyr Cys Val Gln Tyr Leu Glu Asp Val Gly Ser
245 250 255

Phe Ala Tyr Thr Arg His Thr Leu Arg Glu Leu Glu Ala Lys Ala Tyr
260 265 270

Lys Gln Ile Glu Ala Cys Gly Gly Asn Pro Ser Leu Val Ala Leu Val
275 280 285

Lys His Leu Ser Lys Met Phe Thr Glu Glu Asn Lys
290 295 300

<210> 124
<211> 339
<212> PRT
<213> Thalassiosira pseudonana

<400> 124
Met Ala Arg Phe Tyr Phe Leu Asn Ala Leu Leu Met Val Ile Ser Leu
1 5 10 15

Gln Ser Thr Thr Ala Phe Thr Pro Ala Lys Leu Ala Tyr Pro Thr Thr
20 25 30

Thr Thr Ala Leu Asn Val Ala Ser Ala Glu Thr Ser Phe Ser Leu Asp
35 40 45

Glu Tyr Leu Ala Ser Lys Ile Gly Pro Ile Glu Ser Ala Leu Glu Ala
50 55 60

Ser Val Lys Ser Arg Ile Pro Gln Thr Asp Lys Ile Cys Glu Ser Met
65 70 75 80

Ala Tyr Ser Leu Met Ala Gly Gly Lys Arg Ile Arg Pro Val Leu Cys

```
                85                  90                  95

Ile Ala Ala Cys Glu Met Phe Gly Gly Ser Gln Asp Val Ala Met Pro
            100                 105                 110

Thr Ala Val Ala Leu Glu Met Ile His Thr Met Ser Leu Ile His Asp
        115                 120                 125

Asp Leu Pro Ser Met Asp Asn Asp Asp Leu Arg Arg Gly Lys Pro Thr
    130                 135                 140

Asn His Val Val Phe Gly Glu Asp Val Ala Ile Leu Ala Gly Asp Ser
145                 150                 155                 160

Leu Leu Ser Thr Ser Phe Glu His Val Ala Arg Glu Thr Lys Gly Val
                165                 170                 175

Ser Ala Glu Lys Ile Val Asp Val Ile Ala Arg Leu Gly Lys Ser Val
            180                 185                 190

Gly Ala Glu Gly Leu Ala Gly Gly Gln Val Met Asp Leu Glu Cys Glu
        195                 200                 205

Ala Lys Pro Gly Thr Thr Leu Asp Asp Leu Lys Trp Ile His Ile His
    210                 215                 220

Lys Thr Ala Thr Leu Leu Gln Val Ala Val Ala Ser Gly Ala Val Leu
225                 230                 235                 240

Gly Gly Ala Thr Pro Glu Glu Val Ala Ala Cys Glu Leu Phe Ala Met
                245                 250                 255

Asn Ile Gly Leu Ala Phe Gln Val Ala Asp Asp Ile Leu Asp Val Thr
            260                 265                 270

Ala Ser Ser Glu Asp Leu Gly Lys Thr Ala Gly Lys Asp Glu Ala Thr
        275                 280                 285

Asp Lys Thr Thr Tyr Pro Lys Leu Leu Gly Leu Glu Glu Ser Lys Ala
    290                 295                 300

Tyr Ala Arg Gln Leu Ile Asp Glu Ala Lys Glu Ser Leu Ala Pro Phe
305                 310                 315                 320

Gly Asp Arg Ala Ala Pro Leu Leu Ala Ile Ala Asp Phe Ile Ile Asp
                325                 330                 335
```

Arg Lys Asn

<210> 125
<211> 355
<212> PRT
<213> Streptomyces clavuligerus

<400> 125
Met His Leu Ala Pro Arg Arg Val Pro Arg Gly Arg Arg Ser Pro Pro
1 5 10 15

Asp Arg Val Pro Glu Arg Gln Gly Ala Leu Gly Arg Arg Arg Gly Ala
20 25 30

Gly Ser Thr Gly Cys Ala Arg Ala Ala Ala Gly Val His Arg Arg Arg
35 40 45

Gly Gly Gly Glu Ala Asp Pro Ser Ala Ala Val His Arg Gly Trp Gln
50 55 60

Ala Gly Gly Gly Thr Gly Leu Pro Asp Glu Val Val Ser Thr Ala Ala
65 70 75 80

Ala Leu Glu Met Phe His Ala Phe Ala Leu Ile His Asp Asp Ile Met
85 90 95

Asp Asp Ser Ala Thr Arg Arg Gly Ser Pro Thr Val His Arg Ala Leu
100 105 110

Ala Asp Arg Leu Gly Ala Ala Leu Asp Pro Asp Gln Ala Gly Gln Leu
115 120 125

Gly Val Ser Thr Ala Ile Leu Val Gly Asp Leu Ala Leu Thr Trp Ser
130 135 140

Asp Glu Leu Leu Tyr Ala Pro Leu Thr Pro His Arg Leu Ala Ala Val
145 150 155 160

Leu Pro Leu Val Thr Ala Met Arg Ala Glu Thr Val His Gly Gln Tyr
165 170 175

Leu Asp Ile Thr Ser Ala Arg Arg Pro Gly Thr Asp Thr Ser Leu Ala
180 185 190

Leu Arg Ile Ala Arg Tyr Lys Thr Ala Ala Tyr Thr Met Glu Arg Pro
195 200 205

Leu His Ile Gly Ala Ala Leu Ala Gly Ala Arg Pro Glu Leu Leu Ala

210 215 220

Gly Leu Ser Ala Tyr Ala Leu Pro Ala Gly Glu Ala Phe Gln Leu Ala
225 230 235 240

Asp Asp Leu Leu Gly Val Phe Gly Asp Pro Arg Arg Thr Gly Lys Pro
245 250 255

Asp Leu Asp Asp Leu Arg Gly Gly Lys His Thr Val Leu Val Ala Leu
260 265 270

Ala Arg Glu His Ala Thr Pro Glu Gln Arg His Thr Leu Asp Thr Leu
275 280 285

Leu Gly Thr Pro Gly Leu Asp Arg Gln Gly Ala Ser Arg Leu Arg Cys
290 295 300

Val Leu Val Ala Thr Gly Ala Arg Ala Glu Ala Glu Arg Leu Ile Thr
305 310 315 320

Glu Arg Arg Asp Gln Ala Leu Thr Ala Leu Asn Ala Leu Thr Leu Pro
325 330 335

Pro Pro Leu Ala Glu Ala Leu Ala Arg Leu Thr Leu Gly Ser Thr Ala
340 345 350

His Pro Ala
355

<210> 126
<211> 330
<212> PRT
<213> Sulfulobus acidicaldarius

<400> 126
Met Ser Tyr Phe Asp Asn Tyr Phe Asn Glu Ile Val Asn Ser Val Asn
1 5 10 15

Asp Ile Ile Lys Ser Tyr Ile Ser Gly Asp Val Pro Lys Leu Tyr Glu
20 25 30

Ala Ser Tyr His Leu Phe Thr Ser Gly Gly Lys Arg Leu Arg Pro Leu
35 40 45

Ile Leu Thr Ile Ser Ser Asp Leu Phe Gly Gly Gln Arg Glu Arg Ala
50 55 60

Tyr Tyr Ala Gly Ala Ala Ile Glu Val Leu His Thr Phe Thr Leu Val
65 70 75 80

```
His Asp Asp Ile Met Asp Gln Asp Asn Ile Arg Arg Gly Leu Pro Thr
                85                  90                  95

Val His Val Lys Tyr Gly Leu Pro Leu Ala Ile Leu Ala Gly Asp Leu
            100                 105                 110

Leu His Ala Lys Ala Phe Gln Leu Leu Thr Gln Ala Leu Arg Gly Leu
        115                 120                 125

Pro Ser Glu Thr Ile Ile Lys Ala Phe Asp Ile Phe Thr Arg Ser Ile
    130                 135                 140

Ile Ile Ile Ser Glu Gly Gln Ala Val Asp Met Glu Phe Glu Asp Arg
145                 150                 155                 160

Ile Asp Ile Lys Glu Gln Glu Tyr Leu Asp Met Ile Ser Arg Lys Thr
                165                 170                 175

Ala Ala Leu Phe Ser Ala Ser Ser Ser Ile Gly Ala Leu Ile Ala Gly
            180                 185                 190

Ala Asn Asp Asn Asp Val Arg Leu Met Ser Asp Phe Gly Thr Asn Leu
        195                 200                 205

Gly Ile Ala Phe Gln Ile Val Asp Asp Ile Leu Gly Leu Thr Ala Asp
    210                 215                 220

Glu Lys Glu Leu Gly Lys Pro Val Phe Ser Asp Ile Arg Glu Gly Lys
225                 230                 235                 240

Lys Thr Ile Leu Val Ile Lys Thr Leu Glu Leu Cys Lys Glu Asp Glu
                245                 250                 255

Lys Lys Ile Val Leu Lys Ala Leu Gly Asn Lys Ser Ala Ser Lys Glu
            260                 265                 270

Glu Leu Met Ser Ser Ala Asp Ile Ile Lys Lys Tyr Ser Leu Asp Tyr
        275                 280                 285

Ala Tyr Asn Leu Ala Glu Lys Tyr Tyr Lys Asn Ala Ile Asp Ser Leu
    290                 295                 300

Asn Gln Val Ser Ser Lys Ser Asp Ile Pro Gly Lys Ala Leu Lys Tyr
305                 310                 315                 320

Leu Ala Glu Phe Thr Ile Arg Arg Arg Lys
```

325 330

<210> 127
<211> 297
<212> PRT
<213> Synechococcus sp.

<400> 127
Met Val Ala Gln Thr Phe Asn Leu Asp Thr Tyr Leu Ser Gln Arg Gln
1 5 10 15

Gln Gln Val Glu Glu Ala Leu Ser Ala Ala Leu Val Pro Ala Tyr Pro
20 25 30

Glu Arg Ile Tyr Glu Ala Met Arg Tyr Ser Leu Leu Ala Gly Gly Lys
35 40 45

Arg Leu Arg Pro Ile Leu Cys Leu Ala Ala Cys Glu Leu Ala Gly Gly
50 55 60

Ser Val Glu Gln Ala Met Pro Thr Ala Cys Ala Leu Glu Met Ile His
65 70 75 80

Thr Met Ser Leu Ile His Asp Asp Leu Pro Ala Met Asp Asn Asp Asp
85 90 95

Phe Arg Arg Gly Lys Pro Thr Asn His Lys Val Phe Gly Glu Asp Ile
100 105 110

Ala Ile Leu Ala Gly Asp Ala Leu Leu Ala Tyr Ala Phe Glu His Ile
115 120 125

Ala Ser Gln Thr Arg Gly Val Pro Pro Gln Leu Val Leu Gln Val Ile
130 135 140

Ala Arg Ile Gly His Ala Val Ala Ala Thr Gly Leu Val Gly Gly Gln
145 150 155 160

Val Val Asp Leu Glu Ser Glu Gly Lys Ala Ile Ser Leu Glu Thr Leu
165 170 175

Glu Tyr Ile His Ser His Lys Thr Gly Ala Leu Leu Glu Ala Ser Val
180 185 190

Val Ser Gly Gly Ile Leu Ala Gly Ala Asp Glu Glu Leu Leu Ala Arg
195 200 205

Leu Ser His Tyr Ala Arg Asp Ile Gly Leu Ala Phe Gln Ile Val Asp
210 215 220

Asp Ile Leu Asp Val Thr Ala Thr Ser Glu Gln Leu Gly Lys Thr Ala
225 230 235 240

Gly Lys Asp Gln Ala Ala Ala Lys Ala Thr Tyr Pro Ser Leu Leu Gly
245 250 255

Leu Glu Ala Ser Arg Gln Lys Ala Glu Glu Leu Ile Gln Ser Ala Lys
260 265 270

Glu Ala Leu Arg Pro Tyr Gly Ser Gln Ala Glu Pro Leu Leu Ala Leu
275 280 285

Ala Asp Phe Ile Thr Arg Arg Gln His
290 295

<210> 128
<211> 371
<212> PRT
<213> Arabidopsis thaliana

<400> 128
Met Ala Ser Val Thr Leu Gly Ser Trp Ile Val Val His His His Asn
1 5 10 15

His His His Pro Ser Ser Ile Leu Thr Lys Ser Arg Ser Arg Ser Cys
20 25 30

Pro Ile Thr Leu Thr Lys Pro Ile Ser Phe Arg Ser Lys Arg Thr Val
35 40 45

Ser Ser Ser Ser Ser Ile Val Ser Ser Ser Val Val Thr Lys Glu Asp
50 55 60

Asn Leu Arg Gln Ser Glu Pro Ser Ser Phe Asp Phe Met Ser Tyr Ile
65 70 75 80

Ile Thr Lys Ala Glu Leu Val Asn Lys Ala Leu Asp Ser Ala Val Pro
85 90 95

Leu Arg Glu Pro Leu Lys Ile His Glu Ala Met Arg Tyr Ser Leu Leu
100 105 110

Ala Gly Gly Lys Arg Val Arg Pro Val Leu Cys Ile Ala Ala Cys Glu
115 120 125

Leu Val Gly Gly Glu Glu Ser Thr Ala Met Pro Ala Ala Cys Ala Val
130 135 140

```
Glu Met Ile His Thr Met Ser Leu Ile His Asp Asp Leu Pro Cys Met
145                 150                 155                 160

Asp Asn Asp Asp Leu Arg Arg Gly Lys Pro Thr Asn His Lys Val Phe
                165                 170                 175

Gly Glu Asp Val Ala Val Leu Ala Gly Asp Ala Leu Leu Ser Phe Ala
            180                 185                 190

Phe Glu His Leu Ala Ser Ala Thr Ser Ser Asp Val Val Ser Pro Val
        195                 200                 205

Arg Val Val Arg Ala Val Gly Glu Leu Ala Lys Ala Ile Gly Thr Glu
    210                 215                 220

Gly Leu Val Ala Gly Gln Val Val Asp Ile Ser Ser Glu Gly Leu Asp
225                 230                 235                 240

Leu Asn Asp Val Gly Leu Glu His Leu Glu Phe Ile His Leu His Lys
                245                 250                 255

Thr Ala Ala Leu Leu Glu Ala Ser Ala Val Leu Gly Ala Ile Val Gly
            260                 265                 270

Gly Gly Ser Asp Asp Glu Ile Glu Arg Leu Arg Lys Phe Ala Arg Cys
        275                 280                 285

Ile Gly Leu Leu Phe Gln Val Val Asp Asp Ile Leu Asp Val Thr Lys
    290                 295                 300

Ser Ser Lys Glu Leu Gly Lys Thr Ala Gly Lys Asp Leu Ile Ala Asp
305                 310                 315                 320

Lys Leu Thr Tyr Pro Lys Ile Met Gly Leu Glu Lys Ser Arg Glu Phe
                325                 330                 335

Ala Glu Lys Leu Asn Arg Glu Ala Arg Asp Gln Leu Leu Gly Phe Asp
            340                 345                 350

Ser Asp Lys Val Ala Pro Leu Leu Ala Leu Ala Asn Tyr Ile Ala Tyr
        355                 360                 365

Arg Gln Asn
    370
```

<210> 129
<211> 787

<212> PRT
<213> Stevia rebaudiana

<400> 129

```
Met Lys Thr Gly Phe Ile Ser Pro Ala Thr Val Phe His His Arg Ile
1               5                   10                  15

Ser Pro Ala Thr Thr Phe Arg His His Leu Ser Pro Ala Thr Thr Asn
            20                  25                  30

Ser Thr Gly Ile Val Ala Leu Arg Asp Ile Asn Phe Arg Cys Lys Ala
        35                  40                  45

Val Ser Lys Glu Tyr Ser Asp Leu Leu Gln Lys Asp Glu Ala Ser Phe
    50                  55                  60

Thr Lys Trp Asp Asp Asp Lys Val Lys Asp His Leu Asp Thr Asn Lys
65                  70                  75                  80

Asn Leu Tyr Pro Asn Asp Glu Ile Lys Glu Phe Val Glu Ser Val Lys
                85                  90                  95

Ala Met Phe Gly Ser Met Asn Asp Gly Glu Ile Asn Val Ser Ala Tyr
            100                 105                 110

Asp Thr Ala Trp Val Ala Leu Val Gln Asp Val Asp Gly Ser Gly Ser
        115                 120                 125

Pro Gln Phe Pro Ser Ser Leu Glu Trp Ile Ala Asn Asn Gln Leu Ser
    130                 135                 140

Asp Gly Ser Trp Gly Asp His Leu Leu Phe Ser Ala His Asp Arg Ile
145                 150                 155                 160

Ile Asn Thr Leu Ala Cys Val Ile Ala Leu Thr Ser Trp Asn Val His
                165                 170                 175

Pro Ser Lys Cys Glu Lys Gly Leu Asn Phe Leu Arg Glu Asn Ile Cys
            180                 185                 190

Lys Leu Glu Asp Glu Asn Ala Glu His Met Pro Ile Gly Phe Glu Val
        195                 200                 205

Thr Phe Pro Ser Leu Ile Asp Ile Ala Lys Lys Leu Asn Ile Glu Val
    210                 215                 220

Pro Glu Asp Thr Pro Ala Leu Lys Glu Ile Tyr Ala Arg Arg Asp Ile
225                 230                 235                 240
```

```
Lys Leu Thr Lys Ile Pro Met Glu Val Leu His Lys Val Pro Thr Thr
                245                 250                 255

Leu Leu His Ser Leu Glu Gly Met Pro Asp Leu Glu Trp Glu Lys Leu
            260                 265                 270

Leu Lys Leu Gln Cys Lys Asp Gly Ser Phe Leu Phe Ser Pro Ser Ser
        275                 280                 285

Thr Ala Phe Ala Leu Met Gln Thr Lys Asp Glu Lys Cys Leu Gln Tyr
    290                 295                 300

Leu Thr Asn Ile Val Thr Lys Phe Asn Gly Gly Val Pro Asn Val Tyr
305                 310                 315                 320

Pro Val Asp Leu Phe Glu His Ile Trp Val Val Asp Arg Leu Gln Arg
                325                 330                 335

Leu Gly Ile Ala Arg Tyr Phe Lys Ser Glu Ile Lys Asp Cys Val Glu
            340                 345                 350

Tyr Ile Asn Lys Tyr Trp Thr Lys Asn Gly Ile Cys Trp Ala Arg Asn
        355                 360                 365

Thr His Val Gln Asp Ile Asp Asp Thr Ala Met Gly Phe Arg Val Leu
    370                 375                 380

Arg Ala His Gly Tyr Asp Val Thr Pro Asp Val Phe Arg Gln Phe Glu
385                 390                 395                 400

Lys Asp Gly Lys Phe Val Cys Phe Ala Gly Gln Ser Thr Gln Ala Val
                405                 410                 415

Thr Gly Met Phe Asn Val Tyr Arg Ala Ser Gln Met Leu Phe Pro Gly
            420                 425                 430

Glu Arg Ile Leu Glu Asp Ala Lys Lys Phe Ser Tyr Asn Tyr Leu Lys
        435                 440                 445

Glu Lys Gln Ser Thr Asn Glu Leu Leu Asp Lys Trp Ile Ile Ala Lys
    450                 455                 460

Asp Leu Pro Gly Glu Val Gly Tyr Ala Leu Asp Ile Pro Trp Tyr Ala
465                 470                 475                 480

Ser Leu Pro Arg Leu Glu Thr Arg Tyr Tyr Leu Glu Gln Tyr Gly Gly
                485                 490                 495
```

Glu Asp Asp Val Trp Ile Gly Lys Thr Leu Tyr Arg Met Gly Tyr Val
500 505 510

Ser Asn Asn Thr Tyr Leu Glu Met Ala Lys Leu Asp Tyr Asn Asn Tyr
515 520 525

Val Ala Val Leu Gln Leu Glu Trp Tyr Thr Ile Gln Gln Trp Tyr Val
530 535 540

Asp Ile Gly Ile Glu Lys Phe Glu Ser Asp Asn Ile Lys Ser Val Leu
545 550 555 560

Val Ser Tyr Tyr Leu Ala Ala Ala Ser Ile Phe Glu Pro Glu Arg Ser
565 570 575

Lys Glu Arg Ile Ala Trp Ala Lys Thr Thr Ile Leu Val Asp Lys Ile
580 585 590

Thr Ser Ile Phe Asp Ser Ser Gln Ser Ser Lys Glu Asp Ile Thr Ala
595 600 605

Phe Ile Asp Lys Phe Arg Asn Lys Ser Ser Ser Lys Lys His Ser Ile
610 615 620

Asn Gly Glu Pro Trp His Glu Val Met Val Ala Leu Lys Lys Thr Leu
625 630 635 640

His Gly Phe Ala Leu Asp Ala Leu Met Thr His Ser Gln Asp Ile His
645 650 655

Pro Gln Leu His Gln Ala Trp Glu Met Trp Leu Thr Lys Leu Gln Asp
660 665 670

Gly Val Asp Val Thr Ala Glu Leu Met Val Gln Met Ile Asn Met Thr
675 680 685

Ala Gly Arg Trp Val Ser Lys Glu Leu Leu Thr His Pro Gln Tyr Gln
690 695 700

Arg Leu Ser Thr Val Thr Asn Ser Val Cys His Asp Ile Thr Lys Leu
705 710 715 720

His Asn Phe Lys Glu Asn Ser Thr Thr Val Asp Ser Lys Val Gln Glu
725 730 735

Leu Val Gln Leu Val Phe Ser Asp Thr Pro Asp Asp Leu Asp Gln Asp

740 745 750

Met Lys Gln Thr Phe Leu Thr Val Met Lys Thr Phe Tyr Tyr Lys Ala
755 760 765

Trp Cys Asp Pro Asn Thr Ile Asn Asp His Ile Ser Lys Val Phe Glu
770 775 780

Ile Val Ile
785

<210> 130
<211> 527
<212> PRT
<213> Streptomyces clavuligerus

<400> 130
Met Pro Asp Ala His Asp Ala Pro Pro Pro Gln Ile Arg Gln Arg Thr
1 5 10 15

Leu Val Asp Glu Ala Thr Gln Leu Leu Thr Glu Ser Ala Glu Asp Ala
20 25 30

Trp Gly Glu Val Ser Val Ser Glu Tyr Glu Thr Ala Arg Leu Val Ala
35 40 45

His Ala Thr Trp Leu Gly Gly His Ala Thr Arg Val Ala Phe Leu Leu
50 55 60

Glu Arg Gln His Glu Asp Gly Ser Trp Gly Pro Pro Gly Gly Tyr Arg
65 70 75 80

Leu Val Pro Thr Leu Ser Ala Val His Ala Leu Leu Thr Cys Leu Ala
85 90 95

Ser Pro Ala Gln Asp His Gly Val Pro His Asp Arg Leu Leu Arg Ala
100 105 110

Val Asp Ala Gly Leu Thr Ala Leu Arg Arg Leu Gly Thr Ser Asp Ser
115 120 125

Pro Pro Asp Thr Ile Ala Val Glu Leu Val Ile Pro Ser Leu Leu Glu
130 135 140

Gly Ile Gln His Leu Leu Asp Pro Ala His Pro His Ser Arg Pro Ala
145 150 155 160

Phe Ser Gln His Arg Gly Ser Leu Val Cys Pro Gly Gly Leu Asp Gly
165 170 175

```
Arg Thr Leu Gly Ala Leu Arg Ser His Ala Ala Ala Gly Thr Pro Val
            180                 185                 190

Pro Gly Lys Val Trp His Ala Ser Glu Thr Leu Gly Leu Ser Thr Glu
        195                 200                 205

Ala Ala Ser His Leu Gln Pro Ala Gln Gly Ile Ile Gly Gly Ser Ala
    210                 215                 220

Ala Ala Thr Ala Thr Trp Leu Thr Arg Val Ala Pro Ser Gln Gln Ser
225                 230                 235                 240

Asp Ser Ala Arg Arg Tyr Leu Glu Glu Leu Gln His Arg Tyr Ser Gly
                245                 250                 255

Pro Val Pro Ser Ile Thr Pro Ile Thr Tyr Phe Glu Arg Ala Trp Leu
            260                 265                 270

Leu Asn Asn Phe Ala Ala Ala Gly Val Pro Cys Glu Ala Pro Ala Ala
        275                 280                 285

Leu Leu Asp Ser Leu Glu Ala Ala Leu Thr Pro Gln Gly Ala Pro Ala
    290                 295                 300

Gly Ala Gly Leu Pro Pro Asp Ala Asp Asp Thr Ala Ala Val Leu Leu
305                 310                 315                 320

Ala Leu Ala Thr His Gly Arg Gly Arg Arg Pro Glu Val Leu Met Asp
                325                 330                 335

Tyr Arg Thr Asp Gly Tyr Phe Gln Cys Phe Ile Gly Glu Arg Thr Pro
            340                 345                 350

Ser Ile Ser Thr Asn Ala His Val Leu Glu Thr Leu Gly His His Val
        355                 360                 365

Ala Gln His Pro Gln Asp Arg Ala Arg Tyr Gly Ser Ala Met Asp Thr
    370                 375                 380

Ala Ser Ala Trp Leu Leu Ala Ala Gln Lys Gln Asp Gly Ser Trp Leu
385                 390                 395                 400

Asp Lys Trp His Ala Ser Pro Tyr Tyr Ala Thr Val Cys Cys Thr Gln
                405                 410                 415

Ala Leu Ala Ala His Ala Ser Pro Ala Thr Ala Pro Ala Arg Gln Arg
```

420 425 430

Ala Val Arg Trp Val Leu Ala Thr Gln Arg Ser Asp Gly Gly Trp Gly
435 440 445

Leu Trp His Ser Thr Val Glu Glu Thr Ala Tyr Ala Leu Gln Ile Leu
450 455 460

Ala Pro Pro Ser Gly Gly Gly Asn Ile Pro Val Gln Gln Ala Leu Thr
465 470 475 480

Arg Gly Arg Ala Arg Leu Cys Gly Ala Leu Pro Leu Thr Pro Leu Trp
485 490 495

His Asp Lys Asp Leu Tyr Thr Pro Val Arg Val Val Arg Ala Ala Arg
500 505 510

Ala Ala Ala Leu Tyr Thr Thr Arg Asp Leu Leu Leu Pro Pro Leu
515 520 525

<210> 131
<211> 516
<212> PRT
<213> Bradyrhizobium japonicum

<400> 131
Met Asn Ala Leu Ser Glu His Ile Leu Ser Glu Leu Arg Arg Leu Leu
1 5 10 15

Ser Glu Met Ser Asp Gly Gly Ser Val Gly Pro Ser Val Tyr Asp Thr
20 25 30

Ala Gln Ala Leu Arg Phe His Gly Asn Val Thr Gly Arg Gln Asp Ala
35 40 45

Tyr Ala Trp Leu Ile Ala Gln Gln Gln Ala Asp Gly Gly Trp Gly Ser
50 55 60

Ala Asp Phe Pro Leu Phe Arg His Ala Pro Thr Trp Ala Ala Leu Leu
65 70 75 80

Ala Leu Gln Arg Ala Asp Pro Leu Pro Gly Ala Ala Asp Ala Val Gln
85 90 95

Thr Ala Thr Arg Phe Leu Gln Arg Gln Pro Asp Pro Tyr Ala His Ala
100 105 110

Val Pro Glu Asp Ala Pro Ile Gly Ala Glu Leu Ile Leu Pro Gln Phe
115 120 125

```
Cys Gly Glu Ala Ala Trp Leu Leu Gly Gly Val Ala Phe Pro Arg His
    130                 135                 140

Pro Ala Leu Leu Pro Leu Arg Gln Ala Cys Leu Val Lys Leu Gly Ala
145                 150                 155                 160

Val Ala Met Leu Pro Ser Gly His Pro Leu Leu His Ser Trp Glu Ala
                165                 170                 175

Trp Gly Thr Ser Pro Thr Thr Ala Cys Pro Asp Asp Asp Gly Ser Ile
            180                 185                 190

Gly Ile Ser Pro Ala Ala Thr Ala Ala Trp Arg Ala Gln Ala Val Thr
        195                 200                 205

Arg Gly Ser Thr Pro Gln Val Gly Arg Ala Asp Ala Tyr Leu Gln Met
    210                 215                 220

Ala Ser Arg Ala Thr Arg Ser Gly Ile Glu Gly Val Phe Pro Asn Val
225                 230                 235                 240

Trp Pro Ile Asn Val Phe Glu Pro Cys Trp Ser Leu Tyr Thr Leu His
                245                 250                 255

Leu Ala Gly Leu Phe Ala His Pro Ala Leu Ala Glu Ala Val Arg Val
            260                 265                 270

Ile Val Ala Gln Leu Glu Ala Arg Leu Gly Val His Gly Leu Gly Pro
        275                 280                 285

Ala Leu His Phe Ala Ala Asp Ala Asp Asp Thr Ala Val Ala Leu Cys
    290                 295                 300

Val Leu His Leu Ala Gly Arg Asp Pro Ala Val Asp Ala Leu Arg His
305                 310                 315                 320

Phe Glu Ile Gly Glu Leu Phe Val Thr Phe Pro Gly Glu Arg Asn Ala
                325                 330                 335

Ser Val Ser Thr Asn Ile His Ala Leu His Ala Leu Arg Leu Leu Gly
            340                 345                 350

Lys Pro Ala Ala Gly Ala Ser Ala Tyr Val Glu Ala Asn Arg Asn Pro
        355                 360                 365

His Gly Leu Trp Asp Asn Glu Lys Trp His Val Ser Trp Leu Tyr Pro
```

370 375 380

Thr Ala His Ala Val Ala Ala Leu Ala Gln Gly Lys Pro Gln Trp Arg
385 390 395 400

Asp Glu Arg Ala Leu Ala Ala Leu Leu Gln Ala Gln Arg Asp Asp Gly
405 410 415

Gly Trp Gly Ala Gly Arg Gly Ser Thr Phe Glu Glu Thr Ala Tyr Ala
420 425 430

Leu Phe Ala Leu His Val Met Asp Gly Ser Glu Glu Ala Thr Gly Arg
435 440 445

Arg Arg Ile Ala Gln Val Val Ala Arg Ala Leu Glu Trp Met Leu Ala
450 455 460

Arg His Ala Ala His Gly Leu Pro Gln Thr Pro Leu Trp Ile Gly Lys
465 470 475 480

Glu Leu Tyr Cys Pro Thr Arg Val Val Arg Val Ala Glu Leu Ala Gly
485 490 495

Leu Trp Leu Ala Leu Arg Trp Gly Arg Arg Val Leu Ala Glu Gly Ala
500 505 510

Gly Ala Ala Pro
515

<210> 132
<211> 784
<212> PRT
<213> Stevia rebaudiana

<400> 132
Met Asn Leu Ser Leu Cys Ile Ala Ser Pro Leu Leu Thr Lys Ser Asn
1 5 10 15

Arg Pro Ala Ala Leu Ser Ala Ile His Thr Ala Ser Thr Ser His Gly
20 25 30

Gly Gln Thr Asn Pro Thr Asn Leu Ile Ile Asp Thr Thr Lys Glu Arg
35 40 45

Ile Gln Lys Gln Phe Lys Asn Val Glu Ile Ser Val Ser Ser Tyr Asp
50 55 60

Thr Ala Trp Val Ala Met Val Pro Ser Pro Asn Ser Pro Lys Ser Pro
65 70 75 80

```
Cys Phe Pro Glu Cys Leu Asn Trp Leu Ile Asn Asn Gln Leu Asn Asp
                85                  90                  95

Gly Ser Trp Gly Leu Val Asn His Thr His Asn His Asn His Pro Leu
            100                 105                 110

Leu Lys Asp Ser Leu Ser Ser Thr Leu Ala Cys Ile Val Ala Leu Lys
        115                 120                 125

Arg Trp Asn Val Gly Glu Asp Gln Ile Asn Lys Gly Leu Ser Phe Ile
    130                 135                 140

Glu Ser Asn Leu Ala Ser Ala Thr Glu Lys Ser Gln Pro Ser Pro Ile
145                 150                 155                 160

Gly Phe Asp Ile Ile Phe Pro Gly Leu Leu Glu Tyr Ala Lys Asn Leu
                165                 170                 175

Asp Ile Asn Leu Leu Ser Lys Gln Thr Asp Phe Ser Leu Met Leu His
            180                 185                 190

Lys Arg Glu Leu Glu Gln Lys Arg Cys His Ser Asn Glu Met Asp Gly
        195                 200                 205

Tyr Leu Ala Tyr Ile Ser Glu Gly Leu Gly Asn Leu Tyr Asp Trp Asn
    210                 215                 220

Met Val Lys Lys Tyr Gln Met Lys Asn Gly Ser Val Phe Asn Ser Pro
225                 230                 235                 240

Ser Ala Thr Ala Ala Ala Phe Ile Asn His Gln Asn Pro Gly Cys Leu
                245                 250                 255

Asn Tyr Leu Asn Ser Leu Leu Asp Lys Phe Gly Asn Ala Val Pro Thr
            260                 265                 270

Val Tyr Pro His Asp Leu Phe Ile Arg Leu Ser Met Val Asp Thr Ile
        275                 280                 285

Glu Arg Leu Gly Ile Ser His His Phe Arg Val Glu Ile Lys Asn Val
    290                 295                 300

Leu Asp Glu Thr Tyr Arg Cys Trp Val Glu Arg Asp Glu Gln Ile Phe
305                 310                 315                 320

Met Asp Val Val Thr Cys Ala Leu Ala Phe Arg Leu Leu Arg Ile Asn
```

325 330 335

Gly Tyr Glu Val Ser Pro Asp Pro Leu Ala Glu Ile Thr Asn Glu Leu
340 345 350

Ala Leu Lys Asp Glu Tyr Ala Ala Leu Glu Thr Tyr His Ala Ser His
355 360 365

Ile Leu Tyr Gln Glu Asp Leu Ser Ser Gly Lys Gln Ile Leu Lys Ser
370 375 380

Ala Asp Phe Leu Lys Glu Ile Ile Ser Thr Asp Ser Asn Arg Leu Ser
385 390 395 400

Lys Leu Ile His Lys Glu Val Glu Asn Ala Leu Lys Phe Pro Ile Asn
405 410 415

Thr Gly Leu Glu Arg Ile Asn Thr Arg Arg Asn Ile Gln Leu Tyr Asn
420 425 430

Val Asp Asn Thr Arg Ile Leu Lys Thr Thr Tyr His Ser Ser Asn Ile
435 440 445

Ser Asn Thr Asp Tyr Leu Arg Leu Ala Val Glu Asp Phe Tyr Thr Cys
450 455 460

Gln Ser Ile Tyr Arg Glu Glu Leu Lys Gly Leu Glu Arg Trp Val Val
465 470 475 480

Glu Asn Lys Leu Asp Gln Leu Lys Phe Ala Arg Gln Lys Thr Ala Tyr
485 490 495

Cys Tyr Phe Ser Val Ala Ala Thr Leu Ser Ser Pro Glu Leu Ser Asp
500 505 510

Ala Arg Ile Ser Trp Ala Lys Asn Gly Ile Leu Thr Thr Val Val Asp
515 520 525

Asp Phe Phe Asp Ile Gly Gly Thr Ile Asp Glu Leu Thr Asn Leu Ile
530 535 540

Gln Cys Val Glu Lys Trp Asn Val Asp Val Asp Lys Asp Cys Cys Ser
545 550 555 560

Glu His Val Arg Ile Leu Phe Leu Ala Leu Lys Asp Ala Ile Cys Trp
565 570 575

Ile Gly Asp Glu Ala Phe Lys Trp Gln Ala Arg Asp Val Thr Ser His
580 585 590

Val Ile Gln Thr Trp Leu Glu Leu Met Asn Ser Met Leu Arg Glu Ala
595 600 605

Ile Trp Thr Arg Asp Ala Tyr Val Pro Thr Leu Asn Glu Tyr Met Glu
610 615 620

Asn Ala Tyr Val Ser Phe Ala Leu Gly Pro Ile Val Lys Pro Ala Ile
625 630 635 640

Tyr Phe Val Gly Pro Lys Leu Ser Glu Glu Ile Val Glu Ser Ser Glu
645 650 655

Tyr His Asn Leu Phe Lys Leu Met Ser Thr Gln Gly Arg Leu Leu Asn
660 665 670

Asp Ile His Ser Phe Lys Arg Glu Phe Lys Glu Gly Lys Leu Asn Ala
675 680 685

Val Ala Leu His Leu Ser Asn Gly Glu Ser Gly Lys Val Glu Glu Glu
690 695 700

Val Val Glu Glu Met Met Met Met Ile Lys Asn Lys Arg Lys Glu Leu
705 710 715 720

Met Lys Leu Ile Phe Glu Glu Asn Gly Ser Ile Val Pro Arg Ala Cys
725 730 735

Lys Asp Ala Phe Trp Asn Met Cys His Val Leu Asn Phe Phe Tyr Ala
740 745 750

Asn Asp Asp Gly Phe Thr Gly Asn Thr Ile Leu Asp Thr Val Lys Asp
755 760 765

Ile Ile Tyr Asn Pro Leu Val Leu Val Asn Glu Asn Glu Glu Gln Arg
770 775 780

<210> 133
<211> 784
<212> PRT
<213> Stevia rebaudiana

<400> 133
Met Asn Leu Ser Leu Cys Ile Ala Ser Pro Leu Leu Thr Lys Ser Ser
1 5 10 15

Arg Pro Thr Ala Leu Ser Ala Ile His Thr Ala Ser Thr Ser His Gly

```
            20                  25                  30

Gly Gln Thr Asn Pro Thr Asn Leu Ile Ile Asp Thr Thr Lys Glu Arg
        35                  40                  45

Ile Gln Lys Leu Phe Lys Asn Val Glu Ile Ser Val Ser Ser Tyr Asp
    50                  55                  60

Thr Ala Trp Val Ala Met Val Pro Ser Pro Asn Ser Pro Lys Ser Pro
65                  70                  75                  80

Cys Phe Pro Glu Cys Leu Asn Trp Leu Ile Asn Asn Gln Leu Asn Asp
                85                  90                  95

Gly Ser Trp Gly Leu Val Asn His Thr His Asn His Asn His Pro Leu
            100                 105                 110

Leu Lys Asp Ser Leu Ser Ser Thr Leu Ala Cys Ile Val Ala Leu Lys
        115                 120                 125

Arg Trp Asn Val Gly Glu Asp Gln Ile Asn Lys Gly Leu Ser Phe Ile
    130                 135                 140

Glu Ser Asn Leu Ala Ser Ala Thr Asp Lys Ser Gln Pro Ser Pro Ile
145                 150                 155                 160

Gly Phe Asp Ile Ile Phe Pro Gly Leu Leu Glu Tyr Ala Lys Asn Leu
                165                 170                 175

Asp Ile Asn Leu Leu Ser Lys Gln Thr Asp Phe Ser Leu Met Leu His
            180                 185                 190

Lys Arg Glu Leu Glu Gln Lys Arg Cys His Ser Asn Glu Ile Asp Gly
        195                 200                 205

Tyr Leu Ala Tyr Ile Ser Glu Gly Leu Gly Asn Leu Tyr Asp Trp Asn
    210                 215                 220

Met Val Lys Lys Tyr Gln Met Lys Asn Gly Ser Val Phe Asn Ser Pro
225                 230                 235                 240

Ser Ala Thr Ala Ala Ala Phe Ile Asn His Gln Asn Pro Gly Cys Leu
                245                 250                 255

Asn Tyr Leu Asn Ser Leu Leu Asp Lys Phe Gly Asn Ala Val Pro Thr
            260                 265                 270
```

```
Val Tyr Pro Leu Asp Leu Tyr Ile Arg Leu Ser Met Val Asp Thr Ile
        275                 280                 285

Glu Arg Leu Gly Ile Ser His His Phe Arg Val Glu Ile Lys Asn Val
    290                 295                 300

Leu Asp Glu Thr Tyr Arg Cys Trp Val Glu Arg Asp Glu Gln Ile Phe
305                 310                 315                 320

Met Asp Val Val Thr Cys Ala Leu Ala Phe Arg Leu Leu Arg Ile His
                325                 330                 335

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

Ala Phe Lys Asp Glu Tyr Ala Ala Leu Glu Thr Tyr His Ala Ser Gln
        355                 360                 365

Ile Leu Tyr Gln Glu Asp Leu Ser Ser Gly Lys Gln Ile Leu Lys Ser
    370                 375                 380

Ala Asp Phe Leu Lys Gly Ile Leu Ser Thr Asp Ser Asn Arg Leu Ser
385                 390                 395                 400

Lys Leu Ile His Lys Glu Val Glu Asn Ala Leu Lys Phe Pro Ile Asn
                405                 410                 415

Thr Gly Leu Glu Arg Ile Asn Thr Arg Arg Asn Ile Gln Leu Tyr Asn
            420                 425                 430

Val Asp Asn Thr Arg Ile Leu Lys Thr Thr Tyr His Ser Ser Asn Ile
        435                 440                 445

Ser Asn Thr Tyr Tyr Leu Arg Leu Ala Val Glu Asp Phe Tyr Thr Cys
    450                 455                 460

Gln Ser Ile Tyr Arg Glu Glu Leu Lys Gly Leu Glu Arg Trp Val Val
465                 470                 475                 480

Gln Asn Lys Leu Asp Gln Leu Lys Phe Ala Arg Gln Lys Thr Ala Tyr
                485                 490                 495

Cys Tyr Phe Ser Val Ala Ala Thr Leu Ser Ser Pro Glu Leu Ser Asp
            500                 505                 510

Ala Arg Ile Ser Trp Ala Lys Asn Gly Ile Leu Thr Thr Val Val Asp
        515                 520                 525
```

```
Asp Phe Phe Asp Ile Gly Gly Thr Ile Asp Glu Leu Thr Asn Leu Ile
    530                 535                 540

Gln Cys Val Glu Lys Trp Asn Val Asp Val Asp Lys Asp Cys Cys Ser
545                 550                 555                 560

Glu His Val Arg Ile Leu Phe Leu Ala Leu Lys Asp Ala Ile Cys Trp
                565                 570                 575

Ile Gly Asp Glu Ala Phe Lys Trp Gln Ala Arg Asp Val Thr Ser His
            580                 585                 590

Val Ile Gln Thr Trp Leu Glu Leu Met Asn Ser Met Leu Arg Glu Ala
        595                 600                 605

Ile Trp Thr Arg Asp Ala Tyr Val Pro Thr Leu Asn Glu Tyr Met Glu
    610                 615                 620

Asn Ala Tyr Val Ser Phe Ala Leu Gly Pro Ile Val Lys Pro Ala Ile
625                 630                 635                 640

Tyr Phe Val Gly Pro Lys Leu Ser Glu Glu Ile Val Glu Ser Ser Glu
                645                 650                 655

Tyr His Asn Leu Phe Lys Leu Met Ser Thr Gln Gly Arg Leu Leu Asn
            660                 665                 670

Asp Ile His Ser Phe Lys Arg Glu Phe Lys Glu Gly Lys Leu Asn Ala
        675                 680                 685

Val Ala Leu His Leu Ser Asn Gly Glu Ser Gly Lys Val Glu Glu Glu
    690                 695                 700

Val Val Glu Glu Met Met Met Met Ile Lys Asn Lys Arg Lys Glu Leu
705                 710                 715                 720

Met Lys Leu Ile Phe Glu Glu Asn Gly Ser Ile Val Pro Arg Ala Cys
                725                 730                 735

Lys Asp Ala Phe Trp Asn Met Cys His Val Leu Asn Phe Phe Tyr Ala
            740                 745                 750

Asn Asp Asp Gly Phe Thr Gly Asn Thr Ile Leu Asp Thr Val Lys Asp
        755                 760                 765

Ile Ile Tyr Asn Pro Leu Val Leu Val Asn Glu Asn Glu Glu Gln Arg
    770                 775                 780
```

<210> 134
<211> 590
<212> PRT
<213> Zea mays

<400> 134
```
Met Ala Met Pro Val Lys Leu Thr Pro Ala Ser Leu Ser Leu Lys Ala
1               5                   10                  15

Val Cys Cys Arg Phe Ser Ser Gly Gly His Ala Leu Arg Phe Gly Ser
            20                  25                  30

Ser Leu Pro Cys Trp Arg Arg Thr Pro Thr Gln Arg Ser Thr Ser Ser
        35                  40                  45

Ser Thr Thr Arg Pro Ala Ala Glu Val Ser Ser Gly Lys Ser Lys Gln
    50                  55                  60

His Asp Gln Glu Ala Ser Glu Ala Thr Ile Arg Gln Gln Leu Gln Leu
65                  70                  75                  80

Val Asp Val Leu Glu Asn Met Gly Ile Ser Arg His Phe Ala Ala Glu
                85                  90                  95

Ile Lys Cys Ile Leu Asp Arg Thr Tyr Arg Ser Trp Leu Gln Arg His
            100                 105                 110

Glu Glu Ile Met Leu Asp Thr Met Thr Cys Ala Met Ala Phe Arg Ile
        115                 120                 125

Leu Arg Leu Asn Gly Tyr Asn Val Ser Ser Asp Glu Leu Tyr His Val
    130                 135                 140

Val Glu Ala Ser Gly Leu His Asn Ser Leu Gly Gly Tyr Leu Asn Asp
145                 150                 155                 160

Thr Arg Thr Leu Leu Glu Leu His Lys Ala Ser Thr Val Ser Ile Ser
                165                 170                 175

Glu Asp Glu Ser Ile Leu Asp Ser Ile Gly Ser Arg Ser Arg Thr Leu
            180                 185                 190

Leu Arg Glu Gln Leu Glu Ser Gly Gly Ala Leu Arg Lys Pro Ser Leu
        195                 200                 205

Phe Lys Glu Val Glu His Ala Leu Asp Gly Pro Phe Tyr Thr Thr Leu
    210                 215                 220
```

```
Asp Arg Leu His His Arg Trp Asn Ile Glu Asn Phe Asn Ile Ile Glu
225                 230                 235                 240

Gln His Met Leu Glu Thr Pro Tyr Leu Ser Asn Gln His Thr Ser Arg
                245                 250                 255

Asp Ile Leu Ala Leu Ser Ile Arg Asp Phe Ser Ser Ser Gln Phe Thr
            260                 265                 270

Tyr Gln Gln Glu Leu Gln His Leu Glu Ser Trp Val Lys Glu Cys Arg
        275                 280                 285

Leu Asp Gln Leu Gln Phe Ala Arg Gln Lys Leu Ala Tyr Phe Tyr Leu
    290                 295                 300

Ser Ala Ala Gly Thr Met Phe Ser Pro Glu Leu Ser Asp Ala Arg Thr
305                 310                 315                 320

Leu Trp Ala Lys Asn Gly Val Leu Thr Thr Ile Val Asp Asp Phe Phe
                325                 330                 335

Asp Val Ala Gly Ser Lys Glu Glu Leu Glu Asn Leu Val Met Leu Val
            340                 345                 350

Glu Met Trp Asp Glu His His Lys Val Glu Phe Tyr Ser Glu Gln Val
        355                 360                 365

Glu Ile Ile Phe Ser Ser Ile Tyr Asp Ser Val Asn Gln Leu Gly Glu
    370                 375                 380

Lys Ala Ser Leu Val Gln Asp Arg Ser Ile Thr Lys His Leu Val Glu
385                 390                 395                 400

Ile Trp Leu Asp Leu Leu Lys Ser Met Met Thr Glu Val Glu Trp Arg
                405                 410                 415

Leu Ser Lys Tyr Val Pro Thr Glu Lys Glu Tyr Met Ile Asn Ala Ser
            420                 425                 430

Leu Ile Phe Gly Leu Gly Pro Ile Val Leu Pro Ala Leu Tyr Phe Val
        435                 440                 445

Gly Pro Lys Ile Ser Glu Ser Ile Val Lys Asp Pro Glu Tyr Asp Glu
    450                 455                 460

Leu Phe Lys Leu Met Ser Thr Cys Gly Arg Leu Leu Asn Asp Val Gln
465                 470                 475                 480
```

```
Thr Phe Glu Arg Glu Tyr Asn Glu Gly Lys Leu Asn Ser Val Ser Leu
                485                 490                 495

Leu Val Leu His Gly Gly Pro Met Ser Ile Ser Asp Ala Lys Arg Lys
            500                 505                 510

Leu Gln Lys Pro Ile Asp Thr Cys Arg Arg Asp Leu Leu Ser Leu Val
        515                 520                 525

Leu Arg Glu Glu Ser Val Val Pro Arg Pro Cys Lys Glu Leu Phe Trp
    530                 535                 540

Lys Met Cys Lys Val Cys Tyr Phe Phe Tyr Ser Thr Thr Asp Gly Phe
545                 550                 555                 560

Ser Ser Gln Val Glu Arg Ala Lys Glu Val Asp Ala Val Ile Asn Glu
                565                 570                 575

Pro Leu Lys Leu Gln Gly Ser His Thr Leu Val Ser Asp Val
            580                 585                 590


<210> 135
<211> 743
<212> PRT
<213> Populus trichocarpa

<400> 135
Met Gln Asn Phe His Gly Thr Lys Glu Arg Ile Lys Lys Met Phe Asp
1               5                   10                  15

Lys Ile Glu Leu Ser Val Ser Ser Tyr Asp Thr Ala Trp Val Ala Met
            20                  25                  30

Val Pro Ser Pro Asp Cys Pro Glu Thr Pro Cys Phe Pro Glu Cys Thr
        35                  40                  45

Lys Trp Ile Leu Glu Asn Gln Leu Gly Asp Gly Ser Trp Ser Leu Pro
    50                  55                  60

His Gly Asn Pro Leu Leu Val Lys Asp Ala Leu Ser Ser Thr Leu Ala
65                  70                  75                  80

Cys Ile Leu Ala Leu Lys Arg Trp Gly Ile Gly Glu Glu Gln Ile Asn
                85                  90                  95

Lys Gly Leu Arg Phe Ile Glu Leu Asn Ser Ala Ser Val Thr Asp Asn
            100                 105                 110
```

```
Glu Gln His Lys Pro Ile Gly Phe Asp Ile Ile Phe Pro Gly Met Ile
        115                 120                 125

Glu Tyr Ala Ile Asp Leu Asp Leu Asn Leu Pro Leu Lys Pro Thr Asp
    130                 135                 140

Ile Asn Ser Met Leu His Arg Arg Ala Leu Glu Leu Thr Ser Gly Gly
145                 150                 155                 160

Gly Lys Asn Leu Glu Gly Arg Arg Ala Tyr Leu Ala Tyr Val Ser Glu
                165                 170                 175

Gly Ile Gly Lys Leu Gln Asp Trp Glu Met Ala Met Lys Tyr Gln Arg
            180                 185                 190

Lys Asn Gly Ser Leu Phe Asn Ser Pro Ser Thr Thr Ala Ala Ala Phe
        195                 200                 205

Ile His Ile Gln Asp Ala Glu Cys Leu His Tyr Ile Arg Ser Leu Leu
    210                 215                 220

Gln Lys Phe Gly Asn Ala Val Pro Thr Ile Tyr Pro Leu Asp Ile Tyr
225                 230                 235                 240

Ala Arg Leu Ser Met Val Asp Ala Leu Glu Arg Leu Gly Ile Asp Arg
                245                 250                 255

His Phe Arg Lys Glu Arg Lys Phe Val Leu Asp Glu Thr Tyr Arg Phe
            260                 265                 270

Trp Leu Gln Gly Glu Glu Glu Ile Phe Ser Asp Asn Ala Thr Cys Ala
        275                 280                 285

Leu Ala Phe Arg Ile Leu Arg Leu Asn Gly Tyr Asp Val Ser Leu Glu
    290                 295                 300

Asp His Phe Ser Asn Ser Leu Gly Gly Tyr Leu Lys Asp Ser Gly Ala
305                 310                 315                 320

Ala Leu Glu Leu Tyr Arg Ala Leu Gln Leu Ser Tyr Pro Asp Glu Ser
                325                 330                 335

Leu Leu Glu Lys Gln Asn Ser Arg Thr Ser Tyr Phe Leu Lys Gln Gly
            340                 345                 350

Leu Ser Asn Val Ser Leu Cys Gly Asp Arg Leu Arg Lys Asn Ile Ile
        355                 360                 365
```

```
Gly Glu Val His Asp Ala Leu Asn Phe Ser Asp His Ala Asn Leu Gln
    370                 375                 380

Arg Leu Ala Ile Arg Arg Arg Ile Lys His Tyr Ala Thr Asp Asp Thr
385                 390                 395                 400

Arg Ile Leu Lys Thr Ser Tyr Arg Cys Ser Thr Ile Gly Asn Gln Asp
                405                 410                 415

Phe Leu Lys Leu Ala Val Glu Asp Phe Asn Ile Cys Gln Ser Ile Gln
            420                 425                 430

Arg Glu Glu Phe Lys His Ile Glu Arg Trp Val Val Glu Arg Arg Leu
        435                 440                 445

Asp Lys Leu Lys Phe Ala Arg Gln Lys Glu Ala Tyr Cys Tyr Phe Ser
    450                 455                 460

Ala Ala Ala Thr Leu Phe Ala Pro Glu Leu Ser Asp Ala Arg Met Ser
465                 470                 475                 480

Trp Ala Lys Asn Gly Val Leu Thr Thr Val Val Asp Asp Phe Phe Asp
                485                 490                 495

Val Gly Gly Ser Glu Glu Glu Leu Val Asn Leu Ile Glu Leu Ile Glu
            500                 505                 510

Arg Trp Asp Val Asn Gly Ser Ala Asp Phe Cys Ser Glu Glu Val Glu
        515                 520                 525

Ile Ile Tyr Ser Ala Ile His Ser Thr Ile Ser Glu Ile Gly Asp Lys
    530                 535                 540

Ser Phe Gly Trp Gln Gly Arg Asp Val Lys Ser Gln Val Ile Lys Ile
545                 550                 555                 560

Trp Leu Asp Leu Leu Lys Ser Met Leu Thr Glu Ala Gln Trp Ser Ser
                565                 570                 575

Asn Lys Ser Val Pro Thr Leu Asp Glu Tyr Met Thr Thr Ala His Val
            580                 585                 590

Ser Phe Ala Leu Gly Pro Ile Val Leu Pro Ala Leu Tyr Phe Val Gly
        595                 600                 605

Pro Lys Leu Ser Glu Glu Val Ala Gly His Pro Glu Leu Leu Asn Leu
```

610 615 620

Tyr Lys Val Thr Ser Thr Cys Gly Arg Leu Leu Asn Asp Trp Arg Ser
625 630 635 640

Phe Lys Arg Glu Ser Glu Glu Gly Lys Leu Asn Ala Ile Ser Leu Tyr
645 650 655

Met Ile His Ser Gly Gly Ala Ser Thr Glu Glu Glu Thr Ile Glu His
660 665 670

Phe Lys Gly Leu Ile Asp Ser Gln Arg Arg Gln Leu Leu Gln Leu Val
675 680 685

Leu Gln Glu Lys Asp Ser Ile Ile Pro Arg Pro Cys Lys Asp Leu Phe
690 695 700

Trp Asn Met Ile Lys Leu Leu His Thr Phe Tyr Met Lys Asp Asp Gly
705 710 715 720

Phe Thr Ser Asn Glu Met Arg Asn Val Val Lys Ala Ile Ile Asn Glu
725 730 735

Pro Ile Ser Leu Asp Glu Leu
740

<210> 136
<211> 983
<212> PRT
<213> Phomopsis amygdali

<400> 136
Met Glu Phe Asp Glu Pro Leu Val Asp Glu Ala Arg Ser Leu Val Gln
1 5 10 15

Arg Thr Leu Gln Asp Tyr Asp Asp Arg Tyr Gly Phe Gly Thr Met Ser
20 25 30

Cys Ala Ala Tyr Asp Thr Ala Trp Val Ser Leu Val Thr Lys Thr Val
35 40 45

Asp Gly Arg Lys Gln Trp Leu Phe Pro Glu Cys Phe Glu Phe Leu Leu
50 55 60

Glu Thr Gln Ser Asp Ala Gly Gly Trp Glu Ile Gly Asn Ser Ala Pro
65 70 75 80

Ile Asp Gly Ile Leu Asn Thr Ala Ala Ser Leu Leu Ala Leu Lys Arg
85 90 95

```
His Val Gln Thr Glu Gln Ile Ile Gln Pro Gln His Asp His Lys Asp
            100                 105                 110

Leu Ala Gly Arg Ala Glu Arg Ala Ala Ala Ser Leu Arg Ala Gln Leu
        115                 120                 125

Ala Ala Leu Asp Val Ser Thr Thr Glu His Val Gly Phe Glu Ile Ile
    130                 135                 140

Val Pro Ala Met Leu Asp Pro Leu Glu Ala Glu Asp Pro Ser Leu Val
145                 150                 155                 160

Phe Asp Phe Pro Ala Arg Lys Pro Leu Met Lys Ile His Asp Ala Lys
                165                 170                 175

Met Ser Arg Phe Arg Pro Glu Tyr Leu Tyr Gly Lys Gln Pro Met Thr
            180                 185                 190

Ala Leu His Ser Leu Glu Ala Phe Ile Gly Lys Ile Asp Phe Asp Lys
        195                 200                 205

Val Arg His His Arg Thr His Gly Ser Met Met Gly Ser Pro Ser Ser
    210                 215                 220

Thr Ala Ala Tyr Leu Met His Ala Ser Gln Trp Asp Gly Asp Ser Glu
225                 230                 235                 240

Ala Tyr Leu Arg His Val Ile Lys His Ala Ala Gly Gln Gly Thr Gly
                245                 250                 255

Ala Val Pro Ser Ala Phe Pro Ser Thr His Phe Glu Ser Ser Trp Ile
            260                 265                 270

Leu Thr Thr Leu Phe Arg Ala Gly Phe Ser Ala Ser His Leu Ala Cys
        275                 280                 285

Asp Glu Leu Asn Lys Leu Val Glu Ile Leu Glu Gly Ser Phe Glu Lys
    290                 295                 300

Glu Gly Gly Ala Ile Gly Tyr Ala Pro Gly Phe Gln Ala Asp Val Asp
305                 310                 315                 320

Asp Thr Ala Lys Thr Ile Ser Thr Leu Ala Val Leu Gly Arg Asp Ala
                325                 330                 335

Thr Pro Arg Gln Met Ile Lys Val Phe Glu Ala Asn Thr His Phe Arg
```

```
            340                 345                 350

Thr Tyr Pro Gly Glu Arg Asp Pro Ser Leu Thr Ala Asn Cys Asn Ala
        355                 360                 365

Leu Ser Ala Leu Leu His Gln Pro Asp Ala Ala Met Tyr Gly Ser Gln
    370                 375                 380

Ile Gln Lys Ile Thr Lys Phe Val Cys Asp Tyr Trp Trp Lys Ser Asp
385                 390                 395                 400

Gly Lys Ile Lys Asp Lys Trp Asn Thr Cys Tyr Leu Tyr Pro Ser Val
                405                 410                 415

Leu Leu Val Glu Val Leu Val Asp Leu Val Ser Leu Leu Glu Gln Gly
            420                 425                 430

Lys Leu Pro Asp Val Leu Asp Gln Glu Leu Gln Tyr Arg Val Ala Ile
        435                 440                 445

Thr Leu Phe Gln Ala Cys Leu Arg Pro Leu Leu Asp Gln Asp Ala Glu
    450                 455                 460

Gly Ser Trp Asn Lys Ser Ile Glu Ala Thr Ala Tyr Gly Ile Leu Ile
465                 470                 475                 480

Leu Thr Glu Ala Arg Arg Val Cys Phe Phe Asp Arg Leu Ser Glu Pro
                485                 490                 495

Leu Asn Glu Ala Ile Arg Arg Gly Ile Ala Phe Ala Asp Ser Met Ser
            500                 505                 510

Gly Thr Glu Ala Gln Leu Asn Tyr Ile Trp Ile Glu Lys Val Ser Tyr
        515                 520                 525

Ala Pro Ala Leu Leu Thr Lys Ser Tyr Leu Leu Ala Ala Arg Trp Ala
    530                 535                 540

Ala Lys Ser Pro Leu Gly Ala Ser Val Gly Ser Ser Leu Trp Thr Pro
545                 550                 555                 560

Pro Arg Glu Gly Leu Asp Lys His Val Arg Leu Phe His Gln Ala Glu
                565                 570                 575

Leu Phe Arg Ser Leu Pro Glu Trp Glu Leu Arg Ala Ser Met Ile Glu
            580                 585                 590
```

```
Ala Ala Leu Phe Thr Pro Leu Leu Arg Ala His Arg Leu Asp Val Phe
        595                 600                 605

Pro Arg Gln Asp Val Gly Glu Asp Lys Tyr Leu Asp Val Val Pro Phe
    610                 615                 620

Phe Trp Thr Ala Ala Asn Asn Arg Asp Arg Thr Tyr Ala Ser Thr Leu
625                 630                 635                 640

Phe Leu Tyr Asp Met Cys Phe Ile Ala Met Leu Asn Phe Gln Leu Asp
                645                 650                 655

Glu Phe Met Glu Ala Thr Ala Gly Ile Leu Phe Arg Asp His Met Asp
            660                 665                 670

Asp Leu Arg Gln Leu Ile His Asp Leu Leu Ala Glu Lys Thr Ser Pro
        675                 680                 685

Lys Ser Ser Gly Arg Ser Ser Gln Gly Thr Lys Asp Ala Asp Ser Gly
    690                 695                 700

Ile Glu Glu Asp Val Ser Met Ser Asp Ser Ala Ser Asp Ser Gln Asp
705                 710                 715                 720

Arg Ser Pro Glu Tyr Asp Leu Val Phe Ser Ala Leu Ser Thr Phe Thr
                725                 730                 735

Lys His Val Leu Gln His Pro Ser Ile Gln Ser Ala Ser Val Trp Asp
            740                 745                 750

Arg Lys Leu Leu Ala Arg Glu Met Lys Ala Tyr Leu Leu Ala His Ile
        755                 760                 765

Gln Gln Ala Glu Asp Ser Thr Pro Leu Ser Glu Leu Lys Asp Val Pro
    770                 775                 780

Gln Lys Thr Asp Val Thr Arg Val Ser Thr Ser Thr Thr Thr Phe Phe
785                 790                 795                 800

Asn Trp Val Arg Thr Thr Ser Ala Asp His Ile Ser Cys Pro Tyr Ser
                805                 810                 815

Phe His Phe Val Ala Cys His Leu Gly Ala Ala Leu Ser Pro Lys Gly
            820                 825                 830

Ser Asn Gly Asp Cys Tyr Pro Ser Ala Gly Glu Lys Phe Leu Ala Ala
        835                 840                 845
```

```
Ala Val Cys Arg His Leu Ala Thr Met Cys Arg Met Tyr Asn Asp Leu
    850                 855                 860

Gly Ser Ala Glu Arg Asp Ser Asp Glu Gly Asn Leu Asn Ser Leu Asp
865                 870                 875                 880

Phe Pro Glu Phe Ala Asp Ser Ala Gly Asn Gly Gly Ile Glu Ile Gln
                885                 890                 895

Lys Ala Ala Leu Leu Arg Leu Ala Glu Phe Glu Arg Asp Ser Tyr Leu
            900                 905                 910

Glu Ala Phe Arg Arg Leu Gln Asp Glu Ser Asn Arg Val His Gly Pro
        915                 920                 925

Ala Gly Gly Asp Glu Ala Arg Leu Ser Arg Arg Arg Met Ala Ile Leu
    930                 935                 940

Glu Phe Phe Ala Gln Gln Val Asp Leu Tyr Gly Gln Val Tyr Val Ile
945                 950                 955                 960

Arg Asp Ile Ser Ala Arg Ile Pro Lys Asn Glu Val Glu Lys Lys Arg
                965                 970                 975

Lys Leu Asp Asp Ala Phe Asn
            980
```

<210> 137
<211> 881
<212> PRT
<213> Physcomitrella patens

<400> 137

```
Met Ala Ser Ser Thr Leu Ile Gln Asn Arg Ser Cys Gly Val Thr Ser
1               5                   10                  15

Ser Met Ser Ser Phe Gln Ile Phe Arg Gly Gln Pro Leu Arg Phe Pro
            20                  25                  30

Gly Thr Arg Thr Pro Ala Ala Val Gln Cys Leu Lys Lys Arg Arg Cys
        35                  40                  45

Leu Arg Pro Thr Glu Ser Val Leu Glu Ser Ser Pro Gly Ser Gly Ser
    50                  55                  60

Tyr Arg Ile Val Thr Gly Pro Ser Gly Ile Asn Pro Ser Ser Asn Gly
65                  70                  75                  80
```

```
His Leu Gln Glu Gly Ser Leu Thr His Arg Leu Pro Ile Pro Met Glu
                85                  90                  95

Lys Ser Ile Asp Asn Phe Gln Ser Thr Leu Tyr Val Ser Asp Ile Trp
            100                 105                 110

Ser Glu Thr Leu Gln Arg Thr Glu Cys Leu Leu Gln Val Thr Glu Asn
        115                 120                 125

Val Gln Met Asn Glu Trp Ile Glu Glu Ile Arg Met Tyr Phe Arg Asn
    130                 135                 140

Met Thr Leu Gly Glu Ile Ser Met Ser Pro Tyr Asp Thr Ala Trp Val
145                 150                 155                 160

Ala Arg Val Pro Ala Leu Asp Gly Ser His Gly Pro Gln Phe His Arg
                165                 170                 175

Ser Leu Gln Trp Ile Ile Asp Asn Gln Leu Pro Asp Gly Asp Trp Gly
            180                 185                 190

Glu Pro Ser Leu Phe Leu Gly Tyr Asp Arg Val Cys Asn Thr Leu Ala
        195                 200                 205

Cys Val Ile Ala Leu Lys Thr Trp Gly Val Gly Ala Gln Asn Val Glu
    210                 215                 220

Arg Gly Ile Gln Phe Leu Gln Ser Asn Ile Tyr Lys Met Glu Glu Asp
225                 230                 235                 240

Asp Ala Asn His Met Pro Ile Gly Phe Glu Ile Val Phe Pro Ala Met
                245                 250                 255

Met Glu Asp Ala Lys Ala Leu Gly Leu Asp Leu Pro Tyr Asp Ala Thr
            260                 265                 270

Ile Leu Gln Gln Ile Ser Ala Glu Arg Glu Lys Lys Met Lys Lys Ile
        275                 280                 285

Pro Met Ala Met Val Tyr Lys Tyr Pro Thr Thr Leu Leu His Ser Leu
    290                 295                 300

Glu Gly Leu His Arg Glu Val Asp Trp Asn Lys Leu Leu Gln Leu Gln
305                 310                 315                 320

Ser Glu Asn Gly Ser Phe Leu Tyr Ser Pro Ala Ser Thr Ala Cys Ala
                325                 330                 335
```

```
Leu Met Tyr Thr Lys Asp Val Lys Cys Phe Asp Tyr Leu Asn Gln Leu
            340                 345                 350

Leu Ile Lys Phe Asp His Ala Cys Pro Asn Val Tyr Pro Val Asp Leu
        355                 360                 365

Phe Glu Arg Leu Trp Met Val Asp Arg Leu Gln Arg Leu Gly Ile Ser
    370                 375                 380

Arg Tyr Phe Glu Arg Glu Ile Arg Asp Cys Leu Gln Tyr Val Tyr Arg
385                 390                 395                 400

Tyr Trp Lys Asp Cys Gly Ile Gly Trp Ala Ser Asn Ser Ser Val Gln
                405                 410                 415

Asp Val Asp Asp Thr Ala Met Ala Phe Arg Leu Leu Arg Thr His Gly
            420                 425                 430

Phe Asp Val Lys Glu Asp Cys Phe Arg Gln Phe Phe Lys Asp Gly Glu
        435                 440                 445

Phe Phe Cys Phe Ala Gly Gln Ser Ser Gln Ala Val Thr Gly Met Phe
    450                 455                 460

Asn Leu Ser Arg Ala Ser Gln Thr Leu Phe Pro Gly Glu Ser Leu Leu
465                 470                 475                 480

Lys Lys Ala Arg Thr Phe Ser Arg Asn Phe Leu Arg Thr Lys His Glu
                485                 490                 495

Asn Asn Glu Cys Phe Asp Lys Trp Ile Ile Thr Lys Asp Leu Ala Gly
            500                 505                 510

Glu Val Glu Tyr Asn Leu Thr Phe Pro Trp Tyr Ala Ser Leu Pro Arg
        515                 520                 525

Leu Glu His Arg Thr Tyr Leu Asp Gln Tyr Gly Ile Asp Asp Ile Trp
    530                 535                 540

Ile Gly Lys Ser Leu Tyr Lys Met Pro Ala Val Thr Asn Glu Val Phe
545                 550                 555                 560

Leu Lys Leu Ala Lys Ala Asp Phe Asn Met Cys Gln Ala Leu His Lys
                565                 570                 575

Lys Glu Leu Glu Gln Val Ile Lys Trp Asn Ala Ser Cys Gln Phe Arg
            580                 585                 590
```

```
Asp Leu Glu Phe Ala Arg Gln Lys Ser Val Glu Cys Tyr Phe Ala Gly
        595                 600                 605

Ala Ala Thr Met Phe Glu Pro Glu Met Val Gln Ala Arg Leu Val Trp
    610                 615                 620

Ala Arg Cys Cys Val Leu Thr Thr Val Leu Asp Asp Tyr Phe Asp His
625                 630                 635                 640

Gly Thr Pro Val Glu Glu Leu Arg Val Phe Val Gln Ala Val Arg Thr
                645                 650                 655

Trp Asn Pro Glu Leu Ile Asn Gly Leu Pro Glu Gln Ala Lys Ile Leu
            660                 665                 670

Phe Met Gly Leu Tyr Lys Thr Val Asn Thr Ile Ala Glu Glu Ala Phe
        675                 680                 685

Met Ala Gln Lys Arg Asp Val His His His Leu Lys His Tyr Trp Asp
    690                 695                 700

Lys Leu Ile Thr Ser Ala Leu Lys Glu Ala Glu Trp Ala Glu Ser Gly
705                 710                 715                 720

Tyr Val Pro Thr Phe Asp Glu Tyr Met Glu Val Ala Glu Ile Ser Val
                725                 730                 735

Ala Leu Glu Pro Ile Val Cys Ser Thr Leu Phe Phe Ala Gly His Arg
            740                 745                 750

Leu Asp Glu Asp Val Leu Asp Ser Tyr Asp Tyr His Leu Val Met His
        755                 760                 765

Leu Val Asn Arg Val Gly Arg Ile Leu Asn Asp Ile Gln Gly Met Lys
    770                 775                 780

Arg Glu Ala Ser Gln Gly Lys Ile Ser Ser Val Gln Ile Tyr Met Glu
785                 790                 795                 800

Glu His Pro Ser Val Pro Ser Glu Ala Met Ala Ile Ala His Leu Gln
                805                 810                 815

Glu Leu Val Asp Asn Ser Met Gln Gln Leu Thr Tyr Glu Val Leu Arg
            820                 825                 830

Phe Thr Ala Val Pro Lys Ser Cys Lys Arg Ile His Leu Asn Met Ala
```

835 840 845

Lys Ile Met His Ala Phe Tyr Lys Asp Thr Asp Gly Phe Ser Ser Leu
850 855 860

Thr Ala Met Thr Gly Phe Val Lys Lys Val Leu Phe Glu Pro Val Pro
865 870 875 880

Glu

<210> 138
<211> 513
<212> PRT
<213> Stevia rebaudiana

<400> 138
Met Asp Ala Val Thr Gly Leu Leu Thr Val Pro Ala Thr Ala Ile Thr
1 5 10 15

Ile Gly Gly Thr Ala Val Ala Leu Ala Val Ala Leu Ile Phe Trp Tyr
20 25 30

Leu Lys Ser Tyr Thr Ser Ala Arg Arg Ser Gln Ser Asn His Leu Pro
35 40 45

Arg Val Pro Glu Val Pro Gly Val Pro Leu Leu Gly Asn Leu Leu Gln
50 55 60

Leu Lys Glu Lys Lys Pro Tyr Met Thr Phe Thr Arg Trp Ala Ala Thr
65 70 75 80

Tyr Gly Pro Ile Tyr Ser Ile Lys Thr Gly Ala Thr Ser Met Val Val
85 90 95

Val Ser Ser Asn Glu Ile Ala Lys Glu Ala Leu Val Thr Arg Phe Gln
100 105 110

Ser Ile Ser Thr Arg Asn Leu Ser Lys Ala Leu Lys Val Leu Thr Ala
115 120 125

Asp Lys Thr Met Val Ala Met Ser Asp Tyr Asp Asp Tyr His Lys Thr
130 135 140

Val Lys Arg His Ile Leu Thr Ala Val Leu Gly Pro Asn Ala Gln Lys
145 150 155 160

Lys His Arg Ile His Arg Asp Ile Met Met Asp Asn Ile Ser Thr Gln
165 170 175

```
Leu His Glu Phe Val Lys Asn Asn Pro Glu Gln Glu Glu Val Asp Leu
            180                 185                 190

Arg Lys Ile Phe Gln Ser Glu Leu Phe Gly Leu Ala Met Arg Gln Ala
        195                 200                 205

Leu Gly Lys Asp Val Glu Ser Leu Tyr Val Glu Asp Leu Lys Ile Thr
    210                 215                 220

Met Asn Arg Asp Glu Ile Phe Gln Val Leu Val Val Asp Pro Met Met
225                 230                 235                 240

Gly Ala Ile Asp Val Asp Trp Arg Asp Phe Phe Pro Tyr Leu Lys Trp
                245                 250                 255

Val Pro Asn Lys Lys Phe Glu Asn Thr Ile Gln Gln Met Tyr Ile Arg
            260                 265                 270

Arg Glu Ala Val Met Lys Ser Leu Ile Lys Glu His Lys Lys Arg Ile
        275                 280                 285

Ala Ser Gly Glu Lys Leu Asn Ser Tyr Ile Asp Tyr Leu Leu Ser Glu
    290                 295                 300

Ala Gln Thr Leu Thr Asp Gln Gln Leu Leu Met Ser Leu Trp Glu Pro
305                 310                 315                 320

Ile Ile Glu Ser Ser Asp Thr Thr Met Val Thr Thr Glu Trp Ala Met
                325                 330                 335

Tyr Glu Leu Ala Lys Asn Pro Lys Leu Gln Asp Arg Leu Tyr Arg Asp
            340                 345                 350

Ile Lys Ser Val Cys Gly Ser Glu Lys Ile Thr Glu Glu His Leu Ser
        355                 360                 365

Gln Leu Pro Tyr Ile Thr Ala Ile Phe His Glu Thr Leu Arg Arg His
    370                 375                 380

Ser Pro Val Pro Ile Ile Pro Leu Arg His Val His Glu Asp Thr Val
385                 390                 395                 400

Leu Gly Gly Tyr His Val Pro Ala Gly Thr Glu Leu Ala Val Asn Ile
                405                 410                 415

Tyr Gly Cys Asn Met Asp Lys Asn Val Trp Glu Asn Pro Glu Glu Trp
```

420 425 430

Asn Pro Glu Arg Phe Met Lys Glu Asn Glu Thr Ile Asp Phe Gln Lys
435 440 445

Thr Met Ala Phe Gly Gly Gly Lys Arg Val Cys Ala Gly Ser Leu Gln
450 455 460

Ala Leu Leu Thr Ala Ser Ile Gly Ile Gly Arg Met Val Gln Glu Phe
465 470 475 480

Glu Trp Lys Leu Lys Asp Met Thr Gln Glu Glu Val Asn Thr Ile Gly
485 490 495

Leu Thr Thr Gln Met Leu Arg Pro Leu Arg Ala Ile Ile Lys Pro Arg
500 505 510

Ile

<210> 139
<211> 509
<212> PRT
<213> Arabidopsis thaliana

<400> 139
Met Ala Phe Phe Ser Met Ile Ser Ile Leu Leu Gly Phe Val Ile Ser
1 5 10 15

Ser Phe Ile Phe Ile Phe Phe Phe Lys Lys Leu Leu Ser Phe Ser Arg
20 25 30

Lys Asn Met Ser Glu Val Ser Thr Leu Pro Ser Val Pro Val Val Pro
35 40 45

Gly Phe Pro Val Ile Gly Asn Leu Leu Gln Leu Lys Glu Lys Lys Pro
50 55 60

His Lys Thr Phe Thr Arg Trp Ser Glu Ile Tyr Gly Pro Ile Tyr Ser
65 70 75 80

Ile Lys Met Gly Ser Ser Ser Leu Ile Val Leu Asn Ser Thr Glu Thr
85 90 95

Ala Lys Glu Ala Met Val Thr Arg Phe Ser Ser Ile Ser Thr Arg Lys
100 105 110

Leu Ser Asn Ala Leu Thr Val Leu Thr Cys Asp Lys Ser Met Val Ala
115 120 125

```
Thr Ser Asp Tyr Asp Asp Phe His Lys Leu Val Lys Arg Cys Leu Leu
    130                 135                 140

Asn Gly Leu Leu Gly Ala Asn Ala Gln Lys Arg Lys Arg His Tyr Arg
145                 150                 155                 160

Asp Ala Leu Ile Glu Asn Val Ser Ser Lys Leu His Ala His Ala Arg
                165                 170                 175

Asp His Pro Gln Glu Pro Val Asn Phe Arg Ala Ile Phe Glu His Glu
            180                 185                 190

Leu Phe Gly Val Ala Leu Lys Gln Ala Phe Gly Lys Asp Val Glu Ser
        195                 200                 205

Ile Tyr Val Lys Glu Leu Gly Val Thr Leu Ser Lys Asp Glu Ile Phe
    210                 215                 220

Lys Val Leu Val His Asp Met Met Glu Gly Ala Ile Asp Val Asp Trp
225                 230                 235                 240

Arg Asp Phe Phe Pro Tyr Leu Lys Trp Ile Pro Asn Lys Ser Phe Glu
                245                 250                 255

Ala Arg Ile Gln Gln Lys His Lys Arg Arg Leu Ala Val Met Asn Ala
            260                 265                 270

Leu Ile Gln Asp Arg Leu Lys Gln Asn Gly Ser Glu Ser Asp Asp Asp
        275                 280                 285

Cys Tyr Leu Asn Phe Leu Met Ser Glu Ala Lys Thr Leu Thr Lys Glu
    290                 295                 300

Gln Ile Ala Ile Leu Val Trp Glu Thr Ile Ile Glu Thr Ala Asp Thr
305                 310                 315                 320

Thr Leu Val Thr Thr Glu Trp Ala Ile Tyr Glu Leu Ala Lys His Pro
                325                 330                 335

Ser Val Gln Asp Arg Leu Cys Lys Glu Ile Gln Asn Val Cys Gly Gly
            340                 345                 350

Glu Lys Phe Lys Glu Glu Gln Leu Ser Gln Val Pro Tyr Leu Asn Gly
        355                 360                 365

Val Phe His Glu Thr Leu Arg Lys Tyr Ser Pro Ala Pro Leu Val Pro
```

370 375 380

Ile Arg Tyr Ala His Glu Asp Thr Gln Ile Gly Gly Tyr His Val Pro
385 390 395 400

Ala Gly Ser Glu Ile Ala Ile Asn Ile Tyr Gly Cys Asn Met Asp Lys
405 410 415

Lys Arg Trp Glu Arg Pro Glu Asp Trp Trp Pro Glu Arg Phe Leu Asp
420 425 430

Asp Gly Lys Tyr Glu Thr Ser Asp Leu His Lys Thr Met Ala Phe Gly
435 440 445

Ala Gly Lys Arg Val Cys Ala Gly Ala Leu Gln Ala Ser Leu Met Ala
450 455 460

Gly Ile Ala Ile Gly Arg Leu Val Gln Glu Phe Glu Trp Lys Leu Arg
465 470 475 480

Asp Gly Glu Glu Glu Asn Val Asp Thr Tyr Gly Leu Thr Ser Gln Lys
485 490 495

Leu Tyr Pro Leu Met Ala Ile Ile Asn Pro Arg Arg Ser
500 505

<210> 140
<211> 525
<212> PRT
<213> Gibberella fujikoroi

<400> 140
Met Ser Lys Ser Asn Ser Met Asn Ser Thr Ser His Glu Thr Leu Phe
1 5 10 15

Gln Gln Leu Val Leu Gly Leu Asp Arg Met Pro Leu Met Asp Val His
20 25 30

Trp Leu Ile Tyr Val Ala Phe Gly Ala Trp Leu Cys Ser Tyr Val Ile
35 40 45

His Val Leu Ser Ser Ser Ser Thr Val Lys Val Pro Val Val Gly Tyr
50 55 60

Arg Ser Val Phe Glu Pro Thr Trp Leu Leu Arg Leu Arg Phe Val Trp
65 70 75 80

Glu Gly Gly Ser Ile Ile Gly Gln Gly Tyr Asn Lys Phe Lys Asp Ser
85 90 95

```
Ile Phe Gln Val Arg Lys Leu Gly Thr Asp Ile Val Ile Ile Pro Pro
            100                 105                 110

Asn Tyr Ile Asp Glu Val Arg Lys Leu Ser Gln Asp Lys Thr Arg Ser
        115                 120                 125

Val Glu Pro Phe Ile Asn Asp Phe Ala Gly Gln Tyr Thr Arg Gly Met
    130                 135                 140

Val Phe Leu Gln Ser Asp Leu Gln Asn Arg Val Ile Gln Gln Arg Leu
145                 150                 155                 160

Thr Pro Lys Leu Val Ser Leu Thr Lys Val Met Lys Glu Glu Leu Asp
                165                 170                 175

Tyr Ala Leu Thr Lys Glu Met Pro Asp Met Lys Asn Asp Glu Trp Val
            180                 185                 190

Glu Val Asp Ile Ser Ser Ile Met Val Arg Leu Ile Ser Arg Ile Ser
        195                 200                 205

Ala Arg Val Phe Leu Gly Pro Glu His Cys Arg Asn Gln Glu Trp Leu
    210                 215                 220

Thr Thr Thr Ala Glu Tyr Ser Glu Ser Leu Phe Ile Thr Gly Phe Ile
225                 230                 235                 240

Leu Arg Val Val Pro His Ile Leu Arg Pro Phe Ile Ala Pro Leu Leu
                245                 250                 255

Pro Ser Tyr Arg Thr Leu Leu Arg Asn Val Ser Ser Gly Arg Arg Val
            260                 265                 270

Ile Gly Asp Ile Ile Arg Ser Gln Gln Gly Asp Gly Asn Glu Asp Ile
        275                 280                 285

Leu Ser Trp Met Arg Asp Ala Ala Thr Gly Glu Glu Lys Gln Ile Asp
    290                 295                 300

Asn Ile Ala Gln Arg Met Leu Ile Leu Ser Leu Ala Ser Ile His Thr
305                 310                 315                 320

Thr Ala Met Thr Met Thr His Ala Met Tyr Asp Leu Cys Ala Cys Pro
                325                 330                 335

Glu Tyr Ile Glu Pro Leu Arg Asp Glu Val Lys Ser Val Val Gly Ala
```

340 345 350

Ser Gly Trp Asp Lys Thr Ala Leu Asn Arg Phe His Lys Leu Asp Ser
355 360 365

Phe Leu Lys Glu Ser Gln Arg Phe Asn Pro Val Phe Leu Leu Thr Phe
370 375 380

Asn Arg Ile Tyr His Gln Ser Met Thr Leu Ser Asp Gly Thr Asn Ile
385 390 395 400

Pro Ser Gly Thr Arg Ile Ala Val Pro Ser His Ala Met Leu Gln Asp
405 410 415

Ser Ala His Val Pro Gly Pro Thr Pro Pro Thr Glu Phe Asp Gly Phe
420 425 430

Arg Tyr Ser Lys Ile Arg Ser Asp Ser Asn Tyr Ala Gln Lys Tyr Leu
435 440 445

Phe Ser Met Thr Asp Ser Ser Asn Met Ala Phe Gly Tyr Gly Lys Tyr
450 455 460

Ala Cys Pro Gly Arg Phe Tyr Ala Ser Asn Glu Met Lys Leu Thr Leu
465 470 475 480

Ala Ile Leu Leu Leu Gln Phe Glu Phe Lys Leu Pro Asp Gly Lys Gly
485 490 495

Arg Pro Arg Asn Ile Thr Ile Asp Ser Asp Met Ile Pro Asp Pro Arg
500 505 510

Ala Arg Leu Cys Val Arg Lys Arg Ser Leu Arg Asp Glu
515 520 525

<210> 141
<211> 499
<212> PRT
<213> Trametes versicolor

<400> 141
Met Glu Asp Pro Thr Val Leu Tyr Ala Cys Leu Ala Ile Ala Val Ala
1 5 10 15

Thr Phe Val Val Arg Trp Tyr Arg Asp Pro Leu Arg Ser Ile Pro Thr
20 25 30

Val Gly Gly Ser Asp Leu Pro Ile Leu Ser Tyr Ile Gly Ala Leu Arg
35 40 45

```
Trp Thr Arg Arg Gly Arg Glu Ile Leu Gln Glu Gly Tyr Asp Gly Tyr
    50                  55                  60

Arg Gly Ser Thr Phe Lys Ile Ala Met Leu Asp Arg Trp Ile Val Ile
65                  70                  75                  80

Ala Asn Gly Pro Lys Leu Ala Asp Glu Val Arg Arg Arg Pro Asp Glu
                85                  90                  95

Glu Leu Asn Phe Met Asp Gly Leu Gly Ala Phe Val Gln Thr Lys Tyr
            100                 105                 110

Thr Leu Gly Glu Ala Ile His Asn Asp Pro Tyr His Val Asp Ile Ile
        115                 120                 125

Arg Glu Lys Leu Thr Arg Gly Leu Pro Ala Val Leu Pro Asp Val Ile
    130                 135                 140

Glu Glu Leu Thr Leu Ala Val Arg Gln Tyr Ile Pro Thr Glu Gly Asp
145                 150                 155                 160

Glu Trp Val Ser Val Asn Cys Ser Lys Ala Ala Arg Asp Ile Val Ala
                165                 170                 175

Arg Ala Ser Asn Arg Val Phe Val Gly Leu Pro Ala Cys Arg Asn Gln
            180                 185                 190

Gly Tyr Leu Asp Leu Ala Ile Asp Phe Thr Leu Ser Val Val Lys Asp
        195                 200                 205

Arg Ala Ile Ile Asn Met Phe Pro Glu Leu Leu Lys Pro Ile Val Gly
    210                 215                 220

Arg Val Val Gly Asn Ala Thr Arg Asn Val Arg Arg Ala Val Pro Phe
225                 230                 235                 240

Val Ala Pro Leu Val Glu Glu Arg Arg Arg Leu Met Glu Glu Tyr Gly
                245                 250                 255

Glu Asp Trp Ser Glu Lys Pro Asn Asp Met Leu Gln Trp Ile Met Asp
            260                 265                 270

Glu Ala Ala Ser Arg Asp Ser Ser Val Lys Ala Ile Ala Glu Arg Leu
        275                 280                 285

Leu Met Val Asn Phe Ala Ala Ile His Thr Ser Ser Asn Thr Ile Thr
```

290 295 300

His Ala Leu Tyr His Leu Ala Glu Met Pro Glu Thr Leu Gln Pro Leu
305 310 315 320

Arg Glu Glu Ile Glu Pro Leu Val Lys Glu Glu Gly Trp Thr Lys Ala
325 330 335

Ala Met Gly Lys Met Trp Trp Leu Asp Ser Phe Leu Arg Glu Ser Gln
340 345 350

Arg Tyr Asn Gly Ile Asn Ile Val Ser Leu Thr Arg Met Ala Asp Lys
355 360 365

Asp Ile Thr Leu Ser Asp Gly Thr Phe Leu Pro Lys Gly Thr Leu Val
370 375 380

Ala Val Pro Ala Tyr Ser Thr His Arg Asp Asp Ala Val Tyr Ala Asp
385 390 395 400

Ala Leu Val Phe Asp Pro Phe Arg Phe Ser Arg Met Arg Ala Arg Glu
405 410 415

Gly Glu Gly Thr Lys His Gln Phe Val Asn Thr Ser Val Glu Tyr Val
420 425 430

Pro Phe Gly His Gly Lys His Ala Cys Pro Gly Arg Phe Phe Ala Ala
435 440 445

Asn Glu Leu Lys Ala Met Leu Ala Tyr Ile Val Leu Asn Tyr Asp Val
450 455 460

Lys Leu Pro Gly Asp Gly Lys Arg Pro Leu Asn Met Tyr Trp Gly Pro
465 470 475 480

Thr Val Leu Pro Ala Pro Ala Gly Gln Val Leu Phe Arg Lys Arg Gln
485 490 495

Val Ser Leu

<210> 142
<211> 525
<212> PRT
<213> Stevia rebaudiana

<400> 142
Met Gly Leu Phe Pro Leu Glu Asp Ser Tyr Ala Leu Val Phe Glu Gly
1 5 10 15

```
Leu Ala Ile Thr Leu Ala Leu Tyr Tyr Leu Leu Ser Phe Ile Tyr Lys
            20                  25                  30

Thr Ser Lys Lys Thr Cys Thr Pro Pro Lys Ala Ser Gly Glu Ile Ile
        35                  40                  45

Pro Ile Thr Gly Ile Ile Leu Asn Leu Leu Ser Gly Ser Ser Gly Leu
    50                  55                  60

Pro Ile Ile Leu Ala Leu Ala Ser Leu Ala Asp Arg Cys Gly Pro Ile
65                  70                  75                  80

Phe Thr Ile Arg Leu Gly Ile Arg Arg Val Leu Val Val Ser Asn Trp
                85                  90                  95

Glu Ile Ala Lys Glu Ile Phe Thr Thr His Asp Leu Ile Val Ser Asn
            100                 105                 110

Arg Pro Lys Tyr Leu Ala Ala Lys Ile Leu Gly Phe Asn Tyr Val Ser
        115                 120                 125

Phe Ser Phe Ala Pro Tyr Gly Pro Tyr Trp Val Gly Ile Arg Lys Ile
    130                 135                 140

Ile Ala Thr Lys Leu Met Ser Ser Ser Arg Leu Gln Lys Leu Gln Phe
145                 150                 155                 160

Val Arg Val Phe Glu Leu Glu Asn Ser Met Lys Ser Ile Arg Glu Ser
                165                 170                 175

Trp Lys Glu Lys Lys Asp Glu Glu Gly Lys Val Leu Val Glu Met Lys
            180                 185                 190

Lys Trp Phe Trp Glu Leu Asn Met Asn Ile Val Leu Arg Thr Val Ala
        195                 200                 205

Gly Lys Gln Tyr Thr Gly Thr Val Asp Asp Ala Asp Ala Lys Arg Ile
    210                 215                 220

Ser Glu Leu Phe Arg Glu Trp Phe His Tyr Thr Gly Arg Phe Val Val
225                 230                 235                 240

Gly Asp Ala Phe Pro Phe Leu Gly Trp Leu Asp Leu Gly Gly Tyr Lys
                245                 250                 255

Lys Thr Met Glu Leu Val Ala Ser Arg Leu Asp Ser Met Val Ser Lys
```

```
            260                 265                 270

Trp Leu Asp Glu His Arg Lys Lys Gln Ala Asn Asp Asp Lys Lys Glu
        275                 280                 285

Asp Met Asp Phe Met Asp Ile Met Ile Ser Met Thr Glu Ala Asn Ser
    290                 295                 300

Pro Leu Glu Gly Tyr Gly Thr Asp Thr Ile Ile Lys Thr Thr Cys Met
305                 310                 315                 320

Thr Leu Ile Val Ser Gly Val Asp Thr Thr Ser Ile Val Leu Thr Trp
                325                 330                 335

Ala Leu Ser Leu Leu Leu Asn Asn Arg Asp Thr Leu Lys Lys Ala Gln
            340                 345                 350

Glu Glu Leu Asp Met Cys Val Gly Lys Gly Arg Gln Val Asn Glu Ser
        355                 360                 365

Asp Leu Val Asn Leu Ile Tyr Leu Glu Ala Val Leu Lys Glu Ala Leu
    370                 375                 380

Arg Leu Tyr Pro Ala Ala Phe Leu Gly Gly Pro Arg Ala Phe Leu Glu
385                 390                 395                 400

Asp Cys Thr Val Ala Gly Tyr Arg Ile Pro Lys Gly Thr Cys Leu Leu
                405                 410                 415

Ile Asn Met Trp Lys Leu His Arg Asp Pro Asn Ile Trp Ser Asp Pro
            420                 425                 430

Cys Glu Phe Lys Pro Glu Arg Phe Leu Thr Pro Asn Gln Lys Asp Val
        435                 440                 445

Asp Val Ile Gly Met Asp Phe Glu Leu Ile Pro Phe Gly Ala Gly Arg
    450                 455                 460

Arg Tyr Cys Pro Gly Thr Arg Leu Ala Leu Gln Met Leu His Ile Val
465                 470                 475                 480

Leu Ala Thr Leu Leu Gln Asn Phe Glu Met Ser Thr Pro Asn Asp Ala
                485                 490                 495

Pro Val Asp Met Thr Ala Ser Val Gly Met Thr Asn Ala Lys Ala Ser
            500                 505                 510
```

Pro Leu Glu Val Leu Leu Ser Pro Arg Val Lys Trp Ser
515 520 525

<210> 143
<211> 476
<212> PRT
<213> Stevia rebaudiana

<400> 143
Met Ile Gln Val Leu Thr Pro Ile Leu Leu Phe Leu Ile Phe Phe Val
1 5 10 15

Phe Trp Lys Val Tyr Lys His Gln Lys Thr Lys Ile Asn Leu Pro Pro
20 25 30

Gly Ser Phe Gly Trp Pro Phe Leu Gly Glu Thr Leu Ala Leu Leu Arg
35 40 45

Ala Gly Trp Asp Ser Glu Pro Glu Arg Phe Val Arg Glu Arg Ile Lys
50 55 60

Lys His Gly Ser Pro Leu Val Phe Lys Thr Ser Leu Phe Gly Asp Arg
65 70 75 80

Phe Ala Val Leu Cys Gly Pro Ala Gly Asn Lys Phe Leu Phe Cys Asn
85 90 95

Glu Asn Lys Leu Val Ala Ser Trp Trp Pro Val Pro Val Arg Lys Leu
100 105 110

Phe Gly Lys Ser Leu Leu Thr Ile Arg Gly Asp Glu Ala Lys Trp Met
115 120 125

Arg Lys Met Leu Leu Ser Tyr Leu Gly Pro Asp Ala Phe Ala Thr His
130 135 140

Tyr Ala Val Thr Met Asp Val Val Thr Arg Arg His Ile Asp Val His
145 150 155 160

Trp Arg Gly Lys Glu Glu Val Asn Val Phe Gln Thr Val Lys Leu Tyr
165 170 175

Ala Phe Glu Leu Ala Cys Arg Leu Phe Met Asn Leu Asp Asp Pro Asn
180 185 190

His Ile Ala Lys Leu Gly Ser Leu Phe Asn Ile Phe Leu Lys Gly Ile
195 200 205

Ile Glu Leu Pro Ile Asp Val Pro Gly Thr Arg Phe Tyr Ser Ser Lys

210 215 220

Lys Ala Ala Ala Ala Ile Arg Ile Glu Leu Lys Lys Leu Ile Lys Ala
225 230 235 240

Arg Lys Leu Glu Leu Lys Glu Gly Lys Ala Ser Ser Ser Gln Asp Leu
245 250 255

Leu Ser His Leu Leu Thr Ser Pro Asp Glu Asn Gly Met Phe Leu Thr
260 265 270

Glu Glu Glu Ile Val Asp Asn Ile Leu Leu Leu Leu Phe Ala Gly His
275 280 285

Asp Thr Ser Ala Leu Ser Ile Thr Leu Leu Met Lys Thr Leu Gly Glu
290 295 300

His Ser Asp Val Tyr Asp Lys Val Leu Lys Glu Gln Leu Glu Ile Ser
305 310 315 320

Lys Thr Lys Glu Ala Trp Glu Ser Leu Lys Trp Glu Asp Ile Gln Lys
325 330 335

Met Lys Tyr Ser Trp Ser Val Ile Cys Glu Val Met Arg Leu Asn Pro
340 345 350

Pro Val Ile Gly Thr Tyr Arg Glu Ala Leu Val Asp Ile Asp Tyr Ala
355 360 365

Gly Tyr Thr Ile Pro Lys Gly Trp Lys Leu His Trp Ser Ala Val Ser
370 375 380

Thr Gln Arg Asp Glu Ala Asn Phe Glu Asp Val Thr Arg Phe Asp Pro
385 390 395 400

Ser Arg Phe Glu Gly Ala Gly Pro Thr Pro Phe Thr Phe Val Pro Phe
405 410 415

Gly Gly Gly Pro Arg Met Cys Leu Gly Lys Glu Phe Ala Arg Leu Glu
420 425 430

Val Leu Ala Phe Leu His Asn Ile Val Thr Asn Phe Lys Trp Asp Leu
435 440 445

Leu Ile Pro Asp Glu Lys Ile Glu Tyr Asp Pro Met Ala Thr Pro Ala
450 455 460

Lys Gly Leu Pro Ile Arg Leu His Pro His Gln Val
465 470 475

<210> 144
<211> 525
<212> PRT
<213> Arabidopsis thaliana

<400> 144
Met Glu Ser Leu Val Val His Thr Val Asn Ala Ile Trp Cys Ile Val
1 5 10 15

Ile Val Gly Ile Phe Ser Val Gly Tyr His Val Tyr Gly Arg Ala Val
20 25 30

Val Glu Gln Trp Arg Met Arg Arg Ser Leu Lys Leu Gln Gly Val Lys
35 40 45

Gly Pro Pro Pro Ser Ile Phe Asn Gly Asn Val Ser Glu Met Gln Arg
50 55 60

Ile Gln Ser Glu Ala Lys His Cys Ser Gly Asp Asn Ile Ile Ser His
65 70 75 80

Asp Tyr Ser Ser Ser Leu Phe Pro His Phe Asp His Trp Arg Lys Gln
85 90 95

Tyr Gly Arg Ile Tyr Thr Tyr Ser Thr Gly Leu Lys Gln His Leu Tyr
100 105 110

Ile Asn His Pro Glu Met Val Lys Glu Leu Ser Gln Thr Asn Thr Leu
115 120 125

Asn Leu Gly Arg Ile Thr His Ile Thr Lys Arg Leu Asn Pro Ile Leu
130 135 140

Gly Asn Gly Ile Ile Thr Ser Asn Gly Pro His Trp Ala His Gln Arg
145 150 155 160

Arg Ile Ile Ala Tyr Glu Phe Thr His Asp Lys Ile Lys Gly Met Val
165 170 175

Gly Leu Met Val Glu Ser Ala Met Pro Met Leu Asn Lys Trp Glu Glu
180 185 190

Met Val Lys Arg Gly Gly Glu Met Gly Cys Asp Ile Arg Val Asp Glu
195 200 205

Asp Leu Lys Asp Val Ser Ala Asp Val Ile Ala Lys Ala Cys Phe Gly

210 215 220

Ser Ser Phe Ser Lys Gly Lys Ala Ile Phe Ser Met Ile Arg Asp Leu
225 230 235 240

Leu Thr Ala Ile Thr Lys Arg Ser Val Leu Phe Arg Phe Asn Gly Phe
245 250 255

Thr Asp Met Val Phe Gly Ser Lys Lys His Gly Asp Val Asp Ile Asp
260 265 270

Ala Leu Glu Met Glu Leu Glu Ser Ser Ile Trp Glu Thr Val Lys Glu
275 280 285

Arg Glu Ile Glu Cys Lys Asp Thr His Lys Lys Asp Leu Met Gln Leu
290 295 300

Ile Leu Glu Gly Ala Met Arg Ser Cys Asp Gly Asn Leu Trp Asp Lys
305 310 315 320

Ser Ala Tyr Arg Arg Phe Val Val Asp Asn Cys Lys Ser Ile Tyr Phe
325 330 335

Ala Gly His Asp Ser Thr Ala Val Ser Val Ser Trp Cys Leu Met Leu
340 345 350

Leu Ala Leu Asn Pro Ser Trp Gln Val Lys Ile Arg Asp Glu Ile Leu
355 360 365

Ser Ser Cys Lys Asn Gly Ile Pro Asp Ala Glu Ser Ile Pro Asn Leu
370 375 380

Lys Thr Val Thr Met Val Ile Gln Glu Thr Met Arg Leu Tyr Pro Pro
385 390 395 400

Ala Pro Ile Val Gly Arg Glu Ala Ser Lys Asp Ile Arg Leu Gly Asp
405 410 415

Leu Val Val Pro Lys Gly Val Cys Ile Trp Thr Leu Ile Pro Ala Leu
420 425 430

His Arg Asp Pro Glu Ile Trp Gly Pro Asp Ala Asn Asp Phe Lys Pro
435 440 445

Glu Arg Phe Ser Glu Gly Ile Ser Lys Ala Cys Lys Tyr Pro Gln Ser
450 455 460

Tyr Ile Pro Phe Gly Leu Gly Pro Arg Thr Cys Val Gly Lys Asn Phe
465 470 475 480

Gly Met Met Glu Val Lys Val Leu Val Ser Leu Ile Val Ser Lys Phe
485 490 495

Ser Phe Thr Leu Ser Pro Thr Tyr Gln His Ser Pro Ser His Lys Leu
500 505 510

Leu Val Glu Pro Gln His Gly Val Val Ile Arg Val Val
515 520 525

<210> 145
<211> 529
<212> PRT
<213> Vitis vinifera

<400> 145
Met Tyr Phe Leu Leu Gln Tyr Leu Asn Ile Thr Thr Val Gly Val Phe
1 5 10 15

Ala Thr Leu Phe Leu Ser Tyr Cys Leu Leu Leu Trp Arg Ser Arg Ala
20 25 30

Gly Asn Lys Lys Ile Ala Pro Glu Ala Ala Ala Ala Trp Pro Ile Ile
35 40 45

Gly His Leu His Leu Leu Ala Gly Gly Ser His Gln Leu Pro His Ile
50 55 60

Thr Leu Gly Asn Met Ala Asp Lys Tyr Gly Pro Val Phe Thr Ile Arg
65 70 75 80

Ile Gly Leu His Arg Ala Val Val Val Ser Ser Trp Glu Met Ala Lys
85 90 95

Glu Cys Ser Thr Ala Asn Asp Gln Val Ser Ser Ser Arg Pro Glu Leu
100 105 110

Leu Ala Ser Lys Leu Leu Gly Tyr Asn Tyr Ala Met Phe Gly Phe Ser
115 120 125

Pro Tyr Gly Ser Tyr Trp Arg Glu Met Arg Lys Ile Ile Ser Leu Glu
130 135 140

Leu Leu Ser Asn Ser Arg Leu Glu Leu Leu Lys Asp Val Arg Ala Ser
145 150 155 160

Glu Val Val Thr Ser Ile Lys Glu Leu Tyr Lys Leu Trp Ala Glu Lys

```
                165                 170                 175

Lys Asn Glu Ser Gly Leu Val Ser Val Glu Met Lys Gln Trp Phe Gly
            180                 185                 190

Asp Leu Thr Leu Asn Val Ile Leu Arg Met Val Ala Gly Lys Arg Tyr
        195                 200                 205

Phe Ser Ala Ser Asp Ala Ser Glu Asn Lys Gln Ala Gln Arg Cys Arg
    210                 215                 220

Arg Val Phe Arg Glu Phe Phe His Leu Ser Gly Leu Phe Val Val Ala
225                 230                 235                 240

Asp Ala Ile Pro Phe Leu Gly Trp Leu Asp Trp Gly Arg His Glu Lys
                245                 250                 255

Thr Leu Lys Lys Thr Ala Ile Glu Met Asp Ser Ile Ala Gln Glu Trp
            260                 265                 270

Leu Glu Glu His Arg Arg Arg Lys Asp Ser Gly Asp Asp Asn Ser Thr
        275                 280                 285

Gln Asp Phe Met Asp Val Met Gln Ser Val Leu Asp Gly Lys Asn Leu
    290                 295                 300

Gly Gly Tyr Asp Ala Asp Thr Ile Asn Lys Ala Thr Cys Leu Thr Leu
305                 310                 315                 320

Ile Ser Gly Gly Ser Asp Thr Thr Val Val Ser Leu Thr Trp Ala Leu
                325                 330                 335

Ser Leu Val Leu Asn Asn Arg Asp Thr Leu Lys Lys Ala Gln Glu Glu
            340                 345                 350

Leu Asp Ile Gln Val Gly Lys Glu Arg Leu Val Asn Glu Gln Asp Ile
        355                 360                 365

Ser Lys Leu Val Tyr Leu Gln Ala Ile Val Lys Glu Thr Leu Arg Leu
    370                 375                 380

Tyr Pro Pro Gly Pro Leu Gly Gly Leu Arg Gln Phe Thr Glu Asp Cys
385                 390                 395                 400

Thr Leu Gly Gly Tyr His Val Ser Lys Gly Thr Arg Leu Ile Met Asn
                405                 410                 415
```

Leu Ser Lys Ile Gln Lys Asp Pro Arg Ile Trp Ser Asp Pro Thr Glu
420 425 430

Phe Gln Pro Glu Arg Phe Leu Thr Thr His Lys Asp Val Asp Pro Arg
435 440 445

Gly Lys His Phe Glu Phe Ile Pro Phe Gly Ala Gly Arg Arg Ala Cys
450 455 460

Pro Gly Ile Thr Phe Gly Leu Gln Val Leu His Leu Thr Leu Ala Ser
465 470 475 480

Phe Leu His Ala Phe Glu Phe Ser Thr Pro Ser Asn Glu Gln Val Asn
485 490 495

Met Arg Glu Ser Leu Gly Leu Thr Asn Met Lys Ser Thr Pro Leu Glu
500 505 510

Val Leu Ile Ser Pro Arg Leu Ser Leu Asn Cys Phe Asn Leu Met Lys
515 520 525

Ile

<210> 146
<211> 479
<212> PRT
<213> Medicago trunculata

<400> 146
Met Glu Pro Asn Phe Tyr Leu Ser Leu Leu Leu Leu Phe Val Thr Phe
1 5 10 15

Ile Ser Leu Ser Leu Phe Phe Ile Phe Tyr Lys Gln Lys Ser Pro Leu
20 25 30

Asn Leu Pro Pro Gly Lys Met Gly Tyr Pro Ile Ile Gly Glu Ser Leu
35 40 45

Glu Phe Leu Ser Thr Gly Trp Lys Gly His Pro Glu Lys Phe Ile Phe
50 55 60

Asp Arg Met Arg Lys Tyr Ser Ser Glu Leu Phe Lys Thr Ser Ile Val
65 70 75 80

Gly Glu Ser Thr Val Val Cys Cys Gly Ala Ala Ser Asn Lys Phe Leu
85 90 95

Phe Ser Asn Glu Asn Lys Leu Val Thr Ala Trp Trp Pro Asp Ser Val

```
            100                 105                 110

Asn Lys Ile Phe Pro Thr Thr Ser Leu Asp Ser Asn Leu Lys Glu Glu
        115                 120                 125

Ser Ile Lys Met Arg Lys Leu Leu Pro Gln Phe Phe Lys Pro Glu Ala
    130                 135                 140

Leu Gln Arg Tyr Val Gly Val Met Asp Val Ile Ala Gln Arg His Phe
145                 150                 155                 160

Val Thr His Trp Asp Asn Lys Asn Glu Ile Thr Val Tyr Pro Leu Ala
                165                 170                 175

Lys Arg Tyr Thr Phe Leu Leu Ala Cys Arg Leu Phe Met Ser Val Glu
            180                 185                 190

Asp Glu Asn His Val Ala Lys Phe Ser Asp Pro Phe Gln Leu Ile Ala
        195                 200                 205

Ala Gly Ile Ile Ser Leu Pro Ile Asp Leu Pro Gly Thr Pro Phe Asn
    210                 215                 220

Lys Ala Ile Lys Ala Ser Asn Phe Ile Arg Lys Glu Leu Ile Lys Ile
225                 230                 235                 240

Ile Lys Gln Arg Arg Val Asp Leu Ala Glu Gly Thr Ala Ser Pro Thr
                245                 250                 255

Gln Asp Ile Leu Ser His Met Leu Leu Thr Ser Asp Glu Asn Gly Lys
            260                 265                 270

Ser Met Asn Glu Leu Asn Ile Ala Asp Lys Ile Leu Gly Leu Leu Ile
        275                 280                 285

Gly Gly His Asp Thr Ala Ser Val Ala Cys Thr Phe Leu Val Lys Tyr
    290                 295                 300

Leu Gly Glu Leu Pro His Ile Tyr Asp Lys Val Tyr Gln Glu Gln Met
305                 310                 315                 320

Glu Ile Ala Lys Ser Lys Pro Ala Gly Glu Leu Leu Asn Trp Asp Asp
                325                 330                 335

Leu Lys Lys Met Lys Tyr Ser Trp Asn Val Ala Cys Glu Val Met Arg
            340                 345                 350
```

Leu Ser Pro Pro Leu Gln Gly Gly Phe Arg Glu Ala Ile Thr Asp Phe
355 360 365

Met Phe Asn Gly Phe Ser Ile Pro Lys Gly Trp Lys Leu Tyr Trp Ser
370 375 380

Ala Asn Ser Thr His Lys Asn Ala Glu Cys Phe Pro Met Pro Glu Lys
385 390 395 400

Phe Asp Pro Thr Arg Phe Glu Gly Asn Gly Pro Ala Pro Tyr Thr Phe
405 410 415

Val Pro Phe Gly Gly Gly Pro Arg Met Cys Pro Gly Lys Glu Tyr Ala
420 425 430

Arg Leu Glu Ile Leu Val Phe Met His Asn Leu Val Lys Arg Phe Lys
435 440 445

Trp Glu Lys Val Ile Pro Asp Glu Lys Ile Ile Val Asp Pro Phe Pro
450 455 460

Ile Pro Ala Lys Asp Leu Pro Ile Arg Leu Tyr Pro His Lys Ala
465 470 475

<210> 147
<211> 710
<212> PRT
<213> Stevia rebaudiana

<400> 147
Met Gln Ser Asp Ser Val Lys Val Ser Pro Phe Asp Leu Val Ser Ala
1 5 10 15

Ala Met Asn Gly Lys Ala Met Glu Lys Leu Asn Ala Ser Glu Ser Glu
20 25 30

Asp Pro Thr Thr Leu Pro Ala Leu Lys Met Leu Val Glu Asn Arg Glu
35 40 45

Leu Leu Thr Leu Phe Thr Thr Ser Phe Ala Val Leu Ile Gly Cys Leu
50 55 60

Val Phe Leu Met Trp Arg Arg Ser Ser Ser Lys Lys Leu Val Gln Asp
65 70 75 80

Pro Val Pro Gln Val Ile Val Val Lys Lys Lys Glu Lys Glu Ser Glu
85 90 95

Val Asp Asp Gly Lys Lys Lys Val Ser Ile Phe Tyr Gly Thr Gln Thr

```
            100                 105                 110

Gly Thr Ala Glu Gly Phe Ala Lys Ala Leu Val Glu Glu Ala Lys Val
        115                 120                 125

Arg Tyr Glu Lys Thr Ser Phe Lys Val Ile Asp Leu Asp Asp Tyr Ala
    130                 135                 140

Ala Asp Asp Asp Glu Tyr Glu Glu Lys Leu Lys Lys Glu Ser Leu Ala
145                 150                 155                 160

Phe Phe Phe Leu Ala Thr Tyr Gly Asp Gly Glu Pro Thr Asp Asn Ala
                165                 170                 175

Ala Asn Phe Tyr Lys Trp Phe Thr Glu Gly Asp Asp Lys Gly Glu Trp
            180                 185                 190

Leu Lys Lys Leu Gln Tyr Gly Val Phe Gly Leu Gly Asn Arg Gln Tyr
        195                 200                 205

Glu His Phe Asn Lys Ile Ala Ile Val Val Asp Asp Lys Leu Thr Glu
    210                 215                 220

Met Gly Ala Lys Arg Leu Val Pro Val Gly Leu Gly Asp Asp Asp Gln
225                 230                 235                 240

Cys Ile Glu Asp Asp Phe Thr Ala Trp Lys Glu Leu Val Trp Pro Glu
                245                 250                 255

Leu Asp Gln Leu Leu Arg Asp Glu Asp Asp Thr Ser Val Thr Thr Pro
            260                 265                 270

Tyr Thr Ala Ala Val Leu Glu Tyr Arg Val Val Tyr His Asp Lys Pro
        275                 280                 285

Ala Asp Ser Tyr Ala Glu Asp Gln Thr His Thr Asn Gly His Val Val
    290                 295                 300

His Asp Ala Gln His Pro Ser Arg Ser Asn Val Ala Phe Lys Lys Glu
305                 310                 315                 320

Leu His Thr Ser Gln Ser Asp Arg Ser Cys Thr His Leu Glu Phe Asp
                325                 330                 335

Ile Ser His Thr Gly Leu Ser Tyr Glu Thr Gly Asp His Val Gly Val
            340                 345                 350
```

```
Tyr Ser Glu Asn Leu Ser Glu Val Val Asp Glu Ala Leu Lys Leu Leu
        355                 360                 365

Gly Leu Ser Pro Asp Thr Tyr Phe Ser Val His Ala Asp Lys Glu Asp
    370                 375                 380

Gly Thr Pro Ile Gly Gly Ala Ser Leu Pro Pro Pro Phe Pro Pro Cys
385                 390                 395                 400

Thr Leu Arg Asp Ala Leu Thr Arg Tyr Ala Asp Val Leu Ser Ser Pro
                405                 410                 415

Lys Lys Val Ala Leu Leu Ala Leu Ala Ala His Ala Ser Asp Pro Ser
            420                 425                 430

Glu Ala Asp Arg Leu Lys Phe Leu Ala Ser Pro Ala Gly Lys Asp Glu
        435                 440                 445

Tyr Ala Gln Trp Ile Val Ala Asn Gln Arg Ser Leu Leu Glu Val Met
    450                 455                 460

Gln Ser Phe Pro Ser Ala Lys Pro Pro Leu Gly Val Phe Phe Ala Ala
465                 470                 475                 480

Val Ala Pro Arg Leu Gln Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Pro
                485                 490                 495

Lys Met Ser Pro Asn Arg Ile His Val Thr Cys Ala Leu Val Tyr Glu
            500                 505                 510

Thr Thr Pro Ala Gly Arg Ile His Arg Gly Leu Cys Ser Thr Trp Met
        515                 520                 525

Lys Asn Ala Val Pro Leu Thr Glu Ser Pro Asp Cys Ser Gln Ala Ser
    530                 535                 540

Ile Phe Val Arg Thr Ser Asn Phe Arg Leu Pro Val Asp Pro Lys Val
545                 550                 555                 560

Pro Val Ile Met Ile Gly Pro Gly Thr Gly Leu Ala Pro Phe Arg Gly
                565                 570                 575

Phe Leu Gln Glu Arg Leu Ala Leu Lys Glu Ser Gly Thr Glu Leu Gly
            580                 585                 590

Ser Ser Ile Phe Phe Phe Gly Cys Arg Asn Arg Lys Val Asp Phe Ile
        595                 600                 605
```

Tyr Glu Asp Glu Leu Asn Asn Phe Val Glu Thr Gly Ala Leu Ser Glu
610 615 620

Leu Ile Val Ala Phe Ser Arg Glu Gly Thr Ala Lys Glu Tyr Val Gln
625 630 635 640

His Lys Met Ser Gln Lys Ala Ser Asp Ile Trp Lys Leu Leu Ser Glu
645 650 655

Gly Ala Tyr Leu Tyr Val Cys Gly Asp Ala Lys Gly Met Ala Lys Asp
660 665 670

Val His Arg Thr Leu His Thr Ile Val Gln Glu Gln Gly Ser Leu Asp
675 680 685

Ser Ser Lys Ala Glu Leu Tyr Val Lys Asn Leu Gln Met Ser Gly Arg
690 695 700

Tyr Leu Arg Asp Val Trp
705 710

<210> 148
<211> 692
<212> PRT
<213> Arabidopsis thaliana

<400> 148
Met Thr Ser Ala Leu Tyr Ala Ser Asp Leu Phe Lys Gln Leu Lys Ser
1 5 10 15

Ile Met Gly Thr Asp Ser Leu Ser Asp Asp Val Val Leu Val Ile Ala
20 25 30

Thr Thr Ser Leu Ala Leu Val Ala Gly Phe Val Val Leu Leu Trp Lys
35 40 45

Lys Thr Thr Ala Asp Arg Ser Gly Glu Leu Lys Pro Leu Met Ile Pro
50 55 60

Lys Ser Leu Met Ala Lys Asp Glu Asp Asp Asp Leu Asp Leu Gly Ser
65 70 75 80

Gly Lys Thr Arg Val Ser Ile Phe Phe Gly Thr Gln Thr Gly Thr Ala
85 90 95

Glu Gly Phe Ala Lys Ala Leu Ser Glu Glu Ile Lys Ala Arg Tyr Glu
100 105 110

```
Lys Ala Ala Val Lys Val Ile Asp Leu Asp Asp Tyr Ala Ala Asp Asp
        115                 120                 125

Asp Gln Tyr Glu Glu Lys Leu Lys Lys Glu Thr Leu Ala Phe Phe Cys
    130                 135                 140

Val Ala Thr Tyr Gly Asp Gly Glu Pro Thr Asp Asn Ala Ala Arg Phe
145                 150                 155                 160

Tyr Lys Trp Phe Thr Glu Glu Asn Glu Arg Asp Ile Lys Leu Gln Gln
                165                 170                 175

Leu Ala Tyr Gly Val Phe Ala Leu Gly Asn Arg Gln Tyr Glu His Phe
            180                 185                 190

Asn Lys Ile Gly Ile Val Leu Asp Glu Glu Leu Cys Lys Lys Gly Ala
        195                 200                 205

Lys Arg Leu Ile Glu Val Gly Leu Gly Asp Asp Asp Gln Ser Ile Glu
    210                 215                 220

Asp Asp Phe Asn Ala Trp Lys Glu Ser Leu Trp Ser Glu Leu Asp Lys
225                 230                 235                 240

Leu Leu Lys Asp Glu Asp Asp Lys Ser Val Ala Thr Pro Tyr Thr Ala
                245                 250                 255

Val Ile Pro Glu Tyr Arg Val Val Thr His Asp Pro Arg Phe Thr Thr
            260                 265                 270

Gln Lys Ser Met Glu Ser Asn Val Ala Asn Gly Asn Thr Thr Ile Asp
        275                 280                 285

Ile His His Pro Cys Arg Val Asp Val Ala Val Gln Lys Glu Leu His
    290                 295                 300

Thr His Glu Ser Asp Arg Ser Cys Ile His Leu Glu Phe Asp Ile Ser
305                 310                 315                 320

Arg Thr Gly Ile Thr Tyr Glu Thr Gly Asp His Val Gly Val Tyr Ala
                325                 330                 335

Glu Asn His Val Glu Ile Val Glu Glu Ala Gly Lys Leu Leu Gly His
            340                 345                 350

Ser Leu Asp Leu Val Phe Ser Ile His Ala Asp Lys Glu Asp Gly Ser
        355                 360                 365
```

```
Pro Leu Glu Ser Ala Val Pro Pro Pro Phe Pro Gly Pro Cys Thr Leu
    370                 375                 380

Gly Thr Gly Leu Ala Arg Tyr Ala Asp Leu Leu Asn Pro Pro Arg Lys
385                 390                 395                 400

Ser Ala Leu Val Ala Leu Ala Ala Tyr Ala Thr Glu Pro Ser Glu Ala
                405                 410                 415

Glu Lys Leu Lys His Leu Thr Ser Pro Asp Gly Lys Asp Glu Tyr Ser
            420                 425                 430

Gln Trp Ile Val Ala Ser Gln Arg Ser Leu Leu Glu Val Met Ala Ala
        435                 440                 445

Phe Pro Ser Ala Lys Pro Pro Leu Gly Val Phe Phe Ala Ala Ile Ala
    450                 455                 460

Pro Arg Leu Gln Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Pro Arg Leu
465                 470                 475                 480

Ala Pro Ser Arg Val His Val Thr Ser Ala Leu Val Tyr Gly Pro Thr
                485                 490                 495

Pro Thr Gly Arg Ile His Lys Gly Val Cys Ser Thr Trp Met Lys Asn
            500                 505                 510

Ala Val Pro Ala Glu Lys Ser His Glu Cys Ser Gly Ala Pro Ile Phe
        515                 520                 525

Ile Arg Ala Ser Asn Phe Lys Leu Pro Ser Asn Pro Ser Thr Pro Ile
    530                 535                 540

Val Met Val Gly Pro Gly Thr Gly Leu Ala Pro Phe Arg Gly Phe Leu
545                 550                 555                 560

Gln Glu Arg Met Ala Leu Lys Glu Asp Gly Glu Glu Leu Gly Ser Ser
                565                 570                 575

Leu Leu Phe Phe Gly Cys Arg Asn Arg Gln Met Asp Phe Ile Tyr Glu
            580                 585                 590

Asp Glu Leu Asn Asn Phe Val Asp Gln Gly Val Ile Ser Glu Leu Ile
        595                 600                 605

Met Ala Phe Ser Arg Glu Gly Ala Gln Lys Glu Tyr Val Gln His Lys
    610                 615                 620
```

Met Met Glu Lys Ala Ala Gln Val Trp Asp Leu Ile Lys Glu Glu Gly
625 630 635 640

Tyr Leu Tyr Val Cys Gly Asp Ala Lys Gly Met Ala Arg Asp Val His
645 650 655

Arg Thr Leu His Thr Ile Val Gln Glu Gln Glu Gly Val Ser Ser Ser
660 665 670

Glu Ala Glu Ala Ile Val Lys Lys Leu Gln Thr Glu Gly Arg Tyr Leu
675 680 685

Arg Asp Val Trp
690

<210> 149
<211> 713
<212> PRT
<213> Giberella fujikuroi

<400> 149
Met Ala Glu Leu Asp Thr Leu Asp Ile Val Val Leu Gly Val Ile Phe
1 5 10 15

Leu Gly Thr Val Ala Tyr Phe Thr Lys Gly Lys Leu Trp Gly Val Thr
20 25 30

Lys Asp Pro Tyr Ala Asn Gly Phe Ala Ala Gly Gly Ala Ser Lys Pro
35 40 45

Gly Arg Thr Arg Asn Ile Val Glu Ala Met Glu Glu Ser Gly Lys Asn
50 55 60

Cys Val Val Phe Tyr Gly Ser Gln Thr Gly Thr Ala Glu Asp Tyr Ala
65 70 75 80

Ser Arg Leu Ala Lys Glu Gly Lys Ser Arg Phe Gly Leu Asn Thr Met
85 90 95

Ile Ala Asp Leu Glu Asp Tyr Asp Phe Asp Asn Leu Asp Thr Val Pro
100 105 110

Ser Asp Asn Ile Val Met Phe Val Leu Ala Thr Tyr Gly Glu Gly Glu
115 120 125

Pro Thr Asp Asn Ala Val Asp Phe Tyr Glu Phe Ile Thr Gly Glu Asp
130 135 140

```
Ala Ser Phe Asn Glu Gly Asn Asp Pro Pro Leu Gly Asn Leu Asn Tyr
145                 150                 155                 160

Val Ala Phe Gly Leu Gly Asn Asn Thr Tyr Glu His Tyr Asn Ser Met
                165                 170                 175

Val Arg Asn Val Asn Lys Ala Leu Glu Lys Leu Gly Ala His Arg Ile
            180                 185                 190

Gly Glu Ala Gly Glu Gly Asp Asp Gly Ala Gly Thr Met Glu Glu Asp
        195                 200                 205

Phe Leu Ala Trp Lys Asp Pro Met Trp Glu Ala Leu Ala Lys Lys Met
    210                 215                 220

Gly Leu Glu Glu Arg Glu Ala Val Tyr Glu Pro Ile Phe Ala Ile Asn
225                 230                 235                 240

Glu Arg Asp Asp Leu Thr Pro Glu Ala Asn Glu Val Tyr Leu Gly Glu
                245                 250                 255

Pro Asn Lys Leu His Leu Glu Gly Thr Ala Lys Gly Pro Phe Asn Ser
            260                 265                 270

His Asn Pro Tyr Ile Ala Pro Ile Ala Glu Ser Tyr Glu Leu Phe Ser
        275                 280                 285

Ala Lys Asp Arg Asn Cys Leu His Met Glu Ile Asp Ile Ser Gly Ser
    290                 295                 300

Asn Leu Lys Tyr Glu Thr Gly Asp His Ile Ala Ile Trp Pro Thr Asn
305                 310                 315                 320

Pro Gly Glu Glu Val Asn Lys Phe Leu Asp Ile Leu Asp Leu Ser Gly
                325                 330                 335

Lys Gln His Ser Val Val Thr Val Lys Ala Leu Glu Pro Thr Ala Lys
            340                 345                 350

Val Pro Phe Pro Asn Pro Thr Thr Tyr Asp Ala Ile Leu Arg Tyr His
        355                 360                 365

Leu Glu Ile Cys Ala Pro Val Ser Arg Gln Phe Val Ser Thr Leu Ala
    370                 375                 380

Ala Phe Ala Pro Asn Asp Asp Ile Lys Ala Glu Met Asn Arg Leu Gly
385                 390                 395                 400
```

Ser Asp Lys Asp Tyr Phe His Glu Lys Thr Gly Pro His Tyr Tyr Asn
405 410 415

Ile Ala Arg Phe Leu Ala Ser Val Ser Lys Gly Glu Lys Trp Thr Lys
420 425 430

Ile Pro Phe Ser Ala Phe Ile Glu Gly Leu Thr Lys Leu Gln Pro Arg
435 440 445

Tyr Tyr Ser Ile Ser Ser Ser Ser Leu Val Gln Pro Lys Lys Ile Ser
450 455 460

Ile Thr Ala Val Val Glu Ser Gln Gln Ile Pro Gly Arg Asp Asp Pro
465 470 475 480

Phe Arg Gly Val Ala Thr Asn Tyr Leu Phe Ala Leu Lys Gln Lys Gln
485 490 495

Asn Gly Asp Pro Asn Pro Ala Pro Phe Gly Gln Ser Tyr Glu Leu Thr
500 505 510

Gly Pro Arg Asn Lys Tyr Asp Gly Ile His Val Pro Val His Val Arg
515 520 525

His Ser Asn Phe Lys Leu Pro Ser Asp Pro Gly Lys Pro Ile Ile Met
530 535 540

Ile Gly Pro Gly Thr Gly Val Ala Pro Phe Arg Gly Phe Val Gln Glu
545 550 555 560

Arg Ala Lys Gln Ala Arg Asp Gly Val Glu Val Gly Lys Thr Leu Leu
565 570 575

Phe Phe Gly Cys Arg Lys Ser Thr Glu Asp Phe Met Tyr Gln Lys Glu
580 585 590

Trp Gln Glu Tyr Lys Glu Ala Leu Gly Asp Lys Phe Glu Met Ile Thr
595 600 605

Ala Phe Ser Arg Glu Gly Ser Lys Lys Val Tyr Val Gln His Arg Leu
610 615 620

Lys Glu Arg Ser Lys Glu Val Ser Asp Leu Leu Ser Gln Lys Ala Tyr
625 630 635 640

Phe Tyr Val Cys Gly Asp Ala Ala His Met Ala Arg Glu Val Asn Thr

645 650 655

Val Leu Ala Gln Ile Ile Ala Glu Gly Arg Gly Val Ser Glu Ala Lys
660 665 670

Gly Glu Glu Ile Val Lys Asn Met Arg Ser Ala Asn Gln Tyr Gln Val
675 680 685

Cys Ser Asp Phe Val Thr Leu His Cys Lys Glu Thr Thr Tyr Ala Asn
690 695 700

Ser Glu Leu Gln Glu Asp Val Trp Ser
705 710

<210> 150
<211> 453
<212> PRT
<213> Arabidopsis thaliana

<400> 150
Met Gly Gly Leu Lys Phe His Val Leu Met Tyr Pro Trp Phe Ala Thr
1 5 10 15

Gly His Met Thr Pro Phe Leu Phe Leu Ala Asn Lys Leu Ala Glu Lys
20 25 30

Gly His Thr Val Thr Phe Leu Leu Pro Lys Lys Ser Leu Lys Gln Leu
35 40 45

Glu His Phe Asn Leu Phe Pro His Asn Ile Val Phe Arg Ser Val Thr
50 55 60

Val Pro His Val Asp Gly Leu Pro Val Gly Thr Glu Thr Ala Ser Glu
65 70 75 80

Ile Pro Val Thr Ser Thr Asp Leu Leu Met Ser Ala Met Asp Leu Thr
85 90 95

Arg Asp Gln Val Glu Ala Val Val Arg Ala Val Glu Pro Asp Leu Ile
100 105 110

Phe Phe Asp Phe Ala His Trp Ile Pro Glu Val Ala Arg Asp Phe Gly
115 120 125

Leu Lys Thr Val Lys Tyr Val Val Val Ser Ala Ser Thr Ile Ala Ser
130 135 140

Met Leu Val Pro Gly Gly Glu Leu Gly Val Pro Pro Pro Gly Tyr Pro
145 150 155 160

```
Ser Ser Lys Val Leu Leu Arg Lys Gln Asp Ala Tyr Thr Met Lys Lys
                165                 170                 175

Leu Glu Pro Thr Asn Thr Ile Asp Val Gly Pro Asn Leu Leu Glu Arg
            180                 185                 190

Val Thr Thr Ser Leu Met Asn Ser Asp Val Ile Ala Ile Arg Thr Ala
        195                 200                 205

Arg Glu Ile Glu Gly Asn Phe Cys Asp Tyr Ile Glu Lys His Cys Arg
    210                 215                 220

Lys Lys Val Leu Leu Thr Gly Pro Val Phe Pro Glu Pro Asp Lys Thr
225                 230                 235                 240

Arg Glu Leu Glu Glu Arg Trp Val Lys Trp Leu Ser Gly Tyr Glu Pro
                245                 250                 255

Asp Ser Val Val Phe Cys Ala Leu Gly Ser Gln Val Ile Leu Glu Lys
            260                 265                 270

Asp Gln Phe Gln Glu Leu Cys Leu Gly Met Glu Leu Thr Gly Ser Pro
        275                 280                 285

Phe Leu Val Ala Val Lys Pro Pro Arg Gly Ser Ser Thr Ile Gln Glu
    290                 295                 300

Ala Leu Pro Glu Gly Phe Glu Glu Arg Val Lys Gly Arg Gly Leu Val
305                 310                 315                 320

Trp Gly Gly Trp Val Gln Gln Pro Leu Ile Leu Ser His Pro Ser Val
                325                 330                 335

Gly Cys Phe Val Ser His Cys Gly Phe Gly Ser Met Trp Glu Ser Leu
            340                 345                 350

Leu Ser Asp Cys Gln Ile Val Leu Val Pro Gln Leu Gly Asp Gln Val
        355                 360                 365

Leu Asn Thr Arg Leu Leu Ser Asp Glu Leu Lys Val Ser Val Glu Val
    370                 375                 380

Ala Arg Glu Glu Thr Gly Trp Phe Ser Lys Glu Ser Leu Cys Asp Ala
385                 390                 395                 400

Val Asn Ser Val Met Lys Arg Asp Ser Glu Leu Gly Asn Leu Val Arg
```

405 410 415

Lys Asn His Thr Lys Trp Arg Glu Thr Val Ala Ser Pro Gly Leu Met
420 425 430

Thr Gly Tyr Val Asp Ala Phe Val Glu Ser Leu Gln Asp Leu Val Ser
435 440 445

Gly Thr Thr His Asp
450

<210> 151
<211> 1362
<212> DNA
<213> Arabidopsis thaliana

<400> 151
atgggtggtt tgaagtttca tgtacttatg tatccatggt tcgcaacagg ccatatgacc 60

ccgttccttt ttcttgccaa caaattggct gagaaaggtc atacggtcac tttcttgctt 120

cccaagaaat ctctgaaaca gttggaacat ttcaatctgt ttccacacaa cattgtcttt 180

cgctctgtca ccgtccctca tgtggatggt ctccccgttg gcacagagac agcctctgag 240

atccctgtga catcaactga tctcttgatg tctgctatgg atctcacacg tgatcaagtt 300

gaagctgtgg tccgagccgt tgaaccggac ctgatcttct ttgactttgc tcattggatt 360

ccagaagtag ctagggactt cggccttaag actgtaaagt acgtcgtggt gtctgcatcg 420

actatagcta gtatgcttgt cccaggtggt gagttaggtg ttcctccacc gggatatcca 480

tcatcaaagg tgctgcttcg taaacaagat gcttacacta tgaagaaact ggagcctaca 540

aatacaatcg atgtcggacc aaacctcttg gaacgagtca ctacaagtct tatgaactct 600

gatgtcattg cgataaggac agccagagaa atcgaaggaa acttttgcga ctatatagaa 660

aaacattgca ggaaaaaggt tctcttgaca ggtccggtgt tccctgagcc agacaagact 720

agagagctag aggaacgatg ggttaagtgg ctaagtgggt atgaaccaga ctcagtggtg 780

ttttgtgcac tgggctcaca agtcatttta gagaaagatc aattccaaga actctgctta 840

ggaatggagc taacaggttc accgtttctt gtagcggtta agccccctag aggctcatca 900

acgattcaag aagcacttcc tgaaggattc gaagagcggg ttaaaggaag aggccttgtt 960

tggggaggat gggttcaaca accattgata ttgtctcatc catcagtcgg gtgctttgtg 1020

agccattgtg ggtttggatc aatgtgggag tctttgctga gtgattgtca gatagtctta 1080

gtaccacagt tgggtgatca agtcctgaac acaagattgc tgagtgacga actcaaggtt 1140

tcggttgaag tggcaagaga ggaaacagga tggttctcga aagagagctt gtgcgatgct 1200

gtcaatagtg tgatgaaaag ggacagcgag ctcgggaacc tggtgaggaa gaatcacacc 1260

aagtggaggg agacagtagc tagtcctgga ctaatgactg gttatgtcga tgctttcgta 1320

gagtcattgc aggatcttgt ctctgggacc acccatgact ga 1362

<210> 152
<211> 462
<212> PRT
<213> Oryza sativa

<400> 152
Met Asp Ser Gly Tyr Ser Ser Ser Tyr Ala Ala Ala Ala Gly Met His
 1   5   10   15
Val Val Ile Cys Pro Trp Leu Ala Phe Gly His Leu Leu Pro Cys Leu
            20   25   30
Asp Leu Ala Gln Arg Leu Ala Ser Arg Gly His Arg Val Ser Phe Val
        35   40   45
Ser Thr Pro Arg Asn Ile Ser Arg Leu Pro Pro Val Arg Pro Ala Leu
    50   55   60
Ala Pro Leu Val Ala Phe Val Ala Leu Pro Leu Pro Arg Val Glu Gly
65   70   75   80
Leu Pro Asp Gly Ala Glu Ser Thr Asn Asp Val Pro His Asp Arg Pro
                85   90   95
Asp Met Val Glu Leu His Arg Arg Ala Phe Asp Gly Leu Ala Ala Pro
            100   105   110
Phe Ser Glu Phe Leu Gly Thr Ala Cys Ala Asp Trp Val Ile Val Asp
        115   120   125
Val Phe His His Trp Ala Ala Ala Ala Ala Leu Glu His Lys Val Pro
    130   135   140
Cys Ala Met Met Leu Leu Gly Ser Ala His Met Ile Ala Ser Ile Ala
145   150   155   160
Asp Arg Arg Leu Glu Arg Ala Glu Thr Glu Ser Pro Ala Ala Ala Gly
                165   170   175
Gln Gly Arg Pro Ala Ala Ala Pro Thr Phe Glu Val Ala Arg Met Lys
            180   185   190
Leu Ile Arg Thr Lys Gly Ser Ser Gly Met Ser Leu Ala Glu Arg Phe
        195   200   205
Ser Leu Thr Leu Ser Arg Ser Ser Leu Val Val Gly Arg Ser Cys Val
    210   215   220
Glu Phe Glu Pro Glu Thr Val Pro Leu Leu Ser Thr Leu Arg Gly Lys
225   230   235   240
Pro Ile Thr Phe Leu Gly Leu Met Pro Pro Leu His Glu Gly Arg Arg
                245   250   255
Glu Asp Gly Glu Asp Ala Thr Val Arg Trp Leu Asp Ala Gln Pro Ala
            260   265   270
Lys Ser Val Val Tyr Val Ala Leu Gly Ser Glu Val Pro Leu Gly Val
        275   280   285
Glu Lys Val His Glu Leu Ala Leu Gly Leu Glu Leu Ala Gly Thr Arg
    290   295   300
Phe Leu Trp Ala Leu Arg Lys Pro Thr Gly Val Ser Asp Ala Asp Leu
305   310   315   320
Leu Pro Ala Gly Phe Glu Glu Arg Thr Arg Gly Arg Gly Val Val Ala
                325   330   335
Thr Arg Trp Val Pro Gln Met Ser Ile Leu Ala His Ala Ala Val Gly
            340   345   350
Ala Phe Leu Thr His Cys Gly Trp Asn Ser Thr Ile Glu Gly Leu Met
        355   360   365
Phe Gly His Pro Leu Ile Met Leu Pro Ile Phe Gly Asp Gln Gly Pro
    370   375   380
Asn Ala Arg Leu Ile Glu Ala Lys Asn Ala Gly Leu Gln Val Ala Arg
385   390   395   400
Asn Asp Gly Asp Gly Ser Phe Asp Arg Glu Gly Val Ala Ala Ala Ile
                405   410   415

Arg Ala Val Ala Val Glu Glu Glu Ser Ser Lys Val Phe Gln Ala Lys
            420   425   430
Ala Lys Lys Leu Gln Glu Ile Val Ala Asp Met Ala Cys His Glu Arg
        435   440   445
Tyr Ile Asp Gly Phe Ile Gln Gln Leu Arg Ser Tyr Lys Asp
    450   455   460

<210> 153
<211> 1389
<212> DNA
<213> Oryza sativa

<400> 153
atggactccg gctactcctc ctcctacgcc gccgccgccg ggatgcacgt cgtgatctgc 60
ccgtggctcg ccttcggcca cctgctcccg tgcctcgacc tcgcccagcg cctcgcgtcg 120
cggggccacc gcgtgtcgtt cgtctccacg ccgcggaaca tatcccgcct cccgccggtg 180
cgccccgcgc tcgcgccgct cgtcgccttc gtggcgctgc cgctcccgcg cgtcgagggg 240
ctccccgacg gcgccgagtc caccaacgac gtcccccacg acaggccgga catggtcgag 300
ctccaccgga gggccttcga cgggctcgcc gcgcccttct cggagttctt gggcaccgcg 360
tgcgccgact gggtcatcgt cgacgtcttc caccactggg ccgcagccgc cgctctcgag 420
cacaaggtgc catgtgcaat gatgttgttg ggctctgcac atatgatcgc ttccatagca 480
gacagacggc tcgagcgcgc ggagacagag tcgcctgcgg ctgccgggca gggacgccca 540
gcggcggcgc caacgttcga ggtggcgagg atgaagttga tacgaaccaa aggctcatcg 600
ggaatgtccc tcgccgagcg cttctccttg acgctctcga ggagcagcct cgtcgtcggg 660
cggagctgcg tggagttcga gccggagacc gtcccgctcc tgtcgacgct ccgcggtaag 720
cctattacct tccttggcct tatgccgccg ttgcatgaag gccgccgcga ggacggcgag 780
gatgccaccg tccgctggct cgacgcgcag ccggccaagt ccgtcgtgta cgtcgcgcta 840
ggcagcgagg tgccactggg agtggagaag gtccacgagc tcgcgctcgg gctggagctc 900
gccgggacgc gcttcctctg ggctcttagg aagcccactg gcgtctccga cgccgacctc 960
ctccccgccg gcttcgagga gcgcacgcgc ggccgcggcg tcgtggcgac gagatgggtt 1020
cctcagatga gcatactggc gcacgccgcc gtgggcgcgt tcctgaccca ctgcggctgg 1080
aactcgacca tcgaggggct catgttcggc cacccgctta tcatgctgcc gatcttcggc 1140
gaccagggac cgaacgcgcg gctaatcgag gcgaagaacg ccggattgca ggtggcaaga 1200
aacgacggcg atggatcgtt cgaccgagaa ggcgtcgcgg cggcgattcg tgcagtcgcg 1260
gtggaggaag aaagcagcaa agtgtttcaa gccaaagcca agaagctgca ggagatcgtc 1320
gcggacatgg cctgccatga gaggtacatc gacggattca ttcagcaatt gagatcttac 1380
aaggattga   1389

<210> 154
<211> 1389
<212> DNA
<213> Artificial Sequence

<220>
<223> Oryza sativa sequence that has been codon
      optimized for yeast

<400> 154
atggatagtg gctactcctc atcttatgct gctgccgctg gtatgcacgt tgtgatctgc 60
ccttggttgg cctttggtca cctgttacca tgtctggatt tagcccaaag actggcctca 120
agaggccata gagtatcatt tgtgtctact cctagaaata tctctcgttt accaccagtc 180
agacctgctc tagctcctct agttgcattc gttgctcttc cacttccaag agtagaagga 240
ttgccagacg gcgctgaatc tactaatgac gtaccacatg atagacctga catggtcgaa 300
ttgcatagaa gagcctttga tggattggca gctccatttt ctgagttcct gggcacagca 360
tgtgcagact gggttatagt cgatgtattt catcactggg ctgctgcagc cgcattggaa 420
cataaggtgc cttgtgctat gatgttgtta gggtcagcac acatgatcgc atccatagct 480
gatagaagat tggaaagagc tgaaacagaa tccccagccg cagcaggaca aggtaggcca 540
gctgccgccc caacctttga agtggctaga atgaaattga ttcgtactaa aggtagttca 600
gggatgagtc ttgctgaaag gttttctctg acattatcta gatcatcatt agttgtaggt 660
agatcctgcg tcgagttcga acctgaaaca gtacctttac tatctacttt gagaggcaaa 720
cctattactt tccttggtct aatgcctcca ttacatgaag gaaggagaga agatggtgaa 780
gatgctactg ttaggtggtt agatgcccaa cctgctaagt ctgttgttta cgttgcattg 840

ggttctgagg taccactagg ggtggaaaag gtgcatgaat tagcattagg acttgagctg 900
gccggaacaa gattcctttg ggctttgaga aaaccaaccg gtgtttctga cgccgacttg 960
ctaccagctg ggttcgaaga gagaacaaga ggccgtggtg tcgttgctac tagatgggtc 1020
ccacaaatga gtattctagc tcatgcagct gtaggggcct ttctaaccca ttgcggttgg 1080
aactcaacaa tagaaggact gatgtttggt catccactta ttatgttacc aatctttggc 1140
gatcagggac ctaacgcaag attgattgag gcaaagaacg caggtctgca ggttgcacgt 1200
aatgatggtg atggttcctt tgatagagaa ggcgttgcag ctgccatcag agcagtcgcc 1260
gttgaggaag agtcatctaa agttttccaa gctaaggcca aaaaattaca agagattgtg 1320
gctgacatgg cttgtcacga aagatacatc gatggtttca tccaacaatt gagaagttat 1380
aaagactaa 1389

<210> 155
<211> 2506
<212> DNA
<213> Arabidopsis thaliana

<400> 155
cgtcagtcat caaggctaat tcgtcgcgag ttgctacgac gccgtttcgg ttgcttctgg 60
tttctttatg tctatcaacc ttcgctcctc cggttgttcg tctccgatct cagctacttt 120
ggaacgagga ttggactcag aagtacagac aagagctaac aatgtgagct ttgagcaaac 180
aaaggagaag attaggaaga tgttggagaa agtggagctt tctgtttcgg cctacgatac 240
tagttgggta gcaatggttc catcaccgag ctcccaaaat gctccacttt tcccacagtg 300
tgtgaaatgg ttattggata atcaacatga agatggatct tggggacttg ataaccatga 360
ccatcaatct cttaagaagg atgtgttatc atctacactg gctagtatcc tcgcgttaaa 420
gaagtgggga attggtgaaa gacaaataaa caagggtctc cagtttattg agctgaattc 480
tgcattagtc actgatgaaa ccatacagaa accaacaggg tttgatatta tatttcctgg 540
gatgattaaa tatgctagag atttgaatct gacgattcca ttgggctcag aagtggtgga 600
tgacatgata cgaaaaagag atctggatct taaatgtgat agtgaaaagt tttcaaaggg 660
aagagaagca tatctggcct atgttttaga ggggacaaga aacctaaaag attgggattt 720
gatagtcaaa tatcaaagga aaaatgggtc actgtttgat tctccagcca caacagcagc 780
tgcttttact cagtttggga atgatggttg tctccgttat ctctgttctc tccttcagaa 840
attcgaggct gcagttcctt cagtttatcc atttgatcaa tatgcacgcc ttagtataat 900
tgtcactctt gaaagcttag gaattgatag agatttcaaa accgaaatca aaagcatatt 960
ggatgaaacc tatagatatt ggcttcgtgg ggatgaagaa atatgtttgg acttggccac 1020
ttgtgctttg gctttccgat tattgcttgc tcatggctat gatgtgtctt acgatccgct 1080
aaaaccattt gcagaagaat ctggtttctc tgatactttg gaaggatatg ttaagaatac 1140
gttttctgtg ttagaattat ttaaggctgc tcaaagttat ccacatgaat cagctttgaa 1200
gaagcagtgt tgttggacta aacaatatct ggagatggaa ttgtccagct gggttaagac 1260
ctctgttcga gataaatacc tcaagaaaga ggtcgaggat gctcttgctt ttccctccta 1320
tgcaagccta gaaagatcag atcacaggag aaaaatactc aatggttctg ctgtggaaaa 1380
caccagagtt acaaaaacct catatcgttt gcacaatatt tgcacctctg atatcctgaa 1440
gttagctgtg gatgacttca atttctgcca gtccatacac cgtgaagaaa tggaacgtct 1500
tgataggtgg attgtggaga atagattgca ggaactgaaa tttgccagac agaagctggc 1560
ttactgttat ttctctgggg ctgcaacttt attttctcca gaactatctg atgctcgtat 1620
atcgtgggcc aaaggtggag tacttacaac ggttgtagac gacttctttg atgttggagg 1680
gtccaaagaa gaactggaaa acctcataca cttggtcgaa aagtgggatt tgaacggtgt 1740
tcctgagtac agctcagaac atgttgagat catattctca gttctaaggg acaccattct 1800
cgaaacagga gacaaagcat tcacctatca aggacgcaat gtgacacacc acattgtgaa 1860
aatttggttg gatctgctca agtctatgtt gagagaagcc gagtggtcca gtgacaagtc 1920
aacaccaagc ttggaggatt acatggaaaa tgcgtacata tcatttgcat taggaccaat 1980
tgtcctccca gctacctatc tgatcggacc tccacttcca gagaagacag tcgatagcca 2040
ccaatataat cagctctaca agctcgtgag cactatgggt cgtcttctaa atgacataca 2100
aggttttaag agagaaagcg cggaagggaa gctgaatgcg gtttcattgc acatgaaaca 2160
cgagagagac aatcgcagca aagaagtgat catagaatcg atgaaaggtt tagcagagag 2220
aaagagggaa gaattgcata agctagtttt ggaggagaaa ggaagtgtgg ttccaaggga 2280
atgcaaagaa gcgttcttga aaatgagcaa agtgttgaac ttattttaca ggaaggacga 2340
tggattcaca tcaaatgatc tgatgagtct tgttaaatca gtgatctacg agcctgttag 2400
cttacagaaa gaatctttaa cttgatccaa gttgatctgg caggtaaact cagtaaatga 2460
aaataagact ttggtcttct tctttgttgc ttcagaacaa gaagag 2506

<210> 156
<211> 785
<212> PRT

<213> Arabidopsis thaliana

<400> 156

```
Met Ser Ile Asn Leu Arg Ser Ser Gly Cys Ser Ser Pro Ile Ser Ala
 1               5                  10                  15
Thr Leu Glu Arg Gly Leu Asp Ser Glu Val Gln Thr Arg Ala Asn Asn
            20                  25                  30
Val Ser Phe Glu Gln Thr Lys Glu Lys Ile Arg Lys Met Leu Glu Lys
        35                  40                  45
Val Glu Leu Ser Val Ser Ala Tyr Asp Thr Ser Trp Val Ala Met Val
    50                  55                  60
Pro Ser Pro Ser Ser Gln Asn Ala Pro Leu Phe Pro Gln Cys Val Lys
65                  70                  75                  80
Trp Leu Leu Asp Asn Gln His Glu Asp Gly Ser Trp Gly Leu Asp Asn
                85                  90                  95
His Asp His Gln Ser Leu Lys Lys Asp Val Leu Ser Ser Thr Leu Ala
            100                 105                 110
Ser Ile Leu Ala Leu Lys Lys Trp Gly Ile Gly Glu Arg Gln Ile Asn
        115                 120                 125
Lys Gly Leu Gln Phe Ile Glu Leu Asn Ser Ala Leu Val Thr Asp Glu
    130                 135                 140
Thr Ile Gln Lys Pro Thr Gly Phe Asp Ile Ile Phe Pro Gly Met Ile
145                 150                 155                 160
Lys Tyr Ala Arg Asp Leu Asn Leu Thr Ile Pro Leu Gly Ser Glu Val
                165                 170                 175
Val Asp Asp Met Ile Arg Lys Arg Asp Leu Asp Leu Lys Cys Asp Ser
            180                 185                 190
Glu Lys Phe Ser Lys Gly Arg Glu Ala Tyr Leu Ala Tyr Val Leu Glu
        195                 200                 205
Gly Thr Arg Asn Leu Lys Asp Trp Asp Leu Ile Val Lys Tyr Gln Arg
    210                 215                 220
Lys Asn Gly Ser Leu Phe Asp Ser Pro Ala Thr Thr Ala Ala Ala Phe
225                 230                 235                 240
Thr Gln Phe Gly Asn Asp Gly Cys Leu Arg Tyr Leu Cys Ser Leu Leu
                245                 250                 255
Gln Lys Phe Glu Ala Ala Val Pro Ser Val Tyr Pro Phe Asp Gln Tyr
            260                 265                 270
Ala Arg Leu Ser Ile Ile Val Thr Leu Glu Ser Leu Gly Ile Asp Arg
        275                 280                 285
Asp Phe Lys Thr Glu Ile Lys Ser Ile Leu Asp Glu Thr Tyr Arg Tyr
    290                 295                 300
Trp Leu Arg Gly Asp Glu Glu Ile Cys Leu Asp Leu Ala Thr Cys Ala
305                 310                 315                 320
Leu Ala Phe Arg Leu Leu Leu Ala His Gly Tyr Asp Val Ser Tyr Asp
                325                 330                 335
Pro Leu Lys Pro Phe Ala Glu Glu Ser Gly Phe Ser Asp Thr Leu Glu
            340                 345                 350
Gly Tyr Val Lys Asn Thr Phe Ser Val Leu Glu Leu Phe Lys Ala Ala
        355                 360                 365
Gln Ser Tyr Pro His Glu Ser Ala Leu Lys Lys Gln Cys Cys Trp Thr
    370                 375                 380
Lys Gln Tyr Leu Glu Met Glu Leu Ser Ser Trp Val Lys Thr Ser Val
385                 390                 395                 400
Arg Asp Lys Tyr Leu Lys Lys Glu Val Glu Asp Ala Leu Ala Phe Pro
                405                 410                 415
Ser Tyr Ala Ser Leu Glu Arg Ser Asp His Arg Arg Lys Ile Leu Asn
            420                 425                 430
Gly Ser Ala Val Glu Asn Thr Arg Val Thr Lys Thr Ser Tyr Arg Leu
        435                 440                 445
His Asn Ile Cys Thr Ser Asp Ile Leu Lys Leu Ala Val Asp Asp Phe
    450                 455                 460
Asn Phe Cys Gln Ser Ile His Arg Glu Glu Met Glu Arg Leu Asp Arg
465                 470                 475                 480
```

Trp Ile Val Glu Asn Arg Leu Gln Glu Leu Lys Phe Ala Arg Gln Lys
485 490 495
Leu Ala Tyr Cys Tyr Phe Ser Gly Ala Ala Thr Leu Phe Ser Pro Glu
500 505 510
Leu Ser Asp Ala Arg Ile Ser Trp Ala Lys Gly Gly Val Leu Thr Thr
515 520 525
Val Val Asp Asp Phe Phe Asp Val Gly Gly Ser Lys Glu Glu Leu Glu
530 535 540
Asn Leu Ile His Leu Val Glu Lys Trp Asp Leu Asn Gly Val Pro Glu
545 550 555 560
Tyr Ser Ser Glu His Val Glu Ile Ile Phe Ser Val Leu Arg Asp Thr
565 570 575
Ile Leu Glu Thr Gly Asp Lys Ala Phe Thr Tyr Gln Gly Arg Asn Val
580 585 590
Thr His His Ile Val Lys Ile Trp Leu Asp Leu Leu Lys Ser Met Leu
595 600 605
Arg Glu Ala Glu Trp Ser Ser Asp Lys Ser Thr Pro Ser Leu Glu Asp
610 615 620
Tyr Met Glu Asn Ala Tyr Ile Ser Phe Ala Leu Gly Pro Ile Val Leu
625 630 635 640
Pro Ala Thr Tyr Leu Ile Gly Pro Pro Leu Pro Glu Lys Thr Val Asp
645 650 655
Ser His Gln Tyr Asn Gln Leu Tyr Lys Leu Val Ser Thr Met Gly Arg
660 665 670
Leu Leu Asn Asp Ile Gln Gly Phe Lys Arg Glu Ser Ala Glu Gly Lys
675 680 685
Leu Asn Ala Val Ser Leu His Met Lys His Glu Arg Asp Asn Arg Ser
690 695 700
Lys Glu Val Ile Ile Glu Ser Met Lys Gly Leu Ala Glu Arg Lys Arg
705 710 715 720
Glu Glu Leu His Lys Leu Val Leu Glu Glu Lys Gly Ser Val Val Pro
725 730 735
Arg Glu Cys Lys Glu Ala Phe Leu Lys Met Ser Lys Val Leu Asn Leu
740 745 750
Phe Tyr Arg Lys Asp Asp Gly Phe Thr Ser Asn Asp Leu Met Ser Leu
755 760 765
Val Lys Ser Val Ile Tyr Glu Pro Val Ser Leu Gln Lys Glu Ser Leu
770 775 780
Thr
785

<210> 157
<211> 2490
<212> DNA
<213> Zea mays

<400> 157
atggttttgt cttcttcttg tactacagta ccacacttat cttcattagc tgtcgtgcaa 60
cttggtcctt ggagcagtag gattaaaaag aaaaccgata ctgttgcagt accagccgct 120
gcaggaaggt ggagaagggc cttggctaga gcacagcaca catcagaatc cgcagctgtc 180
gcaaagggca gcagtttgac ccctatagtg agaactgacg ctgagtcaag gagaacaaga 240
tggccaaccg atgacgatga cgccgaacct ttagtggatg agatcagggc aatgcttact 300
tccatgtctg atggtgacat ttccgtgagc gcatacgata cagcctgggt cggattggtt 360
ccaagattag acggcggtga aggtcctcaa tttccagcag ctgtgagatg gataagaaat 420
aaccagttgc ctgacggaag ttggggcgat gccgcattat tctctgccta tgacaggctt 480
atcaataccc ttgcctgcgt tgtaactttg acaaggtggt ccctagaacc agagatgaga 540
ggtagaggac tatcttttt gggtaggaac atgtggaaat tagcaactga agatgaagag 600
tcaatgccta ttggcttcga attagcattt ccatctttga tagagcttgc taagagccta 660
ggtgtccatg acttccctta tgatcaccag gccctacaag gaatctactc ttcaagagag 720
atcaaaatga agaggattcc aaaagaagtg atgcataccg ttccaacatc aatattgcac 780
agtttggagg gtatgcctgg cctagattgg gctaaactac ttaaactaca gagcagcgac 840
ggaagttttt tgttctcacc agctgccact gcatatgctt taatgaatac cggagatgac 900

aggtgtttta gctacatcga tagaacagta aagaaattca acggcggcgt ccctaatgtt 960
tatccagtgg atctatttga acatatttgg gccgttgata gacttgaaag attaggaatc 1020
tccaggtact tccaaaagga gatcgaacaa tgcatggatt atgtaaacag gcattggact 1080
gaggacggta tttgttgggc aaggaactct gatgtcaaag aggtggacga cacagctatg 1140
gcctttagac ttcttaggtt gcacggctac agcgtcagtc ctgatgtgtt taaaaacttc 1200
gaaaaggacg gtgaattttt cgcatttgtc ggacagtcta atcaagctgt taccggtatg 1260
tacaacttaa acagagcaag ccagatatcc ttcccaggcg aggatgtgct tcatagagct 1320
ggtgccttct catatgagtt cttgaggaga aaagaagcag agggagcttt gagggacaag 1380
tggatcattt ctaaagatct acctggtgaa gttgtgtata ctttggattt tccatggtac 1440
ggcaacttac ctagagtcga ggccagagac tacctagagc aatacggagg tggtgatgac 1500
gtttggattg gcaagacatt gtataggatg ccacttgtaa acaatgatgt atatttggaa 1560
ttggcaagaa tggatttcaa ccactgccag gctttgcatc agttagagtg gcaaggacta 1620
aaaagatggt atactgaaaa taggttgatg gactttggtg tcgcccaaga agatgccctt 1680
agagcttatt ttcttgcagc cgcatctgtt tacgagcctt gtagagctgc cgagaggctt 1740
gcatgggcta gagccgcaat actagctaac gccgtgagca cccacttaag aaatagccca 1800
tcattcagag aaaggttaga gcattctctt aggtgtagac ctagtgaaga gacagatggc 1860
tcctggttta actcctcaag tggctctgat gcagttttag taaaggctgt cttaagactt 1920
actgattcat tagccaggga agcacagcca atccatggag gtgacccaga agatattata 1980
cacaagttgt taagatctgc ttgggccgag tgggttaggg aaaaggcaga cgctgccgat 2040
agcgtgtgca atggtagttc tgcagtagaa caagagggat caagaatggt ccatgataaa 2100
cagacctgtc tattattggc tagaatgatc gaaatttctg ccggtagggc agctggtgaa 2160
gcagccagtg aggacggcga tagaagaata attcaattaa caggctccat ctgcgacagt 2220
cttaagcaaa aaatgctagt ttcacaggac cctgaaaaaa atgaagagat gatgtctcac 2280
gtggatgacg aattgaagtt gaggattaga gagttcgttc aatatttgct tagactaggt 2340
gaaaaaaaga ctggatctag cgaaaccagg caaacatttt taagtatagt gaaatcatgt 2400
tactatgctg ctcattgccc acctcatgtc gttgatagac acattagtag agtgattttc 2460
gagccagtaa gtgccgcaaa gtaaccgcgg 2490

<210> 158
<211> 827
<212> PRT
<213> Zea mays

<400> 158
Met Val Leu Ser Ser Ser Cys Thr Thr Val Pro His Leu Ser Ser Leu
 1   5   10   15
Ala Val Val Gln Leu Gly Pro Trp Ser Ser Arg Ile Lys Lys Lys Thr
            20   25   30
Asp Thr Val Ala Val Pro Ala Ala Ala Gly Arg Trp Arg Arg Ala Leu
        35   40   45
Ala Arg Ala Gln His Thr Ser Glu Ser Ala Ala Val Ala Lys Gly Ser
    50   55   60
Ser Leu Thr Pro Ile Val Arg Thr Asp Ala Glu Ser Arg Arg Thr Arg
65   70   75   80
Trp Pro Thr Asp Asp Asp Asp Ala Glu Pro Leu Val Asp Glu Ile Arg
                85   90   95
Ala Met Leu Thr Ser Met Ser Asp Gly Asp Ile Ser Val Ser Ala Tyr
            100   105   110
Asp Thr Ala Trp Val Gly Leu Val Pro Arg Leu Asp Gly Gly Glu Gly
        115   120   125
Pro Gln Phe Pro Ala Ala Val Arg Trp Ile Arg Asn Asn Gln Leu Pro
    130   135   140
Asp Gly Ser Trp Gly Asp Ala Ala Leu Phe Ser Ala Tyr Asp Arg Leu
145   150   155   160
Ile Asn Thr Leu Ala Cys Val Val Thr Leu Thr Arg Trp Ser Leu Glu
                165   170   175
Pro Glu Met Arg Gly Arg Gly Leu Ser Phe Leu Gly Arg Asn Met Trp
            180   185   190
Lys Leu Ala Thr Glu Asp Glu Glu Ser Met Pro Ile Gly Phe Glu Leu
        195   200   205
Ala Phe Pro Ser Leu Ile Glu Leu Ala Lys Ser Leu Gly Val His Asp
    210   215   220
Phe Pro Tyr Asp His Gln Ala Leu Gln Gly Ile Tyr Ser Ser Arg Glu

225 230 235 240
Ile Lys Met Lys Arg Ile Pro Lys Glu Val Met His Thr Val Pro Thr
245 250 255
Ser Ile Leu His Ser Leu Glu Gly Met Pro Gly Leu Asp Trp Ala Lys
260 265 270
Leu Leu Lys Leu Gln Ser Ser Asp Gly Ser Phe Leu Phe Ser Pro Ala
275 280 285
Ala Thr Ala Tyr Ala Leu Met Asn Thr Gly Asp Asp Arg Cys Phe Ser
290 295 300
Tyr Ile Asp Arg Thr Val Lys Lys Phe Asn Gly Gly Val Pro Asn Val
305 310 315 320
Tyr Pro Val Asp Leu Phe Glu His Ile Trp Ala Val Asp Arg Leu Glu
325 330 335
Arg Leu Gly Ile Ser Arg Tyr Phe Gln Lys Glu Ile Glu Gln Cys Met
340 345 350
Asp Tyr Val Asn Arg His Trp Thr Glu Asp Gly Ile Cys Trp Ala Arg
355 360 365
Asn Ser Asp Val Lys Glu Val Asp Asp Thr Ala Met Ala Phe Arg Leu
370 375 380
Leu Arg Leu His Gly Tyr Ser Val Ser Pro Asp Val Phe Lys Asn Phe
385 390 395 400
Glu Lys Asp Gly Glu Phe Phe Ala Phe Val Gly Gln Ser Asn Gln Ala
405 410 415
Val Thr Gly Met Tyr Asn Leu Asn Arg Ala Ser Gln Ile Ser Phe Pro
420 425 430
Gly Glu Asp Val Leu His Arg Ala Gly Ala Phe Ser Tyr Glu Phe Leu
435 440 445
Arg Arg Lys Glu Ala Glu Gly Ala Leu Arg Asp Lys Trp Ile Ile Ser
450 455 460
Lys Asp Leu Pro Gly Glu Val Val Tyr Thr Leu Asp Phe Pro Trp Tyr
465 470 475 480
Gly Asn Leu Pro Arg Val Glu Ala Arg Asp Tyr Leu Glu Gln Tyr Gly
485 490 495
Gly Gly Asp Asp Val Trp Ile Gly Lys Thr Leu Tyr Arg Met Pro Leu
500 505 510
Val Asn Asn Asp Val Tyr Leu Glu Leu Ala Arg Met Asp Phe Asn His
515 520 525
Cys Gln Ala Leu His Gln Leu Glu Trp Gln Gly Leu Lys Arg Trp Tyr
530 535 540
Thr Glu Asn Arg Leu Met Asp Phe Gly Val Ala Gln Glu Asp Ala Leu
545 550 555 560
Arg Ala Tyr Phe Leu Ala Ala Ala Ser Val Tyr Glu Pro Cys Arg Ala
565 570 575
Ala Glu Arg Leu Ala Trp Ala Arg Ala Ala Ile Leu Ala Asn Ala Val
580 585 590
Ser Thr His Leu Arg Asn Ser Pro Ser Phe Arg Glu Arg Leu Glu His
595 600 605
Ser Leu Arg Cys Arg Pro Ser Glu Glu Thr Asp Gly Ser Trp Phe Asn
610 615 620
Ser Ser Ser Gly Ser Asp Ala Val Leu Val Lys Ala Val Leu Arg Leu
625 630 635 640
Thr Asp Ser Leu Ala Arg Glu Ala Gln Pro Ile His Gly Gly Asp Pro
645 650 655
Glu Asp Ile Ile His Lys Leu Leu Arg Ser Ala Trp Ala Glu Trp Val
660 665 670
Arg Glu Lys Ala Asp Ala Ala Asp Ser Val Cys Asn Gly Ser Ser Ala
675 680 685
Val Glu Gln Glu Gly Ser Arg Met Val His Asp Lys Gln Thr Cys Leu
690 695 700
Leu Leu Ala Arg Met Ile Glu Ile Ser Ala Gly Arg Ala Ala Gly Glu
705 710 715 720
Ala Ala Ser Glu Asp Gly Asp Arg Arg Ile Ile Gln Leu Thr Gly Ser
725 730 735

```
Ile Cys Asp Ser Leu Lys Gln Lys Met Leu Val Ser Gln Asp Pro Glu
            740                 745                 750
Lys Asn Glu Glu Met Met Ser His Val Asp Asp Glu Leu Lys Leu Arg
        755                 760                 765
Ile Arg Glu Phe Val Gln Tyr Leu Leu Arg Leu Gly Glu Lys Lys Thr
    770                 775                 780
Gly Ser Ser Glu Thr Arg Gln Thr Phe Leu Ser Ile Val Lys Ser Cys
785                 790                 795                 800
Tyr Tyr Ala Ala His Cys Pro Pro His Val Val Asp Arg His Ile Ser
                805                 810                 815
Arg Val Ile Phe Glu Pro Val Ser Ala Ala Lys
            820                 825
```

<210> 159
<211> 2570
<212> DNA
<213> Arabidopsis thaliana

<400> 159

```
cttcttcact aaatacttag acagagaaaa cagagctttt taaagccatg tctcttcagt 60
atcatgttct aaactccatt ccaagtacaa cctttctcag ttctactaaa acaacaatat 120
cttcttcttt ccttaccatc tcaggatctc ctctcaatgt cgctagagac aaatccagaa 180
gcggttccat acattgttca aagcttcgaa ctcaagaata cattaattct caagaggttc 240
aacatgattt gcctctaata catgagtggc aacagcttca aggagaagat gctcctcaga 300
ttagtgttgg aagtaatagt aatgcattca aagaagcagt gaagagtgtg aaaacgatct 360
tgagaaacct aacggacggg gaaattacga tatcggctta cgatacagct tgggttgcat 420
tgatcgatgc cggagataaa actccggcgt ttccctccgc cgtgaaatgg atcgccgaga 480
accaactttc cgatggttct tggggagatg cgtatctctt ctcttatcat gatcgtctca 540
tcaataccct tgcatgcgtc gttgctctaa gatcatggaa tctctttcct catcaatgca 600
acaaaggaat cacgtttttc cgggaaaata ttgggaagct agaagacgaa aatgatgagc 660
atatgccaat cggattcgaa gtagcattcc catcgttgct tgagatagct cgaggaataa 720
acattgatgt accgtacgat tctccggtct taaaagatat atacgccaag aaagagctaa 780
agcttacaag gataccaaaa gagataatgc acaagatacc aacaacattg ttgcatagtt 840
tggaggggat gcgtgattta gattgggaaa agctcttgaa acttcaatct caagacggat 900
ctttcctctt ctctccttcc tctaccgctt ttgcattcat gcagacccga gacagtaact 960
gcctcgagta tttgcgaaat gccgtcaaac gtttcaatgg aggagttccc aatgtctttc 1020
ccgtggatct tttcgagcac atatggatag tggatcggtt acaacgttta gggatatcga 1080
gatactttga agaagagatt aaagagtgtc ttgactatgt ccacagatat tggaccgaca 1140
atggcatatg ttgggctaga tgttcccatg tccaagacat cgatgataca gccatggcat 1200
ttaggctctt aagacaacat ggataccaag tgtccgcaga tgtattcaag aactttgaga 1260
aagagggaga gtttttctgc tttgtggggc aatcaaacca agcagtaacc ggtatgttca 1320
acctataccg ggcatcacaa ttggcgtttc caagggaaga gatattgaaa aacgccaaag 1380
agttttctta taattatctg ctagaaaaac gggagagaga ggagttgatt gataagtgga 1440
ttataatgaa agacttacct ggcgagattg ggtttgcgtt agagattcca tggtacgcaa 1500
gcttgcctcg agtagagacg agattctata ttgatcaata tggtggagaa aacgacgttt 1560
ggattggcaa gactctttat aggatgccat acgtgaacaa taatggatat ctggaattag 1620
caaaacaaga ttacaacaat tgccaagctc agcatcagct cgaatgggac atattccaaa 1680
agtggtatga agaaaatagg ttaagtgagt ggggtgtgcg cagaagtgag cttctcgagt 1740
gttactactt agcggctgca actatatttg aatcagaaag gtcacatgag agaatggttt 1800
gggctaagtc aagtgtattg gttaaagcca tttcttcttc ttttggggaa tcctctgact 1860
ccagaagaag cttctccgat cagtttcatg aatacattgc caatgctcga cgaagtgatc 1920
atcactttaa tgacaggaac atgagattgg accgaccagg atcggttcag gccagtcggc 1980
ttgccggagt gttaatcggg actttgaatc aaatgtcttt tgacctttтc atgtctcatg 2040
gccgtgacgt taacaatctc ctctatctat cgtggggaga ttggatggaa aaatggaaac 2100
tatatggaga tgaaggagaa ggagagctca tggtgaagat gataattcta atgaagaaca 2160
atgacctaac taacttcttc acccacactc acttcgttcg tctcgcggaa atcatcaatc 2220
gaatctgtct tcctcgccaa tacttaaagg caaggagaaa cgatgagaag gagaagacaa 2280
taaagagtat ggagaaggag atggggaaaa tggttgagtt agcattgtcg gagagtgaca 2340
catttcgtga cgtcagcatc acgtttcttg atgtagcaaa agcattttac tactttgctt 2400
tatgtggcga tcatctccaa actcacatct ccaaagtctt gtttcaaaaa gtctagtaac 2460
ctcatcatca tcatcgatcc attaacaatc agtggatcga tgtatccata gatgcgtgaa 2520
taatatttca tgtagagaag gagaacaaat tagatcatgt agggttatca            2570
```

<210> 160
<211> 802
<212> PRT
<213> Arabidopsis thaliana

<400> 160
Met Ser Leu Gln Tyr His Val Leu Asn Ser Ile Pro Ser Thr Thr Phe
 1   5   10   15
Leu Ser Ser Thr Lys Thr Thr Ile Ser Ser Ser Phe Leu Thr Ile Ser
            20   25   30
Gly Ser Pro Leu Asn Val Ala Arg Asp Lys Ser Arg Ser Gly Ser Ile
        35   40   45
His Cys Ser Lys Leu Arg Thr Gln Glu Tyr Ile Asn Ser Gln Glu Val
    50   55   60
Gln His Asp Leu Pro Leu Ile His Glu Trp Gln Gln Leu Gln Gly Glu
65   70   75   80
Asp Ala Pro Gln Ile Ser Val Gly Ser Asn Ser Asn Ala Phe Lys Glu
                85   90   95
Ala Val Lys Ser Val Lys Thr Ile Leu Arg Asn Leu Thr Asp Gly Glu
            100   105   110
Ile Thr Ile Ser Ala Tyr Asp Thr Ala Trp Val Ala Leu Ile Asp Ala
        115   120   125
Gly Asp Lys Thr Pro Ala Phe Pro Ser Ala Val Lys Trp Ile Ala Glu
    130   135   140
Asn Gln Leu Ser Asp Gly Ser Trp Gly Asp Ala Tyr Leu Phe Ser Tyr
145   150   155   160
His Asp Arg Leu Ile Asn Thr Leu Ala Cys Val Val Ala Leu Arg Ser
                165   170   175
Trp Asn Leu Phe Pro His Gln Cys Asn Lys Gly Ile Thr Phe Phe Arg
            180   185   190
Glu Asn Ile Gly Lys Leu Glu Asp Glu Asn Asp Glu His Met Pro Ile
        195   200   205
Gly Phe Glu Val Ala Phe Pro Ser Leu Leu Glu Ile Ala Arg Gly Ile
    210   215   220
Asn Ile Asp Val Pro Tyr Asp Ser Pro Val Leu Lys Asp Ile Tyr Ala
225   230   235   240
Lys Lys Glu Leu Lys Leu Thr Arg Ile Pro Lys Glu Ile Met His Lys
                245   250   255
Ile Pro Thr Thr Leu Leu His Ser Leu Glu Gly Met Arg Asp Leu Asp
            260   265   270
Trp Glu Lys Leu Leu Lys Leu Gln Ser Gln Asp Gly Ser Phe Leu Phe
        275   280   285
Ser Pro Ser Ser Thr Ala Phe Ala Phe Met Gln Thr Arg Asp Ser Asn
    290   295   300
Cys Leu Glu Tyr Leu Arg Asn Ala Val Lys Arg Phe Asn Gly Gly Val
305   310   315   320
Pro Asn Val Phe Pro Val Asp Leu Phe Glu His Ile Trp Ile Val Asp
                325   330   335
Arg Leu Gln Arg Leu Gly Ile Ser Arg Tyr Phe Glu Glu Glu Ile Lys
            340   345   350
Glu Cys Leu Asp Tyr Val His Arg Tyr Trp Thr Asp Asn Gly Ile Cys
        355   360   365
Trp Ala Arg Cys Ser His Val Gln Asp Ile Asp Asp Thr Ala Met Ala
    370   375   380
Phe Arg Leu Leu Arg Gln His Gly Tyr Gln Val Ser Ala Asp Val Phe
385   390   395   400
Lys Asn Phe Glu Lys Glu Gly Glu Phe Phe Cys Phe Val Gly Gln Ser
                405   410   415
Asn Gln Ala Val Thr Gly Met Phe Asn Leu Tyr Arg Ala Ser Gln Leu
            420   425   430
Ala Phe Pro Arg Glu Glu Ile Leu Lys Asn Ala Lys Glu Phe Ser Tyr
        435   440   445

Asn Tyr Leu Leu Glu Lys Arg Glu Arg Glu Glu Leu Ile Asp Lys Trp
450 455 460
Ile Ile Met Lys Asp Leu Pro Gly Glu Ile Gly Phe Ala Leu Glu Ile
465 470 475 480
Pro Trp Tyr Ala Ser Leu Pro Arg Val Glu Thr Arg Phe Tyr Ile Asp
485 490 495
Gln Tyr Gly Gly Glu Asn Asp Val Trp Ile Gly Lys Thr Leu Tyr Arg
500 505 510
Met Pro Tyr Val Asn Asn Asn Gly Tyr Leu Glu Leu Ala Lys Gln Asp
515 520 525
Tyr Asn Asn Cys Gln Ala Gln His Gln Leu Glu Trp Asp Ile Phe Gln
530 535 540
Lys Trp Tyr Glu Glu Asn Arg Leu Ser Glu Trp Gly Val Arg Arg Ser
545 550 555 560
Glu Leu Leu Glu Cys Tyr Tyr Leu Ala Ala Ala Thr Ile Phe Glu Ser
565 570 575
Glu Arg Ser His Glu Arg Met Val Trp Ala Lys Ser Ser Val Leu Val
580 585 590
Lys Ala Ile Ser Ser Ser Phe Gly Glu Ser Ser Asp Ser Arg Arg Ser
595 600 605
Phe Ser Asp Gln Phe His Glu Tyr Ile Ala Asn Ala Arg Arg Ser Asp
610 615 620
His His Phe Asn Asp Arg Asn Met Arg Leu Asp Arg Pro Gly Ser Val
625 630 635 640
Gln Ala Ser Arg Leu Ala Gly Val Leu Ile Gly Thr Leu Asn Gln Met
645 650 655
Ser Phe Asp Leu Phe Met Ser His Gly Arg Asp Val Asn Asn Leu Leu
660 665 670
Tyr Leu Ser Trp Gly Asp Trp Met Glu Lys Trp Lys Leu Tyr Gly Asp
675 680 685
Glu Gly Glu Gly Glu Leu Met Val Lys Met Ile Ile Leu Met Lys Asn
690 695 700
Asn Asp Leu Thr Asn Phe Phe Thr His Thr His Phe Val Arg Leu Ala
705 710 715 720
Glu Ile Ile Asn Arg Ile Cys Leu Pro Arg Gln Tyr Leu Lys Ala Arg
725 730 735
Arg Asn Asp Glu Lys Glu Lys Thr Ile Lys Ser Met Glu Lys Glu Met
740 745 750
Gly Lys Met Val Glu Leu Ala Leu Ser Glu Ser Asp Thr Phe Arg Asp
755 760 765
Val Ser Ile Thr Phe Leu Asp Val Ala Lys Ala Phe Tyr Tyr Phe Ala
770 775 780
Leu Cys Gly Asp His Leu Gln Thr His Ile Ser Lys Val Leu Phe Gln
785 790 795 800
Lys Val

<210> 161
<211> 2859
<212> DNA
<213> Gibberella fujikuroi

<400> 161
atgcctggta aaattgaaaa tggtacccca aaggacctca agactggaaa tgattttgtt 60
tctgctgcta agagtttact agatcgagct ttcaaaagtc atcattccta ctacggatta 120
tgctcaactt catgtcaagt ttatgataca gcttgggttg caatgattcc aaaaacaaga 180
gataatgtaa aacagtggtt gtttccagaa tgtttccatt acctcttaaa aacacaagcc 240
gcagatggct catggggttc attgcctaca acacagacag cgggtatcct agatacagcc 300
tcagctgtgc tggcattatt gtgccacgca caagagcctt tacaaatatt ggatgtatct 360
ccagatgaaa tggggttgag aatagaacac ggtgtcacat ccttgaaacg tcaattagca 420
gtttggaatg atgtggagga caccaaccat attggcgtcg agtttatcat accagcctta 480
ctttccatgc tagaaaagga attagatgtt ccatcttttg aatttccatg taggtccatc 540

ttagagagaa tgcacgggga gaaattaggt catttcgacc tggaacaagt ttacggcaag 600
ccaagctcat tgttgcactc attggaagca tttctcggta agctagattt tgatcgacta 660
tcacatcacc tataccacgg cagtatgatg gcatctccat cttcaacggc tgcttatctt 720
attggggcta caaaatggga tgacgaagcc gaagattacc taagacatgt aatgcgtaat 780
ggtgcaggac atgggaatgg aggtatttct ggtacatttc caactactca tttcgaatgt 840
agctggatta tagcaacgtt gttaaaggtt ggctttactt tgaagcaaat tgacggcgat 900
ggcttaagag gtttatcaac catcttactt gaggcgcttc gtgatgagaa tggtgtcata 960
ggctttgccc ctagaacagc agatgtagat gacacagcca aagctctatt ggccttgtca 1020
ttggtaaacc agccagtgtc acctgatatc atgattaagg tctttgaggg caaagaccat 1080
tttaccactt ttggttcaga aagagatcca tcattgactt ccaacctgca cgtcctttta 1140
tctttactta aacaatctaa cttgtctcaa taccatcctc aaatcctcaa aacaacatta 1200
ttcacttgta gatggtggtg gggttccgat cattgtgtca aagacaaatg gaatttgagt 1260
cacctatatc caactatgtt gttggttgaa gccttcactg aagtgctcca tctcattgac 1320
ggtggtgaat tgtctagtct gtttgatgaa tcctttaagt gtaagattgg tcttagcatc 1380
tttcaagcgg tacttagaat aatcctcacc caagacaacg acggctcttg gagaggatac 1440
agagaacaga cgtgttacgc aatattggct ttagttcaag cgagacatgt atgctttttc 1500
actcacatgg ttgacagact gcaatcatgt gttgatcgag gtttctcatg gttgaaatct 1560
tgctcttttc attctcaaga cctgacttgg acctctaaaa cagcttatga agtgggtttc 1620
gtagctgaag catataaact agctgcttta caatctgctt ccctggaggt tcctgctgcc 1680
accattggac attctgtcac gtctgccgtt ccatcaagtg atcttgaaaa atacatgaga 1740
ttggtgagaa aaactgcgtt attctctcca ctggatgagt ggggtctaat ggcttctatc 1800
atcgaatctt catttttcgt accattactg caggcacaaa gagttgaaat ataccctaga 1860
gataatatca aggtggacga agataagtac ttgtctatta tcccattcac atgggtcgga 1920
tgcaataata ggtctagaac tttcgcaagt aacagatggc tatacgatat gatgtacctt 1980
tcattactcg gctatcaaac cgacgagtac atggaagctg tagctgggcc agtgtttggg 2040
gatgtttcct tgttacatca aacaattgat aaggtgattg ataatacaat gggtaacctt 2100
gcgagagcca atggaacagt acacagtggt aatggacatc agcacgaatc tcctaatata 2160
ggtcaagtcg aggacacctt gactcgtttc acaaattcag tcttgaatca caaagacgtc 2220
cttaactcta gctcatctga tcaagatact ttgagaagag agtttagaac attcatgcac 2280
gctcatataa cacaaatcga agataactca cgattcagta agcaagcctc atccgatgcg 2340
ttttcctctc ctgaacaatc ttactttcaa tgggtgaact caactggtgg ctcacatgtc 2400
gcttgcgcct attcatttgc cttctctaat tgcctcatgt ctgcaaattt gttgcagggt 2460
aaagacgcat ttccaagcgg aacgcaaaag tacttaatct cctctgttat gagacatgcc 2520
acaaacatgt gtagaatgta taacgacttt ggctctattg ccagagacaa cgctgagaga 2580
aatgttaata gtattcattt tcctgagttt actctctgta acggaacttc tcaaaaccta 2640
gatgaaagga aggaaagact tctgaaaatc gcaacttacg aacaagggta tttggataga 2700
gcactagagg ccttggaaag acagagtaga gatgatgccg gagacagagc tggatctaaa 2760
gatatgagaa agttgaaaat cgttaagtta ttctgtgatg ttacggactt atacgatcag 2820
ctctacgtta tcaaagattt gtcatcctct atgaagtaa 2859

<210> 162
<211> 952
<212> PRT
<213> Gibberella fujikuroi

<400> 162
Met Pro Gly Lys Ile Glu Asn Gly Thr Pro Lys Asp Leu Lys Thr Gly
 1   5   10   15
Asn Asp Phe Val Ser Ala Ala Lys Ser Leu Leu Asp Arg Ala Phe Lys
            20   25   30
Ser His His Ser Tyr Tyr Gly Leu Cys Ser Thr Ser Cys Gln Val Tyr
        35   40   45
Asp Thr Ala Trp Val Ala Met Ile Pro Lys Thr Arg Asp Asn Val Lys
    50   55   60
Gln Trp Leu Phe Pro Glu Cys Phe His Tyr Leu Leu Lys Thr Gln Ala
65   70   75   80
Ala Asp Gly Ser Trp Gly Ser Leu Pro Thr Thr Gln Thr Ala Gly Ile
                85   90   95
Leu Asp Thr Ala Ser Ala Val Leu Ala Leu Leu Cys His Ala Gln Glu
            100   105   110
Pro Leu Gln Ile Leu Asp Val Ser Pro Asp Glu Met Gly Leu Arg Ile
        115   120   125
Glu His Gly Val Thr Ser Leu Lys Arg Gln Leu Ala Val Trp Asn Asp

130 135 140

Val Glu Asp Thr Asn His Ile Gly Val Glu Phe Ile Ile Pro Ala Leu
145 150 155 160
Leu Ser Met Leu Glu Lys Glu Leu Asp Val Pro Ser Phe Glu Phe Pro
165 170 175
Cys Arg Ser Ile Leu Glu Arg Met His Gly Glu Lys Leu Gly His Phe
180 185 190
Asp Leu Glu Gln Val Tyr Gly Lys Pro Ser Ser Leu Leu His Ser Leu
195 200 205
Glu Ala Phe Leu Gly Lys Leu Asp Phe Asp Arg Leu Ser His His Leu
210 215 220
Tyr His Gly Ser Met Met Ala Ser Pro Ser Ser Thr Ala Ala Tyr Leu
225 230 235 240
Ile Gly Ala Thr Lys Trp Asp Asp Glu Ala Glu Asp Tyr Leu Arg His
245 250 255
Val Met Arg Asn Gly Ala Gly His Gly Asn Gly Gly Ile Ser Gly Thr
260 265 270
Phe Pro Thr Thr His Phe Glu Cys Ser Trp Ile Ile Ala Thr Leu Leu
275 280 285
Lys Val Gly Phe Thr Leu Lys Gln Ile Asp Gly Asp Gly Leu Arg Gly
290 295 300
Leu Ser Thr Ile Leu Leu Glu Ala Leu Arg Asp Glu Asn Gly Val Ile
305 310 315 320
Gly Phe Ala Pro Arg Thr Ala Asp Val Asp Asp Thr Ala Lys Ala Leu
325 330 335
Leu Ala Leu Ser Leu Val Asn Gln Pro Val Ser Pro Asp Ile Met Ile
340 345 350
Lys Val Phe Glu Gly Lys Asp His Phe Thr Thr Phe Gly Ser Glu Arg
355 360 365
Asp Pro Ser Leu Thr Ser Asn Leu His Val Leu Leu Ser Leu Leu Lys
370 375 380
Gln Ser Asn Leu Ser Gln Tyr His Pro Gln Ile Leu Lys Thr Thr Leu
385 390 395 400
Phe Thr Cys Arg Trp Trp Trp Gly Ser Asp His Cys Val Lys Asp Lys
405 410 415
Trp Asn Leu Ser His Leu Tyr Pro Thr Met Leu Leu Val Glu Ala Phe
420 425 430
Thr Glu Val Leu His Leu Ile Asp Gly Gly Glu Leu Ser Ser Leu Phe
435 440 445
Asp Glu Ser Phe Lys Cys Lys Ile Gly Leu Ser Ile Phe Gln Ala Val
450 455 460
Leu Arg Ile Ile Leu Thr Gln Asp Asn Asp Gly Ser Trp Arg Gly Tyr
465 470 475 480
Arg Glu Gln Thr Cys Tyr Ala Ile Leu Ala Leu Val Gln Ala Arg His
485 490 495
Val Cys Phe Phe Thr His Met Val Asp Arg Leu Gln Ser Cys Val Asp
500 505 510
Arg Gly Phe Ser Trp Leu Lys Ser Cys Ser Phe His Ser Gln Asp Leu
515 520 525
Thr Trp Thr Ser Lys Thr Ala Tyr Glu Val Gly Phe Val Ala Glu Ala
530 535 540
Tyr Lys Leu Ala Ala Leu Gln Ser Ala Ser Leu Glu Val Pro Ala Ala
545 550 555 560
Thr Ile Gly His Ser Val Thr Ser Ala Val Pro Ser Ser Asp Leu Glu
565 570 575
Lys Tyr Met Arg Leu Val Arg Lys Thr Ala Leu Phe Ser Pro Leu Asp
580 585 590
Glu Trp Gly Leu Met Ala Ser Ile Ile Glu Ser Ser Phe Phe Val Pro
595 600 605
Leu Leu Gln Ala Gln Arg Val Glu Ile Tyr Pro Arg Asp Asn Ile Lys
610 615 620
Val Asp Glu Asp Lys Tyr Leu Ser Ile Ile Pro Phe Thr Trp Val Gly
625 630 635 640

```
Cys Asn Asn Arg Ser Arg Thr Phe Ala Ser Asn Arg Trp Leu Tyr Asp
                645                 650                 655
Met Met Tyr Leu Ser Leu Leu Gly Tyr Gln Thr Asp Glu Tyr Met Glu
            660                 665                 670
Ala Val Ala Gly Pro Val Phe Gly Asp Val Ser Leu Leu His Gln Thr
        675                 680                 685
Ile Asp Lys Val Ile Asp Asn Thr Met Gly Asn Leu Ala Arg Ala Asn
    690                 695                 700
Gly Thr Val His Ser Gly Asn Gly His Gln His Glu Ser Pro Asn Ile
705                 710                 715                 720
Gly Gln Val Glu Asp Thr Leu Thr Arg Phe Thr Asn Ser Val Leu Asn
                725                 730                 735
His Lys Asp Val Leu Asn Ser Ser Ser Ser Asp Gln Asp Thr Leu Arg
            740                 745                 750
Arg Glu Phe Arg Thr Phe Met His Ala His Ile Thr Gln Ile Glu Asp
        755                 760                 765
Asn Ser Arg Phe Ser Lys Gln Ala Ser Ser Asp Ala Phe Ser Ser Pro
    770                 775                 780
Glu Gln Ser Tyr Phe Gln Trp Val Asn Ser Thr Gly Gly Ser His Val
785                 790                 795                 800
Ala Cys Ala Tyr Ser Phe Ala Phe Ser Asn Cys Leu Met Ser Ala Asn
                805                 810                 815
Leu Leu Gln Gly Lys Asp Ala Phe Pro Ser Gly Thr Gln Lys Tyr Leu
            820                 825                 830
Ile Ser Ser Val Met Arg His Ala Thr Asn Met Cys Arg Met Tyr Asn
        835                 840                 845
Asp Phe Gly Ser Ile Ala Arg Asp Asn Ala Glu Arg Asn Val Asn Ser
    850                 855                 860
Ile His Phe Pro Glu Phe Thr Leu Cys Asn Gly Thr Ser Gln Asn Leu
865                 870                 875                 880
Asp Glu Arg Lys Glu Arg Leu Leu Lys Ile Ala Thr Tyr Glu Gln Gly
                885                 890                 895
Tyr Leu Asp Arg Ala Leu Glu Ala Leu Glu Arg Gln Ser Arg Asp Asp
            900                 905                 910
Ala Gly Asp Arg Ala Gly Ser Lys Asp Met Arg Lys Leu Lys Ile Val
        915                 920                 925
Lys Leu Phe Cys Asp Val Thr Asp Leu Tyr Asp Gln Leu Tyr Val Ile
    930                 935                 940
Lys Asp Leu Ser Ser Ser Met Lys
945                 950
```

<210> 163
<211> 1503
<212> DNA
<213> Stevia rebaudiana

<400> 163

```
atggaggctt catatctata catttccatt cttctgcttc tagcttcgta cctcttcacc 60
acccaacttc gtcgtaaatc cgccaatcta ccgccgacgg tgttcccatc catccccata 120
atcggtcatc tctacctcct caaaaaacca ctctatagaa cactagccaa aattgccgcc 180
aaatacggcc ctatcctcca actccaacta gggtaccgcc gtgtcctcgt aatctcctcc 240
ccttccgccg ccgaagaatg cttcaccaac aacgacgtta tcttcgccaa ccgtccgaag 300
acgctattcg gaaaaattgt agggggtacc agcctcgggt cgctgtcgta cggcgaccag 360
tggcgcaacc tccgccgcgt tgcatccatc gagattctat cagtccaccg gctcaacgag 420
tttcacgaca tacgtgttga cgaaaaccgg cttctgatcc gtaaactaag atccagttct 480
tctccggtga ctctgataac ggtgttttac gcactaacgt taaatgtgat tatgagaatg 540
atctccggaa agaggtattt cgactcgggt gatcgggaat tggaggagga agggaagcga 600
ttccgggaga tactcgatga gacacttttg ctcgcgggtg cttctaatgt tggggattac 660
ttgccgattc tgaattggtt gggggtgaag agcttggaga agaagctaat cgcattgcag 720
aaaaagagag atgatttctt tcagggactg atcgagcaag ttcggaaatc tagaggggct 780
aaagtgggaa aaggaaggaa gacgatgatc gagttgttgt tatccctaca agaatctgaa 840
cctgagtatt acactgacgc catgatccga tcatttgtgc tgggtttatt agcagcaggg 900
```

agtgatacat cggctggaac tatggaatgg gcgatgtctc ttttggtaaa ccacccgcac 960
gtattaaaaa aggcacaggc tgaaattgat cgagtcatcg gcaacaaccg tctaattgat 1020
gagtccgaca tagggaatat accttacatt ggttgcatca taaacgagac gctcagattg 1080
taccctgcgg gcccgttgct atttccccat gagtcatcag cggactgtgt tatcagcggg 1140
tacaacatcc ctcgtgggac gatgcttatt gtcaaccaat gggcgataca tcatgaccca 1200
aaggtgtggg acgaccctga gacattcaaa ccggaaagat tccaagggct tgaagggaca 1260
cgagacgggt ttaagctgat gccttttggg tccggaagga ggggttgtcc aggggaggga 1320
ttggcgattc gtttgcttgg gatgactctc gggtcggtta tccaatgctt tgattgggaa 1380
cgagtcggtg acgagatggt tgatatgacc gaaggtcttg gggtcacgtt gcctaaagct 1440
gtaccattag tcgccaagtg caagccgcgt tccgaaatga cgaatctact ctctgagcta 1500
tga 1503

<210> 164
<211> 500
<212> PRT
<213> Stevia rebaudiana

<400> 164
Met Glu Ala Ser Tyr Leu Tyr Ile Ser Ile Leu Leu Leu Leu Ala Ser
 1   5   10   15
Tyr Leu Phe Thr Thr Gln Leu Arg Arg Lys Ser Ala Asn Leu Pro Pro
            20   25   30
Thr Val Phe Pro Ser Ile Pro Ile Ile Gly His Leu Tyr Leu Leu Lys
        35   40   45
Lys Pro Leu Tyr Arg Thr Leu Ala Lys Ile Ala Ala Lys Tyr Gly Pro
    50   55   60
Ile Leu Gln Leu Gln Leu Gly Tyr Arg Arg Val Leu Val Ile Ser Ser
65   70   75   80
Pro Ser Ala Ala Glu Glu Cys Phe Thr Asn Asn Asp Val Ile Phe Ala
                85   90   95
Asn Arg Pro Lys Thr Leu Phe Gly Lys Ile Val Gly Gly Thr Ser Leu
            100   105   110
Gly Ser Leu Ser Tyr Gly Asp Gln Trp Arg Asn Leu Arg Arg Val Ala
        115   120   125
Ser Ile Glu Ile Leu Ser Val His Arg Leu Asn Glu Phe His Asp Ile
    130   135   140
Arg Val Asp Glu Asn Arg Leu Leu Ile Arg Lys Leu Arg Ser Ser Ser
145   150   155   160
Ser Pro Val Thr Leu Ile Thr Val Phe Tyr Ala Leu Thr Leu Asn Val
                165   170   175
Ile Met Arg Met Ile Ser Gly Lys Arg Tyr Phe Asp Ser Gly Asp Arg
            180   185   190
Glu Leu Glu Glu Glu Gly Lys Arg Phe Arg Glu Ile Leu Asp Glu Thr
        195   200   205
Leu Leu Leu Ala Gly Ala Ser Asn Val Gly Asp Tyr Leu Pro Ile Leu
    210   215   220
Asn Trp Leu Gly Val Lys Ser Leu Glu Lys Lys Leu Ile Ala Leu Gln
225   230   235   240
Lys Lys Arg Asp Asp Phe Phe Gln Gly Leu Ile Glu Gln Val Arg Lys
                245   250   255
Ser Arg Gly Ala Lys Val Gly Lys Gly Arg Lys Thr Met Ile Glu Leu
            260   265   270
Leu Leu Ser Leu Gln Glu Ser Glu Pro Glu Tyr Tyr Thr Asp Ala Met
        275   280   285
Ile Arg Ser Phe Val Leu Gly Leu Leu Ala Ala Gly Ser Asp Thr Ser
    290   295   300
Ala Gly Thr Met Glu Trp Ala Met Ser Leu Leu Val Asn His Pro His
305   310   315   320
Val Leu Lys Lys Ala Gln Ala Glu Ile Asp Arg Val Ile Gly Asn Asn
                325   330   335
Arg Leu Ile Asp Glu Ser Asp Ile Gly Asn Ile Pro Tyr Ile Gly Cys
            340   345   350
Ile Ile Asn Glu Thr Leu Arg Leu Tyr Pro Ala Gly Pro Leu Leu Phe

```
        355                 360                 365
Pro His Glu Ser Ser Ala Asp Cys Val Ile Ser Gly Tyr Asn Ile Pro
    370                 375                 380
Arg Gly Thr Met Leu Ile Val Asn Gln Trp Ala Ile His His Asp Pro
385                 390                 395                 400
Lys Val Trp Asp Asp Pro Glu Thr Phe Lys Pro Glu Arg Phe Gln Gly
                405                 410                 415
Leu Glu Gly Thr Arg Asp Gly Phe Lys Leu Met Pro Phe Gly Ser Gly
            420                 425                 430
Arg Arg Gly Cys Pro Gly Glu Gly Leu Ala Ile Arg Leu Leu Gly Met
        435                 440                 445
Thr Leu Gly Ser Val Ile Gln Cys Phe Asp Trp Glu Arg Val Gly Asp
    450                 455                 460
Glu Met Val Asp Met Thr Glu Gly Leu Gly Val Thr Leu Pro Lys Ala
465                 470                 475                 480
Val Pro Leu Val Ala Lys Cys Lys Pro Arg Ser Glu Met Thr Asn Leu
                485                 490                 495
Leu Ser Glu Leu
            500
```

<210> 165
<211> 1503
<212> DNA
<213> Stevia rebaudiana

<400> 165

```
atggaagcct cttacctata catttctatt ttgcttttac tggcatcata cctgttcacc 60
actcaactta gaaggaagag cgctaatcta ccaccaaccg tgtttccatc aataccaatc 120
attggacact tatacttact caaaaagcct ctttatagaa ctttagcaaa aattgccgct 180
aagtacggac caatactgca attacaactc ggctacagac gtgttctggt gatttcctca 240
ccatcagcag cagaagagtg ctttaccaat aacgatgtaa tcttcgcaaa tagacctaag 300
acattgtttg gcaaaatagt gggtggaaca tcccttggca gtttatccta cggcgatcaa 360
tggcgtaatc taaggagagt agcttctatc gaaatcctat cagttcatag gttgaacgaa 420
tttcatgata tcagagtgga tgagaacaga ttgttaatta gaaaacttag aagttcatct 480
tctcctgtta ctcttataac agtcttttat gctctaacat tgaacgtcat tatgagaatg 540
atctctggca aaagatattt cgacagtggg gatagagaat tggaggagga aggtaagaga 600
tttcgagaaa tcttagacga aacgttgctt ctagccggtg cttctaatgt tggcgactac 660
ttaccaatat tgaactggtt gggagttaag tctcttgaaa agaaattgat cgctttgcag 720
aaaaagagag atgacttttt ccagggtttg attgaacagg ttagaaaatc tcgtggtgct 780
aaagtaggca aaggtagaaa aacgatgatc gaactcttat tatctttgca agagtcagaa 840
cctgagtact atacagatgc tatgataaga tcttttgtcc taggtctgct ggctgcaggt 900
agtgatactt cagcgggcac tatggaatgg gccatgagct tactggtcaa tcacccacat 960
gtattgaaga aagctcaagc tgaaatcgat agagttatcg gtaataacag attgattgac 1020
gagtcagaca ttggaaatat cccttacatc gggtgtatta tcaatgaaac tctaagactc 1080
tatccagcag ggccattgtt gttcccacat gaaagttctg ccgactgcgt tatttccggt 1140
tacaatatac ctagaggtac aatgttaatc gtaaaccaat gggcgattca tcacgatcct 1200
aaagtctggg atgatcctga aacctttaaa cctgaaagat ttcaaggatt agaaggaact 1260
agagatggtt tcaaacttat gccattcggt tctgggagaa gaggatgtcc aggtgaaggt 1320
ttggcaataa ggctgttagg gatgacacta ggctcagtga tccaatgttt tgattgggag 1380
agagtaggag atgagatggt tgacatgaca gaaggtttgg gtgtcacact tcctaaggcc 1440
gttccattag ttgccaaatg taagccacgt tccgaaatga ctaatctcct atccgaactt 1500
taa                                                               1503
```

<210> 166
<211> 691
<212> PRT
<213> Saccharomyces cerevisiae

<400> 166

```
Met Pro Phe Gly Ile Asp Asn Thr Asp Phe Thr Val Leu Ala Gly Leu
 1               5                  10                  15
Val Leu Ala Val Leu Leu Tyr Val Lys Arg Asn Ser Ile Lys Glu Leu
```

```
            20                  25                  30
Leu Met Ser Asp Asp Gly Asp Ile Thr Ala Val Ser Ser Gly Asn Arg
        35                  40                  45
Asp Ile Ala Gln Val Val Thr Glu Asn Asn Lys Asn Tyr Leu Val Leu
    50                  55                  60
Tyr Ala Ser Gln Thr Gly Thr Ala Glu Asp Tyr Ala Lys Lys Phe Ser
65                  70                  75                  80
Lys Glu Leu Val Ala Lys Phe Asn Leu Asn Val Met Cys Ala Asp Val
                85                  90                  95
Glu Asn Tyr Asp Phe Glu Ser Leu Asn Asp Val Pro Val Ile Val Ser
            100                 105                 110
Ile Phe Ile Ser Thr Tyr Gly Glu Gly Asp Phe Pro Asp Gly Ala Val
        115                 120                 125
Asn Phe Glu Asp Phe Ile Cys Asn Ala Glu Ala Gly Ala Leu Ser Asn
    130                 135                 140
Leu Arg Tyr Asn Met Phe Gly Leu Gly Asn Ser Thr Tyr Glu Phe Phe
145                 150                 155                 160
Asn Gly Ala Ala Lys Lys Ala Glu Lys His Leu Ser Ala Ala Gly Ala
                165                 170                 175
Ile Arg Leu Gly Lys Leu Gly Glu Ala Asp Asp Gly Ala Gly Thr Thr
            180                 185                 190
Asp Glu Asp Tyr Met Ala Trp Lys Asp Ser Ile Leu Glu Val Leu Lys
        195                 200                 205
Asp Glu Leu His Leu Asp Glu Gln Glu Ala Lys Phe Thr Ser Gln Phe
    210                 215                 220
Gln Tyr Thr Val Leu Asn Glu Ile Thr Asp Ser Met Ser Leu Gly Glu
225                 230                 235                 240
Pro Ser Ala His Tyr Leu Pro Ser His Gln Leu Asn Arg Asn Ala Asp
                245                 250                 255
Gly Ile Gln Leu Gly Pro Phe Asp Leu Ser Gln Pro Tyr Ile Ala Pro
            260                 265                 270
Ile Val Lys Ser Arg Glu Leu Phe Ser Ser Asn Asp Arg Asn Cys Ile
        275                 280                 285
His Ser Glu Phe Asp Leu Ser Gly Ser Asn Ile Lys Tyr Ser Thr Gly
    290                 295                 300
Asp His Leu Ala Val Trp Pro Ser Asn Pro Leu Glu Lys Val Glu Gln
305                 310                 315                 320
Phe Leu Ser Ile Phe Asn Leu Asp Pro Glu Thr Ile Phe Asp Leu Lys
                325                 330                 335
Pro Leu Asp Pro Thr Val Lys Val Pro Phe Pro Thr Pro Thr Thr Ile
            340                 345                 350
Gly Ala Ala Ile Lys His Tyr Leu Glu Ile Thr Gly Pro Val Ser Arg
        355                 360                 365
Gln Leu Phe Ser Ser Leu Ile Gln Phe Ala Pro Asn Ala Asp Val Lys
    370                 375                 380
Glu Lys Leu Thr Leu Leu Ser Lys Asp Lys Asp Gln Phe Ala Val Glu
385                 390                 395                 400
Ile Thr Ser Lys Tyr Phe Asn Ile Ala Asp Ala Leu Lys Tyr Leu Ser
                405                 410                 415
Asp Gly Ala Lys Trp Asp Thr Val Pro Met Gln Phe Leu Val Glu Ser
            420                 425                 430
Val Pro Gln Met Thr Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Ser Leu
        435                 440                 445
Ser Glu Lys Gln Thr Val His Val Thr Ser Ile Val Glu Asn Phe Pro
    450                 455                 460
Asn Pro Glu Leu Pro Asp Ala Pro Pro Val Val Gly Val Thr Thr Asn
465                 470                 475                 480
Leu Leu Arg Asn Ile Gln Leu Ala Gln Asn Asn Val Asn Ile Ala Glu
                485                 490                 495
Thr Asn Leu Pro Val His Tyr Asp Leu Asn Gly Pro Arg Lys Leu Phe
            500                 505                 510
Ala Asn Tyr Lys Leu Pro Val His Val Arg Arg Ser Asn Phe Arg Leu
        515                 520                 525
```

Pro Ser Asn Pro Ser Thr Pro Val Ile Met Ile Gly Pro Gly Thr Gly
530 535 540
Val Ala Pro Phe Arg Gly Phe Ile Arg Glu Arg Val Ala Phe Leu Glu
545 550 555 560
Ser Gln Lys Lys Gly Gly Asn Asn Val Ser Leu Gly Lys His Ile Leu
565 570 575
Phe Tyr Gly Ser Arg Asn Thr Asp Asp Phe Leu Tyr Gln Asp Glu Trp
580 585 590
Pro Glu Tyr Ala Lys Lys Leu Asp Gly Ser Phe Glu Met Val Val Ala
595 600 605
His Ser Arg Leu Pro Asn Thr Lys Lys Val Tyr Val Gln Asp Lys Leu
610 615 620
Lys Asp Tyr Glu Asp Gln Val Phe Glu Met Ile Asn Asn Gly Ala Phe
625 630 635 640
Ile Tyr Val Cys Gly Asp Ala Lys Gly Met Ala Lys Gly Val Ser Thr
645 650 655
Ala Leu Val Gly Ile Leu Ser Arg Gly Lys Ser Ile Thr Thr Asp Glu
660 665 670
Ala Thr Glu Leu Ile Lys Met Leu Lys Thr Ser Gly Arg Tyr Gln Glu
675 680 685
Asp Val Trp
690

<210> 167
<211> 335
<212> PRT
<213> Saccharomyces cerevisiae

<400> 167
Met Glu Ala Lys Ile Asp Glu Leu Ile Asn Asn Asp Pro Val Trp Ser
1 5 10 15
Ser Gln Asn Glu Ser Leu Ile Ser Lys Pro Tyr Asn His Ile Leu Leu
20 25 30
Lys Pro Gly Lys Asn Phe Arg Leu Asn Leu Ile Val Gln Ile Asn Arg
35 40 45
Val Met Asn Leu Pro Lys Asp Gln Leu Ala Ile Val Ser Gln Ile Val
50 55 60
Glu Leu Leu His Asn Ser Ser Leu Leu Ile Asp Asp Ile Glu Asp Asn
65 70 75 80
Ala Pro Leu Arg Arg Gly Gln Thr Thr Ser His Leu Ile Phe Gly Val
85 90 95
Pro Ser Thr Ile Asn Thr Ala Asn Tyr Met Tyr Phe Arg Ala Met Gln
100 105 110
Leu Val Ser Gln Leu Thr Thr Lys Glu Pro Leu Tyr His Asn Leu Ile
115 120 125
Thr Ile Phe Asn Glu Glu Leu Ile Asn Leu His Arg Gly Gln Gly Leu
130 135 140
Asp Ile Tyr Trp Arg Asp Phe Leu Pro Glu Ile Ile Pro Thr Gln Glu
145 150 155 160
Met Tyr Leu Asn Met Val Met Asn Lys Thr Gly Gly Leu Phe Arg Leu
165 170 175
Thr Leu Arg Leu Met Glu Ala Leu Ser Pro Ser Ser His His Gly His
180 185 190
Ser Leu Val Pro Phe Ile Asn Leu Leu Gly Ile Ile Tyr Gln Ile Arg
195 200 205
Asp Asp Tyr Leu Asn Leu Lys Asp Phe Gln Met Ser Ser Glu Lys Gly
210 215 220
Phe Ala Glu Asp Ile Thr Glu Gly Lys Leu Ser Phe Pro Ile Val His
225 230 235 240
Ala Leu Asn Phe Thr Lys Thr Lys Gly Gln Thr Glu Gln His Asn Glu
245 250 255
Ile Leu Arg Ile Leu Leu Leu Arg Thr Ser Asp Lys Asp Ile Lys Leu

```
            260                 265                 270
Lys Leu Ile Gln Ile Leu Glu Phe Asp Thr Asn Ser Leu Ala Tyr Thr
        275                 280                 285
Lys Asn Phe Ile Asn Gln Leu Val Asn Met Ile Lys Asn Asp Asn Glu
    290                 295                 300
Asn Lys Tyr Leu Pro Asp Leu Ala Ser His Ser Asp Thr Ala Thr Asn
305                 310                 315                 320
Leu His Asp Glu Leu Leu Tyr Ile Ile Asp His Leu Ser Glu Leu
                325                 330                 335


<210> 168
<211> 712
<212> PRT
<213> Arabidopsis thaliana

<400> 168
Met Ser Ser Ser Ser Ser Ser Ser Thr Ser Met Ile Asp Leu Met Ala
 1               5                  10                  15
Ala Ile Ile Lys Gly Glu Pro Val Ile Val Ser Asp Pro Ala Asn Ala
            20                  25                  30
Ser Ala Tyr Glu Ser Val Ala Ala Glu Leu Ser Ser Met Leu Ile Glu
        35                  40                  45
Asn Arg Gln Phe Ala Met Ile Val Thr Thr Ser Ile Ala Val Leu Ile
    50                  55                  60
Gly Cys Ile Val Met Leu Val Trp Arg Arg Ser Gly Ser Gly Asn Ser
65                  70                  75                  80
Lys Arg Val Glu Pro Leu Lys Pro Leu Val Ile Lys Pro Arg Glu Glu
                85                  90                  95
Glu Ile Asp Asp Gly Arg Lys Lys Val Thr Ile Phe Phe Gly Thr Gln
            100                 105                 110
Thr Gly Thr Ala Glu Gly Phe Ala Lys Ala Leu Gly Glu Glu Ala Lys
        115                 120                 125
Ala Arg Tyr Glu Lys Thr Arg Phe Lys Ile Val Asp Leu Asp Asp Tyr
    130                 135                 140
Ala Ala Asp Asp Asp Glu Tyr Glu Glu Lys Leu Lys Lys Glu Asp Val
145                 150                 155                 160
Ala Phe Phe Phe Leu Ala Thr Tyr Gly Asp Gly Glu Pro Thr Asp Asn
                165                 170                 175
Ala Ala Arg Phe Tyr Lys Trp Phe Thr Glu Gly Asn Asp Arg Gly Glu
            180                 185                 190
Trp Leu Lys Asn Leu Lys Tyr Gly Val Phe Gly Leu Gly Asn Arg Gln
        195                 200                 205
Tyr Glu His Phe Asn Lys Val Ala Lys Val Val Asp Asp Ile Leu Val
    210                 215                 220
Glu Gln Gly Ala Gln Arg Leu Val Gln Val Gly Leu Gly Asp Asp Asp
225                 230                 235                 240
Gln Cys Ile Glu Asp Asp Phe Thr Ala Trp Arg Glu Ala Leu Trp Pro
                245                 250                 255
Glu Leu Asp Thr Ile Leu Arg Glu Glu Gly Asp Thr Ala Val Ala Thr
            260                 265                 270
Pro Tyr Thr Ala Ala Val Leu Glu Tyr Arg Val Ser Ile His Asp Ser
        275                 280                 285
Glu Asp Ala Lys Phe Asn Asp Ile Thr Leu Ala Asn Gly Asn Gly Tyr
    290                 295                 300
Thr Val Phe Asp Ala Gln His Pro Tyr Lys Ala Asn Val Ala Val Lys
305                 310                 315                 320
Arg Glu Leu His Thr Pro Glu Ser Asp Arg Ser Cys Ile His Leu Glu
                325                 330                 335
Phe Asp Ile Ala Gly Ser Gly Leu Thr Met Lys Leu Gly Asp His Val
            340                 345                 350
Gly Val Leu Cys Asp Asn Leu Ser Glu Thr Val Asp Glu Ala Leu Arg
        355                 360                 365
```

Leu Leu Asp Met Ser Pro Asp Thr Tyr Phe Ser Leu His Ala Glu Lys
370 375 380
Glu Asp Gly Thr Pro Ile Ser Ser Ser Leu Pro Pro Pro Phe Pro Pro
385 390 395 400
Cys Asn Leu Arg Thr Ala Leu Thr Arg Tyr Ala Cys Leu Leu Ser Ser
405 410 415
Pro Lys Lys Ser Ala Leu Val Ala Leu Ala Ala His Ala Ser Asp Pro
420 425 430
Thr Glu Ala Glu Arg Leu Lys His Leu Ala Ser Pro Ala Gly Lys Asp
435 440 445
Glu Tyr Ser Lys Trp Val Val Glu Ser Gln Arg Ser Leu Leu Glu Val
450 455 460
Met Ala Glu Phe Pro Ser Ala Lys Pro Pro Leu Gly Val Phe Phe Ala
465 470 475 480
Gly Val Ala Pro Arg Leu Gln Pro Arg Phe Tyr Ser Ile Ser Ser Ser
485 490 495
Pro Lys Ile Ala Glu Thr Arg Ile His Val Thr Cys Ala Leu Val Tyr
500 505 510
Glu Lys Met Pro Thr Gly Arg Ile His Lys Gly Val Cys Ser Thr Trp
515 520 525
Met Lys Asn Ala Val Pro Tyr Glu Lys Ser Glu Lys Leu Phe Leu Gly
530 535 540
Arg Pro Ile Phe Val Arg Gln Ser Asn Phe Lys Leu Pro Ser Asp Ser
545 550 555 560
Lys Val Pro Ile Ile Met Ile Gly Pro Gly Thr Gly Leu Ala Pro Phe
565 570 575
Arg Gly Phe Leu Gln Glu Arg Leu Ala Leu Val Glu Ser Gly Val Glu
580 585 590
Leu Gly Pro Ser Val Leu Phe Phe Gly Cys Arg Asn Arg Arg Met Asp
595 600 605
Phe Ile Tyr Glu Glu Glu Leu Gln Arg Phe Val Glu Ser Gly Ala Leu
610 615 620
Ala Glu Leu Ser Val Ala Phe Ser Arg Glu Gly Pro Thr Lys Glu Tyr
625 630 635 640
Val Gln His Lys Met Met Asp Lys Ala Ser Asp Ile Trp Asn Met Ile
645 650 655
Ser Gln Gly Ala Tyr Leu Tyr Val Cys Gly Asp Ala Lys Gly Met Ala
660 665 670
Arg Asp Val His Arg Ser Leu His Thr Ile Ala Gln Glu Gln Gly Ser
675 680 685
Met Asp Ser Thr Lys Ala Glu Gly Phe Val Lys Asn Leu Gln Thr Ser
690 695 700
Gly Arg Tyr Leu Arg Asp Val Trp
705 710

<210> 169
<211> 709
<212> PRT
<213> Stevia rebaudiana

<400> 169
Met Gln Ser Glu Ser Val Glu Ala Ser Thr Ile Asp Leu Met Thr Ala
1 5 10 15
Val Leu Lys Asp Thr Val Ile Asp Thr Ala Asn Ala Ser Asp Asn Gly
20 25 30
Asp Ser Lys Met Pro Pro Ala Leu Ala Met Met Phe Glu Ile Arg Asp
35 40 45
Leu Leu Leu Ile Leu Thr Thr Ser Val Ala Val Leu Val Gly Cys Phe
50 55 60
Val Val Leu Val Trp Lys Arg Ser Ser Gly Lys Lys Ser Gly Lys Glu
65 70 75 80
Leu Glu Pro Pro Lys Ile Val Val Pro Lys Arg Arg Leu Glu Gln Glu

```
                85                  90                  95
Val Asp Asp Gly Lys Lys Lys Val Thr Ile Phe Phe Gly Thr Gln Thr
            100                 105                 110
Gly Thr Ala Glu Gly Phe Ala Lys Ala Leu Phe Glu Glu Ala Lys Ala
        115                 120                 125
Arg Tyr Glu Lys Ala Ala Phe Lys Val Ile Asp Leu Asp Asp Tyr Ala
    130                 135                 140
Ala Asp Leu Asp Glu Tyr Ala Glu Lys Leu Lys Lys Glu Thr Tyr Ala
145                 150                 155                 160
Phe Phe Phe Leu Ala Thr Tyr Gly Asp Gly Glu Pro Thr Asp Asn Ala
                165                 170                 175
Ala Lys Phe Tyr Lys Trp Phe Thr Glu Gly Asp Glu Lys Gly Val Trp
            180                 185                 190
Leu Gln Lys Leu Gln Tyr Gly Val Phe Gly Leu Gly Asn Arg Gln Tyr
        195                 200                 205
Glu His Phe Asn Lys Ile Gly Ile Val Val Asp Asp Gly Leu Thr Glu
    210                 215                 220
Gln Gly Ala Lys Arg Ile Val Pro Val Gly Leu Gly Asp Asp Asp Gln
225                 230                 235                 240
Ser Ile Glu Asp Asp Phe Ser Ala Trp Lys Glu Leu Val Trp Pro Glu
                245                 250                 255
Leu Asp Leu Leu Leu Arg Asp Glu Asp Asp Lys Ala Ala Ala Thr Pro
            260                 265                 270
Tyr Thr Ala Ala Ile Pro Glu Tyr Arg Val Val Phe His Asp Lys Pro
        275                 280                 285
Asp Ala Phe Ser Asp Asp His Thr Gln Thr Asn Gly His Ala Val His
    290                 295                 300
Asp Ala Gln His Pro Cys Arg Ser Asn Val Ala Val Lys Lys Glu Leu
305                 310                 315                 320
His Thr Pro Glu Ser Asp Arg Ser Cys Thr His Leu Glu Phe Asp Ile
                325                 330                 335
Ser His Thr Gly Leu Ser Tyr Glu Thr Gly Asp His Val Gly Val Tyr
            340                 345                 350
Cys Glu Asn Leu Ile Glu Val Val Glu Glu Ala Gly Lys Leu Leu Gly
        355                 360                 365
Leu Ser Thr Asp Thr Tyr Phe Ser Leu His Ile Asp Asn Glu Asp Gly
    370                 375                 380
Ser Pro Leu Gly Gly Pro Ser Leu Gln Pro Pro Phe Pro Pro Cys Thr
385                 390                 395                 400
Leu Arg Lys Ala Leu Thr Asn Tyr Ala Asp Leu Leu Ser Ser Pro Lys
                405                 410                 415
Lys Ser Thr Leu Leu Ala Leu Ala Ala His Ala Ser Asp Pro Thr Glu
            420                 425                 430
Ala Asp Arg Leu Arg Phe Leu Ala Ser Arg Glu Gly Lys Asp Glu Tyr
        435                 440                 445
Ala Glu Trp Val Val Ala Asn Gln Arg Ser Leu Leu Glu Val Met Glu
    450                 455                 460
Ala Phe Pro Ser Ala Arg Pro Pro Leu Gly Val Phe Phe Ala Ala Val
465                 470                 475                 480
Ala Pro Arg Leu Gln Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Pro Lys
                485                 490                 495
Met Glu Pro Asn Arg Ile His Val Thr Cys Ala Leu Val Tyr Glu Lys
            500                 505                 510
Thr Pro Ala Gly Arg Ile His Lys Gly Ile Cys Ser Thr Trp Met Lys
        515                 520                 525
Asn Ala Val Pro Leu Thr Glu Ser Gln Asp Cys Ser Trp Ala Pro Ile
    530                 535                 540
Phe Val Arg Thr Ser Asn Phe Arg Leu Pro Ile Asp Pro Lys Val Pro
545                 550                 555                 560
Val Ile Met Ile Gly Pro Gly Thr Gly Leu Ala Pro Phe Arg Gly Phe
                565                 570                 575
Leu Gln Glu Arg Leu Ala Leu Lys Glu Ser Gly Thr Glu Leu Gly Ser
            580                 585                 590
```

Ser Ile Leu Phe Phe Gly Cys Arg Asn Arg Lys Val Asp Tyr Ile Tyr
595 600 605
Glu Asn Glu Leu Asn Asn Phe Val Glu Asn Gly Ala Leu Ser Glu Leu
610 615 620
Asp Val Ala Phe Ser Arg Asp Gly Pro Thr Lys Glu Tyr Val Gln His
625 630 635 640
Lys Met Thr Gln Lys Ala Ser Glu Ile Trp Asn Met Leu Ser Glu Gly
645 650 655
Ala Tyr Leu Tyr Val Cys Gly Asp Ala Lys Gly Met Ala Lys Asp Val
660 665 670
His Arg Thr Leu His Thr Ile Val Gln Glu Gln Gly Ser Leu Asp Ser
675 680 685
Ser Lys Ala Glu Leu Tyr Val Lys Asn Leu Gln Met Ser Gly Arg Tyr
690 695 700
Leu Arg Asp Val Trp
705

<210> 170
<211> 660
<212> PRT
<213> Stevia rebaudiana

<400> 170
Met Gln Ser Asn Ser Val Lys Ile Ser Pro Leu Asp Leu Val Thr Ala
1 5 10 15
Leu Phe Ser Gly Lys Val Leu Asp Thr Ser Asn Ala Ser Glu Ser Gly
20 25 30
Glu Ser Ala Met Leu Pro Thr Ile Ala Met Ile Met Glu Asn Arg Glu
35 40 45
Leu Leu Met Ile Leu Thr Thr Ser Val Ala Val Leu Ile Gly Cys Val
50 55 60
Val Val Leu Val Trp Arg Arg Ser Ser Thr Lys Lys Ser Ala Leu Glu
65 70 75 80
Pro Pro Val Ile Val Val Pro Lys Arg Val Gln Glu Glu Glu Val Asp
85 90 95
Asp Gly Lys Lys Lys Val Thr Val Phe Phe Gly Thr Gln Thr Gly Thr
100 105 110
Ala Glu Gly Phe Ala Lys Ala Leu Val Glu Glu Ala Lys Ala Arg Tyr
115 120 125
Glu Lys Ala Val Phe Lys Val Ile Asp Leu Asp Asp Tyr Ala Ala Asp
130 135 140
Asp Asp Glu Tyr Glu Glu Lys Leu Lys Lys Glu Ser Leu Ala Phe Phe
145 150 155 160
Phe Leu Ala Thr Tyr Gly Asp Gly Glu Pro Thr Asp Asn Ala Ala Arg
165 170 175
Phe Tyr Lys Trp Phe Thr Glu Gly Asp Ala Lys Gly Glu Trp Leu Asn
180 185 190
Lys Leu Gln Tyr Gly Val Phe Gly Leu Gly Asn Arg Gln Tyr Glu His
195 200 205
Phe Asn Lys Ile Ala Lys Val Val Asp Asp Gly Leu Val Glu Gln Gly
210 215 220
Ala Lys Arg Leu Val Pro Val Gly Leu Gly Asp Asp Asp Gln Cys Ile
225 230 235 240
Glu Asp Asp Phe Thr Ala Trp Lys Glu Leu Val Trp Pro Glu Leu Asp
245 250 255
Gln Leu Leu Arg Asp Glu Asp Asp Thr Thr Val Ala Thr Pro Tyr Thr
260 265 270
Ala Ala Val Ala Glu Tyr Arg Val Val Phe His Glu Lys Pro Asp Ala
275 280 285
Leu Ser Glu Asp Tyr Ser Tyr Thr Asn Gly His Ala Val His Asp Ala
290 295 300
Gln His Pro Cys Arg Ser Asn Val Ala Val Lys Lys Glu Leu His Ser

305 310 315 320
Pro Glu Ser Asp Arg Ser Cys Thr His Leu Glu Phe Asp Ile Ser Asn
325 330 335
Thr Gly Leu Ser Tyr Glu Thr Gly Asp His Val Gly Val Tyr Cys Glu
340 345 350
Asn Leu Ser Glu Val Val Asn Asp Ala Glu Arg Leu Val Gly Leu Pro
355 360 365
Pro Asp Thr Tyr Ser Ser Ile His Thr Asp Ser Glu Asp Gly Ser Pro
370 375 380
Leu Gly Gly Ala Ser Leu Pro Pro Pro Phe Pro Pro Cys Thr Leu Arg
385 390 395 400
Lys Ala Leu Thr Cys Tyr Ala Asp Val Leu Ser Ser Pro Lys Lys Ser
405 410 415
Ala Leu Leu Ala Leu Ala Ala His Ala Thr Asp Pro Ser Glu Ala Asp
420 425 430
Arg Leu Lys Phe Leu Ala Ser Pro Ala Gly Lys Asp Glu Tyr Ser Gln
435 440 445
Trp Ile Val Ala Ser Gln Arg Ser Leu Leu Glu Val Met Glu Ala Phe
450 455 460
Pro Ser Ala Lys Pro Ser Leu Gly Val Phe Phe Ala Ser Val Ala Pro
465 470 475 480
Arg Leu Gln Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Pro Lys Met Ala
485 490 495
Pro Asp Arg Ile His Val Thr Cys Ala Leu Val Tyr Glu Lys Thr Pro
500 505 510
Ala Gly Arg Ile His Lys Gly Val Cys Ser Thr Trp Met Lys Asn Ala
515 520 525
Val Pro Met Thr Glu Ser Gln Asp Cys Ser Trp Ala Pro Ile Tyr Val
530 535 540
Arg Thr Ser Asn Phe Arg Leu Pro Ser Asp Pro Lys Val Pro Val Ile
545 550 555 560
Met Ile Gly Pro Gly Thr Gly Leu Ala Pro Phe Arg Gly Phe Leu Gln
565 570 575
Glu Arg Leu Ala Leu Lys Glu Ala Gly Thr Asp Leu Gly Leu Ser Ile
580 585 590
Leu Phe Phe Gly Cys Arg Asn Arg Lys Val Asp Phe Ile Tyr Glu Asn
595 600 605
Glu Leu Asn Asn Phe Val Glu Thr Gly Ala Leu Ser Glu Leu Ile Val
610 615 620
Ala Phe Ser Arg Glu Gly Pro Thr Lys Glu Tyr Val Gln His Lys Met
625 630 635 640
Ser Glu Lys Ala Ser Asp Ile Trp Asn Leu Leu Ser Glu Gly Ala Tyr
645 650 655
Leu Tyr Val Cys
660

<210> 171
<211> 2079
<212> DNA
<213> Artificial Sequence

<220>
<223> Arabidopsis thaliana sequence that has been codon optimized for yeast

<400> 171
atgacttctg cactttatgc ctccgatctt ttcaaacaat tgaaaagtat catgggaacg 60
gattctttgt ccgatgatgt tgtattagtt attgctacaa cttctctggc actggttgct 120
ggtttcgttg tcttattgtg gaaaaagacc acggcagatc gttccggcga gctaaagcca 180
ctaatgatcc ctaagtctct gatggcgaaa gatgaggatg atgacttaga tctaggttct 240
ggaaaaacga gagtctctat cttcttcggc acacaaaccg gaacagccga aggattcgct 300
aaagcacttt cagaagagat caaagcaaga tacgaaaagg cggctgtaaa agtaatcgat 360

ttggatgatt acgctgccga tgatgaccaa tatgaggaaa agttgaaaaa ggaaacattg 420
gctttctttt gtgtagccac gtatggtgat ggtgaaccaa ccgataacgc cgcaagattc 480
tacaagtggt ttactgaaga gaacgaaaga gatatcaagt tgcagcaact tgcttacggc 540
gtttttgcct taggtaacag acaatacgag cactttaaca agataggtat tgtcttagat 600
gaagagttat gcaaaaaggg tgcgaagaga ttgattgaag tcggtttagg agatgatgat 660
caatctatcg aggatgactt taatgcatgg aaggaatctt tgtggtctga attagataag 720
ttacttaagg acgaagatga taaatccgtt gccactccat acacagccgt cattccagaa 780
tatagagtag ttactcatga tccaagattc acaacacaga aatcaatgga aagtaatgtg 840
gctaatggta atactaccat cgatattcat catccatgta gagtagacgt tgcagttcaa 900
aaggaattgc acactcatga atcagacaga tcttgcatac atcttgaatt tgatatatca 960
cgtactggta tcacttacga aacaggtgat cacgtgggtg tctacgctga aaaccatgtt 1020
gaaattgtag aggaagctgg aaagttgttg ggccatagtt tagatcttgt tttctcaatt 1080
catgccgata aagaggatgg ctcaccacta gaaagtgcag tgcctccacc atttccagga 1140
ccatgcaccc taggtaccgg tttagctcgt tacgcggatc tgttaaatcc tccacgtaaa 1200
tcagctctag tggccttggc tgcgtacgcc acagaacctt ctgaggcaga aaaactgaaa 1260
catctaactt caccagatgg taaggatgaa tactcacaat ggatagtagc tagtcaacgt 1320
tctttactag aagttatggc tgctttccca tccgctaaac ctcctttggg tgttttcttc 1380
gccgcaatag cgcctagact gcaaccaaga tactattcaa tttcatcctc acctagactg 1440
gcaccatcaa gagttcatgt cacatccgct ttagtgtacg gtccaactcc tactggtaga 1500
atccataagg gcgtttgttc aacatggatg aaaaacgcgg ttccagcaga gaagtctcac 1560
gaatgttctg gtgctccaat ctttatcaga gcctccaact tcaaactgcc ttccaatcct 1620
tctactccta ttgtcatggt cggtcctggt acaggtcttg ctccattcag aggtttctta 1680
caagagagaa tggccttaaa ggaggatggt gaagagttgg gatcttcttt gttgtttttc 1740
ggctgtagaa acagacaaat ggatttcatc tacgaagatg aactgaataa ctttgtagat 1800
caaggagtta tttcagagtt gataatggct ttttctagag aaggtgctca gaaggagtac 1860
gtccaacaca aaatgatgga aaaggccgca caagtttggg acttaatcaa agaggaaggc 1920
tatctatatg tctgtggtga tgcaaagggt atggcaagag atgttcacag aacacttcat 1980
actatagtcc aggaacagga aggcgttagt tcttctgaag cggaagcaat tgtgaaaaag 2040
ttacaaacag agggaagata cttgagagat gtgtggtaa 2079

<210> 172
<211> 2139
<212> DNA
<213> Artificial Sequence

<220>
<223> Arabidopsis thaliana sequence that has been codon
optimized for yeast

<400> 172
atgtcttcct cttcctcttc cagtacctct atgattgatt tgatggctgc tattattaaa 60
ggtgaaccag ttatcgtctc cgacccagca aatgcctctg cttatgaatc agttgctgca 120
gaattgtctt caatgttgat cgaaaacaga caattcgcca tgatcgtaac tacatcaatc 180
gctgttttga tcggttgtat tgtcatgttg gtatggagaa gatccggtag tggtaattct 240
aaaagagtcg aacctttgaa accattagta attaagccaa gagaagaaga aatagatgac 300
ggtagaaaga aagttacaat atttttcggt acccaaactg gtacagctga aggttttgca 360
aaagccttag gtgaagaagc taaggcaaga tacgaaaaga ctagattcaa gatagtcgat 420
ttggatgact atgccgctga tgacgatgaa tacgaagaaa agttgaagaa agaagatgtt 480
gcatttttct ttttggcaac ctatggtgac ggtgaaccaa ctgacaatgc agccagattc 540
tacaaatggt ttacagaggg taatgatcgt ggtgaatggt tgaaaaactt aaagtacggt 600
gttttcggtt tgggtaacag acaatacgaa catttcaaca aagttgcaaa ggttgtcgac 660
gatattttgg tcgaacaagg tgctcaaaga ttagtccaag taggtttggg tgacgatgac 720
caatgtatag aagatgactt tactgcctgg agagaagctt tgtggcctga attagacaca 780
atcttgagag aagaaggtga caccgccgtt gctaccccat atactgctgc agtattagaa 840
tacagagttt ccatccatga tagtgaagac gcaaagttta atgatatcac tttggccaat 900
ggtaacggtt atacagtttt cgatgcacaa cacccttaca aagctaacgt tgcagtcaag 960
agagaattac atacaccaga atccgacaga agttgtatac acttggaatt tgatatcgct 1020
ggttccggtt taaccatgaa gttgggtgac catgtaggtg ttttatgcga caatttgtct 1080
gaaactgttg atgaagcatt gagattgttg gatatgtccc ctgacactta ttttagtttg 1140
cacgctgaaa aagaagatgg tacaccaatt tccagttctt taccacctcc attccctcca 1200
tgtaacttaa gaacagcctt gaccagatac gcttgcttgt tatcatcccc taaaaagtcc 1260
gccttggttg ctttagccgc tcatgctagt gatcctactg aagcagaaag attgaaacac 1320
ttagcatctc cagccggtaa agatgaatat tcaaagtggg tagttgaatc tcaaagatca 1380

ttgttagaag ttatggcaga atttccatct gccaagcctc cattaggtgt cttctttgct 1440
ggtgtagcac ctagattgca accaagattc tactcaatca gttcttcacc taagatcgct 1500
gaaactagaa ttcatgttac atgtgcatta gtctacgaaa agatgccaac cggtagaatt 1560
cacaagggtg tatgctctac ttggatgaaa aatgctgttc cttacgaaaa atcagaaaag 1620
ttgttcttag gtagaccaat cttcgtaaga caatcaaact tcaagttgcc ttctgattca 1680
aaggttccaa taatcatgat aggtcctggt acaggtttag ccccattcag aggtttcttg 1740
caagaaagat tggctttagt tgaatctggt gtcgaattag gtccttcagt tttgttcttt 1800
ggttgtagaa acagaagaat ggatttcatc tatgaagaag aattgcaaag attcgtcgaa 1860
tctggtgcat tggccgaatt atctgtagct ttttcaagag aaggtccaac taaggaatac 1920
gttcaacata agatgatgga taaggcatcc gacatatgga acatgatcag tcaaggtgct 1980
tatttgtacg tttgcggtga cgcaaagggt atggccagag atgtccatag atctttgcac 2040
acaattgctc aagaacaagg ttccatggat agtaccaaag ctgaaggttt cgtaaagaac 2100
ttacaaactt ccggtagata cttgagagat gtctggtga 2139

<210> 173
<211> 2130
<212> DNA
<213> Stevia rebaudiana

<400> 173
atgcaatcgg aatccgttga agcatcgacg attgatttga tgactgctgt tttgaaggac 60
acagtgatcg atacagcgaa cgcatctgat aacggagact caaagatgcc gccggcgttg 120
gcgatgatgt tcgaaattcg tgatctgttg ctgattttga ctacgtcagt tgctgttttg 180
gtcggatgtt tcgttgtttt ggtgtggaag agatcgtccg ggaagaagtc cggcaaggaa 240
ttggagccgc cgaagatcgt tgtgccgaag aggcggctgg agcaggaggt tgatgatggt 300
aagaagaagg ttacgatttt cttcggaaca caaactggaa cggctgaagg tttcgctaag 360
gcacttttcg aagaagcgaa agcgcgatat gaaaaggcag cgtttaaagt gattgatttg 420
gatgattatg ctgctgattt ggatgagtat gcagagaagc tgaagaagga aacatatgct 480
ttcttcttct tggctacata tggagatggt gagccaactg ataatgctgc caaattttat 540
aaatggttta ctgagggaga cgagaaaggc gtttggcttc aaaaacttca atatggagta 600
tttggtcttg gcaacagaca atatgaacat ttcaacaaga ttggaatagt ggttgatgat 660
ggtctcaccg agcagggtgc aaaacgcatt gttcccgttg gtcttggaga cgacgatcaa 720
tcaattgaag acgatttttc ggcatggaaa gagttagtgt ggcccgaatt ggatctattg 780
cttcgcgatg aagatgacaa agctgctgca actccttaca cagctgcaat ccctgaatac 840
cgcgtcgtat ttcatgacaa acccgatgcg ttttctgatg atcatactca aaccaatggt 900
catgctgttc atgatgctca acatccatgc agatccaatg tggctgttaa aaaagagctt 960
catactcctg aatccgatcg ttcatgcaca catcttgaat ttgacatttc tcacactgga 1020
ttatcttatg aaactgggga tcatgttggt gtatactgtg aaaacctaat tgaagtagtg 1080
gaagaagctg ggaaattgtt aggattatca acagatactt atttctcgtt acatattgat 1140
aacgaagatg gttcaccact tggtggacct tcattacaac ctccttttcc tccttgtact 1200
ttaagaaaag cattgactaa ttatgcagat ctgttaagct ctcccaaaaa gtcaactttg 1260
cttgctctag ctgctcatgc ttccgatccc actgaagctg atcgtttaag atttcttgca 1320
tctcgcgagg gcaaggatga atatgctgaa tgggttgttg caaaccaaag aagtcttctt 1380
gaagtcatgg aagctttccc gtcagctaga ccgccacttg gtgttttctt tgcagcggtt 1440
gcaccgcgtt tacagcctcg ttactactct atttcttcct ccccaaagat ggaaccaaac 1500
aggattcatg ttacttgcgc gttggtttat gaaaaaactc ccgcaggtcg tatccacaaa 1560
ggaatctgct caacctggat gaagaacgct gtacctttga ccgaaagtca agattgcagt 1620
tgggcaccga tttttgttag aacatcaaac ttcagacttc caattgaccc gaaagtcccg 1680
gttatcatga ttggtcctgg aaccgggttg gctccattta ggggttttct tcaagaaaga 1740
ttggctctta aagaatccgg aaccgaactc gggtcatcta ttttattctt cggttgtaga 1800
aaccgcaaag tggattacat atatgagaat gaactcaaca actttgttga aaatggtgcg 1860
ctttctgagc ttgatgttgc tttctcccgc gatggcccga cgaaagaata cgtgcaacat 1920
aaaatgaccc aaaaggcttc tgaaatatgg aatatgcttt ctgagggagc atatttatat 1980
gtatgtggtg atgctaaagg catggctaaa gatgtacacc gtacacttca caccattgtg 2040
caagaacagg gaagtttgga ctcgtctaaa gcggagttgt atgtgaagaa tctacaaatg 2100
tcaggaagat acctccgtga tgtttggtaa 2130

<210> 174
<211> 2124
<212> DNA
<213> Stevia rebaudiana

<400> 174

atgcaatcta actccgtgaa gatttcgccg cttgatctgg taactgcgct gtttagcggc 60
aaggttttgg acacatcgaa cgcatcggaa tcgggagaat ctgctatgct gccgactata 120
gcgatgatta tggagaatcg tgagctgttg atgatactca caacgtcggt tgctgtattg 180
atcggatgcg ttgtcgtttt ggtgtggcgg agatcgtcta cgaagaagtc ggcgttggag 240
ccaccggtga ttgtggttcc gaagagagtg caagaggagg aagttgatga tggtaagaag 300
aaagttacgg ttttcttcgg cacccaaact ggaacagctg aaggcttcgc taaggcactt 360
gttgaggaag ctaaagctcg atatgaaaag gctgtcttta aagtaattga tttggatgat 420
tatgctgctg atgacgatga gtatgaggag aaactaaaga aagaatcttt ggccttttc 480
tttttggcta cgtatggaga tggtgagcca acagataatg ctgccagatt ttataaatgg 540
tttactgagg gagatgcgaa aggagaatgg cttaataagc ttcaatatgg agtatttggt 600
ttgggtaaca gacaatatga acattttaac aagatcgcaa aagtggttga tgatggtctt 660
gtagaacagg gtgcaaagcg tcttgttcct gttggacttg gagatgatga tcaatgtatt 720
gaagatgact tcaccgcatg gaaagagtta gtatggccgg agttggatca attacttcgt 780
gatgaggatg acacaactgt tgctactcca tacacagctg ctgttgcaga atatcgcgtt 840
gtttttcatg aaaaaccaga cgcgctttct gaagattata gttatacaaa tggccatgct 900
gttcatgatg ctcaacatcc atgcagatcc aacgtggctg tcaaaaagga acttcatagt 960
cctgaatctg accggtcttg cactcatctt gaatttgaca tctcgaacac cggactatca 1020
tatgaaactg gggaccatgt tggagtttac tgtgaaaact tgagtgaagt tgtgaatgat 1080
gctgaaagat tagtaggatt accaccagac acttactcct ccatccacac tgatagtgaa 1140
gacgggtcgc cacttggcgg agcctcattg ccgcctcctt tcccgccatg cactttaagg 1200
aaagcattga cgtgttatgc tgatgttttg agttctccca agaagtcggc tttgcttgca 1260
ctagctgctc atgccaccga tcccagtgaa gctgatagat tgaaatttct tgcatccccc 1320
gccggaaagg atgaatattc tcaatggata gttgcaagcc aaagaagtct ccttgaagtc 1380
atggaagcat tcccgtcagc taagccttca cttggtgttt tctttgcatc tgttgccccg 1440
cgcttacaac caagatacta ctctatttct tcctcaccca agatggcacc ggataggatt 1500
catgttacat gtgcattagt ctatgagaaa acacctgcag gccgcatcca caaaggagtt 1560
tgttcaactt ggatgaagaa cgcagtgcct atgaccgaga gtcaagattg cagttgggcc 1620
ccaatatacg tccgaacatc caatttcaga ctaccatctg accctaaggt cccggttatc 1680
atgattggac ctggcactgg tttggctcct tttagaggtt tccttcaaga gcggttagct 1740
ttaaaggaag ccggaactga cctcggttta tccattttat tcttcggatg taggaatcgc 1800
aaagtggatt tcatatatga aaacgagctt aacaactttg tggagactgg tgctctttct 1860
gagcttattg ttgctttctc ccgtgaaggc ccgactaagg aatatgtgca acacaagatg 1920
agtgagaagg cttcggatat ctggaacttg ctttctgaag gagcatattt atacgtatgt 1980
ggtgatgcca aaggcatggc caaagatgta catcgaaccc tccacacaat tgtgcaagaa 2040
cagggatctc ttgactcgtc aaaggcagaa ctctacgtga agaatctaca aatgtcagga 2100
agatacctcc gtgacgtttg gtaa 2124

<210> 175
<211> 2427
<212> DNA
<213> Arabidopsis thaliana

<400> 175
atggcaaacg ctgaacgtat gataacgcgc gtccacagcc aacgtgagcg tttgaacgaa 60
acgcttgttt ctgagagaaa cgaagtcctt gccttgcttt ccagggttga agccaaaggt 120
aaaggtattt tacaacaaaa ccagatcatt gctgaattcg aagctttgcc tgaacaaacc 180
cggaagaaac ttgaaggtgg tcctttcttt gaccttctca aatccactca ggaagcaatt 240
gtgttgccac catgggttgc tctagctgtg aggccaaggc ctggtgtttg ggaatactta 300
cgagtcaatc tccatgctct tgtcgttgaa gaactccaac ctgctgagtt tcttcatttc 360
aaggaagaac tcgttgatgg agttaagaat ggtaatttca ctcttgagct tgatttcgag 420
ccattcaatg cgtctatccc tcgtccaaca ctccacaaat acattggaaa tggtgttgac 480
ttccttaacc gtcatttatc ggctaagctc ttccatgaca aggagagttt gcttccattg 540
cttaagttcc ttcgtcttca cagccaccag ggcaagaacc tgatgttgag cgagaagatt 600
cagaacctca acactctgca acacaccttg aggaaagcag aagagtatct agcagagctt 660
aagtccgaaa cactgtatga agagtttgag gccaagtttg aggagattgg tcttgagagg 720
ggatggggag acaatgcaga gcgtgtcctt gacatgatac gtcttctttt ggaccttctt 780
gaggcgcctg atccttgcac tcttgagact tttcttggaa gagtaccaat ggtgttcaac 840
gttgtgatcc tctctccaca tggttacttt gctcaggaca atgttcttgg ttaccctgac 900
actggtggac aggttgttta cattcttgat caagttcgtg ctctggagat agagatgctt 960
caacgtatta agcaacaagg actcaacatt aaaccaagga ttctcattct aactcgactt 1020
ctacctgatg cggtaggaac tacatgcggt gaacgtctcg agagagttta tgattctgag 1080
tactgtgata ttcttcgtgt gcccttcaga acagagaagg gtattgttcg caaatggatc 1140
tcaaggttcg aagtctggcc atatctagag acttacaccg aggatgctgc ggttgagcta 1200

tcgaaagaat tgaatggcaa gcctgacctt atcattggta actacagtga tggaaatctt 1260
gttgcttctt tattggctca caaacttggt gtcactcagt gtaccattgc tcatgctctt 1320
gagaaaacaa agtacccgga ttctgatatc tactggaaga agcttgacga caagtaccat 1380
ttctcatgcc agttcactgc ggatattttc gcaatgaacc acactgattt catcatcact 1440
agtactttcc aagaaattgc tggaagcaaa gaaactgttg ggcagtatga aagccacaca 1500
gcctttactc ttcccggatt gtatcgagtt gttcacggga ttgatgtgtt tgatcccaag 1560
ttcaacattg tctctcctgg tgctgatatg agcatctact tcccttacac agaggagaag 1620
cgtagattga ctaagttcca ctctgagatc gaggagctcc tctacagcga tgttgagaac 1680
aaagagcact tatgtgtgct caaggacaag aagaagccga ttctcttcac aatggctagg 1740
cttgatcgtg tcaagaactt gtcaggtctt gttgagtggt acgggaagaa cacccgcttg 1800
cgtgagctag ctaacttggt tgttgttgga ggagacagga ggaaagagtc aaaggacaat 1860
gaagagaaag cagagatgaa gaaaatgtat gatctcattg aggaatacaa gctaaacggt 1920
cagttcaggt ggatctcctc tcagatggac cgggtaagga acggtgagct gtaccggtac 1980
atctgtgaca ccaagggtgc ttttgtccaa cctgcattat atgaagcctt tgggttaact 2040
gttgtggagg ctatgacttg tggtttaccg actttcgcca cttgcaaagg tggtccagct 2100
gagatcattg tgcacggtaa atcgggtttc cacattgacc cttaccatgg tgatcaggct 2160
gctgatactc ttgctgattt cttcaccaag tgtaaggagg atccatctca ctgggatgag 2220
atctcaaaag gagggcttca gaggattgag gagaaataca cttggcaaat ctattcacag 2280
aggctcttga cattgactgg tgtgtatgga ttctggaagc atgtctcgaa ccttgaccgt 2340
cttgaggctc gccgttacct tgaaatgttc tatgcattga agtatcgccc attggctcag 2400
gctgttcctc ttgcacaaga tgattga 2427

<210> 176
<211> 2421
<212> DNA
<213> Stevia rebaudiana

<400> 176
atggcggaac gtgtactcac tcgtgttcac gaacttcgtg agcgtctcga ttcaactctc 60
gcaactcatc gtaatgaaat cctcttgttt ctttcaagga ttgaaagcca tggaaaagga 120
atattgaagc ctcatcaagt tatgactgaa tttgaagcta tctgcaaaga agatcagagc 180
aaactctctg atggtgcttt ttatgaagtt cttaaatgca cacaggaagc aatagtgcaa 240
cctccatggg ttgcactcgc gatccgtctt cgacccggtg tttgggaata tgttagagtc 300
aatgttaatg ttttggtggt tgaagaatta agtgttcctg aatatcttca cttcaaagaa 360
gaattggtta atggaacatc gaatggcaac ttcgtgttgg aactggattt tgaacctttt 420
accgcatcgt ttcctcgacc aactttaacc aagtctattg gtaatggtgt tgagtttcta 480
aacagacatt tatctgctaa aatgtttcat gataaggata gcatgcaccc tcttcttgat 540
ttcctacgga ctcaccacta taagggaaag acaatgatgt tgaatgatag aatccaaaac 600
ctcaatgctc tacaatcggt gttgcgaaag gcgtcagagt acttatcaac actcgacgcg 660
gcaacaccgt actctcagtt tgaacataag tttcaagaaa tcgggttgga gagaggttgg 720
ggtgataaag cggaggtcgt aatggagatg atccacatgc ttctagacct tctagaagca 780
cccgacgcat gcacactcga gaagtttctc ggaagaatcc caatggtttt caatgttgtc 840
attctttcgc ctcatggcta cttcgcgcaa gaaaatgtgt tgggatatcc cgacactggc 900
ggtcaggttg tttacatctt ggatcaagtt cccgctctgg aacgcgagat gctcaaaagg 960
attaaggagc aaggactcga tatcattcct cgtatattga ttgttacgag gcttcttcct 1020
gacgcggttg ggaccacatg cgggcaacgt ttagagaaag tgtttggagc cgaacactcg 1080
catattcttc gggtcccgtt tagaaccgaa aagggtattc ttcgtaaatg gatctctcgt 1140
tttgaggtgt ggccttacat cgagactttc accgaggatg ttgctaaaga agttacagca 1200
gagttgcaag caaaaccaga tttgatcatt ggaaactata gtgaaggaaa tttggttgca 1260
tctttgctag ctcacaagtt gggtgtcact cagtgtacca ttgctcatgc tttggagaaa 1320
actaaatacc cggattctga tatctactgg aagaactttg aggagaaata tcatttctct 1380
tcgcagttta ccgctgatct tatcgctatg aaccataccg acttcatcat caccagtact 1440
ttccaagaaa ttgctggaag taaggacacg gttggacagt acgagagtca taccgcgttc 1500
acaatgccgg gattgtatcg ggtggttcac gggatcgatg tttttgaccc caaattcaat 1560
attgtttcac ccggggccga tatgggaatt tactactcgt ataccgagaa agaaaagagg 1620
ctcactgcgc ttcaccctga aatcgatgaa cttctcttta gttccgtcga aaacgaagaa 1680
cacttatgtg tgttgaagga taagagtaaa ccaatcttgt tcacaatggc gcgattggat 1740
aatgtgaaga atttaaccgg actggttgaa tggtacgcta aaaacgaccg ccttcgtgag 1800
ctcgtgaacc tcgtggtcgt cggtggtgac cgaaggaaag agtcgaaaga tcttgaagag 1860
caagctcaga tgaagaagat gcatgaactt atcgaaacct acaaactcaa cggtcagttc 1920
aggtggatat cctcacaaat gaaccgcgtg aggaacggtg agttgtatcg cgttattgct 1980
gacacgcgag gtgcgtttat ccagcctgcg ttttacgagg cgtttgggtt gacggttgtg 2040
gaggccatga cttgtggcct gccgacattc gcgacacttc atggtgggcc cgctgagatt 2100

attgttcacg ggaaatccgg gttccatatt gacccgtatc acggtgacca ggtcaccgag 2160
ttgctggtca atttctttga gaaaactaaa caagacccgg gtcattggga ggccatttcc 2220
aagggtggtc tgcaacgtat tcaggagaaa tacacgtggc agatttattc agataggttg 2280
ttgacgcttg ccggagttta tggattctgg aagcatgtgt cgaagcttga caggctcgag 2340
atccgtcgtt atcttgaaat gttttacgcg ctcaagtatc gcaaactggc tgaatctgtt 2400
ccattggctg ttgatgagtg a 2421

<210> 177
<211> 2421
<212> DNA
<213> Coffea arabica

<400> 177
atggccgaac gtgttctgac ccgtgttcac agcctccgcg aacgccttga tgctactttg 60
gctgcccacc gcaacgatgt tttgctgttt atgtcgaggc ttgaaaccca tgggaaaggg 120
atcctgaaac cccaccaact tttggctgag tttgaagaaa ttaacaagga tggtaaacaa 180
aaaattcatg atcatgcctt tgaagaagtc ctgaagtcca cacaggaagc aattgtgttg 240
ccccccctggg ttgcacttgc tattcgtctc agacctggtg tctgggagta tgttcgagtc 300
aatgtccatg cactcgttgt tgaggagtta accgtgccag agtacctgca tttcaaggaa 360
gaactcgttg atggaagcaa aaatgggaat tttgttttgg aactggactt cgaaccattt 420
acagcctctt ttcccaagcc aactctaact aagtacatag gcgacggagt tgagttcctc 480
aacaggcacc tctctgccaa aatgttccat gacaaggaga gcatggcccc tctccttgat 540
tttctccgtg ttcaccaata caagggcaag acaatgatgc ttaacgacag gatcaaggac 600
cttaacactc tccaagcagt tctgaggaag gcagaggagt acctaacaac actctctgca 660
gatacaccat actctgaatt cgagcacaaa ttccaagaaa ttggactgga gagaggttgg 720
ggtgatactg ctgagcgtgt cttggaaatg atctgcatgc ttctggatct tctgggggct 780
cctgactcgt gcacactaga gaaattccta gggagaatcc ctatggtatt caatgttgtt 840
attctttccc cccatggata ctttgcccag gaaaacgtat tgggttatcc tgataccggt 900
ggccaggttg tttacatatt ggatcaagtt cctgccttgg agcgtgagat gctgaagagg 960
ataaaggaac aaggacttga tgtcaagcca cgcattctaa ttataactag gctgcttcct 1020
gatgcccctg gaaccacttg tggtcaacgg cttgagaaag tatatggatc agagtactcc 1080
catatactca gagtcccctt cagaactgag aagggagttg ttcgcaaatg gatctctcgc 1140
tttgaagttt ggccctacat ggaaacattt actgaggatg ttgcaaaaga agtcactgca 1200
gaattacagg caaagccaga tttggttatt ggtaactaca gtgagggtaa ccttgttgcc 1260
tccttgcttg ctcacaagtt aggtgtaaca cagtgtacca ttgctcatgc tttggaaaaa 1320
accaagtatc ctgattctga tatttatttg agcaaatttg atgaaaagta ccacttctca 1380
tgtcagttca ctgcggatct tatcgcaatg aaccatacag atttcattat cactagcact 1440
ttccaagaaa tagctggaag caaggacact gttgggcaat atgaaagcca tatggccttc 1500
acaatgccag gattatacag agttgtacat ggcattgatg tttttgatcc aaaattcaac 1560
attgtctcac ctggagctga tacaaacctc tacttcccac acacagagaa ggaaaagaga 1620
ttgacatctt tccatcctga aattgaggag ttgcttttca gcgatgtgga gaatgaggaa 1680
cacctatgtg tgctaaaaga caaaaagaag cctatcttat tcaccatggc aagactggat 1740
cgcgtaaaga atttgacagg gcttgttgaa ttgtatgcta agaacccaaa actaagggaa 1800
ttggtcaatc ttgtcgtggt tggtggagac cgaaggaagg aatccaaaga tttggaagaa 1860
caagctgaga tgaagaaaat gtattcattg atagaaactt acaacttgaa cggccaattc 1920
agatggattt cttctcagat gaacagggtt agaaatggtg aactctatcg gtacattgct 1980
gacaccaggg gagcattcgt gcaacctgca ttttatgagg catttgggtt gactgtggtc 2040
gaggccatga catgtggttt gccaacattt gcaaccaacc atggtggtcc tgctgagatc 2100
attattcatg ggaaatctgg tttccacatt gatccatacc atggtgagca ggtcagcgag 2160
ctccttgcca atttctttga aaggtgcaag aaagagcctt cttactggga caccatttca 2220
gccggtggct tgaagcgtat ccaggaaaag tacacctggc aaatttactc agatcggttg 2280
ctgacgctgg ctggagttta tggattctgg aaatgtgttt ccaagcttga tcgccaggag 2340
atccgccgtt atctggaaat gttttatgct ctcaagtatc gcaagttggc tgaagctgtt 2400
ccattggctg ttgatcagta a 2421

<210> 178
<211> 481
<212> PRT
<213> Arabidopsis thaliana

<400> 178
Met His Ile Thr Lys Pro His Ala Ala Met Phe Ser Ser Pro Gly Met
 1   5   10   15

```
Gly His Val Ile Pro Val Ile Glu Leu Gly Lys Arg Leu Ser Ala Asn
            20                  25                  30
Asn Gly Phe His Val Thr Val Phe Val Leu Glu Thr Asp Ala Ala Ser
        35                  40                  45
Ala Gln Ser Lys Phe Leu Asn Ser Thr Gly Val Asp Ile Val Lys Leu
    50                  55                  60
Pro Ser Pro Asp Ile Tyr Gly Leu Val Asp Pro Asp Asp His Val Val
65                  70                  75                  80
Thr Lys Ile Gly Val Ile Met Arg Ala Ala Val Pro Ala Leu Arg Ser
                85                  90                  95
Lys Ile Ala Ala Met His Gln Lys Pro Thr Ala Leu Ile Val Asp Leu
            100                 105                 110
Phe Gly Thr Asp Ala Leu Cys Leu Ala Lys Glu Phe Asn Met Leu Ser
        115                 120                 125
Tyr Val Phe Ile Pro Thr Asn Ala Arg Phe Leu Gly Val Ser Ile Tyr
    130                 135                 140
Tyr Pro Asn Leu Asp Lys Asp Ile Lys Glu Glu His Thr Val Gln Arg
145                 150                 155                 160
Asn Pro Leu Ala Ile Pro Gly Cys Glu Pro Val Arg Phe Glu Asp Thr
                165                 170                 175
Leu Asp Ala Tyr Leu Val Pro Asp Glu Pro Val Tyr Arg Asp Phe Val
            180                 185                 190
Arg His Gly Leu Ala Tyr Pro Lys Ala Asp Gly Ile Leu Val Asn Thr
        195                 200                 205
Trp Glu Glu Met Glu Pro Lys Ser Leu Lys Ser Leu Leu Asn Pro Lys
    210                 215                 220
Leu Leu Gly Arg Val Ala Arg Val Pro Val Tyr Pro Ile Gly Pro Leu
225                 230                 235                 240
Cys Arg Pro Ile Gln Ser Ser Glu Thr Asp His Pro Val Leu Asp Trp
                245                 250                 255
Leu Asn Glu Gln Pro Asn Glu Ser Val Leu Tyr Ile Ser Phe Gly Ser
            260                 265                 270
Gly Gly Cys Leu Ser Ala Lys Gln Leu Thr Glu Leu Ala Trp Gly Leu
        275                 280                 285
Glu Gln Ser Gln Gln Arg Phe Val Trp Val Val Arg Pro Pro Val Asp
    290                 295                 300
Gly Ser Cys Cys Ser Glu Tyr Val Ser Ala Asn Gly Gly Gly Thr Glu
305                 310                 315                 320
Asp Asn Thr Pro Glu Tyr Leu Pro Glu Gly Phe Val Ser Arg Thr Ser
                325                 330                 335
Asp Arg Gly Phe Val Val Pro Ser Trp Ala Pro Gln Ala Glu Ile Leu
            340                 345                 350
Ser His Arg Ala Val Gly Gly Phe Leu Thr His Cys Gly Trp Ser Ser
        355                 360                 365
Thr Leu Glu Ser Val Val Gly Gly Val Pro Met Ile Ala Trp Pro Leu
    370                 375                 380
Phe Ala Glu Gln Asn Met Asn Ala Ala Leu Leu Ser Asp Glu Leu Gly
385                 390                 395                 400
Ile Ala Val Arg Leu Asp Asp Pro Lys Glu Asp Ile Ser Arg Trp Lys
                405                 410                 415
Ile Glu Ala Leu Val Arg Lys Val Met Thr Glu Lys Glu Gly Glu Ala
            420                 425                 430
Met Arg Arg Lys Val Lys Lys Leu Arg Asp Ser Ala Glu Met Ser Leu
        435                 440                 445
Ser Ile Asp Gly Gly Gly Leu Ala His Glu Ser Leu Cys Arg Val Thr
    450                 455                 460
Lys Glu Cys Gln Arg Phe Leu Glu Arg Val Val Asp Leu Ser Arg Gly
465                 470                 475                 480
Ala
```

<210> 179
<211> 513

<212> PRT
<213> Arabidopsis thaliana

<400> 179
```
Met Gly Ala Tyr Glu Thr Glu Lys Pro Thr Lys Asp Ala Ala Ala Leu
 1               5                  10                  15
Glu Thr Gln Ser Pro Glu Asp Phe Asp Gln Pro Ser Pro Leu Arg Lys
            20                  25                  30
Ile Ile Ser Val Ala Ser Ile Ala Ala Gly Val Gln Phe Gly Trp Ala
        35                  40                  45
Leu Gln Leu Ser Leu Leu Thr Pro Tyr Val Gln Leu Leu Gly Ile Pro
    50                  55                  60
His Lys Trp Ser Ser Leu Ile Trp Leu Cys Gly Pro Val Ser Gly Met
65                  70                  75                  80
Ile Val Gln Pro Ile Val Gly Phe His Ser Asp Arg Cys Arg Ser Lys
                85                  90                  95
Phe Gly Arg Arg Arg Pro Phe Ile Ala Thr Gly Ala Ala Leu Val Ala
            100                 105                 110
Val Ala Val Phe Leu Ile Gly Tyr Ala Ala Asp Phe Gly Tyr Lys Met
        115                 120                 125
Gly Asp Lys Leu Glu Glu Lys Val Lys Val Arg Ala Ile Gly Ile Phe
    130                 135                 140
Ala Leu Gly Phe Trp Ile Leu Asp Val Ala Asn Asn Thr Leu Gln Gly
145                 150                 155                 160
Pro Cys Arg Ala Phe Leu Ala Asp Leu Ala Ala Gly Asp Ala Lys Arg
                165                 170                 175
Thr Arg Val Ala Asn Ala Phe Phe Ser Phe Phe Met Ala Val Gly Asn
            180                 185                 190
Val Leu Gly Tyr Ala Ala Gly Ser Tyr Thr Asn Leu His Lys Met Phe
        195                 200                 205
Pro Phe Thr Met Thr Lys Ala Cys Asp Ile Tyr Cys Ala Asn Leu Lys
    210                 215                 220
Thr Cys Phe Phe Leu Ser Ile Thr Leu Leu Leu Ile Val Thr Val Thr
225                 230                 235                 240
Ser Leu Trp Tyr Val Asn Asp Lys Gln Trp Ser Pro Pro Pro Arg Asn
                245                 250                 255
Ala Asp Asp Asp Glu Lys Thr Ser Ser Val Pro Leu Phe Gly Glu Ile
            260                 265                 270
Phe Gly Ala Phe Lys Val Met Lys Arg Pro Met Trp Met Leu Leu Ile
        275                 280                 285
Val Thr Ala Leu Asn Trp Ile Ala Trp Phe Pro Phe Leu Leu Phe Asp
    290                 295                 300
Thr Asp Trp Met Gly Arg Glu Val Phe Gly Gly Asp Ser Asp Gly Asn
305                 310                 315                 320
Glu Arg Ser Lys Lys Leu Tyr Ser Leu Gly Val Gln Ser Gly Ala Met
                325                 330                 335
Gly Leu Met Phe Asn Ser Ile Val Leu Gly Phe Met Ser Leu Gly Val
            340                 345                 350
Glu Trp Ile Gly Arg Lys Leu Gly Gly Ala Lys Arg Leu Trp Gly Ile
        355                 360                 365
Val Asn Phe Ile Leu Ala Ala Gly Leu Ala Met Thr Val Leu Val Thr
    370                 375                 380
Lys Phe Ala Glu Asp His Arg Lys Thr Ala Gly Asp Leu Ala Gly Pro
385                 390                 395                 400
Ser Ala Ser Val Lys Ala Gly Ala Leu Ser Leu Phe Ala Val Leu Gly
                405                 410                 415
Ile Pro Leu Ala Ile Thr Phe Ser Thr Pro Phe Ala Leu Ala Ser Ile
            420                 425                 430
Phe Ser Ser Cys Ser Gly Ala Gly Gln Gly Leu Ser Leu Gly Val Leu
        435                 440                 445
Asn Leu Ala Ile Val Ile Pro Gln Met Ile Val Ser Leu Gly Gly Gly
    450                 455                 460
Pro Phe Asp Ala Leu Phe Gly Gly Gly Asn Leu Pro Ala Phe Ile Val
```

465 470 475 480
Ala Ala Ile Ala Ala Ala Ile Ser Gly Val Leu Ala Leu Thr Val Leu
485 490 495
Pro Ser Pro Pro Pro Asp Ala Pro Lys Ala Thr Thr Met Gly Gly Phe
500 505 510
His

<210> 180
<211> 806
<212> PRT
<213> Coffea arabica

<400> 180
Met Ala Glu Arg Val Leu Thr Arg Val His Ser Leu Arg Glu Arg Leu
1 5 10 15
Asp Ala Thr Leu Ala Ala His Arg Asn Asp Val Leu Leu Phe Met Ser
20 25 30
Arg Leu Glu Thr His Gly Lys Gly Ile Leu Lys Pro His Gln Leu Leu
35 40 45
Ala Glu Phe Glu Glu Ile Asn Lys Asp Gly Lys Gln Lys Ile His Asp
50 55 60
His Ala Phe Glu Glu Val Leu Lys Ser Thr Gln Glu Ala Ile Val Leu
65 70 75 80
Pro Pro Trp Val Ala Leu Ala Ile Arg Leu Arg Pro Gly Val Trp Glu
85 90 95
Tyr Val Arg Val Asn Val His Ala Leu Val Val Glu Glu Leu Thr Val
100 105 110
Pro Glu Tyr Leu His Phe Lys Glu Glu Leu Val Asp Gly Ser Lys Asn
115 120 125
Gly Asn Phe Val Leu Glu Leu Asp Phe Glu Pro Phe Thr Ala Ser Phe
130 135 140
Pro Lys Pro Thr Leu Thr Lys Tyr Ile Gly Asp Gly Val Glu Phe Leu
145 150 155 160
Asn Arg His Leu Ser Ala Lys Met Phe His Asp Lys Glu Ser Met Ala
165 170 175
Pro Leu Leu Asp Phe Leu Arg Val His Gln Tyr Lys Gly Lys Thr Met
180 185 190
Met Leu Asn Asp Arg Ile Lys Asp Leu Asn Thr Leu Gln Ala Val Leu
195 200 205
Arg Lys Ala Glu Glu Tyr Leu Thr Thr Leu Ser Ala Asp Thr Pro Tyr
210 215 220
Ser Glu Phe Glu His Lys Phe Gln Glu Ile Gly Leu Glu Arg Gly Trp
225 230 235 240
Gly Asp Thr Ala Glu Arg Val Leu Glu Met Ile Cys Met Leu Leu Asp
245 250 255
Leu Leu Gly Ala Pro Asp Ser Cys Thr Leu Glu Lys Phe Leu Gly Arg
260 265 270
Ile Pro Met Val Phe Asn Val Val Ile Leu Ser Pro His Gly Tyr Phe
275 280 285
Ala Gln Glu Asn Val Leu Gly Tyr Pro Asp Thr Gly Gly Gln Val Val
290 295 300
Tyr Ile Leu Asp Gln Val Pro Ala Leu Glu Arg Glu Met Leu Lys Arg
305 310 315 320
Ile Lys Glu Gln Gly Leu Asp Val Lys Pro Arg Ile Leu Ile Ile Thr
325 330 335
Arg Leu Leu Pro Asp Ala Pro Gly Thr Thr Cys Gly Gln Arg Leu Glu
340 345 350
Lys Val Tyr Gly Ser Glu Tyr Ser His Ile Leu Arg Val Pro Phe Arg
355 360 365
Thr Glu Lys Gly Val Val Arg Lys Trp Ile Ser Arg Phe Glu Val Trp
370 375 380

Pro Tyr Met Glu Thr Phe Thr Glu Asp Val Ala Lys Glu Val Thr Ala
385 390 395 400
Glu Leu Gln Ala Lys Pro Asp Leu Val Ile Gly Asn Tyr Ser Glu Gly
405 410 415
Asn Leu Val Ala Ser Leu Leu Ala His Lys Leu Gly Val Thr Gln Cys
420 425 430
Thr Ile Ala His Ala Leu Glu Lys Thr Lys Tyr Pro Asp Ser Asp Ile
435 440 445
Tyr Leu Ser Lys Phe Asp Glu Lys Tyr His Phe Ser Cys Gln Phe Thr
450 455 460
Ala Asp Leu Ile Ala Met Asn His Thr Asp Phe Ile Ile Thr Ser Thr
465 470 475 480
Phe Gln Glu Ile Ala Gly Ser Lys Asp Thr Val Gly Gln Tyr Glu Ser
485 490 495
His Met Ala Phe Thr Met Pro Gly Leu Tyr Arg Val Val His Gly Ile
500 505 510
Asp Val Phe Asp Pro Lys Phe Asn Ile Val Ser Pro Gly Ala Asp Thr
515 520 525
Asn Leu Tyr Phe Pro His Thr Glu Lys Glu Lys Arg Leu Thr Ser Phe
530 535 540
His Pro Glu Ile Glu Glu Leu Leu Phe Ser Asp Val Glu Asn Glu Glu
545 550 555 560
His Leu Cys Val Leu Lys Asp Lys Lys Lys Pro Ile Leu Phe Thr Met
565 570 575
Ala Arg Leu Asp Arg Val Lys Asn Leu Thr Gly Leu Val Glu Leu Tyr
580 585 590
Ala Lys Asn Pro Lys Leu Arg Glu Leu Val Asn Leu Val Val Val Gly
595 600 605
Gly Asp Arg Arg Lys Glu Ser Lys Asp Leu Glu Glu Gln Ala Glu Met
610 615 620
Lys Lys Met Tyr Ser Leu Ile Glu Thr Tyr Asn Leu Asn Gly Gln Phe
625 630 635 640
Arg Trp Ile Ser Ser Gln Met Asn Arg Val Arg Asn Gly Glu Leu Tyr
645 650 655
Arg Tyr Ile Ala Asp Thr Arg Gly Ala Phe Val Gln Pro Ala Phe Tyr
660 665 670
Glu Ala Phe Gly Leu Thr Val Val Glu Ala Met Thr Cys Gly Leu Pro
675 680 685
Thr Phe Ala Thr Asn His Gly Gly Pro Ala Glu Ile Ile Ile His Gly
690 695 700
Lys Ser Gly Phe His Ile Asp Pro Tyr His Gly Glu Gln Val Ser Glu
705 710 715 720
Leu Leu Ala Asn Phe Phe Glu Arg Cys Lys Lys Glu Pro Ser Tyr Trp
725 730 735
Asp Thr Ile Ser Ala Gly Gly Leu Lys Arg Ile Gln Glu Lys Tyr Thr
740 745 750
Trp Gln Ile Tyr Ser Asp Arg Leu Leu Thr Leu Ala Gly Val Tyr Gly
755 760 765
Phe Trp Lys Cys Val Ser Lys Leu Asp Arg Gln Glu Ile Arg Arg Tyr
770 775 780
Leu Glu Met Phe Tyr Ala Leu Lys Tyr Arg Lys Leu Ala Glu Ala Val
785 790 795 800
Pro Leu Ala Val Asp Gln
805

<210> 181
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Hairpin loop forming sequence

<400> 181
tgaattcgtt aacgaattc 19

<210> 182
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Hairpin loop forming sequence

<400> 182
tgaattcgtt aacgaactc 19

<210> 183
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Hairpin loop forming sequence

<400> 183
tgaattcgtt aacgaagtc 19

<210> 184
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Hairpin loop forming sequence

<400> 184
tgaattcgtt aacgaaatt 19

<210> 185
<211> 287
<212> DNA
<213> Saccharomyces cerevisiae

<400> 185
gcgttggttg gtggatcaag cccacgcgta ggcaatcctc gagcagatcc gccaggcgtg 60
tatatatagc gtggatggcc aggcaacttt agtgctgaca catacaggca tatatatatg 120
tgtgcgacga cacatgatca tatggcatgc atgtgctctg tatgtatata aaactcttgt 180
tttcttcttt tctctaaata ttctttcctt atacattagg acctttgcag cataaattac 240
tatacttcta tagacacaca aacacaaata cacacactaa attaata 287

<210> 186
<211> 491
<212> DNA
<213> Saccharomyces cerevisiae

<400> 186
ggtcagcagc tctgatgtag atacacgtat ctcgacatgt tttattttta ctatacatac 60
ataaaagaaa taaaaaatga taacgtgtat attattattc atataatcaa tgagggtcat 120
tttctgaaac gcaaaaaacg gtaaatggaa aaaaaataaa gatagaaaaa gaaaacaaac 180
aaaggaaagg ttagcatatt aaataactga gctgatactt caacagcatc gctgaagaga 240
acagtattga aaccgaaaca ttttctaaag gcaaacaagg tactccatat ttgctggacg 300
tgttctttct ctcgtttcat atgcataatt ctgtcataag cctgttcttt ttcctggctt 360
aaacatcccg ttttgtaaaa gagaaatcta ttccacatat ttcattcatt cggctaccat 420
actaaggata aactaatccc gttgtttttt ggcctcgtca cataattata aactactaac 480

ccattatcag a 491

<210> 187
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Hairpin loop forming sequence

<400> 187
cacaaataca cacactaaat taataatg 28

<210> 188
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Hairpin loop forming sequence

<400> 188
cacaaataca cacactaaat taatatgaat tcgttaacga attcatg 47

<210> 189
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Hairpin loop forming sequence

<400> 189
cacaaataca cacactaaat taatatgaat tcgttaacga actcatg 47

<210> 190
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Hairpin loop forming sequence

<400> 190
cacaaataca cacactaaat taatatgaat tcgttaacga agtcatg 47

<210> 191
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Hairpin loop forming sequence

<400> 191
cacaaataca cacactaaat taatatgaat tcgttaacga aattatg 47

<210> 192
<211> 444
<212> PRT
<213> Saccharomyces cerevisiae

<400> 192

```
Met Gly Lys Leu Leu Gln Leu Ala Leu His Pro Val Glu Met Lys Ala
 1               5                  10                  15
Ala Leu Lys Leu Lys Phe Cys Arg Thr Pro Leu Phe Ser Ile Tyr Asp
            20                  25                  30
Gln Ser Thr Ser Pro Tyr Leu Leu His Cys Phe Glu Leu Leu Asn Leu
        35                  40                  45
Thr Ser Arg Ser Phe Ala Ala Val Ile Arg Glu Leu His Pro Glu Leu
    50                  55                  60
Arg Asn Cys Val Thr Leu Phe Tyr Leu Ile Leu Arg Ala Leu Asp Thr
65                  70                  75                  80
Ile Glu Asp Asp Met Ser Ile Glu His Asp Leu Lys Ile Asp Leu Leu
                85                  90                  95
Arg His Phe His Glu Lys Leu Leu Leu Thr Lys Trp Ser Phe Asp Gly
            100                 105                 110
Asn Ala Pro Asp Val Lys Asp Arg Ala Val Leu Thr Asp Phe Glu Ser
        115                 120                 125
Ile Leu Ile Glu Phe His Lys Leu Lys Pro Glu Tyr Gln Glu Val Ile
    130                 135                 140
Lys Glu Ile Thr Glu Lys Met Gly Asn Gly Met Ala Asp Tyr Ile Leu
145                 150                 155                 160
Asp Glu Asn Tyr Asn Leu Asn Gly Leu Gln Thr Val His Asp Tyr Asp
                165                 170                 175
Val Tyr Cys His Tyr Val Ala Gly Leu Val Gly Asp Gly Leu Thr Arg
            180                 185                 190
Leu Ile Val Ile Ala Lys Phe Ala Asn Glu Ser Leu Tyr Ser Asn Glu
        195                 200                 205
Gln Leu Tyr Glu Ser Met Gly Leu Phe Leu Gln Lys Thr Asn Ile Ile
    210                 215                 220
Arg Asp Tyr Asn Glu Asp Leu Val Asp Gly Arg Ser Phe Trp Pro Lys
225                 230                 235                 240
Glu Ile Trp Ser Gln Tyr Ala Pro Gln Leu Lys Asp Phe Met Lys Pro
                245                 250                 255
Glu Asn Glu Gln Leu Gly Leu Asp Cys Ile Asn His Leu Val Leu Asn
            260                 265                 270
Ala Leu Ser His Val Ile Asp Val Leu Thr Tyr Leu Ala Gly Ile His
        275                 280                 285
Glu Gln Ser Thr Phe Gln Phe Cys Ala Ile Pro Gln Val Met Ala Ile
    290                 295                 300
Ala Thr Leu Ala Leu Val Phe Asn Asn Arg Glu Val Leu His Gly Asn
305                 310                 315                 320
Val Lys Ile Arg Lys Gly Thr Thr Cys Tyr Leu Ile Leu Lys Ser Arg
                325                 330                 335
Thr Leu Arg Gly Cys Val Glu Ile Phe Asp Tyr Tyr Leu Arg Asp Ile
            340                 345                 350
Lys Ser Lys Leu Ala Val Gln Asp Pro Asn Phe Leu Lys Leu Asn Ile
        355                 360                 365
Gln Ile Ser Lys Ile Glu Gln Phe Met Glu Glu Met Tyr Gln Asp Lys
    370                 375                 380
Leu Pro Pro Asn Val Lys Pro Asn Glu Thr Pro Ile Phe Leu Lys Val
385                 390                 395                 400
Lys Glu Arg Ser Arg Tyr Asp Asp Glu Leu Val Pro Thr Gln Gln Glu
                405                 410                 415
Glu Glu Tyr Lys Phe Asn Met Val Leu Ser Ile Ile Leu Ser Val Leu
            420                 425                 430
Leu Gly Phe Tyr Tyr Ile Tyr Thr Leu His Arg Ala
        435                 440
```

<210> 193
<211> 460
<212> PRT
<213> Schizosaccharomyces pombe

<400> 193

```
Met Ser Leu Ala Asn Arg Ile Glu Glu Ile Arg Cys Leu Cys Gln Tyr
 1               5                  10                  15
Lys Leu Trp Asn Asp Leu Pro Ser Tyr Gly Glu Asp Glu Asn Val Pro
            20                  25                  30
Gln Asn Ile Arg Arg Cys Tyr Gln Leu Leu Asp Met Thr Ser Arg Ser
        35                  40                  45
Phe Ala Val Val Ile Lys Glu Leu Pro Asn Gly Ile Arg Glu Ala Val
    50                  55                  60
Met Ile Phe Tyr Leu Val Leu Arg Gly Leu Asp Thr Val Glu Asp Asp
65                  70                  75                  80
Met Thr Leu Pro Leu Asp Lys Lys Leu Pro Ile Leu Arg Asp Phe Tyr
                85                  90                  95
Lys Thr Ile Glu Val Glu Gly Trp Thr Phe Asn Glu Ser Gly Pro Asn
            100                 105                 110
Glu Lys Asp Arg Gln Leu Leu Val Glu Phe Asp Val Val Ile Lys Glu
        115                 120                 125
Tyr Leu Asn Leu Ser Glu Gly Tyr Arg Asn Val Ile Ser Asn Ile Thr
    130                 135                 140
Lys Glu Met Gly Asp Gly Met Ala Tyr Tyr Ala Ser Leu Ala Glu Lys
145                 150                 155                 160
Asn Asp Gly Phe Ser Val Glu Thr Ile Glu Asp Phe Asn Lys Tyr Cys
                165                 170                 175
His Tyr Val Ala Gly Leu Val Gly Ile Gly Leu Ser Arg Leu Phe Ala
            180                 185                 190
Gln Ser Lys Leu Glu Asp Pro Asp Leu Ala His Ser Gln Ala Ile Ser
        195                 200                 205
Asn Ser Leu Gly Leu Phe Leu Gln Lys Val Asn Ile Ile Arg Asp Tyr
    210                 215                 220
Arg Glu Asp Phe Asp Asp Asn Arg His Phe Trp Pro Arg Glu Ile Trp
225                 230                 235                 240
Ser Lys Tyr Thr Ser Ser Phe Gly Asp Leu Cys Leu Pro Asp Asn Ser
                245                 250                 255
Glu Lys Ala Leu Glu Cys Leu Ser Asp Met Thr Ala Asn Ala Leu Thr
            260                 265                 270
His Ala Thr Asp Ala Leu Val Tyr Leu Ser Gln Leu Lys Thr Gln Glu
        275                 280                 285
Ile Phe Asn Phe Cys Ala Ile Pro Gln Val Met Ala Ile Ala Thr Leu
    290                 295                 300
Ala Ala Val Phe Arg Asn Pro Asp Val Phe Gln Thr Asn Val Lys Ile
305                 310                 315                 320
Arg Lys Gly Gln Ala Val Gln Ile Ile Leu His Ser Val Asn Leu Lys
                325                 330                 335
Asn Val Cys Asp Leu Phe Leu Arg Tyr Thr Arg Asp Ile His Tyr Lys
            340                 345                 350
Asn Thr Pro Lys Asp Pro Asn Phe Leu Lys Ile Ser Ile Glu Cys Gly
        355                 360                 365
Lys Ile Glu Gln Val Ser Glu Ser Leu Phe Pro Arg Arg Phe Arg Glu
    370                 375                 380
Met Tyr Glu Lys Ala Tyr Val Ser Lys Leu Ser Glu Gln Lys Lys Gly
385                 390                 395                 400
Asn Gly Thr Gln Lys Ala Ile Leu Asn Asp Glu Gln Lys Glu Leu Tyr
                405                 410                 415
Arg Lys Asp Leu Gln Lys Leu Gly Ile Ser Ile Leu Phe Val Phe Phe
            420                 425                 430
Ile Ile Leu Val Cys Leu Ala Val Ile Phe Tyr Val Phe Asn Ile Arg
        435                 440                 445
Ile His Trp Ser Asp Phe Lys Glu Leu Asn Leu Phe
    450                 455                 460
```

<210> 194
<211> 445

<212> PRT
<213> Yarrowia lipolytica

<400> 194

```
Met Gly Lys Leu Ile Glu Leu Leu Leu His Pro Ser Glu Leu Ser Ala
 1               5                  10                  15
Ala Ile His Tyr Lys Leu Trp Arg Gln Pro Leu His Pro Arg Asp Leu
            20                  25                  30
Ser Lys Glu Ser Thr Glu Leu Arg Arg Cys Tyr Glu Leu Leu Asp Val
        35                  40                  45
Cys Ser Arg Ser Phe Ala Ala Val Ile Arg Glu Leu His Pro Glu Val
    50                  55                  60
Arg Asp Ala Val Met Leu Phe Tyr Leu Ile Leu Arg Ala Leu Asp Thr
65                  70                  75                  80
Ile Glu Asp Asp Met Thr Leu Ser Arg Asp Ile Lys Ile Pro Ile Leu
                85                  90                  95
Arg Asp Phe Thr Lys Cys Met Lys Thr Pro Gly Trp Lys Phe Thr Asp
            100                 105                 110
Ser Asp Pro Asn Glu Arg Asp Arg Val Val Leu Gln Glu Phe Pro Val
        115                 120                 125
Val Met Thr Glu Phe Asn Lys Leu Lys Pro Lys Tyr Gln Glu Val Ile
    130                 135                 140
Tyr Asp Ile Thr Asp Arg Met Gly Asn Gly Met Ala Asp Tyr Val Ile
145                 150                 155                 160
Asp Asp Asp Phe Asn Asn Asn Gly Val Asp Thr Ile Ala Ala Tyr Asp
                165                 170                 175
Leu Tyr Cys His His Val Ala Gly Ile Val Gly Glu Gly Leu Thr Arg
            180                 185                 190
Ile Thr Ile Leu Ala Gly Phe Gly Thr Asp Val Leu His Glu Asn Pro
        195                 200                 205
Arg Leu Gln Glu Ser Met Gly Leu Phe Leu Gln Lys Val Asn Ile Ile
    210                 215                 220
Arg Asp Tyr Arg Glu Asp Ile Asp Val Asn Arg Ala Phe Trp Pro Arg
225                 230                 235                 240
Glu Ile Trp His Lys Tyr Ala Glu Glu Met Arg Asp Phe Lys Asp Pro
                245                 250                 255
Lys Tyr Ser Lys Lys Ala Leu His Cys Thr Ser Asp Leu Val Ala Asn
            260                 265                 270
Ala Leu Gly His Ala Thr Asp Cys Leu Asp Tyr Leu Asp Asn Val Thr
        275                 280                 285
Asp Pro Ser Thr Phe Thr Phe Cys Ala Ile Pro Gln Val Met Ala Ile
    290                 295                 300
Ala Thr Leu Asp Leu Val Tyr Arg Asn Pro Asp Val Phe Gln Lys Asn
305                 310                 315                 320
Val Lys Leu Arg Lys Gly Thr Thr Val Ser Leu Ile Leu Glu Ala Ser
                325                 330                 335
Asn Val Ser Gly Val Cys Asp Ile Phe Thr Arg Tyr Ala Arg Lys Val
            340                 345                 350
Tyr Lys Lys Ser Asp Pro Asn Asp Pro Asn Tyr Phe Arg Val Ser Val
        355                 360                 365
Leu Cys Gly Lys Ile Glu Gln His Ala Ala Leu Ile Lys Arg Gln Arg
    370                 375                 380
Gly Pro Pro Ala Lys Thr Ile Ala Gln Leu Glu Gly Glu Arg Lys Glu
385                 390                 395                 400
Met Ala Leu Ser Leu Ile Val Cys Leu Ala Val Ile Phe Ser Met Ser
                405                 410                 415
Gly Leu Met Ala Tyr Ile Ala Tyr Val Ser Gly Phe Arg Trp Ser Pro
            420                 425                 430
Arg Glu Ile Phe Asp Ser Lys Met Phe Pro Leu Arg Asp
        435                 440                 445
```

<210> 195

<211> 443
<212> PRT
<213> Candida glabrata

<400> 195

```
Met Gly Lys Val Leu Asp Leu Ala Leu His Pro Leu Glu Leu Arg Ala
 1               5                  10                  15
Ala Leu Lys Leu Lys Phe Ile Arg Gln Pro Leu Phe Ser Thr Asn Asp
            20                  25                  30
Thr Arg Ala Thr Pro Gln Leu Glu Arg Cys Tyr Glu Leu Leu Asn Leu
        35                  40                  45
Thr Ser Arg Ser Phe Ala Ala Val Ile Met Glu Leu His Pro Glu Leu
    50                  55                  60
Arg Asn Val Ile Met Val Phe Tyr Leu Ile Leu Arg Ala Leu Asp Thr
65                  70                  75                  80
Val Glu Asp Asp Met Thr Ile Asp Pro Gln Leu Lys Val Lys Val Leu
                85                  90                  95
Arg Glu Phe Asp Ser Lys Leu Asp Thr Thr Asp Trp Ser Phe Asp Gly
            100                 105                 110
Asn Asp Leu Lys Glu Lys Asp Arg Val Val Leu Thr Glu Phe Pro Cys
        115                 120                 125
Ile Leu Gly Glu Tyr His Lys Leu Lys Pro Glu Tyr Gln Lys Val Ile
    130                 135                 140
Lys Arg Ile Thr Gly Leu Met Gly Asn Gly Met Ala Asp Tyr Ile Leu
145                 150                 155                 160
Asp Glu Asn Phe Asn Leu Asn Gly Val Gln Thr Val Lys Asp Tyr Asp
                165                 170                 175
Lys Tyr Cys His Tyr Val Ala Gly Leu Val Gly Asp Gly Leu Thr Glu
            180                 185                 190
Leu Ile Val Leu Ala Gly Phe Gly Ser Asp Asp Leu Tyr His Gly Lys
        195                 200                 205
Asn Ser Phe Gln Leu Tyr Glu Ser Met Gly Leu Phe Leu Gln Lys Thr
    210                 215                 220
Asn Ile Ile Arg Asp Tyr Ala Glu Asp Leu Asp Asp Gly Arg Ser Phe
225                 230                 235                 240
Trp Pro Lys Glu Ile Trp Ser Glu Tyr Ala Thr Lys Leu Thr Asp Phe
                245                 250                 255
Arg Asp Pro Lys Asn Thr Gln Lys Gly Val Asp Cys Ile Asn His Leu
            260                 265                 270
Val Leu Asn Ala Leu Thr His Val Ile Asp Val Leu Thr Tyr Leu Ser
        275                 280                 285
Ser Ile His Glu Gln Ser Ser Phe Gln Phe Cys Ala Ile Pro Gln Val
    290                 295                 300
Met Ala Ile Ala Thr Leu Ala Lys Val Phe Asn Asn Pro Glu Val Leu
305                 310                 315                 320
Arg Lys Asn Val Lys Ile Arg Lys Gly Thr Thr Cys Asp Leu Ile Leu
                325                 330                 335
Asn Ser Arg Thr Leu Lys Gly Cys Val Asp Ile Phe Gln Tyr Tyr Leu
            340                 345                 350
Arg Asp Met Lys Gln Arg Leu Pro Val Glu Asp Pro Asn Tyr Leu Lys
        355                 360                 365
Phe Asn Ile Gln Val Ala Lys Ile Glu Gln Phe Ile Glu Glu Met Phe
    370                 375                 380
Gln Asp Asn Leu Pro Ala Gly Val Glu Pro Arg Glu Thr Met Ile Tyr
385                 390                 395                 400
Leu Lys Val Gln Glu Arg Leu Lys Trp Asp Thr Gln Val Ile Pro Arg
                405                 410                 415
Val Gln Glu Glu Asp Tyr Lys Phe Asn Met Ala Leu Ser Val Val Phe
            420                 425                 430
Cys Val Leu Leu Ser Phe Tyr Phe Phe Thr Lys
        435                 440
```

<210> 196
<211> 441
<212> PRT
<213> Ashbya gossypii

<400> 196
Met Gly Lys Val Val Gln Leu Phe Thr His Pro Leu Glu Leu Lys Ala
 1   5   10   15
Ala Leu Lys Leu Lys Phe Leu Arg Glu Pro Leu Tyr Pro Ala Asp Asp
            20   25   30
Thr Gln Gly Ser Ala Glu Leu Lys Arg Cys Tyr Gln Leu Leu Gln Arg
        35   40   45
Thr Ser Arg Ser Phe Ala Ala Val Ile Met Glu Leu His Pro Glu Leu
    50   55   60
Arg Asn Ala Val Met Leu Phe Tyr Leu Ile Leu Arg Ala Leu Asp Thr
65   70   75   80
Val Glu Asp Asp Met Thr Ile Ser Pro Lys Val Lys Val Pro Leu Leu
                85   90   95
Arg Glu Phe Asp Gln Lys Leu Lys Leu Asp Thr Trp Ser Phe Asp Gly
            100   105   110
Asn Ala Lys Thr Glu Lys Asp Arg Asp Val Leu Val Glu Phe Ser Thr
        115   120   125
Ile Leu Ala Glu Phe His Lys Leu Lys Pro Glu Tyr Gln Gln Val Ile
    130   135   140
Ala Asp Ile Thr His Lys Met Gly Asn Gly Met Ala Asp Tyr Ile Leu
145   150   155   160
Asp Glu Lys Phe Asn Leu Ser Gly Leu Glu Thr Ile Gln Asp Tyr Asp
                165   170   175
Arg Tyr Cys His Tyr Val Ala Gly Leu Val Gly Asp Gly Leu Thr His
            180   185   190
Leu Ile Met Leu Ala Lys Phe Ser Ser Pro Gly Leu Tyr Tyr Asp Ser
        195   200   205
Pro Asp Leu Tyr Glu Ser Met Gly Leu Phe Leu Gln Lys Thr Asn Ile
    210   215   220
Ile Arg Asp Tyr Ala Glu Asp Leu Ala Asp Gly Arg Ser Phe Trp Pro
225   230   235   240
Lys Glu Ile Trp Ser His Tyr Ala Asp Asp Leu Ala Ser Phe Ser Lys
                245   250   255
Pro Glu Asn Ala Thr Ala Gly Val Tyr Cys Ile Asn His Leu Val Leu
            260   265   270
Asn Ala Leu Gly His Val Gln His Val Leu Thr Tyr Leu Ala Ser Leu
        275   280   285
Arg Glu Gln Ser Ser Phe Gln Phe Cys Ala Ile Pro Gln Val Met Ala
    290   295   300
Ile Ala Thr Leu Ala Leu Val Phe Gly Asn Glu Arg Val Leu Gln Thr
305   310   315   320
Ser Val Lys Ile Arg Lys Gly Thr Thr Cys Tyr Leu Ile Leu Lys Ser
                325   330   335
Arg Thr Phe Gln Gly Cys Val Glu Ile Phe Glu His Tyr Leu Arg Asp
            340   345   350
Ile Arg Lys Arg Leu Thr Val Ala Asp Pro Asn Tyr Leu Lys Leu Asn
        355   360   365
Ile Glu Ile Ala Lys Leu Asp Lys Phe Ile Glu Glu Met Tyr Gln Asp
    370   375   380
Lys Leu Pro Val Gly Ala Lys Pro Gln Glu Thr Glu Ile Tyr Lys Lys
385   390   395   400
Val Arg Glu Arg Ser Ala Tyr Asp Leu Glu Val Leu Pro Arg Glu Gln
                405   410   415
Glu Glu Glu Phe Lys Phe Asn Val Leu Leu Ser Ile Leu Phe Thr Val
            420   425   430
Phe Gly Ala Leu Tyr Trp Tyr Ala Lys
        435   440

<210> 197
<211> 443
<212> PRT
<213> Cyberlindnera jadinii

<400> 197
Met Gly Lys Leu Leu Gln Leu Ala Leu His Pro Asp Glu Leu Ala Ser
 1   5   10   15
Ile Val Gln Phe Lys Leu Phe Arg Lys Asn Glu Asn Ala Arg Asn Pro
            20   25   30
Ala Thr Glu Ser Ala Glu Leu Ile Arg Cys Tyr Glu Leu Leu Asn Leu
        35   40   45
Thr Ser Arg Ser Phe Ala Ala Val Ile Glu Glu Leu His Pro Glu Leu
    50   55   60
Arg Asn Val Ile Met Val Phe Tyr Leu Val Leu Arg Ala Leu Asp Thr
65   70   75   80
Val Glu Val Asp Met Ser Ile Glu Asn Ser Val Lys Leu Pro Val Leu
                85   90   95
Arg Gln Phe His Glu Lys Leu Asp Thr Lys Asp Trp Thr Phe Asp Gly
            100   105   110
Asn Ser Pro Asn Glu Lys Asp Arg Cys Val Leu Val Glu Phe Asp Arg
        115   120   125
Ile Leu Gly Gln Tyr His Glu Leu Lys Pro Gln Tyr Gln Lys Val Ile
    130   135   140
Lys Glu Ile Thr Glu Lys Met Gly Asn Gly Met Ala Asp Tyr Ile Glu
145   150   155   160
Asn Glu Asn Phe Asn Ser Asn Gly Leu Leu Thr Ile Glu Asp Tyr Asp
                165   170   175
Leu Tyr Cys Tyr Tyr Val Ala Gly Leu Val Gly Asp Gly Leu Thr Gln
            180   185   190
Leu Ile Val Leu Ala Lys Phe Gly Asn Ser Glu Leu Ser Val Asn Lys
        195   200   205
Gln Leu Phe Lys Ser Met Gly Leu Phe Leu Gln Lys Thr Asn Ile Ile
    210   215   220
Arg Asp Tyr Glu Glu Asp Gln Val Asp Gly Arg Ala Phe Trp Pro Lys
225   230   235   240
Glu Ile Trp Gly Lys Tyr Ala Asn Glu Leu Ser Asp Phe Met Lys Pro
                245   250   255
Glu Asn Gln Ser Gln Gly Leu Trp Cys Ile Ser Glu Leu Val Cys Asn
            260   265   270
Ala Leu Asp His Val Ile Asp Val Leu Gln Tyr Leu Ala Leu Val Glu
        275   280   285
Glu Gln Thr Ser Phe Asn Phe Cys Ala Ile Pro Gln Val Met Ala Ile
    290   295   300
Ala Thr Leu Glu Leu Val Phe Gln Asn Pro Gln Val Leu Thr Gln His
305   310   315   320
Val Lys Ile Arg Lys Gly Thr Thr Val Ser Leu Ile Leu Glu Ser Arg
                325   330   335
Thr Leu Glu Gly Cys Ala Arg Ile Phe Arg Arg Tyr Leu Arg Lys Ile
            340   345   350
His His Lys Ser His Pro Ser Asp Pro Asn Tyr Leu Arg Leu Gly Ile
        355   360   365
Thr Ile Gly Lys Ile Glu Gln Phe Leu Asp Gly Met Tyr Pro His Tyr
    370   375   380
Val Pro Lys Gly Ile Thr Pro Gln Thr Thr Ser Ile Arg Thr Gln Val
385   390   395   400
Val Lys Arg Leu Gln Leu Asp Glu Pro Met Lys Arg Asp Ile Asp Glu
                405   410   415
Glu Ile Leu Lys Thr Arg Ile Leu Leu Leu Ser Leu Gly Val Ala Val
            420   425   430
Phe Gly Val Val Tyr Gly Val Val Arg Ile Ile
        435   440

<210> 198
<211> 448
<212> PRT
<213> Candida albicans

<400> 198
Met Gly Lys Phe Leu Gln Leu Leu Ser His Pro Thr Glu Leu Lys Ala
 1   5   10   15
Val Ile Gln Leu Phe Gly Phe Arg Gln Pro Leu His Pro Gly Lys Arg
            20   25   30
Asp Val Asn Asp Lys Glu Leu Gly Arg Cys Tyr Glu Leu Leu Asn Leu
        35   40   45
Thr Ser Arg Ser Phe Ala Ala Val Ile Glu Glu Leu His Pro Glu Leu
    50   55   60
Arg Asp Ala Val Met Ile Phe Tyr Leu Val Leu Arg Ala Leu Asp Thr
65   70   75   80
Ile Glu Asp Asp Met Thr Ile Lys Ser Ser Ile Lys Ile Pro Leu Leu
                85   90   95
Arg Glu Phe Asp Thr Lys Leu Asn Thr Lys Asn Trp Thr Phe Asp Gly
            100   105   110
Tyr Gly Pro Asn Glu Lys Asp Arg Thr Val Leu Val Glu Phe Asp Lys
        115   120   125
Ile Leu Asn Val Tyr His Arg Leu Lys Pro Gln Tyr Gln Asp Ile Ile
    130   135   140
Lys Ser Ile Thr Phe Lys Met Gly Asn Gly Met Ala Asp Tyr Ile Leu
145   150   155   160
Asp Glu Glu Phe Asn Val Tyr Gly Val Ala Thr Val Glu Asp Tyr Asn
                165   170   175
Leu Tyr Cys His Tyr Val Ala Gly Leu Val Gly Glu Gly Leu Thr Asn
            180   185   190
Leu Phe Val Leu Ala Asn Phe Gly Asp Lys Thr Leu Thr Glu Asn Asn
        195   200   205
Phe Ala Lys Ala Asp Ser Met Gly Leu Phe Leu Gln Lys Thr Asn Ile
    210   215   220
Ile Arg Asp Tyr His Glu Asp Leu Gln Asp Gly Arg Ser Phe Trp Pro
225   230   235   240
Arg Glu Ile Trp Ser Lys Tyr Thr Glu Asn Leu Gln Asp Phe His Lys
                245   250   255
Val Lys Thr Pro Ala Lys Glu Phe Ala Gly Val Ser Cys Ile Asn Glu
            260   265   270
Leu Val Leu Asn Ala Leu Gly His Val Thr Asp Cys Leu Asp Tyr Leu
        275   280   285
Ser Leu Val Lys Asp Pro Ser Ser Phe Ser Phe Cys Ala Ile Pro Gln
    290   295   300
Val Met Ala Val Ala Thr Leu Ala Glu Val Tyr Asn Asn Pro Lys Val
305   310   315   320
Leu His Gly Val Val Lys Ile Arg Lys Gly Thr Thr Cys Arg Leu Ile
                325   330   335
Leu Glu Ser Arg Thr Leu Pro Gly Val Val Lys Ile Phe Lys Glu Tyr
            340   345   350
Ile Gln Val Ile Asn His Lys Ser Ser Val Arg Asp Pro Asn Tyr Leu
        355   360   365
Lys Ile Gly Ile Lys Cys Gly Glu Ile Glu Gln Tyr Cys Glu Met Ile
    370   375   380
Tyr Pro Asn Lys Gln Ala Leu Pro Pro Ser Met Lys Ser Leu Pro Glu
385   390   395   400
Asn Lys Phe Thr Lys Ile Val Ala Ser Arg Glu Ser Ile Asp Leu Ser
                405   410   415
Val Gln Arg Arg Ile Glu Pro Gly Asn Phe Asn Cys Asn Val Val Leu
            420   425   430
Phe Gly Ile Gly Ala Leu Ile Leu Ser Leu Ile Tyr Phe Val Leu Tyr

435 440 445

<210> 199
<211> 731
<212> PRT
<213> Saccharomyces cerevisiae

<400> 199
Met Thr Glu Phe Tyr Ser Asp Thr Ile Gly Leu Pro Lys Thr Asp Pro
1 5 10 15
Arg Leu Trp Arg Leu Arg Thr Asp Glu Leu Gly Arg Glu Ser Trp Glu
20 25 30
Tyr Leu Thr Pro Gln Gln Ala Ala Asn Asp Pro Pro Ser Thr Phe Thr
35 40 45
Gln Trp Leu Leu Gln Asp Pro Lys Phe Pro Gln Pro His Pro Glu Arg
50 55 60
Asn Lys His Ser Pro Asp Phe Ser Ala Phe Asp Ala Cys His Asn Gly
65 70 75 80
Ala Ser Phe Phe Lys Leu Leu Gln Glu Pro Asp Ser Gly Ile Phe Pro
85 90 95
Cys Gln Tyr Lys Gly Pro Met Phe Met Thr Ile Gly Tyr Val Ala Val
100 105 110
Asn Tyr Ile Ala Gly Ile Glu Ile Pro Glu His Glu Arg Ile Glu Leu
115 120 125
Ile Arg Tyr Ile Val Asn Thr Ala His Pro Val Asp Gly Gly Trp Gly
130 135 140
Leu His Ser Val Asp Lys Ser Thr Val Phe Gly Thr Val Leu Asn Tyr
145 150 155 160
Val Ile Leu Arg Leu Leu Gly Leu Pro Lys Asp His Pro Val Cys Ala
165 170 175
Lys Ala Arg Ser Thr Leu Leu Arg Leu Gly Gly Ala Ile Gly Ser Pro
180 185 190
His Trp Gly Lys Ile Trp Leu Ser Ala Leu Asn Leu Tyr Lys Trp Glu
195 200 205
Gly Val Asn Pro Ala Pro Pro Glu Thr Trp Leu Leu Pro Tyr Ser Leu
210 215 220
Pro Met His Pro Gly Arg Trp Trp Val His Thr Arg Gly Val Tyr Ile
225 230 235 240
Pro Val Ser Tyr Leu Ser Leu Val Lys Phe Ser Cys Pro Met Thr Pro
245 250 255
Leu Leu Glu Glu Leu Arg Asn Glu Ile Tyr Thr Lys Pro Phe Asp Lys
260 265 270
Ile Asn Phe Ser Lys Asn Arg Asn Thr Val Cys Gly Val Asp Leu Tyr
275 280 285
Tyr Pro His Ser Thr Thr Leu Asn Ile Ala Asn Ser Leu Val Val Phe
290 295 300
Tyr Glu Lys Tyr Leu Arg Asn Arg Phe Ile Tyr Ser Leu Ser Lys Lys
305 310 315 320
Lys Val Tyr Asp Leu Ile Lys Thr Glu Leu Gln Asn Thr Asp Ser Leu
325 330 335
Cys Ile Ala Pro Val Asn Gln Ala Phe Cys Ala Leu Val Thr Leu Ile
340 345 350
Glu Glu Gly Val Asp Ser Glu Ala Phe Gln Arg Leu Gln Tyr Arg Phe
355 360 365
Lys Asp Ala Leu Phe His Gly Pro Gln Gly Met Thr Ile Met Gly Thr
370 375 380
Asn Gly Val Gln Thr Trp Asp Cys Ala Phe Ala Ile Gln Tyr Phe Phe
385 390 395 400
Val Ala Gly Leu Ala Glu Arg Pro Glu Phe Tyr Asn Thr Ile Val Ser
405 410 415
Ala Tyr Lys Phe Leu Cys His Ala Gln Phe Asp Thr Glu Cys Val Pro
420 425 430

```
Gly Ser Tyr Arg Asp Lys Arg Lys Gly Ala Trp Gly Phe Ser Thr Lys
        435                 440                 445
Thr Gln Gly Tyr Thr Val Ala Asp Cys Thr Ala Glu Ala Ile Lys Ala
    450                 455                 460
Ile Ile Met Val Lys Asn Ser Pro Val Phe Ser Glu Val His His Met
465                 470                 475                 480
Ile Ser Ser Glu Arg Leu Phe Glu Gly Ile Asp Val Leu Leu Asn Leu
                485                 490                 495
Gln Asn Ile Gly Ser Phe Glu Tyr Gly Ser Phe Ala Thr Tyr Glu Lys
            500                 505                 510
Ile Lys Ala Pro Leu Ala Met Glu Thr Leu Asn Pro Ala Glu Val Phe
        515                 520                 525
Gly Asn Ile Met Val Glu Tyr Pro Tyr Val Glu Cys Thr Asp Ser Ser
    530                 535                 540
Val Leu Gly Leu Thr Tyr Phe His Lys Tyr Phe Asp Tyr Arg Lys Glu
545                 550                 555                 560
Glu Ile Arg Thr Arg Ile Arg Ile Ala Ile Glu Phe Ile Lys Lys Ser
                565                 570                 575
Gln Leu Pro Asp Gly Ser Trp Tyr Gly Ser Trp Gly Ile Cys Phe Thr
            580                 585                 590
Tyr Ala Gly Met Phe Ala Leu Glu Ala Leu His Thr Val Gly Glu Thr
        595                 600                 605
Tyr Glu Asn Ser Ser Thr Val Arg Lys Gly Cys Asp Phe Leu Val Ser
    610                 615                 620
Lys Gln Met Lys Asp Gly Gly Trp Gly Glu Ser Met Lys Ser Ser Glu
625                 630                 635                 640
Leu His Ser Tyr Val Asp Ser Glu Lys Ser Leu Val Val Gln Thr Ala
                645                 650                 655
Trp Ala Leu Ile Ala Leu Leu Phe Ala Glu Tyr Pro Asn Lys Glu Val
            660                 665                 670
Ile Asp Arg Gly Ile Asp Leu Leu Lys Asn Arg Gln Glu Glu Ser Gly
        675                 680                 685
Glu Trp Lys Phe Glu Ser Val Glu Gly Val Phe Asn His Ser Cys Ala
    690                 695                 700
Ile Glu Tyr Pro Ser Tyr Arg Phe Leu Phe Pro Ile Lys Ala Leu Gly
705                 710                 715                 720
Met Tyr Ser Arg Ala Tyr Glu Thr His Thr Leu
                725                 730
```

<210> 200
<211> 417
<212> PRT
<213> Homo sapiens

<400> 200

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

```
    130                 135                 140
Asp Ile Cys Arg Arg Met Gly Ile Gly Met Ala Glu Phe Leu Asp Lys
145                 150                 155                 160
His Val Thr Ser Glu Gln Glu Trp Asp Lys Tyr Cys His Tyr Val Ala
                165                 170                 175
Gly Leu Val Gly Ile Gly Leu Ser Arg Leu Phe Ser Ala Ser Glu Phe
            180                 185                 190
Glu Asp Pro Leu Val Gly Glu Asp Thr Glu Arg Ala Asn Ser Met Gly
        195                 200                 205
Leu Phe Leu Gln Lys Thr Asn Ile Ile Arg Asp Tyr Leu Glu Asp Gln
    210                 215                 220
Gln Gly Gly Arg Glu Phe Trp Pro Gln Glu Val Trp Ser Arg Tyr Val
225                 230                 235                 240
Lys Lys Leu Gly Asp Phe Ala Lys Pro Glu Asn Ile Asp Leu Ala Val
                245                 250                 255
Gln Cys Leu Asn Glu Leu Ile Thr Asn Ala Leu His His Ile Pro Asp
            260                 265                 270
Val Ile Thr Tyr Leu Ser Arg Leu Arg Asn Gln Ser Val Phe Asn Phe
        275                 280                 285
Cys Ala Ile Pro Gln Val Met Ala Ile Ala Thr Leu Ala Ala Cys Tyr
    290                 295                 300
Asn Asn Gln Gln Val Phe Lys Gly Ala Val Lys Ile Arg Lys Gly Gln
305                 310                 315                 320
Ala Val Thr Leu Met Met Asp Ala Thr Asn Met Pro Ala Val Lys Ala
                325                 330                 335
Ile Ile Tyr Gln Tyr Met Glu Glu Ile Tyr His Arg Ile Pro Asp Ser
            340                 345                 350
Asp Pro Ser Ser Ser Lys Thr Arg Gln Ile Ile Ser Thr Ile Arg Thr
        355                 360                 365
Gln Asn Leu Pro Asn Cys Gln Leu Ile Ser Arg Ser His Tyr Ser Pro
    370                 375                 380
Ile Tyr Leu Ser Phe Val Met Leu Leu Ala Ala Leu Ser Trp Gln Tyr
385                 390                 395                 400
Leu Thr Thr Leu Ser Gln Val Thr Glu Asp Tyr Val Gln Thr Gly Glu
                405                 410                 415
His
```

<210> 201
<211> 416
<212> PRT
<213> Mus musculus

<400> 201

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

```
Asp Ile Cys His Arg Met Gly Cys Gly Met Ala Glu Phe Val Asp Lys
145                 150                 155                 160
Asp Val Thr Ser Lys Gln Asp Trp Asp Lys Tyr Cys His Tyr Val Ala
                165                 170                 175
Gly Leu Val Gly Ile Gly Leu Ser Arg Leu Phe Ser Ala Ser Glu Phe
            180                 185                 190
Glu Asp Pro Ile Val Gly Glu Asp Ile Glu Cys Ala Asn Ser Met Gly
        195                 200                 205
Leu Phe Leu Gln Lys Thr Asn Ile Ile Arg Asp Tyr Leu Glu Asp Gln
    210                 215                 220
Gln Glu Gly Arg Lys Phe Trp Pro Gln Glu Val Trp Gly Arg Tyr Ile
225                 230                 235                 240
Lys Lys Leu Glu Asp Phe Ala Lys Pro Glu Asn Val Asp Val Ala Val
                245                 250                 255
Gln Cys Leu Asn Glu Leu Ile Thr Asn Thr Leu Gln His Ile Pro Asp
            260                 265                 270
Val Leu Thr Tyr Leu Ser Arg Leu Arg Asn Gln Ser Val Phe Asn Phe
        275                 280                 285
Cys Ala Ile Pro Gln Val Met Ala Ile Ala Thr Leu Ala Ala Cys Tyr
    290                 295                 300
Asn Asn Gln Gln Val Phe Lys Gly Val Val Lys Ile Arg Lys Gly Gln
305                 310                 315                 320
Ala Val Thr Leu Met Met Asp Ala Thr Asn Met Pro Ala Val Lys Ala
                325                 330                 335
Ile Ile Tyr Gln Tyr Ile Glu Glu Ile Tyr His Arg Ile Pro Asn Ser
            340                 345                 350
Asp Pro Ser Ser Ser Lys Thr Lys Gln Val Ile Ser Lys Ile Arg Thr
        355                 360                 365
Gln Asn Leu Pro Asn Cys Gln Leu Ile Ser Arg Ser His Tyr Ser Pro
    370                 375                 380
Ile Tyr Leu Ser Phe Ile Met Leu Leu Ala Ala Leu Ser Trp Gln Tyr
385                 390                 395                 400
Leu Ser Thr Leu Ser Gln Val Thr Glu Asp Tyr Val Gln Arg Glu His
                405                 410                 415


<210> 202
<211> 416
<212> PRT
<213> Rattus norvegicus

<400> 202
Met Glu Phe Val Lys Cys Leu Gly His Pro Glu Glu Phe Tyr Asn Leu
 1                5                  10                  15
Leu Arg Phe Arg Met Gly Gly Arg Arg Asn Phe Ile Pro Lys Met Asp
            20                  25                  30
Arg Asn Ser Leu Ser Asn Ser Leu Lys Thr Cys Tyr Lys Tyr Leu Asp
        35                  40                  45
Gln Thr Ser Arg Ser Phe Ala Ala Val Ile Gln Ala Leu Asp Gly Asp
    50                  55                  60
Ile Arg His Ala Val Cys Val Phe Tyr Leu Ile Leu Arg Ala Met Asp
65                  70                  75                  80
Thr Val Glu Asp Asp Met Ala Ile Ser Val Glu Lys Lys Ile Pro Leu
                85                  90                  95
Leu Arg Asn Phe His Thr Phe Leu Tyr Glu Pro Glu Trp Arg Phe Thr
            100                 105                 110
Glu Ser Lys Glu Lys His Arg Val Val Leu Glu Asp Phe Pro Thr Ile
        115                 120                 125
Ser Leu Glu Phe Arg Asn Leu Ala Glu Lys Tyr Gln Thr Val Ile Ala
    130                 135                 140
Asp Ile Cys His Arg Met Gly Cys Gly Met Ala Glu Phe Leu Asn Lys
145                 150                 155                 160
Asp Val Thr Ser Lys Gln Asp Trp Asp Lys Tyr Cys His Tyr Val Ala
```

```
                165                 170                 175
Gly Leu Val Gly Ile Gly Leu Ser Arg Leu Phe Ser Ala Ser Glu Phe
            180                 185                 190
Glu Asp Pro Ile Val Gly Glu Asp Thr Glu Cys Ala Asn Ser Met Gly
        195                 200                 205
Leu Phe Leu Gln Lys Thr Asn Ile Ile Arg Asp Tyr Leu Glu Asp Gln
    210                 215                 220
Gln Glu Gly Arg Gln Phe Trp Pro Gln Glu Val Trp Gly Lys Tyr Val
225                 230                 235                 240
Lys Lys Leu Glu Asp Phe Val Lys Pro Glu Asn Val Asp Val Ala Val
                245                 250                 255
Lys Cys Leu Asn Glu Leu Ile Thr Asn Ala Leu Gln His Ile Pro Asp
            260                 265                 270
Val Ile Thr Tyr Leu Ser Arg Leu Arg Asn Gln Ser Val Phe Asn Phe
        275                 280                 285
Cys Ala Ile Pro Gln Val Met Ala Ile Ala Thr Leu Ala Ala Cys Tyr
    290                 295                 300
Asn Asn His Gln Val Phe Lys Gly Val Val Lys Ile Arg Lys Gly Gln
305                 310                 315                 320
Ala Val Thr Leu Met Met Asp Ala Thr Asn Met Pro Ala Val Lys Ala
                325                 330                 335
Ile Ile Tyr Gln Tyr Ile Glu Glu Ile Tyr His Arg Val Pro Asn Ser
            340                 345                 350
Asp Pro Ser Ala Ser Lys Ala Lys Gln Leu Ile Ser Asn Ile Arg Thr
        355                 360                 365
Gln Ser Leu Pro Asn Cys Gln Leu Ile Ser Arg Ser His Tyr Ser Pro
    370                 375                 380
Ile Tyr Leu Ser Phe Ile Met Leu Leu Ala Ala Leu Ser Trp Gln Tyr
385                 390                 395                 400
Leu Ser Thr Leu Ser Gln Val Thr Glu Asp Tyr Val Gln Arg Glu His
                405                 410                 415
```

<210> 203
<211> 396
<212> PRT
<213> Aspergilus nidulans

<400> 203

```
Met Thr Ser Asp Ser His Phe His Pro Pro His Ala Ile Pro Pro Arg
 1               5                  10                  15
Ile Ser Ser Asn Arg Met Ser Gly Ala Ser Thr Arg Asp Lys Ala Ala
            20                  25                  30
Leu Met Gly Asn Phe Glu Lys Asp Trp Leu Ser Lys Gly Asp Lys Leu
        35                  40                  45
Gln Thr Asn Thr Asp Leu Ser Lys Arg His Thr Arg Asn Gln Ser Ser
    50                  55                  60
Leu Asp Gly Thr Lys Tyr Lys Asp Gly Lys Trp Ser Gln Glu Asn Glu
65                  70                  75                  80
Glu Val Ile Met Gly Pro Tyr Asp Tyr Met Leu Gln His Pro Gly Lys
                85                  90                  95
Asp Leu Arg Arg Gln Met Ile Asn Ala Phe Asn Val Trp Leu Lys Val
            100                 105                 110
Pro Ser Glu Ser Leu Ala Ile Ile Thr Lys Val Val Ala Met Leu His
        115                 120                 125
Thr Ala Ser Leu Leu Ile Asp Asp Val Glu Asp Asn Ser Leu Leu Arg
    130                 135                 140
Arg Gly Ile Pro Val Ala His Ser Ile Tyr Gly Thr Ala Gln Thr Ile
145                 150                 155                 160
Asn Ser Ala Asn Tyr Val Tyr Phe Leu Ala Leu Gln Glu Val Gln Lys
                165                 170                 175
Leu Lys Ser Pro Ala Ala Ile Asp Ile Tyr Val Gln Glu Leu Leu Asn
            180                 185                 190
```

```
Leu His Arg Gly Gln Gly Met Asp Leu Phe Trp Arg Asp Thr Leu Thr
        195                 200                 205
Cys Pro Ser Glu Asp Glu Tyr Leu Glu Met Val Gly Asn Lys Thr Gly
    210                 215                 220
Gly Leu Phe Arg Leu Ala Val Lys Leu Met Gln Ala Glu Ser Ser Thr
225                 230                 235                 240
Gly Lys Asp Cys Val Ala Leu Val Asn Val Leu Gly Leu Val Phe Gln
                245                 250                 255
Ile Cys Asp Asp Tyr Leu Asn Leu Ser Asp Thr Thr Tyr Thr Gln Asn
            260                 265                 270
Lys Gly Leu Cys Glu Asp Leu Thr Glu Gly Lys Phe Ser Phe Pro Ile
        275                 280                 285
Ile His Ser Ile Arg Ser Asn Pro Gly Asn His Gln Leu Ile Asn Ile
    290                 295                 300
Leu Arg Gln Arg Thr Lys Asp Glu Glu Val Lys Arg Tyr Ala Leu Gln
305                 310                 315                 320
Tyr Met Glu Ser Thr Gly Ser Phe Lys His Thr Gln Asp Val Val Arg
                325                 330                 335
Gln Leu Arg Ala Arg Ala Leu Gln Leu Ile Glu Glu Ile Glu Asn Ser
            340                 345                 350
Glu Asn Gly Glu Gln Pro Glu Glu His Asn Asp Gly Thr Met Val Arg
        355                 360                 365
Ala Ile Leu Asp Lys Ile Thr Glu Ser Thr Leu Ala Asp Thr Asn Thr
    370                 375                 380
Thr Thr Arg Asp Ile Asn Gly Asn Cys Ala Thr Arg
385                 390                 395
```