METHODS FOR MAKING HIGH INTENSITY SWEETENERS

BACKGROUND

Field

[0001] The present disclosure relates to methods, systems and compositions for producing sweet tasting compounds, as well as compositions comprising the sweet tasting compounds.

Background Description

[0002] The taste system provides sensory information about the chemical composition of the external world. Taste transduction is one of the most sophisticated forms of chemical-triggered sensation in animals. Signaling of taste is found throughout the animal kingdom, from simple metazoans to the most complex of vertebrates. Mammals are believed to have five basic taste modalities: sweet, bitter, sour, salty, and umami (the taste of monosodium glutamate, a.k.a. savory taste).

[0003] For centuries, various natural and unnatural compositions and/or compounds have been added to ingestible compositions, including foods and beverages, and/or orally administered medicinal compositions to improve their taste. Although it has long been known that there are only a few basic types of "tastes," the biological and biochemical basis of taste perception was poorly understood, and most taste improving or taste modifying agents have been discovered largely by simple trial and error processes.

[0004] With respect to the sweet taste, diabetes, and cardiovascular disease are health concerns on the rise globally, but are growing at alarming rates in the United States. Sugar and calories are key components that can be limited to render a positive nutritional effect on health. High-intensity sweeteners can provide the sweetness of sugar, with various taste qualities. Because they are many times sweeter than sugar, much less of the sweetener is required to replace the sugar.

[0005] High-intensity sweeteners have a wide range of chemically distinct structures and hence possess varying properties, such as, without limitation, odor, flavor, mouthfeel, and

aftertaste. These properties, particularly flavor and aftertaste, are well known to vary over the time of tasting, such that each temporal profile is sweetener-specific.

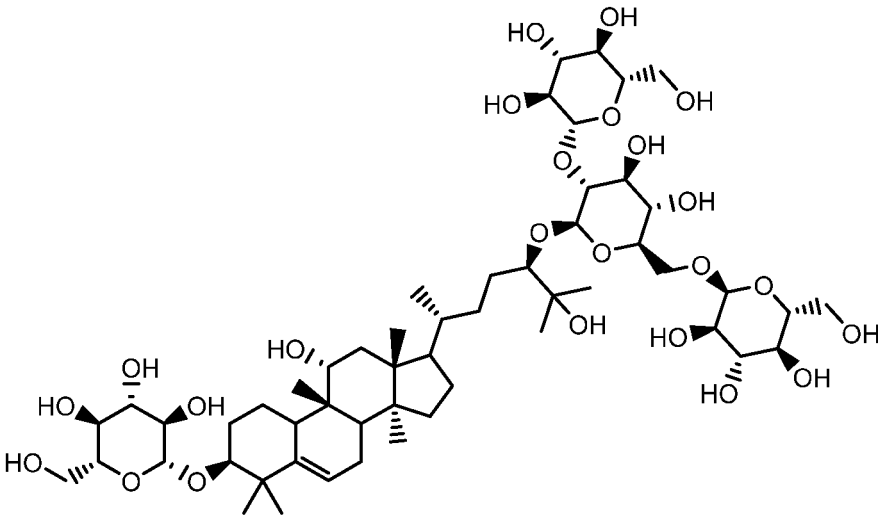
[0006] There has been significant recent progress in identifying useful natural flavoring agents, such as for example sweeteners such as sucrose, fructose, glucose, erythritol, isomalt, lactitol, mannitol, sorbitol, xylitol, certain known natural terpenoids, flavonoids, or protein sweeteners. *See, e.g,* Kinghom, et al., "Noncariogenic Intense Natural Sweeteners," Med. Res. Rev. 18 (5) 347-360 (1998) (discussing discovered natural materials that are much more intensely sweet than common natural sweeteners such as sucrose, fructose, and the like.) Similarly, there has been recent progress in identifying and commercializing new artificial sweeteners, such as aspartame, saccharin, acesulfame-K, cyclamate, sucralose, and the like. *See, e.g.,* Ager, et al., Angew. Chem. Int. Ed. 37, 1802-1817 (1998). The entire contents of the references identified above are hereby incorporated herein by reference in their entirety.

[0007] Sweeteners such as saccharin and 6-methyl-1,2,3-oxathiazin-4(3H)-one-2,2-dioxide potassium salt (acesulfame potassium) are commonly characterized as having bitter and/or metallic aftertastes. Products prepared with 2,4-dihydroxybenzoic acid are claimed to display reduced undesirable aftertastes associated with sweeteners, and do so at concentrations below those concentrations at which their own tastes are perceptible. Also, high intensity sweeteners such as sucralose and aspartame are reported to have sweetness delivery problems, i.e., delayed onset and lingering of sweetness. *See* S. G. Wiet, et al., J. Food Sci., 58(3):599-602, 666 (1993).

[0008] There is a need for new sweetening compounds, sweet taste enhancers, and compositions containing such compounds and enhancers, having improved taste and delivery characteristics. In addition, there is a need for foods containing new sweetening compounds and/or sweet taste enhancers with such desirable characteristics.

SUMMARY

[0009] Provided herein include a method of producing Compound 1 having the structure of:



(1).

[0010] In some embodiments, the method comprises contacting mogroside IIIE with a first enzyme capable of catalyzing production of Compound 1 from mogroside IIIE. In some embodiments, contacting mogroside IIIE with the first enzyme comprises contacting mogroside IIIE with a recombinant host cell that comprises a first gene encoding the first enzyme. The first gene can be, for example, heterologous to the recombinant host cell.

[0011] In some embodiments, the mogroside IIIE contacts with the first enzyme in a recombinant host cell that comprises a first polynucleotide encoding the first enzyme. The mogroside IIIE can be, for example, provided to the recombinant cell, present in the recombinant host cell, produced by the recombinant host cell, or any combination thereof. In some embodiments, the method comprises cultivating the recombinant host cell in a culture medium under conditions in which the first enzyme is expressed. The first enzyme can be, for example, one or more of UDP glycosyltransferases, cyclomaltodextrin glucanotransferases (CGTases), glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0012] In some embodiments, the first enzyme is a CGTase. In some embodiments, the CGTase comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of any one of SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the CGTase comprises the amino acid sequence of any one of SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the CGTase consists of the amino acid sequence of SEQ ID NOs: 1, 3, 78-101, 148, and 154.

[0013] In some embodiments, the first enzyme is a dextransucrase. For example, the dextransucrase can comprise an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of the sequences set forth in SEQ ID NOs: 2, 103, 106-110, 156, 159-162, and 896. In some embodiments, the dextransucrase is encoded by a nucleic acid sequence having at least 70% sequence identity to any one of SEQ ID NOs: 104, 105, 157, 158, and 895.

[0014] In some embodiments, the first enzyme is a transglucosidase. For example, the transglucosidase can comprise an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of any one of SEQ ID NOs: 163-291 and 723. In some embodiments, the transglucosidase comprises an amino acid sequence of any one of SEQ ID NOs: 163-291 and 723.

[0015] In some embodiments, the first enzyme is a beta-glucosidase. For example, the beta-glucosidase can comprise an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence set forth in any one of SEQ ID NOs: 102, 292, 354-374, and 678-741.

[0016] In some embodiments, the method comprises contacting mogroside IIA with an enzyme capable of catalyzing a production of mogroside IIIE from mogroside IIA. In some embodiments, contacting mogroside IIA with the enzyme comprises contacting the mogroside IIA with the recombinant host cell to produce mogroside IIIE, and wherein the recombinant host cell comprises a gene encoding the enzyme capable of catalyzing production of mogroside IIIE from mogroside IIA. The mogroside IIA can be, for example, provided to the recombinant host cell, produced by the recombinant host cell, present in the recombinant host cell, or any combination thereof. The enzyme capable of catalyzing the production of mogroside IIIE from mogroside IIA can be, for example, one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0017] In some embodiments, the second enzyme is a uridine diphosphate-glucosyl transferase (UGT). For example, the UGT can be UGT73C3 (SEQ ID NO: 4), UGT73C6 (SEQ ID NO: 5), UGT 85C2 (SEQ ID NO: 6), UGT73C5 (SEQ ID NO: 7), UGT73E1 (SEQ ID NO:

8), UGT98 (SEQ ID NO: 9 or 407) , UGT1576 (SEQ ID NO:15), UGT SK98 (SEQ ID NO:16), UGT430 (SEQ ID NO:17), UGT1697 (SEQ ID NO:18), UGT11789 (SEQ ID NO:19). In some embodiments, the UGT comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 4-9, 15-19, 125, 126, 128, 129, 293-307, 407, 409, 411, 413, 439, 441 and 444. In some embodiments, the UGT is encoded by a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of the sequences set forth in UGT1495 (SEQ ID NO:10), UGT1817 (SEQ ID NO: 11), UGT5914 (SEQ ID NO: 12), UGT8468 (SEQ ID NO:13), UGT10391 (SEQ ID NO:14), SEQ ID NOs: 116-124, 127, 130, 408, 410, 412, 414, 440, 442, 443, and 445.

[0018] In some embodiments, the method comprises contacting mogrol with one or more enzyme capble of catalyzing a production of mogroside IIIE and/or IIE from mogrol. In some embodiments, contact mogrol with the one or more enzymes comprises contacting mogrol with the recombinant host cell to produce mogroside IIIE and/or mogroside IIE, wherein the recombinant host cell comprises one or more genes encoding one or more enzymes capable of catalyzing production of mogroside IIIE and/or mogroside IIE from mogrol. The mogrol can be, for example, provided to the recombinant host cell, produced by the recombinant host cell, present in the recombinant host cell, or any combination thereof.

[0019] In some embodiments, at least one of the one or more enzymes capable of catalyzing production of mogroside IIE and/or mogroside IIIE from mogrol comprise a sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of the sequences set forth in or is encoded by a sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of the sequences in SEQ ID NOs: 315, 316, 420, 422, 424, 426, 430, 431, 446, 871, 845-949, and 951-1012. In some embodiments, the one or more enzymes capable of catalyzing production of mogroside IIIE from mogrol comprises one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, at least one of the one or more enzymes is a uridine

diphosphate-glucosyltransferase (UGT). For example, the UGT can be UGT73C3, UGT73C6, 85C2, UGT73C5, UGT73E1, UGT98, UGT1495, UGT1817, UGT5914, UGT8468, UGT10391, UGT1576, UGT SK98, UGT430, UGT1697, or UGT11789, or comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 4-9, 15-19, 125, 126, 128, 129, 293-307, 405, 406, 407, 409, 411, 413, 439, 441, and 444.

[0020] In some embodiments, the method comprises contacting a mogroside compound with one or more enzymes capable of catalyzing a production of mogroside IIIE from a mogroside compound to produce mogroside IIIE, wherein the mogroside compound is one or more of mogroside IA1, mogroside IE1, mogroside IIA1, mogroside IIE, mogroside IIA, mogroside IIIA1, mogroside IIIA2, mogroside III, mogroside IV, mogroside IVA, mogroside V, and siamenoside. In some embodiments, contacting the mogroside compound with the one or more enzymes capable of catalyzing the product of mogroside IIIE from the mogroside compound comprises contacting the mogroside compound with the recombinant host cell, wherein the recombinant host cell comprises one or more genes encoding the one or more enzymes capable of catalyzing production of mogroside IIIE from the mogroside compound. The mogroside compound can be, for example, provided to the recombinant host cell, produced by the recombinant host cell, present in the recombinant host cell, and any combination thereof. In some embodiments, the one or more enzymes capable of catalyzing production of mogroside IIIE from the mogroside compound comprises one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the mogroside compound is mogroside IIE.

[0021] In some embodiments, the one or more enzymes capable of catalyzing production of Mogroside IIIE from the mogroside compound comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 1, 3, 78-101, 106-109, 147, 154, 163-303, 405, 411, 354-405, 447-723, 770, 776, and 782. In some embodiments, the mogroside compound is morgroside IIA or mogroside IIE. In some embodiments, contacting

with one or more enzymes produces one or more of mogroside IIIA, mogroside IVE and mogroside V.

[0022] In some embodiments, the one or more enzymes comprises an amino acid having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 304, 405, 411, 872, 874, 978, 880, 882, 884, 886, 888, 890, 892, 894 and 896. In some embodiments, the one or more enzymes is encoded by a sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 305, 406, 412, 873, 875, 877, 879, 881, 883, 885, 887, 889, 891, 893 and 895.

[0023] In some embodiments, the method comprises contacting mogroside IA1 with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding UGT98 or UGT SK98 enzyme. In some embodiments, the UGT98 or UGT SK98 enzyme comprises an amino acid sequence having at least 70% sequence identity to SEQ ID NO: 9, 407, 16 or 306. In some embodiments, the UGT98 is encoded by a sequence set forth in SEQ ID NO: 307. In some embodiments, the contacting results in production of mogroside IIA in the cell.

[0024] In some embodiments, the method comprises contacting 11-hydroxy-24,25 epoxy cucurbitadienol with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding an epoxide hydrolase. In some embodiments, the 11-hydroxy-24,25 epoxy cucurbitadienol is provided to the recombinant host cell, present in the recombinant host cell, produced by the recombinant host cell, and any combination thereof.

[0025] In some embodiments, the method comprises contacting 11-hydroxy cucurbitadienol with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding a cytochrome P450 or an epoxide hydrolase. The 11-hydroxy cucurbitadienol can be provided to, produced by, and/or present in, the recombinant host cell.

[0026] In some embodiments, the method comprises contacting 3,24,25-trihydroxy cucurbitadienol with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding a cytochrome P450. In some embodiments, the 3,24,25-trihydroxy cucurbitadienol is provided to the recombinant host cell, present in the recombinant host cell,produced by the recombinant host cell, or any combination thereof. In some embodiments, the contacting results in production of mogrol in the recombinant host cell. In some

embodiments, the cytochrome P450 comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 20, 49, 308, 315, 430, 872, 874, 876, 878, 880, 882, 884, 886, 888, 889, and 892; or the cytochrome P450 is encoded by a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, 890, and 891.

[0027] In some embodiments, the epoxide hydrolase comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 21-30 and 309-314; or the epoxide hydrolase is encoded by a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID Nos: 114 and 115.

[0028] In some embodiments, the method comprises contacting cucurbitadienol with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding cytochrome P450. In some embodiments, the contacting results in production of 11-hydroxy cucurbitadienol. In some embodiments, the cucurbitadienol is provided to, produced by, and/or present in the recombinant host cell. In some embodiments, the cytochrome P450 is encoded by a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, and 892. In some embodiments, the cytochrome P450 comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NOs: 20, 31, 49, 308, 315, 430, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, and 891. In some embodiments, the method further comprises contacting one or more of 2,3-oxidosqualene, dioxidosqualene and diepoxysqualene with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding a polypeptide having cucurbitadienol synthase activity.

[0029] In some embodiments, the method comprises contacting a mogroside intermediate with the recombinant host cell, wherein the recombinant host cell comprises a gene

encoding a polypeptide having cucurbitadienol synthase activity. In some embodiments, the polypeptide having cucurbitadienol synthase activity is a fusion protein comprising one or more fusion domain fused to a cucurbitadienol synthase. In some embodiments, the fusion protein comprises a fusion domain fused to the N-terminus, the C-terminus, or both of the cucurbitadienol synthase. In some embodiments, the fusion domain is about 3 to about 1000 amino acids long. In some embodiments, the fusion domain is about 5 to about 50 amino acids long. In some embodiments, the fusion domain is a substantial portion or the entire sequence of a functional protein.

[0030] In some embodiments, the fusion polypeptide having cucurbitadienol synthase activity comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327-333, 417, 420, 422, 424, 426, 446, 902, 904 or 906. In some embodiments, the cucurbitadienol synthase is encoded by a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903 and 905.

[0031] In some embodiments, the contacting results in production of cucurbitadienol. In some embodiments, the 2,3-oxidosqualene and diepoxysqualene is provided to, produced by, and/or present in the recombinant host cell. In some embodiments, one or more of the 2,3-oxidosqualene and diepoxysqualene is produced by an enzyme comprising a sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 898 or 900. In some embodiments, the production of one or more of 2,3-oxidosqualene, diepoxysqualene, and diepoxysqualene involves an enzyme encoded by a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 897 or 899. In some embodiments, the cucurbitadienol synthase comprises an amino acid sequence having at least 70% sequence identity to any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327, 329-333, 420, 422, 424, 426, 446, 902, 904, and 906. In some embodiments, the polypeptide comprising cucurbitadienol synthase activity is encoded by a gene comprising a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%,

or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903 and 905.

[0032] In some embodiments, the 11-hydroxy cucurbitadienol is provided to, produced by, and/or present in the recombinant host cell. In some embodiments, the 11-hydroxy cucurbitadienol is expressed in a cell (for example the recombinant host cell) comprising a gene encoding CYP87D18 and/or SgCPR protein. In some embodiments, the CYP87D18 or SgCPR protein comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 872 or 874. In some embodiments, the CYP87D18 or SgCPR protein is encoded by a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 871 or 873.

[0033] In some embodiments, the method comprises contacting squalene with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding a squalene epoxidase. In some embodiments, the contacting results in production of 2,3-oxidosqualene. In some embodiments, the squalene is provided to, produced by, and/or present in the recombinant host cell. In some embodiments, the squalene epoxidase comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 50-56, 60, 61, 334 or 335. In some embodiments, squalene epoxidase is encoded by a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 335.

[0034] In some embodiments, the method comprises contacting farnesyl pyrophosphate with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding a squalene synthase. In some embodiments, the contacting results in production of squalene. In some embodiments, the farnesyl pyrophosphate is provided to, produced by, and/or present in the recombinant host cell. In some embodiments, the squalene synthase comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 69 and 336. In some embodiments, the squalene synthase is encoded by a sequence comprising a

nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 337.

[0035] In some embodiments, the method comprises contacting geranyl-PP with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding farnesyl-PP synthase. In some embodiments, the contacting results in production of farnesyl-PP. In some embodiments, the geranyl-PP is provided to, produced by, and/or present in the recombinant host cell. In some embodiments, the farnesyl-PP synthase comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 338. In some embodiments, the farnesyl-PP synthase is encoded by a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 339.

[0036] In some embodiments, one or more of the genes encoding (1) the first enzyme capable of catalyzing the production of Compound I from mogroside IIIE, (2) the enzyme capable of catalyzing the production of mogroside IIIE from mogroside IIA, (3) the epoxide hydrolase, (4) the cytochrome P450, (5) the polypeptide having cucurbitadienol synthase activity, (6) squalene epoxidase, (7) farnesyl-PP synthase is operably linked to a heterologous promoter. In some embodiments, the heterologous promoter is a CMV, EF1a, SV40, PGK1, human beta actin, CAG, GAL1, GAL10, TEF1, GDS, ADH1, CaMV35S, Ubi, T7, T7lac, Sp6, araBAD, trp, lac, Ptac, pL promoter, or a combination thereof. In some embodiments, the promoter is an inducible, repressible, or constitutive promoter. In some embodiments, production of one or more of pyruvate, acetyl-CoA, citrate, and TCA cycle intermediates have been upregulated in the recombinant host cell. In some embodiments, cytosolic localization has been upregulated in the recombinant host cell. In some embodiments, one or more of the first, second third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth gene comprises at least one sequence encoding a 2A self-cleaving peptide.

[0037] In some embodiments, the recombinant host cell is a plant, bivalve, fish, fungus, bacteria, or mammalian cell. In some embodiments, the plant is selected from the group consisting of Siraitia, Momordica, Gynostemma, Cucurbita, Cucumis, Arabidopsis, Artemisia, Stevia, Panax, Withania, Euphorbia, Medicago, Chlorophytum, Eleutherococcus, Aralia, Morus,

Medicago, Betula, Astragalus, Jatropha, Camellia, Hypholoma, Aspergillus, Solanum, Huperzia, Pseudostellaria, Corchorus, Hedera, Marchantia, and Morus. In some embodiments, the fungus is selected from the group consisting of *Trichophyton, Sanghuangporus, Taiwanofungus, Moniliophthora, Marssonina, Diplodia, Lentinula, Xanthophyllomyces, Pochonia, Colletotrichum, Diaporthe, Histoplasma, Coccidioides, Histoplasma, Sanghuangporus, Aureobasidium, Pochonia, Penicillium, Sporothrix, Metarhizium, Aspergillus, Yarrowia,* and *Lipomyces*. In some embodiments, the fungus is *Aspergillus nidulans, Yarrowia lipolytica*, or *Rhodosporin toruloides*. In some embodiments, the recombinant host cell is a yeast cell. In some embodiments, the yeast is selected from the group consisting of Candida, Sacccharaomyces, Saccharomycotina, Taphrinomycotina, Schizosaccharomycetes, Komagataella, Basidiomycota, Agaricomycotina, Tremellomycetes, Pucciniomycotina, Aureobasidium, Coniochaeta, Rhodosporidium, and Microboryomycetes. In some embodiments, the bacteria is selected from the group consisting of Frankia, Actinobacteria, Streptomyces, and Enterococcus. In some embodiments, the recombinant host cell is a *Saccharomyces cerevisiae* cell or a *Yarrowia lipolytica* cell. In some embodiments, one or more of the first, second third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth gene is a codon optimized gene for expression in a bacterial, mammalian, plant, fungal and/or insect cell. In some embodiments, one or more of the first, second third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth gene comprises a functional mutation to increased activity of the encoded enzyme. In some embodiments, cultivating the recombinant host cell comprises monitoring the cultivating for pH, dissolved oxygen level, nitrogen level, or a combination thereof of the cultivating conditions.

[0038] In some embodiments, the method comprises isolating Compound 1. In some embodiments, isolating Compound 1 comprises lysing the recombinant host cell, and/or isolating Compound 1 from the culture medium. In some embodiments, the method comprises purifying Compound 1. In some embodiments, purifying Compound 1 comprises HPLC, solid phase extraction or a combination thereof. In some embodiments, the purifying comprises harvesting the recombinant host cells; saving the supernatant; and lysing the recombinant host cells. In some embodiments, the lysing comprises subjecting the cells to shear force or detergent washes thereby obtaining a lysate. The shear force can be, for example, from a sonication method, french pressurized cells, or beads. In some embodiments, the lysate is subjected to filtering and

purification steps. In some embodiments, the lysate is filtered and purified by solid phase extraction.

[0039] In some embodiments, the method further comprises contacting a first mogroside with one or more hydrolase to produce Mogroside IIIE before contacting the Mogroside IIIE with the first enzyme. The hydrolase can be, for example, a β-glucan hydrolase. In some embodiments, the hydrolase is EXG1 or EXG2. In some embodiments, the hydrolases comprises an amino acid sequence comprising having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs. 292, 366-368, 372, 376-398, 447-520, 524-845, 1013, 1014, and 1023.

[0040] In some embodiments, the first mogroside is mogroside IV, mogroside V, mogroside VI, or a combination thereof. In some embodiments, the first mogroside is mogroside V, siamenoside I, mogroside IVE, mogroside VI, mogroside IVA, or a combination thereof.

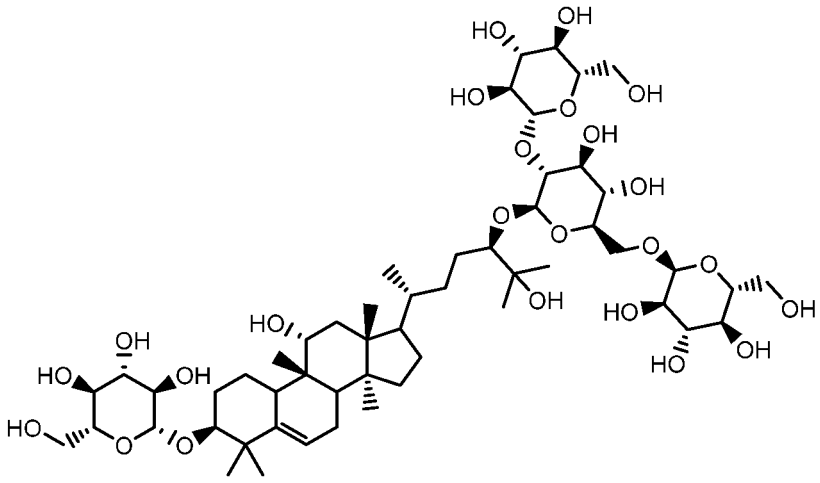
[0041] In some embodiments, contacting the first mogroside with the one or more hyrdrolase comprises contacting the first mogroside with a host cell that comprises a gene encoding the hydrolase. In some embodiments, the first mogroside contacts the one or more hyrdrolase in a host cell that comprises a gene encoding the hydrolase and a gene encoding the enzyme capable of catalyzing production of Compound 1. The host cell can be the recombinant host cell that comprises the first gene encoding the first enzyme capable of catalyzing the production of Compund 1 from mogroside IIIE. In some embodiments, the host cell is not the recombinant host cell comprising the first gene encoding the first enzyme capable of catalyzing the production of Compund 1 from mogroside IIIE. In some embodiments, the first mogroside is provided to, produced by, and/or present in the host cell. The gene encoding the hydrolase can be heterologouis or homologous to the host cell. In some embodiments, the gene encoding the hydrolase is expressed at a normal level in the host cell. In some embodiments, the gene encoding the hydrolase is overexpressed in the host cell.

[0042] In some embodiments, the recombinant host cell comprises an oxidosqualene cyclase such as a cycloartenol synthase or a beta-amyrin synthase or a nucleic acid sequence encoding an oxidosqualene cyclase such as a cycloartenol synthase or a beta-amyrin synthase, and wherein the oxidosqualene cyclase, cycloartenol synthase, or beta-amyrin synthase are modified to produce cucurbitadienol or epoxycucurbitadienol. The oxidosqualene cyclase can,

for example, comprise or consists of a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 341, 343 and 346-347.

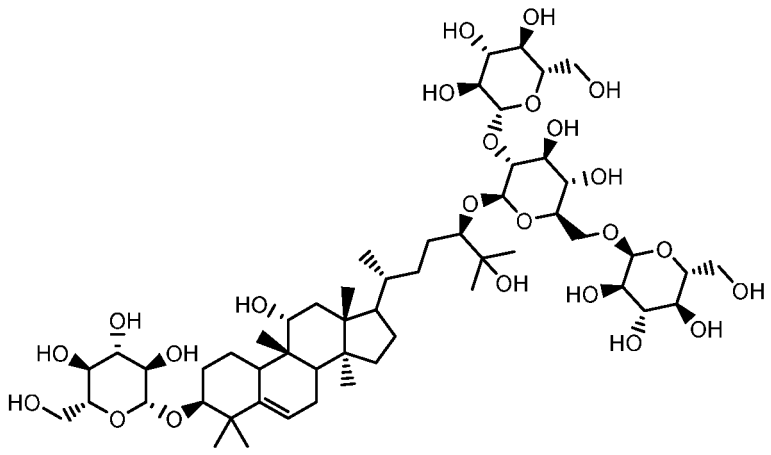
[0043] In some embodiments, the recombinant host cell comprises cytochrome P450 reductase or a gene encoding cytochrome P450 reductase. In some embodiments, the cytochrome P450 reductase comprises, or consists of, a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 318.

[0044] Disclosed herein include a compound having the structure of Compound 1,



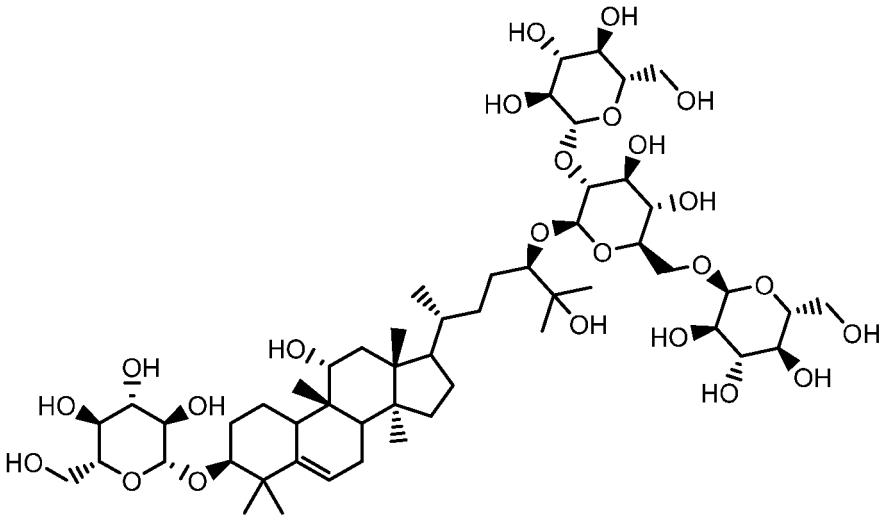
(1), wherein the compound is produced by any of the methods disclosed herein.

[0045] Disclosed herein include a cell lysate comprising Compound 1 having the structure:



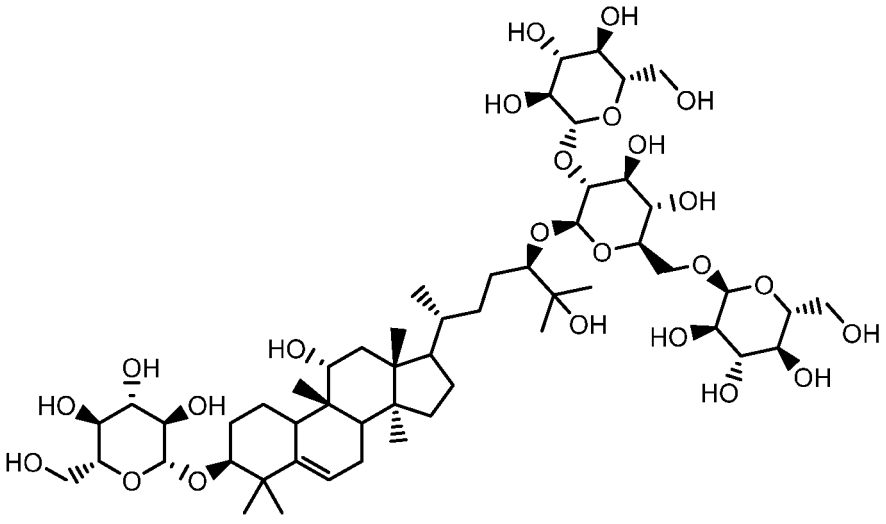
(1).

[0046] Also disclosed herein include a recombinant cell comprising: Compound 1 having the structure:



(1), and a gene encoding an enzyme capable of catalyzing production of Compound 1 from mogroside IIIE. In some embodiments, the gene is a heterologous gene to the recombinant cell.

[0047] Disclosed herein include a recombinant cell comprising a first gene encoding a first enzyme capable of catalyzing production of Compound 1 having the structure:



(1) from mogroside IIIE. The first enzyme can be, for example, one or more of UDP glycosyltransferases, cyclomaltodextrin glucanotransferases (CGTases), glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the first enzyme is a CGTase. For example, the CGTase can comprise an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of

these numbers, sequence identity to the sequence of any one of SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the CGTase comprises, or consists of, the amino acid sequence of any one of SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the first enzyme is a dextransucrase. For example, the dextransucrase can comprise an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of the sequences set forth in SEQ ID NOs: 2, 103, 106-110, 156 and 896. In some embodiments, the dextransucrase is encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 104, 105, 157, 158, and 895. In some embodiments, the first enzyme is a transglucosidase. The transglucosidase can, for example, comprise an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of any one of SEQ ID NOs: 163-291 and 723. In some embodiments, the transglucosidase comprises, or consists of, an amino acid sequence of any one of SEQ ID NOs: 3, 95-102 and 163-291 and 723. In some embodiments, the first enzyme is a beta-glucosidase. In some embodiments, the beta-glucosidase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 102, 292, 354-376, and 678-741.

[0048] In some embodiments, the cell further comprises a second gene encoding a uridine diphosphate-glucosyl transferase (UGT). In some embodiments, the UGT comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 4-9, 15-19, 125, 126, 128, 129, 293-307, 407, 409, 411, 413, 439, 441, and 444. In some embodiments, the UGT is encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 116-124, 127, 130, 408, 410, 412, 414, 440, 442, 443, and 445. In some embodiments, the UGT is encoded by a sequence set forth in UGT1495 (SEQ ID NO: 10), UGT1817 (SEQ ID NO: 11), UGT5914 (SEQ ID NO: 12), UGT8468 (SEQ ID NO: 13), or UGT10391 (SEQ ID NO: 14). In some embodiments, the cell further comprises a third gene encoding UGT98 or UGT SK98. In some embodiments, the UGT98 or UGT SK98 comprises an amino acid sequence

having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 9, 407, 16 or 306. In some embodiments, the UGT98 is encoded by a nucleic acid sequence set forth in SEQ ID NO: 307.

[0049] In some embodiments, the cell comprises a fourth gene encoding an epoxide hydrolase. In some embodiments, the epoxide hydrolase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 21-30 and 309-314; or is encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 114 and 115.

[0050] In some embodiments, the cell comprises a fifth sequence encoding P450. In some embodiments, the P450 comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 20, 49, 308, 315, 430, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, and 891; or is encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, and 892. In some embodiments, the P450 is encoded by a gene comprising or consisting of a sequence set forth in any one of SEQ ID NOs: 31-48, 316 and 318.

[0051] In some embodiments, the cell comprises a sixth sequence encoding a polypeptide having cucurbitadienol synthase activity. In some embodiments, the polypeptide having cucurbitadienol synthase activity is a fusion protein. In some embodiments, the fusion protein comprises one or more fusion domains fused to the N-terminus, the C-terminus, or both of a cucurbitadienol synthase. The length of the fusion domain can vary, for example from about 3 to about 1000 amino acids long, or about 5 to about 50 amino acids long. In some embodiments, the fusion domain is a substantial portion or the entire sequence of a functional protein. In some embodiments, the polypeptide having cucurbitadienol synthase activity comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327-333, 417, 420, 422, 424, 426, 446, 902, 904, 906,

906, 851, 854, 856, 1024, 859, 862, 865, 867, 915, 920, 924, 928, 932, 936, 940, 944, 948, 952, 956, 959, 964, 967, 971, 975, 979, 983, 987, 991, 995, 999, 1003, 1007, and 1011. In some embodiments, the polypeptide having cucurbitadienol synthase activity is encoded by a nucleic acid sequence having at least 70% sequence identity to any one of SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903 and 905.

[0052] In some embodiments, the cell further comprises a seventh gene encoding a squalene epoxidase. In some embodiments, the squalene epoxidase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 50-56, 60, 61, 334, and 335. In some embodiments, the squalene epoxidase is encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 335.

[0053] In some embodiments, the cell comprises an eighth gene encoding a squalene synthase. In some embodiments, the squalene synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 69 or 336. In some embodiments, the squalene synthase is encoded by a sequence comprising, or consisting of, a nucleic acid sequence set forth in SEQ ID NO: 337.

[0054] In some embodiments, the cell further comprises a ninth gene encoding a farnesyl-PP synthase. In some embodiments, the farnesyl-PP synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 338. In some embodiments, the farnesyl-PP synthase is encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 339.

[0055] In some embodiments, the cell is a mammalian, plant, bacterial, fungal, or insect cell. For example, the cell can be a yeast cell. In some embodiments, the yeast is selected from Candida, Saccharaomyces, Saccharomycotina, Taphrinomycotina, Schizosaccharomycetes, Komagataella, Basidiomycota, Agaricomycotina, Tremellomycetes, Pucciniomycotina, Aureobasidium, Coniochaeta, and Microboryomycetes. In some embodiments, the plant is

selected from the group consisting of Siraitia, Momordica, Gynostemma, Cucurbita, Cucumis, Arabidopsis, Artemisia, Stevia, Panax, Withania, Euphorbia, Medicago, Chlorophytum, Eleutherococcus, Aralia, Morus, Medicago, Betula, Astragalus, Jatropha, Camellia, Hypholoma, Aspergillus, Solanum, Huperzia, Pseudostellaria, Corchorus, Hedera, Marchantia, and Morus. In some embodiments, the fungus is selected from Trichophyton, Sanghuangporus, Taiwanofungus, Moniliophthora, Marssonina, Diplodia, Lentinula, Xanthophyllomyces, Pochonia, Colletotrichum, Diaporthe, Histoplasma, Coccidioides, Histoplasma, Sanghuangporus, Aureobasidium, Pochonia, Penicillium, Sporothrix, and Metarhizium.

[0056] In some embodiments, the cell comprises a gene encoding at least one hydrolytic enzyme capable of hydrolyzing Mogroside V. In some embodiments, Compound 1 displays tolerance to hydrolytic enzymes in the recombinant cell, wherein the hydrolytic enzymes display capabilities of hydrolyzing Mogroside VI, Mogroside V, Mogroside IV to Mogroside IIIE.

[0057] In some embodiments, the recombinant cell comprises an oxidosqualene cyclase such as a cycloartenol synthase or a beta-amyrin synthase or a nucleic acid sequence encoding an oxidosqualene cyclase such as a cycloartenol synthase or a beta-amyrin synthase, and wherein the oxidosqualene cyclase, cycloartenol synthase, or beta-amyrin synthase are modified to produce cucurbitadienol or epoxycucurbitadienol.

[0058] In some embodiments, the oxidosqualene cyclase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 341, 343 and 346-347. In some embodiments, the cell comprises cytochrome P450 reductase or a gene encoding cytochrome P450 reductase. In some embodiments, the cytochrome P450 reductase regenerates cytochrome P450 activity. In some embodiments, the cytochrome P450 reductase comprises a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 318.

[0059] In some embodiments, the cell comprises a sequence set forth in any one of SEQ ID NOs: 897, 899, 909, 911, 913, 418, 421, 423, 425, 427, 871, 873, 901, 903, and 905. In some embodiments, the cell comprises an enzyme comprising a sequence set forth in or is

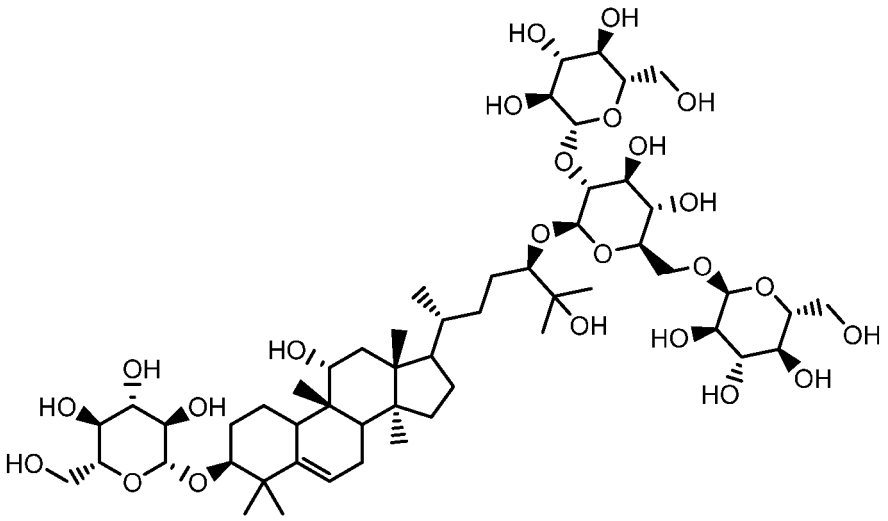
encoded by any one sequence of SEQ ID NOs: 315, 316, 420, 422, 424, 426, 430, 431, 446, 871, 845-949, and 951-1012.

[0060] In some embodiments, the cell comprises a gene encoding a hydrolase capable of hydrolyzing a first mogroside to produce Mogroside IIIE. In some embodiments, the hydrolase is a β-glucan hydrolase. In some embodiments, the hydrolase is EXG1 or EXG2. In some embodiments, the hydrolase comprises an amino acid sequence comprising having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs. 292, 366-368, 372, 376-398, 447-520, 524-845, 1013, 1014, and 1023.

[0061] In some embodiments, the first mogroside is mogroside IV, mogroside V, mogroside VI, or a combination thereof. In some embodiments, the first mogroside is mogroside V, siamenoside I, mogroside IVE, mogroside VI, mogroside IVA, or a combination thereof.

[0062] The gene encoding the hydrolase can be heterologous or homologous to the recombinant host cell. The gene encoding the hydrolase can be expressed at a normal level or overexpressed in the recombinant host cell. In some embodiments, the cell is a yeast cell. In some embodiments, the cell is *Saccharomyces cerevisiae* or *Yarrowia lipolytica*.

[0063] Also disclosed herein include a method of producing Compound 1 having the structure of:



(1). In some embodiments, the method comprises: contacting a first mogroside with one or more hydrolase to produce mogroside IIIE; and contacting the mgroside IIIE with an enzyme capable of catalyzing production of Compound 1 from mogroside IIIE.

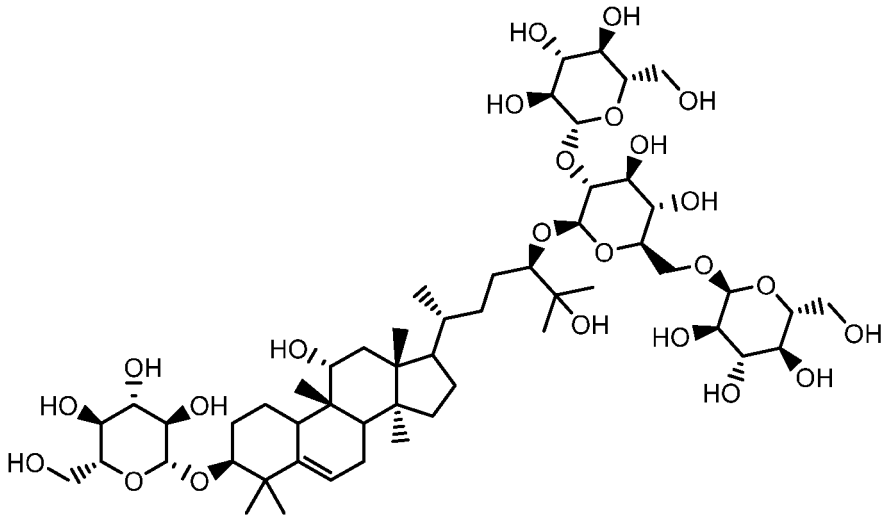
[0064] In some embodiments, the hydrolase is a β-glucan hydrolase. In some embodiments, the hydrolase is EXG1, for example an EXG1 comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 1012 or 1014. In some embodiments, the hydrolase is EXG2, for example an EXG2 comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 1023.

[0065] In some embodiments, the first mogroside is mogroside IV, a mogroside V, a mogroside VI, or a combination thereof. In some embodiments, the first mogroside is mogroside V, siamenoside I, mogroside IVE, mogroside VI, mogroside IVA, or a combination thereof.

[0066] In some embodiments, contacting the first mogroside with the one or more hyrdrolase comprises contacting the first mogroside with a recombinant host cell that comprises a first gene encoding the hydrolase and a second gene encoding the enzyme capable of catalyzing production of Compound 1. In some embodiments, the first mogroside contacts the one or more hyrdrolase in a recombinant host cell that comprises a first gene encoding the hydrolase and a second gene encoding the enzyme capable of catalyzing production of Compound 1. In some embodiments, the first mogroside is produced by the recombinant host cell. In some embodiments, the first gene is native to the recombinant host cell. In some embodiments, the first gene is expressed at a normal level. In some embodiments, the first gene is heterologous to the recombinant host cell. In some embodiments, the first gene is overexpressed. In some embodiments, the second gene is heterologous to the recombinant host cell. In some embodiments, the enzyme capable of catalyzing production of Compound 1 is one or more of UDP glycosyltransferases, cyclomaltodextrin glucanotransferases (CGTases), glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0067] Disclosed herein includes a recombinant cell, comprising: a first gene encoding a hydrolase capable of hydrolyzing a first mogroside to produce mogroside IIIE; and a second gene encoding an enzyme capable of catalyzing production of Compound 1 from

Mogroside IIIE, wherein Compound 1 has the structure:



(1).

[0068] In some embodiments, the hydrolase is a β-glucan hydrolase. For example, the hydrolase can be EXG1 or EXG2. In some embodiments, the EXG2 protein comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 1023. In some embodiments, the first mogroside is mogroside IV, mogroside V, mogroside VI, or a combination thereof. In some embodiments, the first mogroside is mogroside V, siamenoside I, mogroside IVE, mogroside VI, mogroside IVA, or a combination thereof. The first gene can be heterologous or homologous to the recombinant host cell. The first gene can be overexpressed or expressed at a normal level in the host cell. The cell can be, for example, a yeast cell. In some embodiments, the cell is *Saccharomyces cerevisiae* or *Yarrowia lipolytica*.

[0069] Also disclosed herein include a fusion polypeptide having cucurbitadienol synthase activity, wherein the fusion polypeptide comprises a fusion domain fused to a cucurbitadienol synthase. The fusion domain can be fused to the the N-terminus, the C-terminus, or both of the cucurbitadienol synthase. The fusion domain can be, for example, about 3 to about 1000 amino acids long, or about 5 to about 50 amino acids long. In some embodiments, the fusion domain comprises a substantial portion or the entire sequence of a functional protein. In some embodiments, the fusion domain is a substantial portion or the entire sequence of a functional protein. In some embodiments, the fusion domain comprises a portion or the entire sequence of a yeast protein. In some embodiments, the fusion domain is a portion or the entire

sequence of a yeast protein. In some embodiments, the fusion polypeptide comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 851, 854, 856, 1024, 859, 862, 865, 867, 915, 920, 924, 928, 932, 936, 940, 944, 948, 952, 956, 959, 964, 967, 971, 975, 979, 983, 987, 991, 995, 999, 1003, 1007, and 1011. In some embodiments, the fusion polypeptide comprises, or consists of, an amino acid sequence set forth in any one of SEQ ID NOs: 851, 854, 856, 1024, 859, 862, 865, 867, 915, 920, 924, 928, 932, 936, 940, 944, 948, 952, 956, 959, 964, 967, 971, 975, 979, 983, 987, 991, 995, 999, 1003, 1007, and 1011. In some embodiments, the fusion domain of the fusion polypeptide comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 866, 870, 917, 921, 925, 929, 933, 937, 941, 945, 949, 953, 957, 961, 968, 972, 976, 980, 984, 988, 992, 996, 1000, 1004, 1008, and 1012. In some embodiments, the fusion domain of the fusion polypeptide comprises, or consists of, a sequence set forth in any one of SEQ ID NOs: 866, 870, 917, 921, 925, 929, 933, 937, 941, 945, 949, 953, 957, 961, 968, 972, 976, 980, 984, 988, 992, 996, 1000, 1004, 1008, and 1012.

[0070] In some embodiments, the cucurbitadienol synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327, 329-333, 420, 422, 424, 426, 446, 902, 904, and 906. In some embodiments, the cucurbitadienol synthase comprises, or consists of, an amino acid sequence set forth in any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327, 329-333, 420, 422, 424, 426, 446, 902, 904, and 906. In some embodiments, the cucurbitadienol synthase is encoded by a gene comprising a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903, and 905. In some embodiments, the cucurbitadienol synthase is encoded by a gene comprising, or consisting of, a nucleic acid sequence set forth in any one of SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903, and 905.

[0071] Disclosed herein include a recombinant nucleic acid molecule which comprises a nucleic acid sequence encoding any of the fusion polypeptides disclosed herein and

having cucurbitadienol synthase activity. Disclosed herein include a recombinant cell comprising any of the fusion polypeptide disclosed herein and having cucurbitadienol synthase activity and/or any recombinant nucleic acid molecules disclosed herein encoding a fusion polypeptide having cucurbitadienol synthase activity. Also discloshed herein include a method using any of the fusion polypeptide disclosed herein and having cucurbitadienol synthase activity. The method can comprise contacting a substrate for cucurbitadienol synthase with a fusion polypeptide having cucurbitadienol synthase activity. In some embodiments, the contacting results in a production of curcurbitadienol, 24,25-epoxy curcurbitadienol, or a combination thereof. In some embodiments, the substrate for cucurbitadienol synthase comprises one or more of 2,3-oxidosqualene, dioxidosqualene and diepoxysqualene. In some embodiments, the contacting comprises contacting the substrate with a recombinant host cell which comprises a nucleic acid sequence encoding the fusion polypeptide. The recombinant host cell can, for example, express the fusion polypeptide. In some embodiments, the substrate is provided to, present in, and/or produced by the recombinant host cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0072] **Figure 1** shows HPLC data and mass spectroscopy data (inset) of Compound 1 production after treatment of Mogroside IIIE with CGTase.

[0073] **Figure 2** shows HPLC data and mass spectroscopy data (inset) of Compound 1 production after treatment of Mogroside IIIE with *Streptococcus mutans* Clarke ATCC 25175 Dextransucrase.

[0074] **Figure 3** shows HPLC data and mass spectroscopy data (inset) of mogroside glycosylation reaction after treatment with Celluclast in the presence of xylan.

[0075] **Figures 4** and **5** shows HPLC data and mass spectroscopy data (inset) of mogroside glycosylation reaction after treatment with UDP-glycosyltransferase.

[0076] **Figure 6** shows HPLC data and mass spectroscopy data (inset) of Mogrol after treatment with UDP-glycosyltransferase UGT73C6 to Mogroside I.

[0077] **Figures 7-9** show HPLC data and mass spectroscopy data (inset) of Mogrol after treatment with UDP-glycosyltransferase (338) (SEQ ID NO: 405) to the products Mogroside I, Mogroside IIA, and 2 different Mogroside III products.

[0078] **Figures 10** and **11** show HPLC data and mass spectroscopy data (inset) of Mogroside IIIE after treatment with UDP-glycosyltransferase to produce Siamenoside I and Mogroside V products.

[0079] **Figures 12-14** show HPLC data and mass spectroscopy data (inset) of products of the reaction of Mogrol, Siamenoside I or Compound 1 after treatment with UDP-glycosyltransferase (339) (SEQ ID NO:409) to produce Mogroside I, Isomogroside V and Compound 1 derivative, respectively.

[0080] **Figures 15-20** show HPLC data and mass spectroscopy data (inset) of Mogroside IIIA, Mogroside IVE, Mogroside V, respectively which were produced treating Mogroside IIA, Mogroside IIE, Mogroside IIIE, Mogroside IVA, or Mogroside IVE with UDP-glycosyltransferase (330) (SEQ ID NO: 411) .

[0081] **Figures 21** and **22** show mass spectroscopy profile of reaction products Mogroside IVE and Mogroside V.

[0082] **Figure 23** shows production of cucurbitadienol with cucurbitadienol synthase (SgCbQ) (SEQ ID NO: 417).

[0083] **Figure 24** shows production of cucurbitadienol using the enzyme Cpep2 (SEQ ID NO: 420).

[0084] **Figure 25** shows production of cucurbitadienol using the enzyme Cpep4 (SEQ ID NO: 422).

[0085] **Figure 26** shows production of dihydroxycucurbitadienol from catalysis by epoxide hydrolase (SEQ ID NO: 428).

[0086] **Figures 27A-B** show tolerance of Compound 1 to hydrolysis by microbial enzymes

[0087] **Figure 28** shows UPLC chromatogram of α-mogroside isomers mixture from Hilic_80_20_method.

[0088] **Figure 29** shows purity of the sample from UPLC analysis on Hilic_80_20_method.

[0089] **Figure 30** shows a flow chart showing a non-limiting exemplary pathway for producing Compound 1.

[0090] **Figure 31** shows the UV absorbance of the diepoxysqulene of Step 1 shown in Figure 30 for boosting oxidosqualene.

[0091] **Figure 32** shows production of cucurbitadienol in step 2 using enzymes from *Cucumis melo* and *Cucurbita maxima*

[0092] **Figure 33** shows production of cucurbitadienol in step 2 using enzyme from *Pisum sativum.*

[0093] **Figure 34** shows production of cucurbitadienol in step 2 using enzyme from *Dictyostelium sp..*

[0094] **Figure 35** shows the intermediates of Step 3 of the pathway shown in Figure 30.

[0095] **Figure 36** shows the mass spectroscopy data of the intermediates of step 4 of the pathyway shown in Figure 30, mogrol synthesis.

[0096] **Figure 37** shows the intermediates of step 7 of the pathway shown in Figure 30, synthesis of Compound 1.

[0097] **Figure 38** is a schematic illustration showing the production of hyper-glycosylated mogrosides through glycosolation enzymes, which may then be hydrolyzed back to Mogroside IIIE.

[0098] **Figure 39** is a schematic illustration showing how hydrolysis can be used to hydrolyze hyper-glycosylated mogrosides to produce Mogroside IIIE, which can then be converted to Compound 1.

[0099] **Figures 40A-B** show that after 2 days of incubation, substantially all of the mogrosides were converted to Mogroside IIIE in *S. cerevisiae* or *Y. lipolytica*.

[0100] **Figure 41** shows that no hydrolysis product from Compound 1 was detected from in *S. cerevisiae* or *Y. lipolytica*.

[0101] **Figure 42** shows production of Compound 1 in *S. cerevisiae* modified to overexpress a dextransucrase (DexT).

[0102] **Figure 43** shows enzymatic reactions catalyzed by the 311 enzyme (UDP-glycosyltransferases.

DETAILED DESCRIPTION

Definitions

[0103] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art to which this disclosure belongs. All patents, applications, published applications, and other publications are incorporated by reference in their entirety. In the event that there is a plurality of definitions for a term herein, those in this section prevail unless stated otherwise.

[0104] "Solvate" refers to the compound formed by the interaction of a solvent and a compound described herein or salt thereof. Suitable solvates are physiologically acceptable solvates including hydrates.

[0105] A "sweetener", "sweet flavoring agent", "sweet flavor entity", "sweet compound," or "sweet tasting compound," as used herein refers to a compound or physiologically acceptable salt thereof that elicits a detectable sweet flavor in a subject.

[0106] As used herein, the term "operably linked" is used to describe the connection between regulatory elements and a gene or its coding region. Typically, gene expression is placed under the control of one or more regulatory elements, for example, without limitation, constitutive or inducible promoters, tissue-specific regulatory elements, and enhancers. A gene or coding region is said to be "operably linked to" or "operatively linked to" or "operably associated with" the regulatory elements, meaning that the gene or coding region is controlled or influenced by the regulatory element. For instance, a promoter is operably linked to a coding sequence if the promoter effects transcription or expression of the coding sequence.

[0107] The term "regulatory element" and "expression control element" are used interchangeably and refer to nucleic acid molecules that can influence the expression of an operably linked coding sequence in a particular host organism. These terms are used broadly to and cover all elements that promote or regulate transcription, including promoters, 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, intrans, core elements required for basic interaction of RNA polymerase and transcription factors, upstream elements, enhancers, response elements (*see, e.g.*, Lewin, "Genes V" (Oxford University Press, Oxford) pages 847-873), and any

combination thereof. Exemplary regulatory elements in prokaryotes include promoters, operator sequences and a ribosome binding sites. Regulatory elements that are used in eukaryotic cells can include, without limitation, transcriptional and translational control sequences, such as promoters, enhancers, splicing signals, polyadenylation signals, terminators, protein degradation signals, internal ribosome-entry element (IRES), 2A sequences, and the like, that provide for and/or regulate expression of a coding sequence and/or production of an encoded polypeptide in a host cell. In some embodiments herein, the recombinant cell described herein comprises a genes operably linked to regulatory elements.

[0108] As used herein, 2A sequences or elements refer to small peptides introduced as a linker between two proteins, allowing autonomous intraribosomal self-processing of polyproteins (*See e.g.*, de Felipe. Genetic Vaccines and Ther. 2:13 (2004); deFelipe *et al.* Traffic 5:616-626 (2004)). These short peptides allow co-expression of multiple proteins from a single vector. Many 2A elements are known in the art. Examples of 2A sequences that can be used in the methods and system disclosed herein, without limitation, include 2A sequences from the foot-and-mouth disease virus (F2A), equine rhinitis A virus (E2A), *Thosea asigna* virus (T2A), and porcine teschovirus-1 (P2A) as described in U.S. Patent Publication No. 20070116690.

[0109] As used herein, the term "promoter" is a nucleotide sequence that permits binding of RNA polymerase and directs the transcription of a gene. Typically, a promoter is located in the 5' non-coding region of a gene, proximal to the transcriptional start site of the gene. Sequence elements within promoters that function in the initiation of transcription are often characterized by consensus nucleotide sequences. Examples of promoters include, but are not limited to, promoters from bacteria, yeast, plants, viruses, and mammals (including humans). A promoter can be inducible, repressible, and/or constitutive. Inducible promoters initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, such as a change in temperature.

[0110] As used herein, the term "enhancer" refers to a type of regulatory element that can increase the efficiency of transcription, regardless of the distance or orientation of the enhancer relative to the start site of transcription.

[0111] As used herein, the term "transgene" refers to any nucleotide or DNA sequence that is integrated into one or more chromosomes of a target cell by human intervention.

In some embodiment, the transgene comprises a polynucleotide that encodes a protein of interest. The protein-encoding polynucleotide is generally operatively linked to other sequences that are useful for obtaining the desired expression of the gene of interest, such as transcriptional regulatory sequences. In some embodiments, the transgene can additionally comprise a nucleic acid or other molecule(s) that is used to mark the chromosome where it has integrated.

[0112] "Percent (%) sequence identity" with respect to polynucleotide or polypeptide sequences is used herein as the percentage of bases or amino acid residues in a candidate sequence that are identical with the bases or amino acid residues in another sequence, after aligning the two sequences. Gaps can be introduced into the sequence alignment, if necessary, to achieve the maximum percent sequence identity. Conservative substitutions are not considered as part of the sequence identity. Alignment for purposes of determining percent (%) sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer methods and programs such as BLAST, BLAST-2, ALIGN, FASTA (available in the Genetics Computing Group (GCG) package, from Madison, Wisconsin, USA), or Megalign (DNASTAR). Those of skill in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

[0113] For instance, percent (%) amino acid sequence identity values may be obtained by using the WU-BLAST-2 computer program described in, for example, Altschul et al., Methods in Enzymology, 1996, 266:460-480. Many search parameters in the WU-BLAST-2 computer program can be adjusted by those skilled in the art. For example, some of the adjustable parameters can be set with the following values: overlap span = 1, overlap fraction = 0.125, word threshold (T) = 11, and scoring matrix = BLOSUM62. When WU-BLAST-2 is used, a % amino acid sequence identity value is determined by dividing (a) the number of matching identical amino acid residues between the amino acid sequence of a first protein of interest and the amino acid sequence of a second protein of interest as determined by WU-BLAST-2 by (b) the total number of amino acid residues of the first protein of interest.

[0114] Percent amino acid sequence identity may also be determined using the sequence comparison program NCBI-BLAST2 described in, for example, Altschul et al., Nucleic Acids Res., 1997, 25:3389-3402. The NCBI-BLAST2 sequence comparison program may be

downloaded from http://www.ncbi.nlm.nih.gov or otherwise obtained from the National Institute of Health, Bethesda, MD. NCBI-BLAST2 uses several adjustable search parameters. The default values for some of those adjustable search parameters are, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, drop-off for final gapped alignment = 25 and scoring matrix = BLOSUM62.

[0115] In situations where NCBI-BLAST2 is used for amino acid sequence comparisons, the % amino acid sequence identity of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % amino acid sequence identity to, with, or against a given amino acid sequence B) is calculated as follows:

100 times the fraction X/Y

where X is the number of amino acid residues scored as identical matches by the sequence alignment program NCBI-BLAST2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % amino acid sequence identity of A to B will not equal the % amino acid sequence identity of B to A.

[0116] As used herein, "isolated" means that the indicated compound has been separated from its natural milieu, such that one or more other compounds or biological agents present with the compound in its natural state are no longer present.

[0117] As used herein, "purified" means that the indicated compound is present at a higher amount relative to other compounds typically found with the indicated compound (e.g., in its natural environment). In some embodiments, the relative amount of purified a purified compound is increased by greater than 1%, 5%, 10%, 20%, 30%, 40%, 50%, 80%, 90%, 100%, 120%, 150%, 200%, 300%, 400%, or 1000%. In some embodiments, a purified compound is present at a weight percent level greater than 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or 99.5% relative to other compounds combined with the compound. In some embodiments, the compound 1 produced from the embodiments herein is present at a weight percent level greater than 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%,

80%, 90%, 95%, 98%, 99%, or 99.5% relative to other compounds combined with the compound after production.

[0118] "Purification" as described herein, can refer to the methods for extracting Compound 1 from the cell lysate and/or the supernatant, wherein the cell is excreting the product of Compound 1. "Lysate" as described herein, comprises the cellular content of a cell after disruption of the cell wall and cell membranes and can include proteins, sugars, and mogrosides, for example. Purification can involve ammonium sulfate precipitation to remove proteins, salting to remove proteins, hydrophobic separation (HPLC), and use of an affinity column. In view of the products produced by the methods herein, affinity media is contemplated for the removal of specific mogrosides with an adsorbent resin.

[0119] "HPLC" as described herein is a form of liquid chromatography that can be used to separate compounds that are dissolved in solution. Without being limiting the HPLC instruments can comprise of a reservoir of mobile phase, a pump, an injector, a separation column, and a detector. Compounds can then be separated by injecting a sample mixture onto the column. The different components in the mixture pass can pass through the column at different rates due to differences in their partitioning behavior between the mobile liquid phase and the stationary phase. There are several columns that can be used. Without being limiting the columns can be nomal phase columns, reverse phase columns, size exclusion type of columns, and ion exchange columns.

[0120] Also contemplated is the use of solid phase extraction and fractionation, which is useful for desalting proteins and sugar samples. Other methods can include the use of HPLC, liquid chromatography for analyzing samples, and liquid-liquid extraction, described in Auréa Andrade-Eiroa *et al.* (TrAC Trends in Analytical Chemistry Volume 80, June 2016, Pages 641–654; incorporated by reference in its entirety herein.

[0121] "Solid phase extraction" (SPE) for purification, as described herein, refers to a sample preparation process in which compounds that are dissolved or suspended in a liquid mixture are separated from other compounds in the mixture according to their physical and chemical properties. For example, analytical laboratories can use solid phase extraction to concentrate and purify samples for analysis. Solid phase extraction can also be used to isolate analytes of interest from a wide variety of matrices, including urine, blood, water, beverages,

soil, and animal tissue, for example. In the embodiments herein, Compound 1 that is in cell lysate or in the cell media can be purified by solid phase extraction.

[0122] SPE uses the affinity of solutes dissolved or suspended in a liquid (known as the mobile phase) for a solid through which the sample is passed (known as the stationary phase) to separate a mixture into desired and undesired components. SPE can also be used and applied directly in gas–solid phase and liquid–solid phase, or indirectly to solid samples by using, e.g., thermodesorption with subsequent chromatographic analysis. This can result in either the desired analytes of interest or undesired impurities in the sample are retained on the stationary phase. The portion that passes through the stationary phase can be collected or discarded, depending on whether it contains the desired analytes or undesired impurities. If the portion retained on the stationary phase includes the desired analytes, they can then be removed from the stationary phase for collection in an additional step, in which the stationary phase is rinsed with an appropriate eluent.

[0123] Ways that the solid phase extraction can be performed are not limited. Without being limiting, the procedures may include: Normal phase SPE procedure, Reversed phase SPE, Ion exchange SPE, Anion exchange SPE, Cation exchange, and Solid-phase microextraction. Solid phase extraction is described in Sajid *et al.,* and Płotka-Wasylka J *et al.* (Anal Chim Acta. 2017 May 1;965:36-53, Crit Rev Anal Chem. 2017 Apr 11:1-11; incorporated by reference in its entirety).

[0124] In some embodiments, the compound 1 that is produced by the cell is purified by solid phase extraction. In some embodiments, the purity of compound 1, for example purified by solid phase extraction is 70%, 80%, 90% or 100% pure or any level of purity defined by any aforementioned values.

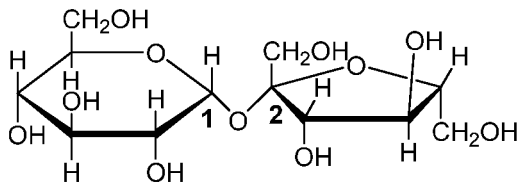
[0125] "Fermentation" as described herein, refers broadly to the bulk growth of host cells in a host medium to produce a specific product. In the embodiments herein, the final product produced is Compound 1. This can also include methods that occur with or without air and can be carried out in an anaerobic environment, for example. The whole cells (recombinant host cells) may be in fermentation broth or in a reaction buffer.

[0126] Compound 1 and intermediate mogroside compound for the production of Compound 1 can be isolated by collection of intermediate mogroside compounds and Compound

1 from the recombinant cell lysate or from the supernatant. The lysate can be obtained after harvesting the cells and subjecting the cells to lysis by shear force (French press cell or sonication) or by detergent treatment. The lysate can then be filtered and treated with ammonium sulfate to remove proteins, and fractionated on a C18 HPLC (5 X 10 cm Atlantis prep T3 OBD column, 5 um, Waters) and by injections using an A/B gradient (A = water B = acetonitrile) of 10 → 30% B over 30 minutes, with a 95% B wash, followed by re-equilibration at 1% (total run time = 42 minutes). The runs can be collected in tared tubes (12 fractions/plate, 3 plates per run) at 30 mL/fraction. The lysate can also be centrifuged to remove solids and particulate matter.

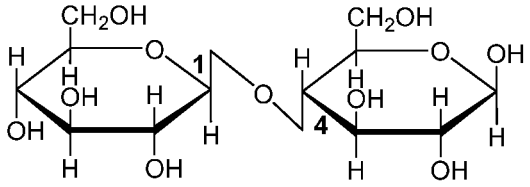
[0127] Plates can then be dried in the Genevac HT12/HT24. The desired compound is expected to be eluted in Fraction 21 along with other isomers. The pooled Fractions can be further fractionated in 47 runs on fluoro-phenyl HPLC column (3 X 10 cm, Xselect fluoro-phenyl OBD column, 5 um, Waters) using an A/B gradient (A = water, B = acetonitrile) of 15 → 30% B over 35 minutes, with a 95% B wash, followed by re-equilibration at 15% (total run time = 45 minutes). Each run was collected in 12 tared tubes (12 fractions/plate, 1 plate per run) at 30 mL/fraction. Fractions containing the desired peak with the desired purity can be pooled based on UPLC analysis and dried under reduced pressure to give a whitish powdery solid. The pure compound can be re-suspended/dissolved in 10 mL of water and lyophilized to obtain at least a 95% purity.

[0128] As used herein, a "glycosidic bond" refers to a covalent bond connecting two furanose and/or pyranose groups together. Generally, a glycosidic bond is the bond between the anomeric carbon of one furanose or pyranose moiety and an oxygen of another furanose or pyranose moiety. Glycosidic bonds are named using the numbering of the connected carbon atoms, and the alpha/beta orientation. α- and β-glycosidic bonds are distinguished based on the relative stereochemistry of the anomeric position and the stereocenter furthest from C1 in the ring. For example, sucrose is a disaccharide composed of one molecule of glucose and one molecule of fructose connected through an alpha 1-2 glycosidic bond, as shown below.



Alpha-glucopyranose Beta-fructofuranose

[0129] An example of a beta 1-4 glycosidic bond can be found in cellulose:



[0130] As used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "an aromatic compound" includes mixtures of aromatic compounds.

[0131] Often, ranges are expressed herein as from "about" one particular value, and/or to "about" another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint.

[0132] "Codon optimization" as described herein, refers to the design process of altering codons to codons known to increase maximum protein expression efficiency. In some alternatives, codon optimization for expression in a cell is described, wherein codon optimization can be performed by using algorithms that are known to those skilled in the art so as to create synthetic genetic transcripts optimized for high mRNA and protein yield in humans. Codons can be optimized for protein expression in a bacterial cell, mammalian cell, yeast cell, insect cell, or plant cell, for example. Programs containing algorithms for codon optimization in humans are readily available. Such programs can include, for example, OptimumGene™ or GeneGPS® algorithms. Additionally codon optimized sequences can be obtained commercially, for example, from Integrated DNA Technologies. In some of the embodiments herein, a recombinant cell for the production of Compound 1 comprises genes encoding enzymes for synthesis, wherein the genes are codon optimized for expression. In some embodiments, the genes are codon optimized for expression in bacterial, yeast, fungal or insect cells.

[0133] As used herein, the terms "nucleic acid," "nucleic acid molecule," and "polynucleotide" are interchangeable and refer to any nucleic acid, whether composed of phosphodiester linkages or modified linkages such as phosphotriester, phosphoramidate,

siloxane, carbonate, carboxymethylester, acetamidate, carbamate, thioether, bridged phosphoramidate, bridged methylene phosphonate, bridged phosphoramidate, bridged phosphoramidate, bridged methylene phosphonate, phosphorothioate, methylphosphonate, phosphorodithioate, bridged phosphorothioate or sultone linkages, and combinations of such linkages. The terms "nucleic acid" and "polynucleotide" also specifically include nucleic acids composed of bases other than the five biologically occurring bases (adenine, guanine, thymine, cytosine and uracil).

[0134] Non-limiting examples of polynucleotides include deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), oligonucleotides, fragments generated by the polymerase chain reaction (PCR), and fragments generated by any of ligation, scission, endonuclease action, and exonuclease action. Nucleic acid molecules can be composed of monomers that are naturally-occurring nucleotides (such as DNA and RNA), or analogs of naturally-occurring nucleotides (e.g., enantiomeric forms of naturally-occurring nucleotides), or a combination of both. Modified nucleotides can have alterations in sugar moieties and/or in pyrimidine or purine base moieties. Sugar modifications include, for example, replacement of one or more hydroxyl groups with halogens, alkyl groups, amines, and azido groups, or sugars can be functionalized as ethers or esters. Moreover, the entire sugar moiety can be replaced with sterically and electronically similar structures, such as aza-sugars and carbocyclic sugar analogs. Examples of modifications in a base moiety include alkylated purines and pyrimidines, acylated purines or pyrimidines, or other well-known heterocyclic substitutes. Nucleic acid monomers can be linked by phosphodiester bonds or analogs of such linkages. Analogs of phosphodiester linkages include phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoranilidate, phosphoramidate, and the like. The term "nucleic acid molecule" also includes so-called "peptide nucleic acids," which comprise naturally-occurring or modified nucleic acid bases attached to a polyamide backbone. Nucleic acids can be either single stranded or double stranded. In some alternatives, a nucleic acid sequence encoding a fusion protein is provided. In some alternatives, the nucleic acid is RNA or DNA. In some embodiments, the nucleic acid comprises any one of SEQ ID NOs: 1-1023.

[0135] "Coding for" or "encoding" are used herein, and refers to the property of specific sequences of nucleotides in a polynucleotide, such as a gene, a cDNA, or an mRNA, to

serve as templates for synthesis of other macromolecules such as a defined sequence of amino acids. Thus, a gene codes for a protein if transcription and translation of mRNA corresponding to that gene produces the protein in a cell or other biological system. In some embodiments herein, a recombinant cell is provided, wherein the recombinant cell comprises genes encoding for enzymes such as dextransucrase, UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, dextranases, and/or UGT. In some embodiments, the transglucosidases comprises an amino acid sequence set forth by any one of SEQ ID NOs: 163-290 and 723. In some embodiments, the CGTases are encoded by or have the sequence of any one of SEQ ID NOs: 1, 3, 78-101, 147 and 154. In some embodiments, the genes encoding the enzymes such as dextransucrase, UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, dextranases, and/or UGT are codon optimized for expression in the host cell. A "nucleic acid sequence coding for a polypeptide" includes all nucleotide sequences that are degenerate versions of each other and that code for the same amino acid sequence.

[0136] Optimization can also be performed to reduce the occurrence of secondary structure in a polynucleotide. In some alternatives of the method, optimization of the sequences in the vector can also be performed to reduce the total GC/AT ratio. Strict codon optimization can lead to unwanted secondary structure or an undesirably high GC content that leads to secondary structure. As such, the secondary structures affect transcriptional efficiency. Programs such as GeneOptimizer can be used after codon usage optimization, for secondary structure avoidance and GC content optimization. These additional programs can be used for further optimization and troubleshooting after an initial codon optimization to limit secondary structures that can occur after the first round of optimization. Alternative programs for optimization are readily available. In some alternatives of the method, the vector comprises sequences that are optimized for secondary structure avoidance and/or the sequences are optimized to reduce the total GC/AT ratio and/or the sequences are optimized for expression in a bacterial or yeast cell.

[0137] "Vector," "Expression vector" or "construct" is a nucleic acid used to introduce heterologous nucleic acids into a cell that has regulatory elements to provide expression of the heterologous nucleic acids in the cell. Vectors include but are not limited to

plasmid, minicircles, yeast, and viral genomes. In some alternatives, the vectors are plasmid, minicircles, yeast, or genomes. In some alternatives, the vector is for protein expression in a bacterial system such as *E. coli.* In some alternatives, the vector is for protein expression in a bacterial system, such as *E. coli.* In some alternatives, the vector is for protein expression in a yeast system. In some embodiments, the vector for expression is a viral vector. In some embodiments the vector is a recombinant vector comprising promoter sequences for upregulation of expression of the genes. "Regulatory elements" can refer to the nucleic acid that has nucleotide sequences that can influence the transcription or translation initiation and rate, stability and mobility of a transcription or translation product.

[0138] "Recombinant host" or "recombinant host cell" as described herein is a host, the genome of which has been augmented by at least one incorporated DNA sequence. Said incorporated DNA sequence may be a heterologous nucleic acid encoding one or more polypeptides. 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 nonrecombinant host. In some embodiments, the recombinant host cell is used to prevent expression problems such as codon-bias. There are commercial hosts for expression of proteins, for example, BL21-CodonPlus™ cells, tRNA-Supplemented Host Strains for Expression of Heterologous Genes, Rosetta™ (DE3) competent strains for enhancing expression of proteins, and commercial yeast expression systems in the genera *Saccharomyces*, *Pichia*, *Kluyveromyces*, *Hansenula* and *Yarrowia.*

[0139] The recombinant host may be a commercially available cell such as Rosetta cells for expression of enzymes that may have rare codons.

[0140] In some embodiments, the recombinant cell comprises a recombinant gene for the production of cytochrome P450 polypeptide comprising the amino acid sequence of any one of CYP533, CYP937, CYP1798, CYP1994, CYP2048, CYP2740, CYP3404, CYP3968, CYP4112, CYP4149, CYP4491,CYP5491, CYP6479, CYP7604, CYP8224, CYP8728, CYP10020, and CYP10285. In some embodiments, the P450 polypeptide is encoded in genes comprising any one of the sequences set forth in SEQ ID Nos: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, and 891.

[0141] In some embodiments, the P450 enzyme is aided by at least one CYP activator, such as CPR4497. In some embodiments, the recombinant host cell further comprises a gene encoding CPR4497, wherein the gene comprises a nucleic acid sequence set forth in SEQ ID NO: 112. In some embodiments, the recombinant host cell further comprises a gene encoding CPR4497, wherein the amino acid sequence of CPR4497 is set forth in SEQ ID NO: 113.

[0142] In some embodiments, wherein the recombinant host cell is a yeast cell, the recombinant cell has a deletion of EXG1 gene and/or the EXG2 gene to prevent reduction of glucanase activity which may lead to deglucosylation of mogrosides.

[0143] The type of host cell can vary. For example, the host cell can be selected from a group consisting of *Agaricus, Aspergillus, Bacillus, Candida, corynebacterium, Escherichia, Fusarium/Gibberella, Kluyveromyces, Laetiporus, Lentinus, Phaffia, Phanerochaete, Pichia, Physcomitrella, Rhodoturula, Saccharomyces, Schizosaccharomyces, Sphaceloma, Xanthophyllomyces, Yarrowia, Lentinus tigrinus, Laetiporus sulphureus, Phanerochaete chrysosporium, Pichia pastoris, Physcomitrella patens, Rhodoturula glutinis, Rhodoturula mucilaginosa, Phaffia rhodozyma, Xanthophyllomyces dendrorhous, Fusarium fujikuroi/Gibberella fujikuroi, Candida utilis, Yarrowia lipolytica, Siraitia, Momordica, Gynostemma, Cucurbita, Cucumis, Arabidopsis, Artemisia, Stevia, Panax, Withania, Euphorbia, Medicago, Chlorophytum, Eleutherococcus, Aralia, Morus, Medicago, Betula, Astragalus, Jatropha, Camellia, Hypholoma, Aspergillus, Solanum, Huperzia, Pseudostellaria, Corchorus, Hedera, Marchantia, and Morus, Trichophyton, Sanghuangporus, Taiwanofungus, Moniliophthora, Marssonina, Diplodia, Lentinula, Xanthophyllomyces, Pochonia, Colletotrichum, Diaporthe, Histoplasma, Coccidioides, Histoplasma, Sanghuangporus, Aureobasidium, Pochonia, Penicillium, Sporothrix, Metarhizium, Aspergillus, Yarrowia, Lipomyces, Aspergillus nidulans, Yarrowia lipolytica, Rhodosporin toruloides, Candida, Sacccharaomyces, Saccharomycotina, Taphrinomycotina, Schizosaccharomycetes, Komagataella, Basidiomycota, Agaricomycotina, Tremellomycetes, Pucciniomycotina, Aureobasidium, Coniochaeta, Rhodosporidium,* and *Microboryomycetes, Gibberella fujikuroi, Kluyveromyces lactis, Schizosaccharomyces pombe, Aspergillus niger, Saccharomyces cerevisiae, Escherichia coli, Rhodobacter sphaeroides,* and *Rhodobacter capsulatus.* Methods to

enhance product yield have been described, for example, in *S. cerevisiae.* Methods are known for making recombinant microorganisms.

[0144] Methods to prepare recombinant host cells from *Aspergillus spp.* is described in WO2014086842, incorporated by reference in its entirety herein. Nucleotide sequences of the genomes can be obtained through gene data libraries available publicly and can allow for rational design and modifications of the pathways to enhance and improve product yield.

[0145] "Culture media" as described herein, can be a nutrient rich broth for the growth and maintenance of cells during their production phase. A yeast culture for maintaining and propagating various strains, can require specific formulations of complex media for use in cloning and protein expression, and can be appreciated by those of skill in the art. Commercially available culture media can be used from ThermoFisher for example. The media can be YPD broth or can have a yeast nitrogen base. Yeast can be grown in YPD or synthetic media at 30°C.

[0146] Lysogeny broth (LB) is typically used for bacterial cells. The bacterial cells used for growth of the enzymes and mogrosides can have antibiotic resistance to prevent the growth of other cells in the culture media and contamination. The cells can have an antibiotic gene cassettes for resistance to antibiotics such as chloramphenicol, penicillin, kanamycin and ampicillin, for example.

[0147] As described herien, a "fusion protein" is a protein created through the joining of two or more nucleic acid sequences that originally coded for a portion or entire amino acid sequence of separate proteins. For example, a fusion protein can contain a functional protein (e.g., an enzyme (including, but not limited to, cucurbitadienol synthase)) and one or more fusion domains. A fusion domain, as describe herein, can be a full length or a portion/fragment of a protein (e.g., a functional protein including but not limited to, an enzyme, a transcription factor, a toxin, and translation factor). The location of the one or more fusion domains in the fusion protein can vary. For example, the one or more fusion domains can be at the N- and/or C-terminal regions (e.g., N- and/or C- termini) of the fusion protein. The one or more fusion domains can also be at the central region of the fusion protein. The fusion domain is not required to be located at the terminus of the fusion protein. A fusion domain can be selected so as to confer a desired property. For example, a fusion domain may affect (e.g., increase or decrese) the enzymatic activity of an enzyme that it is fused to, or affect (e.g., incrase or decrease) the

stability of a protein that it is fused to. A fusion domain may be a multimerizing (e.g., dimerizing and tetramerizing) domain and/or functional domains. In some embodiments, the fusion domain may enhance or decrease the multimerization of the protein that it is fused to. As a non-limiting example, a fusion protein can contain a full length protein A and a fusion domain fused to the N-terminal region and/or C-terminal region of the full length protein A. In some examples, a fusion protein contains a partial sequence of protein A and a fusion domain fused to the N-terminal region and/or C-terminal region (e.g., the N-terminus and C-terminus) of the partial sequnce of protein A. The fusion domain can be, for example, a portion or the entire sequence of protein A, or a portion or the entire sequence of a protein different from protein A. In some embodiments, one or more of the enzymes suitable for use in the methods, systems and compositions disclosed herein can be a fusion protein. In some embodiments, the fusion protein is encoded by a nucleic acid sequence having at least 70%, 80%, 90%, 95%, or 99% sequence identity to one of the nucleic acid sequences listed in Table 1. In some embodiments, the fusion protein comprise an amino acid sequence having at least 70%, 80%, 90%, 95%, or 99% sequence identity to one of the amino acid sequences listed in Table 1. In some embodiments, the fusion protein comprises an amino acid protein sequence having at least 80%, 90%, 95%, or 99% sequence identity to one of the amino acid sequences listed in Table 1, and a fusion domain at N-, C-, or both terminal regions of the fusion protein. In some embodiments, the fusion protein comprises one of the amino acid protein sequences listed in Table 1, and a fusion domain located at N-, C-, or both termnal regions of the fusion protein.

[0148] The length of the fusion domain can vary, for example, from 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, or a range between any of these two numbers, amino acids. In some embodiments, the fusion domain is about 3, 4, 5, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 40, 50, or a range between any two of these numbers, amino acids in length. In some embodiments, the fusion domain is a substantial portion or the entire sequence of a functional protein (for example, an enzyme, a transcription factor, or a translation factor). In some embodiments, the fusion protein is a protein having cucurbitadienol synthase activity.

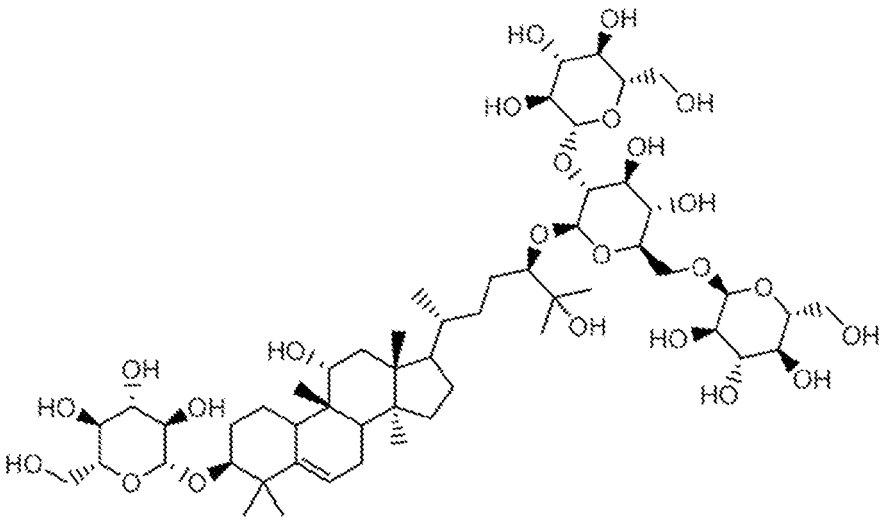
[0149] Optimizing cell growth and protein expression techniques in culture media are also contemplated. For growth in culture media, cells such as yeast can be sensitive to low pH

(Narendranath *et al.,* Appl Environ Microbiol. 2005 May; 71(5): 2239–2243; incorporated by reference in its entirety). During growth, yeast must maintain a constant intracellular pH. There are many enzymes functioning within the yeast cell during growth and metabolism. Each enzyme works best at its optimal pH, which is acidic because of the acidophilic nature of the yeast itself. When the extracellular pH deviates from the optimal level, the yeast cell needs to invest energy to either pump in or pump out hydrogen ions in order to maintain the optimal intracellular pH. As such media containing buffers to control for the pH would be optimal. Alternatively, the cells can also be transferred into a new media if the monitored pH is high.

[0150] Growth optimization of bacterial and yeast cells can also be achieved by the addition of nutrients and supplements into a culture media. Alternatively, the cultures can be grown in a fermenter designed for temperature, pH control and controlled aeration rates. Dissolved oxygen and nitrogen can flowed into the media as necessary.

[0151] The term "Operably linked" as used herein refers to functional linkage between a regulatory sequence and a heterologous nucleic acid sequence resulting in expression of the latter.

[0152] "Mogrosides" and "mogroside compounds" are used interchangeably herein and refer to a family of triterpene glycosides. Non-limiting exemplary examples of mogrosides include such as Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III, which have been identified from the fruits of Siraitia grosvenorii (Swingle) that are responsible for the sweetness of the fruits. In the embodiments herein, mogroside intermediates can be used in the *in vivo*, *ex vivo*, or *in vitro* production of Compound 1 having the structure of:



(1)

In some embodiments, a recombinant cell for producing Compound 1, further produces mogrosides and comprises genes encoding enzymes for the production of mogrosides. Recombinant cells capable of the production of mogrosides are further described in WO2014086842, incorporated by reference in its entirety herein. In some embodiments, the recombinant cell is grown in a media to allow expression of the enzymes and production of Compound 1 and mogroside intermediates. In some embodiments, Compound 1 is obtained by lysing the cell with shear force (i.e. French press cell or sonication) or by detergent lysing methods. In some embodiments, the cells are supplemented in the growth media with precursor molecules such as mogrol to boost production of Compound 1.

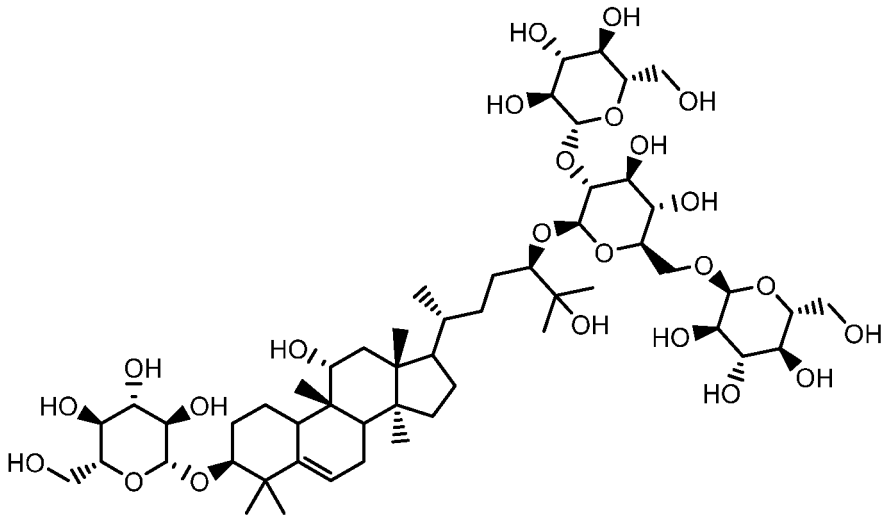
[0153] "Promoter" as used herein refers to a nucleotide sequence that directs the transcription of a structural gene. In some alternatives, a promoter is located in the 5' non-coding region of a gene, proximal to the transcriptional start site of a structural gene. Sequence elements within promoters that function in the initiation of transcription are often characterized by consensus nucleotide sequences. Without being limiting, these promoter elements can include RNA polymerase binding sites, TATA sequences, CAAT sequences, differentiation-specific elements (DSEs; McGehee *et al.*, Mol. Endocrinol. 7:551 (1993); hereby expressly incorporated by reference in its entirety), cyclic AMP response elements (CREs), serum response elements (SREs; Treisman *et al.*, Seminars in Cancer Biol. 1:47 (1990); incorporated by reference in its entirety), glucocorticoid response elements (GREs), and binding sites for other transcription

factors, such as CRE/ATF (O'Reilly *et al.*, J. Biol. Chem. 267:19938 (1992); incorporated by reference in its entirety), AP2 (Ye *et al.*, J. Biol. Chem. 269:25728 (1994); incorporated by reference in its entirety), SP1, cAMP response element binding protein (CREB; Loeken *et al.*, Gene Expr. 3:253 (1993); hereby expressly incorporated by reference in its entirety) and octamer factors (see, in general, Watson *et al.*, eds., Molecular Biology of the Gene, 4th ed. (The Benjamin/Cummings Publishing Company, Inc. 1987; incorporated by reference in its entirety)), and Lemaigre and Rousseau, Biochem. J. 303:1 (1994); incorporated by reference in its entirety). As used herein, a promoter can be constitutively active, repressible or inducible. If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent. In contrast, the rate of transcription is not regulated by an inducing agent if the promoter is a constitutive promoter.

[0154] A "ribosome skip sequence" as described herein refers to a sequence that during translation, forces the ribosome to "skip" the ribosome skip sequence and translate the region after the ribosome skip sequence without formation of a peptide bond. Several viruses, for example, have ribosome skip sequences that allow sequential translation of several proteins on a single nucleic acid without having the proteins linked via a peptide bond. As described herein, this is the "linker" sequence. In some alternatives of the nucleic acids provided herein, the nucleic acids comprise a ribosome skip sequence between the sequences for the genes for the enzymes described herein, such that the proteins are co-expressed and not linked by a peptide bond. In some alternatives, the ribosome skip sequence is a P2A, T2A, E2A or F2A sequence. In some alternatives, the ribosome skip sequence is a T2A sequence.

Compound 1

[0155] As disclosed herein, Compound 1 is a compound having the structure of:



(**1**), or a salt thereof.

[0156] Compound 1 is a high-intensity sweetener the can be used in a wide variety of products in which a sweet taste is desired. Compound 1 provides a low-calorie advantage to other sweeteners such as sucrose or fructose.

[0157] In some embodiments, Compound 1 is in an isolated and purified form. In some embodiments, Compound 1 is present in a composition in which Compound 1 is substantially purified.

[0158] In some embodiments, Compound 1 or salts thereof are isolated and is in solid form. In some embodiments, the solid form is amorphous. In some embodiments, the solid form is crystalline. In some embodiments, the compound is in the form of a lyophile. In some embodiments, Compound 1 is isolated and within a buffer.

[0159] The skilled artisan will recognize that some structures described herein may be resonance forms or tautomers of compounds that may be fairly represented by other chemical structures, even when kinetically; the artisan recognizes that such structures may only represent a very small portion of a sample of such compound(s). Such compounds are considered within the scope of the structures depicted, though such resonance forms or tautomers are not represented herein.

[0160] Isotopes may be present in Compound 1. Each chemical element as represented in a compound structure may include any isotope of said element. For example, in a compound structure a hydrogen atom may be explicitly disclosed or understood to be present in the compound. At any position of the compound that a hydrogen atom may be present, the

hydrogen atom can be any isotope of hydrogen, including but not limited to hydrogen-1 (protium) and hydrogen-2 (deuterium). Thus, reference herein to a compound encompasses all potential isotopic forms unless the context clearly dictates otherwise. In some embodiments, compounds described herein are enriched in one or more isotopes relative to the natural prevalence of such isotopes. In some embodiments, the compounds described herein are enriched in deuterium. In some embodiments, greater than 0.0312% of hydrogen atoms in the compounds described herein are deuterium. In some embodiments, greater than 0.05%, 0.08%, or 0.1% of hydrogen atoms in the compounds described herein are deuterium.

[0161] In some embodiments, Compound 1 is capable of forming acid and/or base salts by virtue of the presence of amino and/or carboxyl groups or groups similar thereto.

[0162] In some embodiments, Compound 1 is substantially isolated. In some embodiments, Compound 1 is substantially purified. In some embodiments, the compound is in the form of a lyophile. In some embodiments, the compound is crystalline. In some embodiments, the compound is amorphous.

Production Compositions

[0163] In some embodiments, the production composition contains none, or less than a certain amount, of undesirable compounds. In some embodiments, the composition contains, or does not contain, one or more isomers of Mogroside I, Mogroside II, and Mogroside III. In some embodiments, the composition contains a weight percent of less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, or 100 ppm of all isomers of Mogroside I, Mogroside II, and Mogroside III. In some embodiments, the composition contains, or does not contain, one or more of Mogroside IIIE, 11-oxo-Mogroside IIIE, Mogroside IIIA2, Mogroside IE, Mogroside IIE, and 11-oxo-mogrol. In some embodiments, the composition contains a weight percent of less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, or 100 ppm of one or more of Mogroside IIIE, 11-oxo-Mogroside IIIE, Mogroside IIIA2, Mogroside IE, Mogroside IIE, and 11-oxo-mogrol. In some embodiments, the composition contains a weight percent of less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, or 100 ppm of Mogroside IIIE. In some embodiments, the composition contains a weight percent of less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, or

100 ppm of 11-oxo-Mogroside IIIE. In some embodiments, the composition contains a weight percent of less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, or 100 ppm of 11-oxo-mogrol.

[0164] In some embodiments, the production composition is in solid form, which may by crystalline or amorphous. In some embodiments, the composition is in particulate form. The solid form of the composition may be produced using any suitable technique, including but not limited to re-crystallization, filtration, solvent evaporation, grinding, milling, spray drying, spray agglomeration, fluid bed agglomeration, wet or dry granulation, and combinations thereof. In some embodiments, a flowable particulate composition is provided to facilitate use in further food manufacturing processes. In some such embodiments, a particle size between 50 μm and 300 μm, between 80 μm and 200 μm, or between 80 μm and 150 μm is generated.

[0165] Some embodiments provide a production composition comprising Compound 1 that is in solution form. For example, in some embodiments a solution produced by one of the production processes described herein is used without further purification. In some embodiments, the concentration of Compound 1 in the solution is greater than 300 ppm, 500 ppm, 800 ppm, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, or 20% by weight. In some embodiments, the concentration of all isomers of Mogroside I, Mogroside II, and Mogroside III is less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, or 100 ppm. In some embodiments, the concentration of one or more of Mogroside IIIE, 11-oxo-Mogroside IIIE, Mogroside IIIA2, Mogroside IE, Mogroside IIE, and 11-oxo-mogrol in the production composition is less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, 100 ppm, 50 ppm, 30 ppm, 20 ppm, 10 ppm, 5 ppm, 1 ppm, or 0.1 ppm of one or more of Mogroside IIIE, 11-oxo-Mogroside IIIE, Mogroside IIIA2, Mogroside IE, Mogroside IIE, and 11-oxo-mogrol. In some embodiments, the concentration of Mogroside IIIE is less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, 100 ppm, 50 ppm, 30 ppm, 20 ppm, 10 ppm, 5 ppm, 1 ppm, or 0.1 ppm. In some embodiments, the concentration of 11-oxo-Mogroside IIIE is less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, 100 ppm, 50 ppm, 30 ppm, 20 ppm, 10 ppm, 5 ppm, 1 ppm, or 0.1 ppm. In some embodiments, the concentration of 11-oxo-mogrol is less than 5%, 3%, 2%, 1%, 0.5%, 0.3%, 0.1%, 800 ppm, 500 ppm, 200 ppm, 100 ppm, 50 ppm, 30 ppm, 20 ppm, 10 ppm, 5 ppm, 1 ppm, or 0.1 ppm.

Methods of Producing Compound 1 and Intermediate Mogroside Compounds

[0166] In some embodiments, Compound 1 is produced by contact of various starting and/or intermediate compounds with one or more enzymes. The contact can be *in vivo* (e.g., in a recombinant cell) or *in vitro*. The starting and intermediate compounds for producing Compound 1 can include, but are not limited to, Mogroside V, Mogroside IIE, Mogroside III_E, Siamenoside I, Mogroside VI isomer, Mogroside II_A, Mogroside IV_E, or Mogroside IV_A.

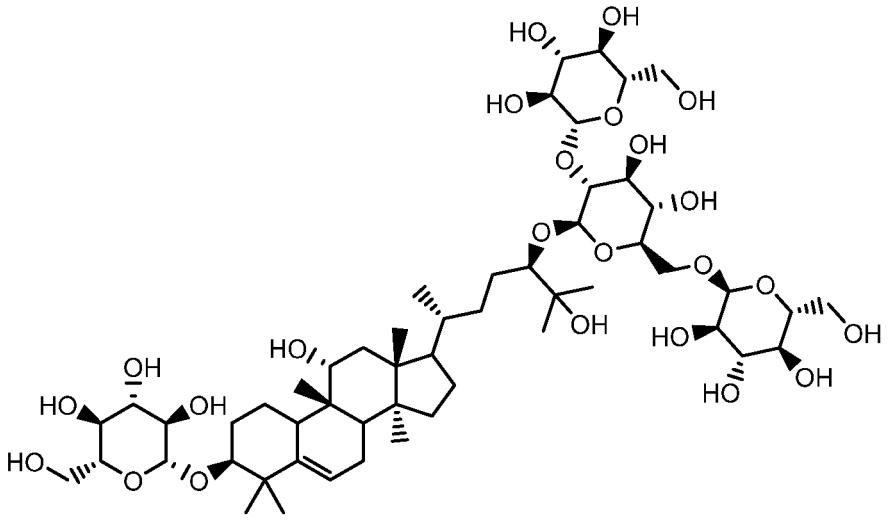
[0167] In some embodiments, Compound 1 as disclosed herein is produced in recombinant host cells *in vivo* as described herein or by modification of these methods. Ways of modifying the methodology include, among others, temperature, solvent, reagents etc., known to those skilled in the art. The methods shown and described herein are illustrative only and are not intended, nor are they to be construed, to limit the scope of the claims in any manner whatsoever. Those skilled in the art will be able to recognize modifications of the disclosed methods and to devise alternate routes based on the disclosures herein; all such modifications and alternate routes are within the scope of the claims.

[0168] In some embodiments, Compound 1 disclosed herein is obtained by purification and/or isolation from a recombinant bacterial cell, yeast cell, plant cell, or insect cell. In some embodiments, the recombinant cell is from *Siraitia grosvenorii*. In some such embodiments, an extract obtained from *Siraitia grosvenorii* may be fractionated using a suitable purification technique. In some embodiments, the extract is fractionated using HPLC and the appropriate fraction is collected to obtain the desired compound in isolated and purified form.

[0169] In some embodiments, Compound 1 is produced by enzymatic modification of a compound isolated from *Siraitia grosvenorii*. For example, in some embodiments, Compound 1 isolated from *Siraitia grosvenorii* is contacted with one or more enzymes to obtain the desired compounds. The contact can be *in vivo* (e.g., in a recombinant cell) or *in vitro*. The starting and intermediate compounds for producing Compound 1 can include, but are not limited to, Mogroside V, Mogroside IIE, Mogroside III_E, Siamenoside I, Mogroside VI isomer, Mogroside II_A, Mogroside IV_E, or Mogroside IV_A. One or more of these compounds can be made *in vivo*. Enzymes suitable for use to generate compounds described herein can include, but are not limited to, a pectinase, a β-galactosidase (e.g., Aromase), a cellulase (e.g., Celluclast), a

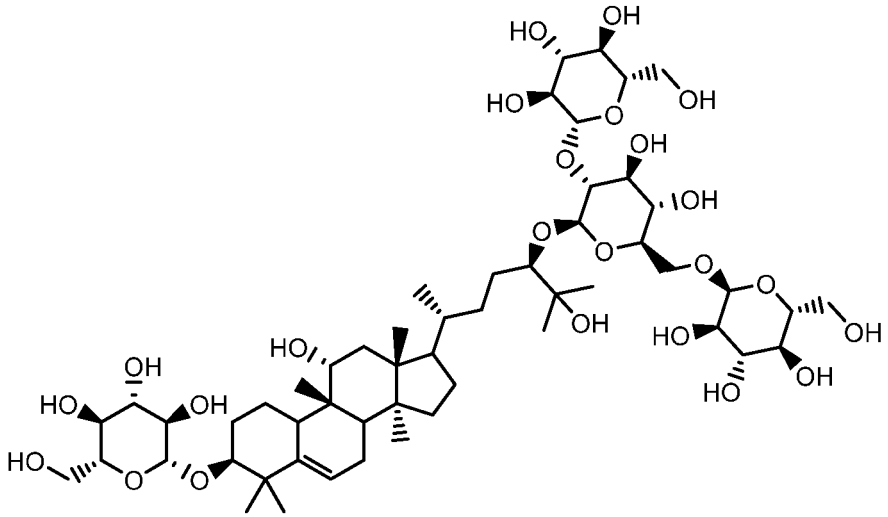
clyclomatlodextrin glucanotransferase (e.g., Toruzyme), an invertase, a glucostransferase (e.g., UGT76G1), a dextransucrase, a lactase, an arabanse, a xylanase, a hemicellulose, an amylase, or a combination thereof. In some embodiments, the enzyme is a Toruzyme comprises an amino acid sequence set forth in any one of SEQ ID NO: 89-94.

[0170] Some embodiments provide a method of making Compound 1,



(1), the method comprises fractionating an extract of *Siraitia grosvenorii* on an HPLC column and collecting an eluted fraction comprising Compound 1.

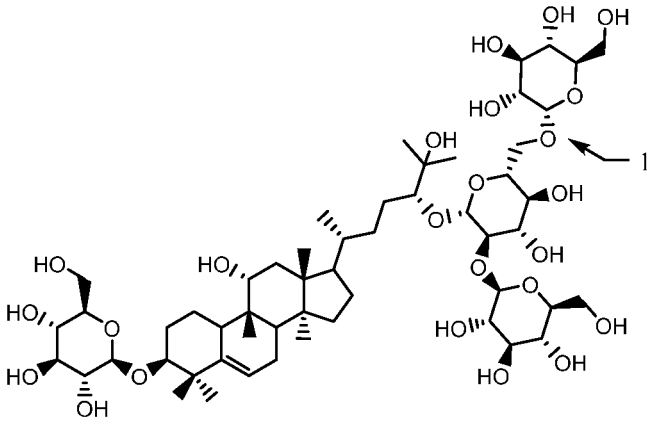
[0171] Some embodiments provide a method of making Compound 1,



, wherein the method comprises treating Mogroside III_E with the glucose transferase enzyme UGT76G1. In some

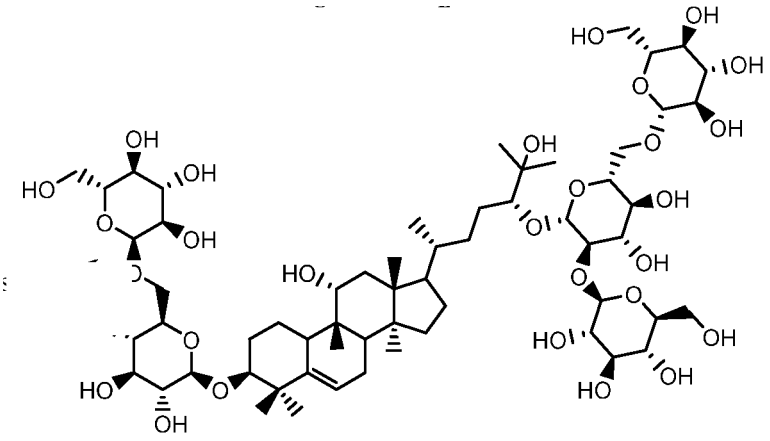
embodiments, UGT76G1 is encoded by a sequence set forth in SEQ ID NO: 440. In some embodiments, UGT76G1 comprises an amino acid sequence set forth in SEQ ID NO: 439.

[0172] Various mogroside compounds can be used as intermediate compounds for producing Compound 1. A non-limiting example of such mogroside compounds is Compound 3 having the structure of:



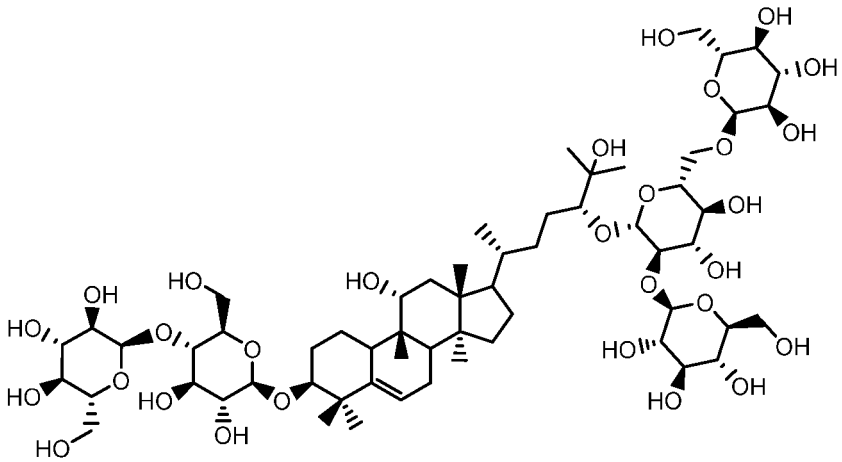
(**3**). In some embodiments, a method for producing Compound 3 comprises contacting mogroside III_E with a cell (e.g., a recombinant host cell) that expresses one or more cyclomaltodextrin glucanotransferases. In some embodiments, the cyclomaltodextrin glucanotransferase comprises an amino acid sequence set forth in SEQ ID NO: 95.

[0173] Various mogroside compounds can be used as intermediate compounds for producing Compound 1. One non-limiting example of such mogroside compounds is Compound 12 having the structure of:



(**12**). In some embodiments, a method for producing Compound 12 comprises contacting mogroside VI with a cell (e.g., a recombinant host cell) that expresses one or more invertase.

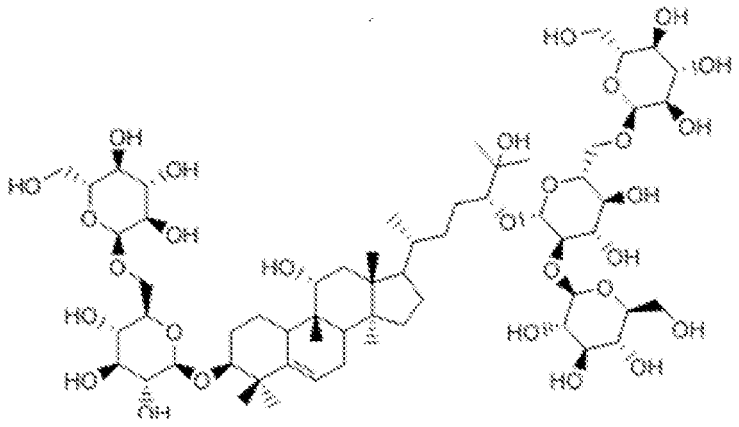
[0174] Various mogroside compounds can be used as intermediate compounds for producing Compound 1. One non-limiting example of such mogroside compounds is Compound 5 having the structure of:



(**5**),

In some embodiments, a method for producing Compound 5 comprises contacting mogroside III_E with a cell (e.g., a recombinant host cell) that expresses one or more cyclomaltodextrin glucanotransferase. In some embodiments, the method is performed in the presence of starch.

[0175] Various mogroside compounds can be used as intermediate compounds for producing Compound 1. One non-limiting example of such mogroside compounds is Compound 4 having the structure of:



(**4**),

In some embodiments, a method for producing Compound 4 comprises contacting mogroside III_E with a cell (e.g., a recombinant host cell) that expresses one or more cyclomaltodextrin glucanotransferase. In some embodiments, the method is performed in the presence of starch.

Hydrolysis of Hyper-glycosylated Mogrosides

[0176] In some embodiments, one or more hyper-glycosylated mogrosides are hydrolyzed to Mogroside III_E by contact with one or more hydrolase enzymes. In some embodiments, the hyper-glycosylated mogrosides are selected from a mogroside IV, a mogroside V, a mogroside VI, and combinations thereof. In some embodiments, the hyper-glycosylated mogrosides are selected from Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside IV_E, and combinations thereof.

[0177] It has been surprisingly discovered that Compound 1 displays tolerance to hydrolysis by certain hydrolyzing enzymes, even though such enzymes display capabilities of hydrolyzing hyper-glycosylated mogrosides to Mogroside IIIE. The alpha-linked glycoside present in Compound 1 provides a unique advantage over other mogrosides (e.g., beta-linked glycosides) due to its tolerance to hydrolysis. In some embodiments, during microbial production of Compound 1, the microbial host will hydrolyze unwanted beta-linked mogrosides back to Mogroside IIIE. This will improve the purity of Compound 1 due to the following: 1) Reduction of unwanted Mogroside VI, Mogroside V, and Mogroside IV levels, 2) The hydrolysis will increase the amount of Mogroside IIIE available to be used as a precursor for production of Compound 1.

[0178] Figure 38 illustrates the production of hyper-glycosylated mogrosides through glycosolation enzymes, which may then be hydrolyzed back to Mogroside IIIE. The result is a mixture of mogrosides with a lower than desirable yield of hyper-glycosylated mogrosides. The hydrolase enzymes can be removed, but a mixture of mogrosides are still obtained and the lifespan of the producing organism may be reduced. However, because Compound 1 is resistant to hydrolysis, the hydrolysis can be used to drive hyper-glycosylated mogrosides to Mogroside IIIE, which can then be converted to Compound 1 (as shown in Figure 39).

[0179] In some embodiments, the hydrolase is a β-glucan hydrolase. In some embodiments, the hydrolase is EXG1. The EXG1 protein can comprise an amino acid sequence having at least 70%, 80%, 90%, 95%, 98%, 99%, or more sequence identity to SEQ ID NO: 1013 or 1014. In some embodiments, the EXG1 protein comprises, or consists of, an amino acid sequence set forth in SEQ ID NO: 1013 or 1014. In some embodiments, the hydrolase is EXG2. The EXG2 protein can comprise an amino acid sequence having at least 70%, 80%, 90%, 95%, 98%, 99%, or more sequence identity to SEQ ID NO: 1023. In some embodiments, the EXG2 protein comprises, or consists of, an amino acid sequence set forth in SEQ ID NO: 1023. The hydrolase can be, for example, any one of the hydrolases disclosed herein.

Production of Compound 1 from Mogroside IIIE

[0180] Compound 1 can be produced from Mogroside IIIE by contact with one or more enzymes capable of converting Mogroside IIIE to Compound 1. In some embodiments, the enzyme capable of catalyzing production of Compound 1 is one or more of UDP glycosyltransferases, cyclomaltodextrin glucanotransferases (CGTases), glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0181] In some embodiments, the enzyme capable of catalyzing the production of Compound 1 is a CGTase. In some embodiments, the CGTase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of any one of SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the CGTase comprises, or consists of, the amino acid sequence of SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the enzyme capable of catalyzing the production of Compound 1 is a dextransucrase. In some embodiments, the dextransucrase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity y to the sequence of any one of SEQ ID NOs: 2, 103, 106-110, 156 and 896. In some embodiments, the dextransucrase comprises, or consists of, an amino acid sequence of any one of SEQ ID NOs: 2, 103, 106-110, 156 and 896. In some embodiments, the dextransucrase is encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between

any two of these numbers, sequence identity to any one of SEQ ID NOs: 104, 105, 157, 158, and 895. In some embodiments, the dextransucrase is encoded by a nucleic acid sequence comprising, or consisting of, any one of SEQ ID NOs: 104, 105, 157, 158, and 895. In some embodiments, the enzyme capable of catalyzing the production of Compound 1 is a transglucosidase. In some embodiments, the transglucosidase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 163-290 and 723. In some embodiments, the transglucosidase comprises, or consists of, an amino acid sequence of any one of SEQ ID NOs: 163-292 and 723. Parameters for determining the percent sequence identity can be performed with ClustalW software of by Blast searched (ncbi.nih.gov). The use of these programs can determine conservation between protein homologues.

[0182] In some embodiments, the enzyme capable of catalyzing the production of Compound 1 is a uridine diphosphate-glucosyl transferase (UGT). The UGT can comprise, for example, an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 4-9, 15-19, 125, 126, 128, 129, 293-307, 407, 409, 411, 413, 439, 441, and 444. In some embodiments, UGT comprises, or consists of, the amino acid sequence of any one of SEQ ID NOs: 4-9, 10-14, 125, 126, 128, 129, 293-304, 306, 407, 409, 411, 413, 439, 441, and 444. In some embodiments, the UGT is encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to UGT1495 (SEQ ID NO: 10), UGT1817 (SEQ ID NO: 11), UGT5914 (SEQ ID NO: 12), UGT8468 (SEQ ID NO:13), UGT10391 (SEQ ID NO:14), and SEQ ID NOs: 116-124, 127, 130, 408, 410, 412, 414, 440, 442, 443, and 445. In some embodiments, the UGT is encoded by a nucleic acid sequence comprising, consisting of, any one of the nucleic acid sequence of UGT1495 (SEQ ID NO: 10), UGT1817(SEQ ID NO: 11), UGT5914 (SEQ ID NO: 12), UGT8468 (SEQ ID NO:13), UGT10391 (SEQ ID NO:14), SEQ ID NOs: 116-124, 127, 130, 408, 410, 412, 414, 440, 442, 443, and 445. In some embodiments, the enzyme can be UGT98 or UGT SK98. For example, as described herein, a recombinant host cell capable of producing Compound 1 can comprises a third gene encoding UGT98 and/or UGT SK98. In some embodiments, the UGT98 or UGT SK98 comprises an amino acid sequence having at least,

70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 9 or 16. In some embodiments, the UGT comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to UGT73C3 (SEQ ID NO: 4), UGT73C6 (SEQ ID NO: 5), UGT85C2 (SEQ ID NO: 6), UGT73C5 (SEQ ID NO: 7), UGT73E1 (SEQ ID NO: 8), UGT98 (SEQ ID NO: 9), UGT1576 (SEQ ID NO: 15), UGT SK98 (SEQ ID NO: 16), UGT430 (SEQ ID NO: 17), UGT1697 (SEQ ID NO: 18), and UGT11789 (SEQ ID NO: 19). In some embodiments, the UGT comprises, or consists of, an amino acid sequence of any one of UGT73C3 (SEQ ID NO: 4), UGT73C6 (SEQ ID NO: 5), 85C2 (SEQ ID NO: 6), UGT73C5 (SEQ ID NO: 7), UGT73E1 (SEQ ID NO: 8), UGT98 (SEQ ID NO: 9), UGT1576 (SEQ ID NO: 15), UGT SK98 (SEQ ID NO:16), UGT430 (SEQ ID NO:17), UGT1697 (SEQ ID NO: 18), and UGT11789 (SEQ ID NO:19). In some embodiments, the UGT is encoded by a nucleic acid sequence at least 70%, 80%, 90%, 95%, 98%, 99% or more sequence identity to UGT1495 (SEQ ID NO: 10), UGT1817 (SEQ ID NO: 11), UGT5914 (SEQ ID NO: 12), UGT8468 (SEQ ID NO:13) or UGT10391 (SEQ ID NO: 14). In some embodiments, the UGT is encoded by a nucleic acid sequence comprising, or consisting of, any one of the sequences of UGT1495 (SEQ ID NO: 10), UGT1817 (SEQ ID NO: 11), UGT5914 (SEQ ID NO: 12), UGT8468 (SEQ ID NO: 13), and UGT10391 (SEQ ID NO: 14). As disclosed herein, the enzyme capable of catalyzing the production of Compound 1 can comprises an amino acid sequence having at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or more sequence identity to any one of the UGT enzymes disclosed herein. Furthermore, a recombinant host cell capable of producing Compound 1 can comprises an enzyme comprising, or consisting of a sequence having at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or more sequence identity to any one of the UGT enzymes disclosed herein. In some embodiments, the recombinant host cell comprises an enzyme comprising, or consisting of a sequence of any one of the UGT enzymes disclosed herein.

[0183] In some embodiments, the method of producing Compound 1 comprises treating Mogroside III_E with the glucose transferase enzyme UGT76G1, for example the UGT76G1 of SEQ ID NO: 439 and the UGT76G1 encoded by the nucleic acid sequence of SEQ ID NO: 440.

Enzymes for the production of mogroside compounds and Compound 1

[0184] As described herein, the enzymes of UDP glycosyltransferases, cyclomaltodextrin glucanotransferases (CGTases), glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases can comprise the amino acid sequences described in the table of sequences herein and can also be encoded by the nucleic acid sequences described in the Table of sequences. Additionally the enzymes can also include functional homologues with at least 70% sequence identity to the amino acid sequences described in the table of sequences. Parameters for determining the percent sequence identity can be performed with ClustalW software of by Blast searched (ncbi.nih.gov). The use of these programs can determine conservation between protein homologues.

[0185] In some embodiments, the transglucosidases comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 163-290 and 723. In some embodiments, the CGTase comprises, or consists of, an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 1, 3, 78-101, and 154. In some embodiments, the transglucosidases comprise an amino acid sequence or is encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 163-290 and 723.

[0186] The methods herein also include incorporating genes into the recombinant cells for producing intermediates such as pyruvate, acetyl-coa, citrate, and other TCA intermediates (Citric acid cycle). Intermediates can be further used to produce mogroside compounds for producing Compound 1. Methods for increasing squalene content are described in Gruchattka *et al.* and Rodriguez *et al.* (PLoS One. 2015 Dec 23; 10(12; Microb Cell Fact. 2016 Mar 3; 15:48; incorporated by reference in their entireties herein).

[0187] Expression of enzymes to produce oxidosqualene and diepoxysqualene are further contemplated. The use of enzymes to produce oxidosqualene and diepoxysqualene can be used to boost squalene synthesis by the way of squalene synthase and/or squalene epoxidase. For example, Su *et al.* describe the gene encoding SgSQS, a 417 amino acid protein from Siraitia

grosvenorii for squalene synthase (Biotechnol Lett. 2017 Mar 28; incorporated by reference in its entirety herein). Genetically engineering the recombinant cell for expression of HMG CoA reductase is also useful for squalene synthesis (Appl Environ Microbiol. 1997 Sep; 63(9):3341-4.; Front Plant Sci. 2011 Jun 30; 2:25; FEBS J. 2008 Apr; 275(8):1852-9.; all incorporated by reference in their entireties herein. In some embodiments, the 2, 3-oxidosqualene or diepoxysqualene is produced by an enzyme comprising a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 898 or 900. In some embodiments, the 2,3-oxidosqualene or diepoxysqualene is produced by an enzyme encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 897 or 899.

[0188] Expression of enzymes to produce cucurbitadienol/epoxycucurbitadienol are also contemplated. Examples of curubitadienol synthases from C pepo, S grosvenorii, C sativus, C melo, C moschata, and C maxim are contemplated for engineering into the recombinant cells by a vector for expression. Oxidosqualene cyclases for titerpene biosynthesis is also contemplated for expression in the recombinant cell, which would lead to the cyclization of an acyclic substrate into various polycyclic triterpenes which can also be used as intermediates for the production of Compound 1 (Org Biomol Chem. 2015 Jul 14;13(26):7331-6; incorporated by reference in its entirety herein).

[0189] Expression of enzymes that display epoxide hydrolase activities to make hydroxy-cucurbitadienols are also contemplated. In some embodiments herein, the recombinant cells for the production of Compound 1 further comprises genes that encode enzymes that display epoxide hydrolase activities to make hydroxy-cucurbitadienols are provided. Such enzymes are provided in Itkin *et al.* which is incorporated by reference in its entirety herein. The enzymes described in Itkin *et al.* are provided in Table 1, table of sequences, provided herein. Ikin *et al.,* also describes enzymes for making key mogrosides, UGS families, glycosyltransferases and hydrolases that can be genetically modified for reverse reactions such as glycosylations.

[0190] The expression of enzymes in recombinant cells to that hydroxylate mogroside compounds to produce mogrol are also contemplated. These enzymes can include

proteins of the CAZY family, UDP glycosyltransferases, CGTases, Glycotransferases, Dextransucrases, Cellulases, B-glucosidases, Transglucosidases, Pectinases, Dextranases, yeast and fungal hydrolyzing enzymes. Such enzymes can be used for example for hydrolyzing Mogroside V to Mogroside IIIE, in which Mogroside IIIE can be further processed to produce Compound 1, for example *in vivo*. In some embodiments, fungal lactases comprise an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to anyone of SEQ ID NO: 678-722.

[0191] In some embodiments, a mogrol precursor such as squalene or oxidosqualine, mogrol or mogroside is produced. The mogrol precursor can be used as a precursor in the production of Compound 1. Squalene can be produced from famesyl pyrophosphate using a squalene synthase, and oxidosqualene can be produced from squalene using a squalene epoxidase. The squalene synthase can be, for example, squalene synthase from Gynostemma pentaphyllum (protein accession number C4P9M2), a cucurbitaceae family plant. The squalene synthase can also comprise a squalene synthase from *Arabidopsis thaliana* (protein accession number C4P9M3), *Brassica napus*, *Citrus macrophylla*, *Euphorbia tirucalli* (protein accession number B9WZW7), *Glycine max, Glycyrrhiza glabra* (protein accession number Q42760, Q42761), *Glycrrhiza uralensis* (protein accession number D6QX40, D6QX41 , D6QX42, D6QX43, D6QX44, D6QX45, D6QX47, D6QX39, D6QX55, D6QX38, D6QX53, D6QX37, D6QX35, B5AID5, B5AID4, B5AID3, C7EDD0, C6KE07, C6KE08, C7EDC9), *Lotusjaponicas* (protein accession number Q84LE3), *Medicago truncatula* (protein accession number Q8GSL6), *Pisum sativum, Ricinus communis* (protein accession number B9RHC3). Various squalene synthases have described in WO 2016/050890, the content of which is incorporated herein by reference in its entirety.

Recombinant host cells

[0192] Any one of the enzymes disclosed herein can be produced *in vitro*, *ex vivo*, or *in vivo*. For example, a nucleic acid sequence encoding the enzyme (including but not limited to any one of UGTs, CGTases, glycotransferases, dextransucrases, celluases, beta-glucosidases, amylases, transglucosidases, pectinases, dextranases, cytochrome P450, epoxide hydrolases, cucurbitadienol synthases, squalene epoxidases, squalene synthases, hydrolases, and

oxidosqualene cyclases) can introduced to a host recombinant cell, for example in the form of an expression vector containing the coding nucleic acid sequence, *in vivo*. The expression vectors can be introduced into the host cell by, for example, standard transformation techniques (e.g., heat transformation) or by transfection. The expression systems can produce the enzymes for mogroside and Compound 1 production, in order to produce Compound 1 in the cell *in vivo*. Useful expression systems include, but are not limited to, bacterial, yeast and insect cell systems. For example, insect cell systems can be infected with a recombinant virus expression system for expression of the enzymes of interest. In some embodiments, the genes are codon optimized for expression in a particular cell. In some embodiments, the genes are operably linked to a promoter to drive transcription and translation of the enzyme protein. As described herein, codon optimization can be obtained, and the optimized sequence can then be engineered into a vector for transforming a recombinant host cell.

[0193] Expression vectors can further comprise transcription or translation regulatory sequences, coding sequences for transcription or translation factors, or various promoters (e.g., GPD1 promoters) and/or enhancers, to promote transcription of a gene of interest in yeast cells.

[0194] The recombinant cells as described herein are, in some embodiments, genetically modified to produce Compound 1 *in vivo*. Additionally, a cell can be fed a mogrol precursor or mogroside precursor during cell growth or after cell growth to boost rate of the production of a particular intermediate for the pathway for producing Compound 1 *in vivo*. The cell can be in suspension or immobilized. The cell can be in fermentation broth or in a reaction buffer. In some embodiments, a permeabilizing agent is used for transfer of a mogrol precursor or mogroside precursor into a cell. In some embodiments, a mogrol precursor or mogroside precursor can be provided in a purified form or as part of a composition or an extract.

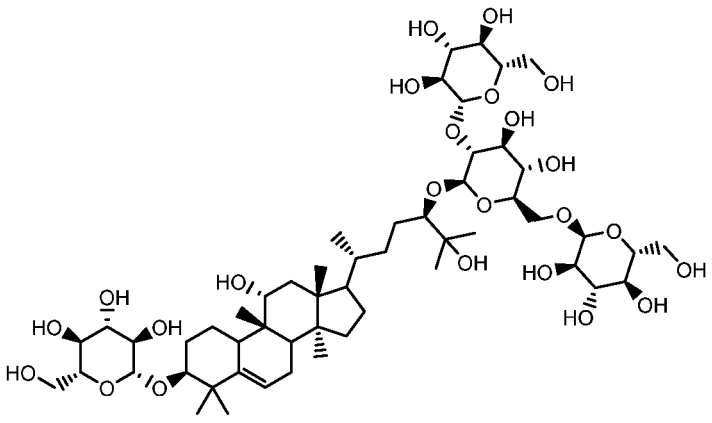
[0195] The recombinant host cell can be, for example a plant, bivalve, fish, fungus, bacteria or mammalian cell. For example, the plant can be selected from Siraitia, Momordica, Gynostemma, Cucurbita, Cucumis, Arabidopsis, Artemisia, Stevia, Panax, Withania, Euphorbia, Medicago, Chlorophytum, Eleutherococcus, Aralia, Morus, Medicago, Betula, Astragalus, Jatropha, Camellia, Hypholoma, Aspergillus, Solanum, Huperzia, Pseudostellaria, Corchorus, Hedera, Marchantia, and Morus. The fungus can be selected from Trichophyton, Sanghuangporus, Taiwanofungus, Moniliophthora, Marssonina, Diplodia, Lentinula,

Xanthophyllomyces, Pochonia, Colletotrichum, Diaporthe, Histoplasma, Coccidioides, Histoplasma, Sanghuangporus, Aureobasidium, Pochonia, Penicillium, Sporothrix, Metarhizium, Aspergillus, Yarrowia, and Lipomyces. In some embodiments, the fungus is *Aspergillus nidulans, Yarrowia lipolytica,* or *Rhodosporin toruloides.* In some embodiments, the recombinant host cell is a yeast cell. In some embodiments, the yeast is selected from Candida, Sacccharaomyces, Saccharomycotina, Taphrinomycotina, Schizosaccharomycetes, Komagataella, Basidiomycota, Agaricomycotina, Tremellomycetes, Pucciniomycotina, Aureobasidium, Coniochaeta, Rhodosporidium, Yarrowia, and Microboryomycetes. In some embodiments, the bacteria is selected from Frankia, Actinobacteria, Streptomyces, and Enterococcus. In some embodiments, the bacteria is *Enterococcus faecalis.*

[0196] In some embodiments, the recombinant genes are codon optimized for expression in a bacterial, mammalian, plant, fungal or insect cell. In some embodiments, one or more of genes comprises a functional mutation to increased activity of the encoded enzyme. In some embodiments, cultivating the recombinant host cell comprises monitoring the cultivating for pH, dissolved oxygen level, nitrogen level, or a combination thereof of the cultivating conditions.

Producing Mogrol from Squalene

[0197] Some embodiments of the method of producing Compound 1 comprises producing an intermediate for use in the production of Compound 1. The compound having the structure of:



(**1**) is produced *in vivo* in a recombinant host. In some embodiments, the compound is in the recombinant host cell, is

secreted into the medium in which the recombinant cell is growing, or both. In some embodiments, the recombinant cell further produces intermediates such as mogroside compounds in vivo. The recombinant cell can be grown in a culture medium, under conditions in which the genes disclosed herein are expressed. Some embodiments of methods of growing the cell are described herein.

[0198] In some embodiments, the intermediate is, or comprises, at least one of squalene, oxidosqualene, curubitadienol, mogrol and mogrosides. In some embodiments, the mogroside is Mogroside IIE. As described herein, mogrosides are a family of glycosides that can be naturally isolated from a plant or a fruit, for example. As contemplated herein, the mogrosides can be produced by a recombinant host cell.

[0199] In some alternatives of the methods described herein, the recombinant host cell comprises a polynucleotide or a sequence comprising one or more of the following:

a gene encoding squalene epoxidase;

a gene encoding cucurbitadienol synthase;

a gene encoding cytochrome P450;

a gene encoding cytochrome P450 reductase; and

a gene encoding epoxide hydrolase.

[0200] In some embodiments, the squalene epoxidase comprises a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 54. In some embodiments, the squalene epoxidase comprises a sequence from *Arabidopsis thaliana* (the protein accession numbers : Q9SM02, 065403, 065402, 065404, 081000, or Q9T064), *Brassica napus* (protein accession number 10 065727, 065726), *Euphorbia tirucalli* (protein accession number A7VJN1), *Medicago truncatula* (protein accession number Q8GSM8, Q8GSM9), *Pisum sativum,* and *Ricinus communis* (protein accession number B9R6VO, B9S7W5, B9S6Y2, B9TOY3, B9S7TO, B9SX91) and functional homologues of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. In some embodiments, the squalene epoxidase comprises, or consists of an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NOs: 50-56, 60, 61, 334 or 335.

[0201] In some embodiments, the cell comprises genes encoding ERG7 (lanosterol synthase). In some embodiments, lanosterol synthase comprises a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 111. In some embodiments, the P450 polypeptide is encoded in genes comprising a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of Claims: 31-48. In some embodiments, the sequences can be separated by ribosome skip sequences to produce separated proteins.

[0202] In some embodiments, the recombinatnt host cell comprises a gene encoding a polypeptide having cucurbitadienol synthase activity. In some embodiments, the polypeptide having cucurbitadienol synthase activity comprises an amino acid sequence as set forth in any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904, and 906. In some embodiments, the polypeptide having cucurbitadienol synthase activity comprises a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of any one of SEQ ID NO: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904 and 906. In some embodiments, the polypeptide having cucurbitadienol synthase activity comprises a C-terminal portion comprising the sequence set forth in SEQ ID NO: 73. In some embodiments, the gene encoding the polypeptide having cucurbitadienol synthase activity is codon optimized. In some embodiments, the codon optimized gene comprises the nucleic acid sequence set forth in SEQ ID NO: 74.

[0203] In some embodiments, the polypeptide having cucurbitadienol synthase activity is a fusion polypeptide comprising a fusion domain fused to a cucurbitadienol synthase. The fusion domain can be fused to, for example, N-terminus or C-terminus of a cucurbitadienol synthase. The fusion domain can be located, for example, at the N-terminal region or the C-terminal region of the fusion polypeptide. The length of the fusion domain can vary. For example, the fusion domain can be, or be about, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 40, 50, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, or a range between any two of these numbers amino acids long. In some embodiments, the fusion domain is 3 to 1000 amino acids long. In some embodiments, the fusion domain is 5 to 50 amino acids long. In some

embodiments, the fusion domain comprises a substantial portion or the entire sequence of a functional protein. In some embodiments, the fusion domain comprises a portion or the entire sequence of a yeast protein. For example, the fusion polypeptide having cucurbitadienol synthase activity can comprise an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 851, 854, 856, 1024, 859, 862, 865, 867, 915, 920, 924, 928, 932, 936, 940, 944, 948, 952, 956, 959, 964, 967, 971, 975, 979, 983, 987, 991, 995, 999, 1003, 1007, and 1011. In some embodiments, the fusion polypeptide comprises, or consists of, an amino acid sequence set forth in any one of SEQ ID NOs: 851, 854, 856, 1024, 859, 862, 865, 867, 915, 920, 924, 928, 932, 936, 940, 944, 948, 952, 956, 959, 964, 967, 971, 975, 979, 983, 987, 991, 995, 999, 1003, 1007, and 1011. In some embodiments, the fusion domain of the fusion polypeptide comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 866, 870, 917, 921, 925, 929, 933, 937, 941, 945, 949, 953, 957, 961, 968, 972, 976, 980, 984, 988, 992, 996, 1000, 1004, 1008, and 1012. In some embodiments, the fusion domain of the fusion polypeptide comprises, or consists of, an amino acid sequence set forth in any one of SEQ ID NOs: 866, 870, 917, 921, 925, 929, 933, 937, 941, 945, 949, 953, 957, 961, 968, 972, 976, 980, 984, 988, 992, 996, 1000, 1004, 1008, and 1012. In some embodiments, the cucurbitadienol synthase fused with the fusion domain comprises an amino acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327, 329-333, 420, 422, 424, 426, 446, 902, 904, and 906. In some embodiments, the cucurbitadienol synthase fused with the fusion domain comprises, or consists of, an amino acid sequence set forth in any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327, 329-333, 420, 422, 424, 426, 446, 902, 904, and 906. In some embodiments, the cucurbitadienol synthase fused with the fusion domain is encoded by a gene comprising a nucleic acid sequence having, or having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903, and 905. In some embodiments, the cucurbitadienol synthase fused with the fusion domain is encoded by a gene comprising, or consists of, a nucleic acid sequence set forth in any one of SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903, and 905. Disclosed herein

include a recombinant nucleic acid molecule comprising a nucleic acid sequence encoding a fusion polypeptide having cucurbitadienol synthase activity. Also disclosed include a recombinant cell comprising a fusion polypeptide having cucurbitadienol synthase activity or a recombinant nucleic acid molecule encoding the fusion polypeptide.

[0204] The fusion polypeptides having cucurbitadienol synthase activity disclosed herein can be used to catalyze enzymatic reactions as cucurbitadienol synthases. For example, a substrate for cucurbitadienol synthase can be contacted with one or more of these fusion polypeptide to produce reaction products. Non-limiting examples of the reaction product include curcurbitadienol, 24,25-epoxy curcurbitadienol, and any combination thereof. Non-limiting examples of the substrate for cucurbitadienol synthase include 2,3-oxidosqualene, dioxidosqualene, diepoxysqualene, and any combination thereof. In some embodiments, the substrate can be contacted with a recombinant host cell which comprises a nucleic acid sequence encoding one or more fusion polypeptides having cucurbitadienol synthase activity. The substrate can be provided to the recombinant host cells, present in the recombinant host cell, produced by the recombinant host cell, or any combination thereof.

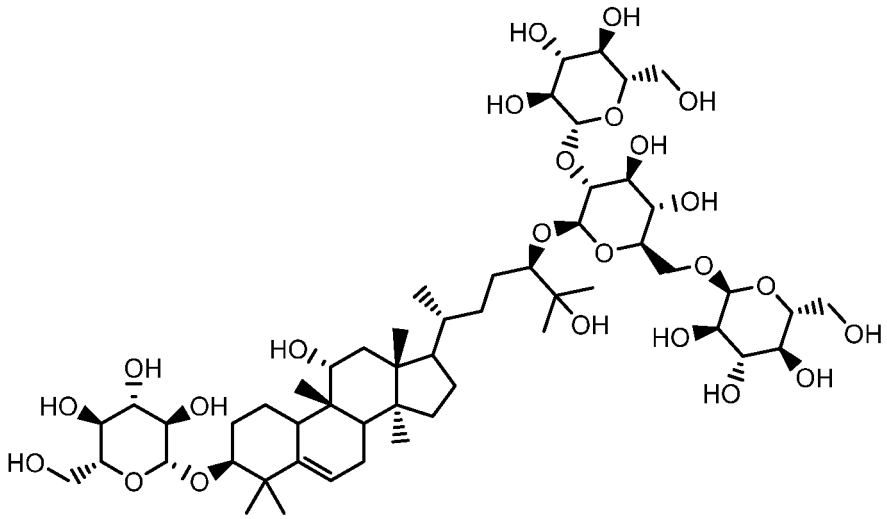
[0205] In some embodiments, the cytochrome P450 is a CYP5491. In some embodiments, the cytochrome P540 comprises an amino acid sequence having, or having at least, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence set forth in SEQ ID NO: 44 and/or SEQ ID NO:74. In some embodiments, the P450 reductase polypeptide comprises an amino acid sequence having, or having at least, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 46. In some embodiments, the P450 polypeptide is encoded by a gene comprising a sequence having, or having at least, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 100% or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, and 891.

[0206] In some embodiments, the epoxide hydrolase comprises an sequence having, or having at least, 70%, 80%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 38 or 40. In some embodiments, the epoxide hydrolase comprises, or consists of, the sequence set forth in SEQ ID NO: 38 or 40.

Some methods of producing squalene for mogrol production

[0207] Squalene is a natural 30 carbon organic molecule that can be produced in plants and animals and is a biochemical precursor to the family of steroids. Additionally, squalene can be used as precursor in mogrol syntheses *in vivo* in a host recombinant cell. Oxidation (via squalene monooxygenase) of one of the terminal double bonds of squalene yields 2,3-squalene oxide, which undergoes enzyme-catalyzed cyclization to afford lanosterol, which is then elaborated into cholesterol and other steroids. As described in Gruchattka *et al.* ("In Vivo Validation of In Silico Predicted Metabolic Engineering Strategies in Yeast: Disruption of α-Ketoglutarate Dehydrogenase and Expression of ATP-Citrate Lyase for Terpenoid Production." PLOS ONE December 23, 2015; incorporated by reference in its entirety herein), synthesis of squalene can occur initially from precursors of the glycolysis cycle to produce squalene. Squalene in turn can be upregulated by the overexpression of ATP-citrate lyase to increase the production of squalene. Some embodiments disclosed herein include enzymes for producing squalene and/or boosting the production of squalene in recombinant host cells, for example recombinant yeast cells. ATP citrate lyase can also mediate acetyl CoA synthesis which can be used for squalene and mevalonate production, which was seen in yeast, *S. cerevisiae* (Rodrigues *et al.* "ATP citrate lyase mediated cytosolic acetyl-CoA biosynthesis increases mevalonate production in *Saccharomyces cerevisiae*" Microb Cell Fact. 2016; 15: 48.; incorporated by reference in its entirety). On example of the gene encoding an enzyme for mediating the acetyl CoA synthesis is set forth in SEQ ID NO: 130. In some embodiments herein, the recombinant cell comprises sequences for mediating acetyl CoA synthesis.

[0208] Some embodiments disclosed herein provide methods for producing Compound 1 having the structure of:



(**1**).

[0209] In some embodiments, the methods further comprises producing intermediates in the pathway for the production of compound 1 *in vivo*. In some embodiments, the recombinant host cell that produces Compound 1 comprises at least one enzyme capable for converting dioxidosqualene to produce 24,25 epoxy cucurbitadienol, converting oxidosqualene to cucurbitadienol, catalyzing the hydroxylation of 24,25, epoxy cucurbitadienol to 11-hydroxy-24,25 epoxy cucurbitadienol, enzyme for catalyzing the hydroxylation of cucurbitadienol to 11-hydroxy-cucurbitadienol, enzyme for the epoxidation of cucurbitadienol to 24,25 epoxy cucurbitadienol, enzymes capable of catalyzing epoxidation of 11-hydroxy-cucurbitadienol to produce 11-hydroxy-24,25 epoxy cucurbitadienol, enzymes for the conversion of 11-hydroxy-cucurbtadienol to 11-hydroxy-24,25 epoxy cucurbitadienol, enzymes for catalyzing the conversion of 11-hydroxy-24,25 epoxy cucurbitadienol to produce mogrol and/or enzymes for catalyzing the glycosylation of a mogroside precursor to produce a mogroside compound. In some embodiments, the enzyme for glycosylation is encoded by a sequence set forth in any one of SEQ ID NOs: 121, 122, 123, and 124.

[0210] In some embodiments, the enzyme for catalyzing the hydroxylation of 24,25 epoxy cucurbitadienol to form 11-hydroxy-24,25 epoxy cucurbitadinol is CYP5491. In some embodiments, the CYP5491 comprises a sequence set forth in SEQ ID NO: 49. In some embodiments, the squalene epoxidase comprises a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of SEQ ID NO: 54.

[0211] In some embodiments, the enzyme capable of epoxidation of 11-hydroxycucurbitadientol comprises an amino acid sequence set forth in SEQ ID NO: 74.

[0212] In some alternatives, the recombinant cell comprises genes for expression of enzymes capable of converting dioxidosqualent to 24,25 epoxy cucurbitadienol, converting oxidosqualene to cucurbitadienol, hydroxylation of 24,25 epoxy cucurbitadienol to 11-hydroxy-24,25 epoxy cucurbitadienol, hydroxylation of cucurbitadienol to produce 11-hydroxy-cucurbitadienol, epoxidation of cucurbitadienol to produce 24,25 epoxy cucurbitadienol, and/or epoxidation of 11-hydroxycucurbitadienol to produce 11-hydroxy-24,25 epoxy cucurbitadienol. In these embodiments herein, the intermediates and mogrosides are produced *in vivo*.

[0213] In some embodiments, a method of producing Compound 1 further comprises producing one or more of mogroside compounds and intermediates, such as oxidosqualene, dixidosqualene, cucurbitdienol, 24,25 epoxy cucurbitadienol, 11-hydrosy-cucurbitadienol, 11-hydroxy 24,25 epoxy cucurbitadienol, mogrol, and mogroside compounds.

Methods for the production of Mogroside compounds

[0214] Described herein include methods of producing a mogroside compound, for example, one of the mogroside compounds described in WO2014086842 (incorporated by reference in its entirety herein). The mogroside compound can be used as an intermediate by a cell to further produce Compound 1 disclosed herein.

[0215] Recombinant hosts such as microorganisms, plant cells, or plants can be used to express polypeptides useful for the biosynthesis of mogrol (the triterpene core) and various mogrol glycosides (mogrosides).

[0216] In some embodiments, the production method can comprise one or more of the following steps in any orders:

(1) enhancing levels of oxido-squalene

(2) enhanicing levels of dioxido-squalene

(3) Oxido-squalene -> cucurbitadienol

(4) Dioxido-squalene -> 24,25 epoxy cucurbitadienol

(5) Cucurbitadienol -> 11-hydroxy-cucurbitadienol

(6) 24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol

(7) 11-hydroxy-cucurbitadienol -> mogrol

(8) 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol

(9) mogrol -> various mogroside compounds.

[0217] In the embodiments herein, the oxido-squalene, dioxido-squalene, cucurbitadienol, 24,25 epoxy cucurbitadienol or mogrol may be also produced by the recombinant cell. The method can include growing the recombinant microorganism in a culture medium under conditions in which one or more of the enzymes catalyzing step(s) of the methods of the invention, e.g. synthases, hydrolases, CYP450s and/or UGTs are expressed. The recombinant microorganism may be grown in a fed batch or continuous process. Typically, the recombinant microorganism is grown in a fermenter at a defined temperature(s) for a desired period of time in order to increase the yield of Compound 1.

[0218] In some embodiments, mogroside compounds can be produced using whole cells that are fed raw materials that contain precursor molecules to increase the yield of Compound 1. 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 in fermentation broth or in a reaction buffer.

[0219] In some embodiments, the recombinant host cell can comprise heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalyzing Oxido-squalene to cucurbitadienol, Cucurbitadienol to11-hydroxycucurbitadienol, 11-hydroxy-cucurbitadienol to mogrol, and/or mogrol to mogroside. In some embodiments, the cell can further comprise Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalyzing Dioxido-squalene to 24,25 epoxy cucurbitadienol, 24,25 epoxy cucurbitadienol to hydroxy-24,25 epoxy cucurbitadienol, 11-hydroxy-24,25 epoxy cucurbitadienol to mogrol, and/or mogrol to mogroside

[0220] The host cell can comprises a recombinant gene encoding a cucurbitadienol synthase and/or a recombinant gene encoding a cytochrome P450 polypeptide.

[0221] In some embodiments, the cell comprises a protein having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904, and 906 (curcurbitadienol synthase).

[0222] In some embodiments, the conversion of Oxido-squalene to cucurbitadienol is catalyzed by cucurbitadienol synthase of any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904, and 906, or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. In some embodiments, the cucurbitadienol synthase polypeptide comprises a C–terminal portion comprising the sequence set forth in SEQ ID NO: 73. In some embodiments, the gene encoding the cucurbitadienol synthase polypeptide is codon optimized. In some embodiments, the codon optimized gene comprises the nucleic acid sequence set forth in SEQ ID NO: 74.

[0223] In some embodiments, the conversion of Cucurbitadienol to 11-hydroxy-cucurbitadienol is catalyzed CYP5491 of SEQ ID NO: 49 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

[0224] In some embodiments, the conversion of 11-hydroxy-cucurbitadienol to mogrol comprises a polypeptide selected from the group consisting of Epoxide hydrolase 1 of SEQ ID NO: 29, Epoxide hydrolase 2 of SEQ ID NO: 30 and functional homologues of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. In some embodiments, the genes encoding epoxide hydrolase 1 and epoxide hydrolase 2 are codon optimized for expression. In some embodiments, the codon optimized genes for epoxide hydrolase comprise a nucleic acid sequence set forth in SEQ ID NO: 114 or 115.

[0225] In some embodiments, the epoxide hydrolase comprises an amino acid sequence as set forth in any one of SEQ ID NOs: 21-28 (Itkin et al, incorporated by reference in its entirety herein).

[0226] In some embodiments, the conversion of mogrol to mogroside is catalyzed in the host recombinant cell by one or more UGTs selected from the group consisting of UGT1576 of SEQ ID NO: 15, UGT98 of SEQ ID NO: 9, UGT SK98 of SEQ ID NO: 68 and functional homologues of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

[0227] In some embodiments, the host recombinant cell comprises a recombinant gene encoding a cytochrome P450 polypeptide is encoded by any one of the sequences in SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, and 891.

[0228] In some embodiments, the host recombinant cell comprises a recombinant gene encoding squalene epoxidase polypeptide comprising the sequence in SEQ ID No: 50.

[0229] In some embodiments, the host recombinant cell comprises a recombinant gene encoding cucurbitadienol synthase polypeptide of any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904, and 906. In some embodiments, the cucurbitadienol synthase polypeptide comprises a C–terminal portion comprising the sequence set forth in SEQ ID NO: 73. In some embodiments, the gene encoding the cucurbitadienol synthase polypeptide is codon optimized. In some embodiments, the codon optimized gene comprises the nucleic acid sequence set forth in SEQ ID NO: 74.

Production of mogroside compounds from mogrol

[0230] In some embodiments, the method of producing Compound 1 comprises contacting mogroside IIIE with a first enzyme capable of catalyzing production of Compound 1 from mogroside IIIE. In some embodiments, the method is performed *in vivo*, wherein a recombinant cell comprises a gene encoding the first enzyme capable of catalyzing production of Compound 1 from mogroside IIIE. In some embodiments, the cell further comprises a gene encoding an enzyme capable of catalyzing production of mogroside IE1 from mogrol. In some embodiments, the enzyme comprises a sequence set forth in any one of SEQ ID NOs: 4-8.

[0231] In some embodiments, the cell further comprises enzymes to convert mogroside IIE to mogroside IV, mogroside V, 11-oxo-mogroside V, and siamenoside I. In some embodiments, the enzymes for converting mogroside IIE to mogroside IV, mogroside V, 11-oxo-mogroside V, and siamenoside I are encoded by genes that comprise the nucleic acid sequences set forth in SEQ ID NOs: 9-14 and 116-120. In some embodiments, the method of producing Compound 1 comprises treating Mogroside III_E with the glucose transferase enzyme UGT76G1.

[0232] In some embodiments, the method comprises fractionating lysate from a recombinant cell on an HPLC column and collecting an eluted fraction comprising Compound 1.

[0233] In some embodiments, the method comprises contacting mogroside IIIE with a first enzyme capable of catalyzing production of Compound 1 from mogroside IIIE. In some embodiments, contacting mogroside IIIE with the first enzyme comprises contacting mogroside IIIE with a recombinant host cell that comprises a first gene encoding the first enzyme. In some embodiments, the first gene is heterologous to the recombinant host cell. In some embodiments, the mogroside IIIE contacts with the first enzyme in a recombinant host cell that comprises a first polynucleotide encoding the first enzyme. In some embodiments, the mogroside IIIE is present in the recombinant host cell. In some embodiments, the mogroside IIIE is produced by the recombinant host cell. In some embodiments, the method comprises cultivating the recombinant host cell in a culture medium under conditions in which the first enzyme is expressed. In some embodiments, the first enzyme is one or more of UDP glycosyltransferases, cyclomaltodextrin glucanotransferases (CGTases), glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the first enzyme is a CGTase. In some embodiments, the CGTase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of any one of SEQ ID NOs: 1, 3, 78-101, 148 and 154. In some embodiments, the transglucosidases are encoded by any one of SEQ ID NOs: 163-290 and 723. In some embodiments, the CGTases comprises, or consists of, a sequence set forth in any one of SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the first enzyme is a dextransucrase. In some embodiments, the dextransucrase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of any one of SEQ ID NOs: 2, 103, 106-110, 156 and 896. In some embodiments, the DexT comprises an amino acid sequence any one of SEQ ID NOs: 2, 103, 106-110, 156 and 896. In some embodiments, the DexT comprises a nucleic acid sequence set forth in SEQ ID NO: 104 or 105. In some embodiments, the dextransucrase comprises an amino acid sequence of SEQ ID NO: 2 or 106-110. In some embodiments, the first enzyme is a transglucosidase. In some embodiments, the transglucosidase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of any one of SEQ ID NOs: 3163-291 and 723. In some embodiments, the transglucosidase comprises

an amino acid sequence of SEQ ID NOs: 163-291 and 723. In some embodiments, the transglucosidases are encoded by any one of SEQ ID NOs: 163-291 and 723. In some embodiments, the transglucosidases comprises an amino acid sequence set forth by any one of SEQ ID NOs: 163-290 and 723. In some embodiments, the genes encode a CGTase comprising any one of the sequence set forth in SEQ ID NOs: 1, 3, 78-101, and 154.

[0234] In some embodiments, the method comprises contacting Mogroside IIA with the recombinant host cell to produce mogroside IIIE, wherein the recombinant host cell further comprises a second gene encoding a second enzyme capable of catalyzing production of Mogroside IIIE from Mogroside IIA. In some embodiments, the mogroside IIA is produced by the recombinant host cell. In some embodiments, the second enzyme is one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the second enzyme is a uridine diphosphate-glucosyl transferase (UGT). In some embodiments, the transglucosidases comprises an amino acid sequence set forth by any one of SEQ ID NOs: 163-290 and 723. In some embodiments, the genes encode a CGTase comprising an amino acid sequences set forth in SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the UGT is UGT73C3 (SEQ ID NO: 4), UGT73C6 (SEQ ID NO:5, 444 or 445), 85C2 (SEQ ID NO: 6), UGT73C5 (SEQ ID NO: 7), UGT73E1 (SEQ ID NO: 8), UGT98 (SEQ ID NO: 9 or 407) , UGT1576 (SEQ ID NO: 15), UGT SK98 (SEQ ID NO: 16), UGT430 (SEQ ID NO: 17), UGT1697 (SEQ ID NO: 18), or UGT11789 (SEQ ID NO: 19) or any one of SEQ ID NOs: 4, 5, 7-9, 15-19, 125, 126, 128, 129, 293-304, 306, 307, 407, 439, 441, and 444. In some embodiments, the UGT is encoded by a gene set forth in UGT1495 (SEQ ID NO: 10), UGT1817 (SEQ ID NO: 11), UGT5914 (SEQ ID NO: 12), UGT8468 (SEQ ID NO: 13) or UGT10391 (SEQ ID NO: 14).

[0235] In some embodiments, the method comprises contacting mogrol with the recombinant host cell to produce mogroside IIIE, wherein the recombinant host cell further comprises one or more genes encoding one or more enzymes capable of catalyzing production of mogroside IIIE from mogrol. In some embodiments, the mogrol is produced by the recombinant host cell. In some embodiments, the one or more enzymes comprises one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases,

amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the second enzyme is a uridine diphosphate-glucosyl transferase (UGT). In some embodiments, the UGT is UGT73C3, UGT73C6, 85C2, UGT73C5, UGT73E1, UGT98, UGT1495, UGT1817, UGT5914, UGT8468, UGT10391, UGT1576, UGT SK98, UGT430, UGT1697, or UGT11789.

[0236] In some embodiments, the method comprises contacting a mogroside compound with the recombinant host cell to produce mogroside IIIE, wherein the recombinant host cell further comprises one or more genes encoding one or more enzymes capable of catalyzing production of mogroside IIIE from the mogroside compound, wherein the mogroside compound is one or more of mogroside IA1, mogroside IE1, mogroside IIA1, mogroside IIE, mogroside IIIA1, mogroside IIIA2, mogroside III, mogroside IV, mogroside IVA, mogroside V, or siamenoside. In some embodiments, the mogroside compound is produced by the recombinant host cell. In some embodiments, the one or more enzymes comprises one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the transglucosidases comprises an amino acid sequence set forth by any one of SEQ ID NOs: 163-290 and 723. In some embodiments, the genes encode a CGTase comprising an amino acid sequences set forth in SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the method comprises contacting Mogroside IA1 with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding UGT98 or UGT SK98. In some embodiments, the UGT98 or UGT SK98 enzyme comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 9, 407 or 16. In some embodiments, the contacting results in production of Mogroside IIA in the cell. In some embodiments, the one or more enzymes comprises an amino acid set forth by any one of SEQ ID NOs: 1, 3, 78-101, 106-109, 147, 154, 163-303, 405, 411, 354-405, 447-723, 770, 776, and 782.

[0237] In some embodiments, the method further comprises contacting 11-hydroxy-24,25 epoxy cucurbitadienol with the recombinant host cell, wherein the recombinant host cell further comprises a third gene encoding an epoxide hydrolase. In some embodiments, the 11-hydroxy-24,25 epoxy cucurbitadienol is produced by the recombinant host cell. In some embodiments, the method further comprises contacting 11-hydroxy-cucurbitadienol with the

recombinant host cell, wherein the recombinant host cell comprises a fourth gene encoding a cytochrome P450 or an epoxide hydrolase. In some embodiments, the P450 polypeptide is encoded in genes comprising the sequence set forth in any one of SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, and 891. In some embodiments, the 11-hydroxy-cucurbitadienol is produced by the recombinant host cell.

[0238] In some embodiments, the method further comprises contacting 3, 24, 25 trihydroxy cucurbitadienol with the recombinant host cell, wherein the recombinant host cell further comprises a fifth gene encoding a cytochrome P450. In some embodiments, the P450 polypeptide is encoded in genes comprising the sequence set forth in any one of SEQ ID NOs: 31-48, 316 and 318. In some embodiments, the 3, 24, 25 trihydroxy cucurbitadienol is produced by the recombinant host cell. In some embodiments, the contacting results in production of Mogrol in the recombinant host cell. In some embodiments, the cytochrome P450 comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 20, 308 or 315. In some embodiments, the P450 polypeptide is encoded in genes comprising the sequence set forth in any one SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, and 891. In some embodiments, the epoxide hydrolase comprises an amino acid sequence having at least 70% of sequence identity to any one of SEQ ID NOs: 21-30 and 309-314.

[0239] In some embodiments, the method further comprises contacting cucurbitadienol with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding cytochrome P450. In some embodiments, contacting results in production of 11-cucurbitadienol. In some embodiments, the 11-hydroxy cucurbitadienol is expressed in cells comprising a gene encoding CYP87D18 or SgCPR protein. In some embodiments, CYP87D18 or SgCPR comprises a sequence set forth in SEQ ID NO: 315, 872 or 874. In some embodiments, the CYP87D18 or SgCPR is encoded by SEQ ID NO: 316, 871 or 873. In some embodiments, the cucurbitadienol is produced by the recombinant host cell. In some embodiments, the gene encoding cytochrome P450 comprises a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID Nos: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, and 891. In some embodiments, the cytochrome P450 comprises an amino

acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, and 891. In some embodiments, the P450 polypeptide is encoded in genes comprising the sequence set forth in any one of SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, and 891.

[0240] In some embodiments, the method further comprises contacting 2,3-oxidosqualene with the recombinant host cell, wherein the recombinant host cell comprises a seventh gene encoding cucurbitadienol synthase. In some embodiments, he cucurbitadienol synthase comprises an amino acid sequence set forth in SEQ ID NO: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904 or 906. In some embodiments, the cucurbitadienol synthase is encoded by any one sequence set forth in SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903 and 905. In some embodiments, the contacting results in production of cucurbitadienol. In some embodiments, the 2,3-oxidosqualene is produced by the recombinant host cell. In some embodiments, the 2,3-oxidosqualene or diepoxysqualene is produced by an enzyme comprising a sequence set forth in SEQ ID NO: 898 or 900. In some embodiments, the 2,3-oxidosqualene or diepoxysqualene is produced by an enzyme encoded by a nucleic acid sequence set forth in SEQ ID NO: 897 or 899.

[0241] In some embodiments, the cucurbitadienol synthase is encoded by a gene comprising a sequence set forth in SEQ ID NO: 74. In some embodiments, the cucurbitadienol synthase is encoded by a gene comprising a nucleic acid sequence set forth in any one of SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903, and 905. In some embodiments, 11-hydroxy cucurbitadienol is produced by the cell. In some embodiments, 11-OH cucurbitadienol is expressed in cells comprising a gene encoding CYP87D18 or SgCPR protein. In some embodiments, CYP87D18 or SgCPR comprises a sequence set forth in SEQ ID NO: 315, 872 or 874. In some embodiments, the CYP87D18 or SgCPR is encoded by SEQ ID NO: 316, 871 or 873. In some embodiments, the cucurbitadienol synthase polypeptide comprises a C-terminal portion comprising the sequence set forth in SEQ ID NO: 73. In some embodiments, the gene encoding the cucurbitadienol synthase polypeptide is codon optimized. In some embodiments, the codon optimized gene comprises the nucleic acid sequence set forth in SEQ ID NO: 74. In some embodiments, the cucurbitadienol synthase comprises an amino acid sequence

having at least 70% sequence identity to any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904, and 906 (which include, for example, cucurbitadienol synthases from *C pepo*, *S grosvenorii*, *C sativus, C melo, C moschata,* and *C maxim*). In some embodiments, the cucurbitadienol synthase polypeptide comprises a C – terminal portion comprising the sequence set forth in SEQ ID NO: 73. In some embodiments, the gene encoding the cucurbitadienol synthase polypeptide is codon optimized. In some embodiments, the codon optimized gene comprises the nucleic acid sequence set forth in SEQ ID NO: 74. In some embodiments, the cucurbitadienol synthase comprises an amino acid comprising the polypeptide from *Lotus japonicas* (BAE53431), *Populus trichocarpa* (XP_002310905), *Actaea racemosa* (ADC84219), *Betula platyphylla* (BAB83085), *Glycyrrhiza glabra* (BAA76902), *Vitis vinifera* (XP_002264289), *Centella asiatica* (AAS01524), *Panax ginseng* (BAA33460), and *Betula platyphylla* (BAB83086), as described in WO 2016/050890, incorporated by reference in its entirety herein.

[0242] In some embodiments, the method comprises contacting squalene with the recombinant host cell, wherein the recombinant host cell comprises an eighth gene encoding a squalene epoxidase. In some embodiments, the contacting results in production of 2, 3-oxidosqualene. In some embodiments, the squalene is produced by the recombinant host cell. In some embodiments, the squalene epoxidase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NOs: 50-56, 60, 61, 334 or 335.

[0243] In some embodiments, the method comprises contacting farnesyl pyrophosphate with the recombinant host cell, wherein the recombinant host cell comprises a ninth gene encoding a squalene synthase. In some embodiments, the contacting results in production of squalene. In some embodiments, the farnesyl pyrophosphate is produced by the recombinant host cell. In some embodiments, the squalene synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 69 or 336.

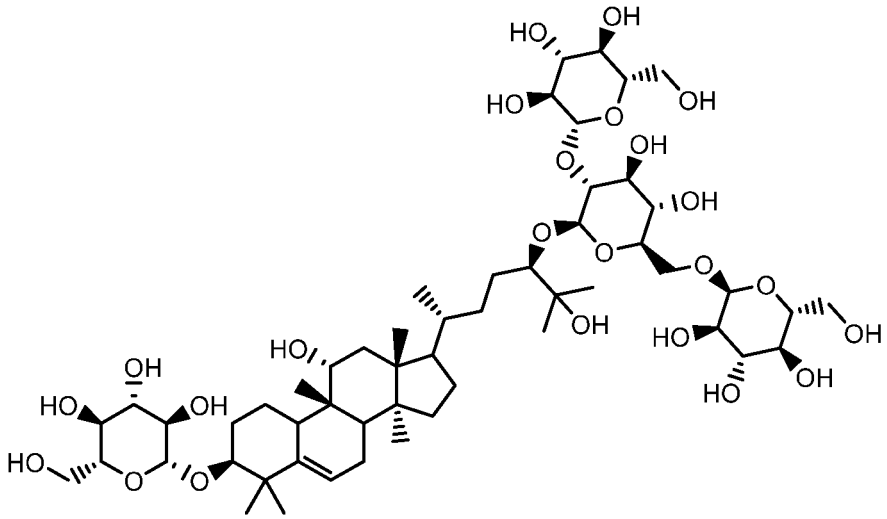
[0244] In some embodiments, the method further comprises contacting geranyl-PP with the recombinant host cell, wherein the recombinant host cell comprises a tenth gene encoding farnesyl-PP synthase. In some embodiments, the contacting results in production of

farnesyl-PP. In some embodiments, the geranyl-PP is produced by the recombinant host cell. In some embodiments, the farnesyl-PP synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 338. In some embodiments, one or more of the first, second third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth gene is operably linked to a heterologous promoter. In some embodiments, the heterologous promoter is a CMV, EF1a, SV40, PGK1, human beta actin, CAG, GAL1, GAL10, TEF1, GDS, ADH1, CaMV35S, Ubi, T7, T7lac, Sp6, araBAD, trp, Lac, Ptac, pL promoter, or a combination thereof. In some embodiments, the promoter is an inducible, repressible, or constitutive promoter. In some embodiments, production of one or more of pyruvate, acetyl-CoA, citrate, and TCA cycle intermediates have been upregulated in the recombinant host cell. In some embodiments, cytosolic localization has been upregulated in the recombinant host cell. In some embodiments, one or more of the first, second, third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth gene comprises at least one sequence encoding a 2A self-cleaving peptide. As used herein, the terms the first, the second, the third, the fourth, the fifth, the sixth, the seventh, the eighth, the ninth, the tenth, and alike do not infer particular order and/or a requirement for presence of the earlier number. For example, the recombinant host cell described herein can comprise the first gene and the third gene, but not the second gene. As another example, the recombinant host cell can comprise the first gene, the fifth gene, and the tenth gene, but not the second gene, the third gene, the fourth gene, the sixth gene, the seventh gene, the eighth gene, and the ninth gene.

[0245] The recombinant host cell can be, for examole, a plant, bivalve, fish, fungus, bacteria or mammalian cell. For example, the plant is selected from Siraitia, Momordica, Gynostemma, Cucurbita, Cucumis, Arabidopsis, Artemisia, Stevia, Panax, Withania, Euphorbia, Medicago, Chlorophytum, Eleutherococcus, Aralia, Morus, Medicago, Betula, Astragalus, Jatropha, Camellia, Hypholoma, Aspergillus, Solanum, Huperzia, Pseudostellaria, Corchorus, Hedera, Marchantia, and Morus. In some embodiments, fungus is selected from Trichophyton, Sanghuangporus, Taiwanofungus, Moniliophthora, Marssonina, Diplodia, Lentinula, Xanthophyllomyces, Pochonia, Colletotrichum, Diaporthe, Histoplasma, Coccidioides, Histoplasma, Sanghuangporus, Aureobasidium, Pochonia, Penicillium, Sporothrix, Metarhizium, Aspergillus, Yarrowia, and Lipomyces. In some embodiments, the fungus is *Aspergillus*

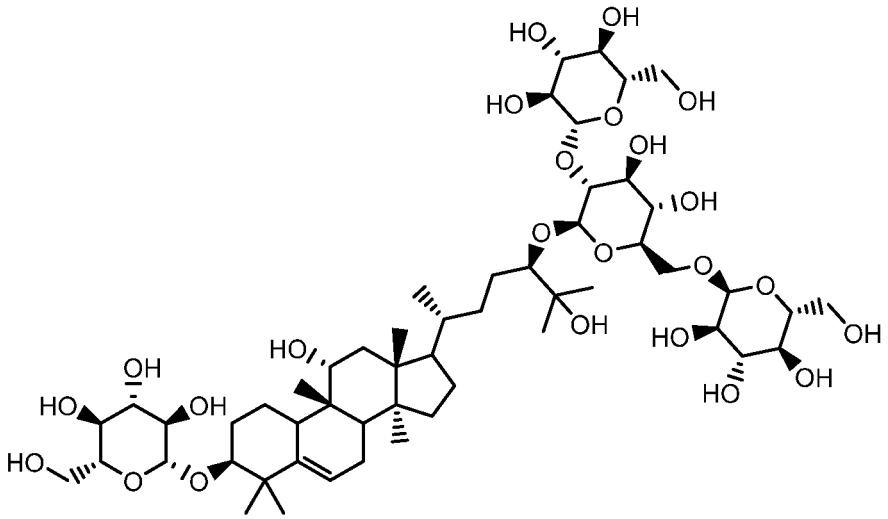
nidulans, Yarrowia lipolytica, or *Rhodosporin toruloides*. In some embodiments, the recombinant host cell is a yeast cell. In some embodiments, the yeast is selected from Candida, Sacccharaomyces, Saccharomycotina, Taphrinomycotina, Schizosaccharomycetes, Komagataella, Basidiomycota, Agaricomycotina, Tremellomycetes, Pucciniomycotina, Aureobasidium, Coniochaeta, Rhodosporidium, and Microboryomycetes. In some embodiments, the bacteria is selected from Frankia, Actinobacteria, Streptomyces, Enterococcus, In some embodiments, the bacteria is *Enterococcus faecalis*. In some embodiments, one or more of the first, second third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth genes has been codon optimized for expression in a bacterial, mammalian, plant, fungal or insect cell. In some embodiments, one or more of the first, second third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth genes comprises a functional mutation to increased activity of the encoded enzyme. In some embodiments, cultivating the recombinant host cell comprises monitoring the cultivating for pH, dissolved oxygen level, nitrogen level, or a combination thereof of the cultivating conditions. In some embodiments, the method comprises isolating Compound 1. In some embodiments, isolating Compound 1 comprises lysing the recombinant host cell. In some embodiments, isolating Compound 1 comprises isolating Compound 1 from the culture medium. In some embodiments, the method comprises purifying Compound 1. In some embodiments, purifying Compound 1 comprises HPLC, solid phase extraction or a combination thereof. In some embodiments, the purifying comprises harvesting the recombinant cells, saving the supernatant and lysing the cells. In some embodiments, the lysing comprises subjecting the cells to shear force or detergent washes thereby obtaining a lysate. In some embodiments, the shear force is from a sonication method, french pressurized cells, or beads. In some embodiments, the lysate is subjected to filtering and purification steps. In some embodiments, the lysate is filtered and purified by solid phase extraction.

[0246] In some embodiments, a compound having the structure of Compound 1,



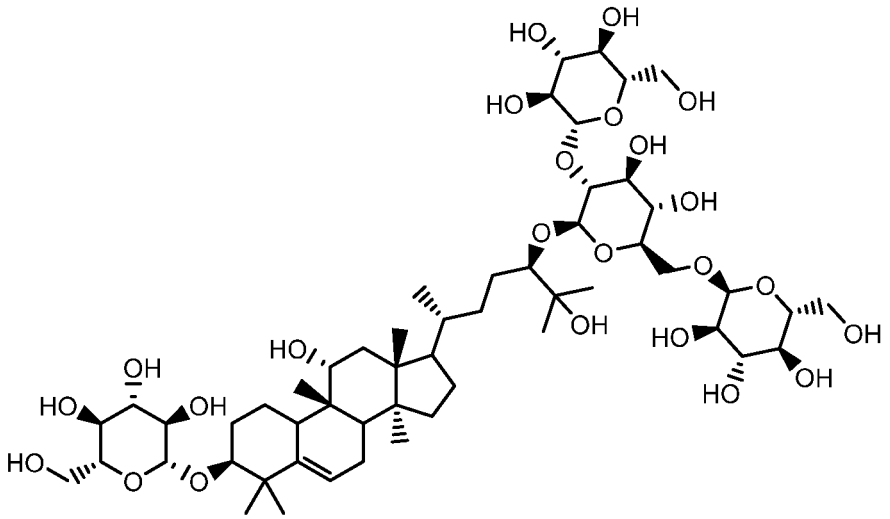
(**1**) is provided, wherein the compound is produced by the method of any one of the alternative methods provided herein.

[0247] In some embodiments, a cell lysate comprising Compound 1 having the structure:



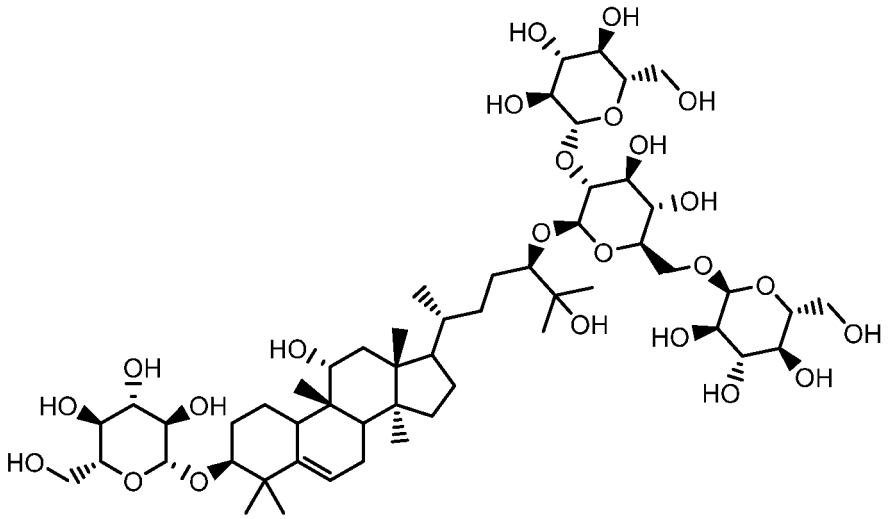
(**1**) is provided.

[0248] In some embodiments, a recombinant cell comprising: Compound 1 having the structure:



(**1**), is provided, and a gene encoding an enzyme capable of catalyzing production of Compound 1 from mogroside IIIE. In some embodiments, the gene is a heterologous gene to the recombinant cell.

[0249] In some embodiments, a recombinant cell comprising a first gene encoding a first enzyme capable of catalyzing production of Compound 1 having the structure:



(**1**) from mogroside IIIE is provided. In some embodiments, the first enzyme comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 1, 3, 78-101, 148, or 154 (CGTase). In some embodiments, the first enzyme comprises the amino acid sequence of SEQ ID NOs: 1, 3, 78-101, 148, or 154 (CGTase). In some embodiments, the first enzyme is a dextransucrase. In some embodiments, the dextransucrase comprises, or consists of, an amino acid sequence having

at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 2, 103, 106-110, 156, and 896. In some embodiments, the dextransucrase comprises, or consists of, the amino acid sequence of SEQ ID NO: 2, 103, 104, or 105. In some embodiments, the dextransucrase comprises, or consists of, the amino acid sequence of any one of SEQ ID NO: 2, 103-110 and 156-162 and 896. In some embodiments, the DexT comprises a nucleic acid sequence set forth in SEQ ID NO: 104 or 105. In some embodiments, the first enzyme is a transglucosidase. In some embodiments, the transglucosidase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of SEQ ID NO: 201 or SEQ ID NO: 291. In some embodiments, the recombinant cell further comprises a second gene encoding a uridine diphosphate-glucosyl transferase (UGT). In some embodiments, the UGT comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 4, 5, 6, 7, 8, 9, 15, 16, 17, 18, and 19. In some embodiments, UGT comprises, or consists of, the amino acid sequence of any one of SEQ ID NOs: 4, 5, 6, 7, 8, 9, 15, 16, 17, and 18. In some embodiments, the UGT is encoded by a sequence set forth in UGT1495 (SEQ ID NO: 10), UGT1817 (SEQ ID NO: 11), UGT5914 (SEQ ID NO: 12), UGT8468 (SEQ ID NO: 13), or UGT10391 (SEQ ID NO: 14). In some embodiments, the cell comprises a third gene encoding UGT98 or UGT SK98. In some embodiments, the UGT98 or UGT SK98 comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 9, 407 or 16. In some embodiments, the cell comprises a fourth gene encoding an epoxide hydrolase. In some embodiments, the epoxide hydrolase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 21-30 and 309-314. In some embodiments, the cell comprises a fifth sequence encoding P450. In some embodiments, the P450 comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NOs: 20, 49, 308, 315 or 317. In some embodiments, P450 is encoded by a gene comprising a sequence set forth in any one of SEQ ID NOs: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887,

and 891. In some embodiments, further comprises a sixth sequence encoding cucurbitadienol synthase. In some embodiments, the cucurbitadienol synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904, and 906. In some embodiments, the cucurbitadienol synthase polypeptide comprises a C–terminal portion comprising the sequence set forth in SEQ ID NO: 73. In some embodiments, the gene encoding the cucurbitadienol synthase polypeptide is codon optimized. In some embodiments, the codon optimized gene comprises the nucleic acid sequence set forth in SEQ ID NO: 74. In some embodiments, the cell further comprises a seventh gene encoding a squalene epoxidase. In some embodiments, the squalene epoxidase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 50-56, 60, 61, 334, and 335. In some embodiments, the cell further comprises an eighth gene encoding a squalene synthase. In some embodiments, the eighth gene comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 69 or SEQ ID NO: 336. In some embodiments, the cell further comprises a ninth gene encoding a farnesyl-PP synthase. In some embodiments, the farnesyl-PP synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 338. In some embodiments, the cell is a mammalian, bacterial, fungal, or insect cell. In some embodiments, the cell is a yeast cell. Non-limiting examples of the yeast include Candida, Sacccharaomyces, Saccharomycotina, Taphrinomycotina, Schizosaccharomycetes, Komagataella, Basidiomycota, Agaricomycotina, Tremellomycetes, Pucciniomycotina, Aureobasidium, Coniochaeta, and Microboryomycetes. In some embodiments, the plant is selected from the group consisting of Siraitia, Momordica, Gynostemma, Cucurbita, Cucumis, Arabidopsis, Artemisia, Stevia, Panax, Withania, Euphorbia, Medicago, Chlorophytum, Eleutherococcus, Aralia, Morus, Medicago, Betula, Astragalus, Jatropha, Camellia, Hypholoma, Aspergillus, Solanum, Huperzia, Pseudostellaria, Corchorus, Hedera, Marchantia, and Morus. In some embodiments, the fungus is Trichophyton, Sanghuangporus, Taiwanofungus, Moniliophthora, Marssonina, Diplodia, Lentinula, Xanthophyllomyces, Pochonia,

Colletotrichum, Diaporthe, Histoplasma, Coccidioides, Histoplasma, Sanghuangporus, Aureobasidium, Pochonia, Penicillium, Sporothrix, or Metarhizium.

[0250] In some embodiments, the cell comprises a sequence of an enzyme set forth in any one of SEQ ID NO: 897, 899, 909, 911, 913, 418, 421, 423, 425, 427, 871, 873, 901, 903 or 905. In some embodiments, the enzyme comprises a sequence set forth in or is encoded by a sequence in SEQ ID NO: 420, 422, 424, 426, 446, 872, 874-896, 898, 900, 902, 904, 906, 908, 910, 912, and 951-1012.

[0251] In some embodiments, DNA can be obtained through gene synthesis. This can be performed by either through Genescript or IDT, for example. DNA can be cloned through standard molecular biology techniques into an overexpression vector such as: pQE1, pGEX-4t3, pDest-17, pET series, pFASTBAC, for example. *E. coli* host strains can be used to produce enzyme (i.e., Top10 or BL21 series +/- codon plus) using 1mM IPTG for induction at OD600 of 1. *E. coli* strains can be propagated at 37C, 250 rpm and switched to room temperature or 30C (150rpm) during induction. When indicated, some enzymes can also be expressed through SF9 insect cell lines using pFASTBAC and optimized MOI. Crude extract containing enzymes can be generated through sonication and used for the reactions described herein. All UDP-glycosyltransferase reactions contain sucrose synthase, and can be obtained from *A. thaliana* via gene synthesis and expressed in *E. coli*.

Hydrolysis of hyper-glycosylated mogrosides to produce Compound 1

[0252] In some embodiments, hyper-glycosylated mogrosides can be hydrolyzed to produce Compound 1. Non-limiting examples of hyper-glycosylated mogrosides include Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III. Enzymes capable of catalyzing the hydrolysis process to produce Compound 1 can be, for example, CGTases (e.g., displays hydrolysis without starch), cellulases, β-glucosidases, transglucosidases, amylases, pectinases, dextranases, and fungal lactases. The

amino acid sequences of some of these enzymes and the nucleic acid sequences encoding some of these enzymes can be found in Table 1.

[0253] In some embodiments, Compound 1 displays tolerance to hydrolytic enzymes in the recombinant cell, wherein the hydrolytic enzymes display capabilities of hydrolyzing Mogroside VI, Mogroside V, Mogroside IV to Mogroside IIIE. The alpha-linked glycoside present in Compound 1 provides a uniqe advantage over other Mogrosides (beta-linked glycosides) due to its tolerance to hydrolysis. During microbial production of Compound 1, the recombinant host cells (e.g., microbial host cells) can hydrolyze unwanted beta-linked Mogrosides back to Mogroside IIIE. Without being bound by any particular theory, it is believed that the hydrolysis by the host cells can improve the purity of Compound 1 due to: 1) Reduction of unwanted Mogroside VI, Mogroside V, and Mogroside IV levels, and/or 2) The hydrolysis will increase the amount of Mogroside IIIE available to be used as a precursor for production of Compound 1.

Purification of mogroside compounds

[0254] Some embodiments comprise isolating mogroside compounds, for example Compound 1. In some embodiments, isolating Compound 1 comprises lysing the recombinant host cell. In some embodiments, isolating Compound 1 comprises isolating Compound 1 from the culture medium. In some embodiments, the method further comprises purifying Compound 1. In some embodiments, purifying Compound 1 comprises HPLC, solid phase extraction or a combination thereof. In some embodiments, the purifying comprises harvesting the recombinant cells, saving the supernatant and lysing the cells. In some embodiments, the lysing comprises subjecting the cells to shear force or detergent washes thereby obtaining a lysate. In some embodiments, the shear force is from a sonication method, french pressurized cells, or beads. In some embodiments, the lysate is subjected to filtering and purification steps. In some embodiments, the lysate is filtered and purified by solid phase extraction. The lysate can then be filtered and treated with ammonium sulfate to remove proteins, and fractionated on a C18 HPLC (5 X 10 cm Atlantis prep T3 OBD column, 5 um, Waters) and by injections using an A/B gradient (A = water B = acetonitrile) of 10 → 30% B over 30 minutes, with a 95% B wash, followed by re-equilibration at 1% (total run time = 42 minutes). The runs can be collected in

tared tubes (12 fractions/plate, 3 plates per run) at 30 mL/fraction. The lysate can also be centrifuged to remove solids and particulate matter. Plates can then be dried in the Genevac HT12/HT24. The desired compound is expected to be eluted in Fraction 21 along with other isomers. The pooled Fractions can be further fractionated in 47 runs on fluoro-phenyl HPLC column (3 X 10 cm, Xselect fluoro-phenyl OBD column, 5 um, Waters) using an A/B gradient (A = water, B = acetonitrile) of 15 → 30% B over 35 minutes, with a 95% B wash, followed by re-equilibration at 15% (total run time = 45 minutes). Each run was collected in 12 tared tubes (12 fractions/plate, 1 plate per run) at 30 mL/fraction. Fractions containing the desired peak with the desired purity can be pooled based on UPLC analysis and dried under reduced pressure to give a whitish powdery solid. The pure compound can be re-suspended/dissolved in 10 mL of water and lyophilized to obtain at least a 95% purity.

[0255] For purification of Compound 1, in some embodiments, the compound can be purified by solid phase extraction, which may remove the need to HPLC. Compound 1 can be purified, for example, to or to about 70%, 80%, 90%, 95%, 98%, 99%, or 100% purity or any level of purity within a range described by any two aforementioned value.s In some embodiments, compound 1 that is purified by solid phase extraction is, or is substantially, identical to the HPLC purified material. In some embodiments, the method comprises fractionating lysate from a recombinant cell on an HPLC column and collecting an eluted fraction comprising Compound 1.

Fermentation

[0256] Host cells can be fermented as described herein for the production of Compound 1. This can also include methods that occur with or without air and can be carried out in an anaerobic environment, for example. The whole cells (e.g., recombinant host cells) may be in fermentation broth or in a reaction buffer.

[0257] Monk fruit (*Siraitia grosvenorii*) extract can also be used to contact the cells in order to produce Compound 1. In some embodiments, a method of producing Compound 1 is provided. The method can comprise contacting monk fruit extract with a first enzyme capable of catalyzing production of Compound 1 from a mogroside such as such as Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-

Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III. In some embodiments, the contacting comprises contacting the mogrol fruit extract with a recombinant host cell that comprises a first gene encoding the first enzyme. In some embodiments, the first gene is heterologous to the recombinant host cell. In some embodiments, the mogrol fruit extract contacts with the first enzyme in a recombinant host cell that comprises a first polynucleotide encoding the first enzyme. In some embodiments, mogroside IIIE is in the mogrol fruit extract. In some embodiments, mogroside IIIE is also produced by the recombinant host cell. In some embodiments, the method further comprises cultivating the recombinant host cell in a culture medium under conditions in which the first enzyme is expressed. In some embodiments, the first enzyme is one or more of UDP glycosyltransferases, cyclomaltodextrin glucanotransferases (CGTases), glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the first enzyme is a CGTase. For example, the CGTase can comprise an amino acid sequence having at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or more sequence identity to the sequence of any one of SEQ ID NO: SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the CGTase comprises the amino acid sequence of any one of SEQ ID NOs: SEQ ID NOs: 1, 3, 78-101, 148, and 154. In some embodiments, the CGTase comprises the amino acid sequence of any one of SEQ ID NOs: 78-101. In some embodiments, the first enzyme is a dextransucrase. In some embodiments, the dextransucrase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of the sequences set forth in SEQ ID NOs: 2, 103, 106-110, 156 and 896. In some embodiments, the dextransucrase comprises an amino acid sequence of any one of SEQ ID NOs: 2, 103, 106-110, 156 and 896. In some embodiments, the first enzyme is a transglucosidase. In some embodiments, the transglucosidase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to the sequence of any one of SEQ ID NOs: 163-290 and 723. In some embodiments, the transglucosidase comprises an amino acid sequence of any one of SEQ ID NOs: 163-290 and 723. In some embodiments, the first enzyme

is a beta-glucosidase. In some embodiments, the beta glucosidase comprises an amino acid sequence set forth in SEQ ID NO: 292, or an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 292. In some embodiments, the mogrol fruit extract comprises Mogroside IIA and the recombinant host cell comprises a second gene encoding a second enzyme capable of catalyzing production of Mogroside IIIE from Mogroside IIA. In some embodiments, mogroside IIA is also produced by the recombinant host cell. In some embodiments, the second enzyme is one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the second enzyme is a uridine diphosphate-glucosyl transferase (UGT). In some embodiments, the UGT is UGT73C3 (SEQ ID NO: 4), UGT73C6 (SEQ ID NO:5, 444, or 445), 85C2 (SEQ ID NO: 6), UGT73C5 (SEQ ID NO: 7), UGT73E1 (SEQ ID NO: 8), UGT98 (SEQ ID NO: 9 or 407) , UGT1576(SEQ ID NO:15), UGT SK98 (SEQ ID NO: 16), UGT430 (SEQ ID NO: 17), UGT1697 (SEQ ID NO: 18), UGT11789 (SEQ ID NO: 19), or comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to UGT73C3 (SEQ ID NO: 4), UGT73C6 (SEQ ID NO:5, 444 or 445), 85C2 (SEQ ID NO: 6), UGT73C5 (SEQ ID NO: 7), UGT73E1 (SEQ ID NO: 8), UGT98 (SEQ ID NO: 9 or 407) , UGT1576 (SEQ ID NO:15), UGT SK98 (SEQ ID NO:16), UGT430 (SEQ ID NO:17), UGT1697 (SEQ ID NO:18), UGT11789 (SEQ ID NO:19). In some embodiments, the UGT is encoded by a gene set forth in UGT1495 (SEQ ID NO: 10), UGT1817 (SEQ ID NO: 11), UGT5914 (SEQ ID NO: 12), UGT8468 (SEQ ID NO: 13) or UGT10391 (SEQ ID NO:14). In some embodiments, the monk fruit extract comprises mogrol. In some embodiments, the method further comprises contacting the mogrol of the monk fruit extract wherein the recombinant host cell further comprises one or more genes encoding one or more enzymes capable of catalyzing production of Mogroside IIIE from mogrol. In some embodiments, mogrol is also produced by the recombinant host cell. In some embodiments, the one or more enzymes comprises one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the second enzyme is a uridine diphosphate-glucosyl transferase (UGT). In some embodiments, the UGT is

UGT73C3, UGT73C6, 85C2, UGT73C5, UGT73E1, UGT98, UGT1495, UGT1817, UGT5914, UGT8468, UGT10391, UGT1576, UGT SK98, UGT430, UGT1697, or UGT11789, or comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to those UGTs. In some embodiments, the method further comprises contacting the monk fruit extract with the recombinant host cell to produce mogroside IIIE, wherein the recombinant host cell further comprises one or more genes encoding one or more enzymes capable of catalyzing production of Mogroside IIIE from the mogroside compound, wherein the mogroside compound is one or more of mogroside IA1, mogroside IE1, mogroside IIA1, mogroside IIE, mogroside IIA, mogroside IIIA1, mogroside IIIA2, mogroside III, mogroside IV, mogroside IVA, mogroside V, or siamenoside. In some embodiments, a mogroside compound is also produced by the recombinant host cell. In some embodiments, the one or more enzymes comprises one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. In some embodiments, the mogroside compound is Mogroside IIE. In some embodiments, the one or more enzymes is comprises an amino acid set forth by any one of SEQ ID NOs: 293-303. In some embodiments, the mogroside compound is Morgroside IIA or Mogroside IIE, and wherein contacting the monk fruit extract with the recombinant cell expressing the one or more enzymes produces Mogroside IIIA, Mogroside IVE and Mogroside V. In some embodiments, the one or more enzymes comprise an amino acid set forth in SEQ ID NO: 304. In some embodiments, the one or more enzymes is encoded by a sequence set forth in SEQ ID NO: 305. In some embodiments, the monk fruit extract comprises Mogroside IA1. In some embodiments, the method further comprises contacting the monk fruit extract with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding UGT98 or UGT SK98. In some embodiments, the UGT98 or UGT SK98 enzyme comprises an amino acid sequence having having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 9, 407, 16 or 306.. In some embodiments, the UGT98 is encoded by a sequence set forth in SEQ ID NO: 307. In some embodiments, the contacting results in production of Mogroside IIA in the cell. In some embodiments, the monk fruit extract comprises 11-hydroxy-24,25 epoxy cucurbitadienol. In some embodiments, the method further comprises contacting

monk fruit extract with the recombinant host cell, wherein the recombinant host cell further comprises a third gene encoding an epoxide hydrolase. In some embodiments, the 11-hydroxy-24,25 epoxy cucurbitadienol is also produced by the recombinant host cell. In some embodiments, the method further comprises contacting monk fruit extract with the recombinant host cell, wherein the recombinant host cell comprises a fourth gene encoding a cytochrome P450 or an epoxide hydrolase. In some embodiments, the 11-hydroxy-cucurbitadienol is also produced by the recombinant host cell. In some embodiments, the monk fruit extract comprises 3, 24, 25 trihydroxy cucurbitadienol. In some embodiments, the method further comprises contacting monk fruit extract with the recombinant host cell, wherein the recombinant host cell further comprises a fifth gene encoding a cytochrome P450. In some embodiments, the 3, 24, 25 trihydroxy cucurbitadienol is also produced by the recombinant host cell. In some embodiments, the contacting with mogrol fruit extract results in production of Mogrol in the recombinant host cell. In some embodiments, the cytochrome P450 comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 20 or 308. In some embodiments, the epoxide hydrolase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 21-30 and 309-314. In some embodiments, the monk fruit extract comprises cucurbitadienol. In some embodiments, the method further comprises contacting cucurbitadienol with the recombinant host cell, wherein the recombinant host cell comprises a gene encoding cytochrome P450. In some embodiments, the contacting results in production of 11-hydroxy cucurbitadienol. In some embodiments, the 11-hydroxy cucurbitadienol is expressed in cells comprising a gene encoding CYP87D18 or SgCPR protein. In some embodiments, CYP87D18 or SgCPR comprises a sequence set forth in SEQ ID NO: 315, 872 or 874. In some embodiments, the CYP87D18 or SgCPR is encoded by SEQ ID NO: 316, 871 or 873. In some embodiments, the cucurbitadienol is also produced by the recombinant host cell. In some embodiments, the gene endocing cytochrome P450 comprises a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID Nos: 31-48, 316, 431, 871, 873, 875, 877, 879, 881, 883, 885, 887, and 891. In some embodiments, the cytochrome P450 comprises an amino

acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 20 or 49. In some embodiments, the monk fruit extract comprises 2, 3-oxidosqualene. In some embodiments, the method further comprises contacting 2, 3-oxidosqualene of the monk fruit extract with the recombinant host cell, wherein the recombinant host cell comprises a seventh gene encoding cucurbitadienol synthase. In some embodiments, he cucurbitadienol synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904 or 906. In some embodiments, the cucurbitadienol synthase is encoded by a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903, or 905. In some embodimemts, the monk fruit extract comprises mogroside intermediates such as Mogroside V, Siamenoside I, Mogroside IV_E, Isomogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III. In some embodiments, the method further comprises contacting a mogroside intermediate with the recombinant host cell, wherein the recombinant host cell comprises a seventh gene encoding cucurbitadienol synthase. In some embodiments, he cucurbitadienol synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904, or 906. In some embodiments, the cucurbitadienol synthase is encoded by a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903, or 905. In some embodiments, the contacting results in production of cucurbitadienol. In some embodiments, the 2,3-oxidosqualene and diepoxysqualene is also produced by the recombinant host cell. In some embodiments, the 2, 3-oxidosqualene or diepoxysqualene is produced by an enzyme comprising a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these

numbers, sequence identity to SEQ ID NO: 898 or 900, or comprising a sequence set forth in SEQ ID NO: 898 or 900. In some embodiments, the 2,3-oxidosqualene or diepoxysqualene is produced by an enzyme encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 897 or 899; or encoded by a nucleic acid set forth in SEQ ID NO: 897 or 899.

[0258] In some embodiments, the cucurbitadienol synthase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 70-73, 75-77, 319, 321, 323, 325, 327-333, 420, 422, 424, 426, 446, 902, 904, and 906. In some embodiments, the cucurbitadienol synthase is a cucurbitadienol synthase from *C pepo, S grosvenorii, C sativus, C melo, C moschata,* or *C maxim.* In some embodiments, the cucurbitadienol synthase is encoded by a gene comprising a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903, and 905, or comprising a nucleic acid sequence set forth in any one of SEQ ID NOs: 74, 320, 322, 324, 326, 328, 418, 421, 423, 425, 427, 897, 899, 901, 903 and 905. In some embodiments, 11-OH cucurbitadienol is produced by the cell. In some embodiments, 11-OH cucurbitadienol is expressed in cells comprising a gene encoding CYP87D18 or SgCPR. In some embodiments, CYP87D18 or SgCPR comprises a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 315, 872, or 874, or a sequence set forth in SEQ ID NO: 315, 872 or 874. In some embodiments, the CYP87D18 or SgCPR is encoded by a sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to SEQ ID NO: 316, 871 or 873, or a sequence set forth in SEQ ID NO: 316, 871 or 873. In some embodiments, the monk fruit extract comprises squalene. In some embodiments, the 2,3-oxidosqualene or diepoxysqualene is produced by an enzyme comprising a sequence having at least 70%, 80%, 85%, 90%, 95%, 98%, 99%, or more sequence identity to SEQ ID NO: 898 or 900, or a sequence set forth in SEQ ID NO: 898 or 900. In some embodiments, the 2, 3-oxidosqualene or diepoxysqualene is produced by an enzyme encoded by a nucleic acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any

two of these numbers, sequence identity to SEQ ID NO; 897 or 899, or a sequence set forth in SEQ ID NO: 897 or 899. In some embodiments, the method further comprises contacting squalene with the recombinant host cell, wherein the recombinant host cell comprises an eighth gene encoding a squalene epoxidase. In some embodiments, the contacting results in production of 2,3-oxidosqualene. In some embodiments, the squalene is also produced by the recombinant host cell. In some embodiments, the squalene epoxidase comprises an amino acid sequence having at least, 70%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, or a range between any two of these numbers, sequence identity to any one of SEQ ID NOs: 50-56, 60, 61, 334 or 335. In some embodiments, squalene epoxide is encoded by a nucleic acid sequence set forth in SEQ ID NO: 335. In some embodiments, the monk fruit extract comprises farnesyl pyrophosphate. In some embodiments, the method further comprises contacting farnesyl pyrophosphate with the recombinant host cell, wherein the recombinant host cell comprises a ninth gene encoding a squalene synthase. In some embodiments, the contacting results in production of squalene. In some embodiments, the farnesyl pyrophosphate is also produced by the recombinant host cell. In some embodiments, the squalene synthase comprises an amino acid sequence having at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or more of sequence identity to any one of SEQ ID NO: 69 and 336. In some embodiments, the squalene synthase is encoded by a sequence comprising the nucleic acid sequence set forth in SEQ ID NO: 337. In some embodiments, the monk fruit extract comprises geranyl-PP. In some embodiments, the method further comprises contacting geranyl-PP with the recombinant host cell, wherein the recombinant host cell comprises a tenth gene encoding farnesyl-PP synthase. In some embodiments, the contacting results in production of farnesyl-PP. In some embodiments, the geranyl-PP is also produced by the recombinant host cell. In some embodiments, the farnesyl-PP synthase comprises an amino acid sequence having at least 70%, at least 80%, at least 90% , at least 95%, at least 98%, at least 99%, or more sequence identity to SEQ ID NO: 338. In some embodiments, the farnesyl-PP synthase is encoded by a nucleic acid sequence set forth in SEQ ID NO: 339. In some embodiments, one or more of the first, second, third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth gene is operably linked to a heterologous promoter. In some embodiments, the heterologous promoter is a CMV, EF1a, SV40, PGK1, human beta actin, CAG, GAL1, GAL10, TEF1, GDS, ADH1, CaMV35S, Ubi, T7, T7lac, Sp6, araBAD, trp, lac,

Ptac, pL promoter, or a combination thereof. In some embodiments, the promoter is an inducible, repressible, or constitutive promoter. In some embodiments, production of one or more of pyruvate, acetyl-CoA, citrate, and TCA cycle intermediates have been upregulated in the recombinant host cell. In some embodiments, cytosolic localization has been upregulated in the recombinant host cell. In some embodiments, one or more of the first, second third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth gene comprises at least one sequence encoding a 2A self-cleaving peptide. In some embodiments, the recombinant host cell is a plant, bivalve, fish, fungus, bacteria or mammalian cell. In some embodiments, the plant is selected from Siraitia, Momordica, Gynostemma, Cucurbita, Cucumis, Arabidopsis, Artemisia, Stevia, Panax, Withania, Euphorbia, Medicago, Chlorophytum, Eleutherococcus, Aralia, Morus, Medicago, Betula, Astragalus, Jatropha, Camellia, Hypholoma, Aspergillus, Solanum, Huperzia, Pseudostellaria, Corchorus, Hedera, Marchantia, and Morus. In some embodiments, the fungus is selected from Trichophyton, Sanghuangporus, Taiwanofungus, Moniliophthora, Marssonina, Diplodia, Lentinula, Xanthophyllomyces, Pochonia, Colletotrichum, Diaporthe, Histoplasma, Coccidioides, Histoplasma, Sanghuangporus, Aureobasidium, Pochonia, Penicillium, Sporothrix, Metarhizium, Aspergillus, Yarrowia, and Lipomyces. In some embodiments, the fungus is *Aspergillus nidulans, Yarrowia lipolytica,* or *Rhodosporin toruloides*. In some embodiments, the recombinant host cell is a yeast cell. In some embodiments, the yeast is selected from Candida, Sacccharaomyces, Saccharomycotina, Taphrinomycotina, Schizosaccharomycetes, Komagataella, Basidiomycota, Agaricomycotina, Tremellomycetes, Pucciniomycotina, Aureobasidium, Coniochaeta, Rhodosporidium, and Microboryomycetes. In some embodiments, the bacteria is selected from the group consisting of Frankia, Actinobacteria, Streptomyces, and Enterococcus. In some embodiments, the bacteria is *Enterococcus faecalis*. In some embodiments, one or more of the first, second third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth gene has been codon optimized for expression in a bacterial, mammalian, plant, fungal or insect cell. In some embodiments, one or more of the first, second third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth genes comprises a functional mutation to increased activity of the encoded enzyme. In some embodiments, cultivating the recombinant host cell comprises monitoring the cultivating for pH, dissolved oxygen level, nitrogen level, or a combination thereof of the cultivating conditions. In some embodiments, the method comprises isolating

Compound 1. In some embodiments, isolating Compound 1 comprises lysing the recombinant host cell. In some embodiments, isolating Compound 1 comprises isolating Compound 1 from the culture medium. In some embodiments, the method further comprises purifying Compound 1. In some embodiments, purifying Compound 1 comprises HPLC, solid phase extraction or a combination thereof. In some embodiments, the purifying further comprises harvesting the recombinant cells, saving the supernatant and lysing the cells. In some embodiments, the lysing comprises subjecting the cells to shear force or detergent washes thereby obtaining a lysate. In some embodiments, the shear force is from a sonication method, french pressurized cells, or beads. In some embodiments, the lysate is subjected to filtering and purification steps. In some embodiments, the lysate is filtered and purified by solid phase extraction. In some embodiments, the method further comprises second or third additions of monk fruit extract to the growth media of the recombinant host cells. Additionally the method can be performed by contacting the monk fruit extract with the recombinant cell lysate, wherein the recombinant cell lysate comprises the expressed enzymes listed herein.

[0259] In general, compounds as disclosed and described herein, individually or in combination, can be provided in a composition, such as, e.g., an ingestible composition. In one embodiment, compounds as disclosed and described herein, individually or in combination, can provide a sweet flavor to an ingestible composition. In other embodiments, the compounds disclosed and described herein, individually or in combination, can act as a sweet flavor enhancer to enhance the sweeteness of another sweetener. In other embodiments, the compounds disclosed herein impart a more sugar-like temporal profile and/or flavor profile to a sweetener composition by combining one or more of the compounds as disclosed and described herein with one or more other sweeteners in the sweetener composition. In another embodiment, compounds as disclosed and described herein, individually or in combination, can increase or enhance the sweet taste of a composition by contacting the composition thereof with the compounds as disclosed and described herein to form a modified composition. In another embodiment, compounds as disclosed and described herein, individually or in combination, can be in a composition that modulates the sweet receptors and/or their ligands expressed in the body other than in the taste buds.

[0260] As used herein, an "ingestible composition" includes any composition that, either alone or together with another substance, is suitable to be taken by mouth whether intended for consumption or not. The ingestible composition includes both "food or beverage products" and "non-edible products". By "Food or beverage products", it is meant any edible product intended for consumption by humans or animals, including solids, semi-solids, or liquids (e.g., beverages) and includes functional food products (e.g., any fresh or processed food claimed to have a health-promoting and/or disease-preventing properties beyond the basic nutritional function of supplying nutrients). The term "non-food or beverage products" or "noncomestible composition" includes any product or composition that can be taken into the mouth by humans or animals for purposes other than consumption or as food or beverage. For example, the non-food or beverage product or noncomestible composition includes supplements, nutraceuticals, pharmaceutical and over the counter medications, oral care products such as dentifrices and mouthwashes, and chewing gum.

Compositions comprising mogroside compounds

[0261] Also disclosed herein include compostions, e.g., ingestible compositions, comprising one or more of the mogroside compounds disclosed herein, incluing but not limited to Compoun 1 and the compounds shown in Figure 43. In some embodiments, an ingestible composition can be a beverage. For example, the beverage can be selected from the group consisting of enhanced sparkling beverages, colas, lemon-lime flavored sparkling beverages, orange flavored sparkling beverages, grape flavored sparkling beverages, strawberry flavored sparkling beverages, pineapple flavored sparkling beverages, ginger-ales, root beers, fruit juices, fruit-flavored juices, juice drinks, nectars, vegetable juices, vegetable-flavored juices, sports drinks, energy drinks, enhanced water drinks, enhanced water with vitamins, near water drinks, coconut waters, tea type drinks, coffees, cocoa drinks, beverages containing milk components, beverages containing cereal extracts and smoothies. In some embodiments, the beverage can be a soft drink.

[0262] An "ingestibly acceptable ingredient" is a substance that is suitable to be taken by mouth and can be combined with a compound described herein to form an ingestible composition. The ingestibly acceptable ingredient may be in any form depending on the

intended use of a product, e.g., solid, semi-solid, liquid, paste, gel, lotion, cream, foamy material, suspension, solution, or any combinations thereof (such as a liquid containing solid contents). The ingestibly acceptable ingredient may be artificial or natural. Ingestibly acceptable ingredients includes many common food ingredients, such as water at neutral, acidic, or basic pH, fruit or vegetable juices, vinegar, marinades, beer, wine, natural water/fat emulsions such as milk or condensed milk, edible oils and shortenings, fatty acids and their alkyl esters, low molecular weight oligomers of propylene glycol, glyceryl esters of fatty acids, and dispersions or emulsions of such hydrophobic substances in aqueous media, salts such as sodium chloride, wheat flours, solvents such as ethanol, solid edible diluents such as vegetable powders or flours, or other liquid vehicles; dispersion or suspension aids; surface active agents; isotonic agents; thickening or emulsifying agents, preservatives; solid binders; lubricants and the like.

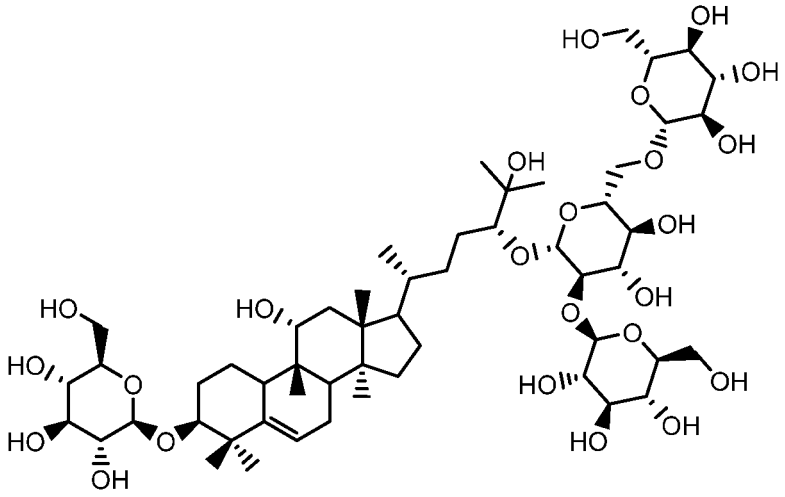
[0263] Additional ingestibly acceptable ingredients include acids, including but are not limited to, citric acid, phosphoric acid, ascorbic acid, sodium acid sulfate, lactic acid, or tartaric acid; bitter ingredients, including, for example caffeine, quinine, green tea, catechins, polyphenols, green robusta coffee extract, green coffee extract, whey protein isolate, or potassium chloride; coloring agents, including, for example caramel color, Red #40, Yellow #5, Yellow #6, Blue #1, Red #3, purple carrot, black carrot juice, purple sweet potato, vegetable juice, fruit juice, beta carotene, turmeric curcumin, or titanium dioxide; preservatives, including, for example sodium benzoate, potassium benzoate, potassium sorbate, sodium metabisulfate, sorbic acid, or benzoic acid; antioxidants including, for example ascorbic acid, calcium disodium EDTA, alpha tocopherols, mixed tocopherols, rosemary extract, grape seed extract, resveratrol, or sodium hexametaphosphate; vitamins or functional ingredients including, for example resveratrol, Co-Q10, omega 3 fatty acids, theanine, choline chloride (citocoline), fibersol, inulin (chicory root), taurine, panax ginseng extract, guanana extract, ginger extract, L-phenylalanine, L-carnitine, L-tartrate, D-glucoronolactone, inositol, bioflavonoids, Echinacea, ginko biloba, yerba mate, flax seed oil, garcinia cambogia rind extract, white tea extract, ribose, milk thistle extract, grape seed extract, pyrodixine HCl (vitamin B6), cyanoobalamin (vitamin B12), niacinamide (vitamin B3), biotin, calcium lactate, calcium pantothenate (pantothenic acid), calcium phosphate, calcium carbonate, chromium chloride, chromium polynicotinate, cupric sulfate, folic acid, ferric pyrophosphate, iron, magnesium lactate, magnesium carbonate,

magnesium sulfate, monopotassium phosphate, monosodium phosphate, phosphorus, potassium iodide, potassium phosphate, riboflavin, sodium sulfate, sodium gluconate, sodium polyphosphate, sodium bicarbonate, thiamine mononitrate, vitamin D3, vitamin A palmitate, zinc gluconate, zinc lactate, or zinc sulphate; clouding agents, including, for example ester gun, brominated vegetable oil (BVO), or sucrose acetate isobutyrate (SAIB); buffers, including, for example sodium citrate, potassium citrate, or salt; flavors, including, for example propylene glycol, ethyl alcohol, glycerine, gum Arabic (gum acacia), maltodextrin, modified corn starch, dextrose, natural flavor, natural flavor with other natural flavors (natural flavor WONF), natural and artificial flavors, artificial flavor, silicon dioxide, magnesium carbonate, or tricalcium phosphate; and stabilizers, including, for example pectin, xanthan gum, carboxylmethylcellulose (CMC), polysorbate 60, polysorbate 80, medium chain triglycerides, cellulose gel, cellulose gum, sodium caseinate, modified food starch, gum Arabic (gum acacia), or carrageenan.

EXAMPLES

[0264] Some aspects of the embodiments discussed above are disclosed in further detail in the following examples, which are not in any way intended to limit the scope of the present disclosure.

Example 1: Production of Siamenoside I

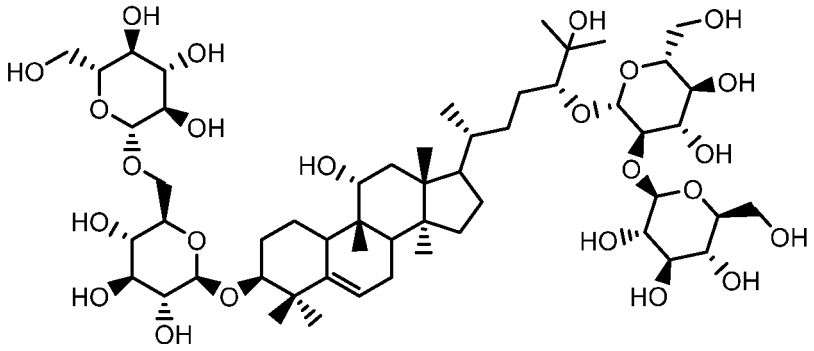


[0265] As disclosed herein, siamenoside I can be an intermediate mogroside compound for the production of Compound 1 disclosed herein. For example, siamenoside I may be hydrolyzed to produce mogroside IIIE which can then be used to produce Compound 1. For

example, a method for producing siamenoside I can comprises: contacting mogrol with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a recombinant cell expressing pectinase from *Aspergillus aculeatus* can be used.

[0266] As another example, the method for producing siamenoside I can comprises: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a pectinase from *Aspergillus aculeatus* can be used.

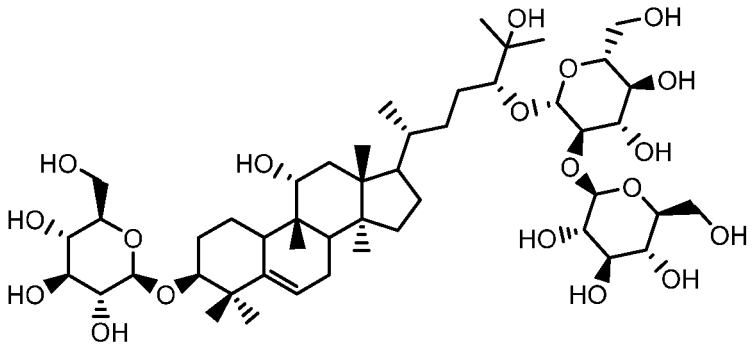
Example 2: Production of Mogroside IV_E



[0267] As disclosed herein, Mogroside IV_E can be an intermediate mogroside compound for the production of Compound 1 disclosed herein. For example, Mogroside IV_E from mogroside V can then be used to produce Compound 1. For example, a method for producing Mogroside IV_E can comprises: contacting mogroside V with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. As another example the recombinant cell can comprises a gene encoding pectinase. The pectinase can be encoded by a gene from *Aspergillus aculeatus.*

[0268] As another example, the method for producing Mogroside IV_E can comprises: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a pectinase from *Aspergillus aculeatus* can be used.

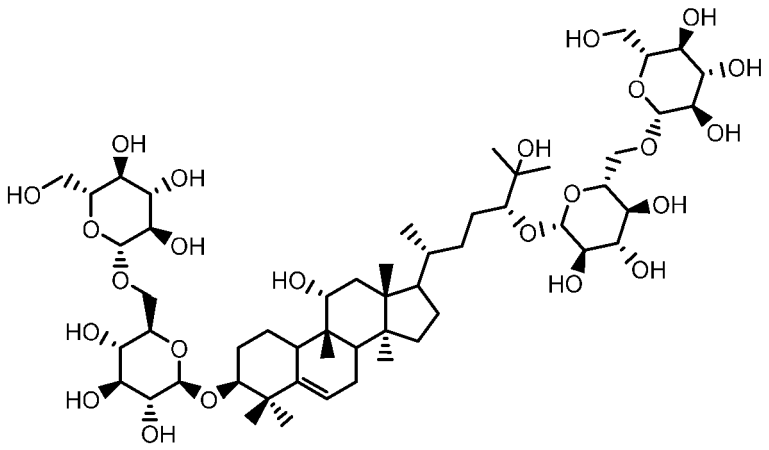
Example 3: Production of Mogroside III_E



[0269] As disclosed herein, Mogroside III_E can be an intermediate mogroside compound for the production of Compound 1 disclosed herein. For example, Mogroside II_A may be glycosylated to produce mogroside IIIE which can then be used to produce Compound 1.

[0270] As another example, the method for producing Mogroside III_E can comprises: contacting one or more of Mogroside V, Mogroside II_A, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a pectinase from *Aspergillus aculeatus* can be encoded by a gene within the recombinant host cell.

Example 4: Production of Mogroside IV_A

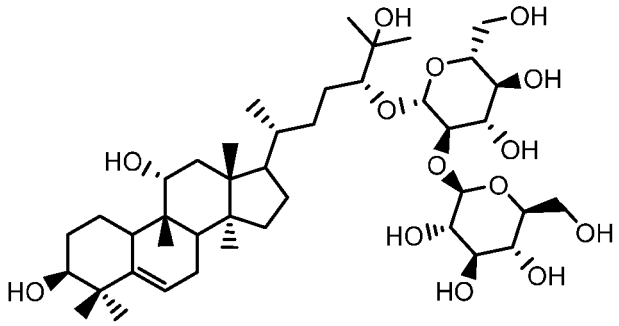


[0271] As disclosed herein, Mogroside IV_A can be an intermediate mogroside compound for the production of Compound 1 disclosed herein. For example, Mogroside IV_A from mogroside V can then be used to produce Compound 1.

[0272] For example, a method for producing Mogroside IV_A can comprises: contacting Mogroside V with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can also be a β-galactosidase from *Aspergillus oryzae*, for example.

[0273] As another example, the method for producing Mogroside IV_A can comprises: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a β-galactosidase from *Aspergillus oryzae* can be used in the method.

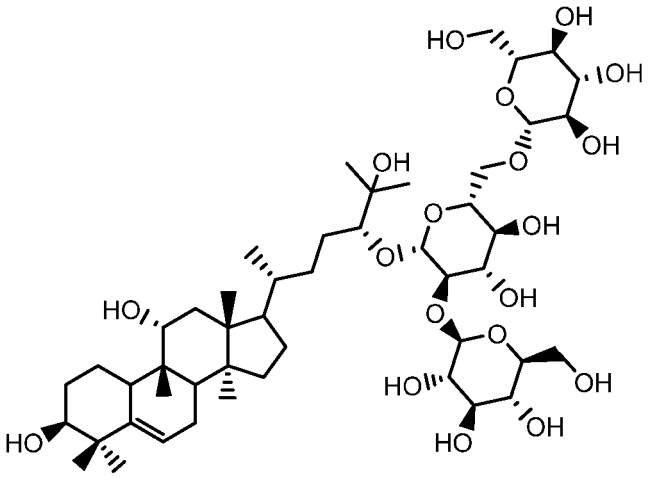
Example 5: Production of Mogroside II_A



[0274] As disclosed herein, Mogroside II_A can be an intermediate mogroside compound for the production of Compound 1 disclosed herein. For example, a method for producing Mogroside II_A can comprise: contacting Mogroside IA_1 with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

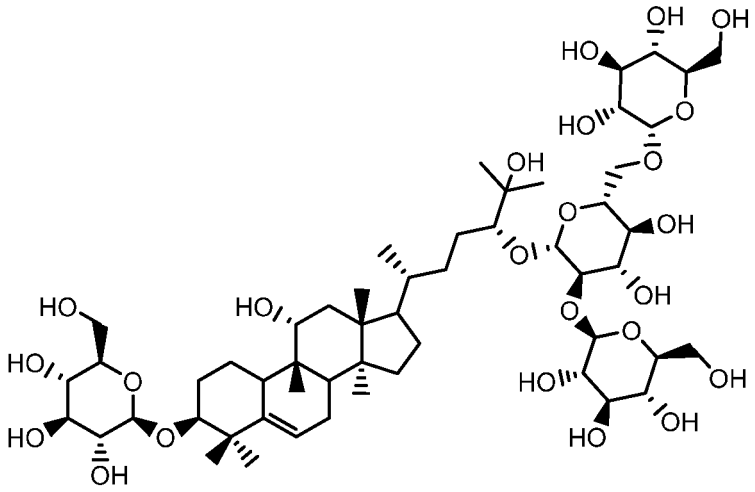
[0275] As another example, the method for producing Mogroside II_A can comprises: contacting one or more of Mogroside IA_1, Mogroside V, Siamenoside I, Mogroside IV_E, Isomogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a celluclast can also be used.

Example 6: Production of Mogroside III_{A1} from Aromase



[0276] As disclosed herein, Mogroside III_{A1} can be an intermediate mogroside compound for the production of Compound 1 disclosed herein. For example, Mogroside III_{A1} can be an intermediate to produce mogroside IVA which can then be used as an intermediate to produce Compound 1. For example, a method for producing Mogroside III_{A1} I can comprise contacting Siamenoside I with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can also be Aromase, for example. As another example, the method for producing Mogroside III_{A1} I can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

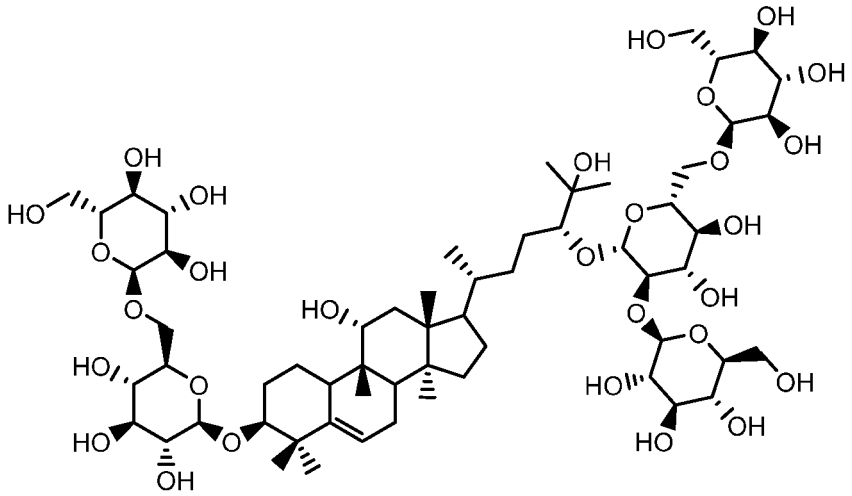
Example 7: Production of Compound 3



[0277] As disclosed herein, Compound 3 can be an intermediate mogroside compound that is produced with Compound 1 disclosed herein. For example, a method for producing Compound 3 can comprises: contacting mogrol with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be Cyclomaltodextrin glucanotransferase from Bacillus lichenformis and/or Toruzyme.

[0278] As another example, the method for producing Compound 3 can comprises: contacting one or more of Mogroside V, Siamenoside I, Mogroside IVE, Iso-mogroside V, Mogroside IIIE, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IVA, Mogroside IIA, Mogroside IIA1, Mogroside IIA2, Mogroside IA, 11-oxo-Mogroside VI, 11-oxo-Mogroside IIIE, 11-oxo-Mogroside IVE, Mogroside IIIE, Mogroside IE, Mogrol, 11-oxo-mogrol, Mogroside IIE, Mogroside IIIA2, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a CTGase enzyme can be used.

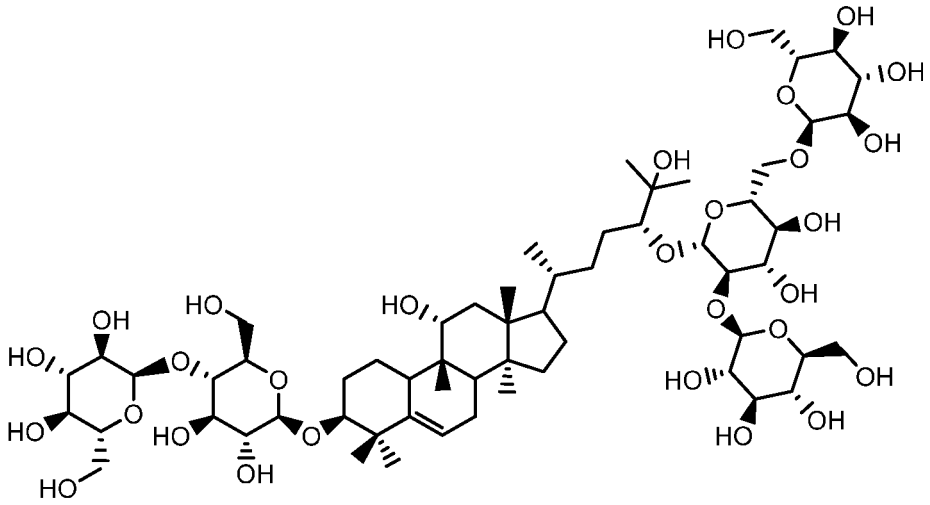
Example 8: Production of Compound 4



[0279] As disclosed herein, Compound 4 produced during the production of Compound 1 disclosed herein. For example, a method for producing Compound 4 can also lead to the production of Compound 1, the method can comprise contacting Mogroside IIIE with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0280] As another example, the method for producing Compound 4 can comprises: contacting one or more of Mogroside V, Siamenoside I, Mogroside IVE, Iso-mogroside V, Mogroside IIIE, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IVA, Mogroside IIA, Mogroside IIA1, Mogroside IIA2, Mogroside IA, 11-oxo-Mogroside VI, 11-oxo-Mogroside IIIE, 11-oxo-Mogroside IVE, Mogroside IE, Mogrol, 11-oxo-mogrol, Mogroside IIE, Mogroside IIIA2, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be Cyclomaltodextrin glucanotransferase from Bacillus lichenformis and/or Toruzyme, for example.

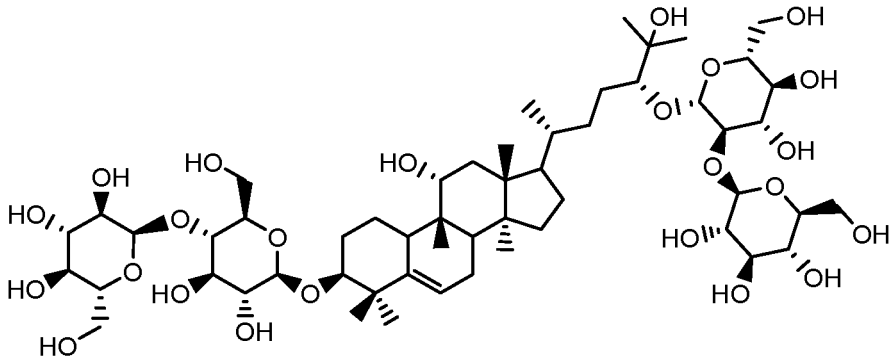
Example 9: Production of Compound 5



[0281] Compound 5 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. For example, a method for producing Compound 5 can also lead to the production of Compound 1, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0282] As another example, the method for producing Compound 5 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a CGTase from *Bacillus lichenformis* or Toruzyme can be used.

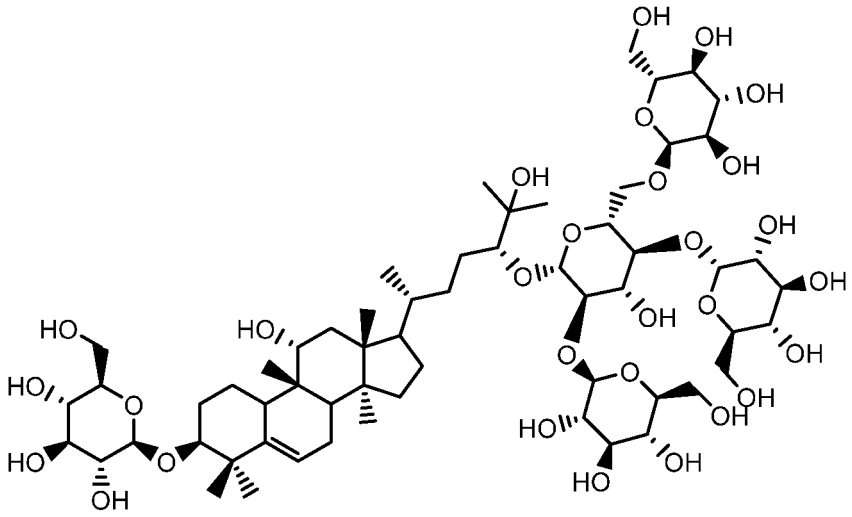
Example 10: Production of Compound 6



[0283] As disclosed herein, Compound 6 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. For example, a method for producing Compound 6 can also lead to the production of Compound 1, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0284] As another example, the method for producing Compound 6 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a CGTase from *Bacillus lichenformis* or Toruzyme can be used.

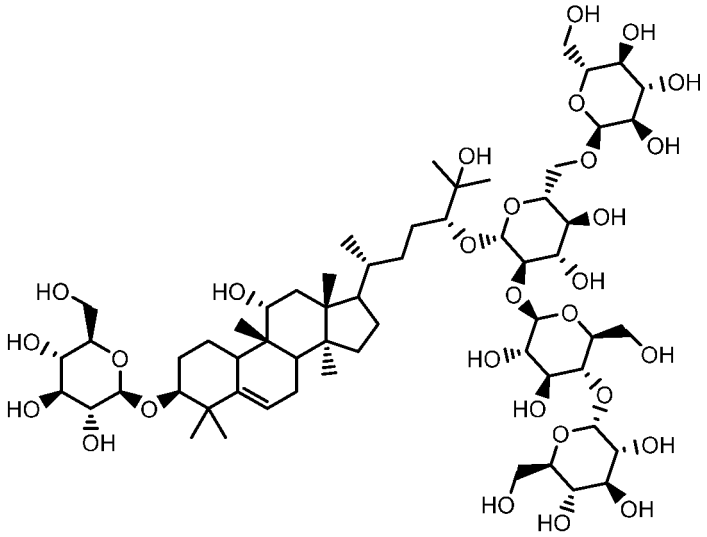
Example 11: Production of Compound 7



[0285] As disclosed herein, Compound 7 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. For example, a method for producing Compound 7 can also lead to the production of Compound 1, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

As another example, the method for producing Compound 7 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a CGTase from *Bacillus lichenformis* or Toruzyme can be used.

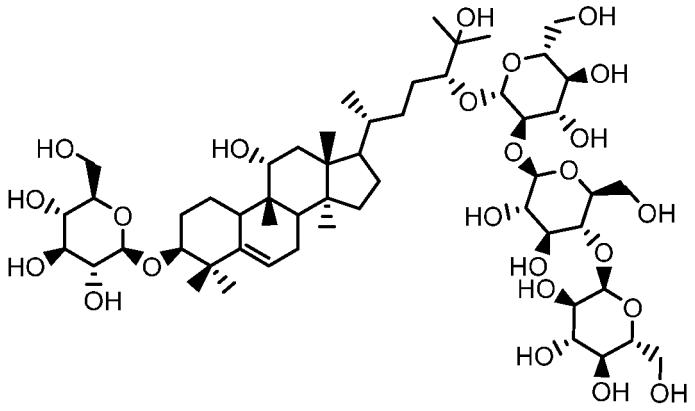
Example 12: Production of Compound 8



[0286] As disclosed herein, Compound 8 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. For example, a method for producing Compound 8 can also lead to the production of Compound 1, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0287] As another example, the method for producing Compound 8 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a CGTase from *Bacillus lichenformis* or Toruzyme can be used.

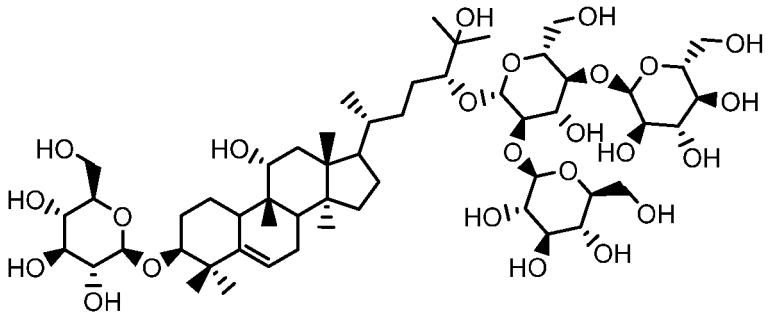
Example 13: Production of Compound 9



[0288] As disclosed herein, Compound 9 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. For example, a method for producing Compound 9 can also lead to the production of Compound 1, the method can comprise contacting Mogroside IIIE with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0289] As another example, the method for producing Compound 9 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IVE, Iso-mogroside V, Mogroside IIIE, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IVA, Mogroside IIA, Mogroside IIA1, Mogroside IIA2, Mogroside IA, 11-oxo-Mogroside VI, 11-oxo-Mogroside IIIE, 11-oxo-Mogroside IVE, Mogroside IE, Mogrol, 11-oxo-mogrol, Mogroside IIE, Mogroside IIIA2, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a CGTase from Bacillus lichenformis or Toruzyme can be used.

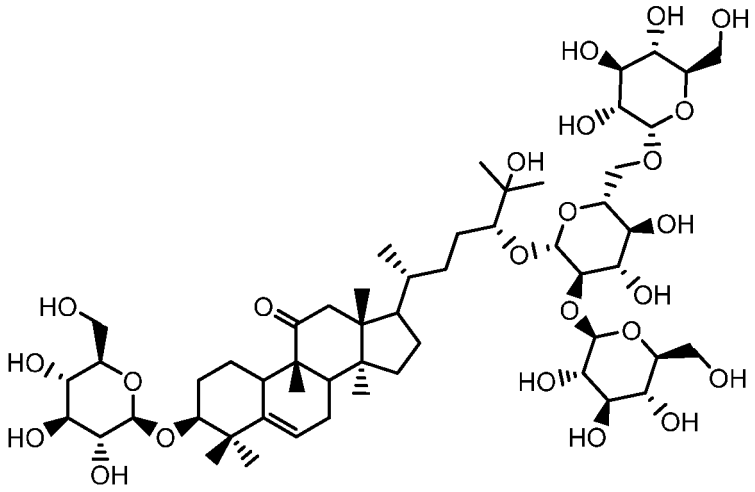
Example 14: Production of Compound 10



[0290] As disclosed herein, Compound 10 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. For example, a method for producing Compound 10 can also lead to the production of Compound 1, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0291] As another example, the method for producing Compound 10 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a CGTase from *Bacillus lichenformis* or Toruzyme can be used.

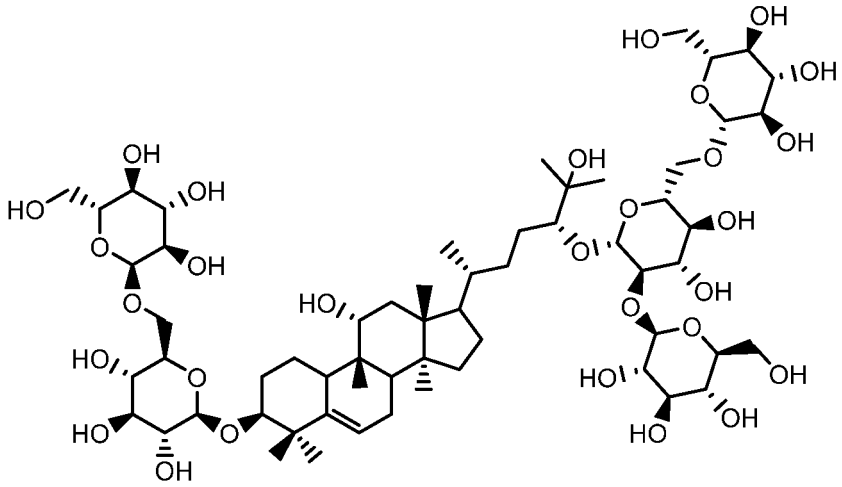
Example 15: Production of Compound 11



[0292] As disclosed herein, Compound 11 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. For example, a method for producing Compound 11 can also lead to the production of Compound 1, the method can comprise contacting Mogroside III_E or 11-oxo-$MIII_E$ with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0293] As another example, the method for producing Compound 11 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a CGTase from *Bacillus lichenformis* or Toruzyme can be used.

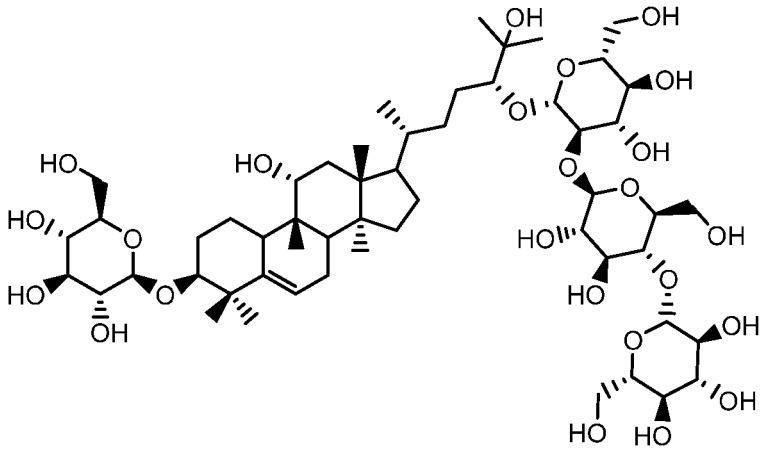
Example 16: Production of Compound 12



[0294] As disclosed herein, Compound 12 can be an intermediate mogroside compound that can be used in the production of Compound 1, disclosed herein. For example, a method for producing Compound 12 can also lead to the production of Compound 1, the method can comprise contacting Mogroside VI isomer with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, invertases and dextranases. The enzyme can be an invertase enzyme from baker's yeast, for example.

[0295] As another example, the method for producing Compound 12 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, Mogroside VI isomer and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, invertases and dextranases.

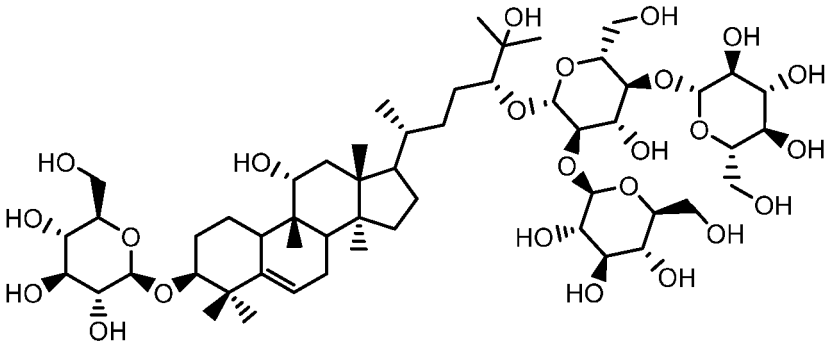
Example 17: Production of Compound 13



[0296] As disclosed herein, Compound 13 can be an intermediate mogroside produced during the production of Compound 1 disclosed herein. For example, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme expressed can also be a celluclast, for example.

[0297] As another example, the method for producing Compound 13 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a celluclast can be used.

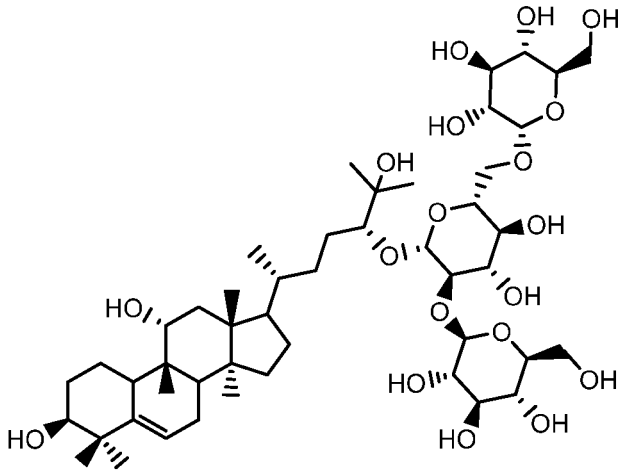
Example 18: Production of Compound 14



[0298] As disclosed herein, Compound 14 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. For example, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme expressed can also be a celluclast, for example.

[0299] As another example, the method for producing Compound 14 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a celluclast can be used. The method can also require the presence of a sugar, such as α-lactose, for example.

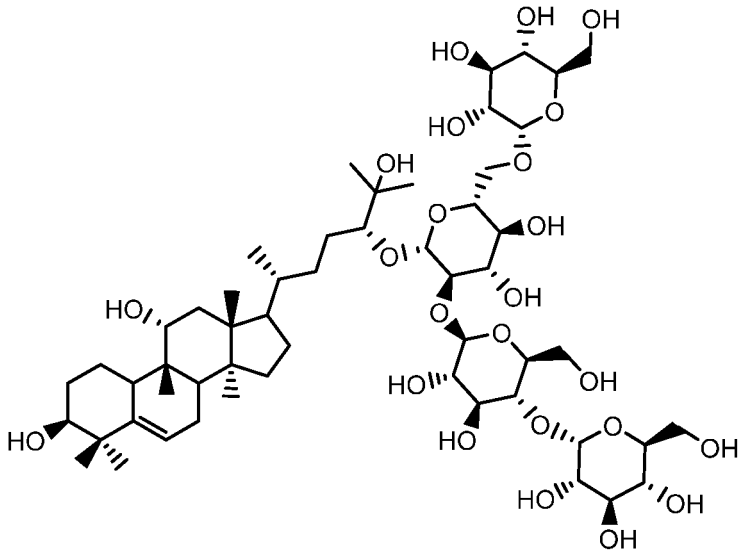
Example 19: Production of Compound 15



[0300] As disclosed herein, Compound 15 can be an intermediate mogroside compound that can be used for the production of Compound 1 disclosed herein. For example, the method can comprise contacting mogroside II_A with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0301] As another example, the method for producing Compound 15 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a toruzyme can be used.

Example 20: Production of Compound 16

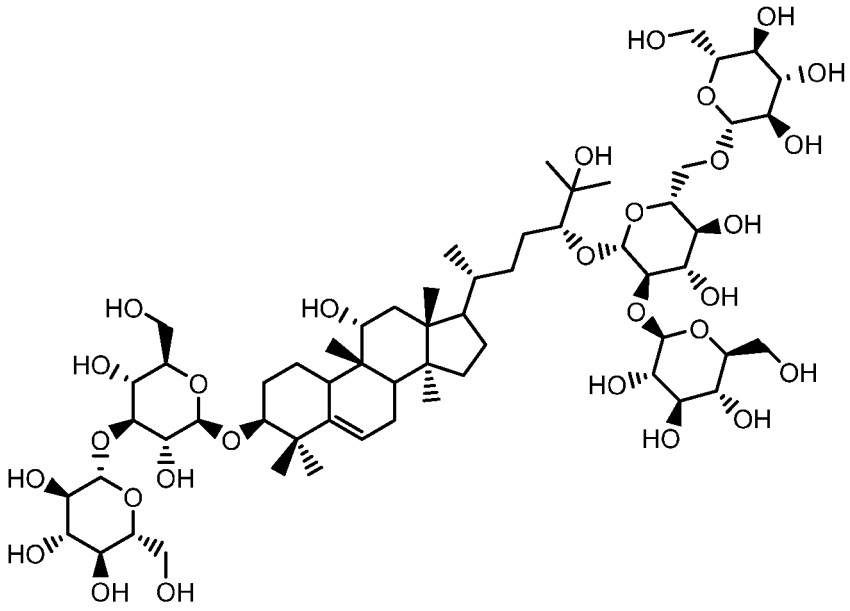


[0302] As disclosed herein, Compound 16 can be an intermediate mogroside compound that can be used for the production of Compound 1 disclosed herein. For example, the method can comprise contacting mogroside II_A with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

[0303] As another example, the method for producing Compound 16 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a toruzyme can be used.

[0304] The enzyme can be Toruzyme, for example. The recombinant cell can further comprise a gene encoding a clyclomatlodextrin glucanotransferase (e.g., Toruzyme), an invertase, a glucostransferase (e.g., UGT76G1), for example.

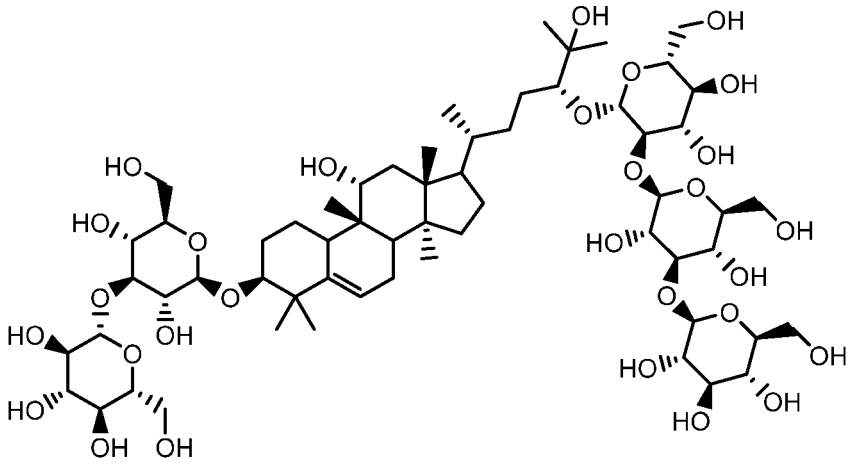
Example 21: Production of Compound 17



[0305] As disclosed herein, Compound 17 can be an intermediate mogroside compound for the production of Compound 1 disclosed herein. For example, Compound 17 may be hydrolyzed to produce mogroside IIIE which can then be used to produce Compound 1. For example, a method for producing Compound 17 can comprises: contacting Siamenoside I with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, transglucosidases, sucrose synthases, pectinases, and dextranases. For example, a recombinant cell expressing a UDP glycosyltransferase can be used.

[0306] As another example, the method for producing Compound 17 can comprises: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. For example, a UDP glycosyltransferases can be used.

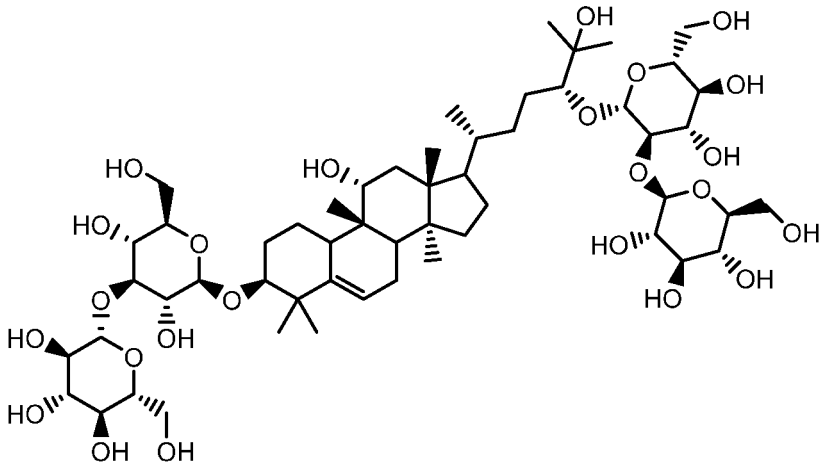
Example 22: Production of Compound 18



[0307] As disclosed herein, Compound 18 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. For example, Compound 18 may be hydrolyzed to produce mogroside IIIE which can then be used to produce Compound 1.For example, a method for producing Compound 18 can also lead to the production of Compound 1, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzymes can be Sus1 and UGT76G1 for example.

[0308] As another example, the method for producing Compound 18 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be UGT76G1, for example.

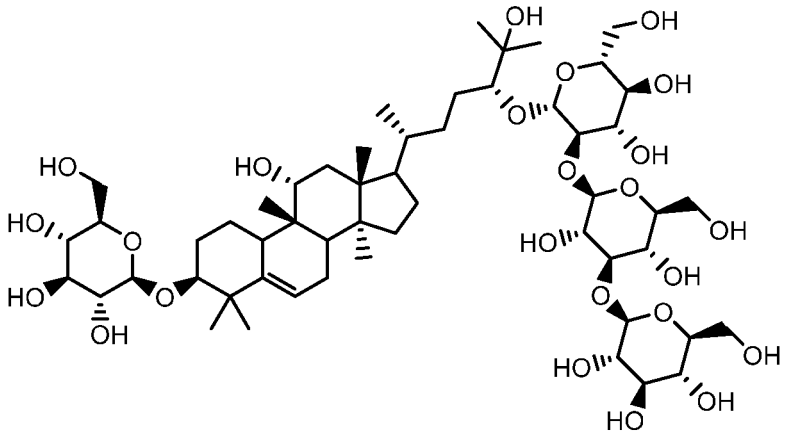
Example 23: Production of Compound 19



[0309] As disclosed herein, Compound 19 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Compound 19 can be further hydrolyzed to produce Compound 1, for example. For example, a method for producing Compound 18 can also lead to the production of Compound 1, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzymes can be Sus1 and UGT76G1 for example.

[0310] As another example, the method for producing Compound 19 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be UGT76G1, for example. The enzyme can also be sucrose synthase Sus1, for example.

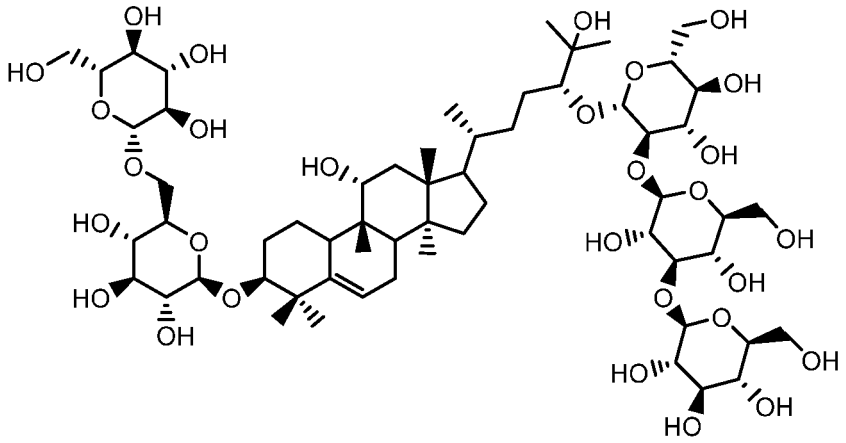
Example 24: Production of Compound 20



[0311] As disclosed herein, Compound 20 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Compound 20 can be further hydrolyzed to produce Compound 1, for example. For example, a method for producing Compound 20 can also lead to the production of Compound 1, the method can comprise contacting Mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzymes can be Sus1 and UGT76G1 for example.

[0312] As another example, the method for producing Compound 20 can comprise: contacting one or more of Mogroside V, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be UGT76G1, for example. The enzyme can also be sucrose synthase Sus1, for example. The enzyme can be sucrose synthase Sus1 and UGT76G1, for example.

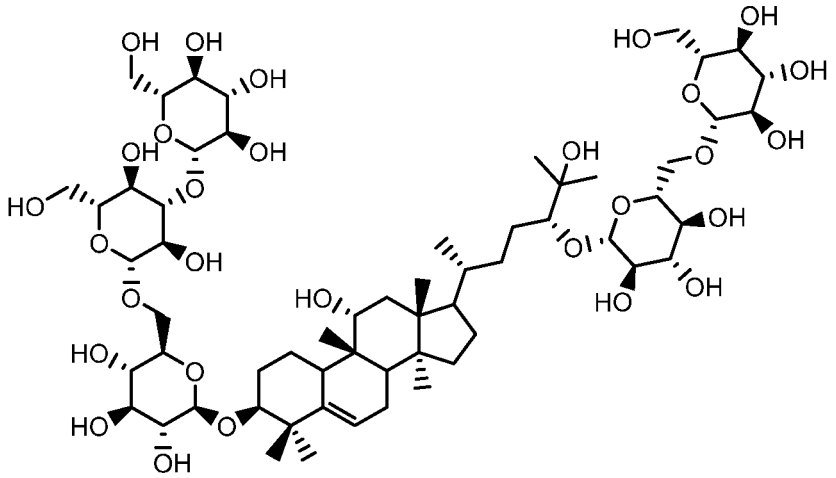
Example 25: Production of Compound 21



[0313] As disclosed herein, Compound 21 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Compound 21 can be further hydrolyzed to produce Compound 1, for example. For example, a method for producing Compound 21 can also lead to the production of Compound 1, the method can comprise contacting Mogroside IV_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzymes can be Sus1 and UGT76G1 for example.

[0314] As another example, the method for producing Compound 21 can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Isomogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be UGT76G1, for example. The enzyme can also be sucrose synthase Sus1, for example. The enzymes can be sucrose synthase Sus1 and GT76G1, for example.

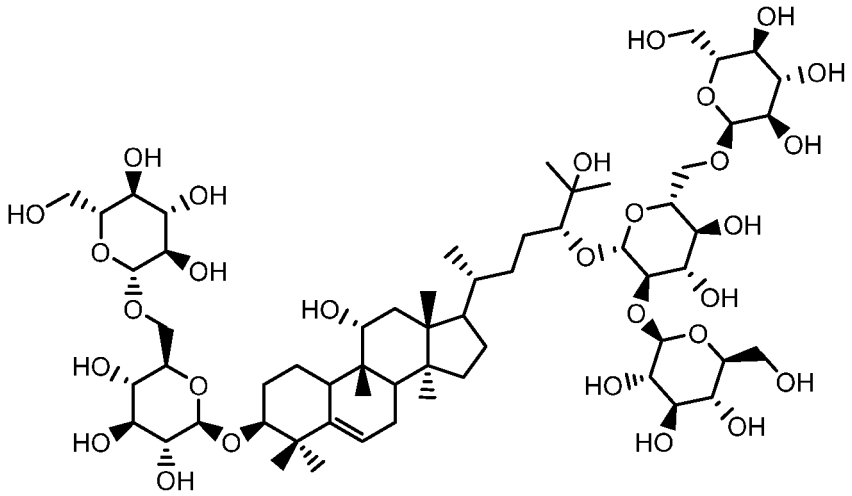
Example 26: Production of Compound 22



[0315] As disclosed herein, Compound 22 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Compound 22 can be further hydrolyzed to produce Compound 1, for example. For example, a method for producing Compound 22 can also lead to the production of Compound 1, the method can comprise contacting Mogroside IVA with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzymes can be Sus1 and UGT76G1 for example.

[0316] As another example, the method for producing Compound 22 can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be UGT76G1, for example. The enzyme can also be sucrose synthase Sus1, for example. The enzymes can be sucrose synthase Sus1 and GT76G1, for example.

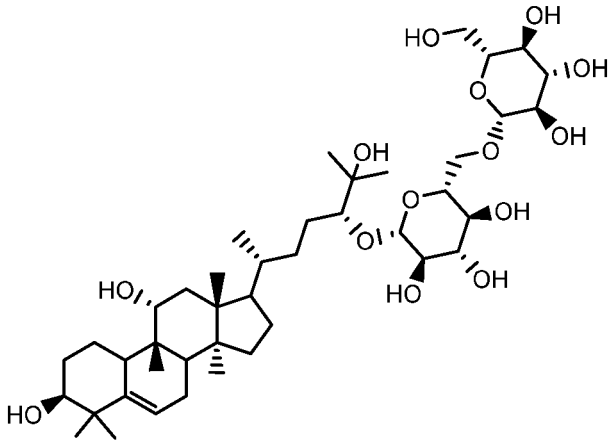
Example 27: Production of Compound 23



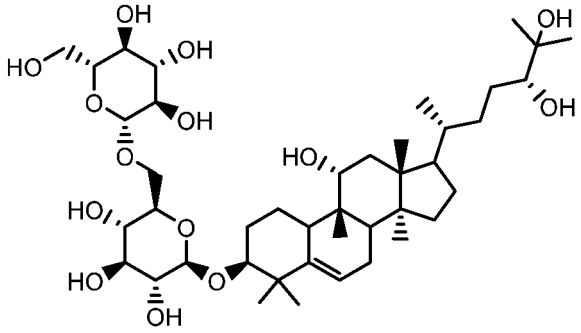
[0317] As disclosed herein, Compound 23 can be an intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Compound 23 can be further hydrolyzed to produce Compound 1, for example. For example, a method for producing Compound 22 can also lead to the production of Compound 1, the method can comprise contacting Mogroside IV_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be dextransucrase, for example.

[0318] As another example, the method for producing Compound 23 can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Isomogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be detransucrase, for example, which will hydrolyze the hyper glycosylated mogroside IV_E isomers to the desired mogroside V isomer.

Examples 28 and 29: Production of Mogroside II_{A1} and Mogroside II_{A2} from Fungal lactase



Mogroside II_{A1}

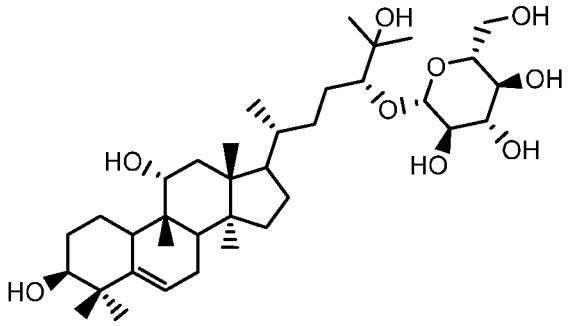


Mogroside II_{A2}

[0319] As disclosed herein, Mogroside II_{A1} and Mogroside II_{A2} can be intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Mogroside II_{A1} and Mogroside II_{A2} can be further hydrolyzed to produce Compound 1, for example. For example, a method for producing Mogroside II_{A1} and Mogroside II_{A2} can also lead to the production of Compound 1, the method can comprise contacting Mogroside IV_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be a lactase from a fungus, for example.

[0320] As another example, the method for producing Mogroside II_{A1} and Mogroside II_{A2} can include: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

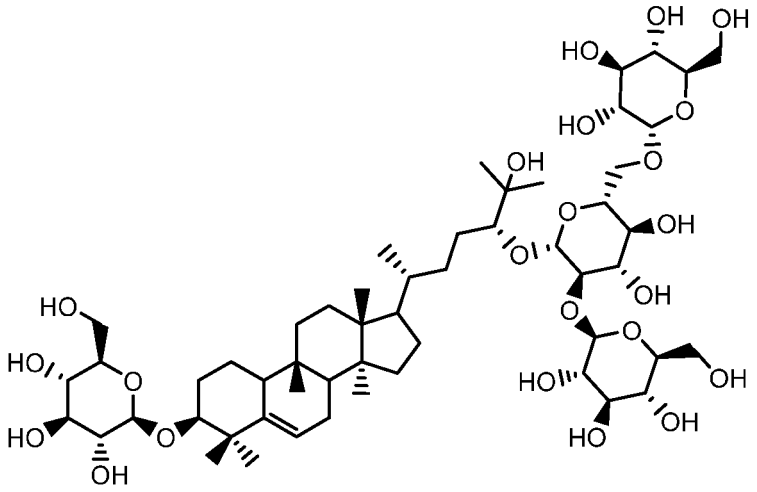
Example 30: Production of Mogroside $_{IA}$ from Viscozyme



[0321] As disclosed herein, Mogroside $_{IA}$ can be intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Mogroside $_{IA}$ can be further hydrolyzed to produce Compound 1, for example. A method for producing Mogroside $_{IA}$ can also lead to the production of Compound 1, the method can comprise contacting Mogroside IIA with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be Viscozyme, for example.

[0322] As another example, the method for producing Mogroside $_{IA}$ can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be Viscozyme, for example.

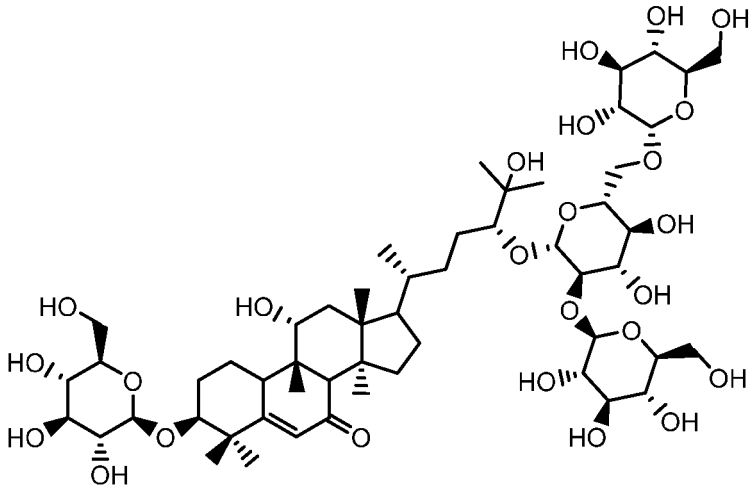
Example 31: Production of Compound 24



[0323] As disclosed herein, Compound 24 can be intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Compound 24 can be further hydrolyzed to produce Compound 1, for example. A method for producing Compound 24_can also lead to the production of Compound 1, the method can comprise contacting mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be dextransucrase DexT, for example.

[0324] As another example, the method for producing Compound 24 can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Isomogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be dextransucrase DexT, for example.

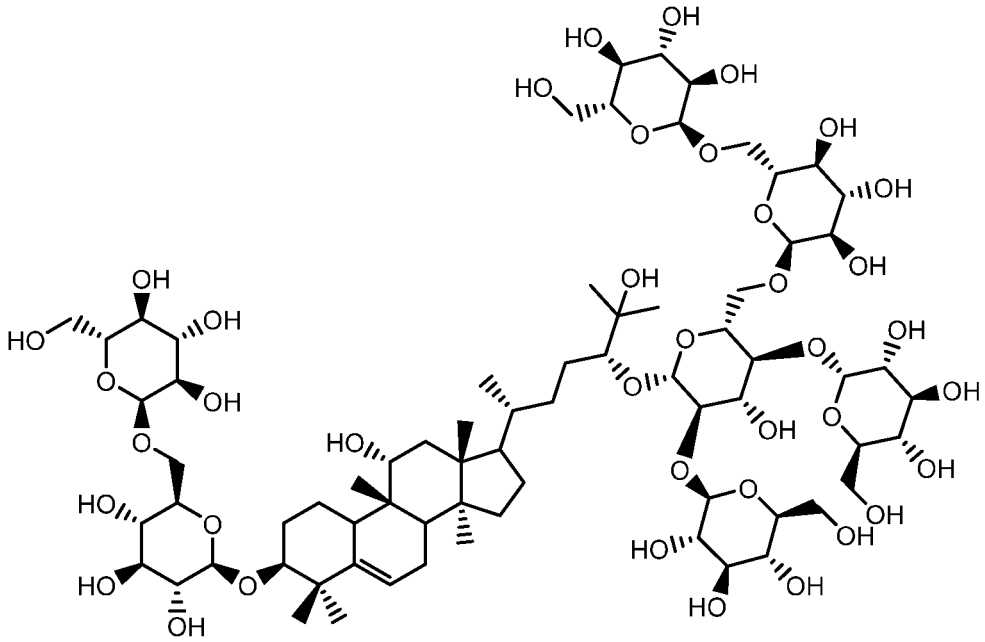
Example 32: Production of Compound 25



[0325] As disclosed herein, Compound 25 can be intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Compound 25 can be further hydrolyzed to produce Compound 1, for example. A method for producing Compound 25_can also lead to the production of Compound 1, the method can comprise contacting mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be dextransucrase DexT, for example.

[0326] As another example, the method for producing Compound 25 can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be dextransucrase DexT, for example.

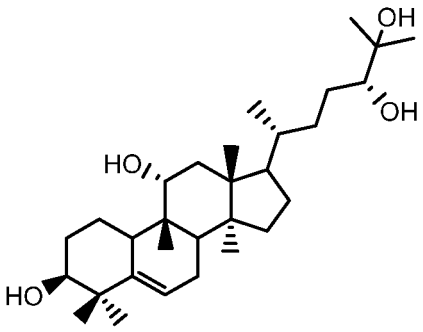
Example 33: Production of Compound 26



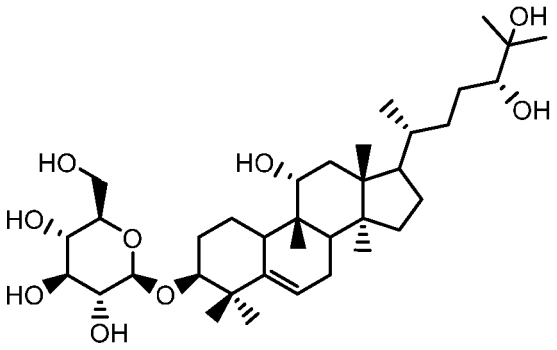
[0327] As disclosed herein, Compound 26 can be intermediate mogroside compound produced during the production of Compound 1 disclosed herein. Compound 26 can be further hydrolyzed to produce Compound 1, for example. A method for producing Compound 26 can also lead to the production of Compound 1, the method can comprise contacting mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be dextransucrase DexT, for example.

[0328] As another example, the method for producing Compound 26 can include: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be dextransucrase DexT, for example.

Examples 34 and 35: Production of Mogrol and Mogroside I_E from pectinase



Mogrol

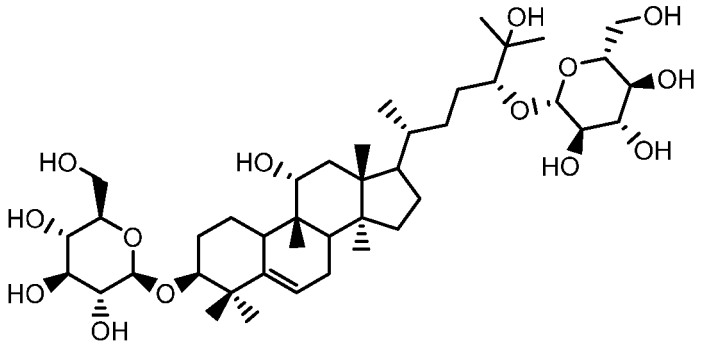


Mogroside $_{IE}$

[0329] As disclosed herein, Mogrol and Mogroside I_E can be intermediate mogroside compounds produced during the production of Compound 1 disclosed herein. Mogrol can be used as a substrate for producing Mogroside IA1, which is further hydrolyzed to form Compound 1 and Mogroside I_E can be further hydrolyzed to produce Compound 1, for example. A method for producing Mogrol and Mogroside can also lead to the production of Compound 1, the method can comprise contacting mogroside V with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be pectinase enzyme from *Aspergillus aculeatus*, for example.

[0330] As another example, the method for producing Mogrol and Mogroside I_E can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

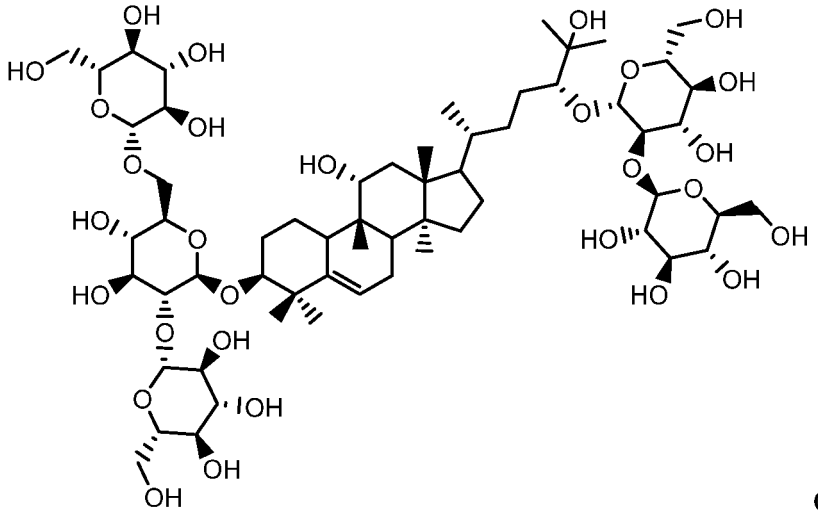
Example 36: Production of Mogroside IIE



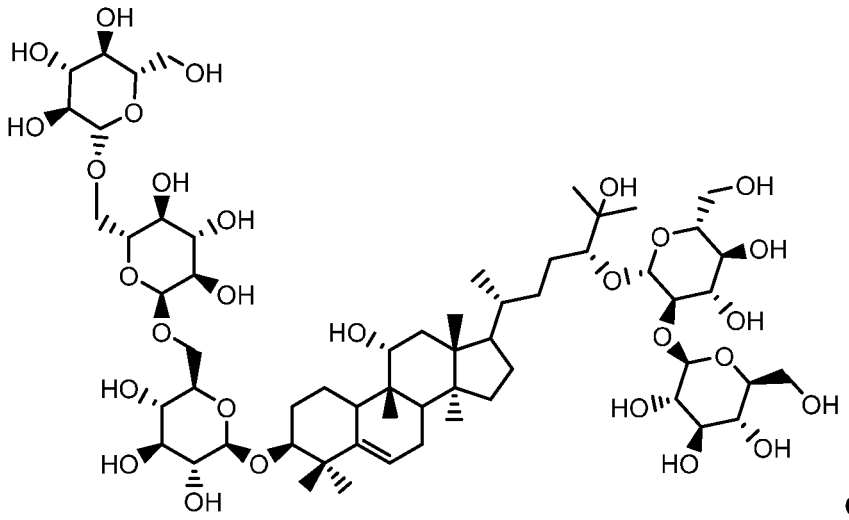
[0331] As disclosed herein, Mogroside IIE can be intermediate mogroside compounds produced during the production of Compound 1 disclosed herein. Mogroside IIE can be further hydrolyzed to produce Compound 1, for example. A method for producing Mogroside IIE can also lead to the production of Compound 1, the method can comprise contacting mogroside V with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be pectinase enzyme from *aspergillus aculeatus*, for example.

[0332] As another example, the method for producing Mogroside IIE can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

Examples 37 and 38: Production of Compounds 32 and 33



Compound 32



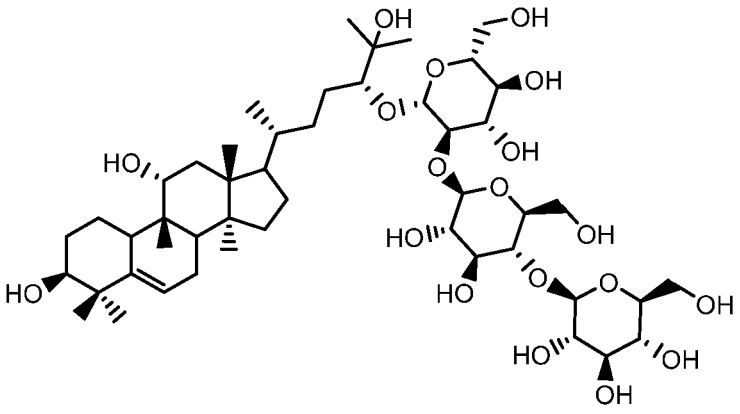
Compound 33

[0333] As disclosed herein, Compounds 32 and 33 can be intermediate mogroside compounds produced during the production of Compound 1 disclosed herein. Compounds 32 and 33 can be further hydrolyzed to produce Compound 1, for example. A method for producing Compounds 32 and 33 can also lead to the production of Compound 1, the method can comprise contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell

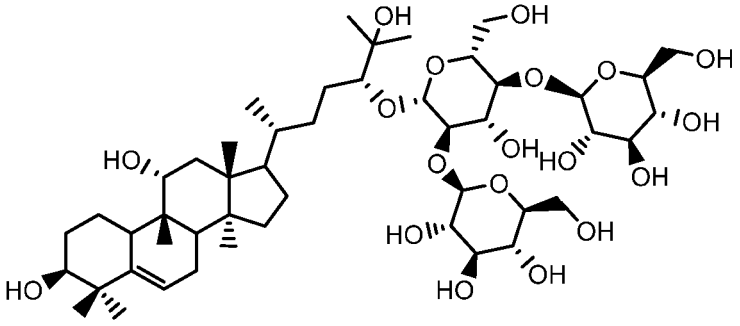
expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be pectinase enzyme from *aspergillus aculeatus*, for example.

[0334] As another example, the method for producing Compound 32 and 33 can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

Examples 39 and 40: Production of Compounds 34 and 35



Compound 34



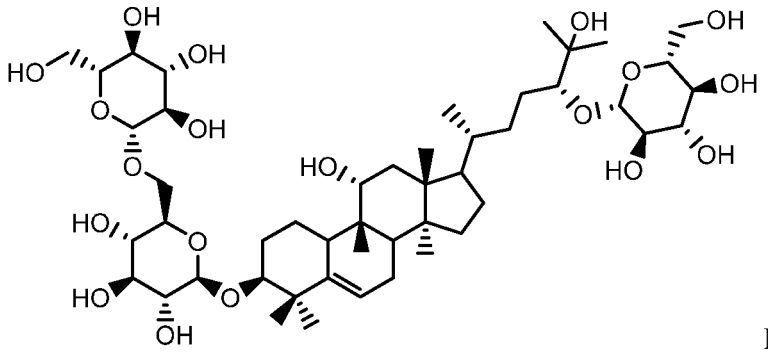
Compound 35

[0335] As disclosed herein, Compounds 34 and 35 can be intermediate mogroside compounds produced during the production of Compound 1 disclosed herein. Compounds 32

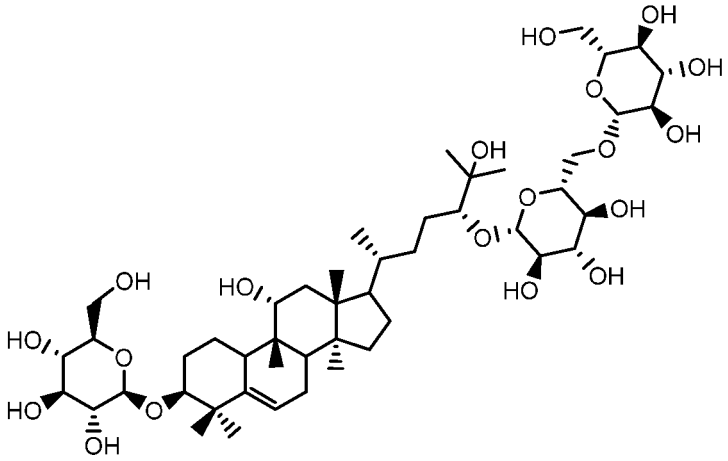
and 33 can be further hydrolyzed to produce Compound 1, for example. A method for producing Compounds 34 and 35 can also lead to the production of Compound 1, the method can comprise contacting mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be celluclast, for example.

[0336] As another example, the method for producing Compounds 34 and 35 can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

Examples 41 and 42: Production of Mogroside III_{A2} and Mogroside III



Mogroside III_{A2}



Mogroside III

[0337] As disclosed herein, Mogroside III_{A2} and Mogroside III can be intermediate mogroside compounds produced during the production of Compound 1 disclosed herein. Mogroside III_{A2} and Mogroside III can be further hydrolyzed to produce Compound 1, for example.

[0338] For example Mogroside III_{A2} and Mogroside III can be also contact UGT to form Mogroside IVA, another mogroside compound that can be used to make Mogroside IIIE, which is further hydrolyzed to form Compound 1.

[0339] A method for producing Mogroside III_{A2} and Mogroside III can also lead to the production of Compound 1, the method can comprise contacting mogroside III_E with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases, glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases. The enzyme can be celluclast, for example.

[0340] As another example, the method for producing Mogroside III_{A2} and Mogroside III can comprise: contacting one or more of Mogroside V, Mogroside IV_E, Siamenoside I, Mogroside IV_E, Iso-mogroside V, Mogroside III_E, 11-Deoxy-mogroside V, 11-Oxo-mogroside V, Mogroside VI, Mogroside IV_A, Mogroside II_A, Mogroside II_{A1}, Mogroside II_{A2}, Mogroside I_A, 11-oxo-Mogroside VI, 11-oxo-Mogroside III_E, 11-oxo-Mogroside IV_E, Mogroside I_E, Mogrol, 11-oxo-mogrol, Mogroside II_E, Mogroside III_{A2}, and Mogroside III with a recombinant host cell expressing one or more of UDP glycosyltransferases, CGTases,

glycotransferases, dextransucrases, cellulases, β-glucosidases, amylases, transglucosidases, pectinases, and dextranases.

Example 43: Use of CGT-SL enzyme to produce Compound 1

[0341] In 1ml reaction volume, 5mg of Mogroside IIIE, 50mg of soluble starch, 0.1M NaOAC pH 5.0, 125ul of CGT-SL enzyme (from *Geobaccilus thermophillus*) and water was mixed and with a stir bar and incubated at 50C. Time point samples were taken for HPLC.

[0342] HPLC Data: Mass spec of Compound 1 production as shown in Figure 1. In some embodiments, CTG-SL can comprise a sequence set forth in SEQ ID NO: 3, 148 or154.

Example 44: Cloning: Gene encoding for dextransucrase enzyme was PCR amplified from *Leuconostoc citreum* ATCC11449 and cloned into pET23a

[0343] Growth conditions: BL21 Codon Plus RIL strain was grown in 2xYT at 37C, 250 rpm until OD600 of 1. 10mM of lactose was added for induction, incubated at room temperature, 150rpm overnight. Crude extract used for the reaction was obtained either by sonication or osmotic shock.

[0344] In some embodiments, the dextransucrase comprises, or consists of, an amino acid sequence of any one of SEQ ID NOs: 2, 103, 106-110, 156, and 896. In some embodiments, the DexT can comprises an amino acid sequence set forth in SEQ ID NO: 103. In some embodiments, the DexT comprises a nucleic acid sequence set forth in SEQ ID NO: 104 or 105.

Example 45: Reaction of Mogroside IIIE with S mutans Clarke ATCC25175 Dextransucrase to produce Compound 1

[0345] Growth conditions: The strain indicated above was grown anaerobically with glucose supplementation as indicated in Wenham, Henessey and Cole (1979) to stimulate dextransucrase production. 5mg/ml Mogroside IIIE was added to the growth media. Time point samples were taken for HPLC. HPLC Data is presented as mass spec of Compound 1 production in Figure 2.

Example 46: Reaction of Mogroside IIIE with CGTase

[0346] In 1ml reaction volume, 5mg of Mogroside IIIE, 50mg of soluble starch, 0.1M NaOAC pH 5.0, 125ul of enzyme and water was mixed and with a stir bar and incubated at 50C. Time point samples were taken for HPLC. The enzyme used was CGTase. The product of Compound 1 is seen in the HPLC data and mass spectroscopy data as shown in Figure 1. Mass peaks correspond to the size of Compound 1.

Example 47: Reaction of Mogroside IIIE with Celluclast

[0347] Celluclast xylosylation were performed with mogroside IIIE with celluclast from the native host: *Trichoderma reesei*

[0348] Reaction conditions: 5mg of Mogroside IIIE, 100mg xylan, 50ul Celluclast were mixed in a total volume of 1ml with 0.1M sodium acetate pH 5.0, incubated at 50C with stirring. Time point samples were taken for HPLC.

[0349] Xylosylated product is highlighted in Figure 3. Products from xylosylation can be used as intermediates in production of Compound 1. The sequences for Celluclast are included in Table 1 and is used herein for the production of xylosylated products.

Example 48: Glycosyltransferases (Maltotriosyl transferase) (native host: Geobacillus sp. APC9669)

[0350] In this example, glycosytransferase AGY15763.1 (Amano Enzyme U.S.A. Co., Ltd., Elgin, IL; SEQ ID NO: 434, see Table 1) was used. 20 mL dI water, 0.6ml 0.5M MES pH 6.5, 6g soluble starch, 150mg Mogroside IIIE, and 3 ml enzyme were added to a 40ml flat-bottom screw cap vial. The vial was sealed with black cap, incubated at 30°C and stirred at 500rpm using magnetic bar. 3 more identical reactions were set up for a total of 600mg Mogroside IIIE used as starting material. The reaction was stopped after 24 hours. Insoluble starch was removed by centrifugation (4000 rpm for 5 min, Eppendorf). The supernatant was heated to 80 °C for 30 minutes with stirring (500 rpm), followed by centrifugation (4000 rpm for 10 min, Eppendorf). The supernatant was filtered through a 250ml, 0.22 micron PES and checked by LC-MS (Sweet Naturals 2016-Enzymatic_2016Q4_A.SPL, line 1254) to obtain HPLC data

[0351] The AGY15763.1 protein (SEQ ID NO: 434) can be encoded by the native gDNA (SEQ ID NO: 437) or codon optimized (for *E. coli*) DNA sequence (SEQ ID NO: 438)

[0352] An example of additional glycosytransferase expected to perform similarly is the UGT76G1 protein from *Stevia rebaudiana* (SEQ ID NO: 439), which can be expressed in *E. coli*. The native coding sequence for UGT76G1 (SEQ ID NO: 439) is provided in SEQ ID NO: 440).

Example 49: UDP-glycosyltransferases UGT73C5 in the presence of Mogrol

[0353] Mogrol was reacted with UDP-glycosyltransferases which produced Mogroside I and Mogroside II. 1mg/ml of Mogrol was reacted with 200ul crude extract containing UGT73C5 (*A. thaliana*)(334), 2ul crude extract containing sucrose synthase, 5mM UDP, 1x M221 protease inhibitor, 200mM sucrose, 0.5mg/ml spectinomycin, in 0.1M Tris-HCl pH7.0, incubated at 30C. Samples were taken after 2 days for HPLC. The reaction products were from Mogrol to Mogroside I and Mogroside II as shown in Figures 4 and 5.

[0354] The protein sequence of UGT73C5 is shown in SEQ ID NO: 441, the native DNA coding sequence for UGT73C5 (SEQ ID NO: 441) is shown in SEQ ID NO: 442, and the UGT73C5 coding sequence (Codon optimized for *E. coli*) is shown in SEQ ID NO: 443.

Example 50: UDP-glycosyltransferases (UGT73C6) in the presence of Mogrol to produce Mogroside I

[0355] Reaction conditions: 1mg/ml of Mogrol was reacted with 200ul crude extract containing UGT73C6, 2ul crude extract containing sucrose synthase, 5mM UDP, 1x M221 protease inhibitor, 200mM sucrose, 0.5mg/ml spectinomycin, in 0.1M Tris-HCl pH7.5, incubated at 30C. Samples were taken after 2 days for HPLC. The reaction product was Mogroside I from Mogrol. As shown in the HPLC data and Mass spectroscopy data of Figure 6.

[0356] The protein and gDNA sequence encoding *A. thaliana* UGT73C6 is shown in SEQ ID NO: 444 and SEQ ID NO: 445, respectively.

Example 51: UDP-glycosyltransferases (338) in the presence of Mogrol to produce Mogroside I, Mogroside IIA and two different Mogroside III products

[0357] Reaction conditions: 1mg/ml of Mogrol or Mogroside IIA was reacted with 200ul crude extract containing 338, 2ul crude extract containing sucrose synthase, 5mM UDP, 1x M221 protease inhibitor, 200mM sucrose, 0.5mg/ml spectinomycin, in 0.1M Tris-HCl pH8.5, incubated at 30C. Samples were taken after 2 days for HPLC

[0358] Mogrol reaction with *Bacillus* sp. UDP-glycotransferase (338) (described in Pandey et al., 2014; incorporated by reference in its entirety herein) led to the reaction products: Mogroside I, Mogroside IIA, and 2 different Mogroside III products. Figures 7-9 show the HPLC and mass spectroscopy data for the products obtained after the reaction. Figure 8 shows the peaks which correlate to the size of Mogrol IIA.

[0359] The protein and gDNA sequence encoding UGT 338 is provided in SEQ ID NO: 405 and SEQ ID NO: 406, respectively.

Example 52: UDP-glycosyltransferases (301 (UGT98)) in the presence of Mogroside IIIE to produce Siamenoside I and Mogroside V

[0360] Reaction conditions: 1mg/ml of Mogroside IIIE was reacted with 200ul crude extract containing 301, 2ul crude extract containing sucrose synthase, 5mM UDP, 1x M221 protease inhibitor, 200mM sucrose, 0.5mg/ml spectinomycin, in 0.1M Tris-HCl pH7.0, incubated at 30C. Samples were taken after 2 days for HPLC and mass spec analysis. The reaction products from Mogroside IIIE were Siamenoside I and Mogroside V as shown in Figures 10-11.

[0361] The protein and gDNA sequence encoding *S. grosvenorii* 301 UGT98 is provided in SEQ ID NO: 407 and SEQ ID NO: 408, respectively.

Example 53: UDP-glycosyltransferases (339) in the presence of Mogrol, Siamenoside I or Compound 1 to produce Mogroside I from Mogrol, Isomogroside V from Siamenoside I and Compound 1 derivative from Compound 1

[0362] Reaction conditions: 1mg/ml of Mogrol, Siamenoside I or Compound 1 was reacted with 200ul crude extract containing 339 (described in Itkin *et al.*, incorporated by

reference in its entirety herein), 2ul crude extract containing sucrose synthase, 5mM UDP, 1x M221 protease inhibitor, 200mM sucrose, 0.5mg/ml spectinomycin, in 0.1M Tris-HCl pH7.0, incubated at 30C. Samples were taken after 2 days for HPLC

[0363] The reaction products from Mogrol lead to Mogroside I, Siamenoside I lead to Isomogroside V, and Compound 1 led to a Compound 1 derivative (Figures 12-14).

[0364] The protein and DNA sequence encoding *S. grosvenorii* UGT339 is provided in SEQ ID NO: 409 and SEQ ID NO: 410, respectively.

Example 54: UDP-glycosyltransferases (330) in the presence of Mogroside IIA, Mogroside IIE, Mogroside IIIE, Mogroside IVA, or Mogroside IVE to produce Mogroside IIIA, Mogroside IVE, and Mogroside V

[0365] As described herein, the use of UDP-glycotransferase (330) as described in Noguchi *et al.* 2008 (incorporated by reference in its entirety herein) led to the reaction products Mogroside IIIA, Mogroside IVE, Mogroside V. The native host is *Sesamum indicum*, and the production host was SF9. For the reaction, 1mg/ml of Mogroside IIA, Mogroside IIE, Mogroside IIIE, Mogroside IVA, or Mogroside IVE was reacted with 200ul crude extract containing 330, 2ul crude extract containing sucrose synthase, 5mM UDP, 1x M221 protease inhibitor, 200mM sucrose, 0.5mg/ml spectinomycin, in 0.1M Tris-HCl pH7.0, incubated at 30C. Samples were taken after 2 days for HPLC.

[0366] The reaction led to surprising products such as Mogroside IIIA, Mogroside IVE, Mogroside V. As shown in Figures 15-20, the sizes of the compounds produced correspond to Mogroside IIIA, Mogroside IVE, and Mogroside V.

[0367] The protein and gDNA sequence encoding the *S. grosvenorii* UGT330 protein is provided in SEQ ID NO: 411 and SEQ ID NO: 412, respectively.

Example 55: UDP-glycosyltransferases (328) (described in Itkin et al) in the presence of Mogroside IIA, Mogroside IIE, Mogroside IIIE, Mogroside IVA, or Mogroside IVE to produce Mogroside IIIA, Mogroside IVE, and Mogroside V

[0368] Reaction conditions: 1mg/ml of Mogroside IIIE was reacted with 200ul crude extract containing 330, 2ul crude extract containing sucrose synthase, 5mM UDP, 1x M221

protease inhibitor, 200mM sucrose, 0.5mg/ml spectinomycin, in 0.1M Tris-HCl pH7.0, incubated at 30C. Samples were taken after 2 days for HPLC

[0369] The reaction products were Mogroside IVE and Mogroside V. As shown in Figures 21-22, the size of the products in the mass spectroscopy data corresponds to Mogroside IVE and Mogroside V. *S. grosvenorii* UGT328 protein (glycosyltransferase) and coding sequence thereof is provided in SEQ ID NO: 413 and 414, respectively.

[0370] The sucrose synthase AtSus1 protein sequence and the gDNA encodes the AtSus1 protein are provided in SEQ ID NO: 415 and 416, respectively.

Example 56: Mogrol production in Yeast

[0371] DNA was obtained through gene synthesis either through Genescript or IDT. For some of the cucurbitadienol synthases, cDNA or genomic DNA was obtained through 10 – 60 day old seedlings followed by PCR amplification using specific and degenerate primers. DNA was cloned through standard molecular biology techniques into one of the following overexpression vectors: pESC-Ura, pESC-His, or pESC-LEU. Saccharomyces cerevisiae strain YHR072 (heterozygous for erg7) was purchased from GE Dharmacon. Plasmids (pESC vectors) containing Mogrol synthesis genes were transformed/co-transformed by using Zymo Yeast Transformation Kit II. Strains were grown in standard media (YPD or SC) containing the appropriate selection with 2% glucose or 2% galactose for induction of heterologous genes at 30C, 220rpm. When indicated, lanosterol synthase inhibitor, Ro 48-8071 (Cayman Chemicals) was added (50ug/ml). Yeast production of mogrol and precursors were prepared after 2 days induction, followed by lysis (Yeast Buster), ethyl acetate extraction, drying, and resuspension in methanol. Samples were analyzed through HPLC.

[0372] Production of cucurbitadienol was catalyzed by cucurbitadienol synthase *S. grosvernorii* SgCbQ in growth conditions with no inhibitor.

[0373] Production of cucurbitadienol is shown in the HPLC and mass spectroscopy data which show mass peaks for the indicated product (Figure 23). The protein sequence and DNA sequence encoding *S. grosvernorii* SgCbQ are provided in SEQ ID NO: 446 and 418, respectively.

[0374] Cpep2 was also used for the production of cucurbitadienol in yeast. As shown in Figure 24, is the mass spectroscopy profile which shows peaks and characteristic fragments

that correspond with cucurbitadienol. Protein sequence of Cpep2 and DNA sequence encoding Cpep2 protein is provided in SEQ ID NO: 420 and 421, respectively.

[0375] *Cucurbita pepo* (Jack O' Lantern) Cpep4 was also used in the production of cucurbitadienol under growth conditions with no inhibitor. Production of cucurbitadienol is shown in the mass spectral data shown in Figures 24 and 25. As shown the peaks and fragments correspond to cucurbitadienol. The protein sequence and DNA sequence encoding Cpep4 are provided in SEQ ID NOs: 422 and 423, respectively.

[0376] A putative cucurbitadienol synthase protein sequence representing Cmax was obtained from native host *Cucurbita maxima.* The deduced coding DNA sequence will be used for gene synthesis and expression. The cucurbitadienol synthase sequences for the protein and DNA encoding the cucurbitadienol synthase is shown below:

[0377] Proteins and DNA coding sequences below were obtained through alignment of genomic DNA PCR product sequence with known cucurbitadienol synthase sequences available through public databases (Pubmed). It is expected that any one of these Cmax proteins may be used in the methods, systems, compositions (e.g., host cells) disclosed herein to produce Compound 1. A non-limiting exemplary Cmax protein is Cmax1 (protein) (SEQ ID NO: 424) encoded by Cmax1 (DNA) (SEQ ID NO: 425).

[0378] A putative cucurbitadienol synthase protein sequence representing Cmos1 was obtained from native host *Cucurbita moschata.* The deduced coding DNA sequence is used for gene synthesis and expression. Protein(s) and DNA coding sequence(s) shown below were obtained through alignment of genomic DNA PCR product sequence with known cucurbitadienol synthase sequences available through public databases (Pubmed). Any one of these Cmos proteins may be used in the methods, systems, compositions (e.g., host cells) disclosed herein to produce Compound 1. A non-limiting exemplary Cmos1 protein is Cmos1 (protein) (SEQ ID NO: 426) encoded by Cmos1 (DNA) (SEQ ID NO: 427).

Example 57: Production of dihydroxycucurbitadienol in yeast (cucurbitadienol synthase & epoxide hydrolase)

[0379] The production of dihydroxycucurbitadienol in yeast was considered using cucurbitadienol synthase & epoxide hydrolase. The native host for these enzymes is *S. Grosvenorii.*

[0380] Growth conditions: SgCbQ was co-expressed with an epoxide hydrolase (EPH) in the presence of lanosterol synthase inhibitor.

[0381] Possible dihydroxycucurbitadienol product is shown in Figure 26.

[0382] EPH protein sequence and a DNA encoding EPH protein (codon optimized *S. cerevisiae*) is provided in SEQ ID NO: 428 and 429, respectively.

Example 58: Production of Mogrol from cucurbitadienol synthase, epoxide hydrolase, cytochrome P450 and cytochrome P450 reductase

[0383] Four enzymes, including Cucurbitadienol synthase, epoxide hydrolase, cytochrome P450, and cytochrome P450 reductase are co-expressed in *S. cerevisiae.* For the growth conditions SgCbQ, EPH, CYP87D18 and AtCPR (cytochrome P450 reductase from *A. thaliana*) are co-expressed in the presence of lanosterol synthase inhibitor. Production of mogrol by *S. cerevisiae* is expected. The protein sequence and DNA sequence encoding SgCbQ, EPH, CYP87D18 and AtCPR (cytochrome P450 reductase from *A. thaliana*) are: CYP87D18 (protein) (SEQ ID NO: 430), and CYP87D18 (DNA) (SEQ ID NO: 431); and AtCPR (protein) (SEQ ID NO: 432), and AtCPR (DNA) (SEQ ID NO: 433).

Example 59: Compound 1 is tolerant to microbial hydrolysis

[0384] Yeast strains *Saccharomyces cerevisiae*, *Yarrowia lipolytica* and *Candida bombicola*, were incubated in YPD supplemented with 1mg/ml Mogroside V or Compound 1. After 3 days, supernatants were analyzed by HPLC.

[0385] As shown in the HPLC data, epoxide hydrolase hydrolyzed Mogroside V to Mogroside IIIE. There was no hydrolysis products observed with Compound 1 (Figure 37).

Example 60: *Streptococcus mutans* Clarke ATCC 25175 Dextransucrase

[0386] *Streptococcus mutans* Clarke can be grown anaerobically with glucose supplementation. An example of growth conditions can be found in Wenham, Henessey and Cole (1979), in which the method is used to stimulate dextransucrase production, for example. 5mg/ml Mogroside IIIE was added to the growth media. Time point samples to monitor production can be taken for HPLC, for example. Sequences for various dextransucrase can be found in the Table 1, which include protein sequences for dextransucrases and nucleic acid sequences that encode dextransucrases (for example, SEQ ID NOs: 157-162). In some embodiments, the dextransucrase comprises, or consists of, an amino acid sequence of any one of SEQ ID NOs: 2, 103, 106-110, 156, and 896. In some embodiments, the DexT can comprises an amino acid sequence set forth in SEQ ID NO: 103. In some embodiments, the DexT comprises a nucleic acid sequence set forth in SEQ ID NO: 104 or 105. In some embodiments, herein the recombinant cell encodes a protein comprising the sequence set forth in any one of SEQ ID NO: 156-162 and/or comprises a nucleic acid encoding dextransucrase comprising a nucleic acid sequence set forth in any one of SEQ ID NOs: 157-162. This example is used to produce Compound 1.

Example 61: 90% pure Compound 1 production procedure and sensory evaluation

[0387] A fraction containing the mixture of 3 α-mogroside isomers is obtained by treating mogroside III_E ($MIII_E$) with Dextransucrase/dextranase enzymes reaction followed by SPE fractionation. Based on UPLC analysis this mixture has 3 isomers, 11-oxo-Compound 1, Compound 1 and mogroside V isomer in 5:90:5% ratios respectively. These 3 isomers are characterized from the purification of a different fraction/source by LC-MS, 1D and 2D NMR spectra and by the comparison of closely related isomers in mogrosides series reported in the literature. This sample is further evaluated in sensory by comparing with pure Compound 1 sample using a triangle test.

Enzyme reaction and purification procedure

[0388] 100 mL of pH 5.5 1M sodium acetate buffer, 200g sucrose, 100 mL dextransucrase DexT (1mg/ml crude extract, pET23a, BL21-Codon Plus-RIL, grown in 2X YT), 12.5g of Mogroside III_E and 600 mL water were added to a 2.8L shake flask, and the flask was shaken at 30°C, 200 rpm. The progress of the reaction was monitored periodically by LC-MS.

After 72 hours, the reaction was treated with 2.5 mL of dextranase (Amano) and continued shaking the flask at 30°C. After 24 hours the reaction mixture was quenched by heating at 80°C and centrifuged at 5000 rpm for 5 minutes and the supernatant was filtered and loaded directly onto a 400g C18 SPE column and fractionated using MeOH: H_2O 5/25/50/75/100 step-gradient. Each step in the gradient was collected in 6 jars, with 225 mL in each jar. The desired products were eluted in the second jar of the 75% MeOH fraction (SPE 75_2) and dried under reduced pressure. It was further re-suspended/dissolved in 7 mL of H_2O, freezed and lyophilized the vial for 3 days to get 1.45g of white solid.

[0389] As per the UPLC analysis (Figures 28 and 29), the mixture has 3 characterized α-mogroside isomers; 11-oxo-Compound 1, Compound 1 and mogroside V isomer in 5:90:5 % ratios, respectively. No residual solvent and/or structurally unrelated impurities were observed based on ^{1}H and ^{13}C-NMR (Pyridine-d_5 + D_2O) analysis.

Sensory *evaluation*

[0390] Triangle testing for pure Compound 1 vs. 90% pure Compound 1 was performed on November 10, 2016. Two different compositions: (1) LSB + 175ppm pure Compound 1 (standard) and (2) LSB + 175ppm 90% pure Compound 1 were tested. All samples of compositions were made with Low Sodium Buffer (LSB) pH ~7.1 and contain 0% ethanol.

[0391] Conclusions: Panelists found that composition (1) LSB + 175ppm pure_Compound 1 (standard) was not significantly different than composition (2) LSB + 175ppm 90% pure Compound **1** (test) ($p>0.05$). Some of the testing analytical results are shown in Tables 2-4.

Table 2. Frequency of panelists that correctly selected the different sample. n = 38 (19 panelists x 2 reps).

Samples	Total
Incorrect	24
Correct	14
Total	38
Correct Sample Selected (p-value)	0.381

Table 3. Analytical Results: Test Day

Theoretical # (μM)	Observed (μM)
175ppm (155.51uM) pure_compound 1 (standard)	132.20 ± 1.54 (n=2)
175ppm (155.56uM) 90%_pure_compound 1 (test)	157.62 ± 0.63 (n=2)

Table 4. Analytical results: the day before the testing day

Theoretical # (μM)	Observed (μM)
175ppm (155.51uM) pure_compound 1 (standard)	134.48 ± 7.31 (n=2)
175ppm (155.56uM) 90%_pure_compound 1 (test)	140.69 ± 4.34 (n=2)

Example 62: Gene expression in recombinant yeast cells

[0392] DNA was obtained through gene synthesis either through Genescript, IDT, or Genewiz. For some of the cucurbitadienol synthases, cDNA or genomic DNA was obtained through 10 – 60 day old seedlings followed by PCR amplification using specific and degenerate primers. DNA was cloned through standard molecular biology techniques or through yeast gap repair cloning (Joska et al., 2014) into one of the following overexpression vectors: pESC-Ura, pESC-His, or pESC-LEU. Gene expression was regulated by one of the following promoters; Gal1, Gal10, Tef1, or GDS. Yeast transformation was performed using Zymo Yeast Transformation Kit II. Yeast strains were grown in standard media (YPD or SC) containing the appropriate selection with 2% glucose or 2% galactose for induction of heterologous genes. Yeast strains were grown in shake flask or 96 well plates at 30°C, 140-250 rpm. When indicated, lanosterol synthase inhibitor, Ro 48-8071 (Cayman Chemicals) was added (50ug/ml). Yeast production of mogrol and precursors were prepared through lysis (Yeast Buster), ethyl acetate extraction, drying, and resuspension in methanol. Samples were analyzed through LCMS methods described below using A/B gradient (A = H_2O, B = acetonitrile):

[0393] For analyzing diepoxysqualene, the LCMS method included the use of C18 2.1 x 50 mm column, 5% B for 1.5min, gradient 5% to 95% B or 5.5 min, 95% B for 6 min, 100% B for 3 min, 5% B for 1.5, and all at flow rate of 0.3 ml/min.

[0394] For analyzing cucurbitadienol, the first LCMS method included the use of C4 2.1 x 100 mm column, gradient 1 to 95% B for 6 minutes, and at flow rate of 0.55 ml/min; and the second LCMS method included the use of Waters Acquity UPLC Protein BEH C4

2.1x100mM, 1.7um, with guard, 62 to 67% B for 2 min, 100% B for 1 min, and at flow rate of 0.9 ml/min.

[0395] For analyzing 11-OH cucurbitadienol, the LCMS method included the use of C8 2.1 x 100 mm column, gradient 60 to 90% B for 6 minutes at flow rate of 0.55 ml/min

[0396] For analyzing Mogrol, the LCMS method included the use of C8 2.1 x 100 mm column, gradient 50 to 90% B for 6 minutes at flow rate of 0.55 ml/min

[0397] For analyzing Mogroside IIIE & Compound 1, the LCMS method included the use of Fluoro-phenyl 2.1 x 100 mm column, gradient 15 to 30% B for 6 minutes, at flow rate of 0.55ml/min.

Example 63:

Step1. Boosting oxidosqualene availability

[0398] *Saccharomyces cerevisiae* strain YHR072 (heterozygous for lanosterol synthase *erg7*) was purchased from GE Dharmacon. Expression of active *erg7* gene was reduced by replacing the promoter with that of *cup1* (Peng et al., 2015). A truncated yeast HMG-CoA reductase (tHMG-CoA) under control of GDS promoter and yeast squalene epoxidase (*erg1*) under the control of Tef1 promoter was integrated into the genome. Oxidosqualene boost was monitored by the production of diepoxysqualene as shown in the HPLC and UV absorbance (Figure 31).

tHMG-CoA (protein) SEQ ID NO:898 (pathway 1)

tHMG-CoA (DNA) SEQ ID NO:897 (pathway 1)

Erg1 (protein) SEQ ID NO: 900; Erg1 (DNA) SEQ ID NO: 899

[0399] In some embodiments, tHMG-CoA enzyme is used for the production of diepoxysqualene.

[0400] Genes encoding for putative squalene epoxidases in *S. grosvenorii* (Itkins et al., 2016) were selected to test for boosting oxidosqualene/diepoxysqualene production. The sequences of 3 squalene epoxidases can be found in Table 1 for their amino acids and the coding sequence (SEQ ID NO: 50-56, 60, 61, 334 or 335). Additional sequences for squalene epoxidases suitable to use in the methods, systems and compositions disclosed herein for producing oxidosqualene and/or diepoxysqualene, and for boosting the production of

oxidosqualene and/or diepoxysqualene include: SQE1 (protein) SEQ ID NO: 908, SQE1 (DNA) SEQ ID NO: 909; SQE2 (protein) SEQ ID NO: 910, SQE2 (DNA) SEQ ID NO: 911; SQE3 (protein) SEQ ID NO: 912, and SQE3 (DNA) SEQ ID NO: 913.

Step2. Cucurbitadienol production

Cucurbitadienol synthase enzymes

[0401] Plasmids containing *S. grosvernorii* cucurbitadienol synthase gene (SgCbQ) were transformed into yeast strain with oxidosqualene boost. Strains were grown 1-3 days at 30 C, 150-250rpm. Production of cucurbitadienol is shown in the HPLC and mass spectroscopy data which show mass peaks for the indicated product (Figure 23). The SgCbQ protein and gDNA encoding SgCbQ is provided in SEQ ID NO: 446 and SEQ ID NO: 418, respectively. *Cucurbita pepo* (Jack O' Lantern) protein Cpep2 was also used for the production of cucurbitadienol in yeast. Figure 24 shows the mass spectroscopy profile which contains peaks and characteristic fragments that correspond with cucurbitadienol. The Cpep2 protein and DNA encoding Cpep2 is provided in SEQ ID NO: 420 and SEQ ID NO: 421, respectively. *Cucurbita pepo* (Jack O' Lantern) protein Cpep4 was also used in the production of cucurbitadienol. The host cells were cultiveated under the growth conditions with no inhibitor. Production of cucurbitadienol is demonstrated in the mass spectral data shown in Figure 25. As shown, the peaks and fragments correspond to cucurbitadienol. The Cpep4 protein and DNA encoding Cpep4 is provided in SEQ ID NO: 422 and SEQ ID NO: 423, respectively. The *Cucurbita maxima* protein Cmax was also used for the production of cucurbitadienol in yeast. Figure 32 shows the mass spectroscopy profile which contains peaks and characteristic fragments that correspond with cucurbitadienol. The Cmax1 protein sequence is provided in SEQ ID NO: 424, and the coding sequence for Cmax1 (DNA) is provided in SEQ ID NO: 425. *Cucumis melo* protein Cmelo was also used for the production of cucurbitadienol in yeast. Figure 32 shows the mass spectroscopy profile which contains peaks and characteristic fragments that correspond with cucurbitadienol. The Cmelo protein sequence is provided in SEQ ID NO: 902, and the coding sequence for Cmelo (DNA) is provided in SEQ ID NO: 901. It is expected that *Cucurbita moschata* protein Cmos1 can also be used for the production of cucrbitodienol in recombinant host cells, for example yeast cells. Cmos1 sequences Cmos1 (protein) (SEQ ID NO: 426) and Cmos1 (DNA) (SEQ ID NO: 427) were obtained through alignment of genomic DNA PCR product sequence with known

cucurbitadienol synthase sequences available through public databases (Pubmed). It is expected that Cmost 1 protein (SEQ ID NO: 426) can be used for the production of cucurbitadienol in recombinant host cells, for example yeast cells.

Converting other oxidosqualene cyclases into a cucurbitadienol synthase

[0402] Plasmids containing modified oxidosqualene genes were transformed into yeast strain with oxidosqualene boost. Strains were grown 1-3 days at 30C, 150-250rpm.

[0403] The protein PSX Y118L from the native host *Pisum sativum* was also used for the production of cucurbitadienol in yeast. Figure 33 shows the mass spectroscopy profile which contain peaks and characteristic fragments that correspond with cucurbitadienol when the tyrosine at position 118 is converted into leucine. The sequences for the protein and DNA encoding the modified oxidosqualene cyclase are: PSXY118L (protein) (SEQ ID NO: 904) and PSXY118L (DNA, codon optimized) (SEQ ID NO: 903).

The oxidosqualene cyclase from *Dictyostelium sp.* was also used for the production of cucurbitadienol in yeast. As shown in Figure 34, the HPLC peak of cucurbitadienol is shown when the tyrosine at position 80 is converted into leucine. The sequences for the protein and DNA encoding the modified oxidosqualene cyclase are: DdCASY80L (protein) (SEQ ID NO: 906) and DdCASY80L (DNA) (SEQ ID NO: 905).

Improving cucurbitadienol synthase activities

[0404] The gene encoding for a cucurbitadienol synthase form *Cucumis melo* was codon optimized (SEQ ID: 907) and used as a starting point for generating a library of modifications. Modifications were introduced through standard molecular biology techniques consisting of fusion peptides at the N-terminus (i.e., 5') or C-terminus (i.e., 3') end of the enzyme. Plasmids libraries of modified cucurbitadienol synthase genes were transformed into a yeast strain with oxidosqualene boost. Enzyme activities were measured by ratios of peak heights or areas of 409 and 427 positive mass fragments at the expected retention times for cucurbitadienol vs. an internal standard using LCMS method 2 described above. Enzyme performance were scored as average % activities over the average activities of the parent enzyme (n=8). Step 1 sequences of the enzymes and the sequences that encode the enzyme can be found in SEQ ID NOs: 951-1012. Step 1 sequence also include the fusions SS2c-G10, SS2e-A7b, SS2d-G11, SS2e-A7a, SS4d-G5, SS4d-C7, SS3b-D8, and SS2c-A10a as described in Table 1.

Step 3. Production of 11-OH cucurbitadienol

[0405] CYP87D18 (CYP450, *S. grosvenorii*) and SgCPR (CYP450 reductase, *S. grosvenorii*) were expressed in *S. cerevisiae* strain producing cucurbitadienol. 11-OH cucurbitadienol (i.e., 11-hydroxy cucurbitadienol) was observed using HPLC and mass spectroscopy data (Figure 36). *S. grosvenorii* CYP87D18 protein sequence is shown in SEQ ID NO: 872, and CYP87D18 (codon optimized DNA) coding sequence is shown in SEQ ID NO: 871. *S. grosvenorii* SgCPR protein sequence is shown in SEQ ID NO: 874, and SgCPR1 (codon optimized DNA) coding sequence is shown in SEQ ID NO: 873.

[0406] Additional CYP450s from *S. grosvenorii* and *Glycyrrhiza* (CYP88D6) were expressed in *S. cerevisiae* strain producing cucurbitadienol. Protein sequences and DNA coding sequences for the enzymes are provided in SEQ ID NOs: 875-890.

Step 4. Production of mogrol

[0407] CYP1798 (CYP450 enzyme, *S. grosvenorii*) and EPH2A (epoxide hydrolase, *S. grosvenorii*) were expressed in *S. cerevisiae* strain producing 11-OH cucurbitadienol. Mogrol was observed using HPLC and mass spectroscopy data (Figure 36). For sequences, DNA coding and protein sequences for the enzymes are provided in SEQ ID NOs: 891-894.

Epoxidation of cucurbitadienol and/or 11-OH cucurbitadienol

[0408] Additional CYP450s and SQEs from *S. grosvenorii* and *Glycyrrhiza* (CYP88D6) were also expressed in *S. cerevisiae* strain producing cucurbitadienol or 11-OH cucurbitadienol to test for epoxidation.

[0409] For SQEs, protein and DNA coding sequences for the enzymes are provided in SEQ ID NOs: 882-888. For CYP450s, protein and DNA coding sequences for the enzymes are provided in SEQ ID NOs: 875-890.

Step 7: Production of Compound 1 from mogroside IIIE in S. cerevisiae.

[0410] *S. cerevisiae* strain expressing a truncated dextransucrase (tDexT) was incubated in YPD (30C, 250rpm) containing 7mg/ml Mogroside V for 1-2 day resulting in hydrolysis to Mogroside IIIE. The *S. cerevisiae* cells were harvested, lysed, and then mixed back with the YPD supernatant containing Mogroside IIIE. To initiate the dextransucrase reaction, sucrose was added to a final concentration of 200g/L, followed by incubation at 30C, 250rpm for 2 days. Production of Compound 1 was observed using HPLC (Figures 37). Protein

sequence for tDexT is shonw in SEQ ID NO: 896, and the DNA coding sequence for tDexT is shown in SEQ ID NO. 895.

Example 64

[0411] *S. cerevisiae* or *Y. lipolytica* was grown in the presence of Mogroside V to allow the hydrolytic enzymes in the yeast to generate Mogroside IIIE. After 1 or 2 days, the cells were lysed in analyzed by HPLC to determine the mogroside content. After 1 day of incubation, *S. cerevisiae* produced a mixture of Mogroside V, Mogrosides IV, and Mogroside IIIE. After 2 days of incubation, substantially all of the mogrosides were converted to Mogroside IIIE as shown in Figure 40A.

[0412] Similarly, after 2 days of incubation *Y. lipolytica* produced mostly Mogroside IIIE (shown in Figure 40B).

Example 65

[0413] *S. cerevisiae* or *Y. lipolytica* was grown in the presence of Compound 1. Unlike other mogrosides (*see* Example 64), no hydrolysis products due to hydrolysis of Compound 1 was observed as shown in Figure 41.

Example 66

[0414] *S. cerevisiae* was modified to overexpress a dextransucrase (DexT). This modified strain was grown in the presence of a mogrosides mixture to allow the hydrolytic enzymes in *S. cerevisiae* to generate Mogroside IIIE. After 2 days of incubation, the cells were lysed to release the DexT enzyme and supplemented with sucrose. After 24 hours, significant amounts of Compound 1 was produced (shown in Figure 42)

Example 67: Generation of fusion proteins having cucurbitadienol synthase activity

[0415] A collection or library of *S. cerevisiae* in-frame fusion polynucleotdies for a cucurbitadienol synthase gene (DNA coding sequence provided in SEQ ID NO: 907, and protein sequence provided in SEQ ID NO: 902) was prepared. The in-frame fusion polynucleotides were cloned into a yeast vector molecule to generate fusion proteins.

[0416] Various fusion proteins were generated and tested for cucurbitadienol synthase activities. The testing results for some of the fusion protein generated in this example are shown in Table 2.

Table 2. Cucurbitadienol synthase activities for the fusion proteins

SEQ ID NO for fusion protein	Activity (as compared to the parent)	SEQ ID NO for fusion protein	Activity (as compared to the parent)
1024	166%	851	142%
854	135%	856	123%
859	105%	862	102%
865	125%	867	145%
915	124%	920	124%
924	121%	928	117%
932	128%	936	126%
940	109%	944	107%
948	102%	952	90%
956	85%	959	46%
964	74%	967	72%
971	89%	975	35%
979	96%	983	80%
987	111%	991	114%
995	124%	999	103%
1003	118%	1007	97%

Example 68: UDP-glycosyltransferases (311 enzyme, SEQIDs: 436-438) in the presence of Mogroside IIIE, Mogroside IVE or Mogroside IVA to produce Mogroside IV and Mogroside V isomers

[0417] Reaction conditions: To a 50ml Falcon tube with 17ml water, 3 ml of pH 7.0 1M Tris-HCl, 0.12g UDP (Carbosynth), 3g sucrose, 300ul of protease inhibitor 100x M221, 150ul of Kanamycin (50mg/ml), 1.185ml sucrose synthase Sus1 (1mg/ml crude extract), 150mg of starting Mogrosides, and 6ml 311 enzyme (1mg/ml crude extract) were added and incubated at 30°C, 150rpm. The progress of the reaction was monitored periodically by LC-MS. After 3 days, the reaction was stopped by heating to 80°C for 30 minutes with stirring (500 rpm). The reaction was then centrifuged (4000 rpm for 10 min, Eppendorf) and the supernatant was filtered through a 50 ml, 0.22 micron PVDF. The reaction products identified are depicted in Figure 43.

Table 1: Some protein and DNA sequences disclosed herein

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Cyclomaltodextrin glucanotransferase (CGTase; *Bacillus*)	MKEKDKLKVNRNNVNFSKDIIYQIVTDRFHNGCPSYNPKGGLYDESRKNKKKYFGGDWIGIIEK LNTNYFTELGVTSLWISQPVENIFTPINDLVGSTSYHGYWARDFKRTNPFFGTFGDFQTLITTA HAKDIKIIMDFAPNHTSPALHDDATYAENGRLYDNGLLLGGYDNDYNHYFHHNGGTDFEEYEDG VYRNLFDLADLNHQNIAIDLYFKEAIKLWLDQGIDGIRVDAVKHMSYGWQKSWLNSIYNYRPVF IFGEWYINPNEYDHRNVHFANNSGMSLLDFSFAHKVREVFRDGMDSMHGLHKMIEETYQIYNDV NNLVTFIDNHDMDRFHINGQSKRRIEQSLVFLLTSRGIPSVYYGTEQYMVGNGDPNNRGQMESF DVNTDNFKIIQSLSSLRSLNYALPYGNTKERYITNDIYVYERYFGSDVVLIALNRNLTEGYEIK DVKTILPSRKYKDILDGLLDGEAIRVENNNIDSLWLGPGSGQVWHHKGVNSIPLIGTVGHKMTT VGQIICIEGCGFTSKKGSVLFEEKEAEVVSWSHTSIKVKVPAVNDGKYEITVVTDTGTRSNIYK HIEVLNTKQVCIRFVIENGYEIPESEVFIMGNTYSLGNMNPCKAVGPFFNQIMYQFPTGYFDIS VPADTLLEFKFIRKINNTLLIEGGENHKYRTPSFGTGEVVVKWQTAEKTILVES	1	
DexT protein	MPANAPDKQSVTNAPVVPPKHDTDQQDDSLEKQQVLEPSVNSNIPKKQTNQQLAVVTAPANSAP QTKTTAEISAGTELDTMPNVKHVDGKVYFYGDDGQPKKNFTTIIDGKPYYFDKDTGALSNNDKQ YVSELFSIGNKHNAVYNTSSDNFTQLEGHLTASSWYRPKDILKNGKRWAPSTVTDFRPLLMAWW PDKSTQVTYLNYMKDQGLLSGTHHFSDNENMRTLTAAAMQAQVNIEKKIGQLGNTDWLKTAMTQ YIDAQPNWNIDSEAKGDDHLQGGALLYTNSDMSPKANSDYRKLSRTPKNQKGQIADKYKQGGFE LLLANDVDNSNPVVQAEQLNWLHYMMNIGSILQNDDQANFDGYRVDAVDNVDADLLQIAGEYAK AAYGVDKNDARANQHLSILEDWGDEDPDYVKAHGNQQITMDFPLHLAIKYALNMPNDKRSGLEP TREHSLVKRITDDKENVAQPNYSFIRAHDSEVQTIIADIIKDKINPASTGLDSTVTLDQIKQAF DIYNADELKADKVYTPYNIPASYALLLTNKDTIPRVYYGDMFTDDGQYMAKQSPYYQAIDALLK ARIKYAAGGQTMKMNYFPDEQSVMTSVRYGKGAMTASDSGNQETRYQGIGLVVNNRPDLKLSDK DEVKMDMGAAHKNQDYRPVLLTTKSGLKVYSTDANAPVVRTDANGQLTFKADMVYGVNDPQVSG YIAAWVPVGASENQDARTKSETTQSTDGSVYHSNAALDSQVIYEGFSNFQDFPTTPDEFTNIKI AQNVNLFKDWGITSFEMAPQYRASSDKSFLDAIVQNGYAFTDRYDIGYNTPTKYGTADNLLDAL RALHGQGIQAINDWVPDQIYNLPDEQLVTAIRTDGSGDHTYGSVIDHTLYASKTVGGGIYQQQY GGAFLEQLKTQYPQLFQQKQISTDQPMNPDIQIKSWEAKYFNGSNIQGRGAWYVLKDWGTQQYF NVSDAQTFLPKQLLGEKAKTGFVTRGKETSFYSTSGYQAKSAFICDNGNWYYFDDKGKMVVGNQ VINGINYYFLPNGIELQDAYLVHDGMYYYYNNIGKQLHNTYYQDKQKNFHYFFEDGHMAQGIVT IIQSDGTPVTQYFDENGKQQKGVAVKGSDGHLHYFDGASGNMLFKSWGRLADGSWLYVDEKGNA VTGKQTINNQTVYFNDDGRQIKNNFKELADGSWLYLNNKGVAVTGEQIINGQTLYFGNDGRQFK GTTHINATGESRYYDPDSGNMITDRFERVGDNQWAYFGYDGVAVTGDRIIKGQKLYFNQNGIQM KGHLRLENGIMRYYDADTGELVRNRFVLLSDGSWVYFGQDGVPVTGVQVINGQTLYFDADGRQV KGQQRVIGNQRYWMDKDNGEMKKITYAAALEHHHHHH	2	
CGTase CGT-SL	MKRWLSVVLSMSLVFSAFFLVSDTQKVTVEAAGNLNKVNFTSDIVYQIVVDRFVDGNTSNNPSG SLFSSGCTNLRKYCGGDWQGIINKINDGYLTEMGVTAIWISQPVENVFAVMNDADGSTSYHGYW ARDFKKTNPFFGTLSDFQRLVDAAHAKGIKVIIDFAPNHTSPASETNPSYMENGRLYDNGTLIG	3	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GYTNDTNSYFHHNGGTTFSNLEDGIYRNLFDLADFNHQNQFIDKYLKDAIKLWLDMGIDGIRMD AVKHMPFGWQKSFMDEVYDYRPVFTFGEWFLSENEVDSNNHFFANESGMSLLDFRFGQKLRQVL RNNSDDWYGFNQMIQDTASAYDEVIDQVTFIDNHDMDRFMADEGDPRKVDIALAVLLTSRGVPN IYYGTEQYMTGNGDPNNRKMMTSFNKNTRAYQVIQKLSSLRRSNPALSYGDTEQRWINSDVYIY ERQFGKDVVLVAVNRSLSKSYSITGLFTALPSGTYTDQLGALLDGNTIQVGSNGAVNAFNLGPG EVGVWTYSAAESVPIIGHIGPMMGQVGHKLTIDGEGFGTNVGTVKFGNTVASVVSWSNNQITVT VPNIPAGKYNITVQTSGGQVSAAYDNFEVLTNDQVSVRFVVNNANTNWGENIYLVGNVHELGNW NTSKAIGPLFNQVIYSYPTWYVDVSVPEGKTIEFKFIKKDGSGNVIWESGSNHVYTTPTSTTGT VNVNWQY		
UGT73C3 protein	MATEKTHQFHPSLHFVLFPFMAQGHMIPMIDIARLLAQRGVTITIVTTPHNAARFKNVLNRAIE SGLAINILHVKFPYQEFGLPEGKENIDSLDSTELMVPFFKAVNLLEDPVMKLMEEMKPRPSCLI SDWCLPYTSIIAKNFNIPKIVFHGMGCFNLLCMHVLRRNLEILENVKSDEEYFLVPSFPDRVEF TKLQLPVKANASGDWKEIMDEMVKAEYTSYGVIVNTFQELEPPYVKDYKEAMDGKVWSIGPVSL CNKAGADKAERGSKAAIDQDECLQWLDSKEEGSVLYVCLGSICNLPLSQLKELGLGLEESRRSF IWVIRGSEKYKELFEWMLESGFEERIKERGLLIKGWAPQVLILSHPSVGGFLTHCGWNSTLEGI TSGIPLITWPLFGDQFCNQKLVVQVLKAGVSAGVEEVMKWGEEDKIGVLVDKEGVKKAVEELMG DSDDAKERRRRVKELGELAHKAVEKGGSSHSNITLLLQDIMQLAQFKN	4	SEQ ID NO:21 in WO2016050890 (which is incorporated by reference in its entirety)
UGT73C6 protein	MAFEKNNEPFPLHFVLFPFMAQGHMIPMVDIARLLAQRGVLITIVTTPHNAARFKNVLNRAIES GLPINLVQVKFPYQEAGLQEGQENMDLLTTMEQITSFFKAVNLLKEPVQNLIEEMSPRPSCLIS DMCLSYTSEIAKKFKIPKILFHGMGCFCLLCVNVLRKNREILDNLKSDKEYFIVPYFPDRVEFT RPQVPVETYVPAGWKEILEDMVEADKTSYGVIVNSFQELEPAYAKDFKEARSGKAWTIGPVSLC NKVGVDKAERGNKSDIDQDECLEWLDSKEPGSVLYVCLGSICNLPLSQLLELGLGLEESQRPFI WVIRGWEKYKELVEWFSESGFEDRIQDRGLLIKGWSPQMLILSHPSVGGFLTHCGWNSTLEGIT AGLPMLTWPLFADQFCNEKLVVQILKVGVSAEVKEVMKWGEEEKIGVLVDKEGVKKAVEELMGE SDDAKERRRRAKELGESAHKAVEEGGSSHSNITFLLQDIMQLAQSNN	5	SEQ ID NO: 23 in WO2016050890
UGT85C2 sequence	MDAMATTEKKPHVIFIPFPAQSHIKAMLKLAQLLHHKGLQITFVNTDFIHNQFLESSGPHCLDG APGFRFETIPDGVSHSPEASIPIRESLLRSIETNFLDRFIDLVTKLPDPPTCIISDGFLSVFTI DAAKKLGIPVMMYWTLAACGFMGFYHIHSLIEKGFAPLKDASYLTNGYLDTVIDWVPGMEGIRL KDFPLDWSTDLNDKVLMFTTEAPQRSHKVSHHIFHTFDELEPSIIKTLSLRYNHIYTIGPLQLL LDQIPEEKKQTGITSLHGYSLVKEEPECFQWLQSKEPNSVVYVNFGSTTVMSLEDMTEFGWGLA NSNHYFLWIIRSNLVIGENAVLPPELEEHIKKRGFIASWCSQEKVLKHPSVGGFLTHCGWGSTI ESLSAGVPMICWPYSWDQLTNCRYICKEWEVGLEMGTKVKRDEVKRLVQELMGEGGHKMRNKAK DWKEKARIAIAPNGSSSLNIDKMVKEITVLARN	6	SEQ ID NO: 25 in WO2016050890
UGT73C5 protein	MVSETTKSSPLHFVLFPFMAQGHMIPMVDIARLLAQRGVIITIVTTPHNAARFKNVLNRAIESG LPINLVQVKFPYLEAGLQEGQENIDSLDTMERMIPFFKAVNFLEEPVQKLIEEMNPRPSCLISD FCLPYTSKIAKKFNIPKILFHGMGCFCLLCMHVLRKNREILDNLKSDKELFTVPDFPDRVEFTR TQVPVETYVPAGDWKDIFDGMVEANETSYGVIVNSFQELEPAYAKDYKEVRSGKAWTIGPVSLC NKVGADKAERGNKSDIDQDECLKWLDSKKHGSVLYVCLGSICNLPLSQLKELGLGLEESQRPFI	7	SEQ ID NO: 22 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	WVIRGWEKYKELVEWFSESGFEDRIQDRGLLIKGWSPQMLILSHPSVGGFLTHCGWNSTLEGIT AGLPLLTWPLFADQFCNEKLVVEVLKAGVRSGVEQPMKWGEEEKIGVLVDKEGVKKAVEELMGE SDDAKERRRRAKELGDSAHKAVEEGGSSHSNISFLLQDIMELAEPNN		
UGT73E1 protein	MSPKMVAPPTNLHFVLFPLMAQGHLVPMVDIARILAQRGATVTIITTPYHANRVRPVISRAIAT NLKIQLLELQLRSTEAGLPEGCESFDQLPSFEYWKNISTAIDLLQQPAEDLLRELSPPPDCIIS DFLFPWTTDVARRLNIPRLVFNGPGCFYLLCIHVAITSNILGENEPVSSNTERVVLPGLPDRIE VTKLQIVGSSRPANVDEMGSWLRAVEAEKASFGIVVNTFEELEPEYVEEYKTVKDKKMWCIGPV SLCNKTGPDLAERGNKAAITEHNCLKWLDERKLGSVLYVCLGSLARISAAQAIELGLGLESINR PFIWCVRNETDELKTWFLDGFEERVRDRGLIVHGWAPQVLILSHPTIGGFLTHCGWNSTIESIT AGVPMITWPFFADQFLNEAFIVEVLKIGVRIGVERACLFGEEDKVGVLVKKEDVKKAVECLMDE DEDGDQRRKRVIELAKMAKIAMAEGGSSYENVSSLIRDVTETVRAPH	8	SEQ ID NO: 24 in WO2016050890
UGT98 protein	MDAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRRNFHIYFCSTSVNLDAIKPKLPSSSSSDSI QLVELCLPSSPDQLPPHLHTTNALPPHLMPTLHQAFSMAAQHFAAILHTLAPHLLIYDSFQPWA PQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFPISEFVLHDYWKAMYSAAGGAVTKKDHK IGETLANCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIKN WLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAIEDALPKGFL ERVGERGMVVKGWAPQAKILKHWSTGGFVSHCGWNSVMESMMFGVPIIGVPMHLDQPFNAGLAE EAGVGVEAKRDSDGKIQREEVAKSIKEVVIEKTREDVRKKAREMGEILRSKGDEKIDELVAEIS LLRKKAPCSI	9	SEQ ID NO: 53 in WO2016050890
UGT1495 gene sequence	ATGCTTCCATGGCTGGCTCACGGCCATGTCTCCCCTTTCTTCGAGCTCGCCAAGTTGCTCGCCG CTAGAAACTTCCACATATTCTTCTGCTCCACCGCCGTAAACCTCCGCTCCGTCGAACCAAAACT CTCTCAGAAGCTCTCCTCCCACGTGGAGCTGGTGGAGCTCAACCTACCGCCCTCGCCGGAGCTC CCTCCGCACCGCCACACCACCGCCGGCCTTCCACCGCACCTCATGTTCTCGCTCAAGCGAGCTT TCGACATGGCCGCTCCCGCCTTCGCCGCCATCCTCCGCGACCTGAACCCGGACTTGCTCATCTA CGACTTCCTGCAGCCGTGGGCGGCGGCGGAGGCTCTGTCGGCGGATATTCCGGCCGTGATGTTC AAAAGCACGGGTGCGCTCATGGCGGCCATGGTCGCGTACGAGCTGACGTTTCCGAACTCTGATT TTTTCTCGCTTTTCCCTGAGATTCGTCTCTCCGAGTGCGAGATTAAACAGCTGAAGAACTTGTT TCAATGTTCTGTGAATGATGCGAAAGACAAGCAAAGGATTAAGGGATGTTATGAGAGATCTTGC GGCATGATTTTGGTGAAATCTTTCAGAGAAATCGAAGGCAAATATATTGATTTTCTCTCTACTC TGCTGGGCAAGAAGGTTGTTCCAGTTGGTCCACTTGTTCAACAAACAGAAGACGACGTCGTATC AGGAAGTTTTGACGAATGGCTAAATGGAAAAGATAGATCGTCTTCCATACTCGTGTCTTTCGGA AGCGAGTTCTACCTGTCCAGAGAAGACATGGAAGAGATCGCGCATGGCTTAGAGCTGAGCCAGG TGAACTTCATATGGGTCGTCAGGTTTCCGGCGGGAGGAGAGAGAAACACGACAAAGGTGGAAGA AGAACTGCCAAAAGGGTTTCTAGAGAGAGTTAGAGAGAGAGGGATGGTGGTGGAGGGCTGGGCG CCGCAGGCTCAGATCTTGAAACATCCAAGCGTCGGCGGATTCCTCAGCCACTGCGGGTGGAGCT CCGTCGTGGAGAGCATGAAATTCGGCGTTCCGATCATCGCCATGCCGATGCACCTCGACCAGCC GCTGAATTCCCGGCTGGTCGAGCGGCTCGGCGTCGGCGTAGTGGTGGAGAGAGACGGCCGCCTC CGGGGAGAGGTGGAGAGAGTTGTCAGAGAGGTGGTGGTGGAGAAAAGTGGAGAGAGAGTGAGGA	10	SEQ ID NO: 27 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AGAAGGTGGAGGAGTTTGCAGAGATCATGAAGAAGAAAAAAGACAATGAAGAGATGGACGTAGT CGTGGAAGAGTTGGTGACGCTCTGCAGGAAGAAGAAGAAGGAGGAGGATTTACAGAGTAATTAT TGGTGCAGAACCGCCATTGATGACCATTGTTCTGAAGTCGTGAAGATTGAAGATGCTGCAGCAG CCGACGAGGAGCCTCTTTGCAAATAA		
UGT1817 gene sequence	ATGGCTGTCACTTACAGCCTGCACATAGCAATGTACCCTTGGTTTGCTTTCGGCCACTTGACTC CATTTCTCCAAGTCTCCAACAAGCTTGCCAAGGAAGGCCACAAAATCTCCTTCTTCATCCCAAC GAAAACGCTAACCAAATTGCAGCCTTTCAATCTCTTTCCAGATCTCATTACCTTTGTCCCCATC ACTGTTCCTCATGTTGATGGTCTCCCTCTTGGAGCTGAGACTACTGCTGATGTTTCTCACCCTT CACAGCTCAGTCTCATCATGACTGCTATGGATTGCACCCAACCCGAAATCGAGTGTCTTCTTCG AGACATAAAACCTGATGCCATCTTCTTCGATTTCGCGCACTGGGTGCCAAAATTGGCATGTGGA TTGGGCATTAAGTCGATTGATTACAGTGTCTGTTCTGCAGTATCAATTGGTTATGTTTTGCCCC TATTAAGGAAAGTTTGTGGACAAGATTTATTAACTGAAGATGATTTTATGCAGCCATCTCCTGG CTACCCGAGTTCCACCATCAATCTTCAAGCTCATGAGGCTCGATATTTTGCATCTCTGAGCCGC TGGAGGTTTGGCAGTGATGTCCCTTTCTTTAGTCGCCATCTTACTGCACTTAATGAATGCAATG CTTTAGCATTCAGGTCATGTAGGGAGATTGAAGGGCCTTTTATAGACTATCCAGAAAGTGAATT AAAAAAGCCTGTGTTGCTTTCCGGAGCAGTGGATCTACAACCGCCAACCACAACTGTAGAAGAA AGATGGGCAAAATGGCTATCAGGGTTCAACACCGACTCGGTCGTATATTGTGCATTTGGAAGTG AGTGTACCTTAGCAAAAGACCAATTCCAAGAACTGCTGTTGGGTTTTGAGCTTTCAAATATGCC ATTCTTTGCTGCACTTAAACCACCTTTTGGTGTTGACTCGGTTGAAGCAGCCTTGCCTGAAGGT TTTGAACAGAGAGTTCAGGGAAGAGGGGTGGTCTATGGGGGATGGGTCCAACAGCAGCTCATTT TGGAGCACCCATCAATTGGATGCTTTGTTACACATTGTGGATCAGGCTCCTTATCAGAGGCGTT AGTGAAGAAGTGTCAATTAGTGTTGTTACCTCGTATCGGTGACCACTTTTTCCGAGCAAGAATG TTGAGCAATTATTTGAAAGTTGGTGTGGAGGTAGAGAAAGGAGAAGGAGATGGATCTTTTACAA AGGAAAGTGTGTGGAAGGCAGTGAAGACAGTGATGGATGAAGAGAATGAAACTGGGAAAGAGTT CAGAGCGAACCGTGCCAAGATAAGAGAGCTATTGCTCGACGAAGATCTCGAGGAGTCTTATATC AACAATTTCATCCACAGCCTGCATACTTTGAATGCATGA	11	SEQ ID NO: 28 in WO2016050890
UGT5914 gene sequence	ATGGAAGCTAAGAACTGCAAAAAGGTTCTGATGTTCCCATGGCTGGCGCATGGTCACATATCAC CATTTGTAGAGCTGGCCAAGAAGCTCACAGACAACAACTTCGCCGTTTTTCTATGTTCTTCCCC TGCAAATCTTCAAAACGTCAAGCCAAAACTCCCCCATCACTACTCTGATTCCATTGAACTCGTG GAGCTCAACCTTCCATCGTCGCCGGAGCTTCCCCCTCATATGCACACCACCAATGGCCTCCCTT TGCATTTAGTTCCCACCCTCGTTGACGCCTTGGACATGGCCGCTCCGCACTTCTCCGCCATTTT ACAGGAACTGAATCCAGATTTTCTCATATTCGACATCTTCCAACCCTGGGCGGCTGAAATCGCT TCCTCCTTCGGCGTTCCTGCTATTTTGTTGCTTATCGTTGGATCTGCTATAACCGCTTTAGGGG TTCATTTTGTCCGGAGCTCCGGTACGGAATTCCCCTTTCCCGAGCTTACTAAATCATTCAAGAA GGAGGACGACCGAAAACCTCCAGGAGATTCCGGCAACGATAGAGGAAAACGGCTATTCAAATGT CTGCTGGACCTGGAACATTCTTCAGAGACTATTTTGGTGAACAGTTTTACAGAGATAGAGGGCA AATATATGGACTATCTCTCGGTCTTACTGAAGAAGAAGATCCTTCCGATTGGTCCTTTGGTTCA GAAAATTGGCTCCGATGACGATGAATCGGGAATCCTCCGGTGGCTTGACAAGAAGAAACCGAAT	12	SEQ ID NO: 30 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TCAACTGTGTACGTTTCGTTCGGGAGTGAGTACTATTTGAGCAAAGAAGACATAGCAGAGCTTG CGCATGGTCTGGAAATCAGCGGCGTCAATTTCATCTGGATTGTTCGGTTTCCAAAGGGAGAGAA AATCGCCATTGAAGAGGCATTACCAGATGAATTTCTTGAAAGAGTCGGAGAGAGAGGCGTCGTC GTTGATGGATGGGCGCCGCAGATGAAAATATTAGGGCATTCGAGCGTCGGCGGGTTTCTGTCTC ACTGCGGATGGAACTCTGTGCTGGAGAGTCTGGTGCTCGGCGTGCCGATCATATCCCTGCCGAT ACACCTCGAACAGCCGTGGAACGCCTTGGTAGCGGAGCACGTCGGCGTTTGTGTGAGGGCGAAG AGAGACGACGGAGGAAATCTTCAAAGAGAGTTGGTGGCGGAGGCCATTAAAGAAGTGGTGGTTG AGGAAACAGGAGCGGAACTGAGAAGCAAAGCAAGAGTAATTAGTGAAATCTTGAAAAATAAAGA AGCTGAAACAATACAAGATTTGGTGGCTGAGCTTCACCGGCTTTCTGACGCAAGAAGAGCTTGT TGA		
UGT8468 (gene sequence)	ATGGAAAAAAATCTTCACATAGTGATGCTTCCATGGTCGGCGTTCGGCCATCTCATACCATTTT TTCACCTCTCCATAGCCTTAGCCAAAGCCAAAGTTTATATCTCCTTCGTCTCCACTCCAAGAAA TATTCAGAGACTYCCCCAAATCCCGCCGGACTTAGCTTCTTTCATAGATTTGGTGGCCATTCCC TTGCCGAGACTCGACGACGATCTGTTGCTAGAATCTGCAGAGGCCACTTCTGATATTCCGATCG ACAAGATTCAGTATTTGAAGCGAGCCGTCGACCTCCTCCGCCACCCCTTCAAGAAGTTTGTCGC CGAACAATCGCCGGACTGGGTCGTCGTTGATTTTCATGCTTATTGGGCCGGCGAGATCTACCAG GAGTTTCAAGTTCCCGTCGCCTACTTCTGTATTTTCTCGGCCATCTGTTTGCTTTATCTTGGAC CTCCAGACGTGTATTCGAAGGATCCTCAGATCATGGCACGAATATCTCCCGTTACCATGACGGT GCCGCCGGAGTGGGTCGGTTTTCCGTCCGCCGTAGCCTACAACTTGCATGAGGCGACGGTCATG TACTCTGCTCTCTATGAAACAAATGGGTCTGGAATAAGCGACTGCGAGAGGATTCGCCGGCTCG TCCTTTCCTGTCAAGCCGTGGCCATTCGAAGCTGCGAGGAGATTGAAGGCGAATACCTTAGGTT ATGTAAGAAACTGATTCCACCGCAGGGGATTGCCGTCGGCTTGCTTCCGCCGGAAAAGCCACCA AAATCAGATCACGAGCTCATCAAATGGCTTGACGAGCAAAAGCTCCGATTCGTCGTGTACGTGA CATTCGGCAGCGAATGCAACCTGACGAAGGACCAAGTTCACGAGATAGCCCACGGGCTGGAACT GTCGGAGCTGCCATTTTTATGGGCACTGAGGAAACCCAGCTGGGCAGCTGAGGAAGACGATGGG CTGCCGTCTGGGTTTCGTGAGAGAACGTCCGGGAGAGGGGTGGTGAGCATGGAGTGGGTGCCGC AGTTGGAGATTCTGGCGCACCAGGCCATCGGCGTCTCTTTAGTTCACGGGGGCTGGGGCTCTAT TATCGAGTCGCTACAAGCTGGGCACTGTCTGGTTGTGCTGCCGTTTATCATCGACCAGCCGCTG AACTCAAAGCTTTTGGTGGAGAAAGGGATGGCGCTTGAGATCAGAAGGAACGGTTCTGATGGAT GGTTTAGTAGAGAAGACATCGCCGGAACTTTGAGAGAAGCTATGCGGTCGTCTGAGGAAGGCGG GCAGCTGAGGAGCCGTGCAAAAGAGGCGGCGGCCATCGTTGGAGATGAGAAGCTGCAGTGGGAA CAATACTTCGGCGCGTTCGTACAGTTTCTGAGGGACAAGTCTTGA	13	SEQ ID NO: 31 in WO2016050890
UGT10391 (gene sequence)	ATGTCCGAGGAGAAAGGCAGAGGGCACAGCTCGTCGACGGAGAGACACACTGCTGCCGCCATGA ACGCCGAGAAACGAAGCACCAAAATCTTGATGCTCCCATGGCTGGCTCACGGCCACATATCTCC ATACTTCGAGCTCGCCAAGAGGCTCACCAAGAAAAACTGCCACGTTTACTTGTGTTCTTCGCCT GTAAATCTCCAAGGCATCAAGCCGAAACTCTCTGAAAATTACTCTTCCTCCATTGAACTTGTGG AGCTTCATCTTCCATCTCTCCCCGACCTTCCTCCCCATATGCACACGACCAAAGGCATCCCTCT ACATCTACAATCCACCCTCATCAAAGCCTTCGACATGGCCGCCCCTGATTTTTCCGACCTGTTG	14	SEQ ID NO: 32 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CAGAAACTCGAGCCGGATCTCGTCATTTCCGATCTCTTCCAGCCATGGGCAGTTCAATTAGCGT CGTCTCGGAACATTCCCGTCGTCAATTTCGTTGTCACCGGAGTCGCTGTTCTTAGTCGTTTGGC TCACGTGTTTTGCAACTCCGTTAAGGAATTCCCTTTCCCGGAACTCGATCTAACCGACCATTGG ATCTCCAAGAGCCGCCGCAAAACGTCCGACGAATTAGGTCGCGAGTGCGCGATGCGATTTTTCA ACTGCATGAAACAATCTTCAAACATCACTCTAGCCAACACTTTCCCCGAGTTCGAAGAAAAATA CATCGATTATCTCTCTTCCTCGTTTAAGAAAAAGATTCTTCCGGTTGCTCCTCTAGTTCCTGAA ATCGACGCAGACGACGAGAAATCGGAAATTATCGAGTGGCTTGACAAGAAGAAACCGAAATCGA CTGTTTACGTTTCGTTTGGGAGTGAGTATTATCTGACGAAAGAAGACAGGGAAGAGCTCGCCCA TGGCTTAGAAAAGAGCGGCGTGAATTTCATCTGGGTTATTAGGTTTCCAAAGGGCGAGAAGATC ACCATTGAAGAGGCTTTACCAGAAGGATTTCTCGAGAGAGTAGGGGACAGGGGAGTGATTATCG ACGGGTGGGCGCCGCAGTTGAAAATATTGAGGCATTCAAGCGTGGGCGGGTTCGTGTGCCACTG CGGGTGGAACTCTGTGGTGGAGAGCGTGGTGTTTGGGGTGCCGATCATAGCCTTGCCGATGCAG CTCGATCAGCCATGGCATGCGAAGGTGGCGGAGGACGGCGGCGTCTGTGCGGAGGCGAAGAGAG ACGTTGAAGGGAGCGTTCAGAGAGAAGAGGTGGCGAAGGCCATTAAAGAGGTGGTGTTTGAGAA GAAGGGGGGGGTTCTGAGTGGAAAAGCAAGAGAGATCAGCGAGGCCTTGAGAAAGAGGGAAGGG GAAATCATAGAGGAATTGGTTGCTGAGTTTCACCAGCTCTGTGAAGCTTGA		
UGT1576 protein	MASPRHTPHFLLFPFMAQGHMIPMIDLARLLAQRGVIITIIITTPHNAARYHSVLARAIDSGLHI HVLQLQFPCKEGGLPEGCENVDLLPSLASIPRFYRAASDLLYEPSEKLFEELIPRPTCIISDMC LPWTMRIALKYHVPRLVFYSLSCFFLLCMRSLKNNLALISSKSDSEFVTFSDLPDPVEFLKSEL PKSTDEDLVKFSYEMGEADRQSYGVILNLFEEMEPKYLAEYEKERESPERVWCVGPVSLCNDNK LDKAERGNKASIDEYKCIRWLDGQQPSSVVYVSLGSLCNLVTAQIIELGLGLEASKKPFIWVIR RGNITEELQKWLVEYDFEEKIKGRGLVILGWAPQVLILSHPAIGCFLTHCGWNSSIEGISAGVP MVTWPLFADQVFNEKLIVQILRIGVSVGTETTMNWGEEEEKGVVVKREKVREAIEIVMDGDERE ERRERCKELAETAKRAIEEGGSSHRNLTMLIEDIIHGGGLSYEKGSCR	15	SEQ ID NO: 48 in WO2016050890
UGT SK98 protein	MDAQRGHTTTILMLPWVGYGHLLPFLELAKSLSRRKLFHIYFCSTSVSLDAIKPKLPPSISSDD SIQLVELRLPSSPELPPHLHTTNGLPSHLMPALHQAFVMAAQHFQVILQTLAPHLLIYDILQPW APQVASSLNIPAINFSTTGASMLSRTLHPTHYPSSKFPISEFVLHNHWRAMYTTADGALTEEGH KIEETLANCLHTSCGVVLVNSFRELETKYIDYLSVLLNKKVVPVGPLVYEPNQEGEDEGYSSIK NWLDKKEPSSTVFVSFGTEYFPSKEEMEEIAYGLELSEVNFIWVLRFPQGDSTSTIEDALPKGF LERAGERAMVVKGWAPQAKILKHWSTGGLVSHCGWNSMMEGMMFGVPIIAVPMHLDQPFNAGLL EEAGVGVEAKRGSDGKIQREEVAKSIKEVVIEKTREDVRKKAREMGEILRSKGDEKIDELVAEI SLLRKKAPCSI	16	SEQ ID NO: 50 in WO2016050890
UGT430 protein	MEQAHDLLHVLLFPYPAKGHIKPFLCLAELLCNAGLNVTFLNTDYNHRRLHNLHLLAACFPSLH FESISDGLQPDQPRDILDPKFYISICQVTKPLFRELLLSYKRTSSVQTGRPPITCVITDVIFRF PIDVAEELDIPVFSFCTFSARFMFLYFWIPKLIEDGQLPYPNGNINQKLYGVAPEAEGLLRCKD LPGHWAFADELKDDQLNFVDQTTASLRSSGLILNTFDDLEAPFLGRLSTIFKKIYAVGPIHALL NSHHCGLWKEDHSCLAWLDSRAARSVVFVSFGSLVKITSRQLMEFWHGLLNSGTSFLFVLRSDV VEGDGEKQVVKEIYETKAEGKWLVVGWAPQEKVLAHEAVGGFLTHSGWNSILESIAAGVPMISC	17	SEQ ID NO: 62 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	PKIGDQSSNCTWISKVWKIGLEMEDQYDRATVEAMVRSIMKHEGEKIQKTIAELAKRAKYKVSK DGTSYRNLEILIEDIKKIKPN		
UGT1697 protein	MVQPRVLLFPFPALGHVKPFLSLAELLSDAGIDVVFLSTEYNHRRISNTEALASRFPTLHFETI PDGLPPNESRALADGPLYFSMREGTKPRFRQLIQSLNDGRWPITCIITDIMLSSPIEVAEEFGI PVIAFCPCSARYLSIHFFIPKLVEEGQIPYADDDPIGEIQGVPLFEGLLRRNHLPGSWSDKSAD ISFSHGLINQTLAAGRASALILNTFDELEAPFLTHLSSIFNKIYTIGPLHALSKSRLGDSSSSA SALSGFWKEDRACMSWLDCQPPRSVVFVSFGSTMKMKADELREFWYGLVSSGKPFLCVLRSDVV SGGEAAELIEQMAEEEGAGGKLGMVVEWAAQEKVLSHPAVGGFLTHCGWNSTVESIAAGVPMMC WPILGDQPSNATWIDRVWKIGVERNNREWDRLTVEKMVRALMEGQKRVEIQRSMEKLSKLANEK VVRGGLSFDNLEVLVEDIKKLKPYKF	18	SEQ ID NO: 68 in WO2016050890
UGT11789 protein	MDAKEESLKVFMLPWLAHGHISPYLELAKRLAKRKFLVYFCSTPVNLEAIKPKLSKSYSDSIQL MEVPLESTPELPPHYHTAKGLPPHLMPKLMNAFKMVAPNLESILKTLNPDLLIVDILLPWMLPL ASSLKIPMVFFTIFGAMAISFMIYNRTVSNELPFPEFELHECWKSKCPYLFKDQAESQSFLEYL DQSSGVILIKTSREIEAKYVDFLTSSFTKKVVTTGPLVQQPSSGEDEKQYSDIIEWLDKKEPLS TVLVSFGSEYYLSKEEMEEIAYGLESASEVNFIWIVRFPMGQETEVEAALPEGFIQRAGERGKV VEGWAPQAKILAHPSTGGHVSHNGWSSIVECLMSGVPVIGAPMQLDGPIVARLVEEIGVGLEIK RDEEGRITRGEVADAIKTVAVGKTGEDFRRKAKKISSILKMKDEEEVDTLAMELVRLCQMKRGQ ESQD	19	SEQ ID NO: 72 in WO2016050890
CYP1798 protein	MEMSSSVAATISIWMVVVCIVGVGWRVVNWVWLRPKKLEKRLREQGLAGNSYRLLFGDLKERAA MEEQANSKPINFSHDIGPRVFPSMYKTIQNYGKNSYMWLGPYPRVHIMDPQQLKTVFTLVYDIQ KPNLNPLIKFLLDGIVTHEGEKWAKHRKIINPAFHLEKLKDMIPAFFHSCNEIVNEWERLISKE GSCELDVMPYLQNLAADAISRTAFGSSYEEGKMIFQLLKELTDLVVKVAFGVYIPGWRFLPTKS NNKMKEINRKIKSLLLGIINKRQKAMEEGEAGQSDLLGILMESNSNEIQGEGNNKEDGMSIEDV IEECKVFYIGGQETTARLLIWTMILLSSHTEWQERARTEVLKVFGNKKPDFDGLSRLKVVTMIL NEVLRLYPPASMLTRIIQKETRVGKLTLPAGVILIMPIILIHRDHDLWGEDANEFKPERFSKGV SKAAKVQPAFFPFGWGPRICMGQNFAMIEAKMALSLILQRFSFELSSSYVHAPTVVFTTQPQHG AHIVLRKL	20	SEQ ID NO: 74 in WO2016050890
EPH1 epoxide hydrolase	MEKIEHSTIATNGINMHVASAGSGPAVLFLHGFPELWYSWRHQLLYLSSLGYRAIAPDLRGFGD TDAPPSPSSYTAHHIVGDLVGLLDQLGVDQVFLVGDWGAMMAWYFCLFRPDRVKALVNLSVHFT PRNPAISPLDGFRLMLGDDFYVCKFQEPGVAEADFGSVDTATMFKKFLTMRDPRPPIIPNGFRS LATPEALPSWLTEEDIDYFAAKFAKTGFTGGFNYYRAIDLTWELTAPWSGSEIKVPTKFIVGDL DLVYHFPGVKEYIHGGGFKKDVPFLEEVVVMEGAAHFINQEKADEINSLIYDFIKQF	21	Disclosed in the suppl. of Itkin et al.(Proc Natl Acad Sci USA 2016, 22; 113(47):E7619 -E7628)
EPH2 epoxide hydrolase	MEKIEHTTISTNGINMHVASIGSGPAVLFLHGFPELWYSWRHQLLFLSSMGYRAIAPDLRGFGD TDAPPSPSSYTAHHIVGDLVGLLDQLGIDQVFLVGHDWGAMMAWYFCLFRPDRVKALVNLSVHF	22	Disclosed in the

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	LRRHPSIKFVDGFRALLGDDFYFCQFQEPGVAEADFGSVDVATMLKKFLTMRDPRPPMIPKEKG FRALETPDPLPAWLTEEDIDYFAGKFRKTGFTGGFNYYRAFNLTWELTAPWSGSEIKVAAKFIV GDLDLVYHFPGAKEYIHGGGFKKDVPLLEEVVVVDGAAHFINQERPAEISSLIYDFIKKF		supplement of Itkin et al
EPH3 epoxide hydrolase	MDQIEHITINTNGIKMHIASVGTGPVVLLLHGFPELWYSWRHQLLYLSSVGYRAIAPDLRGYGD TDSPASPTSYTALHIVGDLVGALDELGIEKVFLVGHDWGAIIAWYFCLFRPDRIKALVNLSVQF IPRNPAIPFIEGFRTAFGDDFYMCRFQVPGEAEEDFASIDTAQLFKTSLCNRSSAPPCLPKEIG FRAIPPPENLPSWLTEEDINYYAAKFKQTGFTGALNYYRAFDLTWELTAPWTGAQIQVPVKFIV GDSDLTYHFPGAKEYIHNGGFKKDVPLLEEVVVVKDACHFINQERPQEINAHIHDFINKF	23	Disclosed in the supplement of Itkin et al
EPH4 epoxide hydrolase	MENIEHTTVQTNGIKMHVAAIGTGPPVLLLHGFPELWYSWRHQLLYLSSAGYRAIAPDLRGYGD TDAPPSPSSYTALHIVGDLVGLLDVLGIEKVFLIGHDWGAIIAWYFCLFRPDRIKALVNLSVQF FPRNPTTPFVKGFRAVLGDQFYMVRFQEPGKAEEEFASVDIREFFKNVLSNRDPQAPYLPNEVK FEGVPPPALAPWLTPEDIDVYADKFAETGFTGGLNYYRAFDRTWELTAPWTGARIGVPVKFIVG DLDLTYHFPGAQKYIHGEGFKKAVPGLEEVVVMEDTSHFINQERPHEINSHIHDFFSKFC	24	Disclosed in the supplement of Itkin et al
EPH5 epoxide hydrolase	MEKESEIHSIRHTTVSVNGINMHVAEKGEGPLVLFIHGFPELWYSWRHQILDLASLGYRAVAPD LRGYGDSDAPPSASSYTSFHIVGDLIALLDAIVGVEEKVFVVAHDWGAIIAWYLCLYRPDRIKA LVNLSVAFIRRNPKGKPVEWIRALYGDDHYMCRCQEPGEIEGEFAEIGTERVLTQFLTYHSPKP LMLPKGKAFGHPLDTPIPLPPWLSHQDIEYYASKFDKKGFTGPVNYYRNLDRNWELNAPFTRAQ VKVPVKFIVGDLDLTYHSFGTKEYIHSGEMKKDVPFLQEVVVMEGVGHFIQSEKPHEISDHIYQ FIKKF	25	Disclosed in the supplement of Itkin et al
EPH6 epoxide hydrolase	MEKIEHTIITTNGINMHVASIGTGPAVLFLHGFPELWYSWRHQLLSFSSLGYRAIAPDLRGYGD SDAPPSPSSYTVFHIVGDLVGLLDQLGIDQVFLVGHDWGASIAWYFSLLRPDRIKALVNLSVQY FPRNPARNTVEALRALFGDDYYVCRFQEPGEMEEDFASIDTAVIFKIFLSSRDPRPPCIPKAVG FRAFPVPDSLPSWLSEEDISYYASKFSKKGFTGGLNYYRALALNWELTAPWTGTQIKVPTKFIV GDLDLTYHIPGSKEYIHKGGFERDVPSLEEVVVIEGAAHFVNQERPEEISKHIYDFIKKF	26	Disclosed in the supplement of Itkin et al
EPH7 epoxide hydrolase	MDAIEHRTVSVNGINMHVAEKGEGPVVLLLHGFPELWYSWRHQILALSSLGYRAVAPDLRGYGD TDAPGSISSYTCFHIVGLVALVESLGVDRVFVVAHDWGAMIAWCLCLFRPEMVKAFVCLSVPFR QRNPKMKPVQSMRAFFGDDYYICRFQNPGEIEEEMAQVGAREVLRGILTSRRPGPPILPKGQAF RARPGASTALPSWLSEKDLSFFASKYDQKGFTGPLNYYRAMDLNWELTASWTGVQVKVPVKYIV GDVDMVFTTPGVKEYVNGGGFKKDVPFLQEVVIMEGVGHFINQEKPEEISSHIHDFISRF	27	Disclosed in the supplement of Itkin et al
EPH8 epoxide hydrolase	MDQIQHKFIDIRGLKLHIAEIGTGSPAVVFLHGFPEIWYSWRHQMVAAAAVGYRAISPDLRGYG FSDPHPQPQNASFDDFVEDTLAILDFLHIPKAFLVGKDFGSWPVYLFSLVHPTRVAGIVSLGVP FLPPNPKRYRDLPEGFYIFRWKESGRAEADFGRFDVKTVLRRIYTLFSRSEIPIAEKDQEIMDM VDESTPPPPWLTDEDLAAYATAYEHSGFESALQVPYRRRHQELGMSNPRVDVPVLLIIGGKDYF LKFPGIEDYIKSEKMREIVPDLEVADLADGTHFMQEQFPAQVNHLLISFLGKRNT	28	Disclosed in the supplement of Itkin et al
EH1 epoxide hydrolase 1	MDAIEHRTVSVNGINMHVAEKGEGPVVLLLHGFPELWYSWRHQILALSSLGYRAVAPDLRGYGD TDAPGSISSYTCFHIVGDLVALVESLGMDRVFVVAHDWGAMIAWCLCLFRPEMVKAFVCLSVPF RQRNPKMKPVQSMRAFFGDDYYICRFQNPGEIEEEMAQVGAREVLRGILTSRRPGPPILPKGQA	29	SEQ ID NO: 38 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	FRARPGASTALPSWLSEKDLSFFASKYDQKGFTGPLNYYRAMDLNWELTASWTGVQVKVPVKYI VGDVDMVFTTPGVKEYVNGGGFKKDVPFLQEVVIMEGVGHFINQEKPEEISSHIHDFISKF		
EH2 epoxide hydrolase	MDEIEHITINTNGIKMHIASVGTGPVVLLLHGFPELWYSWRHQLLYLSSVGYRAIAPDLRGYGD TDSPASPTSYTALHIVGDLVGALDELGIEKVFLVGHDWGAIIAWYFCLFRPDRIKALVNLSVQF IPRNPAIPFIEGFRTAFGDDFYICRFQVPGEAEEDFASIDTAQLFKTSLCNRSSAPPCLPKEIG FRAIPPPENLPSWLTEEDINFYAAKFKQTGFTGALNYYRAFDLTWELTAPWTGAQIQVPVKFIV GDSDLTYHFPGAKEYIHNGGFKRDVPLLEEVVVVKDACHFINQERPQEINAHIHDFINKF	30	SEQ ID NO: 40 in WO2016050890
CYP533 gene (coding sequence)	ATGGAACTCTTCTCTACCAAAACTGCAGCCGAGATCATCGCTGTTGTCTTGTTTTTCTACGCTC TCATCCGGCTATTATCTGGAAGATTCAGCTCTCAACAGAAGAGACTGCCACCTGAAGCCGGTGG CGCCTGGCCACTGATCGGCCATCTCCATCTCCTAGGTGGGTCGGAACCTGCACATAAAACCTTG GCGAACATGGCGGACGCCTACGGACCAGTTTTTACGTTGAAACTGGGCATGCATACAGCTTTGG TTATGAGCAGTTGGGAAATAGCGAGAGAGTGCTTTACTAAAAACGACAGAATCTTTGCCTCCCG CCCCATAGTCACTGCCTCAAAGCTTCTCACCTATAACCATACCATGTTTGGGTTCAGCCAATAT GGTCCATTCTGGCGCCATATGCGCAAAATAGCCACGCTTCAACTCCTCTCAAACCACCGCCTCG AGCAGCTCCAACACATCAGAATATCGGAGGTCCAGACTTCGATTAAGAAACTGTACGAGTTGTG GGTCAACAGCAGAAATAATGGAGGCGAGAAAGTGTTGGTGGAGATGAAGACGTGGTTCGGAGGC ATAACCTTGAACACCATATTCAGGATGGTGGTCGGAAAGCGATTCTCGACTGCTTTCGAAGGCA GTGGTGGCGAACGGTATCGGAAGGCGTTGAGGGATTCTCTTGAATGGTTTGGGGCATTCGTTCC GTCAGATTCATTCCCGTTTTTAAGATGGTTGGATTTGGGAGGATATGAGAAGGCGATGAAGAAG ACGGCGAGTGTGCTGGACGAGGTGCTTGATAAATGGCTCAAAGAGCATCAGCAGAGGAGAAACT CCGGTGAACTGGAGACGGAGGAGCACGACTTCATGCACGTGATGCTGTCTATTGTTAAGGATGA TGAAGAACTATCCGGCTACGATGCCGATACAGTCACAAAAGCTACATGTTTGAATTTAATAGTT GGTGGATTCGACACTACACAAGTAACTATGACATGGGCTCTTTCTTTGCTTCTCAACAATGAAG AGGTATTAAAAAAGGCCCAACTTGAACTAGACGAACAAGTTGGAAGAGAGAGGTTTGTGGAAGA GTCCGATGTTAAAAATCTGTTATATCTCCAGGCCATCGTGAAGGAAACTTTGCGTTTGTACCCT TCAGCGCCAATCTCGACATTTCATGAGGCCATGGAAGATTGCACTGTTTCTGGCTACCACATCT TTTCAGGGACGCGTTTGATGGTGAATCTTCAAAAGCTTCAAAGAGATCCACTTGCATGGGAGGA TCCATGTGACTTTCGACCGGAGAGATTTCTGACAACTCATAAGGATTTCGATCTTAGAGGACAT AGTCCTCAATTGATACCATTTGGGAGTGGTCGAAGAATATGCCCTGGCATCTCGTTTGCCATTC AAGTTTTGCATCTTACGCTTGCAAATCTACTTCATGGGTTTGACATTGGAAGGCCATCTCATGA ACCAATCGATATGCAGGAGAGTAAAGGACTAACGAGTATTAAAACAACTCCACTTGAGGTTGTT TTAGCTCCACGCCTTGCTGCTCAAGTTTATGAGTGA	31	SEQ ID NO: 3 in WO2016050890
CYP937 gene(coding seuqence)	ATGCCGATCGCAGAAGGTGCAGTCTCTGATTTGTTTGGTCGCCCACTCTTCTTTGCACTATATG ATTGGTTCTTAGAGCATGGATCTGTTTATAAACTTGCCTTTGGACCAAAAGCCTTTGTTGTTGT ATCAGATCCCATTGTGGCAAGATATATTCTTCGAGAAAATGCATTTGGTTATGACAAGGGAGTG CTTGCTGATATTTTAGAACCGATAATGGGTAAAGGACTAATACCAGCTGACCTTGGCACTTGGA AGCAGAGGAGACGAGTTATTGCTCCAGGATTCCATGCCTTGTACTTGGAAGCTATGACCAAAGT ATTTGCCAATTGTTCAGAACGATCAATATTGAAATTGGAGAAGCTTCTAGGAGAAGGTGAACTA	32	SEQ ID NO: 4 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CAGGAGAATAAAACCATTGAGTTGGATATGGAAGCAGAGTTTTCAAGTTTGGCTCTTGATATCA TTGGACTCGGTGTTTTCAACTATGATTTTGGTTCTGTAACCAAAGAATCTCCGGTGATTAAGGC TGTATATGGGACTCTTTTTGAAGCAGAGCATAGATCGACTTTCTATATCCCATATTGGAAAGTA CCTTTGGCAAGGTGGATAGTCCCAAGGCAGCGTAAATTCCATGGTGACCTTAAGGTTATTAATG AGTGTCTTGATGGCCTAATACGCAACGCAAGAGAAACCCGAGACGAAACGGATGTTGAGAAATT GCAGCAAAGGGACTACTTAAATCTCAAGGATGCCAGTCTTTTGCGTTTCTTAGTTGATATGCGG GGAGCTGATGTTGATGATCGCCAGCTTAGGGACGATCTGATGACGATGCTTATTGCTGGCCATG AAACAACTGCTGCTGTGCTTACATGGGCTGTTTTTTTGCTTGCACAAAATCCTTCAAAAATGAA AAAAGCGCAAGCAGAGATTGATTTGGTTCTTGGCATGGGGAGGCCAACTTTTGAATCATTTAAA GCATTGAAGTACATCAGACTTATCGTTGCAGAGACTCTTCGTTTGTTTCCTCAGCCTCCATTGC TGATAAGACGAGCTCTCAAATCAGATATATTACCAGGAGGATACAATGGTGACAAAACTGGATA TGCAATTCCTGCAGGGACTGACATCTTCATCTCTGTTTACAATCTCCACAGATCTCCCTACTTC TGGGATAATCCTCAAGAATTTGAACCAGAGAGATTTCAAGTAAAGAGGGCAAGCGAGGGAATTG AAGGATGGGATGGTTTCGACCCATCTAGAAGCCCTGGAGCTCTATACCCGAATGAGATTGTAGC AGACTTTTCCTTCTTACCATTTGGTGGAGGCCCTAGAAAATGTGTGGGAGATCAATTTGCTCTA ATGGAGTCAACTATAGCATTGGCCATGTTACTGCAGAAGTTTGATGTGGAGCTAAAAGGAAGTC CAGAATCTGTAGAACTAGTTACTGGAGCCACAATACATACCAAAAGTGGGTTGTGGTGCAAACT GAGAAGAAGATCACAAGTAAACTGA		
CYP1798 gene(coding sequence, codon optimized)	ATGGAAATGTCCTCAAGTGTCGCAGCCACAATCAGTATCTGGATGGTCGTCGTATGTATCGTAG GTGTAGGTTGGAGAGTCGTAAATTGGGTTTGGTTGAGACCAAAGAAATTGGAAAAGAGATTGAG AGAACAAGGTTTGGCCGGTAATTCTTACAGATTGTTGTTCGGTGACTTGAAGGAAAGAGCTGCA ATGGAAGAACAAGCAAATTCAAAGCCTATAAACTTCTCCCATGACATCGGTCCAAGAGTTTTCC CTTCAATGTACAAGACCATCCAAAACTACGGTAAAAACTCCTACATGTGGTTAGGTCCATACCC TAGAGTCCACATCATGGATCCACAACAATTGAAGACCGTTTTTACTTTGGTCTACGACATTCAA AAGCCAAATTTGAACCCTTTGATTAAATTCTTGTTAGATGGTATCGTTACACATGAAGGTGAAA AGTGGGCTAAGCACAGAAAGATTATTAACCCAGCATTCCATTTGGAAAAGTTGAAGGATATGAT ACCTGCTTTCTTTCACTCATGTAATGAAATCGTCAACGAATGGGAAAGATTGATTTCAAAAGAA GGTTCCTGCGAATTGGATGTAATGCCTTATTTGCAAAATTTGGCCGCTGACGCCATTTCAAGAA CCGCTTTTGGTTCTTCATACGAAGAAGGTAAAATGATCTTCCAATTGTTGAAGGAATTGACTGA TTTGGTTGTCAAGGTAGCTTTTGGTGTTTATATTCCAGGTTGGAGATTCTTGCCTACAAAGAGT AACAACAAAATGAAGGAAATTAATAGAAAAATCAAGTCTTTGTTGTTGGGTATCATTAACAAGA GACAAAAGGCAATGGAAGAAGGTGAAGCCGGTCAATCTGATTTGTTGGGTATATTAATGGAAAG TAATTCTAACGAAATCCAAGGTGAAGGTAATAACAAGGAAGATGGCATGTCTATTGAAGACGTC ATCGAAGAGTGTAAGGTATTTTATATAGGTGGTCAAGAAACTACAGCAAGATTATTGATCTGGA CTATGATATTGTTGTCCAGTCATACAGAATGGCAAGAAAGAGCCAGAACCGAAGTCTTGAAGGT ATTTGGTAATAAGAAACCAGATTTCGACGGTTTGTCAAGATTGAAGGTAGTTACTATGATCTTG AACGAAGTTTTAAGATTGTACCCACCTGCTTCCATGTTGACAAGAATCATCCAAAAGGAAACAA GAGTTGGTAAATTAACCTTGCCAGCAGGTGTTATCTTGATAATGCCTATCATCTTGATACATAG	33	SEQ ID NO: 5 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AGATCACGACTTGTGGGGTGAAGATGCTAACGAGTTTAAACCAGAAAGATTCAGTAAAGGTGTT TCTAAGGCAGCCAAAGTCCAACCAGCCTTTTTCCCTTTTGGTTGGGGTCCTAGAATTTGCATGG GTCAAAACTTCGCTATGATCGAAGCTAAGATGGCATTGAGTTTGATCTTGCAAAGATTTTCTTT CGAATTGTCTTCATCCTACGTTCATGCACCAACTGTCGTCTTCACTACACAACCACAACACGGT GCCCACATCGTTTTGAGAAAGTTATGA		
CYP1994 gene (coding sequence)	ATGGAACCACAACCAAGTGCGGAATTCAACTGGAATCACAGCCTAAGCACCGTCGCTATCGGTG TCATTGCCATTATTTTCTTCCGTTTTCTCGTCAAAAGAGTCACCGGCGCCGGTGAGCGAAAGGG TCCGAAGCCGCCAAAAGTAGCCGGAGGGTGGCCTCTAATTGGCCACCTCCCTCTCCTCGGAGGA CCTGAACTGCCCCATGTCAAACTGGGTGGTTTGGCTGATAAATATGGTCCAATCTTCTCGATCC GGCTGGGTGTCCACTCCGCCGTCGTGATAAACAGTTGGGAGGCGGCGAAACAGTTATTAACCAA CCATGACGTCGCCGTCTCTTCCCGCCCCCAAATGCTCGGCGGAAAACTCCTGGGCTACAACTAC GCCGTGTTTGGTTTCGGACCCTACGGCTCTTACTGGCGCAACATGCGCAAGATAACCACGCAAG AGCTTCTATCCAATAGCAGAATCCAGCTCCTAAGAGACGTTCGAGCGTCAGAAGTGAACCAAGG CATAAAAGAGCTCTACCAGCACTGGAAAGAAAGAAGAGACGGTCACGACCAAGCCTTGGTGGAA CTGCAGCAGTGGGTCGGGGACTTGACTATGAATCTGATTCTCGGAGTCATCGCCGGGAAAAGGT TCTTTGGAGCTGCAGCAACGGTAGACGAGGAAGAGGCGCGACGGAGCCATAAAGCATTGAAGGA GTTGTTACATTATATGGGGCTTTTTCTACTGGGTGATGCTGTTCCATATCTAGGATGGTTGGAC GTCGGCGGCCATGTGAAGGCGATGAAGAAAACTTCAAAAGAATTGGACCGTATGTTAACACAGT GGTTGGAGGAGCACAAGAAGGAAGGACCCAAGAAAGATCATAAAGACTTCATGGACGTGATGCT TTCAGTTCTCAATGAAACATCCGATGTTCTTTCAGATAAGACCCATGGCTTCGATGCTGATACC ATCATCAAAGCTACATGTATGACGATGGTTTTAGGAGGGAGTGATACGACGGCGGTGGTTGTGA TATGGGCAATCTCGCTGCTGCTGAATAATCGCCCTGCGTTGAGAAAAGTGCAAGAAGAACTGGA AGCCCATATCGGCCGAGACAGAGAACTGGAGGAATCGGATCTCGGTAAGCTAGTGTATTTGCAG GCAGTCGTGAAGGAGACATTGCGGCTGTACGGAGCCGGAGGCCTTTTCTTTCGTGAAACCACAG AGGATGTCACCATCGACGGATTCCATGTCGAGAAAGGGACATGGCTGTTCGTGAACGTGGGGAA GATCCACAGAGATGGGAAGGTGTGGCCGGAGCCAACGGAGTTCAAACCGGAGAGGTTTCTGACG ACCCACAAAGATTTTGATCTGAAGGGCCAGCGGTTTGAGCTCATCCCTTTCGGGGGAGGAAGAA GATCGTGCCCTGGAATGTCTTTTGGGCTCCAAATGCTACAGCTTATTTTGGGTAAACTGCTTCA GGCTTTTGATATATCGACGCCGGGGGACGCCGCCGTTGATATGACCGGATCCATTGGACTGACG AACATGAAAGCCACTCCATTGGAAGTGCTCATCACCCCGCGCTTGCCTCTTTCGCTTTACGATT GA	34	SEQ ID NO: 6 in WO2016050890
CYP2048 gene (coding sequence)	ATGGAGACTCTTCTTCTTCATCTTCAATCGTTATTTCATCCAATTTCCTTCACTGGTTTCGTTG TCCTCTTTAGCTTCCTGTTCCTGCTCCAGAAATGGTTACTGACACGTCCAAACTCTTCATCAGA AGCCTCACCCCCTTCTCCACCAAAGCTTCCCATCTTCGGACACCTTCTAAACCTGGGTCTGCAT CCCCACATCACCCTCGGAGCCTACGCTCGCCGCTATGGCCCTCTCTTCCTCCTCCACTTCGGCA GCAAGCCCACCATCGTCGTCTCTTCTGCCGAAATCGCTCGCGATATCATGAAGACCCACGACCT CGTCTTCGCCAACCGTCCTAAATCAAGCATCAGCGAAAAGATTCTTTACGGCTCCAAAGATTTA GCCGCATCTCCTTACGGCGAATACTGGAGGCAGATGAAAAGCGTTGGCGTGCTTCATCTTTTGA	35	SEQ ID NO: 7 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GCAACAAAAGGGTTCAATCCTTTCGCTCTGTCAGAGAAGAAGAAGTCGAACTGATGATCCAGAA GATCCAACAGAACCCCCTATCAGTTAATTTAAGCGAAATATTCTCTGGACTGACGAACGACATA GTTTGCAGGGTGGCTTTAGGGAGAAAGTATGGCGTGGGAGAAGACGGAAAGAAGTTCCGGTCTC TTCTGCTGGAGTTTGGGGAAGTATTGGGAAGTTTCAGTACGAGAGACTTCATCCCGTGGCTGGG TTGGATTGATCGTATCAGTGGGCTGGACGCCAAAGCCGAGAGGGTAGCCAAAGAGCTCGATGCT TTCTTTGACAGAGTGATCGAAGATCACATCCATCTAAACAAGAGAGAGAATAATCCCGATGAGC AGAAGGACTTGGTGGATGTGCTGCTTTGTGTACAGAGAGAAGACTCCATCGGGTTTCCCCTTGA GATGGATAGCATAAAAGCTTTAATCTTGGACATGTTTGCTGCAGGCACAGACACGACATACACG GTGTTGGAGTGGGCAATGTCCCAACTGTTGAGACACCCAGAAGCGATGAAGAAACTGCAGAGGG AGGTCAGAGAAATAGCAGGTGAGAAAGAACACGTAAGTGAGGATGATTTAGAAAAGATGCATTA CTTGAAGGCAGTAATCAAAGAAACGCTGCGGCTACACCCACCAATCCCACTCCTCGTCCCCAGA GAATCAACCCAAGACATCAGGTTGAGGGGGTACGATATCAGAGGCGGCACCCGGGTTATGATCA ATGCATGGGCCATCGGAAGA		
CYP2740 gene (coding sequence)	ATGTCGATGAGTAGTGAAATTGAAAGCCTCTGGGTTTTCGCGCTGGCTTCTAAATGCTCTGCTT TAACTAAAGAAAACATCCTCTGGTCTTTACTCTTCTTTTTCCTAATCTGGGTTTCTGTTTCCAT TCTCCACTGGGCCCATCCGGGCGGCCCGGCTTGGGGCCGCTACTGGTGGCGCCGCCGCCGCAGC AATTCCACCGCCGCTGCTATTCCCGGCCCGAGAGGCCTCCCCCTCGTCGGCAGCATGGGCTTGA TGGCCGACTTGGCCCACCACCGGATTGCCGCCGTGGCTGACTCCTTAAACGCCACCCGCCTCAT GGCCTTTTCGCTCGGCGACACTCGCGTGATCGTCACATGCAACCCCGACGTCGCCAAAGAGATT CTCAACAGCTCCCTCTTCGCCGACCGCCCCGTTAAGGAGTCCGCTTACTCCTTGATGTTCAACC GCGCCATTGGGTTCGCCCCCTATGGCCTTTACTGGCGGACCCTCCGCCGCATCGCTTCCCACCA CCTCTTCTGCCCCAAGCAAATCAAGTCCTCCCAGTCCCAGCGCCGCCAAATCGCTTCCCAAATG GTCGCAATGTTCGCAAACCGCGATGCCACACAGAGCCTCTGCGTTCGCGACTCTCTCAAGCGGG CTTCTCTCAACAACATGATGGGCTCTGTTTTCGGCCGAGTTTACGACCTCTCTGACTCGGCTAA CAATGACGTCCAAGAACTCCAGAGCCTCGTCGACGAAGGCTACGACTTGCTGGGCCTCCTCAAC TGGTCCGACCATCTCCCATGGCTCGCCGACTTCGACTCTCAGAAAATCCGGTTCAGATGCTCCC GACTCGTCCCCAAGGTGAACCACTTCGTCGGCCGGATCATCGCCGAACACCGCGCCAAATCCGA CAACCAAGTCCTAGATTTCGTCGACGTTTTGCTCTCTCTCCAAGAAGCCGACAAACTCTCTGAC TCCGATATGATCGCCGTTCTTTGGGAAATGATTTTTCGTGGGACGGACACGGTGGCAGTTTTAA TCGAGTGGATACTGGCCAGGATGGTACTTCACAACGATATCCAAAGGAAAGTTCAAGAGGAGCT AGATAACGTGGTTGGGAGTACACGCGCCGTCGCGGAATCCGACATTCCGTCGCTGGTGTATCTA ACGGCTGTGGTTAAGGAAGTTCTGAGGTTACATCCGCCGGGCCCACTCCTGTCGTGGGCCCGCC TAGCCATCACTGATACAATCATCGATGGGCATCACGTGCCCCGGGGGACCACCGCTATGGTTAA CATGTGGTCGATAGCGCGGGACCCACAGGTCTGGTCGGACCCACTCGAATTTATGCCCCAGAGG TTTGTGTCCGACCCCGGTGACGTGGAGTTCTCGGTCATGGGTTCGGATCTCCGGCTGGCTCCGT TCGGGTCGGGCAGAAGGACCTGCCCCGGGAAGGCCTTCGCCTGGACAACTGTCACCTTCTGGGT GGCCACGCTTTTACACGACTTCAAATGGTCGCCGTCCGATCAAAACGACGCCGTCGACTTGTCG GAGGTCCTCAAGCTCTCCTGCGAGATGGCCAATCCCCTCACCGTTAAAGTACACCCAAGGCGCA	36	SEQ ID NO: 8 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GTTTAAGCTTTTAA		
CYP3404 gene (coding sequence)	ATGGATGGTTTTCTTCCAACAGTGGCGGCGAGCGTGCCTGTGGGAGTGGGTGCAATATTGTTCA CGGCGTTGTGCGTCGTCGTGGGAGGGGTTTTGGTTTATTTCTATGGACCTTACTGGGGAGTGAG AAGGGTGCCTGGTCCACCAGCTATTCCACTGGTCGGACATCTTCCCTTGCTGGCTAAGTACGGC CCAGACGTTTTCTCTGTCCTTGCCACCCAATATGGCCCTATCTTCAGGTTCCATATGGGTAGGC AGCCATTGATAATTATAGCAGACCCTGAGCTTTGTAAAGAAGCTGGTATTAAGAAATTCAAGGA CATCCCAAATAGAAGTGTCCCTTCTCCAATATCAGCTTCCCCTCTTCATCAGAAGGGTCTTTTC TTCACAAGGGATGCAAGATGGTCGACAATGCGGAACACGATATTATCGGTCTATCAGTCCTCCC ATCTAGCGAGACTAATACCTACTATGCAATCAATCATTGAAACTGCAACTCAAAATCTCCATTC CTCTGTCCAGGAAGACATCCCTTTCTCCAATCTCTCCCTCAAATTGACCACCGATGTGATTGGA ACAGCAGCCTTCGGTGTCAACTTTGGGCTCTCTAATCCACAGGCAACCAAAACTTGTGCTACCA ACGGCCAAGACAACAAAAATGACGAAGTTTCAGACTTCATCAATCAACACATCTACTCCACAAC GCAGCTCAAGATGGATTTATCAGGTTCCTTCTCAATCATACTTGGACTGCTTGTCCCTATACTC CAAGAACCATTTAGACAAGTCCTAAAGAGAATACCATTCACCATGGACTGGAAAGTGGACCGGA CAAATCAGAAATTAAGTGGTCGGCTTAATGAGATTGTGGAGAAGAGAATGAAGTGTAACGATCA AGGTTCAAAAGACTTCTTATCGCTCATTTTGAGAGCAAGAGAGTCAGAGACAGTATCAAGGAAT GTCTTCACTCCAGACTACATCAGTGCAGTTACGTATGAACACCTACTTGCTGGGTCGGCTACCA CGGCGTTTACGTTGTCTTCTATTGTATATTTAGTTGCTGGGCATCCAGAAGTCGAGAAGAAGTT GCTAGAAGAGATTGACAACTTTGGTCCATCCGATCAGATACCAACAGCTAATGATCTTCATCAG AAGTTTCCATATCTTGATCAGGTGATTAAAGAGGCTATGAGGTTCTACACTGTTTCCCCTCTAG TAGCCAGAGAAACAGCTAAAGATGTGGAGATTGGTGGATATCTTCTTCCAAAGGGGACATGGGT TTGGTTAGCACTTGGAGTTCTTGCCAAGGATCCAAAGAACTTTCCAGAACCAGATAAATTCAAA CCAGAGAGGTTTGATCCAAATGAAGAAGAGGAGAAACAAAGGCATCCTTATGCTTTAATCCCCT TTGGAATTGGTCCTCGAGCATGCATTGGTAAAAAATTCGCCCTTCAGGAGTTGAAGCTCTCGTT GATTCATTTGTACAGGAAGTTTGTATTTCGGCAT	37	SEQ ID NO: 9 in WO2016050890
CYP3968 gene (coding sequence)	ATGGAAATCATTTTATCATATCTCAACAGCTCCATAGCTGGACTCTTCCTCTTGCTTCTCTTCT CGTTTTTTGTTTTGAAAAAGGCTAGAACCTGTAAACGCAGACAGCCTCCTGAAGCAGCCGGCGG ATGGCCGATCATCGGCCACCTGAGACTGCTCGGGGGTTCGCAACTTCCCCATGAAACCTTGGGA GCCATGGCCGACAAGTATGGACCAATCTTCAGCATCCGAGTTGGTGTCCACCCATCTCTTGTTA TAAGCAGTTGGGAAGTGGCTAAAGAGTGCTACACCACCCTCGACTCAGTTGTCTCTTCTCGTCC CAAGAGTTTGGGTGGAAAGTTGTTGGGCTACAACTTCGCCGCTTTTGGGTTCAGGCCTTATGAT TCCTTTTACCGGAGTATCCGCAAAACCATAGCCTCCGAGGTGCTGTCGAACCGCCGTCTGGAGT TGCAGAGACACATTCGAGTTTCTGAGGTGAAGAGATCGGTGAAGGAGCTTTACAATCTGTGGAC GCAGAGAGAGGAAGGCTCAGACCACATACTTATTGATGCGGATGAATGGATTGGTAATATTAAT TTGAACGTGATTCTGATGATGGTTTGTGGGAAGCGGTTTCTTGGCGGTTCTGCCAGCGATGAGA AGGAGATGAGGCGGTGTCTCAAAGTCTCGAGAGATTTCTTCGATTTGACAGGGCAGTTTACGGT GGGAGATGCCATTCCTTTCCTGCGATGGCTGGATTTGGGTGGATATGCGAAGGCGATGAAGAAA ACTGCAAAAGAAATGGACTGTCTCGTTGAGGAATGGCTGGAAGAACACCGCCGGAAGAGAGACT	38	SEQ ID NO: 10 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CCGGCGCCACCGACGGTGAACGTGACTTCATGGATGTGATGCTTTCGATTCTTGAAGAGATGGA CCTTGCTGGCTACGACGCTGACACAGTCAACAAAGCCACATGCCTGAGCATTATTTCTGGGGGA ATCGATACTATAACGCTAACTCTGACATGGGCGATCTCGTTATTGCTGAACAATCGAGAGGCAC TGCGAAGGGTTCAAGAGGAGGTGGACATCCATGTCGGAAACAAAAGGCTTGTGGATGAATCAGA CTTGAGCAAGCTGGTGTATCTCCAAGCCGTCGTGAAAGAGACATTAAGGTTGTACCCAGCAGGG CCGCTGTCGGGAGCTCGAGAGTTCAGTCGGGACTGCACGGTCGGAGGGTATGACGTGGCCGCCG GCACACGGCTCATCACAAACCTTTGGAAGATACAGACGGACCCTCGGGTGTGGCCGGAGCCACT TGAGTTCAGGCCGGAGAGGTTTCTGAGCAGCCACCAGCAGTTGGATGTGAAGGGCCAGAACTTT GAACTGGCCCCATTTGGTTGTGGAAGAAGAGTGTGCCCTGGGGCGGGGCTTGGGGTTCAGATGA CGCAGTTGGTGCTGGCGAGTCTGATTCATTCGGTGGAACTTGGAACTCGCTCCGATGAAGCGGT GGACATGGCTGCTAAGTTTGGACTCACAATGTACAGAGCCACCCCTCTTCAGGCTCTCGTCAAG CCACGCCTCCAAGCCGGTGCTTATTCATGA		
CYP4112 gene (coding sequence)	ATGGGTGTATTGTCCATTTTATTATTCAGATATTCCGTCAAGAAGAAGCCATTAAGATGCGGTC ACGATCAAAGAAGTACCACAGATAGTCCACCTGGTTCAAGAGGTTTGCCATTGATAGGTGAAAC TTTGCAATTCATGGCTGCTATTAATTCTTTGAACGGTGTATACGATTTCGTTAGAATAAGATGT TTGAGATACGGTAGATGCTTTAAGACAAGAATCTTCGGTGAAACCCATGTTTTTGTCTCAACTA CAGAATCCGCTAAGTTGATCTTGAAGGATGGTGGTGAAAAATTCACCAAAAAGTACATCAGATC AATCGCTGAATTGGTTGGTGACAGAAGTTTGTTATGTGCATCTCATTTGCAACACAAGAGATTG AGAGGTTTGTTGACTAATTTGTTTTCTGCCACATTCTTGGCTTCTTTCGTAACTCAATTCGATG AACAAATCGTTGAAGCTTTTAGATCATGGGAATCCGGTAGTACCATAATCGTTTTGAACGAAGC ATTGAAGATCACTTGTAAGGCCATGTGCAAAATGGTCATGTCCTTAGAAAGAGAAAACGAATTG GAAGCTTTGCAAAAGGAATTGGGTCATGTTTGTGAAGCTATGTTGGCATTTCCATGCAGATTCC CTGGTACAAGATTTCACAATGGTTTGAAGGCAAGAAGAAGAATCATTAAAGTTGTCGAAATGGC CATTAGAGAAAGAAGAAGATCTGAAGCTCCTAGAGAAGATTTCTTGCAAAGATTGTTGACAGAA GAAAAGGAAGAAGAAGACGGTGGTGGTGTTTTAAGTGATGCCGAAATTGGTGACAACATATTGA CAATGATGATCGCAGGTCAAGATACCACTGCCTCTGCTATTACCTGGATGGTCAAGTTTTTGGA AGAAAACCAAGATGTATTGCAAAACTTAAGAGACGAACAATTCGAAATCATGGGTAAACAAGAA GGTTGTGGTTCATGCTTCTTGACATTAGAAGATTTGGGTAATATGTCCTATGGTGCAAAAGTAG TTAAGGAATCATTGAGATTAGCCTCCGTCGTACCATGGTTTCCTAGATTGGTTTTACAAGATTC TTTGATCCAAGGTTACAAAATTAAAAAGGGTTGGAACGTCAACATAGACGTAAGATCTTTACAT TCAGATCCATCCTTGTATAATGACCCAACAAAGTTTAACCCTAGTAGATTCGATGACGAAGCTA AACCTTACTCATTTTTGGCATTCGGTATGGGTGGTAGACAATGTTTGGGTATGAACATGGCAAA GGCCATGATGTTGGTTTTCTTGCACAGATTGGTCACCTCATTCAGATGGAAGGTTATAGATTCC GACTCTTCAATCGAAAAATGGGCTTTGTTCTCTAAGTTGAAGTCAGGTTGCCCTATCGTAGTTA CCCACATCGGTTCCTAA	39	SEQ ID NO: 11 in WO2016050890
CYP4149 gene (coding sequence)	ATGGATTTCTACTGGATCTGTGTTCTTCTGCTTTGCTTCGCATGGTTTTCCATTTTATCCCTTC ACTCGAGAACAAACAGCAGCGGCACTTCCAAACTTCCTCCCGGACCGAAACCCTTGCCGATCAT CGGAAGCCTTTTGGCTCTCGGCCACGAGCCCCACAAGTCTTTGGCTAATCTCGCTAAATCTCAT	40	SEQ ID NO: 12 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GGCCCTCTTATGACCTTAAAGCTCGGCCAAATCACCACCGTCGTAGTTTCCTCCGCTGCCATGG CTAAGCAAGTTCTCCAAACGCACGACCAGTTTCTGTCCAGCAGGACCGTTCCAGACGCAATGAC CTCTCACAACCACGATGCTTTCGCACTCCCATGGATTCCGGTTTCACCCCTCTGGCGAAACCTT CGACGAATATGCAACAACCAGTTGTTTGCCGGCAAGATTCTCGACGCCAACGAGAATCTCCGGC GAACCAAAGTGGCCGAGCTCGTATCCGATATCTCGAGAAGTGCATTGAAAGGTGAGATGGTGGA TTTTGGAAACGTGGTGTTCGTCACTTCGCTCAATCTGCTTTCCAATACGATTTTCTCGGTGGAT TTCTTCGACCCAAATTCTGAAATTGGGAAAGAGTTCAGGCACGCAGTACGAGGCCTCATGGAAG AAGCTGCCAAACCAAATTTGGGGGATTATTTCCCTCTGCTGAAGAAGATAGATCTTCAAGGAAT AAAGAGGAGACAGACCACTTACTTCGATCGGGTTTTTAATGTTTTGGAGCACATGATCGACCAG CGTCTTCAGCAGCAGAAGACGACGTCTGGTTCTACCTCCAACAACAACAACGACTTACTGCACT ACCTTCTCAACCTCAGCAACGAAAATAGCGACATGAAATTGGGGAAACTTGAGCTGAAACACTT CTTATTGGTGCTATTCGTCGCTGGGACTGAAACGAGTTCTGCAACACTGCAATGGGCAATGGCA GAACTACTAAGAAACCCAGAAAAGTTAGCAAAAGCTCAAGCGGAGACCAGGCGGGTGATTGGGA AAGGGAACCCAATTGAAGAATCAGACATTTCGAGGCTGCCTTATCTGCAAGCAGTGGTGAAAGA AACTTTCAGATTGCACACACCAGCGCCATTTCTACTGCCGCGCAAAGCACTACAGGACGTGGAA ATTGCAGGTTTCACAGTCCCAAAGGACGCTCAGGTACTGGTAAATTTATGGGCTATGAGCAGAG ATTCAAGCATCTGGGAGAACCCAGAGTGGTTCGAGCCAGAAAGGTTTTTGGAGTCGGAGCTGGA CGTTAGAGGGAGAGATTTTGAGCTGATCCCGTTCGGCGGTGGGCGGAGGATTTGCCCCGGTCTG CCGTTGGCGATGAGAATGTTGCATTTGATTTTGGGTTCTCTCATCCACTTCTTTGATTGGAAGC TTGAAGATGGGTGTCGGCCGGAAGACGTGAAAATGGACGAAAAGCTTGGCCTCACTCTGGAGTT GGCTTTTCCCCTCACAGCCTTGCCTGTCCTTGTCTAA		
CYP4491 gene (coding sequence)	ATGTCCTCCTGCGGTGGTCCAACTCCTTTGAATGTTATCGGTATCTTATTACAATCAGAATCCT CCAGAGCCTGCAACTCAGACGAAAACTCAAGAATTTTGAGAGATTTCGTAACAAGAGAAGTTAA CGCTTTCTTATGGTTGTCCTTGATCACTATCACAGCAGTTTTGATCAGTAAAGTTGTCGGTTTG TTTAGATTGTGGTCTAAGGCAAAGCAATTGAGAGGTCCACCTTGTCCATCATTCTACGGTCATT CTAAGATCATCTCAAGACAAAATTTGACTGATTTGTTATATGACTCCCACAAAAAGTACGGTCC AGTAGTTAAATTGTGGTTAGGTCCTATGCAATTGTTAGTCTCCGTAAAGGAACCAAGTTTGTTG AAGGAAATATTGGTTAAAGCTGAGGATAAGTTGCCTTTAACAGGTAGAGCCTTTAGATTGGCTT TCGGTAGATCTTCATTATTTGCATCCAGTTTCGAAAAGGTTCAAAACAGAAGACAAAGATTGGC CGAAAAGTTGAATAAGATCGCATTCCAAAGAGCCAACATCATTCCAGAAAAGGCCGTAGCTTGT TTCATGGGTAGAGTTCAAGATTTGATGATAGAAGAATCTGTCGACTGTAATAAGGTTTCTCAAC ATTTGGCTTTTACTTTGTTAGGTTGCACATTGTTTGGTGACGCCTTCTTAGGTTGGTCTAAGGC TACAATCTATGAAGAATTGTTGATGATGATCGCTAAGGACGCATCCTTTTGGGCTAGTTATAGA GTTACCCCAATCTGGAAGCAAGGTTTCTGGAGATACCAAAGATTGTGTATGAAGTTGAAGTGCT TGACTCAAGATATCGTTCAACAATACAGAAAGCATTACAAGTTGTTTTCTCACTCACAAAACCA AAACTTACACAACGAAACCAAGTCAACTGGTGTTGAAGTCGCTTTTGATATTCCACCTTGTCCT GCTGCAGACGTTAGAAATTCTTGCTTTTTCTACGGTTTGAACGATCATGTTAACCCAAACGAAG AACCTTGTGGTAATATTATGGGTGTCATGTTTCACGGTTGCTTGACTACAACCTCTTTGATCGC	41	SEQ ID NO: 13 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ATCAATCTTGGAAAGATTGGCCACTAACCCAGAAATCCAAGAAAAGATTAATTCTGAATTGAAC TTAGTTCAAAAGGGTCCAGTCAAGGATCATAGAAAGAATGTTGACAACATGCCTTTGTTATTGG CAACAATCTATGAATCAGCTAGATTATTGCCAGCAGGTCCTTTATTGCAAAGATGTCCTTTGAA GCAAGATTTGGTTTTGAAAACAGGTATCACCATTCCAGCTGGTACCTTGGTCGTAGTTCCTATT AAATTGGTTCAAATGGATGACTCTTCATGGGGTTCAGATGCCAATGAGTTTAATCCATACAGAT TCTTGTCCATGGCTTGTAATGGTATTGACATGATACAAAGAACCCCTTTAGCTGGTGAAAACAT TGGTGACCAAGGTGAAGGTTCATTTGTCTTGAATGACCCAATTGGTAACGTAGGTTTCTTACCT TTTGGTTTCGGTGCAAGAGCCTGCGTTGGTCAAAAGTTTATAATCCAAGGTGTCGCTACTTTGT TCGCAAGTTTGTTGGCCCATTACGAAATTAAATTGCAATCCGAGAGTAAGAATGATTCTAAACC ATCCAGTAACACCTCTGCCAGTCAAATCGTCCCAAACTCAAAAATCGTATTCGTAAGAAGAAAC TCATAA		
CYP5491 gene (coding sequence)	ATGTGGACTGTCGTGCTCGGTTTGGCGACGCTGTTTGTCGCCTACTACATCCATTGGATTAACA AATGGAGAGATTCCAAGTTCAACGGAGTTCTGCCGCCGGGCACCATGGGTTTGCCGCTCATCGG AGAGACGATTCAACTGAGTCGACCCAGTGACTCCCTCGACGTTCACCCTTTCATCCAGAAAAAA GTTGAAAGATACGGGCCGATCTTCAAAACATGTCTGGCCGGAAGGCCGGTGGTGGTGTCGGCGG ACGCAGAGTTCAACAACTACATAATGCTGCAGGAAGGAAGAGCAGTGGAAATGTGGTATTTGGA TACGCTCTCCAAATTTTTCGGCCTCGACACCGAGTGGCTCAAAGCTCTGGGCCTCATCCACAAG TACATCAGAAGCATTACTCTCAATCACTTCGGCGCCGAGGCCCTGCGGGAGAGATTTCTTCCTT TTATTGAAGCATCCTCCATGGAAGCCCTTCACTCCTGGTCTACTCAACCTAGCGTCGAAGTCAA AAATGCCTCCGCTCTCATGGTTTTTAGGACCTCGGTGAATAAGATGTTCGGTGAGGATGCGAAG AAGCTATCGGGAAATATCCCTGGGAAGTTCACGAAGCTTCTAGGAGGATTTCTCAGTTTACCAC TGAATTTTCCCGGCACCACCTACCACAAATGCTTGAAGGATATGAAGGAAATCCAGAAGAAGCT AAGAGAGGTTGTAGACGATAGATTGGCTAATGTGGGCCCTGATGTGGAAGATTTCTTGGGGCAA GCCCTTAAAGATAAGGAATCAGAGAAGTTCATTTCAGAGGAGTTCATCATCCAACTGTTGTTTT CTATCAGTTTTGCTAGCTTTGAGTCCATCTCCACCACTCTTACTTTGATTCTCAAGCTCCTTGA TGAACACCCAGAAGTAGTGAAAGAGTTGGAAGCTGAACACGAGGCGATTCGAAAAGCTAGAGCA GATCCAGATGGACCAATTACTTGGGAAGAATACAAATCCATGACTTTTACATTACAAGTCATCA ATGAAACCCTAAGGTTGGGGAGTGTCACACCTGCCTTGTTGAGGAAAACAGTTAAAGATCTTCA AGTAAAAGGATACATAATCCCGGAAGGATGGACAATAATGCTTGTCACCGCTTCACGTCACAGA GACCCAAAAGTCTATAAGGACCCTCATATCTTCAATCCATGGCGTTGGAAGGACTTGGACTCAA TTACCATCCAAAAGAACTTCATGCCTTTTGGGGGAGGCTTAAGGCATTGTGCTGGTGCTGAGTA CTCTAAAGTCTACTTGTGCACCTTCTTGCACATCCTCTGTACCAAATACCGATGGACCAAACTT GGGGGAGGAAGGATTGCAAGAGCTCATATATTGAGTTTTGAAGATGGGTTACATGTGAAGTTCA CACCCAAGGAATGA	42	SEQ ID NO: 14 in WO2016050890
CYP6479 gene (coding sequence)	ATGAAGATGAAGATGGAATCCATGCGCACCTCCCTGGATATCTCCGACCATGACATACTTCCAA GGGTTTATCCTCATGTTCACCTATGGATCAACAAATATGGGAAAAACTTCATTCAGTGGAATGG CAACGTAGCTCAGTTGATTGTTTCGGATCCTGACACGATCAAGGAGATACTCCAAAACCGAGAA CAAGCTGTTCCCAAAATAGATCTCAGCGGAGATGCACGGAGGATATTCGGGAATGGGCTTTCGA	43	SEQ ID NO: 15 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTTCTGACGGTGAAAAATGGGCTAAGGCTCGAAGAATCGCTGATTACGCTTTCCACGGGGATCT CCTAAGAAATATGGGGCCAACCATGGTTTCCTGTGCTGAGGCAATGGTGGAAAAGTGGAAGCAT CATCAAGGCAAAGAGCTTGATTTGTTCGAAGAGTTTAAGGTGCTCACTTCAGATATCATTGCAC ATACAGCCTTTGGAAGCAGTTATTTGGAAGGGAAAGTTATTTTTCAGACTCTAAGTAAGCTGAG CATGATATTATTTAAGAATCAGTTCAAACGAAGGATTCCTGTTATCAGCAAGTTCTTCAGATCA AAGGATGCGAGGGAGGGAGAGGAGCTGGAAAGAAGGTTGAAAAATTCCATAATTTCAATAATGG AAAAGAGAGAAGAGAAGGTGATAAGTGGTGAAGCAGATAACTATGGTAATGATTTTCTTGGATT ACTTTTGAAGGCAAAGAATGAGCCTGACCAGAGGCAGAGGATTTCTGTTGATGATGTAGTGGAT GAATGCAAAACAGTTTACTTCGCTGGGCAAGAAACTACAAGTGTTTTGCTTGCTTGGACCGCCT TTCTTTTAGCAACTCATGAGCATTGGCAAGAAGAAGCAAGAAAGGAAGTGCTGAATATGTTTGG CAACAAGAATCCAACTTTAGAAGGCATCACAAAATTAAAGATTATGAGCATGATCATCAAGGAA TCTCTAAGATTATATCCTCCAGCCCCGCCCATGTCAAGGAAGGTTAAAAAGGAAGTCAGATTGG GGAAGCTGGTTCTCCCCCCCCAACATTCAAGTAAGCATCTCAACTATTGCAGTTCATCATGATAC TGCAATATGGGGTGAAGATGCCCATGTATTCAAACCAGAAAGATTTTCTGAAGGAACAGCTAAA GATATCCCATCAGCTGCATACATCCCATTTGGCTTTGGTCCTCGAAACTGCATCGGCAATATCT TGGCCATCAACGAAACTAAGATTGCACTGTCGATGATTCTACAACGATTTTCTTTCACCATCTC CCCGGCCTACGTCCACGCACCTTTCCAGTTCCTCACTATCTGCCCCCAACACGGGGTTCAGGTA AAGCTTCAGTCCCTATTAAGTGAAAGGTGA		
CYP7604 gene (coding sequence)	ATGGAAGCTGAATTTGGTGCCGGTGCTACTATGGTATTATCCGTTGTCGCAATCGTCTTCTTTT TCACATTTTTACACTTGTTTGAATCTTTCTTTTTGAAGCCAGATAGATTGAGATCTAAGTTGAG AAAGCAAGGTATTGGTGGTCCATCTCCTTCATTTTTGTTGGGTAATTTGTCAGAAATTAAATCC ATCAGAGCTTTGTCTTCACAAGCTAAGAACGCAGAAGATGCCTCTGCTGGTGGTGGTGGTGGTT CCGCCAGTATAGCTCATGGTTGGACTTCAAATTTGTTTCCTCACTTAGAACAATGGAGAAACAG ATATGGTCCAATTTTCGTATACTCCAGTGGTACAATCCAAATCTTGTGTATCACAGAAATGGAA ACCGTTAAGGAAATCTCTTTGTCAACCTCCTTGAGTTTAGGTAAACCTGCTCATTTGTCTAAGG ATAGAGGTCCATTGTTAGGTTTGGGTATCTTAGCCTCTTCAGGTCCTATTTGGGTTCACCAAAG AAAGATCATCGCTCCACAATTGTATTTGGATAAAGTAAAGGGTATGACCTCATTGATGGTTGAA AGTGCAAATTCTATGTTAAGATCCTGGGAAACTAAAGTTGAAAATCATGGTGGTCAAGCCGAAA TTAACGTCGATGGTGACTTGAGAGCATTAAGTGCCGATATCATTTCTAAGGCTTGCTTTGGTTC AAACTATTCCGAAGGTGAAGAAATTTTCTTGAAGTTGAGAGCATTGCAAGTTGTCATGAGTAAG GGTTCTATTGGTATACCTGGTTTTAGATACATACCAACTAAAAATAACAGAGAAATGTGGAAGT TGGAAAAGGAAATCGAATCAATGATCTTGAAGGTTGCCAACGAAAGAACACAACATTCCAGTCA CGAACAAGATTTGTTGCAAATGATTTTGGAAGGTGCAAAGTCTTTGGGTGAAGACAATAAGAGT ATGAACATATCAAGAGACAAGTTTATTGTTGACAATTGTAAGAACATCTATTTCGCTGGTCATG AAACTACAGCTATAACCGCATCTTGGTGCTTGATGTTGTTAGCTGCACACCCTGATTGGCAAGC AAGAGCCAGATCTGAAGTTTTACAATGTTGCGATGACAGACCAATCGATGCAGACACAGTCAAA AATATGAAGACCTTGACTATGGTAATTCAAGAAACTTTGAGATTGTACCCACCTGCTGTATTCG TTACAAGACAAGCATTAGAAGATATCAGATTCAAAAACATCACAATACCAAAGGGTATGAACTT	44	SEQ ID NO: 16 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TCATATACCAATCCCTATGTTGCAACAAGACTTCCACTTATGGGGTCCTGATGCTTGTTCATTT GACCCACAAAGATTCTCCAATGGTGTCTTAGGTGCATGCAAAAACCCACAAGCCTATATGCCTT TTGGTGTTGGTCCAAGAGTCTGTGCCGGTCAACATTTCGCTATGATCGAATTGAAAGTCATCGT ATCATTGGTTTTGTCCAGATTCGAATTTTCTTTGTCACCTTCCTACAAGCATTCACCAGCCTTC AGATTAGTTGTCGAACCAGAAAACGGTGTCATATTGCATGTCAGAAAGTTGTGA		
CYP8224 gene (coding sequence)	ATGGAAGTGGATATCAATATCTTCACCGTCTTTTCCTTCGTATTATGCACAGTCTTCCTCTTCT TTCTATCCTTCTTGATCCTCCTCCTCCTCCGAACGCTCGCCGGAAAATCCATAACGAGCTCCGA GTACACGCCAGTGTACGGCACCGTCTACGGTCAGGCTTTCTATTTCAACAACCTGTACGATCAT CTAACGGAGGTGGCCAAGAGACATCGAACCTTCCGGCTGCTTGCGCCGGCATACAGCGAGATAT ACACGACCGATCCGAGAAACATCGAGCATATGTTGAAGACGAAATTCGATAAGTATTCGAAAGG AAGCAAGGATCAAGAAATCGTTGGGGATCTGTTTGGAGAGGGGATATTTGCAGTCGATGGAGAT AAGTGGAAGCAGCAGAGGAAGCTGGCTAGCTATGAATTCTCGACGAGGATTCTTAGGGATTTTA GCTGCTCGGTTTTCAGACGAAGTGCTGCTAAACTTGTTGGAGTTGTTTCGGAGTTTTCCAGCAT GGGTCGGGTTTTTGATATCCAGGATTTGCTAATGCGGTGCGCTTTGGACTCCATTTTCAAAGTG GGGTTCGGGGTTGATTTGAATTGCTTGGAGGAATCAAGCAAAGAAGGGAGCGATTTCATGAAAG CCTTCGATGATTCTAGCGCTCAGATTTTTTGGCGCTATATCGATCCCTTCTGGAAATTGAAGAG ATTGCTTAACATCGGTTCCGAAGCTTCGTTTAGGAACAACATAAAAACCATAGATGCTTTTGTG CACCAGTTGATCAGAGACAAGAGAAAATTGCTTCAGCAACCGAATCACAAGAATGACAAAGAGG ACATACTTTGGAGGTTTCTGATGGAAAGTGAGAAGGATCCAACAAGAATGAATGATCAATATCT AAGGGATATAGTCCTCAATTTCATGTTGGCTGGCAAAGATTCAAGTGGAGGAACTCTGTCCTGG TTCTTCTACATGCTATGCAAGAACCCTTTAATACAGGAAAAAGTTGCAGAAGAAGTGAGGCAAA TTGTTGCGTTTGAAGGGGAAGAAGTTGACATCAATTTGTTCATACAAAACTTAACTGATTCAGC TCTTGACAAAATGCATTATCTTCATGCAGCATTGACCGAGACTCTGAGGCTATATCCTGCAGTC CCTTTGGATGGAAGGACTGCAGAAATAGATGACATTCTTCCTGATGGCTATAAACTAAGAAAAG GGGATGGAGTATACTACATGGCCTATTCCATGGGCAGGATGTCCTCCCTTTGGGGAGAAGATGC TGAAGATTTTAAACCCGAAAGATGGCTTGAAAGTGGAACTTTTCAACCCGAATCACCTTTCAAA TTCATCGCTTTTCATGCGGGTCCTCGAATGTGTTTGGGAAAAGAGTTTGCTTATCGACAAATGA AGATAGTATCTGCTGCTTTGCTTCAATTTTTTCGATTCAAAGTAGCTGATACAACGAGGAATGT GACTTATAGGATCATGCTTACCCTTCACATTGATGGAGGTCTCCCTCTTCTTGCAATTCCGAGA ATTAGAAAATTTACCTAA	45	SEQ ID NO: 17 in WO2016050890
CYP8728 gene sequence	TTGGATAGTGGAGTTAAAAGAGTGAAACGGCTAGTTGAAGAGAAACGGCGAGCAGAATTGTCTG CCCGGATTGCCTCTGGAGAATTCACAGTCGAAAAAGCTGGTTTTCCATCTGTATTGAGGAGTGG CTTATCAAAGATGGGTGTTCCCAGTGAGATTCTGGACATATTATTTGGTTTCGTTGATGCTCAA GAAGAATATCCCAAGATTCCCGAAGCAAAAGGATCAGTAAATGCAATTCGTAGTGAGGCCTTCT TCATACCTCTCTATGAGCTTTATCTCACATATGGTGGAATATTTAGGTTGACTTTTGGGCCAAA GTCATTCTTGATAGTTTCTGATCCTTCCATTGCTAAACATATACTGAAGGATAATCCGAGGAAT TATTCTAAGGGTATCTTAGCTGAAATTCTAGAGTTTGTCATGGGGAAGGGACTTATACCAGCTG ACGAGAAGATATGGCGTGTACGAAGGCGGGCTATAGTCCCATCTTTGCATCTGAAGTATGTAGG	46	SEQ ID NO: 18 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TGCTATGATTAATCTTTTTGGAGAAGCTGCAGATAGGCTTTGCAAGAAGCTAGATGCTGCAGCA TCTGATGGGGTTGATGTGGAAATGGAGTCCCTGTTCTCCCGTTTGACTTTAGATATCATTGGCA AGGCAGTTTTTAACTATGACTTTGATTCACTTACAAATGACACTGGCATAGTTGAGGCTGTTTA CACTGTGCTAAGAGAAGCAGAGGATCGCAGTGTTGCACCAATTCCAGTATGGGAAATTCCAATT TGGAAGGATATTTCACCACGGCAAAAAAAGGTCTCTAAAGCCCTCAAATTGATCAACGACACCC TCGATCAACTAATTGCTATATGCAAGAGGATGGTTGATGAGGAGGAGCTGCAGTTTCATGAGGA ATACATGAATGAGCAAGATCCAAGCATCCTTCATTTCCTTTTGGCATCAGGAGATGATGTTTCA AGCAAGCAGCTTCGTGATGACTTGATGACTATGCTTATAGCTGGGCATGAAACATCTGCTGCAG TTTTAACATGGACCTTTTATCTTCTTTCCAAGGAGCCGAGGATCATGTCCAAGCTCCAGGAGGA GGTTGATTCAGTCCTTGGGGATCGGTTTCCAACTATTGAAGATATGAAGAACCTCAAATATGCC ACACGAATAATTAACGAATCCTTGAGGCTTTACCCACAGCCACCAGTTTTAATACGTCGATCTC TTGACAATGATATGCTCGGGAAGTACCCCATTAAAAAGGGTGAGGACATATTCATTTCTGTTTG GAACTTGCATCGCAGTCCAAAACTCTGGGATGATGCGGATAAATTTAATCCTGAAAGGTGGCCT CTGGATGGACCCAATCCAAATGAGACAAATCAAAATTTCAGATATTTACCTTTTGGTGGCGGAC CACGGAAATGTGTGGGAGACATGTTTGCTTCGTACGAGACTGTTGTAGCACTTGCAATGCTTGT TCGGCGATTTGACTTCCAAATGGCACTTGGAGCACCTCCTGTAAAAATGACAACTGGAGCTACA ATTCACACAACAGATGGATTGAAAATGACAGTTACACGAAGAATGAGACCTCCAATCATACCCA CATTAGAGATGCCTGCAGTGGTCGTTGACTCGTCTGTCGTGGACTCGTCCGTCGCCATTTTGAA AGAAGAAACACAAATTGGTTAG		
DNA sequence encoding CYP10020	CAGTTCCTCTCCTGGTCCTCCCAGTTTGGCAAGAGGTTCATCTTCTGGAATGGGATCGAGCCCA GAATGTGCCTCACCGAGACCGATTTGATCAAAGAGCTTCTCTCTAAGTACAGCGCCGTCTCCGG TAAGTCATGGCTTCAGCAACAGGGCTCCAAGCACTTCATCGGCCGCGGTCTCTTAATGGCCAAC GGCCAAAACTGGTACCACCAGCGTCACATCGTCGCGCCGGCCTTCATGGGAGACAGACTCAAGA GTTACGCCGGGTACATGGTGGAATGCACAAAGGAGATGCTTCAGTCAATTGAAAACGAGGTCAA CTCGGGGCGATCCGAGTTCGAAATCGGTGAGTATATGACCAGACTCACCGCCGATATAATATCA CGAACCGAGTTCGAAAGCAGCTACGAAAAGGGAAAGCAAATTTTCCATTTGCTCACCGTTTTAC AGCATCTCTGCGCTCAGGCGAGCCGCCACCTCTGCCTTCCTGGAAGCCGGTTTTTTCCGAGTAA ATACAACAGAGAGATAAAGGCATTGAAGACGAAGGTGGAGGGGTTGTTAATGGAGATAATACAG AGCAGAAGAGACTGTGTGGAGGTGGGGAGGAGCAGTTCGTATGGAAATGATCTGTTGGGAATGT TGCTGAATGAGATGCAGAAGAAGAAAGATGGGAATGGGTTGAGCTTGAATTTGCAGATTATAAT GGATGAATGCAAGACCTTCTTCTTCGCCGGCCATGAAACCACTGCTCTTTTGCTCACTTGGACT GTAATGTTATTGGCCAGCAACCCTTCTTGGCAACACAAGGTTCGAGCCGAAGTTATGGCCGTCT GCAATGGAGGAACTCTCTCTCTTGAACATCTCTCCAAGCTCTCTCTGTTGAGTATGGTGATAAA TGAATCGTTGAGGCTATACCCGCCAGCAAGTATTCTTCCAAGAATGGCATTTGAAGATATAAAG CTGGGAGATCTTGAGATCCCAAAAGGGCTGTCGATATGGATCCCAGTGCTTGCAATTCACCACA GTGAAGAGCTATGGGGCAAAGATGCAAATGAGTTCAACCCAGAAAGATTTGCAAATTCAAAAGC CTTCACTTCGGGGAGATTCATTCCCTTTGCTTCTGGCCCTCGCAACTGCGTTGGCCAATCATTT GCTCTCATGGAAACCAAGATCATTTTGGCTATGCTCATCTCCAAGTTTTCCTTCACCATCTCTG	47	SEQ ID NO: 19 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ACAATTATCGCCATGCACCCGTGGTCGTCCTCACTATAAAACCCAAATACGGAGTCCAAGTTTG CTTGAAGCCTTTCAATTAA		
DNA sequence encoding CYP10285	ATGGAAGACACCTTCCTACTCTATCCTTCCCTCTCTCTTCTCTTTCTTCTTTTTGCTTTCAAGC TCATCCGTCGATCCGGAGGAGTTCGCAGGAACTTACCGCCGAGTCCGCCCTCTCTTCCGGTTAT CGGCCACCTCCATCTCTTGAAAAAGCCACTCCACCGGACTTTCCAGAAACTTTCCGCCAAATAT GGTCCTGTTATGTCCCTCCGCCTCGGGTCTCGCCTCGCAGTCATTGTATCGTCGTCGTCGGCGG TGGACGAGTGTTTCACTAAAAACGACGTCGTGCTCGCCAACCGTCCTCGTTTGCTAATTGGCAA ACACCTCGGCTACAACTACACTACCATGGTTGGGGCTCCCTACGGCGACCACTGGCGTAGCCTC CGCCGCATCGGTGCCCTCGAAATCTTCTCTTCATCTCGCCTCAACAAATTCGCCGACATCCGAA GGGATGAAGTAGAGGGATTGCTTCGCAAACTCTCACGCAATTCGCTCCATCAATTCTCGAAAGT GGAAGTTCAATCGGCCTTGTCGGAGCTGACGTTCAACATCTCGATGAGAATGGCGGCAGGGAAA CGGTATTACGGAGATGACGTGACGGACGAGGAAGAGGCGAGAAAGTTCAGAGAGTTAATTAAAC AGATAGTGGCGCTGGGCGGAGTATCAAATCCAGGGGATTTCGTCCCGATTCTGAATTGGATTCC GAACGGTTTCGAGAGGAAGTTGATCGAGTGTGGGAAGAAGACGGATGCGTTCTTGCAGGGGCTG ATCGAGGACCACCGGAGAAAGAAGGAAGAGGGTAGGAACACGATGATCGATCACCTGCTCTCTC TGCAAGAATCGGAGCCTGCTCACTACGGAGACCAAATAATCAAAGGATTTATACTGGTGTTACT GACGGCGGGGACCGATACATCGGCCGTGACAATGGAGTGGGCGCTATCTCATCTCCTGAACAAT CCTGAAGTGCTAAAGAAGGCAAGAGATGAGGTCGACACTGAAATTGGACAAGAACGACTTGTCG AAGAATCAGACGTAGTATCTAAGTTACCCTATCTTCAAGGGATCATCTCCGAGACTCTCCGGCT GAATCCCGCCGCTCCGATGTTGTTGCCCCATTACGCCTCGGACGACTGCACGATATGTGGATAC GACGTGCCACGTGACACAATCGTAATGGTCAATGCATGGGCCATACATAGGGATCCAAACGAAT GGGAGGAGCCCACGTGTTTCAGACCAGAACGATATGAAAAGTCGTCGTCGGAAGCGGAGGTACA CAAGTCGGTGAGTTTCGGGGTGGGAAGGCGAGCTTGTCCTGGGTCTGGCATGGCGCAGAGGGTG ATGGGCTTGACTTTGGCGGCACTGGTTCAGTGCTTCGAGTGGGAGAGAGTTGGAGAAGAAGAAG TGGACATGAACGAAGGCTCAGGTGCCACAATGCCCAAGATGGTGCCATTGGAGGCCATGTGCAG AGCTCGTCCCATCGTCCACAACCTTCTTTACTGA	48	SEQ ID NO: 20 in WO2016050890
CYP5491 protein	MWTVVLGLATLFVAYYIHWINKWRDSKFNGVLPPGTMGLPLIGETIQLSRPSDSLDVHPFIQKK VERYGPIFKTCLAGRPVVVSADAEFNNYIMLQEGRAVEMWYLDTLSKFFGLDTEWLKALGLIHK YIRSITLNHFGAEALRERFLPFIEASSMEALHSWSTQPSVEVKNASALMVFRTSVNKMFGEDAK KLSGNIPGKFTKLLGGFLSLPLNFPGTTYHKCLKDMKEIQKKLREVVDDRLANVGPDVEDFLGQ ALKDKESEKFISEEFIIQLLFSISFASFESISTTLTLILKLLDEHPEVVKELEAEHEAIRKARA DPDGPITWEEYKSMTFTLQVINETLRLGSVTPALLRKTVKDLQVKGYIIPEGWTIMLVTASRHR DPKVYKDPHIFNPWRWKDLDSITIQKNFMPFGGGLRHCAGAEYSKVYLCTFLHILCTKYRWTKL GGGRIARAHILSFEDGLHVKFTPKE	49	SEQ ID NO: 44 in WO2016050890
Squalene epoxidase (*S. cerevisiae*)	MSAVNVAPELINADNTITYDAIVIGAGVIGPCVATGLARKGKKVLIVERDWAMPDRIVGELMQP GGVRALRSLGMIQSINNIEAYPVTGYTVFFNGEQVDIPYPYKADIPKVEKLKDLVKDGNDKVLE DSTIHIKDYEDDERERGVAFVHGRFLNNLRNITAQEPNVTRVQGNCIEILKDEKNEVVGAKVDI DGRGKVEFKAHLTFICDGIFSRFRKELHPDHVPTVGSSFVGMSLFNAKNPAPMHGHVILGSDHM	50	SEQ ID NO: 54 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	PILVYQISPEETRILCAYNSPKVPADIKSWMIKDVQPFIPKSLRPSFDEAVSQGKFRAMPNSYL PARQNDVTGMCVIGDALNMRHPLTGGGMTVGLHDVVLLIKKIGDLDFSDREKVLDELLDYHFER KSYDSVINVLSVALYSLFAADSDNLKALQKGCFKYFQRGGDCVNKPVEFLSGVLPKPLQLTRVF FAVAFYTIYLNMEERGFLGLPMALLEGIMILITAIRVFTPFLFGELIG		
Squalene epoxidase (*Gynostemma pentaphyllum*)	MVDQFSLAFIFASVLGAVAFYYLFLRNRIFRVSREPRRESLKNIATTNGECKSSYSDGDIIIVG AGVAGSALAYTLGKDGRRVHVIERDLTEPDRTVGELLQPGGYLKLTELGLEDCVNEIDAQRVYG YALFKDGKDTKLSYPLEKFHSDVSGRSFHNGRFIQRMREKAATLPNVRLEQGTVTSLLEENGII KGVQYKSKTGQEMTAYAPLTIVCDGCFSNLRRSLCNPKVDVPSCFVALVLENCELPHANYGHVI LADPSPILFYPISSTEVRCLVDVPGQKVPSISNGEMANYLKSVVAPQIPPQIYDALRSCYDKGN IRTMPNRSMPADPYPTPGALLMGDAFNMRHPLTGGGMTVALSDIVVLRDLLKPLRDLHDAPILS NYLEAFYTLRKPVASTINTLAGALYKVFCASPDQARREMRQACFDYLSLGGVFSNGPVSLLSGL NPRPLSLVLHFFAVAIYGVGRLLIPFPSPRRVWIGARLISGASGIIFPIIKAEGVRQIFFPATL PAYYRAPPLVRGR	51	SEQ ID NO: 88 in WO2016050890
Squaiene epoxidase 1 (*Arabidopsis thaliana*)	MESQLWNWILPLLISSLLISFVAFYGFFVKPKRNGLRHDRKTVSTVTSDVGSVNITGDTVADVI VVGAGVAGSALAYTLGKDKRRVHVIERDLSEPDRIVGELLQPGGYLKLLELGIEDCVEEIDAQR VYGYALFKNGKRIRLAYPLEKFHEDVSGRSFHNGRFIQRMREKAASLPNVQLEQGTVLSLLEEN GTIKGVRYKNKAGEEQTAFAALTIVCDGCFSNLRRSLCNPQVEVPSCFVGLVLENCNLPYANHG HVVLADPSPILMYPISSTEVRCLVDVPGQKVPSIANGEMKNYLKTVVAPQMPHEVYDSFIAAVD KGNIKSMPNRSMPASPYPTPGALLMGDAFNMRHPLTGGGMTVALADIVVLRNLLRPLRDLSDGA SLCKYLESFYTLRKPVAATINTLANALYQVFCSSENEARNEMREACFDYLGLGGMCTSGPVSLL SGLNPRPLTLVCHFFAVAVYGVIRLLIPFPSPKRIWLGAKLISGASGIIFPIIKAEGVRQMFFP ATVPAYYYKAPTVGETKCS	52	SEQ ID NO: 89 in WO2016050890
Squalene epoxidase 4 (*Arabidopsis thaliana*)	MTYAWLWTLLAFVLTWMVFHLIKMKKAATGDLEAEAEARRDGATDVIIVGAGVAGASLAYALAK DGRRVHVIERDLKEPQRFMGELMQAGGRFMLAQLGLEDCLEDIDAQEAKSLAIYKDGKHATLPF PDDKSFPHEPVGRLLRNGRLVQRLRQKAASLSNVQLEEGTVKSLIEEEGVVKGVTYKNSAGEEI TAFAPLTVVCDGCYSNLRRSLVDNTEEVLSYMVGYVTKNSRLEDPHSLHLIFSKPLVCVIYQIT SDEVRCVAEVPADSIPSISNGEMSTFLKKSMAPQIPETGNLREIFLKGIEEGLPEIKSTATKSM SSRLCDKRGVIVLGDAFNMRHPIIASGMMVALSDICILRNLLKPLPNLSNTKKVSDLVKSFYII RKPMSATVNTLASIFSQVLVATTDEAREGMRQGCFNYLARGDFKTRGLMTILGGMNPHPLTLVL HLVAITLTSMGHLLSPFPSPRRFWHSLRILAWALQMLGAHLVDEGFKEMLIPTNAAAYRRNYIA TTTV	53	SEQ ID NO: 90 in WO2016050890
Squaiene epoxidase 6 (*Arabidopsis thaliana*)	MAFTHVCLWTLVAFVLTWTVFYLTNMKKKATDLADTVAEDQKDGAADVIIVGAGVGGSALAYAL AKDGRRVHVIERDMREPERMMGEFMQPGGRLMLSKLGLQDCLEDIDAQKATGLAVYKDGKEADA PFPVDNNNFSYEPSARSFHNGRFVQQLRRKAFSLSNVRLEEGTVKSLLEEKGVVKGVTYKNKEG EETTALAPLTVVCDGCYSNLRRSLNDDNNAEIMSYIVGYISKNCRLEEPEKLHLILSKPSFTMV YQISSTDVRCGFEVLPENFPSIANGEMSTFMKNTIVPQVPPKLRKIFLKGIDEGAHIKVVPAKR MTSTLSKKKGVIVLGDAFNMRHPVVASGMMVLLSDILILRRLLQPLSNLGDANKVSEVINSFYD IRKPMSATVNTLGNAFSQVLIGSTDEAKEAMRQGVYDYLCSGGFRTSGMMALLGGMNPRPLSLV	54	SEQ ID NO: 91 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	YHLCAITLSSIGQLLSPFPSPLRIWHSLKLFGLAMKMLVPNLKAEGVSQMLFPANAAAYHKSYM AATTL		
Squalene epoxidase 5 (*Arabidopsis thaliana*)	MAFTNVCLWTLLAFMLTWTVFYVTNRGKKATQLADAVVEEREDGATDVIIVGAGVGGSALAYAL AKDGRRVHVIERDLREPERIMGEFMQPGGRLMLSKLGLEDCLEGIDAQKATGMTVYKDGKEAVA SFPVDNNNFPFDPSARSFHNGRFVQRLRQKASSLPNVRLEEGTVKSLIEEKGVIKGVTYKNSAG EETTALAPLTVVCDGCYSNLRRSLNDNNAEVLSYQVGFISKNCQLEEPEKLKLIMSKPSFTMLY QISSTDVRCVFEVLPNNIPSISNGEMATFVKNTIAPQVPLKLRKIFLKGIDEGEHIKAMPTKKM TATLSEKKGVILLGDAFNMRHPAIASGMMVLLSDILILRRLLQPLSNLGNAQKISQVIKSFYDI RKPMSATVNTLGNAFSQVLVASTDEAKEAMRQGCYDYLSSGGFRTSGMMALLGGMNPRPISLIY HLCAITLSSIGHLLSPFPSPLRIWHSLRLFGLAMKMLVPHLKAEGVSQMLFPVNAAAYSKSYMA ATAL	55	SEQ ID NO: 92 in WO2016050890
Squalene epoxidase 2 (*Arabidopsis thaliana*)	MKPFVIRNLPRFQSTLRSSLLYTNHRPSSRFSLSTRRFTTGATYIRRWKATAAQTLKLSAVNST VMMKPAKIALDQFIASLFTFLLLYILRRSSNKNKKNRGLVVSQNDTVSKNLETEVDSGTDVIIV GAGVAGSALAHTLGKEGRRVHVIERDFSEQDRIVGELLQPGGYLKLIELGLEDCVKKIDAQRVL GYVLFKDGKHTKLAYPLETFDSDVAGRSFHNGRFVQRMREKALILSNVRLEQGTVTSLLEEHGT IKGVRYRTKEGNEFRSFAPLTIVCDGCFSNLRRSLCKPKVDVPSTFVGLVLENCELPFANHGHV VLGDPSPILMYPISSSEVRCLVDVPGQKLPPIANGEMAKYLKTRVAPQVPTKVREAFITAVEKG NIRTMPNRSMPADPIPTPGALLLGDAFNMRHPLTGGGMTVALADIVVLRDLLRPIRNLNDKEAL SKYIESFYTLRKPVASTINTLADALYKVFLASSDEARTEMREACFDYLSLGGVFSSGPVALLSG LNPRPLSLVLHFFAVAIYAVCRLMLPFPSIESFWLGARIISSASSIIFPIIKAEGVRQMFFPRT IPAIYRAPP	56	SEQ ID NO: 93 in WO2016050890
Squalene epoxidase 3 (*Arabidopsis thaliana*)	MAPTIFVDHCILTTTFVASLFAFLLLYVLRRRSKIIHGSVNVRNGTLTVKSGTDVDIIIVGAGV AGAALAHTLGKEGRRVHVIERDLTEPDRIVGELLQPGGYLKLIELGLEDCVKDIDAQRVLGYAL FKDGKHTKLSYPLDQFDSDVAGRSFHNGRFVQRMREKASLLPNVRMEQGTVTSLVEENGIIKGV QYKTKDGQELKSFAPLTIVCDGCFSNLRRSLCKPKVEVPSNFVGLVLENCELPFPNHGHVVLGD PSPILFYPISSSEVRCLVDVPGSKLPSVASGEMAHHLKTMVAPQVPPQIRDAFISAVEKGNIRT MPNRSMPADPIHTPGALLLGDAFNMRHPLTGGGMTVALSDIVILRDLLNPLVDLTNKESLSKYI ESFYTLRKPVASTINTLAGALYKVFLASPDDARSEMRRACFDYLSLGGVCSSGPVALLSGLNPR PMSLVLHFFAVAIFGVGRLLVPLPSVKRLWLGARLISSASGIIFPIIKAEGVRQMFFPRTIPAI YRAPPIPSSSSPQ	57	SEQ ID NO: 94 in WO2016050890
Squalene monooxygenase 1,1 (*Brassica napus*)	MDLAFPHVCLWTLLAFVLTWTVFYVNNRRKKVAKLPDAATEVRRDGDADVIIVGAGVGGSALAY ALAKDGRRVHVIERDMREPVRMMGEFMQPGGRLLLSKLGLEDCLEGIDEQIATGLAVYKDGQKA LVSFPEDNDFPYEPTGRAFYNGRFVQRLRQKASSLPTVQLEEGTVKSLIEEKGVIKGVTYKNSA GEETTAFAPLTVVCDGCYSNLRRSVNDNNAEVISYQVGYVSKNCQLEDPEKLKLIMSKPSFTML YQISSTDVRCVMEIFPGNIPSISNGEMAVYLKNTMAPQVPPELRKIFLKGIDEGAQIKAMPTKR MEATLSEKQGVIVLGDAFNMRHPAIASGMMVVLSDILILRRLLQPLRNLSDANKVSEVIKSFYV IRKPMSATVNTLGNAFSQVLIASTDEAKEAMRQGCFDYLSSGGFRTSGMMALLGGMNPRPLSLI FHLCGITLSSIGQLLSPFPSPLGIWHSLRLFGAEGVSQMLSPAYAAAYRKSYMTATAL	58	SEQ ID NO: 95 in WO2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Squalene monooxygenase 1,2 (*Brassica napus*)	MDMAFVEVCLRMLLVFVLSWTIFHVNNRKKKKAIKLADLATEERKEGGPDVIIVGAGVGGSALA YALAKDGRRVHVIERDMREPVRMMGEFMQPGGRLMLSKLGLQDCLEEIDAQKSTGIRLFKDGKE TVACFPVDTNFPYEPSGRFFHNGRFVQRLRQKASSLPNVRLEEGTVRSLIEEKGVVKGVTYKNS SGEETISFAPLTVVCDGCHSNLRRSLNDNNAEVIAYEIGYISRNCRLEQPDKLHLIMAKPSFAM LYQVSSIDVRCNFELLSKNLPSVSNGEMTSFVRNSIAPQVPLKLRKTFLKGLDEGSHIKITQAK RIPATLSRKKGVIVLGDAFNMRHPVIASGMMVLLSDILILSRLLKPLGNLGDENKVSEVMKSFY ALRKPMSATVNTLGNSFWQVLIASTDEAKEAMRQGCFDYLSSGGFRTSGLMALIGGMNPRPLSL FYHLFVISLSSIGQLLSPFPTPLRVWHSLRLLDLSLKMLVPHLKAEGIGQMLSPTNAAAYRKSY MAATVV	59	SEQ ID NO: 96 in WO2016050890
Squalene epoxidase (*Euphorbia tirucalli*)	MEVIFDTYIFGTFFASLCAFLLLFILRPKVKKMGKIREISSINTQNDTAITPPKGSGTDVIIVG AGVAGAALACTLGKDGRRVHVIERDLKEPDRIVGELLQPGGYLKLVELGLQDCVEEIDAQRIVG YALFMDGNNTKLSYPLEKFDAEVSGKSFHNGRFIQRMREKAASLPNVQLEQGTVTSLLEENGTI KGVQYKIKDGQEHKAYAPLTVVCDGCFSNLRRSLCKPKVDVPSHFVGLVLENCDLPFANHGHVI LADPSPILFYPISSTEVRCLVDVPGQKLPSIASGEMAKYLKTMVAKQIPPVLHDAFVSAIDKGN IRTMPNRSMPADPLPTPGALLMGDAFNMRHPLTGGGMTVALADIVLLRDLLKPLRDLNDAPALA KYLESFYTLRKPVASTINTLAGALYKVFSASPDEARKEMRQACFDYLSLGGECAMGPVSLLSGL NPSPLTLVLHFFGVAIYGVGRLLIPFPTPKGMWIGARIISSASGIIFPIIKAEGVRQVFFPATV PAIYRNPPVNGKSVEVPKS	60	SEQ ID NO: 97 in WO2016050890
Squalene epoxidase (*Medicago truncatula*)	MIDPYGFGWITCTLITLAALYNFLFSRKNHSDSTTTENITTATGECRSFNPNGDVDIIIVGAGV AGSALAYTLGKDGRRVLIIERDLNEPDRIVGELLQPGGYLKLIELGLDDCVEKIDAQKVFGYAL FKDGKHTRLSYPLEKFHSDIAGRSFHNGRFILRMREKAASLPNVRLEQGTVTSLLEENGTIKGV QYKTKDAQEFSACAPLTIVCDGCFSNLRRSLCNPKVEVPSCFVGLVLENCELPCADHGHVILGD PSPVLFYPISSTEIRCLVDVPGQKVPSISNGEMAKYLKTVVAPQVPPELHAAFIAAVDKGHIRT MPNRSMPADPYPTPGALLMGDAFNMRHPLTGGGMTVALSDIVVLRNLLKPLRDLNDASSLCKYL ESFYTLRKPVASTINTLAGALYKVFCASPDPARKEMRQACFDYLSLGGLFSEGPVSLLSGLNPC PLSLVLHFFAVAIYGVGRLLLPFPSPKRLWIGIRLIASASGIILPIIKAEGIRQMFFPATVPAY YRAPPDA	61	SEQ ID NO: 98 in WO2016050890
Squalene monooxygenase (*Medicago truncatula*)	MDLYNIGWILSSVLSLFALYNLIFAGKKNYDVNEKVNQREDSVISIDAGEIKSDKLNGDADVII VGAGIAGAALAHTLGKDGRRVHIIERDLSEPDRIVGELLQPGGYLKLVELGLQDCVDNIDAQRV FGYALFKDGKHTRLSYPLEKFHSDVSGRSFHNGRFIQRMREKAASLPNVNMEQGTVISLLEEKG TIKGVQYKNKDGQALTAYAPLTIVCDGCFSNLRRSLCNPKVDNPSCFVGLILENCELPCANHGH VILGDPSPILFYPISSTEIRCLVDVPGTKVPSISNGDMTKYLKTTVAPQVPPELYDAFIAAVDK GNIRTMPNRSMPADPRPTPGAVLMGDAFNMRHPLTGGGMTVALSDIVVLRNLLKPMRDLNDAPT LCKYLESFYTLRKPVASTINTLAGALYKVFSASPDEARKEMRQACFDYLSLGGLFSEGPISLLS GLNPRPLSLVLHFFAVAVFGVGRLLLPFPSPKRVWIGARLLSGASGIILPIIKAEGIRQMFFPA TVPAYYRAPPVNAF	62	SEQ ID NO: 99 in WO2016050890
Squalene monooxygenase	MADNYLLGWILCSIIGLFGLYYMVYLVVKREEEDNNRKALLQARSDSAKTMSAVSQNGECRSDN PADADIIIVGAGVAGSALAHTLGKDGRRVHVIERDLTEPDRIVGELLQPGGYLKLIELGLEDCV	63	SEQ ID NO: 100 in

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
(*Ricinus communis*)	EEIDAQRVFGYALFMDGKHTQLSYPLEKFHSDVAGRSFHNGRFIQRMREKASSIPNVRLEQGTV TSLIEEKGIIRGVVYKIKTGEELTAFAPLTIVCDGCFSNLRRSLCNPKVDVPSCFVGLVLEDCK LPYQYHGHVVLADPSPILFYQISSTEVRCLVDVPGQKVPSISNGEMAKYLKNVVAPQVPPEIYD SFVAAVDKGNIRTMPNRSMPASPYPTPGALLMGDAFNMRHPLTGGGMTVALSDIVVLRELLKPL RDLHDAPTLCRYLESFYTLRKPVASTINTLAGALYKVFCASSDEARNEMRQACFDYLSLGGVFS TGPISLLSGLNPRPLSLVVHFFAVAIYGVGRLLLPFPSPKRVWVGARLISGASGIIFPIIKAEG VRQMFFPATVPAYYRAPPVECN		WO2016050890
Squalene monooxygenase (*Ricinus communis*)	MEYKLAVAGIIASLWALFMLCSLKRKKNITRASFNNYTDETLKSSSKEICQPEIVASPDIIIVG AGVAGAALAYALGEDGRQVHVIERDLSEPDRIVGELLQPGGYLKLIELGLEDCVEKIDAQQVFG YAIFKDGKSTKLSYPLDGFQTNVSGRSFHNGRFIQRMREKATSLPNLILQQGTVTSLVEKKGTV KGVNYRTRNGQEMTAYAPLTIVCDGCFSNLRRSLCNPKVEIPSCFVALVLENCDLPYANHGHVI LADPSPILFYPISSTEVRCLVDIPGQKVPSISNGELAQYLKSTVAKQIPSELHDAFISAIEKGN IRTMPNRSMPASPHPTPGALLVGDAFNMRHPLTGGGMTVALSDIVLLRNLLRPLENLNDASVLC KYLESFYILRKPMASTINTLAGALYKVFSASTDRARSEMRQACFDYLSLGGVFSNGPIALLSGL NPRPLNLVLHFFAVAVYGVGRLILPFPSPKSIWDGVKLISGASSVIFPIMKAEGIGQIFFPITK PPNHKSQTW	64	SEQ ID NO: 101 in WO2016050890
Squalene monooxygenase (*Ricinus communis*)	MGVSREENARDEKCHYYENGISLSEKSMSTDIIIVGAGVAGSALAYTLGKDGRRVHVIERDLSL QDRIVGELLQPGGYLKLIELGLEDCVEEIDAQQVFGYALYKNGRSTKLSYPLESFDSDVSGRSF HNGRFIQRMREKAASLPNVRLEEGTVTSLLEVKGTIKGVQYKTKNGEELTASAPLTIVCDGCFS NLRRSLCNPKVDIPSCFVALILENSGQKLPSISNGDMANYLKSVVAPQIPPVLSEAFISAIEKG KIRTMPNRSMPAAPHPTPGALLLGDAFNMRHPLTGGGMTVALSDIVVLRNLLKPLHDLTDASAL CEYLKSFYSLRKPVASTINTLAGALYKVFSASHDPARNEMRQACFDYLSLGGVFSNGPIALLSG LNPRPLSLVAHFFAVAIYGVGRLIFPLPSAKGMWMGARMIKVASGIIFPIIRAEGVQHMFFSKT LSAFSRSQTS	65	SEQ ID NO: 102 in WO2016050890
Squalene monooxygenase (*Ricinus communis*)	MEYQYFVGGIIASALLFVLVCRLAGKRQRRALRDTVDRDEISQNSENGISQSEKNMNTDIIIVG AGVAGSTLAYTLGKDGRRVRVIERDLSLQDRIVGELLQPGGYLKLIELGLEDCVEEIDALQVFG YALYKNGRSTKLSYPLDSFDSDVSGRSFHNGRFIQRMREKAASLPNVRMEGGTVTSLLEVKGTI KGVQYKNKNGEELIACAPLTIVCDGCFSNLRRSLCNSKVDIPFCFVALILENCELPYPNHGHVI LADPSPILFYRISISEIRCLVDIPAGQKLPSISNGEMANYLKSVVAPQIPPELSNAFLSAIEKG KIRTMPKRSMPAAPHPTPGALLLGDAFNMRHPLTGGVMTVALSDIVVLRSLLRPLHDLTDASAL CEYLKSFYSLRKPMVSTINTLAGALYRVFSASQDPARDEMRQACFDYLSLGGVFSNGPIALLSG LNPRPLSLIVHFFAVAVYGVGRLIFPLPSAKRMWMQE	66	SEQ ID NO: 103 in WO2016050890
Squalene monooxygenase (*Ricinus communis*)	MEYQYLMGGGIMTLLFVLSYRLKRETRASVENARDEVLQNSENGISQSEKAMNTDIKLLLEQIV QKIAMLNSIRLEEGTVTSLLEVKRDIKGVQYKTKNGEELTACAPLTIVSHGCFSNLRLHVTPST SKFKSFIGLEVDIPSSFAALILGNCELPFPNHGHVILADPSSILFYRISSSEICCLVDVPAGQK LPSISNGEMANYLKSVVAHQAFKVGLAY	67	SEQ ID NO: 104 in WO2016050890
Squalene	MSPISIQLPPRPQLYRSLISSLSLSTYKQPPSPPSFSLTIANSPPQPQPQATVSSKTRTITRLS	68	SEQ ID NO:

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
monooxygenase (*Ricinus communis*)	NSSNRVNLLQAEQHPQEPSSDLSYSSSPPHCVSGGYNIKLMEVGTDNYAVIIILGTFFASLFAF VFLSILRYNFKNKNKAKIHDETTLKTQNDNVRLPDNGSGNDVIIVGAGVAGAALAYTLGKDGRR VHVIERDLTEPDRIVGELLQPGGYLKLIELGLEDCVQEIDAQRVLGYALFKDGKNTRLSYPLEK FHADVAGRSFHNGRFIQRMREKAASLPNVKLEQGTVTSLLEENGTIKGVQYKTKDGQEIRAYAP LTIVCDGCFSNLRRSLCNPKVDVPSCFVGLVLENCQLPFANHGHVVLADPSPILFYPISSTEVR CLVDVPGQKVPSIANGEMAKYLKNVVAPQIPPVLHDAFISAIDKGNIRTMPNRSMPADPHPTPG ALLMGDAFNMRHPLTGGGMTVALSDIVVLRDLLKPLRDLNDATSLTKYLESFYTLRKPVASTIN TLAGALYKVFSASPDQARKEMRQACFDYLSLGGIFSSGPVALLSGLNPRPLSLVMHFFAVAIYG VGRLLLPFPSPKSVWIGARLISSASGIIFPIIKAEGVRQMFFPATIPAIYRPPPVKDTSDDEQK SR		105 in WO2016050890
ERG9 protein (*S. cerevisiae*)	MGKLLQLALHPVEMKAALKLKFCRTPLFSIYDQSTSPYLLHCFELLNLTSRSFAAVIRELHPEL RNCVTLFYLILRALDTIEDDMSIEHDLKIDLLRHFHEKLLLTKWSFDGNAPDVKDRAVLTDFES ILIEFHKLKPEYQEVIKEITERMGNGMADYILDENYNLNGLQTVHDYDVYCHYVAGLVGDGLTR LIVIAKFANESLYSNEQLYESMGLFLQKTNIIRDYNEDLVDGRSFWPKEIWSQYAPQLKDFMKP ENEQLGLDCINHLVLNALSHVIDVLTYLAGIHEQSTFQFCAIPQVMAIATLALVFNNREVLHGN VKIRKGTTCYLILKSRTLRGCVEIFDYYLRDIKSKLAVQDPNFLKLNIQISKIEQFMEEMYQDK LPPNVKPNETPIFLKVKERSRYDDELVPTQQEEEYKFNMVLSIILSVLLGFYYIYTLHRA	69	SEQ ID NO: 87 in WO2016050890
Cucurbitadienol synthase (*S. grosvenorii*)	MWRLKVGAESVGENDEKWLKSISNHLGRQVWEFCPDAGTQQQLLQVHKARKAFHDDRFHRKQSS DLFITIQYGKEVENGGKTAGVKLKEGEEVRKEAVESSLERALSFYSSIQTSDGNWASDLGGPMF LLPGLVIALYVTGVLNSVLSKHHRQEMCRYVYNHQNEDGGWGLHIEGPSTMFGSALNYVALRLL GEDANAGAMPKARAWILDHGGATGITSWGKLWLSVLGVYEWSGNNPLPPEFWLFPYFLPFHPGR MWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYAVPYHEIDWNKSRNTCAKEDLYYPHPKM QDILWGSLHHVYEPLFTRWPAKRLREKALQTAMQHIHYEDENTRYICLGPVNKVLNLLCCWVED PYSDAFKLHLQRVHDYLWVAEDGMKMQGYNGSQLWDTAFSIQAIVSTKLVDNYGPTLRKAHDFV KSSQIQQDCPGDPNVWYRHIHKGAWPFSTRDHGWLISDCTAEGLKAALMLSKLPSETVGESLER NRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTSATMEALTLF KKLHPGHRTKEIDTAIVRAANFLENMQRTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCLA IRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAERDPTPLH RAARLLINSQLENGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE	70	SEQ ID NO: 43 in WO2016050890
Cucurbitadienol synthase (UniProtKB-Q6BE24)	MWRLKVGAESVGEEDEKWVKSVSNHLGRQVWEFCADAAADTPHQLLQIQNARNHFHHNRFHRKQ SSDLFLAIQYEKEIAKGAKGGAVKVKEGEEVGKEAVKSTLERALGFYSAVQTRDGNWASDLGGP LFLLPGLVIALHVTGVLNSVLSKHHRVEMCRYLYNHQNEDGGWGLHIEGTSTMFGSALNYVALR LLGEDADGGDGGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLP FHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTIPYHEIDWNKSRNTCAKEDLYY PHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQAAMKHIHYEDENSRYICLGPVNKVLNMLC CWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTLRK AHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPLSTRDHGWLISDCTAEGLKASLMLSKLPSTMVG EPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTAATME	71	Disclosed in Takase *et al.* (Org Biomol Chem. 2015 Jul 14;13(26):7331-6) which is incorporated by reference

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGRTY NSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGERD PAPLHRAARLLMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE		in its entirety
Cucurbitadienol synthase (*C. pepo*)	MWRLKVGAESVGEEDEKWVKSVSNHLGRQVWEFCADAAADTPHQLLQIQNARNHFHHNRFHRKQ SSDLFLAIQYEKEIAKGAKGGAVKVKEGEEVGKEAVKSTLERALGFYSAVQTRDGNWASDLGGP LFLLPGLVIALHVTGVLNSVLSKHHRVEMCRYLYNHQNEDGGWGLHIEGTSTMFGSALNYVALR LLGEDADGGDGGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLP FHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTIPYHEIDWNKSRNTCAKEDLYY PHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQAAMKHIHYEDENSRYICLGPVNKVLNMLC CWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTLRK AHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPLSTRDHGWLISDCTAEGLKASLMLSKLPSTMVG EPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTAATME ALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGRTY NSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGERD PAPLHRAARLLMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE	72	SEQ ID NO: 1 of WO2014/086842 which is incorporated by reference in its entirety.
C-terminal portion of S. Grosvenorrii cucurbitadienol synthase	LERNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTSATMEAL TLFKKLHPGHRTKEIDTAIVRAANFLENMQRTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNN CLAIRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAERDPT PLHRAARLLINSQLENGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE	73	SEQ ID NO: 2 in WO2014/086842
Codon optimized cucurbitadienol synthase gene from *Siraitia grosvenorii*	ATGTGGAGATTGAAAGTAGGTGCTGAATCCGTAGGTGAAAACGACGAAAAGTGGTTGAAAAGTA TAAGTAATCATTTGGGTAGACAAGTCTGGGAATTTTGTCCAGATGCAGGTACACAACAACAATT GTTGCAAGTACATAAGGCTAGAAAGGCATTTCATGATGACAGATTCCACAGAAAGCAATCTTCA GATTTGTTCATCACCATCCAATACGGCAAGGAAGTAGAAAACGGTGGCAAGACTGCTGGTGTTA AATTGAAGGAAGGTGAAGAAGTTAGAAAAGAAGCAGTTGAATCCAGTTTGGAAAGAGCCTTGTC TTTCTACTCTTCAATCCAAACCTCTGATGGTAATTGGGCATCAGACTTGGGTGGTCCAATGTTC TTGTTACCTGGTTTGGTCATTGCCTTGTACGTAACTGGTGTTTTGAACTCTGTATTGTCAAAGC ATCACAGACAAGAAATGTGTAGATACGTTTACAACCATCAAAACGAAGATGGTGGTTGGGGTTT GCACATTGAAGGTCCATCCACTATGTTTGGTAGTGCATTGAATTATGTCGCCTTAAGATTGTTA GGTGAAGATGCAAACGCCGGTGCTATGCCTAAGGCAAGAGCCTGGATATTAGACCATGGTGGTG CTACTGGTATCACATCCTGGGGTAAATTGTGGTTAAGTGTCTTAGGTGTATATGAATGGTCTGG TAATAACCCATTGCCACCTGAATTTTGGTTGTTCCCTTACTTTTTACCATTCCATCCTGGTAGA ATGTGGTGTCACTGCAGAATGGTTTACTTGCCAATGTCTTACTTGTACGGCAAGAGATTCGTTG GTCCAATAACACCTATCGTCTTGTCATTGAGAAAGGAATTGTACGCAGTTCCTTACCATGAAAT CGATTGGAACAAGTCCAGAAACACCTGTGCTAAGGAAGATTTGTATTACCCACACCCTAAAATG CAAGACATTTTGTGGGGTAGTTTACATCACGTTTACGAACCATTATTTACTAGATGGCCTGCTA AAAGATTGAGAGAAAAGGCATTACAAACAGCCATGCAACATATCCACTACGAAGATGAAAACAC CAGATACATCTGCTTGGGTCCAGTTAACAAGGTCTTGAACTTGTTGTGTTGCTGGGTTGAAGAT CCTTATTCTGACGCTTTCAAGTTGCATTTGCAAAGAGTACACGATTACTTGTGGGTTGCAGAAG	74	SEQ ID NO: 42 in WO2014/086842

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ACGGTATGAAAATGCAAGGTTACAATGGTTCACAATTGTGGGATACAGCTTTTTCCATTCAAGC AATAGTCAGTACTAAGTTGGTAGATAACTACGGTCCAACATTAAGAAAAGCTCATGACTTCGTA AAGTCCAGTCAAATACAACAAGATTGTCCAGGTGACCCTAATGTTTGGTATAGACATATCCACA AAGGTGCATGGCCATTTTCTACCAGAGATCATGGTTGGTTGATTTCAGACTGTACTGCTGAAGG TTTGAAGGCTGCATTGATGTTGTCTAAGTTGCCATCAGAAACTGTTGGTGAATCCTTGGAAAGA AATAGATTATGCGATGCCGTTAACGTCTTGTTGAGTTTGCAAAACGACAACGGTGGTTTCGCTT CTTACGAATTGACTAGATCATACCCATGGTTGGAATTAATTAATCCTGCTGAAACATTCGGTGA TATCGTCATTGACTATCCATACGTAGAATGTACCTCCGCTACTATGGAAGCATTGACCTTGTTC AAGAAGTTGCATCCTGGTCACAGAACAAAGGAAATCGATACCGCAATTGTTAGAGCCGCTAATT TCTTGGAAAACATGCAAAGAACAGACGGTTCTTGGTATGGTTGTTGGGGTGTTTGCTTTACCTA CGCTGGTTGGTTCGGTATTAAAGGTTTAGTCGCAGCCGGTAGAACATACAATAACTGTTTGGCC ATAAGAAAAGCTTGCGATTTCTTGTTATCTAAGGAATTACCAGGTGGTGGTTGGGGTGAATCCT ACTTGAGTTGTCAAAACAAGGTTTACACTAATTTGGAAGGCAACAGACCTCATTTAGTTAACAC AGCCTGGGTCTTGATGGCTTTAATCGAAGCCGGTCAAGCTGAAAGAGATCCAACTCCTTTGCAT AGAGCTGCAAGATTGTTGATCAACTCACAATTGGAAAACGGTGATTTTCCACAACAAGAAATCA TGGGTGTTTTCAACAAGAACTGCATGATAACATATGCCGCTTACAGAAACATTTTTCCTATATG GGCTTTGGGTGAATACTGCCACAGAGTCTTGACCGAATAA		
Cycloartenol synthase [*Lotus japonicus*] GenBank Accession No. BAE53431.1	MWKLKIAEGGNPWLRSTNSHVGRQVWEFDPKLGSPQDLAEIETARNNFHDNRFSHKHSSDLLMR IQFSKENPIGEVLPKVKVKDVEDVTEEAVVTTLRRAISFHSTLQSHDGHWPGDYGGPMFLMPDL VITLSITGALNAVLTDEHRKEMCRYLYNHQNKDGGWGLHIEGPSTMFGSVLNYVTLRLLGEGPN DGQGDMEKARDWILGHGGATYITSWGKMWLSVLGVFEWSGNNPLPPEIWLLPYALPFHPGRMWC HCRMVYLPMSYLYGKRFVGPITPTILSLRKELFTIPYHDIDWNQARNLCAKEDLYYPHPLVQDI LWASLHKVVEPVLMQWPGKKLREKAINSVMEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNS EAFKLHLPRIYDYLWIAEDGMKMQGYNGSQLWDTAFAAQAIISTNLIEEYGPTLRKAHTFIKNS QVLEDCPGDLNKWYRHISKGAWPFSTADHGWPISDCTAEGLKAILSLSKIAPDIVGEPLDAKRL YDAVNVILSLQNEDGGLATYELTRSYSWLELINPAETFGDIVIDYPYVECTSAAIQALTSFRKL YPGHRREEIQHSIEKAAAFIEKIQSSDGSWYGSWGVCFTYGTWFGVKGLIAAGKSFSNCSSIRK ACEFLLSKQLPSGGWGESYLSCQNKVYSNLEGNRPHAVNTGWAMLALIEAEQAKRDPTPLHRAA LYLINSQMENGDFPQQEIMGVFNKNCMITYAAYRSIFPIWALGEYRCRVLQAR	75	Disclosed in WO2014/086842
Hypothetical protein POPTR_0007s15200g [*Populus trichocarpa*]	MWKLTIGAESVHDNGQSSSWLKSVNNHLGRQVWEFCPQLGSPDELLQLQNVRLSFQAQRFDKKH SADLLMRFQFEKENPCVNLPQIKVKDDEDVTEEAVTTTLRRAVNFYRKIQAHDGHWPGDYGGPM FLLPGLIITLSITGALNAVLSKEHQREMCRYLYNHQNRDGGWGLHIEGPSTMFGTCLNYVTLRL LGEGAEGGDGEMEKGRKWILDHGGATEITSWGKMWLSVLGVHEWSGNNPLPPEVWLCPYLLPMH PGRMWCHCRMVYLPMSYLYGKRFVGPITPTIQSLRKEIYTVPYHEVDWNTARNTCAKEDLYYPH PLVQDILWASLHYAYEPILTRWPLNRLREKALHKVMQHIHYEDENTQYICIGPVNKVLNMLCCW VEDPHSEAFKLHLPRVFDYLWIAEDGMKMQGYNGSQLWDTAFAVQAIVSTNLAEEYSGTLRKAH KYLKDSQVLEDCPGDLNFWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKLPTEMVGDP LGVERLRDAVNVILSLQNADGGFATYELTRSYQWLELINPAETFGDIVIDYPYVECTSAAIQAL	76	Disclosed in WO2014/086842

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ASFKKLYPGHRREEIDNCIAEAANFIEKIQATDGSWYGSWGVCFTYAGWFGIKGLVAAGMTYNS SSSIRKACDYMLSKELAGGGWGESYLSCQNKVYTNLKDDRPHIVNTGWAMLALIEAGQAERDPI PLHRAARVLINSQMENGDFPQEEIMGVFNKNCMISY SAYRNIFPIWALGEYRCQVLQAL		
putative 2,3 oxidosqualene cyclase [*Actaea racemosa*]	MWKLKIAEGEDPWLRSVNNHVGRQVWEFDRNLGTPEELIEVEKAREDFSNHKFEKKHSSDLLMR LQLAKENPCSIDLPRVQVKDTEEVTEEAVTTTLRRGLSFYSTIQGHDGHWPGDYGGPLFLMPGL VIALSVTGALNAVLSSEHQRETRRYIYNHQNEDGGWGLHIEGSSTMFITTLNYVTLRLLGEGAD DGEGAMEKARKWILNHGSATATTSWGKMWLSVLGVFEWSGNNPLPPEMWLLPYCLPFHPGRMWC HCRMVYLPMSYLYGKRFVGPITPTIESLRKELYSVPYHEIDWNQARNLCAKEDLYYPHPLVQDI LWTSLHYGVEPILTRWPANKLREKSLLTTMQHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNS EAFKLHIPRIYDYLWVAEDGMKMQGYNGSQLWDTAFAVQAIISTNLFEDYAPTLRKAHKYIKDS QVLDDCPGDLNFWYRHISKGAWPFSTADHGWPISDCTAEGLKAALLLSKIPSKSVGDPINAKQL YDAVNVILSLQNGDGGFATYELTRSYPWLELINPAETFGDIVIDYPYVECTAAAIQALTSFKKL YPGHRREDIENCVEKAVKFLKEIQAPDGSWYGSWGVCFTYGIWFGIKGLVAAGETFTNSSSIRK ACDFLLSKELDSGGWGESYLSCQNKVYTNLKGNRPHLVNTGWAMSALIDAGQAERDPKPLHRAA RVLINSQMDNGDFPQEEIMGVFNRNCMISYSAYRNIFPIWALGEYRCRVLKAP	77	Disclosed in WO2014/086842
CGTase AAA22298.1	MKSRYKRLTSLALSLSMALGISLPAWASPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNNPAGD AFSGDRSNLKLYFGGDWQGIIDKINDGYLTGMGVTALWISQPVENITSVIKYSGVNNTSYHGYW ARDFKQTNDAFGDFADFQNLIDTLTLITSRSDRLRPQPHVSGRAGTNPGFAENGALYDNGSLLG AYSNDTAGLFHHNGGTDFSTIEDGIYKNLYDLADINHNNNAMDAYFKSAIDLWLGMGVDGIRFD AVKQYPFGWQKSFVSSIYGGDHPVFTFGEWYLGADQTDGDNIKFANESGMNLLDFEYAQEVREV FRDKTETMKDLYEVLASTESQYDYINNMVTFIDNHDMDRFQVAGSGTRATEQALALTLTSRGVP AIYYGTEQYMTGDGDPNNRAMMTSFNTGTTAYKVIQALAPLRKSNPAIAYGTTTERWVNNDVLI IERKFGSSAALVAINRNSSAAYPISGLLSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNFTLAA GGTAVWQYTAPETSPAIGNVGPTMGQPGNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSWEDTQ IKAVIPKVAAGKTGVSVKTSSGTASNTFKSFNVLTGDQVTVRFLVNQANTNYGTNVYLVGNAAE LGTWDPNKAIGPMYNQVIAKYPSWYYDVSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTTPASG VGTVTVDWQN	78	
CGTase 3WMS_A	MKYLLPTAAAGLLLLAAQPAMAMDIGINSDPSPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNN PAGDAFSGDRSNLKLYFGGDWQGIIDKINDGYLTGMGVTALWISQPVENITSVIKYSGVNNTSY HGYWARDFKQTNDAFGDFADFQNLIDTAHAHNIKVVIDFAPNHTSPADRDNPGFAENGALYDNG SLLGAYSNDTAGLFHHNGGTDFSTIEDGIYKNLIDLADINHNNNAMDAYFKSAIDLWLGMGVDG IRFDAVKHMPFGWQKSFVSSIYGGDHPVFTFGEWYLGADQTDGDNIKFANESGMNLLDFEYAQE VREVFRDKTETMKDLYEVLASTESQYDYINNMVTFIDNHDMDRFQVAGSGTRATEQALALTLTS RGVPAIYYGTEQYMTGDGDPNNRAMMTSFNTGTTAYKVIQALAPLRKSNPAIAYGTTTERWVNN DVLIIERKFGSSAALVAINRNSSAAYPISGLLSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNF TLAAGGTAVWQYTAPETSPAIGNVGPTMGQPGNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSW EDTQIKAVIPKVAAGKTGVSVKTSSGTASNTFKSFNVLTGDQVTMRFLVNQANTNYGTNVYLVG	79	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	NAAELGSWDPNKAIGPMYNQVIAKYPSWYYDVSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTT PASSVGTVTVDWQNLE		
CGTase 4JCL_A	SPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNNPAGDAFSGDRSNLKLYFGGDWQGIIDKINDG YLTGMGVTALWISQPVENITSVIKYSGVNNTSYHGYWARDFKQTNDAFGDFADFQNLIDTAHAH NIKVVIDFAPNHTSPADRDNPGFAENGGMYDNGSLLGAYSNDTAGLFHHNGGTDFSTIEDGIYK NLYDLADINHNNNAMDAYFKSAIDLWLGMGVDGIRFDAVKHMPFGWQKSFVSSIYGGDHPVFTF GEWYLGADQTDGDNIKFANESGMNLLDFEYAQEVREVFRDKTETMKDLYEVLASTESQYDYINN MVTFIDNHDMDRFQVAGSGTRATEQALALTLTSRGVPAIYYGTEQYMTGDGDPNNRAMMTSFNT GTTAYKVIQALAPLRKSNPAIAYGTTTERWVNNDVLIIERKFGSSAALVAINRNSSAAYPISGL LSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNFTLAAGGTAVWQYTAPETSPAIGNVGPTMGQP GNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSWEDTQIKAVIPKVAAGKTGVSVKTSSGTASNT FKSFNVLTGDQVTVRFLVNQANTNYGTNVYLVGNAAELGSWDPNKAIGPMYNQVIAKYPSWYYD VSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTTPASGVGTVTVDWQN	80	
CGTase WP_036618292.1	MKSRYKRLTSLALSLSMALGISLPAWASPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNNPAGD AFSGDRSNLKLYFGGDWQGIIDKINDGYLTGMGVTALWISQPVENITSVIKYSGVNNTSYHGYW ARDFKQTNDAFGDFADFQNLIDTAHAHNIKVVIDFAPNHTSPADRDNPGFAENGALYDNGSLLG AYSNDTAGLFHHNGGTDFSTIEDGIYKNLYDLADINHNNNAMDAYFKSAIDLWLGMGVDGIRFD AVKHMPFGWQKSFVSSIYGGDHPVFTFGEWYLGADQTDGDNIKFANESGMNLLDFEYAQEVREV FRDKTETMKDLYEVLASTESQYDYINNMVTFIDNHDMDRFQVAGSGTRATEQALALTLTSRGVP AIYYGTEQYMTGDGDPNNRAMMTSFNTGTTAYKVIQALAPLRKSNPAIAYGTTTERWVNNDVLI IERKFGSSAALVAINRNSSAAYPISGLLSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNFTLAA GGTAVWQYTAPETSPAIGNVGPTMGQPGNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSWEDTQ IKAVIPKVAAGKTGVSVKTSSGTASNTFKSFNVLTGDQVTVRFLVNQANTNYGTNVYLVGNAAE LGSWDPNKAIGPMYNQVIAKYPSWYYDVSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTTPASG VGTVTVDWQN	81	
CGTase P04830.2	MKSRYKRLTSLALSLSMALGISLPAWASPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNNPAGD AFSGDRSNLKLYFGGDWQGIIDKINDGYLTGMGVTALWISQPVENITSVIKYSGVNNTSYHGYW ARDFKQTNDAFGDFADFQNLIDTAHAHNIKVVIDFAPNHTSPADRDNPGFAENGGMYDNGSLLG AYSNDTAGLFHHNGGTDFSTIEDGIYKNLYDLADINHNNNAMDAYFKSAIDLWLGMGVDGIRFD AVKHMPFGWQKSFVSSIYGGDHPVFTFGEWYLGADQTDGDNIKFANESGMNLLDFEYAQEVREV FRDKTETMKDLYEVLASTESQYDYINNMVTFIDNHDMDRFQVAGSGTRATEQALALTLTSRGVP AIYYGTEQYMTGDGDPNNRAMMTSFNTGTTAYKVIQALAPLRKSNPAIAYGTTTERWVNNDVLI IERKFGSSAALVAINRNSSAAYPISGLLSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNFTLAA GGTAVWQYTAPETSPAIGNVGPTMGQPGNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSWEDTQ IKAVIPKVAAGKTGVSVKTSSGTASNTFKSFNVLTGDQVTVRFLVNQANTNYGTNVYLVGNAAE LGSWDPNKAIGPMYNQVIAKYPSWYYDVSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTTPASG VGTVTVDWQN	82	
CGTase	MKSRYKRLTSLALSLSMALGISLPAWASPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNNPAGD	83	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
AAC04359.1	AFSGDRSNLKLYFGGDWQGIIDKINDGYLTGMGVTALWISQPVENITSVIKYSGVNNTSYHGYW ARDFKQTNDAFGDFADFQNLIDTAHAHNIKVVIDFAPNHTSPADRDNPGFAENGALYDNGSLLG AYSNDTAGLFHHNGGTDFSTIEDGIYKNLYDLADINHNNNAMDAYFKSAIDLWLGMGVDGIRFD AVKHMPFGWQKSFVSSIYGGDHPVFTFGEWYLGADQTDGDNIKFANESGMNLLDFEYAQEVREV FRDKTETMKDLYEVLASTESQYDYINNMVTFIDNHDMDRFQVAGSGTRATEQALALTLTSRGVP AIYYGTEQYMTGDGDPNNRAMMTSFNTGTTAYKVIQALAPLRKSNPAIAYGTTTERWVNNDVLI IERKFGSSAALVAINRNSSAAYPISGLLSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNFTLAA GGTAVWQYTAPETSPAIGNVGPTMGQPGNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSWEDTQ IKAVIPKVAAGKTGVSVKTSSGTASNTFKSFNVLTGDQVTVRFLVNQANTNYGTNVYLVGNAAE LGSWDPNKAIGPMYNQVIAKYPSWYYDVSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTTPASG VGTVTVDWQN		
CGTase CAA41773.1	MKSRYKRLTSLALSLSMALGISLPAWASPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNNPAGD AFSGDRSNLKLYFGGDWQGIIDKINDGYLTGMGVTALWISQPVENITSVIKYSGVNNTSYHGYW ARDFKQTNDAFGDFADFQNLIDTAHAHNIKVVIDFAPNHTSPADRDNPGFAENGGMYDNGSLLG AYSNDTAGLFHHNGGTDFSTIEDGIYKNLYDLADINHNNNAMDAYFKSAIDLWLGMGVDGIRFD AVKHMPFGWQKSFVSSIYGGDHPVFTFGEWYLGADQTDGDNIKFANESGMNLLDFEYAQEVREV FRDKTETMKDLYEVLASTESQYDYINNMVTFIDNHDMDRFQVAGSGTRATEQALALTLTSRGVP AIYYGTEQYMTGDGDPNNRAMMTSFNTGTTAYKVIQALAPLRKSNPAIAYGTTTERWVNNDVLI IERKFGSSAALVAINRNSSAAYPISGLLSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNFTLAA GGTAVWQYTAPETSPAIGNVGPTMGQPGNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSWEDTQ IKAVIPKVAAGKTGVSVKTSSGTASNTFKSFNVLTGDQVTVRFLVNQANTNYGTNVYLVGNAAE LGSWDPNKAIGPMYNQVIAKYPSWYYDVSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTTPASG VGTVTVDWQN	84	
CGTase AGT21379.1	SPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNNPAGDAFSGDRSNLKLYFGGDWQGIIDKINDG YLTGMGVTALWISQPVENITSVIKYSGVNNTSYHGYWARDFKQTNDAFGDFADFQNLIDTAHAH NIKVVIDFAPNHTSPADRDNPGFAENGALYDNGSLLGAYSNDTAGLFHHNGGTDFSTIEDGIYK NLYDLADINHNNNAMDAYFKSAIDLWLGMGVDGIRFDAVKHMPFGWQKSFVSSIYGGDHPVFTF GEWYLGADQTDGDNIKFANESGMNLLDFEYAQEVREVFRDKTETMKDLYEVLASTESQYDYINN MVTFIDNHDMDRFQVAGSGTRATEQALALTLTSRGVPAIYYGTEQYMTGDGDPNNRAMMTSFNT GTTAYKVIQALAPLRKSNPAIAYGTTTERWVNNDVLIIERKFGSSAALVAINRNSSAAYPISGL LSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNFTLAAGGTAVWQYTAPETSPAIGNVGPTMGQP GNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSWEDTQIKAVIPKVAAGKTGVSVKTSSGTASNT FKSFNVLTGDQVTMRFLVNQANTNYGTNVYLVGNAAELGSWDPNKAIGPMYNQVIAKYPSWYYD VSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTTPASSVGTVTVDWQN	85	
CGTase AGT95840.1	SPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNNPAGDAFSGDRSNLKLYFGGDWQGIIDKINDG YLTGMGVTALWISQPVENITSVIKYSGVNNTSYHGYWARDFKQTNDAFGDFADFQNLIDTAHAH NIKVVIDFAPNHTSPADRDNPGFAENGALYDNGSLPGAYSNDTAGLFHHNGGTDFSTIEDGIYK NLYDLADINHNNNAMDAYFKSAIDLWLGMGVDGIRFDAVKHMPFGWQKSFVSSIYGGDHPVFTF	86	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GEWYLGADQTDGDNIKFANESGMNLLDFEYAQEVREVFRDKTETMKDLYEVLASTESQYDYINN MVTFIDNHDMDRFQVAGSGTRATEQALALTLTSRGVPAIYYGTEQYMTGDGDPNNRAMMTSFNT GTTAYKVIQALAPLRKSNPAIAYGTTTERWVNNDVLIIERKFGSSAALVAINRNSSAAYPISGL LSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNFTLAXGXTAVWQYTAPETSPAIGDVGPTMGQP GNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSWEDTQIKAVIPKVAAGKTGVSVKTSSGTASNT FKSFNVLTGDQVTVRFLVNQANTNYGTNVYLVGNAAELDSWDPNKAIGPMYNQVIAKYPSWYYD VSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTTPASGVGTVTADWQN		
CGTase P31835.1	MKKQVKWLTSVSMSVGIALGAALPVWASPDTSVNNKLNFSTDTVYQIVTDRFVDGNSANNPTGA AFSSDHSNLKLYFGGDWQGITNKINDGYLTGMGITALWISQPVENITAVINYSGVNNTAYHGYW PRDFKKTNAAFGSFTDFSNLIAAAHSHNIKVVMDFAPNHTNPASSTDPSFAENGALYNNGTLLG KYSNDTAGLFHHNGGTDFSTTESGIYKNLYDLADINQNNNTIDSYLKESIQLWLNLGVDGIRFD AVKHMPQGWQKSYVSSIYSSANPVFTFGEWFLGPDEMTQDNINFANQSGMHLLDFAFAQEIREV FRDKSETMTDLNSVISSTGSSYNYINNMVTFIDNHDMDRFQQAGASTRPTEQALAVTLTSRGVP AIYYGTEQYMTGNGDPNNRGMMTGFDTNKTAYKVIKALAPLRKSNPALAYGSTTQRWVNSDVYV YERKFGSNVALVAVNRSSTTAYPISGALTALPNGTYTDVLGGLLNGNSITVNGGTVSNFTLAAG GTAVWQYTTTESSPIIGNVGPTMGKPGNTITIDGRGFGTTKNKVTFGTTAVTGANIVSWEDTEI KVKVPNVAAGNTAVTVTNAAGTTSAAFNNFNVLTADQVTVRFKVNNATTALGQNVYLTGNVAEL GNWTAANAIGPMYNQVEASYPTWYFDVSVPANTALQFKFIKVNGSTVTWEGGNNHTFTSPSSGV ATVTVDWQN	87	
CGTase KFM94552.1	MKSRYKRLTSLALSLSMALGISLPAWASPDTSVDNKVNFSTDVIYQIVTDRFADGDRTNNPAGD AFSGDRSNLKLYFGGDWQGIIDKINDGYLTGMGVTALWISQPVENITSVIKYSGVNNTSYHGYW ARDFKQTNDAFGDFADFQNLIDTAHAHNIKVVIDFAPNHTSPADRDNPGFAENGALYDNGSLLG AYSNDTAGLFHHNGGTDFSTIEDGIYKNLYDLADINHNNNAMDAYFKSAIDLWLGMGVDGIRFD AVKHMPFGWQKSFVSSIYGGDHPVFTFGEWYLGADQTDGDNIKFANESGMNLLDFEYAQEVREV FRDKTETMKDLYEVLASTESQYDYINNMVTFIDNHDMDRFQVAGSGTRATEQALALTLTSRGVP AIYYGTEQYMTGDGDPNNRAMMTSFNTGTTAYKVIQALAPLRKSNPAIAYGTTTERWVNNDVLI IERKFGSSAALVAINRNSSAAYPISGLLSSLPAGTYSDVLNGLLNGNSITVGSGGAVTNFTLAA GGTAVWQYTAPETSPAIGNVGPTMGQPGNIVTIDGRGFGGTAGTVYFGTTAVTGSGIVSWEDTQ IKAVIPKVAAGKTGVSVKTSSGTASNTFKSFNVLTGDQVTVRFLVNQANTNYGTNVYLVGNAAE LGSWDPNKAIGPMYNQVIAKYPSWYYDVSVPAGTKLDFKFIKKGGGTVTWEGGGNHTYTTPASG VGTVTVDWQN	88	
Toruzyme AJE25826.1	MKKTLKLLSILLITIALLFSTIPSVPAAPDTSVSNVVNYSTDVIYQIVTDRFLDGNPSNNPTGD LYDPTHTSLKKYFGGDWQGIINKINDGYLTGMGITAIWISQPVENIYAVLPDSTFGGSTSYHGY WARDFKKTNPFFGSFTDFQNLIATAHAHNIKVIIDFAPNHTSPASETDPTYGENGRLYDNGELL GGYTNDTNGYFHHYGGTNFSSYEDGIYRNLFDLADLDQQNNTIDSYLKAAIKLWLDMGIDGIRM DAVKHMAFGWQKNFMDSILSYRPVFTFGEWYLGTNEVDPNNTYFANESGMNLLDFRFAQKVRQV FRDNTDTMYGLDSMIQSTAADYNFINDMVTFIDNHDMDRFYTGGSTRPVEQALAFTLTSRGVPA IYYGTEQYMTGNGDPYNRAMMTSFNTNTTAYNVIKKLAPLRKSNPAIAYGTQKQRWVNNDVYIY	89	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ERQFGNNVALIAINRNLSTSYNITGLYTALPAGTYSDVLGGLLNGNSITVSSNGSVTSFTLAPG AVAVWQYVSTTNPPLIGHVGPTMTKAGQTITIDGRGFGTTAGQVLFGTTPATIVSWEDTEVKVK VPALTPGKYNITLKTASEVTSNSYNNINVLTGNQVCVRFVVNNATTVWGENVYLTGNVAELGNW DTSKAIGPMFNQVVYQYPTWYYDVSVPAGTTIEFKFIKKNGSTVTWEGGYNHVYTTPTSGTATV IVNWQN		
Toruzyme KHO62967.1	MRKNVKLFAAIILFFSLLLTSCGSKDTSSNITPKSDVIYQVMIDRFYNGDKSNDDPKISKGMFD PTYTNWRMYWGGDLKGLTEKIPYIKGMGVTAIWISPVVDNINKPAIYNGEINAPYHGYWARDFK RVEEHFGSWEDFDNFVKTAHANGIKVILDFAPNHTSPADKNNPDFAENGALYDDGNLLGTYSND VNKLFHHNGGITNWNNLKDLQDKNLFDLADLDQSNPIVDKYLKDSIKLWFSHGIDGVRLDAVKH MPMEWVKSFADTIYGVNKDAILFGEWMLNGPTDPLYGYNIQFANTSGFSVLDFMLNSAIKDVFE KGYGFDRLNDTIEETNKDYDNPYKLVTFVDNHDMPRFLSVNDDKDKLHEAIAFIMTSRGIPAIY YGTEQYLHNDTNGGNDPYNRPMMEKFDENTTAYVLIRELSNLRKATQALQYGKTVSRYVSNDVY IYERQYGKDIVVVAINKGEETTVKNIETSLRKGKYSDYLKGLLKGGNLKVERGNSENDILSITL PKDSVSIWTNVKVK	90	
Toruzyme KHO61869.1	MKKTLKLLSILLITIALLFSSIPSVPAAPDTSVSNVVNYSTDVIYQIVTDRFLDGNPNNNPTGD LYDPTHTSLKKYFGGDWQGIINKINDGYLTGMGITAIWISQPVENIYAVLPDSTFGGSTSYHGY WARDFKKTNPFFGSFTDFQNLIATAHAHNIKVIIDFAPNHTSPASETDPTYGENGRLYDNGVLL GGYTNDTNGYFHHYGGTNFSSYEDGIYRNLFDLADLDQQNNTIDSYLKAAIKLWLDMGIDGIRM DAVKHMAFGWQKNFMDSILSYRPVFTFGEWYLGTNEVDPNNTYFANESGMSLLDFRFAQKVRQV FRDNTDTMYGLDSMLQSTAADYNFINDMVTFIDNHDMDRFYTGGSTRPVEQALAFTLTSRGVPA IYYGTEQYMTGNGDPYNRAMMTSFDTTTTAYNVIKKLAPLRKSNPAIAYGTQKQRWINNDVYIY ERQFGNNVALVAINRNLSTSYYITGLYTALPAGTYSDVLGGLLNGNNISVASDGSVTPFTLAPG EVAVWQYVSTTNPPLIGHVGPTMTKAGQTITIDGRGFGTTAGQVLFGTTPATIVSWEDTEVKVK VPALTPGKYNVTLKTASGVTSNSYNNINVLTGNQVCVRFVVNNASTVWGENVYLTGNVAELGSW DTSKAIGPMFNQVVYQYPTWYYDVSVPAGTTIEFKFIKKNGSTVTWEGGYNHVYTTPTSGTATV IVNWQN	91	
Toruzyme KHO61665.1	MKKTLKLLSILLITIALLFSSIPSVPAAPDTSVSNVVNYSTDVIYQIVTDRFLDGNPSNNPTGD LYDPTHTSLKKYFGGDWQGIINKINDGYLTGMGITAIWISQPVENIYAVLPDSTFGGSTSYHGY WARDFKKTNPFFGSFTDFQNLIATAHAHNIKVIIDFAPNHTSPASETDPTYGENGRLYDNGVLL GGYTNDTNGYFHHYGGTNFSSYEDGIYRNLFDLADLDQQNNTIDSYLKVAIKLWLNMGIDGIRM DAVKHMAFGWQKNFMDSILSYRPVFTFGEWYLGTNEVDPNNTYFANESGMSLLDFRFAQKVRQV FRDNTDTMYGLDSMLQSTAADYNFINDMVTFIDNHDMDRFYTGGSTRPVEQALAFTLTSRGVPA IYYGTEQYMTGNGDPYNRAMMTSFDTTTTAYNVIKKLAPLRKSNPAIAYGTQKQRWINNDVYIY ERQFGNNVALVAINRNLSTSYYITGLYTALPAGTYSDVLGGLLNGNNISVASDGSVTPFTLAPG EVAVWQYVSTTNPPLIGHVGPTMTKAGQTITIDGRGFGTTAGQVLFGTTPATIVSWEDTEVKVK VPALTPGKYNVTLKTASGVTSNSYNNINVLTGNQVCVRFVVNNASTVWGENVYLTGNVAELGSW DTSKAIGPMFNQVVYQYPTWYYDVSVPAGTTIEFKFIKKNGSTVTWEGGYNHVYTTPTSGTATV IVNWQN	92	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Toruzyme WP_042834654.1	MKKTLKLLSILLITIALLFSSIPSVPAAPDTSVSNVVNYSTDVIYQIVTDRFLDGNPNNNPTGD LYDPTHTSLKKYFGGDWQGIINKINDGYLTGMGITAIWISQPVENIYAVLPDSTFGGSTSYHGY WARDFKKTNPFFGSFTDFQNLIATAHAHNIKVIIDFAPNHTSPASETDPTYGENGRLYDNGVLL GGYTNDTNGYFHHYGGTNFSSYEDGIYRNLFDLADLDQQNNTIDSYLKAAIKLWLDMGIDGIRM DAVKHMAFGWQKNFMDSILSYRPVFTFGEWYLGTNEVDPNNTYFANESGMSLLDFRFAQKVRQV FRDNTDTMYGLDSMLQSTAADYNFINDMVTFIDNHDMDRFYTGGSTRPVEQALAFTLTSRGVPA IYYGTEQYMTGNGDPYNRAMMTSFDTTTTAYNVIKKLAPLRKSNPAIAYGTQKQRWINNDVYIY ERQFGNNVALVAINRNLSTSYYITGLYTALPAGTYSDVLGGLLNGNNISVASDGSVTPFTLAPG EVAVWQYVSTTNPPLIGHVGPTMTKAGQTITIDGRGFGTTAGQVLFGTTPATIVSWEDTEVKVK VPALTPGKYNVTLKTASGVTSNSYNNINVLTGNQVCVRFVVNNASTVWGENVYLTGNVAELGSW DTSKAIGPMFNQVVYQYPTWYYDVSVPAGTTIEFKFIKKNGSTVTWEGGYNHVYTTPTSGTATV IVNWQN	93	
Toruzyme WP_042834464.1	MKKTLKLLSILLITIALLFSSIPSVPAAPDTSVSNVVNYSTDVIYQIVTDRFLDGNPSNNPTGD LYDPTHTSLKKYFGGDWQGIINKINDGYLTGMGITAIWISQPVENIYAVLPDSTFGGSTSYHGY WARDFKKTNPFFGSFTDFQNLIATAHAHNIKVIIDFAPNHTSPASETDPTYGENGRLYDNGVLL GGYTNDTNGYFHHYGGTNFSSYEDGIYRNLFDLADLDQQNNTIDSYLKVAIKLWLNMGIDGIRM DAVKHMAFGWQKNFMDSILSYRPVFTFGEWYLGTNEVDPNNTYFANESGMSLLDFRFAQKVRQV FRDNTDTMYGLDSMLQSTAADYNFINDMVTFIDNHDMDRFYTGGSTRPVEQALAFTLTSRGVPA IYYGTEQYMTGNGDPYNRAMMTSFDTTTTAYNVIKKLAPLRKSNPAIAYGTQKQRWINNDVYIY ERQFGNNVALVAINRNLSTSYYITGLYTALPAGTYSDVLGGLLNGNNISVASDGSVTPFTLAPG EVAVWQYVSTTNPPLIGHVGPTMTKAGQTITIDGRGFGTTAGQVLFGTTPATIVSWEDTEVKVK VPALTPGKYNVTLKTASGVTSNSYNNINVLTGNQVCVRFVVNNASTVWGENVYLTGNVAELGSW DTSKAIGPMFNQVVYQYPTWYYDVSVPAGTTIEFKFIKKNGSTVTWEGGYNHVYTTPTSGTATV IVNWQN	94	
Cyclomaltodextrin glucanotransferase [*Geobacillus stearothermophilus*] GenBank: CAA41770.1	MRRWLSLVLSMSFVFSAIFIVSDTQKVTVEAAGNLNKVNFTSDVVYQIVVDRFVDGNTSNNPSG ALFSSGCTNLRKYCGGDWQGIINKINDGYLTDMGVTAIWISQPVENVFSVMNDASGSASYHGYW ARDFKKPNPFFGTLSDFQRLVDAAHAKGIKVIIDFAPNHTSPASETNPSYMENGRLYDNGTLLG GYTNDANMYFHHNGGTTFSSLEDGIYRNLFDLADLNHQNPVIDRYLKDAVKMWIDMGIDGIRMD AVKHMPFGWQKSLMDEIDNYRPVFTFGEWFLSENEVDANNHYFANESGMSLLDFRFGQKLRQVL RNNSDNWYGFNQMIQDTASAYDEVLDQVTFIDNHDMDRFMIDGGDPRKVDMALAVLLTSRGVPN IYYGTEQYMTGNGDPNNRKMMSSFNKNTRAYQVIQKLSSLRRNNPALAYGDTEQRWINGDVYVY ERQFGKDVVLVAVNRSSSSNYSITGLFTALPAGTYTDQLGGLLDGNTIQVGSNGSVNAFDLGPG EVGVWAYSATESTPIIGHVGPMMGQVGHQVTIDGEGFGTNTGTVKFGTTAANVVSWSNNQIVVA VPNVSPGKYNITVQSSSGQTSAAYDNFEVLTNDQVSVRFVVNNATTNLGQNIYIVGNVYELGNW DTSKAIGPMFNQVVYSYPTWYIDVSVPEGKTIEFKFIKKDSQGNVTWESGSNHVYTTPTNTTGK IIVDWQN	95	
cyclomaltodextrin glucanotransferase	MRRWLSLVLSMSFVFSAIFIVSDTQKVTVEAAGNLNKVNFTSDVVYQIVVDRFVDGNTSNNPSG ALFSSGCTNLRKYCGGDWQGIINKINDGYLTDMGVTAIWISQPVENVFSVMNDASGSASYHGYW	96	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
[*Geobacillus stearothermophilus*] GenBank: CAA41771.1	ARDFKKPNPFFGTLSDFQRLVDAAHAKGIKVIIDFAPNHTSPASETNPSYMENGRLYDNGTLLG GYTNDANMYFHHNGGTTFSSLEDGIYRNLFDLADLNHQNPVIDRYLKDAVKMWIDMGIDGIRMD AVKHMPFGWQKSLMDEIDNYRPVFTFGEWFLSENEVDANNHYFANESGMSLLDFRFGQKLRQVL RNNSDNWYGFNQMIQDTASAYDEVLDQVTFIDNHDMDRFMIDGGDPRKVDMALAVLLTSRGVPN IYYGTEQYMTGNGDPNNRKMMSSFNKNTRAYQVIQKLSSLRRNNPALAYGDTEQRWINGDVYVY ERQFGKDVVLVAVNRSSSSNYSITGLFTALPAGTYTDQLGGLLDGNTIQVGSNGSVNAFDLGPG EVGVWAYSATESTPIIGHVGPMMGQVGHQVTIDGEGFGTNTGTVKFGTTAANVVSWSNNQIVVA VPNVSPGKYNITVQSSSGQTSAAYDNFEVLTNDQVSVRFVVNNATTNLGQNIYIVGNVYELGNW DTSKAIGPMFNQVVYSYPTWYIDVSVPEGKTIEFKFIKKDSQGNVTWESGSNHVYTTPTNTTGK IIVDWQN		
cyclomaltodextrin glucanotransferase [*Geobacillus stearothermophilus*] GenBank: CAA41772.1	MRRWLSLVLSMSFVFSAIFIVSDTQKVTVEAAGNLNKVNFTSDVVYQIVVDRFVDGNTSNNPSG ALFSSGCTNLRKYCGGDWQGIINKINDGYLTDMGVTAIWISQPVENVFSVMNDASGSASYHGYW ARDFKKPNPFFGTLSDFQRLVDAAHAKGIKVIIDFAPNHTSPASETNPSYMENGRLYDNGTLLG GYTNDANMYFHHNGGTTFSSLEDGIYRNLFDLADLNHQNPVIDRYLKDAVKMWIDMGIDGIRMD AVKHMPFGWQKSLMDEIDNYRPVFTFGEWFLSENEVDANNHYFANESGMSLLDFRFGQKLRQVL RNNSDNWYGFNQMIQDTASAYDEVLDQVTFIDNHDMDRFMIDGGDPRKVDMALAVLLTSRGVPN IYYGTEQYMTGNGDPNNRKMMSSFNKNTRAYQVIQKLSSLRRNNPALAYGDTEQRWINGDVYVY ERQFGKDVVLVAVNRSSSSNYSITGLFTALPAGTYTDQLGGLLDGNTIQVGSNGSVNAFDLGPG EVGVWAYSATESTPIIGHVGPMMGQVGHQVTIDGEGFGTNTGTVKFGTTAANVVSWSNNQIVVA VPNVSPGKYNITVQSSSGQTSAAYDNFEVLTNDQVSVRFVVNNATTNLGQNIYIVGNVYELGNW DTSKAIGPMFNQVVYSYPTWYIDVSVPEGKTIEFKFIKKDSQGNVTWESGSNHVYTTPTNTTGK IIVDWQN	97	
Chain A, Cyclodextrin Glucanotransferase (E.C.2.4.1.19, CGTase) PDB: 1CYG_A	AGNLNKVNFTSDVVYQIVVDRFVDGNTSNNPSGALFSSGCTNLRKYCGGDWQGIINKINDGYLT DMGVTAIWISQPVENVFSVMNDASGSASYHGYWARDFKKPNPFFGTLSDFQRLVDAAHAKGIKV IIDFAPNHTSPASETNPSYMENGRLYDNGTLLGGYTNDANMYFHHNGGTTFSSLEDGIYRNLFD LADLNHQNPVIDRYLKDAVKMWIDMGIDGIRMDAVKHMPFGWQKSLMDEIDNYRPVFTFGEWFL SENEVDANNHYFANESGMSLLDFRFGQKLRQVLRNNSDNWYGFNQMIQDTASAYDEVLDQVTFI DNHDMDRFMIDGGDPRKVDMALAVLLTSRGVPNIYYGTEQYMTGNGDPNNRKMMSSFNKNTRAY QVIQKLSSLRRNNPALAYGDTEQRWINGDVYVYERQFGKDVVLVAVNRSSSSNYSITGLFTALP AGTYTDQLGGLLDGNTIQVGSNGSVNAFDLGPGEVGVWAYSATESTPIIGHVGPMMGQVGHQVT IDGEGFGTNTGTVKFGTTAANVVSWSNNQIVVAVPNVSPGKYNITVQSSSGQTSAAYDNFEVLT NDQVSVRFVVNNATTNLGQNIYIVGNVYELGNWDTSKAIGPMFNQVVYSYPTWYIDVSVPEGKT IEFKFIKKDSQGNVTWESGSNHVYTTPTNTTGKIIVDWQN	98	
Cyclomaltodextrin glucanotransferase (also known as Cyclodextrin-glycosyltransferas	MRRWLSLVLSMSFVFSAIFIVSDTQKVTVEAAGNLNKVNFTSDVVYQIVVDRFVDGNTSNNPSG ALFSSGCTNLRKYCGGDWQGIINKINDGYLTDMGVTAIWISQPVENVFSVMNDASGSASYHGYW ARDFKKPNPFFGTLSDFQRLVDAAHAKGIKVIIDFAPNHTSPASETNPSYMENGRLYDNGTLLG GYTNDANMYFHHNGGTTFSSLEDGIYRNLFDLADLNHQNPVIDRYLKDAVKMWIDMGIDGIRMD AVKHMPFGWQKSLMDEIDNYRPVFTFGEWFLSENEVDANNHYFANESGMSLLDFRFGQKLRQVL	99	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
e; CGTase) UniProtKB/Swiss-Prot: P31797.1	RNNSDNWYGFNQMIQDTASAYDEVLDQVTFIDNHDMDRFMIDGGDPRKVDMALAVLLTSRGVPN IYYGTEQYMTGNGDPNNRKMMSSFNKNTRAYQVIQKLSSLRRNNPALAYGDTEQRWINGDVYVY ERQFGKDVVLVAVNRSSSSNYSITGLFTALPAGTYTDQLGGLLDGNTIQVGSNGSVNAFDLGPG EVGVWAYSATESTPIIGHVGPMMGQVGHQVTIDGEGFGTNTGTVKFGTTAANVVSWSNNQIVVA VPNVSPGKYNITVQSSSGQTSAAYDNFEVLTNDQVSVRFVVNNATTNLGQNIYIVGNVYELGNW DTSKAIGPMFNQVVYSYPTWYIDVSVPEGKTIEFKFIKKDSQGNVTWESGSNHVYTTPTNTTGK IIVDWQN		
hypothetical protein AA906_05840 [*Geobacillus stearothermophilus*]	MSRNGAVTPDWQFTVEVQEGETITYKYVKGGSWDQEGLADHTREDDNDDDVSYYGYGAIGTDLK VTVHNEGNNTMIVQDRILRWIDMPVVIEEVQKQGSQVTIKGNAIKNGVLTINGERVPIDGRMAF SYTFTPASHQKEVSIHIEPSAESKTAIFNNDGGAIAKNTKDYVLNLETKQLREGKLTTPPSNGD SPESDWPGSETPSHDGGATPGNGTSPGSGGPSDGTSPGGSVPPGGTAPPGNEAPPSRPPQKPSP SKPKEKPRKPTTPPGQVKKVYWDGVELKKGQIGRLTVQKPINLWKRTKDGRLVFVRILQPGEVY RVYGYDVRFGGQYAVGGGYYVTDIDTHIRYETPSKEKLKLVNGE	100	
Maltodextrin glucosidase [*Geobacillus stearothermophilus*] GenBank: KYD32676.1	MSRNGAVTPDWQFTVEVQEGETITYKYVKGGSWDQEGLADHTREDDNDDDVSYYGYGAIGTDLK VTVHNEGNNTMIVQDRILRWIDMPVVIEEVQKQGSQVTIKGNAIKNGVLTINGERVPIDGRMAF SYTFTPASHQKEVSIHIEPSAESKTAIFNNDGGAIAKNTKDYVLNLETKQLREGKLTTPPSNGD SPESDWPGSETPSHDGGATPGNGTSPGSGGPSDGTSPGGSVPPGGTAPPGNGAPPSAPPQKPSP SKPKEKPRKPTTPPSQVKKVYWDGVELKKGQIGRLTVQKPINLWKRAKDGRLVFVRILQPGEVY RVYGYDARFGGQYAVGGGYYVTDIDTHIRYETPSKEKLKLVNGE	101	
Beta-glucosidase (Almonds)	MAMQLRSLLLCVLLLLLGFALADTNAAARIHPPVVCANLSRANFDTLVPGFVFGAATASYQVEG AANLDGRGPSIWDTFTHKHPEKIADGSNGDVAIDQYHRYKEDVAIMKDMGLESYRFSISWSRVL PNGTLSGGINKKGIEYYNNLINELLHNGIEPLVTLFHWDVPQTLEDEYGGFLSNRIVNDFEEYA ELCFKKFGDRVKHWTTLNEPYTFSSHGYAKGTHAPGRCSAWYNQTCFGGDSATEPYLVTHNLLL AHAAAVKLYKTKYQAYQKGVIGITVVTPWFEPASEAKEDIDAVFRALDFIYGWFMDPLTRGDYP QSMRSLVGERLPNFTKKESKSLSGSFDYIGINYYSARYASASKNYSGHPSYLNDVNVDVKTELN GVPIGPQAASSWLYFYPKGLYDLLCYTKEKYNDPIIYITENGVDEFNQPNPKLSLCQLLDDSNR IYYYYHHLCYLQAAIKEGVKVKGYFAWSLLDNFEWDNGYTVRFGINYVDYDNGLKRHSKHSTHW FKSFLKKSSRNTKKIRRCGNNNTSATKFVF	102	
DexT protein [*Leuconostoc citreum*]	MPANAPDKQSVTNAPVVPPKHDTDQQDDSLEKQQVLEPSVNSNIPKKQTNQQLAVVTAPANSAP QTKTTAEISAGTELDTMPNVKHVDGKVYFYGDDGQPKKNFTTIIDGKPYYFDKDTGALSNNDKQ YVSELFSIGNKHNAVYNTSSDNFTQLEGHLTASSWYRPKDILKNGKRWAPSTVTDFRPLLMAWW PDKSTQVTYLNYMKDQGLLSGTHHFSDNENMRTLTAAAMQAQVNIEKKIGQLGNTDWLKTAMTQ YIDAQPNWNIDSEAKGDDHLQGGALLYTNSDMSPKANSDYRKLSRTPKNQKGQIADKYKQGGFE LLLANDVDNSNPVVQAEQLNWLHYMMNIGSILQNDDQANFDGYRVDAVDNVDADLLQIAGEYAK AAYGVDKNDARANQHLSILEDWGDEDPDYVKAHGNQQITMDFPLHLAIKYALNMPNDKRSGLEP TREHSLVKRITDDKENVAQPNYSFIRAHDSEVQTIIADIIKDKINPASTGLDSTVTLDQIKQAF DIYNADELKADKVYTPYNIPASYALLLTNKDTIPRVYYGDMFTDDGQYMAKQSPYYQAIDALLK ARIKYAAGGQTMKMNYFPDEQSVMTSVRYGKGAMTASDSGNQETRYQGIGLVVNNRPDLKLSDK	103	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	DEVKMDMGAAHKNQDYRPVLLTTKSGLKVYSTDANAPVVRTDANGQLTFKADMVYGVNDPQVSG YIAAWVPVGASENQDARTKSETTQSTDGSVYHSNAALDSQVIYEGFSNFQDFPTTPDEFTNIKI AQNVNLFKDWGITSFEMAPQYRASSDKSFLDAIVQNGYAFTDRYDIGYNTPTKYGTADNLLDAL RALHGQGIQAINDWVPDQIYNLPDEQLVTAIRTDGSGDHTYGSVIDHTLYASKTVGGGIYQQQY GGAFLEQLKTQYPQLFQQKQISTDQPMNPDIQIKSWEAKYFNGSNIQGRGAWYVLKDWGTQQYF NVSDAQTFLPKQLLGEKAKTGFVTRGKETSFYSTSGYQAKSAFICDNGNWYYFDDKGKMVVGNQ VINGINYYFLPNGIELQDAYLVHDGMYYYYNNIGKQLHNTYYQDKQKNFHYFFEDGHMAQGIVT IIQSDGTPVTQYFDENGKQQKGVAVKGSDGHLHYFDGASGNMLFKSWGRLADGSWLYVDEKGNA VTGKQTINNQTVYFNDDGRQIKNNFKELADGSWLYLNNKGVAVTGEQIINGQTLYFGNDGRQFK GTTHINATGESRYYDPDSGNMITDRFERVGDNQWAYFGYDGVAVTGDRIIKGQKLYFNQNGIQM KGHLRLENGIMRYYDADTGELVRNRFVLLSDGSWVYFGQDGVPVTGVQVINGQTLYFDADGRQV KGQQRVIGNQRYWMDKDNGEMKKITYAAALEHHHHHH		
DexT gene sequence	ATGCCAGCAAATGCCCCAGATAAACAATCAGTGACTAATGCACCAGTAGTGCCGCCAAAGCATG ATACGGACCAGCAGGACGATTCACTAGAAAAACAGCAAGTATTAGAACCGAGCGTAAATAGTAA TATACCAAAAAAGCAGACAAATCAACAGTTAGCGGTTGTTACAGCACCAGCAAATTCAGCACCT CAAACCAAAACAACAGCAGAAATTTCTGCTGGTACAGAGTTAGACACGATGCCTAATGTTAAGC ATGTAGATGGCAAAGTTTATTTTTATGGAGATGATGGCCAACCAAAAAAGAATTTTACTACTAT TATAGATGGTAAACCTTACTACTTTGATAAAGATACAGGGGCACTATCTAATAACGATAAGCAA TATGTATCGGAATTATTCAGTATTGGCAATAAACATAACGCCGTCTATAACACATCATCAGATA ATTTTACGCAATTAGAAGGACATCTGACGGCAAGTAGTTGGTATCGTCCAAAAGATATTTTGAA AAATGGTAAACGTTGGGCACCTTCAACAGTGACTGATTTCAGACCATTATTGATGGCCTGGTGG CCGGATAAGAGTACGCAAGTCACTTATCTGAATTACATGAAAGATCAGGGCCTCTTGTCTGGTA CTCATCACTTTTCCGATAATGAAAATATGCGGACCTTAACGGCAGCTGCCATGCAGGCACAGGT AAACATTGAGAAAAAAATTGGGCAACTTGGCAATACGGATTGGTTGAAAACGGCGATGACGCAA TACATTGATGCCCAGCCCAATTGGAATATTGACAGTGAGGCGAAAGGAGATGATCATCTACAAG GTGGTGCACTACTTTATACAAATAGTGATATGTCGCCAAAGGCCAATTCTGATTATCGTAAGCT GAGCCGTACGCCTAAAAATCAAAAAGGTCAAATTGCTGATAAATATAAGCAAGGTGGGTTTGAA TTATTACTAGCAAACGATGTCGATAATTCTAATCCAGTTGTGCAAGCAGAACAACTTAATTGGT TACATTATATGATGAATATCGGTAGTATTTTACAAAATGATGACCAAGCTAATTTTGATGGTTA CCGTGTTGATGCTGTCGATAATGTGGACGCTGACTTACTACAGATTGCTGGTGAATATGCTAAG GCTGCCTATGGTGTTGACAAAAATGACGCGAGAGCGAATCAACATTTATCAATTTTGGAAGACT GGGGAGATGAAGATCCAGACTATGTCAAAGCACATGGCAACCAGCAAATTACAATGGATTTCCC CTTGCATTTAGCGATTAAATACGCGCTCAACATGCCTAATGATAAGCGGAGTGGCCTTGAGCCA ACCCGTGAACACAGTTTAGTCAAACGAATTACAGATGATAAAGAAAATGTTGCACAACCAAATT ATTCATTTATCCGAGCTCATGACAGTGAAGTACAAACGATTATTGCTGATATTATTAAAGATAA AATCAACCCGGCGTCAACAGGGCTAGATTCAACAGTGACTTTGGATCAAATTAAGCAGGCTTTT GACATCTATAATGCTGATGAATTGAAAGCAGATAAAGTTTACACACCTTACAATATTCCAGCAT CATACGCTTTGTTATTGACTAATAAAGACACAATTCCACGTGTTTATTATGGGGATATGTTCAC	104	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GGATGATGGCCAATACATGGCTAAACAATCACCTTACTATCAAGCGATTGATGCGTTGTTGAAA GCTCGTATCAAGTATGCTGCTGGTGGTCAAACCATGAAAATGAACTATTTTCCAGATGAACAAT CTGTTATGACATCAGTTCGTTATGGTAAGGGTGCAATGACGGCAAGTGACTCTGGTAACCAAGA GACACGCTATCAAGGTATTGGACTTGTTGTCAACAATCGCCCAGATTTGAAACTATCTGACAAA GATGAAGTCAAAATGGATATGGGTGCGGCACATAAAAACCAAGATTATCGCCCAGTTTTGTTGA CGACAAAATCAGGATTAAAAGTCTACAGCACTGATGCAAATGCACCTGTCGTTCGAACTGACGC CAATGGCCAATTAACTTTTAAGGCAGACATGGTATATGGTGTAAACGACCCACAAGTGTCAGGG TACATTGCGGCTTGGGTACCAGTAGGGGCTTCAGAAAATCAAGATGCTCGAACGAAAAGTGAAA CAACGCAGTCAACTGACGGGAGTGTTTATCATTCTAATGCAGCGTTAGATTCGCAAGTCATTTA TGAAGGCTTTTCAAATTTTCAAGACTTTCCAACAACACCCGATGAGTTTACGAACATTAAAATT GCTCAAAATGTTAACTTATTTAAGGATTGGGGTATTACTAGCTTTGAAATGGCGCCACAATATC GCGCCAGCTCAGATAAAAGTTTCTTAGATGCTATCGTACAAAATGGTTATGCATTTACAGATCG ATATGATATTGGTTACAACACACCAACAAAGTATGGGACAGCAGATAATTTGTTAGATGCTTTA CGTGCATTGCATGGTCAGGGTATTCAAGCGATTAACGACTGGGTACCAGATCAAATTTATAATC TACCCGATGAACAGTTAGTCACGGCTATTCGAACAGACGGTTCAGGTGATCATACTTATGGTTC AGTTATTGACCATACTTTGTATGCATCAAAGACAGTTGGCGGGGGCATTTATCAGCAACAATAT GGTGGGGCCTTCTTGGAACAATTAAAAACACAGTACCCGCAACTTTTCCAGCAAAAACAGATTT CCACAGATCAGCCAATGAACCCAGATATTCAAATTAAGTCATGGGAAGCCAAGTATTTCAACGG TTCGAACATTCAGGGGCGTGGGGCTTGGTATGTTTTGAAGGACTGGGGCACACAACAGTATTTT AATGTGTCAGATGCGCAGACCTTCCTTCCAAAGCAATTATTGGGTGAAAAGGCCAAAACTGGTT TTGTTACGCGTGGTAAGGAGACTTCATTCTATTCCACTAGTGGCTATCAAGCAAAATCTGCCTT TATTTGTGATAACGGTAATTGGTACTACTTTGATGACAAAGGGAAAATGGTTGTTGGAAACCAA GTTATCAATGGCATCAATTATTACTTTTTACCGAATGGTATCGAATTACAAGATGCCTATCTAG TACATGATGGTATGTACTATTATTATAATAATATTGGCAAGCAACTGCACAACACATATTACCA AGATAAACAAAAAAATTTCCATTACTTCTTTGAAGATGGGCACATGGCACAGGGTATTGTCACC ATCATTCAAAGTGATGGCACCCCAGTCACACAGTACTTTGATGAGAATGGTAAGCAACAAAAAG GCGTGGCGGTCAAAGGATCAGATGGTCATTTGCATTACTTTGACGGTGCGTCAGGGAATATGCT CTTTAAATCATGGGGTAGACTAGCAGATGGCTCTTGGCTATATGTAGACGAGAAAGGTAATGCG GTTACAGGCAAACAAACCATTAATAATCAAACGGTTTACTTTAATGATGATGGTCGTCAAATCA AAAATAACTTTAAAGAATTAGCAGATGGTTCTTGGCTTTATCTTAACAATAAAGGTGTTGCAGT AACAGGAGAGCAAATAATTAATGGGCAGACACTTTATTTTGGTAACGATGGTCGTCAATTTAAA GGGACAACACATATAAATGCTACTGGTGAAAGCCGTTACTATGACCCAGACTCAGGTAATATGA TAACTGATCGTTTTGAACGTGTTGGTGATAATCAATGGGCTTATTTTGGTTATGATGGTGTTGC AGTAACAGGGGACCGAATCATTAAAGGGCAAAAACTCTATTTCAACCAAAATGGTATCCAAATG AAAGGCCACTTACGTCTTGAAAATGGTATCATGCGTTATTACGATGCTGATACTGGCGAATTAG TTCGTAATCGATTTGTATTGCTATCTGATGGTTCATGGGTTTACTTTGGCCAAGATGGCGTACC CGTAACTGGCGTGCAAGTGATTAATGGCCAAACATTATATTTTGACGCAGATGGTAGGCAAGTC AAAGGGCAGCAACGTGTAATCGGCAATCAACGCTATTGGATGGATAAAGACAATGGTGAAATGA		

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAAAAATAACATACGCGGCCGCACTCGAGCACCACCACCACCACCACTGA		
DexT gene (DNA sequence cloned into pET23a)	ATGCAAAACGGCGAAGTGTGTCAGCGTAAAAAACTGTACAAGTCAGGGAAGATATTAGTTACAG CAAGTATTTTTGCTGTTATGGGTTTTGGTACTGCCATGTCACAAGCAAACGCGAGCAGTAGTGA TAATGATAGCAAAACACAAACTATTTCAAAAATAGTAAAAAGTAAAGTCGAACCGGCAACTGTT CAACCAGCGAAACCAGCGGAACCTACTAATAAAATAGTTGACCAAGCAGATATGCATACGGTCA GCGGGCAAAACAGCGTGCCACCAGTAGTGACTAATCAATCCAATTAACAGGCTGCAAAACCAAC TACACCTGTTACCGATGTCACAGATACGCATAAAATCGAAGCAAACAACGTCCCTGCTGATGTT ATGCCAGCAAATGCCCCAGATAAACAATCAGTGACTAATGCACCAGTAGTGCCGCCAAAGCATG ATACGGACCAGCAGGACGATTCACTAGAAAAACAGCAAGTATTAGAACCGAGCGTAAATAGTAA TATACCAAAAAAGCAGACAAATCAACAGTTAGCGGTTGTTACAGCACCAGCAAATTCAGCACCT CAAACCAAAACAACAGCAGAAATTTCTGCTGGTACAGAGTTAGACACGATGCCTAATGTTAAGC ATGTAGATGGCAAAGTTTATTTTTATGGAGATGATGGCCAACCAAAAAAGAATTTTACTACTAT TATAGATGGTAAACCTTACTACTTTGATAAAGATACAGGGGCACTATCTAATAACGATAAGCAA TATGTATCGGAATTATTCAGTATTGGCAATAAACATAACGCCGTCTATAACACATCATCAGATA ATTTTACGCAATTAGAAGGACATCTGACGGCAAGTAGTTGGTATCGTCCAAAAGATATTTTGAA AAATGGTAAACGTTGGGCACCTTCAACAGTGACTGATTTCAGACCATTATTGATGGCCTGGTGG CCGGATAAGAGTACGCAAGTCACTTATCTGAATTACATGAAAGATCAGGGCCTCTTGTCTGGTA CTCATCACTTTTCCGATAATGAAAATATGCGGACCTTAACGGCAGCTGCCATGCAGGCACAGGT AAACATTGAGAAAAAAATTGGGCAACTTGGCAATACGGATTGGTTGAAAACGGCGATGACGCAA TACATTGATGCCCAGCCCAATTGGAATATTGACAGTGAGGCGAAAGGAGATGATCATCTACAAG GTGGTGCACTACTTTATACAAATAGTGATATGTCGCCAAAGGCCAATTCTGATTATCGTAAGCT GAGCCGTACGCCTAAAAATCAAAAAGGTCAAATTGCTGATAAATATAAGCAAGGTGGGTTTGAA TTATTACTAGCAAACGATGTCGATAATTCTAATCCAGTTGTGCAAGCAGAACAACTTAATTGGT TACATTATATGATGAATATCGGTAGTATTTTACAAAATGATGACCAAGCTAATTTTGATGGTTA CCGTGTTGATGCTGTCGATAATGTGGACGCTGACTTACTACAGATTGCTGGTGAATATGCTAAG GCTGCCTATGGTGTTGACAAAAATGACGCGAGAGCGAATCAACATTTATCAATTTTGGAAGACT GGGGAGATGAAGATCCAGACTATGTCAAAGCACATGGCAACCAGCAAATTACAATGGATTTCCC CTTGCATTTAGCGATTAAATACGCGCTCAACATGCCTAATGATAAGCGGAGTGGCCTTGAGCCA ACCCGTGAACACAGTTTAGTCAAACGAATTACAGATGATAAAGAAAATGTTGCACAACCAAATT ATTCATTTATCCGAGCTCATGACAGTGAAGTACAAACGATTATTGCTGATATTATTAAAGATAA AATCAACCCGGCGTCAACAGGGCTAGATTCAACAGTGACTTTGGATCAAATTAAGCAGGCTTTT GACATCTATAATGCTGATGAATTGAAAGCAGATAAAGTTTACACACCTTACAATATTCCAGCAT CATACGCTTTGTTATTGACTAATAAAGACACAATTCCACGTGTTTATTATGGGGATATGTTCAC GGATGATGGCCAATACATGGCTAAACAATCACCTTACTATCAAGCGATTGATGCGTTGTTGAAA GCTCGTATCAAGTATGCTGCTGGTGGTCAAACCATGAAAATGAACTATTTTCCAGATGAACAAT CTGTTATGACATCAGTTCGTTATGGTAAGGGTGCAATGACGGCAAGTGACTCTGGTAACCAAGA GACACGCTATCAAGGTATTGGACTTGTTGTCAACAATCGCCCAGATTTGAAACTATCTGACAAA GATGAAGTCAAAATGGATATGGGTGCGGCACATAAAAACCAAGATTATCGCCCAGTTTTGTTGA	105	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CGACAAAATCAGGATTAAAAGTCTACAGCACTGATGCAAATGCACCTGTCGTTCGAACTGACGC CAATGGCCAATTAACTTTTAAGGCAGACATGGTATATGGTGTAAACGACCCACAAGTGTCAGGG TACATTGCGGCTTGGGTACCAGTAGGGGCTTCAGAAAATCAAGATGCTCGAACGAAAAGTGAAA CAACGCAGTCAACTGACGGGAGTGTTTATCATTCTAATGCAGCGTTAGATTCGCAAGTCATTTA TGAAGGCTTTTCAAATTTTCAAGACTTTCCAACAACACCCGATGAGTTTACGAACATTAAAATT GCTCAAAATGTTAACTTATTTAAGGATTGGGGTATTACTAGCTTTGAAATGGCGCCACAATATC GCGCCAGCTCAGATAAAAGTTTCTTAGATGCTATCGTACAAAATGGTTATGCATTTACAGATCG ATATGATATTGGTTACAACACACCAACAAAGTATGGGACAGCAGATAATTTGTTAGATGCTTTA CGTGCATTGCATGGTCAGGGTATTCAAGCGATTAACGACTGGGTACCAGATCAAATTTATAATC TACCCGATGAACAGTTAGTCACGGCTATTCGAACAGACGGTTCAGGTGATCATACTTATGGTTC AGTTATTGACCATACTTTGTATGCATCAAAGACAGTTGGCGGGGGCATTTATCAGCAACAATAT GGTGGGGCCTTCTTGGAACAATTAAAAACACAGTACCCGCAACTTTTCCAGCAAAAACAGATTT CCACAGATCAGCCAATGAACCCAGATATTCAAATTAAGTCATGGGAAGCCAAGTATTTCAACGG TTCGAACATTCAGGGGCGTGGGGCTTGGTATGTTTTGAAGGACTGGGGCACACAACAGTATTTT AATGTGTCAGATGCGCAGACCTTCCTTCCAAAGCAATTATTGGGTGAAAAGGCCAAAACTGGTT TTGTTACGCGTGGTAAGGAGACTTCATTCTATTCCACTAGTGGCTATCAAGCAAAATCTGCCTT TATTTGTGATAACGGTAATTGGTACTACTTTGATGACAAAGGGAAAATGGTTGTTGGAAACCAA GTTATCAATGGCATCAATTATTACTTTTTACCGAATGGTATCGAATTACAAGATGCCTATCTAG TACATGATGGTATGTACTATTATTATAATAATATTGGCAAGCAACTGCACAACACATATTACCA AGATAAACAAAAAAATTTCCATTACTTCTTTGAAGATGGGCACATGGCACAGGGTATTGTCACC ATCATTCAAAGTGATGGCACCCCAGTCACACAGTACTTTGATGAGAATGGTAAGCAACAAAAAG GCGTGGCGGTCAAAGGATCAGATGGTCATTTGCATTACTTTGACGGTGCGTCAGGGAATATGCT CTTTAAATCATGGGGTAGACTAGCAGATGGCTCTTGGCTATATGTAGACGAGAAAGGTAATGCG GTTACAGGCAAACAAACCATTAATAATCAAACGGTTTACTTTAATGATGATGGTCGTCAAATCA AAAATAACTTTAAAGAATTAGCAGATGGTTCTTGGCTTTATCTTAACAATAAAGGTGTTGCAGT AACAGGAGAGCAAATAATTAATGGGCAGACACTTTATTTTGGTAACGATGGTCGTCAATTTAAA GGGACAACACATATAAATGCTACTGGTGAAAGCCGTTACTATGACCCAGACTCAGGTAATATGA TAACTGATCGTTTTGAACGTGTTGGTGATAATCAATGGGCTTATTTTGGTTATGATGGTGTTGC AGTAACAGGGGACCGAATCATTAAAGGGCAAAAACTCTATTTCAACCAAAATGGTATCCAAATG AAAGGCCACTTACGTCTTGAAAATGGTATCATGCGTTATTACGATGCTGATACTGGCGAATTAG TTCGTAATCGATTTGTATTGCTATCTGATGGTTCATGGGTTTACTTTGGCCAAGATGGCGTACC CGTAACTGGCGTGCAAGTGATTAATGGCCAAACATTATATTTTGACGCAGATGGTAGGCAAGTC AAAGGGCAGCAACGTGTAATCGGCAATCAACGCTATTGGATGGATAAAGACAATGGTGAAATGA AAAAAATAACATACGCGGCCGCACTCGAGCACCACCACCACCACCACTGA		
Dextransucrase (also known as 6-glucosyltransferase)	MEKKVRFKLRKVKKRWVTVSVASAVVTLTSLSGSLVKADSTDDRQQAVTESQASLVTTSEAAKE TLTATDTSTATSATSQPTATVTDNVSTTNQSTNTTANTANFDVKPTTTSEQSKTDNSDKIIATS KAVNRLTATGKFVPANNNTAHSRTVTDKIVPIKPKIGKLKQPSSLSQDDIAALGNVKNIRKVNG KYYYYKEDGTLQKNYALNINGKTFFFDETGALSNNTLPSKKGNITNNDNTNSFAQYNQVYSTDA	106	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
UniProtKB/Swiss-Prot: P13470.2	ANFEHVDHYLTAESWYRPKYILKDGKTWTQSTEKDFRPLLMTWWPDQETQRQYVNYMNAQLGIH QTYNTATSPLQLNLAAQTIQTKIEEKITAEKNTNWLRQTISAFVKTQSAWNSDSEKPFDDHLQK GALLYSNNSKLTSQANSNYRILNRTPTNQTGKKDPRYTADRTIGGYEFLLANDVDNSNPVVQAE QLNWLHFLMNFGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAKGIHKNDKAANDHLS ILEAWSYNDTPYLHDDGDNMINMDNRLRLSLLYSLAKPLNQRSGMNPLITNSLVNRTDDNAETA AVPSYSFIRAHDSEVQDLIRNIIRAEINPNVVGYSFTMEEIKKAFEIYNKDLLATEKKYTHYNT ALSYALLLTNKSSVPRVYYGDMFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAMRNQQVGN SEIITSVRYGKGALKATDTGDRTTRTSGVAVIEGNNPSLRLKASDRVVVNMGAAHKNQAYRPLL LTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGVWVPVGAAADQDVRV AASTAPSTDGKSVHQNAALDSRVMFEGFSNFQAFATKKEEYTNVVIAKNVDKFAEWGVTDFEMA PQYVSSTDGSFLDSVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ MYALPEKEVVTATRVDKYGTPVAGSQIKNTLYVVDGKSSGKDQQAKYGGAFLEELQAKYPELFA RKQISTGVPMDPSVKIKQWSAKYFNGTNILGRGAGYVLKDQATNTYFSLVSDNTFLPKSLVNPN HGTSSSVTGLVFDGKGYVYYSTSGNQAKNAFISLGNNWYYFDNNGYMVTGAQSINGANYYFLSN GIQLRNAIYDNGNKVLSYYGNDGRRYENGYYLFGQQWRYFQNGIMAVGLTRIHGAVQYFDASGF QAKGQFITTADGKLRYFDRDSGNQISNRFVRNSKGEWFLFDHNGVAVTGTVTFNGQRLYFKPNG VQAKGEFIRDADGHLRYYDPNSGNEVRNRFVRNSKGEWFLFDHNGIAVTGTRVVNGQRLYFKSN GVQAKGELITERKGRIKYYDPNSGNEVRNRYVRTSSGNWYYFGNDGYALIGWHVVEGRRVYFDE NGVYRYASHDQRNHWDYDYRRDFGRGSSSAVRFRHSRNGFFDNFFRF		
Dextransucrase (also known as Sucrose 6-glucosyltransferase) UniProtKB/Swiss-Prot: P08987.3	MDKKVRYKLRKVKKRWVTVSVASAVMTLTTLSGGLVKADSNESKSQISNDSNTSVVTANEESNV TTEVTSKQEAASSQTNHTVTTISSSTSVVNPKEVVSNPYTVGETASNGEKLQNQTTTVDKTSEA AANNISKQTTEADTDVIDDSNAANLQILEKLPNVKEIDGKYYYYDNNGKVRTNFTLIADGKILH FDETGAYTDTSIDTVNKDIVTTRSNLYKKYNQVYDRSAQSFEHVDHYLTAESWYRPKYILKDGK TWTQSTEKDFRPLLMTWWPSQETQRQYVNYMNAQLGINKTYDDTSNQLQLNIAAATIQAKIEAK ITTLKNTDWLRQTISAFVKTQSAWNSDSEKPFDDHLQNGAVLYDNEGKLTPYANSNYRILNRTP TNQTGKKDPRYTADNTIGGYEFLLANDVDNSNPVVQAEQLNWLHFLMNFGNIYANDPDANFDSI RVDAVDNVDADLLQIAGDYLKAAKGIHKNDKAANDHLSILEAWSDNDTPYLHDDGDNMINMDNK LRLSLLFSLAKPLNQRSGMNPLITNSLVNRTDDNAETAAVPSYSFIRAHDSEVQDLIRDIIKAE INPNVVGYSFTMEEIKKAFEIYNKDLLATEKKYTHYNTALSYALLLTNKSSVPRVYYGDMFTDD GQYMAHKTINYEAIETLLKARIKYVSGGQAMRNQQVGNSEIITSVRYGKGALKATDTGDRTTRT SGVAVIEGNNPSLRLKASDRVVVNMGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDR GELIFTAADIKGYANPQVSGYLGVWVPVGAAADQDVRVAASTAPSTDGKSVHQNAALDSRVMFE GFSNFQAFATKKEEYTNVVIAKNVDKFAEWGVTDFEMAPQYVSSTDGSFLDSVIQNGYAFTDRY DLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQMYAFPEKEVVTATRVDKFGKPVEGSQ IKSVLYVADSKSSGKDQQAKYGGAFLEELQAKYPELFARKQISTGVPMDPSVKIKQWSAKYFNG TNILGRGAGYVLKDQATNTYFNISDNKEINFLPKTLLNQDSQVGFSYDGKGYVYYSTSGYQAKN TFISEGDKWYYFDNNGYMVTGAQSINGVNYYFLSNGLQLRDAILKNEDGTYAYYGNDGRRYENG YYQFMSGVWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFVTTADGKIRYFDKQSGNMYRNR	107	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	FIENEEGKWLYLGEDGAAVTGSQTINGQHLYFRANGVQVKGEFVTDRYGRISYYDSNSGDQIRN RFVRNAQGQWFYFDNNGYAVTGARTINGQHLYFRANGVQVKGEFVTDRHGRISYYDGNSGDQIR NRFVRNAQGQWFYFDNNGYAVTGARTINGQHLYFRANGVQVKGEFVTDRYGRISYYDSNSGDQI RNRFVRNAQGQWFYFDNNGYAVTGARTINGQHLYFRANGVQVKGEFVTDRYGRISYYDANSGER VRIN		
Glucosyltransferase-S (also known as GTF-S, Dextransucrase, Sucrose 6-glucosyltransferase) UniProtKB/Swiss-Prot: P49331.3	METKRRYKMYKVKKHWVTIAVASGLITLGTTTLGSSVSAETEQQTSDKVVTQKSEDDKAASESS QTDAPKTKQAQTEQTQAQSQANVADTSTSITKETPSQNITTQANSDDKTVTNTKSEEAQTSEER TKQAEEAQATASSQALTQAKAELTKQRQTAAQENKNPVDLAAIPNVKQIDGKYYYIGSDGQPKK NFALTVNNKVLYFDKNTGALTDTSQYQFKQGLTKLNNDYTPHNQIVNFENTSLETIDNYVTADS WYRPKDILKNGKTWTASSESDLRPLLMSWWPDKQTQIAYLNYMNQQGLGTGENYTADSSQESLN LAAQTVQVKIETKISQTQQTQWLRDIINSFVKTQPNWNSQTESDTSAGEKDHLQGGALLYSNSD KTAYANSDYRLLNRTPTSQTGKPKYFEDNSSGGYDFLLANDIDNSNPVVQAEQLNWLHYLMNYG SIVANDPEANFDGVRVDAVDNVNADLLQIASDYLKAHYGVDKSEKNAINHLSILEAWSDNDPQY NKDTKGAQLPIDNKLRLSLLYALTRPLEKDASNKNEIRSGLEPVITNSLNNRSAEGKNSERMAN YIFIRAHDSEVQTVIAKIIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKKYTQSNIPTAY ALMLSNKDSITRLYYGDMYSDDGQYMATKSPYYDAIDTLLKARIKYAAGGQDMKITYVEGDKSH MDWDYTGVLTSVRYGTGANEATDQGSEATKTQGMAVITSNNPSLKLNQNDKVIVNMGTAHKNQE YRPLLLTTKDGLTSYTSDAAAKSLYRKTNDKGELVFDASDIQGYLNPQVSGYLAVWVPVGASDN QDVRVAASNKANATGQVYESSSALDSQLIYEGFSNFQDFVTKDSDYTNKKIAQNVQLFKSWGVT SFEMAPQYVSSEDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSGIQVIAD WVPDQIYNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNGKDYQAKYGGAFLSELAAKY PSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVGYVLKDNASDKYFELKGNQTYLPKQ MTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYYFDNNGHMVYGLQHLNGEVQYFLS NGVQLRESFLENADGSKNYFGHLGNRYSNGYYSFDNDSKWRYFDASGVMAVGLKTINGNTQYFD QDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYLNSDGIALVGVQTINGKTYYF GQDGKQIKGKIITDNGKLKYFLANSGELARNIFATDSQNNWYYFGSDGVAVTGSQTIAGKKLYF ASDGKQVKGSFVTYNGKVHYYHADSGELQVNRFEADKDGNWYYLDSNGEALTGSQRINGQRVFF TREGKQVKGDVAYDERGLLRYYDKNSGNMVYNKVVTLANGRRIGIDRWGIARYY	108	
Dextransucrase (EC 2.4.1.5) precursor - Streptococcus mutans PIR: A45866	METKRRYKMHKVKKHWVTVAVASGLITLGTTTLGSSVSAETEQQTSDKVVTQKSEDDKAASESS QTDAPKTKQAQTEQTQAQSQANVADTSTSITKETPSQNITTQANSDDKTVTNTKSEEAQTSEER TKQSEEAQTTASSQALTQAKAELTKQRQTAAQENKNPVDLAAIPNVKQIDGKYYYIGSDGQPKK NFALTVNNKVLYFDKNTGALTDTSQYQFKQGLTKLNNDYTPHNQIVNFENTSLETIDNYVTADS WYRPKDILKNGKTWTASSESDLRPLLMSWWPDKQTQIAYLNYMNQQGLGTGENYTADSSQESLN LAAQTVQVKIETKISQTQQTQWLRDIINSFVKTQPNWNSQTESDTSAGEKDHLQGGALLYSNSD KTAYANSDYRLLNRTPTSQTGKPKYFEDNSSGGYDFLLANDIDNSNPVVQAEQLNWLHYLMNYG SIVANDPEANFDGVRVDAVDNVNADLLQIASDYLKAHYGVDKSEKNAINHLSILEAWSDNDPQY NKDTKGAQLPIDNKLRLSLLYALTRPLEKDASNKNEIRSGLEPVITNSLNNRSAEGKNSERMAN YIFIRAHDSEVQTVIAKIIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKKYTQSNIPTAY	109	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ALMLSNKDSITRLYYGDMYSDDGQYMATKSPYYDAIDTLLKARIKYAAGGQDMKITYVEGDKSH MDWDYTGVLTSVRYGTGANEATDQGSEATKTQGMAVITSNNPSLKLNQNDKVIVNMGAAHKNQE YRPLLLTTKDGLTSYTSDAAAKSLYRKTNDKGELVFDASDIQGYLNPQVSGYLAVWVPVGASDN QDVRVAASNKANATGQVYESSSALDSQLIYEGFSNFQDFVTKDSDYTNKKIAQNVQLFKSWGVT SFEMAPQYVSSEDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSGIQVIAD WVPDQIYNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNGKDYQAKYGGAFLSELAAKY PSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVGYVLKDNASDKYFELKGNQTYLPKQ MTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYYFDNNGHMVYGLQQLNGEVQYFLS NGVQLRESFLENADGSKNYFGHLGNRYSNGYYSFDNDSKWRYFDASGVMAVGLKTINGNTQYFD QDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYLNSDGIALVGVQTINGKTYYF GQDGKQIKGKIITDNGKLKYFLANSGELARNIFATDSQNNWYYFGSDGVAVTGSQTIAGKKLYF ASDGKQVKGSFVTYNGKVHYYHADSGELQVNRFEADKDGNWYYLDSNGEALTGSQRINDQRVFF TREGKQVKGDVAYDERGLLRYYD		
Dextranase	MFSAVLLGWLLFQPTVGHAIRQRAGNHTVCNSQLCTWWHDNGEINTASMVQLGNVRQSHKYLVQ VSIAGVNDFYDSFAYESIPRNGRGRIYSPWDPPNSDTLGSDVDDGITIETSAGINMAWSQFEYS TGVDVKILTRDGSRLPDPSGVKIRPTAISYDIRSSSDGGIVIRVPHDPNGRRFSVEFDNDLYTY RSDGSRYVSSGGSIVGVEPRNALVIFASPFLPDNMVPRIDGPDTKVMTPGPINQGDWGSSGILY FPPGVYWMNSNQQGQTPKIGENHIRLHPNTYWAYLAPGAYVKGAIEYSTKSDFYATGHGVLSGE HYVYQANPATYYQALKSDATSLRMWWHNNLGGGQTWYCQGPTINAPPFNTMDFHGSSDITTRIS DYKQVGAFFFQTDGPQMYPNSQVHDVFYHVNDDAIKTYYSGVTVTRATIWKAHNDPIIQMGWDT RDVTGVTLQDLYIIHTRYIKSETYVPSAIIGASPFYMPGRSVDPAKSISMTISNLVCEGLCPAL MRITPLQNYRDFRIQNVAFPDGLQANSIGTGKSIVPASSGLKFGVAISNWTVGGEQVTMSNFQS DSLGQLDIDVSYWGQWVIR	110	
Lanosterol synthase [*S. cerevisiae*]	MTEFYSDTIGLPKTDPRLWRLRTDELGRESWEYLTPQQAANDPPSTFTQWLLQDPKFPQPHPER NKHSPDFSAFDACHNGASFFKLLQEPDSGIFPCQYKGPMFMTIGYVAVNYIAGIEIPEHERIEL IRYIVNTAHPVDGGWGLHSVDKSTVFGTVLNYVILRLLGLPKDHPVCAKARSTLLRLGGAIGSP HWGKIWLSALNLYKWEGVNPAPPETWLLPYSLPMHPGRWWVHTRGVYIPVSYLSLVKFSCPMTP LLEELRNEIYTKPFDKINFSKNRNTVCGVDLYYPHSTTLNIANSLVVFYEKYLRNRFIYSLSKK KVYDLIKTELQNTDSLCIAPVNQAFCALVTLIEEGVDSEAFQRLQYRFKDALFHGPQGMTIMGT NGVQTWDCAFAIQYFFVAGLAERPEFYNTIVSAYKFLCHAQFDTECVPGSYRDKRKGAWGFSTK TQGYTVADCTAEAIKAIIMVKNSPVFSEVHHMISSERLFEGIDVLLNLQNIGSFEYGSFATYEK IKAPLAMETLNPAEVFGNIMVEYPYVECTDSSVLGLTYFHKYFDYRKEEIRTRIRIAIEFIKKS QLPDGSWYGSWGICFTYAGMFALEALHTVGETYENSSTVRKGCDFLVSKQMKDGGWGESMKSSE LHSYVDSEKSLVVQTAWALIALLFAEYPNKEVIDRGIDLLKNRQEESGEWKFESVEGVFNHSCA IEYPSYRFLFPIKALGMYSRAYETHTL	111	SEQ ID NO: 55 in WO 2016/050890
DNA sequence encoding S. *grosvenorii*	ATGAAGGTCTCTCCATTTGAGTTCATGTCGGCAATAATTAAGGGCAGGATGGACCCGTCCAATT CTTCATTTGAGTCGACTGGCGAGGTTGCCTCAGTTATTTTCGAGAACCGTGAGCTGGTTGCGAT CTTAACCACCTCGATCGCCGTCATGATTGGCTGCTTCGTTGTTCTCATGTGGCGAAGAGCCGGC	112	SEQ ID NO: 45 of WO 2016/050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
CPR4497	AGTCGGAAAGTTAAGAACGTGGAGCTACCTAAGCCGTTGATTGTGCACGAGCCGGAGCCCGAAG TTGAAGACGGCAAGAAGAAGGTTTCAATCTTCTTCGGTACACAGACAGGCACCGCCGAAGGATT TGCAAAGGCTCTAGCTGACGAGGCGAAAGCACGATACGAGAAGGCCACATTTAGAGTTGTTGAT TTGGATGATTATGCAGCTGATGACGATCAGTATGAAGAGAAGTTGAAGAACGAGTCTTTCGCTG TCTTCTTATTGGCAACGTATGGCGATGGAGAGCCCACTGATAATGCCGCAAGATTCTATAAATG GTTCGCGGAGGGGAAAGAGAGAGGGGAGTGGCTTCAGAACCTTCATTATGCGGTCTTTGGCCTT GGCAACCGACAGTACGAGCATTTTAATAAGATTGCAAAGGTGGCAGATGAGCTGCTTGAGGCAC AGGGAGGCAACCGCCTTGTTAAAGTTGGTCTTGGAGATGACGATCAGTGCATAGAGGATGACTT CAGTGCCTGGAGAGAATCATTGTGGCCTGAGTTGGATATGTTGCTTCGAGATGAGGATGATGCA ACAACAGTGACCACCCCTTACACAGCTGCCGTATTAGAATATCGAGTTGTATTCCATGATTCTG CAGATGTAGCTGCTGAGGACAAGAGCTGGATCAATGCAAACGGTCATGCTGTACATGATGCTCA GCATCCCTTCAGATCTAATGTGGTTGTGAGGAAGGAGCTCCATACGTCCGCATCTGATCGCTCC TGTAGTCATCTAGAATTTAATATTTCTGGGTCTGCACTCAATTATGAAACAGGGGATCATGTCG GTGTTTACTGTGAAAACTTAACTGAGACTGTGGACGAGGCACTAAACTTATTGGGTTTGTCTCC TGAAACGTATTTCTCCATATATACTGATAACGAGGATGGCACTCCACTTGGTGGAAGCTCTTTA CCACCTCCTTTTCCATCCTGCACCCTCAGAACAGCATTGACTCGATATGCAGATCTCTTGAATT CACCCAAGAAGTCAGCTTTGCTTGCATTAGCAGCACATGCTTCAAATCCAGTAGAGGCTGACCG ATTAAGATATCTTGCATCACCTGCCGGGAAGGATGAATACGCCCAGTCTGTGATTGGTAGCCAG AAAAGCCTTCTTGAGGTCATGGCTGAATTTCCTTCTGCCAAGCCCCCACTTGGTGTCTTCTTCG CAGCTGTTGCACCGCGCTTGCAGCCTCGATTCTACTCCATATCATCATCTCCAAGGATGGCTCC ATCTAGAATTCATGTTACTTGTGCTTTAGTCTATGACAAAATGCCAACAGGACGTATTCATAAA GGAGTGTGCTCAACTTGGATGAAGAATTCTGTGCCCATGGAGAAAAGCCATGAATGCAGTTGGG CTCCAATTTTCGTGAGACAATCAAACTTCAAGCTTCCTGCAGAGAGTAAAGTGCCCATTATCAT GGTTGGTCCTGGAACTGGATTGGCTCCTTTCAGAGGTTTCTTACAGGAAAGATTAGCTTTGAAG GAATCTGGAGTAGAATTGGGGCCTTCCATATTGTTCTTTGGATGCAGAAACCGTAGGATGGATT ACATATACGAGGATGAGCTGAACAACTTTGTTGAGACTGGTGCTCTCTCTGAGTTGGTTATTGC CTTCTCACGCGAAGGGCCAACTAAGGAATATGTGCAGCATAAAATGGCAGAGAAGGCTTCGGAT ATCTGGAATTTGATATCAGAAGGGGCTTACTTATATGTATGTGGTGATGCAAAGGGCATGGCTA AGGATGTCCACCGAACTCTCCATACTATCATGCAAGAGCAGGGATCTCTTGACAGCTCAAAAGC TGAGAGCATGGTGAAGAATCTGCAAATGAATGGAAGGTATCTGCGTGATGTCTGGTGA		
CPR4497 protein [S. *grosvenorii*]	MKVSPFEFMSAIIKGRMDPSNSSFESTGEVASVIFENRELVAILTTSIAVMIGCFVVLMWRRAG SRKVKNVELPKPLIVHEPEPEVEDGKKKVSIFFGTQTGTAEGFAKALADEAKARYEKATFRVVD LDDYAADDDQYEEKLKNESFAVFLLATYGDGEPTDNAARFYKWFAEGKERGEWLQNLHYAVFGL GNRQYEHFNKIAKVADELLEAQGGNRLVKVGLGDDDQCIEDDFSAWRESLWPELDMLLRDEDDA TTVTTPYTAAVLEYRVVFHDSADVAAEDKSWINANGHAVHDAQHPFRSNVVVRKELHTSASDRS CSHLEFNISGSALNYETGDHVGVYCENLTETVDEALNLLGLSPETYFSIYTDNEDGTPLGGSSL PPPFPSCTLRTALTRYADLLNSPKKSALLALAAHASNPVEADRLRYLASPAGKDEYAQSVIGSQ KSLLEVMAEFPSAKPPLGVFFAAVAPRLQPRFYSISSSPRMAPSRIHVTCALVYDKMPTGRIHK	113	SEQ ID NO: 46 of WO 2016/050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GVCSTWMKNSVPMEKSHECSWAPIFVRQSNFKLPAESKVPIIMVGPGTGLAPFRGFLQERLALK ESGVELGPSILFFGCRNRRMDYIYEDELNNFVETGALSELVIAFSREGPTKEYVQHKMAEKASD IWNLISEGAYLYVCGDAKGMAKDVHRTLHTIMQEQGSLDSSKAESMVKNLQMNGRYLRDVW		
Codon optimized coding sequence of Epoxide hydrolase 1 from S grosvenorii	ATGGACGCGATTGAACATAGAACCGTAAGTGTTAATGGTATCAATATGCATGTGGCAGAAAAGG GAGAGGGACCTGTCGTGTTGTTGCTTCATGGTTTCCCAGAATTGTGGTACAGTTGGAGACATCA AATATTGGCTCTTTCCTCTTTAGGTTACAGAGCTGTCGCACCAGACTTACGAGGCTACGGGGAT ACAGATGCCCCAGGGTCAATTTCATCATACACATGCTTTCACATCGTAGGAGATCTCGTGGCTC TAGTTGAGTCTCTGGGTATGGACAGGGTTTTTGTTGTAGCCCACGATTGGGGTGCCATGATCGC TTGGTGTTTGTGTCTGTTTAGACCTGAAATGGTTAAAGCTTTTGTTTGTCTCTCCGTCCCATTC AGACAGAGAAACCCTAAGATGAAACCAGTTCAAAGTATGAGAGCCTTTTTCGGCGATGATTACT ATATTTGCAGATTTCAAAATCCTGGGGAAATCGAAGAGGAGATGGCTCAAGTGGGTGCAAGGGA AGTCTTAAGAGGAATTCTAACATCTCGTCGTCCTGGACCACCAATCTTACCAAAAGGGCAAGCT TTTAGAGCAAGACCAGGAGCATCCACTGCATTGCCATCTTGGCTATCTGAAAAAGATCTGTCAT TTTTCGCTTCTAAGTATGATCAAAAGGGCTTTACAGGCCCACTAAACTACTACAGAGCCATGGA TCTTAATTGGGAATTGACTGCGTCATGGACTGGTGTCCAAGTTAAAGTACCTGTCAAATACATC GTGGGTGACGTTGACATGGTTTTTACGACTCCTGGTGTAAAGGAATATGTCAACGGCGGTGGTT TCAAAAAGGACGTTCCATTTTTACAGGAAGTGGTAATCATGGAAGGCGTTGGTCATTTCATTAA TCAGGAAAAACCTGAGGAGATTTCATCTCATATACACGATTTCATAAGCAAATTCTAA	114	SEQ ID NO: 37 of WO 2016/050890
Codon optimized coding sequence of *S. grosvenorii* Epoxide hydrolase 2	ATGGATGAAATCGAACATATTACCATCAATACAAATGGAATCAAAATGCATATTGCGTCAGTCG GCACAGGACCAGTTGTTCTCTTGCTACACGGCTTTCCAGAATTATGGTACTCTTGGAGACACCA ACTACTTTACCTGTCCTCCGTTGGGTACAGAGCAATAGCTCCAGATTTGAGAGGCTATGGCGAT ACTGACAGTCCAGCTAGTCCTACCTCTTATACTGCTCTTCATATTGTAGGTGACCTGGTCGGCG CATTAGACGAATTGGGAATAGAAAAGGTCTTTTTAGTGGGTCATGACTGGGGTGCTATTATCGC ATGGTACTTTTGTTTGTTTAGACCAGATAGAATTAAAGCACTTGTGAATTTGTCTGTCCAGTTT ATCCCACGTAACCCAGCAATACCTTTTATAGAAGGTTTCAGAACAGCTTTTGGTGATGACTTCT ACATTTGTAGATTTCAAGTACCTGGGGAAGCTGAAGAGGATTTCGCGTCTATCGATACTGCTCA ATTGTTTAAAACTTCATTATGCAATAGAAGCTCAGCCCCTCCTTGTTTGCCTAAAGAGATTGGT TTTAGGGCTATCCCACCACCAGAAAATCTGCCATCTTGGCTCACAGAGGAAGATATCAACTTCT ACGCAGCCAAGTTTAAACAAACTGGTTTTACTGGTGCCCTTAACTATTATAGAGCATTCGACTT GACATGGGAATTAACAGCCCCATGGACAGGAGCCCAGATCCAAGTTCCTGTAAAGTTCATAGTT GGTGATTCAGATCTCACGTACCATTTCCCTGGTGCTAAGGAATACATCCACAACGGAGGGTTTA AAAGAGATGTGCCACTATTAGAGGAAGTTGTTGTGGTAAAAGATGCCTGCCACTTCATTAACCA AGAGCGACCACAAGAGATTAATGCTCATATTCATGACTTCATCAATAAGTTCTAA	115	SEQ ID NO: 39 of WO 2016/050890
UGT3494 [*S. grosvenorii*]	ATGGCGGATCGGAAAGAGAGCGTTGTGATGTTCCCGTTCATGGGGCAGGGCCATATCATCCCTT TTCTAGCTTTGGCCCTCCAGATTGAGCACAGAAACAGAAACTACGCCATATACTTGGTAAATAC TCCTCTCAACGTTAAGAAAATGAGATCTTCTCTCCCTCCAGATTGA	116	SEQ ID NO: 29 of WO 2016/050890
Fragment of *S. grosvenorii*	TTCTGCTCCACGCCTGTAAATTTGGAAGCCATTAAACCAAAGCTTTCCAAAAGCTACTCTGATT CGATCCAACTAATGGAGGTTCCTCTCGAATCGACGCCGGAGCTTCCTCCTCACTATCATACAGC	117	SEQ ID NO: 33 of WO

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
UGT11789 gene sequence	CAAAGGCCTTCCGCCGCATTTAATGCCCAAACTCATGAATGCCTTTAAAATGGTTGCTCCCAAT CTCGAATCGATCCTAAAAAACCCTAAACCCAGATCTGCTCATCGTCGACATTCTCCTTCCATGGA TGCTTCCACTCGCTTCATCGCTCAAAATTCCGATGGTTTTCTTCACTATTTTCGGTGCCATGGC CATCTCCTTTATGATTTATAATCGAACCGTCTCGAACGAGCTTCCATTTCCAGAATTTGAACTT CACGAGTGCTGGAAATCGAAGTGCCCCTATTTGTTCAAGGACCAAGCGGAAAGTCAATCGTTCT TAGAATACTTGGATCAATCTTCAGGCGTAATTTTGATCAAAACTTCCAGAGAGATTGAGGCTAA GTATGTAGACTTTCTCACTTCGTCGTTTACGAAGAAGGTTGTGACCACCGGTCCCCTGGTTCAG CAACCTTCTTCCGGCGAAGACGAGAAGCAGTACTCCGATATCATCGAATGGCTAGACAAGAAGG AGCCGTTATCGACGGTGCTCGTTTCGTTTGGGAGCGAGTATTATCTGTCAAAGGAAGAGATGGA AGAAATCGCCTACGGGCTGGAGAGCGCCAGCGAGGTGAATTTCATCTGGATTGTTAGGTTTCCG ATGGGACAGGAAACGGAGGTCGAGGCGGCGCTGCCGGAGGGGTTCATCCAGAGGGCAGGAGAGA GAGGGAAAGTGGTCGAGGGCTGGGCTCCGCAGGCGAAAATATTGGCGCATCCGAGCACCGGCGG CCATGTGAGCCACAACGGGTGGAGCTCGATTGTGGAGTGCTTGATGTCCGGTGTACCGGTGATC GGCGCGCCGATGCAACTTGACGGGCCAATCGTCGCAAGGCTGGTGGAGGAGATCGGCGTGGGTT TGGAAATCAAGAGAGATGAGGAAGGGAGAATCACGAGGGGCGAAGTTGCCGATGCAATCAAGAC GGTGGCGGTGGGCAAAACCGGGGAAGATTTTAGAAGGAAAGCAAAAAAAATCAGCAGCATTTTG AAGATGAAAGATGAAGAAGAGGTTGACACTTTGGCAATGGAATTAGTGAGGTTATGCCAAATGA AAAGAGGGCAGGAGTCTCAGGACTAA		2016/050890
UGT11999 gene sequence [*S. grosvenorii*]	TCCCGGTCAACGGTAGAGGACTTCACGGAGCTTCGAGAGTGGATGCCTTCTGGATCGAACATGG TCTACCGGTACCACGAGATTAAAAAATCCTTAGATGGAGCAACCGGCAACGAATCGGGGACGTC TGATTCGGTCCGATTCGGAATTGTGATTGAGGAGAGTGTTGCTGTGGCTGTAAGAAGCTCCCCT GAACTGGAACCGGAATGGTTCGATTTGCTCGCGAAGCTTTACCAGAAGCCAGTTGTTCCGGTAG GATTTCTACCTCCAGTAATTGAAGATGCGGAAGAATTGAGCAGCGATATCAAGGAATGGTTAGA CAAACAGAGCTCAAACTCGGTCCTTTACGTCGCATTCGGGACCGAGGCGACTCTGAGTCAAGAT GACGTCACTGAGTTAGCCATGGGGCTTGAGCAATCTGGGATACCATTTTTCTGGGTACTGAGAA CCTCACCTCGGGACGAGTCAGACATGTTACCGGCCGGGTTCAAGGAGCGAGTCGAAGGTCGAGG AAGTGTTCACGTGGGATGGGTCTCGCAGGTGAAGATACTGAGTCACGACTCGGTTGGCGGTTGT TTGACACACTGTGGATGGAACTCGATCATAGAGGGGCTCGGATTCGGGCGCGTTATGGTATTGT TTCCAGTCGTGAACGACCAGGGATTGAACGCTAGATTGTTGGGGGAGAAGAAGCTCGGGATAGA GATAGAAAGGGACGAGCGAGATGGATCGTTCACACGCGACTCGGTGTCGGAATCGGTGAGGTCG GCAATGGCGGAAAGTTCAGGCGAGGCCTTGAGAGTGAGGGCCAGGGAAATGAAGGGGTTGTTTG GAAACGGAGATGAGAACGAGCATCAACTGAACAAGTTTGTACAATTTCTCGAGGCAAACAGGAA TAGGCAGTCCGAGTAA	118	SEQ ID NO: 34 of WO 2016/050890
Partial UGT13679 gene sequence [*Siraitia grosvenorii*]	CTGCTGCCGATTCCGCTGCCGAAACCGGCCGCCGATCTCTTGCCGGAAGGTGCAGAGGCGACGG TGGATATTCCGTCCGACAAGATTCCGTATCTGAAATTGGCCCTCGATCTCGCCGAGCAGCCGTT TCGGAAGTTCGTCGTTGATCGTCCGCCGGATTGGATGATCGTCGATTTTAATGCTACTTGGGTC TGCGATATTTCTCGGGAGCTTCAAATCCCAATCGTTTTCTTTCGTGTTCTTTCGCCTGGATTTC TTGCTTTCTTTGCGCATGTTCTTGGGAGTGGTCTGCCGCTGTCGGAGATCGAAAGCCTGATGAC	119	SEQ ID NO: 35 of WO 2016/050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TCCGCCGGTGATCGACGGGTCGACGGTGGCGTACCGCCGGCATGAAGCTGCCGTTATTTGTGCT GGGTTTTTTGAGAAGAACGCTTCTGGTATGAGTGATCGCGATCGGGTAACCAAAATTCTCTCTG CCAGTCAAGCAATCGCAGTTCGTTCTTGCTACGAATTTGACGTTGAGTATTTGAAATTGTACGA GAAATATTGTGGAAAAAGAGTGATTCCTCTAGGGTTTCTCCCTCCAGAAAAGCCCCAAAAGTCC GAGTTCGCCGCCGATTCGCCATGGAAACCGACCTTCGAGTGGCTTGACAAACAAAAGCCCCGAT CAGTGGTGTTCGTCGGATTCGGCAGCGAATGCAAACTCACGAAAGATGATGTTTACGAGATAGC GCGCGGGGTGGAGCTGTCGGAGCTGCCATTTTTGTGGGCTCTGAGAAAACCGATCTGGGCGGCG GCGGACGATTCCGACGCTCTGCCTGCCGGATTCCTCGAGCGGACGGCGGAGAGAGGGATTGTGA GCATGGGGTGGGCGCCGCAGATGGAGATTTTAACGCACCCGTCGATTGGCGGCTCTCTGTTTCA CGCCGGGTGGGGATCCGCCATTGAAGCTCTGCAATTCGGGCATTGCCTTGTTCTGTTGCCATTC ATCGTGGATCAGCCACTGAATGCAAGGCTTCTGGTGGAGAAGGGTGTTGCAGTCGAAGTTGGAA GAAAGGAAGACGGGTCTTTTAGTGGAGAAGACATAGCTAAAGCTCTGAGAGAAGCTATGGTTTC AGAAGAAGGTGAGCAGATGAGG AGGCAAGCGAGAAAG		
Partial sequence of *S. grosvenorii* UGT 15423	ATGGAAAACGACGGCGTTTTGCACGTGGTGGTATTCCCATGGCTAGCCTTGGGTCATCTCATTC CTTTCGCTCGACTCGCCACCTGCTTAGCCCACAAGGGTCTCAGGGTTTCGTTCGTATCAACCAC AAGGAACCTGAGCAGAATTCCCAAAATACCCCCACATCTCTCCTCCTCCGTCAACCTCGTCGGC TTTCCTCTGCCCCACGTCGACGGCCTTCCGGACGCCGCCGAGGCTTCCTCCGACGTGCCTTACA ACAAGCAACAGTTACTGAAGAAGGCCTTCGACTCTCTGGAATCACCGCTCGCCGATTTGCTTCG TGATTTGAATCCCGATTGGATTATCTACGATTACGCCTCTCATTGGCTTCCGCAGCTCGCGGCG GAGCTCCGTATCTCGTCTGTTTTCTTCAGCCTCTTCACCGCGGCGTTTCTTGCTTTTCTTGGCC CACCGTCGGCGTTGTCCGGCGACGGCAGTTCCCGGTGA	120	SEQ ID NO: 36 of WO 2016/050890
UGT1576 gene sequence [*S. grosvenorii*]	ATGGCTTCTCCTCGCCACACTCCTCACTTTCTGCTCTTCCCTTTCATGGCTCAAGGCCACATGA TCCCCATGATTGACCTTGCCAGGCTTCTGGCTCAGCGAGGAGTTATCATCACTATTATCACCAC GCCCCACAATGCTGCTCGCTACCACTCTGTTCTTGCTCGCGCCATCGATTCTGGGTTACACATC CATGTCCTCCAACTGCAGTTTCCATGTAAGGAAGGTGGGCTGCCAGAAGGGTGCGAGAATGTGG ACTTGCTACCTTCACTTGCTTCCATACCCAGATTCTACAGAGCAGCAAGTGATCTCCTTTACGA ACCATCTGAAAAACTGTTTGAGGAACTCATCCCCCGGCCGACCTGCATAATCTCCGATATGTGC CTGCCCTGGACCATGCGAATTGCTCTGAAATATCACGTCCCAAGGCTCGTTTTCTACAGTTTGA GCTGCTTCTTTCTTCTCTGTATGCGGAGTTTAAAAAACAATCTAGCGCTTATAAGCTCCAAGTC TGATTCTGAGTTCGTAACTTTCTCTGACTTGCCTGATCCAGTCGAGTTTCTCAAGTCGGAGCTA CCTAAATCCACCGATGAAGACTTGGTGAAGTTTAGTTATGAAATGGGGGAGGCCGATCGGCAGT CATACGGCGTTATTTTAAATCTATTTGAGGAGATGGAACCAAAGTATCTTGCAGAATATGAAAA GGAAAGAGAATCGCCGGAAAGAGTCTGGTGCGTCGGCCCAGTTTCGCTTTGCAACGACAACAAA CTCGACAAAGCTGAAAGAGGCAACAAAGCCTCCATCGACGAATACAAATGCATCAGGTGGCTCG ACGGGCAGCAGCCATCTTCGGTGGTTTACGTCTCTTTAGGAAGCTTGTGCAATCTGGTGACGGC GCAGATCATAGAGCTGGGTTTGGGTTTGGAGGCATCAAAGAAACCCTTCATTTGGGTCATAAGA AGAGGAAACATAACAGAGGAGTTACAGAAATGGCTTGTGGAGTACGATTTCGAGGAGAAAATTA AAGGGAGAGGGCTGGTGATTCTTGGCTGGGCTCCCCAAGTTCTGATACTGTCACACCCTGCAAT	121	SEQ ID NO: 47 of WO 2016050890

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CGGATGCTTTTTGACGCACTGCGGTTGGAACTCAAGCATCGAAGGGATATCGGCCGGCGTGCCA ATGGTCACCTGGCCGCTTTTTGCGGATCAAGTCTTCAACGAGAAGCTAATTGTACAAATACTCA GAATCGGCGTAAGTGTAGGCACGGAAACTACTATGAACTGGGGAGAGGAAGAGGAGAAAGGGGT GGTTGTGAAGAGAGAGAAAGTGAGGGAAGCCATAGAAATAGTGATGGATGGAGATGAGAGAGAA GAGAGGAGAGAGAGATGCAAAGAGCTTGCTGAAACGGCGAAGAGAGCTATAGAAGAAGGGGGCT CGTCTCACCGGAACCTCACGATGTTGATTGAAGATATAATTCATGGAGGAGGTTTGAGTTATGA GAAAGGAAGTTGTCGCTGA		
UGT SK98 gene sequence [*S. grosvenorii*]	ATGGATGCCCAGCGAGGTCACACCACCACCATTTTGATGCTTCCATGGGTCGGCTACGGCCATC TCTTGCCTTTCCTCGAGCTGGCCAAAAGCCTCTCCAGGAGGAAATTATTCCACATCTACTTCTG TTCAACGTCTGTTAGCCTCGACGCCATTAAACCAAAGCTTCCTCCTTCTATCTCTTCTGATGAT TCCATCCAACTTGTGGAACTTCGTCTCCCTTCTTCTCCTGAGTTACCTCCTCATCTTCACACAA CCAACGGCCTTCCCTCTCACCTCATGCCCGCTCTCCACCAAGCCTTCGTCATGGCCGCCCAACA CTTTCAGGTCATTTTACAAACACTTGCCCCGCATCTCCTCATTTATGACATTCTCCAACCTTGG GCTCCTCAAGTGGCTTCATCCCTCAACATTCCAGCCATCAACTTCAGTACTACCGGAGCTTCAA TGCTTTCTCGAACGCTTCACCCTACTCACTACCCAAGTTCTAAATTCCCAATCTCAGAGTTTGT TCTTCACAATCACTGGAGAGCCATGTACACCACCGCCGATGGGGCTCTTACAGAAGAAGGCCAC AAAATTGAAGAAACACTTGCGAATTGCTTGCATACTTCTTGCGGGGTAGTTTTGGTCAATAGTT TCAGAGAGCTTGAGACGAAA6TATATCGATTATCTCTCTGTTCTCTTGAACAAGAAAGTTGTTC CGGTCGGTCCTTTGGTTTACGAACCGAATCAAGAAGGGGAAGATGAAGGTTATTCAAGCATCAA AAATTGGCTTGACAAAAAGGAACCGTCCTCAACCGTCTTCGTTTCATTTGGAACCGAATACTTC CCGTCAAAGGAAGAAATGGAAGAGATAGCGTATGGGTTAGAGCTGAGCGAGGTTAATTTCATCT GGGTCCTTAGATTTCCTCAAGGAGACAGCACCAGCACCATTGAAGACGCCTTGCCGAAGGGGTT T CTGGAGAGAGCGGGAGAGAGGGCGATGGTGGTGAAGGGTTGGGCTCCTCAGGCGAAGATACTGA AGCATTGGAGCACAGGGGGGCTTGTGAGTCACTGTGGATGGAACTCGATGATGGAGGGCATGAT GTTTGGCGTACCCATAATAGCGGTCCCGATGCATCTGGACCAGCCCTTTAACGCCGGACTCTTG GAAGAAGCTGGCGTCGGCGTGGAAGCCAAGCGAGGTTCGGACGGCAAAATTCAAAGAGAAGAAG TTGCAAAGTCGATCAAAGAAGTGGTGATTGAGAAAACCAGGGAAGACGTGAGGAAGAAAGCAAG AGAAATGGGTGAGATTTTGAGGAGTAAAGGAGATGAGAAAATTGATGAGTTGGTGGCTGAAATT TCTCTTTTGCGCAAAAAGGCTCCATGTTCAATTTAA	122	SEQ ID NO: 49 of WO 21589
S grosvenorii UG98 gene sequence	ATGGATGCCCAGCGAGGTCACACCACAACCATTTTGATGTTTCCATGGCTCGGCTATGGCCATC TTTCGGCTTTCCTAGAGTTGGCCAAAAGCCTCTCAAGGAGGAACTTCCATATCTACTTCTGTTC AACCTCTGTTAACCTCGACGCCATTAAACCAAAGCTTCCTTCTTCTTCCTCTTCTGATTCCATC CAACTTGTGGAACTTTGTCTTCCATCTTCTCCTGATCAGCTCCCTCCTCATCTTCACACAACCA ACGCCCTCCCCCCTCACCTCATGCCCACTCTCCACCAAGCCTTCTCCATGGCTGCCCAACACTT TGCTGCCATTTTACACACACTTGCTCCGCATCTCCTCATTTACGACTCTTTCCAACCTTGGGCT CCTCAACTAGCTTCATCCCTCAACATTCCAGCCATCAACTTCAATACTACGGGAGCTTCAGTCC TGACCCGAATGCTTCACGCTACTCACTACCCAAGTTCTAAATTCCCAATTTCAGAGTTTGTTCT	123	SEQ ID NO: 51 of WO 2001606050089 00

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CCACGATTATTGGAAAGCCATGTACAGCGCCGCCGGTGGGGCTGTTACAAAAAAAGACCACAAA ATTGGAGAAACACTTGCGAATTGCTTGCATGCTTCTTGTAGTGTAATTCTAATCAATAGTTTCA GAGAGCTCGAGGAGAAATATATGGATTATCTCTCCGTTCTCTTGAACAAGAAAGTTGTTCCGGT TGGTCCTTTGGTTTACGAACCGAATCAAGACGGGGAAGATGAAGGTTATTCAAGCATCAAAAAT TGGCTTGACAAAAAGGAACCGTCCTCCACCGTCTTCGTTTCATTTGGAAGCGAATACTTCCCGT CAAAGGAAGAAATGGAAGAGATAGCCCATGGGTTAGAGGCGAGCGAGGTTCATTTCATCTGGGT CGTTAGGTTTCCTCAAGGAGACAACACCAGCGCCATTGAAGATGCCTTGCCGAAGGGGTTTCTG GAGAGGGTGGGAGAGAGAGGGATGGTGGTGAAGGGTTGGGCTCCTCAGGCGAAGATACTGAAGC ATTGGAGCACAGGGGGATTCGTGAGCCACTGTGGATGGAACTCGGTGATGGAAAGCATGATGTT TGGCGTTCCCATAATAGGGGTTCCGATGCATCTGGACCAGCCCTTTAACGCCGGACTCGCGGAA GAAGCTGGCGTCGGCGTGGAAGCCAAGCGAGATTCGGACGGCAAAATTCAAAGAGAAGAAGTTG CAAAGTCGATCAAAGAAGTGGTGATTGAGAAAACCAGGGAAGACGTGAGGAAGAAAGCAAGAGA AATGGGTGAGATTTTGAGGAGTAAAGGAGATGAGAAAATTGATGAGTTGGTGGCTGAAATTTCT CTTTTGCGCAAAAAGGCTCCATGTTCAATTTAA		
Codon optimized coding sequence for UGT SK98	CATTTGTCTGCTTTTTTGGAATTGGCCAAGTCCTTGTCTAGAAGAAACTTCCATATCTACTTTT GCTCCACCTCCGTTAATTTGGATGCTATTAAGCCAAAGTTGCCATCCTCTTCATCCTCCGATTC TATTCAATTGGTTGAATTGTGCTTGCCATCTTCCCCAGATCAATTGCCACCACACTTGCATACA ACTAATGCTTTACCACCACATTTGATGCCAACATTGCATCAAGCTTTTTCTATGGCTGCTCAAC ATTTTGCTGCTATCTTGCATACTTGGCTCCTCATTTGTTGATCTACGATTCTTTTCAACCATGG GCTCCACAATTGGCTTCATCTTTGAATATTCCAGCCATCAACTTCAACACTACTGGTGCTTCAG TTTTGACCAGAATGTTGCATGCTACTCATTACCCA	124	SEQ ID NO: 52 of WO 2016050890
UGT protein also referred to as UGT94A9, A9 or UGT94-289-1	MDAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRRNFHIYFCSTSVNLDAIKPKLPSSSSSDSI QLVELCLPSSPDQLPPHLHTTNALPPHLMPTLHQAFSMAAQHFAAILHTLAPHLLIYDSFQPWA PQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFPISEFVLHDYWKAMYSAAGGAVTKKDHK IGETLANCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIKN WLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAIEDALPKGFL ERVGERGMVVKGWAPQAKILKHWSTGGFVSHCGWNSVMESMMFGVPIIGVPMHLDQPFNAGLAE EAGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEKMDEMVAAIS LFLKI	125	Seq ID NO: 6 of WO 2016038617
UGT protein is also referred to as EH1, EPH1 and contig 73966.	MENIEHTTVQTNGIKMHVAAIGTGPPVLLLHGFPELWYSWRHQLLYLSSAGYRAIAPDLRGYGD TDAPPSPSSYTALHIVGDLVGLLDVLGIEKVFLIGHDWGAIIAWYFCLFRPDRIKALVNLSVQF FPRNPTTPFVKGFRAVLGDQFYMVRFQEPGKAEEEFASVDIREFFKNVLSNRDPQAPYLPNEVK FEGVPPPALAPWLTPEDIDVYADKFAETGFTGGLNYYRAFDRTWELTAPWTGARIGVPVKFIVG DLDLTYHFPGAQKYIHGEGFKKAVPGLEEVVVMEDTSHFINQERPHEINSHIHDFFSKFC	126	Seq ID NO: 18 in WO2016038617
UCT85E5 gene coding sequence	ATGGTGCAACCTCGGGTACTGCTGTTTCCTTTCCCGGCACTGGGCCACGTGAAGCCCTTCTTAT CACTGGCGGAGCTGCTTTCCGACGCCGGCATAGACGTCGTCTTCCTCAGCACCGAGTATAACCA CCGTCGGATCTCCAACACTGAAGCCCTAGCCTCCCGCTTCCCGACGCTTCATTTCGAAACTATA CCGGATGGCCTGCCGCCTAATGAGTCGCGCGCTCTTGCCGACGGCCCACTGTATTTCTCCATGC	127	Seq ID NO: 33 of WO2016038617

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GTGAGGGAACTAAACCGAGATTCCGGCAACTGATTCAATCTCTTAACGACGGTCGTTGGCCCAT CACCTGTATTATCACTGACATCATGTTATCTTCTCCGATTGAAGTAGCGGAAGAATTTGGGATT CCAGTAATTGCCTTCTGCCCCTGCAGTGCTCGCTACTTATCGATTCACTTTTTTATACCGAAGC TCGTTGAGGAAGGTCAAATTCCATACGCAGATGACGATCCGATTGGAGAGATCCAGGGGGTGCC CTTGTTCGAAGGTCTTTTGCGACGGAATCATTTGCCTGGTTCTTGGTCTGATAAATCTGCAGAT ATATCTTTCTCGCATGGCTTGATTAATCAGACCCTTGCAGCTGGTCGAGCCTCGGCTCTTATAC TCAACACCTTCGACGAGCTCGAAGCTCCATTTCTGACCCATCTCTCTTCCATTTTCAACAAAAT CTACACCATTGGACCCCTCCATGCTCTGTCCAAATCAAGGCTCGGCGACTCCTCCTCCTCCGCT TCTGCCCTCTCCGGATTCTGGAAAGAGGATAGAGCCTGCATGTCCTGGCTCGACTGTCAGCCGC CGAGATCTGTGGTTTTCGTCAGTTTCGGGAGTACGATGAAGATGAAAGCCGATGAATTGAGAGA GTTCTGGTATGGGTTGGTGAGCAGCGGGAAACCGTTCCTCTGCGTGTTGAGATCCGACGTTGTT TCCGGCGGAGAAGCGGCGGAATTGATCGAACAGATGGCGGAGGAGGAGGGAGCTGGAGGGAAGC TGGGAATGGTAGTGGAGTGGGCAGCGCAAGAGAAGGTCCTGAGCCACCCTGCCGTCGGTGGGTT TTTGACGCACTGCGGGTGGAAC TCAACGGTGGAAAGCATTGCCGCGGGAGTTCCGATGATGTGCTGGCCGATTCTCGGCGACCAAC CCAGCAACGCCACTTGGATCGACAGAGTGTGGAAAATTGGGGTTGAAAGGAACAATCGTGAATG GGACAGGTTGACGGTGGAGAAGATGGTGAGAGCATTGATGGAAGGCCAAAAGAGAGTGGAGATT CAGAGATCAATGGAGAAGCTTTCAAAGTTGGCAAATGAGAAGGTTGTCAGGGGTGGGTTGTCTT TTGATAACTTGGAAGTTCTCGTTGAAGACATCAAAAAATTGAAACCATATAAATTTTAA		
UGT protein	MVQPRVLLFPFPALGHVKPFLSLAELLSDAGIDVVFLSTEYNHRRISNTEALASRFPTLHFETI PDGLPPNESRALADGPLYFSMREGTKPRFRQLIQSLNDGRWPITCIITDIMLSSPIEVAEEFGI PVIAFCPCSARYLSIHFFIPKLVEEGQIPYADDDPIGEIQGVPLFEGLLRRNHLPGSWSDKSAD ISFSHGLINQTLAAGRASALILNTFDELEAPFLTHLSSIFNKIYTIGPLHALSKSRLGDSSSSA SALSGFWKEDRACMSWLDCQPPRSVVFVSFGSTMKMKADELREFWYGLVSSGKPFLCVLRSDVV SGGEAAELIEQMAEEEGAGGKLGMVVEWAAQEKVLSHPAVGGFLTHCGWNSTVESIAAGVPMMC WPILGDQPSNATWIDRVWKIGVERNNREWDRLTVE	128	Seq ID NO: 34 of WO2016038617
UGT protein referred to as UGT94C9 and UGT94-289-3	MDAAQQGDTTTILMLPWLGYGHLSAFLELAKSLSRRNFHIYFCSTSVNLDAIKPKLPSSFSDSI QFVELHLPSSPEFPPHLHTTNGLPPTLMPALHQAFSMAAQHFESILQTLAPHLLIYDSLQPWAP RVASSLKIPAINFNTTGVFVISQGLHPIHYPHSKFPFSEFVLHNHWKAMYSTADGASTERTRKR GEAFLYCLHASCSVILINSFRELEGKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIKNW LDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVNFIWVVRFPQGDNTSGIEDALPKGFLE RAGERGMVVKGWAPQAKILKHWSTGGFVSHCGWNSVMESMMFGVPIIGVPMHVDQPFNAGLVEE AGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEKFDEMVAEISL LLKI	129	Seq ID NO: 38 of WO2016038617
Coding sequence for enzymes for acetyl CoA synthesis	TCACAACGTTTAGCTTTGACAGATGGTATCGCCGTGACAGACGGGGTCAAAGAAGTTTGGCTTT CATAAAAAGAGCTAGACATTCAAGCAAAAACTTGAATGTCTAGCTCTTTTGTTGATTGAATCGG GGGATTTAAATACTTAGTTTCGATAAGAGCGAACGGTATTATTAATAGCAGAAATACTATATTT TAATTCATCTTCTAACGTTTGATCAATATCTGTGTCTAAAGTTTCTGCAAACATGGCTTCATAT	130	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TCAGCGATAGAAAGTTCTGTCCGATTATCTAGCAGTGCTAAATGAGTTTCTTTTTGTAAATGAT TTTGATAACCAGCTACTAATTCACCAGTGAAAAATTCAGCGACAGCACCAGAACCATAACTGAA TAACCCAATTTGATTGCCTGCGGTTAAAGTCGTTGCATTTTCTAAAAGGGAAATGAGTCCCAGA TAAAGTGAACCCGTATACAAGTTTCCTACGCGACGACTATAGATGATGCTTTCTTCATAACGGG CTAAAATTCGTTCCTGTTCTGCTTCAGTTTGGTCGGAGATTTTTGCTAATAAGGCTTTTTTGCC CATTTTTGTGTAAGGAATATGGAACGCTAAAGCATCATAATCTGCAAAATCAAGACCGGTTCTT TTTTTATGTTCATCCCAGACTTGGGCAAAAGATTGGATGTAGGTTTCGTTTGACAAAGGACCAT CGACCATAGGATACGGATGGCCTGTTGGACGCCAAAAGTCATAGATATCTTGCGTCAGCATCAC ATTATCCTCTTTTAAAGCCAAGATGCGCGGTTCACTAGCAACTAACATTGCAACCGCCCCAGCT CCTTGTGTAGGCTCACCGCCAGAATTTAATCCATATTTTGCAATATCTGCTGCTACAACCAAGA CTTTTTTATCTGGATGTAAGGCTACGTGATTCTTAGCTAACTGTAAGCCTGCTGTTGCTCCGTA ACAAGCTTCCTTGATTTCGAAAGAGCGAGCGAAAGGTTGAATCCCCATTAAACGATGTAAGACA ACTGCGGCCGCTTTTGACTCATCGATACTGGACTCAGTCCCGACAATCACCATATCAATGGCCT CTTTATCTTCTTTGGTCAAGATCGCTTCTGCGGCATTGGCTGCAAATGTCACAATATCTTGGCT GATTGGGTTCACCGCCATTTGGTCTTGCCCAATACCAATATGAAATTTTCCAGGGTCTACATTT CTGGCTTCAGCCAGTGCCGTCATATCAATATAATAAGGGGGCACAAAAAAACTAATTTTATCAA TCCCAATTGTCATTTCTTTAACTCCTTTACGATAAATAGATTCATTATATAAAATAGCACGAAA TGAACCAAAATGGGGAATTTTTGTATTAACTTCATAGATTATTAAAAAATATCTTATAAGTCTG TTAACATTCAGTAATTGGCACTTGTGATTCTGGGATTTTATGATATATTTCAAGATGGAGGTGC ATTTAGTTGAAAACAGTAGTTATTATTGATGCATTACGAACACCAATTGGAAAATATAAAGGCA GCTTAAGTCAAGTAAGTGCCGTAGACTTAGGAACACATGTTACAACACAACTTTTAAAAAGACA TTCCACTATTTCTGAAGAAATTGATCAAGTAATCTTTGGAAATGTTTTACAAGCTGGAAATGGC CAAAATCCCGCACGACAAATAGCAATAAACAGCGGTTTATCTCATGAAATTCCCGCAATGACAG TTAATGAGGTCTGCGGATCAGGAATGAAGGCCGTTATTTTGGCGAAACAATTGATTCAATTAGG AGAAGCGGAAGTTTTAATTGCTGGCGGGATTGAGAATATGTCCCAAGCACCTAAATTACAACGA TTTAATTACGAAACAGAAAGCTATGATGCGCCTTTTTCTAGTATGATGTACGATGGGTTAACGG ATGCCTTTAGTGGTCAAGCAATGGGCTTAACTGCTGAAAATGTGGCCGAAAAGTATCATGTAAC TAGAGAAGAGCAAGATCAATTTTCTGTACATTCACAATTAAAAGCAGCTCAAGCACAAGCAGAA GGGATATTCGCTGACGAAATAGCCCCATTAGAAGTATCAGGAACGCTTGTGGAGAAAGATGAAG GGATTCGCCCTAATTCGAGCGTTGAGAAGCTAGGAACGCTTAAAACAGTTTTTAAAGAAGACGG TACTGTAACAGCAGGGAATGCATCAACCATTAATGATGGGGCTTCTGCTTTGATTATTGCTTCA CAAGAATATGCCGAAGCACACGGTCTTCCTTATTTAGCTATTATTCGAGACAGTGTGGAAGTCG GTATTGATCCAGCCTATATGGGAATTTCGCCGATTAAAGCCATTCAAAAACTGTTAGCGCGCAA TCAACTTACTACGGAAGAAATTGATCTGTATGAAATCAACGAAGCATTTGCAGCAACTTCAATC GTGGTCCAAAGAGAACTGGCTTTACCAGAGGAAAAGGTCAACATTTATGGTGGCGGTATTTCAT TAGGTCATGCGATTGGTGCCACAGGTGCTCGTTTATTAACGAGTTTAAGTTATCAATTAAATCA AAAAGAAAAGAAATATGGAGTGGCTTCTTTATGTATCGGCGGTGGCTTAGGACTCGCTATGCTA CTAGAGAGACCTCAGCAAAAAAAAAACAGCCGATTTATCAAATGAGTCCTGAGGAACGCCTGGC		

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TTCTCTTCTTAATGAAGGCCAGATTTCTGCTGATACAAAAAAAGAATTTGAAAATACGGCTTTA TCTTCGCAGATTGCCAATCATATGATTGAAAATCAAATCAGTGAAACAGAAGTGCCGATGGGCG TTGGCTTACATTTAACAGTGGACGAAACTGATTATTTGGTACCAATGGCGACAGAAGAGCCCTC AGTGATTGCGGCTTTGAGTAATGGTGCAAAAATAGCACAAGGATTTAAAACAGTGAATCAACAA CGTTTAATGCGTGGACAAATCGTTTTTTACGATGTTGCAGACGCCGAGTCATTGATTGATGAAC TACAAGTAAGAGAAACGGAAATTTTTCAACAAGCAGAGTTAAGTTATCCATCTATCGTTAAACG CGGCGGCGGCTTAAGAGATTTGCAATATCGTGCTTTTGATGAATCATTTGTATCTGTCGACTTT TTAGTAGATGTTAAGGATGCAATGGGGGCAAATATCGTTAACGCTATGTTGGAAGGTGTGGCCG AGTTGTTCCGTGAATGGTTTGCGGAGCAAAAGATTTTATTCAGTATTTTAAGTAATTATGCCAC GGAGTCGGTTGTTACGATGAAAACGGCTATTCCAGTTTCACGTTTAAGTAAGGGGAGCAATGGC CGGGAAATTGCTGAAAAAATTGTTTTAGCTTCACGCTATGCTTCATTAGATCCTTATCGGGCAG TCACGCATAACAAAGGGATCATGAATGGCATTGAAGCTGTCGTTTTAGCTACAGGAAATGATAC ACGCGCTGTTAGCGCTTCTTGTCATGCTTTTGCGGTGAAGGAAGGTCGCTACCAAGGTTTGACT AGTTGGACGCTGGATGGCGAACAACTAATTGGTGAAATTTCAGTTCCGCTTGCGTTAGCCACGG TTGGCGGTGCCACAAAAGTCTTACCTAAATCTCAAGCAGCTGCTGATTTGTTAGCAGTGACGGA TGCAAAAGAACTAAGTCGAGTAGTAGCGGCTGTTGGTTTGGCACAAAATTTAGCGGCGTTACGG GCCTTAGTCTCTGAAGGAATTCAAAAAGGACACATGGCTCTACAAGCACGTTCTTTAGCGATGA CGGTCGGAGCTACTGGTAAAGAAGTTGAGGCAGTCGCTCAACAATTAAAACGTCAAAAAACGAT GAACCAAGACCGAGCCTTGGCTATTTTAAATGATTTAAGAAAACAATAAAAAAACAGTTCAGCA GAAATTATTCTGCTGAACTGTTTTTTTTCACATTAGGTAGCCGTTTCAGGCCACGAATTGGTTT TACTTTTAAGACATCTAAGAAGAAAGTGAA		
CGT-SL glucotransferases AAD00555.1	MKRWLSVVLSMSLVFSAFFLVSDTQKVTVEAAGNLNKVNFTSDIVYQIVVDRFVDGNTSNNPSG SLFSSGCTNLRKYCGGDWQGIINKINDGYLTEMGVTAIWISQPVENVFAVMNDADGSTSYHGYW ARDFKKTNPFFGTLSDFQRLVDAAHAKGIKVIIDFAPNHTSPASETNPSYMENGRLYDNGTLIG GYTNDTNSYFHHNGGTTFSNLEDGIYRNLFDLADFNHQNQFIDKYLKDAIKLWLDMGIDGIRMD AVKHMPFGWQKSFMDEVYDYRPVFTFGEWFLSENEVDSNNHFFANESGMSLLDFRFGQKLRQVL RNNSDDWYGFNQMIQDTASAYDEVIDQVTFIDNHDMDRFMADEGDPRKVDIALAVLLTSRGVPN IYYGTEQYMTGNGDPNNRKMMTSFNKNTRAYQVIQKLSSLRRSNPALSYGDTEQRWINSDVYIY ERQFGKDVVLVAVNRSLSKSYSITGLFTALPSGTYTDQLGALLDGNTIQVGSNGAVNAFNLGPG EVGVWTYSAAESVPIIGHIGPMMGQVGHKLTIDGEGFGTNVGTVKFGNTVASVVSWSNNQITVT VPNIPAGKYNITVQTSGGQVSAAYDNFEVLTNDQVSVRFVVNNANTNWGENIYLVGNVHELGNW NTSKAIGPLFNQVIYSYPTWYVDVSVPEGKTIEFKFIKKDGSGNVIWESGSNHVYTTPTSTTGT VNVNWQY	148	
CGT-SL glucotransferases KMY60644.1	MSRNGAVTPDWQFTVEVQEGETITYKYVKGGSWDQEGLADHTREDDNDDDVSYYGYGAIGTDLK VTVHNEGNNTMIVQDRILRWIDMPVVIEEVQKQGSQVTIKGNAIKNGVLTINGERVPIDGRMAF SYTFTPASHQKEVSIHIEPSAESKTAIFNNDGGAIAKNTKDYVLNLETKQLREGKLTTPPSNGD SPESDWPGSETPSHDGGATPGNGTSPGSGGPSDGTSPGGSVPPGGTAPPGNEAPPSRPPQKPSP SKPKEKPRKPTTPPGQVKKVYWDGVELKKGQIGRLTVQKPINLWKRTKDGRLVFVRILQPGEVY	154	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	RVYGYDVRFGGQYAVGGGYYVTDIDTHIRYETPSKEKLKLVNGE		
DexT protein [*Leuconostoc citreum*]	MPANAPDKQSVTNAPVVPPKHDTDQQDDSLEKQQVLEPSVNSNIPKKQTNQQLAVVTAPANSAP QTKTTAEISAGTELDTMPNVKHVDGKVYFYGDDGQPKKNFTTIIDGKPYYFDKDTGALSNNDKQ YVSELFSIGNKHNAVYNTSSDNFTQLEGHLTASSWYRPKDILKNGKRWAPSTVTDFRPLLMAWW PDKSTQVTYLNYMKDQGLLSGTHHFSDNENMRTLTAAAMQAQVNIEKKIGQLGNTDWLKTAMTQ YIDAQPNWNIDSEAKGDDHLQGGALLYTNSDMSPKANSDYRKLSRTPKNQKGQIADKYKQGGFE LLLANDVDNSNPVVQAEQLNWLHYMMNIGSILQNDDQANFDGYRVDAVDNVDADLLQIAGEYAK AAYGVDKNDARANQHLSILEDWGDEDPDYVKAHGNQQITMDFPLHLAIKYALNMPNDKRSGLEP TREHSLVKRITDDKENVAQPNYSFIRAHDSEVQTIIADIIKDKINPASTGLDSTVTLDQIKQAF DIYNADELKADKVYTPYNIPASYALLLTNKDTIPRVYYGDMFTDDGQYMAKQSPYYQAIDALLK ARIKYAAGGQTMKMNYFPDEQSVMTSVRYGKGAMTASDSGNQETRYQGIGLVVNNRPDLKLSDK DEVKMDMGAAHKNQDYRPVLLTTKSGLKVYSTDANAPVVRTDANGQLTFKADMVYGVNDPQVSG YIAAWVPVGASENQDARTKSETTQSTDGSVYHSNAALDSQVIYEGFSNFQDFPTTPDEFTNIKI AQNVNLFKDWGITSFEMAPQYRASSDKSFLDAIVQNGYAFTDRYDIGYNTPTKYGTADNLLDAL RALHGQGIQAINDWVPDQIYNLPDEQLVTAIRTDGSGDHTYGSVIDHTLYASKTVGGGIYQQQY GGAFLEQLKTQYPQLFQQKQISTDQPMNPDIQIKSWEAKYFNGSNIQGRGAWYVLKDWGTQQYF NVSDAQTFLPKQLLGEKAKTGFVTRGKETSFYSTSGYQAKSAFICDNGNWYYFDDKGKMVVGNQ VINGINYYFLPNGIELQDAYLVHDGMYYYYNNIGKQLHNTYYQDKQKNFHYFFEDGHMAQGIVT IIQSDGTPVTQYFDENGKQQKGVAVKGSDGHLHYFDGASGNMLFKSWGRLADGSWLYVDEKGNA VTGKQTINNQTVYFNDDGRQIKNNFKELADGSWLYLNNKGVAVTGEQIINGQTLYFGNDGRQFK GTTHINATGESRYYDPDSGNMITDRFERVGDNQWAYFGYDGVAVTGDRIIKGQKLYFNQNGIQM KGHLRLENGIMRYYDADTGELVRNRFVLLSDGSWVYFGQDGVPVTGVQVINGQTLYFDADGRQV KGQQRVIGNQRYWMDKDNGEMKKITYAAALE	156	
DexT gene (coding sequence)	ATGCCAGCAAATGCCCCAGATAAACAATCAGTGACTAATGCACCAGTAGTGCCGCCAAAGCATG ATACGGACCAGCAGGACGATTCACTAGAAAAACAGCAAGTATTAGAACCGAGCGTAAATAGTAA TATACCAAAAAAGCAGACAAATCAACAGTTAGCGGTTGTTACAGCACCAGCAAATTCAGCACCT CAAACCAAAACAACAGCAGAAATTTCTGCTGGTACAGAGTTAGACACGATGCCTAATGTTAAGC ATGTAGATGGCAAAGTTTATTTTTATGGAGATGATGGCCAACCAAAAAAGAATTTTACTACTAT TATAGATGGTAAACCTTACTACTTTGATAAAGATACAGGGGCACTATCTAATAACGATAAGCAA TATGTATCGGAATTATTCAGTATTGGCAATAAACATAACGCCGTCTATAACACATCATCAGATA ATTTTACGCAATTAGAAGGACATCTGACGGCAAGTAGTTGGTATCGTCCAAAAGATATTTTGAA AAATGGTAAACGTTGGGCACCTTCAACAGTGACTGATTTCAGACCATTATTGATGGCCTGGTGG CCGGATAAGAGTACGCAAGTCACTTATCTGAATTACATGAAAGATCAGGGCCTCTTGTCTGGTA CTCATCACTTTTCCGATAATGAAAATATGCGGACCTTAACGGCAGCTGCCATGCAGGCACAGGT AAACATTGAGAAAAAAATTGGGCAACTTGGCAATACGGATTGGTTGAAAACGGCGATGACGCAA TACATTGATGCCCAGCCCAATTGGAATATTGACAGTGAGGCGAAAGGAGATGATCATCTACAAG GTGGTGCACTACTTTATACAAATAGTGATATGTCGCCAAAGGCCAATTCTGATTATCGTAAGCT GAGCCGTACGCCTAAAAATCAAAAAGGTCAAATTGCTGATAAATATAAGCAAGGTGGGTTTGAA	157	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TTATTACTAGCAAACGATGTCGATAATTCTAATCCAGTTGTGCAAGCAGAACAACTTAATTGGT TACATTATATGATGAATATCGGTAGTATTTTACAAAATGATGACCAAGCTAATTTTGATGGTTA CCGTGTTGATGCTGTCGATAATGTGGACGCTGACTTACTACAGATTGCTGGTGAATATGCTAAG GCTGCCTATGGTGTTGACAAAAATGACGCGAGAGCGAATCAACATTTATCAATTTTGGAAGACT GGGGAGATGAAGATCCAGACTATGTCAAAGCACATGGCAACCAGCAAATTACAATGGATTTCCC CTTGCATTTAGCGATTAAATACGCGCTCAACATGCCTAATGATAAGCGGAGTGGCCTTGAGCCA ACCCGTGAACACAGTTTAGTCAAACGAATTACAGATGATAAAGAAAATGTTGCACAACCAAATT ATTCATTTATCCGAGCTCATGACAGTGAAGTACAAACGATTATTGCTGATATTATTAAAGATAA AATCAACCCGGCGTCAACAGGGCTAGATTCAACAGTGACTTTGGATCAAATTAAGCAGGCTTTT GACATCTATAATGCTGATGAATTGAAAGCAGATAAAGTTTACACACCTTACAATATTCCAGCAT CATACGCTTTGTTATTGACTAATAAAGACACAATTCCACGTGTTTATTATGGGGATATGTTCAC GGATGATGGCCAATACATGGCTAAACAATCACCTTACTATCAAGCGATTGATGCGTTGTTGAAA GCTCGTATCAAGTATGCTGCTGGTGGTCAAACCATGAAAATGAACTATTTTCCAGATGAACAAT CTGTTATGACATCAGTTCGTTATGGTAAGGGTGCAATGACGGCAAGTGACTCTGGTAACCAAGA GACACGCTATCAAGGTATTGGACTTGTTGTCAACAATCGCCCAGATTTGAAACTATCTGACAAA GATGAAGTCAAAATGGATATGGGTGCGGCACATAAAAACCAAGATTATCGCCCAGTTTTGTTGA CGACAAAATCAGGATTAAAAGTCTACAGCACTGATGCAAATGCACCTGTCGTTCGAACTGACGC CAATGGCCAATTAACTTTTAAGGCAGACATGGTATATGGTGTAAACGACCCACAAGTGTCAGGG TACATTGCGGCTTGGGTACCAGTAGGGGCTTCAGAAAATCAAGATGCTCGAACGAAAAGTGAAA CAACGCAGTCAACTGACGGGAGTGTTTATCATTCTAATGCAGCGTTAGATTCGCAAGTCATTTA TGAAGGCTTTTCAAATTTTCAAGACTTTCCAACAACACCCGATGAGTTTACGAACATTAAAATT GCTCAAAATGTTAACTTATTTAAGGATTGGGGTATTACTAGCTTTGAAATGGCGCCACAATATC GCGCCAGCTCAGATAAAAGTTTCTTAGATGCTATCGTACAAAATGGTTATGCATTTACAGATCG ATATGATATTGGTTACAACACACCAACAAAGTATGGGACAGCAGATAATTTGTTAGATGCTTTA CGTGCATTGCATGGTCAGGGTATTCAAGCGATTAACGACTGGGTACCAGATCAAATTTATAATC TACCCGATGAACAGTTAGTCACGGCTATTCGAACAGACGGTTCAGGTGATCATACTTATGGTTC AGTTATTGACCATACTTTGTATGCATCAAAGACAGTTGGCGGGGGCATTTATCAGCAACAATAT GGTGGGGCCTTCTTGGAACAATTAAAAACACAGTACCCGCAACTTTTCCAGCAAAAACAGATTT CCACAGATCAGCCAATGAACCCAGATATTCAAATTAAGTCATGGGAAGCCAAGTATTTCAACGG TTCGAACATTCAGGGGCGTGGGGCTTGGTATGTTTTGAAGGACTGGGGCACACAACAGTATTTT AATGTGTCAGATGCGCAGACCTTCCTTCCAAAGCAATTATTGGGTGAAAAGGCCAAAACTGGTT TTGTTACGCGTGGTAAGGAGACTTCATTCTATTCCACTAGTGGCTATCAAGCAAAATCTGCCTT TATTTGTGATAACGGTAATTGGTACTACTTTGATGACAAAGGGAAAATGGTTGTTGGAAACCAA GTTATCAATGGCATCAATTATTACTTTTTACCGAATGGTATCGAATTACAAGATGCCTATCTAG TACATGATGGTATGTACTATTATTATAATAATATTGGCAAGCAACTGCACAACACATATTACCA AGATAAACAAAAAAATTTCCATTACTTCTTTGAAGATGGGCACATGGCACAGGGTATTGTCACC ATCATTCAAAGTGATGGCACCCCAGTCACACAGTACTTTGATGAGAATGGTAAGCAACAAAAAG GCGTGGCGGTCAAAGGATCAGATGGTCATTTGCATTACTTTGACGGTGCGTCAGGGAATATGCT		

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTTTAAATCATGGGGTAGACTAGCAGATGGCTCTTGGCTATATGTAGACGAGAAAGGTAATGCG GTTACAGGCAAACAAACCATTAATAATCAAACGGTTTACTTTAATGATGATGGTCGTCAAATCA AAAATAACTTTAAAGAATTAGCAGATGGTTCTTGGCTTTATCTTAACAATAAAGGTGTTGCAGT AACAGGAGAGCAAATAATTAATGGGCAGACACTTTATTTTGGTAACGATGGTCGTCAATTTAAA GGGACAACACATATAAATGCTACTGGTGAAAGCCGTTACTATGACCCAGACTCAGGTAATATGA TAACTGATCGTTTTGAACGTGTTGGTGATAATCAATGGGCTTATTTTGGTTATGATGGTGTTGC AGTAACAGGGGACCGAATCATTAAAGGGCAAAAACTCTATTTCAACCAAAATGGTATCCAAATG AAAGGCCACTTACGTCTTGAAAATGGTATCATGCGTTATTACGATGCTGATACTGGCGAATTAG TTCGTAATCGATTTGTATTGCTATCTGATGGTTCATGGGTTTACTTTGGCCAAGATGGCGTACC CGTAACTGGCGTGCAAGTGATTAATGGCCAAACATTATATTTTGACGCAGATGGTAGGCAAGTC AAAGGGCAGCAACGTGTAATCGGCAATCAACGCTATTGGATGGATAAAGACAATGGTGAAATGA AAAAAATAACATACGCGGCCGCACTCGAGCACCACCACCACCACCACTGA		
DexT gene (coding sequence is cloned into pET23a)	ATGCAAAACGGCGAAGTGTGTCAGCGTAAAAAACTGTACAAGTCAGGGAAGATATTAGTTACAG CAAGTATTTTTGCTGTTATGGGTTTTGGTACTGCCATGTCACAAGCAAACGCGAGCAGTAGTGA TAATGATAGCAAAACACAAACTATTTCAAAAATAGTAAAAAGTAAAGTCGAACCGGCAACTGTT CAACCAGCGAAACCAGCGGAACCTACTAATAAAATAGTTGACCAAGCAGATATGCATACGGTCA GCGGGCAAAACAGCGTGCCACCAGTAGTGACTAATCAATCCAATTAACAGGCTGCAAAACCAAC TACACCTGTTACCGATGTCACAGATACGCATAAAATCGAAGCAAACAACGTCCCTGCTGATGTT ATGCCAGCAAATGCCCCAGATAAACAATCAGTGACTAATGCACCAGTAGTGCCGCCAAAGCATG ATACGGACCAGCAGGACGATTCACTAGAAAAACAGCAAGTATTAGAACCGAGCGTAAATAGTAA TATACCAAAAAAGCAGACAAATCAACAGTTAGCGGTTGTTACAGCACCAGCAAATTCAGCACCT CAAACCAAAACAACAGCAGAAATTTCTGCTGGTACAGAGTTAGACACGATGCCTAATGTTAAGC ATGTAGATGGCAAAGTTTATTTTTATGGAGATGATGGCCAACCAAAAAAGAATTTTACTACTAT TATAGATGGTAAACCTTACTACTTTGATAAAGATACAGGGGCACTATCTAATAACGATAAGCAA TATGTATCGGAATTATTCAGTATTGGCAATAAACATAACGCCGTCTATAACACATCATCAGATA ATTTTACGCAATTAGAAGGACATCTGACGGCAAGTAGTTGGTATCGTCCAAAAGATATTTTGAA AAATGGTAAACGTTGGGCACCTTCAACAGTGACTGATTTCAGACCATTATTGATGGCCTGGTGG CCGGATAAGAGTACGCAAGTCACTTATCTGAATTACATGAAAGATCAGGGCCTCTTGTCTGGTA CTCATCACTTTTCCGATAATGAAAATATGCGGACCTTAACGGCAGCTGCCATGCAGGCACAGGT AAACATTGAGAAAAAAATTGGGCAACTTGGCAATACGGATTGGTTGAAAACGGCGATGACGCAA TACATTGATGCCCAGCCCAATTGGAATATTGACAGTGAGGCGAAAGGAGATGATCATCTACAAG GTGGTGCACTACTTTATACAAATAGTGATATGTCGCCAAAGGCCAATTCTGATTATCGTAAGCT GAGCCGTACGCCTAAAAATCAAAAAGGTCAAATTGCTGATAAATATAAGCAAGGTGGGTTTGAA TTATTACTAGCAAACGATGTCGATAATTCTAATCCAGTTGTGCAAGCAGAACAACTTAATTGGT TACATTATATGATGAATATCGGTAGTATTTTACAAAATGATGACCAAGCTAATTTTGATGGTTA CCGTGTTGATGCTGTCGATAATGTGGACGCTGACTTACTACAGATTGCTGGTGAATATGCTAAG GCTGCCTATGGTGTTGACAAAAATGACGCGAGAGCGAATCAACATTTATCAATTTTGGAAGACT GGGGAGATGAAGATCCAGACTATGTCAAAGCACATGGCAACCAGCAAATTACAATGGATTTCCC	158	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTTGCATTTAGCGATTAAATACGCGCTCAACATGCCTAATGATAAGCGGAGTGGCCTTGAGCCA ACCCGTGAACACAGTTTAGTCAAACGAATTACAGATGATAAAGAAAATGTTGCACAACCAAATT ATTCATTTATCCGAGCTCATGACAGTGAAGTACAAACGATTATTGCTGATATTATTAAAGATAA AATCAACCCGGCGTCAACAGGGCTAGATTCAACAGTGACTTTGGATCAAATTAAGCAGGCTTTT GACATCTATAATGCTGATGAATTGAAAGCAGATAAAGTTTACACACCTTACAATATTCCAGCAT CATACGCTTTGTTATTGACTAATAAAGACACAATTCCACGTGTTTATTATGGGGATATGTTCAC GGATGATGGCCAATACATGGCTAAACAATCACCTTACTATCAAGCGATTGATGCGTTGTTGAAA GCTCGTATCAAGTATGCTGCTGGTGGTCAAACCATGAAAATGAACTATTTTCCAGATGAACAAT CTGTTATGACATCAGTTCGTTATGGTAAGGGTGCAATGACGGCAAGTGACTCTGGTAACCAAGA GACACGCTATCAAGGTATTGGACTTGTTGTCAACAATCGCCCAGATTTGAAACTATCTGACAAA GATGAAGTCAAAATGGATATGGGTGCGGCACATAAAAACCAAGATTATCGCCCAGTTTTGTTGA CGACAAAATCAGGATTAAAAGTCTACAGCACTGATGCAAATGCACCTGTCGTTCGAACTGACGC CAATGGCCAATTAACTTTTAAGGCAGACATGGTATATGGTGTAAACGACCCACAAGTGTCAGGG TACATTGCGGCTTGGGTACCAGTAGGGGCTTCAGAAAATCAAGATGCTCGAACGAAAAGTGAAA CAACGCAGTCAACTGACGGGAGTGTTTATCATTCTAATGCAGCGTTAGATTCGCAAGTCATTTA TGAAGGCTTTTCAAATTTTCAAGACTTTCCAACAACACCCGATGAGTTTACGAACATTAAAATT GCTCAAAATGTTAACTTATTTAAGGATTGGGGTATTACTAGCTTTGAAATGGCGCCACAATATC GCGCCAGCTCAGATAAAAGTTTCTTAGATGCTATCGTACAAAATGGTTATGCATTTACAGATCG ATATGATATTGGTTACAACACACCAACAAAGTATGGGACAGCAGATAATTTGTTAGATGCTTTA CGTGCATTGCATGGTCAGGGTATTCAAGCGATTAACGACTGGGTACCAGATCAAATTTATAATC TACCCGATGAACAGTTAGTCACGGCTATTCGAACAGACGGTTCAGGTGATCATACTTATGGTTC AGTTATTGACCATACTTTGTATGCATCAAAGACAGTTGGCGGGGGCATTTATCAGCAACAATAT GGTGGGGCCTTCTTGGAACAATTAAAAACACAGTACCCGCAACTTTTCCAGCAAAAACAGATTT CCACAGATCAGCCAATGAACCCAGATATTCAAATTAAGTCATGGGAAGCCAAGTATTTCAACGG TTCGAACATTCAGGGGCGTGGGGCTTGGTATGTTTTGAAGGACTGGGGCACACAACAGTATTTT AATGTGTCAGATGCGCAGACCTTCCTTCCAAAGCAATTATTGGGTGAAAAGGCCAAAACTGGTT TTGTTACGCGTGGTAAGGAGACTTCATTCTATTCCACTAGTGGCTATCAAGCAAAATCTGCCTT TATTTGTGATAACGGTAATTGGTACTACTTTGATGACAAAGGGAAAATGGTTGTTGGAAACCAA GTTATCAATGGCATCAATTATTACTTTTTACCGAATGGTATCGAATTACAAGATGCCTATCTAG TACATGATGGTATGTACTATTATTATAATAATATTGGCAAGCAACTGCACAACACATATTACCA AGATAAACAAAAAAATTTCCATTACTTCTTTGAAGATGGGCACATGGCACAGGGTATTGTCACC ATCATTCAAAGTGATGGCACCCCAGTCACACAGTACTTTGATGAGAATGGTAAGCAACAAAAAG GCGTGGCGGTCAAAGGATCAGATGGTCATTTGCATTACTTTGACGGTGCGTCAGGGAATATGCT CTTTAAATCATGGGGTAGACTAGCAGATGGCTCTTGGCTATATGTAGACGAGAAAGGTAATGCG GTTACAGGCAAACAAACCATTAATAATCAAACGGTTTACTTTAATGATGATGGTCGTCAAATCA AAAATAACTTTAAAGAATTAGCAGATGGTTCTTGGCTTTATCTTAACAATAAAGGTGTTGCAGT AACAGGAGAGCAAATAATTAATGGGCAGACACTTTATTTTGGTAACGATGGTCGTCAATTTAAA GGGACAACACATATAAATGCTACTGGTGAAAGCCGTTACTATGACCCAGACTCAGGTAATATGA		

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TAACTGATCGTTTTGAACGTGTTGGTGATAATCAATGGGCTTATTTTGGTTATGATGGTGTTGC AGTAACAGGGGACCGAATCATTAAAGGGCAAAAACTCTATTTCAACCAAAATGGTATCCAAATG AAAGGCCACTTACGTCTTGAAAATGGTATCATGCGTTATTACGATGCTGATACTGGCGAATTAG TTCGTAATCGATTTGTATTGCTATCTGATGGTTCATGGGTTTACTTTGGCCAAGATGGCGTACC CGTAACTGGCGTGCAAGTGATTAATGGCCAAACATTATATTTTGACGCAGATGGTAGGCAAGTC AAAGGGCAGCAACGTGTAATCGGCAATCAACGCTATTGGATGGATAAAGACAATGGTGAAATGA AAAAAATAACATACGCGGCCGCACTCGAGCACCACCACCACCACCACTGA		
Transglucosidase CAA25303.1 GI:2343	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR EYTVPQACGTSTATVTDTWR	163	
Glucoamylase G1 1008149A GI:224027	ATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVASPSTDNPTYFYTRDSGLVLKTLVDL FRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGLGEPKFNVDETAYTGSWGRPQRDGPAL RATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQYWNQTGYDLWEEVNGSSFFTIAVQHR ALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSFILANFDSSRSGKDANTLLGSIHTFDP EAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYYNGNPWFLCTL AAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAATGTYSSSSSTYSSIVDAVKTFADGFVS IVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSASSVPGTCA ATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVTSTSKTTATASKTSTSTSSTSCTTPTA VAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALSADKYTSSDPLWYVTVTLPAGESFEYK FIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	164	
Transglucosidase	SLLAPSQPQFXIPASAAVGAQLIANIDDPQAADAQSVCPGYKASKVQHNSRGFTASLQLAGRPC NVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTXASWYFLSENLVPRPKASLXASVSQSDLFV SWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFVTALPEEYNLYGLGEHITQFRLQRNA XLTIYPSDDGTPIDQNLYGQHPFYLDTR YYKGDRQ	165	
Transglucosidase AAB23581.1 GI:257187	SQDYISLSHGVFLRNSHGLEILLRSQKLIWRTLGGGIDLTFYSGPAPADVTRQYLTSTVGLPAM QQYNTLGFHQCRWGYNXWSDLADVVANFEKFEIPLEYIWTDIDYMHGYRNFDNDQHRFSYSEGD EFLSKLHESGRYYVPIVDAALYIPNPEXASDAYATYDRGAADDVFLKNPDGSLYIGAVWPGYTV FPDWHHPKAVDFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGXLTLNPAHPSFLLPGEP GDIIYDYPEAFXITXATEAASAXAGASXQAAATATTXXXXVSYLRTTPXPGVRNVEHPPYVINH DQEGHDLSVHAVSPXATHVDGVEEYDVHGLYGHQGLXATYQGLLEVWSHKRRPFIIGRSTFAGS	166	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADTCGFNGNSDEELCNRWMQLSAFFPF YRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTTGSTVMRALSWEFPNDP TLAAVETQFMVGPAIMVVPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKPGVXTTISAPL GHIPVYVRGGNILPMQEPALTTREARQTPWALLAALGSXGTASGQLYLDDGEXIYPXATLHVDF TASRSSLRSSAQGRWKERNPLAMVTVLGVNKEPSAVTLNGQAVFPGSVTYXSTSQVLFVGGLQX LTKGGAWAENWVLEW		
Transglucosidase CAA25219.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	167	
Transglucosidase BAA23616.1	MVKLTHLLARAWLVPLAYGASQSLLSTTAPSQPQFTIPASADVGAQLIANIDDPQAADAQSVCP GYKASKVQHNSRGFTASLQLAGRPCNVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTNASWY FLSENLVPRPKASLNASVSQSDLFVSWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFV TALPEEYNLYGLGEHITQFRLQRNANLTIYPSDDGTPIDQNLYGQHPFYLDTRYYKGDRQNGSY IPVKSSEADASQDYISLSHGVFLRNSHGLEILLRSQKLIWRTLGGGIDLTFYSGPAPADVTRQY LTSTVGLPAMQQYNTLGFHQCRWGYNNWSDLADVVANFEKFEIPLEYIWTDIDYMHGYRNFDND QHRFSYSEGDEFLSKLHESGRYYVPIVDAALYIPNPENASDAYATYDRGAADDVFLKNPDGSLY IGAVWPGYTVFPDWHHPKAVDFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGNLTLNPA HPSFLLPGEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTSVSYLRTTPTPGVRN VEHPPYVINHDQEGHDLSVHAVSPNATHVDGVEEYDVHGLYGHQGLNATYQGLLEVWSHKRRPF IIGRSTFAGSGKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADTCGFNGNSDEELCNR WMQLSAFFPFYRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTTGSTVMR ALSWEFPNDPTLAAVETQFMVGPAIMVVPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKP GVNTTISAPLGHIPVYVRGGNILPMQEPALTTREARQTPWALLAALGSNGTASGQLYLDDGESI YPNATLHVDFTASRSSLRSSAQGRWKERNPLANVTVLGVNKEPSAVTLNGQAVFPGSVTYNSTS QVLFVGGLQNLTKGGAWAENWVLEW	168	
Transglucosidase P56526.1	MVKLTHLLARAWLVPLAYGASQSLLSTTAPSQPQFTIPASADVGAQLIANIDDPQAADAQSVCP GYKASKVQHNSRGFTASLQLAGRPCNVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTNASWY FLSENLVPRPKASLNASVSQSDLFVSWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFV TALPEEYNLYGLGEHITQFRLQRNANLTIYPSDDGTPIDQNLYGQHPFYLDTRYYKGDRQNGSY IPVKSSEADASQDYISLSHGVFLRNSHGLEILLRSQKLIWRTLGGGIDLTFYSGPAPADVTRQY LTSTVGLPAMQQYNTLGFHQCRWGYNNWSDLADVVANFEKFEIPLEYIWTDIDYMHGYRNFDND	169	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	QHRFSYSEGDEFLSKLHESGRYYVPIVDAALYIPNPENASDAYATYDRGAADDVFLKNPDGSLY IGAVWPGYTVFPDWHHPKAVDFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGNLTLNPA HPSFLLPGEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTSVSYLRTTPTPGVRN VEHPPYVINHDQEGHDLSVHAVSPNATHVDGVEEYDVHGLYGHQGLNATYQGLLEVWSHKRRPF IIGRSTFAGSGKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADTCGFNGNSDEELCNR WMQLSAFFPFYRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTTGSTVMR ALSWEFPNDPTLAAVETQFMVGPAIMVVPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKP GVNTTISAPLGHIPVYVRGGNILPMQEPALTTREARQTPWALLAALGSNGTASGQLYLDDGESI YPNATLHVDFTASRSSLRSSAQGRWKERNPLANVTVLGVNKEPSAVTLNGQAVFPGSVTYNSTS QVLFVGGLQNLTKGGAWAENWVLEW		
Transglucosidase AAP04499.1	MSFRSLLALSGLVCTGLANVISKRATWDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSFI LANFDSSRSAKDANTLLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDATG TYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLTA NNRRNVVPSASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVTS TSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALSA DKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	170	
Transglucosidase AAM18050.2	SSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWDTSDGIALSADKYTSSNPLWYVTVT LPAGESFEYKFIRIESDDSVEWESDPNREYTVPQVCGESTATVTDTWR	171	
Transglucosidase AAT67041.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFRQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSGDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	172	
Transglucosidase P69328.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT	173	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR		
Transglucosidase BAF37801.1	MVKLTHLLARAWLVPLAYGASQSLLSTTAPSQPQFTIPASADVGAQLIANIDDPQAADAQSVCP GYKASKVQHNSRGFTASLQLAGRPCNVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTNASWY FLSENLVPRPKASLNASVSQSDLFVSWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFV TALPEEYNLYGLGEHITQFRLQRNANLTIYPSDDGTPIDQNLYGQHPFYLDTRYYKGDRQNGSY IPVKSSEADASQDYISLSHGVFLRNSHGLEILLRSQKLIWRTLGGGIDLTFYSGPAPADVTRQY LTSTVGLPAMQQYNTLGFHQCRWGYNNWSDLADVVANFEKFEIPLEYIWTDIDYMHGYRNFDND QHRFSYSEGDEFLSKLHESGRYYVPIVDAALYIPNPENASDAYATYDRGAADDVFLKNPDGSLY IGAVWPGYTVFPDWHHPKAVDFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGNLTLNPA HPSFLLPGEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTSVSYLRTTPTPGVRN VEHPPYVINHDQEGHDLSVHAVSPNATHVDGVEEYDVHGLYGHQGLNATYQGLLEVWSHKRRPF IIGRSTFAGSGKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADTCGFNGNSDEELCNR WMQLSAFFPFYRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTTGSTVMR ALSWEFPNDPTLAAVETQFMVGPAIMVVPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKP GVNTTISAPLGHIPVYVRGGNILPMQEPALTTREARQTPWALLAALGSNGTASGQLYLDDGESI YPNATLHVDFTASRSSLRSSAQGRWKERNPLANVTVLGVNKEPSAVTLNGQAVFPGSVTYNSTS QVLFVGGLQNLTKGGAWAENWVLEW	174	
Transglucosidase CAK37022.1	MLYAEDNKLIFRFDDHLLWIQPWGENALRVRATKLASMPTEDWALSSKVTSIEPTISIEEHKDS SITNGKIKATVSQRGKITIYNQKGEKLLEEYARNRRDLKDPKCSALEVEARELRPILGGDFHLT MRFESLDPKEKIYGMGQYQQPFLNLKGVDIELAHRNSQASVPFALSSLGYGFLWNNPAIGRAVL GTNTMSFEAYSTKVLDYWVVAGDSPAEIEEAYSKVTGYVPMMPEYGLGFWQCKLRYWNQEQLLD VAREYKRRNIPLDLIVVDFFHWKHQGEWSFDPEFWPDPDAMIKELQSLNVELMVSVWPTVENAS TNYPEMLEKGLLIRHDRGLRVSMQCNGDITHFDATNPSARAYVWSKAKQNYYDKGIKVFWLDEA EPEYSVYDFDLYRYHAGSNLQIGNIFPKEYARGFYEGMESAGQTNIVNLLRCAWAGSQKYGALV WSGDIASSWSSFRNQLAAGLNMGLAGIPWWTTDIGGFHGGDPSDPAFRELFTRWFQWGAFCPVM RLHGDREPKPENRPTDSGSDNEIWSYGEEVYEICKKYIGIREELRDYTRGLMKEAHEKGTPVMR TLFYEFPADKKAWDVETEHLFGSKYLVVPVFEAGKRSVEVYLPAGASWKVWGQEDVIHEGGKEI QVDCPIETMPVFVRV	175	
Transglucosidase CAK37087.1	MSSPQQVYLLPLKDDGSPDVPGGYIYLPAPTNPPYLLRFVIEGSSSICREGALWVNIPEKGESF NRSAFRSFSLSPDFNKNIQIDVPITSAGSFAFYVTFSPLPEFSVLSTPTPEPTRTPTHYIDVSP KLTLRGQDLPLNALSIYSVISKFMGQYPKEWEKHLNGISQRNYNMVHFTPLMKRGASNSPYSIF DQLQFDDAVFPNGEDDVARLISKMENEYGLLSLTDVVWNHTAHNSKWLEEHPEAGYSVETAPWL EAALELDTALLKFGQDLQNLGLPTEFQTVDELMKVMNVMRDKVIAGIRLWEFYAIDVKSDTHKI LDKWKTSKDIDLTDTNWAQLNLQDYKNWTLKQQATFIRDHAIPTSKQVLGRFSRAVDLQFGAAI LTALFGPHNPSTSDTSIVEESLSKILDEVNLPFYEEYDGDVSEIMNQVFNRIKYLRIDDHGPKL	176	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GAVTAQSPLIETYFTRLPLNDVTKKHKKEALALVNNGWIWNADALRDNAGPDSRAYLRREVIVW GDCVKLRYGSCRDDNPFLWDFMTDYTRLMAKYFSGFRIDNCHSTPLVVAEYLLDEARKVRPNLT VFAELFTGSEEADYIFVKRLGINALIREAMQAWSTGELSRLVHRHGGRPIGSFDLDLPSSGSSH AIASSGLDSGKEKVVHIRPTPVQALFMDCTHDNEMPAQKRTAKDTLPNGALVAMCASAIGSVIG YDEVYPRLVDLVHEHRLYFSEFSEAPETGLNSLEGGIGGIKKLLNELHTKMGIEGYDETHIHHD GEYITVHRVHPRTRKGVFLIAHTAFPGQDSRSVLAPTHLVGTQVKHIGTWLLEVDTSQTTKERI QADKSYLRGLPSQVKTFEGTKIEESGKDTIISVLNSFVAGSIALFETSMPSVEHASGLDNYITE GVDHAFSDLSLVDLNFALYRCEAEERDSSKGQDGAYDIPGHGPLVYAGLQGWWSVLENIIKYNE LGHPLCDHLRNGQWALDYIVARLEKLSHKEEHPALGRPAAWLQEKFQAVRQLPSFLLPRYFAII VQVAYNAAWKRGIQLLGPHIQKGQEFIHQLGMVSVQQTGYVNSASLWPTKKVPSLAAGLPHFAV DWARCWGRDVFISLRGLLLCTGRFEDAKEHITAFASVLKHGMIPNLLSSGKLPRYNSRDSVWFF LQSIQDYTEMAPDGLEILDHKVPRRFLPYDDVWFPFDDPRAYSQQSTISEIIQEVFQRHAQGLS FREYNAGPDLDMQMTQDGFQIDVKVDWETGLIFGGSQYNCGTWQDKMGESAKAGNKGVPGTPRD GAAIEITGLVYSALTWVAKLHERGIYKHDGVDIGGGKSISFEDWASRIRANFERCYYVPLQPKD DGQYDIDANIINRRGIYKDLYRSGKPYEDYQLRSNFPIAMTVAPDLFTASKALAALALADEVLV GPVGMATLDPSDLNYRPNYNNSEDSTDFATAKGRNYHQGPEWVWQRGYFLRAFLHFDLARRTTP AERTETYQQITRRLEGCKRALRESPWKGLTELTNKNGAYCADSSPTQAWSAGCLLDLYYDASRH SQS		
Transglucosidase CAK43781.1	MWSSWLLSALLATEALAVPYEEYILAPSSRDLAPASVRQVNGSVTNAAALTGAGGQATFNGVSS VTYDFGINVAGIVSVDVASASSDSAFIGVTFTESSMWISSEACDATQDAGLDTPLWFAVGQGAG LYTVEKKYNRGAFRYMTVVSNTTATVSLNSVKINYTASPTQDLRAYTGYFHSNDELLNRIWYAG AYTLQLCSIDPTTGDALVGLGVITSSETISLPQTDKWWTNYTITNGSSTLTDGAKRDRLVWPGD MSIALESVAVSTEDLYSVRTALESLYALQKPDGRLPYAGKPFFDTVSFTYHLHSLVGAASYYQY TGDRAWLTRYWGQYKKGVQWALSSVDSTGLANITASADWLRFGMGAHNIEANAILYYVLNDAIS LAQTLNDNAPIRNWTTTAARIKTVANELLWDDKNGLYTDNETTTLHPQDGNSWAVKANLTLSAN QSAIVSESLAARWGPYGAPAPEAGATVSPFIGGFELQAHYQAGQPDRALDLLRLQWGFMLDDPR MTNSTFIEGYSTDGSLAYAPYTNTPRVSHAHGWATGPTSALTIYTAGLRVTGPAGATWLYKPQP GNLTQVEAGFSTRLGSFASSFSRSGGRYQELSFSTPNGTTGSVELGDVSGQLVSDRGVKVQLVG GKASGLQGGKWKLSNN	177	
Transglucosidase CAK37133.1	MSSDSQLSRSHFLAPPTVIPAPSYIASSAAAQIITADQEFNAADFVADDEGHDSSASALVTPEA LSSLNAFLDNILFNILAAAKSTQLVKIRPAVAEVLKPRLAKEMVSAADDELSEYLGGPEDEQLE FRSGQTSIGEFDLVRSWKLTRLRCMVYTRLGDMEEDDEEEYINQEIIGEDGGGLRRLASHVGHI TPAASIFLTSIIEHMGEQALIIAGEIARSRLSANLEDEDDLAGTGANRASMDRLVVEDHDMERL ALNPTLGRLWRTWRKRVRGSNLSRAVSRESLRNRQSLVFGPGSRKSSAITIDEISPRTASSRSV NEPLPETEDEVDPASVPLPMSEHDIQEIEIPCFLPELDTGDIQTMQAVVAHKVRPHSLMVLTLP SPRSPSSNGNSPITPRLVNIKSPRHVRSRSLPNTAPADEQPSEVEQPAERTSPTPSEERRRLET MYEHDEDDERHGEAATKPEAVEQNEPVVPSAGQGAAATPSTSVASVEVAMSDASSTPVSSPSLS DRDYPETDEVEKHERVEKAQLAPGVETAPGPLAPRTQGVVDSTPAQPTAAADQDASKAADCDQS	178	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TPEDSTPPTVPSSAEDTAVEKASRPVSTSGESAISDSSRSLPGKRGSSVPGVQHQYGRSSPGIA SVSSGVERAAVQRLPARPSTSVASSVYSKSRRSGSFSSSREKRPVTAGSTTSQVSSKLKGLIGR PADTGSLRLRTSSEVSRVSTRESAYDDTSGLDELIRSEETIHFTLTPRSMREMELPDSPRWRAQ QASTDPTDLPKSVEPIPDDMSRSRHSTTSSKSTVDLPPVPKYIQSKPKSIEIPTTGLQQKPAVG QARDAKHSMESTRDFANFLKSTGPNTPTTPATVDGSPAKSSRLRRLSDATEISKKLSRPASSTV SVANSARSGPRLEARSAVAPRGDQTSDLIDFIREGPPTAGAHRIPRTVAPFRDTMDSDELQAIE PGRTAKGAPSVASTQSVAETSLVSVGSRTGLLESTSRTSTPTALAKETKTTFAAPVSVSDDHRP PRTRRRVPDPYAIDLDDDDELDELLEEPKPKRDEESLIDFLRNVPPPEPTPPQQPLAATANSRR GSASVKARLRRNTASEKTLMAKPSKTSLHQQPDNYMGGASNYTVKVGMERNAGAMNGAYDLKTP SVRQTETSALADFLKNTGPPEPPVTKAPAATKSKDSGFSRLFMRRKKVEA		
Transglucosidase CAK37219.1	MAALVQTIPQQSGTVSVLQTRPSSSSGTFTTSSQPGQQQNPRNSTMSWNPYNNSGSSGNYRVGH QVVAPYAFTSTPNPSNPTNMQSRQSLSPHLRPEHRTSSAPSVPQGSASPANVGVNSRFAHPAAG SVSTSSSNSSVHSYMSKDDSAIPTRQIRTDAPLRPLSTVNLPSPSSSNFMNISSPTVARPSPDR YRRGNRRPENAAGAQPASTQPNGPAPARSATLATDDSSLHMSTPGLAGVSLDAPRRPGHVRVPS ADDTTRADKPQTELAKRYRRRSWGNIDNAGLINMQLHLPTSSPTPTAGGHDYFDQSMRPRSAQS HREVQGSIPSAHSSTSSVRDAGHSESASSSKSGPKTDDSKRPNKPSPLSQPVEVDAKPPTPKAP QPSTTPQPAPTESLATQRLAEITKGDPKRPGKSRLRRAFSFGSASELLKASSQNKREAMATERA RRELLQEELGPEQAAIAEQQEASGLGESIYSSHHQGRIFNSSTDNLSVSSTASSASIMLRKMGK GMKRSTRSLVGLFRPKSVVSTSSTDGVMIEPMAPQLSVVNIEAERKSTAVTSDSQDHALGSSLF SKVETDAANAVSHEDGALDKSRKSIVGGDRERAEVLAAVRKGILKKTYSDPANQTYVLKSSENL NSNDSPHSSVPSTPEDQTRSGNRRSDAVKIAGEDDYLSEGRFQTSESKSAPITPQAMMPKSLVF SPRIQFHETWPSGEYDRRGDIATCNRLTPLLAQQIKEELNNFKMVRNLLPLLPST	179	
Transglucosidase CAK37226.1	MFVYCSNCSFALLFLMLLSLLSFTASRTLAFTTSITQDGLLCFPSALEFLLRTEITVPYWAPSG SILRPTAALHAYDCSVCLDPPSKIQEARKSVLMSANALSEIIDIPVSDGGFIHGVIRFYARGDH LRWLQPPTRDAKFLAPDPYLHSLMIESWRQTLGEMHFWTRAGYLFDVVLAEVKRSEPDNYEFLN WSGTNYMPTCPYYYVSMSPMVQPVVRSAQGSVDAAQRSSQISQDPSNLVQTPLHSPEFSDDSTR SSFNTAQTASSCQSFSSPVSQSPCHEDVIQQNVQTSQVSLPFVRMDPSVTLDDPFVIEPISAEG SWSMQDHIADMKRQFRLPGPMVRNASPSFDSPTPSTTERISAREINRRRDSEKPYDPTPLANDT SASCDDETWSMEDASEKDASESSFKDAVEAHSDSASSTATGPVVASKNDDDQSLPKCNTNDTQP TTCTTVNPSLLMFESSHKTYPTIEPSYEVAASRPRSLSPVQNLENELQVGSIGGKDAEDAGSLS FNDEMKEGSEMDLFSASLDQYTAEQLASRQLTRTPELEESNPNEYGLGFGFQHNLFDGFDFFLP EDQSELPLESNMIM	180	
Transglucosidase CAK37273.1	MLGSLLLLLPLVGAAVIGPRANSQSCPGYKASNVQKQARSLTADLTLAGTPCNSYGKDLEDLKL LVEYQTDERLHVMIYDADEEVYQVPESVLPRVGSDEDSEDSVLEFDYVEEPFSFTISKGDEVLF DSSASPLVFQSQYVNLRTWLPDDPYVYGLGEHSDPMRLPTYNYTRTLWNRDAYGTPNNTNLYGS HPVYYDHRGKSGTYGVFLLNSNGMDIKINQTTDGKQYLEYNLLGGVLDFYFFYGEDPKQASMEY SKIVGLPAMQSYWTFGVCPPPPNPITVRVVVYNYSQAKIPLETMWTDIDYMDKRRVFTLDPQRF PLEKMRELVTYLHNHDQHYIVMVDPAVSVSNNTAYITGVRDDVFLHNQNGSLYEGAVWPGVTVF	181	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	PDWFNEGTQDYWTAQFQQFFDPKSGVDIDALWIDMNEASNFCPYPCLDPAAYAISADLPPAAPP VRPSSPIPLPGFPADFQPSSKRSVKRAQGDKGKKVGLPNRNLTDPPYTIRNAAGVLSMSTIETD LIHAGEGYAEYDTHNLYGTRLVMSSASRTAMQARRPDVRPLVITRSTFAGAGAHVGHWLGDNFS DWVHYRISIAQILSFASMFQIPMVGADVCGFGSNTTEELCARWASLGAFYTFYRNHNELGDISQ EFYRWPTVAESARKAIDIRYKLLDYIYTALHRQSQTGEPFLQPQFYLYPEDSNTFANDRQFFYG DALLVSPVLNEGSTSVDAYFPDDIFYDWYTGAVVRGHGENITLSNINITHIPLHIRGGNIIPVR TSSGMTTTEVRKQGFELIIAPDLDDTASGSLYLDDGDSLNPSSVTELEFTYSKGELHVKGTFGQ KAVPKVEKCTLLGKSARTFKGFALDAPVNFKLK		
Transglucosidase CAK96369.1	MSLSFSSDVALNATEAAVFLSERDVAGQIPINFVTTSAVSLRAACFGDNIYDRDAAGRCISNLL VVGYRRFLVDLYWSSDQRDWMFCPLSLSPDVPVVTVSSISPASSTTTTATSGITATTTATTTTT TTSETIKATVTAVARSSGSVLYELGPYRCSLDFDLSDLINVFRGFFQAYSSELTVFTRYISLNL HAAGSATSPDEPASTVTGSQLPTSSEFVSYQFDEHLSSYIYTPSSLASERANLNQSWYQVEDGY KPITEYFTIHEEPNGDQSTPDGWPCVKYLQLAQEKRLLIDYGTIDSQLQDYNFSYMSDVIFPPN YLTSTVSVSLDSDGSVDTGCFYDSGATTVSQANNSWAISDYIPIPEGLSENSTIAAMSLVASNL TACGLTPALNNTLFNQTADTHPQPYTDISLSSSWAWSIGQPANADSSSASFSATDRCAVIDLTN SGHWRAINCSQVRYAACRVGNNPFTWQLSPTPYTFRDAYDHGCPENTSMAVPRTGLENTYLYQY LLTRTDVLDPTSAIPNKTKVWLNMNCIDVESCWVTGGPDQECPYASDPQQLERRTVLVAAIAGI VICIIAALTLFVKCNANRRNSRRNKRVIKGWEYEGVPS	182	
Transglucosidase CAK96386.1	MADSKPKPSSIPPWQQSNNASNTDNSSESTSSPTSDDTSRSTLIEQASKFLEDESIRDAPTDRK VSFLQSKGLREDEINSLLGISATSTASDTTEEEKAASPDTTTPSSTEPAPAPEPTDNASASSNQ STPSSSITTPTPSPSTTTPKTNNTRDVPPIITYPEFLSTPTKPPPLVTLRSVLYTLYGAAGISA SFYGASEYLIKPMLSNLTSARQELASTATSNLQKLNEKLEQNVSVIPESLKNKTANVENDSSST DTESITSDPTELFHRDVATQTATSDFAATYNNSNKTGTDKDTPADPTAAVTDHLKRLESIRSQL RECSDTEKESGTLESGMRTRLNELHHYLDGLIYSKPGFNPLSGYGMYSTPGIDSGSGAATGVGK GEEDAIANFRAEIRGVKGALLSARNFPAGRGGRIGGVAGSIPTGLMRMNRVVNGIGSARPKKER YKHSPTTFRYYQ	183	
Transglucosidase CAK47557.1	MGVGDYVHSKEAGQPRPRTTEVSNQSRQAVAAQARIDVPPTNLVAPVPLPINKSIPLEHYSTPA FSEQMPQAPAENGVHRDMFDTDVEGIDESTIAATSVMGAEDAPLQFQLRPATVPQYQEAAPVVD ERPLHPSRLPRRAYDGKWYENFGDKAMKSAGFDSEDADDASQLTSMAGDDERSDTTEDANYARR YRSSTEEPLSKRLQSFWSASRRSYKNPEPQAYPEPSKTAASAAPPLLRQSTSDARLSKQALPNR KVTLPRSMTATPRTRFSPPKPSLLEQLDITPTRRTSGPRPQPGKEPGITSTTTHQHHRHNSDDN HLFNTSRDSLPPLSTFDMTNIDDLDVDDDNDPINDPFARRNSVQRIVSDPDFQPNKSTITSSSS QNKRRNLESDYPPEILRQKSFKDLQSEPFDHTPTATASAPVKTTTPAPTPGPNATSDEKMDFLM NSEDKDRRDFFSTLTMNEWEDYGDLLIDQFSDALSKMKDLRHARRKTAALFEAEIARRNEVVEE QSADLTRKLEEMRSGGAEVLKGRTP	184	
Transglucosidase CAK47704.1	MQAIEQAGSIFTGWISSCLFCLSGRGDDESFHHQQAMKQKGVEREMRVCHTQPHLVPPMNLTDY DDLPSPSSQPRVSSLQSWVVEGRTRASRASNRASMSLKRKSTAPVRISGPSEFRRVSMFLTELD EYRPLELSFNTPGNRLPDLPRFEDFPLDHDRKQVISRPPRALSSTEMGRISQPRTHRPSSSFQL	185	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ARKPVGSGSRRSSLPTQEQLQLLEKNTPITSPLIPHFSQRSSAVTGLTANAVPSTTPRLDLSGG STSLARNELHRDTHEAPTSTVPRTPTKPSLQDRPLPSIPTEEDSPSSGSTYHPPTTPSESRPPT TPSENPNQTPTRSGRVTQWLFQTPNKPNFLFSNPSKISDKGPFRIRSRTLSGSTLASTTTNITG GHKTTPSLASGTTVAPTMQASSTESRNFDLPLGSPFSPKQTFPTVTEEHTYPTIHEGEQQQHQE PEVFDDMLTQYYEYRHSAVGLAF		
Transglucosidase CAK44239.1	MKIFILLAIWLLASLGYATSFVGNMAEVYHEITDVLRGPHHAAHFAKLNGGKPKPDKPKGVGKT LDDVVTLTVCKTDVSLAPTVGTFSLPTIVTAATLAPTVVVTGVVPTEISIGLPTEFAPEIQTGV LTGESHSTDTTVVPTNSVVSGLSSFLTGSQTVSEITGSTENHTITTEINASAVTSTFAHTTFPT ANAGNDHEGMASSAVALLVALIFSLVRI	186	
Transglucosidase CAK44326.1	MRTRSQQASPGGFVSLDENAPRRTRSAKNAAQQEPATSEQPPTRSKSQRAPKKTTTTTTKKST AKAQTRKATTTKRTTRQSTRKTDQPVSNEDAQTTHTEEDFATDNTTAEKNTPREVPETVDPTPA SSENENPDRESLPHVSHPFMVPQPPKESQEIDCFDGPDPRGIGAASCLKSLIDELSSVGSPLSE RSKTPSWTSEDGTEAALAPRPAQESGATNVETSTTTERVSLPTPAAEERTVEPPAEPTVAEPRG VAIVTSSEQFNTVSSASAAGSGGVVVVEDERVGALIASFARLSLDDLAPRSSNEAAATLMESTT ASFGEPVEPAHVTRRSLRASRQEWILGWAQQVPSTGYFHPITGQLVEGPAASVELVGDPASNRG PPTRRIGVRDYILRRRRREVQVMSPLQEEPQSSPGPGSRAVALNAVPPRGIRAQRAQNTKRLTK PPVTRKRARTESSDEEAPGPQTPAANKRRNLGPPGSTPYRPATRPRSLTANITPYSERLRRRAA EKDGRIHSTSLRVSQLLAQQEADRRRQAAESSAPPCSELPRTTFDFSLDDAHETSQGQEQSQSL QEQSSTPQPPATPERQSGWNIRGLLNSVPRTFTRILPSFRRTPEPTQVQAPPEPSSERISRTQP PQSSSISQSQAQNSRRSSEEPPQKRRRKSWSLFAQPFDRSLYLGDIPKKDSATSSSAPLESRPV AKLSAEATTPQESATSDAKKDVAAEGEDSRGREIEEQKQKKRKRSPSPDVIPNPPGCSYGLDLD YFCYSSESEDEQEPPLPRTEPNKFGRLTRTAVRGALRSERHSSKKVRFDASPEDTPSKLRLRAR ATDPYRGRHFIGMGNDSEIATPDSPTPAPHAADESSSRRPGFVPNVQGTFQLDYDAFSDDSDSS GASASANVSASAPIPAPSSATVTQASISESVPSTESRQTPRQAAPAPSTPAKIDEEALARARSQ AEKYKPKTPSGLRTASRYSSPMTATPDTVSAPVIAPAITPTPSTSQTAPASAPEPEQQTTEDFG DDEFAREAQWLYENCPSGDLNDLVWPQPITYEEEGFSPEVIDLVNEIWDPSTVDYAYTNIWTPG LDAFKRELETGASEAAQA	187	
Transglucosidase CAK47737.1	MAKSASQIHRAWWKECSVYQIWPASYKDSNDDGIGDIPGIISKLDYIKNIGVDIVWLCPSYKSP QVDMGYDIADYYSIADEYGTVADVEKLIQGCHERGMKLLMDLVVNHTSDQNEWFKQSRSSKDNK YRNWYVWKPARYDEQGNRHPPNNWVSHFQGSAWEWDEHTGEYYLHLYATEQPDLNWEHPPVRKA VHDIMRFWLDKGADGFRMDVINFISKDQRFPDAPVKDPRTPWQWGDKYYANGPRLHEYLQDLGK ILKEYDAFSVGEMPFVRDTEEVLRAVRYDRNEINMIFNFEHVDIDHGTYDKFEPGSWKLTDLKA FFETWQKFMYNNDGWNALYWENHDQPRSIDRYAQAKEEFRTEAGKMLATVLALQSGTPFVYQGQ EIGMRNVPVEWDMNEYKDIDCLNHWHRLLKHRPDDIEAQKSARQEYQKKSRDNGRTPVQWSSAP NGGFTGPNAKPWMSVSPDYVRFNAEAQVNDPNSIYHYWAAVLGLRKKYLDIFVYGDYDLVDKDS QEIFAYARQYENKKALVLTNWTEKTLEWDATTNGVKGVKDVLLNSYESAEAAKGRFSGQKWSLR PYEAVVLLVEA	188	
Transglucosidase	MAYYEPQGWQAPAARQASWEQPAPPSRSGSSSVSQRDEIPAFSSQFDEVDRAIDNLVKSGKLWA	189	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
CAK47819.1	APRRDSMPMMMGRPYPDYDPRMVNSMSQRHHSISEFDSRMHPSPNVQGFYASQRFQGRPNEVEQ MMQAKRRMAAQRERELRNYHQEQQYNRSLLAEMSGNKSDRSLSPAAMSEESRRELLARQHRALY GNDSPAFFPPAGLADDGTRSESQAGGTPTSSTGVRGASPRNVDPFGLAQTPVQAGADSLGQTAA SAASLQSPSRANSTSSPSSAINPVFGKYDSADQPVTSTSSPGGADSPSSRQAPSKSMAGPIGSV GPIGTRPLPQPHAGQVSNPALNKRSTTPLPSPLGFGFTPGDAASDRSVPSVSTAPTTAAATASV KDTSGGVGLGWGNGSGVWGSKNGLGVQASVWG		
Transglucosidase CAK49181.1	MLSKMQLAQLAAFAMTLATSEAAYQGFNYGNKFSDESSKFQADFEAEFKAAKNLVGTSGFTSAR LYTMIQAYSTSDVIEAIPAAIAQDTSLLLGLWASGGGMDNEITALKTAISQYGEELGKLVVGIS VGSEDLYRNSVEGAEADAGVGVNPDELVEYIKEVRSVIAGTALADVSIGHVDTWDSWTNSSNSA VVEAVDWLGFDGYPFFQSSMANSIDNAKTLFEESVAKTKAVAGDKEVWITETGWPVSGDSQGDA VASIANAKTFWDEVGCPLFGNVNTWWYILQDASPTTPNPSFGIVGSTLSTTPLFDLSCKNSTTS SSSAVVSAAASSAAGSKAVGSSQASSGAAAWATSASGSAKPTFTVGRPGVNGTVFGNGTYPLRP SGSASARPSAGAISSGSGSSSSGSGSSGSTGTSATSGQSSSSGSSAAAGSSSPAAFSGASTLSG SLFGAVVAVFMTLAAL	190	
Transglucosidase CAK49185.1	MPCVQAAAETDKSFVQIANADIEELIKQLTLDEKVALLTGDDFWHTVPIPRLGIPSIRLSDGPN GVRGTRFFGSVPAACLPCGTAIGATFDRNLAVQVGHLLAAEAKAKGAHVILGPTINIQRGPLGG RGFESFSEDPLLSGIIAGHYCKGLKEDNIVATLKHFVCNDQEHERMAVNSILTDRALREIYLLP FMIAIALGKPEAIMTAYNKVNGLHASESPALLQGILREEWGWEGLLMSDWFGTYSTSEAIHAGL DLEMPGPTRWRGGALTHAITANKIPMATVNARVRAVLRLVQQASRSGIPERALELQLNRAEDRQ LLRKIASEAVVLLKNDDNILPLDKTKKIAVIGPNSKIATYCGGGSAALNPYEAVTPFEGISNSA SGGVEFAQGIYGHQNLPLLGKRLRTQDGLTGFTLRIFNDPPTVANRVPLEERHETDSMVFFLDY NHPKLQPVWFADAEGYFVPEESGMYDFGLCVQGTGKLFVDGKLLVNNANVQRPGPSFLGSGTME ERGTLELTAGRQYKVHVQWGCAKTSTFKVPGVVDFGHGGFRFGACRQLSPHTGIEEAVQLAASV DQVVLVAGLSAEWESEGEDRTSMGLPPHTDELISRVLEVNPDTVVVLQSGTPVEMPWIQNAKAV LHAWYGGNETGNGLADVIFGDVNPSGKLPLTFPRHVKNNPTYFNHRSEGGRVLYGEDVYVGYRF YDEIEIDPLFPFGHGLSYTTFELSGLSFERDSNSLHAICTLRNTGSRAGAEVIQLYVAPVSPPI KRPQKELKEFRKVWLEPGAEDVVQIPLDLVRATSFWDEKSSSWCSHSGTYRIMLGTSSRGAFLE SPIELSETTFWSGL	191	
Transglucosidase CAK38411.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	192	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Transglucosidase CAK47899.1	MDPNTSDRLKRLQMENLGTARREYISTADDRRHKKARLEDIQAIRMTNVPGSAAQRMATVNQGR LEDWANVHKTIGDTEDLENLDSLLDGQSHRLQLSAIIRESGGQKYIGSQRRSGSAPATRSLHPV GRGGGVIGTRGRGTRQSSLPPSNPTPRGHVTQPPKRSHGDAALDDNDFYRTAREGNASRKEEAM RAQNRRSSVRRPTSSTTRPRRPVSQVDYSSMLSQPQSFLAAARSLVSARTTPAAPTPASQISRD GGRSEASRKSSSPMDTTDRPTVQKTQQEPKPQMAEPPKPFTRPVVQLPAIPPRCTAVQQESTTK ADEPLPVLPSSAVLDATSQDQGSASMSLGSTEDGQSISGIPESQTGTQEKVNSDLSTAAAPVPD IKDTSPKAATDVKEAILLDFSYTPPEQSIHGQSPAPTEVLTPSLEDLRGLDFKQDIHPKFPTRR RVDFDMSSDKREASTNVHDLMPTKQYDKAEASEDLHRQINMLCELLQSTSLSGEHRESLKQCKT ALEGKLHGAYDSTGKRTQTQGDPFLGKPVLETLEAAAEPDEQPQSTISGLGIQNVSMDNFTPNK AETIVEPDTQATPSVGEMIMPTSVENARLAMASPSPSSRLNVTAPPFVPKTPFRAQSNSFSSDS NATCVPETPCPHRRVSMPEGHIIGDHLLPGRRRETISSGTEPLAAKQPPANEVTEPRFKFSIPP KISRKLTIKTPVREGFGKEETPGPVAPRLASGNIPKPAPKPSAALQQSVHAPKAKPSSVLGGLE SSRYASPSSNKPFR	193	
Transglucosidase CAK38738.1	MPPSTVFAYWRREHRRSSASPVSPSLQPTSKAPVTSNPPQLPGLSSTRPNNLTALGASSVESSS PQVPNNPHEDYYDATKKTIAVSVAPANAPGSSSANLAIPSSSSDSHTRPLSISDEQDVTTTTSQ SNYSQSSIAPPRSDQSDGDSPKPSSPFRLSLGKSLLNSHNLTSDHYNKRSSTPGLSSSGHFRFR TSPDISPGDRMALSHKDKDKEYKYEGAGNRRSADRDGSSEQAHHKSGRTRLHLLNPMSLLARRR SSNLASLRTEDTRVGARNIVPAIPDDYDPRIRGNIVHDFSAPRPRRNLSTAPVLMHDVNNQSSS ADVTYNGTGNFAHGNDQSAQSGEQRKRHTQYSPVFREHFEDDQKVLQVESKAYLQSSLLTAQTN AENDPHTLPVFARKLPSKIPEQEVSPEVPSDQTTLKQDSQHSPPNNSRELAQEDTDTIEVIPHQ PSGLPKHLKSNASRFSFDMNGVESSAQEKLLEEKHKEKEAARRAKARMEGTSFSDGEDDFDEDL LDDMDDLEEKIPGVNVDADEDDDFSGFSGPGNALNKPWLAPELSPIIASPLPTGSTNSQNVQEL AQGPLAGISAPLPVSDPAVSDVTTNFQALSVATIAPNNAPQVAMGSHPPAPQPIEDDDDLYFDD GEFGDLSTEDMGEKFDESIFDDETSHLYERKPVVQQPVPAPVPPPDNGTGSTNPLDVTAEHDEF TPEPDYDGGLRHVPSMASDYRKGSIRVYGQTRESLANLGSAKAQGGVLSEHNLEAFHNALAKAA SEAAASDRFGREASISEQSLGQESTAQTMDTPSGLVSDDSRLSQTVDMAAFEEVFEDFSYDDND DALFDDPIIAAANAEALENDDEGFYGQEFGFYAQAHGGCNGELTNGGYFGPRGVEGVNRSFSSR GKFREPSLTPITERSEWSTRNSVISLTAHGAAHSNPIASPGLAQLVDLGAMDDEMSLSALMKLR RGAWGGSNGSLRSSSGSPPLLHSTSNRASFISDASPTVYTAPPDAFGGSATESPIRESDKFRWS LNNTEQRVGQSAAGEREP	194	
Transglucosidase CAK38790.1	MLVEPLIRTDWPVWACKPHPHLVGPEAVAKNRNRSALQPSLAPSQSLVVLAQSNLPPAFQPSAL SHGSVFGWPCILPWRSSGNAGDEPPSGPYSYSGWPTPLTSSNQPSPSRREHAVQPPPLTTSLGG HQFQGLGLALGSGYSSTPLSSTSLSSPFTQGQSPAVGSPGGAAIGSSPMASRQYNVPYNPQDWG PVGSGSMNAGQATYTPPNSMLRIVSQPRSTGPHSDVSLSPPPPPYSPPSQQHQRENVSQNTSSM GSTSPSISSSYNGAVRAGVDAPSEYRQRRLPRTRPLSMFVGSESSHNRRVSLPPPPPLPPGLSS RSSSQNRSETYREPASVMAGPGPHIVVSPDNLHSTQLSDDSNMLEPTQPFDTDRPPAARRAVSA GPAVNSASSSRAHSQSGARSPPGTSWEPGMPLPPPPPGPPPATRSQSVNGLSDSSSSRNSQGPV RGGRARPPPVLGTSLDSIPPTPAGWVDETIDVKPRTERQPLTIDTATTSNTNGPESLESSRASH	195	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	NPNSGGLFRSPAIKDPNAKGIRERRIERRNRQSQVLDSLSAVSMSSNPWAEALEQLKPSNLVLG ESSVDIDNGRNPASAKAAPLSTRSISSDGPQITSRSRASSGGLFSNRSCSTPKPEPSPQAPTSN SRFAQTPPFSPGTERSSAFPKRTSPALPPKALPTPPLQSGSETTPSRPGSKEERPVSHILHLPN EPVTTVSPLAPRRVSAQQNPSLDSVIKRDDDYVRNAIQRHREFIEKEAGTMDEKEALRLFADFI ISESQIRRERYAKVWDLDSFDVESVRRKLFVSPPKTAPVPQTSQVVPGPSSRRASNPTAPKLDI PQVRPESAWWNNYQPCLSPIASLGLSNDEMSSRGRAPSRWWESKTGSSSEGGERRVQRSKRETK YMGLSRGALLWEESQGSSDTGNAGTSNGGNQYAAYGPDEYPPEKVGWHEEPALEDYSNNVRLGS SRRFEEVQRMDVSRLITLPPPYPRHYPAVNNSHPDLVTYRTLVRSITDLSEIKTARQRHQTEMD GLFQDHQARVREGRRQFKANIQSQIQQGSITFAEAAEAEAALIVEENRLERDLIKGGLDTYQES VFKPMRAILADRIDRATACIDELRGRLFDDARSETPDQTQEEGDEKPELLEKLTQLKWLFEARE QLHREVFDLISDRDEKYRAVVLLPYKQASNEDKVRETNEFFVKDALDRRVDYEANALARLESFL DVIEGNVARGVEIQLSAFWDIAPSLSELVQQIPEKLRGFTVQIPANEYEENPSYRAHPLQYLYT LVSHAEKSSYQYIESQINLFCLLHEVKSAVMRASCNLMEAERIRLGESEGKVQQEMQETRTDEE RTLTSDLKDKVATVEGQWAEALGSGIQRLRERVKEQLMVEDGWEDLEQLEQA		
Transglucosidase CAK38810.1	MVTQSSLLDRVWLYTHKRSPILLALHPPSQSSLISSEGHLEEQTEPTVQSRRYIPNTIMNTNMH PQSLDSVRSGEEAKSENEDSNTRSGAISLIRARTISRSSTPRDTQEGCSSEQADDSQQAVSIPR IVVMEPPGDSKAKRNKMKKNKYKKKKLLLGDSESETSGSQGSGLNGKPTAESNTSNDTSSHMKE IEDLSHSAWMPAKGLNSGGCANKEKPMSGSNVSERCTVDSDKKRDLIESLSKLDIKGKQRVWQV NAVTGNDTLEEINTQRGPQPVDRARRTVLAPSYATVLSGKRASIEGPSSMPNNSDSLLPTTMEF PKLTDKTSAGQDSSPEARSGHAKLSPIPEISGEFAGDNSNDIPLSQTDLETAGSSSLDNPISTP VSTVSTGWSSTALQISPQTTEPSSEPSSSKAVSHRHATSLHHAHPLPPTPPSSTHSHSLSTANA TITNAQGTLSAQKPEGFFWQLDSHGFPCAKAHCEKRCNLWDGATVICPRCGPYSEVRYCSRAHL LEDIKPHWLYCGQMVFQHPCRETSIPRRVRAGPPLIPCLHHYDMPERHRQAVHFNMNAREGDYF IFTDWLDFVTAGLPGDKTAIRCSNRIMYVVKFDDAAEKDRFRRVLAACLFMTIELPDLTDYLYR LIRDKLRLANAPNHIEPSLRYQFLQEFNVTIQERITGSRHACETDWDGRNRRNCQDPVCRAEYR RLLGSVGGRGYSRMIETLESTYWILRAARTTHPSVKDAMKRMMGEGYAEVAEEDRRAFRRGDGW DGAGSGDMEIEGFNEGDE	196	
Transglucosidase CAK38817.1	MLATPMTPQASHPSSSNMVCSLASTTTTTSSSSSSSSSSSSSSATQQTTISSRPKLTLQTTSLP RTFGTSSTGLSLSIAAGTASPTVRNTFKNAYEVTGPSSATASPSSKHPSNLRFSKPSSPFTTHN PYQLPLGVKSILRNSPLEPTCRRRAGSVATTGPNGGPSARRVFFPAKKQVSYRNPLEEEIQTVH YTARHSDLHDDPEPALEPQSQPQQPEVTSSDEDSDSNASGCPSDTSTSEDEPETGLGKTTSSPI KRKKRKHSNAERQVRAVALMDGIAGPSNPDSLTPQTPRRKRAKRRCEWRWTLGPLENRDKLLHP VQDETGPTSSASQPETIPHESETETPSSDPPLSSASTTLYHSSPSSSVSSDVETENDEWQTHTT HELECAHADQ	197	
Transglucosidase CAK38846.1	MAFWGVAEREVIERAVALEWADAAQVDERKESPNIRGVLSAGPSQPSRGDASEIKPGFGFSSAL LWGAIFGAFGWTRVLRPVGRIPTRDSCSDRSDGTSWKRYLDLTLLSLDEPPTKGTKELEGQRKS QRARETKWALGSRGEKWALPELIILDD	198	
Transglucosidase	MVKLTHLLARAWLVPLAYGASQSLLSTTAPSQPQFTIPASADVGAQLIANIDDPQAADAQSVCP	199	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
CAK44692.1	GYKASKVQHNSRGFTASLQLAGRPCNVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTNASWY FLSENLVPRPKASLNASVSQSDLFVSWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFV TALPEEYNLYGLGEHITQFRLQRNANLTIYPSDDGTPIDQNLYGQHPFYLDTRYYKGDRQNGSY IPVKSSEADASQDYISLSHGVFLRNSHGLEILLRSQKLIWRTLGGGIDLTFYSGPAPADVTRQY LTSTVGLPAMQQYNTLGFHQCRWGYNNWSDLADVVANFEKFEIPLEYIWTDIDYMHGYRNFDND QHRFSYSEGDEFLSKLHESGRYYVPIVDAALYIPNPENASDAYATYDRGAADDVFLKNPDGSLY IGAVWPGYTVFPDWHHPKAVDFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGNLTLNPA HPSFLLPGEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTSVSYLRTTPTPGVRN VEHPPYVINHDQEGHDLSVHAVSPNATHVDGVEEYDVHGLYGHQGLNATYQGLLEVWSHKRRPF IIGRSTFAGSGKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADTCGFNGNSDEELCNR WMQLSAFFPFYRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTTGSTVMR ALSWEFPNDPTLAAVETQFMVGPAIMVVPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKP GVNTTISAPLGHIPVYVRGGNILPMQEPALTTREARQTPWALLAALGSNGTASGQLYLDDGESI YPNATLHVDFTASRSSLRSSAQGRWKERNPLANVTVLGVNKEPSAVTLNGQAVFPGSVTYNSTS QVLFVGGLQNLTKGGAWAENWVLEW		
Transglucosidase CAK44966.1	MLANSLVVLAAIVASILNPVLGAPALDVGVTEPQAEPKYVFAHFMVGIVENYQLEDWITDMKAA QAIGIDAFALNCASIDKYTPTQLALAYQAAQQVNFKVFISFDFAYWSNGDTGKITAYMQQYANH PAQMQYRGGAVVSTFVGDSFNWSPVKQATSHPIHAVPNLQDPAAASSNSQRGADGAFSWYAWPT DGGNSIIKGPMTTIWDDRYIKDLAGTTYMAPVSPWFATHFNSKNWVFICENLPTLRWEQMLSLK PSLVEIISWNDYGESHYIGPYSANHSDDGSSKWANGMPHDAWRDLYKPYIAAYKSGDSKPTIPQ EGLVYWYRPTPKGVNCPEDNMPAPNGFQMLSDSIFVATMLSSPATLTVTSGSLGPVKVDVPAGI VTTNVTMGIGAQTFQISRNGQVILSGKGGLDVADRSKYYNFNVFVGSVMGSSAAGNASRMLLLL HTTLLKVLLSGDKQVNVCSTTGSKGVICHLLIPDQVTSLIIPKFLQHIVQGKKYN	200	
Transglucosidase CAK47997.1	MKRLMYLLVVLLLSYVVCALPYDDHGKKRDLGPLSDLPGGDVIVWVDQAGNALANNVVGGGNSD PTATADNSPTTLPPILSTLDGDLDLSPAVPLPASTNLPKTGNYRRFGISYSPYNNDGSCKSQDQ VDEDLDKLAQYGFVRIYGVDCDQTNKVTKAARQRNLKVFAGVFDLQNFPSSLDYITGAANGDWS VFHTINIGNELVNDGKNSAADVVNAVNTARSKLRAAGYQGPVVTVDAFSVMIQHPELCQASDYC AANCHAFFDNNNTPDKAGQYVKDQANKVSKAARGKKTLISESGWPHNGQPNGKAVPSSLNQQKA IASLQQTFTGEDELVLFTAFDDLWKQDSSGTFGAEKFWGIQKH	201	
Transglucosidase CAK4847.1	MPHEERVSSHVRQLLQSLTLEEKVALLAGKNMWETVNIDRLHIPSLKMTDGPAGVRGSKWTYGS LTTWIPCGISLAATFDPAMVEQVGSVLGQEARRKGCQVLLAPTMNLSRSPLGGRNFESYGEDPY LVGVIATAMIRGIQAHGVGACMKHFILNDTETRRFNVDQTIDERTLREVYMKPFTMVLNDPAST PWTAMVSYPKINGLHADISPHILPRLLRQELQFDRLVMSDWGGLNSTAESLRATTDLEMPGPAV RRGERLLAAIRAGEVEVAAHVDPSVRRFLQLLERTGLLGDATKSAEHSEAATDDPIFHRIARDA AQSGLVLLKNDKGILPLKPTTLQRVAIIGPNACQPTAGGAGSAAVNPFYVSTPESCLRDVLHAA NSELQVSYEPGIPSSLRPPLLGKLLTVPDGSRKGWQVSFFEGHALEGPVVASSMWDDSLIYLFS DGDVPAVLDDRPYSYRATGVVTPQESGRYTWSLANTGKAKLFVNDELLIDNTEWTGLTGGFLGC SSADKTASVYLEAGRAYQLRVDNVVTLPVVEAFDNTLFPRISGVRVGLALEQDEPEMLAQAVAA	202	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ARQADVAVVVVGHNKDSEGEGGDRATMQLPGRTDELVAAVCAANPNTVVVVQSASAVAMPWVDA ASGLVMAWYQGQENGHALAAALLGDCDFSGRLPITFPRRLKDHGSHAWFPGEAAQDRNTFGEGV RVGYRHFDAQGIPALWPFGFGLSYTRFQLTNIRVCGRVEGRSPESQPVLIQARVCNVGGRDGQE VVQVYVAPSAGIREAGEMSFPKTLGGFCKISVPAGDSREVSIPIRGSELSWYDARVAQWRLDAG KYACWVGRSSSHIDAELEIEVAEGEDTRQGTLE		
Transglucosidase CAK39248.1	MQVLRCGIHGFHEVLGIDVDEIRFYWTIETDDKHASQLAYRVVLSTDEAAVQGDAIVESKLAWD SGRVMSNEQRNIICKPDNGFQSTCSYYWRVTLWDQSERPHHSAVNHFFTAYPRSHLLPPYSMNQ TYMPHTSLIFRSWFEDEPNRWKAVWIGDGGDKPIYLRKAFDLAQPPARAIMFASGLGHFNMTVN GSPASDHRLDPGWTNYHRRVQFTAYDVTAQLQTGANVLGAHLGNGFYAGDKGEDRFFWPMYEDN TYVRYGNELCFFSELHLFYPDGSHTTMISDPSWRVRRSATSLANIYASENHDRRQYPTGWDTPD FDDADWAFAKPLTGPRGHIYYQTQPPVVLHETFQPVKITEPRPGIVCYDLGQNASTMVRVEVEG PRGSEIIVRYSETIQEDGTVLMPDPLFKEYETGVFSRIHLAGTGAPETWEPDFSFTSARYIQVE GVSLDGSDGRPVIRSVVGRHISSAARRLGTMQTDKEDVNQLLSALSWTFSSNLFSYHTDCPQIE KFGWLEVTHLLAPATQYVRDMEALYTKILDDILDTQEPSGLVPTMAPEIRYMCGPLHDTITWGC AVCLLPDILREYYGSTHVIAKVFPAAVRYMEYMRTKERRGGLIEHGLGDWGRGIAFGNNQANIE TAIYYRCLQCVAMMARELGEMQKAKEFEQWAARIYAVYNRHLLVTDDASRPYAYYTSLDNYPAR DRDAIAQAMALQFGLVPEQHRKDVMAAFLDDVADGRIRAGEIGLRFLFNTLADAKRPDLVLQMA RQEEHPSYMRFLRRGETTLLEFWQDECRSKCHDMLGTIYEWFYAAVLGLKPTGPAYRTFVVDPP YNAEFKHVKGSVDCPYGTIAVEFTRNEQGQAVVNVRVPFGTTAIVKLPRSGKSSAYCREGEESR AVDGGEVSLSHGVYSIIEG	203	
Transglucosidase CAK39259.1	MPSTYLGALATLAVFPCLGQARSTWPLGSGLELSYQASQHQISIHQDNQTIFSTIPGQPFLSAS AGKDQFVEDSGNFNITNVNQARCRGQNITQLAGIPRSDSVKNQVAVRGYLLDCGGEDIAYGMNF WVPRRFSDRVAFEASVDSEANASVPVDRLYLTFASHALEDFYGLGAQASFASMKNRSIPIFSRE QGVGRGDQPYTAIEDSQGFFSGGDQYTTYTAIPQYVSSDGRVFYLDENDTAYAVFDFQRSDAVT VRYDSLSVHGHLMQADTMLDAITMLTEYTGRMPTLPEWVDHGALLGIQGGQEKVNRIVKQGFEH DCPVAGVWLQDWSGTHLQSAPYGNMNISRLWWNWESDTSLYPTWAEFVQTLREQHGVRTLAYVN PFLANVSSKSDGYRRNLFLEASQHRYMVQNTTTNSTAIISSGKGIDAGILDLTNEDTRAWFADV LRTQVWSANISGCMWDFGEYTPITPDTSLANISTSAFFYHNQYPRDWAAYQRSVAAEMPLFHEM VTFHRSASMGANRHMNLFWVGDQATLWTRNDGIKSVVTIQGQMGISGYAHSHSDIGGYTTVFEP PTTSNSSGAIPRSAELLGRWGELGAVSSAVFRSHEGNVPSVNAQFYSNSTTYAYFAYNARLFRS LGPYRRRILNTESQRRGWPLLRMPVLYHPEDLRARQISYESFFLGRDLYVAPVLDEGHKSVEVY FPGHSANRTYTHVWTGQTYRAGQTAKVSAPFGKPAVFLVDGASSPELDVFLDFVRKENGTVLYA	204	
Transglucosidase CAK39383.1	MAGVNRSFSYSRGDDALLRDDEREISPLRSAEDGLYSTSYGDVSPLSAGVQAQNRPFDRGLVSV PEGQTLERHMTSTPGMDNLGPASVGGGISDVPVRNLPAERDFNTTGSDNPYIPAPPDGDIYPSS EAVRYRDSYSSHTGLGAGAPFAEHSTPGTTPSQRSFFDSPYQGVDAGPYQRHSAYSSHDYPLVI NPDDIADDGDDGFPVHPKGAADYRSNANVPGTGVAGAAAAGGFLGKFRALFKREEPSPFYDSDI GGGLGGAEKAQGGRHIIGGGSRKRGWIVGLILAAVIVAAIVGGAVGGILGHQEHDGDTSSSSSS SSSSGTGSGGSDKGDGLLDKDSDEIKALMNNKNLHKVFPGVDYTPWGVQYPLCLQYPPSQNNVT	205	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	RDLAVLTQLTNTIRLYGTDCNQTEMVLEAIDRLQLTNMKLWLGVWIDTNTTTTDRQISQLYKIV ENANDTSIFKGAIVGNEALYRAGSDVASAETNLIGYINDVKDHFKDKNIDLPVGTSDLGDNWNA QLVSAADFVMSNIHPFFGGVEIDDAASWTWTFWQTHDTPLTAGTNKQQIISEVGWPTGGGNDCG SDNKCQNDKQGAVAGIDELNQFLSEWVCQALDNGTEYFWFEAFDEPWKVQYNTPGQEWEDKWGL MDSARNLKPGVKIPDCGGKTIT		
Transglucosidase CAK96650.1	MSRNFHPVNTNFPSTTTLTADPDIIPSPEDNRNNYTSTQSYFCRQDVSTNPTFNSNDQALLDNC SNIDVSQLVTEANCFWGSRSSTLPKNEGYPSVEEYPSTYLMGNHGHGYADENMHETSAPYGPCD HLQVAGAGMDMEMRMTGNVMGKVDETYEANQAAMLAALGMTHLDEGVSHAADDDAKTSSSVTVD DDPEWKEWKKWKEREANGGVRLPPEERMNQADVAAWLGMDSKEEHGVFQTQEEDDYDEDETISY VSDDDMMEMEEGMEDDEVVVNVVEGPSRQVLISSQTGAHAEYDTASFGCDFEEQEEQEDSEEQE NEERWGGDGQPPPSLFPRFYSDNAETGLIIELSSADSDAEETMSDPSPSSLSSSSSSSPSETPL QPHLQEAYALPWTTSPSNTSTITSTTSTPTDTSPVPTPFQPDPHPHPHPHPTTQHYLPPPSSQT TPPFTLLTDLHTHLTHTPQHRASHLRNLASEISNTMYFVNSHTISGDFSAEDAAPVDRVMRIVR EGELKMRYQERYERQKGKLVKRERRVAEREDRVSKREEMVSRREMGVAGWWEVWRERGREVWEG VEGEMMGTMEGNGYRRNGMDTLQRTVRVTREVVRRMDRIVDEGEVDGDGDVEGGVEVRGAYKID YNCDDDIVGSTSYRIEICNPPQRSSAS	206	
Transglucosidase CAK40060.1	MSSTPDDKPQRATAAQLANRKIKEVRRRRPNSAAPSAPAPSFGGGPFASIDPNTVSSTPASSQT ASNGFTFGQSQNQSFPGASSAPSQNGGTPFAFGSGGGGSSSSSFNFSSSFGGTSSASNPFASMN TTTSDQSSKPASFSGFQGSLFNIPPGGNQSPAQQPLPSGSIFGTSSQSNASTGGLFGASTNNGP SASGAATPASGSIFGQNNAAAAASTPSTNLFGQSSVKKPSPFGQSSAFSGDDSMQTSPDAKGSG SQQKPSIFSSAPPAQPSFAGAGSTSLFGASASSAAETPSKPVFDFKATTPSTSLFGGAATPTPS ASTPAPAAVTTAAPASSSLFGAASPAKPSTPFQNPFQSSNLFQTTPSTSQKPDEEKKEEPKPAD SQPKSPFQFSASTTTPGSLFAKSDAPASPAPAAPSTGLFQTSSTKNLFEPKPPATADAEQTKAP ANPFGGLFAKPATPSKPAGEQQPLPSSSTPFQGLFSKPSTSNDAAKTSEPEKQATPGPMSFAPS SGGFSQTSNLFSPKPAASPAPAAETQATAASTSAAPVDSPIKVNGANSSPSVFTNGNTAPSTFG QMQTPKLAGKTTDPKTSEDAEMLYRMRTLNECFQRELAKLDPSSQNFDAAVQFYMRVRATLGAS VGSKRKASGEAEDAVATKKARPFGIPSEKADTPKENSTTPAVSAQSSTPFKGFGTSQASPASSK RKSIDEGDDNSPAKRVNGDSSTANIFAQSFSKSKTEAEKPEPSVVKPSTPESTKPALFSTTPTT APAKPLFSLSDSASKGSASTSLFSSSMSSATSTSAASGGNAPKNPFVLKPTSSEGSSTGSAGGT DFFAQFKARSFEDAEKEKEKRKAEDFDSEEEDEAEWERRDAEEQRKKQEQFGTQTQKRAKFVPG KGFVFEDESNESPAKKAEDSSSTTPGAGTIFSSQNNTAVKSNNIFGHLSATPSEAEDNDNDADD TEEASTPGDESDDAAENAVAADKKAESADSSAKEPEAGGRSLFDRVQYGEDGKPKRQGEEEPKG NVSTLFGSSNFSSSFNTPTSLTPASSGESNLAAPKPATTNLFGAPSTTSSIFGTPLSGSGNSTP SIFNAAQNATKSTGDNTWKPDSPIKFASDSASASSSKPDSGSATPALEAPKPFSNLFGAPPSLT KSSTSKDAQPSLGFTFGTPGQSSPSVFAPSTLTSAAPSRSTTPGGASDTGAEESGDGDGAESLP QLDLTRGGAGEENEDLVAESRARAMKHTTGTGWESQGVGFLRVLKDRTTSRGRIVVRADPSGKV ILNTRLMKEIRYSVAKNSVQFLVPQSEGPPQMWALRVKTNADAERLCKSMEETKN	207	
Transglucosidase	MSNRWTLLLSLVILLGCLVIPGVTVKHENFKTCSQSGFCKRNRAFADDAAAQGSSWASPYELDS	208	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
CAK40395.1	SSIQFKDGQLHGTILKSVSPNEKVKLPLVVSFLESGAARVVVDEEKRMNGDIQLRHDSKARKER YNEAEKWVLVGGLELSKTATLRPETESGFTRVLYGPDNQFEAVIRHAPFSADFKRDGQTHVQLN NKGYLNMEHWRPKVEVEGEGEQQTQEDESTWWDESFGGNTDTKPRGPESVGLDITFPGYKHVFG IPEHADSLSLKETRGGEGNHEEPYRMYNADVFEYELSSPMTLYGAIPFMQAHRKDSTVGVFWLN AAETWVDIVKSTSSPNPLALGVGATTDTQSHWFSESGQLDVFVFLGPTPQEISKTYGELTGYTQ LPQHFAIAYHQCRWNYITDEDVKEVDRNFDKYQIPYDVIWLDIEYTDDRKYFTWDPLSFPDPIS MEEQLDESERKLVVIIDPHIKNQDKYSIVQEMKSKDLATKNKDGEIYDGWCWPGSSHWIDTFNP AAIKWWVSLFKFDKFKGTLSNVFIWNDMNEPSVFNGPETTMPKDNLHHGNWEHRDIHNVHGITL VNATYDALLERKKGEIRRPFILTRSYYAGAQRMSAMWTGDNQATWEHLAASIPMVLNNGIAGFP FAGADVGGFFQNPSKELLTRWYQAGIWYPFFRAHAHIDTRRREPYLIAEPHRSIISQAIRLRYQ LLPAWYTAFHEASVNGMPIVRPQYYAHPWDEAGFAIDDQLYLGSTGLLAKPVVSEEATTADIYL ADDEKYYDYFDYTVYQGAGKRHTVPAPMETVPLLMQGGHVIPRKDRPRRSSALMRWDPYTLVVV LDKNGQADGSLYVDDGETFDYERGAYIHRRFRFQESALVSEDVGTKGPKTAEYLKTMANVRVER VVVVDPPKEWQGKTSVTVIEDGASAASTASMQYHSQPDGKAAYAVVKNPNVGIGKTWRIEF		
Transglucosidase CAK96888.1	MPLRPPASPLSLETISPSPPDLDSDPLIASDDDLDDDDRAARDQRIEKLAQAYCHGTPLFILSA SLRGPFENGWANPWKKDRRTTGGGHVGIHSEHSERPIIPETIPQKRPLYRESLGISRSKSAVPL SDPTYSSKKRESQGGTAGEPSSKRPRDSRGRTSSSNNTPKPIALHKRTIETSGHSDALTPFQHT EQSWLKRDRAEINFRKVDPPTSPTTTVSSRHREGGHYTIQVPGTDYRVTTKNILRARTDSRDGT TFDSHVPDSVKAQLTSRHPGVRPETEDHENSICVLSSTSHLSKFEFRRRKRSADPEHGTTSPVP MQLENEQVSHQTTVVEDSRVSSQPAPVPPNTTSMQAIEVDMQVNEDNSHHPNVSERSGTVDHQG NSQSRKVSSTTTAEGVNISDKYPSAQRVSVNPALAENVTSLQTISAVKPNSECDNDTIPDLHFN TQAALLHAQKSFQNDLESQVPNPGETHNQPSSPANDITPFHRMNTSRIGKYSRAIHPGTAHMPM STQCIIDAVTPFTFSTEKKARSRFISPQMPSSSRIDRGTATPNTGSPLSSESEDEDEDEDPTIL PHKSPAAQQQTPDDTQEGSALPMALSHSHPTTVQDGQGVAPGPDSFNLSQAIADAGSWLQQSFD INHEIQQCRSSAKSRPSSSAGISRSASGTILGHSTLDSVLLC	209	
Transglucosidase CAK45960.1	MPGHSRSRDRLSPSSELDDADPVYSPSVYQREHYYNNDSLFDSADDDYTRTPRNVYSYETHDEY HDDDDDDDDVHEHDHDHEYDDKFEEPWVPLRAQVEGDQWREGFETAIPKEEDVTQAKEYQYQMS GALGDDGPPPLPSDALGRGKGKKRLDRETRRQRRKERLAAFFKHKNGSASAGLVSGDALAKLLG SQDGDEDCLSHLGTERADSMSQKNLEGGRQRKLPVLSEEPMMLRPFPAVAPTGQTQGRVVSGAQ LEEGGPGMEMRHRGGGGPPAEGLLQKEGDWDGSTKGSSTSARPSFWKRYHKTFIFFAILIVLAA IAIPVGIIEARRLHGTSGGDNSSNSNLKGISRDSIPAYARGTYLDPFTWYDTTDFNVTFTNATV GGLSIMGLNSTWNDSAQANENVPPLNEKFPYGSQPIRGVNLGGWLSIEPFIVPSLFDTYTSSEG IIDEWTLSEKLGDSAASVIEKHYATFITEQDFADIRDAGLDHVRIQFSYWAIKTYDGDPYVPKI AWRYLLRAIEYCRKYGLRVNLDPHGIPGSQNGWNHSGRQGTIGWLNGTDGELNRQRSLEMHDQL SQFFAQDRYKNVVTIYGLVNEPLMLSLPVEKVLNWTTEATNLVQKNGIKAWVTVHDGFLNLDKW DKMLKTRPSNMMLDTHQYTVFNTGEIVLNHTRRVELICESWYSMIQQINITSTGWGPTICGEWS QADTDCAQYVNNVGRGTRWEGTFSLTDSTQYCPTASEGTCSCTQANAVPGVYSEGYKTFLQTYA EAQMSAFESAMGWFYWTWATESAAQWSYRTAWKNGYMPKKAYSPSFKCGDTIPSFGNLPEYY	210	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Transglucosidase CAK40856.1	STSYGGTRTPDSSSTDVSRPSDLRTGPATRAGSGLTPSLDPSSRPLASRPANRDRIPPPPPKSH HGKRIAPSPGVTPSLTQTTPGKATNRFSFHGSPSEPSYSPRPPQSGSDYFSAKPKDEPPSTEQS TESLRRSQSQHKRPPTPPLSRRHSQMRRSKTTMSKVNPLRLSIHVAQASSAASSSSSPPPSPSG WSLNPARTRESRTGSTPSEEPMHTATSTLRPEPSAAAPVSPTETSQSTSTGSSTKRTSLYNPLP PPPPPRRSRGSSNHSIDSSGQSLRSGKPADETTAAPAAPAAQDEFVPHPSNAHDILADLSRLQK EVDDLRGHYESRKASQ	211	
Transglucosidase CAK40944.1	MYISKVLLVTCAAFAPFASAAVQAKPTDTPVPVSSTHVASSPLAPTPTPVSPSPLHTASSSVII SSSSSSVRFHPSSSASPSHMASSSRRISSSAISSSAIASSSASFTRSYITKASARPTTTTSTDA DSKSNSNSGSDSESATAAAASATHSGAAAPAVQLSGGMAAGVLAAAGFIML	212	
Transglucosidase CAK41060.1	MPRSTVDQTSSAEAPYSGPRKLVLCLDGTGNQFMGFERDSNIVKIYQMLEKNTPGQFHYYQPGI GTYVEGQSSSSGLLRYPRKLQSNIITTIDQGVGTTFESHVLAAYRFIMRYYSPGDHIYIFGFSR GAYTARFLAEMIHELGLLSQGNEEMIHFAWETFSNFQQARGKTDRTAKDEALISYMKKFNTTFC RPQVQIHFLGLFDCVNSVGQFEIPFHRKSHQYLVSPAARHIRHAVSIHERRLKFKPALVLLDKT KPVDLKEVWFAGNHGDVGGGWSLAPGQFHLLSDTPLNWMLQEVLHLENSESKLSFHTLNVADVV ERENAFPGKEEPGTTAYDVRKRTNQPHDMLMFNRGATFLMVIFWWILEILPLFTRLELEHGKWV PRQWPPNMGAPRDIPEDAVIHQSVHEMVRAGILDPKSIPPRGGNNSHLPSTARITGAWKAMRKN QEKQISSLLQKKPAGALRKEFDGKAD	213	
Transglucosidase CAK41144.1	MDPANEYCGLEDYGLVGDMHTCALVSKNGSVDSMCWPVFDSPSIFCRILDKEKGGHFSITPDRR LKNPLSKQRYRPYTNMLETRWIHEEGVMNILDYFPIAKPKPHVACRSGMVRKAECVRGEMEIEI ELFPAFNYARDSHVAQQSSASDDAIQVYHFQAESQNLVVSVLGDRGDISGDDSDLSIEFELSDR PGHLGPGLVGKVTLKEGQSITMLLHDQESITCNVEDLAPYLQQIERTTGDFWSDWTSKCTFRGH YREQVERSLLVLKLLTYKPTGAIVAAPTFSLPEHIGGSRNWDYRYSWVRDAAFTVYVFLKNGYP EEAESYINFIFERIFPPMDKNPKPGEPFLPIMITIHGEREIPEMELEHLEGYRGSRPVRIGNGA ATHIQLDIYGELMDSIYLYNKHAADISYDQWRAIRRMIDFVIQIRHQPDQSIWEVRGPPQNFVY SKIMLWVALDRGLRLAEKRSNLPCPDRARWMHERDALYDEIMTKGYNSEKGFFCMSYENQDAMD AAVLIAPLVFFVAPNDPRLLSTIQKITEVPAKGGLSVANMVSRYDTGKVDDGVGGNEGAFLMVT FWLVEAMMRQLRKTATSQFDSILSFANHLGMFSEEVATSGEQIGNMPQAFSHLACQYMFNVIVK RKSLETCEYHIALDVEKRILQICYNEDISIFNKWLSISGFTYRQSFTISQLVVV	214	
Transglucosidase CAK46428.1	MPIKLPKGFARRKSSSNALEEVQNPTQSSFRVFERPTGDKKSFSDGNLVAKRLSEGQPLDSPSE DDNNIFALHNSQPARHQYEPPLSPDEYLTPFAHPDGPQPESQSPHTRNLYDIPIPPLSGAIRAA GRTFSFGGRFSKASAPTPPPQPSTPGPSRSRGMTTSTSSTATPPKLPDTELRIGKIDDDFQNMF GDIGKRYYGSKDASLDQPVDLDSSGPSSRPDALPRKDERVSRPTPIDTDRSREVEPSPYSWDSR HSEEGLLTTLDSPNEQPATQPYQRNIDPVSVGDRRKSIPLPGTTPLATTSHRSLAKPRTTADKG LRRSGVYSNRRDSVPVEDEDAKLIMESLYSKRSSQVPFMADHGASDAENDGPLFDQPGANSSHP DRRESIQKDSLSSPVLSDHLDPSIAAHARLAAQYEKAQPVTVSSTNKVMTPSQFEHYRQQQELR RSNSDASKSENSAESDYDDDDEVEKDREAERQ RRKQEAHLSVYRQQMMKVTGQESPAPAMRTELDQASKSAPNLLQPGTTLGSGKSSDGDDDEEIP LGILAAHGFPNRNRPPSRLAPSSSIPNLRASFQQPYLSSPSPASVAERDPNNRGSLPVFARNLP	215	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	RDPYFGASLVNQSNRESLALGGGASVHGGPSPALPPGGLVGVIATEERARAMRRGSPNTQAMYD YQGGMPVPPVHPRGIPRPYTMMSLSSPNAGGVQPTISATEQAQIELSQQMTQMVQMQMQWMQQM IHMQGGQSTQLMPPGGPPPTLGANLNARPSSMPSAGNMNNPHAGYSGDQRTLSMLDPNVSSRLN SAAMPHVSGGLRPSTPAGQGYAPSIAPSERSNVGLAPRYRPVSTIQPDLGNVGSPSIPKSWSDE NRKSSLSAAVPAAPQASQMSHRPMPSNSKPIRASKLNVGADQDDEDDDEGWAEMMKKRENKRNN WKMKKETSSFGDLLNAVH		
Transglucosidase CAK41498.1	MYGSQSGHQSSAPPQPEWRLPSSTQQPASSRHHPPQPSWRASSPPPPPPPPRPTTTTTSSSSPF NPTVYGQISNPPSTVNNYPGTSVSPVSAVAGSETTSWGVKYNRHQLHAQSPPPLPPRPSSTAQS PQAQSPVVSPLDPNKPLPAAPGWATQSADNTSYQQWPSNPPYAPQQSVSASSLQPPPPPPAIST GYQSSSAQQSNPWQQPPAVPPPPYSGVPLGQYQDSSVQQPATLPSNQQITGHNAPNPAIPQAPS PKPSTGLHYESQPLPGPPQAPVAATLSPTHGTPPVVPPPVPPKTSPITVPTSASVLATGGPSDW EHLSPIPGSIDDLGAFGSRPQDGSSSEPLSQASQSRPPNIGEPVRKDESVSPITPPNNAPQMTS QTEGPLASQTVVRGNPHQPVRMGSTGSVSSDISTSETPESIDGIIEAWNRPISSQPSAEQNPQS SASGVRAGILPPSRKQSPIGTPTPRQESIIPR KQVRSGSSSVESSSVTNGPTTDKRTILPAFVPLDPYDDLDPWSKSSLERYVAMLRKEAVADSDA ERYNIFTAFMAKETKLREILFNIEPESTRVGENPKVSSRQPTPILRASTSVSNDDTESGLIPVE TEGGHVVSTTDDADSEDGSYSPGGRPILPRIQTPGATKLQRSASHTVSNKYNTDHVAHATSSRA TSVPPSMLGDARHEHALPPLTTNPPQPIYIPFRYTEGPQRGSDVLVFDRPAYQAYSDLRQASAE SGRVMSNAPAPTPGERPDSAVPSRRNEHDETFIGLIREKSVAYRKRAPRKTSSPPPLPAALRHG KPASPVDDLRSMASSPLSKQSESSWNMTTRKDLENYSSDFSYIREAVKSWEISSKSRREQLDKE RIHRQEVSEKRIDALFNGKEIGYADINLLEEEFRQKEARAQLDEERQELDKFVAEVFEPLDQRL KEEIAALQALYEAALAQLDHENGRTKSATTDR YNLSHTMRTVNEIYRKLELRYQKRLEIALDRERRRKKAERRPLVFMGDSVALKGVDQEFDQMEK RNILEAARERDHRANRLMDSFDDAIMHGLGENQSLLDEVAAKVAKVDTATIRSSGLPESEVEQL LKSVYNLIESLRKDSESILHNFNMADSVLNDADYSVSVAEARYSDADADVFRRLDDEKRKEDTK IQTDLKTKLESIRSGPANIVTSINGLLESLGKPPIIDQTGPSSQMPADTPASVSQHLPAEIAPQ KPQEDPEHQERLRKALENAKRRNAARVNTEISRP	216	
Transglucosidase CAK41767.1	MCNKSNYSSPKWWKESVVYQVYPASFNCGKSTTNTNGWGDVTGIIEKVPYLESLGVDISQTSRE QCLTSLSLVYTSPQVDMGYDIADYESIDPRYGTLADVDLLIKTLKDHDMKLMMDLVVNHTSDQH SWFVESANSKDSPKRDWYIWRPAKGFDEAGNPVPPNNWAQILGDTLSAWTWHAETQEFYLTLHT SAQAELNWENPDVVTAVYDVMEFWLRRGICGFRMDVINFISKDQSFPDAPIIDPASKYQPGEQF YTNGPRFHEFMHGIYDNVLSKYDTITVGETPYVTDMKEIIKTVGSTAKELNMAFNFDHMEIEDI KTKGESKWSLRDWKLTELKGILSGWQKRMREWDGWNAIFLECHDQARSVSRYTNDSDEFRDRGA KLLALLETTLGGTIFLYQGQEIGMRNFPVEWDPDTEYKDIESVNFWKKSKELHPVGSEGLAQAR TLLQKKARDHARTPMQWSADPHAGFTVPDATPWMRVNDDYGTVNVEAQMSFPWEMKGELSVWQY WQQALQRRKLHKGAFVYGDFEDLDYHNELVFAYSRTSADGKETWLVAMNWTTDAVEWTVPSGIH VTRWVSSTLQTAPLMAGQSTVTLRALEGVVGCCS	217	
Transglucosidase	MPLFNAKTLLAGLCAASIVSPSLGLPQSSHPGSSAVANTQGRASSESHPSWTLESDSTAATVST	218	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
CAK41979.1	SNTPLYSPSSSATASLGATQGSLTDDHDGASKSSTRSYGVTTISYSSAPVNNPQSAHDASASPS TPSGPHSHTTTLSSSSAGVPAQSQTTSSSRSHSVVSKETSTRSPSSLFTPSASTETPTNPSHTP STYTISTHSFSSSEHASSESATSFHAVSTSKHTHTHTPTSSTSSSNTPTRSSALTQHETSTSSS TPTRSHTHTASTPASSKANTSSSIKTHTTHSHTEDTTSSSASHTPTSSKSSSSSAEDVSSSPAS HSPTPSTHSITTTTNTDTSQSASITSGPSTTPNSTITTTSSTTTSSVDVYAIIKHLYKLVKDTY PVIKKWKEDPKSVKASDLIKPLKRVIPVADDVLDVLGAPSSLSSSSGDSESVLDSCSSGGGLLG DLIGIASCISSTADEAVSILGSSSDSDSSDESTLSSYFDAFETEGSSLSAVGVTATGSTASSTG TTTTTSSGDTSSKSTKTATSTNTDSDTSTQSTKTKTSTKSTTTSDSSSSAGSGGHSASASASPT STKSSTSTKESSTSTKTKTKSNTESSSKAASSSAAASKTDSSSSSAKSTSTDSTSTKKTTSTKS TATSSSAPLSASSSAHTSSIATTNTTSTSTNSKSSTSTDTTTIIIHTHSGTASGTTTHHTSTVT PSPSRNQTATTLVTTTSSYTPPLCYNHADPDNGAGNVCICTRSNGDYTTLSELPSGSGCSYTSI PTPTTTSTTKTTSTKTTSDPPFTVTELNSDVIVCATSTLSYFSTFTYTQCAGSSSTIYTAPTPT PTAQVVIAYLSDVYSFWSFFTPDIGSSIDFCNDAEAGELEASGSIKVIDPPYPDGTKELDFEIH DMKDCVYKGTSDEPGTFTCPDLAKTVDCESYGENKVHDCYGALSDGGVVYEEGIDCASIGSVEV QAENGI		
Transglucosidase CAL00956.1	MHLSKISAILTPVLNAAAVLSSQAPADDLSVLSSEVARANNQSLLWGPYKPNLYFGVRPRIPNS LFAGLMWAKVDNYATAQQNFRHTCEQNEGMAGYGWDEYDIRKGGRQTIHDAGNSLDLTIDFVKV PGGQHGGSWAARVKGVPRGDADPDQPTSVLFYAGLEGLGNLGVEGEPEDPRGFTGDVKLGGFTT DLGDFSIDVTSGPESNEYPEHGHPTYDEKPLDRTLVSSLTMHPEQLWQTKVIMFTQMKKEVDEM VEKYGSENPPPPYQLFTIKNEPGDGNMHLVQKVFKGSFEFDILFSSASSPQPMTSELLTEQISS ASLEFSERFESVHPPQAPFDTAEYTEFSKSMLSNLVGGIGFFHGTDIVDRSAAPEYDEENEGFW EETAEARGRAQPILEGPKDLFTCVPSRPFFPRGFLWDEGFHLIPVIDWDTDLALEIVKSWLSLM DEDGWIAREQILGSEARSKVPPEFTIQYPHYANPPTLFIILEAFIDKLDAKKNASMQTYADSGV TGNLRSIFVDQPELGEAFIRSIYPLLKKHYYWYRSTQKGDIKSYDREAYSTREAYRWRGRSIQH ILTSGLDDYPRPQPPHPGELHVDLMSWMGMMTRALRRIAVTIGETEDAEVFKTYETAIERNIDD LHWDDDARTYCDATIDEYEEHVHVCHKGYISIFPPFLTGMLGPDSPRLKAILDLIGDPEELWSDY GIRSLSKKDQFYGTAENYWRSPIWVNINYLVLKNLYDIAIVSGPHKEQARELYSNLRKNLVENV FQEWKKTGFAWEQYNPETGSGQRTQHFTGWTSMVVKMMSMPDLPASEQKGHDEL	219	
Transglucosidase CAL00976.1	MEVMDMPKRNSPVSQPTAVAMASTSPHPEKKASDAPYIVDDDFPGDDDDDDDVSISPISERAPP WSGTRWARFFPELSSHFSLASPTNSTNPPFPQPLTKGPPHIDGPSQQPERRSKGPSSLSSEDVA DNRSSSYTSRSSLTSQGSEATSPVHKLVDSLHIKSPTKAGVFDESKFAHQIPPPFPSRSSIAQS KDKPLPQEPPIELTPLSIRHKTPQIPDRPGYLSRLDPPPRSKKHASHHHPTLSQACTDLERTLA GLAEQQHSPAQLSPRSPLQILDGPLQISRGNMDMVATRPAPRPPASVHDNRQIHKAKSREDMKQ TKKLLKNKPSFSFTVPAFGRKLSRVHHRSTSNTSSKSEPESYRASVLHQPAVAELGDSEVAELQ GSSVIGFRERPSSAGGEKELRMRLPRLQTKEMGAPGRKRDNIHSTHEGPEQLRRPNGRARGASV GEKYFVSYSKLDGMPVHSTRQHQPTSQTSCMVYELEGGSTQPPAELQGDTTSPIDVVPVRISVG VSSVGAMPGTLPDRVILTVLEHITSLDDLFNVAVSRKDFYRVFKMHELKLIRIAVFAMSAPAWE LREMSPPWDTEWHFVLDPDAPVPEYTPSCYLKRYAEDIYTLAHLKSLILARCGTFLRPETIRGL	220	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	SGTDDIRAAEVDDAFWRVWTFCRLFGSGKGREGDIAGQLDWLRGGEVARNRRVSEFTSIADPYD ANSVLFEPPTGFGDGNNGGLSKEQLLDMTEIWTCLGVLLQPMHEEWAYYILTLGLSAVLVLGSI HPYDNTTAVFQRAHSMGLTNWEASDTGASRSSFLREAVSKACQPRGSSTSQASMRSSGFSSQPS GSHDVSQTSNVRGEREPSPDFHRRRQAAYSAQLRIQRQQQPSPPNPMLAEERPISHYATIMSRL EGLPPAPQPPMSVSRIEIPPTTHSYMTNVSYMQPLQSVTPVYYPPQVRDPVDHAIDIMVRELGF GEEDAKWALKITDSGEGINVNAAISLLTRERKTHEQSSRGFSLRKRKSFLSSVINSPESRHSGW KWA		
Transglucosidase CAK42352.1	MEPLRRSQSSRSMRRSHHSSQSTEPFDPELARFQATTAASRAMLRSKCSYDVLGGPSKMAVPQR QHRPAGAALNATNAPVEDVDLRRSVLDKTSDLSPPAGLPSIREFGRLDAGIATLPSSYRRLRKT RSMFTNWQRSSHVPRGLSSPGCPTHNILTRREPQDVLRAPGTLRRSMSFFRGDTQNSDSLRYAR GQDVAIEMARSHYQQPEIYPTELRKSSLTVPKSRPFKKTLRSVVSDAESASVSPAIQRSTNVIS YGKARSLSSSLKKGLKKVLELSRPSSARISLGKSSSNDQQRIQGSPSTISAKHSDPLNSGVEGN ALTHSPDTEGTVVYTGVKKSESSESLATSRSRVTSWADSTIANTVITYRADDHSSLSVIDEHDS SCLKPSSSEDVSLTTCRTPKPNCTIDSQRLYSALMKRIDGNKTENASKEIVLGHVREHRAIPTP VSSMYTRRSRKTIRLIASDESLQSPGSYTTADVGTVTPCEPAQRQAQRTHGQKHLQDIRFGSAN LASSKHTIKEDSRDETGNMAMGRSQSPEEDEDSPSMYSRSTGGTSPKTTDPKMGESDPEVANEP GVATIYASQRAIYSSPKRNADQGPEAMQRPSADWQQWVRTQMERIEYLTPTRRHYREDAQIQDE TADLAYRTPSRDRRGFWSGSPDEDLRTTCKVTARNNFSRPFSRSSSVRTTVIAPKEQADTLVPP PPPSDSTPKVLSSSSGRSLFINQVETRPDGMALSPVPALLNNRYRAPESPTPRRDATDKARWRA GGRRYGRQPSRLLPEAQDSKASQIRSSRVPQENRRLTDENVRLENGYQEVASKDSQLQNMYSPI SSKRMVEMFLESRRRRMGTEMSDGAPSKDDGTKLSSDSVYDRHHIESLFYSLAPMYETPTN	221	
Transglucosidase CAK42453.1	MANIIWLALVLVALSIHVQAKDVFAHFILANAENFTQTHWTRDISAAKAAQIDAFALNTGYGAA NTDQLLTDAFTVAAAHDFKLFLSLDYSGDGHWPPDQVLKVLQGYANHTAYYRVDNKHPLVSTFE GYEALADWSTIKEKLPNIYFMPEWSVRTPQELASEDAVDGLLSWSAWPYGTTPMNTSTDEQYIS ALKAKDKPYIMPVSPWFYTDMVRYHKNWVWQGDGLWHTRWKQVLDLQPQFVEILTWNDFGESHY IGPLHENELGIFSFGQAPFNYASGMVHDAWREFLPYVVGEYKNGSGKGVIDKEGVVVWYRVTPA WACKAGLTTGNSVTQGQQTMPPGQVLKDEVFFQALLEDTADVEVSIGGGENKSVGWTDTPSGGS SGGGRGLYFGSVPMDNRTGEVVVTLSRNGKFVAQMIGEKITTQCPDKLTNWNAWVGTAMSNVSN ASTSRASLSEENGAASVRVGGGRGMDMWMGALWMVVVVGIRADRSIPTKWACGSQINEEVLIQR RERLPLVEGDRVYLDSSSKAFRRRRRTCTR	222	
Transglucosidase CAK42457.1	MKVPADHALLLSSLLLAPSVGASTCQEPINHPGEPFSFVQPLNTSILTPYGGSPPVFPSPETKG KGGWEKAMAQAKNWVSQLTVEEKAWMATGQPGPCVGNILPIPRLNFTGLCLQNGPQCIQQGDYS SVFVSGVSAAASWDRKLLYDRGYAMATEHKGKGTHVVLGPIGGPLGRSPYDGRTWEGFAADPYL TGVCMEETILGIQDAGVQANAKHFIANEQETQRNPTYAPDANATTYIQDSVSSNLDDRTLHEIY MWPFANAARARVASFMCSYNRVNGSHSCQNSYLLNHLLKTELGFQGYVMSDWGATHSGVASAES GMDMTMPGGFTVYGELWTEGSYFGKNLTEAINNGTITTDRIDDMIVRIMTPYFWLGQDKNYPSV DASVGPLNVDSPPDTWLYDWKFTGPSNRDVRGNNSAMIREHGAASTVLLKNERNALPLRKPRNI VIVGNDAGSDTQGPSTQTDFEYGVLANAGGSGTCRFSYLSTPQDAITTRARQYGGRVQTWLNNT	223	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	LITEKSMPELWNPEQPDVCLVFLKSWSEENVDRTYLTLDWNGNAVVEAVAKYCNNTVVVTHSAG VNVLPFADHPNVTAILAAHYPGEEAGNAIADLLYGDANPSAKLPYVIAYNESDYNAPLTTAVAT NGTYDWQSWFDEELEVGYRYFDAHNIPVRYEFGFGLSYTTYNLTKLVAAKPVASNLTALPEQRA VQPGGNPALWDTVYTLTAQVSNTGSVDGYAIPQLYVGFPDTAPAGTPPSQLRGFDKIWLEAGET KKVTFELMRRDVSYWDVTAQDWRIPAGEFTFKAGFSSRDFHANATATFFRK		
Transglucosidase CAK42741.1	MHLRRIFVLTVLSYVTALPSDINLGVALRGCDVEACDMECRMAGSIGGNCGGNPALNLLGLPLL STNTPNSSSAGPVTLAATETLTDVESTTTTKTTTDRESVTATETTTDIESTTATQTVTDTHLLT VTKSITEKQPTTATQTATDTKFLRTTQTINNTLTATQTTTDIESLRVTKTKNNTITATQTTTDI EPTTATQTINNTLTATQTTTDTESYTAMEITTATASFTTTQTTTDTESITATQNITDTQTIHHT ESLTATRTITDTDSVTATATPTTVTDTQTSISTTTATQTATPTPEVGACFCCTEQVRLPYELNG NCDNIPVSNSTDGCPSGDDPKRNHLLCCDSSGYCTQLS	224	
Transglucosidase CAK46804.1	MGSYTFTWPYNANEVFVTGTFDDWGKTVKLDRVGDVFEKEVPLPVTDEKVHYKFVVDGIWTTDN RAPEEDDGSSNINNVLYPDQILKDSTTPLLNGTAAMAGVTPGSTTAALAAGVPKESSSKHGQNG YYPTISSAAPGSTTAALGQDVPLEQRANVPGSFPVTPASEADKFSVNPIPASSGAGNPIKLNPG EKVPDSSTFNTNTISSTARTDRAGYEQGTSGGFPGSPAYDASAFAIPPVSKNMIPESSLPMGEN QGATEPTYTIQSAAPTSTTAGLAAAVPLESQRQTSSGAPTRDVPDVVRQSMSEAHRDPEAATNK EAVDEKKEMEEELRRKVPVDNSTGAPAPTTVAGLGTSSGLGFTAGAAPSTNLGPSTGLDVATGM GTTTGLDSVSGPTAAQSFQKETTSGLPAHDVPDVVKQSISEAHKDPEAAGVEEAVGEKREVEEE LQQKVPVSNQSGTPAPVITAATSETAPGSGAE PASERAPRATGGGPASAQISPRATTPTDGPTVTTGVATSKAPEESGPGASGREETTEIPTKPAA GATGASATKTVDSGVESGIAPEDTTSAPTAGATGASATKTADPTETSGAPTSGASKPAESAPTN NAAATSKPATNNAAGAATNGKEEKKKKGFFSRLKEKLKSV	225	
Transglucosidase CAK97412.1	MARVDFWHTASIPRLNIPALRMSDGPNGVRGTRFFNGIPAACFPCATALGATWDAHLLHEVGQL MGDESIAKGSHIVLGPTINIQRSPLGGRGFESFAEDGVLSGILAGNYCKGLQEKGVAATLKHFV CNDQEHERLAVSSIVTMRALREIYLLPFQLAMRICPTACVMTAYNKVNGTHVSENKELITDILR KEWNWDGLVMSDWFGTYTTSDAINAGLDLEMPGKTRWRGSALAHAVSSNKVAEFVLDDRVRNIL NLVNWVEPLGIPEHAPEKALNRPQDRDLLRRAAAESVVLMKNEDNILPLRKDKPILVIGPNAQI AAYCGGGSASLDPYYTVSPFEGVTAKATSEVQFSQGVYSHKELPLLGPLLKTQDGKPGFTFRVY NEPPSHKDRTLVDELHLLRSSGFLMDYINPKIHSFTFFVDMEGYFTPTESGVYDFGVTVVGTGR LLIDNETVVDNTKNQRQGTAFFGNATVEERGSKHLNAGQTYKVVLEFGSAPTSDLDTRGIVVFG PGGFRFGAARQVSQEELISNAVSQASQASQVIIFAGLTSEWETEGNDREHMDLPPGTDEMISRV LDANPDNTVVCLQSGTPVTMPWVHKAKALVHAWFGGNECGNGIADVLFGDVNPSAKLPVTFPVR LQDNPSYLNFRSERGRVLYGEDVYVGYRYYEKTNVKPLYPFGHGLSYTTFSRSDLKITTSPEKS TLTDGEPITATVQVKNTGTVAGAEIVQLWVLPPKTEVNRPVRELKGFTKVFLQPGEEKQVEIVV EKKLATSWWDEQRGKWASEKGTYGVSVTGTGEEELSGEFGVERTRYWVGL	226	
Transglucosidase CAK43189.1	MAVRRSARLRSRQATEPEAPADPVVTDNNAPCDTNNHNNSENTSEIDTTMARLGKQPERLPPVV EHEEPADAAKDVPVQRSRKKTKTETASKRRSKVEAKEPVAEVTPVIAESQSNTDTTEPAVAETE KKPTPKAVPEPAVAETEKNPTPKALPETASKLSTPKKSIPTLKGTPVHRNTPVHRSTPVHKSTP	227	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TRTPSSTLVRPSHQEMHPSKVRQSTTKQADSGLILGFKPIKKDAEGKVIKDTLADNTPTKAKAS PAPYYGTPAFEFKFSCESQLSDEAKKLMETVREDAAKIKAQMTLEPDQNRAEAADRKIVQPKGK ASRFSDVHMAEFKKMDSIAGHASAFRATPGRFQPVVKTLKRTNSKARLDESDRNSPSPSKIARP SPAIVAPASNKRVKHDKADDASTRRPTAASPPKPVQPRPRSTVRSSLMTPTRSSAARASSVTAR PPRTSMIPSLVRSPAAKPADVPRTPQTEFNPRLKSNLPTLGNLKSILRRHQPLFSKDPSKIAAG THVAAPDFTSNLLFGSRGTTEEPAQTPSPKKRVEFTPSVKARHEEVMFSPSPSKVPVASPSRTT SDVVYPTLPVLTPEQNRVSAKSPAQATTPTIRHVRPSDVHANPLPEVAGVPHGIGHKKRTRESG EDTKTNDLPEVAGVPHGIGQKKRNRAALEDETDTENVPPVDLTADARSAKRMKMTSPSPLKAPT LSARKVAAPSPTKAATPSPTKPRSHTPLRSATTSRMSTPGISTPASVRARNRGVLSVSRLNMLA QPKNRG		
Transglucosidase CAK43257.1	MSLRWRKKTHWPPSPCVEDEVVSLSRELHGLSQIREMPGLEGVCSRGSVDQYPVLVDVFSYSSY EETTVIYEDFSRDSSSEDNVGPPTPVDEKQDPMLYLVGDDQAVSLSAPLATREQSQDPSKTPAD NEQGSTTRGRPRADTRAQRDSPSKKDTAQSSRDASRASNIRTPAVTQSKSTPSLPRRFGSVKHG RSADALTTKSGYQSDSATVKSKAKPETVDKSAAPQTDKKSPTGLTVAERLEEKIRQRQELRAKE SSGDAPKTPSPPSTDQPSVPVGRAAEPSITPAATAAPKSTAPKTRPRSTSTPKEPTNDHRVDGP EASSSAAAPALQLPPRPGLASKPAGRSVSSNDAKSTAQLPARRAVSFLDDVPQRSSSLPRTPED IPEPPPLPRRRSSSQDAVRHRSSSQDAVRQTSPKRPFFLPPCPRSTPIAGYQDWHTVKGLPHLN ICPSCMKQMRKSEFRDHFVLASPRSRGEKIRCSMSEPWTRLAWMQTLKKQLDHLELLHQITRPP LSIKPCPGRIITEQHWYRIVDPETNMYLPQFNVCSACVRNLRVLMPQHRDTFKRSSTKQERACD FLTDSPRFVRYIDYLDIAANRADQENMLRPDVTEFLSYARRKVVLRDCRRDRRILSTWHYMPQL PELTVCEDCYDDVVWPLVRAKQPIARKFSTSMRLLPGDGPSRCREASCQLYSPRMRAKFAEAVQ SNDLMYLKQVALRRRDAEQRYRDREEELLEDASRGYDVEGEMRRNVEEWKRNE	228	
Transglucosidase CAK97469.1	MSLLPVILVTFFLVFCCAAGPIAPFASKRDLESLSPSPSLTTPYVHVASSSVDDDDDNEINALT VVPVTPSSLPQTASSSSTTIEPALSSSAAFIQESSIVAATTSSLSESSSAVTHSFAPSTSSDST VNTLSQTLTSTTTTTTTTSSPTTLGEPSSAPFSPLSSAVRSSHTTSSSSHIMHITTPSTTLSES SQIVTPSIIIPGGPMEASSSHAPGTATSSHITHETSSSAVRVSHSSSAAAEMSKSSPSHQGTLN SSSRLLHSSTAALLPSSTAPESAPETSSTRTSETTTSTSLTIGVIVPLPEASTSPSTMLEMSSS TSQSSESITTTADDTSSASSTKVLSTPTESETTTPTSHTASPFIGVTIPTTTKTTQPADPADIT TTTTTPTSEAEDATTTSAPTPVVVLVTPEGSTTVIGTSSFIAPNPDITSTSTSTSSLTTTIEPT PTTSTTEPPTTLHATSTTIQTVYVVITDTPTPEPTWDSTTIATAIITVYDKSTSETVPTTTTTS RAGEEMESTTTTPLDTEQTSTQSTEDEIPTSTPIADTETATAIITSYPSTSTLSGETGDVIRIV PVTPTGPITVTVTVTEKERETVTKTETVTERVTETESVTT	229	
Transglucosidase CAK97480.1	MPQKEFVPKTYQESSTGAQSSSSVHLRSSPEERSFDFSFEPIRENLFRVTFSSQDHPLPPYPSV TKPATSLDGVHVSATGGSNQKTIEVGDVTASVEWSNTPVVSLSWKGTEKPLYRDLPLRSYVADS TGIAHYTEHDRDCLHVGLGEKRAPMDLTGRHFQLSATDSFGYDVYNTDPLYKHIPLLIKASPDG CVAIFSTTHGRGTWSVGSEVDGLWGHFKVYRQDYGGLEQYLIVGKTLKDVVRSYAELVGLPILV PRWAYGYISGGYKYTMLDDPPAHEALMEFADKLEEHGIPCSAHQMSSGYSIAETEPKVRNVFTW NKYRFPNPEEWIAKYHGRGIRLLSNIKPFLLASHPDFQKLIDGNGFFKDPESSKPGYMRLWSAG	230	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GATGGDGCHIDFSSAVAFKWWYDGVQSLKRAGIDAMWNDNNEYTLPDDDWKLALDEPTVSDAVK KGVENSVGQWGRAMHTELMGKASHDALLNIEPNHRPFVLTRSATAGTMRYAASTWSGDNVTSWE GMKGANALSLSAGISLLQCCGHDIGGFEGPQPSPELLLRWIQLGIHSPRFAINCFKTSPGNSSV GDVIEPWMYPEITPLVRDTIKRRYEILPYIYSLGLESHLTASPPQRWVGWGYESDPEVWTKALK SGDEQFWFGDTIMVGGVYEPGVSVAKLYLPRKANDQFDFGYVNMNEPYNYLASGQWVEVPSEWR KSIPLLARIGGAIPVGKPVHTRVPGDDTPASVAVKEVDDYRGVEIFPPLGSSHGQVFSTTWFED DGISLEARISEYTVTYSSTEEKVIVGFSRDEKSGFVPAWTDLDIILHNGDERRVVSDIGKTVEY KGKGSRGRVVYTLKN		
Transglucosidase CAK47332.1	MNEGRLAHPQFNQYSFKAGASTVQAEAAPALNYEDASTHNAAKNATKRSGRKGDQTYTYSIPPE LAEAARLVAEASPQPVPTDYGVDISLVVSKYRKYDNNDTNVPKQKYVEPNGLDGYVHTGQPEDS PEIHTELKKRATTDFWLTQMGDSGSSPYAPDGYKVWRNVRDYGAKGDGITDDTAAINKAISDGG RCGAECGSSTIYPAFVYFPAGKYLVSSPIIQYYNTEFYGNPFDYPTILAASSFVGLGVITSDVY TGDDTEWYINQNNFLRSIRNFKMDITRTDPNAYVCAIHWQVAQGTSLENIEFYMMQDGLTTQQG IYMENGSGGFLTNLTFVGGNFGYVYAFSQRCTPLSDLPSGHTLATQFTSTSLTFMNCKTALQVH WDWAWTMQDVVVENCTNGIVIVGGAGGPKSTGQSVGSLILVDAVIAHTQTGIVTTLLAENSTSF LLQGVVFIEVDTAILDSAQGKTLMAGGSNVPVFSWGFGRVVTTGAESTFYNGQDIPRTNRSVPL TTIGYIEPNFYLRRRPTYRDIGMSQVINVKDWGAAGDGKTDDTAVLNSILDRAANMSSIVFFPY GVYIIRDTLRVPVNSRIMGQVWSQIMATGPKFQDEQNPHIAVQVGQVGDRGIVEIQSLMFTVSG PTAGAVLMEWNVHQVIQGSAGMWDSHFRVGGATGSQLQADECPKGSGVVLPACKAASLLLHLTS QSSAYLENIWLWVADHDLDLQDQAQIDVYSARGLLVESQGPTWLYGTASEHNVLYQYQVSQARD LYMGMIQTESPYFQNVPPAPSPFSPGLFPNDPTFSDCDSDSQTCPVSWALRIIDSTSVYSMGAG IYSWFSAYSQDCLDTESCQQHAVGISQSTNTWLYNLVTKGIAEMVTPTNEHPTLSADNVNGFMS SILAWVRLANTTIGARKFPGFQLYQPKWLDGLTDTCKTALSQKILCHPYLEMKFSNPGIGQYID NNTLADEVCDQGCGESLQMWTTNVANSCLNQTIDDTDPVAAGGYIYAGYNLTCLRDPHTKKYCP DVLSHFTIVDSVRSMTLAEMCSYCFTTSLEMRQASPYAAYTDVDKDALETVNAECGLSGPTDLH KPLYTEDEVDRPICMSGITHTTSEGDTCDLLAYKYHVASAVIQLANPMLVNDCSELIPGRQLCM PLSCDTQYTLQDNDTCLSIEWAQPIGFGEVRRYNPWLNVDCTNLQTTRQVHGSVLCLSPQGGSH NVTGTGSPCPGISDGYTNVVQYAPTNSTIAKGTTCYCGKWYTVQQGDSCATICIKQGIPSSLFL AVNPSLSTSDCDTSLQVGYTYCVGPDTHWDDTDNFWGEFACEAY	231	
Transglucosidase 1ACZ_A	CTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALSADKYTSSDPLWYVTVTLPAG ESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWX	232	
Transglucosidase ACF60497.1	MSNRWTLLLSLVILLGCLVIPGVTVKHENFKTCSQSGFCKRNRAFADDAAAQGSSWASPYELDS SSIQFKDGQLHGTILKSVSPNEKVKLPLVVSFLESGAARVVVDEEKRMNGDIQLRHDSKARKER YNEAEKWVLVGGLELSKTATLRPETESGFTRVLYGPDNQFEAVIRHAPFSADFKRDGQTHVQLN NKGYLNMEHWRPKVEVEGEGEQQTQEDESTWWDESFGGNTDTKPRGPESVGLDITFPGYKHVFG IPEHADSLSLKETRGGEGNHEEPYRMYNADVFEYELSSPMTLYGAIPFMQAHRKDSTVGVFWLN AAETWVDIVKSTSSPNPLALGVGATTDTQSHWFSESGQLDVFVFLGPTPQEISKTYGELTGYTQ LPQHFAIAYHQCRWNYITDEDVKEVDRNFDKYQIPYDVIWLDIEYTDDRKYFTWDPLSFPDPIS	233	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	MEEQLDESERKLVVIIDPHIKNQDKYSIVQEMKSKDLATKNKDGEIYDGWCWPGSSHWIDTFNP AAIKWWVSLFKFDKFKGTLSNVFIWNDMNEPSVFNGPETTMPKDNLHHGNWEHRDIHNVHGITL VNATYDALLERKKGEIRRPFILTRSYYAGAQRMSAMWTGDNQATWEHLAASIPMVLNNGIAGFP FAGADVGGFFQNPSKELLTRWYQAGIWYPFFRAHAHIDTRRREPYLIAEPHRSIISQAIRLRYQ LLPAWYTAFHEASVNGMPIVRPQYYAHPWDEAGFAIDDQLYLGSTGLLAKPVVSEEATTADIYL ADDEKYYDYFDYTVYQGAGKRHTVPAPMETVPLLMQGGHVIPRKDRPRRSSALMRWDPYTLVVV LDKNGQADGSLYVDDGETFDYKRGAYIHRRFRFQESALVSEDVGTKGPKTAEYLKTMANVRVER VVVVDPPKEWQGKTSVTVIEDGASAASTASMQYHSQPDGKAAYAVVKNPNVGIGKTWRIEF		
Transglucosidase CAY05387.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFRQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSGDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	234	
Transglucosidase CAY05391.1	MSFRSLLALSGLVCTGLANVISKRATWDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSFI LANFDSSRSAKDANTLLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDATG TYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLTA NNRRNVVPSASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVTS TSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALSA DKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	235	
Transglucosidase 1AC0_A	CTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALSADKYTSSDPLWYVTVTLPAG ESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	236	
Transglucosidase CAS97680.1	MSNRWTLLLSLVILLGCLVIPGVTVKHENFKTCSQSGFCKRNRAFADDAAAQGSSWASPYELDS SSIQFKDGQLHGTILKSVSPNEKVKLPLVVSFLESGAARVVVDEEKRMNGDIQLRHDSKARKER YNEAEKWVLVGGLELSKTATLRPETESGFTRVLYGPDNQFEAVIRHAPFSADFKRDGQTHVQLN NKGYLNMEHWRPKVEVEGEGEQQTQEDESTWWDESFGGNTDTKPRGPESVGLDITFPGYKHVFG IPEHADSLSLKETRGGEGNHEEPYRMYNADVFEYELSSPMTLYGAIPFMQAHRKDSTVGVFWLN AAETWVDIVKSTSSPNPLALGVGATTDTQSHWFSESGQLDVFVFLGPTPQEISKTYGELTGYTQ LPQHFAIAYHQCRWNYITDEDVKEVDRNFDKYQIPYDVIWLDIEYTDDRKYFTWDPLSFPDPIS MEEQLDESERKLVVIIDPHIKNQDKYSIVQEMKSKDLATKNKDGEIYDGWCWPGSSHWIDTFNP	237	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAIKWWVSLFKFDKFKGTLSNVFIWNDMNEPSVFNGPETTMPKDNLHHGNWEHRDIHNVHGITL VNATYDALLERKKGEIRRPFILTRSYYAGAQRMSAMWTGDNQATWEHLAASIPMVLNNGIAGFP FAGADVGGFFQNPSKELLTRWYQAGIWYPFFRAHAHIDTRRREPYLIAEPHRSIISQAIRLRYQ LLPAWYTAFHEASVNGMPIVRPQYYAHPWDEAGFAIDDQLYLGSTGLLAKPVVSEEATTADIYL ADDEKYYDYFDYTVYQGAGKRHTVPAPMETVPLLMQGGHVIPRKDRPRRSSALMRWDPYTLVVV LDKNGQADGSLYVDDGETFDYERGAYIHRRFRFQESALVSEDVGTKGPKTAEYLKTMANVRVER VVVVDPPKEWQGKTSVTVIEDGASAASTASMQYHSQPDGKAAYAVVKNPNVGIGKTWRIEF		
Transglucosidase 1KUL_A	CTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALSADKYTSSDPLWYVTVTLPAG ESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	239	
Transglucosidase 1KUM_A	CTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALSADKYTSSDPLWYVTVTLPAG ESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	240	
Transglucosidase ADO32576.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILSNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNEDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTPLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	241	
Transglucosidase ADX86749.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	242	
Transglucosidase AEE60909.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT	243	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR		
Transglucosidase AFJ52556.1	CTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALSADKYTSSDPLWYVTVTLPAG ESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	244	
Transglucosidase CCO73840.1	MSNRWTLLLSLVILLGCLVIPGVTVKHENFKTCSQSGFCKRNRAFADDAAAQGSSWASPYELDS SSIQFKDGQLHGTILKSVSPNEKVKLPLVVSFLESGAARVVVDEEKRMNGDIQLRHDSKARKER YNEAEKWVLVGGLELSKTATLRPETESGFTRVLYGPDNQFEAVIRHAPFSADFKRDGQTHVQLN NKGYLNMEHWRPKVEVEGEGEQQTQEDESTWWDESFGGNTDTKPRGPESVGLDITFPGYKHVFG IPEHADSLSLKETRGGEGNHEEPYRMYNADVFEYELSSPMTLYGAIPFMQAHRKDSTVGVFWLN AAETWVDIVKSTSSPNPLALGVGATTDTQSHWFSESGQLDVFVFLGPTPQEISKTYGELTGYTQ LPQHFAIAYHQCRWNYITDEDVKEVDRNFDKYQIPYDVIWLDIEYTDDRKYFTWDPLSFPDPIS MEEQLDESERKLVVIIDPHIKNQDKYSIVQEMKSKDLATKNKDGEIYDGWCWPGSSHWIDTFNP AAIKWWVSLFKFDKFKGTLSNVFIWNDMNEPSVFNGPETTMPKDNLHHGNWEHRDIHNVHGITL VNATYDALLERKKGEIRRPFILTRSYYAGAQRMSAMWTGDNQATWEHLAASIPMVLNNGIAGFP FAGADVGGFFQNPSKELLTRWYQAGIWYPFFRAHAHIDTRRREPYLIAEPHRSIISQAIRLRYQ LLPAWYTAFHEASVNGMPIVRPQYYAHPWDEAGFAIDDQLYLGSTGLLAKPVVSEEATTADIYL ADDEKYYDYFDYTVYQGAGKRHTVPAPMETVPLLMQGGHVIPRKDRPRRSSALMRWDPYTLVVV LDKNGQADGSLYVDDGETFDYERGAYIHRRFRFQESALVSEDVGTKGPKTAEYLKTMANVRVER VVVVDPPKEWQGKTSVTVIEDGASAASTASMQYHSQPDGKAAYAVVKNPNVGIGKTWRIEF	245	
Transglucosidase BAM72725.1	MVKLTHLLARAWLVPLAYGASQSLLSTTAPSQPQFTIPASADVGAQLIANIDDPQAADAQSVCP GYKASKVQHNSRGFTASLQLAGRPCNVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTNASWY FLSENLVPRPKASLNASVSQSDLFVSWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFV TALPEEYNLYGLGEHITQFRLQRNANLTIYPSDDGTPIDQNLYGQHPFYLDTRYYKGDRQNGSY IPVKSSEADASQDYISLSHGVFLRNSHGLEILLRSQKLIWRTLGGGIDLTFYSGPAPADVTRQY LTSTVGLPAMQQYNTLGFHQCRWGYNNWSDLADVVANFEKFEIPLEYIWTDIDYMHGYRNFDND QHRFSYSEGDEFLSKLHESGRYYVPIVDAALYIPNPENASDAYATYDRGAADDVFLKNPDGSLY IGAVWPGYTVFPDWHHPKAVDFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGNLTLNPA HPSFLLPGEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTSVSYLRTTPTPGVRN VEHPPYVINHDQEGHDLSVHAVSPNATHVDGVEEYDVHGLYGHQGLNATYQGLLEVWSHKRRPF IIGRSTFAGSGKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADTCGFNGNSDEELCNR WMQLSAFFPFYRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTTGSTVMR ALSWEFPNDPTLAAVETQFMVGPAIMVVPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKP GVNTTISAPLGHIPVYVRGGNILPMQEPALTTREARQTPWALLAALGSNGTASGQLYLDDGESI YPNATLHVDFTASRSSLRSSAQGRWKERNPLANVTVLGVNKEPSAVTLNGQAVFPGSVTYNSTS QVLFVGGLQNLTKGGAWAENWVLEW	246	
Transglucosidase	MWSSWLLSALLATEALAVPYEEYILAPSSRDLAPASVRQVNGSVTNAAALTGAGGQATFNGVSS	247	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
AGN92963.1	VTYDFGINVAGIVSVDVASASSESAFIGVTFTESSMWISNEACDATQDAGLDTPLWFAVGQGAG VYSVGKKYTRGAFRYMTVVSNTTATVSLNSVKINYTASPIQDLRAYTGYFHSSDELLNRIWYAG AYTLQLCSIDPTTGDALVGLGAITSSETITLPQTDKWWTNYTITNGSSTLTDGAKRDRLVWPGD MSIALESVAVSTEDLYSVRTALESLYALQKADGQLPYAGKPFYDTVSFTYHLHSLVGAASYYQY TGDRAWLTRYWGQYKKGVQWALSGVDSTGLANITASADWLRFGMGAHNIEANAILYYVLNDAIS LAQSLNDNAPIRNWTATAARIKTVANELLWDDKNGLYTDNETTTLHPQDGNSWAVKANLTLSAN QSAIISESLAARWGPYGAPAPEAGATVSPFIG GFELQAHYQAGQPDRALDLLRLQWGFMLDDPRMTNSTFIEGYSTDGSLVYAPYTNRPRVSHAHG WSTGPTSALTIYTAGLRVTGPAGATWLYKPQPGNLTQVEAGFSTRLGSFASSFSRSGGRYQELS FTTPNGTTGSVELGDVSGQLVSEGGVKVQLVGGKASGLQGGKWRLNV		
Transglucosidase AIY23066.1	MVKLTHLLARAWLVPLAYGASQSLLSTTAPSQPQFTIPASADVGAQLIANIDDPQAADAQSVYP GYKASKVQHNSRGFTASLQLAGRPCNVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTNASWY FLSENLVPRPKASLNASVSQSDLFVSWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFV TALPEEYNLYGLGEHITQFRLQRNANLTIYPSDDGTPIDQNLYGQHPFYLDTRYYKGDRQNGSY IPVKSSEADASQDYISLSHGVFLRNSHGLEILLRSQKLIWRTLGGGIDLTFYSGPAPADVTRQY LTSTVGLPAMQQYNTLGFHQCRWGYNNWSDLADVVANFEKFEIPLEYIWTDIDYMHGYHNFDND QHRFSYSEGDEFLSKLHESGRYYVPIVDAALYIPNPENASDAYATYDRGAADDVFLKNPDGSLY IGAVWPGYTVFPDWHHPKAVDFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGNLTLNPA HPSFLLPGEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTSVSYLRTTPTPGVRN VEHPPYVINHDQEGHDLSVHAVSPNATHVDGVEEYDVHGLYGHQGLNATYQGLLEVWSHKRRPF IIGRSTFAGSGKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADICGFNGNSDEELCNR WMQLSAFFPFYRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTSGSTVMR ALSWEFPNDPTLAAVETQFMVGPAIMVVPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKP GVTTTISAPLGHIPVYVRGGNILPMQEPALTTREARQTPWALLAALGSNGAASGQLYLDDGESI YPNATLHVDFTASRSSLRSSAQGRWKERNPLANVTVLGVNKEPSAVTLNGQAVFPGSVTYNSTS QVLFVGGLQNLTKGGAWAENWVLEW	248	
Transglucosidase AIY23067.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAASSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	249	
Transglucosidase GAQ47522.1	MDPANEYCGLEDYGLVGDMHTCALVSKNGSVDSMCWPVFDSPSIFCRILDKEKGGHFSITPDRR LKNPLSKQRYRPYTNMLETRWIHEEGVVNILDYFPIAKPKPHVVEKGLPQWCRCYQNKGSAQQE	250	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ACRSGMVRKAECVRGEMEIEIELFPAFNYARDSHIAQQSSASDDGIQVYHFQSESQNLVVSVLG DKGDISEDDSDLSIEFELSDRPGHLGPGLIGKVTLKEGQSVTMLLHDQESITCNAQDLAPYLQQ IERTTGDFWSDWTSKCTFRGHYREQVERSLLVLKLLTYKPTGAIVAAPTFSLPEHIGGSRNWDY RYSWVRDAAFTVYVFLKNGYPEEAESYINFIFERIFPPMDKNPKPGEPFLPIMITIHGEREIPE MELDHLEGYRGSRPVRIGNGAATHIQLDIYGELMDSIYLYNKHAADISYDQWRAIRRMIDFVIQ IRHQPDQSIWEVRGPPQNFVYSKIMLWVALDRGVRLAEKRSNLPCPDRARWMHERDALYDEIMT KGYNAEKGFFCMSYENQDAMDAAVLIAPLVFFVAPNDPRLLSTIQKITDVPAKGGLSVANMVSR YDTGKVDDGVGGNEGAFLMVTFWLVEAMMRAAKSKAYLPHDPFFQQLRKTATSQFDSILSFANH LGMFSEEVATSGEQIGNMPQAFSHLACVSAAMNLGGGGDR		
Transglucosidase GAQ47133.1	MSFRSLLALSGLVCSGLASVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVIKTLVDLFRNGDTDLLSTIENYISSQAIVQGISNPSGDLSSGGLG EPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLLTGQDNGYTSAATEIVWPLVRNDLSY VAQYWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPQILCYLQSFWT GEYILANFDSSRSGKDTNTLLGSIHTFDPEAGCDDSTFQPCSPRALANHKEVVDSFRSIYTLND GLSDSEAVAVGRYPEDSYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEITDVSLDFFQALYSD AATGTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSLSEQYDKSDGDELSARDLTWSYAA LLTANNRRNSVVPPSWGETSASSVPGTCAATSASGTYSSVTVTSWPSIVATGGTTTTATTTGSG SVTSTSKTTTTASKTSTTTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWDTSDGI ALSADKYTSSNPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGESTATVTD TWR	251	
Transglucosidase GAQ46031.1	MAKSASQIHRAWWKECSVYQIWPASYKDSNDDGIGDIPGIISKLDYIKNIGVDIVWLCPSYKSP QVDMGYDIADYYSIADEYGTVADVEKLIQGCHERGMKLLMDLVVNHTSDQNEWFKQSRSSKDNE YRNWYVWKPARYDEQGNRHPPNNWVSHFQGSAWEWDEHTQEYYLHLYAVEQPDLNWEHPPVRKA VHDIMRFWLDKGADGFRMDVINFVSKDQRFPDAPVKDPRTPWQWGDKYYANGPRLHEYLADLGK ILKEYDAFSVGEMPFVRDTEEVLRAVRYDRNEINMIFNFEHVDIDHGTYDKFEPGSWKLTDLKA FFETWQKFMYNNDGWNALYWENHDQPRSIDRYAQAKEEFRTEAGKMLATVLALQSGTPFVYQGQ EIGMRNVPIEWDMNEYKDIDCLNHWHRLLKHRPDDIEAQKSARQEYQKKSRDNGRTPVQWSSAP NGGFTGPNAKPWMSVNPDYVRFNAEAQVNDPNSIYHYWAAVLGLRKKYLDIFVYGDYDLVDKDS QEIFAYSRQYEDQKALVLTNWTENTLEWDATANGVKGVKDVVLNSYESAEAAKGRFSGQKWSLR PYEAVVLLVEA	252	
Transglucosidase GAQ44395.1	MRCHRLLSGVLAFLPLSVAQSCWRNTTCSGPTDSAFSGPWEKNIFAPSSRTLNPEKLFLITQPD KTEDYIPFALHGNGSLVVYDFGKEVGGIVSVNFSSTGSGALGVAFTEAKNYIGEWSDSSNGGFK GPDGALYGNFTEAGSHYYVMPDKSLRGGFRYLTLFLITSDNSTIHIEDVSLEIGFQPTWSNLKA YQGYFHSNDDLLNKIWYTGAYTLQTNEVPTDTGRQIPAMAEGWANNCTLGPGDTIIVDGAKRDR AVWPGDMGIAVPSAFVSLGDLDSVKNALQVMYDTQDNSTGAFDESGPPLSQKDSDTYHMWTMVG TYNYMLYTNDSDFLEQNWEGYQKAMDYIYGKVTYPSGLLNVTGTRDWARWQQGYNNSEAQMILY HTLNTGAELATWAGDSGDLSSTWTSRAEKLRQAINEYCWDESYGAFKDNATDTTLHPQDANSMA LLFGVVDADRAASISERLTDNWTPIGAVAPELPENISPFISSFEIQGHLTVGQPQRALELIRRS	253	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	WGWYYNNANGTQSTVIEGYLQNGTFGYRGSRGYYYDTAYVSHSHGWSSGPTSALTNYIVGITVT SPLGATWRIAPQFVDLQSAEGGFTTSLGKFQAGWSKTDKGYTLDFTVPHGTQGNLTLPFVGTAK PSIKIDGTEITRGVQYANSTATVTVSGGGTYKVVVQ		
Transglucosidase GAQ43928.1	MRFRDGMWLVDPSKSLQYAEDIYSINASPDNRSLNLLCPTRHIFSRGNTLNLSTLHINLESHFD GVISLEVQHWLGARKGTPDFELFPDGEGPKLSEERIGISKSERGTTLKSGALSVTVSPDQHDFS IRFHSSDDYDWEVTSLLNRSVGLAYDPPISNGKQVEDLVQGQSGSRKHYIFTQTELDIGESVHG LGERFGPFNRLGQHVEIWNEDGGTSSDQAYKNVSFWMSSKGYGVFIDTPEKVDLEIGSERCCRV QTSVEGQRLKWYIIYGPSPKEVLTKYSVLTGRAPMVPAWSFGLWLTTSFTTNYDEATVTDFLQQ MSDRSIPVEVFHYDSFWMRAFHWCDFVFSPDHFPDPKGSIARIKHAGLTNKVCVWINPYLGQAS PVFLEAAEKGYLLKRTNGDVWQWDLWQTGMGLVDFTNPEAVRWYEGCLERLFDVGIESIKTDFG ERIPTKGVKWHDESVDPARMHNYYAFIYNKIVYNALTRRYGDGQAVLFARSACAGVQRFPLCWG GDCESTPAALAESVRGGLSIGLSSFSFWSCDIGGFEGTPPPWIYKRWVAFGLLCSHSRLHGSDS YRVPWLIDNDDAGPQGSTAVLRTFVRLKRRLMPYLYTQAVQSTRMGWPLSLRATALEFPHDPTA WAACDRQFFVGENLLVAPVFTEHGDVEFYLPEGQWTSLWDEKKVVSGPGWRREKHGFGTLPIYV REGAVIVMGKEQGEGGFAYDWCEAPEVRLYQTKQGDCATVVDASGKEVGTLTVQDDGSLKGLEC FRGDVTVRRIE	254	
Transglucosidase GAQ43980.1	MDFFQTFWSSSPLSKIPSSSFNQTFMCTMCPKSNYTTPKWWKEAVVYQVYPASFNCGKPTSKTN GWGDVTGIIDKVPYLKSLGVDIVWLSPIYTSPHVDMGYDIADYKSIDPRYGTLADVDLLIKALR NHDMKLMMDLVVNHTSDQHSWFVESASSKYSPKRDWYIWRPAKGFDDDGNPVPPNNWAQILGDA LSAWTWHEETREFYLTLHTSAQVELNWENPEVVAAVYDVMEFWLRRGICGFRMDVINLISKDQS FPDAPIIDPTSKYQPGEQFYTNGPRFHEFMHGIYDNVLSKYDTITVGETPYVTDIEEIIKTVGS TAKELNMAFNFDHMEIEDVKTKGDSKWSLRDWKLTELKGILSGWQKRMKKWDGWNAIFLECHDQ ARSVSRYTIDSDEFRERGAKLLALLETTLGGTIFLYQGQEIGMRNFPLEWDPDIEYKDVESVNF LNKSKELHPVGTEGLAKARTLLQKKARDHARTPMQWSAAPHAGFTVPDATPWMRVNDDYETVNV ETQMSFPWQSKGELSVWQYWQQAIQHRKLYKNAFVYGGFEDLDYHNEKVFAYLRTSADGNDSWL VAMNWTTSAVEWTVPSDIHVTRWVSSTLQTAPPVASKTVITLRAFEGVLGCCN	255	
Transglucosidase GAQ43844.1	MAGTRPMSNRWTLLLSLVILLGCLVIPGVTVKHENFKTCSQSGFCKRNRAFADDASAQGPSWTS PYELDSSSIQFKDGQLHGTILKSVSANEKVKLPLVVSFLESGAARVVVDEEKRLNGEIQLRHDS KARKERYNEAEKWVLVGGLELSKTATLKPETETGFTRVLYGPDNQFEAVIRHAPFSADFKRDGQ THVQLNNKGYLNMEHWRPKVEVEGEGEQQTQEDESTWWDESFGGNTDTKPRGPESVGLDITFPG YKHVFGIPEHADSLSLKETRGGEGNHEEPYRMYNADVFEYELNSPMTLYGAIPFMQAHRKDSTV GVFWLNAAETWVDIVKSTSSPNPLALGVGATTDTQSHWFSESGQLDVFVFLGPTPQEISKTYGE LTGYTQLPQHFAIAYHQCRWNYITDEDVKEVDRNFDKYQIPYDVIWLDIEYTDDRKYFTWDPLT FPDPISMEEQLDESERKLVVIIDPHIKNQDKYSISQEMTSKDLATKNKDGEIYDGWCWPGSSHW IDTFNPAAIKWWISLFKFDKFKGTLSNVFIWNDMNEPSVFNGPETTMPKDNLHHGNWEHRDIHN VHGITLVNATYDALLERKKGEVRRPFILTRSYYAGAQRMSAMWTGDNQATWEHLAASIPMVLNN GIAGFPFAGADVGGFFHNPSKELLTRWYQAGIWYPFFRAHAHIDTRRREPYLIAEPHRSIISQA IRLRYQLLPAWYTAFHEASVNGMPIVRPQYYAHPTDEAGFAIDDQLYLGSTGLLAKPVVSEEAT	256	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TADIYLADDEKYYDYFDYTVYQGAGKRHTVPAPMETVPLLMQGGHVIPRKDRPRRSSALMRWDP YTLVVVLDKNGQADGSLYVDDGETFDYERGAYIHRRFRFQESALVSEDVGTKGPKTAEYLKTMA NVRVERVVVVDPPKEWQGKTSVTVIEDGASAASTAPMQYHSQSDGKAAYAVVKNPNVGIGKTWR IEF		
Transglucosidase GAQ42954.1	MLLHLLAYAALSSVVTAASLQPRLQDGLALTPQMGWNTYNHYSCSPNETIVRSNAQALVDLGLS SLGYRYVTTDCGWTVADRLPDGSLTWNETLFPQGFPAMGDFLHDLGLLFGVYQDSGILLCGSPP NETGSLYHEAQDARTFASWNVDSLKYDNCYSDAATNYPNVNYAPSTSPEPRFANMSHALLQQNR TILFQICEWGISFPAGWAPALGHSWRIGNDIIPAWRTIFRIINQAAPQTDFAGPGQWPDLDMLE VGNNIFSLPEEQTHFSLWAILKSPLIIGAALKDELTAINDASLAVLKQKDVVAFNQDALGKSAS LRRRWTEEGYEVWSGPLSNGRTVAAVINWRNESRDLTLDLPDIGLQHAGTVKNIWDGTTAQNVV TSYTATVAGHGTMLLELQNTTAVGVYPRDVFGESSGQTTTFENIYAVTTSAKYTVSVYFSQPAS SAETISIGSNANQSIISVQVPASSTLVSANIPLTAGSSNTVTINTSIPIDAIHITAPNGTYYPC TNFTLAGSTTLTTCGSGYCQPVGSKIGYISPSGTAKATISATTSGSKYLEIDWINNEIAFDSSW GWGSNSRNLTVTVNSEEPVRIEVPLSGRHSELFGPGLGWWDTATLGLLTSGWKEGLNEVIVGNV GGDEGFQSYGADFEILRSNFEKMKVSLTLYAAALQLADAAVVQKRTVDVAELEHYWSYGRSEPV YPTPETSGSGDWEEAFTKAKSLVAQMTNDEKNNITYGYTSTTNGCSGMSGGVPRLGYPGMCLQD AASGVRGTDMVNAYASGLHIGASWNRDLAYEHAHYMGAEFKRKGANVALGPVVGPLGRMARGGR NWEGYSNDPYLSGSLVQNTIRGLQESVIACVKHFIGNEQETNRNTPQLLEDSYNQSVSSNIDDK TIHELYLWPFQDAVKAGAGAVMCSYNRINNSYGCQNSKNLNGLLKGELGFQGFVVSDWNAQQSG IASAAAGLDMVMPDSVYWENGNLSLAVRNGSLSSTRLDDMATRIVAAWYKYAELEDPGFGMPIS LLEPHDPVDARDPASKATILQEAIEGHVLVKNTDNALPLKEPKFLSLFGYDAIAAQRNTMDDLS WSLWTMGLDNTLSYPNGTAVDPSHLKYMFLSSTNPSENGPGVSLNGTMISGGGSGASTPSYIDA PFDAFQRQAYEDNTFLAWDFASQSPVVNPASEACLVFINEAAAEGWDRPYVADPYSDTLVENVA SQCNNTMVIIHNAGIRLVDRWVDNPNVTAVIYGHLPGQDSGRALVEIMYGKQSPSGRLPYTVAK NASDYGALLSPVVPEGTKDLYYPQDNFTEGVYIDYKAFEQKNITPRYEFGYGLTYSTFDYSGLK ISIHTGVNTDYLPPNSTIEEGGIPALWDVVATVTCSVANTGSVAAAEVAQLYLGIPGGPAKVLR GFEKKLIQPGHHTKVQFDLTRRDLSSWDVVNQAWVLQKGDYSVYVGKSVLDTQLTGTLTI	257	
Transglucosidase GAQ42198.1	MAVSASSPEPLGANIDERTPLNSSAQHATPSANTPDYSSITKGLTSDVHSSHGSDEEQPLINPP ESPGKDVTALTSISTVIGVLLLGEFISNADATLVMAATGRISSEFNRLRDASWLSTAYTLGLCA AQPMYGKLSDIYGRKPLLLWAYFLFGVGCVISGIGPDMATVILGRAISGIGGAGTMAMGSIIIT DIVPRRDVAHWRAYINIAMTLGRSAGGPVGGWLTDTIGWRWSFIIQGPLAAVAALLVVWKLKLV HPVTEKSIRRVDFLGTFLLATGIITITVIMDQAGQSFAWASLSTAILSTLSLSAFVAFVLVELY VAPEPIFELRMLRKPNVTPSYLIGSLQITAQVGMMFSVPLYFQVTSKASATVAGGHLVPAVIGN TLGGLIAGAFIRRTGQFKVLLILAGLVASVAYLLLFLRWNGHTGFWESLYIIPGGMGTGFCSAA AFVSMTAFLMPQEVAMATGGYFLLFSFAMTAGVTVTNSLLGTVFKRQMEQHLTGPGAKKIIERA LSDTSYINGLQGHVRDVVVKGYVAGLRYTYPMRVSISSLALSVYLFGKLALGLSAAEWRTQSIY FLLTDRFGRADNSTTATCDTGDQIYCGGSWQGIINHLDYIQGMGFTAIWISPITEQLPQDTSDG EAYHGYWQQKIYDVNSNFGTADDLKSLSDALHARGMYLMVDVVPNHMGYAGNGNDVDYSVFDPF	258	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	DSSSYFHPYCLITDWDNLTMVQDCWEGDTIVSLPDLNTTETAVRTIWYDWVADLVSNYSVDGLR IDSVLEVEPDFFPGYQEAAGVYCVGEVDNGNPALDCPYQDYLDGVLNYPIYWQLLYAFESSSGS ISDLYNMIKSVASDCSDPTLLGNFIENHDNPRFASYTSDYSQAKNVLSYIFLSDGIPIVYAGEE QHYSGGDVPYNREATWLSGYDTSAELYTWIATTNAIRKLAISADSDYITYANDPIYTDSNTIAM RKGTSGSQVITVLSNKGSSGSSYTLSLSGSGYTSGTELIEAYTCTSVTVDSNGDIPVPMASGLP RVLLPASVVDSSSLCGGSGSTTTTAATSTSTSKATSSTTTTTAITTTSSSCTATSTTLPITFEE LVTTTYGEEIYLSGSISQLGEWDTSDAVKLSADDYTSSNPEWYVTVSLPVGTTFEYKFIKVEED GSVTWESDPNREYTVPECGSGETVVDTWR		
Transglucosidase GAQ39994.1	MPLTYLGALAMLTALPSLGQARSTWPLGSGLELSYQASQHQISIHQDNQTIFSTLPGQPFLSAG AGKDQIVEDSGNFNITNVAQARCQGQNITQLAGIPRRDSVKNQVAVRGYLLDCGGEDIAYAMNF WVPKTLSDRVAFEATVDSDANASVPVERLYLTFASHAREDFYGLGAQASFASMKNRSIPIFSRE QGVGRGDQPYTAIEDSQGFFSGGDQYTTYTAIPQYVSSDGRVFYLDENDTAYAVFDFQRPDAVT VRYDSITVHGHLMQADNMLDAITMLTEYTGRMPALPEWVDHGALLGIQGGQEKVNRIVKQGFEH DCPVAGVWLQDWSGTHLQSAPYGNMNISRLWWNWESDTSLYPTWAEFVQALREQHGVRTLAYVN PFLADVSSKSDGYRRNLFQEASKHRYMVQNTTTNSTAIISSGKGIDAGILDLTNEETRAWFADV LRTQVWSANISGCMWDFGEYTPITADTSLANISTSAFFYHNQYPRDWAAYQRSVAAEMPLFHEM VTFHRSASMGANRHMNLFWVGDQATLWTPNDGIKSVVTIQGQMGISGYAHSHSDIGGYTTVFEP PTTSNSSGAIPRSAELLGRWGELGAVSSAVFRSHEGNVPSVNAQFYSNSTTYAYFAYNARMFRS LGPYRRRILNTESQRRGWPLLRMPVLYHPEDLRARQISYESFFLGRDLYVAPVLDEGRKSVEVY FPGHSANRTYTHVWSGQTYRGGQTAQVSAPFGKPAVFVVDGASSPELDVFLDFVRKENGTVLRA	259	
Transglucosidase GAQ38166.1	MVKLTDLLARAWLVPLAYGASQSRLSTTTSSQPQFTIPASADVGAQLIANIDDPQAANAQSVCP GYKASKVQHNSRGFTASLQLAGRPCNVYGTDVDSLTLSVEYQDSDRLNIQILPTHVDSTNASWY FLSENLVPRPKASLNASVSDSDFSVSWSNEPSFNFKVIRKATGDALFSTEGTVLVYEDQFIEFV TALPEEYNLYGLGEHITQFRLQRDANLTIYPSDDGTPIDKNIYGQHPFYLDTRYYKGDRQNGSY VPVKSSETDASQKYISLSHGVFLRNSHGLEILLRPQKLIWRTLGGGIDLTFYSGPNPADVTRQY LTSTVGLPAMQQYSTLGFHQCRWGYNNWSDLADVVANFEKFEIPLEYIWTDIDYMHGYRNFDND QNRFSYSEGDEFLSKLHESGRYYVPIVDAALYIPNPENASDAYATYDRGAADDVFLKNPDGSLY IGAVWPGYTVFPDWHHPKAVEFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGNLTLNPA HPPFLLPGEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATSTSTSVSYLRTTPTPGVRN VEHPPYVINHDQEGHDLSVHAVSPNATHIDGVEEYDVHGLYGHQGLNATYHGLLEVWSHERRPF IIGRSTFAGSGKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADTCGFTGNSDEELCNR WMQLSAFFPFYRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTTGSTVMR ALSWEFPNDPTLAAVETQFMVGPAIMVIPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKP GVNTTISAPLGHIPVYVRGGNILPMQEPALTTRGARQTPWALLAALGSNGTASGQLYLDDGESI YPNATLRVGFTASRSSLRSSAQGRWKERNPLANVTVLGVNKEPSAVTLNGKTVSPGSITYNSTS QVLFVGGLQNLTNGGAWAENWVLEW	260	
Transglucosidase GAQ36312.1	MLGSLLFLLPLVGAAVIGPRAGSQSCPGYKASNVQKSARSLTADLTLAGAPCNSYGKDVEDLKL LVEYQTDERLHVMIYDADEEVYQVPESVLPRVGSDKDSQDSVLEFDYVEEPFSFTISKGDEVLF	261	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	DSSASTLIFQSQYVRLRTWLPDDPYVYGLGEHSDPMRLPTYNYTRTLWNRDAYGTPNNTNLYGS HPVYYDHRGKSGTHGVFLLNSNGMDIKINQTTDGKQYLEYNLLGGVLDFYFFYGEDPKQASMEY SKIVGLPAMQSYWTFGFHQCRYGYRDVYELAEVVYNYSQAKIPLETMWTDIDYMDKRRVFTLDP QRFPLEKMRELVTYLHNHDQHYIVMVDPAVSVSNNSAYLTGVRDNVFLHNQNGSLYEGAVWPGV TVFPDWFNEDTQDYWTAQFQQFFDPKSGVDIDALWIDMNEASNFCPYPCLDPAAFAISDDLPPA APPVRPSSPIPLPGFPADFQPSSKRSVKRAQGDKGKKVGLPNRNLTDPPYTIRNAAGVLSMSTI ETDLIHAGEGYAEYDTHNLYGTMMSSASRTAMQARRPDVRPLVITRSTFAGAGAHVGHWLGDNL SDWVHYRISIAQILSFASMFQIPMVGADVCGFGSNTTEELCGRWASLGAFYTFYRNHNELGDIP QEFYRWPTVAESARKAIDIRYRLLDYIYTALHRQSQTGEPFLQPQFYLYPEDSNTFANDRQFFY GDALLVSPVLNEGSTSVDAYFPDDIFYDWYTGAVVRGHGENITLSNINITHIPLHIRGGNIIPV RMSSGMTTTEVRKQGFELIIAPDLDGTASGSLYLDDGDSLNPSSVTELEFTYSNGELHVQGTFG QKAVPKVEKCTLLGKSARTFKGFALDAPVNLKLK		
Transglucosidase GAQ33831.1	MSSPQQVYLLPLKDDGSPDVPGGYLYLPSPTDPPYLLRFVIEGSSSICREGALWVNIPEKGESF NRSAFRSFSLSPDFNKNIQIDIPITSAGSFAFYVTFSPLPEFSVLSTPTPEPTRTPTHYIDVSP KLTLRGQDLPLNALSIYSVISKFMGQYPKDWEKHLNGISQRNYNMVHFTPLMKRGASNSPYSIF DQLQFDDAVFPNGEDDVARLVSKMEDEYGLLSLTDVVWNHTAHNSKWLEEHPEAGYSVETAPWL EAALELDTALLKFGQELSTLGLPTEFHTVDELMEVMNAMRDKVISGIRLWEFYAIDVKADTQRI LDQWKTSKDLNLTDKKWAQLNLSDYKNWTLKQQATFIREYAIPTSKQVLGRFSRAVDLHFGAAI LTALFGPHDSPTSDTNTVEESLSKILDEVNLPFYEEYDGDVSEIMNQVFNRIKYLRIDDHGPKL GAVTAQSPLIETYFTRLPLNDVTKKHKKGALALVNNGWIWNADALRDNAGPDSRAYLRREVIVW GDCVKLRYGSCRDDNPFLWDFMTDYTRLMAKYFSGFRIDNCHSTPLVVAEYLLDEARKVRPNLT VFAELFTGSEEADYIFVKRLGINALIREAMQAWSTGELSRLVHRHGGRPIGSFDLDLPSSGSSH AIASSGLDSGKEKVAHIRPTPVQALFMDCTHDNEMPAQKRTAKDTLPNGALVAMCASAIGSVIG YDEVYPRLVDLVHEHRLYFSEFSEAPETGLNSLEGGIGGIKKLLNDLHTRMGVEEYDETHIHHD GEYITVHRVHPRTRKGVFLIAHTAFSGQDGKSVLAPTHLVGTHVKHIGTWLLEVDASQTTKERI QTDKSYLRGLPSQVKTFEGTKIEESGKDTIISVLDSFVAGSIALFETSMPSVEHASGLDNYITE GVDHAFSDLSLVDLNFALYRCEAEERDSSKGQDGVYDIPGHGPLVYAGLQGWWSVLENIIKYNE LGHPLCDHLRNGQWALDYIVARLEKLGHTDEHTALGRPAAWLQEKFQAVRQLPSFLLPRYFAII VQVAYNAAWKRGIQLLGPHIQNGQEFIHQLGMVSVQQTGYVNSASLWPTKKVPSLAAGLPHFAV DWARCWGRDVFISLRGLLLCTSRFEDAKEHITAFASVLKHGMIPNLLSSGKLPRYNSRDSVWFF LQSIQDYTKMAPDGLRLLDHNVPRRFLPYDDVWFPYDDPRAYSQHSTISEIIQEVLQRHAQGLS FREYNAGPDLDMQMTQEGFQIDVKVDWETGLIFGGSQYNCGTWQDKMGESAKAGNKGVPGTPRD GAAIEITGLVYSALTWVAELHERGLYKHDGVDIGDGKSISFKEWASRIQANFERCYYVPLQPKD DGQYDIDANIINRRGIYKDLYRSGKPYEDYQLRSNFPIAMTVAPDLFTSSKALAALALADEVLV GPVGMATLDPSDLNYRPNYNNSEDSTDFATAKGRNYHQGPEWVWQRGYFLRAFLHFDLARRTTP AERTETYQQITRRLEGCKRALRESPWKGLTELTNKNGAHCADSNICFWSVMSLTESAAAAMLTA AVVVDSPIDVHEPRWDDQTRGRDLYTQLVISLLVGLSAFFSFCVLRPKWTELYAARRRQRCAAS YLPELPDSFFGWIPVLYRITDEQVLESAGLDAFVFLTFLKFAIRFLSAIFFFALVIILPTHYKN	262	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TGKSGVPGWDDDDDETFDGDKDKKKIISDPNYLWMYVIFTYIFTGLAVYMLIQETNKVIRTRQK YLGSQTSTTDRTIRLSGIPPDLGTEEKIKDFMEGLKVGKVESVTLCRDWRELDHLIDERLKLLR NLERAWTRHLGYKRVKASPNALTLMHQQPRGSSIVSDGESERIQLLSEGGRDHVTDYAHKRPTV RIWYGPFKLRYKNIDAIDYYEEKLRRLDEKIQVARQKEYPPTEVAFVTMESIAASQMVVQAILD PHPMQLLARLAPAPADVVWKNTYLPRSRRMMQSWFITVVIGFLTVFWSVLLIPVAYLLEYETLH KVFPQLADALARNPLAKSLVQTGLPTLVLSLLTVAVPYLYNWLSNQQGMMSRGDIELSVISKTF FFSFFNLFLVFTVFGTATTFYGFWENLRDAFKDATTIAFALAKTLENFAPFYINFLCLQGIGLF PFRLLEFGSVAMYPINFLAAKTPRDYAELSTPPTFSYGYSIPQTVLSLIICVVYSVFPSSWLIC LFGLIYFTIGKFIYKYQLLYAMDHQQHSTGRAWPMICSRILMGLIVFQLAMIGVLALRRAITRS LLIVPLLMATVWFSYFFARTYEPLMKFIALKSIDRERPGGGDISPSPSSTFSPPSGLDRDSFPI RIGGQELGLRLRKYVNPSLILPLHDAWLPGRTMVPELQGELEHRNSENNAADESV		
Transglucosidase GAQ33901.1	MWSSWLLSALLATEALAVPYEEYILAPSSRDLAPASVRQVNGSVTNAAALTGAGGQATFNGVSS VTYDFGINVAGIVSVDVASASSESAFIGVTFTESSMWISNEACDATQDAGLDTPLWFAVGQGAG VYSVGKKYTRGAFRYMTVVSNTTATVSLNSVKINYTASPIQDLRAYTGYFHSSDELLNRIWYAG AYTLQLCSIDPTTGDALVGLGVITSSETITLPQTDKWWTNYTITNGSSTLTDGAKRDRLVWPGD MSIALESVAVSTEDLYSVRTALESLYALQKADGQLPYAGKPFYDTVSFTYHLHSLVGAASYYQY TGDRAWLTRYWGQYKKGVQWALSSVDSTGLANITASADWLRFGMGAHNIEANAILYYVLNDAIS LAQSLNDNAPIRNWTATAARIKTVANELLWDDKNGLYTDNETTTLHPQDGNSWAVKANLTLSAN QSAIISESLAARWGPYGAPAPEAGATVSPFIGGFELQAHYQAGQPDRALDLLRLQWGFMLDDPR MTNSTFIEGYSTDGSLVYAPYTNRPRVSHAHGWSTGPTSALTIYTAGLRVTGPAGATWLYKPQP GNLTQVEAGFSTRLGSFASSFSRSGGRYQELSFTTPNGTTGSVELGDVSGQLVSEGGVKVQLVG GKASGLQGGKWRLNV	263	
Transglucosidase EHA19108.1	MRWHKLLPGVLALLPLSVAQSCWRNTTCSGPTESAFSGPWEKNIFAPSSRTVNPEKLFLITQPD KTEEYSPFALHGNGSLVVYDFGKEVGGIVSVNFSSTGSGALGVAFTEAKNWIGEWSDSSNGGFK GPDGALYGNFTEAGSHYYVMPDKSLRGGFRYLTLFLITSDNSTIQIEDVNLEIGFQPTWSNLKA YQGYFHSNDDLLNKIWYTGAYTLQTNEVPTDTGRQIPAMAVGWANNCTLGPGDTIIVDGAKRDR AVWPGDMGIAVPSAFVSLGDLDSVKNALQVMYDTQNNSTGAFDESGPPLSQKDSDTYHMWTMVG TYNYMLFTNDSDFLERNWEGYQKAMDYIYGKVTYPSGLLNVTGTRDWARWQQGYNNSEAQMILY HTLNTGAELATWAGDSGDLSSTWTSRAEKLRQAINEYCWDDSYGAFKDNATDTTLHPQDANSMA LLFGVVDADRAASISERLTDNWTPIGAVAPELPENISPFISSFEIQGHLTVGQPQRALELIRRS WGWYYNNANGTQSTVIEGYLQNGTFGYRSDRGYYYDTAYVSHSHGWSSGPTSALTNYIVGISVT SPLGATWRIAPQFVDLQSAEGGFTTSLGKFQAGWSKTDKGYTLDFTVPHGTQGNLTLPFVSAAK PSIKIDGTEISRGVQYANSTATVTVSGGGTYKVEVQ	264	
Transglucosidase EHA19157.1	MPQKEFVPKTYQESSTGAQSSSSVHLRSSPEERSFDFSFEPIRENLFRVTFSSQDHPLPPYPSV TKPATSLDGVHVSATGGSNQKTIEVGDVTASVEWSNTPVVSLSWKGTEKPLYRDLPLRSYVADS TGIAHYTEHDRDCLHVGLGEKRAPMDLTGRHFQLSATDSFGYDVYNTDPLYKHIPLLIKASPDG CVAIFSTTHGRGTWSVGSEVDGLWGHFKVYRQDYGGLEQYLIVGKTLKDVVRSYAELVGLPILV PRWAYGYISGGYKYTMLDDPPAHEALMEFADKLEEHGIPCSAHQMSSGYSIAETEPKVRNVFTW	265	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	NKYRFPNPEEWIAKYHGRGIRLLSNIKPFLLASHPDFQKLIDGNGFFKDPESSKPGYMRLWSAG GATGGDGCHIDFSSAVAFKWWYDGVQSLKRAGIDAMWNDNNEYTLPDDDWKLALDEPTVSDAVK KGVENSVGQWGRAMHTELMGKASHDALLNIEPNHRPFVLTRSATAGTMRYAASTWSGDNVTSWE GMKGANALSLSAGISLLQCCGHDIGGFEGPQPSPELLLRWIQLGIHSPRFAINCFKTSPGNSSV GDVIEPWMYPEITPLVRDTIKRRYEILPYIYSLGLESHLTASPPQRWVGWGYESDPEVWTKALK SGDEQFWFGDTIMVGGVYEPGVSVAKLYLPRKANDQFDFGYVNMNEPYNYLASGQWVEVPSEWR KSIPLLARIGGAIPVGKPVHTRVPGDDTPASVAVKEVDDYRGVEIFPPLGSSHGQVFSTTWFED DGISLEARISEYTVTYSSTEEKVIVGFSRDEKSGFVPAWTDLDIILHNGDERRVVSDIGKTVEY KGKGSRGRVVYTLKN		
Transglucosidase EHA19519.1	MRLSTSSLLLSVSLLGKLALGLSAAEWRTQSIYFLLTDRFGRTDNSTTATCNTGDQIYCGGSWQ GIINHLDYIQGMGFTAIWISPITEQLPQDTADGEAYHGYWQQKIYDVNSNFGTADDLKSLSDAL HARGMYLMVDVVPNHMGYAGNGNDVDYSVFDPFDSSSYFHPYCLITDWDNLTMVQDCWEGDTIV SLPDLNTTETAVRTIWYDWVADLVSNYSVDGLRIDSVLEVEPDFFPGYQEAAGVYCVGEVDNGN PALDCPYQEYLDGVLNYPIYWQLLYAFESSSGSISDLYNMIKSVASDCSDPTLLGNFIENHDNP RFASYTSDYSQAKNVLSYIFLSDGIPIVYAGEEQHYSGGKNDAFYTDSNTIAMRKGTSGSQVIT VLSNKGSSGSSYTLTLSGSGYTSGTKLIEAYTCTSVTVDSSGDIPVPMASGLPRVLLPASVVDS SSLCGGSGSNSSTTTTTTATSSSTATSKSASTSSTSTACTATSTSLAVTFEELVTTTYGEEIYL SGSISQLGDWDTSDAVKMSADDYTSSNPEWSVTVTLPVGTTFEYKFIKVESDGTVTWESDPNRE YTVPECGSGETVVDTWR	266	
Transglucosidase EHA20839.1	MVKLTHLLARAWLVPLAYGASQSLLSTTAPSQPQFTIPASADVGAQLIANIDDPQAADAQSVCP GYKASKVQHNSRGFTASLQLAGRPCNVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTNASWY FLSENLVPRPKASLNASVSQSDLFVSWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFV TALPEEYNLYGLGEHITQFRLQRNANLTIYPSDDGTPIDQNLYGQHPFYLDTRYYKGDRQNGSY IPVKSSEADASQDYISLSHGVFLRNSHGLEILLRSQKLIWRTLGGGIDLTFYSGPAPADVTRQY LTSTVGLPAMQQYNTLGFHQCRWGYNNWSDLADVVANFEKFEIPLEYIWTDIDYMHGYRNFDND QHRFSYSEGDEFLSKLHESGRYYVPIVDAALYIPNPENASDAYATYDRGAADDVFLKNPDGSLY IGAVWPGYTVFPDWHHPKAVDFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGNLTLNPA HPSFLLPGEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTSVSYLRTTPTPGVRN VEHPPYVINHDQEGHDLSVHAVSPNATHVDGVEEYDVHGLYGHQGLNATYQGLLEVWSHKRRPF IIGRSTFAGSGKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADTCGFNGNSDEELCNR WMQLSAFFPFYRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTTGSTVMR ALSWEFPNDPTLAAVETQFMVGPAIMVVPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKP GVNTTISAPLGHIPVYVRGGNILPMQEPALTTREARQTPWALLAALGSNGTASGQLYLDDGESI YPNATLHVDFTASRSSLRSSAQGRWKERNPLANVTVLGVNKEPSAVTLNGQAVFPGSVTYNSTS QVLFVGGLQNLTKGGAWAENWVLEW	267	
Transglucosidase EHA21384.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ	268	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR		
Transglucosidase EHA23512.1	MAKSASQIHRAWWKECSVYQIWPASYKDSNDDGIGDIPGIISKLDYIKNIGVDIVWLCPSYKSP QVDMGYDIADYYSIADEYGTVADVEKLIQGCHERGMKLLMDLVVNHTSDQNEWFKQSRSSKDNE YRNWYVWKPARYDEQGNRHPPNNWVSHFQGSAWEWDEHTGEYYLHLYATEQPDLNWEHPPVRKA VHDIMRFWLDKGADGFRMDVINFISKDQRFPDAPVKDPRTPWQWGDKYYANGPRLHEYLQDLGK ILKEYDAFSVGEMPFVRDTEEVLRAVRYDRNEINMIFNFEHVDIDHGTYDKFEPGSWKLTDLKA FFETWQKFMYNNDGWNALYWENHDQPRSIDRYAQAKEEFRTEAGKMLATVLALQSGTPFVYQGQ EIGMRNVPVEWDMNEYKDIDCLNHWHRLLKHRPDDIEAQKSARQEYQKKSRDNGRTPVQWSSAP NGGFTGPNAKPWMSVNPDYVRFNAEAQVNDPNSIYHYWAAVLGLRKKYLDIFVYGDYDLVDKDS QEVFAYARQFENQKALVLTNWTEKTLEWDATANGVKGIKDVLLNSYESAEAAKERFTGQKWSLR PYEAVVLLVEA	269	
Transglucosidase EHA23680.1	MPSTYLGALATLAVFPCLGQARSTWPLGSGLELSYQASQHQISIHQDNQTIFSTIPGQPFLSAS AGKDQFVEDSGNFNITNVNQARCRGQNITQLAGIPRSDSVKNQVAVRGYLLDCGGEDIAYGMNF WVPRRFSDRVAFEATVDSEANASVPVDRLYLTFASHALEDFYGLGAQASFASMKNRSIPIFSRE QGVGRGDQPYTAIEDSQGFFSGGDQYTTYTAIPQYVSSDGRVFYLDENDTAYAVFDFQRSDAVT VRYDSLSVHGHLMQADTMLDAITMLTEYTGRMPTLPEWVDHGALLGIQGGQEKVNRIVKQGFEH DCPVAGVWLQDWSGTHLQSAPYGNMNISRLWWNWESDTSLYPTWAEFVQTLREQHGVRTLAYVN PFLANVSSKSDGYRRNLFLEASQHRYMVQNTTTNSTAIISSGKGIDAGILDLTNEDTRAWFADV LRTQVWSANISGCMWDFGEYTPITPDTSLANISTSAFFYHNQYPRDWAAYQRSVAAEMPLFHEM VTFHRSASMGANRHMNLFWVGDQATLWTRNDGIKSVVTIQGQMGISGYAHSHSDIGGYTTVFEP PTTSNSSGAIPRSAELLGRWGELGAVSSAVFRSHEGNVPSVNAQFYSNSTTYAYFAYNARLFRS LGPYRRRILNTESQRRGWPLLRMPVLYHPEDLRARQISYESFFLGRDLYVAPVLDEGHKSVDVY FPGHGANRTYTHVWTGQTYRAGQTAKVSAPFGKPAVFLVNGASSPELDVFLNFVRKENGTVLHA	270	
Transglucosidase EHA25759.1	MDPANEYCGLEDYGLVGDMHTCALVSKNGSVDSMCWPVFDSPSIFCRILDKEKGGHFSITPDRR LKNPLSKQRYRPYTNMLETRWIHEEGVMNILDYFPIAKPKPHVVEKGLPQWCRCYQNKGSAQYQ ACRSGMVRKAECVRGEMEIEIELFPAFNYARDSHVAQQSSASDDAIQVYHFQAESQNLVVSVLG DKGDISEDDSDLSIEFELSDRPGHLGPGLVGKVILKEGQSITMLLHDQESITCDVDDLAPYLQQ IERTTGDFWSDWTSKCTFRGHYREQVERSLLVLKLLTYKPTGAIVAAPTFSLPEHIGGSRNWDY RYSWVRDAAFTVYVFLKNGYPEEAESYINFIFERIFPPMDKNPKPGEPFLPIMITIHGEREIPE MELDHLEGYRGSRPVRIGNGAATHIQLDIYGELMDSIYLYNKHAADISYDQWRAIRRMIDFVIQ IRHQPDQSIWEVRGPPQNFVYSKIMLWVALDRGLRLAEKRSNLPCPDRARWMHERDALYDEIMT KGYNAEKGFFCMSYENQDAMDAAVLIAPLVFFVAPNDPRLLSTIQRITEVPAKGGLSVANMVSR	271	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	YDTGKVDDGVGGNEGAFLMVTFWLVEAMMRAARSKSYLPHDPFFQQLRKTATSQFDSILSFANH LGMFSEEVATSGEQIGNMPQAFSHLACVSAAMNLGGGGDR		
Transglucosidase EHA26514.1	MSSPQQVYLLPLKDDGSPDVPGGYIYLPAPTNPPYLLRFVIEGSSSICREGALWVNIPEKGESF NRSAFRSFSLSPDFNKNIQIDVPITSAGSFAFYVTFSPLPEFSVISTPTPEPTRTPTHYIDVSP KLTLRGQDLPLNALSIYSVISKFMGQYPKEWEKHLNGISQRNYNMVHFTPLMKRGASNSPYSIF DQLQFDDAVFPNGEDDVARLISKMENEYGLLSLTDVVWNHTAHNSKWLEEHPEAGYSVETAPWL EAALELDTALLKFGQDLQNLGLPTEFQTVDELMKVMNVMRDKVIAGIRLWEFYAIDVKSDTHKI LDKWKTSKDIDLTDTNWAQLNLQDYKNWTLKQQATFIRDHAIPTSKQVLDRFSRAVDLQFGAAI LTALFGPHNPSTSDTSIVEESLSKILDEVNLPFYEEYDGDVSEIMNQVFNRIKYLRIDDHGPKL GAVTAQSPLIETYFTRLPLNDVTKKHKKEALALVNNGWIWNADALRDNAGPDSRAYLRREVIVW GDCVKLRYGSCRDDNPFLWDFMTDYTRLMAKYFSGFRIDNCHSTPLVVAEYLLDEARKVRPNLT VFAELFTGSEEADYIFVKRLGINALIREAMQAWSTGELSRLVHRHGGRPIGSFDLDLPSSGSSH AIASSGLDSGKEKVVHIRPTPVQALFMDCTHDNEMPAQKRTAKDTLPNGALVAMCASAIGSVIG YDEVYPRLVDLVHEHRLYFSEFSEAPETGLNSLEGGIGGIKKLLNELHTKMGIEGYDETHIHHD GEYITVHRVHPRTRKGVFLIAHTAFPGQDSRSVLAPTHLVGTQVKHIGTWLLEVDTSQTTKERI QADKSYLRGLPSQVKTFEGTKIEESGKDTIISVLNSFVAGSIALFETSMPSVEHASGLDNYITE GVDHAFSDLSLVDLNFALYRCEAEERDSSKGQDGAYDIPGHGPLVYAGLQGWWSVLENIIKYNE LGHPLCDHLRNGQWALDYIVARLEKLSHKEEHPALGRPAAWLQEKFQAVRQLPSFLLPRYFAII VQVAYNAAWKRGIQLLGSHIQKGQEFIHQLGMVSVQQTGYVNSASLWPTKKVPSLAAGLPHFAV DWARCWGRDVFISLRGLLLCTGRFEDAKEHITAFASVLKHGMIPNLLSSGKLPRYNSRDSVWFF LQSIQDYTEMAPDGLEILDHKVPRRFLPYDDVWFPFDDPRAYSQQSTISEIIQEVFQRHAQGLS FREYNAGPDLDMQMTQDGFQIDVKVDWETGLIFGGSQYNCGTWQDKMGESAKAGNKGVPGTPRD GAAIEITGLVYSALTWVAKLHERGIYKHDGVDIGGGKSISFEDWASRIRANFERCYYVPLQPKD DGQYDIDANIINRRGIYKDLYRSGKPYEDYQLRSNFPIAMTVAPDLFTASKALAALALADEVLV GPVGMATLDPSDLNYRPNYNNSEDSTDFATAKGRNYHQGPEWVWQRGYFLRAFLHFDLARRTTP AERTETYQQITRRLEGCKRALRESPWKGLTELTNKNGAYCADSSPTQAWSAGCLLDLYYDASRH SQS	272	
Transglucosidase EHA26552.1	MWSSWLLSALLATEALAVPYEEYILAPSSRDLAPASVRQVNGSVTNAAALTGAGGQATFNGVSS VTYDFGINVAGIVSVDVASASSESAFIGVTFTESSMWISSEACDATQDAGLDTPLWFAVGQGAG LYTVEKKYNRGAFRYMTVVSNTTATVSLNSVKINYTASPTQDLRAYTGYFHSNDELLNRIWYAG AYTLQLCSIDPTTGDALVGLGVITSSETISLPQTDKWWTNYTITNGSSTLTDGAKRDRLVWPGD MSIALESVAVSTEDLYSVRTALESLYALQKPDGRLPYAGKPFFDTVSFTYHLHSLVGAASYYQY TGDRAWLTRYWGQYKKGVQWALSSVDSTGLANITASADWLRFGMGAHNIEANAILYYVLNDAIS LAQTLNDNAPIRNWTTTAARIKTVANELLWDDKNGLYTDNETTTLHPQDGNSWAVKANLTLSAN QSAIVSESLAARWGPYGAPAPEAGATVSPFIGGFELQAHYQAGQPDRALDLLRLQWGFMLDDPR MTNSTFIEGYSTDGSLAYAPYTNTPRVSHAHGWATGPTSALTIYTAGLRVTGPAGATWLYKPQP GNLTQVEAGFSTRLGSFASSFSRSGGRYQELSFSTPNGTTGSVELGDVSGQLVSDRGVKVQLVG GKASGLQGGKWKLSNN	273	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Transglucosidase EHA26885.1	MLGSLLLLLPLVGAAVIGPRANSQSCPGYKASNVQKQARSLTADLTLAGTPCNSYGKDLEDLKL LVEYQTDERLHVMIYDADEEVYQVPESVLPRVGSDEDSEDSVLEFDYVEEPFSFTISKGDEVLF DSSASPLVFQSQYVNLRTWLPDDPYVYGLGEHSDPMRLPTYNYTRTLWNRDAYGTPNNTNLYGS HPVYYDHRGKSGTYGVFLLNSNGMDIKINQTTDGKQYLEYNLLGGVLDFYFFYGEDPKQASMEY SKIVGLPAMQSYWTFGFHQCRYGYRDVYELAEVVYNYSQAKIPLETMWTDIDYMDKRRVFTLDP QRFPLEKMRELVTYLHNHDQHYIVMVDPAVSVSNNTAYITGVRDDVFLHNQNGSLYEGAVWPGV TVFPDWFNEGTQDYWTAQFQQFFDPKSGVDIDALWIDMNEASNFCPYPCLDPAAYAISADLPPA	274	
	APPVRPSSPIPLPGFPADFQPSSKRSVKRAQGDKGKKVGLPNRNLTDPPYTIRNAAGVLSMSTI ETDLIHAGEGYAEYDTHNLYGTTHIPMVGADVCGFGSNTTEELCARWASLGAFYTFYRNHNELG DISQEFYRWPTVAESARKAIDIRYKLLDYIYTALHRQSQSGEPFLQPQFYLYPEDSNTFANDRQ FFYGDALLVSPVLNEGSTSVDAYFPDDIFYDWYTGAVVRGHGENITLSNINITHIPLHIRGGNI IPVRTSSGMTTTEVRKQGFELIIAPDLDDTASGSLYLDDGDSLNPSSVTELEFTYSKGELHVKG TFGQKAVPKVEKCTLLGKSART		
Transglucosidase EHA27488.1	MCHKSNYSSPKWWKESVVYQVYPASFNCGKSTTTTNGWGDVTGIIEKVPYLKSLGVDIVWLSPI YTSPQVDMGYDIADYKSIDPRYGTLADVDLLIKSLKDHDMRLMMDLVVNHTSDQHSWFVESASS KDSPKRDWYIWRPAKGFDEAGNPVPPNNWAQILGDTLSAWTWHEETQEFYLTLHTSAQAELNWE NPDVVTAVYDVMEFWLRRGICGFRMDVINFISKDQSFPDAPIIDPASKYQPGEQFYTNGPRFHE FMHGIYDNVLSKYDTITVGETPYVTDMKEIIKTVGSTAKELNMAFNFDHMEIEDIKTKGESKWS	275	
	LRDWKLTELKGILSGWQKRMREWDGWNAIFLECHDQARSVSRYTNDSDEFRDRGAKLLALLETT LGGTIFLYQGQEIGMRNFPVEWGPDTEYKDIESVNFWKKSKELHPVGSEGLAQARTLLQKKARD HARTPMQWSADPHAGFTVPDATPWMRVNDDYRTVNVEAQMSFPWEMKGELSVWQYWQQALQRRK LHKGAFVYGDFEDLDYHNESVFAYSRTSADGKETWLPVPSWSLPKKRTLS		
Transglucosidase EHA28539.1	MSNRWTLLLSLVILLGCLVIPGVTVKHENFKTCSQSGFCKRNRAFADDAAAQGSSWASPYELDS SSIQFKDGQLHGTILKSVSPNEKVKLPLVVSFLESGAARVVVDEEKRMNGDIQLRHDSKARKER YNEAEKWVLVGGLELSKTATLRPETESGFTRVLYGPDNQFEAVIRHAPFSADFKRDGQTHVQLN NKGYLNMEHWRPKVEVEGEGEQQTQEDESTWWDESFGGNTDTKPRGPESVGLDITFPGYKHVFG IPEHADSLSLKETRGGEGNHEEPYRMYNADVFEYELSSPMTLYGAIPFMQAHRKDSTVGVFWLN	276	
	AAETWVDIVKSTSSPNPLALGVGATTDTQSHWFSESGQLDVFVFLGPTPQEISKTYGELTGYTQ LPQHFAIAYHQCRWNYITDEDVKEVDRNFDKYQIPYDVIWLDIEYTDDRKYFTWDPLSFPDPIS MEEQLDESERKLVVIIDPHIKNQDKYSIVQEMKSKDLATKNKDGEIYDGWCWPGSSHWIDTFNP AAIKWWVSLFKFDKFKGTLSNVFIWNDMNEPSVFNGPETTMPKDNLHHGNWEHRDIHNVHGITL VNATYDALLERKKGEIRRPFILTRSYYAGAQRMSAMWTGDNQATWEHLAASIPMVLNNGIAGFP FAGADVGGFFQNPSKELLTRWYQAGIWYPFFRAHAHIDTRRREPYLIAEPHRSIISQAIRLRYQ LLPAWYTAFHEASVNGMPIVRPQYYAHPWDEAGFAIDDQLYLGSTGLLAKPVVSEEATTADIYL ADDEKYYDYFDYTVYQGAGKRHTVPAPMETVPLLMQGGHVIPRKDRPRRSSALMRWDPYTLVVV LDKNGQADGSLYVDDGETFDYKRGAYIHRRFRFQESALVSEDVGTKGPKTAEYLKTMANVRVER VVVVDPPKEWQGKTSVTVIEDGASAASTASMQYHSQPDGKAAYAVVKNPNVGIGKTWRIEF		
Transglucosidase	MWSSWLLSALLATEALAVPYEEYILAPSSRDLAPASVRQVNGSVTNAAALTGAGGQATFNGVSS	277	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
XP_001389086.1	VTYDFGINVAGIVSVDVASASSDSAFIGVTFTESSMWISSEACDATQDAGLDTPLWFAVGQGAG LYTVEKKYNRGAFRYMTVVSNTTATVSLNSVKINYTASPTQDLRAYTGYFHSNDELLNRIWYAG AYTLQLCSIDPTTGDALVGLGVITSSETISLPQTDKWWTNYTITNGSSTLTDGAKRDRLVWPGD MSIALESVAVSTEDLYSVRTALESLYALQKPDGRLPYAGKPFFDTVSFTYHLHSLVGAASYYQY TGDRAWLTRYWGQYKKGVQWALSSVDSTGLANITASADWLRFGMGAHNIEANAILYYVLNDAIS LAQTLNDNAPIRNWTTTAARIKTVANELLWDDKNGLYTDNETTTLHPQDGNSWAVKANLTLSAN QSAIVSESLAARWGPYGAPAPEAGATVSPFIGGFELQAHYQAGQPDRALDLLRLQWGFMLDDPR MTNSTFIEGYSTDGSLAYAPYTNTPRVSHAHGWATGPTSALTIYTAGLRVTGPAGATWLYKPQP GNLTQVEAGFSTRLGSFASSFSRSGGRYQELSFSTPNGTTGSVELGDVSGQLVSDRGVKVQLVG GKASGLQGGKWKLSNN		
Transglucosidase XP_001400455.1	MAKSASQIHRAWWKECSVYQIWPASYKDSNDDGIGDIPGIISKLDYIKNIGVDIVWLCPSYKSP QVDMGYDIADYYSIADEYGTVADVEKLIQGCHERGMKLLMDLVVNHTSDQNEWFKQSRSSKDNK YRNWYVWKPARYDEQGNRHPPNNWVSHFQGSAWEWDEHTGEYYLHLYATEQPDLNWEHPPVRKA VHDIMRFWLDKGADGFRMDVINFISKDQRFPDAPVKDPRTPWQWGDKYYANGPRLHEYLQDLGK ILKEYDAFSVGEMPFVRDTEEVLRAVRYDRNEINMIFNFEHVDIDHGTYDKFEPGSWKLTDLKA FFETWQKFMYNNDGWNALYWENHDQPRSIDRYAQAKEEFRTEAGKMLATVLALQSGTPFVYQGQ EIGMRNVPVEWDMNEYKDIDCLNHWHRLLKHRPDDIEAQKSARQEYQKKSRDNGRTPVQWSSAP NGGFTGPNAKPWMSVSPDYVRFNAEAQVNDPNSIYHYWAAVLGLRKKYLDIFVYGDYDLVDKDS QEIFAYARQYENKKALVLTNWTEKTLEWDATTNGVKGVKDVLLNSYESAEAAKGRFSGQKWSLR PYEAVVLLVEA	278	
Transglucosidase XP_001390530.1	MSFRSLLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIVVAS PSTDNPDYFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGL GEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSTATDIVWPLVRNDLSYVAQ YWNQTGYDLWEEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDSQAPEILCYLQSFWTGSF ILANFDSSRSGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALANHKEVVDSFRSIYTLNDGLS DSEAVAVGRYPEDTYYNGNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSLDFFKALYSDAAT GTYSSSSSTYSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAALLT ANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTTTTATPTGSGSVT STSKTTATASKTSTSTSSTSCTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWETSDGIALS ADKYTSSDPLWYVTVTLPAGESFEYKFIRIESDDSVEWESDPNREYTVPQACGTSTATVTDTWR	279	
Transglucosidase XP_001393899.1	MSNRWTLLLSLVILLGCLVIPGVTVKHENFKTCSQSGFCKRNRAFADDAAAQGSSWASPYELDS SSIQFKDGQLHGTILKSVSPNEKVKLPLVVSFLESGAARVVVDEEKRMNGDIQLRHDSKARKER YNEAEKWVLVGGLELSKTATLRPETESGFTRVLYGPDNQFEAVIRHAPFSADFKRDGQTHVQLN NKGYLNMEHWRPKVEVEGEGEQQTQEDESTWWDESFGGNTDTKPRGPESVGLDITFPGYKHVFG IPEHADSLSLKETRGGEGNHEEPYRMYNADVFEYELSSPMTLYGAIPFMQAHRKDSTVGVFWLN AAETWVDIVKSTSSPNPLALGVGATTDTQSHWFSESGQLDVFVFLGPTPQEISKTYGELTGYTQ LPQHFAIAYHQCRWNYITDEDVKEVDRNFDKYQIPYDVIWLDIEYTDDRKYFTWDPLSFPDPIS MEEQLDESERKLVVIIDPHIKNQDKYSIVQEMKSKDLATKNKDGEIYDGWCWPGSSHWIDTFNP	280	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAIKWWVSLFKFDKFKGTLSNVFIWNDMNEPSVFNGPETTMPKDNLHHGNWEHRDIHNVHGITL VNATYDALLERKKGEIRRPFILTRSYYAGAQRMSAMWTGDNQATWEHLAASIPMVLNNGIAGFP FAGADVGGFFQNPSKELLTRWYQAGIWYPFFRAHAHIDTRRREPYLIAEPHRSIISQAIRLRYQ LLPAWYTAFHEASVNGMPIVRPQYYAHPWDEAGFAIDDQLYLGSTGLLAKPVVSEEATTADIYL ADDEKYYDYFDYTVYQGAGKRHTVPAPMETVPLLMQGGHVIPRKDRPRRSSALMRWDPYTLVVV LDKNGQADGSLYVDDGETFDYERGAYIHRRFRFQESALVSEDVGTKGPKTAEYLKTMANVRVER VVVVDPPKEWQGKTSVTVIEDGASAASTASMQYHSQPDGKAAYAVVKNPNVGIGKTWRIEF		
Transglucosidase XP_001399012.1	MPQKEFVPKTYQESSTGAQSSSSVHLRSSPEERSFDFSFEPIRENLFRVTFSSQDHPLPPYPSV TKPATSLDGVHVSATGGSNQKTIEVGDVTASVEWSNTPVVSLSWKGTEKPLYRDLPLRSYVADS TGIAHYTEHDRDCLHVGLGEKRAPMDLTGRHFQLSATDSFGYDVYNTDPLYKHIPLLIKASPDG CVAIFSTTHGRGTWSVGSEVDGLWGHFKVYRQDYGGLEQYLIVGKTLKDVVRSYAELVGLPILV PRWAYGYISGGYKYTMLDDPPAHEALMEFADKLEEHGIPCSAHQMSSGYSIAETEPKVRNVFTW NKYRFPNPEEWIAKYHGRGIRLLSNIKPFLLASHPDFQKLIDGNGFFKDPESSKPGYMRLWSAG GATGGDGCHIDFSSAVAFKWWYDGVQSLKRAGIDAMWNDNNEYTLPDDDWKLALDEPTVSDAVK KGVENSVGQWGRAMHTELMGKASHDALLNIEPNHRPFVLTRSATAGTMRYAASTWSGDNVTSWE GMKGANALSLSAGISLLQCCGHDIGGFEGPQPSPELLLRWIQLGIHSPRFAINCFKTSPGNSSV GDVIEPWMYPEITPLVRDTIKRRYEILPYIYSLGLESHLTASPPQRWVGWGYESDPEVWTKALK SGDEQFWFGDTIMVGGVYEPGVSVAKLYLPRKANDQFDFGYVNMNEPYNYLASGQWVEVPSEWR KSIPLLARIGGAIPVGKPVHTRVPGDDTPASVAVKEVDDYRGVEIFPPLGSSHGQVFSTTWFED DGISLEARISEYTVTYSSTEEKVIVGFSRDEKSGFVPAWTDLDIILHNGDERRVVSDIGKTVEY KGKGSRGRVVYTLKN	281	
Transglucosidase XP_001402053.1	MVKLTHLLARAWLVPLAYGASQSLLSTTAPSQPQFTIPASADVGAQLIANIDDPQAADAQSVCP GYKASKVQHNSRGFTASLQLAGRPCNVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTNASWY FLSENLVPRPKASLNASVSQSDLFVSWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFV TALPEEYNLYGLGEHITQFRLQRNANLTIYPSDDGTPIDQNLYGQHPFYLDTRYYKGDRQNGSY IPVKSSEADASQDYISLSHGVFLRNSHGLEILLRSQKLIWRTLGGGIDLTFYSGPAPADVTRQY LTSTVGLPAMQQYNTLGFHQCRWGYNNWSDLADVVANFEKFEIPLEYIWTDIDYMHGYRNFDND QHRFSYSEGDEFLSKLHESGRYYVPIVDAALYIPNPENASDAYATYDRGAADDVFLKNPDGSLY IGAVWPGYTVFPDWHHPKAVDFWANELVIWSKKVAFDGVWYDMSEVSSFCVGSCGTGNLTLNPA HPSFLLPGEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTSVSYLRTTPTPGVRN VEHPPYVINHDQEGHDLSVHAVSPNATHVDGVEEYDVHGLYGHQGLNATYQGLLEVWSHKRRPF IIGRSTFAGSGKWAGHWGGDNYSKWWSMYYSISQALSFSLFGIPMFGADICGFNGNSDEELCNR WMQLSAFFPFYRNHNELSTIPQEPYRWASVIEATKSAMRIRYAILPYFYTLFDLAHTTGSTVMR ALSWEFPNDPTLAAVETQFMVGPAIMVVPVLEPLVNTVKGVFPGVGHGEVWYDWYTQAAVDAKP GVNTTISAPLGHIPVYVRGGNILPMQEPALTTREARQTPWALLAALGSNGTASGQLYLDDGESI YPNATLHVDFTASRSSLRSSAQGRWKERNPLANVTVLGVNKEPSAVTLNGQAVFPGSVTYNSTS QVLFVGGLQNLTKGGAWAENWVLEW	282	
Transglucosidase	MFVESAKKALLALSLLAASAQAVPRVRRQGASSSFDYKSQIVRGVNLGGWLVTEPWITPSLYDS	283	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
A2RAR6.1	TGGGAVDEWTLCQILGKDEAQAKLSSHWSSFITQSDFDRMAQAGLNHVRIPIGYWAVAPIDGEP YVSGQIDYLDQAVTWARAAGLKVLVDLHGAPGSQNGFDNSGHRGPIQWQQGDTVNQTMTAFDAL ARRYAQSDTVTAIEAVNEPNIPGGVNEDGLKNYYYGALADVQRLNPSTTLFMSDGFQPVESWNG FMQGSNVVMDTHHYQVFDTGLLSMSIDDHVKTACSLATQHTMQSDKPVVVGEWTGALTDCAKYL NGVGNAARYDGTYMSTTKYGDCTGKSTGSVADFSADEKANTRRYIEAQLEAYEMKSGWLFWTWK TEGAPGWDMQDLLANQLFPTSPTDRQYPHQCS		
Transglucosidase A2QX52.1	MPGHSRSRDRLSPSSELDDADPVYSPSVYQREHYYNNDSLFDSADDDYTRTPRNVYSYETHDEY HDDDDDDDDVHEHDHDHEYDDKFEEPWVPLRAQVEGDQWREGFETAIPKEEDVTQAKEYQYQMS GALGDDGPPPLPSDALGRGKGKKRLDRETRRQRRKERLAAFFKHKNGSASAGLVSGDALAKLLG SQDGDEDCLSHLGTERADSMSQKNLEGGRQRKLPVLSEEPMMLRPFPAVAPTGQTQGRVVSGAQ LEEGGPGMEMRHRGGGGPPAEGLLQKEGDWDGSTKGSSTSARPSFWKRYHKTFIFFAILIVLAA IAIPVGIIEARRLHGTSGGDNSSNSNLKGISRDSIPAYARGTYLDPFTWYDTTDFNVTFTNATV GGLSIMGLNSTWNDSAQANENVPPLNEKFPYGSQPIRGVNLGGWLSIEPFIVPSLFDTYTSSEG IIDEWTLSEKLGDSAASVIEKHYATFITEQDFADIRDAGLDHVRIQFSYWAIKTYDGDPYVPKI AWRYLLRAIEYCRKYGLRVNLDPHGIPGSQNGWNHSGRQGTIGWLNGTDGELNRQRSLEMHDQL SQFFAQDRYKNVVTIYGLVNEPLMLSLPVEKVLNWTTEATNLVQKNGIKAWVTVHDGFLNLDKW DKMLKTRPSNMMLDTHQYTVFNTGEIVLNHTRRVELICESWYSMIQQINITSTGWGPTICGEWS QADTDCAQYVNNVGRGTRWEGTFSLTDSTQYCPTASEGTCSCTQANAVPGVYSEGYKTFLQTYA EAQMSAFESAMGWFYWTWATESAAQWSYRTAWKNGYMPKKAYSPSFKCGDTIPSFGNLPEYY	284	
Transglucosidase XP_001389036.2	MSSPQQVYLLPLKDDGSPDVPGGYIYLPAPTNPPYLLRFVIEGSSSICREGALWVNIPEKGESF NRSAFRSFSLSPDFNKNIQIDVPITSAGSFAFYVTFSPLPEFSVLSTPTPEPTRTPTHYIDVSP KLTLRGQDLPLNALSIYSVISKFMGQYPKEWEKHLNGISQRNYNMVHFTPLMKRGASNSPYSIF DQLQFDDAVFPNGEDDVARLISKMENEYGLLSLTDVVWNHTAHNSKWLEEHPEAGYSVETAPWL EAALELDTALLKFGQDLQNLGLPTEFQTVDELMKVMNVMRDKVIAGIRLWEFYAIDVKSDTHKI LDKWKTSKDIDLTDTNWAQLNLQDYKNWTLKQQATFIRDHAIPTSKQVLGRFSRAVDLQFGAAI LTALFGPHNPSTSDTSIVEESLSKILDEVNLPFYEEYDGDVSEIMNQVFNRIKYLRIDDHGPKL GAVTAQSPLIETYFTRLPLNDVTKKHKKEALALVNNGWIWNADALRDNAGPDSRAYLRREVIVW GDCVKLRYGSCRDDNPFLWDFMTDYTRLMAKYFSGFRIDNCHSTPLVVAEYLLDEARKVRPNLT VFAELFTGSEEADYIFVKRLGINALIREAMQAWSTGELSRLVHRHGGRPIGSFDLDLPSSGSSH AIASSGLDSGKEKVVHIRPTPVQALFMDCTHDNEMPAQKRTAKDTLPNGALVAMCASAIGSVIG YDEVYPRLVDLVHEHRLYFSEFSEAPETGLNSLEGGIGGIKKLLNELHTKMGIEGYDETHIHHD GEYITVHRVHPRTRKGVFLIAHTAFPGQDSRSVLAPTHLVGTQVKHIGTWLLEVDTSQTTKERI QADKSYLRGLPSQVKTFEGTKIEESGKDTIISVLNSFVAGSIALFETSMPSVEHASGLDNYITE GVDHAFSDLSLVDLNFALYRCEAEERDSSKGQDGAYDIPGHGPLVYAGLQGWWSVLENIIKYNE LGHPLCDHLRNGQWALDYIVARLEKLSHKEEHPALGRPAAWLQEKFQAVRQLPSFLLPRYFAII VQVAYNAAWKRGIQLLGPHIQKGQEFIHQLGMVSVQQTGYVNSASLWPTKKVPSLAAGLPHFAV DWARCWGRDVFISLRGLLLCTGRFEDAKEHITAFASVLKHGMIPNLLSSGKLPRYNSRDSVWFF LQSIQDYTEMAPDGLEILDHKVPRRFLPYDDVWFPFDDPRAYSQQSTISEIIQEVFQRHAQGLS	285	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	FREYNAGPDLDMQMTQDGFQIDVKVDWETGLIFGGSQYNCGTWQDKMGESAKAGNKGVPGTPRD GAAIEITGLVYSALTWVAKLHERGIYKHDGVDIGGGKSISFEDWASRIRANFERCYYVPLQPKD DGQYDIDANIINRRGIYKDLYRSGKPYEDYQLRSNFPIAMTVAPDLFTASKALAALALADEVLV GPVGMATLDPSDLNYRPNYNNSEDSTDFATAKGRNYHQGPEWVWQRGYFLRAFLHFDLARRTTP AERTETYQQITRRLEGCKRALRESPWKGLTELTNKNGAYCADSSPTQAWSAGCLLDLYYDASRH SQMRIWYGPFKLRYKNIDAIDYYEEKLRRLDEKIQVARQKEYPPTEVAFVTMESIAASQMVVQA ILDPHPMQLLARLAPAPADVVWKNTYLPRSRRMMQSWFITVVIGFLTVFWSVLLIPVAYLLEYE TLHKVFPQLADALARNPLAKSLVQTGLPTLVLSLLTVAVPYLYNWLSNQQGMMSRGDIELSVIS KTFFFSFFNLFLVFTVFGTATTFYGFWENLRDAFKDATTIAFALAKTLENFAPFYINFLCLQGI GLFPFRLLEFGSVAMYPINFLAAKTPRDYAELSTPPTFSYGYSIPQTVLSLIICVVYSVFPSSW LICLFGLIYFTIGKFIYKYQLLYAMDHQQHSTGRAWPMICSRILMGLMVFQLAMIGVLALRRAI TRSLLIVPLLMATVWFSYFFARTYEPLMKFIALKSIDRERPGGGDISPSPSSTFSPPSGLDRDS FPIRIGGQELGLRLRKYVNPSLILPLHDAWLPGRTMVPELQGELEHRNPGNNAADESV		
Transglucosidase XP_001389510.2	MLGSLLLLLPLVGAAVIGPRANSQSCPGYKASNVQKQARSLTADLTLAGTPCNSYGKDLEDLKL LVEYQTDERLHVMIYDADEEVYQVPESVLPRVGSDEDSEDSVLEFDYVEEPFSFTISKGDEVLF DSSASPLVFQSQYVNLRTWLPDDPYVYGLGEHSDPMRLPTYNYTRTLWNRDAYGTPNNTNLYGS HPVYYDHRGKSGTYGVFLLNSNGMDIKINQTTDGKQYLEYNLLGGVLDFYFFYGEDPKQASMEY SKIVGLPAMQSYWTFGFHQCRYGYRDVYELAEVVYNYSQAKIPLETMWTDIDYMDKRRVFTLDP QRFPLEKMRELVTYLHNHDQHYIVMVDPAVSVSNNTAYITGVRDDVFLHNQNGSLYEGAVWPGV TVFPDWFNEGTQDYWTAQFQQFFDPKSGVDIDALWIDMNEASNFCPYPCLDPAAYAISADLPPA APPVRPSSPIPLPGFPADFQPSSKRSVKRAQGDKGKKVGLPNRNLTDPPYTIRNAAGVLSMSTI ETDLIHAGEGYAEYDTHNLYGTMMSSASRTAMQARRPDVRPLVITRSTFAGAGAHVGHWLGDNF SDWVHYRISIAQILSFASMFQIPMVGADVCGFGSNTTEELCARWASLGAFYTFYRNHNELGDIS QEFYRWPTVAESARKAIDIRYKLLDYIYTALHRQSQTGEPFLQPQFYLYPEDSNTFANDRQFFY GDALLVSPVLNEGSTSVDAYFPDDIFYDWYTGAVVRGHGENITLSNINITHIPLHIRGGNIIPV RTSSGMTTTEVRKQGFELIIAPDLDDTASGSLYLDDGDSLNPSSVTELEFTYSKGELHVKGTFG QKAVPKVEKCTLLGKSARTFKGFALDAPVNFKLK	286	
Transglucosidase XP_001391128.2	MPSTYLGALATLAVFPCLGQARSTWPLGSGLELSYQASQHQISIHQDNQTIFSTIPGQPFLSAS AGKDQFVEDSGNFNITNVNQARCRGQNITQLAGIPRSDSVKNQVAVRGYLLDCGGEDIAYGMNF WVPRRFSDRVAFEASVDSEANASFASMKNRSIPIFSREQGVGRGDQPYTAIEDSQGFFSGGDQY TTYTAIPQYVSSDGRVFYLDENDTAYAVFDFQRSDAVTVRYDSLSVHGHLMQADTMLDAITMLT EYTGRMPTLPEWVDHGALLGIQGGQEKVNRIVKQGFEHDCPVAGVWLQDWSGTHLQSAPYGNMN ISRLWWNWESDTSLYPTWAEFVQTLREQHGVRTLAYVNPFLANVSSKSDGYRRNLFLEASQHRY MVQNTTTNSTAIISSGKGIDAGILDLTNEDTRAWFADVLRTQVWSANISGCMWDFGEYTPITPD TSLANISTSAFFYHNQYPRDWAAYQRSVAAEMPLFHEMVTFHRSASMGANRHMNLFWVGDQATL WTRNDGIKSVVTIQGQMGISGYAHSHSDIGGYTTVFEPPTTSNSSGAIPRSAELLGRWGELGAV SSAVFRSHEGNVPSVNAQFYSNSTTYAYFAYNARLFRSLGPYRRRILNTESQRRGWPLLRMPVL YHPEDLRARQISYESFFLGRDLYVAPVLDEGHKSVEVYFPGHSANRTYTHVWTGQTYRAGQTAK	287	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	VSAPFGKPAVFLVDGASSPELDVFLDFVRKENGTVL YA		
Transglucosidase XP_001395384.2	MDPANEYCGLEDYGLVGDMHTCALVSKNGSVDSMCWPVFDSPSIFCRILDKEKGGHFSITPDRR LKNPLSKQRYRPYTNMLETRWIHEEGVMNILDYFPIAKPKPHVVEKGLPQWCRCYQNKGSAQYQ ACRSGMVRKAECVRGEMEIEIELFPAFNYARDSHVAQQSSASDDAIQVYHFQAESQNLVVSVLG DRGDISGDDSDLSIEFELSDRPGHLGPGLVGKVTLKEGQSITMLLHDQESITCNVEDLAPYLQQ IERTTGDFWSDWTSKCTFRGHYREQVERSLLVLKLLTYKPTGAIVAAPTFSLPEHIGGSRNWDY RYSWVRDAAFTVYVFLKNGYPEEAESYINFIFERIFPPMDKNPKPGEPFLPIMITIHGEREIPE MELEHLEGYRGSRPVRIGNGAATHIQLDIYGELMDSIYLYNKHAADISYDQWRAIRRMIDFVIQ IRHQPDQSIWEVRGPPQNFVYSKIMLWVALDRGLRLAEKRSNLPCPDRARWMHERDALYDEIMT KGYNSEKGFFCMSYENQDAMDAAVLIAPLVFFVAPNDPRLLSTIQKITEVPAKGGLSVANMVSR YDTGKVDDGVGGNEGAFLMVTFWLVEAMMRAARSKAYLPHDPFFQQLRKTATSQFDSILSFANH LGMFSEEVATSGEQIGNMPQAFSHLACVSAAMNLGGGGDR	288	
Transglucosidase XP_001396506.2	MCNKSNYSSPKWWKESVVYQVYPASFNCGKSTTNTNGWGDVTGIIEKVPYLESLGVDIVWLSPI YTSPQVDMGYDIADYESIDPRYGTLADVDLLIKTLKDHDMKLMMDLVVNHTSDQHSWFVESANS KDSPKRDWYIWRPAKGFDEAGNPVPPNNWAQILGDTLSAWTWHAETQEFYLTLHTSAQAELNWE NPDVVTAVYDVMEFWLRRGICGFRMDVINFISKDQSFPDAPIIDPASKYQPGEQFYTNGPRFHE FMHGIYDNVLSKYDTITVGETPYVTDMKEIIKTVGSTAKELNMAFNFDHMEIEDIKTKGESKWS LRDWKLTELKGILSGWQKRMREWDGWNAIFLECHDQARSVSRYTNDSDEFRDRGAKLLALLETT LGGTIFLYQGQEIGMRNFPVEWDPDTEYKDIESVNFWKKSKELHPVGSEGLAQARTLLQKKARD HARTPMQWSADPHAGFTVPDATPWMRVNDDYGTVNVEAQMSFPWEMKGELSVWQYWQQALQRRK LHKGAFVYGDFEDLDYHNELVFAYSRTSADGKETWLVAMNWTTDAVEWTVPSGIHVTRWVSSTL QTAPLMAGQSTVTLRALEGVVGCCS	289	
Transglucosidase XP_001398938.2	MGLCVGWRWILLCVVMGAAVCGTDKTATMRWHKLLPGVLALLPLSVAQSCWRNTTCSGPTESAF SGPWEKNIFAPSSRTVNPEKLFLITQPDKTEEYSPFALHGNGSLVVYDFGKEVGGIVSVNFSST GSGALGVAFTEAKNWIGEWSDSSNGGFKGPDGALYGNFTEAGSHYYVMPDKSLRGGFRYLTLFL ITSDNSTIQIEDVNLEIGFQPTWSNLKAYQGYFHSNDDLLNKIWYTGAYTLQTNEVPTDTGRQI PAMAVGWANNCTLGPGDTIIVDGAKRDRAVWPGDMGIAVPSAFVSLGDLDSVKNALQVMYDTQN NSTGAFDESGPPLSQKDSDTYHMWTMVGTYNYMLFTNDSDFLERNWEGYQKAMDYIYGKVTYPS GLLNVTGTRDWARWQQGYNNSEAQMILYHTLNTGAELATWAGDSGDLSSTWTSRAEKLRQAINE YCWDDSYGAFKDNATDTTLHPQDANSMALLFGVVDADRAASISERLTDNWTPIGAVAPELPENI SPFISSFEIQGHLTVGQPQRALELIRRSWGWYYNNANGTQSTVIEGYLQNGTFGYRSDRGYYYD TAYVSHSHGWSSGPTSALTNYIVGISVTSPLGATWRIAPQFVDLQSAEGGFTTSLGKFQAGWSK TDKGYTLDFTVPHGTQGNLTLPFVSAAKPSIKIDGTEISRGVQYANSTATVTVSGGGTYKVEVQ	290	
Betaglucosidase Protein ID: H9ZGE3|H9ZGE3	MAMQLRSLLLCVLLLLLGFALADTNAAARIHPPVVCANLSRANFDTLVPGFVFGAATASYQVEG AANLDGRGPSIWDTFTHKHPEKIADGSNGDVAIDQYHRYKEDVAIMKDMGLESYRFSISWSRVL PNGTLSGGINKKGIEYYNNLINELLHNGIEPLVTLFHWDVPQTLEDEYGGFLSNRIVNDFEEYA ELCFKKFGDRVKHWTTLNEPYTFSSHGYAKGTHAPGRCSAWYNQTCFGGDSATEPYLVTHNLLL AHAAAVKLYKTKYQAYQKGVIGITVVTPWFEPASEAKEDIDAVFRALDFIYGWFMDPLTRGDYP	292	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	QSMRSLVGERLPNFTKKESKSLSGSFDYIGINYYSARYASASKNYSGHPSYLNDVNVDVKTELN GVPIGPQAASSWLYFYPKGLYDLLCYTKEKYNDPIIYITENGVDEFNQPNPKLSLCQLLDDSNR IYYYYHHLCYLQAAIKEGVKVKGYFAWSLLDNFEWDNGYTVRFGINYVDYDNGLKRHSKHSTHW FKSFLKKSSRNTKKIRRCGNNNTSATKFVF		
UGT73-251_5	MDSPPQKPHFLLFPFMAQGHMIPMIDLAKLLAQRGAIITVVTTPHNAARYHSVLARAIDSGLHI HVLQLQFPCNEGGLPEGCENFDLLPSLGSASTFFRATFLLYEPSEKVFEELIPRPTCIISDMCL PWTVRLAQKYHVPRLVFYSLSCFFLLCMRSLKNNQALISSKSDSELVTFSDLPDPVEFLKSQLP KSNDEEMAKFGYEIGEADRQSHGVIVNVFEEMEPKYLAEYRKERESPEKVWCVGPVSLCNDNKL DKAQRGNKASIDERECIEWLDGQQPSSVVYVSLGSLCNLVTAQLIELGLGLEASNKPFIWVIRK GNITEELQKWLVEYDFEEKTKGRGLVILGWAPQVLILSHPAIGCFLTHCGWNSSIEGISAGMPM ITWPLFADQVFNEKLIVEILRIGVSVGMETAMHWGEEEEKGVVVKREKVREAIERAMDGDEREE RRERCKELAEMAKRAVEEGGSSHRNLTLLTEDILVNGGGQERMDDADDFPTIVN	293	Disclosed in Itkin et al., 2016, and WO 2016/038617
UGT73-251-6	MDSPPHRPHFLLFPFMAQGHMIPMIDLAKLLAQRGAIVTILTTPHNAARTHSVLARAIDSGLQI RVRPLQFPCKEAGLPEGCENLDLLPSLGSASTFFRATCLLYDPSEKLFEELSPRPTCIISDMCL PWTIRLAQKYHVPRLVFYSLSCFFLLCMRSLKNNPALISSKSDSEFVTFSDLPDPVEFLKSELP KSTDEDLVKFSYEMGEADRKSYGVILNIFEEMEPKYLAEYGNERESPEKVWCVGPVSLCNDNKL DKAQRGNKASIDERECIKWLGGQQPSSVVYASLGSLCNLVTAQFIELGLGLEASNKPFIWVIRK GNITEELQKWLVEYDFEEKTKGRGLVILGWAPQVLILSHPSIGCFLTHCGWNSSIEGISAGVPM VTWPLFSDQVFNEKLIVQILRIGVSVGAETAMNWGEEEEKGVVVKREKVREAIERMMDGDEREE RRERCKELAETAKRAIEEGGSSHRNLTLLIEDIGTSLRRL	294	Disclsoed in Itkin et al., 2016, and WO 2016/038617
UGT73-327-2	MGSAGVELKVAFLPFAAPGHMIPLMNIARLFAMHGADVTFITTPATASRFQNVVDSDLRRGHKI KLHTFQLPSAEAGLPPGVESFNECTSKEMTEKLFGAFEMLNGDIEQFLKGAKVDCIVSDTILVW TLDAAARLGIPRIAFRSSGFFSECIHHSLRCHKPHKKVGSDTEPFIFPGLPHKIEITRLNIPQW YSEEGYIQHIEKMKEMDKKSYAVLLNTFYELEADYVEYFESVIGLKTWIVGPVSLWANEGGGKN DSRTENNNAELMEWLDSKQPNSVLYVSFGSMTKFPSAQVLEIAHGLEDSGCHFIWVVRKMNESE AADEEFPEGFEERVRESKRGLIIRDWAPQELILNHAAVGGFVTHCGWNSILESVCAGRPIIAWP LSAEQFFNEKFVTRVLKVGVSIGVRKWWGSTSSETLDVVKRDRIAEAVARLMGDDREVVEMRDG VRELSHAAKRAIKEGGSSHSTLLSLIHELKTMKFKRQSSNVDG	295	Disclosed in Itkin et al., 2016, and WO 2016/038617
UGT74-345-2	MDETTVNGGRRASDVVVFAFPRHGHMSPMLQFSKRLVSKGLRVTFLITTSATESLRLNLPPSSS LDLQVISDVPESNDIATLEGYLRSFKATVSKTLADFIDGIGNPPKFIVYDSVMPWVQEVARGRG LDAAPFFTQSSAVNHILNHVYGGSLSIPAPENTAVSLPSMPVLQAEDLPAFPDDPEVVMNFMTS QFSNFQDAKWIFFNTFDQLECKVVNWMADRWPIKTVGPTIPSAYLDDGRLEDDRAFGLNLLKPE DGKNTRQWQWLDSKDTASVLYISFGSLAILQEEQVKELAYFLKDTNLSFLWVLRDSELQKLPHN FVQETSHRGLVVNWCSQLQVLSHRAVSCFVTHCGWNSTLEALSLGVPMVAIPQWVDQTTNAKFV ADVWRVGVRVKKKDERIVTKEELEASIRQVVQGEGRNEFKHNAIKWKKLAKEAVDEGGSSDKNI EEFVKTIA	296	Itkin et al., 2016, and WO 2016/038617
UGT75-281-2	MMRNHHFLLVCFPSQGYINPSLQLARRLISLGVNVTFATTVLAGRRMKNKTHQTATTPGLSFAT	297	Itkin et al.,

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	FSDGFDDETLKPNGDLTHYFSELRRCGSESLTHLITSAANEGRPITFVIYSLLLSWAADIASTY DIPSALFFAQPATVLALYFYYFHGYGDTICSKLQDPSSYIELPGLPLLTSQDMPSFFSPSGPHA FILPPMREQAEFLGRQSQPKVLVNTFDALEADALRAIDKLKMLAIGPLIPSALLGGNDSSDASF CGDLFQVSSEDYIEWLNSKPDSSVVYISVGSICVLSDEQEDELVHALLNSGHTFLWVKRSKENN EGVKQETDEEKLKKLEEQGKMVSWCRQVEVLKHPALGCFLTHCGWNSTIESLVSGLPVVAFPQQ IDQATNAKLIEDVWKTGVRVKANTEGIVEREEIRRCLDLVMGSRDGQKEEIERNAKKWKELARQ AIGEGGSSDSNLKTFLWEIDLEI		2016, and WO 2016/038617
UGT85-269-4	MAEQAHDLLHVLLFPFPAEGHIKPFLCLAELLCNAGFHVTFLNTDYNHRRLHNLHLLAARFPSL HFESISDGLPPDQPRDILDPKFFISICQVTKPLFRELLLSYKRISSVQTGRPPITCVITDVIFR FPIDVAEELDIPVFSFCTFSARFMFLYFWIPKLIEDGQLPYPNGNINQKLYGVAPEAEGLLRCK DLPGHWAFADELKDDQLNFVDQTTASSRSSGLILNTFDDLEAPFLGRLSTIFKKIYAVGPIHSL LNSHHCGLWKEDHSCLAWLDSRAAKSVVFVSFGSLVKITSRQLMEFWHGLLNSGKSFLFVLRSD VVEGDDEKQVVKEIYETKAEGKWLVVGWAPQEKVLAHEAVGGFLTHSGWNSILESIAAGVPMIS CPKIGDQSSNCTWISKVWKIGLEMEDRYDRVSVETMVRSIMEQEGEKMQKTIAELAKQAKYKVS KDGTSYQNLECLIQDIKKLNQIEGFINNPNFSDLLRV	298	Itkin et al., 2016, and WO 2016/038617
UGT85-269-1	MVQPRVLLFPFPALGHVKPFLSLAELLSDAGIDVVFLSTEYNHRRISNTEALASRFPTLHFETI PDGLPPNESRALADGPLYFSMREGTKPRFRQLIQSLNDGRWPITCIITDIMLSSPIEVAEEFGI PVIAFCPCSARYLSIHFFIPKLVEEGQIPYADDDPIGEIQGVPLFEGLLRRNHLPGSWSDKSAD ISFSHGLINQTLAAGRASALILNTFDELEAPFLTHLSSIFNKIYTIGPLHALSKSRLGDSSSSA SALSGFWKEDRACMSWLDCQPPRSVVFVSFGSTMKMKADELREFWYGLVSSGKPFLCVLRSDVV SGGEAAELIEQMAEEEGAGGKLGMVVEWAAQEKVLSHPAVGGFLTHCGWNSTVESIAAGVPMMC WPILGDQPSNATWIDRVWKIGVERNNREWDRLTVE	300	Itkin et al., 2016, and WO 2016/038617
UGT94-289-1	MDAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRRNFHIYFCSTSVNLDAIKPKLPSSSSSDSI QLVELCLPSSPDQLPPHLHTTNALPPHLMPTLHQAFSMAAQHFAAILHTLAPHLLIYDSFQPWA PQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFPISEFVLHDYWKAMYSAAGGAVTKKDHK IGETLANCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIKN WLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAIEDALPKGFL ERVGERGMVVKGWAPQAKILKHWSTGGFVSHCGWNSVMESMMFGVPIIGVPMHLDQPFNAGLAE EAGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEKMDEMVAAIS LFLKI	301	Itkin et al., 2016, and WO 2016/038617
UGT94-289-2	MDAQQGHTTTILMLPWVGYGHLLPFLELAKSLSRRKLFHIYFCSTSVSLDAIKPKLPPSISSDD SIQLVELRLPSSPELPPHLHTTNGLPSHLMPALHQAFVMAAQHFQVILQTLAPHLLIYDILQPW APQVASSLNIPAINFSTTGASMLSRTLHPTHYPSSKFPISEFVLHNHWRAMYTTADGALTEEGH KIEETLANCLHTSCGVVLVNSFRELETKYIDYLSVLLNKKVVPVGPLVYEPNQEGEDEGYSSIK NWLDKKEPSSTVFVSFGTEYFPSKEEMEEIAYGLELSEVNFIWVLRFPQGDSTSTIEDALPKGF LERAGERAMVVKGWAPQAKILKHWSTGGLVSHCGWNSMMEGMMFGVPIIAVPMHLDQPFNAGLV EEAGVGVEAKRDSDGKIQREEVAKSIKEVVIEKTREDVRKKAREMGEILRSKGDEKIDELVAEI SLLRKKAPCSI	302	Itkin et al., 2016, and WO 2016/038617

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
UGT94-289-3	MDAAQQGDTTTILMLPWLGYGHLSAFLELAKSLSRRNFHIYFCSTSVNLDAIKPKLPSSFSDSI QFVELHLPSSPEFPPHLHTTNGLPPTLMPALHQAFSMAAQHFESILQTLAPHLLIYDSLQPWAP RVASSLKIPAINFNTTGVFVISQGLHPIHYPHSKFPFSEFVLHNHWKAMYSTADGASTERTRKR GEAFLYCLHASCSVILINSFRELEGKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIKNW LDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVNFIWVVRFPQGDNTSGIEDALPKGFLE RAGERGMVVKGWAPQAKILKHWSTGGFVSHCGWNSVMESMMFGVPIIGVPMHVDQPFNAGLVEE AGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEKFDEMVAEISL LLKI	303	Itkin et al., 2016, and WO 2016/038617
UDP-glycotransferase (330)	MDTRKRSIRILMFPWLAHGHISAFLELAKSLAKRNFVIYICSSQVNLNSISKNMSSKDSISVKL VELHIPTTILPPPYHTTNGLPPHLMSTLKRALDSARPAFSTLLQTLKPDLVLYDFLQSWASEEA ESQNIPAMVFLSTGAAAISFIMYHWFETRPEEYPFPAIYFREHEYDNFCRFKSSDSGTSDQLRV SDCVKRSHDLVLIKTFRELEGQYVDFLSDLTRKRFVPVGPLVQEVGCDMENEGNDIIEWLDGKD RRSTVFSSFGSEYFLSANEIEEIAYGLELSGLNFIWVVRFPHGDEKIKIEEKLPEGFLERVEGR GLVVEGWAQQRRILSHPSVGGFLSHCGWSSVMEGVYSGVPIIAVPMHLDQPFNARLVEAVGFGE EVVRSRQGNLDRGEVARVVKKLVMGKSGEGLRRRVEELSEKMREKGEEEIDSLVEELVTVVRRR ERSNLKSENSMKKLNVMDDGE	304	Disclosed in Noguchi et al., 2008)(Plant J. 2008 May;54(3):415-27)
UDP-glycotransferase (330)	ATGGATACAAGAAAGAGAAGCATCAGGATTCTAATGTTCCCATGGCTTGCTCATGGCCATATCT CAGCATTCCTCGAGCTGGCGAAGTCACTTGCCAAAAGAAACTTCGTCATTTACATTTGTTCTTC ACAAGTAAATCTAAATTCCATCAGCAAGAACATGTCATCAAAAGACTCCATTTCCGTAAAACTT GTTGAGCTTCACATTCCCACCACCATACTTCCCCCTCCTTACCACACCACCAATGGCCTCCCAC CCCACCTCATGTCCACCCTCAAGAGAGCCCTCGACAGTGCCCGGCCCGCCTTCTCCACCCTCCT CCAAACCCTCAAGCCCGACTTGGTTTTATACGATTTCCTCCAGTCGTGGGCCTCGGAGGAGGCC GAGTCGCAGAATATACCAGCCATGGTGTTTCTGAGTACCGGAGCTGCAGCGATTTCTTTTATTA TGTACCATTGGTTTGAGACCAGACCGGAGGAGTACCCTTTTCCGGCTATATACTTCCGGGAACA CGAGTATGATAACTTCTGCCGTTTTAAGTCTTCCGACAGCGGTACTAGTGATCAATTGAGAGTC AGCGATTGCGTTAAACGGTCGCACGATTTGGTTCTGATCAAGACATTCCGTGAACTGGAAGGAC AATACGTAGATTTTCTCTCCGACTTGACTCGGAAGAGATTCGTACCAGTTGGCCCCCTTGTTCA GGAGGTAGGTTGTGATATGGAGAATGAAGGAAATGACATCATCGAATGGCTCGACGGGAAAGAC CGTCGTTCGACGGTTTTCTCCTCATTCGGGAGCGAGTACTTCTTGTCTGCCAATGAGATCGAAG AGATAGCTTATGGGCTGGAGCTAAGCGGGCTTAACTTCATCTGGGTTGTTAGGTTTCCTCATGG CGACGAGAAAATCAAGATTGAGGAGAAACTGCCGGAAGGGTTTCTTGAGAGAGTGGAAGGAAGA GGGTTGGTGGTGGAGGGATGGGCACAGCAGAGGAGAATATTGTCACATCCGAGTGTTGGAGGGT TTTTGAGCCACTGTGGGTGGAGTTCTGTGATGGAAGGGGTGTATTCCGGTGTGCCGATTATTGC CGTGCCGATGCATCTTGACCAGCCGTTCAATGCTAGGTTGGTGGAGGCGGTGGGGTTTGGGGAG GAGGTGGTGAGGAGTAGACAAGGAAATCTTGACAGAGGAGAGGTGGCGAGGGTGGTGAAGAAGC TGGTTATGGGGAAAAGTGGGGAGGGGTTACGGCGGAGGGTGGAGGAGTTGAGTGAGAAGATGAG AGAGAAAGGGGAGGAGGAGATTGATTCACTGGTGGAGGAATTGGTGACGGTGGTTAGGAGGAGA GAGAGATCGAATCTCAAGTCTGAGAATTCTATGAAGAAATTGAATGTGATGGATGATGGAGAAT	305	Disclosed in Noguchi et al., 2008

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AG		
UGT98 protein [*S. grosvenorii*]	MDAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRRNFHIYFCSTSVNLDAIKPKLPSSSSSDSI QLVELCLPSSPDQLPPHLHTTNALPPHLMPTLHQAFSMAAQHFAAILHTLAPHLLIYDSFQPWA PQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFPISEFVLHDYWKAMYSAAGGAVTKKDHK IGETLANCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIKN WLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAIEDALPKGFL ERVGERGMVVKGWAPQAKILKHWSTGGFVSHCGWNSVMESMMFGVPIIGVPMHLDQPFNAGLAE EAGVGVEAKRDSDGKIQREEVAKSIKEVVIEKTREDVRKKAREMGEILRSKGDEKIDELVAEIS LLRKKAPCSIAAALEHHHHHH	306	
UGT98 gene [*S. grosvenorii*]	CTCGAATTCATGGATGCCCAGCGAGGTCACACCACAACCATTTTGATGTTTCCATGGCTCGGCT ATGGCCATCTTTCGGCTTTCCTAGAGTTGGCCAAAAGCCTCTCAAGGAGGAACTTCCATATCTA CTTCTGTTCAACCTCTGTTAACCTCGACGCCATTAAACCAAAGCTTCCTTCTTCTTCCTCTTCT GATTCCATCCAACTTGTGGAACTTTGTCTTCCATCTTCTCCTGATCAGCTCCCTCCTCATCTTC ACACAACCAACGCCCTCCCCCCTCACCTCATGCCCACTCTCCACCAAGCCTTCTCCATGGCTGC CCAACACTTTGCTGCCATTTTACACACACTTGCTCCGCATCTCCTCATTTACGACTCTTTCCAA CCTTGGGCTCCTCAACTAGCTTCATCCCTCAACATTCCAGCCATCAACTTCAATACTACGGGAG CTTCAGTCCTGACCCGAATGCTTCACGCTACTCACTACCCAAGTTCTAAATTCCCAATTTCAGA GTTTGTTCTCCACGATTATTGGAAAGCCATGTACAGCGCCGCCGGTGGGGCTGTTACAAAAAAA GACCACAAAATTGGAGAAACACTTGCGAATTGCTTGCATGCTTCTTGTAGTGTAATTCTAATCA ATAGTTTCAGAGAGCTCGAGGAGAAATATATGGATTATCTCTCCGTTCTCTTGAACAAGAAAGT TGTTCCGGTTGGTCCTTTGGTTTACGAACCGAATCAAGACGGGGAAGATGAAGGTTATTCAAGC ATCAAAAATTGGCTTGACAAAAAGGAACCGTCCTCCACCGTCTTCGTTTCATTTGGAAGCGAAT ACTTCCCGTCAAAGGAAGAAATGGAAGAGATAGCCCATGGGTTAGAGGCGAGCGAGGTTCATTT CATCTGGGTCGTTAGGTTTCCTCAAGGAGACAACACCAGCGCCATTGAAGATGCCTTGCCGAAG GGGTTTCTGGAGAGGGTGGGAGAGAGAGGGATGGTGGTGAAGGGTTGGGCTCCCCAGGCGAAGA TACTGAAGCATTGGAGCACAGGGGGATTCGTGAGCCACTGTGGATGGAACTCGGTGATGGAAAG CATGATGTTTGGCGTTCCCATAATAGGGGTTCCGATGCATCTGGACCAGCCCTTTAACGCCGGA CTCGCGGAAGAAGCTGGCGTCGGCGTGGAAGCCAAGCGAGATTCGGACGGCAAAATTCAAAGAG AAGAAGTTGCAAAGTCGATCAAAGAAGTGGTGATTGAGAAAACCAGGGAAGACGTGAGGAAGAA AGCAAGAGAAATGGGTGAGATTTTGAGGAGTAAAGGAGATGAGAAAATTGATGAGTTGGTGGCT GAAATTTCTCTTTTGCGCAAAAAGGCCCCATGTTCAATTGCGGCCGCACTCGAGCACCACCACC ACCACCACTGA	307	
CYP 1798	MEMSSSVAATISIWMVVVCIVGVGWRVVNWVWLRPKKLEKRLREQGLAGNSYRLLFGDLKERAA MEEQANSKPINFSHDIGPRVFPSMYKTIQNYGKNSYMWLGPYPRVHIMDPQQLKTVFTLVYDIQ KPNLNPLIKFLLDGIVTHEGEKWAKHRKIINPAFHLEKLKDMIPAFFHSCNEIVNEWERLISKE GSCELDVMPYLQNLAADAISRTAFGSSYEEGKMIFQLLKELTDLVVKVAFGVYIPGWRFLPTKS NNKMKEINRKIKSLLLGIINKRQKANMEEGEAGQSDLLGILMESNSNEIQGEGNNKEDGMSIED VIEECKVFYIGGQETTARLLIWTMILLSSHTEWQERARTEVLKVFGNKKPDFDGLSRLKVVTMI	308	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	LNEVLRLYPPASMLTRIIQKETRVGKLTLPAGVILIMPIILIHRDHDLWGEDANEFKPERFSKG VSKAAKVQPAFFPFGWGRICMGQNFAMIEAKMALSLILQRFSFELSSSYVHAPTVVFTTQPQHG AHIVLRKL		
Epoxide hydrolase	MDAIEHRTVSVNGINSHVAEKGEGPVVLLLHGFPELWYSWRHQILALSSLGYRAVAPDLRGYGD TDAPGSISSYTCFHIVGDLVALVESLGMDRVFVVAHDWGAMIAWCLCLFRPEMVKAFVCLSVPF RQRNPKMKPVQSMRAFFGDDYYICRFQNPGEIEEEMAQVGAREVLRGILTSRRPGPPILPKGQA FRARPGASTALPSWLSEKDLSFFASKYDQKGFTGPLNYYRAMDLNWELTASWTGVQVKVPVKYI VGDVDMVFTTPGVKEYVNGGGFKKDVPFLQEVVIMEGVGHFINQEKQEISSHIMDFISKF	309	
Epoxide hydrolase	MDEIEHITINTNGIKMHIASVGTGPVVLLLHGFPELWYSWRHQLLYLSSVGYRAIAPDLRGYGD TDSPASPTSYTALHIVGDLVGALDELGIEKVFLVGHDWGAIIAWYFCLFRPDRIKALVNLSVQF IPRNPAIPFIEGFRTAFGDDFYICRFQVPGEAEEDFASIDTAQLFKTSLCNRSSAPPCLPKEIG FRAIPPPENLPSWLTEEDINFYAAKFKQTGFTGALNYYRAFDLTWELTAPWTGAQIQVPVKFIV GDSDLTYHFPGAKEYIHNGGFKRDVPLLEEVVVVKDACHFFNQERPQEINAHIHDFINKF	310	
Epoxide hydrolase	MENIEHTTVQTNGIKMHVAAIGTGPPVLLLHGFPELWYSWRHQLLYLSSAGYRAIAPDLRGYGD TDAPPSPSSYTALHIVGDLVGLLDVLGIEKVFLIGHDWGAIIAWYFCLFRPDRIKALVNLSVQF FPRNPTTPFVKGFRAVLGDQFYMVRFQEPGKAEEEFASVDIREFFKNVLSNRDPQAPYLPNEVK FEGVPPPALAPWLTPEDIDVYADKFAETGFTGGLNYYRAFDRTWELTAPWTGARIGVPVKFIVG DLDLTYHFPGAQKYIHGEGFKKAVPGLEEVVVMEDTSHFINQERPHEINSHIHDFFSKFC	311	
Epoxide hydrolase	MDQIEHITINTNGIKMHIASVGTGPVVLLLHGFPELWYSWRHQLLYLSSVGYRAIAPDLRGYGD TDSPASPTSYTALHIVGDLVGALDELGIEKVFLVGHDWAAIIAWYFCLFRPDRIKALVNLSVQF IPRNPAIPFIEGFRTAFGDDFYMCRFQVPGEAEEDFASIDTAQLFKTSLCNRSSAPPCLPKEIG FRAIPPPENLPSWLTEEDINYYAAKFKQTGFTGALNYYRAFDLTWELTAPWTGAQIQVPVKFIV GDSDLTYHFPGAKEYIHNGGFKKDVPLLEEVVVVKDACHFINQERPQEINAHIHDFINKF	312	
Epoxide hydrolase	MEKIEHSTIATNGINMHVASAGSGPAVLFLHGFPELWYSWRHQLLYLSSLGYRAIAPDLRGFGD TDAPPSPSSYTAHHIVGDLVGLLDQLGVDQVFLVGHDWGAMMAWYFCLFRPDRVKALVNLSVHF TPRNPAISPLDGFRLMLGDDFYVCKFQEPGVAEADFGSVDTATMFKKFLTMRDPRPPIIPNGFR SLATPEALPSWLTEEDIDYFAAKFAKTGFTGGFNYYRAIDLTWELTAPWSGSEIKVPTKFIVGD LDLVYHFPGVKEYIHGGGFKKDVPFLEEVVVMEGAAHFINQEKADEINSLIYDFIKQF	313	
Epoxide hydrolase	MEKIEHTTISTNGINMHVASIGSGPAVLFLHGFPELWYSWRHQLLFLSSMGYRAIAPDLRGFGD TDAPPSPSSYTAHHIVGDLVGLLDQLGIDQVFLVGHDWGAMMAWYFCLFRPDRVKALVNLSVHF LRRHPSIKFVDGFRALLGDDFYFCQFQEPGVAEADFGSVDVATMLKKFLTMRDPRPPPMIPKEKG FRALETPDPLPAWLTEEDIDYFAGKFRKTGFTGGFNYYRAFNLTWELTAPWSGSEIKVAAKFIV GDLDLVYHFPGAKEYIHGGGFKKDVPLLEEVVVVDGAAHFINQERPAEISSLIYDFIKKF	314	
CYP87D18	MWTVVLGLATLFVAYYIHWINKWRDSKFNGVLPPGTMGLPLIGETIQLSRPSDSLDVHPFIQKK VERYGPIFKTCLAGRPVVVSADAEFNNYIMLQEGRAVEMWYLDTLSKFFGLDTEWLKALGLIHK YIRSITLNHFGAEALRERFLPFIEASSMEALHSWSTQPSVEVKNASALMVFRTSVNKMFGEDAK KLSGNIPGKFTKLLGGFLSLPLNFPGTTYHKCLKDMKEIQKKLREVVDDRLANVGPDVEDFLGQ	315	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ALKDKESEKFISEEFIIQLLFSISFASFESISTTLTLILKLLDEHPEVVKELEAEHEAIRKARA DPDGPITWEEYKSMTFTLQVINETLRLGSVTPALLRKTVKDLQVKGYIIPEGWTIMLVTASRHR DPKVYKDPHIFNPWRWKDLDSITIQKNFMPFGGGLRHCAGAEYSKVYLCTFLHILCTKYRWTKL GGGRIARAHILSFEDGLHVKFTPKE		
CYP87D18 gene sequence	ATGTGGACTGTCGTGCTCGGTTTGGCGACGCTGTTTGTCGCCTACTACATCCATTGGATTAACA AATGGAGAGATTCCAAGTTCAACGGAGTTCTGCCGCCGGGCACCATGGGTTTGCCGCTCATCGG AGAGACGATTCAACTGAGTCGACCCAGTGACTCCCTCGACGTTCACCCTTTCATCCAGAAAAAA GTTGAAAGATACGGGCCGATCTTCAAAACATGTCTGGCCGGAAGGCCGGTGGTGGTGTCGGCGG ACGCAGAGTTCAACAACTACATAATGCTGCAGGAAGGAAGAGCAGTGGAAATGTGGTATTTGGA TACGCTCTCCAAATTTTTCGGCCTCGACACCGAGTGGCTCAAAGCTCTGGGCCTCATCCACAAG TACATCAGAAGCATTACTCTCAATCACTTCGGCGCCGAGGCCCTGCGGGAGAGATTTCTTCCTT TTATTGAAGCATCCTCCATGGAAGCCCTTCACTCCTGGTCTACTCAACCTAGCGTCGAAGTCAA AAATGCCTCCGCTCTCATGGTTTTTAGGACCTCGGTGAATAAGATGTTCGGTGAGGATGCGAAG AAGCTATCGGGAAATATCCCTGGGAAGTTCACGAAGCTTCTAGGAGGATTTCTCAGTTTACCAC TGAATTTTCCCGGCACCACCTACCACAAATGCTTGAAGGATATGAAGGAAATCCAGAAGAAGCT AAGAGAGGTTGTAGACGATAGATTGGCTAATGTGGGCCCTGATGTGGAAGATTTCTTGGGGCAA GCCCTTAAAGATAAGGAATCAGAGAAGTTCATTTCAGAGGAGTTCATCATCCAACTGTTGTTTT CTATCAGTTTTGCTAGCTTTGAGTCCATCTCCACCACTCTTACTTTGATTCTCAAGCTCCTTGA TGAACACCCAGAAGTAGTGAAAGAGTTGGAAGCTGAACACGAGGCGATTCGAAAAGCTAGAGCA GATCCAGATGGACCAATTACTTGGGAAGAATACAAATCCATGACTTTTACATTACAAGTCATCA ATGAAACCCTAAGGTTGGGGAGTGTCACACCTGCCTTGTTGAGGAAAACAGTTAAAGATCTTCA AGTAAAAGGATACATAATCCCGGAAGGATGGACAATAATGCTTGTCACCGCTTCACGTCACAGA GACCCAAAAGTCTATAAGGACCCTCATATCTTCAATCCATGGCGTTGGAAGGACTTGGACTCAA TTACCATCCAAAAGAACTTCATGCCTTTTGGGGGAGGCTTAAGGCATTGTGCTGGTGCTGAGTA CTCTAAAGTCTACTTGTGCACCTTCTTGCACATCCTCTGTACCAAATACCGATGGACCAAACTT GGGGGAGGAAGGATTGCAAGAGCTCATATATTGAGTTTTGAAGATGGGTTACATGTGAAGTTCA CACCCAAGGAATGA	316	
AtCPR protein	MTSALYASDLFKQLKSIMGTDSLSDDVVLVIATTSLALVAGFVVLLWKKTTADRSGELKPLMIP KSLMAKDEDDDLDLGSGKTRVSIFFGTQTGTAEGFAKALSEEIKARYEKAAVKDDYAADDDQYE EKLKKETLAFFCVATYGDGEPTDNAARFYKWFTEENERDIKLQQLAYGVFALGNRQYEHFNKIG IVLDEELCKKGAKRLIEVGLGDDDQSIEDDFNAWKESLWSELDKLLKDEDDKSVATPYTAVIPE YRVVTHDPRFTTQKSMESNVANGNTTIDIHHPCRVDVAVQKELHTHESDRSCIHLEFDISRTGI TYETGDHVGVYAENHVEIVEEAGKLLGHSLDLVFSIHADKEDGSPLESAVPPPFPGPCTLGTGL ARYADLLNPPRKSALVALAAYATEPSEAEKLKHLTSPDGKDEYSQWIVASQRSLLEVMAAFPSA KPPLGVFFAAIAPRLQPRYYSISSSPRLAPSRVHVTSALVYGPTPTGRIHKGVCSTWMKNAVPA EKSHECSGAPIFIRASNFKLPSNPSTPIVMVGPGTGLAPFRGFLQERMALKEDGEELGSSLLFF GCRNRQMDFIYEDELNNFVDQGVISELIMAFSREGAQKEYVQHKMMEKAAQVWDLIKEEGYLYV CGDAKGMARDVHRTLHTIVQEQEGVSSSEAEAIVKKLQTEGRYLRDVW.	317	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
AtCPR gene sequence	ATGACTTCTGCTTTGTATGCTTCCGATTTGTTTAAGCAGCTCAAGTCAATTATGGGGACAGATT CGTTATCCGACGATGTTGTACTTGTGATTGCAACGACGTCTTTGGCACTAGTAGCTGGATTTGT GGTGTTGTTATGGAAGAAAACGACGGCGGATCGGAGCGGGGAGCTGAAGCCTTTGATGATCCCT AAGTCTCTTATGGCTAAGGACGAGGATGATGATTTGGATTTGGGATCCGGGAAGACTAGAGTCT CTATCTTCTTCGGTACGCAGACTGGAACAGCTGAGGGATTTGCTAAGGCATTATCCGAAGAAAT CAAAGCGAGATATGAAAAAGCAGCAGTCAAAGATGACTATGCTGCCGATGATGACCAGTATGAA GAGAAATTGAAGAAGGAAACTTTGGCATTTTTCTGTGTTGCTACTTATGGAGATGGAGAGCCTA CTGACAATGCTGCCAGATTTTACAAATGGTTTACGGAGGAAAATGAACGGGATATAAAGCTTCA ACAACTAGCATATGGTGTGTTTGCTCTTGGTAATCGCCAATATGAACATTTTAATAAGATCGGG ATAGTTCTTGATGAAGAGTTATGTAAGAAAGGTGCAAAGCGTCTTATTGAAGTCGGTCTAGGAG ATGATGATCAGAGCATTGAGGATGATTTTAATGCCTGGAAAGAATCACTATGGTCTGAGCTAGA CAAGCTCCTCAAAGACGAGGATGATAAAAGTGTGGCAACTCCTTATACAGCTGTTATTCCTGAA TACCGGGTGGTGACTCATGATCCTCGGTTTACAACTCAAAAATCAATGGAATCAAATGTGGCCA ATGGAAATACTACTATTGACATTCATCATCCCTGCAGAGTTGATGTTGCTGTGCAGAAGGAGCT TCACACACATGAATCTGATCGGTCTTGCATTCATCTCGAGTTCGACATATCCAGGACGGGTATT ACATATGAAACAGGTGACCATGTAGGTGTATATGCTGAAAATCATGTTGAAATAGTTGAAGAAG CTGGAAAATTGCTTGGCCACTCTTTAGATTTAGTATTTTCCATACATGCTGACAAGGAAGATGG CTCCCCATTGGAAAGCGCAGTGCCGCCTCCTTTCCCTGGTCCATGCACACTTGGGACTGGTTTG GCAAGATACGCAGACCTTTTGAACCCTCCTCGAAAGTCTGCGTTAGTTGCCTTGGCGGCCTATG CCACTGAACCAAGTGAAGCCGAGAAACTTAAGCACCTGACATCACCTGATGGAAAGGATGAGTA CTCACAATGGATTGTTGCAAGTCAGAGAAGTCTTTTAGAGGTGATGGCTGCTTTTCCATCTGCA AAACCCCCACTAGGTGTATTTTTTGCTGCAATAGCTCCTCGTCTACAACCTCGTTACTACTCCA TCTCATCCTCGCCAAGATTGGCGCCAAGTAGAGTTCATGTTACATCCGCACTAGTATATGGTCC AACTCCTACTGGTAGAATCCACAAGGGTGTGTGTTCTACGTGGATGAAGAATGCAGTTCCTGCG GAGAAAAGTCATGAATGTAGTGGAGCCCCAATCTTTATTCGAGCATCTAATTTCAAGTTACCAT CCAACCCTTCAACTCCAATCGTTATGGTGGGACCTGGGACTGGGCTGGCACCTTTTAGAGGTTT TCTGCAGGAAAGGATGGCACTAAAAGAAGATGGAGAAGAACTAGGTTCATCTTTGCTCTTCTTT GGGTGTAGAAATCGACAGATGGACTTTATATACGAGGATGAGCTCAATAATTTTGTTGATCAAG GCGTAATATCTGAGCTCATCATGGCATTCTCCCGTGAAGGAGCTCAGAAGGAGTATGTTCAACA TAAGATGATGGAGAAGGCAGCACAAGTTTGGGATCTAATAAAGGAAGAAGGATATCTCTATGTA TGCGGTGATGCTAAGGGCATGGCGAGGGACGTCCACCGAACTCTACACACCATTGTTCAGGAGC AGGAAGGTGTGAGTTCGTCAGAGGCAGAGGCTATAGTTAAGAAACTTCAAACCGAAGGAAGATA CCTCAGAGATGTCTGGTGA	318	
cucurbitadienol synthase [*S. grosvernorii*] Seq 59, SgCbQ protein	MWRLKVGAESVGENDEKWLKSISNHLGRQVWEFCPDAGTQQQLLQVHKARKAFHDDRFHRKQSS DLFITIQYGKEVENGGKTAGVKLKEGEEVRKEAVESSLERALSFYSSIQTSDGNWASDLGGPMF LLPGLVIALYVTGVLNSVLSKHHRQEMCRYVYNHQNEDGGWGLHIEGPSTMFGSALNYVALRLL GEDANAGAMPKARAWILDHGGATGITSWGKLWLSVLGVYEWSGNNPLPPEFWLFPYFLPFHPGR MWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYAVPYHEIDWNKSRNTCAKEDLYYPHPKM	319	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	QDILWGSLHHVYEPLFTRWPAKRLREKALQTAMQHIHYEDENTRYICLGPVNKVLNLLCCWVED PYSDAFKLHLQRVHDYLWVAEDGMKMQGYNGSQLWDTAFSIQAIVSTKLVDNYGPTLRKAHDFV KSSQIQQDCPGDPNVWYRHIHKGAWPFSTRDHGWLISDCTAEGLKAALMLSKLPSETVGESLER NRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTSATMEALTLF KKLHPGHRTKEIDTAIVRAANFLENMQRTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCLA IRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAERDPTPLH RAARLLINSQLENGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE		
cucurbitadienol synthase SgCbQ gene sequence	ATGTGGAGGTTAAAGGTCGGAGCAGAAAGCGTTGGGGAGAATGATGAGAAATGGTTGAAGAGCA TAAGCAATCACTTGGGACGCCAGGTGTGGGAGTTCTGTCCGGATGCCGGCACCCAACAACAGCT CTTGCAAGTCCACAAAGCTCGTAAAGCTTTCCACGATGACCGTTTCCACCGAAAGCAATCTTCC GATCTCTTTATCACTATTCAGTATGGAAAGGAAGTAGAAAATGGTGGAAAGACAGCGGGAGTGA AATTGAAAGAAGGGGAAGAGGTGAGGAAAGAGGCAGTAGAGAGTAGCTTAGAGAGGGCATTAAG TTTCTACTCAAGCATCCAGACAAGCGATGGGAACTGGGCTTCGGATCTTGGGGGGCCCATGTTT TTACTTCCGGGTCTGGTGATTGCCCTCTACGTTACAGGCGTCTTGAATTCTGTTTTATCCAAGC ACCACCGGCAAGAGATGTGCAGATATGTTTACAATCACCAGAATGAAGATGGGGGGTGGGGTCT CCACATCGAGGGCCCAAGCACCATGTTTGGTTCCGCACTGAATTATGTTGCACTCAGGCTGCTT GGAGAAGACGCCAACGCCGGGGCAATGCCAAAAGCACGTGCTTGGATCTTGGACCACGGTGGCG CCACCGGAATCACTTCCTGGGGCAAATTGTGGCTTTCTGTACTTGGAGTCTACGAATGGAGTGG CAATAATCCTCTTCCACCCGAATTTTGGTTATTTCCTTACTTCCTACCATTTCATCCAGGAAGA ATGTGGTGCCATTGTCGAATGGTTTATCTACCAATGTCATACTTATATGGAAAGAGATTTGTTG GGCCAATCACACCCATAGTTCTGTCTCTCAGAAAAGAACTCTACGCAGTTCCATATCATGAAAT AGACTGGAATAAATCTCGCAATACATGTGCAAAGGAGGATCTGTACTATCCACATCCCAAGATG CAAGATATTCTGTGGGGATCTCTCCACCACGTGTATGAGCCCTTGTTTACTCGTTGGCCTGCCA AACGCCTGAGAGAAAAGGCTTTGCAGACTGCAATGCAACATATTCACTATGAAGATGAGAATAC CCGATATATATGCCTTGGCCCTGTCAACAAGGTACTCAATCTGCTTTGTTGTTGGGTTGAAGAT CCCTACTCCGACGCCTTCAAACTTCATCTTCAACGAGTCCATGACTATCTCTGGGTTGCTGAAG ATGGCATGAAAATGCAGGGTTATAATGGGAGCCAGTTGTGGGACACTGCTTTCTCCATCCAAGC AATCGTATCCACCAAACTTGTAGACAACTATGGCCCAACCTTAAGAAAGGCACACGACTTCGTT AAAAGTTCTCAGATTCAGCAGGACTGTCCTGGGGATCCTAATGTTTGGTACCGTCACATTCATA AAGGTGCATGGCCATTTTCAACTCGAGATCATGGATGGCTCATCTCTGACTGTACAGCAGAGGG ATTAAAGGCTGCTTTGATGTTATCCAAACTTCCATCCGAAACAGTTGGGGAATCATTAGAACGG AATCGCCTTTGCGATGCTGTAAACGTTCTCCTTTCTTTGCAAAACGATAATGGTGGCTTTGCAT CATATGAGTTGACAAGATCATACCCTTGGTTGGAGTTGATCAACCCCGCAGAAACGTTTGGAGA TATTGTCATTGATTATCCGTATGTGGAGTGCACCTCAGCCACAATGGAAGCACTGACGTTGTTT AAGAAATTACATCCCGGCCATAGGACCAAAGAAATTGATACTGCTATTGTCAGGGCGGCCAACT TCCTTGAAAATATGCAAAGGACGGATGGCTCTTGGTATGGATGTTGGGGGGTTTGCTTCACGTA TGCGGGGTGGTTTGGCATAAAGGGATTGGTGGCTGCAGGAAGGACATATAATAATTGCCTTGCC ATTCGCAAGGCTTGCGATTTTTTACTATCTAAAGAGCTGCCCGGCGGTGGATGGGGAGAGAGTT	320	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ACCTTTCATGTCAGAATAAGGTATACACAAATCTTGAAGGAAACAGACCGCACCTGGTTAACAC GGCCTGGGTTTTAATGGCCCTCATAGAAGCTGGCCAGGCTGAGAGAGACCCAACACCATTGCAT CGTGCAGCAAGGTTGTTAATCAATTCCCAGTTGGAGAATGGTGATTTCCCCCAACAGGAGATCA TGGGAGTCTTTAATAAAAATTGCATGATCACATATGCTGCATACCGAAACATTTTTCCCATTTG GGCTCTTGGAGAGTATTGCCATCGGGTTTTGACTGAATAA		
cucurbitadienol synthase Cpep2 protein	MWRLKVGAESVGEKDEKWVKSVSNHLGRQVWEFCAADAAAVTPHQLLQIQNARNHFHRNRFHRK QSSDLFLAIQYEKEIAKGGKGKEAVKVKEGEEVGKEAVKSTLERALSFYTAVQTSDGNWASDLG GPMFLLPGLVIALYVTGVLNSVLSKHHRVEMCRYIYNHQNEDGGWGLHIEGTSTMFGSALNYVA LRLLGEDADGGDDGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYS LPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTVPYHEIDWNKSRNTCAKEDL YYPHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQTAMKHIHYEDENSRYICLGPVNKVLNM LCCWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTL RKAHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSTM VGEPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTAAT MEALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGR TYNSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGE RDPAPLHRGARLVMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE.	321	
cucurbitadienol synthase Cpep2 gene sequence	ATGTGGAGGCTGAAGGTGGGAGCAGAGAGCGTTGGGGAGAAGGATGAGAAATGGGTGAAGAGCG TAAGCAATCACTTGGGCCGCCAAGTTTGGGAGTTCTGTGCCGCCGACGCCGCCGCCGTCACTCC TCACCAGTTACTACAAATTCAGAATGCTCGCAACCACTTCCATCGCAATCGTTTCCACCGGAAG CAGTCTTCCGATCTCTTTCTCGCTATTCAGTATGAAAAGGAAATAGCGAAGGGCGGAAAAGGGA AAGAGGCGGTGAAAGTGAAAGAAGGGGAGGAGGTGGGGAAAGAGGCGGTGAAGAGTACGTTAGA GAGGGCACTAAGTTTCTACACAGCCGTGCAGACGAGCGATGGGAATTGGGCCTCGGATCTTGGA GGGCCCATGTTTTTACTTCCGGGTCTCGTGATTGCCCTTTATGTCACAGGCGTGTTGAATTCAG TTTTGTCCAAGCACCACCGCGTAGAGATGTGCAGATATATTTACAATCACCAGAATGAAGATGG AGGGTGGGGTCTACATATTGAGGGCACAAGCACCATGTTTGGTTCGGCACTCAATTATGTTGCA CTTAGGCTGCTTGGAGAAGACGCCGATGGCGGAGACGATGGTGCAATGACAAAAGCACGTGCTT GGATCTTGGAGCGCGGCGGCGCCACTGCGATCACTTCGTGGGGAAAATTGTGGCTGTCCGTGCT TGGAGTGTACGAATGGAGTGGCAACAACCCTCTTCCGCCTGAGTTTTGGCTTCTCCCTTACAGC CTACCATTTCATCCAGGACGAATGTGGTGCCATTGTCGAATGGTTTATCTTCCCATGTCTTACT TATATGGGAAGAGATTTGTTGGCCCAATCACTCCCAAAGTTCTTTCTCTAAGACAAGAGCTCTA CACGGTTCCTTATCATGAAATAGACTGGAATAAATCCCGCAATACATGTGCAAAGGAGGATCTA TACTATCCACATCCCAAGATGCAAGACATACTATGGGGATCTATCTACCATGTATATGAGCCAT TGTTCACTCGTTGGCCTGGGAAACGCCTGAGGGAAAAGGCTTTACAAACTGCAATGAAACATAT TCACTATGAAGATGAAAATAGTCGCTATATATGTCTTGGCCCAGTCAACAAGGTACTCAACATG CTTTGTTGTTGGGTTGAAGATCCCTACTCAGACGCCTTCAAACTTCACCTTCAACGCGTCCATG ACTATCTCTGGGTTGCTGAAGATGGCATGAGAATGCAGGGTTACAATGGCAGCCAGTTGTGGGA CACTGCTTTCTCCATCCAAGCCATTGTAGCTACCAAACTTGTAGACAGCTATGCCCCAACTTTA	322	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AGAAAAGCACATGACTTTGTTAAGGATTCTCAGATCCAGGAGGACTGTCCTGGGGATCCTAATG TTTGGTTCCGTCATATTCATAAAGGTGCTTGGCCATTTTCGACTCGAGATCATGGATGGCTCAT CTCTGACTGCACGGCTGAGGGATTGAAGGCTTCTTTGATGTTATCCAAACTTCCATCCACAATG GTTGGGGAGCCATTAGAAAAGAATCGCCTTTGTGATGCTGTTAATGTTCTCCTTTCTTTGCAAA ATGATAACGGTGGATTTGCATCATACGAGTTGACGAGATCATACCCTTGGTTGGAGTTGATCAA CCCAGCAGAAACATTCGGAGACATTGTCATCGACTATCCGTATGTGGAGTGCACCGCAGCAACA ATGGAAGCACTGACGTTATTTAAGAAGCTACATCCAGGCCATAGGACCAAAGAGATTGACACAG CTATTGGCAAGGCAGCCAACTTCCTTGAGAAAATGCAAAGGGCGGATGGCTCTTGGTATGGGTG TTGGGGGGTTTGTTTCACGTATGCGGGGTGGTTTGGCATCAAGGGATTGGTGGCTGCAGGAAGA ACATATAATAGCTGCCTTGCCATCCGCAAGGCTTGTGAGTTTCTGCTATCTAAAGAGCTGCCCG GCGGTGGATGGGGGGAGAGTTACCTTTCATGTCAGAATAAGGTGTACACCAATCTTGAGGGAAA CAAGCCACACTTGGTTAACACTGCCTGGGTTTTAATGGCTCTCATTGAAGCCGGCCAGGGTGAG AGAGACCCAGCACCATTGCACCGTGGAGCAAGGTTGGTAATGAATTCTCAACTGGAGAATGGTG ATTTCGTGCAACAGGAGATCATGGGAGTGTTCAATAAGAACTGCATGATCACATATGCTGCATA CCGAAACATCTTCCCCATTTGGGCGCTTGGAGAGTATTGCCATCGGGTTCTTACTGAATGA		
cucurbitadienol synthase Cpep4 protein	MWRLKVGAESVGEKDEKWVKSVSNHLGRQVWEFCAADAAAVTPHQLLQIQNARNHFHRNRFHRK QSSDLFLAIQYEKEIAKGGKGKEAVKVKEGEEVGKEAVKSTLERALSFYTAVQTSDGNWASDLG GPMFLLPGLVIALYVTGVLNSVLSKHHRVEMCRYIYNHQNEDGGWGLHIEGTSTMFGSALNYVA LRLLGEDADGGDDGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFLLLPYS LPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTVPYHEIDWNKSRNTCAKEDL YYPHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQTAMKHIHYEDENSRYICLGPVNKVLNM LCCWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTL RKAHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSTM VGEPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYSYVECTAAT MEALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGR TYNSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGE RDPAPLHRAARLVMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE.	323	
cucurbitadienol synthase Cpep4 gene sequence	ATGTGGAGGCTGAAGGTGGGAGCAGAGAGCGTTGGGGAGAAGGATGAGAAATGGGTGAAGAGCG TAAGCAATCACTTGGGCCGCCAAGTTTGGGAGTTCTGTGCCGCCGACGCCGCCGCCGTCACTCC TCACCAGTTACTACAAATTCAGAATGCTCGCAACCACTTCCATCGCAATCGTTTCCACCGGAAG CAGTCTTCCGATCTCTTTCTCGCTATTCAGTATGAAAAGGAAATAGCGAAGGGCGGAAAAGGGA AAGAGGCGGTGAAAGTGAAAGAAGGGGAGGAGGTGGGGAAAGAGGCGGTGAAGAGTACGTTAGA GAGGGCACTAAGTTTCTACACAGCCGTGCAGACGAGCGATGGGAATTGGGCCTCGGATCTTGGA GGGCCCATGTTTTTACTTCCGGGTCTCGTGATTGCCCTTTATGTCACAGGCGTGTTGAATTCAG TTTTGTCCAAGCACCACCGCGTAGAGATGTGCAGATATATTTACAATCACCAGAATGAAGATGG AGGGTGGGGTCTACATATTGAGGGCACAAGCACCATGTTTGGTTCGGCACTCAATTATGTTGCA CTTAGGCTGCTTGGAGAAGACGCCGATGGCGGAGACGATGGTGCAATGACAAAAGCACGTGCTT GGATCTTGGAGCGCGGCGGCGCCACTGCGATCACTTCGTGGGGAAAATTGTGGCTGTCCGTGCT	324	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TGGAGTGTACGAATGGAGTGGCAACAACCCTCTTCCGCCTGAGTTTTTGCTTCTCCCTTACAGC CTACCATTTCATCCAGGACGAATGTGGTGCCATTGTCGAATGGTTTATCTTCCCATGTCTTACT TATATGGGAAGAGATCTGTTCGCCCAATCACTCCCAAAGTTCTTTCTCTAAGACAAGAGCTCTA CACGGTTCCTTATCATGAAATAGACTGGAATAAATCCCGCAATACATGTGCAAAGGAGGATCTA TACTATCCACATCCCAAGATGCAAGACATACTATGGGGATCTATCTACCATGTATATGAGCCAT TGTTCACTCGTTGGCCTGGGAAACGCCTGAGGGAAAAGGCTTTACAAACTGCAATGAAACATAT TCACTATGAAGATGAAAATAGTCGCTATATATGTCTTGGCCCAGTCAACAAGGTACTCAACATG CTTTGTTGTTGGGTTGAAGATCCCTACTCAGACGCCTTCAAACTTCACCTTCAACGCGTCCATG ACTATCTCTGGGTTGCTGAAGATGGCATGAGAATGCAGGGTTACAATGGCAGCCAGTTGTGGGA CACTGCTTTCTCCATCCAAGCCATTGTAGCTACCAAACTTGTAGACAGCTATGCCCCAACTTTA AGAAAAGCACATGACTTTGTTAAGGATTCTCAGATCCAGGAGGACTGTCCTGGGGATCCTAATG TTTGGTTCCGTCATATTCATAAAGGTGCTTGGCCATTTTCGACTCGAGATCATGGATGGCTCAT CTCTGACTGCACGGCTGAGGGATTGAAGGCTTCTTTGATGTTATCCAAACTTCCATCCACAATG GTTGGGGAGCCATTAGAAAAGAATCGCCTTTGTGATGCTGTTAATGTTCTCCTTTCTTTGCAAA ATGATAACGGCGGATTTGCATCATACGAGTTGACGAGATCATACCCTTGGTTGGAGTTGATCAA CCCAGCAGAAACATTCGGAGACATTGTCATCGACTATTCGTATGTGGAGTGCACCGCAGCAACA ATGGAAGCACTGACGTTATTTAAGAAGCTACATCCAGGCCATAGGACCAAAGAGATTGACACAG CTATTGGCAAGGCAGCCAACTTCCTTGAGAAAATGCAAAGGGCGGATGGCTCTTGGTATGGGTG TTGGGGGGTTTGTTTCACGTATGCGGGGTGGTTTGGCATAAAGGGATTGGTGGCTGCAGGAAGA ACATATAATAGCTGTCTTGCCATCCGCAAGGCTTGTGAGTTTCTGCTATCTAAAGAGCTGCCCG GCGGTGGATGGGGGGAGAGTTACCTTTCATGTCAGAATAAGGTGTACACCAATCTTGAGGGAAA CAAGCCACACTTGGTTAACACTGCCTGGGTTTTAATGGCTCTCATTGAAGCTGGCCAGGGTGAG AGAGACCCAGCACCATTGCACCGTGCAGCAAGGTTGGTAATGAATTCTCAACTGGAGAATGGCG ATTTCGTGCAACAGGAGATCATGGGAGTGTTCAATAAGAACTGCATGATCACATATGCTGCATA CCGAAACATCTTCCCCATTTGGGCGCTTGGAGAGTATTGCCATCGGGTTCTTACTGAATGA		
cucurbitadienol synthase Cmax1 protein	MWRLKVGAESVGEKDEKWVKSVSNHLGRQVWEFCADAAADTPHQLLQIQNARNHFHHNRFHRKQ SSDLFLAIQYEKEIAKGAKGGAVKVKEGEEVGKEAVKSTLESALGFYSAVQTSDGNWASDLGGP MFLLPGLVIALHVTGVLNSVLSKHHRVEMCRYLYNHQNEDGGWGLHIEGTSTMFGSALNYVALR LLGEDADGGDGGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLP FHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTIPYHEIDWNKSRNTCAKEDLYY PHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQAAMKHIHYEDENSRYICLGPVNKVLNMLC CWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTLRK AHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSAMVG EPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTAATME ALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGRTY NSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGERD PAPLHRAARLLMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE	325	
cucurbitadienol	ATGTGGAGGCTGAAGGTGGGAGCAGAGAGCGTTGGGGAGGAGGATGAGAAATGGGTGAAGAGCG	326	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
synthase gene	TAAGCAATCACTTGGGCCGCCAAGTTTGGGAGTTCTGTGCCGACGCCGCCGCCGACACTCCTCA CCAGTTACTACAAATTCAGAATGCTCGCAACCACTTCCATCACAATCGTTTCCACCGGAAGCAG TCTTCCGATCTCTTTCTGGCTATTCAATATGAAAAGGAAATAGCAAAGGGCGCAAAAGGTGGAG CGGTGAAAGTGAAAGAAGGGGAGGAGGTGGGGAAAGAGGCGGTGAAGAGTACGTTAGAAAGGGC ACTCGGTTTCTACTCGGCCGTGCAGACAAGAGATGGGAATTGGGCCTCGGATCTTGGAGGGCCC TTGTTTTTACTTCCGGGTCTCGTGATTGCCCTTCATGTCACAGGCGTCTTGAATTCAGTTTTGT CCAAGCACCACCGCGTAGAGATGTGCAGATATCTTTACAATCACCAGAATGAAGATGGAGGGTG GGGTCTACATATTGAGGGCACAAGCACCATGTTTGGTTCGGCACTGAATTACGTTGCACTAAGG CTGCTTGGAGAAGACGCCGATGGCGGAGACGGTGGCGCAATGACAAAAGCACGTGCTTGGATCT TGGAGCGCGGCGGCGCCACTGCGATCACTTCGTGGGGAAAATTGTGGCTGTCCGTACTTGGAGT GTACGAATGGAGTGGCAACAACCCTCTTCCGCCTGAGTTTTGGCTTCTCCCTTACAGCCTACCA TTTCATCCAGGAAGAATGTGGTGCCATTGTCGAATGGTTTATCTTCCAATGTCTTACTTATATG GGAAGAGATTTGTTGGGCCAATCACTCCCAAAGTTCTTTCTCTAAGGCAAGAGCTCTACACAAT TCCTTATCATGAAATAGACTGGAATAAATCCCGCAATACATGTGCAAAGGAGGATCTGTACTAT CCACATCCCAAGATGCAAGACATTCTATGGGGATCCATCTACCATGTATATGAGCCATTGTTCA CTCGTTGGCCTGGGAAACGCCTGAGGGAAAAGGCTTTACAAGCTGCAATGAAACATATTCACTA TGAAGATGAAAATAGTCGATATATATGTCTTGGCCCAGTCAACAAGGTACTCAACATGCTTTGT TGTTGGGTTGAAGATCCCTACTCAGACGCCTTCAAACTTCACCTTCAACGCGTCCATGACTATC TCTGGGTTGCTGAAGATGGCATGAGAATGCAGGGCTACAATGGCAGCCAGTTGTGGGACACTGC TTTCTCCATCCAAGCCATCGTAGCCACCAAACTTGTAGACAGCTATGCCCCAACTTTAAGAAAA GCACATGACTTTGTTAAGGATTCTCAGATCCAGGAGGACTGTCCTGGGGATCCTAATGTTTGGT TCCGTCATATTCATAAAGGTGCTTGGCCACTTTCGACACGAGATCATGGATGGCTCATCTCCGA CTGTACAGCTGAGGGATTGAAGGCTTCTTTGATGTTATCCAAACTTCCATCCACAATGGTTGGG GAGCCATTAGAAAAGAATCGCCTTTGTGATGCTGTTAATGTTCTCCTTTCTTTGCAAAATGATA ATGGTGGATTTGCATCATACGAGTTGACGAGATCATACCCTTGGTTGGAGTTGATCAACCCAGC TGAAACATTCGGAGACATTGTCATTGACTATCCGTATGTGGAGTGCACCGCAGCAACAATGGAA GCACTGACGTTATTTAAGAAGCTACATCCAGGCCATAGGACCAAAGAGATTGACACAGCTATTG GCAAGGCAGCCAACTTCCTTGAGAAAATGCAGAGGGCGGATGGCTCTTGGTACGGGTGTTGGGG GGTTTGTTTTACGTATGCGGGTTGGTTTGGCATAAAGGGATTGGTGGCTGCAGGAAGAACATAT AATAGCTGCCTTGCCATTCGCAAGGCTTGTGAGTTTCTGCTATCTAAAGAGCTGCCCGGCGGTG GATGGGGGGAGAGTTACCTTTCATGTCAGAATAAGGTGTACACCAATCTTGAGGGGAACAAGCC ACACTTGGTTAACACTGCCTGGGTTTTAATGGCTCTCATTGAAGCTGGCCAGGGTGAGAGAGAC CCAGCACCATTGCACCGTGCAGCAAGGTTGCTAATGAATTCCCAATTGGAGAATGGCGATTTCG TGCAACAGGAGATCATGGGAGTGTTCAATAAGAACTGCATGATCACATATGCTGCATACCGAAA CATCTTCCCCATTTGGGCGCTTGGAGAGTATTGCCATCGGGTTCTTACTGAATGA		
cucurbitadienol synthase Cmos1 protein	MWRLKVGAESVGEKDEKWVKSVSNHLGRQVWEFCADAAAAATPRQLLQIQNARNHFHRNRFHRK QSSDLFLAIQYEKEIAEGGKGGAVKVKEEEEVGKEAVKSTLERALSFYSAVQTSDGNWASDLGG PMFLLPGLVIALYVTGVLNSVLSKHHRVEMCRYLYNHQNEDGGWGLHIEGTSTMFGSALNYVAL	327	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	RLLGEDADGGDDGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSL PFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTVPYHEIDWNKSRNTCAKEDLY YPHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQTAMKHIHYEDENSRYICLGPVNKVLNML CCWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTLR KAHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSAMV GEPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTAATM EALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGRT YNSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGER DPAPLHRAARLLMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE		
cucurbitadienol synthase Cmos1 gene sequence	ATGTGGAGGTTGAAGGTGGGAGCAGAGAGCGTTGGGGAGAAGGATGAGAAATGGGTGAAGAGCG TAAGCAATCACTTGGGCCGCCAAGTTTGGGAGTTCTGTGCCGACGCCGCCGCCGCCGCCACTCC TCGCCAGTTACTACAAATTCAGAATGCTCGCAACCACTTCCATCGCAATCGTTTCCACCGGAAG CAGTCTTCCGATCTCTTTCTCGCTATTCAGTATGAAAAGGAAATAGCAGAGGGCGGAAAAGGTG GAGCGGTGAAAGTGAAAGAAGAGGAGGAGGTGGGGAAAGAGGCGGTGAAGAGTACGTTAGAAAG GGCACTAAGTTTCTACTCAGCCGTGCAGACAAGCGATGGGAATTGGGCCTCGGATCTTGGAGGG CCCATGTTTTTACTTCCGGGTCTCGTGATTGCCCTTTATGTCACAGGCGTGTTGAATTCAGTTT TGTCCAAGCACCACCGCGTAGAGATGTGCAGATATCTTTACAATCACCAGAATGAAGATGGAGG GTGGGGTCTACATATTGAGGGCACAAGCACCATGTTTGGTTCGGCACTCAATTACGTTGCACTA AGGCTGCTTGGAGAAGACGCGGATGGCGGAGACGATGGCGCAATGACAAAAGCACGTGCTTGGA TCTTGGAGCGCGGCGGCGCCACTGCGATCACTTCGTGGGGAAAGTTGTGGCTGTCCGTGCTTGG AGTGTACGAATGGAGTGGCAACAACCCTCTTCCGCCTGAGTTTTGGCTTCTCCCTTACAGCCTA CCATTTCATCCAGGAAGAATGTGGTGCCATTGTCGAATGGTTTATCTTCCCATGTCTTACTTAT ATGGGAAGAGATTTGTTGGGCCAATCACTCCCAAAGTTCTATCGCTAAGACAAGAGCTTTACAC GGTTCCTTATCATGAAATAGACTGGAACAAATCCCGCAATACATGTGCAAAGGAGGATCTATAC TATCCACATCCCAAGATGCAAGACATTCTATGGGGATCCATCTACCATGTGTATGAGCCATTGT TCACTCGTTGGCCTGGGAAACGCCTGAGGGAAAAGGCTTTACAAACTGCAATGAAACATATTCA CTATGAAGATGAAAATAGTCGATATATATGTCTTGGCCCAGTCAACAAGGTACTCAACATGCTT TGTTGTTGGGTTGAAGATCCCTACTCAGACGCCTTCAAACTTCACCTTCAACGCGTCCATGACT ATCTCTGGGTTGCTGAAGATGGCATGAGAATGCAGGGCTACAATGGCAGCCAGTTGTGGGACAC TGCTTTCTCCATCCAAGCCATCGTAGCCACCAAACTTGTAGACAGCTATGCCCCAACTTTAAGA AAAGCACATGACTTTGTTAAGGATTCTCAGATCCAGGAGGACTGTCCTGGGGATCCTAATGTTT GGTTCCGTCATATTCATAAAGGTGCTTGGCCATTTTCGACTCGAGATCATGGATGGCTCATCTC CGACTGTACAGCTGAGGGATTGAAGGCTTCTTTGATGTTATCCAAACTTCCATCCGCAATGGTT GGGGAGCCATTAGAAAAGAATCGCCTTTGTGATGCTGTTAATGTTCTCCTTTCTTTGCAAAATG ATAATGGTGGATTTGCATCATACGAGTTGACGAGATCATACCCTTGGTTGGAGTTGATCAACCC AGCAGAAACATTCGGAGACATTGTCATCGACTATCCGTATGTGGAGTGCACCGCAGCAACAATG GAAGCACTGACGTTATTTAAGAAGCTACATCCAGGCCATAGGACCAAAGAGATTGACACAGCTA TTGGCAAGGCAGCCAACTTCCTTGAGAAAATGCAGAGGGCGGATGGCTCTTGGTATGGGTGTTG	328	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GGGGGTTTGTTTCACGTATGCGGGGTGGTTTGGCATAAAGGGATTGGTGGCTGCAGGAAGAACA TATAATAGCTGCCTTGCCATCCGCAAGGCTTGTGAGTTTCTGCTATCTAAAGAGCTGCCCGGCG GTGGATGGGGGGAGAGTTACCTTTCATGTCAGAATAAGGTGTACACCAATCTTGAGGGAAACAA GCCACACTTGGTTAACACTGCCTGGGTTTTAATGGCTCTCATTGAAGCTGGCCAGGGTGAGAGA GACCCAGCACCATTGCACCGTGCAGCAAGGTTGCTAATGAATTCCCAATTGGAGAATGGCGATT TCGTGCAACAGGAGATCATGGGAGTGTTCAATAAGAACTGCATGATCACATATGCTGCATACCG AAACATCTTCCCCATTTGGGCGCTTGGAGAGTATTGCCATCGGGTTCTGACTGAAT		
cucurbitadienol synthase [*Cucumis melo*]	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDM.	329	
cucurbitadienol synthase [*Citrullus colocynthis*]	MWRLKVGAESVGEKEEKWLKSISNHLGRQVWEFCADQPTASPNHLQQIDNARKHFRNNRFHRKQ SSDLFLAIQNEKEIANGTKGGGIKVKEEEDVRKETVKNTVERALSFYSAIQTNDGNWASDLGGP MFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYLYNHQNEDGGWGLHIEGTSTMFGSALNYVALR LLGEDADGGEGGAMTKARGWILDRGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYCLP FHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNKSRNTCAKEDLYY PHPKMQDILWGSIYHLYEPLFTRWPGKRLREKALQMAMKHIHYEDENSRYICLGPVNKVLNMLC CWVEDPYSDAFKFHLQRVPDYLWIAEDGMRMQGYNGSQLWDTAFSVQAIISTKLIDSFGTTLKK AHDFVKDSQIQQDFPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSKIVG EPLEKSRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTSATME ALTLFKKLHPGHRTKEIDTAVAKAANFLENMQRTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTY STCVAIRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAERD PAPLHRAARLLINSQLENGDFPQEEIMGVFNKNCMITYAAYRNIFPIWALGEYFHRVLTE.	330	
cucurbitadienol synthase [*Cucurbita pepo*]	MWRLKVGAESVGEEDEKWVKSVSNHLGRQVWEFCADAAADTPHQLLQIQNARNHFHHNRFHRKQ SSDLFLAIQYEKEIAKGAKGGAVKVKEGEEVGKEAVKSTLERALGFYSAVQTRDGNWASDLGGP LFLLPGLVIALHVTGVLNSVLSKHHRVEMCRYLYNHQNEDGGWGLHIEGTSTMFGSALNYVALR LLGEDADGGDGGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLP FHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTIPYHEIDWNKSRNTCAKEDLYY PHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQAAMKHIHYEDENSRYICLGPVNKVLNMLC	331	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTLRK AHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPLSTRDHGWLISDCTAEGLKASLMLSKLPSTMVG EPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTAATME ALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGRTY NSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGERD PAPLHRAARLLMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE		
cucurbitadienol synthase [*Cucumis sativa*]	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCAENDDDDDDEAVIHVVANSSKHLLQQQRRQ SSFENARKQFRNNRFHRKQSSDLFLTIQYEKEIARNGAKNGGNTKVKEGEDVKKEAVNNTLERA LSFYSAIQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGW GLHIEGSSTMFGSALNYVALRLLGEDANGGECGAMTKARSWILERGGATAITSWGKLWLSVLGV YEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITHMVLSLRKELYTI PYHEIDWNRSRNTCAQEDLYYPHPKMQDILWGSIYHVYEPLFNGWPGRRLREKAMKIAMEHIHY EDENSRYIYLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTA FSIQAILSTKLIDTFGSTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISD CTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPA ETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAALAKAANFLENMQRTDGSWYGCWG VCFTYAGWFGIKGLVAAGRTYNNCVAIRKACHFLLSKELPGGGWGESYLSCQNKVYTNLEGNRP HLVNTAWVLMALIEAGQGERDPAPLHRAARLLINSQLENGDFPQQEIMGVFNKNCMITYAAYRN IFPIWALGEYSHRVLTE	332	
cucurbitadienol synthase [*Citrullus lanatus*] (partial)	DGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYLYNHQNEDGGWGLHIEGTSTM FGSALNYVALRLLGEDADGGEGGAMTKARSWILDRGGATAITSWGKLWLSVLGVYEWSGNNPLP PEFWLLPYCLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRS RNTCAKEDLYYPHPKMQDILWGSIYHLYEPLFTRWPGKRLREKALQMAMKHIHYEDENSRYICL GPVNKVLNMLCCWVEDPYSDAFKFHLQRVPDYLWVAEDGMRMQGYNGSQLWDTAFSVQAIISTK LIDSFGTTLKKAHDFVKDSQIQQDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASL MLSKLPSEIVGEPLEKSRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDY PYVECTSATMEALTLFKKLHPGRRTKEIDIAVARAANFLENMQRTDGSWYGCWGVCFTYAGWFG IKGLVAAGRTYNSCVAIRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLM ALIEAGQAERDPAPLHRAARLLINSQLENGDFPQEEIMGVFNKNCMITYAAYRNIFPIWALGEY FHRVLTE	333	
Squalene epoxidase/ squalene monooxidase	MSAVNVAPELINADNTITYDAIVIGAGVIGPCVATGLARKGKKVLIVERDWAMPDRIVGELMQP GGVRALRSLGMIQSINNIEAYPVTGYTVFFNGEQVDIPYPYKADIPKVEKLKDLVKDGNDKVLE DSTIHIKDYEDDERERGVAFVHGRFLNNLRNITAQEPNVTRVQGNCIEILKDEKNEVVGAKVDI DGRGKVEFKAHLTFICDGIFSRFRKELHPDHVPTVGSSFVGMSLFNAKNPAPMHGHVILGSDHM PILVYQISPEETRILCAYNSPKVPADIKSWMIKDVQPFIPKSLRPSFDEAVSQGKFRAMPNSYL PARQNDVTGMCVIGDALNMRHPLTGGGMTVGLHDVVLLIKKIGDLDFSDREKVLDELLDYHFER KSYDSVINVLSVALYSLFAADSDNLKALQKGCFKYFQRGGDCVNKPVEFLSGVLPKPLQLTRVF FAVAFYTIYLNMEERGFLGLPMALLEGIMILITAIRVFTPFLFGELIG	334	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Squalene epoxidase/ squalene monooxidase gene sequence	ATGTCTGCTGTTAACGTTGCACCTGAATTGATTAATGCCGACAACACAATTACCTACGATGCGA TTGTCATCGGTGCTGGTGTTATCGGTCCATGTGTTGCTACTGGTCTAGCAAGAAAGGGTAAGAA AGTTCTTATCGTAGAACGTGACTGGGCTATGCCTGATAGAATTGTTGGTGAATTGATGCAACCA GGTGGTGTTAGAGCATTGAGAAGTCTGGGTATGATTCAATCTATCAACAACATCGAAGCATATC CTGTTACCGGTTATACCGTCTTTTTCAACGGCGAACAAGTTGATATTCCATACCCTTACAAGGC CGATATCCCTAAAGTTGAAAAATTGAAGGACTTGGTCAAAGATGGTAATGACAAGGTCTTGGAA GACAGCACTATTCACATCAAGGATTACGAAGATGATGAAAGAGAAAGGGGTGTTGCTTTTGTTC ATGGTAGATTCTTGAACAACTTGAGAAACATTACTGCTCAAGAGCCAAATGTTACTAGAGTGCA AGGTAACTGTATTGAGATATTGAAGGATGAAAAGAATGAGGTTGTTGGTGCCAAGGTTGACATT GATGGCCGTGGCAAGGTGGAATTCAAAGCCCACTTGACATTTATCTGTGACGGTATCTTTTCAC GTTTCAGAAAGGAATTGCACCCAGACCATGTTCCAACTGTCGGTTCTTCGTTTGTCGGTATGTC TTTGTTCAATGCTAAGAATCCTGCTCCTATGCACGGTCACGTTATTCTTGGTAGTGATCATATG CCAATCTTGGTTTACCAAATCAGTCCAGAAGAAACAAGAATCCTTTGTGCTTACAACTCTCCAA AGGTCCCAGCTGATATCAAGAGTTGGATGATTAAGGATGTCCAACCTTTCATTCCAAAGAGTCT ACGTCCTTCATTTGATGAAGCCGTCAGCCAAGGTAAATTTAGAGCTATGCCAAACTCCTACTTG CCAGCTAGACAAAACGACGTCACTGGTATGTGTGTTATCGGTGACGCTCTAAATATGAGACATC CATTGACTGGTGGTGGTATGACTGTCGGTTTGCATGATGTTGTCTTGTTGATTAAGAAAATAGG TGACCTAGACTTCAGCGACCGTGAAAAGGTTTTGGATGAATTACTAGACTACCATTTCGAAAGA AAGAGTTACGATTCCGTTATTAACGTTTTGTCAGTGGCTTTGTATTCTTTGTTCGCTGCTGACA GCGATAACTTGAAGGCATTACAAAAAGGTTGTTTCAAATATTTCCAAAGAGGTGGCGATTGTGT CAACAAACCCGTTGAATTTCTGTCTGGTGTCTTGCCAAAGCCTTTGCAATTGACCAGGGTTTTC TTCGCTGTCGCTTTTTACACCATTTACTTGAACATGGAAGAACGTGGTTTCTTGGGATTACCAA TGGCTTTATTGGAAGGTATTATGATTTTGATCACAGCTATTAGAGTATTCACCCCATTTTTGTT TGGTGAGTTGATTGGTTAA	335	
Squalene synthase Erg9	MGKLLQLALHPVEMKAALKLKFCRTPLFSIYDQSTSPYLLHCFELLNLTSRSFAAVIRELHPEL RNCVTLFYLILRALDTIEDDMSIEHDLKIDLLRHFHEKLLLTKWSFDGNAPDVKDRAVLTDFES ILIEFHKLKPEYQEVIKEITEKMGNGMADYILDENYNLNGLQTVHDYDVYCHYVAGLVGDGLTR LIVIAKFANESLYSNEQLYESMGLFLQKTNIIRDYNEDLVDGRSFWPKEIWSQYAPQLKDFMKP ENEQLGLDCINHLVLNALSHVIDVLTYLAGIHEQSTFQFCAIPQVMAIATLALVFNNREVLHGN VKIRKGTTCYLILKSRTLRGCVEIFDYYLRDIKSKLAVQDPNFLKLNIQISKIEQFMEEMYQDK LPPNVKPNETPIFLKVKERSRYDDELVPTQQEEEYKFNMVLSIILSVLLGFYYIYTLHRA	336	
Squalene synthase Erg9 gene sequence	ATGGGAAAGCTATTACAATTGGCATTGCATCCGGTCGAGATGAAGGCAGCTTTGAAGCTGAAGT TTTGCAGAACACCGCTATTCTCCATCTATGATCAGTCCACGTCTCCATATCTCTTGCACTGTTT CGAACTGTTGAACTTGACCTCCAGATCGTTTGCTGCTGTGATCAGAGAGCTGCATCCAGAATTG AGAAACTGTGTTACTCTCTTTTATTTGATTTTAAGGGCTTTGGATACCATCGAAGACGATATGT CCATCGAACACGATTTGAAAATTGACTTGTTGCGTCACTTCCACGAGAAATTGTTGTTAACTAA ATGGAGTTTCGACGGAAATGCCCCCGATGTGAAGGACAGAGCCGTTTTGACAGATTTCGAATCG ATTCTTATTGAATTCCACAAATTGAAACCAGAATATCAAGAAGTCATCAAGGAGATCACCGAGA	337	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAATGGGTAATGGTATGGCCGACTACATCTTAGATGAAAATTACAACTTGAATGGGTTGCAAAC CGTCCACGACTACGACGTGTACTGTCACTACGTAGCTGGTTTGGTCGGTGATGGTTTGACCCGT TTGATTGTCATTGCCAAGTTTGCCAACGAATCTTTGTATTCTAATGAGCAATTGTATGAAAGCA TGGGTCTTTTCCTACAAAAAACCAACATCATCAGAGATTACAATGAAGATTTGGTCGATGGTAG ATCCTTCTGGCCCAAGGAAATCTGGTCACAATACGCTCCTCAGTTGAAGGACTTCATGAAACCT GAAAACGAACAACTGGGGTTGGACTGTATAAACCACCTCGTCTTAAACGCATTGAGTCATGTTA TCGATGTGTTGACTTATTTGGCCGGTATCCACGAGCAATCCACTTTCCAATTTTGTGCCATTCC CCAAGTTATGGCCATTGCAACCTTGGCTTTGGTATTCAACAACCGTGAAGTGCTACATGGCAAT GTAAAGATTCGTAAGGGTACTACCTGCTATTTAATTTTGAAATCAAGGACTTTGCGTGGCTGTG TCGAGATTTTTGACTATTACTTACGTGATATCAAATCTAAATTGGCTGTGCAAGATCCAAATTT CTTAAAATTGAACATTCAAATCTCCAAGATCGAACAGTTTATGGAAGAAATGTACCAGGATAAA TTACCTCCTAACGTGAAGCCAAATGAAACTCCAATTTTCTTGAAAGTTAAAGAAAGATCCAGAT ACGATGATGAATTGGTTCCAACCCAACAAGAAGAAGAGTACAAGTTCAATATGGTTTTATCTAT CATCTTGTCCGTTCTTCTTGGGTTTTATTATATATACACTTTACACAGAGCGTGA		
Farnesyl PP synthase	MASEKEIRRERFLNVFPKLVEELNASLLAYGMPKEACDWYAHSLNYNTPGGKLNRGLSVV61DT YAILSNKTVEQLGQEEYEKVAILGWCIELLQAYFLVADDMMDKSITRRGQPCWYKVPEVGEIAI NDAFMLEAAIYKLLKSHFRNEKYYIDITELFHEVTFQTELGQLMDLITAPEDKVDLSKFSLKKH SFIVTFKTAYYSFYLPVALAMYVAGITDEKDLKQARDVLIPLGEYFQIQDDYLDCFGTPEQIGK IGTDIQDNKCSWVINKALELASAEQRKTLDENYGKKDSVAEAKCKKIFNDLKIEQLYHEYEESI AKDLKAKISQVDESRGFKADVLTAFLNKVYKRSK	338	
Farnesyl PP synthase gene sequence	ATGGCTTCAGAAAAAGAAATTAGGAGAGAGAGATTCTTGAACGTTTTCCCTAAATTAGTAGAGG AATTGAACGCATCGCTTTTGGCTTACGGTATGCCTAAGGAAGCATGTGACTGGTATGCCCACTC ATTGAACTACAACACTCCAGGCGGTAAGCTAAATAGAGGTTTGTCCGTTGTGGACACGTATGCT ATTCTCTCCAACAAGACCGTTGAACAATTGGGGCAAGAAGAATACGAAAAGGTTGCCATTCTAG GTTGGTGCATTGAGTTGTTGCAGGCTTACTTCTTGGTCGCCGATGATATGATGGACAAGTCCAT TACCAGAAGAGGCCAACCATGTTGGTACAAGGTTCCTGAAGTTGGGGAAATTGCCATCAATGAC GCATTCATGTTAGAGGCTGCTATCTACAAGCTTTTGAAATCTCACTTCAGAAACGAAAAATACT ACATAGATATCACCGAATTGTTCCATGAGGTCACCTTCCAAACCGAATTGGGCCAATTGATGGA CTTAATCACTGCACCTGAAGACAAAGTCGACTTGAGTAAGTTCTCCCTAAAGAAGCACTCCTTC ATAGTTACTTTCAAGACTGCTTACTATTCTTTCTACTTGCCTGTCGCATTGGCCATGTACGTTG CCGGTATCACGGATGAAAAGGATTTGAAACAAGCCAGAGATGTCTTGATTCCATTGGGTGAATA CTTCCAAATTCAAGATGACTACTTAGACTGCTTCGGTACCCCAGAACAGATCGGTAAGATCGGT ACAGATATCCAAGATAACAAATGTTCTTGGGTAATCAACAAGGCATTGGAACTTGCTTCCGCAG AACAAAGAAAGACTTTAGACGAAAATTACGGTAAGAAGGACTCAGTCGCAGAAGCCAAATGCAA AAAGATTTTCAATGACTTGAAAATTGAACAGCTATACCACGAATATGAAGAGTCTATTGCCAAG GATTTGAAGGCCAAAATTTCTCAGGTCGATGAGTCTCGTGGCTTCAAAGCTGATGTCTTAACTG CGTTCTTGAACAAAGTTTACAAGAGAAGCAAATAG	339	
cycloartenol	MWKLKVAEGGTPWLRTLNNHVGRQVWEFDPHSGSPQDLDDIETARRNFHDNRFTHKHSDDLLMR	340	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
synthase	LQFAKENPMNEVLPKVKVKDVEDVTEEAVATTLRRGLNFYSTIQSHDGHWPGDYGGPMFLMPGL VITLSVTGALNAVLTDEHRKEMRRYLYNHQNKDGGWGLHIEGPSTMFGSVLCYVTLRLLGEGPN DGEGDMERGRDWILEHGGATYITSWGKMWLSVLGVFEWSGNNPMPPEWLLPYALPVHPGRMWCH CRMVYLPMSYLYGKRFVGPITPTVLSLRKELFTVPYHDIDWNQARNLCAKEDLYYPHPLVQDIL WATLHKFVEPVFMNWPGKKLREKAIKTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNSE AFKLHLPRIYDYLWVAEDGMKMQGYNGSQLWDTAFAAQAIISTNLIDEFGPTLKKAHAFIKNSQ VSEDCPGDLSKWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKIAPEIVGEPLDSKRLY DAVNVILSLQNENGGLATYELTRSYTWLEIINPAETFGDIVIDCPYVECTSAAIQALATFGKLY PGHRREEIQCCIEKAVAFIEKIQASDGSWYGSWGVCFTYGTWFGIKGLIAAGKNFSNCLSIRKA CEFLLSKQLPSGGWAESYLSCQNKVYSNLEGNRSHVVNTGWAMLALIEAEQAKRDPTPLHRAAV CLINSQLENGDFPQEEIMGVFNKNCMITYAAYRCIFPIWALGEYR RVLQAC		
oxidosqualene cylcases	MWKLKVAEGGTPWLRTLNNHVGRQVWEFDPHSGSPQDLDDIETARRNFHDNRFTHKHSDDLLMR LQFAKENPMNEVLPKVKVKDVEDVTEEAVATTLRRGLNFYSTIQSHDGHWPGDLGGPMFLMPGL VITLSVTGALNAVLTDEHRKEMRRYLYNHQNKDGGWGLHIEGPSTMFGSVLCYVTLRLLGEGPN DGEGDMERGRDWILEHGGATYITSWGKMWLSVLGVFEWSGNNPMPPEIWLLPYALPVHPGRMWC HCRMVYLPMSYLYGKRFVGPITPTVLSLRKELFTVPYHDIDWNQARNLCAKEDLYYPHPLVQDI LWATLHKFVEPVFMNWPGKKLREKAIKTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNS EAFKLHLPRIYDYLWVAEDGMKMQGYNGSQLWDTAFAAQAIISTNLIDEFGPTLKKAHAFIKNS QVSEDCPGDLSKWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKIAPEIVGEPLDSKRL YDAVNVILSLQNENGGLATYELTRSYTWLEIINPAETFGDIVIDCPYVECTSAAIQALATFGKL YPGHRREEIQCCIEKAVAFIEKIQASDGSWYGSWGVCFTYGTWFGIKGLIAAGKNFSNCLSIRK ACEFLLSKQLPSGGWAESYLSCQNKVYSNLEGNRSHVVNTGWAMLALIEAEQAKRDPTPLHRAA VCLINSQLENGDFPQEEIMGVFNKNCMITYAAYRCIFPIWALGEYRRVLQAC	341	PMID: 26058429 Takase el al., 2015
oxidosqualene cylcases gene sequence	ATGTGGAGGCTAACAATAGGTGAGGGCGGCGGTCCGTGGCTGAAGTCGAACAATGGCTTCCTTG GCCGCCAAGTGTGGGAGTACGACGCCGATGCCGGCACGCCGGAAGAGCGTGCCGAGGTTGAGAG GGTGCGTGCGGAATTCACAAAGAACAGGTTCCAGAGGAAGGAGTCACAGGACCTTCTTCTACGC TTGCAGTACGCAAAAGACAACCCTCTTCCGGCGAATATTCCGACAGAAGCCAAGCTTGAAAAGA GTACAGAGGTCACTCACGAGACTATCTACGAATCATTGATGCGAGCTTTACATCAATATTCCTC TCTACAAGCAGACGATGGGCATTGGCCTGGTGATTACAGTGGGATTCTCTTCATTATGCCTATC ATTATATTCTCTTTATATGTTACTAGATCACTTGACACCTTTTTATCTCCGGAACATCGTCATG AGATATGTCGCTACATTTACAATCAACAGAATGAAGATGGTGGTTGGGGAAAAATGGTTCTTGG CCCAAGTACCATGTTTGGATCGTGTATGAATTATGCAACCTTAATGATTCTTGGCGAGAAGCGA AATGGTGATCATAAGGATGCATTGGAAAAAGGGCGTTCTTGGATTTTATCTCATGGAACTGCAA CTGCAATACCACAGTGGGGAAAAATATGGTTGTCGATAATTGGCGTTTACGAATGGTCAGGAAA CAATCCTATTATACCTGAATTGTGGTTGGTTCCACATTTTCTTCCGATTCACCCAGGTCGTTTT TGGTGTTTTACCCGGTTGATATACATGTCAATGGCATATCTCTATGGTAAGAAATTTGTTGGGC CTATTAGTCCTACAATATTAGCTCTGCGACAAGACCTCTATAGTATACCTTACTGCAACATTAA TTGGGACAAGGCGCGTGATTATTGTGCAAAGGAGGACCTTCATTACCCACGCTCACGGGCACAA	342	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GATCTTATATCTGGTTGCCTAACGAAAATTGTGGAGCCAATTTTGAATTGGTGGCCAGCAAACA AGCTAAGAGATAGAGCTTTAACTAACCTCATGGAGCATATCCATTATGACGACGAATCAACCAA ATATGTGGGCATTTGCCCTATTAACAAGGCATTGAACATGATTTGTTGTTGGGTAGAAAACCCA AATTCGCCTGAATTCCAACAACATCTTCCACGATTCCATGACTATTTGTGGATGGCGGAGGATG GAATGAAGGCACAGGTATATGATGGATGTCATAGCTGGGAACTAGCGTTCATAATTCATGCCTA TTGTTCCACGGATCTTACTAGCGAGTTTATCCCGACTCTAAAAAAGGCGCACGAGTTCATGAAG AACTCACAGGTTCTTTTCAACCACCCAAATCATGAAAGCTATTATCGCCACAGATCAAAAGGCT CATGGACCCTTTCAAGTGTAGATAATGGTTGGTCTGTATCTGATTGTACTGCGGAAGCTGTTAA GGCATTGCTACTATTATCAAAGATATCCGCTGACCTTGTTGGCGATCCAATAAAACAAGACAGG TTGTATGATGCCATTGATTGCATCCTATCTTTCATGAATACAGATGGAACATTTTCTACCTACG AATGCAAACGGACATTCGCTTGGTTAGAGGTTCTCAACCCTTCTGAGAGTTTTCGGAACATTGT CGTGGACTATCCATCTGTTGAATGCACATCATCTGTGGTTGATGCTCTCATATTATTTAAAGAG ACGAATCCACGATATCGAAGAGCAGAGATAGATAAATGCATTGAAGAAGCTGTTGTATTTATTG AGAACAGTCAAAATAAGGATGGTTCATGGTATGGCTCATGGGGTATATGTTTCGCATATGGATG CATGTTTGCAGTAAGGGCGTTGGTTGCTACAGGAAAAACCTACGACAATTGTGCTTCTATCAGG AAATCATGCAAATTTGTCTTATCAAAGCAACAAACAACAGGTGGATGGGGTGAAGACTATCTTT CTAGTGACAATGGGGAATATATTGATAGCGGTAGGCCTAATGCTGTGACCACCTCATGGGCAAT GTTGGCTTTAATTTATGCTGGACAGGTTGAACGTGACCCAGTACCACTGTATAATGCTGCAAGA CAGCTAATGAATATGCAGCTAGAAACAGGTGACTTCCCCCAACAGGAACACATGGGTTGCTTCA ACTCCTCCTTGAACTTCAACTACGCCAACTACCGCAATCTATACCCGATTATGGCTCTTGGGGA ACTTCGCCGTCGACTTCTTGCGATTAAGAGCTGA		
cycloartenol synthase	MWKLKVAEGGTPWLRTLNNHVGRQVWEFDPHSGSPQDLDDIETARRNFHDNRFTHKHSDDLLMR LQFAKENPMNEVLPKVKVKDVEDVTEEAVATTLRRGLNFYSTIQSHDGHWPGDLGGPMFLMPGL VITLSVTGALNAVLTDEHRKEMRRYLYNHQNKDGGWGLHIEGPSTMFGSVLCYVTLRLLGEGPN DGEGDMERGRDWILEHGGATYITSWGKMWLSVLGVFEWSGNNPMPPEIWLLPYALPVHPGRMWC HCRMVYLPMSYLYGKRFVGPITPTVLSLRKELFTVPYHDIDWNQARNLCAKEDLYYPHPLVQDI LWATLHKFVEPVFMNWPGKKLREKAIKTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNS EAFKLHLPRIYDYLWVAEDGMKMQGYNGSQLWDTAFAAQAIISTNLIDEFGPTLKKAHAFIKNS QVSEDCPGDLSKWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKIAPEIVGEPLDSKRL YDAVNVILSLQNENGGLATYELTRSYTWLEIINPAETFGDIVIDCPYVECTSAAIQALATFGKL YPGHRREEIQCCIEKAVAFIEKIQASDGSWYGSWGVCFTYGTWFGIKGLIAAGKNFSNCLSIRK ACEFLLSKQLPSGGWAESYLSCQNKVYSNLEGNRSHVVNTGWAMLALIEAEQAKRDPTPLHRAA VCLINSQLENGDFPQEEIMGVFNKNCMITYAAYRCIFPIWALGEYRRVLQAC	343	PMID: 26058429 Takase el al., 2015
beta-amyrin synthase	MWRLTIGEGGGPWLKSNNGFLGRQVWEYDADAGTPEERAEVERRAEFTKNRFQRKESQDLLLRL QYAKDNPLPANIPTEAKLEKSTEVTHETIYESLMRALHQYSSLQADDGHWPGDYSGILFIMPII IFSLYVTRSLDTFLSPEHRHEICRYIYNQQNEDGGWGKMVLGPSTMFGSCMNYATLMILGKRNG DHKDALEKGRSWILSHGTATAIPQWGKIWLSIIGVYEWSGNNPIIPELWLVPHFLPIHPGRFWC FTRLIYMSMAYLYGKKFVGPISPTILALRQDLYSIPYCNINWDKARDYCAKEDLHYPRSRAQDL	344	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ISGCLTKIVEPILNWWPANKLRDRALTNLMEHIHYDDESTKYVGICPINKALNMICCWVENPNS PEFQQHLPRFHDYLWMAEDGMKAQVYDGCHSWELAFIIHAYCSTDLTSEFIPTLKKAHEFMKNS QVLFNHPNHESYYRHRSKGSWTLSSVDNGWSVSDCTAEAVKALLLLSKISADLVGDPIKQDRLY DAIDCILSFMNTDGTFSTYECKRTFAWLEVLNPSESFRNIVVDYPSVECTSSVVDALILFKETN PRYRRAEIDKCIEEAVVFIENSQNKDGSWYGSWGICFAYGCMFAVRALVATGKTYDNCASIRKS CKFVLSKQQTTGGWGEDYLSSDNGEYIDSGRPNAVTTSWAMLALIYAGQVERDPVPLYNAARQL MNMQLETGDFPQQEHMGCFNSSLNFNYANYRNLYPIMALGELRRRLLAIKS		
beta-amyrin synthase gene sequence	ATGTGGAGGCTAACAATAGGTGAGGGCGGCGGTCCGTGGCTGAAGTCGAACAATGGCTTCCTTG GCCGCCAAGTGTGGGAGTACGACGCCGATGCCGGCACGCCGGAAGAGCGTGCCGAGGTTGAGAG GGTGCGTGCGGAATTCACAAAGAACAGGTTCCAGAGGAAGGAGTCACAGGACCTTCTTCTACGC TTGCAGTACGCAAAAGACAACCCTCTTCCGGCGAATATTCCGACAGAAGCCAAGCTTGAAAAGA GTACAGAGGTCACTCACGAGACTATCTACGAATCATTGATGCGAGCTTTACATCAATATTCCTC TCTACAAGCAGACGATGGGCATTGGCCTGGTGATTACAGTGGGATTCTCTTCATTATGCCTATC ATTATATTCTCTTTATATGTTACTAGATCACTTGACACCTTTTTATCTCCGGAACATCGTCATG AGATATGTCGCTACATTTACAATCAACAGAATGAAGATGGTGGTTGGGGAAAAATGGTTCTTGG CCCAAGTACCATGTTTGGATCGTGTATGAATTATGCAACCTTAATGATTCTTGGCGAGAAGCGA AATGGTGATCATAAGGATGCATTGGAAAAAGGGCGTTCTTGGATTTTATCTCATGGAACTGCAA CTGCAATACCACAGTGGGGAAAAATATGGTTGTCGATAATTGGCGTTTACGAATGGTCAGGAAA CAATCCTATTATACCTGAATTGTGGTTGGTTCCACATTTTCTTCCGATTCACCCAGGTCGTTTT TGGTGTTTTACCCGGTTGATATACATGTCAATGGCATATCTCTATGGTAAGAAATTTGTTGGGC CTATTAGTCCTACAATATTAGCTCTGCGACAAGACCTCTATAGTATACCTTACTGCAACATTAA TTGGGACAAGGCGCGTGATTATTGTGCAAAGGAGGACCTTCATTACCCACGCTCACGGGCACAA GATCTTATATCTGGTTGCCTAACGAAAATTGTGGAGCCAATTTTGAATTGGTGGCCAGCAAACA AGCTAAGAGATAGAGCTTTAACTAACCTCATGGAGCATATCCATTATGACGACGAATCAACCAA ATATGTGGGCATTTGCCCTATTAACAAGGCATTGAACATGATTTGTTGTTGGGTAGAAAACCCA AATTCGCCTGAATTCCAACAACATCTTCCACGATTCCATGACTATTTGTGGATGGCGGAGGATG GAATGAAGGCACAGGTATATGATGGATGTCATAGCTGGGAACTAGCGTTCATAATTCATGCCTA TTGTTCCACGGATCTTACTAGCGAGTTTATCCCGACTCTAAAAAAGGCGCACGAGTTCATGAAG AACTCACAGGTTCTTTTCAACCACCCAAATCATGAAAGCTATTATCGCCACAGATCAAAAGGCT CATGGACCCTTTCAAGTGTAGATAATGGTTGGTCTGTATCTGATTGTACTGCGGAAGCTGTTAA GGCATTGCTACTATTATCAAAGATATCCGCTGACCTTGTTGGCGATCCAATAAAACAAGACAGG TTGTATGATGCCATTGATTGCATCTATCTTTCATGAATACAGATGGAACATTTTCTACCTACGA ATGCAAACGGACATTCGCTTGGTTAGAGGTTCTCAACCCTTCTGAGAGTTTTCGGAACATTGTC GTGGACTATCCATCTGTTGAATGCACATCATCTGTGGTTGATGCTCTCATATTATTTAAAGAGA CGAATCCACGATATCGAAGAGCAGAGATAGATAAATGCATTGAAGAAGCTGTTGTATTTATTGA GAACAGTCAAAATAAGGATGGTTCATGGTATGGCTCATGGGGTATATGTTTCGCATATGGATGC ATGTTTGCAGTAAGGGCGTTGGTTGCTACAGGAAAAACCTACGACAATTGTGCTTCTATCAGGA AATCATGCAAATTTGTCTTATCAAAGCAACAAACAACAGGTGGATGGGGTGAAGACTATCTTTC	345	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TAGTGACAATGGGGAATATATTGATAGCGGTAGGCCTAATGCTGTGACCACCTCATGGGCAATG TTGGCTTTAATTTATGCTGGACAGGTTGAACGTGACCCAGTACCACTGTATAATGCTGCAAGAC AGCTAATGAATATGCAGCTAGAAACAGGTGACTTCCCCCAACAGGAACACATGGGTTGCTTCAA CTCCTCCTTGAACTTCAACTACGCCAACTACCGCAATCTATACCCGATTATGGCTCTTGGGGAA CTTCGCCGTCGACTTCTTGCGATTAAGAGCTGA		
Modified sequence of B-amyrin synthase from Avena strigosa (AJ311789), which reacts preferentially with diepoxysqualene	MWRLTIGEGGGPWLKSNNGFLGRQVWEYDADAGTPEERAEVERVRAEFTKNRFQRKESQDLLLR LQYAKDNPLPANIPTEAKLEKSTEVTHETIYESLMRALHQYSSLQADDGHWPGDYSGILFIMPI IIFSLYVTRSLDTFLSPEHRHEICRYIYNQQNEDGGWGKMVLGPSTMFGSCMNYATLMILGEKR NGDHKDALEKGRSWILSHGTATAIPQWGKIWLSIIGVYEWSGNNPIIPELWLVPHFLPIHPGRF WCFTRLIYMSMAYLYGKKFVGPISPTILALRQDLYSIPYCNINWDKARDYCAKEDLHYPRSRAQ DLISGCLTKIVEPILNWWPANKLRDRALTNLMEHIHYDDESTKYVGICPINKALNMICCWVENP NSPEFQQHLPRFHDYLWMAEDGMKAQVYDGCHSWELAFIIHAYCSTDLTSEFIPTLKKAHEFMK NSQVLFNHPNHESYYRHRSKGSWTLSSVDNGWSVSDCTAEAVKALLLLSKISADLVGDPIKQDR LYDAIDCILSFMNTDGTFSTYECKRTFAWLEVLNPSESFRNIVVDYPSVECTSSVVDALILFKE TNPRYRRAEIDKCIEEAVVFIENSQNKDGSWYGSWGICFAYGCMFAVRALVATGKTYDNCASIR KSCKFVLSKQQTTGGWGEDYLSSDNGEYIDSGRPNAVTTSWAMLALIYAGQVERDPVPLYNAAR QLMNMQLETGDFPQQEHMGCFNSFLNFNYANYRNLYPIMALGELRRRLLAIKS	346	PMID: 27412861 (Salmon et al., 216)
Modified sequence of from *Arabidopsis thaliana* (AtLup1, Q9C5M3.1), which reacts preferentially with diepoxysqualene	MWKLKIGKGNGEDPHLFSSNNFVGRQTWKFDHKAGSPEERAAVEEARRGFLDNRFRVKGCSDLL WRMQFLREKKFEQGIPQLKATNIEEITYETTTNALRRGVRYFTALQASDGHWPGEITGPLFFLP PLIFCLYITGHLEEVFDAEHRKEMLRHIYCHQNEDGGWGLHIESKSVMFCTVLNYICLRMLGEN PEQDACKRARQWILDRGGVIFIPSWGKFWLSILGVYDWSGTNPTPPELLMLPSFLPIHPGKILC YSRMVSIPMSYLYGKRFVGPITPLILLLREELYLEPYEEINWKKSRRLYAKEDMYYAHPLVQDL LSDTLQNFVEPLLTRWPLNKLVREKALQLTMKHIHYEDENSHYITIGCVEKVLCMLACWVENPN GDYFKKHLARIPDYMWVAEDGMKMQSFGCQLWDTGFAIQALLASNLPDETDDALKRGHNYIKAS QVRENPSGDFRSMYRHISKGAWTFSDRDHGWQVSDCTAEALKCCLLLSMMSADIVGQKIDDEQL YDSVNLLLSLQSGNGGVNAWEPSRAYKWLELLNPTEFMANTMVEREFVECTSSVIQALDLFRKL YPDHRKKEINRSIEKAVQFIQDNQTPDGSWYGNWGVCFIYATWFALGGLAAAGETYNDCLAMRN GVHFLLTTQRDDGGWGESYLSCSEQRYIPSEGERSNLVQTSWAMMALIHTGQAERDLIPLHRAA KLIINSQLENGDFPQQEIVGAFMNPCMLHYATYRNTFPLWALAEYRKVVFIVN	347	PMID: 27412861 (Salmon et al., 2016)
XP_001396506.2	MCNKSNYSSPKWWKESVVYQVYPASFNCGKSTTNTNGWGDVTGIIEKVPYLESLGVDIVWLSPI YTSPQVDMGYDIADYESIDPRYGTLADVDLLIKTLKDHDMKLMMDLVVNHTSDQHSWFVESANS KDSPKRDWYIWRPAKGFDEAGNPVPPNNWAQILGDTLSAWTWHAETQEFYLTLHTSAQAELNWE NPDVVTAVYDVMEFWLRRGICGFRMDVINFISKDQSFPDAPIIDPASKYQPGEQFYTNGPRFHE FMHGIYDNVLSKYDTITVGETPYVTDMKEIIKTVGSTAKELNMAFNFDHMEIEDIKTKGESKWS LRDWKLTELKGILSGWQKRMREWDGWNAIFLECHDQARSVSRYTNDSDEFRDRGAKLLALLETT LGGTIFLYQGQEIGMRNFPVEWDPDTEYKDIESVNFWKKSKELHPVGSEGLAQARTLLQKKARD HARTPMQWSADPHAGFTVPDATPWMRVNDDYGTVNVEAQMSFPWEMKGELSVWQYWQQALQRRK LHKGAFVYGDFEDLDYHNELVFAYSRTSADGKETWLVAMNWTTDAVEWTVPSGIHVTRWVSSTL	352	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	QTAPLMAGQSTVTLRALEGVVGCCS		
beta-glucosidase [*Trichoderma reesei*] GenBank: BAP5915.1	MTSFHDGVKLSTVTCVLSGLVALGSAGPTAASANAQVAAAAAAQAWVPDGYYVPPYYPAPYGGW VEDWQESYTKAKALVDSMTLAEKTNITAGTGIYMGERCAGNTGSAFRVSFPQLCLNDSPAGVRH ADNVTAFPDGITVGATFDKALMYKRGVAIGKENRGKGVNVWLGPTVGPIGRKPKGGRNWEGFGA DPVLQAVGARETIKGVQEQGVIATIKHFIGNEQEMYRMYNPFQYAYSSNIDDRTLHEVYAWPFA EGIRAGVGAVMMAYNAVNGTACSQHPYLMSALLKDEMGFQGFIMTDWLAHMSGVASAIAGLDMD MPGDVQIPFFGGSYWMYELTRSALNGSVPMDRINDAATRIAAAWYKMGQDKGFPATNFDTNSRA AFNPLYPAALPLSPFGITNEFVPVQDDHDVIARQISQEAITLLKNDGDILPLSPSQHLKVFGTD AQKNPDGINSCTDRNCNKGTLGQGWGSGTVDYPYLDDPISAITAEADNVTFYNTDKFPSVGEVS DSDVAIVFVNSDAGENTYTVEGNHGDRDKSGLYAWHDGDKLVQDAASKFSNVIVVIHTVGPLIL EKWIDLPSVKAVLVAHLPGQEAGKSLTNVLFGHASPCGHLPYSITKEEDDLPKSVTTLIDSEFL NQPQDTYTEGLYIDYRWLNKNKTKPRYAFGHGLSYTNFTFKAASIKQVARLSAYPPARPAKGST PDFAQSIPSASEAVAPSGFGKIPRYIYSWLSQGDANRAISDGKTGKYPYPDGYSTTQKPGARAG GGEGGNPALWDVAYSLTVTVQNTGDEYAGKASVQAYLQFPDDIDYDTPIIQLRDFEKTKELKPG ETTTVTLTLTRKDVSVWDVVAQDWKVPAVDGGYKVWIGDASDSLSIVCHTDTLECETGVVGPV	354	
Beta-glucosidase [*Trichodermareesei*] BAP59014.1 GI:690966588	MMGFDVEDVLSQLSQNEKIALLSGIDFWHTYPIPKYNVPSVRLTDGPNGIRGTKFFAGIPAACL PCGTALASTWDKQLLKKAGKLLGDECIAKGAHCWLGPTINTPRSPLGGRGFESFSEDPYLSGIL AASMILGCESTGVISAVKHFVANDQEHERRAVDCLITQRALREVYLRPFQIVARDARPGALMTS YNKVNGKHVADSAEFLQGILRTEWNWDPLIVSDWYGTYTTIDAIKAGLDLEMPGVSRYRGKYIE SALQARLLKQSTIDERARRVLRFAQKASHLKVSEVEQGRDFPEDRVLNRQICGSSIVLLKNENS ILPLPKSVKKVALVGSHVRLPAISGGGSASLVPYYAISLYDAVSEVLAGATITHEVGAYAHQML PVIDAMISNAVIHFYNDPIDVKDRKLLGSENVSSTSFQLMDYNNIPTLNKAMFWGTLVGEFIPT ATGIWEFGLSVFGTADLYIDNELVIENTTHQTRGTAFFGKGTTEKVATRRMVAGSTYKLRLEFG SANTTKMETTGVVNFGGGAVHLGACLKVDPQEMIARAVKAAADADYTIICTGLSGEWESEGFDR PHMDLPPGVDTMISQVLDAAPNAVVVNQSGTPVTMSWAHKAKAIVQAWYGGNETGHGISDVLFG NVNPSGKLSLSWPVDVKHNPAYLNYASVGGRVLYGEDVYVGYKFYDKTEREVLFPFGHGLSYAT FKLPDSTVRTVPETFHPDQPTVAIVKIKNTSSVPGAQVLQLYISAPNSPTHRPVKELHGFEKVY LEAGEEKEVQIPIDQYATSFWDEIESMWKSERGIYDVLVGFSSQEISGKGKLIVPETRFWMGL	355	
beta-glucosidase [*Trichoderma reesei*] AHK23047.1 GI:588294532	MLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNRTAEDIALL KSLGAKSYRFSISWSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFITLFHWDLPEGLHQ RYGGLLNRTEFPLDFENYARVMFRALPKVRNWITFNEPLCSAIPGYGSGTFAPGRQSTSEPWTV GHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYPWDAADPADKEAAERRLEFFTAWFAD PIYLGDYPASMRKQLGDRLPTFTPEERALVHGSNDFYGMNHYTSNYIRHRSSPASADDTVGNVD VLFTNKQGNCIGPETQSPWLRPCAAGFRDFLVWISKRYGYPPIYVTENGTSIKGESDLPKEKIL EDDFRVKYYNEYIRAMVTAVELDGVNVKGYFAWSLMDNFEWADGYVTRFGVTYVDYENGQKRFP KKSAKSLKPLFDELIAAA	356	
beta-glucosidase [*Trichoderma*	MLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNRTAEDIALL KSLGAKSYRFSISWSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFITLFHWDLPEGLHQ	357	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
reesei] BAA74959.1 GI:4249562	RYGGLLNRTEFPLDFENYARVMFRALPKVRNWITFNEPLCSAIPGYGSGTFAPGRQSTSEPWTV GHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYPWDAADPADKEAAERRLEFFTAWFAD PIYLGDYPASMRKQLGDRLPTFTPEERALVHGSNDFYGMNHYTSNYIRHRSSPASADDTVGNVD VLFTNKQGNCIGPETQSPWLRPCAAGFRDFLVWISKRYGYPPIYVTENGTSIKGESDLPKEKIL EDDFRVKYYNEYIRAMVTAVELDGVNVKGYFAWSLMDNFEWADGYVTRFGVTYVDYENGQKRFP KKSAKSLKPLFDELIAAA		
beta-glucosidase [*Trichoderma reesei* RUT C-3] ETS5552.1 GI:572282538	MLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNRTAEDIALL KSLGAKSYRFSISWSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFITLFHWDLPEGLHQ RYGGLLNRTEFPLDFENYARVMFRALPKVRNWITFNEPLCSAIPGYGSGTFAPGRQSTSEPWTV GHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYPWDAADPADKEAAERRLEFFTAWFAD PIYLGDYPASMRKQLGDRLPTFTPEERALVHGSNDFYGMNHYTSNYIRHRSSPASADDTVGNVD VLFTNKQGNCIGPETQSPWLRPCAAGFRDFLVWISKRYGYPPIYVTENGTSIKGESDLPKEKIL EDDFRVKYYNEYIRAMVTAVELDGVNVKGYFAWSLMDNFEWADGYVTRFGVTYVDYENGQKRFP KKSAKSLKPLFDELIAAA	358	
Chain D, Crystal Structure Of Beta-Glucosidase 2 From Fungus *Trichoderma Reesei* In Complex With Tris 3AHY_D GI:303324838	MHHHHHHMLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNRT AEDIALLKSLGAKSYRFSISWSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFITLFHWD LPEGLHQRYGGLLNRTEFPLDFENYARVMFRALPKVRNWITFNEPLCSAIPGYGSGTFAPGRQS TSEPWTVGHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYPWDAADPADKEAAERRLEF FTAWFADPIYLGDYPASMRKQLGDRLPTFTPEERALVHGSNDFYGMNHYTSNYIRHRSSPASAD DTVGNVDVLFTNKQGNCIGPETQSPWLRPCAAGFRDFLVWISKRYGYPPIYVTENGTSIKGESD LPKEKILEDDFRVKYYNEYIRAMVTAVELDGVNVKGYFAWSLMDNFEWADGYVTRFGVTYVDYE NGQKRFPKKSAKSLKPLFDELIAAA	359	
Chain C, Crystal Structure Of Beta-Glucosidase 2 From Fungus *Trichoderma Reesei* In Complex With Tris 3AHY_C GI:33324837	MHHHHHHMLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNRT AEDIALLKSLGAKSYRFSISWSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFITLFHWD LPEGLHQRYGGLLNRTEFPLDFENYARVMFRALPKVRNWITFNEPLCSAIPGYGSGTFAPGRQS TSEPWTVGHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYPWDAADPADKEAAERRLEF FTAWFADPIYLGDYPASMRKQLGDRLPTFTPEERALVHGSNDFYGMNHYTSNYIRHRSSPASAD DTVGNVDVLFTNKQGNCIGPETQSPWLRPCAAGFRDFLVWISKRYGYPPIYVTENGTSIKGESD LPKEKILEDDFRVKYYNEYIRAMVTAVELDGVNVKGYFAWSLMDNFEWADGYVTRFGVTYVDYE NGQKRFPKKSAKSLKPLFDELIAAA	360	
Chain B, Crystal Structure Of Beta-Glucosidase 2 From Fungus *Trichoderma Reesei* In Complex With Tris	MHHHHHHMLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNRT AEDIALLKSLGAKSYRFSISWSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFITLFHWD LPEGLHQRYGGLLNRTEFPLDFENYARVMFRALPKVRNWITFNEPLCSAIPGYGSGTFAPGRQS TSEPWTVGHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYPWDAADPADKEAAERRLEF FTAWFADPIYLGDYPASMRKQLGDRLPTFTPEERALVHGSNDFYGMNHYTSNYIRHRSSPASAD DTVGNVDVLFTNKQGNCIGPETQSPWLRPCAAGFRDFLVWISKRYGYPPIYVTENGTSIKGESD LPKEKILEDDFRVKYYNEYIRAMVTAVELDGVNVKGYFAWSLMDNFEWADGYVTRFGVTYVDYE	361	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
3AHY_B GI:303324836	NGQKRFPKKSAKSLKPLFDELIAAA		
Chain A, Crystal Structure Of Beta-Glucosidase 2 From Fungus Trichoderma Reesei In Complex With Tris 3AHY_A GI:303324835	MHHHHHHMLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNRT AEDIALLKSLGAKSYRFSISWSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFITLFHWD LPEGLHQRYGGLLNRTEFPLDFENYARVMFRALPKVRNWITFNEPLCSAIPGYGSGTFAPGRQS TSEPWTVGHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYPWDAADPADKEAAERRLEF FTAWFADPIYLGDYPASMRKQLGDRLPTFTPEERALVHGSNDFYGMNHYTSNYIRHRSSPASAD DTVGNVDVLFTNKQGNCIGPETQSPWLRPCAAGFRDFLVWISKRYGYPPIYVTENGTSIKGESD LPKEKILEDDFRVKYYNEYIRAMVTAVELDGVNVKGYFAWSLMDNFEWADGYVTRFGVTYVDYE NGQKRFPKKSAKSLKPLFDELIAAA	362	
beta-glucosidase-like protein [*Trichoderma reesei*] BAP59016.1 GI:690966592	MRLKHWKTAAFAAASIVSQVEAGFWNFGRDTSSSTRPPTKDQFIESLISKLTLEDLVLQLHLMF ADDIVGAASHNELYDQTMHLSPKSPIGTIHDWYPMNKSYFNVLQKLQLDNSHVKIPMMLVEECL HGVGSFKQSIFPQNIAMAASFDTDIVYRVGRAIGTEARSIGIHGCFSPVLDLAQDPRWGRVQED FGEDKILTSHIGSAYSSGLSKNKTWSDPDAVFPIMKHFAAHGAAQAGHNTAPFTGLGPRQIKQD LLVPFKANYDLGGARGVMMAYNEIDGVPSCVNPMLYEVLDDWGYDGIVIGDDTAMRNLLTQHRV TTSEADTLQQWYNAGGQIDFYDFDLDSKINITKALVANGTVPLKTLQSHVRKILGVKWDLGLFE NPYIPEHIDPLAVVASHQDVALEAAHKSIILLKNDNRTLPLSSPKKIALIGPFADTINLGDYSG ALGQYPAKYTQTLREGVLRHANKSGHTVRTSWGTNSWEYNNQYVIPGYLLSTNGKPGGLKATYY AHTNFTSPKATRVEVPAQDWGLYPPPGLSSNNFSAVWEGELESPTDLDVNGWIGLAIGPNSTSK LYVDGKLISSKGYSGSGNLLGTIEGYAWTQANSTLPPQGGVEFTFKKNAKHHVRIEFQSWNNYK KTANVNSVNSQLIFWWNLVSPNGKALDQAVSIAKDSDVVILAVGAAWNSDGESGDRGTLGLAPS QDELAREVFALGKPVVLVLEGGRPSAIPDHYGNSSAVLSTFFGGQAGGQAIADVLFGDFNPGAR VPITVPWSVGQIPAYYNYKPSARAAQYLDIPSEPIYPFGYGLSYTTFSTSSPTASVSGSSKRSS VDAQTSQSFGSGDWITFSVTVKNTGSVAGSYVAQVYLLGRVSTITQPVKQLVGFQRVYLEAGQK KTANIQLEVDRYLKIINRKDEWELEKGSYTFALLEHGGSNADTSKNVTLQCVG	363	
extracellularbeta glucosidase 1713235A GI:227874	MRYRTAAALALATGPFARADSHSTSGASAEAVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVS GVGWNGGPCVGNTSPASKISYPSLCLQDGPLGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFI GEEVKASGIHVILGPVAGPLGKTPQGGRNWEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYI LNEQELNRETISSNPDDRTLHELYTWPFADAVQANVASVMCSYNKVNTTWACEDQYTLQTVLKD QLGFPGYVMTDWDAQHTTVQSANSGLDMSMPGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMV TRILAAWYLTGQDQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVVG SAAIIGNHARNSPSCNDKGCDDGALGMGWGSGAVNYPYFVAPYDAINTRASSQGTQVTLSNTDN TSSGASAARGKDVAIVFITADSGEGYITVEGNAGDRNNLDPWHNGNALVQAVAGANSNVIVVVH SVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVS GGSDSFSEGLFIDYKHFDDANITPRYEFGYGLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSD LFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRR RDLSYWDTASQKWVVPSGSFGISVGASSRDIRLTST LSVA	364	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Chain A, Crystal Structure Of A Glycoside Hydrolase Family 3 Beta-glucosidase, Bgl1 From *Hypocrea Jecorina* 3ZZ1_A GI:429544273	VVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVSGVGWNGGPCVGNTSPASKISYPSLCLQDGPL GVRYSTGSTAFTPGVQAASTWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQGGRNWE GFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDRTLHELYTWPFADA VQANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQSANSGLDMSMP GTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMVTRILAAWYLTGQDQAGYPSFNISRNVQGNHK TNVRAIARDGIVLLKNDANILPLKKPASIAVVGSAAIIGNHARNSPSCNDKGCDDGALGMGWGS GAVNYPYFVAPYDAINTRASSQGTQVTLSNTDNTSSGASAARGKDVAIVFITADSGEGYITVEG NAGDRNNLDPWHNGNALVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESGN ALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFGYG LSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQLYITYP SSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRRRDLSYWDTASQKWVVPSGSFGISVGASSRDI RLTSTLSVA	365	
Chain A, Crystal Structure Of A Glycoside Hydrolase Family 3 Beta-glucosidase, Bgl1 from *Hypocrea Jecorina* 3ZYZ_A GI:429544272	VVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVSGVGWNGGPCVGNTSPASKISYPSLCLQDGPL GVRYSTGSTAFTPGVQAASTWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQGGRNWE GFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDRTLHELYTWPFADA VQANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQSANSGLDMSMP GTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMVTRILAAWYLTGQDQAGYPSFNISRNVQGNHK TNVRAIARDGIVLLKNDANILPLKKPASIAVVGSAAIIGNHARNSPSCNDKGCDDGALGMGWGS GAVNYPYFVAPYDAINTRASSQGTQVTLSNTDNTSSGASAARGKDVAIVFITADSGEGYITVEG NAGDRNNLDPWHNGNALVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESGN ALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFGYG LSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQLYITYP SSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRRRDLSYWDTASQKWVVPSGSFGISVGASSRDI RLTSTLSVA	366	
Chain B, Crystal Structure Of Beta-d-glucoside glucohydrolase from *Trichoderma Reesei* 4I8D_B GI:430801090	AVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVSGVGWNGGPCVGNTSPASKISYPSLCLQDGP LGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQGGRNW EGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDRTLHELYTWPFAD AVQANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQSANSGLDMSM PGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMVTRILAAWYLTGQDQAGYPSFNISRNVQGNH KTNVRAIARDGIVLLKNDANILPLKKPASIAVVGSAAIIGNHARNSPSCNDKGCDDGALGMGWG SGAVNYPYFVAPYDAINTRASSQGTQVTLSNTDNTSSGASAARGKDVAIVFITADSGEGYITVE GNAGDRNNLDPWHNGNALVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESG NALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFGY GLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQLYITY PSSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRRRDLSYWDTASQKWVVPSGSFGISVGASSRD IRLTSTLSVA	367	
Chain A, Crystal Structure Of Beta-	AVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVSGVGWNGGPCVGNTSPASKISYPSLCLQDGP LGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQGGRNW	368	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
d-glucoside Glucohydrolase from *Trichoderma Reesei* 4I8D_A GI:430801089	EGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDRTLHELYTWPFAD AVQANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQSANSGLDMSM PGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMVTRILAAWYLTGQDQAGYPSFNISRNVQGNH KTNVRAIARDGIVLLKNDANILPLKKPASIAVVGSAAIIGNHARNSPSCNDKGCDDGALGMGWG SGAVNYPYFVAPYDAINTRASSQGTQVTLSNTDNTSSGASAARGKDVAIVFITADSGEGYITVE GNAGDRNNLDPWHNGNALVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESG NALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFGY GLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQLYITY PSSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRRRDLSYWDTASQKWVVPSGSFGISVGASSRD IRLTSTLSVA		
Cel3d protein [*Trichoderma reesei*] AAP57759.1 GI:31747172	MILGCESTGVISAVKHFVANDQEHERRAVDCLITQRALREVYLRPFQIVARDARPGALMTSYNK VNGKHVADSAEFLQGILRTEWNWDPLIVSDWYGTYTTIDAIKAGLDLEMPGVSRYRGKYIESAL QARLLKQSTIDERARRVLRFAQKASHLKVSEVEQGRDFPEDRVLNRQICGSSIVLLKNENSILP LPKSVKKVALVGSHVRLPAISGGGSASLVPYYAISLYDAVSEVLAGATITHEVGAYAHQMLPVI DAMISNAVIHFYNDPIDVKDRKLLGSENVSSTSFQLMDYNNIPTLNKAMFWGTLVGEFIPTATG IWEFGLSVFGTADLYIDNELVIENTTHQTRGTAFFGKGTTEKVATRRMVAGSTYKLRLEFGSAN TTKMETTGVVNFGGGAVHLGACLKVDPQEMIARAVKAAADADYTIICTGLSGEWESEGFDRPHM DLPPGVDTMISQVLDAAPNAVVVNQSGTPVTMSWAHKAKAIVQAWYGGNETGHGISDVLFGNVN PSGKLSLSWPVDVKHNPAYLNYASVGGRVLYGEDVYVGYKFYDKTEREVLFPFGHGLSYATFKL PDSTVRTVPETFHPDQPTVAIVKIKNTSSVPGAQVLQLYISAPNSPTHRPVKELHGFEKVYLEA GEEKEVQIPIDQYATSFWDEIESMWKSERGIYDVLVGFSSQEISGKGKLIVPETRFWMGL	369	
Cel3c protein [*Trichoderma reesei*] AAP57756.1 GI:31747168	MADIDVEAILKKLTLAEKVDLLAGIDFWHTKALPKHGVPSLRFTDGPNGVRGTKFFNGVPAACF PCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESIGEDPFLAGLG AAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYALPFQIAVRDSQPGAFMTA YNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTEAVVAGLDLEMPGPPRFRGETLK FNVSNGKPFIHVIDQRAREVLQFVKKCAASGVTENGPETTVNNTPETAALLRKVGNEGIVLLKN ENNVLPLSKKKKTLIVGPNAKQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVP PILGEQCLTPDGAPGMRWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGT YTADEDCTYELGLVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFK IEFGSAPTYTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETEG ADRASMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYGGNETGNSIADV VFGDYNPSGKLSLSFPKRLQDNPAFLNFRTEAGRTLYGEDVYVGYRYYEFADKDVNFPFGHGLS YTTFAFSNLSVSHKDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQAAKINRPVKELKGFAKVELQ PGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVSDSSAAKDGVALRGKFTVGETYWWSG V	370	
Cel3b protein [*Trichoderma reesei*]	MKTLSVFAAALLAAVAEANPYPPPHSNQAYSPPFYPSPWMDPSAPGWEQAYAQAKEFVSGLTLL EKVNLTTGVGWMGEKCVGNVGTVPRLGMRSLCMQDGPLGLRFNTYNSAFSVGLTAAASWSRHLW VDRGTALGSEAKGKGVDVLLGPVAGPLGRNPNGGRNVEGFGSDPYLAGLALADTVTGIQNAGTI	371	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
AAP57755.1 GI:31747166	ACAKHFLLNEQEHFRQVGEANGYGYPITEALSSNVDDKTIHEVYGWPFQDAVKAGVGSFMCSYN QVNNSYACQNSKLINGLLKEEYGFQGFVMSDWQAQHTGVASAVAGLDMTMPGDTAFNTGASYFG SNLTLAVLNGTVPEWRIDDMVMRIMAPFFKVGKTVDSLIDTNFDSWTNGEYGYVQAAVNENWEK VNYGVDVRANHANHIREVGAKGTVIFKNNGILPLKKPKFLTVIGEDAGGNPAGPNGCGDRGCDD GTLAMEWGSGTTNFPYLVTPDAALQSQALQDGTRYESILSNYAISQTQALVSQPDAIAIVFANS DSGEGYINVDGNEGDRKNLTLWKNGDDLIKTVAAVNPKTIVVIHSTGPVILKDYANHPNISAIL WAGAPGQESGNSLVDILYGKQSPGRTPFTWGPSLESYGVSVMTTPNNGNGAPQDNFNEGAFIDY RYFDKVAPGKPRSSDKAPTYEFGFGLSWSTFKFSNLHIQKNNVGPMSPPNGKTIAAPSLGSFSK NLKDYGFPKNVRRIKEFIYPYLSTTTSGKEASGDAHYGQTAKEFLPAGALDGSPQPRSAASGEP GGNRQLYDILYTVTATITNTGSVMDDAVPQLYLSHGGPNEPPKVLRGFDRIERIAPGQSVTFKA DLTRRDLSNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP		
beta-D-glucoside glucohydrolase [*Trichoderma reesei*] AAA18473.1 GI:493580	MRYRTAAALALATGPFARADSHSTSGASAEAVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVS GVGWNGGPCVGNTSPASKISYPSLCLQDGPLGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFI GEEVKASGIHVILGPVAGPLGKTPQGGRNWEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYI LNEQELNRETISSNPDDRTLHELYTWPFADAVQANVASVMCSYNKVNTTWACEDQYTLQTVLKD QLGFPGYVMTDWNAQHTTVQSANSGLDMSMPGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMV TRILAAWYLTGQDQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVVG SAAIIGNHARNSPSCNDKGCDDGALGMGWGSGAVNYPYFVAPYDAINTRASSQGTQVTLSNTDN TSSGASAARGKDVAIVFITADSGEGYITVEGNAGDRNNLDPWHNGNALVQAVAGANSNVIVVVH SVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVS GGSDSFSEGLFIDYKHFDDANITPRYEFGYGLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSD LFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRR RDLSYWDTASQKWVVPSGSFGISVGASSRDIRLTST LSVA	372	
putative beta-glucosidase 1 precursor [*Trichoderma reesei* RUT C-30] ETS02983.1 GI:572279864	MTSFHDGVKLSTVTYGYYVPPYYPAPYGGWVEDWQESYTKAKALVDSMTLAEKTNITAGTGIYM GEWRCAGNTGSAFRVSFPQLCLNDSPAGVRHADNVTAFPDGITVGATFDKALMYKRGVAIGKEN RGKGVNVWLGPTVGPIGRKPKGGRNWEGFGADPVLQAVGARETIKGVQEQGVIATIKHFIGNEQ EMYRMYNPFQYAYSSNIDDRTLHEVYAWPFAEGIRAGVGAVMMAYNAVNGTACSQHPYLMSALL KDEMGFQGFIMTDWLAHMSGVASAIAGLDMDMPGDVQIPFFGGSYWMYELTRSALNGSVPMDRI NDAATRIAAAWYKMGQDKGFPATNFDTNSRAAFNPLYPAALPLSPFGITNEFVPVQDDHDVIAR QISQEAITLLKNDGDILPLSPSQHLKVFGTDAQKNPDGINSCTDRNCNKGTLGQGWGSGTVDYP YLDDPISAITAEADNVTFYNTDKFPSVGEVSDSDVAIVFVNSDAGENTYTVEGNHGDRDKSGLY AWHDGDKLVQDAASKFSNVIVVIHTVGPLILEKWIDLPSVKAVLVAHLPGQEAGKSLTNVLFGH ASPCGHLPYSITKEEDDLPKSVTTLIDSEFLNQPQDTYTEGLYIDYRWLNKNKTKPRYAFGHGL SYTNFTFKAASIKQVARLSAYPPARPAKGSTPDFAQSIPSASEAVAPSGFGKIPRYIYSWLSQG DANRAISDGKTGKYPYPDGYSTTQKPGARAGGGEGGNPALWDVAYSLTVTVQNTGDEYAGKASV QAYLQFPDDIDYDTPIIQLRDFEKTKELKPGETTTVTLTLTRKDVSVWDVVAQDWKVPAVDGGY KVWIGDASDSLSIVCHTDTLECETGVVGPV	373	
putative beta-	MVAVKQIALLAGLAHWADAAEKVITNDTHFYGQSPPVYPSPEMTGGNEWEAAYQKAKAFVGQLT	374	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
glucosidase [*Trichoderma reesei* RUT C-30] ETS01786.1 GI:572278616	LEEKVNLTAGVPPNTTCSGVIPAIERLKFPGMCLSDAGNGLRNTDFVSGFPSGIHVGASWSKDL AFRRAVAMGAEFRKKGVNVLLGPVVGPAGRTVRGGRNWEGFSVDPWLAGVLVSETVSGIQEQGV ITSTKHYILNEQETHRMPEANVSAVSSNIDDKTMHEYYLWPFQDAVRAGSGNIMCSYQRINNSY GCSNSKTLNGLLKTELGFQGFVVSDWSAQHAGVASAEAGMDMAMPGPAEFWGEHLVEAVKNGSL PESRITDMATRIIATWYQFDQDNGIPKPGIGMPSNVLDSHEIVDARDPAAVPVLLNGAIEGHVL VKNTKNTLPLKKPRKLSLFGYSATTPDFFSPSRDEQLSDSWIFGKEAYNSNYLSPDGFATFGRN GTTFGGCGSGAITPALAISPFEALKWRAAQDGTATFNNFLSDKPDVDPTSDACIVFGNAYACEG NDRPAIQDDYTDDLIKAVASQCNKTIVVLHNAGIRLVDGFVDHPNITAVIFAHLPGQESGPALT SLLYGETSPSGRLPYTVAKNDTDYGVVLDPAQATGEFAYFPQADFKEGVYLDYRYFDKEGIEPR YEFGFGLSYTTFAYLNLSVDHVSGANTYPWPGGPIVSGGQTDLWDAIATVSVDIRNTGSVASYE VAQLYIGIPGAPAKQLRGFEKPFLRPNESQSVTFHLTRRDLSVWSVERQKWQLQQGTYKIYVGS SSRRLHVNGTLDI		
predicted protein, partial [*Trichoderma reesei* QM6a] EGR51923.1 GI:340521689	HLRSHTVESPNSIVKRGTCAFPTDDPNLVAVTPDAENAGWAMSPDQPCKPGHYCPIACKPGMVM AQWSPDSSYSYPSSMDGGLYCDEDGEVHKPFPNKPYCVEGTGAVVAVNKCGEPMSWCQTVLPGN EAMLIPTLVEDQATIAVPDTSYWCETAAHYYINPPGSSVADCVWGVSSKPVGNWSPYVAGANTD GDGNTFVKLGWNPIWQDSALKSTLPSFGVKIECPDGGCNGLPCEISPNSDGSVDSKESAVGAGN AAFCVVTVPKGGVANIVAYNVDGSSGGSDSDSDSDSGSSSSAAPSSTAHGLKAGGFAALAEKPT STTAAPSSTEVSTTAAASTTEVASTTAAAESTTTAAESTAAETTDASATATTKAAHSTTGGKAS STARARPSVNPGMFHENGTSPHQTTAAPSGPSATQADSAPVTTTTKKGEAGRQQGSTAFAGLIV AFVAAACFL	375	
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] EGR50829.1 GI:340520593	MVAVKQIALLAGLAHWADAAEKVITNDTHFYGQSPPVYPSPEMTGGNEWEAAYQKAKAFVGQLT LEEKVNLTAGVPPNTTCSGVIPAIERLKFPGMCLSDAGNGLRNTDFVSGFPSGIHVGASWSKDL AFRRAVAMGAEFRKKGVNVLLGPVVGPAGRTVRGGRNWEGFSVDPWLAGVLVSETVSGIQEQGV ITSTKHYILNEQETHRMPEANVSAVSSNIDDKTMHEYYLWPFQDAVRAGSGNIMCSYQRINNSY GCSNSKTLNGLLKTELGFQGFVVSDWSAQHAGVASAEAGMDMAMPGPAEFWGEHLVEAVKNGSL PESRITDMATRIIATWYQFDQDNGIPKPGIGMPSNVLDSHEIVDARDPAAVPVLLNGAIEGHVL VKNTKNTLPLKKPRKLSLFGYSATTPDFFSPSRDEQLSDSWIFGKEAYNSNYLSPDGFATFGRN GTTFGGCGSGAITPALAISPFEALKWRAAQDGTATFNNFLSDKPDVDPTSDACIVFGNAYACEG NDRPAIQDDYTDDLIKAVASQCNKTIVVLHNAGIRLVDGFVDHPNITAVIFAHLPGQESGPALT SLLYGETSPSGRLPYTVAKNDTDYGVVLDPAQATGEFAYFPQADFKEGVYLDYRYFDKEGIEPR YEFGFGLSYTTFAYLNLSVDHVSGANTYPWPGGPIVSGGQTDLWDAIATVSVDIRNTGSVASYE VAQLYIGIPGAPAKQLRGFEKPFLRPNESQSVTFHLTRRDLSVWSVERQKWQLQQGTYKIYVGS SSRRLHVNGTLDI	376	
cell wall protein [*Trichoderma reesei* QM6a] EGR50785.1	MLACITRATLPTVVAATPSHHHHHAHRHAKKHAAARVEKRAPDVVTEVVVGATATVFELDGKIV DAATAKAGLAEGEYIIVGETTPTFVPPPPPPPATSSAAPLRAQFVEEPISSPAAPTTTSAPPPP PTTTAQATTSSAPPPPKTSKPAQSSPSSGATGLDADFPSGKISCKTFPSEYGAVALDWLGTGGW SGLQFVPNYSPDAQSISDIITGIAGQTCSKGAMCSYACPPGYQKTQWPKAQGATLQSIGGLYCN EDGFLELTRPDHPKLCEAGAGGVTIKNDLDDSVCTCRTDYPGIESMVIPACTSAGETIELTNPD	377	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
GI:340520549	ETDYYVWDGKTTSAQYYVNKKGYAVEDACVWNSPLDPRGAGNWSPINIGTGKTADGITWLSIFE NLPTSSAKLDFNIEITGDVNSKCSYIDGAWTGGDKGCTTAMPSGGKAVIRYF		
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] EGR49878.1 GI:340519640	MILGCESTGVISAVKHFVANDQEHERRAVDCLITQRALREVYLRPFQIVARDARPGALMTSYNK VNGKHVADSAEFLQGILRTEWNWDPLIVSDWYGTYTTIDAIKAGLDLEMPGVSRYRGKYIESAL QARLLKQSTIDERARRVLRFAQKASHLKVSEVEQGRDFPEDRVLNRQICGSSIVLLKNENSILP LPKSVKKVALVGSHVRLPAISGGGSASLVPYYAISLYDAVSEVLAGATITHEVGAYAHQMLPVI DAMISNAVIHFYNDPIDVKDRKLLGSENVSSTSFQLMDYNNIPTLNKAMFWGTLVGEFIPTATG IWEFGLSVFGTADLYIDNELVIENTTHQTRGTAFFGKGTTEKVATRRMVAGSTYKLRLEFGSAN TTKMETTGVVNFGGGAVHLGACLKVDPQEMIARAVKAAADADYTIICTGLSGEWESEGFDRPHM DLPPGVDTMISQVLDAAPNAVVVNQSGTPVTMSWAHKAKAIVQAWYGGNETGHGISDVLFGNVN PSGKLSLSWPVDVKHNPAYLNYASVGGRVLYGEDVYVGYKFYDKTEREVLFPFGHGLSYATFKL PDSTVRTVPETFHPDQPTVAIVKIKNTSSVPGAQVLQLYISAPNSPTHRPVKELHGFEKVYLEA GEEKEVQIPIDQYATSFWDEIESMWKSERGIYDVLVGFSSQEISGKGKLIVPETRFWMGL	378	
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] EGR49703.1 GI:340519465	MRYRTAAALALATGPFARADSHSTSGASAEAVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVS GVGWNGGPCVGNTSPASKISYPSLCLQDGPLGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFI GEEVKASGIHVILGPVAGPLGKTPQGGRNWEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYI LNEQELNRETISSNPDDRTLHELYTWPFADAVQANVASVMCSYNKVNTTWACEDQYTLQTVLKD QLGFPGYVMTDWNAQHTTVQSANSGLDMSMPGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMV TRILAAWYLTGQDQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVVG SAAIIGNHARNSPSCNDKGCDDGALGMGWGSGAVNYPYFVAPYDAINTRASSQGTQVTLSNTDN TSSGASAARGKDVAIVFITADSGEGYITVEGNAGDRNNLDPWHNGNALVQAVAGANSNVIVVVH SVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVS GGSDSFSEGLFIDYKHFDDANITPRYEFGYGLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSD LFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRR RDLSYWDTASQKWVVPSGSFGISVGASSRDIRLTST LSVA	379	
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] 814 aa protein EGR49559.1 GI:340519320	MTLAEKTNITAGTGIYMGERCAGNTGSAFRVSFPQLCLNDSPAGVRHADNVTAFPDGITVGATF DKALMYKRGVAIGKENRGKGVNVWLGPTVGPIGRKPKGGRNWEGFGADPVLQAVGARETIKGVQ EQGVIATIKHFIGNEQEMYRMYNPFQYAYSSNIDDRTLHEVYAWPFAEGIRAGVGAVMMAYNAV NGTACSQHPYLMSALLKDEMGFQGFIMTDWLAHMSGVASAIAGLDMDMPGDVQIPFFGGSYWMY ELTRSALNGSVPMDRINDAATRIAAAWYKMGQDKGFPATNFDTNSRAAFNPLYPAALPLSPFGI TNEFVPVQDDHDVIARQISQEAITLLKNDGDILPLSPSQHLKVFGTDAQKNPDGINSCTDRNCN KGTLGQGWGSGTVDYPYLDDPISAITAEADNVTFYNTDKFPSVGEVSDSDVAIVFVNSDAGENT YTVEGNHGDRDKSGLYAWHDGDKLVQDAASKFSNVIVVIHTVGPLILEKWIDLPSVKAVLVAHL PGQEAGKSLTNVLFGHASPCGHLPYSITKEEDDLPKSVTTLIDSEFLNQPQDTYTEGLYIDYRW LNKNKTKPRYAFGHGLSYTNFTFKAASIKQVARLSAYPPARPAKGSTPDFAQSIPSASEAVAPS GFGKIPRYIYSWLSQGDANRAISDGKTGKYPYPDGYSTTQKPGARAGGGEGGNPALWDVAYSLT VTVQNTGDEYAGKASVQAYLQFPDDIDYDTPIIQLRDFEKTKELKPGETTTVTLTLTRKDVSVW DVVAQDWKVPAVDGGYKVWIGDASDSLSIVCHTDTLECETGVVGPV	380	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
glycosidehydrolase family 3 protein [*Trichoderma reesei QM6a*] EGR48517.1 GI:340518276	MKTLSVFAAALLAAVAEANPYPPPHSNQAYSPPFYPSPWMDPSAPGWEQAYAQAKEFVSGLTLL EKVNLTTGVGWMGEKCVGNVGTVPRLGMRSLCMQDGPLGLRFNTYNSAFSVGLTAAASWSRHLW VDRGTALGSEAKGKGVDVLLGPVAGPLGRNPNGGRNVEGFGSDPYLAGLALADTVTGIQNAGTI ACAKHFLLNEQEHFRQVGEANGYGYPITEALSSNVDDKTIHEVYGWPFQDAVKAGVGSIMCSYN QVNNSYACQNSKLINGLLKEEYGFQGFVMSDWQAQHTGVASAVAGLDMTMPGDTAFNTGASYFG SNLTLAVLNGTVPEWRIDDMVMRIMAPFFKVGKTVDSLIDTNFDSWTNGEYGYVQAAVNENWEK VNYGVDVRANHANHIREVGAKGTVIFKNNGILPLKKPKFLTVIGEDAGGNPAGPNGCGDRGCDD GTLAMEWGSGTTNFPYLVTPDAALQSQALQDGTRYESILSNYAISQTQALVSQPDAIAIVFANS DSGEGYINVDGNEGDRKNLTLWKNGDDLIKTVAAVNPKTIVVIHSTGPVILKDYANHPNISAIL WAGAPGQESGNSLVDILYGKQSPGRTPFTWGPSLESYGVSVMTTPNNGNGAPQDNFNEGAFIDY RYFDKVAPGKPRSSDKAPTYEFGFGLSWSTFKFSNLHIQKNNVGPMSPPNGKTIAAPSLGSFSK NLKDYGFPKNVRRIKEFIYPYLSTTTSGKEASGDAHYGQTAKEFLPAGALDGSPQPRSAASGEP GGNRQLYDILYTVTATITNTGSVMDDAVPQLYLSHGGPNEPPKVLRGFDRIERIAPGQSVTFKA DLTRRDLSNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP	381	
glycosidehydrolase family 3 protein [*Trichoderma reesei* QM6a] EGR47352.1 GI:340517106	MANSIGGSSADKFDLDPLWQNLDWAIGQMMLMGWDGTQVTPQIRSLIEDHHLGSIILTAKNLKS AHHTALLVQELQMIAKNSGHPQPLLIAVDQENGGVNSLFDEDFVCQFPSAMAIAATGSLELSYE VNKATATEISACGVNLMLGPVLDVLNNARYQVIGVRASGDDPQEVSQYGLAALSGIRDAGVASC GKHFPSYGNLDFLGSNLDVPIITQTLEELSLSALVPFRNAIASGKLDAMFIGGCGISNPSMNVS HACLSDQVVDDLLRNELGFKGVAISECLEMEALSQDLGVQNGVVMAVEAGCDIVLLCRAYDVQL EAIKGLKLGYENGIITKERIFTSLKRIFHLKSTCTSWAKALNPPGINLLSQIRPSHLALSRRAY DDSITIVRDKEKLLPLSLSMHPGEELLLLTPLVKPLPASSLTKSLLESKNDPSLVSTEHDRWNH QIRERSAIMSGEGVFREFGKTLARYRNEKLLHTSYTANGVRPVHENLINRASCIIIFTADANRN LYQAGFTKHVDMMCSMLRSRGQKKQLIVVAVSSPYDFAMDKSIGTYICTYDFTENAMAALVRAL VGDSNPVGTMPGTLRKSKKVLKSRQHWLVEEFDSSRDRKGLNDLIRAVHRASDQDFRYLQTATA DTFLLANQNIKETHFVVRNSSTQALYGFAATYFVQNVGILGALIVDPTKRNMSIGRSLHRRAIK SLTQQRGIKKVQLGSCFPALFLGIPLDIEVTTTKEWFSNSGWDTQFPRRLTNMVIQDLSAWYAP EGLSQSIQRANISFDLIYGVESGDTVMHHVRTHANPEVLELYRTALEESKACGIVRAKDAAGNL LGTIIICRPNSPLARYVPPLVSLGQDIGGLLAPIVPPAPLSTLVLQGLALMGVRQNKGHKATKS VLSWVVDDAYEPLVAMGFDVLQAFEEITNSPETFQT	382	
carbohydrate esterase family 4 protein [*Trichoderma reesei* QM6a] EGR46266.1 GI:340516015	MLPRRMRKSRCCIAVLAVIAIIVMLLAAAGAFGYKKLKITPLDGKSPPWYPTPKGGSVRQWADS YQKAAEMVARMTLPEKVNITTGTGWSMGLAVGNTAPALLVGFPALALQDGPLGIRFADNATALP AGVTVGATWNRHLMYEHGRVHALEARGKGINALLGPCVGPLGRMPAGGRNWEGFGADPYLQGVA GYETIKGIQDQGVMATIKHFVANEQEHFRQAWEWVLPNALSSNIDDRTMHEIYAWPFGDAVKAG VASVMCSYNMVNNSYACGNSKLLNGILKDELGFQGFVMSDWLAQRSGVGSALAGLDMSMPGDGL RWQDGNSLWGPNLSRAVLNGSLPLERLNDMVVRIVAAWYQLGQDDEKLFDRKGPNFSSWTNDRM GVTAPASSSPQEKVVVNQFVNVQANHSILARQIAAEGTVLLKNEGVLPLSVDGLLGGGGGSNST KREGQVRIGIFGEDAGPGKGPNYCEDRSCNQGTLASGWGSGAVEFPYLVSPIEALRKKFNKDKV KLTEHLKNDELDTGVIKSQDICMVFINSDSGEGYRAWEGVRGDRNDLKPQKGGVGLVTHVGLNC	383	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GNGSGTTIVVLHSVGPVVVDPWIDMPGIKALISANLPGQESGNALASILFGEENPSGKLPYTVG KSLSDYGPGGQVMYLPNGAVPQQDFSEGLYIDYRHFDKFNIEPRYEFGFGLSYTNFDYKNLKIT ETKPRSPLPDERPAAEVEPPSFDTTIPQAEEALFPSGIRRLKKYVYPYIESVKDIKEGQYSYPD GYDKEQPLSGAGGGEGGNPSLWDSHVIVSVEVTNTGKLGGKAVPQLYLSYPASETVDFPVRVLR GFDKVYIGKGETKTVEFSLTRRDLSYWDVERQNWVIPEGEYTFAVGESSRDLRVSGTW		
glycosidehydrolase family 3 protein, partial [*Trichoderma reesei* QM6a] EGR44807.1 GI:340514546	QLFAVGFHGREINQEITTLIRDYGVGAIVLFKRNENGLVTRISPPIASQQPGPMTLGAAGSLEY AYEVAKATAEMLRYFGINMNYAPVGDVNSEPLNPVIGVRSPSDQAETVSKFAAACTKGLREHKV VPCIKHFPGHGDTAVDSHYGLPVINKSRADMEKLELIPFRDAVADNIEMVMTAHISLPQLAKDG LPATLSPDTIGILRNEWKYEGVIMTDCLEMDGIRATYGTVEGSLMAFQAGVDNVMICHTFDVQA AAVDYICGAIESGKLSQERVDQSLERLRKLKERYTNWDIALHAEPPEALEAINERGEKLARQVY ADATTLVRAQEGLLPLKATAKIAFVSPGPDVPIGGAVDSGTLPTRVPWIADTFGEQIRKRAPEM SDVRFTSSNLTEEQWEQIDEADVVILATRNARESKYQKELGLEVAKRRGSRPLISIATCAPYDF LDDEEIRTYIAVYEPTVEAFSAAVDILFGDAQPRGKLPVAH	384	
glycosidehydrolase family 3 protein [*Trichoderma reesei* QM6a] EGR44527.1 GI:340514262	MADIDVEAILKKLTLAEKVDLLAGIDFWHTKALPKHGVPSLRFTDGPNGVRGTKFFNGVPAACF PCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESIGEDPFLAGLG AAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYALPFQIAVRDSQPGAFMTA YNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTEAVVAGLDLEMPGPPRFRGETLK FNVSNGKPFIHVIDQRAREVLQFVKKCAASGVTENGPETTVNNTPETAALLRKVGNEGIVLLKN ENNVLPLSKKKKTLIVGPNAKQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTHRF LPILGEQCLTPDGAPGMRWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEG TYTADEDCTYELGLVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKF KIEFGSAPTYTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETE GADRASMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYGGNETGNSIAD VVFGDYNPSGKLSLSFPKRLQDNPAFLNFRTEAGRTLYGEDVYVGYRYYEFADKDVNFPFGHGL SYTTFAFSNLSVSHKDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQAAKINRPVKELKGFAKVEL QPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVSDSSAAKDGVALRGKFTVGETYWWS GV	385	
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] XP_006969529.1 GI:589115013	MADIDVEAILKKLTLAEKVDLLAGIDFWHTKALPKHGVPSLRFTDGPNGVRGTKFFNGVPAACF PCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESIGEDPFLAGLG AAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYALPFQIAVRDSQPGAFMTA YNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTEAVVAGLDLEMPGPPRFRGETLK FNVSNGKPFIHVIDQRAREVLQFVKKCAASGVTENGPETTVNNTPETAALLRKVGNEGIVLLKN ENNVLPLSKKKKTLIVGPNAKQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTHRF LPILGEQCLTPDGAPGMRWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEG TYTADEDCTYELGLVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKF KIEFGSAPTYTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETE GADRASMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYGGNETGNSIAD VVFGDYNPSGKLSLSFPKRLQDNPAFLNFRTEAGRTLYGEDVYVGYRYYEFADKDVNFPFGHGL	386	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	SYTTFAFSNLSVSHKDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQAAKINRPVKELKGFAKVEL QPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVSDSSAAKDGVALRGKFTVGETYWWS GV		
glycoside hydrolase family 3 protein, partial [Trichoderma reesei QM6a] XP_006969215.1 GI:589114385	QLFAVGFHGREINQEITTLIRDYGVGAIVLFKRNENGLVTRISPPIASQQPGPMTLGAAGSLEY AYEVAKATAEMLRYFGINMNYAPVGDVNSEPLNPVIGVRSPSDQAETVSKFAAACTKGLREHKV VPCIKHFPGHGDTAVDSHYGLPVINKSRADMEKLELIPFRDAVADNIEMVMTAHISLPQLAKDG LPATLSPDTIGILRNEWKYEGVIMTDCLEMDGIRATYGTVEGSLMAFQAGVDNVMICHTFDVQA AAVDYICGAIESGKLSQERVDQSLERLRKLKERYTNWDIALHAEPPEALEAINERGEKLARQVY ADATTLVRAQEGLLPLKATAKIAFVSPGPDVPIGGAVDSGTLPTRVPWIADTFGEQIRKRAPEM SDVRFTSSNLTEEQWEQIDEADVVILATRNARESKYQKELGLEVAKRRGSRPLISIATCAPYDF LDDEEIRTYIAVYEPTVEAFSAAVDILFGDAQPRGKLPVAH	387	
carbohydrate esterase family 4 protein [*Trichoderma reesei* QM6a] XP_006967911.1 GI:589111777	MLPRRMRKSRCCIAVLAVIAIIVMLLAAAGAFGYKKLKITPLDGKSPPWYPTPKGGSVRQWADS YQKAAEMVARMTLPEKVNITTGTGWSMGLAVGNTAPALLVGFPALALQDGPLGIRFADNATALP AGVTVGATWNRHLMYEHGRVHALEARGKGINALLGPCVGPLGRMPAGGRNWEGFGADPYLQGVA GYETIKGIQDQGVMATIKHFVANEQEHFRQAWEWVLPNALSSNIDDRTMHEIYAWPFGDAVKAG VASVMCSYNMVNNSYACGNSKLLNGILKDELGFQGFVMSDWLAQRSGVGSALAGLDMSMPGDGL RWQDGNSLWGPNLSRAVLNGSLPLERLNDMVVRIVAAWYQLGQDDEKLFDRKGPNFSSWTNDRM GVTAPASSSPQEKVVVNQFVNVQANHSILARQIAAEGTVLLKNEGVLPLSVDGLLGGGGGSNST KREGQVRIGIFGEDAGPGKGPNYCEDRSCNQGTLASGWGSGAVEFPYLVSPIEALRKKFNKDKV KLTEHLKNDELDTGVIKSQDICMVFINSDSGEGYRAWEGVRGDRNDLKPQKGGVGLVTHVGLNC GNGSGTTIVVLHSVGPVVVDPWIDMPGIKALISANLPGQESGNALASILFGEENPSGKLPYTVG KSLSDYGPGGQVMYLPNGAVPQQDFSEGLYIDYRHFDKFNIEPRYEFGFGLSYTNFDYKNLKIT ETKPRSPLPDERPAAEVEPPSFDTTIPQAEEALFPSGIRRLKKYVYPYIESVKDIKEGQYSYPD GYDKEQPLSGAGGGEGGNPSLWDSHVIVSVEVTNTGKLGGKAKGETKTVEFSLTRRDLSYWDVE RQNWVIPEGEYTFAVGESSRDLRVSGTW	388	
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] XP_006966911.1 GI:589109777	MANSIGGSSADKFDLDPLWQNLDWAIGQMMLMGWDGTQVTPQIRSLIEDHHLGSIILTAKNLKS AHHTALLVQELQMIAKNSGHPQPLLIAVDQENGGVNSLFDEDFVCQFPSAMAIAATGSLELSYE VNKATATEISACGVNLMLGPVLDVLNNARYQVIGVRASGDDPQEVSQYGLAALSGIRDAGVASC GKHFPSYGNLDFLGSNLDVPIITQTLEELSLSALVPFRNAIASGKLDAMFIGGCGISNPSMNVS HACLSDQVVDDLLRNELGFKGVAISECLEMEALSQDLGVQNGVVMAVEAGCDIVLLCRAYDVQL EAIKGLKLGYENGIITKERIFTSLKRIFHLKSTCTSWAKALNPPGINLLSQIRPSHLALSRRAY DDSITIVRDKEKLLPLSLSMHPGEELLLLTPLVKPLPASSLTKSLLESKNDPSLVSTEHDRWNH QIRERSAIMSGEGVFREFGKTLARYRNEKLLHTSYTANGVRPVHENLINRASCIIIFTADANRN LYQAGFTKHVDMMCSMLRSRGQKKQLIVVAVSSPYDFAMDKSIGTYICTYDFTENAMAALVRAL VGDSNPVGTMPGTLRKSKKVLKSRQHWLVEEFDSSRDRKGLNDLIRAVHRASDQDFRYLQTATA DTFLLANQNIKETHFVVRNSSTQALYGFAATYFVQNVGILGALIVDPTKRNMSIGRSLHRRAIK SLTQQRGIKKVQLGSCFPALFLGIPLDIEVTTTKEWFSNSGWDTQFPRRLTNMVIQDLSAWYAP EGLSQSIQRANISFDLIYGVESGDTVMHHVRTHANPEVLELYRTALEESKACGIVRAKDAAGNL	389	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	LGTIIICRPNSPLARYVPPLVSLGQDIGGLLAPIVPPAPLSTLVLQGLALMGVRQNKGHKATKS VLSWVVDDAYEPLVAMGFDVLQAFEEITNSPETFQT		
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] XP_006965281.1 GI:589106517	MKTLSVFAAALLAAVAEANPYPPPHSNQAYSPPFYPSPWMDPSAPGWEQAYAQAKEFVSGLTLL EKVNLTTGVGWMGEKCVGNVGTVPRLGMRSLCMQDGPLGLRFNTYNSAFSVGLTAAASWSRHLW VDRGTALGSEAKGKGVDVLLGPVAGPLGRNPNGGRNVEGFGSDPYLAGLALADTVTGIQNAGTI ACAKHFLLNEQEHFRQVGEANGYGYPITEALSSNVDDKTIHEVYGWPFQDAVKAGVGSIMCSYN QVNNSYACQNSKLINGLLKEEYGFQGFVMSDWQAQHTGVASAVAGLDMTMPGDTAFNTGASYFG SNLTLAVLNGTVPEWRIDDMVMRIMAPFFKVGKTVDSLIDTNFDSWTNGEYGYVQAAVNENWEK VNYGVDVRANHANHIREVGAKGTVIFKNNGILPLKKPKFLTVIGEDAGGNPAGPNGCGDRGCDD GTLAMEWGSGTTNFPYLVTPDAALQSQALQDGTRYESILSNYAISQTQALVSQPDAIAIVFANS DSGEGYINVDGNEGDRKNLTLWKNGDDLIKTVAAVNPKTIVVIHSTGPVILKDYANHPNISAIL WAGAPGQESGNSLVDILYGKQSPGRTPFTWGPSLESYGVSVMTTPNNGNGAPQDNFNEGAFIDY RYFDKVAPGKPRSSDKAPTYEFGFGLSWSTFKFSNLHIQKNNVGPMSPPNGKTIAAPSLGSFSK NLKDYGFPKNVRRIKEFIYPYLSTTTSGKEASGDAHYGQTAKEFLPAGALDGSPQPRSAASGEP GGNRQLYDILYTVTATITNTGSVMDDAVPQLYLSHGGPNEPPKVLRGFDRIERIAPGQSVTFKA DLTRRDLSNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP	390	
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] XP_006964430.1 GI:589104815	MTLAEKTNITAGTGIYMGERCAGNTGSAFRVSFPQLCLNDSPAGVRHADNVTAFPDGITVGATF DKALMYKRGVAIGKENRGKGVNVWLGPTVGPIGRKPKGGRNWEGFGADPVLQAVGARETIKGVQ EQGVIATIKHFIGNEQEMYRMYNPFQYAYSSNIDDRTLHEVYAWPFAEGIRAGVGAVMMAYNAV NGTACSQHPYLMSALLKDEMGFQGFIMTDWLAHMSGVASAIAGLDMDMPGDVQIPFFGGSYWMY ELTRSALNGSVPMDRINDAATRIAAAWYKMGQDKGFPATNFDTNSRAAFNPLYPAALPLSPFGI TNEFVPVQDDHDVIARQISQEAITLLKNDGDILPLSPSQHLKVFGTDAQKNPDGINSCTDRNCN KGTLGQGWGSGTVDYPYLDDPISAITAEADNVTFYNTDKFPSVGEVSDSDVAIVFVNSDAGENT YTVEGNHGDRDKSGLYAWHDGDKLVQDAASKFSNVIVVIHTVGPLILEKWIDLPSVKAVLVAHL PGQEAGKSLTNVLFGHASPCGHLPYSITKEEDDLPKSVTTLIDSEFLNQPQDTYTEGLYIDYRW LNKNKTKPRYAFGHGLSYTNFTFKAASIKQVARLSAYPPARPAKGSTPDFAQSIPSASEAVAPS GFGKIPRYIYSWLSQGDANRAISDGKTGKYPYPDGYSTTQKPGARAGGGEGGNPALWDVAYSLT VTVQNTGDEYAGKASVQAYLQFPDDIDYDTPIIQLRDFEKTKELKPGETTTVTLTLTRKDVSVW DVVAQDWKVPAVDGGYKVWIGDASDSLSIVCHTDTLECETGVVGPV	391	
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] XP_006964076.1 GI:589104107	MRYRTAAALALATGPFARADSHSTSGASAEAVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVS GVGWNGGPCVGNTSPASKISYPSLCLQDGPLGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFI GEEVKASGIHVILGPVAGPLGKTPQGGRNWEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYI LNEQELNRETISSNPDDRTLHELYTWPFADAVQANVASVMCSYNKVNTTWACEDQYTLQTVLKD QLGFPGYVMTDWNAQHTTVQSANSGLDMSMPGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMV TRILAAWYLTGQDQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVVG SAAIIGNHARNSPSCNDKGCDDGALGMGWGSGAVNYPYFVAPYDAINTRASSQGTQVTLSNTDN TSSGASAARGKDVAIVFITADSGEGYITVEGNAGDRNNLDPWHNGNALVQAVAGANSNVIVVVH SVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVS	392	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GGSDSFSEGLFIDYKHFDDANITPRYEFGYGLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSD LFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRR RDLSYWDTASQKWVVPSGSFGISVGASSRDIRLTSTLSVA		
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] XP_006964050.1 GI:589104055	MILGCESTGVISAVKHFVANDQEHERRAVDCLITQRALREVYLRPFQIVARDARPGALMTSYNK VNGKHVADSAEFLQGILRTEWNWDPLIVSDWYGTYTTIDAIKAGLDLEMPGVSRYRGKYIESAL QARLLKQSTIDERARRVLRFAQKASHLKVSEVEQGRDFPEDRVLNRQICGSSIVLLKNENSILP LPKSVKKVALVGSHVRLPAISGGGSASLVPYYAISLYDAVSEVLAGATITHEVGAYAHQMLPVI DAMISNAVIHFYNDPIDVKDRKLLGSENVSSTSFQLMDYNNIPTLNKAMFWGTLVGEFIPTATG IWEFGLSVFGTADLYIDNELVIENTTHQTRGTAFFGKGTTEKVATRRMVAGSTYKLRLEFGSAN TTKMETTGVVNFGGGAVHLGACLKVDPQEMIARAVKAAADADYTIICTGLSGEWESEGFDRPHM DLPPGVDTMISQVLDAAPNAVVVNQSGTPVTMSWAHKAKAIVQAWYGGNETGHGISDVLFGNVN PSGKLSLSWPVDVKHNPAYLNYASVGGRVLYGEDVYVGYKFYDKTEREVLFPFGHGLSYATFKL PDSTVRTVPETFHPDQPTVAIVKIKNTSSVPGAQVLQLYISAPNSPTHRPVKELHGFEKVYLEA GEEKEVQIPIDQYATSFWDEIESMWKSERGIYDVLVGFSSQEISGKGKLIVPETRFWMGL	393	
glycoside hydrolase family 3 protein [*Trichoderma reesei* QM6a] XP_006963375.1 GI:589102705	MVAVKQIALLAGLAHWADAAEKVITNDTHFYGQSPPVYPSPEMTGGNEWEAAYQKAKAFVGQLT LEEKVNLTAGVPPNTTCSGVIPAIERLKFPGMCLSDAGNGLRNTDFVSGFPSGIHVGASWSKDL AFRRAVAMGAEFRKKGVNVLLGPVVGPAGRTVRGGRNWEGFSVDPWLAGVLVSETVSGIQEQGV ITSTKHYILNEQETHRMPEANVSAVSSNIDDKTMHEYYLWPFQDAVRAGSGNIMCSYQRINNSY GCSNSKTLNGLLKTELGFQGFVVSDWSAQHAGVASAEAGMDMAMPGPAEFWGEHLVEAVKNGSL PESRITDMATRIIATWYQFDQDNGIPKPGIGMPSNVLDSHEIVDARDPAAVPVLLNGAIEGHVL VKNTKNTLPLKKPRKLSLFGYSATTPDFFSPSRDEQLSDSWIFGKEAYNSNYLSPDGFATFGRN GTTFGGCGSGAITPALAISPFEALKWRAAQDGTATFNNFLSDKPDVDPTSDACIVFGNAYACEG NDRPAIQDDYTDDLIKAVASQCNKTIVVLHNAGIRLVDGFVDHPNITAVIFAHLPGQESGPALT SLLYGETSPSGRLPYTVAKNDTDYGVVLDPAQATGEFAYFPQADFKEGVYLDYRYFDKEGIEPR YEFGFGLSYTTFAYLNLSVDHVSGANTYPWPGGPIVSGGQTDLWDAIATVSVDIRNTGSVASYE VAQLYIGIPGAPAKQLRGFEKPFLRPNESQSVTFHLTRRDLSVWSVERQKWQLQQGTYKIYVGS SSRRLHVNGTLDI	394	
cell wall protein [Trichoderma reesei QM6a] XP_006963339.1 GI:589102633	MLACITRATLPTVVAATPSHHHHHAHRHAKKHAAARVEKRAPDVVTEVVVGATATVFELDGKIV DAATAKAGLAEGEYIIVGETTPTFVPPPPPPPATSSAAPLRAQFVEEPISSPAAPTTTSAPPPP PTTTAQATTSSAPPPPKTSKPAQSSPSSGATGLDADFPSGKISCKTFPSEYGAVALDWLGTGGW SGLQFVPNYSPDAQSISDIITGIAGQTCSKGAMCSYACPPGYQKTQWPKAQGATLQSIGGLYCN EDGFLELTRPDHPKLCEAGAGGVTIKNDLDDSVCTCRTDYPGIESMVIPACTSAGETIELTNPD ETDYYVWDGKTTSAQYYVNKKGYAVEDACVWNSPLDPRGAGNWSPINIGTGKTADGITWLSIFE NLPTSSAKLDFNIEITGDVNSKCSYIDGAWTGGDKGCTTAMPSGGKAVIRYF	395	
predicted protein, partial [*Trichoderma reesei* QM6a]	HLRSHTVESPNSIVKRGTCAFPTDDPNLVAVTPDAENAGWAMSPDQPCKPGHYCPIACKPGMVM AQWSPDSSYSYPSSMDGGLYCDEDGEVHKPFPNKPYCVEGTGAVVAVNKCGEPMSWCQTVLPGN EAMLIPTLVEDQATIAVPDTSYWCETAAHYYINPPGSSVADCVWGVSSKPVGNWSPYVAGANTD GDGNTFVKLGWNPIWQDSALKSTLPSFGVKIECPDGGCNGLPCEISPNSDGSVDSKESAVGAGN	396	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
XP_006962014.1 GI:589099983	AAFCVVTVPKGGVANIVAYNVDGSSGGSDSDSDSDSGSSSSAAPSSTAHGLKAGGFAALAEKPT STTAAPSSTEVSTTAAASTTEVASTTAAAESTTTAAESTAAETTDASATATTKAAHSTTGGKAS STARARPSVNPGMFHENGTSPHQTTAAPSGPSATQADSAPVTTTTKKGEAGRQQGSTAFAGLIV AFVAAACFL		
hypothetical protein M419DRAFT_70331 [*Trichoderma reesei* RUT C-30] ETS05514.1 GI:572282500	MKVTDVQAALASAVVLLSLPAGSVASSHKRFHQLPNKKHTHLRSHTVESPNSIVKRGTCAFPTD DPNLVAVTPDAENAGWAMSPDQPCKPGHYCPIACKPGMVMAQWSPDSSYSYPSSMDGGLYCDED GEVHKPFPNKPYCVEGTGAVVAVNKCGEPMSWCQTVLPGNEAMLIPTLVEDQATIAVPDTSYWC ETAAHYYINPPGSSVADCVWGVSSKPVGNWSPYVAGANTDGDGNTFVKLGWNPIWQDSALKSTL PSFGVKIECPDGGCNGLPCEISPNSDGSVDSKESAVGAGNAAFCVVTVPKGGVANIVAYNKPTS TTAAPSSTEVSTTAAASTTEVASTTAAAESTTTAAESTAAETTDASATATTKAAHSTTGGKASS TARARPSVNPGMFHENGTSPHQTTAAPSGPSATQADSAPVTTTTKKGEAGRQQGSTAFAGLIVA FVAAACFL	397	
beta-D-glucoside glucohydrolase I [*Trichoderma reesei* RUT C-30] ETS03194.1 GI:572280097	MRYRTAAALALATGPFARADSHSTSGASAEAVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVS GVGWNGGPCVGNTSPASKISYPSLCLQDGPLGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFI GEEVKASGIHVILGPVAGPLGKTPQGGRNWEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYI LNEQELNRETISSNPDDRTLHELYTWPFADAVQANVASVMCSYNKVNTTWACEDQYTLQTVLKD QLGFPGYVMTDWNAQHTTVQSANSGLDMSMPGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMV TRILAAWYLTGQDQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVVG SAAIIGNHARNSPSCNDKGCDDGALGMGWGSGAVNYPYFVAPYDAINTRASSQGTQVTLSNTDN TSSGASAARGKDVAIVFITADSGEGYITVEGNAGDRNNLDPWHNGNALVQAVAGANSNVIVVVH SVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVS GGSDSFSEGLFIDYKHFDDANITPRYEFGYGLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSD LFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRR RDLSYWDTASQKWVVPSGSFGISVGASSRDIRLTSTLSVA	398	
hypotheticalprotein M419DRAFT_122639 [*Trichoderma reesei* RUT C-30] ETS03170.1 GI:572280073	MILGCESTGVISAVKHFVANDQEHERRAVDCLITQRALREVYLRPFQIVARDARPGALMTSYNK VNGKHVADSAEFLQGILRTEWNWDPLIVSDWYGTYTTIDAIKAGLDLEMPGVSRYRGKYIESAL QARLLKQSTIDERARRVLRFAQKASHLKVSEVEQGRDFPEDRVLNRQICGSSIVLLKNENSILP LPKSVKKVALVGSHVRLPAISGGGSASLVPYYAISLYDAVSEVLAGATITHEVGAYAHQMLPVI DAMISNAVIHFYNDPIDVKDRKLLGSENVSSTSFQLMDYNNIPTLNKAMFWGTLVGEFIPTATG IWEFGLSVFGTADLYIDNELVIENTTHQTRGTAFFGKGTTEKVATRRMVAGSTYKLRLEFGSAN TTKMETTGVVNFGGGAVHLGACLKVDPQEMIARAVKAAADADYTIICTGLSGEWESEGFDRPHM DLPPGVDTMISQVLDAAPNAVVVNQSGTPVTMSWAHKAKAIVQAWYGGNETGHGISDVLFGNVN PSGKLSLSWPVDVKHNPAYLNYASVGGRVLYGEDVYVGYKFYDKTEREVLFPFGHGLSYATFKL PDSTVRTVPETFHPDQPTVAIVKIKNTSSVPGAQVLQLYISAPNSPTHRPVKELHGFEKVYLEA GEEKEVQIPIDQYATSFWDEIESMWKSERGIYDVLVGFSSQEISGKGKLIVPETRFWMGL	399	
SUN-domain-containing protein [*Trichoderma*	MLACITRATLPTVVAATPSHHHHHAHRHAKKHAAARVEKRAPDVVTEVVVGATATVFELDGKIV DAATAKAGLAEGEYIIVGETTPTFVPPPPPPPATSSAAPLRAQFVEEPISSPAAPTTTSAPPPP PTTTAQATTSSAPPPPKTSKPAQSSPSSGATGLDADFPSGKISCKTFPSEYGAVALDWLGTGGW	400	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
reesei RUT C-30] ETS01671.1 GI:572278501	SGLQFVPNYSPDAQSISDIITGIAGQTCSKGAMCSYACPPGYQKTQWPKAQGATLQSIGGLYCN EDGFLELTRPDHPKLCEAGAGGVTIKNDLDDSVCTCRTDYPGIESMVIPACTSAGETIELTNPD ETDYYVWDGKTTSAQYYVNKKGYAVEDACVWNSPLDPRGAGNWSPINIGTGKTADGITWLSIFE NLPTSSAKLDFNIEITGDVNSKCSYIDGAWTGGDKGCTTAMPSGGKAVIRYF		
hypotheticalprotein M419DRAFT_25095 [*Trichoderma reesei* RUT C-30] ETS01349.1 GI:572278157	MKTLSVFAAALLAAVAEANPYPPPHSNQAYSPPFYPSPWMDPSAPGWEQAYAQAKEFVSGLTLL EKVNLTTGVGWMGEKCVGNVGTVPRLGMRSLCMQDGPLGLRFNTYNSAFSVGLTAAASWSRHLW VDRGTALGSEAKGKGVDVLLGPVAGPLGRNPNGGRNVEGFGSDPYLAGLALADTVTGIQNAGTI ACAKHFLLNEQEHFRQVGEANGYGYPITEALSSNVDDKTIHEVYGWPFQDAVKAGVGSIMCSYN QVNNSYACQNSKLINGLLKEEYGFQGFVMSDWQAQHTGVASAVAGLDMTMPGDTAFNTGASYFG SNLTLAVLNGTVPEWRIDDMVMRIMAPFFKVGKTVDSLIDTNFDSWTNGEYGYVQAAVNENWEK VNYGVDVRANHANHIREVGAKGTVIFKNNGILPLKKPKFLTVIGEDAGGNPAGPNGCGDRGCDD GTLAMEWGSGTTNFPYLVTPDAALQSQALQDGTRYESILSNYAISQTQALVSQPDAIAIVFANS DSGEGYINVDGNEGDRKNLTLWKNGDDLIKTVAAVNPKTIVVIHSTGPVILKDYANHPNISAIL WAGAPGQESGNSLVDILYGKQSPGRTPFTWGPSLESYGVSVMTTPNNGNGAPQDNFNEGAFIDY RYFDKVAPGKPRSSDKAPTYEFGFGLSWSTFKFSNLHIQKNNVGPMSPPNGKTIAAPSLGSFSK NLKDYGFPKNVRRIKEFIYPYLSTTTSGKEASGDAHYGQTAKEFLPAGALDGSPQPRSAASGEP GGNRQLYDILYTVTATITNTGSVMDDAVPQLYLSHGGPNEPPKVLRGFDRIERIAPGQSVTFKA DLTRRDLSNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP	401	
beta-N-acetylglucosaminidase [*Trichoderma reesei* RUT C-30] ETS00749.1 GI:572277491	MANSIGGSSADKFDLDPLWQNLDWAIGQMMLMGWDGTQVTPQIRSLIEDHHLGSIILTAKNLKS AHHTALLVQELQMIAKNSGHPQPLLIAVDQENGGVNSLFDEDFVCQFPSAMAIAATGSLELSYE VNKATATEISACGVNLMLGPVLDVLNNARYQVIGVRASGDDPQEVSQYGLAALSGIRDAGVASC GKHFPSYGNLDFLGSNLDVPIITQTLEELSLSALVPFRNAIASGKLDAMFIGGCGISNPSMNVS HACLSDQVVDDLLRNELGFKGVAISECLEMEALSQDLGVQNGVVMAVEAGCDIVLLCRAYDVQL EAIKGLKLGYENGIITKERIFTSLKRIFHLKSTCTSWAKALNPPGINLLSQIRPSHLALSRRAY DDSITIVRDKEKLLPLSLSMHPGEELLLLTPLVKPLPASSLTKSLLESKNDPSLVSTEHDRWNH QIRERSAIMSGEGVFREFGKTLARYRNEKLLHTSYTANGVRPVHENLINRASCIIIFTADANRN LYQAGFTKHVDMMCSMLRSRGQKKQLIVVAVSSPYDFAMDKSIGTYICTYDFTENAMAALVRAL VGDSNPVGTMPGTLRKSKKVLKSRQHWLVEEFDSSRDRKGLNDLIRAVHRASDQDFRYLQTATA DTFLLANQNIKETHFVVRNSSTQALYGFAATYFVQNVGILGALIVDPTKRNMSIGRSLHRRAIK SLTQQRGIKKVQLGSCFPALFLGIPLDIEVTTTKEWFSNSGWDTQFPRRLTNMVIQDLSAWYAP EGLSQSIQRANISFDLIYGVESGDTVMHHVRTHANPEVLELYRTALEESKACGIVRAKDAAGNL LGTIIICRPNSPLARYVPPLVSLGQDIGGLLAPIVPPAPLSTLVLQGLALMGVRQNKGHKATKS VLSWVVDDAYEPLVAMGFDVLQAFEEITNSPETFQT	402	
hypothetical protein M419DRAFT_86704 [*Trichoderma reesei* RUT C-30]	MLPRRMRKSRCCIAVLAVIAIIVMLLAAAGAFGYKKLKITPLDGKSPPWYPTPKGGSVRQWADS YQKAAEMVARMTLPEKVNITTGTGWSMGLAVGNTAPALLVGFPALALQDGPLGIRFADNATALP AGVTVGATWNRHLMYEHGRVHALEARGKGINALLGPCVGPLGRMPAGGRNWEGFGADPYLQGVA GYETIKGIQDQGVMATIKHFVANEQEHFRQAWEWVLPNALSSNIDDRTMHEIYAWPFGDAVKAG VASVMCSYNMVNNSYACGNSKLLNGILKDELGFQGFVMSDWLAQRSGVGSALAGLDMSMPGDGL	403	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
ETR99336.1 GI:572275968	RWQDGNSLWGPNLSRAVLNGSLPLERLNDMVVRIVAAWYQLGQDDEKLFDRKGPNFSSWTNDRM GVTAPASSSPQEKVVVNQFVNVQANHSILARQIAAEGTVLLKNEGVLPLSVDGLLGGGGGSNST KREGQVRIGIFGEDAGPGKGPNYCEDRSCNQGTLASGWGSGAVEFPYLVSPIEALRKKFNKDKV KLTEHLKNDELDTGVIKSQDICMVFINSDSGEGYRAWEGVRGDRNDLKPQKGGVGLVTHVGLNC GNGSGTTIVVLHSVGPVVVDPWIDMPGIKALISANLPGQESGNALASILFGEENPSGKLPYTVG KSLSDYGPGGQVMYLPNGAVPQQDFSEGLYIDYRHFDKFNIEPRYEFGFGLSYTNFDYKNLKIT ETKPRSPLPDERPAAEVEPPSFDTTIPQAEEALFPSGIRRLKKYVYPYIESVKDIKEGQYSYPD GYDKEQPLSGAGGGEGGNPSLWDSHVIVSVEVTNTGKLGGKAVPQLYLSYPASETVDFPVRVLR GFDKVYIGKGETKTVEFSLTRRDLSYWDVERQNWVIPEGEYTFAVGESSRDLRVSGTW		
putative beta-N-acetylglucosaminidase [*Trichoderma reesei* RUT C-30] ETR97676.1 GI:572274120	MPSPEQRRKIGQLFAVGFHGREINQEITTLIRDYGVGAIVLFKRNVLDAAQLQALCLGLQKIAR DAGHNQPLFIGIDQENGLVTRISPPIASQQPGPMTLGAAGSLEYAYEVAKATAEMLRYFGINMN YAPVGDVNSEPLNPVIGVRSPSDQAETVSKFAAACTKGLREHKVVPCIKHFPGHGDTAVDSHYG LPVINKSRADMEKLELIPFRDAVADNIEMVMTAHISLPQLAKDGLPATLSPDTIGILRNEWKYE GVIMTDCLEMDGIRATYGTVEGSLMAFQAGVDNVMICHTFDVQAAAVDYICGAIESGKLSQERV DQSLERLRKLKERYTNWDIALHAEPPEALEAINERGEKLARQVYADATTLVRAQEGLLPLKATA KIAFVSPGPDVPIGGAVDSGTLPTRVPWIADTFGEQIRKRAPEMSDVRFTSSNLTEEQWEQIDE ADVVILATRNARESKYQKELGLEVAKRRGSRPLISIATCAPYDFLDDEEIRTYIAVYEPTVEAF SAAVDILFGDAQPRGKLPVAH	404	
UDP-glycotransferase (338)	MGHKHIAIFNIPAHGHINPTLALTASLVKRGYRVTYPVTDEFVKAVEETGAEPLNYRSTLNIDP QQIRELMKNKKDMSQAPLMFIKEMEEVLPQLEALYENDKPDLILFDFMAMAGKLLAEKFGIEAV RLCSTYAQNEHFTFRSISEEFKIELTPEQEDALKNSNLPSFNFEDMFEPAKLNIVFMPRAFQPY GETFDERFSFVGPSLAKRKFQEKETPIISDSGRPVMLISLGTAFNAWPEFYHMCIEAFRDTKWQ VIMAVGTTIDPESFDDIPENFSIHQRVPQLEILKKAELFITHGGMNSTMEGLNAGVPLVAVPQM PEQEITARRVEELGLGKHLQPEDTTAASLREAVSQTDGDPHVLKRIQDMQKHIKQAGGAEKAAD EIEAFLAPAGVK.	405	Pandey et al., 2014
UDP-glycotransferase 338 (gDNA, native)	ATGGGACATAAACATATCGCGATTTTTAATATTCCGGCTCACGGCCATATTAATCCAACGCTAG CTTTAACGGCAAGCCTTGTCAAACGCGGTTATCGGGTAACATATCCGGTGACGGATGAGTTTGT GAAGGCTGTTGAGGAAACTGGGGCAGAGCCGCTCAACTACCGCTCAACTTTAAATATCGATCCG CAGCAAATTCGGGAGCTGATGAAAAATAAAAAAGATATGTCGCAGGCTCCGCTGATGTTTATCA AAGAAATGGAGGAGGTTCTTCCTCAGCTTGAAGCGCTCTATGAGAATGACAAGCCAGACCTTAT CCTTTTTGACTTTATGGCCATGGCGGGAAAACTGCTGGCTGAGAAGTTTGGAATAGAGGCGGTC CGCCTTTGTTCTACATATGCACAGAACGAACATTTTACATTCAGATCCATTTCTGAAGAGTTTA AGATCGAGCTGACGCCTGAGCAAGAGGATGCTTTGAAAAATTCGAATCTTCCGTCATTTAACTT TGAGGATATGTTCGAGCCTGCAAAATTGAACATTGTCTTTATGCCTCGTGCTTTTCAGCCTTAC GGCGAAACGTTTGATGAGCGGTTCTCTTTTGTTGGTCCTTCTCTTGCCAAACGCAAGTTTCAGG AAAAAGAAACGCCGATTATTTCGGACAGCGGCCGTCCTGTCATGCTGATATCTTTAGGGACGGC GTTCAATGCCTGGCCGGAATTTTATCATATGTGCATAGAAGCATTCAGGGACACGAAGTGGCAG GTTATCATGGCTGTTGGCACGACAATCGATCCTGAAAGCTTTGATGACATACCTGAGAACTTTT	406	Pandey et al., 2014

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CGATTCATCAGCGCGTTCCTCAGCTGGAGATCCTGAAGAAAGCGGAGCTGTTCATCACCCATGG GGGTATGAACAGTACGATGGAAGGGTTGAATGCCGGTGTACCGCTCGTTGCCGTTCCGCAAATG CCTGAACAGGAAATCACTGCCCGCCGCGTCGAAGAGCTTGGGCTTGGCAAGCATTTGCAGCCGG AAGACACAACAGCAGCTTCACTGCGGGAAGCCGTCTCTCAGACGGATGGTGACCCGCATGTCCT GAAACGGATACAGGACATGCAAAAGCACATTAAACAAGCCGGAGGGGCCGAGAAAGCCGCAGAT GAAATTGAGGCATTTTTAGCACCCGCAGGAGTAAAATAA		
301 UGT98	MDAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRRNFHIYFCSTSVNLDAIKPKLPSSSSSDSI QLVELCLPSSPDQLPPHLHTTNALPPHLMPTLHQAFSMAAQHFAAILHTLAPHLLIYDSFQPWA PQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFPISEFVLHDYWKAMYSAAGGAVTKKDHK IGETLANCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIKN WLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAIEDALPKGFL ERVGERGMVVKGWAPQAKILKHWSTGGFVSHCGWNSVMESMMFGVPIIGVPMHLDQPFNAGLAE EAGVGVEAKRDSDGKIQREEVAKSIKEVVIEKTREDVRKKAREMGEILRSKGDEKIDELVAEIS LLRKKAPCSIAAALEHHHHHH	407	
301 UGT98 (gDNA, native)	CTCGAATTCATGGATGCCCAGCGAGGTCACACCACAACCATTTTGATGTTTCCATGGCTCGGCT ATGGCCATCTTTCGGCTTTCCTAGAGTTGGCCAAAAGCCTCTCAAGGAGGAACTTCCATATCTA CTTCTGTTCAACCTCTGTTAACCTCGACGCCATTAAACCAAAGCTTCCTTCTTCTTCCTCTTCT GATTCCATCCAACTTGTGGAACTTTGTCTTCCATCTTCTCCTGATCAGCTCCCTCCTCATCTTC ACACAACCAACGCCCTCCCCCCTCACCTCATGCCCACTCTCCACCAAGCCTTCTCCATGGCTGC CCAACACTTTGCTGCCATTTTACACACACTTGCTCCGCATCTCCTCATTTACGACTCTTTCCAA CCTTGGGCTCCTCAACTAGCTTCATCCCTCAACATTCCAGCCATCAACTTCAATACTACGGGAG CTTCAGTCCTGACCCGAATGCTTCACGCTACTCACTACCCAAGTTCTAAATTCCCAATTTCAGA GTTTGTTCTCCACGATTATTGGAAAGCCATGTACAGCGCCGCCGGTGGGGCTGTTACAAAAAAA GACCACAAAATTGGAGAAACACTTGCGAATTGCTTGCATGCTTCTTGTAGTGTAATTCTAATCA ATAGTTTCAGAGAGCTCGAGGAGAAATATATGGATTATCTCTCCGTTCTCTTGAACAAGAAAGT TGTTCCGGTTGGTCCTTTGGTTTACGAACCGAATCAAGACGGGGAAGATGAAGGTTATTCAAGC ATCAAAAATTGGCTTGACAAAAAGGAACCGTCCTCCACCGTCTTCGTTTCATTTGGAAGCGAAT ACTTCCCGTCAAAGGAAGAAATGGAAGAGATAGCCCATGGGTTAGAGGCGAGCGAGGTTCATTT CATCTGGGTCGTTAGGTTTCCTCAAGGAGACAACACCAGCGCCATTGAAGATGCCTTGCCGAAG GGGTTTCTGGAGAGGGTGGGAGAGAGAGGGATGGTGGTGAAGGGTTGGGCTCCCCAGGCGAAGA TACTGAAGCATTGGAGCACAGGGGGATTCGTGAGCCACTGTGGATGGAACTCGGTGATGGAAAG CATGATGTTTGGCGTTCCCATAATAGGGGTTCCGATGCATCTGGACCAGCCCTTTAACGCCGGA CTCGCGGAAGAAGCTGGCGTCGGCGTGGAAGCCAAGCGAGATTCGGACGGCAAAATTCAAAGAG AAGAAGTTGCAAAGTCGATCAAAGAAGTGGTGATTGAGAAAACCAGGGAAGACGTGAGGAAGAA AGCAAGAGAAATGGGTGAGATTTTGAGGAGTAAAGGAGATGAGAAAATTGATGAGTTGGTGGCT GAAATTTCTCTTTTGCGCAAAAAGGCCCCATGTTCAATTGCGGCCGCACTCGAGCACCACCACC ACCACCACTGA	408	
UDP-	MVQPRVLLFPFPALGHVKPFLSLAELLSDAGIDVVFLSTEYNHRRISNTEALASRFPTLHFETI	409	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
glycosyltransferases (339)	PDGLPPNESRALADGPLYFSMREGTKPRFRQLIQSLNDGRWPITCIITDIMLSSPIEVAEEFGI PVIAFCPCSARYLSIHFFIPKLVEEGQIPYADDDPIGEIQGVPLFEGLLRRNHLPGSWSDKSAD ISFSHGLINQTLAAGRASALILNTFDELEAPFLTHLSSIFNKIYTIGPLHALSKSRLGDSSSSA SALSGFWKEDRACMSWLDCQPPRSVVFVSFGSTMKMKADELREFWYGLVSSGKPFLCVLRSDVV SGGEAAELIEQMAEEEGAGGKLGMVVEWAAQEKVLSHPAVGGFLTHCGWNSTVESIAAGVPMMC WPILGDQPSNATWIDRVWKIGVERNNREWDRLTVEKMVRALMEGQKRVEIQRSMEKLSKLANEK VVRGGLSFDNLEVLVEDIKKLKPYKF		
UDP-glycosyltransferases (339) (gDNA)	ATGGTGCAACCTCGGGTACTGCTGTTTCCTTTCCCGGCACTGGGCCACGTGAAGCCCTTCTTAT CACTGGCGGAGCTGCTTTCCGACGCCGGCATAGACGTCGTCTTCCTCAGCACCGAGTATAACCA CCGTCGGATCTCCAACACTGAAGCCCTAGCCTCCCGCTTCCCGACGCTTCATTTCGAAACTATA CCGGATGGCCTGCCGCCTAATGAGTCGCGCGCTCTTGCCGACGGCCCACTGTATTTCTCCATGC GTGAGGGAACTAAACCGAGATTCCGGCAACTGATTCAATCTCTTAACGACGGTCGTTGGCCCAT CACCTGTATTATCACTGACATCATGTTATCTTCTCCGATTGAAGTAGCGGAAGAATTTGGGATT CCAGTAATTGCCTTCTGCCCCTGCAGTGCTCGCTACTTATCGATTCACTTTTTTATACCGAAGC TCGTTGAGGAAGGTCAAATTCCATACGCAGATGACGATCCGATTGGAGAGATCCAGGGGGTGCC CTTGTTCGAAGGTCTTTTGCGACGGAATCATTTGCCTGGTTCTTGGTCTGATAAATCTGCAGAT ATATCTTTCTCGCATGGCTTGATTAATCAGACCCTTGCAGCTGGTCGAGCCTCGGCTCTTATAC TCAACACCTTCGACGAGCTCGAAGCTCCATTTCTGACCCATCTCTCTTCCATTTTCAACAAAAT CTACACCATTGGACCCCTCCATGCTCTGTCCAAATCAAGGCTCGGCGACTCCTCCTCCTCCGCT TCTGCCCTCTCCGGATTCTGGAAAGAGGATAGAGCCTGCATGTCCTGGCTCGACTGTCAGCCGC CGAGATCTGTGGTTTTCGTCAGTTTCGGGAGTACGATGAAGATGAAAGCCGATGAATTGAGAGA GTTCTGGTATGGGTTGGTGAGCAGCGGGAAACCGTTCCTCTGCGTGTTGAGATCCGACGTTGTT TCCGGCGGAGAAGCGGCGGAATTGATCGAACAGATGGCGGAGGAGGAGGGAGCTGGAGGGAAGC TGGGAATGGTAGTGGAGTGGGCAGCGCAAGAGAAGGTCCTGAGCCACCCTGCCGTCGGTGGGTT TTTGACGCACTGCGGGTGGAACTCAACGGTGGAAAGCATTGCCGCGGGAGTTCCGATGATGTGC TGGCCGATTCTCGGCGACCAACCCAGCAACGCCACTTGGATCGACAGAGTGTGGAAAATTGGGG TTGAAAGGAACAATCGTGAATGGGACAGGTTGACGGTGGAGAAGATGGTGAGAGCATTGATGGA AGGCCAAAAGAGAGTGGAGATTCAGAGATCAATGGAGAAGCTTTCAAAGTTGGCAAATGAGAAG GTTGTCAGGGGTGGGTTGTCTTTTGATAACTTGGAAGTTCTCGTTGAAGACATCAAAAAATTGA AACCATATAAATTTTAA	410	
UDP-glycosyltransferases (330) (protein)	MDTRKRSIRILMFPWLAHGHISAFLELAKSLAKRNFVIYICSSQVNLNSISKNMSSKDSISVKL VELHIPTTILPPPYHTTNGLPPHLMSTLKRALDSARPAFSTLLQTLKPDLVLYDFLQSWASEEA ESQNIPAMVFLSTGAAAISFIMYHWFETRPEEYPFPAIYFREHEYDNFCRFKSSDSGTSDQLRV SDCVKRSHDLVLIKTFRELEGQYVDFLSDLTRKRFVPVGPLVQEVGCDMENEGNDIIEWLDGKD RRSTVFSSFGSEYFLSANEIEEIAYGLELSGLNFIWVVRFPHGDEKIKIEEKLPEGFLERVEGR GLVVEGWAQQRRILSHPSVGGFLSHCGWSSVMEGVYSGVPIIAVPMHLDQPFNARLVEAVGFGE EVVRSRQGNLDRGEVARVVKKLVMGKSGEGLRRRVEELSEKMREKGEEEIDSLVEELVTVVRRR ERSNLKSENSMKKLNVMDDGE	411	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
UDP-glycosyltransferases (330) (gDNA, native)	ATGGATACAAGAAAGAGAAGCATCAGGATTCTAATGTTCCCATGGCTTGCTCATGGCCATATCT CAGCATTCCTCGAGCTGGCGAAGTCACTTGCCAAAAGAAACTTCGTCATTTACATTTGTTCTTC ACAAGTAAATCTAAATTCCATCAGCAAGAACATGTCATCAAAAGACTCCATTTCCGTAAAACTT GTTGAGCTTCACATTCCCACCACCATACTTCCCCCTCCTTACCACACCACCAATGGCCTCCCAC CCCACCTCATGTCCACCCTCAAGAGAGCCCTCGACAGTGCCCGGCCCGCCTTCTCCACCCTCCT CCAAACCCTCAAGCCCGACTTGGTTTTATACGATTTCCTCCAGTCGTGGGCCTCGGAGGAGGCC GAGTCGCAGAATATACCAGCCATGGTGTTTCTGAGTACCGGAGCTGCAGCGATTTCTTTTATTA TGTACCATTGGTTTGAGACCAGACCGGAGGAGTACCCTTTTCCGGCTATATACTTCCGGGAACA CGAGTATGATAACTTCTGCCGTTTTAAGTCTTCCGACAGCGGTACTAGTGATCAATTGAGAGTC AGCGATTGCGTTAAACGGTCGCACGATTTGGTTCTGATCAAGACATTCCGTGAACTGGAAGGAC AATACGTAGATTTTCTCTCCGACTTGACTCGGAAGAGATTCGTACCAGTTGGCCCCCTTGTTCA GGAGGTAGGTTGTGATATGGAGAATGAAGGAAATGACATCATCGAATGGCTCGACGGGAAAGAC CGTCGTTCGACGGTTTTCTCCTCATTCGGGAGCGAGTACTTCTTGTCTGCCAATGAGATCGAAG AGATAGCTTATGGGCTGGAGCTAAGCGGGCTTAACTTCATCTGGGTTGTTAGGTTTCCTCATGG CGACGAGAAAATCAAGATTGAGGAGAAACTGCCGGAAGGGTTTCTTGAGAGAGTGGAAGGAAGA GGGTTGGTGGTGGAGGGATGGGCACAGCAGAGGAGAATATTGTCACATCCGAGTGTTGGAGGGT TTTTGAGCCACTGTGGGTGGAGTTCTGTGATGGAAGGGGTGTATTCCGGTGTGCCGATTATTGC CGTGCCGATGCATCTTGACCAGCCGTTCAATGCTAGGTTGGTGGAGGCGGTGGGGTTTGGGGAG GAGGTGGTGAGGAGTAGACAAGGAAATCTTGACAGAGGAGAGGTGGCGAGGGTGGTGAAGAAGC TGGTTATGGGGAAAAGTGGGGAGGGGTTACGGCGGAGGGTGGAGGAGTTGAGTGAGAAGATGAG AGAGAAAGGGGAGGAGGAGATTGATTCACTGGTGGAGGAATTGGTGACGGTGGTTAGGAGGAGA GAGAGATCGAATCTCAAGTCTGAGAATTCTATGAAGAAATTGAATGTGATGGATGATGGAGAAT AG	412	
UDP-glycosyltransferases (328) described in Itkin et al. (protein)	MDAAQQGDTTTILMLPWLGYGHLSAFLELAKSLSRRNFHIYFCSTSVNLDAIKPKLPSSFSDSI QFVELHLPSSPEFPPHLHTTNGLPPTLMPALHQAFSMAAQHFESILQTLAPHLLIYDSLQPWAP RVASSLKIPAINFNTTGVFVISQGXHPIHYPHSKFPFSEFVLHNHWKAMYSTADGASTERTRKR GEAFLYCLHASCSVILINSFRELEGKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIKNW LDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVNFIWVVRFPQGDNTSGIEDALPKGFLE RAGERGMVVKGWAPQAKILKHWSTGGFVSHCGWNSVMESMMFGVPIIGVPMHVDQPFNAGLVEE AGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEKFDEMVAEISL LLKI	413	
UDP-glycosyltransferases (328 (gDNA, native)	ATGGATGCTGCCCAACAAGGTGACACCACAACCATTTTGATGCTTCCATGGCTCGGCTATGGCC ATCTTTCAGCTTTTCTCGAGCTGGCCAAAAGCCTCTCAAGGAGGAACTTCCATATCTACTTCTG TTCAACCTCTGTTAATCTTGACGCCATTAAACCAAAGCTTCCTTCTTCTTTCTCTGATTCCATT CAATTTGTGGAGCTCCATCTCCCTTCTTCTCCTGAGTTCCCTCCTCATCTTCACACAACCAACG GCCTTCCCCCTACCCTCATGCCCGCTCTCCACCAAGCCTTCTCCATGGCTGCCCAGCACTTTGA GTCCATTTTACAAACACTTGCCCCGCACCTTCTCATTTATGACTCTCTTCAACCTTGGGCTCCT CGGGTAGCTTCATCCCTCAAAATTCCGGCCATCAACTTCAATACCACGGGAGTTTTCGTCATTT	414	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTCAAGGGYTTCACCCTATTCACTACCCACATTCTAAATTCCCATTCTCAGAGTTCGTTCTTCA CAATCATTGGAAAGCCATGTACTCCACTGCCGATGGAGCTTCTACCGAAAGAACCCGCAAACGT GGAGAAGCGTTTCTGTATTGCTTGCATGCTTCTTGTAGTGTAATTCTAATCAATAGTTTCAGAG AGCTCGAGGGGAAATATATGGATTATCTCTCTGTTCTCTTGAACAAGAAAGTTGTTCCGGTTGG TCCTTTGGTTTACGAACCGAATCAAGACGGGGAAGATGAAGGTTATTCAAGCATCAAAAATTGG CTTGACAAAAAGGAACCGTCCTCCACCGTCTTCGTGTCATTTGGAAGCGAATACTTCCCGTCAA AGGAAGAAATGGAAGAGATAGCCCATGGGTTAGAGGCGAGCGAGGTTAATTTCATCTGGGTCGT TAGGTTTCCTCAAGGAGACAACACCAGCGGCATTGAAGATGCCTTGCCGAAGGGTTTTCTGGAG AGGGCGGGAGAGAGAGGGATGGTGGTGAAGGGTTGGGCTCCTCAGGCGAAGATACTGAAGCATT GGAGCACAGGGGGATTCGTGAGCCACTGTGGATGGAACTCGGTGATGGAGAGCATGATGTTTGG CGTTCCCATAATAGGGGTTCCGATGCATGTGGACCAGCCCTTTAACGCCGGACTCGTGGAAGAA GCTGGCGTCGGCGTGGAGGCCAAGCGAGATCCAGACGGCAAAATTCAAAGAGACGAAGTTGCAA AGTTGATCAAAGAAGTGGTGGTTGAGAAAACCAGAGAAGATGTGCGGAAGAAAGCAAGAGAAAT GAGTGAGATTTTGAGGAGCAAGGGAGAGGAGAAGTTTGATGAGATGGTCGCTGAAATTTCTCTC TTGCTTAAAATATGA		
AtSus1 protein	MKHHHHHHQLHAGAHAAAGTMANAERMITRVHSQRERLNETLVSERNEVLALLSRVEAKGKGIL QQNQIIAEFEALPEQTRKKLEGGPFFDLLKSTQEAIVLPPWVALAVRPRPGVWEYLRVNLHALV VEELQPAEFLHFKEELVDGVKNGNFTLELDFEPFNASIPRPTLHKYIGNGVDFLNRHLSAKLFH DKESLLPLLKFLRLHSHQGKNLMLSEKIQNLNTLQHTLRKAEEYLAELKSETLYEEFEAKFEEI GLERGWGDNAERVLDMIRLLLDLLEAPDPCTLETFLGRVPMVFNVVILSPHGYFAQDNVLGYPD TGGQVVYILDQVRALEIEMLQRIKQQGLNIKPRILILTRLLPDAVGTTCGERLERVYDSEYCDI LRVPFRTEKGIVRKWISRFEVWPYLETYTEDAAVELSKELNGKPDLIIGNYSDGNLVASLLAHK LGVTQCTIAHALEKTKYPDSDIYWKKLDDKYHFSCQFTADIFAMNHTDFIITSTFQEIAGSKET VGQYESHTAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSIYFPYTEEKRRLTKFHSEIEELLYS DVENKEHLCVLKDKKKPILFTMARLDRVKNLSGLVEWYGKNTRLRELANLVVVGGDRRKESKDN EEKAEMKKMYDLIEEYKLNGQFRWISSQMDRVRNGELYRYICDTKGAFVQPALYEAFGLTVVEA MTCGLPTFATCKGGPAEIIVHGKSGFHIDPYHGDQAADTLADFFTKCKEDPSHWDEISKGGLQR IEEKYTWQIYSQRLLTLTGVYGFWKHVSNLDRLEARRYLEMFYALKYRPLAQAVPLAQDD	415	
AtSus1 (gDNA)	ATGAAACATCACCATCACCATCACCAGCTGCATGCGGGAGCTCATGCGGCCGCGGGTACCATGG CAAACGCTGAACGTATGATAACGCGCGTCCACAGCCAACGTGAGCGTTTGAACGAAACGCTTGT TTCTGAGAGAAACGAAGTCCTTGCCTTGCTTTCCAGGGTTGAAGCCAAAGGTAAAGGTATTTTA CAACAAAACCAGATCATTGCTGAATTCGAAGCTTTGCCTGAACAAACCCGGAAGAAACTTGAAG GTGGTCCTTTCTTTGACCTTCTCAAATCCACTCAGGAAGCAATTGTGTTGCCACCATGGGTTGC TCTAGCTGTGAGGCCAAGGCCTGGTGTTTGGGAATACTTACGAGTCAATCTCCATGCTCTTGTC GTTGAAGAACTCCAACCTGCTGAGTTTCTTCATTTCAAGGAAGAACTCGTTGATGGAGTTAAGA ATGGTAATTTCACTCTTGAGCTTGATTTCGAGCCATTCAATGCGTCTATCCCTCGTCCAACACT CCACAAATACATTGGAAATGGTGTTGACTTCCTTAACCGTCATTTATCGGCTAAGCTCTTCCAT GACAAGGAGAGTTTGCTTCCATTGCTTAAGTTCCTTCGTCTTCACAGCCACCAGGGCAAGAACC	416	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TGATGTTGAGCGAGAAGATTCAGAACCTCAACACTCTGCAACACACCTTGAGGAAAGCAGAAGA GTATCTAGCAGAGCTTAAGTCCGAAACACTGTATGAAGAGTTTGAGGCCAAGTTTGAGGAGATT GGTCTTGAGAGGGGATGGGGAGACAATGCAGAGCGTGTCCTTGACATGATACGTCTTCTTTTGG ACCTTCTTGAGGCGCCTGATCCTTGCACTCTTGAGACTTTTCTTGGAAGAGTACCAATGGTGTT CAACGTTGTGATCCTCTCTCCACATGGTTACTTTGCTCAGGACAATGTTCTTGGTTACCCTGAC ACTGGTGGACAGGTTGTTTACATTCTTGATCAAGTTCGTGCTCTGGAGATAGAGATGCTTCAAC GTATTAAGCAACAAGGACTCAACATTAAACCAAGGATTCTCATTCTAACTCGACTTCTACCTGA TGCGGTAGGAACTACATGCGGTGAACGTCTCGAGAGAGTTTATGATTCTGAGTACTGTGATATT CTTCGTGTGCCCTTCAGAACAGAGAAGGGTATTGTTCGCAAATGGATCTCAAGGTTCGAAGTCT GGCCATATCTAGAGACTTACACCGAGGATGCTGCGGTTGAGCTATCGAAAGAATTGAATGGCAA GCCTGACCTTATCATTGGTAACTACAGTGATGGAAATCTTGTTGCTTCTTTATTGGCTCACAAA CTTGGTGTCACTCAGTGTACCATTGCTCATGCTCTTGAGAAAACAAAGTACCCGGATTCTGATA TCTACTGGAAGAAGCTTGACGACAAGTACCATTTCTCATGCCAGTTCACTGCGGATATTTTCGC AATGAACCACACTGATTTCATCATCACTAGTACTTTCCAAGAAATTGCTGGAAGCAAAGAAACT GTTGGGCAGTATGAAAGCCACACAGCCTTTACTCTTCCCGGATTGTATCGAGTTGTTCACGGGA TTGATGTGTTTGATCCCAAGTTCAACATTGTCTCTCCTGGTGCTGATATGAGCATCTACTTCCC TTACACAGAGGAGAAGCGTAGATTGACTAAGTTCCACTCTGAGATCGAGGAGCTCCTCTACAGC GATGTTGAGAACAAAGAGCACTTATGTGTGCTCAAGGACAAGAAGAAGCCGATTCTCTTCACAA TGGCTAGGCTTGATCGTGTCAAGAACTTGTCAGGTCTTGTTGAGTGGTACGGGAAGAACACCCG CTTGCGTGAGCTAGCTAACTTGGTTGTTGTTGGAGGAGACAGGAGGAAAGAGTCAAAGGACAAT GAAGAGAAAGCAGAGATGAAGAAAATGTATGATCTCATTGAGGAATACAAGCTAAACGGTCAGT TCAGGTGGATCTCCTCTCAGATGGACCGGGTAAGGAACGGTGAGCTGTACCGGTACATCTGTGA CACCAAGGGTGCTTTTGTCCAACCTGCATTATATGAAGCCTTTGGGTTAACTGTTGTGGAGGCT ATGACTTGTGGTTTACCGACTTTCGCCACTTGCAAAGGTGGTCCAGCTGAGATCATTGTGCACG GTAAATCGGGTTTCCACATTGACCCTTACCATGGTGATCAGGCTGCTGATACTCTTGCTGATTT CTTCACCAAGTGTAAGGAGGATCCATCTCACTGGGATGAGATCTCAAAAGGAGGGCTTCAGAGG ATTGAGGAGAAATACACTTGGCAAATCTATTCACAGAGGCTCTTGACATTGACTGGTGTGTATG GATTCTGGAAGCATGTCTCGAACCTTGACCGTCTTGAGGCTCGCCGTTACCTTGAAATGTTCTA TGCATTGAAGTATCGCCCATTGGCTCAGGCTGTTCCTCTTGCACAAGATGATTGA		
SgCbQ protein	MWRLKVGAESVGENDEKWLKSISNHLGRQVWEFCPDAGTQQQLLQVHKARKAFHDDRFHRKQSS DLFITIQYGKEVENGGKTAGVKLKEGEEVRKEAVESSLERALSFYSSIQTSDGNWASDLGGPMF LLPGLVIALYVTGVLNSVLSKHHRQEMCRYVYNHQNEDGGWGLHIEGPSTMFGSALNYVALRLL GEDANAGAMPKARAWILDHGGATGITSWGKLWLSVLGVYEWSGNNPLPPEFWLFPYFLPFHPGR MWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYAVPYHEIDWNKSRNTCAKEDLYYPHPKM QDILWGSLHHVYEPLFTRWPAKRLREKALQTAMQHIHYEDENTRYICLGPVNKVLNLLCCWVED PYSDAFKLHLQRVHDYLWVAEDGMKMQGYNGSQLWDTAFSIQAIVSTKLVDNYGPTLRKAHDFV KSSQIQQDCPGDPNVWYRHIHKGAWPFSTRDHGWLISDCTAEGLKAALMLSKLPSETVGESLER NRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTSATMEALTLF	417	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	KKLHPGHRTKEIDTAIVRAANFLENMQRTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCLA IRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAERDPTPLH RAARLLINSQLENGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE		
SgCbQ (gDNA)	ATGTGGAGGTTAAAGGTCGGAGCAGAAAGCGTTGGGGAGAATGATGAGAAATGGTTGAAGAGCA TAAGCAATCACTTGGGACGCCAGGTGTGGGAGTTCTGTCCGGATGCCGGCACCCAACAACAGCT CTTGCAAGTCCACAAAGCTCGTAAAGCTTTCCACGATGACCGTTTCCACCGAAAGCAATCTTCC GATCTCTTTATCACTATTCAGTATGGAAAGGAAGTAGAAAATGGTGGAAAGACAGCGGGAGTGA AATTGAAAGAAGGGGAAGAGGTGAGGAAAGAGGCAGTAGAGAGTAGCTTAGAGAGGGCATTAAG TTTCTACTCAAGCATCCAGACAAGCGATGGGAACTGGGCTTCGGATCTTGGGGGGCCCATGTTT TTACTTCCGGGTCTGGTGATTGCCCTCTACGTTACAGGCGTCTTGAATTCTGTTTTATCCAAGC ACCACCGGCAAGAGATGTGCAGATATGTTTACAATCACCAGAATGAAGATGGGGGGTGGGGTCT CCACATCGAGGGCCCAAGCACCATGTTTGGTTCCGCACTGAATTATGTTGCACTCAGGCTGCTT GGAGAAGACGCCAACGCCGGGGCAATGCCAAAAGCACGTGCTTGGATCTTGGACCACGGTGGCG CCACCGGAATCACTTCCTGGGGCAAATTGTGGCTTTCTGTACTTGGAGTCTACGAATGGAGTGG CAATAATCCTCTTCCACCCGAATTTTGGTTATTTCCTTACTTCCTACCATTTCATCCAGGAAGA ATGTGGTGCCATTGTCGAATGGTTTATCTACCAATGTCATACTTATATGGAAAGAGATTTGTTG GGCCAATCACACCCATAGTTCTGTCTCTCAGAAAAGAACTCTACGCAGTTCCATATCATGAAAT AGACTGGAATAAATCTCGCAATACATGTGCAAAGGAGGATCTGTACTATCCACATCCCAAGATG CAAGATATTCTGTGGGGATCTCTCCACCACGTGTATGAGCCCTTGTTTACTCGTTGGCCTGCCA AACGCCTGAGAGAAAAGGCTTTGCAGACTGCAATGCAACATATTCACTATGAAGATGAGAATAC CCGATATATATGCCTTGGCCCTGTCAACAAGGTACTCAATCTGCTTTGTTGTTGGGTTGAAGAT CCCTACTCCGACGCCTTCAAACTTCATCTTCAACGAGTCCATGACTATCTCTGGGTTGCTGAAG ATGGCATGAAAATGCAGGGTTATAATGGGAGCCAGTTGTGGGACACTGCTTTCTCCATCCAAGC AATCGTATCCACCAAACTTGTAGACAACTATGGCCCAACCTTAAGAAAGGCACACGACTTCGTT AAAAGTTCTCAGATTCAGCAGGACTGTCCTGGGGATCCTAATGTTTGGTACCGTCACATTCATA AAGGTGCATGGCCATTTTCAACTCGAGATCATGGATGGCTCATCTCTGACTGTACAGCAGAGGG ATTAAAGGCTGCTTTGATGTTATCCAAACTTCCATCCGAAACAGTTGGGGAATCATTAGAACGG AATCGCCTTTGCGATGCTGTAAACGTTCTCCTTTCTTTGCAAAACGATAATGGTGGCTTTGCAT CATATGAGTTGACAAGATCATACCCTTGGTTGGAGTTGATCAACCCCGCAGAAACGTTTGGAGA TATTGTCATTGATTATCCGTATGTGGAGTGCACCTCAGCCACAATGGAAGCACTGACGTTGTTT AAGAAATTACATCCCGGCCATAGGACCAAAGAAATTGATACTGCTATTGTCAGGGCGGCCAACT TCCTTGAAAATATGCAAAGGACGGATGGCTCTTGGTATGGATGTTGGGGGGTTTGCTTCACGTA TGCGGGGTGGTTTGGCATAAAGGGATTGGTGGCTGCAGGAAGGACATATAATAATTGCCTTGCC ATTCGCAAGGCTTGCGATTTTTTACTATCTAAAGAGCTGCCCGGCGGTGGATGGGGAGAGAGTT ACCTTTCATGTCAGAATAAGGTATACACAAATCTTGAAGGAAACAGACCGCACCTGGTTAACAC GGCCTGGGTTTTAATGGCCCTCATAGAAGCTGGCCAGGCTGAGAGAGACCCAACACCATTGCAT CGTGCAGCAAGGTTGTTAATCAATTCCCAGTTGGAGAATGGTGATTTCCCCCAACAGGAGATCA TGGGAGTCTTTAATAAAAATTGCATGATCACATATGCTGCATACCGAAACATTTTTCCCATTTG	418	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GGCTCTTGGAGAGTATTGCCATCGGGTTTTGACTGAATAA		
	ATGTGGAGGTTAAAGGTCGGAGCAGAAAGCGTTGGGGAGAATGATGAGAAATGGTTGAAGAGCA TAAGCAATCACTTGGGACGCCAGGTGTGGGAGTTCTGTCCGGATGCCGGCACCCAACAACAGCT CTTGCAAGTCCACAAAGCTCGTAAAGCTTTCCACGATGACCGTTTCCACCGAAAGCAATCTTCC GATCTCTTTATCACTATTCAGTATGGAAAGGAAGTAGAAAATGGTGGAAAGACAGCGGGAGTGA AATTGAAAGAAGGGGAAGAGGTGAGGAAAGAGGCAGTAGAGAGTAGCTTAGAGAGGGCATTAAG TTTCTACTCAAGCATCCAGACAAGCGATGGGAACTGGGCTTCGGATCTTGGGGGGCCCATGTTT TTACTTCCGGGTCTGGTGATTGCCCTCTACGTTACAGGCGTCTTGAATTCTGTTTTATCCAAGC ACCACCGGCAAGAGATGTGCAGATATGTTTACAATCACCAGAATGAAGATGGGGGGTGGGGTCT CCACATCGAGGGCCCAAGCACCATGTTTGGTTCCGCACTGAATTATGTTGCACTCAGGCTGCTT GGAGAAGACGCCAACGCCGGGGCAATGCCAAAAGCACGTGCTTGGATCTTGGACCACGGTGGCG CCACCGGAATCACTTCCTGGGGCAAATTGTGGCTTTCTGTACTTGGAGTCTACGAATGGAGTGG CAATAATCCTCTTCCACCCGAATTTTGGTTATTTCCTTACTTCCTACCATTTCATCCAGGAAGA ATGTGGTGCCATTGTCGAATGGTTTATCTACCAATGTCATACTTATATGGAAAGAGATTTGTTG GGCCAATCACACCCATAGTTCTGTCTCTCAGAAAAGAACTCTACGCAGTTCCATATCATGAAAT AGACTGGAATAAATCTCGCAATACATGTGCAAAGGAGGATCTGTACTATCCACATCCCAAGATG CAAGATATTCTGTGGGGATCTCTCCACCACGTGTATGAGCCCTTGTTTACTCGTTGGCCTGCCA AACGCCTGAGAGAAAAGGCTTTGCAGACTGCAATGCAACATATTCACTATGAAGATGAGAATAC CCGATATATATGCCTTGGCCCTGTCAACAAGGTACTCAATCTGCTTTGTTGTTGGGTTGAAGAT CCCTACTCCGACGCCTTCAAACTTCATCTTCAACGAGTCCATGACTATCTCTGGGTTGCTGAAG ATGGCATGAAAATGCAGGGTTATAATGGGAGCCAGTTGTGGGACACTGCTTTCTCCATCCAAGC AATCGTATCCACCAAACTTGTAGACAACTATGGCCCAACCTTAAGAAAGGCACACGACTTCGTT AAAAGTTCTCAGATTCAGCAGGACTGTCCTGGGGATCCTAATGTTTGGTACCGTCACATTCATA AAGGTGCATGGCCATTTTCAACTCGAGATCATGGATGGCTCATCTCTGACTGTACAGCAGAGGG ATTAAAGGCTGCTTTGATGTTATCCAAACTTCCATCCGAAACAGTTGGGGAATCATTAGAACGG AATCGCCTTTGCGATGCTGTAAACGTTCTCCTTTCTTTGCAAAACGATAATGGTGGCTTTGCAT CATATGAGTTGACAAGATCATACCCTTGGTTGGAGTTGATCAACCCCGCAGAAACGTTTGGAGA TATTGTCATTGATTATCCGTATGTGGAGTGCACCTCAGCCACAATGGAAGCACTGACGTTGTTT AAGAAATTACATCCCGGCCATAGGACCAAAGAAATTGATACTGCTATTGTCAGGGCGGCCAACT TCCTTGAAAATATGCAAAGGACGGATGGCTCTTGGTATGGATGTTGGGGGGTTTGCTTCACGTA TGCGGGGTGGTTTGGCATAAAGGGATTGGTGGCTGCAGGAAGGACATATAATAATTGCCTTGCC ATTCGCAAGGCTTGCGATTTTTTACTATCTAAAGAGCTGCCCGGCGGTGGATGGGGAGAGAGTT ACCTTTCATGTCAGAATAAGGTATACACAAATCTTGAAGGAAACAGACCGCACCTGGTTAACAC GGCCTGGGTTTTAATGGCCCTCATAGAAGCTGGCCAGGCTGAGAGAGACCCAACACCATTGCAT CGTGCAGCAAGGTTGTTAATCAATTCCCAGTTGGAGAATGGTGATTTCCCCCAACAGGAGATCA TGGGAGTCTTTAATAAAAATTGCATGATCACATATGCTGCATACCGAAACATTTTTCCCATTTG GGCTCTTGGAGAGTATTGCCATCGGGTTTTGACTGAATAA	419	
Cpep2 protein	MWRLKVGAESVGEKDEKWVKSVSNHLGRQVWEFCAADAAAVTPHQLLQIQNARNHFHRNRFHRK	420	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	QSSDLFLAIQYEKEIAKGGKGKEAVKVKEGEEVGKEAVKSTLERALSFYTAVQTSDGNWASDLG GPMFLLPGLVIALYVTGVLNSVLSKHHRVEMCRYIYNHQNEDGGWGLHIEGTSTMFGSALNYVA LRLLGEDADGGDDGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYS LPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTVPYHEIDWNKSRNTCAKEDL YYPHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQTAMKHIHYEDENSRYICLGPVNKVLNM LCCWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTL RKAHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSTM VGEPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTAAT MEALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGR TYNSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGE RDPAPLHRGARLVMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE		
Cpep2 gene sequence	ATGTGGAGGCTGAAGGTGGGAGCAGAGAGCGTTGGGGAGAAGGATGAGAAATGGGTGAAGAGCG TAAGCAATCACTTGGGCCGCCAAGTTTGGGAGTTCTGTGCCGCCGACGCCGCCGCCGTCACTCC TCACCAGTTACTACAAATTCAGAATGCTCGCAACCACTTCCATCGCAATCGTTTCCACCGGAAG CAGTCTTCCGATCTCTTTCTCGCTATTCAGTATGAAAAGGAAATAGCGAAGGGCGGAAAAGGGA AAGAGGCGGTGAAAGTGAAAGAAGGGGAGGAGGTGGGGAAAGAGGCGGTGAAGAGTACGTTAGA GAGGGCACTAAGTTTCTACACAGCCGTGCAGACGAGCGATGGGAATTGGGCCTCGGATCTTGGA GGGCCCATGTTTTTACTTCCGGGTCTCGTGATTGCCCTTTATGTCACAGGCGTGTTGAATTCAG TTTTGTCCAAGCACCACCGCGTAGAGATGTGCAGATATATTTACAATCACCAGAATGAAGATGG AGGGTGGGGTCTACATATTGAGGGCACAAGCACCATGTTTGGTTCGGCACTCAATTATGTTGCA CTTAGGCTGCTTGGAGAAGACGCCGATGGCGGAGACGATGGTGCAATGACAAAAGCACGTGCTT GGATCTTGGAGCGCGGCGGCGCCACTGCGATCACTTCGTGGGGAAAATTGTGGCTGTCCGTGCT TGGAGTGTACGAATGGAGTGGCAACAACCCTCTTCCGCCTGAGTTTTGGCTTCTCCCTTACAGC CTACCATTTCATCCAGGACGAATGTGGTGCCATTGTCGAATGGTTTATCTTCCCATGTCTTACT TATATGGGAAGAGATTTGTTGGCCCAATCACTCCCAAAGTTCTTTCTCTAAGACAAGAGCTCTA CACGGTTCCTTATCATGAAATAGACTGGAATAAATCCCGCAATACATGTGCAAAGGAGGATCTA TACTATCCACATCCCAAGATGCAAGACATACTATGGGGATCTATCTACCATGTATATGAGCCAT TGTTCACTCGTTGGCCTGGGAAACGCCTGAGGGAAAAGGCTTTACAAACTGCAATGAAACATAT TCACTATGAAGATGAAAATAGTCGCTATATATGTCTTGGCCCAGTCAACAAGGTACTCAACATG CTTTGTTGTTGGGTTGAAGATCCCTACTCAGACGCCTTCAAACTTCACCTTCAACGCGTCCATG ACTATCTCTGGGTTGCTGAAGATGGCATGAGAATGCAGGGTTACAATGGCAGCCAGTTGTGGGA CACTGCTTTCTCCATCCAAGCCATTGTAGCTACCAAACTTGTAGACAGCTATGCCCCAACTTTA AGAAAAGCACATGACTTTGTTAAGGATTCTCAGATCCAGGAGGACTGTCCTGGGGATCCTAATG TTTGGTTCCGTCATATTCATAAAGGTGCTTGGCCATTTTCGACTCGAGATCATGGATGGCTCAT CTCTGACTGCACGGCTGAGGGATTGAAGGCTTCTTTGATGTTATCCAAACTTCCATCCACAATG GTTGGGGAGCCATTAGAAAAGAATCGCCTTTGTGATGCTGTTAATGTTCTCCTTTCTTTGCAAA ATGATAACGGTGGATTTGCATCATACGAGTTGACGAGATCATACCCTTGGTTGGAGTTGATCAA CCCAGCAGAAACATTCGGAGACATTGTCATCGACTATCCGTATGTGGAGTGCACCGCAGCAACA	421	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ATGGAAGCACTGACGTTATTTAAGAAGCTACATCCAGGCCATAGGACCAAAGAGATTGACACAG CTATTGGCAAGGCAGCCAACTTCCTTGAGAAAATGCAAAGGGCGGATGGCTCTTGGTATGGGTG TTGGGGGGTTTGTTTCACGTATGCGGGGTGGTTTGGCATCAAGGGATTGGTGGCTGCAGGAAGA ACATATAATAGCTGCCTTGCCATCCGCAAGGCTTGTGAGTTTCTGCTATCTAAAGAGCTGCCCG GCGGTGGATGGGGGGAGAGTTACCTTTCATGTCAGAATAAGGTGTACACCAATCTTGAGGGAAA CAAGCCACACTTGGTTAACACTGCCTGGGTTTTAATGGCTCTCATTGAAGCCGGCCAGGGTGAG AGAGACCCAGCACCATTGCACCGTGGAGCAAGGTTGGTAATGAATTCTCAACTGGAGAATGGTG ATTTCGTGCAACAGGAGATCATGGGAGTGTTCAATAAGAACTGCATGATCACATATGCTGCATA CCGAAACATCTTCCCCATTTGGGCGCTTGGAGAGTATTGCCATCGGGTTCTTACTGAATGA		
Cpep4 protein	MWRLKVGAESVGEKDEKWVKSVSNHLGRQVWEFCAADAAAVTPHQLLQIQNARNHFHRNRFHRK QSSDLFLAIQYEKEIAKGGKGKEAVKVKEGEEVGKEAVKSTLERALSFYTAVQTSDGNWASDLG GPMFLLPGLVIALYVTGVLNSVLSKHHRVEMCRYIYNHQNEDGGWGLHIEGTSTMFGSALNYVA LRLLGEDADGGDDGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFLLLPYS LPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTVPYHEIDWNKSRNTCAKEDL YYPHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQTAMKHIHYEDENSRYICLGPVNKVLNM LCCWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTL RKAHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSTM VGEPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYSYVECTAAT MEALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGR TYNSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGE RDPAPLHRAARLVMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE.	422	
Cpep4 gene sequence	ATGTGGAGGCTGAAGGTGGGAGCAGAGAGCGTTGGGGAGAAGGATGAGAAATGGGTGAAGAGCG TAAGCAATCACTTGGGCCGCCAAGTTTGGGAGTTCTGTGCCGCCGACGCCGCCGCCGTCACTCC TCACCAGTTACTACAAATTCAGAATGCTCGCAACCACTTCCATCGCAATCGTTTCCACCGGAAG CAGTCTTCCGATCTCTTTCTCGCTATTCAGTATGAAAAGGAAATAGCGAAGGGCGGAAAAGGGA AAGAGGCGGTGAAAGTGAAAGAAGGGGAGGAGGTGGGGAAAGAGGCGGTGAAGAGTACGTTAGA GAGGGCACTAAGTTTCTACACAGCCGTGCAGACGAGCGATGGGAATTGGGCCTCGGATCTTGGA GGGCCCATGTTTTTACTTCCGGGTCTCGTGATTGCCCTTTATGTCACAGGCGTGTTGAATTCAG TTTTGTCCAAGCACCACCGCGTAGAGATGTGCAGATATATTTACAATCACCAGAATGAAGATGG AGGGTGGGGTCTACATATTGAGGGCACAAGCACCATGTTTGGTTCGGCACTCAATTATGTTGCA CTTAGGCTGCTTGGAGAAGACGCCGATGGCGGAGACGATGGTGCAATGACAAAAGCACGTGCTT GGATCTTGGAGCGCGGCGGCGCCACTGCGATCACTTCGTGGGGAAAATTGTGGCTGTCCGTGCT TGGAGTGTACGAATGGAGTGGCAACAACCCTCTTCCGCCTGAGTTTTTGCTTCTCCCTTACAGC CTACCATTTCATCCAGGACGAATGTGGTGCCATTGTCGAATGGTTTATCTTCCCATGTCTTACT TATATGGGAAGAGATCTGTTCGCCCAATCACTCCCAAAGTTCTTTCTCTAAGACAAGAGCTCTA CACGGTTCCTTATCATGAAATAGACTGGAATAAATCCCGCAATACATGTGCAAAGGAGGATCTA TACTATCCACATCCCAAGATGCAAGACATACTATGGGGATCTATCTACCATGTATATGAGCCAT TGTTCACTCGTTGGCCTGGGAAACGCCTGAGGGAAAAGGCTTTACAAACTGCAATGAAACATAT	423	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TCACTATGAAGATGAAAATAGTCGCTATATATGTCTTGGCCCAGTCAACAAGGTACTCAACATG CTTTGTTGTTGGGTTGAAGATCCCTACTCAGACGCCTTCAAACTTCACCTTCAACGCGTCCATG ACTATCTCTGGGTTGCTGAAGATGGCATGAGAATGCAGGGTTACAATGGCAGCCAGTTGTGGGA CACTGCTTTCTCCATCCAAGCCATTGTAGCTACCAAACTTGTAGACAGCTATGCCCCAACTTTA AGAAAAGCACATGACTTTGTTAAGGATTCTCAGATCCAGGAGGACTGTCCTGGGGATCCTAATG TTTGGTTCCGTCATATTCATAAAGGTGCTTGGCCATTTTCGACTCGAGATCATGGATGGCTCAT CTCTGACTGCACGGCTGAGGGATTGAAGGCTTCTTTGATGTTATCCAAACTTCCATCCACAATG GTTGGGGAGCCATTAGAAAAGAATCGCCTTTGTGATGCTGTTAATGTTCTCCTTTCTTTGCAAA ATGATAACGGCGGATTTGCATCATACGAGTTGACGAGATCATACCCTTGGTTGGAGTTGATCAA CCCAGCAGAAACATTCGGAGACATTGTCATCGACTATTCGTATGTGGAGTGCACCGCAGCAACA ATGGAAGCACTGACGTTATTTAAGAAGCTACATCCAGGCCATAGGACCAAAGAGATTGACACAG CTATTGGCAAGGCAGCCAACTTCCTTGAGAAAATGCAAAGGGCGGATGGCTCTTGGTATGGGTG TTGGGGGGTTTGTTTCACGTATGCGGGGTGGTTTGGCATAAAGGGATTGGTGGCTGCAGGAAGA ACATATAATAGCTGTCTTGCCATCCGCAAGGCTTGTGAGTTTCTGCTATCTAAAGAGCTGCCCG GCGGTGGATGGGGGGGAGAGTTACCTTTCATGTCAGAATAAGGTGTACACCAATCTTGAGGGAAA CAAGCCACACTTGGTTAACACTGCCTGGGTTTTAATGGCTCTCATTGAAGCTGGCCAGGGTGAG AGAGACCCAGCACCATTGCACCGTGCAGCAAGGTTGGTAATGAATTCTCAACTGGAGAATGGCG ATTTCGTGCAACAGGAGATCATGGGAGTGTTCAATAAGAACTGCATGATCACATATGCTGCATA CCGAAACATCTTCCCCATTTGGGCGCTTGGAGAGTATTGCCATCGGGTTCTTACTGAATGA		
Cmax1 protein	MWRLKVGAESVGEKDEKWVKSVSNHLGRQVWEFCADAAADTPHQLLQIQNARNHFHHNRFHRKQ SSDLFLAIQYEKEIAKGAKGGAVKVKEGEEVGKEAVKSTLESALGFYSAVQTSDGNWASDLGGP MFLLPGLVIALHVTGVLNSVLSKHHRVEMCRYLYNHQNEDGGWGLHIEGTSTMFGSALNYVALR LLGEDADGGDGGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLP FHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTIPYHEIDWNKSRNTCAKEDLYY PHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQAAMKHIHYEDENSRYICLGPVNKVLNMLC CWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTLRK AHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSAMVG EPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTAATME ALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGRTY NSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGERD PAPLHRAARLLMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE	424	
Cmax1 gene sequence	ATGTGGAGGCTGAAGGTGGGAGCAGAGAGCGTTGGGGAGGAGGATGAGAAATGGGTGAAGAGCG TAAGCAATCACTTGGGCCGCCAAGTTTGGGAGTTCTGTGCCGACGCCGCCGCCGACACTCCTCA CCAGTTACTACAAATTCAGAATGCTCGCAACCACTTCCATCACAATCGTTTCCACCGGAAGCAG TCTTCCGATCTCTTTCTGGCTATTCAATATGAAAAGGAAATAGCAAAGGGCGCAAAAGGTGGAG CGGTGAAAGTGAAAGAAGGGGAGGAGGTGGGGAAAGAGGCGGTGAAGAGTACGTTAGAAAGGGC ACTCGGTTTCTACTCGGCCGTGCAGACAAGAGATGGGAATTGGGCCTCGGATCTTGGAGGGCCC TTGTTTTTACTTCCGGGTCTCGTGATTGCCCTTCATGTCACAGGCGTCTTGAATTCAGTTTTGT	425	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CCAAGCACCACCGCGTAGAGATGTGCAGATATCTTTACAATCACCAGAATGAAGATGGAGGGTG GGGTCTACATATTGAGGGCACAAGCACCATGTTTGGTTCGGCACTGAATTACGTTGCACTAAGG CTGCTTGGAGAAGACGCCGATGGCGGAGACGGTGGCGCAATGACAAAAGCACGTGCTTGGATCT TGGAGCGCGGCGGCGCCACTGCGATCACTTCGTGGGGAAAATTGTGGCTGTCCGTACTTGGAGT GTACGAATGGAGTGGCAACAACCCTCTTCCGCCTGAGTTTTGGCTTCTCCCTTACAGCCTACCA TTTCATCCAGGAAGAATGTGGTGCCATTGTCGAATGGTTTATCTTCCAATGTCTTACTTATATG GGAAGAGATTTGTTGGGCCAATCACTCCCAAAGTTCTTTCTCTAAGGCAAGAGCTCTACACAAT TCCTTATCATGAAATAGACTGGAATAAATCCCGCAATACATGTGCAAAGGAGGATCTGTACTAT CCACATCCCAAGATGCAAGACATTCTATGGGGATCCATCTACCATGTATATGAGCCATTGTTCA CTCGTTGGCCTGGGAAACGCCTGAGGGAAAAGGCTTTACAAGCTGCAATGAAACATATTCACTA TGAAGATGAAAATAGTCGATATATATGTCTTGGCCCAGTCAACAAGGTACTCAACATGCTTTGT TGTTGGGTTGAAGATCCCTACTCAGACGCCTTCAAACTTCACCTTCAACGCGTCCATGACTATC TCTGGGTTGCTGAAGATGGCATGAGAATGCAGGGCTACAATGGCAGCCAGTTGTGGGACACTGC TTTCTCCATCCAAGCCATCGTAGCCACCAAACTTGTAGACAGCTATGCCCCAACTTTAAGAAAA GCACATGACTTTGTTAAGGATTCTCAGATCCAGGAGGACTGTCCTGGGGATCCTAATGTTTGGT TCCGTCATATTCATAAAGGTGCTTGGCCACTTTCGACACGAGATCATGGATGGCTCATCTCCGA CTGTACAGCTGAGGGATTGAAGGCTTCTTTGATGTTATCCAAACTTCCATCCACAATGGTTGGG GAGCCATTAGAAAAGAATCGCCTTTGTGATGCTGTTAATGTTCTCCTTTCTTTGCAAAATGATA ATGGTGGATTTGCATCATACGAGTTGACGAGATCATACCCTTGGTTGGAGTTGATCAACCCAGC TGAAACATTCGGAGACATTGTCATTGACTATCCGTATGTGGAGTGCACCGCAGCAACAATGGAA GCACTGACGTTATTTAAGAAGCTACATCCAGGCCATAGGACCAAAGAGATTGACACAGCTATTG GCAAGGCAGCCAACTTCCTTGAGAAAATGCAGAGGGCGGATGGCTCTTGGTACGGGTGTTGGGG GGTTTGTTTTACGTATGCGGGTTGGTTTGGCATAAAGGGATTGGTGGCTGCAGGAAGAACATAT AATAGCTGCCTTGCCATTCGCAAGGCTTGTGAGTTTCTGCTATCTAAAGAGCTGCCCGGCGGTG GATGGGGGGAGAGTTACCTTTCATGTCAGAATAAGGTGTACACCAATCTTGAGGGGAACAAGCC ACACTTGGTTAACACTGCCTGGGTTTTAATGGCTCTCATTGAAGCTGGCCAGGGTGAGAGAGAC CCAGCACCATTGCACCGTGCAGCAAGGTTGCTAATGAATTCCCAATTGGAGAATGGCGATTTCG TGCAACAGGAGATCATGGGAGTGTTCAATAAGAACTGCATGATCACATATGCTGCATACCGAAA CATCTTCCCCATTTGGGCGCTTGGAGAGTATTGCCATCGGGTTCTTACTGAATGA		
Cmos1 protein	MWRLKVGAESVGEKDEKWVKSVSNHLGRQVWEFCADAAAAATPRQLLQIQNARNHFHRNRFHRK QSSDLFLAIQYEKEIAEGGKGGAVKVKEEEEVGKEAVKSTLERALSFYSAVQTSDGNWASDLGG PMFLLPGLVIALYVTGVLNSVLSKHHRVEMCRYLYNHQNEDGGWGLHIEGTSTMFGSALNYVAL RLLGEDADGGDDGAMTKARAWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSL PFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPKVLSLRQELYTVPYHEIDWNKSRNTCAKEDLY YPHPKMQDILWGSIYHVYEPLFTRWPGKRLREKALQTAMKHIHYEDENSRYICLGPVNKVLNML CCWVEDPYSDAFKLHLQRVHDYLWVAEDGMRMQGYNGSQLWDTAFSIQAIVATKLVDSYAPTLR KAHDFVKDSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSAMV GEPLEKNRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTAATM	426	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	EALTLFKKLHPGHRTKEIDTAIGKAANFLEKMQRADGSWYGCWGVCFTYAGWFGIKGLVAAGRT YNSCLAIRKACEFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVLMALIEAGQGER DPAPLHRAARLLMNSQLENGDFVQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE		
Cmos1 gene sequence	ATGTGGAGGTTGAAGGTGGGAGCAGAGAGCGTTGGGGAGAAGGATGAGAAATGGGTGAAGAGCG TAAGCAATCACTTGGGCCGCCAAGTTTGGGAGTTCTGTGCCGACGCCGCCGCCGCCGCCACTCC TCGCCAGTTACTACAAATTCAGAATGCTCGCAACCACTTCCATCGCAATCGTTTCCACCGGAAG CAGTCTTCCGATCTCTTTCTCGCTATTCAGTATGAAAAGGAAATAGCAGAGGGCGGAAAAGGTG GAGCGGTGAAAGTGAAAGAAGAGGAGGAGGTGGGGAAAGAGGCGGTGAAGAGTACGTTAGAAAG GGCACTAAGTTTCTACTCAGCCGTGCAGACAAGCGATGGGAATTGGGCCTCGGATCTTGGAGGG CCCATGTTTTTACTTCCGGGTCTCGTGATTGCCCTTTATGTCACAGGCGTGTTGAATTCAGTTT TGTCCAAGCACCACCGCGTAGAGATGTGCAGATATCTTTACAATCACCAGAATGAAGATGGAGG GTGGGGTCTACATATTGAGGGCACAAGCACCATGTTTGGTTCGGCACTCAATTACGTTGCACTA AGGCTGCTTGGAGAAGACGCGGATGGCGGAGACGATGGCGCAATGACAAAAGCACGTGCTTGGA TCTTGGAGCGCGGCGGCGCCACTGCGATCACTTCGTGGGGAAAGTTGTGGCTGTCCGTGCTTGG AGTGTACGAATGGAGTGGCAACAACCCTCTTCCGCCTGAGTTTTGGCTTCTCCCTTACAGCCTA CCATTTCATCCAGGAAGAATGTGGTGCCATTGTCGAATGGTTTATCTTCCCATGTCTTACTTAT ATGGGAAGAGATTTGTTGGGCCAATCACTCCCAAAGTTCTATCGCTAAGACAAGAGCTTTACAC GGTTCCTTATCATGAAATAGACTGGAACAAATCCCGCAATACATGTGCAAAGGAGGATCTATAC TATCCACATCCCAAGATGCAAGACATTCTATGGGGATCCATCTACCATGTGTATGAGCCATTGT TCACTCGTTGGCCTGGGAAACGCCTGAGGGAAAAGGCTTTACAAACTGCAATGAAACATATTCA CTATGAAGATGAAAATAGTCGATATATATGTCTTGGCCCAGTCAACAAGGTACTCAACATGCTT TGTTGTTGGGTTGAAGATCCCTACTCAGACGCCTTCAAACTTCACCTTCAACGCGTCCATGACT ATCTCTGGGTTGCTGAAGATGGCATGAGAATGCAGGGCTACAATGGCAGCCAGTTGTGGGACAC TGCTTTCTCCATCCAAGCCATCGTAGCCACCAAACTTGTAGACAGCTATGCCCCAACTTTAAGA AAAGCACATGACTTTGTTAAGGATTCTCAGATCCAGGAGGACTGTCCTGGGGATCCTAATGTTT GGTTCCGTCATATTCATAAAGGTGCTTGGCCATTTTCGACTCGAGATCATGGATGGCTCATCTC CGACTGTACAGCTGAGGGATTGAAGGCTTCTTTGATGTTATCCAAACTTCCATCCGCAATGGTT GGGGAGCCATTAGAAAAGAATCGCCTTTGTGATGCTGTTAATGTTCTCCTTTCTTTGCAAAATG ATAATGGTGGATTTGCATCATACGAGTTGACGAGATCATACCCTTGGTTGGAGTTGATCAACCC AGCAGAAACATTCGGAGACATTGTCATCGACTATCCGTATGTGGAGTGCACCGCAGCAACAATG GAAGCACTGACGTTATTTAAGAAGCTACATCCAGGCCATAGGACCAAAGAGATTGACACAGCTA TTGGCAAGGCAGCCAACTTCCTTGAGAAAATGCAGAGGGCGGATGGCTCTTGGTATGGGTGTTG GGGGGTTTGTTTCACGTATGCGGGGTGGTTTGGCATAAAGGGATTGGTGGCTGCAGGAAGAACA TATAATAGCTGCCTTGCCATCCGCAAGGCTTGTGAGTTTCTGCTATCTAAAGAGCTGCCCGGCG GTGGATGGGGGGAGAGTTACCTTTCATGTCAGAATAAGGTGTACACCAATCTTGAGGGAAACAA GCCACACTTGGTTAACACTGCCTGGGTTTTAATGGCTCTCATTGAAGCTGGCCAGGGTGAGAGA GACCCAGCACCATTGCACCGTGCAGCAAGGTTGCTAATGAATTCCCAATTGGAGAATGGCGATT TCGTGCAACAGGAGATCATGGGAGTGTTCAATAAGAACTGCATGATCACATATGCTGCATACCG	427	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAACATCTTCCCCATTTGGGCGCTTGGAGAGTATTGCCATCGGGTTCTGACTGAAT		
EPH protein	MEKIEHSTIATNGINMHVASAGSGPAVLFLHGFPELWYSWRHQLLYLSSLGYRAIAPDLRGFGD TDAPPSPSSYTAHHIVGDLVGLLDQLGVDQVFLVGHDWGAMMAWYFCLFRPDRVKALVNLSVHF TPRNPAISPLDGFRLMLGDDFYVCKFQEPGVAEADFGSVDTATMFKKFLTMRDPRPPIIPNGFR SLATPEALPSWLTEEDIDYFAAKFAKTGFTGGFNYYRAIDLTWELTAPWSGSEIKVPTKFIVGD LDLVYHFPGVKEYIHGGGFKKDVPFLEEVVVMEGAAHFINQEKADEINSLIYDFIKQF.	428	
EPH gene sequence (codon optimized, *E coli*)	ATGGAGAAGATTGAACACTCTACTATCGCTACTAATGGTATCAATATGCACGTTGCCTCTGCTG GTTCTGGTCCAGCTGTTTTGTTTTTGCACGGTTTCCCAGAATTATGGTATTCCTGGAGACACCA ATTGTTGTACTTGTCTTCTTTGGGTTACAGAGCTATTGCTCCAGATTTGAGAGGTTTCGGTGAC ACCGATGCTCCACCATCTCCATCCTCCTACACCGCCCACCACATCGTTGGTGATTTGGTCGGTT TGTTGGATCAATTAGGTGTCGATCAAGTCTTTTTGGTTGGTCATGATTGGGGTGCTATGATGGC CTGGTACTTCTGTTTGTTCCGTCCAGACAGAGTCAAGGCCTTAGTTAATTTATCTGTCCACTTC ACCCCACGTAACCCAGCTATCTCTCCATTAGATGGTTTCCGTTTGATGTTGGGTGATGATTTCT ACGTTTGTAAGTTTCAAGAACCAGGTGTCGCTGAAGCCGATTTCGGTTCTGTTGATACTGCCAC TATGTTTAAAAAGTTCTTGACCATGAGAGATCCACGTCCACCTATTATTCCAAACGGTTTCAGA TCCTTGGCCACCCCAGAAGCTTTGCCATCCTGGTTGACTGAAGAGGATATCGATTACTTTGCTG CCAAATTCGCTAAGACTGGTTTTACTGGTGGTTTCAACTACTACAGAGCTATCGACTTGACCTG GGAGTTGACTGCTCCATGGTCCGGTTCTGAAATCAAGGTTCCAACTAAGTTTATTGTTGGTGAC TTAGACTTGGTTTACCATTTCCCAGGTGTTAAGGAATACATTCACGGTGGTGGTTTCAAGAAGG ACGTTCCATTCTTGGAAGAAGTTGTCGTCATGGAAGGTGCTGCTCATTTTATCAACCAAGAAAA AGCTGACGAAATTAATTCTTTGATCTATGACTTCATTAAACAATTCTAG	429	
CYP87D18 protein	MWTVVLGLATLFVAYYIHWINKWRDSKFNGVLPPGTMGLPLIGETIQLSRPSDSLDVHPFIQKK VERYGPIFKTCLAGRPVVVSADAEFNNYIMLQEGRAVEMWYLDTLSKFFGLDTEWLKALGLIHK YIRSITLNHFGAEALRERFLPFIEASSMEALHSWSTQPSVEVKNASALMVFRTSVNKMFGEDAK KLSGNIPGKFTKLLGGFLSLPLNFPGTTYHKCLKDMKEIQKKLREVVDDRLANVGPDVEDFLGQ ALKDKESEKFISEEFIIQLLFSISFASFESISTTLTLILKLLDEHPEVVKELEAEHEAIRKARA DPDGPITWEEYKSMTFTLQVINETLRLGSVTPALLRKTVKDLQVKGYIIPEGWTIMLVTASRHR DPKVYKDPHIFNPWRWKDLDSITIQKNFMPFGGGLRHCAGAEYSKVYLCTFLHILCTKYRWTKL GGGRIARAHILSFEDGLHVKFTPKE.	430	
CYP87D18 gene sequence	ATGTGGACTGTCGTGCTCGGTTTGGCGACGCTGTTTGTCGCCTACTACATCCATTGGATTAACA AATGGAGAGATTCCAAGTTCAACGGAGTTCTGCCGCCGGGCACCATGGGTTTGCCGCTCATCGG AGAGACGATTCAACTGAGTCGACCCAGTGACTCCCTCGACGTTCACCCTTTCATCCAGAAAAAA GTTGAAAGATACGGGCCGATCTTCAAAACATGTCTGGCCGGAAGGCCGGTGGTGGTGTCGGCGG ACGCAGAGTTCAACAACTACATAATGCTGCAGGAAGGAAGAGCAGTGGAAATGTGGTATTTGGA TACGCTCTCCAAATTTTTCGGCCTCGACACCGAGTGGCTCAAAGCTCTGGGCCTCATCCACAAG TACATCAGAAGCATTACTCTCAATCACTTCGGCGCCGAGGCCCTGCGGGAGAGATTTCTTCCTT TTATTGAAGCATCCTCCATGGAAGCCCTTCACTCCTGGTCTACTCAACCTAGCGTCGAAGTCAA AAATGCCTCCGCTCTCATGGTTTTTAGGACCTCGGTGAATAAGATGTTCGGTGAGGATGCGAAG	431	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAGCTATCGGGAAATATCCCTGGGAAGTTCACGAAGCTTCTAGGAGGATTTCTCAGTTTACCAC TGAATTTTCCCGGCACCACCTACCACAAATGCTTGAAGGATATGAAGGAAATCCAGAAGAAGCT AAGAGAGGTTGTAGACGATAGATTGGCTAATGTGGGCCCTGATGTGGAAGATTTCTTGGGGCAA GCCCTTAAAGATAAGGAATCAGAGAAGTTCATTTCAGAGGAGTTCATCATCCAACTGTTGTTTT CTATCAGTTTTGCTAGCTTTGAGTCCATCTCCACCACTCTTACTTTGATTCTCAAGCTCCTTGA TGAACACCCAGAAGTAGTGAAAGAGTTGGAAGCTGAACACGAGGCGATTCGAAAAGCTAGAGCA GATCCAGATGGACCAATTACTTGGGAAGAATACAAATCCATGACTTTTACATTACAAGTCATCA ATGAAACCCTAAGGTTGGGGAGTGTCACACCTGCCTTGTTGAGGAAAACAGTTAAAGATCTTCA AGTAAAAGGATACATAATCCCGGAAGGATGGACAATAATGCTTGTCACCGCTTCACGTCACAGA GACCCAAAAGTCTATAAGGACCCTCATATCTTCAATCCATGGCGTTGGAAGGACTTGGACTCAA TTACCATCCAAAAGAACTTCATGCCTTTTGGGGGAGGCTTAAGGCATTGTGCTGGTGCTGAGTA CTCTAAAGTCTACTTGTGCACCTTCTTGCACATCCTCTGTACCAAATACCGATGGACCAAACTT GGGGGAGGAAGGATTGCAAGAGCTCATATATTGAGTTTTGAAGATGGGTTACATGTGAAGTTCA CACCCAAGGAATGA		
AtCPR protein	MTSALYASDLFKQLKSIMGTDSLSDDVVLVIATTSLALVAGFVVLLWKKTTADRSGELKPLMIP KSLMAKDEDDDLDLGSGKTRVSIFFGTQTGTAEGFAKALSEEIKARYEKAAVKDDYAADDDQYE EKLKKETLAFFCVATYGDGEPTDNAARFYKWFTEENERDIKLQQLAYGVFALGNRQYEHFNKIG IVLDEELCKKGAKRLIEVGLGDDDQSIEDDFNAWKESLWSELDKLLKDEDDKSVATPYTAVIPE YRVVTHDPRFTTQKSMESNVANGNTTIDIHHPCRVDVAVQKELHTHESDRSCIHLEFDISRTGI TYETGDHVGVYAENHVEIVEEAGKLLGHSLDLVFSIHADKEDGSPLESAVPPPFPGPCTLGTGL ARYADLLNPPRKSALVALAAYATEPSEAEKLKHLTSPDGKDEYSQWIVASQRSLLEVMAAFPSA KPPLGVFFAAIAPRLQPRYYSISSSPRLAPSRVHVTSALVYGPTPTGRIHKGVCSTWMKNAVPA EKSHECSGAPIFIRASNFKLPSNPSTPIVMVGPGTGLAPFRGFLQERMALKEDGEELGSSLLFF GCRNRQMDFIYEDELNNFVDQGVISELIMAFSREGAQKEYVQHKMMEKAAQVWDLIKEEGYLYV CGDAKGMARDVHRTLHTIVQEQEGVSSSEAEAIVKKLQTEGRYLRDVW	432	
AtCPR gene sequence	ATGACTTCTGCTTTGTATGCTTCCGATTTGTTTAAGCAGCTCAAGTCAATTATGGGGACAGATT CGTTATCCGACGATGTTGTACTTGTGATTGCAACGACGTCTTTGGCACTAGTAGCTGGATTTGT GGTGTTGTTATGGAAGAAAACGACGGCGGATCGGAGCGGGGAGCTGAAGCCTTTGATGATCCCT AAGTCTCTTATGGCTAAGGACGAGGATGATGATTTGGATTTGGGATCCGGGAAGACTAGAGTCT CTATCTTCTTCGGTACGCAGACTGGAACAGCTGAGGGATTTGCTAAGGCATTATCCGAAGAAAT CAAAGCGAGATATGAAAAAGCAGCAGTCAAAGATGACTATGCTGCCGATGATGACCAGTATGAA GAGAAATTGAAGAAGGAAACTTTGGCATTTTTCTGTGTTGCTACTTATGGAGATGGAGAGCCTA CTGACAATGCTGCCAGATTTTACAAATGGTTTACGGAGGAAAATGAACGGGATATAAAGCTTCA ACAACTAGCATATGGTGTGTTTGCTCTTGGTAATCGCCAATATGAACATTTTAATAAGATCGGG ATAGTTCTTGATGAAGAGTTATGTAAGAAAGGTGCAAAGCGTCTTATTGAAGTCGGTCTAGGAG ATGATGATCAGAGCATTGAGGATGATTTTAATGCCTGGAAAGAATCACTATGGTCTGAGCTAGA CAAGCTCCTCAAAGACGAGGATGATAAAAGTGTGGCAACTCCTTATACAGCTGTTATTCCTGAA TACCGGGTGGTGACTCATGATCCTCGGTTTACAACTCAAAAATCAATGGAATCAAATGTGGCCA	433	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ATGGAAATACTACTATTGACATTCATCATCCCTGCAGAGTTGATGTTGCTGTGCAGAAGGAGCT TCACACACATGAATCTGATCGGTCTTGCATTCATCTCGAGTTCGACATATCCAGGACGGGTATT ACATATGAAACAGGTGACCATGTAGGTGTATATGCTGAAAATCATGTTGAAATAGTTGAAGAAG CTGGAAAATTGCTTGGCCACTCTTTAGATTTAGTATTTTCCATACATGCTGACAAGGAAGATGG CTCCCCATTGGAAAGCGCAGTGCCGCCTCCTTTCCCTGGTCCATGCACACTTGGGACTGGTTTG GCAAGATACGCAGACCTTTTGAACCCTCCTCGAAAGTCTGCGTTAGTTGCCTTGGCGGCCTATG CCACTGAACCAAGTGAAGCCGAGAAACTTAAGCACCTGACATCACCTGATGGAAAGGATGAGTA CTCACAATGGATTGTTGCAAGTCAGAGAAGTCTTTTAGAGGTGATGGCTGCTTTTCCATCTGCA AAACCCCCACTAGGTGTATTTTTTGCTGCAATAGCTCCTCGTCTACAACCTCGTTACTACTCCA TCTCATCCTCGCCAAGATTGGCGCCAAGTAGAGTTCATGTTACATCCGCACTAGTATATGGTCC AACTCCTACTGGTAGAATCCACAAGGGTGTGTGTTCTACGTGGATGAAGAATGCAGTTCCTGCG GAGAAAAGTCATGAATGTAGTGGAGCCCCAATCTTTATTCGAGCATCTAATTTCAAGTTACCAT CCAACCCTTCAACTCCAATCGTTATGGTGGGACCTGGGACTGGGCTGGCACCTTTTAGAGGTTT TCTGCAGGAAAGGATGGCACTAAAAGAAGATGGAGAAGAACTAGGTTCATCTTTGCTCTTCTTT GGGTGTAGAAATCGACAGATGGACTTTATATACGAGGATGAGCTCAATAATTTTGTTGATCAAG GCGTAATATCTGAGCTCATCATGGCATTCTCCCGTGAAGGAGCTCAGAAGGAGTATGTTCAACA TAAGATGATGGAGAAGGCAGCACAAGTTTGGGATCTAATAAAGGAAGAAGGATATCTCTATGTA TGCGGTGATGCTAAGGGCATGGCGAGGGACGTCCACCGAACTCTACACACCATTGTTCAGGAGC AGGAAGGTGTGAGTTCGTCAGAGGCAGAGGCTATAGTTAAGAAACTTCAAACCGAAGGAAGATA CCTCAGAGATGTCTGGTGA		
AGY15763.1 protein	MWKVPKFIKQSYLVFLLALLLYSSFGFSFSRTEATTSTGALGPVTPKDTIYQIVTDRFFDGDPS NNKPPGFDPTLFDDPDGNNQGNGKDLKLYQGGDFQGIIDKIPYLKNMGITAVWISAPYENRDTV IEDYQSDGSINRWTSFHGYHARNYFATNKHFGTMKDFIRLRDALHQNGIKLVIDFVSNHSSRWQ NPTLNFAPEDGKLYEPDKDANGNYVFDANGEPADYNGDGKVENLLADPHNDVNGFFHGLGDRGN DTSRFGYRYKDLGSLADYSQENALVVEHLEKAAKFWKSKGIDGFRHDATLHMNPAFVKGFKDAI DSDAGGPVTHFGEFFIGRPDPKYDEYRTFPERTGVNNLDFEYFRAATNAFGNFSETMSSFGDMM IKTSNDYIYENQTVTFLDNHDVTRFRYIQPNDKPYHAALAVLMTSRGIPNIYYGTEQYLMPSDS SDIAGRMFMQTSTNFDENTTAYKVIQKLSNLRKNNEAIAYGTTEILYSTNDVLVFKRQFYDKQV IVAVNRQPDQTFTIPELDTTLPVGTYSDVLGGLLYGSSMSVNNVNGQNKISSFTLSGGEVNVWS YNPSLGTLTPRIGDVISTMGRPGNTVYIYGTGLGGSVTVKFGSTVATVVSNSDQMIEAIVPNTN PGIQNITVTKGSVTSDPFRYEVLSGDQVQVIFHVNATTNWGENIYVVGNIPELGSWDPNQSSEA MLNPNYPEWFLPVSVPKGATFEFKFIKKDNNGNVIWESRSNRVFTAPNSSTGTIDTPLYFWDN	434	
AGY15764.1 protein	TTSTGALGPVTPKDTIYQIVTDRFFDGDPSNNKPPGFDPTLFDDPDGNNQGNGKDLKLYQGGDF QGIIDKIPYLKNMGITAVWISAPYENRDTVIEDYQSDGSINRWTSFHGYHARNYFATNKHFGTM KDFIRLRDALHQNGIKLVIDFVSNHSSRWQNPTLNFAPEDGKLYEPDKDANGNYVFDANGEPAD YNGDGKVENLLADPHNDVNGFFHGLGDRGNDTSRFGYRYKDLGSLADYSQENALVVEHLEKAAK FWKSKGIDGFRHDATLHMNPAFVKGFKDAIDSDAGGPVTHFGEFFIGRPDPKYDEYRTFPERTG VNNLDFEYFRAATNAFGNFSETMSSFGDMMIKTSNDYIYENQTVTFLDNHDVTRFRYIQPNDKP	435	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	YHAALAVLMTSRGIPNIYYGTEQYLMPSDSSDIAGRMFMQTSTNFDENTTAYKVIQKLSNLRKN NEAIAYGTTEILYSTNDVLVFKRQFYDKQVIVAVNRQPDQTFTIPELDTTLPVGTYSDVLGGLL YGSSMSVNNVNGQNKISSFTLSGGEVNVWSYNPSLGTLTPRIGDVISTMGRPGNTVYIYGTGLG GSVTVKFGSTVATVVSNSDQMIEAIVPNTNPGIQNITVTKGSVTSDPFRYEVLSGDQVQVIFHV NATTNWGENIYVVGNIPELGSWDPNQSSEAMLNPNYPEWFLPVSVPKGATFEFKFIKKDNNGNV IWESRSNRVFTAPNSSTGTIDTPLYFWDN		
Glycosyltransferase(311)	MDSGYSSSYAAAAGMHVVICPWLAFGHLLPCLDLAQRLASRGHRVSFVSTPRNISRLPPVRPAL APLVAFVALPLPRVEGLPDGAESTNDVPHDRPDMVELHRRAFDGLAAPFSEFLGTACADWVIVD VFHHWAAAAALEHKVPCAMMLLGSAHMIASIADRRLERAETESPAAAGQGRPAAAPTFEVARMK LIRTKGSSGMSLAERFSLTLSRSSLVVGRSCVEFEPETVPLLSTLRGKPITFLGLMPPLHEGRR EDGEDATVRWLDAQPAKSVVYVALGSEVPLGVEKVHELALGLELAGTRFLWALRKPTGVSDADL LPAGFEERTRGRGVVATRWVPQMSILAHAAVGAFLTHCGWNSTIEGLMFGHPLIMLPIFGDQGP NARLIEAKNAGLQVARNDGDGSFDREGVAAAIRAVAVEEESSKVFQAKAKKLQEIVADMACHER YIDGFIQQLRSYKDAAALE	436	
Glycosyltransferase(311) (gDNA native)	ATGGACTCCGGCTACTCCTCCTCCTACGCCGCCGCCGCCGGGATGCACGTCGTGATCTGCCCGT GGCTCGCCTTCGGCCACCTGCTCCCGTGCCTCGACCTCGCCCAGCGCCTCGCGTCGCGGGGCCA CCGCGTGTCGTTCGTCTCCACGCCGCGGAACATATCCCGCCTCCCGCCGGTGCGCCCCGCGCTC GCGCCGCTCGTCGCCTTCGTGGCGCTGCCGCTCCCGCGCGTCGAGGGGCTCCCCGACGGCGCCG AGTCCACCAACGACGTCCCCCACGACAGGCCGGACATGGTCGAGCTCCACCGGAGGGCCTTCGA CGGGCTCGCCGCGCCCTTCTCGGAGTTCTTGGGCACCGCGTGCGCCGACTGGGTCATCGTCGAC GTCTTCCACCACTGGGCCGCAGCCGCCGCTCTCGAGCACAAGGTGCCATGTGCAATGATGTTGT TGGGCTCTGCACATATGATCGCTTCCATAGCAGACAGACGGCTCGAGCGCGCGGAGACAGAGTC GCCTGCGGCTGCCGGGCAGGGACGCCCAGCGGCGGCGCCAACGTTCGAGGTGGCGAGGATGAAG TTGATACGAACCAAAGGCTCATCGGGAATGTCCCTCGCCGAGCGCTTCTCCTTGACGCTCTCGA GGAGCAGCCTCGTCGTCGGGCGGAGCTGCGTGGAGTTCGAGCCGGAGACCGTCCCGCTCCTGTC GACGCTCCGCGGTAAGCCTATTACCTTCCTTGGCCTTATGCCGCCGTTGCATGAAGGCCGCCGC GAGGACGGCGAGGATGCCACCGTCCGCTGGCTCGACGCGCAGCCGGCCAAGTCCGTCGTGTACG TCGCGCTAGGCAGCGAGGTGCCACTGGGAGTGGAGAAGGTCCACGAGCTCGCGCTCGGGCTGGA GCTCGCCGGGACGCGCTTCCTCTGGGCTCTTAGGAAGCCCACTGGCGTCTCCGACGCCGACCTC CTCCCCGCCGGCTTCGAGGAGCGCACGCGCGGCCGCGGCGTCGTGGCGACGAGATGGGTTCCTC AGATGAGCATACTGGCGCACGCCGCCGTGGGCGCGTTCCTGACCCACTGCGGCTGGAACTCGAC CATCGAGGGGCTCATGTTCGGCCACCCGCTTATCATGCTGCCGATCTTCGGCGACCAGGGACCG AACGCGCGGCTAATCGAGGCGAAGAACGCCGGATTGCAGGTGGCAAGAAACGACGGCGATGGAT CGTTCGACCGAGAAGGCGTCGCGGCGGCGATTCGTGCAGTCGCGGTGGAGGAAGAAAGCAGCAA AGTGTTTCAAGCCAAAGCCAAGAAGCTGCAGGAGATCGTCGCGGACATGGCCTGCCATGAGAGG TACATCGACGGATTCATTCAGCAATTGAGATCTTACAAGGATTGA	437	
Glycosyltransferase(311) (gDNA	ATGGATAGCGGTTATAGCAGCAGCTATGCAGCAGCAGCCGGTATGCATGTTGTTATTTGTCCGT GGCTGGCATTTGGTCATCTGCTGCCGTGTCTGGATCTGGCACAGCGTCTGGCAAGCCGTGGTCA	438	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
codon optimized, E. coli)	TCGTGTTAGCTTTGTTAGCACACCGCGTAATATTAGCCGTCTGCCTCCGGTTCGTCCGGCACTG GCACCGCTGGTTGCATTTGTTGCACTGCCGCTGCCTCGTGTTGAAGGTCTGCCGGATGGTGCAG AAAGCACCAATGATGTTCCGCATGATCGTCCGGATATGGTTGAACTGCATCGTCGTGCATTTGA TGGTCTGGCAGCACCGTTTAGCGAATTTCTGGGCACCGCATGTGCAGATTGGGTTATTGTTGAT GTTTTTCATCATTGGGCAGCCGCAGCAGCACTGGAACATAAAGTTCCGTGTGCAATGATGCTGC TGGGTAGCGCACATATGATTGCAAGCATTGCAGATCGTCGTCTGGAACGTGCAGAAACCGAAAG TCCTGCGGCAGCAGGTCAGGGTCGTCCTGCAGCCGCACCGACCTTTGAAGTTGCACGTATGAAA CTGATTCGTACCAAAGGTAGCAGCGGTATGAGCCTGGCAGAACGTTTTAGTCTGACCCTGAGCC GTAGCAGCCTGGTTGTTGGTCGTAGCTGTGTTGAATTTGAACCGGAAACCGTTCCGCTGCTGAG CACCCTGCGTGGTAAACCGATTACCTTTCTGGGTCTGATGCCTCCGCTGCATGAAGGTCGTCGC GAAGATGGTGAAGATGCAACCGTTCGTTGGCTGGATGCACAGCCTGCAAAAAGCGTTGTTTATG TTGCCCTGGGTAGTGAAGTTCCGCTGGGTGTTGAAAAAGTGCATGAACTGGCACTGGGTTTAGA ACTGGCAGGCACCCGTTTTCTGTGGGCACTGCGTAAACCGACCGGTGTTAGTGATGCCGATCTG CTTCCGGCAGGTTTTGAAGAACGTACCCGTGGTCGTGGTGTTGTTGCAACCCGTTGGGTTCCGC AGATGAGCATTCTGGCACATGCAGCAGTGGGTGCATTTCTGACCCATTGTGGTTGGAATAGCAC CATTGAAGGCCTGATGTTTGGCCATCCGCTGATTATGCTGCCGATTTTTGGTGATCAGGGTCCG AATGCACGTCTGATTGAAGCAAAAAATGCAGGTCTGCAGGTTGCCCGTAATGATGGTGATGGTA GCTTTGATCGTGAAGGTGTTGCAGCAGCCATTCGTGCAGTTGCAGTTGAAGAAGAAAGCAGCAA AGTTTTTCAGGCCAAAGCCAAAAAACTGCAAGAAATTGTTGCAGATATGGCCTGCCATGAACGT TATATTGATGGTTTTATTCAGCAGCTGCGTAGCTACAAAGAT		
UGT76G1 protein	MENKTETTVRRRRRIILFPVPFQGHINPILQLANVLYSKGFSITIFHTNFNKPKTSNYPHFTFR FILDNDPQDERISNLPTHGPLAGMRIPIINEHGADELRRELELLMLASEEDEEVSCLITDALWY FAQSVADSLNLRRLVLMTSSLFNFHAHVSLPQFDELGYLDPDDKTRLEEQASGFPMLKVKDIKS AYSNWQILKEILGKMIKQTKASSGVIWNSFKELEESELETVIREIPAPSFLIPLPKHLTASSSS LLDHDRTVFQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLVDSKQSFLWVVRPGFVKGSTW VEPLPDGFLGERGRIVKWVPQQEVLAHGAIGAFWTHSGWNSTLESVCEGVPMIFSDFGLDQPLN ARYMSDVLKVGVYLENGWERGEIANAIRRVMVDEEGEYIRQNARVLKQKADVSLMKGGSSYESL ESLVSYISSL	439	
UGT76G1 gene sequence	ATGGAAAATAAAACGGAGACCACCGTTCGCCGGCGCCGGAGAATAATATTATTCCCGGTACCAT TTCAAGGCCACATTAACCCAATTCTTCAGCTAGCCAATGTGTTGTACTCTAAAGGATTCAGTAT CACCATCTTTCACACCAACTTCAACAAACCCAAAACATCTAATTACCCTCACTTCACTTTCAGA TTCATCCTCGACAACGACCCACAAGACGAACGCATTTCCAATCTACCGACTCATGGTCCGCTCG CTGGTATGCGGATTCCGATTATCAACGAACACGGAGCTGACGAATTACGACGCGAACTGGAACT GTTGATGTTAGCTTCTGAAGAAGATGAAGAGGTATCGTGTTTAATCACGGATGCTCTTTGGTAC TTCGCGCAATCTGTTGCTGACAGTCTTAACCTCCGACGGCTTGTTTTGATGACAAGCAGCTTGT TTAATTTTCATGCACATGTTTCACTTCCTCAGTTTGATGAGCTTGGTTACCTCGATCCTGATGA CAAAACCCGTTTGGAAGAACAAGCGAGTGGGTTTCCTATGCTAAAAGTGAAAGACATCAAGTCT GCGTATTCGAACTGGCAAATACTCAAAGAGATATTAGGGAAGATGATAAAACAAACAAAAGCAT	440	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTTCAGGAGTCATCTGGAACTCATTTAAGGAACTCGAAGAGTCTGAGCTCGAAACTGTTATCCG TGAGATCCCGGCTCCAAGTTTCTTGATACCACTCCCCAAGCATTTGACAGCCTCTTCCAGCAGC TTACTAGACCACGATCGAACCGTTTTTCAATGGTTAGACCAACAACCGCCAAGTTCGGTACTGT ATGTTAGTTTTGGTAGTACTAGTGAAGTGGATGAGAAAGATTTCTTGGAAATAGCTCGTGGGTT GGTTGATAGCAAGCAGTCGTTTTTATGGGTGGTTCGACCTGGGTTTGTCAAGGGTTCGACGTGG GTCGAACCGTTGCCAGATGGGTTCTTGGGTGAAAGAGGACGTATTGTGAAATGGGTTCCACAGC AAGAAGTGCTAGCTCATGGAGCAATAGGCGCATTCTGGACTCATAGCGGATGGAACTCTACGTT GGAAAGCGTTTGTGAAGGTGTTCCTATGATTTTCTCGGATTTTGGGCTCGATCAACCGTTGAAT GCTAGATACATGAGTGATGTTTTGAAGGTAGGGGTGTATTTGGAAAATGGGTGGGAAAGAGGAG AGATAGCAAATGCAATAAGAAGAGTTATGGTGGATGAAGAAGGAGAATACATTAGACAGAATGC AAGAGTTTTGAAACAAAAGGCAGATGTTTCTTTGATGAAGGGTGGTTCGTCTTACGAATCATTA GAGTCTCTAGTTTCTTACATTTCATCGTTGTAA		
UGT73C5 protein	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVK LTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPE MLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKY LKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRGSVSETTKSSPLHFVLFPFMAQGHMIPMVDIA RLLAQRGVIITIVTTPHNAARFKNVLNRAIESGLPINLVQVKFPYLEAGLQEGQENIDSLDTME RMIPFFKAVNFLEEPVQKLIEEMNPRPSCLISDFCLPYTSKIAKKFNIPKILFHGMGCFCLLCM HVLRKNREILDNLKSDKELFTVPDFPDRVEFTRTQVPVETYVPAGDWKDIFDGMVEANETSYGV IVNSFQELEPAYAKDYKEVRSGKAWTIGPVSLCNKVGADKAERGNKSDIDQDECLKWLDSKKHG SVLYVCLGSICNLPLSQLKELGLGLEESQRPFIWVIRGWEKYKELVEWFSESGFEDRIQDRGLL IKGWSPQMLILSHPSVGGFLTHCGWNSTLEGITAGLPLLTWPLFADQFCNEKLVVEVLKAGVRS GVEQPMKWGEEEKIGVLVDKEGVKKAVEELMGESDDAKERRRRAKELGDSAHKAVEEGGSSHSN ISFLLQDIMELAEPNNAAAS	441	
UGT73C5 gene sequence	ATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGG AATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAA CAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAA TTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTC CAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTC GAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAA ATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTAACCC ATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGA TGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTAC TTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCG ACCATCCTCCAAAATCGGATCTGGTTCCGCGTGGATCCGTTTCCGAAACAACCAAATCTTCTCC ACTTCACTTTGTTCTCTTCCCTTTCATGGCTCAAGGCCACATGATTCCCATGGTTGATATTGCA AGGCTCTTGGCTCAGCGTGGTGTGATCATAACAATTGTCACGACGCCTCACAATGCAGCGAGGT TCAAGAATGTCCTAAACCGTGCCATTGAGTCTGGCTTGCCCATCAACTTAGTGCAAGTCAAGTT	442	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TCCATATCTAGAAGCTGGTTTGCAAGAAGGACAAGAGAATATCGATTCTCTTGACACAATGGAG CGGATGATACCTTTCTTTAAAGCGGTTAACTTTCTCGAAGAACCAGTCCAGAAGCTCATTGAAG AGATGAACCCTCGACCAAGCTGTCTAATTTCTGATTTTTGTTTGCCTTATACAAGCAAAATCGC CAAGAAGTTCAATATCCCAAAGATCCTCTTCCATGGCATGGGTTGCTTTTGTCTTCTGTGTATG CATGTTTTACGCAAGAACCGTGAGATCTTGGACAATTTAAAGTCAGATAAGGAGCTTTTCACTG TTCCTGATTTTCCTGATAGAGTTGAATTCACAAGAACGCAAGTTCCGGTAGAAACATATGTTCC AGCTGGAGACTGGAAAGATATCTTTGATGGTATGGTAGAAGCGAATGAGACATCTTATGGTGTG ATCGTCAACTCATTTCAAGAGCTCGAGCCTGCTTATGCCAAAGACTACAAGGAGGTAAGGTCCG GTAAAGCATGGACCATTGGACCCGTTTCCTTGTGCAACAAGGTAGGAGCCGACAAAGCAGAGAG GGGAAACAAATCAGACATTGATCAAGATGAGTGCCTTAAATGGCTCGATTCTAAGAAACATGGC TCGGTGCTTTACGTTTGTCTTGGAAGTATCTGTAATCTTCCTTTGTCTCAACTCAAGGAGCTGG GACTAGGCCTAGAGGAATCCCAAAGACCTTTCATTTGGGTCATAAGAGGTTGGGAGAAGTACAA AGAGTTAGTTGAGTGGTTCTCGGAAAGCGGCTTTGAAGATAGAATCCAAGATAGAGGACTTCTC ATCAAAGGATGGTCCCCTCAAATGCTTATCCTTTCACATCCATCAGTTGGAGGGTTCCTAACAC ACTGTGGTTGGAACTCGACTCTTGAGGGGATAACTGCTGGTCTACCGCTACTTACATGGCCGCT ATTCGCAGACCAATTCTGCAATGAGAAATTGGTCGTTGAGGTACTAAAAGCCGGTGTAAGATCC GGGGTTGAACAGCCTATGAAATGGGGAGAAGAGGAGAAAATAGGAGTGTTGGTGGATAAAGAAG GAGTGAAGAAGGCAGTGGAAGAATTAATGGGTGAGAGTGATGATGCAAAAGAGAGAAGAAGAAG AGCCAAAGAGCTTGGAGATTCAGCTCACAAGGCTGTGGAAGAAGGAGGCTCTTCTCATTCTAAC ATCTCTTTCTTGCTACAAGACATAATGGAACTGGCAGAACCCAATAATGCGGCCGCATCGTGA		
UGT73C5 gene sequence (Codon optimized, *E. coli*)	ATGAGGCATGGATCCGTTAGCGAAACCACCAAAAGCAGTCCGCTGCATTTTGTTCTGTTTCCGT TTATGGCACAGGGTCATATGATTCCGATGGTTGATATTGCACGTCTGCTGGCACAGCGTGGTGT GATTATTACCATTGTTACCACACCGCATAATGCAGCACGCTTTAAAAACGTTCTGAATCGTGCA ATTGAAAGCGGTCTGCCGATTAATCTGGTTCAGGTTAAATTTCCGTATCTGGAAGCAGGTCTGC AAGAAGGTCAAGAAAATATTGATAGCCTGGATACCATGGAACGCATGATTCCGTTTTTCAAAGC CGTGAATTTTCTGGAAGAACCGGTGCAGAAACTGATCGAAGAAATGAATCCGCGTCCGAGCTGT CTGATTAGCGATTTTTGTCTGCCGTATACCAGCAAAATCGCCAAAAAATTCAACATCCCGAAAA TCCTGTTTCATGGTATGGGTTGTTTTTGcctgctgtgtatgcatgttcTGCGTAAAAATCGTGA AATCCTGGATAACCTGAAAAGCGATAAAGAACTGTTTACCGTTCCGGATTTTCCGGATCGTGTG GAATTTACCCGTACACAGGTTCCGGTTGAAACCTATGTTCCGGCAGGCGATTGGAAAGATATTT TTGATGGTATGGTGGAAGCCAACGAAACCAGCTATGGTGTTATTGTGAATAGCTTTCAAGAACT GGAACCGGCATATGCGAAAGATTACAAAGAAGTTCGTAGCGGTAAAGCATGGACCATTGGTCCG GTTAGCCTGTGTAATAAAGTTGGTGCAGATAAAGCAGAACGCGGTAATAAAAGTGATATCGATC AGGATGAATGCCTGAAATGGCTGGATAGCAAAAAACATGGTAGCGTTCTGTATGTTTGTCTGGG TAGCATTTGCAATCTGCCGCTGAGCCAGCTGAAAGAATTAGGTCTGGGTTTAGAAGAAAGCCAG CGTCCGTTTATTTGGGTTATTCGTGGTTGGGAGAAATACAAAGAACTGGTTGAATGGTTTAGCG AAAGCGGTTTTGAAGATCGTATTCAGGATCGTGGCCTGCTGATTAAAGGTTGGAGTCCGCAGAT GCTGATTCTGAGCCATCCGAGCGTTGGTGGCTTTCTGACCCATTGTGGTTGGAATAGCACCCTG	443	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GAAGGTATTACAGCTGGCCTGCCGCTGCTGACCTGGCCTCTGTTTGCAGATCAGTTTTGTAATG AAAAACTGGTGGTGGAAGTTCTGAAAGCCGGTGTGCGTAGCGGTGTTGAACAGCCGATGAAATG GGGTGAAGAAGAAAAAATTGGCGTCCTGGTTGATAAAGAAGGTGTTAAAAAAGCCGTGGAAGAA CTGATGGGTGAAAGTGATGATGCAAAAGAACGTCGTCGTCGTGCAAAAGAGCTGGGCGATAGCG CACATAAAGCAGTTGAAGAAGGTGGTAGCAGCCATAGCAATATTAGCTTTCTGCTGCAGGATAT TATGGAACTGGCAGAACCGAATAACTAAGCGGCCGCTGAA		
UGT73C6 protein	MAFEKNNEPFPLHFVLFPFMAQGHMIPMVDIARLLAQRGVLITIVTTPHNAARFKNVLNRAIES GLPINLVQVKFPYQEAGLQEGQENMDLLTTMEQITSFFKAVNLLKEPVQNLIEEMSPRPSCLIS DMCLSYTSEIAKKFKIPKILFHGMGCFCLLCVNVLRKNREILDNLKSDKEYFIVPYFPDRVEFT RPQVPVETYVPAGWKEILEDMVEADKTSYGVIVNSFQELEPAYAKDFKEARSGKAWTIGPVSLC NKVGVDKAERGNKSDIDQDECLEWLDSKEPGSVLYVCLGSICNLPLSQLLELGLGLEESQRPFI WVIRGWEKYKELVEWFSESGFEDRIQDRGLLIKGWSPQMLILSHPSVGGFLTHCGWNSTLEGIT AGLPMLTWPLFADQFCNEKLVVQILKVGVSAEVKEVMKWGEEEKIGVLVDKEGVKKAVEELMGE SDDAKERRRRAKELGESAHKAVEEGGSSHSNITFLLQDIMQLAQSNN	444	
UGT73C6 (gDNA, native)	ATGGCTTTCGAAAAAAACAACGAACCTTTTCCTCTTCACTTTGTTCTCTTCCCTTTCATGGCTC AAGGCCACATGATTCCCATGGTTGATATTGCAAGGCTCTTGGCTCAGCGAGGTGTGCTTATAAC AATTGTCACGACGCCTCACAATGCAGCAAGGTTCAAGAATGTCCTAAACCGTGCCATTGAGTCT GGTTTGCCCATCAACCTAGTGCAAGTCAAGTTTCCATATCAAGAAGCTGGTCTGCAAGAAGGAC AAGAAAATATGGATTTGCTTACCACGATGGAGCAGATAACATCTTTCTTTAAAGCGGTTAACTT ACTCAAAGAACCAGTCCAGAACCTTATTGAAGAGATGAGCCCGCGACCAAGCTGTCTAATCTCT GATATGTGTTTGTCGTATACAAGCGAAATCGCCAAGAAGTTCAAAATACCAAAGATCCTCTTCC ATGGCATGGGTTGCTTTTGTCTTCTGTGTGTTAACGTTCTGCGCAAGAACCGTGAGATCTTGGA CAATTTAAAGTCTGATAAGGAGTACTTCATTGTTCCTTATTTTCCTGATAGAGTTGAATTCACA AGACCTCAAGTTCCGGTGGAAACATATGTTCCTGCAGGCTGGAAAGAGATCTTGGAGGATATGG TAGAAGCGGATAAGACATCTTATGGTGTTATAGTCAACTCATTTCAAGAGCTCGAACCTGCGTA TGCCAAAGACTTCAAGGAGGCAAGGTCTGGTAAAGCATGGACCATTGGACCTGTTTCCTTGTGC AACAAGGTAGGAGTAGACAAAGCAGAGAGGGGAAACAAATCAGATATTGATCAAGATGAGTGCC TTGAATGGCTCGATTCTAAGGAACCGGGATCTGTGCTCTACGTTTGCCTTGGAAGTATTTGTAA TCTTCCTCTGTCTCAGCTCCTTGAGCTGGGACTAGGCCTAGAGGAATCCCAAAGACCTTTCATC TGGGTCATAAGAGGTTGGGAGAAATACAAAGAGTTAGTTGAGTGGTTCTCGGAAAGCGGCTTTG AAGATAGAATCCAAGATAGAGGACTTCTCATCAAAGGATGGTCCCCTCAAATGCTTATCCTTTC ACATCCTTCTGTTGGAGGGTTCTTAACGCACTGCGGATGGAACTCGACTCTTGAGGGGATAACT GCTGGTCTACCAATGCTTACATGGCCACTATTTGCAGACCAATTCTGCAACGAGAAACTGGTCG TACAAATACTAAAAGTCGGTGTAAGTGCCGAGGTTAAAGAGGTCATGAAATGGGGAGAAGAAGA GAAGATAGGAGTGTTGGTGGATAAAGAAGGAGTGAAGAAGGCAGTGGAAGAACTAATGGGTGAG AGTGATGATGCAAAAGAGAGAAGAAGAAGAGCCAAAGAGCTTGGAGAATCAGCTCACAAGGCTG TGGAAGAAGGAGGCTCCTCTCATTCTAATATCACTTTCTTGCTACAAGACATAATGCAACTAGC ACAGTCCAATAAT	445	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
SgCbQ protein	MWRLKVGAESVGENDEKWLKSISNHLGRQVWEFCPDAGTQQQLLQVHKARKAFHDDRFHRKQSS DLFITIQYGKEVENGGKTAGVKLKEGEEVRKEAVESSLERALSFYSSIQTSDGNWASDLGGPMF LLPGLVIALYVTGVLNSVLSKHHRQEMCRYVYNHQNEDGGWGLHIEGPSTMFGSALNYVALRLL GEDANAGAMPKARAWILDHGGATGITSWGKLWLSVLGVYEWSGNNPLPPEFWLFPYFLPFHPGR MWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYAVPYHEIDWNKSRNTCAKEDLYYPHPKM QDILWGSLHHVYEPLFTRWPAKRLREKALQTAMQHIHYEDENTRYICLGPVNKVLNLLCCWVED PYSDAFKLHLQRVHDYLWVAEDGMKMQGYNGSQLWDTAFSIQAIVSTKLVDNYGPTLRKAHDFV KSSQIQQDCPGDPNVWYRHIHKGAWPFSTRDHGWLISDCTAEGLKAALMLSKLPSETVGESLER NRLCDAVNVLLSLQNDNGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTSATMEALTLF KKLHPGHRTKEIDTAIVRAANFLENMQRTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCLA IRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAERDPTPLH RAARLLINSQLENGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYCHRVLTE	446	
glycoside hydrolase family 5 protein [*Trichoderma reesei* QM6a] GenBank: EGR512.1	MNKSVAPLLLAASILYGGAAAQQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCIPGATTIT TSTRPPSGPTTTTRATSTSSSTPPTSSGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGS NNYPDGIGQMQHFVNDDGMTIFRLPVGWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIV DIHNYARWNGGIIGQGGPTNAQFTSLWSQLASKYASQSRVWFGIMNEPHDVNINTWAATVQEVV TAIRNAGATSQFISLPGNDWQSAGAFISDGSAAALSQVTNPDGSTTNLIFDVHKYLDSDNSGTH AECTTNNIDGAFSPLATWLRQNNRQAILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWG AGSFDSTYVLTETPTGSGNSWTDTSLVSSCLARK	447	
glycoside hydrolase family 61 protein [*T. reesei* QM6a] EGR50392.1 GI:340520155	MKSCAILAALGCLAGSVLGHGQVQNFTINGQYNQGFILDYYYQKQNTGHFPNVAGWYAEDLDLG FISPDQYTTPDIVCHKNAAPGAISATAAAGSNIVFQWGPGVWPHPYGPIVTYVAECSGSCTTVN KNNLRWVKIQEAGINYNTQVWAQQDLINQGNKWTVKIPSSLRPGNYVFRHELLAAHGASSANGM QNYPQCVNIAVTGSGTKALPAGTPATQLYKPTDPGILFNPYTTITSYTIPGPALWQG	448	
glycoside hydrolase family 61 protein, partial [*Trichoderma reesei* QM6a] EGR49821.1 GI:340519583	SAHTTFTTLFIDKKNQGDGTCVRMPYDDKTATNPVKPITSSDMACGRNGGDPVPFICSAKKGSL LTFEFRLWPDAQQPGSIDPGHLGPCAVYLKKVDNMFSDSAAGGGWFKIWEDGYDSKTQKWCVDR LVKNNGLLSVRLPRGLPAGYYIVRPEILALHWAAHRDDPQFYLGCAQIFVDSDVRGPLEIPRRQ QATIPGYVNAKTPGLTFDIYQDKLPPYPMPGPKVYIPPAKGNKPNQDLNAGRLVQTDGLIPKDC LIKKANWCGRPVEPYSSARMCWRAVNDCYAQSKKCRESSPPIGLTNCDRWSDHCGKMDALCEQE KYKGPP	449	
glycoside hydrolase family 7 protein [*T. reesei* QM6a]	MAPSVTLPLTTAILAIARLVAAQQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMH DANYNSCTVNGGVNTTLCPDEATCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVS PRLYLLDSDGEYVMLKLNGQELSFDVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGY CDAQCPVQTWRNGTLNTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYK SYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDTISSCPSASA	450	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
EGR48251.1 GI:340518009	YGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNI RWGDIGSTTNSTAPPPPPASSTTFSTTRRSSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTC QYSNDYYSQCL		
glycoside hydrolase family 45 protein [*T. reesei* QM6a] EGR47058.1 GI:340516811	MKATLVLGSLIVGAVSAYKATTTRYYDGQEGACGCGSSSGAFPWQLGIGNGVYTAAGSQALFDT AGASWCGAGCGKCYQLTSTGQAPCSSCGTGGAAGQSIIVMVTNLCPNNGNAQWCPVVGGTNQYG YSYHFDIMAQNEIFGDNVVVDFEPIACPGQAASDWGTCLCVGQQETDPTPVLGNDTGSTPPGSS PPATSSSPPSGGGQQTLYGQCGGAGWTGPTTCQAPGTCKVQNQWYSQCLP	451	
glycoside hydrolase family 5 protein [*T. reesei* QM6a] EGR44174.1 GI:340513898	MRATSLLAAALAVAGDALAGKIKYLGVAIPGIDFGCDIDGSCPTDTSSVPLLSYKGGDGAGQMK HFAEDDGLNVFRISATWQFVLNNTVDGKLDELNWGSYNKVVNACLETGAYCMIDMHNFARYNGG IIGQGGVSDDIFVDLWVQIAKYYEDNDKIIFGLMNEPHDLDIEIWAQTCQKVVTAIRKAGATSQ MILLPGTNFASVETYVSTGSAEALGKITNPDGSTDLLYFDVHKYLDINNSGSHAECTTDNVDAF NDFADWLRQNKRQAIISETGASMEPSCMTAFCAQNKAISENSDVYIGFVGWGAGSFDTSYILTL TPLGKPGNYTDNKLMNECILDQFTLDEKYRPTPTSISTAAEETATATATSDGDAPSTTKPIFRE ETASPTPNAVTKPSPDTSDSSDDDKDSAASMSAQGLTGTVLFTVAALGYMLVAF	452 453	
glycoside hydrolase family 45 protein [*T. reesei* QM6a] XP_006967072.1 GI:589110099	MKATLVLGSLIVGAVSAYKATTTRYYDGQEGACGCGSSSGAFPWQLGIGNGVYTAAGSQALFDT AGASWCGAGCGKCYQLTSTGQAPCSSCGTGGAAGQSIIVMVTNLCPNNGNAQWCPVVGGTNQYG YSYHFDIMAQNEIFGDNVVVDFEPIACPGQAASDWGTCLCVGQQETDPTPVLGNDTGSTPPGSS PPATSSSPPSGGGQQTLYGQCGGAGWTGPTTCQAPGTCKVQNQWYSQCLP	454	
glycoside hydrolase family 7 protein [*T. reesei* QM6a] XP_006965674.1 GI:589107303	MAPSVTLPLTTAILAIARLVAAQQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMH DANYNSCTVNGGVNTTLCPDEATCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVS PRLYLLDSDGEYVMLKLNGQELSFDVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGY CDAQCPVQTWRNGTLNTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYK SYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDTISSCPSASA YGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNI RWGDIGSTTNSTAPPPPPASSTTFSTTRRSSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTC QYSNDYYSQCL	455	
glycoside hydrolase family 61 protein, partial [*T. reesei* QM6a] XP_006964038.1	SAHTTFTTLFIDKKNQGDGTCVRMPYDDKTATNPVKPITSSDMACGRNGGDPVPFICSAKKGSL LTFEFRLWPDAQQPGSIDPGHLGPCAVYLKKVDNMFSDSAAGGGWFKIWEDGYDSKTQKWCVDR LVKNNGLLSVRLPRGLPAGYYIVRPEILALHWAAHRDDPQFYLGCAQIFVDSDVRGPLEIPRRQ QATIPGYVNAKTPGLTFDIYQDKLPPYPMPGPKVYIPPAKGNKPNQDLNAGRLVQTDGLIPKDC LIKKANWCGRPVEPYSSARMCWRAVNDCYAQSKKCRESSPPIGLTNCDRWSDHCGKMDALCEQE KYKGPP	456	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
GI:589104031			
glycoside hydrolase family 61 protein [*T. reesei* QM6a] XP_006963879.1 GI:589103713	MKSCAILAALGCLAGSVLGHGQVQNFTINGQYNQGFILDYYYQKQNTGHFPNVAGWYAEDLDLG FISPDQYTTPDIVCHKNAAPGAISATAAAGSNIVFQWGPGVWPHPYGPIVTYVAECSGSCTTVN KNNLRWVKIQEAGINYNTQVWAQQDLINQGNKWTVKIPSSLRPGNYVFRHELLAAHGASSANGM QNYPQCVNIAVTGSGTKALPAGTPATQLYKPTDPGILFNPYTTITSYTIPGPALWQG	457	
glycoside hydrolase family 5 protein [*T. reesei* QM6a] XP_006962583.1 GI:589101121	MNKSVAPLLLAASILYGGAAAQQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCIPGATTIT TSTRPPSGPTTTTRATSTSSSTPPTSSGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGS NNYPDGIGQMQHFVNDDGMTIFRLPVGWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIV DIHNYARWNGGIIGQGGPTNAQFTSLWSQLASKYASQSRVWFGIMNEPHDVNINTWAATVQEVV TAIRNAGATSQFISLPGNDWQSAGAFISDGSAAALSQVTNPDGSTTNLIFDVHKYLDSDNSGTH AECTTNNIDGAFSPLATWLRQNNRQAILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWG AGSFDSTYVLTETPTGSGNSWTDTSLVSSCLARK	458	
glycoside hydrolase family 61 protein [*T. reesei* QM6a] XP_006961567.1 GI:589099089	MIQKLSNLLVTALAVATGVVGHGHINDIVINGVWYQAYDPTTFPYESNPPIVVGWTAADLDNGF VSPDAYQNPDIICHKNATNAKGHASVKAGDTILFQWVPVPWPHPGPIVDYLANCNGDCETVDKT TLEFFKIDGVGLLSGGDPGTWASDVLISNNNTWVVKIPDNLAPGNYVLRHEIIALHSAGQANGA QNYPQCFNIAVSGSGSLQPSGVLGTDLYHATDPGVLINIYTSPLNYIIPGPTVVSGLPTSVAQG SSAATATASATVPGGGSGPTSRTTTTARTTQASSRPSSTPPATTSAPAGGPTQTLYGQCGGSGY SGPTRCAPPATCSTLNPYYAQCLN	459	
Endoglucanase-7; also known as Cellulase-61B (Cel61B), Endo-1,4-beta-glucanase (EGVII); Endoglucanase VII; Endoglucanase-61B; Q7Z9M7.3 GI:43314396	MKSCAILAALGCLAGSVLGHGQVQNFTINGQYNQGFILDYYYQKQNTGHFPNVAGWYAEDLDLG FISPDQYTTPDIVCHKNAAPGAISATAAAGSNIVFQWGPGVWPHPYGPIVTYVVECSGSCTTVN KNNLRWVKIQEAGINYNTQVWAQQDLINQGNKWTVKIPSSLRPGNYVFRHELLAAHGASSANGM QNYPQCVNIAVTGSGTKALPAGTPATQLYKPTDPGILFNPYTTITSYTIPGPALWQG	460	
xylanase [*Trichoderma reesei*] CAA49293.1 GI:396564	MVSFTSLLAASPPSRASCRPAAEVESVAVEKRQTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPG GQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGSYNPNGNSYLSVYGWSRNPLIEYYIVENFGTY NPSTGATKLGEVTSDGSVYDIYRTQRVNQPSIIGTATFYQYWSVRRNHRSSGSVNTANHFNAWA QQGLTLGTMDYQIVAVEGYFSSGSASITVS	461	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
xylanase[*T. reesei*] CAA49294.1 GI:396566	MVAFSSLICALTSIASTLAMPTGLEPESSVNVTERGMYDFVLGAHNDHRRRASINYDQNYQTGG QVSYSPSNTGFSVNWNTQDDFVVGVGWTTGSSAPINFGGSFSVNSGTGLLSVYGWSTNPLVEYY IMEDNHNYPAQGTVKGTVTSDGATYTIWENTRVNEPSIQGTATFNQYISVRNSPRTSGTVTVQN HFNAWASLGLHLGQMNYQVVAVEGWGGSGSASQSVSN	462	
beta-xylanase precursor [*Trichoderma reesei*] AAB5278.1 GI:78816	MVSFTSLLAGVAAISGVLAAPAAEVEPVAVEKRQTIQPGTGYNNGYFHSYWNDGHGGVTYTNGP GGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGSYNPNGNSYLSVYGWSRNPLIEYYIVGNFGT YNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSIIGTATFYQYWSVRRNHRSSGSVNTANHFNAW AQQGLTLGTMDYQIVAVEGYFSSGSASITVS	463	
Chain A, Structural Comparison Of Two Major Endo-1,4-Beta-Xylanases From *Trichodrema Reesei* 1XYP_A GI:112721	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	464	
Chain B, Structural Comparison Of Two Major Endo-1,4-Beta-Xylanases From *Trichodrema Reesei* 1XYP_B GI:1127211	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	465	
Chain A, Structural Comparison Of Two Major Endo-1,4-Beta-Xylanases From *Trichodrema Reesei* 1XYO_A GI:1127212	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	466	
Chain B, Structural Comparison Of Two	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	467	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Major Endo-1,4-Beta-Xylanases From *Trichodrema Reesei* 1XYO_B GI:1127213			
Chain A, Structural Comparison Of Two Major Endo-1,4-Beta-Xylanases From *Trichodrema Reesei* 1ENX_A GI:1127272	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	468	
Chain B, Structural Comparison Of Two Major Endo-1,4-Beta-Xylanases From *Trichodrema Reesei* 1ENX_B GI:1127273	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	469	
Chain A, Endo-1,4-beta-xylanase Ii Complex With 4,5-epoxypentyl-beta-D-xyloside 1RED_A GI:1942592	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	470	
Chain B, Endo-1,4-beta-xylanase Ii Complex With 4,5-epoxypentyl-beta-D-xyloside 1RED_B GI:1942593	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	471	
Chain A, Endo-1,4-Beta-Xylanase Ii Complex With 3,4-Epoxybutyl-Beta-D-	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	472	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Xyloside 1REE_A GI:1942594			
Chain B, Endo-1,4-Beta-Xylanase Ii Complex With 3,4-Epoxybutyl-Beta-D-Xyloside 1REE_B GI:1942595	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	473	
Chain A, Endo-1,4-Beta-Xylanase Ii Complex With 2,3-Epoxypropyl-Beta-D-Xyloside 1REF_A GI:1942596	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	474	
Chain B, Endo-1,4-Beta-Xylanase Ii Complex With 2,3-Epoxypropyl-Beta-D-Xyloside 1REF_B GI:1942597 xylanase III [*T. reesei*] BAA89465.2 GI:7328936	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	475	
xylanase, partial [*T. reesei*] AAG01167.1 GI:9858850	SYLSVYGWXTDPLIEYYIVESYGDYNPGSGGTYKGTVTSDGSVYDIYTATRTNAASIQGTATFT QYWSVRR	476	
Transcription factor ACEII [*T. reesei*] AAK69383.1 GI:14581734	MDLRQACDRCHDKKLRCPRISGSPCCSRCAKANVACVFSPPSRPFRPHEPLNHSHEHSHSHSHN HNGVGVSFDWLDLMSLEQQQEQQQGQPQHPPPPVQTLSERLAALLCALDRMLQAVPSSLDMHHV SRQQLREYADTVGTGFDLQSTLDSLLHHAQDLASLYSEAVPASFNKRTTAAEADALCAVPDCVH QDRTSLHTTPLPKLDHALLNLVMACHIRLLDVMDTLAEHGRMCAFMVATLPPDYDPKFAVPEIR VGTFVAPTDTAASMLLSVVVELQTVLVARVKDLVAMVDQVKDDARAAREAKVVRLQCGILLERA ESTLGEWSRFKDGLVSARLLK	477	
xylanase regulator 1, partial [*Trichoderma*	MLSNPLRRYSAYPDISSASFDPNYHGSQSHLHSINVNTFGNSHPYPMQHLAQHAELSSSRMIRA SPVQPKQRQGSLIAARKNSTGTAGPIRRRISRACDQCNQLRTKCDGLHPCAHCIEFGLGCEYVR ERKKRGKASRKDIAAQQAAAAAAQHSGQVQDGPEDQHRKLSRQQSESSRGSAELAQPAHDPPHG	478	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
reesei] AAO33577.1 GI:28194501	HIEGSVSSFSDNGLSQHAAMGGMDGLEDHHGHVGVDPALGRTQLEASSAMGLGAYGEVHPGYES PGMNGHVMVPPSYGAQTTMAGYSGISYAAQAPSPATYSSDGNFRLTGHIHDYPLANGSSPSWGV SLASPSLRYHVLRPVLLDVRNIYPVSLACDQMDMYFSSSSSAQMRPMSPYVEGFVFRKRSFLHP TDPRRCQPALLASMLWVAAQTSEASFLTSLPSARSKVCQKLLELTVGLLQPLIHTGTNSPSPKT SPVVGAAALGVLGVAMPGSLNMDSLAGETGAFGAIGSLDDVIAYVRLATVVSASEYKGASLRWW GAAWSLARELKLGRELPPGNPPANQEDGEGLSEDVDEHDLNRNNTRLGRKRSAKSDAITEEERE ERRRAWWLVYIVDRHLALCYNRPLFLLDSECSDLYHPMDDIKWQAGKFRSHDAGNSSINIDSSM TDEFGDSPRAARGAHYECRGRSIFGYFLSLMTILGEIVDVHHAKSHPRFGVGFRSARDWDEQVA EITRHLDMYEESLKRFVAKHLPLSSKDKEQHEMHDSGAVTDMQSPLSVRTNASSRMTESEIQAS IVVAYSTHVMHVLHILLADKWDPINLLDDDDLWISSEGFVTATSHAVSAAEAISQILEFDPGLE FMPFFFGIYLLQGSFLLLLIADKLQAEASPSVIKACETIVRAHEACVVTLSTEYQRNFSKVMRS ALALIRGRVPEDLAEQQQRRRELLALYRWTGNGTGLAL		
Transcription factor ACEII protein Q96WN6.1 GI:50400614	MDLRQACDRCHDKKLRCPRISGSPCCSRCAKANVACVFSPPSRPFRPHEPLNHSHEHSHSHSHN HNGVGVSFDWLDLMSLEQQQEQQQGQPQHPPPPVQTLSERLAALLCALDRMLQAVPSSLDMHHV SRQQLREYADTVGTGFDLQSTLDSLLHHAQDLASLYSEAVPASFNKRTTAAEADALCAVPDCVH QDRTSLHTTPLPKLDHALLNLVMACHIRLLDVMDTLAEHGRMCAFMVATLPPDYDPKFAVPEIR VGTFVAPTDTAASMLLSVVVELQTVLVARVKDLVAMVDQVKDDARAAREAKVVRLQCGILLERA ESTLGEWSRFKDGLVSARLLK	479	
Chain A, Structure Of Vil-Xylanase 2D97_A GI:112490431	TIQPGTGXNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGS YNPNGNSXLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII GTATFXQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGXFSSGSASITVS	480	
Chain A, Structure Of Vil (Extra KiI2 ADDED)-Xylanase 2D98_A, GI:112490433	TIQPGTGXNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGS YNPNGNSXLSVYGWSRNPLIEYYIVENFGTXNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII GTATFXQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDXQIVAVEGXFSSGSASITVS	481	
Chain A, Xylanase Ii From *Tricoderma Reesei* At 100k 2DFB_A GI:112490475	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	482	
Chain A, Xylanase Ii From T. Reesei At 293k 190 aa protein	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	483	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
2DFC_A GI:112490477			
TPA_inf: chitinase 18-12 [*T. reesei*] DAA05860.1 GI:126032263	MPSLTALAGLLALVPSALAGWNPDSKQNIAVYWGQNSANSQSTQQRLSFYCNDDNINVIEIAFL NGINPPMTNFANAGDRCTPFSDNPWLLSCPEIEADIKTCQANGKTILLSLGGDTYSQGGWASPE AAQDAAAQVWAMFGPVQSDSSAPRPFGDAVVDGFDFDFESTTNNLVAFGAQLRTLSDAAATDSN KKFYLAAAPQCFFPDAAVGPLINAVPMDWIQIQFYNNPCGVSAYTPGSEQQNNYNYQTWEDWAK TSPNPNVKLLVGIPAGPNAGHGYVSDAQLKSVFEYSKKFDTFAGAMMWDMSQLYQNSGFEDQVV DALK	484	
TPA_inf: chitinase 18-12 [*T. reesei*] DAA05860.1 GI:126032263	MPSLTALAGLLALVPSALAGWNPDSKQNIAVYWGQNSANSQSTQQRLSFYCNDDNINVIEIAFL NGINPPMTNFANAGDRCTPFSDNPWLLSCPEIEADIKTCQANGKTILLSLGGDTYSQGGWASPE AAQDAAAQVWAMFGPVQSDSSAPRPFGDAVVDGFDFDFESTTNNLVAFGAQLRTLSDAAATDSN KKFYLAAAPQCFFPDAAVGPLINAVPMDWIQIQFYNNPCGVSAYTPGSEQQNNYNYQTWEDWAK TSPNPNVKLLVGIPAGPNAGHGYVSDAQLKSVFEYSKKFDTFAGAMMWDMSQLYQNSGFEDQVV DALK	485	
TPA_inf: chitinase 18-13 [*T. reesei*] DAA5861.1 GI:126032265	MFFSKALAAAGLLATAAYAAPTMEKRAAGGKLVVYWGAEDDSTTLANVCADSSYDIVNLAFLSR FFAGGGYPELSLSTLGGPSAAQRAAGATNLQDGTSLIPAIQACQAAGKLVILSMGGAVDFSAVT LSSDAQGQQLADTVWNLFLGGTANPTLRPFGSVKLDGVDLDNETGNPTGYLAMAQRFKSNFAKD TSKKYYLTAAPQCPFPDASEPLNVCQLADYIWVQFYNNGNCNIAQSGFNNAVKNWSKSIGNATL FIGALASGADGDQGYVSASSLLSAYQGVSALNLPNIGGIMLWEAQLAVKNGNFQKTVKAGIASG TTPPPPPPTGGCSWAGHCAGASCSTDNDCSDDLTCNGGVCGTAGSTPAPTCSWEGHCLGASCGN DNDCSDPYSCKNGVCSN	486	
TPA_inf: chitinase 18-14 [*T. reesei*] DAA05862.1 GI:126032267	MFFTKAVGGLGLLASLASSAPNPIARRQAPGAQNVVYWGQNGGGTVENNDLSAYCTPTSGIDII VLSFLYQWGQGSSALGGTIGQSCGITTSGEPQNCDALTAAITKCKTAGVKIILSLGGASAFSSF QTADQAAQAGQYLWNAYGGGSGVTRPLGNNVMDGFDLDIESNPGTNENYAALVSALRSNFASDP SRQYVISGAPQCPLPEPNMGVIIQNAQFDYLWVQFYNNNEYPGDPCSLGLPGDAPFNFNNWTTF IQSTPSKDAKVFVGVPAAPLAANGAPSGEVYYATPSQLADIVNDVKSNPAFGGIMMWSAGFSDT NVNDGCNYAQEAKNILLTGSPCSSGPVSVSRPPVSSPTITSSPPGTSPAPPSQTGSVPQWGQCG GNGYTGPTQCVAPFKCVATSEWWSQCE	487	
TPA_inf: chitinase 18-16 [*T. reesei*] DAA05864.1 GI:126032269	MLSRTLLTALGLTTIAAAAPSQTVKTRQAPGGQNAVYWGATNNENDNLSTYCTASSGIDIVILS FLDIYGATGNFPSGNMGNSCYVGTNGVPQLCDDLASSIATCQAAGIKVIISLGGAASSYSLQSQ SQAVAIGQYLWNAYGNSGNTTVQRPFGNVFVNGFDFDIELNAGSQYYQYLISTLRSNFANDPKN TYYITGAPQCPIPEPNMGEIISTSQFDYLWVQFYNNNPVCSLGLPGDAPFNFNDWVSFISTTPS KNAKLFVGAPASTLGANGNAGGAKYYATPEQLAGIVNSVKSSPFFGGIMLWDAGYSDSNVNNGC NYAQEAKNILLTGTACGGESSPPPSTTTTAVPPPASSTPSNPSGGSVPQWGQCGGDGYTGPTQC VAPYKCVATSEWWSSCQ	488	
TPA_inf: chitinase 18-18 [*T. reesei*]	MVSASAGLAAVGLLNGYWGQYTTTEGLRPHCDSGVDSITLGFVNGAPDASGYPSLNFGPNCWAE SYPGNLGLPSKLLSHCMSLQSDIPYCRSKGVKVILSIGGVYNALTSNYFVGDNGTATDFATFLY	489	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
DAA05866.1 GI:126032275	NAFGPYNASYTGPRPFDDITTGLPTSVDGFDFDIEADFPNGPYIKMIETFRSLDSSMLITGAPQ CPTNPQYFVMKDMIQQAAFDKLFIQFYNNPVCDAIPGNTAGDKFNYDDWEAVIAGSAKSKSAKL YIGLPAIQEPNESGYIDPIAMKNLVCQYKDRPHFGGLSLWDLSRGLVNNINGTSFNQWALDALQ YGCNPIPTTTTTTSTVSSTTAASSTTASSTTASTTKASSTSKASSTSKASSTSKASSTSKASSTSKASST SKASSTSKASSTSKASTTSKASTTSKVSTTSKASSTSKASSSTKASTTSKASSTSKASTTSKAS TTSKASTTSKASTTSKASTTSKASSTSKASSSTKASTTSKASSTSKASTTSKASTTSKASTTSK ASTTSKASTTSKASSTSKASSSTKASTTSKASSTSKASTTSKASTTSKASTTSKASTTSKASTT SKASTTSKASTTSKASTTSKASTTSKASTTSKASTTSKASTTSKVSTTSKASTTSKASTTSKAS STSKVSTTSKASTTSKVSAKATTSTKASTTVKPSTTSKASTTSKASTTSKASTTSKASTTSKAS TTSKASTTSKASTTSKAATTSVKPTSKTSTSSKPNVSASSSNVGRDATSLVEASTSTSAAVLYP TTTSRWSNSTITRSSSLTTPIVSDPASLTTSVVYTTSVHTVTKCPAYVTDCPAGGYVTTETIPL YTTVCPISEATQTAAPTVTTEAPQPWTTSTVYTTRVYTITSCAPGVVDCPANQVTTETIPWYTT VCPVTATATPVGPGSVVFPQNTEVGQPSLVGPVVEAAYPTASSSLQTLVKPATSVGVPQGSPAG SSVAPGSSSKPTAPAGPPSYPTGGSGNASPSGSWSGVPVGPSSVPGIPEANAASVMSASLFGLV IVMAAQVFVL		
Chain A, Structural Comparison Of Two Major Endo-1,4-Beta-Xylanases From *T. Reesei* 1XYN_A GI:157834272	ASINYDQNYQTGGQVSYSPSNTGFSVNWNTQDDFVVGVGWTTGSSAPINFGGSFSVNSGTGLLS VYGWSTNPLVEYYIMEDNHNYPAQGTVKGTVTSDGATYTIWENTRVNEPSIQGTATFNQYISVR NSPRTSGTVTVQNHFNAWASLGLHLGQMNYQVVAVEGWGGSGSASQSVSN	490	
xylanase, partial [*Trichoderma reesei*] ACB38137.1 GI:170786291	QTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	491	
Chain A, Xylanase Ii From *T. Reesei* Cocrystallized With Tris-Dipicolinate Europium 3LGR_A GI:319443539	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	492	
Chain A, Crystal Structures Of	TIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGS YNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII	493	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Mutant Endo-1,4-xylanase Ii Complexed With Substrate (1.15 A) And Products (1.6A) 4HK8_A GI:572153255	GTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVQGYFSSGSASITVS		
Chain A, Crystal Structures Of Mutant Endo-beta-1,4-xylanase Ii Complexed With Substrate (1.15 A) And Products (1.6A) 4HK9_A GI:572153256	IQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGHFVGGKGWQPGTKNKVINFSGSY NPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSIIG TATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	494	
Chain A, Crystal Structures Of Mutant Endo-beta-1,4-xylanase Ii Complexed With Substrate (1.15 A) And Products (1.6A) 4HKL_A GI:572153257	TIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGHFVGGKGWQPGTKNKVINFSGS YNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII GTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	495	
Chain A, Crystal Structures Of Mutant Endo-beta-1,4-xylanase Ii (e177q) In Apo Form 4HKO_A GI:572153258	TIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINNPLI EYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSIIGTATFYQYWSVRRNHRSSGS VNTANHFNAWAQQGLTLGTMDYQIVAVQGYFSSGSASITVS	496	
Chain A, Crystal Structures Of	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI	497	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Mutant Endo-beta-1,4-xylanase Ii Complexed With Substrate And Products 4HKW_A GI:572153259	IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS		
Chain A, Joint X-ray/neutron Structure Of Trichoderma Reesei Xylanase Ii In Complex With Mes At Ph 5.7 4S2D_A GI:929984639	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	498	
Chain A, Joint X-ray/neutron Structure Of *T. Reesei* Xylanase Ii At Ph 4.4 4S2F_A GI:929984640	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	499	
Chain A, Joint X-ray/neutron Structure Of T. Reesei Xylanase Ii At Ph 5.8 4S2G_A GI:929984641	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	500	
Chain A, Joint X-ray/neutron Structure Of T. Reesei Xylanase Ii At Ph 8.5 4S2H_A GI:929984642	XTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	501	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Chain A, X-ray Structure Analysis Of Xylanase - N44d 4XQ4_A GI:929984784	TIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGDFVGGKGWQPGTKNKVINFSGS YNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII GTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	502	
Chain B, X-ray Structure Analysis Of Xylanase - N44d 4XQ4_B GI:929984785	TIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGDFVGGKGWQPGTKNKVINFSGS YNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII GTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	503	
Chain A, X-ray Structure Analysis Of Xylanase-wt At Ph4.0 4XQD_A GI:929984786	TIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGS YNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII GTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	504	
Chain B, X-ray Structure Analysis Of Xylanase-wt At Ph4.0 4XQD_B GI:929984787	TIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGS YNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII GTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	505	
Chain A, X-ray Structure Analysis Of Xylanase-n44e With Mes At Ph6.0 4XQW_A GI:929984788	TIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGEFVGGKGWQPGTKNKVINFSGS YNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII GTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	506	
Chain A, Neutron And X-ray Structure Analysis Of Xylanase: N44d At Ph6 4XPV_A GI:931139811	TIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGDFVGGKGWQPGTKNKVINFSGS YNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSII GTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	507	
truncated xylanase 5 [*T. reesei*]	MVSFSSLVVALVGIASSWALWRNPSTQT	508	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
ANW82584.1 GI:1048222282			
truncated xylanase 5 [T. reesei] ANW82585.1 GI:1048222284	MVSFSSLVVALVGIASSWALWRNPSTQT	509	
xylanase[*T. reesei*] ANX99792.1 GI:1049178838	MVSFSSLVVALVGIASSWAAPLEESPNANITERGPSNFVLGGHNAVRRAAINYNQDYTTGGDVV YTHSNTGFAVNWSYPNDFVVGVGWNPGGSAPINFSGNFGVGSGVGLLSVYGWSTNPLVEYYVVE DNFGFSSGGTVKGSVTSDGSSYTIWENTRVNEPSIVGTATFNQYISIRNSKRSSGTVTVANHFN AWKSLGMNLGTMNYQVIAVEGWGGQGGVQQTVSN	510	
xylanase[*T. reesei*] ANX99793.1 GI:1049178840	MVSFSSLVVALVGIASSLAAPLEESLNANITERGPNNFVLGGHNAVRRAAINYNQDYTTGGDVV YTHSNTGFAVNWSYPNDFVVGVGWNPGGSAPINFSGNFGVGSGVGLLSVYGWSTNPLVEYYVVE DNFGFSSGGTVKGSVTSDGSSYTIWENTRVNEPSIVGTATFNQYISIRNSKRSSGTVTIANHFN AWKSLGMNLGTLNYQVIAVEGWGGQGGVQQTVSN	511	
xylanase[*T. reesei*] ANX99794.1 GI:1049178842	MVSFSSLVVALVGIASSLAAPLEESLNANITERGPNNFVLGGHNAVRRAAINYNQDYTTGGDVV YTHSNTGFAVNWSYPNDFVVGVGWNPGGSAPINFSGNFGVGSGVGLLSVYGWSTNPLVEYYVVE DNFGFSSGGTVKGSVTSDGSSYTIWENTRVNEPSIVGTATFNQYISIRNSKRSSGTVTIANHFN AWKSLGMNLGTLNYQVIAVEGWGGQGGVQQTVSN	512	
xylanase[T. reesei] ANX99795.1 GI:1049178844	MVSFSSLVVALVGIASSLAAPLEESLNANITERGPNNFVLGGHNAVRRAAINYNQDYTTGGDVV YTHSNTGFAVNWSYPNDFVVGVGWNPGGSAPINFSGNFGVGSGVGLLSVYGWSTNPLVEYYVVE DNFGFSSGGTVKGSVTSDGSSYTIWENTRVNEPSIVGTATFNQYISIRNSKRSSGTVTIANHFN AWKSLGMNLGTLNYQVIAVEGWGGQGGVQQTVSN	513	
xylanase 2, partial [*T. reesei*] APU51339.1 GI:1130479396	QTIQPGTGYNNGYCYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWCPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	514	
xylanase 2, partial [*T. reesei*] APU51340.1 GI:1130479398	QTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	515	
Chain A, Microed Structure Of Xylanase At 2.3 A Resolution 5K7P_A	QTIQPGTGYNNGYFYSYWNDGHGGVTYTNGPGGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSG SYNPNGNSYLSVYGWSRNPLIEYYIVENFGTYNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSI IGTATFYQYWSVRRNHRSSGSVNTANHFNAWAQQGLTLGTMDYQIVAVEGYFSSGSASITVS	516	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
GI:1175128641			
glycoside hydrolase family 18 protein, chitinase [*T. reesei* QM6a] EGR44650.1 GI:340514387	MFFTKAVGGLGLLASLASSAPNPIARRQAPGAQNVVYWGQNGGGTVENNDLSAYCTPTSGIDII VLSFLYQWGQGSSALGGTIGQSCGITTSGEPQNCDALTAAITKCKTAGVKIILSLGGASAFSSF QTADQAAQAGQYLWNAYGGGSGVTRPLGNNVMDGFDLDIESNPGTNENYAALVSALRSNFASDP SRQYVISGAPQCPLPEPNMGVIIQNAQFDYLWVQFYNNNEYPGDPCSLGLPGDAPFNFNNWTTF IQSTPSKDAKVFVGVPAAPLAANGAPSGEVYYATPSQLADIVNDVKSNPAFGGIMMWSAGFSDT NVNDGCNYAQEAKNILLTGSPCSSGPVSVSRPPVSSPTITSSPPGTSPAPPSQTGSVPQWGQCG GNGYTGPTQCVAPFKCVATSEWWSQCE	517	
glycoside hydrolase family 5 protein[*T. reesei* QM6a] EGR44819.1 GI:340514558	MKSSISVVLALLGHSAAWSYATKSQYRANIKINARQTYQTMIGGGCSGAFGIACQQFGSSGLSP ENQQKVTQILFDENIGGLSIVRNDIGSSPGTTILPTCPATPQDKFDYVWDGSDNCQFNLTKTAL KYNPNLYVYADAWSAPGCMKTVGTENLGGQICGVRGTDCKHDWRQAYADYLVQYVRFYKEEGID ISLLGAWNEPDFNPFTYESMLSDGYQAKDFLEVLYPTLKKAFPKVDVSCCDATGARQERNILYE LQQAGGERYFDIATWHNYQSNPERPFNAGGKPNIQTEWADGTGPWNSTWDYSGQLAEGLQWALY MHNAFVNSDTSGYTHWWCAQNTNGDNALIRLDRDSYEVSARLWAFAQYFRFARPGSVRIGATSD VENVYVTAYVNKNGTVAIPVINAAHFPYDLTIDLEGIKKRKLSEYLTDNSHNVTLQSRYKVSGS SLKVTVEPRAMKTFWLEPQSTFAVI	518	
glycoside hydrolase family 11 [*T. reesei* QM6a] EGR45030.1 GI:340514771	MVSFTSLLAGVAAISGVLAAPAAEVESVAVEKRQTIQPGTGYNNGYFYSYWNDGHGGVTYTNGP GGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGSYNPNGNSYLSVYGWSRNPLIEYYIVENFGT YNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSIIGTATFYQYWSVRRNHRSSGSVNTANHFNAW AQQGLTLGTMDYQIVAVEGYFSSGSASITVS	519	
glycoside hydrolase family 18, chitinase [*T. reesei* QM6a] EGR45486.1 GI:340515230	MLSRTLLTALGLTTIAAAAPSQTVKTRQAPGGQNAVYWGATNNENDNLSTYCTASSGIDIVILS FLDIYGATGNFPSGNMGNSCYVGTNGVPQLCDDLASSIATCQAAGIKVIISLGGAASSYSLQSQ SQAVAIGQYLWNAYGNSGNTTVQRPFGNVFVNGFDFDIELNAGSQYYQYLISTLRSNFANDPKN TYYITGAPQCPIPEPNMGEIISTSQFDYLWVQFYNNNPVCSLGLPGDAPFNFNDWVSFISTTPS KNAKLFVGAPASTLGANGNAGGAKYYATPEQLAGIVNSVKSSPFFGGIMLWDAGYSDSNVNNGC NYAQEAKNILLTGTACGGESSPPPSTTTTAVPPPASSTPSNPSGGSVPQWGQCGGDGYTGPTQC VAPYKCVATSEWWSSCQ	520	
xylanase regulator 1 [*T. reesei* QM6a] EGR48040.1 GI:340517797	MLSNPLRRYSAYPDISSASFDPNYHGSQSHLHSINVNTFGNSHPYPMQHLAQHAELSSSRMIRA SPVQPKQRQGSLIAARKNSTGTAGPIRRRISRACDQCNQLRTKCDGLHPCAHCIEFGLGCEYVR ERKKRGKASRKDIAAQQAAAAAAQHSGQVQDGPEDQHRKLSRQQSESSRGSAELAQPAHDPPHG HIEGSVSSFSDNGLSQHAAMGGMDGLEDHHGHVGVDPALGRTQLEASSAMGLGAYGEHPGYESP GMNGHVMVPPSYGAQTTMAGYSGISYAAQAPSPATYSSDGNFRLTGHIHDYPLANGSSPSWGQS DLRYPVLEPLLPHLGNILPVSLACDLIDLYFSSSSSAQMHPMSPYVLGFVFRKRSFLHPTNPRR CQPALLASMLWVAAQTSEASFLTSLPSARSKVCQKLLELTVGLLQPLIHTGTNSPSPKTSPVVG	521	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAALGVLGVAMPGSLNMDSLAGETGAFGAIGSLDDVITYVHLATVVSASEYKGASLRWWGAAWS LARELKLGRELPPGNPPANQEDGEGLSEDVDEHDLNRNNTRFVTEEEREERRRAWWLVYIVDRH LALCYNRPLFLLDSECSDLYHPMDDIKWQAGKFRSHDAGNSSINIDSSMTDEFGDSPRAARGAH YECRGRSIFGYFLSLMTILGEIVDVHHAKSHPRFGVGFRSARDWDEQVAEITRHLDMYEESLKR FVAKHLPLSSKDKEQHEMHDSGAVTDMQSPLSVRTNASSRMTESEIQASIVVAYSTHVMHVLHI LLADKWDPINLLDDDDLWISSEGFVTATSHAVSAAEAISQILEFDPGLEFMPFFYGVYLLQGSF LLLLIADKLQAEASPSVIKACETIVRAHEACVVTLSTEYQRNFSKVMRSALALIRGRVPEDLAE QQQRRRELLALYRWTGNGTGLAL		
predicted protein, partial [*T. reesei* QM6a] EGR49987.1 GI:340519749	LRILPVGDSITYGFLSDQDGGDGNGYRLQLRQHLSKDRVVFAGTETSGNMTDGYYLIVSSSLHR QAAWNGKTIQYISDHVTPSLEQRPNIILLHAGTNDMNPNGAISREGHDPVAASERLGSLVDKMT TLCPDAVILVAMIIGTCNDEQAPQTKVFQSLIPNVVAPRLESGKHVLAVDFSTFPLDKLRDCIH PTNEGYHLLGYYWYDFIAQIPRDWITAPVGEDPQRPEEQNLAMRLETDL	522	
Transcription factor [*T. reesei* QM6a] EGR51484.1 GI:340521249	MSFSNPRRRTPVTRPGTDCEHGLSLKTTMTLRKGATFHSPTSPSASSAAGDFVPPTLTRSQSAF DDVVDASRRRIAMTLNDIDEALSKASLSDKSPRPKPLRDTSLPVPRGFLEPPVVDPAMNKQEPE RRVLRPRSVRRTRNHASDSGIGSSVVSTNDKAGAADSTKKPQASALTRSAASSTTAMLPSLSHR AVNRIREHTLRPLLEKPTLKEFEPIVLDVPRRIRSKEIICLRDLEKTLIFMAPEKAKSAALYLD FCLTSVRCIQATVEYLTDREQVRPGDRPYTNGYFIDLKEQIYQYGKQLAAIKEKGSLADDMDID PSDEVRLYGGVAENGRPAELIRVKKDGTAYSMATGKIVDMTESPTPLKRSLSEQREDEEEIMRS MARRKKNATPEELAPKKCREPGCTKEFKRPCDLTKHEKTHSRPWKCPIPTCKYHEYGWPTEKEM DRHINDKHSDAPAMYECLFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKAPSQNGSTAQQ TPPLANVSTPSSTPSYSVPTPPQDQVMSTDFPMYPADDDWLATYGAQPNTIDAMDLGLENLSPA SAASSYEQYPPYQNGSTFIINDEDIYAAHVQIPAQLPTPEQVYTKMMPQQMPVYHVQQEPCTTV PILGEPQFSPNAQQNAVLYTPTSLREVDEGFDESYAADGADFQLFPATVDKTDVFQSLFTDMPS ANLGFSQTTQPDIFNQIDWSNLDYQGFQE	523	
Glycoside hydrolase family 10 protein [*T. reesei* QM6a] EGR52056.1 GI:340521822	MKANVILCLLAPLVAALPTETIHLDPELAALRANLTERTADLWDRQASQSIDQLIKRKGKLYFG TATDRGLLQREKNAAIIQADLGQVTPENSMKWQSLENNQGQLNWGDADYLVNFAQQNGKSIRGH TLIWHSQLPAWVNNININNADTLRQVIRTHVSTVVGRYKGKIRAWDVVNEIFNEDGTLRSSVFSRL LGEEFVSIAFRAARDADPSARLYINDYNLDRANYGKVNGLKTYVSKWISQGVPIDGIGSQSHLS GGGGSGTLGALQQLATVPVTELAITELDIQGAPTTDYTQVVQACLSVSKCVGITVWGISDKDSW RASTNPLLFDANFNPKPAYNSIVGILQ	524	
Glycoside hydrolase family 18 protein, chitinase [*T. reesei* QM6a]	MFFSKALAAAGLLATAAYAAPTMEKRAAGGKLVVYWGAEDDSTTLANVCADSSYDIVNLAFLSR FFAGGGYPELSLSTLGGPSAAQRAAGATNLQDGTSLIPAIQACQAAGKLVILSMGGAVDFSAVT LSSDAQGQQLADTVWNLFLGGTANPTLRPFGSVKLDGVDLDNETGNPTGYLAMAQRFKSNFAKD TSKKYYLTAAPQCPFPDASEPLNVCQLADYIWVQFYNNGNCNIAQSGFNNAVKNWSKSIGNATL FIGALASGADGDQGYVSASSLLSAYQGVSALNLPNIGGIMLWEAQLAVKNGNFQKTVKAGIASG	525	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
EGR52465.1 GI:340522232	TTPPPPPPTGGCSWAGHCAGASCSTDNDCSDDLTCNGGVCGTAGSTPAPTCSWEGHCLGASCGN DNDCSDPYSCKNGVCSN		
Glycoside hydrolase family 18 proetin [*T. reesei* QM6a] EGR52759.1 GI:340522526	MPSLTALAGLLALVPSALAGWNPDSKQNIAVYWGQNSANSQSTQQRLSFYCNDDNINVIEIAFL NGINPPMTNFANAGDRCTPFSDNPWLLSCPEIEADIKTCQANGKTILLSLGGDTYSQGGWASPE AAQDAAAQVWAMFGPVQSDSSAPRPFGDAVVDGFDFDFESTTNNLVAFGAQLRTLSDAAATDSN KKFYLAAAPQCFFPDAAVGPLINAVPMDWIQIQFYNNPCGVSAYTPGSEQQNNYNYQTWEDWAK TSPNPNVKLLVGIPAGPNAGHGYVSDAQLKSVFEYSKKFDTFAGAMMWDMSQLYQNSGFEDQVV DALK	526	
Glycoside hydrolase family 11 protein [*T. reesei* QM6a] EGR52985.1 GI:340522752	MVAFSSLICALTSIASTLAMPTGLEPESSVNVTERGMYDFVLGAHNDHRRRASINYDQNYQTGG QVSYSPSNTGFSVNWNTQDDFVVGVGWTTGSSAPINFGGSFSVNSGTGLLSVYGWSTNPLVEYY IMEDNHNYPAQGTVKGTVTSDGATYTIWENTRVNEPSIQGTATFNQYISVRNSPRTSGTVTVQN HFNAWASLGLHLGQMNYQVVAVEGWGGSGSASQSVSN	527	
Glycoside hydrolase family 18 protein [*T. reesei* QM6a] XP_006961069.1 GI:589098093	MPSLTALAGLLALVPSALAGWNPDSKQNIAVYWGQNSANSQSTQQRLSFYCNDDNINVIEIAFL NGINPPMTNFANAGDRCTPFSDNPWLLSCPEIEADIKTCQANGKTILLSLGGDTYSQGGWASPE AAQDAAAQVWAMFGPVQSDSSAPRPFGDAVVDGFDFDFESTTNNLVAFGAQLRTLSDAAATDSN KKFYLAAAPQCFFPDAAVGPLINAVPMDWIQIQFYNNPCGVSAYTPGSEQQNNYNYQTWEDWAK TSPNPNVKLLVGIPAGPNAGHGYVSDAQLKSVFEYSKKFDTFAGAMMWDMSQLYQNSGFEDQVV DALK	528	
Glycoside hydrolase family 18 protein, chitinase [*T. reesei* QM6a] XP_006961376.1 GI:589098707	MFFSKALAAAGLLATAAYAAPTMEKRAAGGKLVVYWGAEDDSTTLANVCADSSYDIVNLAFLSR FFAGGGYPELSLSTLGGPSAAQRAAGATNLQDGTSLIPAIQACQAAGKLVILSMGGAVDFSAVT LSSDAQGQQLADTVWNLFLGGTANPTLRPFGSVKLDGVDLDNETGNPTGYLAMAQRFKSNFAKD TSKKYYLTAAPQCPFPDASEPLNVCQLADYIWVQFYNNGNCNIAQSGFNNAVKNWSKSIGNATL FIGALASGADGDQGYVSASSLLSAYQGVSALNLPNIGGIMLWEAQLAVKNGNFQKTVKAGIASG TTPPPPPPTGGCSWAGHCAGASCSTDNDCSDDLTCNGGVCGTAGSTPAPTCSWEGHCLGASCGN DNDCSDPYSCKNGVCSN	529	
Glycoside hydrolase family 11 protein [*T. reesei* QM6a] XP_006961811.1 GI:589099577	MVAFSSLICALTSIASTLAMPTGLEPESSVNVTERGMYDFVLGAHNDHRRRASINYDQNYQTGG QVSYSPSNTGFSVNWNTQDDFVVGVGWTTGSSAPINFGGSFSVNSGTGLLSVYGWSTNPLVEYY IMEDNHNYPAQGTVKGTVTSDGATYTIWENTRVNEPSIQGTATFNQYISVRNSPRTSGTVTVQN HFNAWASLGLHLGQMNYQVVAVEGWGGSGSASQSVSN	530	
Glycoside	MKANVILCLLAPLVAALPTETIHLDPELAALRANLTERTADLWDRQASQSIDQLIKRKGKLYFG	531	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
hydrolase family 10 [*T. reesei* QM6a] XP_006962419.1 GI:589100793	TATDRGLLQREKNAAIIQADLGQVTPENSMKWQSLENNQGQLNWGDADYLVNFAQQNGKSIRGH TLIWHSQLPAWVNNIINNADTLRQVIRTHVSTVVGRYKGKIRAWDVVNEIFNEDGTLRSSVFSRL LGEEFVSIAFRAARDADPSARLYINDYNLDRANYGKVNGLKTYVSKWISQGVPIDGIGSQSHLS GGGGSGTLGALQQLATVPVTELAITELDIQGAPTTDYTQVVQACLSVSKCVGITVWGISDKDSW RASTNPLLFDANFNPKPAYNSIVGILQ		
Transcription factor protein [*T. reesei* QM6a] XP_006962963.1 GI:589101881	MSFSNPRRRTPVTRPGTDCEHGLSLKTTMTLRKGATFHSPTSPSASSAAGDFVPPTLTRSQSAF DDVVDASRRRIAMTLNDIDEALSKASLSDKSPRPKPLRDTSLPVPRGFLEPPVVDPAMNKQEPE RRVLRPRSVRRTRNHASDSGIGSSVVSTNDKAGAADSTKKPQASALTRSAASSTTAMLPSLSHR AVNRIREHTLRPLLEKPTLKEFEPIVLDVPRRIRSKEIICLRDLEKTLIFMAPEKAKSAALYLD FCLTSVRCIQATVEYLTDREQVRPGDRPYTNGYFIDLKEQIYQYGKQLAAIKEKGSLADDMDID PSDEVRLYGGVAENGRPAELIRVKKDGTAYSMATGKIVDMTESPTPLKRSLSEQREDEEEIMRS MARRKKNATPEELAPKKCREPGCTKEFKRPCDLTKHEKTHSRPWKCPIPTCKYHEYGWPTEKEM DRHINDKHSDAPAMYECLFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKAPSQNGSTAQQ TPPLANVSTPSSTPSYSVPTPPQDQVMSTDFPMYPADDDWLATYGAQPNTIDAMDLGLENLSPA SAASSYEQYPPYQNGSTFIINDEDIYAAHVQIPAQLPTPEQVYTKMMPQQMPVYHVQQEPCTTV PILGEPQFSPNAQQNAVLYTPTSLREVDEGFDESYAADGADFQLFPATVDKTDVFQSLFTDMPS ANLGFSQTTQPDIFNQIDWSNLDYQGFQE	532	
Predicted protein, partial [*T. reesei* QM6a] XP_006964048.1 GI:589104051	LRILPVGDSITYGFLSDQDGGDGNGYRLQLRQHLSKDRVVFAGTETSGNMTDGYYLIVSSSLHR QAAWNGKTIQYISDHVTPSLEQRPNIILLHAGTNDMNPNGAISREGHDPVAASERLGSLVDKMT TLCPDAVILVAMIIGTCNDEQAPQTKVFQSLIPNVVAPRLESGKHVLAVDFSTFPLDKLRDCIH PTNEGYHLLGYYWYDFIAQIPRDWITAPVGEDPQRPEEQNLAMRLETDL	533	
Xylanase regulator 1 protein [*T. reesei* QM6a] XP_006966092.1 GI:589108139	MLSNPLRRYSAYPDISSASFDPNYHGSQSHLHSINVNTFGNSHPYPMQHLAQHAELSSSRMIRA SPVQPKQRQGSLIAARKNSTGTAGPIRRRISRACDQCNQLRTKCDGLHPCAHCIEFGLGCEYVR ERKKRGKASRKDIAAQQAAAAAAQHSGQVQDGPEDQHRKLSRQQSESSRGSAELAQPAHDPPHG HIEGSVSSFSDNGLSQHAAMGGMDGLEDHHGHVGVDPALGRTQLEASSAMGLGAYGEVHPGYES PGMNGHVMVPPSYGAQTTMAGYSGISYAAQAPSPATYSSDGNFRLTGHIHDYPLANGSSPSWGQ SDLRYPVLEPLLPHLGNILPVSLACDLIDLYFSSSSSAQMHPMSPYVLGFVFRKRSFLHPTNPR RCQPALLASMLWVAAQTSEASFLTSLPSARSKVCQKLLELTVGLLQPLIHTGTNSPSPKTSPVV GAAALGVLGVAMPGSLNMDSLAGETGAFGAIGSLDDVITYVHLATVVSASEYKGASLRWWGAAW SLARELKLGRELPPGNPPANQEDGEGLSEDVDEHDLNRNNTRFVTEEEREERRRAWWLVYIVDR HLALCYNRPLFLLDSECSDLYHPMDDIKWQAGKFRSHDAGNSSINIDSSMTDEFGDSPRAARGA HYECRGRSIFGYFLSLMTILGEIVDVHHAKSHPRFGVGFRSARDWDEQVAEITRHLDMYEESLK RFVAKHLPLSSKDKEQHEMHDSGAVTDMQSPLSVRTNASSRMTESEIQASIVVAYSTHVMHVLH ILLADKWDPINLLDDDDLWISSEGFVTATSHAVSAAEAISQILEFDPGLEFMPFFYGVYLLQGS FLLLLIADKLQAEASPSVIKACETIVRAHEACVVTLSTEYQRNFSKVMRSALALIRGRVPEDLA	534	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	EQQQRRRELLALYRWTGNGTGLAL		
Glycoside hydrolase family 18 protein, chitinase [*T. reesei* QM6a] XP_006968673.1 GI:589113301	MLSRTLLTALGLTTIAAAAPSQTVKTRQAPGGQNAVYWGATNNENDNLSTYCTASSGIDIVILS FLDIYGATGNFPSGNMGNSCYVGTNGVPQLCDDLASSIATCQAAGIKVIISLGGAASSYSLQSQ SQAVAIGQYLWNAYGNSGNTTVQRPFGNVFVNGFDFDIELNAGSQYYQYLISTLRSNFANDPKN TYYITGAPQCPIPEPNMGEIISTSQFDYLWVQFYNNNPVCSLGLPGDAPFNFNDWVSFISTTPS KNAKLFVGAPASTLGANGNAGGAKYYATPEQLAGIVNSVKSSPFFGGIMLWDAGYSDSNVNNGC NYAQEAKNILLTGTACGGESSPPPSTTTTAVPPPASSTPSNPSGGSVPQWGQCGGDGYTGPTQC VAPYKCVATSEWWSSCQ	535	
Glycoside hydrolase family 11 [*T. reesei* QM6a] XP_006968947.1 GI:589113849	MVSFTSLLAGVAAISGVLAAPAAEVESVAVEKRQTIQPGTGYNNGYFYSYWNDGHGGVTYTNGP GGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGSYNPNGNSYLSVYGWSRNPLIEYYIVENFGT YNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSIIGTATFYQYWSVRRNHRSSGSVNTANHFNAW AQQGLTLGTMDYQIVAVEGYFSSGSASITVS	536	
Glycoside hydrolase family 5 protein [*T. reesei* QM6a] XP_006969226.1 GI:589114407	MKSSISVVLALLGHSAAWSYATKSQYRANIKINARQTYQTMIGGGCSGAFGIACQQFGSSGLSP ENQQKVTQILFDENIGGLSIVRNDIGSSPGTTILPTCPATPQDKFDYVWDGSDNCQFNLTKTAL KYNPNLYVYADAWSAPGCMKTVGTENLGGQICGVRGTDCKHDWRQAYADYLVQYVRFYKEEGID ISLLGAWNEPDFNPFTYESMLSDGYQAKDFLEVLYPTLKKAFPKVDVSCCDATGARQERNILYE LQQAGGERYFDIATWHNYQSNPERPFNAGGKPNIQTEWADGTGPWNSTWDYSGQLAEGLQWALY MHNAFVNSDTSGYTHWWCAQNTNGDNALIRLDRDSYEVSARLWAFAQYFRFARPGSVRIGATSD VENVYVTAYVNKNGTVAIPVINAAHFPYDLTIDLEGIKKRKLSEYLTDNSHNVTLQSRYKVSGS SLKVTVEPRAMKTFWLEPQSTFAVI	537	
Glycoside hydrolase family 18 protein, chitinase [*T. reesei* QM6a] XP_0069693971. GI:589114749	MFFTKAVGGLGLLASLASSAPNPIARRQAPGAQNVVYWGQNGGGTVENNDLSAYCTPTSGIDII VLSFLYQWGQGSSALGGTIGQSCGITTSGEPQNCDALTAAITKCKTAGVKIILSLGGASAFSSF QTADQAAQAGQYLWNAYGGGSGVTRPLGNNVMDGFDLDIESNPGTNENYAALVSALRSNFASDP SRQYVISGAPQCPLPEPNMGVIIQNAQFDYLWVQFYNNNEYPGDPCSLGLPGDAPFNFNNWTTF IQSTPSKDAKVFVGVPAAPLAANGAPSGEVYYATPSQLADIVNDVKSNPAFGGIMMWSAGFSDT NVNDGCNYAQEAKNILLTGSPCSSGPVSVSRPPVSSPTITSSPPGTSPAPPSQTGSVPQWGQCG GNGYTGPTQCVAPFKCVATSEWWSQCE	538	
Chain A, Crystal Structure Of An Endo-beta-1,4-xylanase (glycoside	XASQSIDQLIKRKGKLYFGTATDRGLLQREKNAAIIQADLGQVTPENSMKWQSLENNQGQLNWG DADYLVNFAQQNGKSIRGHTLIWHSQLPAWVNNINNADTLRQVIRTHVSTVVGRYKGKIRAWDV VNEIFNEDGTLRSSVFSRLLGEEFVSIAFRAARDADPSARLYINDYNLDRANYGKVNGLKTYVS KWISQGVPIDGIGSQSHLSGGGGSGTLGALQQLATVPVTELAITELDIQGAPTTDYTQVVQACL SVSKCVGITVWGISDKDSWRASTNPLLFDANFNPKPAYNSIVGILQ	539	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Hydrolase Family 10/gh10) Enzyme From *T. Reesei* 4XV0_A GI:756143139			
Endo-1,4-beta-xylanase 1 (also known as EX 1; Xylanase 1; 1,4-beta-D-xylan xylanohydrolase 1; Acidic endo-beta-1,4-xylanase) G0R947.1 GI:1042851765	MVAFSSLICALTSIASTLAMPTGLEPESSVNVTERGMYDFVLGAHNDHRRRASINYDQNYQTGG QVSYSPSNTGFSVNWNTQDDFVVGVGWTTGSSAPINFGGSFSVNSGTGLLSVYGWSTNPLVEYY IMEDNHNYPAQGTVKGTVTSDGATYTIWENTRVNEPSIQGTATFNQYISVRNSPRTSGTVTVQN HFNAWASLGLHLGQMNYQVVAVEGWGGSGSASQSVSN	540	
Endo-1,4-beta-xylanase 2 (also known as Xylanase 2; 1,4-beta-D-xylan xylanohydrolase 2; Alkaline endo-beta-1,4-xylanase) GRUP7.1 GI:142851766	MVSFTSLLAGVAAISGVLAAPAAEVESVAVEKRQTIQPGTGYNNGYFYSYWNDGHGGVTYTNGP GGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGSYNPNGNSYLSVYGWSRNPLIEYYIVENFGT YNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSIIGTATFYQYWSVRRNHRSSGSVNTANHFNAW AQQGLTLGTMDYQIVAVEGYFSSGSASITVS	541	
Endo-1,4-beta-xylanase 3 (also known as Xylanase 3; 1,4-beta-D-xylan xylanohydrolase 3) G0RA32.1 GI:1042851767	MKANVILCLLAPLVAALPTETIHLDPELAALRANLTERTADLWDRQASQSIDQLIKRKGKLYFG TATDRGLLQREKNAAIIQADLGQVTPENSMKWQSLENNQGQLNWGDADYLVNFAQQNGKSIRGH TLIWHSQLPAWVNNINNADTLRQVIRTHVSTVVGRYKGKIRAWDVVNEIFNEDGTLRSSVFSRL LGEEFVSIAFRAARDADPSARLYINDYNLDRANYGKVNGLKTYVSKWISQGVPIDGIGSQSHLS GGGGSGTLGALQQLATVPVTELAITELDIQGAPTTDYTQVVQACLSVSKCVGITVWGISDKDSW RASTNPLLFDANFNPKPAYNSIVGILQ	542	
Endo-1,4-beta-xylanase 1 (also known as EX 1; Xylanase 1; 1,4-	MVAFSSLICALTSIASTLAMPTGLEPESSVNVTERGMYDFVLGAHNDHRRRASINYDQNYQTGG QVSYSPSNTGFSVNWNTQDDFVVGVGWTTGSSAPINFGGSFSVNSGTGLLSVYGWSTNPLVEYY IMEDNHNYPAQGTVKGTVTSDGATYTIWENTRVNEPSIQGTATFNQYISVRNSPRTSGTVTVQN HFNAWASLGLHLGQMNYQVVAVEGWGGSGSASQSVSN	543	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
beta-D-xylan xylanohydrolase 1; Acidic endo-beta-1,4-xylanase) P36218.1 GI:549460			
Hypothetical protein M419DRAFT_104468 [*T. reesei* RUT C-30] ETR97430.1 GI:572273844	MFFTKAVGGLGLLASLASSAPNPIARRQAPGAQNVVYWGQNGGGTVENNDLSAYCTPTSGIDII VLSFLYQWGQGSSALGGTIGQSCGITTSGEPQNCDALTAAITKCKTAGVKIILSLGGASAFSSF QTADQAAQAGQYLWNAYGGGSGVTRPLGNNVMDGFDLDIESNPGTNENYAALVSALRSNFASDP SRQYVISGAPQCPLPEPNMGVIIQNAQFDYLWVQFYNNNEYPGDPCSLGLPGDAPFNFNNWTTF IQSTPSKDAKVFVGVPAAPLAANGAPSGEVYYATPSQLADIVNDVKSNPAFGGIMMWSAGFSDT NVNDGCNYAQEAKNILLTGSPCSSGPVSVSRPPVSSPTITSSPPGTSPAPPSQTGSVPQWGQCG GNGYTGPTQCVAPFKCVATSEWWSQCE	544	
Endo beta-1,4-xylanase isotype 2 [*T. reesei* RUT C-30] ETR98398.1 GI:572274931	MVAFSSLICALTSIASTLAMPTGLEPESSVNVTERGMYDFVLGAHNDHRRRASINYDQNYQTGG QVSYSPSNTGFSVNWNTQDDFVVGVGWTTGSSAPINFGGSFSVNSGTGLLSVYGWSTNPLVEYY IMEDNHNYPAQGTVKGTVTSDGATYTIWENTRVNEPSIQGTATFNQYISVRNSPRTSGTVTVQN HFNAWASLGLHLGQMNYQVVAVEGWGGSGSASQSVSN	545	
Hypothetical protein M419DRAFT_114979 [*T. reesei* RUT C-30] ETR98463.1 GI:572274996	MFFSKALAAAGLLATAAYAAPTMEKRAAGGKLVVYWGAEDDSTTLANVCADSSYDIVNLAFLSR FFAGGGYPELSLSTLGGPSAAQRAAGATNLQDGTSLIPAIQACQAAGKLVILSMGGAVDFSAVT LSSDAQGQQLADTVWNLFLGGTANPTLRPFGSVKLDGVDLDNETGNPTGYLAMAQRFKSNFAKD TSKKYYLTAAPQCPFPDASEPLNVCQLADYIWVQFYNNGNCNIAQSGFNNAVKNWSKSIGNATL FIGALASGADGDQGYVSASSLLSAYQGVSALNLPNIGGIMLWEAQLAVKNGNFQKTVKAGIASG TTPPPPPPTGGCSWAGHCAGASCSTDNDCSDDLTCNGGVCGTAGSTPAPTCSWEGHCLGASCGN DNDCSDPYSCKNGVCSN	546	
Hypothetical protein M419DRAFT_133349 [*T. reesei* RUT C-30] ETR98658.1 GI:572275208	MLSRTLLTALGLTTIAAAAPSQTVKTRQAPGGQNAVYWGATNNENDNLSTYCTASSGIDIVILS FLDIYGATGNFPSGNMGNSCYVGTNGVPQLCDDLASSIATCQAAGIKVIISLGGAASSYSLQSQ SQAVAIGQYLWNAYGNSGNTTVQRPFGNVFVNGFDFDIELNAGSQYYQYLISTLRSNFANDPKN TYYITGAPQCPIPEPNMGEIISTSQFDYLWVQFYNNNPVCSLGLPGDAPFNFNDWVSFISTTPS KNAKLFVGAPASTLGANGNAGGAKYYATPEQLAGIVNSVKSSPFFGGIMLWDAGYSDSNVNNGC NYAQEAKNILLTGTACGGESSPPPSTTTTAVPPPASSTPSNPSGGSVPQWGQCGGDGYTGPTQC VAPYKCVATSEWWSSCQ	547	
Glycoside hydrolase [*T.*	MVSASAGLAAVGLLNGYWGQYTTTEGLRPHCDSGVDSITLGFVNGAPDASGYPSLNFGPNCWAE SYPGNLGLPSKLLSHCMSLQSDIPYCRSKGVKVILSIGGVYNALTSNYFVGDNGTATDFATFLY	548	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
reesei RUT C-30] ETS00190.1 GI:572276883	NAFGPYNASYTGPRPFDDITTGLPTSVDGFDFDIEADFPNGPYIKMIETFRSLDSSMLITGAPQ CPTNPQYFVMKDMIQQAAFDKLFIQFYNNPVCDAIPGNTAGDKFNYDDWEAVIAGSAKSKSAKL YIGLPAIQEPNESGYIDPIAMKNLVCQYKDRPHFGGLSLWDLSRGLVNNINGTSFNQWALDALQ YGCNPIPTTTTTTSTASSTSKASTTSKASTTSKASTTSKASTTSKASTTSKASTTSKASTTSKA STTSKASTTSKASTTSKASTTSKASTTSKVSTTSKASTTSKASTTSKASSTSKVSTTSKASTTS KVSAKATTSTKASTTVKPSTTSKASTTSKASTTSKASTTSKASTTSKASTTSKASTTSKASTTS KAATTSVKPTSKTSTSSKPNVSASSSNVGRDATSLVEASTSTSAAVLYPTTTSRWSNSTITRSS SLTTPIVSDPASLTTSVVYTTSVHTVTKCPAYVTDCPAGGYVTTETIPLYTTVCPISEATQTAA PTVTTEAPQPWTTSTVYTTRVYTITSCAPGVVDCPANQVTTETIPWYTTVCPVTATATPVGPGS VVFPQNTEVGQPSLVGPVVEAAYPTASSSLQTLVKPATSVGVPQGSPAGSSVAPGSSSKPTAPA GPPSYPTGGSGNASPSGSWSGVPVGPSSVPGIPEANAASVMSASLFGLVIVMAAQVFVL		
Xylanase regulator 1 [*T. reesei* RUT C-30] ETS02023.1 GI:572278872	MLSNPLRRYSAYPDISSASFDPNYHGSQSHLHSINVNTFGNSHPYPMQHLAQHAELSSSRMIRA SPVQPKQRQGSLIAARKNSTGTAGPIRRRISRACDQCNQLRTKCDGLHPCAHCIEFGLGCEYVR ERKKRGKASRKDIAAQQAAAAAAQHSGQVQDGPEDQHRKLSRQQSESSRGSAELAQPAHDPPHG HIEGSVSSFSDNGLSQHAAMGGMDGLEDHHGHVGVDPALGRTQLEASSAMGLGAYGEVHPGYES PGMNGHVMVPPSYGAQTTMAGYSGISYAAQAPSPATYSSDGNFRLTGHIHDYPLANGSSPSWGQ SDLRYPVLEPLLPHLGNILPVSLACDLIDLYFSSSSSAQMHPMSPYVLGFVFRKRSFLHPTNPR RCQPALLASMLWVAAQTSEASFLTSLPSARSKVCQKLLELTVGLLQPLIHTGTNSPSPKTSPVV GAAALGVLGVAMPGSLNMDSLAGETGAFGAIGSLDDVITYVHLATVVSASEYKGASLRWWGAAW SLARELKLGRELPPGNPPANQEDGEGLSEDVDEHDLNRNNTRFVTEEEREERRRAWWLVYIVDR HLALCYNRPLFLLDSECSDLYHPMDDIKWQAGKFRSHDAGNSSINIDSSMTDEFGDSPRAARGA HYECRGRSIFGYFLSLMTILGEIVDVHHAKSHPRFGVGFRSARDWDEQVAEITRHLDMYEESLK RFVAKHLPLSSKDKEQHEMHDSGAVTDMQSPLSVRTNASSRMTESEIQASIVVAYSTHVMHVLH ILLADKWDPINLLDDDDLWISSEGFVTATSHAVSAAEAISQILEFDPGLEFMPFFYGVYLLQGS FLLLLIADKLQAEASPSVIKACETIVRAHEACVVTLSTEYQRNFSKVMRSALALIRGRVPEDLA EQQQRRRELLALYRWTGNGTGLAL	549	
SGNH hydrolase [*T. reesei* RUT C-30] ETS03411.1 GI:572280314	MLLVQVRPSSSPAIDLIRGTELRILPVGDSITYGFLSDQDGGDGNGYRLQLRQHLSKDRVVFAG TETSGNMTDGYYAAWNGKTIQYISDHVTPSLEQRPNIILLHAGTNDMNPNGAISREGHDPVAAS ERLGSLVDKMTTLCPDAVILVAMIIGTCNDEQAPQTKVFQSLIPNVVAPRLESGKHVLAVDFST FPLDKLRDCIHPTNEGYHLLGYYWYDFIAQIPRDWITAPVGEDPQRPEEQNLAMRLETDLLLLG LLGLLVVLMYA	550	
Xylanase III [*T. reesei* RUT C-30] ETS05245.1 GI:572282231	MKANVILCLLAPLVAALPTETIHLDPELAALRANLTERTADLWDRQASQSIDQLIKRKGKLYFG TATDRGLLQREKNAAIIQADLGQVTPENSMKWQSLENNQGQLNWGDADYLVNFAQQNGKSIRGH TLIWHSQLPAWVNNINNADTLRQVIRTHVSTVVGRYKGKIRAWDVVNEIFNEDGTLRSSVFSRL LGEEFVSIAFRAARDADPSARLYINDYNLDRANYGKVNGLKTYVSKWISQGVPIDGIGSQSHLS GGGGSGTLGALQQLATVPVTELAITELDIQGAPTTDYTQVVQACLSVSKCVGITVWGISDKDSW RASTNPLLFDANFNPKPAYNSIVGILQ	551	
Hypothetical	MPSLTALAGLLALVPSALAGWNPDSKQNIAVYWGQNSANSQSTQQRLSFYCNDDNINVIEIAFL	552	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
protein M419DRAFT_94061 [*T. reesei* RUT C-30] ETS6436.1 GI:572283462	NGINPPMTNFANAGDRCTPFSDNPWLLSCPEIEADIKTCQANGKTILLSLGGDTYSQGGWASPE AAQDAAAQVWAMFGPVQSDSSAPRPFGDAVVDGFDFDFESTTNNLVAFGAQLRTLSDAAATDSN KKFYLAAAPQCFFPDAAVGPLINAVPMDWIQIQFYNNPCGVSAYTPGSEQQNNYNYQTWEDWAK TSPNPNVKLLVGIPAGPNAGHGYVSDAQLKSVFEYSKKFDTFAGAMMWDMSQLYQNSGFEDQVV DALK		
Endo-1,4-beta-xylanase 2 (also known as EX 2; Xylanase 2; 1,4-beta-D-xylan xylanohydrolase 2; Alkaline endo-beta-1,4-xylanase P36217.2 GI:1042782319	MVSFTSLLAGVAAISGVLAAPAAEVESVAVEKRQTIQPGTGYNNGYFYSYWNDGHGGVTYTNGP GGQFSVNWSNSGNFVGGKGWQPGTKNKVINFSGSYNPNGNSYLSVYGWSRNPLIEYYIVENFGT YNPSTGATKLGEVTSDGSVYDIYRTQRVNQPSIIGTATFYQYWSVRRNHRSSGSVNTANHFNAW AQQGLTLGTMDYQIVAVEGYFSSGSASITVS	553	
Endo-1,4-beta-xylanase 3 (also known as Xylanase; 1,4-beta-D-xylan xylanohydrolase 3) A0A024SIB3.1 GI:1042851768	MKANVILCLLAPLVAALPTETIHLDPELAALRANLTERTADLWDRQASQSIDQLIKRKGKLYFG TATDRGLLQREKNAAIIQADLGQVTPENSMKWQSLENNQGQLNWGDADYLVNFAQQNGKSIRGH TLIWHSQLPAWVNNININNADTLRQVIRTHVSTVVGRYKGKIRAWDVVNEIFNEDGTLRSSVFSRL LGEEFVSIAFRAARDADPSARLYINDYNLDRANYGKVNGLKTYVSKWISQGVPIDGIGSQSHLS GGGGSGTLGALQQLATVPVTELAITELDIQGAPTTDYTQVVQACLSVSKCVGITVWGISDKDSW RASTNPLLFDANFNPKPAYNSIVGILQ	554	
TPA_inf: chitinase 18-13 [*T. reesei*] DAA5861.1 GI:1232265	MFFSKALAAAGLLATAAYAAPTMEKRAAGGKLVVYWGAEDDSTTLANVCADSSYDIVNLAFLSR FFAGGGYPELSLSTLGGPSAAQRAAGATNLQDGTSLIPAIQACQAAGKLVILSMGGAVDFSAVT LSSDAQGQQLADTVWNLFLGGTANPTLRPFGSVKLDGVDLDNETGNPTGYLAMAQRFKSNFAKD TSKKYYLTAAPQCPFPDASEPLNVCQLADYIWVQFYNNGNCNIAQSGFNNAVKNWSKSIGNATL FIGALASGADGDQGYVSASSLLSAYQGVSALNLPNIGGIMLWEAQLAVKNGNFQKTVKAGIASG TTPPPPPPTGGCSWAGHCAGASCSTDNDCSDDLTCNGGVCGTAGSTPAPTCSWEGHCLGASCGN DNDCSDPYSCKNGVCSN	580	
GI:572280314		673	
		674	
		675	
		676	
94. RecName: Full=Endo-1,4-	MKANVILCLLAPLVAALPTETIHLDPELAALRANLTERTADLWDRQASQSIDQLIKRKGKLYFG TATDRGLLQREKNAAIIQADLGQVTPENSMKWQSLENNQGQLNWGDADYLVNFAQQNGKSIRGH	677	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
beta-xylanase 3; Short=Xylanase 3; AltName: Full=1,4-beta-D-xylan xylanohydrolase 3; Flags: Precursor 347 aa protein	TLIWHSQLPAWVNNINNADTLRQVIRTHVSTVVGRYKGKIRAWDVVNEIFNEDGTLRSSVFSRL LGEEFVSIAFRAARDADPSARLYINDYNLDRANYGKVNGLKTYVSKWISQGVPIDGIGSQSHLS GGGGSGTLGALQQLATVPVTELAITELDIQGAPTTDYTQVVQACLSVSKCVGITVWGISDKDSW RASTNPLLFDANFNPKPAYNSIVGILQ		
Beta-galactosidase [*Aspergillus niger*] AOV94178.1 GI:1078570522	MKLQSILSCWAILVAQIWATTDGLTDLVAWDPYSLTVNGNRLFVYSGEFHYPRLPVPEMWLDVF QKMRAHGFNAVSLYFFWDYHSPINGTYDFETGAHNIQRLFDYAQEAGIYIIARAGPYCNAEFNG GGLALYLSDGSGGELRTSDATYHQAWTPWIERIGKIIAENSITNGGPVILNQIENELQETTHSA SNTLVEYMEQIEEAFRAAGVDVPFTSNEKGQRSRSWSTDYEDVGGAVNVYGLDSYPGGLSCTNP STGFSVLRNYYQWFQNTSYTQPEYLPEFEGGWFSAWGADSFYDQCTSELSPQFADVYYKNIIGQ RVTLQNLYMLYGGTNWGHLAAPVVYTSYDYSAPLRETRQIRDKLSQTKLVGLFTRVSSGLLGVE MEGNGTSYTSTTSAYTWVLRNPNTTAGFYVVQQDTTSSQTDITFSLNVNTSAGAFTLPNINLQG RQSKVISTDYPLGHSTLLYVSTDIATYGTFGDTDVVVLYARSGQEVSFSFKNTTKLTFEEYGDS VNLTSSSGNRTITSYTYTQGSGTSVVKFSNGAIFYLVETETAFRFWAPPTTTDPYVTAEQQIFV LGPYLVRNVSISGSVVDLVGDNDNATTVEVFAGSSAKAVKWNGKEITVTKTDYGSLVGSIGGAD SSSITIPSLTGWKVRDSLPEIQSSYDDSKWTVCNKTTTLSPVDPLSLPVLFASDYGYYTGIKIY RGRFDGTNVTGANLTAQGGLAFGWNVWLNGDLVASLPGDADETSSNAAIDFSNHTLKQTDNLLT VVIDYTGHDETSTGDGVENPRGLLGATLNGGSFTSWKIQGNAGGAAGAYELDPVRAPMNEGGLL AERQGWHLPGYKAKSSDGWTDGSPLDGLNKSGVAFYLTTFTLDLPKNYDVPLGIQFTSPSTVDP VRIQLFINGYQYGKYVPYLGPQTTFPIPPGIINNRDKNTIGLSLWAQTDAGAKLENIELISYGA YESGFDAGNGTGFDLNGAKLGYQPEWTEARAKYT	678	
Beta-galactosidase [*Aspergillus niger*] AOV94179.1 GI:1078570524	MTRITKLCVLLLSSIGLLAAAQNQTETGWPLHDDGLTTDVQWDHYSFKVHGERIFVFSGEFHYW RIPVPGLWRDILEKIKAAGFTTFAFYSSWAWHAPNNHTVDFSTGARDITPIFELAKELGMYIIV RPGPYINAEASAGGFPLWLTTGDYGTLRNNDSRYTEAWKPYFEKMTEITSRYQITNGHNTFCYQ IENEYGDQWLSDPSERVPNETAIAYMELLESSARENGILVPFTANDPNMNAMAWSRDWSNAGGN VDVVGLDSYPSCWTCDVSQCTSTNGEYVAYQVVEYYDYFLDFSPTMPSFMPEFQGGSYNPWAGP EGGCGDDTGVDFVNLFYRWNIAQRVTAMSLYMLYGGTNWGAIAAPVTATSYDYSSPISEDRSIS SKYYETKLLSLFTRSARDLTMTDLIGNGTQYTNNTAVKAYELRNPTTNAGFYVTLHEDSTVGTN EAFNLRVNTSAGNLIVPRRGGSIRLNGHQSKIIVTDFTFGSETLLYSTAEVLTYAVIDKKPTLV LWVPTGESGEFAVKGAKSGSVVSKCQSCPAINFHQQGGNLIVGFTQFQGMSVVQIDNDIRVVLL DRTAAYKFWAPALTEDPLVPEDEAVLIQGPYLVRSASLEKSTLAIKGDSINETAVEIFAPENVK TITWNGKQLKTSKSSYGSLKATIAAPASIQLPAFTSWKVNDSLPERLPTYDASGPAWVDANHMT TANPSKPATLPVLYADEYGFHNGVRLWRGYFNGTASGVFLNVQGGSAFGFSAYLNGHFLGSYLG NASIEQANQTFLFPNNITHPTTQNTLLVIHDDTGHDETTGALNPRGILEARLLPSDTTNNSTSP EFTHWRIAGTAGGESNLDPVRGAWNEDGLYAERVGWHLPGFDDSTWSSASSSLSFTGATVKFFR	679	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TTIPLDIPRGLDVSISFVLGTPDNAPNAYRAQLFVNGYQYGRFNPYIGNQVVFPVPVGVLDYTG ENTIGVAVWAQTEDGAGITVDWKVNYVADSSLDVSGLETGELRPGWSAERLKFA		
Beta-galactosidase [*Aspergillus niger*] AOV94180.1 GI:1078570526	MKTSFLLAIGLAVEACLGLVSAPNYVRQINATDSSLQDIVTWDEYSIRVRGERILLLLGEFHPF RLPCPGLWLDVFQKVRALGFSAVSFYVDWALLEGERGSIRADGVFALEEFFQAATEAGLYLTAR PGPYINAEVSGGGFPGWLKRVQGRLKTTDQGYLDAITPYMQAIGRIIAKAQITNGGPVILFQPE NEYTACVQDEGYTQVSNYSMPDINSSCLQKEYMAYVEEQYRKAGIVVPFIVNDADPMGNFAPGT GVGAVDIYSFDDYPLQWSTAPSNPSNWSSLISPLLSYNETVHEEQSPTTPFSISEFQGGVPDAW GGVGIETSAAYIGPEFERIFYKINYGFRAAIQNLYMIFGGTNWGNLGHSGGYTSYDVGAAIAED RQVIREKYSELKLQSNFLQASSAYLETHSDNGSYGIYTDATSLAVTRLAGNPTNFYVVRHGELT SRESTSYKLRVNTSAGNLAIPQLSGSLSLHGRDSKIHLVDYNVGNVSLIYSTAELFTWKQAGSK SVVVLYGGEDELHEFAVPANKGKPTSIEGDGLQVQQINSTTVIQWAVQPSRRVVHFSDTLEVHL LWRNEAYNYWVLDLPVPGAIGRHVSRSHTNRSVIVKAGYLLRTAEIIGTSLYLTGDINTTTTIE LISAPQPVTSILFNKNRIPTTITSPGRLTGTLTYHKPNISLPDLTTLDWYYLNTLPEVHDPTYD DHLWTPCTHTTTANPRNLTTPTSLYASDYGYNGGTLLYRGTFTATGNETSLYLLTEGGYAYGHS IWLNNTFLASWPGNPAFLLSNQTITFPSPLTPGTTYKLTILIDHLGNDENFPANGEFMKDPRGI LDYTLHGRDDKSAISWKMTGNFGGESYADLSRGPLNEGALFAERKGYHLPGAPTEQWTKRSPFD GLPEDERPGVGFFATKFDLQIPDGYDVPISVVFENSTMAGDGSGPARFRSELFVNGWQFGKYVN HIGPQLSYPVPEGILNYNGSNYLALTIWAMDEKSFKLDGLRLQANAVVQSGYRKPSLVKGEVYK ERVDSY	680	
Lactase B [*Aspergillus luchuensis*] GAT22890.1 GI:1002328951	MTLQCKLESACSSTPHNAMVSVLQQDQWAGEPAEQQPHLSAVAAMGRDNECTMFEPGLSGHLLR GGHEATQVRNMVIEILF	681	
Lactase B [*Aspergillus luchuensis*] GAT26827.1 GI:1002325961	MTRITKLCALLLSSTGLLAAAQNQTETGWPLYDDGLTTDIQWDHYSFKVHGERIFVFSGEFHYW RIPVPGLWRDILEKIKAAGFTTFSIYSSWAWHAPNNHTVDFSTGARDITPIFELAKELGMYIIV RPGPYINAEASAGGFPLWLTTGDYGTLRNNDSRYTAAWKPYFEKMTEITSRYQVTNGHNTFCYQ IENEYGDQWLSDPSERVPNETAIAYMELLESSARENGILVPFTANDPNMNAMAWSRDWSNAGGN VDVVGLDSYPSCWTCDVSQCTSTNGEYVAYQVVEYYDYFLEFSPTMPSFMPEFQGGSYNPWAGP EGGCGDDTGVDFVNLFYRWNIAQRVTAMSLYMLYGGTNWGAIAAPVTATSYDYSSPISEDRSIS SKYYETKLLSLFTRSARDLTMTDLIGNGTQYTNNTAVKAYELRNPTTNAGFYVTLHEDSTVGTN EAFSLRVNTSAGNLIVPRLGGSIRLNGHQSKIIVTDFTFGSETLLYSTAEVLTYAVLDKKPTLV LWVPTGESGEFAVKGAKSGSVVSKCQDCSAINFHQQGGNLVVGFTQAQGMSIVQIDNDIRVILL DRTAAYEFWAPALTEDPLVPEDEAVLIQGPYLVRSASLEKSTLAIKGDSINETAVEIFAPNDVK TVTWNGKQLKTSKSSYGSLKATIAAPVSIQLPAFTSWKVNDSLPERLPTYDASGLAWVDANHMT TANPSKPATLPVLCADEYGFHNGVRLWRGYFNGTASGVFLNVQGGSAFGFSAYLNGQFLGSYLG NASIEQANQTFVFPTNITHPTTQNTLLIIHDDTGHDETTGALNPRGILEARLLPSTTTDNTASP EFTHWRLAGTAGGESNLDPVRGAWNEDGLYAERVGWHLPGFDDSTWPSVSSSSLSFTGATVKFF	682	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	RTTIPLNIPRGLDVSISFVLGTPDNAPNTYRAQLFVNGYQYGRFNPYIGNQVVFPVPVGVLDYS GENTIGVAVWAQTEDGAAITVDWKVNYVADSSLDVAGLETAGLRPGWSVERLKFA		
Putative Lactase B [*Aspergillus calidoustus*] CEN62581.1 GI:972234022	MPFFMPEFQGGSYNPWDGPEGGCTEDTGAEFANLFYRWNIAQRVTAMSLYMMYGGTNWGGLAAP VTATSYDYSAPISEDRSIGSKYYETKLLALFTRCAKDLTMTDRIDNGTQYTTNAAISATELRNP ETNAAFYVTNHLDTTLGTDESFKLHVDTSEGALTIPKHGGAIRLNGHQSKIIVTDFRLGRETLL YSTAEVLTYAVFDKKPTLVLWVPAGESGEFAIKGAKRGSATTCSDCSPVEFHRSKESLTVSFTQ ADGISIVQLDNGVRVLLLDRPSAYTFWAPALTDDPLVPETESVFVSGPYLVRSAKLSGSTLALR GDSNGKTAIEVFAPKKVNKITWNGRRIKVTKTRYGSLKASLASAPSIELPALDGWKVSDSLPER LPAYDDSGAAWVDADHMTTPNPHKPATLPVLYADEYGFHNGVRLWRGYFNSSASGVFLNIQGGA AFGWSAYLNGHFLDSYLGDASTNQANGTLSFPDDTLNTDGTPNVLLVIHDDTGHDQTTGVLNPR GILEARLLPLDTESDTEAPEFTHWRVAGTAGGESDLDPVRGVYNEDGLFAERVGWHLPGFDDDD WPAANNSLSFTGATVKFFRTVIPPLDIPQGVDVSISFVFSASSGGNSSSSSSSTGGNTRAFRAQ LFVNGYQYGRFNPYVGNQIVYPVPPGILDYNGENTIGVAVWAQTEAGASLELDWRVNYVVDSSL DVANLDVGGLRPEWEEERLSFA	683	
Beta-glucosidase, lactase phlorizinhydrolase [*Aspergillus oryzae* 100-8] KDE76127.1 GI:635504017	MNVNMFKAGDDILQDVDQSCKDRLPAVEELPLPPTFTWGTATAAYQVEGGAFQDGKGKSIWDTF THLDPSRTNGENGDIACDHYNRMAEDVVLMASYGVDVYRFSIAWARILPLGGRGDPINEKGIAF YNNLIDCLLEHNIEPVVTLYHWDVPQGLYDRYGAFLDTTEFRADFEHFARLCFSRFGDRVKRWI TFNEPYIIAIFGHHSGVLAPGRSSATGGDSRTEPWRVGHTIILAHTAAVQAYATDFQPTQKGDI SIVLNGHYYEPWDAGSEEHWLAAQRRLEFYIGWFGDPIFLGKDYPAPMRAQLGSRLPEFTSEEL DLLRRSAPINSFYGMNHYTTKYARALPDPPAEDDCTGNVEEGPTNSEGKTMGPLSGMSWLRVTP AGFRKLLNWVWDRYRRPIVVTENGCPCPGESQMTKEQALDDQFRIRYFGLYLDAISRAIYDDGV KVEGYYVWSLMDNFEWSAGYGPRYGITHVDFTTLVRTPKQSAKYLHHSFNKRRATSLR	684	
Beta-glucosidase, lactase phlorizinhydrolase [*Aspergillus oryzae* 3.042] EIT76661.1 GI:391867415	MGSTSTSTLPPDFLWGFATASYQIEGAVNEDGRGPSIWDTFCKIPGKIAGGANGDVACDSYHRT HEDIALLKACGAKAYRFSLSWSRIIPLGGRNDPINEKGLQYYIKFVDDLHAAGITPLVTLFHWD LPDELDKRYGGLLNKEEFVADFAHYARIVFKAFGSKVKHWITFNEPWCSSVLGYNVGQFAPGRT SDRSKSPVGDSSRECWIVGHSLLVAHGAAVKIYRDEFKASDGGEIGITLNGDWAEPWDPENPAD VEACDRKIEFAISWFADPIYHGKYPDSMVKQLGDRLPKWTPEDIALVHGSNDFYGMNHYCANFI KAKTGEADPNDTAGNLEILLQNRKGEWVGPETQSPWLRPSAIGFRKLLKWLSERYNYPKIYVTE NGTSLKGENDLPLEQLLQDDFRTQYFRDYIGAMADAYTLDGVNVRAYMAWSLMDNFEWAEGYET RFGVTYVDYENNQKRIPKQSAKAIGEIFDQYIEKA	685	
Beta-glucosidase, lactase phlorizinhydrolase [*Aspergillus oryzae* 3.42] EIT82651.1	MNVNMFKAGDDILQDVDQSCKDRLPAVEELPLPPTFTWGTATAAYQVEGGAFQDGKGKSIWDTF THLDPSRTNGENGDIACDHYNRMAEDVVLMASYGVDVYRFSIAWARILPLGGRGDPINEKGIAF YNNLIDCLLEHNIEPVVTLYHWDVPQGLYDRYGAFLDTTEFRADFEHFARLCFSRFGDRVKRWI TFNEPYIIAIFGHHSGVLAPGRSSATGGDSRTEPWRVGHTIILAHTAAVQAYATDFQPTQKGDI SIVLNGHYYEPWDAGSEEHWLAAQRRLEFYIGWFGDPIFLGKDYPAPMRAQLGSRLPEFTSEEL DLLRRSAPINSFYGMNHYTTKYARALPDPPAEDDCTGNVEEGPTNSEGKTMGPLSGMSWLRVTP AGFRKLLNWVWDRYRRPIVVTENGCPCPGESQMTKEQALDDQFRIRYFGLYLDAISRAIYDDGV	686	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
GI:391873626	KVEGYYVWSLMDNFEWSAGYGPRYGITHVDFTTLVRTPKQSAKYLHHSFNKRRATSLR		
Beta-glucosidase, lactase phlorizinhydrolase [*Aspergillus oryzae* 3.042] EIT82651.1 GI:391873626	MNVNMFKAGDDILQDVDQSCKDRLPAVEELPLPPTFTWGTATAAYQVEGGAFQDGKGKSIWDTF THLDPSRTNGENGDIACDHYNRMAEDVVLMASYGVDVYRFSIAWARILPLGGRGDPINEKGIAF YNNLIDCLLEHNIEPVVTLYHWDVPQGLYDRYGAFLDTTEFRADFEHFARLCFSRFGDRVKRWI TFNEPYIIAIFGHHSGVLAPGRSSATGGDSRTEPWRVGHTIILAHTAAVQAYATDFQPTQKGDI SIVLNGHYYEPWDAGSEEHWLAAQRRLEFYIGWFGDPIFLGKDYPAPMRAQLGSRLPEFTSEEL DLLRRSAPINSFYGMNHYTTKYARALPDPPAEDDCTGNVEEGPTNSEGKTMGPLSGMSWLRVTP AGFRKLLNWVWDRYRRPIVVTENGCPCPGESQMTKEQALDDQFRIRYFGLYLDAISRAIYDDGV KVEGYYVWSLMDNFEWSAGYGPRYGITHVDFTTLVRTPKQSAKYLHHSFNKRRATSLR	687	
Lactase B [*Aspergillus kawachii* IFO 4308] GAA82087.1 GI:358365465	MTRITKLCALLLSSTGLLAAAQNQTETGWPLYDDGLTTDIQWDHYSFKVHVPGLWRDILEKIKA AGFTTFSIYSSWAWHAPNNHTVDFSTGARDITPIFELAKELGMYIIVRPGPYINAEASAGGFPL WLTTGDYGTLRNNDSRYTAAWKPYFEKMTEITSRYQVTNGHNTFCYQIENEYGDQWLSDPSERV PNETAIAYMELLESSARENGILVPFTANDPNMNAMAWSRDWSNAGGNVDVVGLDSYPSCWTCDV SQCTSTNGEYVAYQVVEYYDYFLEFSPTMPSFMPEFQGGSYNPWAGPEGGCGDDTGVDFVNLFY RWNIAQRVTAMSLYMLYGGTNWGAIAAPVTATSYDYSSPISEDRSISSKYYETKLLSLFTRSAR DLTMTDLIGNGTQYTNNTAVKAYELRNPTTNAGFYVTLHEDSTVGTNEAFSLRVNTSAGNLIVP RLGGSIRLNGHQSKIIVTDFTFGSETLLYSTAEVLTYAVLDKKPTLVLWVPTGESGEFAVKGAK SGSVVSKCQDCSAINFHQQGGNLVVGFTQAQGMSIVQIDNDIRVILLDRTAAYEFWAPALTEDP LVPEDEAVLIQGPYLVRSASLEKSTLAIKGDSINETAVEIFAPNDVKTVTWNGKQLKTSKSSYG SLKATIAAPVSIQLPAFTSWKVNDSLPERLPTYDASGLAWVDANHMTTANPSKPATLPVLYADE YGFHNGVRLWRGYFNGTASGVFLNVQGGSAFGFSAYLNGQFLGSYLGNASIEQANQTFVFPTNI THPTTQNTLLIIHDDTGHDETTGALNPRGILEARLLPSTTTDNTASPEFTHWRLAGTAGGESNL DPVRGAWNEDGLYAERVGWHLPGFDDSTWPSVSSSSLSFTGATVKFFRTTIPLNIPRGLDVSIS FVLGTPDNAPNTYRAQLFVNGYQYGRFNPYIGNQVVFPVPVGVLDYSGENTIGVAVWAQTEDGA AITVDWKVNYVADSSLDVAGLETAGLRPGWSVERLKFA	688	
Probable beta-galactosidase C (Lactase C) A1CE56.1 GI:00680864	MRILSLLFLLLLGFLAGNRVVSATDHGKTTDVTWDRYSLSVKGERLFVFSGEFHYQRLPVPEMW LDVFQKLRANGFNAISVYFFWGYHSASEGEFDFETGAHNIQRLFDYAKEAGIYVIARAGPYCNA ETTAGGYALWAANGQMGNERTSDDAYYAKWRPWILEVGKIIAANQITNGGPVILNQHENELQET SYEADNTLVVYMKQIARVFQEAGIVVPSSHNEKGMRAVSWSTDHHDVGGAVNIYGLDSYPGGLS CTNPSSGFNLVRTYYQWFQNSSYTQPEYLPEFEGGWFQPWGGHDYDTCATELSPEFADVYYKNN IGSRVTLQNIYMVFGGTNWGHSAAPVVYTSYDYSAPLRETREIRDKLKQTKLIGLFTRVSSDLL KTHMEGNGTGYTSDSSIYTWALHNPDTNAGFYVLAHKTSSSRSVTEFSLNVTTSAGAISIPDIQ LDGRQSKIIVTDYQFGKSSALLYSSAEVLTYANLDVDVLVLYLNVGQKGLFVFKDERSKLSFQT YGNTNVTASVSSHGTQYIYTQAEGVTAVKFSNGVLAYLLDKESAWNFFAPPTTSNPQVAPDEHI LVQGPYLVRGVTINHDTVEIIGDNANTTSLEVYAGNLRVKVVKWNGKAIKSRRTAYGSLVGRAP GAEDARISPPSLDSWSAQDTLPDIQPDYDDSRWTVCNKTASVNAVPLLSLPVLYSGDYGYHAGT KVYRGRFDGRNVTGANVTVQNGVASGWAAWLNGQFVGGVAGAIDLAVTSAVLSFNSSLLHDRDN	689	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	VLTVVTDYTGHDQNSVRPKGTQNPRGILGATLIGGGKFTSWRIQGNAGGEKNIDPVRGPINEGG LYGERMGWHLPGYKAPRSAAKSSPLDGISGAEGRFYTTTFTLKLDRDLDVPIGLQLGAPAGTQA VVQVFMNGYQFGHYLPHIGPQSLFPFFPPGVINNRGENTLAISMWALTDAGAKLDQVELVAYGKY RSGFDFNQDWGYLQPQWKDNRRQYA		
Probable beta-galactosidase B (Lactase B) A1D199.1 GI:00680896	MAHIYRLLLLLLSNLWFSAAAQNQSETEWPLHDNGLSKVVQWDHYSFQVNGQRIFIFSGEFHYW RIPVPELWRDILEKVKATGFTAFAFYSSWAYHAPNNRTVDFSTGARDITPIFELAKELGMYMIV RPGPYVNAEASAGGFPLWLTTGEYGSLRNDDPRYTAAWTPYFANMSQITSKYQVTDGHNTLVYQ IENEYGQQWIGDPKDRNPNKTAVAYMELLEASALENGITVPLTSNDPNMNSKSWGSDWSNAGGN VDVAGLDSYPSCWTCDVSQCTSTNGEYVPYKVIDYYDYFQEVQPTLPSFMPEFQGGSYNPWAGP EGGCPQDTGAEFANLFYRWNIGQRVTAMSLYMLYGGTNWGAIAAPVTATSYDYSAPISEDRSIG AKYSETKLLALFTRTAKDLTMTEAIGNGTQYTTNTAVRAFELRNPQTNAGFYVTFHNDTTVGGN QAFKLHVNTSVGALTVPKNEGVIQLNGHQSKIIVTDFTLGKRTLLYSTAEVLTYAVFENRPTLV LWVPTGESGEFAIKGTKSGKVENGDGCSGINFKREKDYLVVNFSQAKGLSVLRLDNGVRVVLLD KAAAYRFWAPALTDDPIVQETETVLVHGPYLVRSASVSKSTLALRGDSVEKTTLEIFAPHSVRK ITWNGKEVKTSQTPYGSLKATLAAPPTIKLPALTSWRSNDSLPERLPSYDDSGPAWIEANHMTT SNPSPPATLPVLYADEYGFHNGVRLWRGYFNGSASGVFLNIQGGSAFGWSAWLNGHFLDSHLGT ATTSQANKTLTFSSSILNPTENVLLIVHDDTGHDQTTGALNPRGIIEARLLSNDTSSPAPGFTQ WRIAGTAGGESNLDPIRGVFNEDGLFAERMGWHLPGFDDSAWTPENSTTSASSALSFTGATVRF FRTVVPLDIPAGLDVSISFVLSTPSAAPKGYRAQLFVNGYQYGRYNPHIGNQVVFPVPPGILDY QGDNTIGLAVWAQTEEGAGIQVDWKVNYVADSSLSVAGFGKGLRPGWTEERLKFA	690	
Probable beta-galactosidase A (Lactase A) A1D1Z9.1 GI:300680858	MKLLSVCAVALLAAQAAGASIKHKLNGFTIMEHSDPAKRELLQKYVTWDEKSLFVNGERIMIFS GEVHPFRLPVPSLWLDVFQKIKALGFNCVSFYVDWALLEGKPGKYRAEGNFALEPFFDAAKQAG IYLLARPGPYINAEASGGGFPGWLQRVNGTLRTSDPAYLKATDNYIAHVAATVAKGQITNGGPV ILYQPENEYSGACCNATFPDGDYMQYVIDQARNAGIVVPLINNDAWTGGHNAPGTGKGEVDIYG HDSYPLGFDCGHPSVWPKGNLPTTFRTDHLRESPTTPYSLIEFQAGSFDPWGGPGFAACAALVN HEFERVFYKNDLSFGAAILNLYMTFGGTNWGNLGHPGGYTSYDYGSPLTESRNVTREKYSELKL IGNFVKASPSYLLATPGNLTTSGYADTADLTVTPLLGNGTGSYFVVRHTDYTSQASTPYKLSLP TSAGRLTVPQLGGTLTLNGRDSKVHVVDYNVAGTNILYSTAEVFTWKKFGDSKVLVLYGGPGEH HELAVSLKSDVQVVEGSNSEFTSKKVEDVVVVAWDVSASRRIVQIGDLKIFLLDRNSAYNYWVP QLDKDDSSTGYSSEKTTASSIIVKAGYLVRTAYTKGSGLYLTADFNATTPVEVIGAPSNVRNLY INGEKTQFKTDKNGIWSTGVKYSAPKIKLPSMKDLDWKYLDTLPEVQSTYDDSAWPAADLDTTP NTLRPLTMPKSLHSSDYGFHTGYLIYRGHFVADGSETTFDVRTQGGSAFGSSVWLNEAFLGSWT GLNANADYNSTYRLPQVEKGKNYVLTVVIDTMGLNENWVVGTDEMKNPRGILSYKLSGRDASAI TWKLTGNLGGEDYQDKIRGPLNEGGLYAERQGFHQPEPPSKKWKSASPLDGLSKPGIGFYTAQF DLDIPSGWDVPLYFNFGNSTKSAYRVQLYVNGYQYGKFVSNIGPQTSFPVPQGILNYQGTNWVA LTLWALESDGAKLDDFELVNTTPVMTALSKIRPSKQPNYRQRKGAY	691	
Probable beta-galactosidase E	MKFLLRRFIALAAASSVVAAPSVSHLSLQDAANRRELLQDLVTWDQHSLFVRGERLMIFSGEFH PFRLPVPGLWFDVFQKITSLGFNAVSFYTDWGLMEGNPGHVVTDGIWSLDEFFTAASEAGIYLI	692	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
(Lactase E) A1DJ58.1 GI: 300680873	ARPGPYINAETSAGGIPGWVLRLKGIIRSNSEDYLRATDTYMATLGKIIAKAQITNGGPVILVQ PENEYTTWPNVSESEFPTTMNKEVMAYAEKQLRDAGVVVPTVVNDNKNLGYFAPGTGLGETDLY GIDAYPMRYDCGNPYVWPTYRFPRDWQHTHRNHSPTTPFAIMEFQGGSGDGWGGVTEDGCAILV NNEAVRVVYKNNYGFGVGVFNIYMTYGGTNWGNLGYHGGYTSYDYGAAITEDRQIWREKYSEEK LQANFLKVSPAYLTATPGNGVNGSYTGNKDIAVTPLFGNGTTTNFYLVRHADFTSTGSVQYQLS VSTSVGNVTIPQLGGSLSLNGRDSKFHVTDYDVGEFNLIYSSAEIFTWAKGDNKKRVLVLYGGA GELHEFALPKHLPRPTVVDGSDVKMAKKGSAWVVQWEVTAQRRVLRAGKLEIHLLWRNDAYQHW VLELPAKQPIANYSSPSKETVLVKGGYLLRSACITNNKLHLTGDVNATTPLEVISAPKRFDGIV FNGQSLKSTRSKIGNLAATVRYQPPAISLPDLKRLDWKYLDSLPEISPDYSDEGWMSLTNTYTN NTRKFTGPTCLYADDYGYHGGSLIYRGHFKANGDESWVFLNTSGGVGFANSVWLNQTFLGSWTG SGNNMTYPRNISLPHELSPGKPYVFTVVIDHMGQDEEAPGTDAIKFPRGILDYALSGHEVSDLK WKMTGNLGGEQYQDSTRGPLNEGAMYAERRGYHLPNPPTSSWKSSSPINDGLTGAGIGFYATSF SLDLPEGYDIPLSFLFNNSASDARSGTSYRCQLFVNGYQFGKYVNDLGPQTNFPVPEGILNYNG VNYVAVSLWALEPQGALVGGLELVASTPILSAYRKPVPAPQPGWKPRRGAY		
Probable beta-galactosidase C (Lactase C) A1DM65.1 GI: 300680868	MRIFSFLFLLLLGILTGQGLVSGTDNGKTTDVTWDKYSLSVKGQRLFVFSGEFHYQRLPVPELW LDVFQKLRANGFNAISVYFFWSFHSASEGEFDFENGAHDIQRLFDYAKEAGLYVIARAGPYCNA ETSAGGFALWAANGQMGNERTSDEAYYEKWRPWILEVGKIIAKNQITNGGPVILNQHENELTET TYDPNHTLVVYMKQIAQVFEEAGIVVPSSHNEKGMRGVSWSTDYHNVGGAVNIYGLDSYPGGLS CTNPNSGFRLVRTYYQWFQNYSSTQPSYMPEFEGGWFQPWGGSFYDTCATELSPEFPDVYYKNN IGSRVTLHSIYMTYGGTNWGHSAAPVVYTSYDYAAPLRETREIRDKLKQTKLIGLFTRVSTDLL KTYMEGNGTGYTSDSSIYTWSLRNPDTNAGFYVLAHSTSSARDVTTFSLNATTSAGAISIPDIE LNGRQSKIIVTDYNFGTNSTLLFSSAEVLTYANLDVNVLVFYLNVGQKGTFALKDEPKLAFQTY GNSNVTTSESSYGTQYSYTQGEGVTAVKFSNGVLAYLLDKESAWNFFAPPTTSSPQVAPNEHIL VQGPYLVRGASINHGTVEITGDNANTTSIEVYTGNSQVKKVKWNGKTIETRKTAYGSLIGTVPG AEDVKIRLPSLDSWKAQDTLPEIQPDYDDSTWTVCNKTTSVNAIAPLSLPVLYSGDYGYHAGTK VYRGRFDGRNVTGANVTVQNGAAAGWAAWVNGQYAGGSAGSPSLAATSAVLTFNGLSLKDRDNV LTVVTDYTGHDQNSVRPKGTQNPRGILGATLTGGGNFTSWRIQGNAGGEKNIDPVRGPMNEGGL YGERMGWHLPGYKVPKSASKSSPLDGVSGAEGRFYTTTFKLKLDKDLDVPIGLQLGAPEGTKAV VQVFMNGYQFGHYLPHTGPQSLFPFPPGVINNRGENTLAISMWALTDAGAKLDKVELVAYGKYR SGFDFNQDWGYLQPGWKDRSQYA	693	
Probable beta-galactosidase B (Lactase B) A2QA64.2 GI: 300681011	MTRITKLCVLLLSSIGLLAAAQNQTETGWPLHDDGLTTDVQWDHYSFKVHGERIFVFSGEFHYW RIPVPGLWRDILEKIKAAGFTTFAFYSSWAWHAPNNHTVDFSTGARDITPIFELAKELGMYIIV RPGPYINAEASAGGFPLWLTTGDYGTLRNNDSRYTEAWKPYFEKMTEITSRYQITNGHNTFCYQ IENEYGDQWLSDPSERVPNETAIAYMELLESSARENGILVPFTANDPNMNAMAWSRDWSNAGGN VDVVGLDSYPSCWTCDVSQCTSTNGEYVAYQVVEYYDYFLDFSPTMPSFMPEFQGGSYNPWAGP EGGCGDDTGVDFVNLFYRWNIAQRVTAMSLYMLYGGTNWGAIAAPVTATSYDYSSPISEDRSIS SKYYETKLLSLFTRSARDLTMTDLIGNGTQYTNNTAVKAYELRNPTTNAGFYVTLHEDSTVGTN EAFSLRVNTSAGNLIVPRLGGSIRLNGHQSKIIVTDFTFGSETLLYSTAEVLTYAVIDKKPTLV	694	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	LWVPTDESGEFAVKGAKSGSVVSKCQSCPAINFHQQGGNLIVGFTQSQGMSVVQIDNDIRVVLL DRTAAYKFWAPALTEDPLVPEDEAVVLIQGPYLVRSASLEKSTLAIKGDSINETAVEIFAPENV KTITWNGKQLKTSKSSYGSLKATIAAPASIQLPAFTSWKVNDSLPERLPTYDASGPAWVDANHM TTANPSKPATLPVLYADEYGFHNGVRLWRGYFNGTASGVFLNVQGGSAFGFSAYLNGHFLGSYL GNASIEQANQTFLFPNNITHPTTQNTLLVIHDDTGHDETTGALNPRGILEARLLPSDTTNNSTS PEFTHWRIAGTAGGESNLDPVRGAWNEDGLYAERVGWHLPGFDDSTWSSVSSSSSLSFTGATVK FFRTTIPLDIPRGLDVSISFVLGTPDNAPNAYRAQLFVNGYQYGRFNPYIGNQVVFPVPVGVLD YTGENTIGVAVWAQTEDGAGITVDWKVNYVADSSLDVSGLETGELRPGWSAERLKFA		
Probable beta-galactosidase A (Lactase A) A2QAN3.1 GI: 300680857	MKLSSACAIALLAAQAAGASIKHRINGFTLTEHSDPAKRELLQKYVTWDDKSLFINGERIMIFS GEFHPFRLPVKELQLDIFQKVKALGFNCVSFYVDWALVEGKPGEYRADGIFDLEPFFDAASEAG IYLLARPGPYINAESSGGGFPGWLQRVNGTLRSSDKAYLDATDNYVSHVAATIAKYQITNGGPI ILYQPENEYTSGCCGVEFPDPVYMQYVEDQARNAGVVIPLINNDASASGNNAPGTGKGAVDIYG HDSYPLGFDCANPTVWPSGDLPTNFRTLHLEQSPTTPYAIVEFQGGSYDPWGGPGFAACSELLN NEFERVFYKNDFSFQIAIMNLYMIFGGTNWGNLGYPNGYTSYDYGSAVTESRNITREKYSELKL LGNFAKVSPGYLTASPGNLTTSGYADTTDLTVTPLLGNSTGSFFVVRHSDYSSEESTSYKLRLP TSAGSVTIPQLGGTLTLNGRDSKIHVTDYNVSGTNIIYSTAEVFTWKKFADGKVLVLYGGAGEH HELAISTKSNVTVIEGSESGISSKQTSSSVVVGWDVSTTRRIIQVGDLKILLLDRNSAYNYWVP QLATDGTSPGFSTPEKVASSIIVKAGYLVRTAYLKGSGLYLTADFNATTSVEVIGVPSTAKNLF INGDKTSHTVDKNGIWSATVDYNAPDISLPSLKDLDWKYVDTLPEIQSSYDDSLWPAADLKQTK NTLRSLTTPTSLYSSDYGFHTGYLLYRGHFTATGNESTFAIDTQGGSAFGSSVWLNGTYLGSWT GLYANSDYNATYNLPQLQAGKTYVITVVIDNMGLEENWTVGEDLMKTPRGILNFLLAGRPSSAI SWKLTGNLGGEDYEDKVRGPLNEGGLYAERQGFHQPEPPSQNWKSSSPLEGLSEAGIGFYSASF DLDLPKGWDVPLFLNIGNSTTPSPYRVQVYVNGYQYAKYISNIGPQTSFPVPEGILNYRGTNWL AVTLWALDSAGGKLESLELSYTTPVLTALGEVESVDQPKYKKRKGAY	695	
Probable beta-galactosidase C (Lactase C) A2QL84.1 GI: 300680867	MKLQSILSCWAILVAQIWATTDGLTDLVAWDPYSLTVNGNRLFVYSGEFHYPRLPVPEMWLDVF QKMRAHGFNAVSLYFFWDYHSPINGTYDFETGAHNIQRLFDYAQEAGIYIIARAGPYCNAEFNG GGLALYLSDGSGGELRTSDATYHQAWTPWIERIGKIIADNSITNGGPVILNQIENELQETTHSA SNTLVEYMEQIEEAFRAAGVDVPFTSNEKGQRSRSWSTDYEDVGGAVNVYGLDSYPGGLSCTNP STGFSVLRNYYQWFQNTSYTQPEYLPEFEGGWFSAWGADSFYDQCTSELSPQFADVYYKNNIGQ RVTLQNLYMLYGGTNWGHLAAPVVYTSYDYSAPLRETRQIRDKLSQTKLVGLFTRVSSGLLGVE MEGNGTSYTSTTSAYTWVLRNPNTTAGFYVVQQDTTSSQTDITFSLNVNTSAGAFTLPNINLQG RQSKVISTDYPLGHSTLLYVSTDIATYGTFGDTDVVVLYARSGQVVSFAFKNTTKLTFEEYGDS VNLTSSSGNRTITSYTYTQGSGTSVVKFSNGAIFYLVETETAFRFWAPPTTTDPYVTAEQQIFV LGPYLVRNVSISGSVVDLVGDNDNATTVEVFAGSPAKAVKWNGKEITVTKTDYGSLVGSIGGAD SSSITIPSLTGWKVRDSLPEIQSSYDDSKWTVCNKTTTLSPVDPLSLPVLFASDYGYYTGIKIY RGRFDGTNVTGANLTAQGGLAFGWNVWLNGDLVASLPGDADETSSNAAIDFSNHTLKQTDNLLT VVIDYTGHDETSTGDGVENPRGLLGATLNGGSFTSWKIQGNAGGAAGAYELDPVRAPMNEGGLL AERQGWHLPGYKAKSSDGWTDGSPLDGLNKSGVAFYLTTFTLDLPKKYDVPLGIQFTSPSTVDP	696	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	VRIQLFINGYQYGKYVPYLGPQTTFPIPPGIINNRDKNTIGLSLWAQTDAGAKLENIELISYGA YESGFDAGNGTGFDLNGAKLGYQPEWTEARAKYT		
Probable beta-galactosidase A (Lactase A) BXMP7.2 GI: 300681017	MKLLSVCAIALLAAQAAGASIKHMLNGFTLMEHSDPAKRELLQKYVTWDEKSLFVNGERIMIFS GEVHPFRLPVPSLWLDVFQKIKALGFNCVSFYVDWALLEGKPGEYRAEGNFALEPFFDVAKQAG IYLLARPGPYINAEASGGGFPGWLQRVNGTLRTSDPAYLKATDNYIAHVAATIAKGQITNGGPV ILYQPENEYSGACCDATFPDGDYMQYVIDQARNAGIVVPLINNDAWTGGHNAPGTGKGEVDIYG HDSYPLGFDCGHPSVWPKGNLPTTFRTDHLKQSPTTPYSLIEFQAGSFDPWGGPGFAACAALVN HEFERVFYKNDLSFGAAILNLYMTFGGTNWGNLGHPGGYTSYDYGSPLTESRNVTREKYSELKL IGNFVKASPSYLLATPGNLTTSGYADTADLTVTPLLGNGTGSYFVVRHTDYTSQASTPYKLSLP TSAGRLTVPQLGGTLTLNGRDSKIHVVDYNVAGTNIIYSTAEVFTWKNFGDSKVLILYGGPGEH HELAVSFKSDVQVVEGSNSEFKSKKVGDVAVVAWDVSPSRRIVQIGDLKIFLLDRNSVYNYWVP QLDKDDSSTGYSSEKTTASSIIVKAGYLVRTAYTKGSGLYLTADFNATTPVEVIGAPSNVRNLY INGEKTQFKTDKNGIWSTEVKYSAPKIKLPSMKDLDWKYLDTLQEVQSTYDDSAWPAADLDTTP NTLRPLTTPKSLYSSDYGFHTGYLIYRGHFVADGSETTFDVRTQGGSAFGSSVWLNESFLGSWT GLNANADYNSTYKLPQVEQGKNYVLTILIDTMGLNENWVVGTDEMKNPRGILSYKLSGRDASAI TWKLTGNLGGEDYQDKIRGPLNEGGLYAERQGFHQPQPPSQKWKSASPLDGLSKPGIGFYTAQF DLDIPSGWDVPLYFNFGNSTKSAYRVQLYVNGYQYGKFVSNIGPQTSFPVPQGILNYQGTNWVA LTLWALESDGAKLDDFELVNTTPVMTALSKIRPSKQPNYRQRKGAY	697	
Probable beta-galactosidase B (Lactase B) B0XNY2.1 GI: 300680860	MAHIYRLLLLLLSNLWFSTAAQNQSETEWPLHDNGLSKVVQWDHYSFQVNGQRIFIFSGEFHYW RIPVPELWRDILEKVKATGFTAFAFYSSWAYHAPNNSTVDFSTGARDITPIFELAKELGMYMIV RPGPYVNAEASAGGFPLWLMTGEYGSLRNDDPRYTAAWTPYFANMSQITSKYQVTDGHNTLVYQ IENEYGQQWIGDPKNRNPNKTAVAYMELLEASARENGITVPLTSNDPNMNSKSWGSDWSNAGGN VDVAGLDSYPSCWTCDVSQCTSTNGEYVPYKVIDYYDYFQEVQPTLPSFMPEFQGGSYNPWAGP EGGCPQDTSAEFANLFYRWNIGQRVTAMSLYMLYGGTNWGAIAAPVTATSYDYSAPISEDRSIG AKYSETKLLALFTRTAKDLTMTEAIGNGTQYTTNTAVRAFELRNPQTNAGFYVTFHTDTTVGGN QAFKLHVNTSVGALTVPKNEGLIQLNGHQSKIIVTDFTLGKRTLLYSTAEVLTYAVFENRPTLV LWVPTGESGEFAIKGAKSGKVENGDGCSGIKFKREKDYLVVNFSQAKGLSVLRLDNGVRVVLLD KAAAYRFWAPALTDDPNVQETETVLVHGPYLVRSASISKTTLALRGDSVEKTTLEIFAPHSVRK ITWNGKEVQTSHTPYGSLKATLAAPPDIKLPALTSWRSNDSLPERLPSYDDSGPAWIEANHMTT SNPSPPATFPVLYADEYGFHNGVRLWRGYFNGSASGVFLNIQGGSAFGWSAWLNGHFLDSHLGT ATTSQANKTLTFPSSILNPTENVLLIVHDDTGHDQTTGALNPRGILEARLLSNDTSSPPPEFTH WRLAGTAGGESNLDPIRGVFNEDGLFAERMGWHLPGFDDSAWTSENSATSASSALSFTGATVRF FRSVVPLNIPAGLDVSISFVLSTPTAAPKGYRAQLFVNGYQYGRYNPHIGNQVVFPVPPGILDY QGDNTIGLAVWAQTEEGAGIQVDWKVNYVADSSLSVAGFGKGLRPGWTEERLKFA	698	
Probable beta-galactosidase E (Lactase E)	MKSLLKRLIALAAAYSVAAAPSFSHHSSQDAANKRELLQDLVTWDQHSLFVRGERLMIFSGEFH PFRLPVPGLWFDVFQKIKSLGFNAVSFYTDWGLMEGNPGHVVTDGIWSLDEFFTAAREAGLYLI ARPGPYINAETSAGGIPGWVLRRKGIIRSNSEDYLRATDTYMATLGKIIAKAQITNGGPVILVQ PENEYTTWPNVSESEFPTTMNQEVMAYAEKQLRDAGVVVPTVVNDNKNLGYFAPGTGLGETDLY	699	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
B0XXE7.1 GI:300680872	GIDAYPMRYDCGNPYVWPTYRFPRDWQHEHRNHSPTTPFAIMEFQGGSGDGWGGVTEDGCAILV NNEAVRVVYKNNYGFGVRVFNIYMTYGGTNWGNLGYYGGYTSYDYGAAITEDRQIWREKYSEEK LQANFLKVSPAYLTSTPGNGVNGSYTGNKDITVTPLFGNGTTTNLYLVRHADFTSTGSAQYNLS ISTSVGNVTIPQLGGSLSLNGRDSKFHITDYDVGGFNLIYSSAEVFTWAKGDNKKRVLVLYGGA GELHEFALPKHLPRPTVVEGSYVKIAKQGSAWVVQWEVAAQRRVLRAGKLEIHLLWRNDAYQHW VLELPAKQPIANYSSPSKETVIVKGGYLLRSAWITDNDLHLTGDVNVTTPLEVISAPKRFDGIV FNGQSLKSTRSKIGNLAATVHYQPPAISLPDLKRLDWKYIDSLPEISTEYNDEGWTPLTNTYTN NTREFTGPTCLYADDYGYHGGSLIYRGHFTANGDESWVFLNTSGGVGFANSVWLNQTFLGSWTG SGRNMTYPRNISLPHELSPGEPYVFTVVIDHMGQDEEAPGTDAIKFPRGILDYALSGHELSDLR WKMTGNLGGEQYQDLTRGPLNEGAMYAERQGYHLPSPPTSSWKSSNPIKEGLTGAGIGFYATSF SLDLPEGYDIPLSFRFNNSASAARSGTSYRCQLFVNGYQFGKYVNDLGPQTKFPVPEGILNYNG VNYVAVSLWALESQGALIGGLDLVASTPILSGYRKPAPAPQPGWKPRRGAY		
Probable beta-galactosidase C (Lactase C) B0Y752.1 GI:300680865	MRIFSFLFLLLLGILTGQGLVSGTDNGKTTDVTWDKYSLSVKGQRLFVFSGEFHYQRLPVPELW LDVFQKLRANGFNAISVYFFWSFHSASEGEFDFENGAHDIQRLFDYAKEAGLYVIARAGPYCNA ETSAGGFALWAANGQMGNERTSDEAYYEKWRPWILEVGKIIAKNQITNGGPVILNQHENELVET TYDPNHTLVVYMKQIAQVFEEAGIVVPSSHNEKGMRGVSWSTDYHNVGGAVNIYGLDSYPGGLS CTNPNSGFNLVRTYHQWFQNYSFTQPSYLPEFEGGWFQPWGGSFYDTCATELSPEFPDVYYKNN IGSRVTLHSIYMTYGGTNWGHSAAPVVYTSYDYAAPLRETREIRDKLKQTKLIGLFTRVSKDLL KTYMEGNGTGYTSDSSIYTWSLRNPDTNAGFYVLAHSTSSTRDVTTFTLNVTTSAGAISIPDIE LNGRQSKIIVTDYNFGTNSTLLFSSAEVLTYANLDVNVLVFYLNVGQKGTFVFKDEPKLAFQTY GNSNLTTSESSYGTQYSYTQGKGVTAVKFSNGVLAYFLDKESAWNFFAPPTTSSPQVAPNEHIL VQGPYLVRGASVNHGTVEITGDNANTTSIEVYTGNSQVKKIKWNGKTIETRKTAYGSLIGTAPG AEDVKIQLPSLDSWKAQDTLPEIQPDYDDSKWTVCNKTTSVNAIAPLSLPVLYSGDYGYHAGTK VYRGRFDGRNVTGANVTVQNGAAAGWAAWVNGQYAGGSAGSPNLAATSAVLTFNSSSLKDQDNV LTVVTDYTGHDQNSVRPKGTQNPRGILGATLIGGGNFTSWRIQGNAGGEKNIDPVRGPMNEGGL YGERMGWHLPGYKVPKSASKSSPLDGVSGAEGRFYTTTFKLKLDKDLDVPIGLQLGAPEGTKAV VQVFMNGYQFGHYLPHTGPQSLFPFPPGVINNRGENTLAISMWALTDAGAKLDKVELVAYGKYR SGFDFNQDWGYLQPGWKDRSQYA	700	
Probable beta-galactosidase C (Lactase C) B8N2I5.1 GI:300680866	MRLLSFIYLVWLALLTGTPQVSATDNGKTSDVAWDKYSLSVKGERLFVFSGEFHYQRLPVPELW LDVFQKLRANGFNTISVYFFWSYHSASEDVFDFTTGAHDIQRLFDYAKQAGLYVIARAGPYCNA ETSAGGFALWAANGQMGSERTSDEAYYKKWKPWILEVGKIIAANQITNGGPVILNQHENELQET TYDSNDTKVIYMEQVAKAFEEAGVVVPSSHNEKGMRTVSWSTDYKNVGGAVNVYGLDSYPGSLS CANPNSGFNLLRTYYQWFQNYSYTQPEYLAEFEGGWFQPWGGSFYDSCASELSPEFADVYYKNN IGSRVTLHNIYMTFGGTNWGHSAAPVVYTSYDYGSPLRETREIRDKLKQTKLLGLFTRVSKDLL KTYMEGNGTSYTSDDSIYTWALRNPDSDAGFYVVAHNTSSSREVTTFSLNITTSAGALTIPDIE LDGRQSKIIVTDYSIGSESSLLYSSAEVLTYATLDVDVLVFYLNAGQKGAFVFKDAPADLKYQT YGNSNLSALETSQGTQYSYTQGEGVTAVKFSNGVLVYLLDKETAWNFFAPPTVSSPTVAPNEHI LVFGPYLVRGASIKHDTVEIVGDNSNSTSIEIYTGDEHVKKVSWNGNLIDTRATAYGSLIGTVP	701	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GAEDIEISLPSLSSWKAQDTLPEISPDYDDSRWTICNKTTSVNSVAPLSLPVLYSGDYGYHTGT KIYRGRFDGQNATGANVTVQNGVAAGWAAWLNGAYVGGFSGDPDKVASWEVLKFNHSSLRSRDN VLTIITDYTGHDQNSQKPIGTQNPRGIMGATLIGGGNFTLWRIQGNAGGEKNIDPVRGPMNEGG LYGERMGWHLPGYQVPESALDSSPLEGVSGAEGRFYTTSFQLDLEEDLDVPIGLQLSAPAGTEA VVQIFMNGYQFGHYLPHIGPQSLFPFPPGVIYNRGQNSLAISMWALTDAGARLEQVELKAYAKY RSGFDFNRDWTYLQPGWKDRTEYA		
Probable beta-galactosidase A (Lactase A) B8N6V7.1 GI:300680889	MKLLSVAAVALLAAQAAGASIKHRLNGFTILEHPDPAKRDLLQDIVTWDDKSLFINGERIMLFS GEVHPFRLPVPSLWLDIFHKIRALGFNCVSFYIDWALLEGKPGDYRAEGIFALEPFFDAAKEAG IYLIARPGSYINAEVSGGGFPGWLQRVNGTLRSSDEPFLKATDNYIANAAAAVAKAQITNGGPV ILYQPENEYSGGCCGVKYPDADYMQYVMDQARKADIVVPFISNDASPSGHNAPGSGTGAVDIYG HDSYPLGFDCANPSVWPEGKLPDNFRTLHLEQSPSTPYSLLEFQAGAFDPWGGPGFEKCYALVN HEFSRVFYRNDLSFGVSTFNLYMTFGGTNWGNLGHPGGYTSYDYGSPITETRNVTREKYSDIKL LANFVKASPSYLTATPRNLTTGVYTDTSDLAVTPLIGDSPGSFFVVRHTDYSSQESTSYKLKLP TSAGNLTIPQLEGTLSLNGRDSKIHVVDYNVSGTNIIYSTAEVFTWKKFDGNKVLVLYGGPKEH HELAIASKSNVTIIEGSDSGIVSTRKGSSVIIGWDVSSTRRIVQVGDLRVFLLDRNSAYNYWVP ELPTEGTSPGFSTSKTTASSIIVKAGYLLRGAHLDGADLHLTADFNATTPIEVIGAPTGAKNLF VNGEKASHTVDKNGIWSSEVKYAAPEIKLPGLKDLDWKYLDTLPEIKSSYDDSAWVSADLPKTK NTHRPLDTPTSLYSSDYGFHTGYLIYRGHFVANGKESEFFIRTQGGSAFGSSVWLNETYLGSWT GADYAMDGNSTYKLSQLESGKNYVITVVIDNLGLDENWTVGEETMKNPRGILSYKLSGQDASAI TWKLTGNLGGEDYQDKVRGPLNEGGLYAERQGFHQPQPPSESWESGSPLEGLSKPGIGFYTAQF DLDLPKGWDVPLYFNFGNNTQAARAQLYVNGYQYGKFTGNVGPQTSFPVPEGILNYRGTNYVAL SLWALESDGAKLGSFELSYTTPVLTGYGNVESPEQPKYEQRKGAY	702	
Probable beta-galactosidase B (Lactase B) B8NKI4.2 GI:68115	MLISKTVLSGLALGASFVGVSAQQNSTRWPLHDNGLTDTVEWDHYSFLINGQRHFVFSGEFHYW RIPVPELWRDLLEKIKAAGFTAFSIYNHWGYHSPKPGVLDFENGAHNFTSIMTLAKEIGLYMII RPGPYVNAEANAGGLPLWTTTGAYGKLRDNDPRYLEALTPYWANISKIIAPHLITNGGNVILYQ IENEYAEQWLDEETHEPNTSGQEYMQYLEDVARENGIDAPLIHNLPNMNGHSWSKDLSNATGNV DVIGVDSYPTCWTCNVSECASTNGEYIPYKTLIYYDYFKELSPTQPSFMPEFQGGSYNPWGGPQ GGCPDDLGPDFANLFYRNLISQRVSAISLYMLYGGTNWGWHASTDVATSYDYSSPISENRKLIE KYYETKVLTQFTKIAQDLSKVDRLGNSTKYSSNPAVSVAELRNPDTGAAFYVTQHEYTPSGTVE KFTVKVNTSEGALTIPQYGSQITLNGHQSKIIVTDFKFGSKTLLYSTAEVLTYAVIDGKEVLAL WVPTGESGEFTVKGVNSAKFADKGRTANIEIHPGANNVTVSFMQRSGMSLVELGDGTRIVLLDR SAAHVFWSTPLNNDPAEAGNNTVLVHGPYLVRSAKLEGCDLKLTGDIQNSTEVSIFAPKSVCSV NWNGKKTSVKSAKGGVITTTLGGDAKFELPTISGWKSADSLPEIAKDYSATSKAWVVATKTNSS NPTPPAPNNPVLYVDENDIHVGNHIYRATFPSTDEPPTDVYLNITGGRAFSYSVWLNSDFIGSW LGTATTEQNDQTFSFSNATLSTDEDNILVVVMDNSAHDLRDGALNPRGITNATLIGPGSYSFTE WKLAGNAGFEDHLDPVRAPLNEGSLYAERVGIHLPGYEFDEAEEVSSNSTSLTVPGAGIRVFRT VVPLSVPQGLDVSISFRLTAPSNVTFTSAEGYTNQLRALLFVNGYQYGRFNPYIGHQIDFPVPP GVLDYNGDNTIAVTVWSQSVDGAEIKVDWNVDYVHETSFDMNFDGAYLRPGWIEERREYA	703	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Probable beta-galactosidase B (Lactase B) Q0CMF3.2 GI:300681013	MARFPQLLFLLLASIGLLSAAQNHSDSEWPLHDNGLSTVVQWDHYSFHVHGQRIFVFSGEFHYW RIPVPGLWRDILEKIKAAGFTAFAFYSSWGYHAPNNHTVDFSTGARDITPIYELAKELGMYIIV RPGPYVNAEASAGGYPLWVTTGAYGSLRNDDARYTAAWKPYFAKMSEITSQYQVTDGHNTFCYQ IENEYGQQWIGDPVDRNPNQTAVAYMELLEASARENGIVVPLTANDPNMNTKSWGSDWSHAGGN VDVVGLDSYPSCWTCDVTQCTSTNGEYVPYKVMQYYDYFQEVQPTMPGFMPEFQGGSYNPWAGP EGGCPGDTGVDFANLFYRWNIAQRVTAMSLYMLYGGTNWGAIAAPVTATSYDYSSPISEDRSIG SKYYETKLLALFTRSATDLTMTDRIGNGTHYTNNPAVAAYELRNPVTNGAFYVTIHADSTVGTD ESFRLNVNTSAGALTVPSKGSIRLNGHQSKIIVTDFRFGPSHTLLYSTAEVLTHAVMDKKATLV LWVPTGESGEFAVKGAKSGKVERCPQCSNATFTRKKDVLVVNFTQAGGMSVLQLNNGVRVVLLD RAAAYKFWAPPLTDDPFAPETDLVLVQGPYLVRSASLSGSTLALRGDSANETALEVFASKKVHT VTWNGKRIKTSRSSYGSLTASLAAPPAVSLPALSSAQWKSQDSLPERLPSYDDSGPAWVDANHM TTQNPRTPDTLPVLYADEYGFHNGIRLWRGSFTDAASGVYLNVQGGAAFGWSAYLNGHFLGSHL GTATTSQANKTLLFPAGTLRKNTTNTILVIHDDTGHDQTTGALNPRGILAARLLAPSDSSTAPN FTQWRVAGTAGGESDLDPVRGVYNEDGLFAERMGWHLPGFDDADWPANNSTTTRGAQVSLSVTG ATVRFFRAVVPLHLPRGVDASISFMLGTPAGASTAYRAQLFVNGYQYGRFYPHIGNQVVYPVPA GVLDYDGENTIGVAVWAQSEAGAEMSLDWRVNYVADSSLDAVRVAAEGALRPGWSEERLQYA	704	
Probable beta-galactosidase B\ (Lactase B) Q2U6P1.2 GI:300681012	MLISKTVLSGLALGASFVGVSAQQNSTRWPLHDNGLTDTVEWDHYSFLINGQRHFVFSGEFHYW RIPVPELWRDLLEKIKAAGFTAFSIYNHWGYHSPKPGVLDFENGAHNFTSIMTLAKEIGLYMII RPGPYVNAEANAGGLPLWTTTGAYGKLRDNDPRYLEALTPYWANISKIIAPHLITNDGNVILYQ IENEYAEQWLDEETHEPNTSGQEYMQYLEDVARENGIDAPLIHNLPNMNGHSWSKDLSNATGNV DVIGVDSYPTCWTCNVSECASTNGEYIPYLKLTYPQISYFKELSPTQPSFMPEFQGGSYNPWGG PQGGCPDDLGPDFANLFYRNLISQRVSAISLYMLYGGTNWGWHASTDVATSYDYSSPISENRKL IEKYYETKVLTQFTKIAQDLSKVDRLGNSTKYSSNPAVSVAELRNPDTGAAFYVTQHEYTPSGT VEKFTVKVNTSEGALTIPQYGSQITLNGHQSKIIVTDFKFGSKTLLYSTAEVLTYAVIDGKEVL ALWVPTGESGEFTVKGVNSAKFADKGRTANIEIHPGTNNVTVSFMQRSGMSLVELGDGTRIVLL DRSAAHVFWSTPLNNDPAEAGNNTVLVHGPYLVRSAKLEGCDLKLTGDIQNSTEVSIFAPKSVC SVNWNGKKTSVKSAKGGVITTTLGGDAKFELPTISGWKSADSLPEIAKDYSATSKAWVVATKTN SSNPTPPAPNNPVLYVDENDIHVGNHIYRATFPSTDEPPTDVYLNITGGRAFGYSVWLNSDFIG SWLGTATTEQNDQTFSFSNATLSTDEDNILVVVMDNSAHDLRDGALNPRGITNATLIGPGSYSF TEWKLAGNAGFEDHLDPVRAPLNEGSLYAERVGIHLPGYEFDEAEEVSSNSTSLTVPGAGIRVF RTVVPLSVPQGLDVSISFRLTAPSNVTFTSAEGYTNQLRALLFVNGYQYGRFNPYIGHQIDFPV PPGVLDYNGDNTIAVTVWSQSVDGAEIKVDWNVDYVHETSFDMNFDGAYLRPGWIEERREYA	705	
Probable beta-galactosidase A (Lactase A) Q2UCU3.1 GI:121801672	MKLLSVAAVALLAAQAAGASIKHRLNGFTILEHPDPAKRDLLQDIVTWDDKSLFINGERIMLFS GEVHPFRLPVPSLWLDIFHKIRALGFNCVSFYIDWALLEGKPGDYRAEGIFALEPFFDAAKEAG IYLIARPGSYINAEVSGGGFPGWLQRVNGTLRSSDEPFLKATDNYIANAAAAVAKAQITNGGPV ILYQPENEYSGGCCGVKYPDADYMQYVMDQARKADIVVPFISNDASPSGHNAPGSGTGAVDIYG HDSYPLGFDCANPSVWPEGKLPDNFRTLHLEQSPSTPYSLLEFQAGAFDPWGGPGFEKCYALVN HEFSRVFYRNDLSFGVSTFNLYMTFGGTNWGNLGHPGGYTSYDYGSPITETRNVTREKYSDIKL	706	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	LANFVKASPSYLTATPRNLTTGVYTDTSDLAVTPLIGDSPGSFFVVRHTDYSSQESTSYKLKLP TSAGNLTIPQLEGTLSLNGRDSKIHVVDYNVSGTNIIYSTAEVFTWKKFDGNKVLVLYGGPKEH HELAIASKSNVTIIEGSDSGIVSTRKGSSVIIGWDVSSTRRIVQVGDLRVFLLDRNSAYNYWVP ELPTEGTSPGFSTSKTTASSIIVKAGYLLRGAHLDGADLHLTADFNATTPIEVIGAPTGAKNLF VNGEKASHTVDKNGIWSSEVKYAAPEIKLPGLKDLDWKYLDTLPEIKSSYDDSAWVSADLPKTK NTHRPLDTPTSLYSSDYGFHTGYLIYRGHFVANGKESEFFIRTQGGSAFGSSVWLNETYLGSWT GADYAMDGNSTYKLSQLESGKNYVITVVIDNLGLDENWTVGEETMKNPRGILSYKLSGQDASAI TWKLTGNLGGEDYQDKVRGPLNEGGLYAERQGFHQPQPPSESWESGSPLEGLSKPGIGFYTAQF DLDLPKGWDVPLYFNFGNNTQAARAQLYVNGYQYGKFTGNVGPQTSFPVPEGILNYRGTNYVAL SLWALESDGAKLGSFELSYTTPVLTGYGNVESPEQPKYEQRKGAY		
Probable beta-galactosidase C (Lactase C) Q2UMD5.1 GI:121804415	MRLLSFIYLVWLALLTGTPQVSATDNGKTSDVAWDKYSLSVKGERLFVFSGEFHYQRLPVPELW LDVFQKLRANGFNTISVYFFWSYHSASEDVFDFTTGAHDIQRLFDYAKQAGLYVIARAGPYCNA ETSAGGFALWAANGQMGSERTSDEAYYKKWKPWILEVGKIIAANQITNGGPVILNQHENELQET TYDSNDTKVIYMEQVAKAFEEAGVVVPSSHNEKGMRTVSWSTDYKNVGGAVNVYGLDSYPGSLS CANPNSGFNLLRTYYQWFQNYSYTQPEYLAEFEGGWFQPWGGSFYDSCASELSPEFADVYYKNN IGSRVTLHNIYMTFGGTNWGHSAAPVVYTSYDYGSPLRETREIRDKLKQTKLLGLFTRVSKDLL KTYMEGNGTSYTSDDSIYTWALRNPDSDAGFYVVAHNTSSSREVTTFSLNITTSAGAMTIPDIE LDGRQSKIIVTDYSIGSESSLLYSSAEVLTYATLDVDVLVFYLNAGQKGAFVFKDAPADLKYQT YGNSNLSALETSQGTQYSYTQGEGVTAVKFSNGVLVYLLDKETAWNFFAPPTVSSPTVAPNEHI LVFGPYLVRGASIKHDTVEIVGDNSNSTSIEIYTGDEHVKKVSWNGNLIDTRATAYGSLIGTVP GAEDIEISLPSLSSWKAQDTLPEISPDYDDSRWTICNKTTSVNSVAPLSLPVLYSGDYGYHTGT KIYRGRFDGQNATGANVTVQNGVAAGWAAWLNGAYVGGFSGDPDKVASWEVLKFNHSSLRSRDN VLTIITDYTGHDQNSQKPIGTQNPRGIMGATLIGGGNFTLWRIQGNAGGEKNIDPVRGPMNEGG LYGERMGWHLPGYQVPESALDSSPLEGVSGAEGRFYTTSFQLDLEEDLDVPIGLQLSAPAGTEA VVQIFMNGYQFGHYLPHIGPQSLFPFPPGVIKNRGQNSLAISMWALTDAGARLEQVELKAYAKY RSGFDFNRDWTYLQPGWKDRTEYA	707	
Probable beta-galactosidase E (Lactase E) Q4WG05.1 GI:74668464	MKSLLKRLIALAAAYSVAAAPSFSHHSSQDAANKRELLQDLVTWDQHSLFVRGERLMIFSGEFH PFRLPVPGLWFDVFQKIKSLGFNAVSFYTDWGLMEGNPGHVVTDGIWSLDEFFTAAREAGLYLI ARPGPYINAETSAGGIPGWVLRRKGIIRSNSEDYLRATDTYMATLGKIIAKAQITNGGPVILVQ PENEYTTWPNVSESEFPTTMNQEVMAYAEKQLRDAGVVVPTVVNDNKNLGYFAPGTGLGETDLY GIDAYPMRYDCGNPYVWPTYRFPRDWQHEHRNHSPTTPFAIMEFQGGSGDGWGGVTEDGCAILV NNEAVRVVYKNNYGFGVRVFNIYMTYGGTNWGNLGYYGGYTSYDYGAAITEDRQIWREKYSEEK LQANFLKVSPAYLTSTPGNGVNGSYTGNKDITVTPLFGNGTTTNLYLVRHADFTSTGSAQYNLS ISTSVGNVTIPQLGGSLSLNGRDSKFHITDYDVGGFNLIYSSAEVFTWAKGDNKKRVLVLYGGA GELHEFALPKHLPRPTVVEGSYVKIAKQGSAWVVQWEVAAQRRVLRAGKLEIHLLWRNDAYQHW VLELPAKQPIANYSSPSKETVIVKGGYLLRSAWITDNDLHLTGDVNVTTPLEVISAPKRFDGIV FNGQSLKSTRSKIGNLAATVHYQPPAISLPDLKRLDWKYIDSLPEISTEYNDEGWTPLTNTYTN NTREFTGPTCLYADDYGYHGGSLIYRGHFTANGDESWVFLNTSGGVGFANSVWLNQTFLGSWTG	708	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	SGRNMTYPRNISLPHELSPGEPYVFTVVIDHMGQDEEAPGTDAIKFPRGILDYALSGHELSDLR WKMTGNLGGEQYQDLTRGPLNEGAMYAERQGYHLPSPPTSSWKSSNPIKEGLTGAGIGFYATSF SLDLPEGYDIPLSFRFNNSASAARSGTSYRCQLFVNGYQFGKYVNDLGPQTKFPVPEGILNYNG		
Probable beta-galactosidase C (Lactase C) Q4WNE4.1 GI:74671041	MRIFSFLFLLLLGILTGQGLVSGTDNGKTTDVTWDKYSLSVKGQRLFVFSGEFHYQRLPVPELW LDVFQKLRANGFNAISVYFFWSFHSASEGEFDFENGAHDIQRLFDYAKEAGLYVIARAGPYCNA ETSAGGFALWAANGQMGNERTSDEAYYEKWRPWILEVGKIIAKNQITNGGPVILNQHENELVET TYDPNHTLVVYMKQIAQVFEEAGIVVPSSHNEKGMRGVSWSTDYHNVGGAVNIYGLDSYPGGLS CTNPNSGFNLVRTYHQWFQNYSFTQPSYLPEFEGGWFQPWGGSFYDTCATELSPEFPDVYYKNN IGSRVTLHSIYMTYGGTNWGHSAAPVVYTSYDYAAPLRETREIRDKLKQTKLIGLFTRVSKDLL KTYMEGNGTGYTSDSSIYTWSLRNPDTNAGFYVLAHSTSSTRDVTTFTLNVTTSAGAISIPDIE LNGRQSKIIVTDYNFGTNSTLLFSSAEVLTYANLDVNVLVFYLNVGQKGTFVFKDEPKLAFQTY GNSNLTTSESSYGTQYSYTQGKGVTAVKFSNGVLAYFLDKESAWNFFAPPTTSSPQVAPNEHIL VQGPYLVRGASVNHGIVEITGDNANTTSIEVYTGNSQVKKIKWNGKTIETRKTAYGSLIGTAPG AEDVKIQLPSLDSWKAQDTLPEIQPDYDDSKWTVCNKTTSVNAIAPLSLPVLYSGDYGYHAGTK VYRGRFDGRNVTGANVTVQNGAAAGWAAWVNGQYAGGSAGSPNLAATSAVLTFNSSSLKDQDNV LTVVTDYTGHDQNSVRPKGTQNPRGILGATLIGGGNFTSWRIQGNAGGEKNIDPVRGPMNEGGL YGERMGWHLPGYKVPKSASKSSPLDGVSGAEGRFYTTTFKLKLDKDLDVPIGLQLGAPEGTKAV VQVFMNGYQFGHYLPHTGPQSLFPFPPGVINNRGENTLAISMWALTDAGAKLDKVELVAYGKYR SGFDFNQDWGYLQPGWKDRSQYA	709	
Probable beta-galactosidase B (Lactase B) Q4WRD3.1 GI:74672078	MAHIYRLLLLLLSNLWFSTAAQNQSETEWPLHDNGLSKVVQWDHYSFQVNGQRIFIFSGEFHYW RIPVPELWRDILEKVKATGFTAFAFYSSWAYHAPNNSTVDFSTGARDITPIFELAKELGMYMIV RPGPYVNAEASAGGFPLWLMTGEYGSLRNDDPRYTAAWTPYFANMSQITSKYQVTDGHNTLVYQ IENEYGQQWIGDPKNRNPNKTAVAYMELLEASARENGITVPLTSNDPNMNSKSWGSDWSNAGGN VDVAGLDSYPSCWTCDVSQCTSTNGEYVPYKVIDYYDYFQEVQPTLPSFMPEFQGGSYNPWAGP EGGCPQDTSAEFANLFYRWNIGQRVTAMSLYMLYGGTNWGAIAAPVTATSYDYSAPISEDRSIG AKYSETKLLALFTRTAKDLTMTEAIGNGTQYTTNTAVRAFELRNPQTNAGFYVTFHTDTTVGGN QAFKLHVNTSVGALTVPKNEGLIQLNGHQSKIIVTDFTLGKRTLLYSTAEVLTYAVFENRPTLV LWVPTGESGEFAIKGAKSGKVENGDGCSGIKFKREKDYLVVNFSQAKGLSVLRLDNGVRVVLLD KAAAYRFWAPALTDDPNVQETETVLVHGPYLVRSASISKTTLALRGDSVEKTTLEIFAPHSVRK ITWNGKEVQTSHTPYGSLKATLAAPPDIKLPALTSWRSNDSLPERLPSYDDSGPAWIEANHMTT SNPSPPATFPVLYADEYGFHNGVRLWRGYFNGSASGVFLNIQGGSAFGWSAWLNGHFLDSHLGT ATTSQANKTLTFPSSILNPTENVLLIVHDDTGHDQTTGALNPRGILEARLLSNDTSSPPPEFTH WRLAGTAGGESNLDPIRGVFNEDGLFAERMGWHLPGFDDSAWTSENSATSASSALSFTGATVRF FRSVVPLNIPAGLDVSISFVLSTPTAAPKGYRAQLFVNGYQYGRYNPHIGNQVVFPVPPGILDY QGDNTIGLAVWAQTEEGAGIQVDWKVNYVADSSLSVAGFGKGLRPGWTEERLKFA	710	
Probable beta-galactosidase A (Lactase A)	MKLLSVCAIALLAAQAAGASIKHMLNGFTLMEHSDPAKRELLQKYVTWDEKSLFVNGERIMIFS GEVHPFRLPVPSLWLDVFQKIKALGFNCVSFYVDWALLEGKPGEYRAEGNFALEPFFDVAKQAG IYLLARPGPYINAEASGGGFPGWLQRVNGTLRTSDPAYLKATDNYIAHVAATIAKGQITNGGPV	711	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Q4WS33.2 GI:300681010	ILYQPENEYSGACCDATFPDGDYMQYVIDQARNAGIVVPLINNDAWTGGHNAPGTGKGEVDIYG HDSYPLGFDCGHPSVWPKGNLPTTFRTDHLKQSPTTPYSLIEFQAGSFDPWGGPGFAACAALVN HEFERVFYKNDLSFGAAILNLYMTFGGTNWGNLGHPGGYTSYDYGSPLTESRNVTREKYSELKL IGNFVKASPSYLLATPGNLTTSGYADTADLTVTPLLGNGTGSYFVVRHTDYTSQASTPYKLSLP TSAGRLTVPQLGGTLTLNGRDSKIHVVDYNVAGTNIIYSTAEVFTWKNFGDSKVLILYGGPGEH HELAVSLKSDVQVVEGSNSEFKSKKVGDVVVVAWDVSPSRRIVQIGDLKIFLLDRNSVYNYWVP QLDKDDSSTGYSSEKTTASSIIVKAGYLVRTAYTKGSGLYLTADFNATTPVEVIGAPSNVRNLY INGEKTQFKTDKNGIWSTEVKYSAPKIKLPSMKDLDWKYLDTLQEVQSTYDDSAWPAADLDTTP NTLRPLTTPKSLYSSDYGFHTGYLIYRGHFVADGSETTFDVRTQGGSAFGSSVWLNESFLGSWT GLNANADYNSTYKLPQVEQGKNYVLTILIDTMGLNENWVVGTDEMKNPRGILSYKLSGRDASAI TWKLTGNLGGEDYQDKIRGPLNEGGLYAERQGFHQPQPPSQKWKSASPLDGLSKPGIGFYTAQF DLDIPSGWDVPLYFNFGNSTKSAYRVQLYVNGYQYGKFVSNIGPQTSFPVPQGILNYQGTNWVA LTLWALESDGAKLDDFELVNTTPVMTALSKIRPSKQPNYRQRKGAY		
Probable beta-galactosidase A (Lactase A) Q4ZHV7.1 GI:74645200	MKLSSACAIALLAAQAAGASIKHRINGFTLTEHSDPAKRELLQKYVTWDDKSLFINGERIMIFS GEFHPFRLPVKELQLDIFQKVKALGFNCVSFYVDWALVEGEPGEYRADGIFDLEPFFDAASEAG IYLLARPGPYINAESSGGGFPGWLQRVNGTLRSSDKAYLDATDNYVSHVAATIAKYQITNGGPI ILYQPENEYTSGCCGVEFPDPVYMQYVEDQARNAGVVIPLINNDASASGNNAPGTGKGAVDIYG HDSYPLGFDCANPTVWPSGDLPTNFRTLHLEQSPTTPYAIVEFQGGSYDPWGGPGFAACSELLN NEFERVSYKNDFSFQIAIMNLYMIFGGTNWGNLGYPNGYTSYDYGSAVTESRNITREKYSELKL LGNFAKVSPGYLTASPGNLTTSGYADTTDLTVTPLLGNSTGSFFVVRHSDYSSEESTSYKLRLP TSASSVTIPQLGGTLTLNGRDSKIHVTDYNVSGTNIIYSTAEVFTWKKFADGKVLVLYGGAGEH HELAISTKSNVTVIEGSESGISSKQTSSSVVVGWDVSTTRRIIQVGDLKILLLDRNSAYNYWVP QLATDGTSPGFSTPEKVASSIIVKAGYLVRTAYLKGSGLYLTADFNATTSVEVIGVPSTAKNLF INGDKTSHTVDKNGIWSATVDYNAPDISLPSLKDLDWKYVDTLPEIQSSYDDSLWPAADLKQTK NTLRSLTTPTSLYSSDYGFHTGYLLYRGHFTATGNESTFAIDTQGGSAFGSSVWLNGTYLGSWT GLYANSDYNATYNLPQLQAGKTYVITVVINNMGLEENWTVGEDLMKTPRGILNFLLAGRPSSAI SWKLTGNLGGEDYEDKVRGPLNEGGLYAERQGFHQPEPPSQDWKSSSPLEGLSEAGIGFYSASF DLDLPKGWDVPLFLNIGNSTTPSPYRVQVYVNGYQYAKYISNIGPQTSFPVPEGILNYRGTNWL AVTLWALDSAGGKLESLELSYTTPVLTALGEVESVDQPKYKKRKGAY	712	
Probable beta-galactosidase B (Lactase B) Q5BEQ0.2 GI:300681009	MATAFWLLLFLLGSLHVLTAAQNSSQSEWPIHDNGLSKVVQWDHYSFYINGQRIFLFSGEFHYW RIPVPALWRDILEKIKAIGFTGFAFYSSWAYHAPNNQTVDFSTGARDITPIYDLAKELGMYIIV RPGPYVNAEASAGGFPLWLTTGAYGSTRNDDPRYTAAWEPYFAEVSEITSKYQVTDGHYTLCYQ IENEYGQQWIGDPRDRNPNQTAIAYMELLQASARENGITVPLTGNDPNMNTKSWGSDWSDAGGN LDTVGLDSYPSCWSCDVSVCTGTNGEYVPYKVLDYYDYFQEVQPTMPFFMPEFQGGSYNPWDGP EGGCTEDTGADFANLFYRWNIGQRVSAMSLYMMFGGTNWGGIAAPVTASSYDYSAPISEDRSIG SKYYETKLLALFTRCAKDLTMTDRLGNGTQYTDNEAVIASELRNPDTNAAFYVTTHLDTTVGTD ESFKLHVNTSKGALTIPRHGGTIRLNGHHSKIIVTDFNFGSETLLYSTAEVLTYAVFDRKPTLV LWVPTGESGEFAIKGAKSGSVAKCSGCSNIKFHRDSGSLTVAFTQGEGISVLQLDNGVRVVLLD	713	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	RQKAYTFWAPALTDNPLVPEGESVLVSGPYLVRTARLARSTLTLRGDSKGETLEIFAPRKIKKV TWNGKAVEATRTSYGSLKAILAKPPSVELPTLNGWKYSDSLPERFPTYDDSGAAWVEIDANHMT TPNPNKPATLPVLYADEYGFHNGVRLWRGYFNSSASGVYLNIQGGAAFGWSAWLNGHFLGSHLG SASIQQANGTLDFPANTLNTEGTPNVLLVVHDDTGHDQTTGVLNPRGILEARLLSEASDNNDDD SPGFTHWRVAGTAGGESDLDPVRGVYNEDGLYAERVGWHLPGFDDSKWATVNGTSLSFTGATVR FFRTVIPPLSIPENTDVSISFVFSTPNVNNTSAGNTSAFRAQLFVNGYQYGRYNPYVGNQVVYP VPPGILDYNGENTIGVAVWAQTEAGARLNLDWRVNYVLGSSLDAGRLDLSFVAIAYVYIFECLQ L		
Probable beta-galactosidase A (Lactase A) Q5BFC4.2 GI:300681016	MRLLPVWTAALLAAQAAGVALTHKLNGFTITEHPDAEKRELLQKYVTWDDKSLFINGERIMIFG AEIHPWRLPVPSLWRDILQKVKALGFNCVSFYVDWALLEGKPGEYRAEGSFAWEPFFDAASDLG IYLIARPGPYINAEASGGGFPGWLQRLNGTIRSSDQSYLDATENYVSHIGGLIAKYQITNGGPV ILYQPDNEYSGGCCGQEFPNPDYFQYVIDQARRAGIVVPTISNDAWPGGHNAPGTGKGEVDIYG HDNYPLGFDCANPDVWPEGNLPTDYRDLHLEISPSTPYALVEYQVGAFDPWGGPGFEQCAALTG YEFERVFHKNTFSFGVGILSLYMTFGGTNWGNLGHPGGYTSYDYGSPIKETREITREKYSELKL LGNFIKSSPGYLLATPGKLTNTTYTNTADLTVTPLLGNGTGSFFVLRHSDYSSQASTPYKLRLP TSAGQLTIPQLGGSLVLNGRDSKVHLVDYDVAGTKILYSTAEVFTWKKFHDGKVLVLYGGPGEH HELAVSSKAKVKVVEGLGSGISSKQIRGAVVVAWDVEPARRIVQIGDLKIFLLDRNSAYNYWVP QLGTETSIPYATEKAVAASVIVKAGYLVRTAYVKGRDLHLTADFNATTPVEVIGAPKTAENLFI NGKKAHHTVDKNGIWSTEVGYSPPKIVLPVLEDLKWKSIDTLPEIQPSYDDSPWPDANLPTKNT IYPLRTPTSLYASDYGFHTGYLLFRGHFTANGRESNFSIQTQGGQAFGSSVWLSGTYLGSWTGD NDYQDYNATYTLPSLKAGKEYVFTVVVDNMGLNENWIVGQDEMKKPRGILNYELSGHEASDITW KLTGNFGGEDYVDKVRGPLNEGGLYAERHGYHQPYPPTKSKDWKSSTPLTGLSKPGISFYTASF DLDIKSGWDVPIYFEFGNSTTPAPAYRVQLYVNGWQYGKYVNNIGPQTRFPVPEGILNYKGTNW VAVTLWALEGSGAKLDSFKLVHGIPVRTALDVEGVELPRYQSRKGVY	471	
Lactase, partial [*Aspergillus niger*] ABL07484.1 GI:118582212	SIKHRINGFTLTEHSDPAKRELLQKYVTWDDKSLFINGERIMIFSGEFHPFRLPVKELQLDIFQ KVKALGFNCVSFYVDWALVEGKPGEYGADGIFDLEPFFDAASEAGIYLLARPGPYINAESSGGG FPGWLQRVNGTLRTSDKAYLEATDNYVSHIAATIAKYQITNGGPIILYQPENEYTGGCCGVEFP DPVYMQYVEDQARNAGVVIPLINNDASASGHNAPGTGEGAVDIYGHDSYPLGFDCANPTVWPSG DLPTNFRTLHLVQSPTTPYAIVEFQGGSYDPWGGPGFAACSELLNNEFERVFYKNDFSFQIAIM NLYMIFGGTNWGNLGYPNGYTSYDYGSAVTESRNITREKYSELKLLGNFAKVSPGYLTASPGNL TTSGYADTTDLTVTPLLGNSTGSFFVVRHSDYSSEDSTSYKLRLPTSAGTVTIPQLGGTLTLNG RDSKIHVTDYNVSGTNIIYSTAEVFTWKKFADGKVLVLYGGAGEHHELAISTKSNVTVIEGSES GISSKQTSSSVIVGWDVSTTRRIIQVGDLKVLLLDRNSAYNYWVPQLATDGTSPGFSTSETVAS SIIVKAGYLVRTAYLKGSGLYLTADFNATTSVEVIGVPSTAKNLFINGDKTSHTVDKNGIWSAT VEYNAPDISLPSLKDLDWKYVDTLPEIQSSYDDSLWPAADLKQTKNTLRSLTTPTSLYSSDYGF HTGYLLYRGHFTATGNESTFSIDTQGGSAFGSSVWLNGTYLGSWTGLYVNSDYNATYKLPQLQA GKSYVITVVIDNMGLEENWTVGEDLMKTPRGILNFLLAGRPGSAISWKLTGNLGGEDYEDKVRG PLNEGGLYAERQGFHQPEPPSGNWKSSSPLEGLSEAGIGFYSAKFDLDLPKGWDVPLFLNIGNS	715	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TTPSPYRVQVYVNGYQYAKYISNNGPQTSFPVPEGILNYRGTNWLAVTLWALDSAGGKLESLEL SYTTPVLTALGEVESVDQPKYKKRKGAY		
Unnamed protein product [Aspergillus oryzae RIB40] BAE57671.1 GI:83767532	MGSTSTSTLPPDFLWGFATASYQIEGAVNEDGRGPSIWDTFCKIPGKIAGGANGDVACDSYHRT HEDIALLKACGAKAYRFSLSWSRIIPLGGRNDPINEKGLQYYIKFVDDLHAAGITPLVTLFHWD LPDELDKRYGGLLNKEEFVADFAHYARIVFKAFGSKVKHWITFNEPWCSSVLGYNVGQFAPGRT SDRSKSPVGDSSRECWIVGHSLLVAHGAAVKIYRDEFKASDGGEIGITLNGDWAEPWDPENPAD VEACDRKIEFAISWFADPIYHGKYPDSMVKQLGDRLPKWTPEDIALVHGSNDFYGMNHYCANFI KAKTGEADPNDTAGNLEILLQNKKGEWVGPETQSPWLRPSAIGFRKLLKWLSERYNYPKIYVTE NGTSLKGENDLPLEQLLQDDFRTQYFRDYIGAMADAYTLDGVNVRAYMAWSLME	716	
Unnamed protein product [Aspergillus oryzae RIB40] BAE62705.1 GI:83772577	MARVRLKLPADFIWGVSSSSWQIEGGLQLEGRGPSVLDTIGNVLSPEAADRSDANVANMHYFMY EQDIARLAAAGIPYYSFSLSWPRIVPFGVAGSPVNTQGLDHYDDLINTCIKYGVTPIVTLNHVD APTAVQADLDSLPEHFLYYAKIVMTRYADRVPYWVTFNEPNIGVGTLFQKYQDLTSALIAHADV YDWYKNTLGGTGKITMKFANNLAMPLDTQDSSHIAAASRYQDILLGIMSNPLFLGKQYPDAAID TVDMMQPLTDDQIKHIHGKIDFWSFDPYTAQYASPLPQGTEACASNSSDPFWPTCVILSNVQAN GWLMGQASNAYAYLAPQYVRQQLGYIWNTFRPSGILIAEYGFNPFLESNRTLDAQRYDLERTLY YQDFLTETLKAIHEDNVNVIGALAWSIADNNEFGSYEEQYGLQTVNRTNGKFTRTYKRSLFDYV DFFHRHVQSA	717	
Unnamed protein product [Aspergillus oryzae RIB40] BAE63197.1 GI:83773069	MNVNMFKAGDDILQDVDQSCKDRLPAVEELPLPPSFTWGTATAAYQVEGGAFQDGKGKSIWDTF THLDPSRTNGENGDIACDHYNRMAEDVVLMASYGVDVYRFSIAWARILPLGGRGDPINEKGIAF YNNLIDCLLEHNIEPVVTLYHWDVPQGLYDRYGAFLDTTEFRADFEHFARLCFSRFGDRVKRWI TFNEPYIISIFGHHSGVLAPGRSSATGGDSRTEPWRVGHTIILAHTAAVQAYATDFQPTQKGDI SIVLNGHYYEPWDAGSEEHRLAAQRRLEFYIGWFGDPIFLGKDYPAPMRAQLGSRLPEFTSEEL DLLRRSAPINSFYGMNHYTTKYARALPDPPAEDDCTGNVEEGPTNSEGKTMGPLSGMSWLRVTP AGFRKLLNWVWDRYRRPIVVTENGCPCPGESQMTKEQALDDQFRIRYFGLYLDAISRAIYDDGV KVEGYYVWSLMDNFEWSAGYGPRYGITHVDFTTLVRTPKQSAKYLHHSFNKRRATSLR	718	
Beta-galactosidase [Aspergillus fumigatus Af293] XP_753202.1 GI:70996895	MTLSAVPDYENQHILQRNRLKPRAYFLPATSISLNGRWDFHYAASPVSAPEPTWSKGTKNATAE PRRDSNQFSSDGADSKTAWAPITVPGHWQLQGYGRPHYTNVIYPFPVCPPFVPTENPTGTYRRT FHVPAEWDASSQLRLRFDGVDSAYHVWVNGVPIGYSQGSRNPAEFDVSQVVDRDGANELFVRVY QWSDGSYIEDQDQWWLSGIFRDVTLLAFPGQARIEDFFVRTALDKDYVDATLRLSVDLALATAA IVQVTLSNPSTGSTLQTEKYSLGEKQDKLEAELSVSNPNKWTAETPNLYNLCIALYVDGAKDPV QTINHRVGFRQVEIKNGNITVNGVPVMFRGVNRHDHHPRFGRAVPLSFLREDLLIMKRHNVNAL RCSHYPSHPRLYELCDELGLWVMDEADLECHGFYDAIARPLDIPESMDYEERKKLTFGQAAQFT TNNPEWKEAYVDRMAQMVQRDKNHSCIVIWSLGNEAFYGSNHQAMYDYVKQVDPSRPVHYEGDM EAKTVDMYSYMYPSLERLVGFATAEGDEFKKPIVLCEYAHAMGNAPGGLEEYMEAFRTHRRLQG GWVWEWANHGLWDEKKGWYGYGGDFGDTPHDGNFVLDGLLFSDHTPTPGITELKKAYAPVRVWP GEDGTLVVANDYNFVGLEGLQASYKIEVLGDSGRIIATGIIELPPIPAGQNGTIKLPSAPATAI PGEVWLTISFLQKGETAWAGNNYEVAWYQQCLKSSSPRFSLAVPAEALTHSSTKTSHRISGASF SLEFSRETGSLYAWTAGGLSLLDQSSSTGAISPGFWRPPTDNDMSHDLLEWRRFGLDTLTSQLR	719	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	KMHVVQHTPTSVEVTTETYISAPILGWGFFASTSYTISGNGALTVNVHLKPHGPMPADLPRLGL DVLLADELDNTSWFGLGPGEAYPDKKRAQKVGIYNAATAELHTPYEVPQEGGNRMDTRWLRVHD SRGWGLRVTRVKDESDKQPTELFQWLATRYSPEAIEAAKHAPELVPEKRIRLRLDVESCGVGTG ACGPRTLDKYRVKCEERKFGFTLQPVLAELC		
Beta-galactosidase, putative [Aspergillus fumigatus Af293] EAL91164.1 GI:66850838	MTLSAVPDYENQHILQRNRLKPRAYFLPATSISLNGRWDFHYAASPVSAPEPTWSKGTKNATAE PRRDSNQFSSDGADSKTAWAPITVPGHWQLQGYGRPHYTNVIYPFPVCPPFVPTENPTGTYRRT FHVPAEWDASSQLRLRFDGVDSAYHVWVNGVPIGYSQGSRNPAEFDVSQVVDRDGANELFVRVY QWSDGSYIEDQDQWWLSGIFRDVTLLAFPGQARIEDFFVRTALDKDYVDATLRLSVDLALATAA IVQVTLSNPSTGSTLQTEKYSLGEKQDKLEAELSVSNPNKWTAETPNLYNLCIALYVDGAKDPV QTINHRVGFRQVEIKNGNITVNGVPVMFRGVNRHDHHPRFGRAVPLSFLREDLLIMKRHNVNAL RCSHYPSHPRLYELCDELGLWVMDEADLECHGFYDAIARPLDIPESMDYEERKKLTFGQAAQFT TNNPEWKEAYVDRMAQMVQRDKNHSCIVIWSLGNEAFYGSNHQAMYDYVKQVDPSRPVHYEGDM EAKTVDMYSYMYPSLERLVGFATAEGDEFKKPIVLCEYAHAMGNAPGGLEEYMEAFRTHRRLQG GWVWEWANHGLWDEKKGWYGYGGDFGDTPHDGNFVLDGLLFSDHTPTPGITELKKAYAPVRVWP GEDGTLVVANDYNFVGLEGLQASYKIEVLGDSGRIIATGIIELPPIPAGQNGTIKLPSAPATAI PGEVWLTISFLQKGETAWAGNNYEVAWYQQCLKSSSPRFSLAVPAEALTHSSTKTSHRISGASF SLEFSRETGSLYAWTAGGLSLLDQSSSTGAISPGFWRPPTDNDMSHDLLEWRRFGLDTLTSQLR KMHVVQHTPTSVEVTTETYISAPILGWGFFASTSYTISGNGALTVNVHLKPHGPMPADLPRLGL DVLLADELDNTSWFGLGPGEAYPDKKRAQKVGIYNAATAELHTPYEVPQEGGNRMDTRWLRVHD SRGWGLRVTRVKDESDKQPTELFQWLATRYSPEAIEAAKHAPELVPEKRIRLRLDVESCGVGTG ACGPRTLDKYRVKCEERKFGFTLQPVLAELC	720	
Beta-galactosidase [Aspergillus candidus] CAD24293.1 GI:18958133	MKLLSVAAVALLAAQAAGASIKHRLNGFTILEHPDPAKRDLLQDIVTWDDKSLFINGERIMLFS GEVHPFRLPVPSLWLDIFHKIRALGFNCVSFYIDWALLEGKPGDYRAEGIFALEPFFDAAKEAG IYLIARPGSYINAEVSGGGFPGWLQRVNGTLRSSDEPFLKATDNYIANAAAAVAKAQITNGGPV ILYQPENEYSGGCCGVKYPDADYMQYVMDQARKADIVVPFISNDASPSGHNAPGSGTGAVDIYG HDSYPLGFDCANPSVWPEGKLPDNFRTLHLEQSPSTPYSLLEFQAGAFDPWGGPGFEKCYALVN HEFSRVFYRNDLSFGVSTFNLYMTFGGTNWGNLGHPGGYTSYDYGSPITETRNVTREKYSDIKL LANFVKASPSYLTATPRNLTTGVYTDTSDLAVTPLMGDSPGSFFVVRHTDYSSQESTSYKLKLP TSAGNLTIPQLEGTLSLNGRDSKIHVVDYNVSGTNIIYSTAEVFTWKKFDGNKVLVLYGGPKEH HELAIASKSNVTIIEGSDSGIVSTRKGSSVIIGWDVSSTRRIVQVGDLRVFLLDRNSAYNYWVP ELPTEGTSPGFSTSKTTASSIIVKAGYLLRGAHLDGADLHLTADFNATTPIEVIGAPTGAKNLF VNGEKASHTVDKNGIWSSEVKYAAPEIKLPGLKDLDWKYLDTLPEIKSSYDDSAWVSADLPKTK NTHRPLDTPTSLYSSDYGFHTGYLIYRGHFVANGKESEFFIRTQGGSAFGSSVWLNETYLGSWT GADYAMDGNSTYKLSQLESGKNYVITVVIDNLGLDENWTVGEETMKNPRGILSYKLSGQDASAI TWKLTGNLGGEDYQDKVRGPLNEGGLYAERQGFHQPQPPSESWESGSPLEGLSKPGIGFYTAQF DLDLPKGWDVPLYFNFGNNTQAARAQLYVNGYQYGKFTGNVGPQTSFPVPEGILNYRGTNYVAL SLWALESDGAKLGSFELSYTTPVLTGYGDVESPEQPKYEQRKGAY	721	
Beta-galactosidase	MKLSSACAIALLAAQAAGASIKHRINGFTLTEHSDPAKRELLQKYVTWDDKSLFINGERIMIFS	722	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
(Lactase-N; Lactase; Tilactase) P29853.2 GI:461623	GEFHPFRLPVKELQLDIFQKVKALGFNCVSFYVDWALVEGKPGEYRADGIFDLEPFFDAASEAG IYLLARPGPYINAESSGGGFPGWLQRVNGTLRSSDKAYLDATDNYVSHVAATIAKYQITNGGPI ILYQPENEYTSGCSGVEFPDPVYMQYVEDQARNAGVVIPLINNDASASGNNAPGTGKGAVDIYG HDSYPLGFDCANPTVWPSGDLPTNFRTLHLEQSPTTPYAIVEFQGGSYDPWGGPGFAACSELLN NEFERVFYKNDFSFQIAIMNLYMIFGGTNWGNLGYPNGYTSYDYGSAVTESRNITREKYSELKL LGNFAKVSPGYLTASPGNLTTSGYADTTDLTVTPLLGNSTGSFFVVRHSDYSSEESTSYKLRLP TSAGSVTIPQLGGTLTLNGRDSKIHVTDHNVSGTNIIYSTAEVFTWKKFADGKVLVLYGGAGEH HELAISTKSNVTVIEGSESGISSKQTSSSVVVGWDVSTTRRIIQVGDLKILLLDRNSAYNYWVP QLATDGTSPGFSTPEKVASSIIVKAGYLVRTAYLKGSGLYLTADFNATTSVEVIGVPSTAKNLF INGDKTSHTVDKNGIWSATVDYNAPDISLPSLKDLDWKYVDTLPEIQSSYDDSLWPAADLKQTK NTLRSLTTPTSLYSSDYGFHTGYLLYRGHFTATGNESTFAIDTQGGSAFGSSVWLNGTYLGSWT GLYANSDYNATYNLPQLQAGKTYVITVVIDNMGLEENWTVGEDLMKSPRGISTSCLPDGQAAPI SWKLTGNLGGEDYEDKVRGPLNEGGLYAERQGFHQPEPPSQNWKSSSPLEGLSEAGIGFYSASF DLDLPKDGMSHCSSTSVTALRHPRTACRSTSTDIVCEIHKQHRTSDQLPCPRGNPELSRNELVG GDPVALDSAGGKLESLELSYTTPVLTALGEVESVDQPKYKKRKGAY		
Alpha-glucosidase P1 subunit, ANP P1 subunit Aspergillus niger AAB2358.1 GI:257186 (transglucosidase)	SLLAPSQPQFXIPASAAVGAQLIANIDDPQAADAQSVCPGYKASKVQHNSRGFTASLQLAGRPC NVYGTDVESLTLSVEYQDSDRLNIQILPTHVDSTXASWYFLSENLVPRPKASLXASVSQSDLFV SWSNEPSFNFKVIRKATGDALFSTEGTVLVYENQFIEFVTALPEEYNLYGLGEHITQFRLQRNA XLTIYPSDDGTPIDQNLYGQHPFYLDTRYYKGDRQ	723	
Celluclast hypothetical protein M419DRAFT_125268 [T. reesei RUT C-3] ETR97394.1 GI:57227381	MADIDVEAILKKLTLAEKVDLLAGIDFWHTKALPKHGVPSLRFTDGPNGVRGTKFFNGVPAACF PCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESIGEDPFLAGLG AAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYALPFQIAVRDSQPGAFMTA YNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTEAVVAGLDLEMPGPPRFRGETLK FNVSNGKPFIHVIDQRAREVLQFVKKCAASGVTENGPETTVNNTPETAALLRKVGNEGIVLLKN ENNVLPLSKKKKTLIVGPNAKQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTHRF LPILGEQCLTPDGAPGMRWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEG TYTADEDCTYELGLVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKF KIEFGSAPTYTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETE GADRASMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYGGNETGNSIAD VVFGDYNPSGKLSLSFPKRLQDNPAFLNFRTEAGRTLYGEDVYVGYRYYEFADKDVNFPFGHGL SYTTFAFSNLSVSHKDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQAAKINRPVKELKGFAKVEL QPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVSDSSAAKDGVALRGKFTVGETYWWS GV	724	
Velvet complex subunit 2	MPSLIPPIVSASSASNSAALDHLYHHQPPPRLPLGAVPQSPIQSQAPPPPHLHPPSHHFQLHPG HGHHQQPHHERDHRLPPPVASYSAHSHHLQHDPLPQRLESSQPGHPGAAEHRDHPQHALDEPSR	725	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
GRS98.2 GI:1881915	SHDPYPSMATGALVHSESQQPASASLLLPISNVEEATGRRYHLDVVQQPRRARMCGFGDKDRRP ITPPPCVRLIIIDVATGKEIDCNDIDHSMFVLNVDLWNEDGTREVNLVRSSTSSSPSVSSTVTY PYGSISVGESSHTYGQSAHPPSREAPYSVSQTASYAPEYQTQPTYSQGSSAYPSNGTYGPPQQY FPQHQAYRTETGPPGAMQTTVGGFRGYAQDQNALTKMAVVGGQPQGMFTRNLIGSLAASAFRLA DTSEHLGIWFVLQDLSVRTEGPFRLRFSFVNVGPLAGQNGAKVNTGRAPILASCFSEVFNVYSA KKFPGVCESTPLSKTFAAQGIKIPIRKDANLKGGDGEDDYGD		
alpha-L-arabinofuranosidase [T. reesei] CAA93243.1 GI:158814	MLSNARIIAAGCIAAGSLVAAGPCDIYSSGGTPCVAAHSTTRALFSAYTGPLYQVKRGSDGATT AISPLSSGVANAAAQDAFCAGTTCLITIIYDQSGRGNHLTQAPPGGFSGPESNGYDNLASAIGA PVTLNGQKAYGVFVSPGTGYRNNAASGTAKGDAAEGMYAVLDGTHYNGACCFDYGNAETNSRDT GNGHMEAIYFGDSTVWGTGSGKGPWIMADLENGLFSGSSPGNNAGDPSISYRFVTAAIKGQPNQ WAIRGGNAASGSLSTFYSGARPQVSGYNPMSKEGAIILGIGGDNSNGAQGTFYEGVMTSGYPSD ATENSVQANIVAARYAVAPLTSGPALTVGSSISLRATTACCTTRYIAHSGSTVNTQVVSSSSAT ALKQQASWTVRAGLANNACFSFESRDTSGSYIRHSNFGLVLNANDGSKLFAEDATFCTQAGING QGSSIRSWSYPTRYFRHYNNTLYIASNGGVHVFDATAAFNDDVSFVVSGGFA	726	
Beta-xylosidase [Trichoderma reesei] CAA93248.1 GI:2791278	MVNNAALLAALSALLPTALAQNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCD SSAGYVERAQALISLFTLEELILNTQNSGPGVPRLGLPNYQVWNEALHGLDRANFATKGGQFEW ATSFPMPILTTAALNRTLIHQIADIISTQARAFSNSGRYGLDVYAPNVNGFRSPLWGRGQETPG EDAFFLSSAYTYEYITGIQGGVDPEHLKVAATVKHFAGYDLENWNNQSRLGFDAIITQQDLSEY YTPQFLAAARYAKSRSLMCAYNSVNGVPSCANSFFLQTLLRESWGFPEWGYVSSDCDAVYNVFN PHDYASNQSSAAASSLRAGTDIDCGQTYPWHLNESFVAGEVSRGEIERSVTRLYANLVRLGYFD KKNQYRSLGWKDVVKTDAWNISYEAAVEGIVLLKNDGTLPLSKKVRSIALIGPWANATTQMQGN YYGPAPYLISPLEAAKKAGYHVNFELGTEIAGNSTTGFAKAIAAAKKSDAIIYLGGIDNTIEQE GADRTDIAWPGNQLDLIKQLSEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVA LFDILSGKRAPAGRLVTTQYPAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFY TTFKETLASHPKSLKFNTSSILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSN AGPAPYPNKWLVGFDRLADIKPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVK LEFELVGEEVTIENWPLEEQQIKDATPDA	727	
Unnamed protein product [T. reesei] CAW52645.1 GI:219752323	MVNNAALLAALSALLPTALAQNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCD SSAGYVERAQALISLFTLEELILNTQNSGPGVPRLGLPNYQVWNEALHGLDRANFATKGGQFEW ATSFPMPILTTAALNRTLIHQIADIISTQARAFSNSGRYGLDVYAPNVNGFRSPLWGRGQETPG EDAFFLSSAYTYEYITGIQGGVDPEHLKVAATVKHFAGYDLENWNNQSRLGFDAIITQQDLSEY YTPQFLAAARYAKSRSLMCAYNSVNGVPSCANSFFLQTLLRESWGFPEWGYVSSDCDAVYNVFN PHDYASNQSSAAASSLRAGTDIDCGQTYPWHLNESFVAGEVSRGEIERSVTRLYANLVRLGYFD KKNQYRSLGWKDVVKTDAWNISYEAAVEGIVLLKNDGTLPLSKKVRSIALIGPWANATTQMQGN YYGPAPYLISPLEAAKKAGYHVNFELGTEIAGNSTTGFAKAIAAAKKSDAIIYLGGIDNTIEQE GADRTDIAWPGNQLDLIKQLSEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVA LFDILSGKRAPAGRLVTTQYPAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFY TTFKETLASHPKSLKFNTSSILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSN	728	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AGPAPYPNKWLVGFDRLADIKPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVK LEFELVGEEVTIENWPLEEQQIKDATPDA		
Unnamed protein product [T. reesei] CBC2392.1 GI:257341433	MVNNAALLAALSALLPTALAQNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCD SSAGYVERAQALISLFTLEELILNTQNSGPGVPRLGLPNYQVWNEALHGLDRANFATKGGQFEW ATSFPMPILTTAALNRTLIHQIADIISTQARAFSNSGRYGLDVYAPNVNGFRSPLWGRGQETPG EDAFFLSSAYTYEYITGIQGGVDPEHLKVAATVKHFAGYDLENWNNQSRLGFDAIITQQDLSEY YTPQFLAAARYAKSRSLMCAYNSVNGVPSCANSFFLQTLLRESWGFPEWGYVSSDCDAVYNVFN PHDYASNQSSAAASSLRAGTDIDCGQTYPWHLNESFVAGEVSRGEIERSVTRLYANLVRLGYFD KKNQYRSLGWKDVVKTDAWNISYEAAVEGIVLLKNDGTLPLSKKVRSIALIGPWANATTQMQGN YYGPAPYLISPLEAAKKAGYHVNFELGTEIAGNSTTGFAKAIAAAKKSDAIIYLGGIDNTIEQE GADRTDIAWPGNQLDLIKQLSEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVA LFDILSGKRAPAGRLVTTQYPAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFY TTFKETLASHPKSLKFNTSSILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSN AGPAPYPNKWLVGFDRLADIKPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVK LEFELVGEEVTIENWPLEEQQIKDATPDA	729	
Chain A, The Structure Of Hypocrea Jecorina Beta-xylosidase Xyl3a (bxl1) 5A7M_A GI:152244671	XNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCDSSAGYVERAQALISLFTLEE LILNTQNSGPGVPRLGLPNYQVWNEALHGLDRANFATKGGQFEWATSFPMPILTTAALNRTLIH QIADIISTQARAFSNSGRYGLDVYAPNVNGFRSPLWGRGQETPGEDAFFLSSAYTYEYITGIQG GVDPEHLKVAATVKHFAGYDLENWNNQSRLGFDAIITQQDLSEYYTPQFLAAARYAKSRSLMCA YNSVNGVPSCANSFFLQTLLRESWGFPEWGYVSSDCDAVYNVFNPHDYASNQSSAAASSLRAGT DIDCGQTYPWHLNESFVAGEVSRGEIERSVTRLYANLVRLGYFDKKNQYRSLGWKDVVKTDAWN ISYEAAVEGIVLLKNDGTLPLSKKVRSIALIGPWANATTQMQGNYYGPAPYLISPLEAAKKAGY HVNFELGTEIAGNSTTGFAKAIAAAKKSDAIIYLGGIDNTIEQEGADRTDIAWPGNQLDLIKQL SEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVALFDILSGKRAPAGRLVTTQY PAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFYTTFKETLASHPKSLKFNTSS ILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSNAGPAPYPNKWLVGFDRLADI KPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVKLEFELVGEEVTIENWPLE	73	
Chain B, The Structure Of Hypocrea Jecorina Beta-xylosidase Xyl3a (bxl1) 5A7M_B GI:152244672	XNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCDSSAGYVERAQALISLFTLEE LILNTQNSGPGVPRLGLPNYQVWNEALHGLDRANFATKGGQFEWATSFPMPILTTAALNRTLIH QIADIISTQARAFSNSGRYGLDVYAPNVNGFRSPLWGRGQETPGEDAFFLSSAYTYEYITGIQG GVDPEHLKVAATVKHFAGYDLENWNNQSRLGFDAIITQQDLSEYYTPQFLAAARYAKSRSLMCA YNSVNGVPSCANSFFLQTLLRESWGFPEWGYVSSDCDAVYNVFNPHDYASNQSSAAASSLRAGT DIDCGQTYPWHLNESFVAGEVSRGEIERSVTRLYANLVRLGYFDKKNQYRSLGWKDVVKTDAWN ISYEAAVEGIVLLKNDGTLPLSKKVRSIALIGPWANATTQMQGNYYGPAPYLISPLEAAKKAGY HVNFELGTEIAGNSTTGFAKAIAAAKKSDAIIYLGGIDNTIEQEGADRTDIAWPGNQLDLIKQL SEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVALFDILSGKRAPAGRLVTTQY PAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFYTTFKETLASHPKSLKFNTSS ILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSNAGPAPYPNKWLVGFDRLADI	731	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	KPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVKLEFELVGEEVTIENWPLE		
Chain A, The Structure Of Hypocrea Jecorina Beta-xylosidase Xyl3a (Bxl1) In Complex With 4-thioxylobiose 5AE6_A GI:169428461	XNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCDSSAGYVERAQALISLFTLEE LILNTQNSGPGVPRLGLPNYQVWNEALHGLDRANFATKGGQFEWATSFPMPILTTAALNRTLIH QIADIISTQARAFSNSGRYGLDVYAPNVNGFRSPLWGRGQETPGEDAFFLSSAYTYEYITGIQG GVDPEHLKVAATVKHFAGYDLENWNNQSRLGFDAIITQQDLSEYYTPQFLAAARYAKSRSLMCA YNSVNGVPSCANSFFLQTLLRESWGFPEWGYVSSDCDAVYNVFNPHDYASNQSSAAASSLRAGT DIDCGQTYPWHLNESFVAGEVSRGEIERSVTRLYANLVRLGYFDKKNQYRSLGWKDVVKTDAWN ISYEAAVEGIVLLKNDGTLPLSKKVRSIALIGPWANATTQMQGNYYGPAPYLISPLEAAKKAGY HVNFELGTEIAGNSTTGFAKAIAAAKKSDAIIYLGGIDNTIEQEGADRTDIAWPGNQLDLIKQL SEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVALFDILSGKRAPAGRLVTTQY PAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFYTTFKETLASHPKSLKFNTSS ILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSNAGPAPYPNKWLVGFDRLADI KPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVKLEFELVGEEVTIENWPLEE	732	
Chain B, The Structure Of Hypocrea Jecorina Beta-xylosidase Xyl3a (Bxl1) In Complex With 4-thioxylobiose 5AE6_B GI:169428462	XNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCDSSAGYVERAQALISLFTLEE LILNTQNSGPGVPRLGLPNYQVWNEAALNRTLIHQIADIISTQARAFSNSGRYGLDVYAPNVNG FRSPLWGRGQETPGEDAFFLSSAYTYEYITGIQGGVDPEHLKVAATVKHFAGYDLENWNNQSRL GFDAIITQQDLSEYYTPQFLAAARYAKSRSLMCAYNSVNGVPSCANSFFLQTLLRESWGFPEWG YVSSDCDAVYNVFNPHDYASNQSSAAASSLRAGTDIDCGQTYPWHLNESFVAGEVSRGEIERSV TRLYANLVRLGYFDKKNQYRSLGWKDVVKTDAWNISYEAAVEGIVLLKNDGTLPLSKKVRSIAL IGPWANATTQMQGNYYGPAPYLISPLEAAKKAGYHVNFELGTEIAGNSTTGFAKAIAAAKKSDA IIYLGGIDNTIEQEGADRTDIAWPGNQLDLIKQLSEVGKPLVVLQMGGGQVDSSSLKSNKKVNS LVWGGYPGQSGGVALFDILSGKRAPAGRLVTTQYPAEYVHQFPQNDMNLRPDGKSNPGQTYIWY TGKPVYEFGSGLFYTTFKETLASHPKSLKFNTSSILSAPHPGYTYSEQIPVFTFEANIKNSGKT ESPYTAMLFVRTSNAGPAPYPNKWLVGFDRLADIKPGHSSKLSIPIPVSALARVDSHGNRIVYP GKYELALNTDESVKLEFELVGEEVTIENWPLEE	733	
Glycoside hydrolase family 43 [Trichoderma reesei QM6a] EGR49145.1 GI:3451895	MPLIRNPILPGFNADPSIVRVGSDYYIATSTFEWYPGVQIHHSTDLANWELAVRPLSRRSQLDL RGEPDSCGVWAPCLTHDGDKFWLVYTDVKRKDGSFKDTHNYIVTAPRIEGPWSDPVYANSSGFD PSLFHDDDGRKWLVNMVQDHRARPRTFAGIALQEFDPAQGKLVGTRKVVFHGSELGLVEGPHLY KRNGWYYLLTAEGGTGYTHAATLARSRSIWGPYELHPQQHILTSKDHPFAALQRAGHADIVETA DGKTYLVHLAGRPIGQKRRCVLGRETALQEAYWGEDDWLYVKNGPVPSLDVEVPGVRDEEAYWK EKRYEFHDGLHKDFQWLRTPEPERLFAIEDGKLVLTGRESIGSWFEQSLVARRQTHFSFDAETV IDFSPEDERQFAGLTLYYSRYNFFYLAVSAHSDGRREVQILRSEASWPNGKLEDVGANACYVRI PQQGRVKLAATIRGERLQFYYALVAEGGEEQEELQRIGPVLDASIVSDECGGHQAHGSFTGSFV GVACSDVNGTEKRAVFDYFVYRPAHDSTDRYSVSMEGIQRV	734	
Glycoside hydrolase family 3 [T. reesei QM6a]	MVNNAALLAALSALLPTALAQNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCD SSAGYVERAQALISLFTLEELILNTQNSGPGVPRLGLPNYQVWNEALHGLDRANFATKGGQFEW ATSFPMPILTTAALNRTLIHQIADIISTQARAFSNSGRYGLDVYAPNVNGFRSPLWGRGQETPG EDAFFLSSAYTYEYITGIQGGVDPEHLKVAATVKHFAGYDLENWNNQSRLGFDAIITQQDLSEY	735	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
EGR4972.1 GI:34519464	YTPQFLAAARYAKSRSLMCAYNSVNGVPSCANSFFLQTLLRESWGFPEWGYVSSDCDAVYNVFN PHDYASNQSSAAASSLRAGTDIDCGQTYPWHLNESFVAGEVSRGEIERSVTRLYANLVRLGYFD KKNQYRSLGWKDVVKTDAWNISYEAAVEGIVLLKNDGTLPLSKKVRSIALIGPWANATTQMQGN YYGPAPYLISPLEAAKKAGYHVNFELGTEIAGNSTTGFAKAIAAAKKSDAIIYLGGIDNTIEQE GADRTDIAWPGNQLDLIKQLSEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVA LFDILSGKRAPAGRLVTTQYPAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFY TTFKETLASHPKSLKFNTSSILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSN AGPAPYPNKWLVGFDRLADIKPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVK LEFELVGEEVTIENWPLEEQQIKDATPDA		
Glycoside hydrolase family 3 [T. reesei QM6a] EGR586.1 GI:34519849	MRVNVPLHALQIAARSVAAAICKSSSASGRSLRGGKIDQASRINIYSISNPSPNPPLTPSFPDC TRDPLCSNDVCDTTKSIAERAAAIVKPMTLNEKVANVGSSASGSARLGLPAYQWQNEALHGVAG STGVQFQSPLGANFSAATSFPMPILLSAAFDDALVKSVATAISTEARAFANYGFAGLDFWTPNI NPFRDPRWGRGMETPGEDAFRIQGYVLALVDGLQGGIDPDFYRTLSTCKHFAAYDIENGRTANN LSPTQQDMADYYLPMFETCVRDAKVASIMCAYNAVDGVPACADSYLLQDVLRDTYGFTEDFNYV VSDCDAVENVFDPHHYAANLTQAAAMSINAGTDLDCGSSYNVLNASVQAGLTTEATLDKSLIRL YSALVKVGYFDQPAEYNSLGWGNVNTTQSQALAHDAATEGMTLLKNDGTLPLSRTLSNVAVIGP WANVTTQMQGNYAGTAPLLVNPLSVFQQKWRNVKYAQGTAINSQDTSGFNAALSAASSSDVIVY LGGIDISVENEGFDRSSITWPGNQLNLISQLANLGKPLVIVQFGGGQIDDSALLSNSKVNSILW AGYPGQDGGNAIFDVLTGANPPAGRLPVTQYPANYVNNNNIQDMNLRPSNGIPGRTYAWYTGTP VLPFGYGLHYTNFSLSFQSTKTAGSDIATLVNNAGSNKDLATFATIVVNVKNTGGKANLASDYV GLLFLKSTNAGPAPHPNKQLAAYGRVRNVGVGATQQLTLTVNLGSLARADTNGDRWIYPGAYTL ILDVNGPLTFNFTLTGTATKISTLPSRS	736	
Glycoside hydrolase family 3 [T. reesei QM6a] XP_6963621.1 GI:58913197	MRVNVPLHALQIAARSVAAAICKSSSASGRSLRGGKIDQASRINIYSISNPSPNPPLTPSFPDC TRDPLCSNDVCDTTKSIAERAAAIVKPMTLNEKVANVGSSASGSARLGLPAYQWQNEALHGVAG STGVQFQSPLGANFSAATSFPMPILLSAAFDDALVKSVATAISTEARAFANYGFAGLDFWTPNI NPFRDPRWGRGMETPGEDAFRIQGYVLALVDGLQGGIDPDFYRTLSTCKHFAAYDIENGRTANN LSPTQQDMADYYLPMFETCVRDAKVASIMCAYNAVDGVPACADSYLLQDVLRDTYGFTEDFNYV VSDCDAVENVFDPHHYAANLTQAAAMSINAGTDLDCGSSYNVLNASVQAGLTTEATLDKSLIRL YSALVKVGYFDQPAEYNSLGWGNVNTTQSQALAHDAATEGMTLLKNDGTLPLSRTLSNVAVIGP WANVTTQMQGNYAGTAPLLVNPLSVFQQKWRNVKYAQGTAINSQDTSGFNAALSAASSSDVIVY LGGIDISVENEGFDRSSITWPGNQLNLISQLANLGKPLVIVQFGGGQIDDSALLSNSKVNSILW AGYPGQDGGNAIFDVLTGANPPAGRLPVTQYPANYVNNNNIQDMNLRPSNGIPGRTYAWYTGTP VLPFGYGLHYTNFSLSFQSTKTAGSDIATLVNNAGSNKDLATFATIVVNVKNTGGKANLASDYV GLLFLKSTNAGPAPHPNKQLAAYGRVRNVGVGATQQLTLTVNLGSLARADTNGDRWIYPGAYTL ILDVNGPLTFNFTLTGTATKISTLPSRS	737	
glycoside hydrolase family 3 [T. reesei QM6a]	MVNNAALLAALSALLPTALAQNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCD SSAGYVERAQALISLFTLEELILNTQNSGPGVPRLGLPNYQVWNEALHGLDRANFATKGGQFEW ATSFPMPILTTAALNRTLIHQIADIISTQARAFSNSGRYGLDVYAPNVNGFRSPLWGRGQETPG	738	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
XP_696475.1 GI:5891415	EDAFFLSSAYTYEYITGIQGGVDPEHLKVAATVKHFAGYDLENWNNQSRLGFDAIITQQDLSEY YTPQFLAAARYAKSRSLMCAYNSVNGVPSCANSFFLQTLLRESWGFPEWGYVSSDCDAVYNVFN PHDYASNQSSAAASSLRAGTDIDCGQTYPWHLNESFVAGEVSRGEIERSVTRLYANLVRLGYFD KKNQYRSLGWKDVVKTDAWNISYEAAVEGIVLLKNDGTLPLSKKVRSIALIGPWANATTQMQGN YYGPAPYLISPLEAAKKAGYHVNFELGTEIAGNSTTGFAKAIAAAKKSDAIIYLGGIDNTIEQE GADRTDIAWPGNQLDLIKQLSEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVA LFDILSGKRAPAGRLVTTQYPAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFY TTFKETLASHPKSLKFNTSSILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSN AGPAPYPNKWLVGFDRLADIKPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVK LEFELVGEEVTIENWPLEEQQIKDATPDA		
Glycoside hydrolase family 43 [T. reesei QM6a] XP_6964816.1 GI:58915587	MPLIRNPILPGFNADPSIVRVGSDYYIATSTFEWYPGVQIHHSTDLANWELAVRPLSRRSQLDL RGEPDSCGVWAPCLTHDGDKFWLVYTDVKRKDGSFKDTHNYIVTAPRIEGPWSDPVYANSSGFD PSLFHDDDGRKWLVNMVQDHRARPRTFAGIALQEFDPAQGKLVGTRKVVFHGSELGLVEGPHLY KRNGWYYLLTAEGGTGYTHAATLARSRSIWGPYELHPQQHILTSKDHPFAALQRAGHADIVETA DGKTYLVHLAGRPIGQKRRCVLGRETALQEAYWGEDDWLYVKNGPVPSLDVEVPGVRDEEAYWK EKRYEFHDGLHKDFQWLRTPEPERLFAIEDGKLVLTGRESIGSWFEQSLVARRQTHFSFDAETV IDFSPEDERQFAGLTLYYSRYNFFYLAVSAHSDGRREVQILRSEASWPNGKLEDVGANACYVRI PQQGRVKLAATIRGERLQFYYALVAEGGEEQEELQRIGPVLDASIVSDECGGHQAHGSFTGSFV GVACSDVNGTEKRAVFDYFVYRPAHDSTDRYSVSMEGIQRV	739	
Family 43 glycoside hydrolase [T. reesei RUT C-3] ETS2497.1 GI:572279375	MPLIRNPILPGFNADPSIVRVGSDYYIATSTFEWYPGVQIHHSTDLANWELAVRPLSRRSQLDL RGEPDSCGVWAPCLTHDGDKFWLVYTDVKRKDGSFKDTHNYIVTAPRIEGPWSDPVYANSSGFD PSLFHDDDGRKWLVNMVQDHRARPRTFAGIALQEFDPAQGKLVGTRKVVFHGSELGLVEGPHLY KRNGWYYLLTAEGGTGYTHAATLARSRSIWGPYELHPQQHILTSKDHPFAALQRAGHADIVETA DGKTYLVHLAGRPIGQKRRCVLGRETALQEAYWGEDDWLYVKNGPVPSLDVEVPGVRDEEAYWK EKRYEFHDGLHKDFQWLRTPEPERLFAIEDGKLVLTGRESIGSWFEQSLVARRQTHFSFDAETV IDFSPEDERQFAGLTLYYSRYNFFYLAVSAHSDGRREVQILRSEASWPNGKLEDVGANACYVRI PQQGRVKLAATIRGERLQFYYALVAEGGEEQEELQRIGPVLDASIVSDECGGHQAHGSFTGSFV GVACSDVNGTEKRAVFDYFVYRPAHDSTDRYSVSMEGIQRV	740	
Beta-xylosidase [T. reesei RUT C-3] ETS3193.1 GI:5722896	MVNNAALLAALSALLPTALAQNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKNNLVCD SSAGYVERAQALISLFTLEELILNTQNSGPGVPRLGLPNYQVWNEALHGLDRANFATKGGQFEW ATSFPMPILTTAALNRTLIHQIADIISTQARAFSNSGRYGLDVYAPNVNGFRSPLWGRGQETPG EDAFFLSSAYTYEYITGIQGGVDPEHLKVAATVKHFAGYDLENWNNQSRLGFDAIITQQDLSEY YTPQFLAAARYAKSRSLMCAYNSVNGVPSCANSFFLQTLLRESWGFPEWGYVSSDCDAVYNVFN PHDYASNQSSAAASSLRAGTDIDCGQTYPWHLNESFVAGEVSRGEIERSVTRLYANLVRLGYFD KKNQYRSLGWKDVVKTDAWNISYEAAVEGIVLLKNDGTLPLSKKVRSIALIGPWANATTQMQGN YYGPAPYLISPLEAAKKAGYHVNFELGTEIAGNSTTGFAKAIAAAKKSDAIIYLGGIDNTIEQE GADRTDIAWPGNQLDLIKQLSEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVA LFDILSGKRAPAGRLVTTQYPAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFY	741	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TTFKETLASHPKSLKFNTSSILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSN AGPAPYPNKWLVGFDRLADIKPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVK LEFELVGEEVTIENWPLEEQQIKDATPDA		
Glycoside hydrolase [T. reesei RUT C-3] ETS3636.1 GI:57228576	MALFHLAQARTCLPPYQAQTTYQGCYHDPNSPRDLAGPMLTVGNLNSPQYCANICGAAGYQYSG VEFTIQCFCGHRIESTSVKADESQCSSPCPADSSKVCGGGNMINIYSISNPSPNPPLTPSFPDC TRDPLCSNDVCDTTKSIAERAAAIVKPMTLNEKVANVGSSASGSARLGLPAYQWQNEALHGVAG STGVQFQSPLGANFSAATSFPMPILLSAAFDDALVKSVATAISTEARAFANYGFAGLDFWTPNI NPFRDPRWGRGMETPGEDAFRIQGYVLALVDGLQGGIDPDFYRTLSTCKHFAAYDIENGRTANN LSPTQQDMADYYLPMFETCVRDAKVASIMCAYNAVDGVPACADSYLLQDVLRDTYGFTEDFNYV VSDCDAVENVFDPHHYAANLTQAAAMSINAGTDLDCGSSYNVLNASVQAGLTTEATLDKSLIRL YSALVKVGYFDQPAEYNSLGWGNVNTTQSQALAHDAATEGMTLLKNDGTLPLSRTLSNVAVIGP WANVTTQMQGNYAGTAPLLVNPLSVFQQKWRNVKYAQGTAINSQDTSGFNAALSAASSSDVIVY LGGIDISVENEGFDRSSITWPGNQLNLISQLANLGKPLVIVQFGGGQIDDSALLSNSKVNSILW AGYPGQDGGNAIFDVLTGANPPAGRLPVTQYPANYVNNNNIQDMNLRPSNGIPGRTYAWYTGTP VLPFGYGLHYTNFSLSFQSTKTAGSDIATLVNNAGSNKDLATFATIVVNVKNTGGKANLASDYV GLLFLKSTNAGPAPHPNKQLAAYGRVRNVGVGATQQLTLTVNLGSLARADTNGDRWIYPGAYTL ILDVNGPLTFNFTLTGTATKISTLPSRS	742	
Chain B, The Three-Dimensional Crystal Structure Of The Catalytic Core Of Cellobiohydrolase I From T. Reesei 1CEL_B GI:89287	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	743	
Chain A, The Three-Dimensional Crystal Structure Of The Catalytic Core Of Cellobiohydrolase I From T. Reesei 1CEL_A GI:89286	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	744	
Chain A, Three-Dimensional Structure Of Cellobiohydrolase From T. Reesei	SGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLME QTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPA NQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLA NHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDG	745	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
3CBH_A GI:157836775	TSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL		
Chain A, Determination Of The Three-Dimensional Structure Of The C- Terminal Domain Of Cellobiohydrolase I From T. Reesei. 2CBH_A GI:157834734	TQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL	746	
Chain A, Three-Dimensional Structures Of Three Engineered Cellulose-Binding Domains Of Cellobiohydrolase I From T. Reesei, Nmr, 19 Structures 1AZK_A GI:15783159	TQSHYGQCGGIGYSGPTVCASGTTCQVLNPYASQCL	747	
Chain A, Three-Dimensional Structures Of Three Engineered Cellulose-Binding Domains Of Cellobiohydrolase I From T. Reesei, Nmr, 18 Structures 1AZJ_A GI:15783158	TQSHYGQCGGIGYSGPTVCASGTTCQVLNPAYSQCL	748	
Chain A, Three-Dimensional Structures Of Three Engineered	TQSHAGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL	749	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Cellulose-Binding Domains Of Cellobiohydrolase I From T. Reesei, Nmr, 14 Structures 1AZH_A GI:15783156			
Chain A, Three-Dimensional Structures Of Three Engineered Cellulose- Binding Domains Of Cellobiohydrolase I From T. Reesei, Nmr, 23 Structures 1AZ6_A GI:15783153	TQSHAGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL	750	
Cellobiohydrolase II [T. reesei] AAA3421.1 GI:17541	MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPG AASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYY ASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVV YDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGT PKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTG QQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPA PQAGAWFQAYFVQLLTNANPSFL	751	
Cellobiohydrolase II core protein, CBH II cp=3.2.1.91 [T. reesei, Peptide Partial, 38 aa] AAB3868.1 GI:5528	GSASYXGNPFVGVSPWANAYYAXEVXXLAIPXLTGAMA	752	
Chain A, Determination Of The Three-Dimensional Structure Of The C- Terminal Domain Of	TQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL	753	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Cellobiohydrolase I From T. Reesei. 1CBH_A GI:15783535			
cellobiohydrolase II [T. reesei] AAG3998.1 GI:11692747	MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPG AASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYY ASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVV YDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGT PKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGRLLANHGWSNAFFITDQGRSGKQPTG QQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPA PQAGAWFQAYFVQLLTNANPSFL	754	
Chain A, Cellobiohydrolase Cel7a (E223s, A224h, L225v, T226a, D262g) Mutant 1EGN_A GI:14277711	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISSHVAPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCGWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	755	
Chain B, Cellobiohydrolase Cel7a With Loop Deletion 245-252 And Bound Non-Hydrolysable Cellotetraose 1Q2E_B GI:39654596	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGGTCDPDGCDWN PYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGSYSGNELND DYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGA VRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	756	
Chain A, Cellobiohydrolase Cel7a With Loop Deletion 245-252 And Bound Non-Hydrolysable Cellotetraose 1Q2E_A GI:39654595	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGGTCDPDGCDWN PYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGSYSGNELND DYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGA VRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	757	
Chain A, Cellobiohydrolase	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS	758	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Cel7a With Disulphide Bridge Added Across Exo-Loop By Mutations D241c And D249c 1Q2B_A GI:39654594	FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGCGCGGTYSCNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG		
Glycoside hydrolase family 5 [T. reesei QM6a] XP_6969897.1 GI:589115749	MRATSLLAAALAVAGDALAGKIKYLGVAIPGIDFGCDIDGSCPTDTSSVPLLSYKGGDGAGQMK HFAEDDGLNVFRISATWQFVLNNTVDGKLDELNWGSYNKVVNACLETGAYCMIDMHNFARYNGG IIGQGGVSDDIFVDLWVQIAKYYEDNDKIIFGLMNEPHDLDIEIWAQTCQKVVTAIRKAGATSQ MILLPGTNFASVETYVSTGSAEALGKITNPDGSTDLLYFDVHKYLDINNSGSHAECTTDNVDAF NDFADWLRQNKRQAIISETGASMEPSCMTAFCAQNKAISENSDVYIGFVGWGAGSFDTSYILTL TPLGKPGNYTDNKLMNECILDQFTLDEKYRPTPTSISTAAEETATATATSDGDAPSTTKPIFRE ETASPTPNAVTKPSPDTSDSSDDDKDSAASMSAQGLTGTVLFTVAALGYMLVAF	759	
Glycoside hydrolase family 5 [T. reesei QM6a] EGR512.1 GI:3452785	MNKSVAPLLLAASILYGGAAAQQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCIPGATTIT TSTRPPSGPTTTTRATSTSSSTPPTSSGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGS NNYPDGIGQMQHFVNDDGMTIFRLPVGWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIV DIHNYARWNGGIIGQGGPTNAQFTSLWSQLASKYASQSRVWFGIMNEPHDVNINTWAATVQEVV TAIRNAGATSQFISLPGNDWQSAGAFISDGSAAALSQVTNPDGSTTNLIFDVHKYLDSDNSGTH AECTTNNIDGAFSPLATWLRQNNRQAILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWG AGSFDSTYVLTETPTGSGNSWTDTSLVSSCLARK	760	
Glycoside hydrolase family 61 [T. reesei QM6a] EGR52697.1 GI:34522464	MIQKLSNLLVTALAVATGVVGHGHINDIVINGVWYQAYDPTTFPYESNPPIVVGWTAADLDNGF VSPDAYQNPDIICHKNATNAKGHASVKAGDTILFQWVPVPWPHPGPIVDYLANCNGDCETVDKT TLEFFKIDGVGLLSGGDPGTWASDVLISNNNTWVVKIPDNLAPGNYVLRHEIIALHSAGQANGA QNYPQCFNIAVSGSGSLQPSGVLGTDLYHATDPGVLINIYTSPLNYIIPGPTVVSGLPTSVAQG SSAATATASATVPGGGSGPTSRTTTTARTTQASSRPSSTPPATTSAPAGGPTQTLYGQCGGSGY SGPTRCAPPATCSTLNPYYAQCLN	761	
Endo-1,4-beta-glucanase V [T. reesei RUT C-3] ETR998.1 GI:572276454	MKATLVLGSLIVGAVSAYKATTTRYYDGQEGACGCGSSSGAFPWQLGIGNGVYTAAGSQALFDT AGASWCGAGCGKCYQLTSTGQAPCSSCGTGGAAGQSIIVMVTNLCPNNGNAQWCPVVGGTNQYG YSYHFDIMAQNEIFGDNVVVDFEPIACPGQAASDWGTCLCVGQQETDPTPVLGNDTGSTPPGSS PPATSSSPPSGGGQQTLYGQCGGAGWTGPTTCQAPGTCKVQNQWYSQCLP	762	
Chain A, The Structure Of A Glycoside Hydrolase Family 61 Member, Cel61b From The Hypocrea	MKSCAILAALGCLAGSVLGHGQVQNFTINGQYNQGFILDYYYQKQNTGHFPNVAGWYAEDLDLG FISPDQYTTPDIVCHKNAAPGAISATAAAGSNIVFQWGPGVWPHPYGPIVTYVVECSGSCTTVN KNNLRWVKIQEAGINYNTQVWAQQDLINQGNKWTVKIPSSLRPGNYVFRHELLAAHGASSANGM QNYPQCVNIAVTGSGTKALPAGTPATQLYKPTDPGILFNPYTTITSYTIPGPALWQG	763	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Jecorina. 2VTC_A GI:19844312			
Chain B, The Structure Of A Glycoside Hydrolase Family 61 Member, Cel61b From The Hypocrea Jecorina. 2VTC_B GI:198443121	MKSCAILAALGCLAGSVLGHGQVQNFTINGQYNQGFILDYYYQKQNTGHFPNVAGWYAEDLDLG FISPDQYTTPDIVCHKNAAPGAISATAAAGSNIVFQWGPGVWPHPYGPIVTYVVECSGSCTTVN KNNLRWVKIQEAGINYNTQVWAQQDLINQGNKWTVKIPSSLRPGNYVFRHELLAAHGASSANGM QNYPQCVNIAVTGSGTKALPAGTPATQLYKPTDPGILFNPYTTITSYTIPGPALWQG	764	
Endoglucanase VIII [T. reesei RUT C-3] ETR9685.1 GI:572273122	MRATSLLAAALAVAGDALAGKIKYLGVAIPGIDFGCDIDGSCPTDTSSVPLLSYKGGDGAGQMK HFAEDDGLNVFRISATWQFVLNNTVDGKLDELNWGSYNKVVNACLETGAYCMIDMHNFARYNGG IIGQGGVSDDIFVDLWVQIAKYYEDNDKIIFGLMNEPHDLDIEIWAQTCQKVVTAIRKAGATSQ MILLPGTNFASVETYVSTGSAEALGKITNPDGSTDLLYFDVHKYLDINNSGSHAECTTDNVDAF NDFADWLRQNKRQAIISETGASMEPSCMTAFCAQNKAISENSDVYIGFVGWGAGSFDTSYILTL TPLGKPGNYTDNKLMNECILDQFTLDEKYRPTPTSISTAAEETATATATSDGDAPSTTKPIFRE ETASPTPNAVTKPSPDTSDSSDDDKDSAASMSAQGLTGTVLFTVAALGYMLVAF	765	
Putative endoglucanase [T. reesei RUT C-3] ETS3449.1 GI:57228352	MKLWIGLLLLGLACRASAHTTFTTLFIDKKNQGDGTCVRMPYDDKTATNPVKPITSSDMACGRN GGDPVPFICSAKKGSLLTFEFRLWPDAQQPGSIDPGHLGPCAVYLKKVDNMFSDSAAGGGWFKI WEDGYDSKTQKWCVDRLVKNNGLLSVRLPRGLPAGYYIVRPEILALHWAAHRDDPQFYLGCAQI FVDSDVRGPLEIPRRQQATIPGYVNAKTPGLTFDIYQDKLPPYPMPGPKVYIPPAKGNKPNQDL NAGRLVQTDGLIPKDCLIKKANWCGRPVEPYSSARMCWRAVNDCYAQSKKCRESSPPIGLTNCD RWSDHCGKMDALCEQEKYKGPPKFTEKEYVVPAPGKLPEMWNDIFERLEQNGTSTKFF	766	
Endoglucanase VII [T. reesei RUT C-3] ETS3833.1 GI:57228773	MKSCAILAALGCLAGSVLGHGQVQNFTINGQYNQGFILDYYYQKQNTGHFPNVAGWYAEDLDLG FISPDQYTTPDIVCHKNAAPGAISATAAAGSNIVFQWGPGVWPHPYGPIVTYVAECSGSCTTVN KNNLRWVKIQEAGINYNTQVWAQQDLINQGNKWTVKIPSSLRPGNYVFRHELLAAHGASSANGM QNYPQCVNIAVTGSGTKALPAGTPATQLYKPTDPGILFNPYTTITSYTIPGPALWQG	767	
Endoglucanase III [T. reesei RUT C-3] ETS4885.1	MNKSVAPLLLAASILYGGAAAQQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCIPGATTIT TSTRPPSGPTTTTRATSTSSSTPPTSSGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGS NNYPDGIGQMQHFVNDDGMTIFRLPVGWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIV DIHNYARWNGGIIGQGGPTNAQFTSLWSQLASKYASQSRVWFGIMNEPHDVNINTWAATVQEVV TAIRNAGATSQFISLPGNDWQSAGAFISDGSAAALSQVTNPDGSTTNLIFDVHKYLDSDNSGTH AECTTNNIDGAFSPLATWLRQNNRQAILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWG	768	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
GI:572281861	AGSFDSTYVLTETPTGSGNSWTDTSLVSSCLARK		
Putative endoglucanase [T. reesei RUT C-3] ETS538.1 GI:572282294	MKLLSIASLLSLVATAQAHMEVSWPPVFRSKYNPRVPGNLINYDMTSPLNADGSNYPCKGYQVD VGRPEGAPGVTWRAGGTYNLTVAGSATHSGGSCQASLSYDRGRTWVVVHSWIGGCPLTPTWDFT LPNDTPPGEALFAWTWFNRIGNREMYMNCGAVTIRPSGRAARSPADSIYNRPAQFVANVNNGCA TLEGADVLFPSPGPDTDFDSDRTAAPVGKCGASSRRTRPVRA	769	
Endoglucanase-5 (Cellulase V; Endo-1,4-beta-glucanase V; EG V; Endoglucanase V; P43317.1 GI:117136	MKATLVLGSLIVGAVSAYKATTTRYYDGQEGACGCGSSSGAFPWQLGIGNGVYTAAGSQALFDT AGASWCGAGCGKCYQLTSTGQAPCSSCGTGGAAGQSIIVMVTNLCPNNGNAQWCPVVGGTNQYG YSYHFDIMAQNEIFGDNVVVDFEPIACPGQAASDWGTCLCVGQQETDPTPVLGNDTGSTPPGSS PPATSSSPPSGGGQQTLYGQCGGAGWTGPTTCQAPGTCKVQNQWYSQCLP	770	
Chain A, Active-Site Mutant E212q Determined At Ph 6. With No Ligand Bound In The Active Site 2CEL_A GI:194214	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSQMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	771	
Chain B, Active-Site Mutant E212q Determined At Ph 6. With No Ligand Bound In The Active Site 2CEL_B GI:194215	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSQMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	772	
Chain A, Active-Site Mutant E212q Determined At Ph 6. With Cellobiose Bound In The Active Site 3CEL_A GI:157836779	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSQMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	773	
Chain A, Active-site Mutant D214n	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS	774	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Determined At Ph 6. With No Ligand Bound In The Active Site 4CEL_A GI:1941941	FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMNIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG		
Chain B, Active-site Mutant D214n Determined At Ph 6. With No Ligand Bound In The Active Site 4CEL_B GI:1941942	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMNIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	775	
Endoglucanase-4 (Cellulase IV; Cellulase-61A; Cel61A; Endo-1,4-beta-glucanase IV; EGIV; Endoglucanase IV; Endoglucanase-61A) O1445.1 GI:21263647	MIQKLSNLLVTALAVATGVVGHGHINDIVINGVWYQAYDPTTFPYESNPPIVVGWTAADLDNGF VSPDAYQNPDIICHKNATNAKGHASVKAGDTILFQWVPVPWPHPGPIVDYLANCNGDCETVDKT TLEFFKIDGVGLLSGGDPGTWASDVLISNNNTWVVKIPDNLAPGNYVLRHEIIALHSAGQANGA QNYPQCFNIAVSGSGSLQPSGVLGTDLYHATDPGVLINIYTSPLNYIIPGPTVVSGLPTSVAQG SSAATATASATVPGGGSGPTSRTTTTARTTQASSRPSSTPPATTSAPAGGPTQTLYGQCGGSGY SGPTRCAPPATCSTLNPYYAQCLN	776	
Endoglucanase I precursor [T. reesei RUT C-3] ETS775.1 GI:57228411	MAPSVTLPLTTAILAIARLVAAQQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMH DANYNSCTVNGGVNTTLCPDEATCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVS PRLYLLDSDGEYVMLKLNGQELSFDVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGY CDAQCPVQTWRNGTLNTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYK SYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDTISSCPSASA YGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNI RWGDIGSTTNSTAPPPPPASSTTFSTTRRSSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTC QYSNDYYSQCL	777	
Chain A, Cbh1 (E217q) In Complex With Cellohexaose And Cellobiose 7CEL_A GI:157837135	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWQANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	778	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Chain A, Cbh1 (E212q) Cellotetraose Complex 5CEL_A GI:1578372	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSQMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	779	
Chain A, Cbh1 (E212q) Cellopentaose Complex 6CEL_A GI:15783787	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSQMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	780	
Endoglucanase EG-II (EGLII; Cellulase; Endo-1,4-beta-glucanase(P7982.1 GI:121794	MNKSVAPLLLAASILYGGAVAQQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCIPGATTIT TSTRPPSGPTTTTRATSTSSSTPPTSSGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGS NNYPDGIGQMQHFVNEDGMTIFRLPVGWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIV DIHNYARWNGGIIGQGGPTNAQFTSLWSQLASKYASQSRVWFGIMNEPHDVNINTWAATVQEVV TAIRNAGATSQFISLPGNDWQSAGAFISDGSAAALSQVTNPDGSTTNLIFDVHKYLDSDNSGTH AECTTNNIDGAFSPLATWLRQNNRQAILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWG AGSFDSTYVLTETPTSSGNSWTDTSLVSSCLARK	781	
Endoglucanase EG-1; Cellulase; Endo-1,4-beta-glucanase; P7981.1 GI:121788	MAPSVTLPLTTAILAIARLVAAQQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMH DANYNSCTVNGGVNTTLCPDEATCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVS PRLYLLDSDGEYVMLKLNGQELSFDVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGY CDAQCPVQTWRNGTLNTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYK SYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDTISSCPSASA YGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNI RWGDIGSTTNSTAPPPPPASSTTFSTTRRSSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTC QYSNDYYSQCL	782	
endo-1,4-beta-glucanase V (EGV) [T. reesei] CAA83846.1 GI:485864	MKATLVLGSLIVGAVSAYKATTTRYYDGQEGACGCGSSSGAFPWQLGIGNGVYTAAGSQALFDT AGASWCGAGCGKCYQLTSTGQAPCSSCGTGGAAGQSIIVMVTNLCPNNGNAQWCPVVGGTNQYG YSYHFDIMAQNEIFGDNVVVDFEPIACPGQAASDWGTCLCVGQQETDPTPVLGNDTGSTPPGSS PPATSSSPPSGGGQQTLYGQCGGAGWTGPTTCQAPGTCKVQNQWYSQCLP	783	
beta-1,4-glucanase [T. reesei]	MKFLQVLPALIPAALAQTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADW QWSGGQNNVKSYQNSQIAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYS	784	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
ABV71388.1 GI:15777972	GDYELMIWLGKYGDIGPIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFF NYLRDNKGYNAAGQYVLSYQFGTEPFTGSGTLNVASWTASIN		
Cellbiohydrolase II [T. reesei] ADC83999.1 GI:289152138	MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPG AASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYY ASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVV YDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGT PKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTG QQQWGDWCNVTGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPA PQAGAWFQAYFVQLLTNANPSFL	785	
Endo-beta-1,4-glucanase [T. reesei] BAA214.1 GI:2116583	MKFLQVLPALIPAALAQTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADW QWSGGQNNVKSYQNSQIAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYS GDYELMIWLGKYGDIGPIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFF NYLRDNKGYNAAGQYVLSYQFGTEPFTGSGTLNVASWTASIN	786	
Chain A, Crystal Structure Of Cel5a (Eg2) From Hypocrea Jecorina (T. Reesei) 3QR3_A GI:39981273	MGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGSNNYPDGIGQMQHFVNEDGMTIFRLPV GWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIVDIHNYARWNGGIIGQGGPTNAQFTSL WSQLASKYASQSRVWFGIMNEPHDVNINTWAATVQEVVTAIRNAGATSQFISLPGNDWQSAGAF ISDGSAAALSQVTNPDGSTTNLIFDVHKYLDSDNSGTHAECTTNNIDGAFSPLATWLRQNNRQA ILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWGAGSFDSTYVLTETPTSSGNSWTDTSL VSSCLARKGGSGSGHHHHHH	787	
Chain B, Crystal Structure Of Cel5a (Eg2) From Hypocrea Jecorina (T. Reesei) 3QR3_B GI:39981274	MGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGSNNYPDGIGQMQHFVNEDGMTIFRLPV GWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIVDIHNYARWNGGIIGQGGPTNAQFTSL WSQLASKYASQSRVWFGIMNEPHDVNINTWAATVQEVVTAIRNAGATSQFISLPGNDWQSAGAF ISDGSAAALSQVTNPDGSTTNLIFDVHKYLDSDNSGTHAECTTNNIDGAFSPLATWLRQNNRQA ILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWGAGSFDSTYVLTETPTSSGNSWTDTSL VSSCLARKGGSGSGHHHHHH	788	
Cel74a [T. reesei] AAP57752.1 GI:317471	MKVSRVLALVLGAVIPAHAAFSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADD SWTAVTDGIADNAGWHNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFT NLPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPD PSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGK YFPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLD LQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDV TSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVR	789	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	VGNTAGTQQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQRSQFQGSFAS VSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAGTLY VSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGASW TDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSISSA SSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQND YYYQCV		
Cel61b [T. reesei] AAP57753.1 GI:31747162	MKSCAILAALGCLAGSVLGHGQVQNFTINGQYNQGFILDYYYQKQNTGHFPNVAGWYAEDLDLG FISPDQYTTPDIVCHKNAAPGAISATAAAGSNIVFQWGPGVWPHPYGPIVTYVVECSGSCTTVN KNNLRWVKIQEAGINYNTQVWAQQDLINQGNKWTVKIPSSLRPGNYVFRHELLAAHGASSANGM QNYPQCVNIAVTGSGTKALPAGTPATQLYKPTDPGILFNPYTTITSYTIPGPALWQG	790	
Cel5b [T. reesei] AAP57754.1 GI:31747164	MRATSLLAAALAVAGDALAGKIKYLGVAIPGIDFGCDIDGSCPTDTSSVPLLSYKGGDGAGQMK HFAEDDGLNVFRISATWQFVLNNTVDGKLDELNWGSYNKVVNACLETGAYCMIDMHNFARYNGG IIGQGGVSDDIFVDLWVQIAKYYEDNDKIIFGLMNEPHDLDIEIWAQTCQKVVTAIRKAGATSQ MILLPGTNFASVETYVSTGSAEALGKITNPDGSTDLLYFDVHKYLDINNSGSHAECTTDNVDAF NDFADWLRQNKRQAIISETGASMEPSCMTAFCAQNKAISENSDVYIGFVGWGAGSFDTSYILTL TPLGKPGNYTDNKLMNECILDQFTLDEKYRPTPTSISTAAEETATATATSDGDAPSTTKPIFRE ETASPTPNAVTKPSPDTSDSSDDDKDSAASMSAQGLTGTVLFTVAALGYMLVAF	791	
Endoglucanase I 132152A GI:22541	MAPSVTLPLTTAILAIARLVAAQQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMH DANYNSCTVNGGVNTTLCPDEATCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVS PRLYLLDSDGEYVMLKLNGQELSFDVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGY CDAQCPVQTWRNGTLNTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYK SYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDTISSCPSASA YGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNI RWGDIGSTTNSTAPPPPPASSTTFSTTRRSSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTC QYSNDYYSQCL	792	
Endoglucanase IV [Trichoderma reesei] CAA71999.1 GI:2315274	MIQKLSNLLVTALAVATGVVGHGHINDIVINGVWYQAYDPTTFPYESNPPIVVGWTAADLDNGF VSPDAYQNPDIICHKNATNAKGHASVKAGDTILFQWVPVPWPHPGPIVDYLANCNGDCETVDKT TLEFFKIDGVGLLSGGDPGTWASDVLISNNNTWVVKIPDNLAPGNYVLRHEIIALHSAGQANGA QNYPQCFNIAVSGSGSLQPSGVLGTDLYHATDPGVLINIYTSPLNYIIPGPTVVSGLPTSVAQG SSAATATASATVPGGGSGPTSRTTTTARTTQASSRPSSTPPATTSAPAGGPTQTLYGQCGGSGY SGPTRCAPPATCSTLNPYYAQCLN	793	
Endoglucanase I [Trichoderma reesei] ADM8177.1 GI:3329711	MAPSVTLPLTTAILAIARLVAAQQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMH DANYNSCTVNGGVNTTLCPDEATCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVS PRLYLLDSDGEYVMLKLNGQELSFDVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGY CDAQCPVQTWRNGTLNTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFSPYGSGYK SYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDVPSAQPGGDTISSCPSASA YGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNI	794	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	RWGDIGSTTNSTAPPPPPASSTTFSTTRRSSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTC QYSNDYYSQCL		
Endoglucanase II precursor [Trichoderma reesei] AAA34213.1 GI:17549	MNKSVAPLLLAASILYGGAVAQQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCIPGATTIT TSTRPPSGPTTTTRATSTSSSTPPTSSGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGS NNYPDGIGQMQHFVNEDGMTIFRLPVGWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIV DIHNYARWNGGIIGQGGPTNAQFTSLWSQLASKYASQSRVWFGIMNEPHDVNINTWAATVQEVV TAIRNAGATSQFISLPGNDWQSAGAFISDGSAAALSQVTNPDGSTTNLIFDVHKYLDSDNSGTH AECTTNNIDGAFSPLATWLRQNNRQAILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWG AGSFDSTYVLTETPTSSGNSWTDTSLVSSCLARK	795	
Endoglucanase II [Trichoderma reesei] ABA64553.1 GI:77176916	MNKSVAPLLLAASILYGGAVAQQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCIPGATTIT TSTRPPSGPTTTTRATSTSSSTPPTSSGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGS NNYPDGIGQMQHFVNEDGMTIFRLPVGWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIV DIHNYARWNGGIIGQGGPTNAQFTSLWSQLASKYASQSRVWFGIMNEPHDVNINTWAATVQEVV TAIRNAGATSQFISLPGNDWQSAGAFISDGSAAALSQVTNPDGSTTNLIFDVHKYLDSDNSGTH AECTTNNIDGAFSPLATWLRQNNRQAILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWG AGSFDSTYVLTETPTGSGNSWTDTSLVSSCLARK	796	
Chain A, The X-Ray Crystal Structure Of T. Reesei Family 12 Endoglucanase 3, Cel12a, At 1.9 A Resolution 1H8V_A GI:14278359	XTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADWQWSGGQNNVKSYQNSQ IAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYSGDYELMIWLGKYGDIG PIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYV LSYQFGTEPFTGSGTLNVASWTASIN	797	
Chain B, The X-Ray Crystal Structure Of T. Reesei Family 12 Endoglucanase 3, Cel12a, At 1.9 A Resolution 1H8V_B GI:142783	XTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADWQWSGGQNNVKSYQNSQ IAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYSGDYELMIWLGKYGDIG PIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYV LSYQFGTEPFTGSGTLNVASWTASIN	798	
Chain C, The X-Ray Crystal Structure Of T. Reesei	XTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADWQWSGGQNNVKSYQNSQ IAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYSGDYELMIWLGKYGDIG PIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYV	799	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Family 12 Endoglucanase 3, Cel12a, At 1.9 A Resolution 1H8V_C GI:14278361	LSYQFGTEPFTGSGTLNVASWTASIN		
Chain D, The X-Ray Crystal Structure Of T. Reesei Family 12 Endoglucanase 3, Cel12a, At 1.9 A Resolution 1H8V_D GI:14278362	XTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADWQWSGGQNNVKSYQNSQ IAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYSGDYELMIWLGKYGDIG PIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYV LSYQFGTEPFTGSGTLNVASWTASIN	800	
Chain E, The X-Ray Crystal Structure Of The T. Reesei Family 12 Endoglucanase 3, Cel12a, At 1.9 A Resolution 1H8V_E GI:14278363	XTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADWQWSGGQNNVKSYQNSQ IAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYSGDYELMIWLGKYGDIG PIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYV LSYQFGTEPFTGSGTLNVASWTASIN	801	
Chain F, The X-Ray Crystal Structure Of The T. Reesei Family 12 Endoglucanase 3, Cel12a, At 1.9 A Resolution 1H8V_F GI:14278364	XTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADWQWSGGQNNVKSYQNSQ IAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYSGDYELMIWLGKYGDIG PIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYV LSYQFGTEPFTGSGTLNVASWTASIN	802	
Endoglucanase I precursor [T. reesei] AAA34212.1 GI:17547	MAPSVTLPLTTAILAIARLVAAQQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMH DANYNSCTVNGGVNTTLCPDEATCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVS PRLYLLDSDGEYVMLKLNGQELSFDVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGY CDAQCPVQTWRNGTLNTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYK SYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDTISSCPSASA YGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNI RWGDIGSTTNSTAPPPPPASSTTFSTTRRSSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTC QYSNDYYSQCL	803	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Chain A, Solution Structure Of The Cellulose-binding Domain Of Endoglucanase I From T. Reesei And Its Interaction With Cello-oligosaccharides 4BMF_A GI:5743	SCTQTHWGQCGGIGYSGCKTCTSGTTCQYSNDYYSQCL	804	
Chain A, Endoglucanase I From Trichoderma Reesei 1EG1_A GI:239236	XQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMHDANYNSCTVNGGVNTTLCPDEA TCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVSPRLYLLDSDGEYVMLKLNGQEL SFDVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGYCDAQCPVQTWRNGTLNTSHQGF CCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYKSYYGPGDTVDTSKTFTIITQFN TDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDTISSCPSASAYGGLATMGKALSSGMVLVFSIW NDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNIRWGDIGSTT	805	
Chain C, Endoglucanase I From Trichoderma Reesei 1EG1_C GI:239237	XQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMHDANYNSCTVNGGVNTTLCPDEA TCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVSPRLYLLDSDGEYVMLKLNGQEL SFDVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGYCDAQCPVQTWRNGTLNTSHQGF CCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYKSYYGPGDTVDTSKTFTIITQFN TDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDTISSCPSASAYGGLATMGKALSSGMVLVFSIW NDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNIRWGDIGSTT	806	
Endoglucanase [Trichoderma reesei RUT C-3] ETS6856.1 GI:572283882	MRSFSLLGSLSLLTSLSWALPTEGVISKLEGRQSGSSWFLPNIDHTTGAVRGYVPNLFNSAGQQ NFTYPVYKTVASGDSAGFVNALYSDGPSGGQRDNCYLAGEPRVIYLPPGTYTVSSTIFFDTDTV IIGDAANPPTIKAAAGFNGDYLIVGGQGDGDSHPCGGSGGETHFSVMIKNVILDTTANAGSSGF TALSWAVAQNCALVNVKINMPQGVHTGMLVSGGSTISISDVSFNFGNIGLHWNGHQQGQIKGMT FTDCTNGIFIDSGFTISIFAPTCNTVGRCIVLNSGNAWVAVIDGQSINSGDFFTSNVGFPNFML ENISKDTTNSNMVVVGGNVKVGGSTSLGTYVYGNTRGANPVYQTNPTSQPVNRPAALAPGGRYP VINAPQYADKTVANVVNLKDPNQNGGHTLQGDGFTDDTAALQGALNTAASQGKIAYLPFGIYIV KSTITIPPGTELYGEAWSTISGSGSAFSSETNPTPVVQIGATPGQKGVAHVQDIRFTVNEALPG AILLRINMAGNNPGDVAVFNSLNTIGGTRDTSISCSSESNCRAAYLGLHLAAGSSAYIDNFWSW VADHATDQSGKGTRTAVKGGVLVEATAGTWLTGLGSEHNWLYQLSFHNAANVFISLFQSETNYN QGNNGAPLPGTPFDATSIDPNFSWCSGGDTVCRMGLAQYYTGSNSNIFHYAAGSWNFIGLTKVN QGLMNFIQSTISNAHLYGFTSGPNTGETMRLPNGVEFGNGGNDGYGGSWGTLIANIASQS	807	
Endoglucanase, partial [T. reesei]	MKATLVLGSLIVGAVSAYKATTTRYYDGQEGACGCGSSSGAFPWQLGIGNGVYTAAGSQALFDT AGASWCGAGCGKCYQLTSTGQAPCSSCGTGGAAGQSIIVMVTNLCPNNGNAQWCPVVGGTNQYG YSYHFDIMAQNEIFGDNVVVDFEPIACPGQAASDWGTCLCVGQQETDPTPVLGNDTGSTPPGSS PPATSSSPPSGGGQQTLYGQCGGAGWTGPTTCQAPGTCKVQNQWYSQCLP	808	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
AHK2346.1 GI:58829453			
Chain A, Hypocrea Jecorina Cel7a E212q Mutant In Complex With P-nitrophenyl Cellobioside 4UWT_A GI:922664681	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSQMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	809	
Chain A, O-nitrophenyl Cellobioside As An Active Site Probe For Family 7 Cellobiohydrolases 4VZ_A GI:931139719	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQDGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	810	
Chain A, Hypocrea Jecorina Cel7a (wild Type) Soaked With Xylopentaose. 4D5Q_A GI:783282859	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWQANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	811	
Chain A, Cbh1 In Complex With S-propranolol 1DY4_A GI:1284415	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	812	
glycoside hydrolase family 6 [T. reesei QM6a]	MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPG AASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYY ASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVV YDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGT	813	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
XP_696258.1 GI:58911115	PKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTG QQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPA PQAGAWFQAYFVQLLTNANPSFL		
glycoside hydrolase family 7 [T. reesei QM6a] XP_6969224.1 GI:58911443	MYRKLAVISAFLATARAQSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSS TNCYDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLM ASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQ CPRDLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRY YVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLW DDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGN PSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQ CL	814	
glycoside hydrolase family 7 [T. reesei QM6a] EGR44817.1 GI:34514556	MYRKLAVISAFLATARAQSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSS TNCYDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLM ASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQ CPRDLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRY YVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLW DDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGN PSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQ CL	815	
glycoside hydrolase family 6 [T. reesei QM6a] EGR5117.1 GI:3452782	MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPG AASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYY ASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVV YDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGT PKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTG QQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPA PQAGAWFQAYFVQLLTNANPSFL	816	
Chain A, Hypocrea Jecorina Cellobiohydrolase Cel7a E212q Soaked With Xylotriose. 4D5I_A GI:783282849	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSQMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	817	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Chain A, Hypocrea Jecorina Cellobiohydrolase Cel7a E217q Soaked With Xylopentaose. 4D5P_A GI:783282856	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWQANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	818	
Chain A, Hypocrea Jecorina Cel7a In Complex With (R)-Dihydroxy-Phenanthrenolol 2V3I_A GI:19356563	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	819	
Chain A, Hypocrea Jecorina Cel7a In Complex With (S)-Dihydroxy-Phenanthrenolol 2V3R_A GI:19356564	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	820	
Chain A, Michaelis Complex Of Hypocrea Jecorina Cel7a E217q Mutant With Cellononaose Spanning The Active Site 4C4C_A GI:57215318	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWQANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	821	
Chain A, Covalent Glycosyl-enzyme Intermediate Of Hypocrea Jecorina Cel7a E217q Mutant Trapped Using Dnp-2-deoxy-2-fluoro-cellotrioside	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWQANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	822	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
4C4D_A GI:572153181			
Chain A, Cel6a D175a Mutant 1HGW_A GI:1865599	SGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLME QTLADIRTANKNGGNYAGQFVVYDLPDRACAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPA NQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLA NHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDG TSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	823	
Chain B, Cel6a D175a Mutant 1HGW_B GI:1865591	SGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLME QTLADIRTANKNGGNYAGQFVVYDLPDRACAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPA NQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLA NHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDG TSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	824	
Chain A, Cel6a D221a Mutant 1HGY_A GI:18655911	SGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLME QTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLLVIEPASLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPA NQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLA NHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDG TSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	825	
Chain B, Cel6a D221a Mutant 1HGY_B GI:18655912	SGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLME QTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLLVIEPASLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPA NQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLA NHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDG TSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	826	
Cellobiohydrolase, beta glucan 13195A GI:223874	ESACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYTSAYSSZPGGGGGVVIFFKNVGARLYLMASDTTYQEFTLLGNEFSFDVDV SQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPSSN NANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTCDPDGC DWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQDGVTFQQPNAELGSYSGNE LNDDYCTAEEAEFGGSSFSDKGGLTQFXXATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSST PGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSGGNPPGGNPPGTTTTTTTSS SZPPPGAHRRYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL	827	
Chain A, Cel6a (y169f) With A Non-hydrolysable	SGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLME QTLADIRTANKNGGNYAGQFVVFDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPA	828	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Cellotetraose 1QJW_A GI:6137482	NQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLA NHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDG TSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL		
Chain B, Cel6a (y169f) With A Non-hydrolysable Cellotetraose 1QJW_B GI:6137483	SGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLME QTLADIRTANKNGGNYAGQFVVFDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPA NQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLA NHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDG TSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	829	
Chain A, Cel6a In Complex With M-iodobenzyl Beta-d-glucopyranosyl-Beta(1,4)-d-xylopyranoside 1QK_A GI:6137484	TATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQT LADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDI RTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQ DPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANH GWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTS DSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	830	
Chain B, Cel6a In Complex With M-iodobenzyl Beta-d-glucopyranosyl-Beta(1,4)-d-xylopyranoside 1QK_B GI:6137485	TATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQT LADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDI RTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQ DPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANH GWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTS DSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	831	
Chain A, Wild Type Cel6a With A Non-hydrolysable Cellotetraose 1QK2_A GI:6137486	TATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQT LADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDI RTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQ DPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANH GWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTS DSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	832	
Chain B, Wild Type Cel6a With A Non-hydrolysable Cellotetraose 1QK2_B GI:6137487	TATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQT LADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDI RTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQ DPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANH GWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTS DSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	833	
Exoglucanase 2 (also known as	MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPG AASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYY	834	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
1,4-beta-cellobiohydrolase; exocellobiohydrolase II; CBHII; Exoglucanase II P7987.1 GI:121855	ASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVV YDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGT PKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTG QQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPA PQAGAWFQAYFVQLLTNANPSFL		
Exoglucanase 1 (also known as 1,4-beta-cellobiohydrolase; Exocellobiohydrolase I; CBHI; AltName: Full=Exoglucanase I P62694.1 GI:542144	MYRKLAVISAFLATARAQSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSS TNCYDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLM ASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQ CPRDLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRY YVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLW DDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGN PSGGNPPGGNRGTTTTRRPATTTGSSPGPTQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQC L	835	
Unnamed protein product [T. reesei] CAV28333.1 GI:21829938	MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPG AASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYY ASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVV YDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGT PKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTG QQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPA PQAGAWFQAYFVQLLTNANPSFL	836	
Cellobiohydrolase II 134188A GI:225475	MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPG AASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYY ASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVV YDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGT PKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTG QQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPA PQAGAWFQAYFVQLLTNANPSFL	837	
Unnamed protein [Trichoderma reesei]	MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPG AASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYY ASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVV YDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGT	838	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
AAA72922.1 GI:17543	PKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGRLLANHGWSNAFFITDQGRSGKQPTG QQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPA AQAGAWFQAYFVQLLTNANPSFL		
Chain A, Hypocrea Jecorina Cellobiohydrolase Cel7a E217q Soaked With Xylotriose. 4D5J_A GI:783282851	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWQANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	839	
Chain A, Hypocrea Jecorina Cellobiohydrolase Cel7a E212q Soaked With Xylopentaose. 4D5O_A GI:783282853	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSQMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	840	
Chain A, Hypocrea Jecorina Cellobiohydrolase Cel7a E217q Soaked With Xylotetraose. 4D5V_A GI:783282861	XSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTLCPDN ETCAKNCCLDGAAYASTYGVTTSGNSLSIDFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFS FDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGW EPSSNNANTGIGGHGSCCSEMDIWQANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTC DPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELGS YSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTN ETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSG	841	
Chain A, Cellobiohydrolase Ii, Catalytic Domain, Mutant Y169f 1CB2_A GI:182776	SGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLME QTLADIRTANKNGGNYAGQFVVFDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPA NQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLA NHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDG TSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	842	
Chain B, Cellobiohydrolase	SGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLME QTLADIRTANKNGGNYAGQFVVFDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS	843	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Ii, Catalytic Domain, Mutant Y169f 1CB2_B GI:182777	DIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPA NQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLA NHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDG TSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL		
SS2c-G10 fusion protein	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMQPSTATAAPKEKTSSEKKDNYIIKGVFWDPACVIA	846	
SS2c-G10 fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG	848	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGCAACCAT CTACCGCTACCGCCGCTCCAAAAGAAAAGACCAGCAGTGAAAAGAAGGACAACTATATTATCAA AGGTGTCTTCTGGGACCCAGCATGTGTTATTGCTTAG		
SS2c-G10 fusion protein	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMQPSTATAAPKEKTSSEKKDNYIIKGVFWDPACVIA-	1024	
SS2e-A7b fusion protein, coding sequence	ATGTCTGCATCTACTACAAGTTTAGAGGAATATCAAAAAACTTTCCTTGAACTGGGATTAGAAT GCAAAGCACTAAGATTTGGGTCATTCAAGCTGAATTCAGGCAGGCAGTCGCCATATTTTTTCAA TCTTAGTTTGTTCAATTCTGGAAAGCTGTTGGCAAACCTTGCCACCGCGTATGCAACTGCTATC ATTCAATCGGAGCTTAAATTCGATGTTATTTTCGGACCTGCTTACAAAGGGATCCCTTTGGCTG CTATTGTATGCGTTAAACTAGCAGAAATCGGGGGCACTAAATTTCAAGGTATTCAATATGCTTT TAATAGAAAGAAAGTTAAAGACCACGGCGAAGGTGGTATTATTGTTGGAGCATCGCTTGAAGAC AAGAGGGTGTTGATTATCGACGATGTCATGACTGCAGGAACTGCAATCAATGAAGCATTTGAGA	849	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TAATCAGTATTGCTCAAGGTAGGGTAGTGGGTTGTATTGTTGCTTTAGATAGGCAAGAAGTGAT TCATGAATCTGATCCGGAAAGAACAAGTGCTACCCAATCTGTTTCAAAGAGATACAACGTTCCT GTGCTAAGTATTGTATCACTGACTCAAGTGGTACAATTTATGGGAAATAGACTATCACCAGAGC AAAAATCAGCGATTGAAAACTACCGTAAGGCCTATGGTATATGA		
SS2e-A7b fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA	850	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGTCTGCAT CTACTACAAGTTTAGAGGAATATCAAAAAACTTTCCTTGAACTGGGATTAGAATGCAAAGCACT AAGATTTGGGTCATTCAAGCTGAATTCAGGCAGGCAGTCGCCATATTTTTTCAATCTTAGTTTG TTCAATTCTGGAAAGCTGTTGGCAAACCTTGCCACCGCGTATGCAACTGCTATCATTCAATCGG AGCTTAAATTCGATGTTATTTTCGGACCTGCTTACAAAGGGATCCCTTTGGCTGCTATTGTATG CGTTAAACTAGCAGAAATCGGGGGCACTAAATTTCAAGGTATTCAATATGCTTTTAATAGAAAG AAAGTTAAAGACCACGGCGAAGGTGGTATTATTGTTGGAGCATCGCTTGAAGACAAGAGGGTGT TGATTATCGACGATGTCATGACTGCAGGAACTGCAATCAATGAAGCATTTGAGATAATCAGTAT TGCTCAAGGTAGGGTAGTGGGTTGTATTGTTGCTTTAGATAGGCAAGAAGTGATTCATGAATCT GATCCGGAAAGAACAAGTGCTACCCAATCTGTTTCAAAGAGATACAACGTTCCTGTGCTAAGTA TTGTATCACTGACTCAAGTGGTACAATTTATGGGAAATAGACTATCACCAGAGCAAAAATCAGC GATTGAAAACTACCGTAAGGCCTATGGTATATGA		
SS2e-A7b fusion protein	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMSASTTSLEEYQKTFLELGLECKALRFGSFKLNSGRQSPYFFNLSL FNSGKLLANLATAYATAIIQSELKFDVIFGPAYKGIPLAAIVCVKLAEIGGTKFQGIQYAFNRK KVKDHGEGGIIVGASLEDKRVLIIDDVMTAGTAINEAFEIISIAQGRVVGCIVALDRQEVIHES DPERTSATQSVSKRYNVPVLSIVSLTQVVQFMGNRLSPEQKSAIENYRKAYGI-	851	
SS2d-G11 fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT	853	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGACCGAAT TAGATTATCAAGGAACTGCTGAGGCGGCTTCTACCTCGTATAGTCGAAATCAAACGGACCTTAA GCCGTTTCCTTCTGCAGGCAGTGCATCTTCATCAATTAAAACGACGGAACCTGTGAAAGATCAT AGAAGAAGGCGTTCTTCCAGCATAATTTCACATGTGGAACCGGAGACTTTTGAAGATGAAAATG ACCAGCAACTTCTACCAAATATGAATGCTACTTGGGTAGACCAACGCGGCGCTTGGATTATTCA TGTGGTCATTATCATACTGCTGAAACTATTTTATAATTTATTTCCTGGTGTTACCACAGAATGG TCGTGGACTCTGACTAATATGACATATGTTATTGGGTCCTATGTCATGTTCCATCTGATTAAGG GTACCCCTTTCGATTTCAATGGTGGTGCTTATGACAACTTGACGATGTGGGAACAAATTGACGA CGAGACTTTATATACTCCTTCAAGAAAATTTTTGATTAGTGTCCCGATCGCCCTATTCTTAGTT AGTACTCATTATGCTCACTATGATTTGAAATTGTTTTCATGGAATTGTTTTTTGACAACCTTTG		

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GTGCTGTTGTCCCAAAGTTACCTGTTACTCATAGATTAAGGATTTCTATCCCAGGTATCACAGG TCGCGCCCAAATTAGTTGA		
SS2d-G11 fusion protein	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGSGMTELDYQGTAEAASTSYSRNQTDLKPFPSAGSASSSIKTTEPVKDH RRRRSSSIISHVEPETFEDENDQQLLPNMNATWVDQRGAWIIHVVIIILLKLFYNLFPGVTTEW SWTLTNMTYVIGSYVMFHLIKGTPFDFNGGAYDNLTMWEQIDDETLYTPSRKFLISVPIALFLV STHYAHYDLKLFSWNCFLTTFGAVVPKLPVTHRLRISIPGITGRAQIS	854	
SS2e-A7a fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG	855	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGCCAAGAG TAGCTATCATCATTTACACACTATATGGTCACGTTGCTGCCACCGCAGAGGCAGAAAAGAAGGG AATTGAAGCCGCTGGAGGCTCTGCAGACATTTATCAAGTCGAGGAAACGTTGTCTCCAGAAGTT GTTAAGGCGCTTGGCGGTGCTCCAAAGCCAGATTACCCAATTGCCACTCAAGATACGTTGACAG AATATGATGCCTTTTTGTTTGGTATTCCAACTAGATTTGGTAACTTCCCTGCTCAATGGAAGGC TTTCTGGGACCGTACCGGTGGGTTGTGGGCTAAGGGTGCTTTGCATGGTAAGGTCGCTGGTTGT TTCGTCTCCACCGGAACTGGTGGTGGTAATGAAGCCACAATTATGAACTCTTTGTCTACTTTGG CTCATCACGGTATCATTTTTGTCCCATTGGGTTACAAGAATGTTTTCGCTGAATTGACCAATAT GGATGAAGTTCACGGTGGTTCACCATGGGGTGCGGGTACCATTGCAGGCAGTGACGGTTCAAGA TCTCCTTCCGCCTTGGAATTACAAGTACACGAAATTCAAGGCAAGACTTTCTACGAAACCGTTG CAAAGTTTTGA		
SS2e-A7a fusion protein	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA	856	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMPRVAIIIYTLYGHVAATAEAEKKGIEAAGGSADIYQVEETLSPEV VKALGGAPKPDYPIATQDTLTEYDAFLFGIPTRFGNFPAQWKAFWDRTGGLWAKGALHGKVAGC FVSTGTGGGNEATIMNSLSTLAHHGIIFVPLGYKNVFAELTNMDEVHGGSPWGAGTIAGSDGSR SPSALELQVHEIQGKTFYETVAKF-		
SS4d-G5 fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC	858	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGGGAAAGA ACGTTTTGTTGCTAGGATCTGGTTTTGTTGCACAACCTGTTATCGACACATTGGCTGCTAATGA TGACATCAATGTCACTGTCGCATGTAGAACATTAGCCAATGCGCAAGCATTGGCCAAGCCCTCT GGATCCAAGGCTATTTCATTGGATGTTACCGATGACAGTGCCTTAGACAAAGTTCTGGCTGATA ACGATGTTGTCATCTCTTTGATTCCATACACCTTCCATCCAAATGTGGTAAAGAGCGCCATCAG AACAAAGACCGATGTCGTCACTTCCTCTTACATCTCACCTGCCTTAAGAGAATTGGAACCAGAA ATCGTAAAGGCAGGTATTACAGTTATGAACGAAATTGGGTTGGATCCAGGTATCGACCACTTGT ATGCGGTCAAGACTATTGATGAAGTTCACAGAGCTGGTGGTAAGCTAAAGTCATTCTTGTCATA CTGTGGTGGTTTACCAGCTCCTGAAGACTCTGATAATCCATTAGGATACAAATTTTCATGGTCC TCCAGAGGTGTGCTACTGGCTTTAAGAAACTCTGCTAAATACTGGAAAGACGGAAAGATTGAAA CTGTTTCTTCCGAAGACTTAATGGCCACTGCTAAGCCTTACTTCATCTACCCAGGTTATGCATT CGTTTGCTACCCAAATAGAGACTCTACCCTTTTCAAGGATCTTTATCATATTCCAGAAGCCGAA ACGGTCATTAGAGGTACTTTGAGATATCAAGGTTTCCCAGAATTTGTTAAGGCTTTAGTTGACA TGGGTATGTTGAAGGATGATGCTAACGAAATCTTCAGCAAGCCAATTGCCTGGAACGAAGCACT AAAACAATATTTAGGTGCCAAGTCTACTTCTAAAGAAGATTTGATTGCTTCCATTGACTCAAAG GCTACTTGGAAAGATGATGAAGATAGAGAAAGAATCCTTTCCGGGTTTGCTTGGTTAGGCTTGT TCTCTGACGCAAAGATCACACCAAGAGGTAATGCTTTAGACACTCTATGTGCACGTTTAGAAGA ACTAATGCAATATGAAGACAATGAAAGAGATATGGTTGTACTACAACACAAATTCGGTATTGAA TGGGCTGATGGAACTACCGAAACAAGAACATCCACTTTAGTTGACTATGGTAAGGTTGGTGGTT ACAGTTCTATGGCCGCTACTGTTGGTTATCCAGTTGCCATTGCAACGAAATTCGTCTTAGATGG TACAATCAAGGGACCAGGCTTACTAGCGCCATACTCACCAGAGATTAATGATCCAATCATGAAA GAACTAAAGGACAAGTACGGCATCTATCTAAAGGAAAAGACAGTGGCTTAA		
SS4d-G5 fusion protein	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMGKNVLLLGSGFVAQPVIDTLAANDDINVTVACRTLANAQALAKPS	859	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GSKAISLDVTDDSALDKVLADNDVVISLIPYTFHPNVVKSAIRTKTDVVTSSYISPALRELEPE IVKAGITVMNEIGLDPGIDHLYAVKTIDEVHRAGGKLKSFLSYCGGLPAPEDSDNPLGYKFSWS SRGVLLALRNSAKYWKDGKIETVSSEDLMATAKPYFIYPGYAFVCYPNRDSTLFKDLYHIPEAE TVIRGTLRYQGFPEFVKALVDMGMLKDDANEIFSKPIAWNEALKQYLGAKSTSKEDLIASIDSK ATWKDDEDRERILSGFAWLGLFSDAKITPRGNALDTLCARLEELMQYEDNERDMVVLQHKFGIE WADGTTETRTSTLVDYGKVGGYSSMAATVGYPVAIATKFVLDGTIKGPGLLAPYSPEINDPIMK ELKDKYGIYLKEKTVA-		
SS4d-C7 fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT	861	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGTCACGTC TTCCTCTAAAGCAGTTCTTAGCGGATAACCCCAAAAAAGTTCTTGTTCTTGACGGTGGTCAAGG AACAGAACTGGAAAACAGAGGTATCAAAGTTGCAAATCCCGTGTGGTCTACTATTCCATTTATT AGCGAATCATTTTGGTCTGATGAGTCATCTGCTAACAGAAAAATTGTCAAAGAAATGTTCAACG ATTTCTTGAATGCTGGCGCAGAAATATTGATGACTACAACATACCAAACGAGTTATAAATCAGT TTCTGAAAACACCCCAATCAGAACTTTATCCGAGTACAATAACCTTTTAAACAGGATTGTCGAT TTTTCTCGTAATTGTATTGGCGAAGACAAATATTTGATTGGCTGTATTGGCCCATGGGGTGCTC ATATTTGTCGTGAGTTTACAGGCGACTATGGTGCTGAGCCAGAAAATATTGATTTCTACCAATA CTTCAAGCCTCAGTTGGAGAATTTCAATAAAAATGACAAATTGGATTTGATTGGGTTTGAAACC ATTCCTAACATCCATGAACTGAAAGCTATCTTATCTTGGGATGAGAGTATCCTGTCTAGACCCT TCTATATCGGGTTGTCTGTGCATGAGCACGGTGTCTTGAGAGACGGCACTACCATGGAAGAAAT CGCACAAGTTATTAAGGACTTGGGCGACAAAATAAATCCTAACTTCTCGTTCTTAGGAATCAAC TGCGTCAGCTTCAACCAATCACCCGACATTCTTGAGTCTCTACATCAAGCACTACCAAATATGG CCTTGCTTGCTTATCCAAACAGTGGTGAAGTTTATGATACTGAAAAGAAGATATGGTTGCCAAA TAGCGATAAGCTGAACAGTTGGGATACGGTTGTTAAACAGTACATTAGCAGCGGTGCCCGTATC ATTGGTGGTTGTTGCAGAACAAGTCCAAAAGACATCCAAGAGATTTCTGCAGCCGTCAAGAAAT ACACGTAA		
SS4d-C7 fusion protein	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMSRLPLKQFLADNPKKVLVLDGGQGTELENRGIKVANPVWSTIPFI SESFWSDESSANRKIVKEMFNDFLNAGAEILMTTTYQTSYKSVSENTPIRTLSEYNNLLNRIVD FSRNCIGEDKYLIGCIGPWGAHICREFTGDYGAEPENIDFYQYFKPQLENFNKNDKLDLIGFET IPNIHELKAILSWDESILSRPFYIGLSVHEHGVLRDGTTMEEIAQVIKDLGDKINPNFSFLGIN	862	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CVSFNQSPDILESLHQALPNMALLAYPNSGEVYDTEKKIWLPNSDKLNSWDTVVKQYISSGARI IGGCCRTSPKDIQEISAAVKKYT-		
SS3b-D8 fusion protein, coding sequence	ATGTCGCAAGAGTTCGAGACACCGGCGGTTGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAG GAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAG CATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGAC GATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCC GTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGA GATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAG AAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTG ACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCT ATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTAC ATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGT TTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCA TGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCC TGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCAC CCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAG AATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATA GTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCA GAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGG CAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAG GCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTG GACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTT CAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAG GGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAAT TGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCA AGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTT AGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGA TGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGC AGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGG TCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACA GTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGG GCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAG AAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCA TCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAA CTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAAC AAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGG CCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCT AATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAG	864	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACT CCCATCGTGTCTTGGATATGTGA		
SS3b-D8 fusion protein	MSQEFETPAVGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDD DEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVK KEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRY IYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITS WGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPI VLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREK AMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQ GYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPF STRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTR SYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQ KTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQN KVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNK NCMITYAAYRNIFPIWALGEYSHRVLDM-	865	
SS3b-D8 fusion protein, fusion domain	MSQEFETPAV	866	
SS2c-A10a	ATGAATTCGAATGAAGACATCATACCTGAACTATAA	867	
SS2c-A10a fusion protein	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMNSNEDIIPEL-	869	
SS2c-A10a fusion protein, fusion domain	MNSNEDIIPEL	870	
Pathway step 3 sequence A	ATGTGGACAGTTGTGTTGGGACTTGCTACCTTGTTTGTTGCCTATTATATTCATTGGATCAACA AGTGGAGAGATTCCAAGTTCAATGGTGTTCTACCTCCTGGAACTATGGGGCTACCATTGATAGG	871	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
CYP87D18 DNA (codon optimized)	AGAGACAATTCAGTTGTCAAGACCATCTGACAGTTTGGATGTGCATCCCTTTATCCAGAAGAAA GTCGAACGTTATGGTCCGATATTTAAAACCTGTTTGGCAGGCAGACCAGTTGTTGTTTCAGCGG ATGCAGAGTTCAATAATTACATTATGTTACAAGAAGGTAGAGCTGTAGAAATGTGGTATTTGGA CACACTGTCTAAATTCTTCGGGTTGGATACAGAGTGGTTAAAAGCCTTAGGCTTAATCCACAAG TACATAAGATCCATTACCCTAAACCATTTTGGTGCTGAAGCATTGAGAGAAAGATTCTTGCCAT TTATAGAGGCATCGTCTATGGAAGCGTTACATTCTTGGTCCACTCAACCCAGTGTGGAGGTCAA GAATGCAAGTGCTTTGATGGTATTCAGAACGTCTGTAAACAAAATGTTTGGAGAAGATGCTAAG AAATTATCAGGAAATATTCCAGGTAAATTCACAAAGCTGCTGGGTGGCTTTCTATCTCTACCGT TAAATTTTCCCGGCACTACTTATCACAAGTGCTTAAAAGACATGAAAGAAATCCAGAAGAAATT ACGTGAAGTTGTAGATGATAGACTTGCCAATGTTGGGCCAGATGTTGAGGACTTTCTAGGGCAA GCGTTGAAAGACAAAGAATCCGAGAAATTCATAAGCGAAGAATTTATCATCCAATTGCTATTTT CAATAAGCTTTGCTTCGTTCGAATCGATCAGCACGACGTTGACATTGATTTTGAAGCTACTTGA CGAACATCCTGAGGTTGTAAAGGAATTAGAAGCCGAACATGAAGCTATCAGAAAAGCTAGAGCT GATCCAGATGGTCCAATTACCTGGGAAGAATACAAATCTATGACCTTCACACTTCAAGTCATAA ACGAAACACTTAGGTTAGGCTCAGTGACTCCTGCCTTATTGAGGAAAACTGTTAAAGATCTGCA AGTCAAGGGTTACATTATTCCTGAAGGATGGACTATAATGTTGGTAACTGCATCTAGGCATCGT GATCCAAAGGTCTACAAAGATCCGCACATATTCAATCCTTGGAGATGGAAAGACCTGGACTCAA TTACCATTCAAAAGAACTTTATGCCATTCGGTGGTGGTTTAAGGCATTGTGCAGGAGCTGAATA CTCCAAAGTGTATCTGTGTACTTTTCTTCACATTCTTTGCACAAAATATAGGTGGACGAAGTTA GGTGGCGGTAGAATTGCAAGAGCCCATATTTTAAGTTTTGAGGATGGTTTGCACGTCAAGTTTA CTCCTAAAGAGTAA		
Pathway step 3 sequence B CYP87D18 Protein	MWTVVLGLATLFVAYYIHWINKWRDSKFNGVLPPGTMGLPLIGETIQLSRPSDSLDVHPFIQKK VERYGPIFKTCLAGRPVVVSADAEFNNYIMLQEGRAVEMWYLDTLSKFFGLDTEWLKALGLIHK YIRSITLNHFGAEALRERFLPFIEASSMEALHSWSTQPSVEVKNASALMVFRTSVNKMFGEDAK KLSGNIPGKFTKLLGGFLSLPLNFPGTTYHKCLKDMKEIQKKLREVVDDRLANVGPDVEDFLGQ ALKDKESEKFISEEFIIQLLFSISFASFESISTTLTLILKLLDEHPEVVKELEAEHEAIRKARA DPDGPITWEEYKSMTFTLQVINETLRLGSVTPALLRKTVKDLQVKGYIIPEGWTIMLVTASRHR DPKVYKDPHIFNPWRWKDLDSITIQKNFMPFGGGLRHCAGAEYSKVYLCTFLHILCTKYRWTKL GGGRIARAHILSFEDGLHVKFTPKE	872	
Pathway step 3 sequence C SgCPR DNA (codon optimized)	ATGAAGGTCAGTCCATTCGAATTCATGTCCGCTATTATCAAGGGTAGAATGGACCCATCTAACT CCTCATTTGAATCTACTGGTGAAGTTGCCTCCGTTATCTTTGAAAACAGAGAATTGGTTGCCAT CTTGACCACTTCTATTGCTGTTATGATTGGTTGCTTCGTTGTCTTGATGTGGAGAAGAGCTGGT TCTAGAAAGGTTAAGAATGTCGAATTGCCAAAGCCATTGATTGTCCATGAACCAGAACCTGAAG TTGAAGATGGTAAGAAGAAGGTTTCCATCTTCTTCGGTACTCAAACTGGTACTGCTGAAGGTTT TGCTAAGGCTTTGGCTGATGAAGCTAAAGCTAGATACGAAAAGGCTACCTTCAGAGTTGTTGAT TTGGATGATTATGCTGCCGATGATGACCAATACGAAGAAAAATTGAAGAACGAATCCTTCGCCG TTTTCTTGTTGGCTACTTATGGTGATGGTGAACCTACTGATAATGCTGCTAGATTTTACAAGTG GTTCGCCGAAGGTAAAGAAAGAGGTGAATGGTTGCAAAACTTGCACTATGCTGTTTTTGGTTTG	873	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GGTAACAGACAATACGAACACTTCAACAAGATTGCTAAGGTTGCCGACGAATTATTGGAAGCTC AAGGTGGTAATAGATTGGTTAAGGTTGGTTTAGGTGATGACGATCAATGCATCGAAGATGATTT TTCTGCTTGGAGAGAATCTTTGTGGCCAGAATTGGATATGTTGTTGAGAGATGAAGATGATGCT ACTACTGTTACTACTCCATATACTGCTGCTGTCTTGGAATACAGAGTTGTCTTTCATGATTCTG CTGATGTTGCTGCTGAAGATAAGTCTTGGATTAACGCTAATGGTCATGCTGTTCATGATGCTCA ACATCCATTCAGATCTAACGTTGTCGTCAGAAAAGAATTGCATACTTCTGCCTCTGATAGATCC TGTTCTCATTTGGAATTCAACATTTCCGGTTCCGCTTTGAATTACGAAACTGGTGATCATGTTG GTGTCTACTGTGAAAACTTGACTGAAACTGTTGATGAAGCCTTGAACTTGTTGGGTTTGTCTCC AGAAACTTACTTCTCTATCTACACCGATAACGAAGATGGTACTCCATTGGGTGGTTCTTCATTG CCACCACCATTTCCATCATGTACTTTGAGAACTGCTTTGACCAGATACGCTGATTTGTTGAACT CTCCAAAAAAGTCTGCTTTGTTGGCTTTAGCTGCTCATGCTTCTAATCCAGTTGAAGCTGATAG ATTGAGATACTTGGCTTCTCCAGCTGGTAAAGATGAATATGCCCAATCTGTTATCGGTTCCCAA AAGTCTTTGTTGGAAGTTATGGCTGAATTCCCATCTGCTAAACCACCATTAGGTGTTTTTTTTG CTGCTGTTGCTCCAAGATTGCAACCTAGATTCTACTCCATTTCATCCTCTCCAAGAATGGCTCC ATCTAGAATCCATGTTACTTGTGCTTTGGTTTACGATAAGATGCCAACTGGTAGAATTCATAAG GGTGTTTGTTCTACCTGGATGAAGAATTCTGTTCCAATGGAAAAGTCCCATGAATGTTCTTGGG CTCCAATTTTCGTTAGACAATCCAATTTTAAGTTGCCAGCCGAATCCAAGGTTCCAATTATCAT GGTTGGTCCAGGTACTGGTTTGGCTCCTTTTAGAGGTTTTTTACAAGAAAGATTGGCCTTGAAA GAATCCGGTGTTGAATTGGGTCCATCCATTTTGTTTTTCGGTTGCAGAAACAGAAGAATGGATT ACATCTACGAAGATGAATTGAACAACTTCGTTGAAACCGGTGCTTTGTCCGAATTGGTTATTGC TTTTTCTAGAGAAGGTCCTACCAAAGAATACGTCCAACATAAGATGGCTGAAAAGGCTTCTGAT ATCTGGAACTTGATTTCTGAAGGTGCTTACTTGTACGTTTGTGGTGATGCTAAAGGTATGGCTA AGGATGTTCATAGAACCTTGCATACCATCATGCAAGAACAAGGTTCTTTGGATTCTTCCAAAGC TGAATCCATGGTCAAGAACTTGCAAATGAATGGTAGATACTTAAGAGATGTTTGGTAA		
Pathway step 3 sequence D SgCPR Protein	MKVSPFEFMSAIIKGRMDPSNSSFESTGEVASVIFENRELVAILTTSIAVMIGCFVVLMWRRAG SRKVKNVELPKPLIVHEPEPEVEDGKKKVSIFFGTQTGTAEGFAKALADEAKARYEKATFRVVD LDDYAADDDQYEEKLKNESFAVFLLATYGDGEPTDNAARFYKWFAEGKERGEWLQNLHYAVFGL GNRQYEHFNKIAKVADELLEAQGGNRLVKVGLGDDDQCIEDDFSAWRESLWPELDMLLRDEDDA TTVTTPYTAAVLEYRVVFHDSADVAAEDKSWINANGHAVHDAQHPFRSNVVVRKELHTSASDRS CSHLEFNISGSALNYETGDHVGVYCENLTETVDEALNLLGLSPETYFSIYTDNEDGTPLGGSSL PPPFPSCTLRTALTRYADLLNSPKKSALLALAAHASNPVEADRLRYLASPAGKDEYAQSVIGSQ KSLLEVMAEFPSAKPPLGVFFAAVAPRLQPRFYSISSSPRMAPSRIHVTCALVYDKMPTGRIHK GVCSTWMKNSVPMEKSHECSWAPIFVRQSNFKLPAESKVPIIMVGPGTGLAPFRGFLQERLALK ESGVELGPSILFFGCRNRRMDYIYEDELNNFVETGALSELVIAFSREGPTKEYVQHKMAEKASD IWNLISEGAYLYVCGDAKGMAKDVHRTLHTIMQEQGSLDSSKAESMVKNLQMNGRYLRDVW	874	
Pathway step 3 sequence E CYP51G1 (codon	ATGGAACCTGAAAACAAGTTCTTCAATGTTGGGTTATTGATCGTAGTTACGTTGGTTTTGGCTA AACTAATTTCTGCGGTCATTAATTCCAGGTCTAAGAAGAGAGTACCTCCAACCGTCAAAGGTTT TCCACTTGTAGGTGGCTTGGTTAGATTTCTTAAAGGGCCAATTGTGATGTTGAGAGAAGAATAT	875	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
optimized)	CCCAAACATGGATCCGTATTCACTCTGAATTTACTACATAAGAAGATTACCTTTCTGATTGGAC CAGAAGTTTCTGCACATTTCTTTAAGGCTTCAGAGAGTGATTTATCACAGCAAGAAGTCTACCA ATTTAACGTGCCCACTTTTGGTCCGGGCGTTGTTTTCGATGTCGACTACTCGGTAAGGCAAGAA CAATTCAGATTCTTTACCGAAGCATTGAGAGTTACAAAACTGAAGGGCTATGTTGACCAAATGG TGAAAGAAGCAGAAGATTACTTTTCAAAATGGGGTGATTCAGGAGAGGTTGATCTAAAATGCGA ACTTGAACACTTGATCATATTAACCGCATCTAGATGTTTGTTGGGAAGAGAAGTTCGTGACCAG TTATTTGCTGATGTAAGTGCCCTATTTCATGACTTGGATAACGGTATGCTGCCAATATCCGTGA TGTTCCCATACTTGCCTATACCCGCTCATAGGAGAAGAGATCAAGCGAGATCAAAATTGGCTGA TATCTTTGTCAACATCATATCCTCTCGTAAATGTACTGGCACTTCTGAAAATGACATGTTACAA TGCTTTATAAACTCTAAATACAAAGATGGCAGACCAACTACTGATTCTGAAATCACAGGGTTAT TGATAGCCGCATTATTCGCTGGGCAACATACGAGCTCGATTACTAGCACATGGACAGGCGCATA TTTGTTATGTCACAAAGAGTATATGAGTGCCGTTCTTGAAGAGCAGCAGAAACAAATGGAGAAG CATGGTGACGAAATTGATCACGATATTCTATCCGAAATGGACAATTTGTACCGTTGCATCAAAG AAGCCCTAAGACTACATCCACCCTTGATTATGCTTATGAGGTCGAGTCATACCGATTTTAGCGT TACGACAAGAGAAGGAAAAGAGTATGATATTCCGAAGGGACATATTATAGCCACAAGTCCAGCT TTCGCAAATCGTTTACCTCACGTGTATAAAGACCCTGACAGATTTGATCCAGATAGGTTTGCTC CAGGTAGAGATGAGGATAAGGCTGCTGGACCTTTCTCCTACATATCATTTGGTGGTGGTAGACA CGGTTGTTTAGGTGAACCTTTTGCGTATTTACAAATCAAGGCAATCTGGTCACACTTACTGAGA AATTTTGAGTTAGAGTTGATTAGTCCTTTCCCGGAAATTGACTGGAATGCCATGGTTGTGGGTG TCAAGGGTAAAGTGATGGTCAGGTATAAGAGAAGAAAGCTTAGCGTATCTTAG		
Pathway step 3 sequence F CYP51G1 Protein	MEPENKFFNVGLLIVVTLVLAKLISAVINSRSKKRVPPTVKGFPLVGGLVRFLKGPIVMLREEY PKHGSVFTLNLLHKKITFLIGPEVSAHFFKASESDLSQQEVYQFNVPTFGPGVVFDVDYSVRQE QFRFFTEALRVTKLKGYVDQMVKEAEDYFSKWGDSGEVDLKCELEHLIILTASRCLLGREVRDQ LFADVSALFHDLDNGMLPISVMFPYLPIPAHRRRDQARSKLADIFVNIISSRKCTGTSENDMLQ CFINSKYKDGRPTTDSEITGLLIAALFAGQHTSSITSTWTGAYLLCHKEYMSAVLEEQQKQMEK HGDEIDHDILSEMDNLYRCIKEALRLHPPLIMLMRSSHTDFSVTTREGKEYDIPKGHIIATSPA FANRLPHVYKDPDRFDPDRFAPGRDEDKAAGPFSYISFGGGRHGCLGEPFAYLQIKAIWSHLLR NFELELISPFPEIDWNAMVVGVKGKVMVRYKRRKLSVS.	876	
Pathway step 3 sequence G CYP71B97 (codon optimized)	ATGTTATCGTTGGCCATTTGGGTTTCACTTTTGTTCTTGTTGTCATCATTGCTTCTTTTAAAGA CGAAGAAGAAAGTTGCTCCACAAAAGAAGAAGAAGCAATTTCCACCTGGACCTCCCAAACTACC ATTGTTAGGCCATCTGCACTTATTGGGTTCTTTGCCTCATTGCTCCTTATGTGAACTGTCTAGA AAATATGGTCCTGTCATGTTGTTAAAATTAGGCTCAGTACCTACCGTAGTCATATCTAGCGCTG CAGCCGCTAGAGAGGTGTTGAAAGTACACGATCTAGCATGTTGCTCTCGTCCGAGATTGGCTGC TTCCGGTAGATTCTCGTACAATTTTCTGGATCTGAACTTAAGCCCATATGGTGAGAGATGGAGA GAACTGAGGAAAATTTGCGTATTGGTTTTGCTGAGTGCTAGACGTGTTCAGAGCTTCCAACAGA TAAGAGAAGAAGAGGTGGGATTATTACTTAAATCCATTAGTCAAGTTTCCAGTAGTGCCACTCC AGTTGATCTATCTGAGAAATCCTATTCTTTGACAGCTAACATTATCACTAGAATCGCGTTTGGG	877	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAGTCATTCAGAGGTGGCGAATTAGACAATGAAAACTTTCAACAAGTCATCCACAGAGCATCGA TTGCCTTAGGTTCCTTTTCTGTGACAAACTTCTTTCCTTCAGTAGGGTGGATTATCGACAGATT AACCGGTGTACATGGCAGATTGGAGAAGAGTTTTGCTGAATTAGACACCTTCTTTCAGCATATC ATTGATGATCGTATCAATTTTGTCGCAACAAGCCAAACCGAAGAAAACATTATAGACGTACTAT TGAAAATGGAAAGAGAACGTTCAAAATTTGATGTCCTACAACTGAATAGGGACTGCATAAAAGC CTTGATAATGGATATATTTCTTGCCGGTGTAGATACTGGAGCAGGGACAATTGTGTGGGCATTG ACTGAATTGGTGAGAAATCCCAGAGTGATGAAGAAGTTGCAAGACGAAATAAGGTCGTGTGTGA AAGAGGATCAAGTCAAGGAACGTGATTTAGAGAAACTTCAGTACTTAAAGATGGTCGTTAAAGA AGTTTTAAGATTGCATGCTCCAGTTCCTTTGTTATTGCCGAGAGAGACAATGTCTCATTTCAAA CTAAATGGTTATGACATTGATCCGAAAACTCACTTGCATGTCAATGTTTGGGCGATTGGTAGGG ACCCAGATTCTTGGTCTGATCCAGAAGAATTCTTCCCAGAAAGATTCGCAGGATCAAGTATTGA TTACAAAGGACATAATTTTGAATTGCTGCCATTTGGTGGTGGCAGAAGGATCTGTCCCGGTATG AACATGGGGACAGTTGCGGTTGAACTTGCACTAACGAACCTATTACTTTGTTTTGATTGGACTC TACCTGATGGCATGAAAGAGGAAGATGTTGACATGGAAGAAGATGGTGGACTTGCTATTGCTAA GAAATCTCCCCTAAAATTAGTTCCAGTTAGGTGTCTTAATTAG		
Pathway step 3 sequence H CYP71B97 Protein	MLSLAIWVSLLFLLSSLLLLKTKKKVAPQKKKKQFPPGPPKLPLLGHLHLLGSLPHCSLCELSR KYGPVMLLKLGSVPTVVISSAAAAREVLKVHDLACCSRPRLAASGRFSYNFLDLNLSPYGERWR ELRKICVLVLLSARRVQSFQQIREEEVGLLLKSISQVSSSATPVDLSEKSYSLTANIITRIAFG KSFRGGELDNENFQQVIHRASIALGSFSVTNFFPSVGWIIDRLTGVHGRLEKSFAELDTFFQHI IDDRINFVATSQTEENIIDVLLKMERERSKFDVLQLNRDCIKALIMDIFLAGVDTGAGTIVWAL TELVRNPRVMKKLQDEIRSCVKEDQVKERDLEKLQYLKMVVKEVLRLHAPVPLLLPRETMSHFK LNGYDIDPKTHLHVNVWAIGRDPDSWSDPEEFFPERFAGSSIDYKGHNFELLPFGGGRRICPGM NMGTVAVELALTNLLLCFDWTLPDGMKEEDVDMEEDGGLAIAKKSPLKLVPVRCLN.	878	
Pathway step 3 sequence I CYP73A152 (codon optimized)	ATGGATTTGCTTTTGTTGGAAAAGACGTTGTTGGGTCTATTTATCGCTGTCGTATTGGCAATAG CCATTAGCAAATTAAGGGGTAAAAGGTTTAAACTGCCACCAGGTCCGTTACCTGTCCCTATCTT TGGCAACTGGTTACAGGTTGGTGATGATTTGAACCACAGAAATCTAACGGGTTTAGCCAAGAAA TTTGGGGATATTTTCTTGTTAAGAATGGGCCAAAGAAACTTAGTGGTAGTTTCATCTCCTGAAC TTGCCAAAGAAGTGCTTCATACACAAGGAGTGGAGTTTGGATCTAGAACAAGAAATGTAGTGTT CGACATATTTACCGGAAAAGGTCAAGATATGGTTTTCACAGTATATGGTGAACATTGGCGTAAA ATGCGTAGAATAATGACTGTACCATTCTTCACCAACAAGGTTGTCCAACAATATAGGCATGGAT GGGAAGCAGAAGCAGCTAGCGTTGTTGAAGATGTGAAGAAGAATCCGGAATCTGCTACTACTGG TATTGTGTTACGTCGTAGACTTCAATTGATGATGTACAATAACATGTATCGTATAATGTTTGAC AGAAGATTTGAGTCCGAGGATGATCCCCTATTTCACAAATTGAGAGCACTGAATGGTGAGAGAT CTAGGTTGGCTCAATCGTTCGAGTACAACTATGGAGACTTCATCCCTATTTTAAGACCTTTCTT GAGAGGCTATTTGAAAATTTGCAAGGAAGTCAAGGACACTAGGTTACAGTTGTTTAAAGACTAC TTTGTTGAAGAAAGAAAGAAATTGGCGAACGTGAAAACTACCACAAATGAGGGCTTAAAATGTG CGATCGATCACATTCTGGACGCACAACAGAAAGGTGAAATCAATGAAGATAACGTTTTATACAT TGTTGAGAATATTAATGTAGCTGCCATTGAAACTACGTTGTGGTCGATAGAATGGGGAATTGCA	879	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GAGCTTGTCAATCATCCTGAAATCCAAAGAAAGCTGAGAAATGAGATGGATACAGTCTTAGGCT CAGGTGTTCCTATCACTGAACCAGATACACATAAGTTGCCCTATTTACAAGCTGTCATAAAAGA AACTCTTAGACTTAGAATGGCTATACCCTTGCTAGTTCCACATATGAATCTACATGATGCCAAA CTGGGTGGTTACGACATTCCAGCAGAATCCAAGATTCTAGTAAACGCTTGGTGGTTAGCCAATA ATCCAGCTAATTGGAAGAATCCAGAAGAATTCAGACCAGAGAGATTCTTGGAAGAAGAATCCAA AGTTGAAGCTAATGGGAACGACTTTAGATATTTACCGTTCGGTGTAGGAAGAAGGAGTTGTCCA GGGATAATTTTAGCGCTACCTATCCTAGCTATCACCATAGGCAGACTGGTTCAGAACTTTGAAT TGTTACCTCCACCAGGGCAAAGTAAGCTGGATACAAGTGAGAAGGGTGGTCAGTTTTCATTGCA TATTCTTAAACACTCAACCATTGTCGTTAAACCCAGGGCATTTTAG		
Pathway step 3 sequence J CYP73A152 Protein	MDLLLLEKTLLGLFIAVVLAIAISKLRGKRFKLPPGPLPVPIFGNWLQVGDDLNHRNLTGLAKK FGDIFLLRMGQRNLVVVSSPELAKEVLHTQGVEFGSRTRNVVFDIFTGKGQDMVFTVYGEHWRK MRRIMTVPFFTNKVVQQYRHGWEAEAASVVEDVKKNPESATTGIVLRRRLQLMMYNNMYRIMFD RRFESEDDPLFHKLRALNGERSRLAQSFEYNYGDFIPILRPFLRGYLKICKEVKDTRLQLFKDY FVEERKKLANVKTTTNEGLKCAIDHILDAQQKGEINEDNVLYIVENINVAAIETTLWSIEWGIA ELVNHPEIQRKLRNEMDTVLGSGVPITEPDTHKLPYLQAVIKETLRLRMAIPLLVPHMNLHDAK LGGYDIPAESKILVNAWWLANNPANWKNPEEFRPERFLEEESKVEANGNDFRYLPFGVGRRSCP GIILALPILAITIGRLVQNFELLPPPGQSKLDTSEKGGQFSLHILKHSTIVVKPRAF.	880	
Pathway step 3 sequence K CYP80C13 (codon optimized)	ATGTTAAAAGATCCCTTTTGCTTTCCCTTTCTACCTCTGTTGAGTTTGGCTGTTCTTCTGTTCT TACTATTGAGAAGGATCTGCTCTAAATCTAAGCCTAGACCTTTGCCTCCGGGTCCTACTCCATG GCCTGTGGTCGGAAATCTATTGCAAATAGGCACAAATCCCCATATTTCGATCACTCAATTTTCT CAAACTTACGGTCCGTTGATTTCCTTGCGTTTGGGAACTAGCTTATTGGTCGTTGCATCGTCAC CAGCTGCTGCTACTGCCGTTCTTAGAACACATGATAGATTACTTAGTGCGAGATATATGTTCCA GACGATTCCTGACAAACGTAAACATGCCCAATTGTCCTTATCTACATCGCCATTCTGCGATGAC CATTGGAAGTCATTGAGAAGCATTTGTAGAGCAAACTTATTCACGTCCAAGGCTATAGAGTCAC AAGGAGGTCTTAGAAGAAGAAAGATGAAAGAGATGGTGGAATTTCTACAATCCAAACAAGGTAC GGTTGTAGGTGTTAGGGACTTAGTGTTTACCACCGTTTTCAACATCTTATCCAACTTGGTGTTC TCAAGAGACTTAGTTGGCTATGTAGGTGAAGGTTTCAATGGGATTAAGTCATCTTTTCACCGTT CTATGAAATTAGGGTTAACACCTAATCTGGCAGACTTTTATCCAATACTGGAAGGGTTCGATCT TCAAGGACTACAGAAGAAGGCTGTACTATATAACAAAGGAGTTGATTCTACATGGGAAATCCTA GTCAAAGAAAGGAGAGAATTACACAGGAACAACTTGGTAGTTTCACCGAATGACTTCTTGGATG TTTTGATACAGAATCAATTCAGTGATGATCAGATCAACTACTTGATTACCGAGGTTCTAACAGC TGGTATTGATACAACCACTTCTACCGTTGAATGGGCTATGGCGGAACTGTTAAAGAATAAGGAT TTAACTGAGAAAGTCAGGGTCGAATTGGAAAGAGAGATGAAAATCAAGGAAAATGCGATTGATG AGAGTCAGATTAGTCAATTTCAGTTTCTTCAACAGTGTGTCAAAGAAACTTTGAGACTTTATCC ACCAGTGCCATTTCTGTTACCAAGACTAGCACCAGAACCTTGTGAAGTGATGGGTTACAGTATT CCGAAAGATACCTCGATATTTGTTAACGCATGGGGCATTGGTAGAGATCCATCTATATGGGAGG AACCCTCAGCATTCAAACCAGAAAGATTTGTCAATTCAGACTTAGACTTTAAAGCCTATGATTA CAGATTCTTGCCTTTTGGTGGAGGCAGAAGATCTTGTCCAGGCCTTTTGATGACAACTGTACAA	881	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GTACCATTGATAATTGCCACGTTAATCCACAATTTTGACTGGAGCCTACCTAATGGCGGTGATT TGGCCCAATTGGATTTAAGCGGTCAAATGGGTGTATCCTTACAAAAGGAAAAGCCACTGTTGCT TATTCCCAGGAAACGTACTTAG		
Pathway step 3 sequence L CYP80C13 Protein	MLKDPFCFPFLPLLSLAVLLFLLLRRICSKSKPRPLPPGPTPWPVVGNLLQIGTNPHISITQFS QTYGPLISLRLGTSLLVVASSPAAATAVLRTHDRLLSARYMFQTIPDKRKHAQLSLSTSPFCDD HWKSLRSICRANLFTSKAIESQGGLRRRKMKEMVEFLQSKQGTVVGVRDLVFTTVFNILSNLVF SRDLVGYVGEGFNGIKSSFHRSMKLGLTPNLADFYPILEGFDLQGLQKKAVLYNKGVDSTWEIL VKERRELHRNNLVVSPNDFLDVLIQNQFSDDQINYLITEVLTAGIDTTTSTVEWAMAELLKNKD LTEKVRVELEREMKIKENAIDESQISQFQFLQQCVKETLRLYPPVPFLLPRLAPEPCEVMGYSI PKDTSIFVNAWGIGRDPSIWEEPSAFKPERFVNSDLDFKAYDYRFLPFGGGRRSCPGLLMTTVQ VPLIIATLIHNFDWSLPNGGDLAQLDLSGQMGVSLQKEKPLLLIPRKRT	882	
Pathway step 3 sequence M CYP92A127 (codon optimized)	ATGGAAGCTCCCTCGTGGGTGTCTTATGCCGCAGCTTGGGTTGCAACATTGGCTCTATTGTTAC TTAGTAGGCGTTTGAGAAGAAGAAAATTGAATTTGCCACCTGGACCTAAACCCTGGCCATTAAT TGGCAATTTAAACCTAATAGGTTCTTTACCGCATCAATCCATCCATCAATTGTCCCAAAAGTAT GGCCCAATAATGCACTTGAGATTTGGATCATTTCCTGTTGTAGTTGGCAGTTCTGTGGATATGG CCAAGATCTTCTTGAAAACTCAGGATCTAACCTTCGTTTCACGTCCAAAGACAGCAGCTGGCAA ATACACCACTTACAATTATAGCAATATAACGTGGTCACAATATGGTCCTTATTGGAGACAAGCG AGGAAAATGTGTTTGATGGAATTGTTCTCTGCTAGAAGATTGGACAGTTATGAATACATTAGGA AAGAAGAGATGAATGCCTTGCTTAAGGAAATTTGCAAAAGTTCGGGAAAAGTCATCAAACTAAA GGACTACCTATCTACAGTTTCCTTGAACGTGATAAGCAGGATGGTCTTAGGGAAGAAATACACT GACGAGTCAGAAGATGCAATCGTTAGTCCAGACGAATTTAAGAAAATGCTTGACGAATTGTTTC TTCTATCTGGTGTATTGAACATCGGTGATTCGATACCGTGGATTGATTTCTTAGATCTACAGGG TTACGTGAAACGTATGAAAGCTTTGTCCAAGAAATTCGACAGATTTCTGGAGCATGTTTTAGAC GAGCATAATGAGAGAAGAAAAGGTGTCAAAGATTATGTAGCTAAAGACATGGTCGATGTACTGT TACAACTGGCAGATGATCCGGATCTTGAGGTGAAATTGGAACGTCACGGTGTTAAGGCGTTCAC ACAAGACTTAATAGCCGGTGGTACAGAATCTTCCGCTGTCACTGTAGAATGGGCAATGAGCGAA CTTCTAAAGAAACCAGAGATGTTCGAAAAGGCCTCTGAAGAGTTAGATAGAGTGATTGGTAGGG AAAGATGGGTTGAGGAAAAGGATATCGCGAATTTACCCTATATTGACGCAATTGCTAAAGAAAC CATGAGGTTACATCCTGTGGCACCAATGTTGGTACCTAGATTATGCAGAGAAGATTGTCAGATT GCTGGCTACGATATAGCAAAGGGCACTAGAGTTCTTGTCAACGTTTGGACAATTGGAAGAGATC CAACTGTTTGGGAAAATCCGGATGAATTTAACCCAGAAAGATTTCTTGGGAAATCAATTGATGT CAAAGGGCAAGACTTTGAGTTGTTACCCTTTGGAAGTGGTAGAAGAATGTGTCCTGGATATTCA CTGGGTTTAAAAGTTATTCAGTCATCACTAGCCAACTTATTGCATGGGTTTTCCTGGAAGCTGG CTGGTGATACCAAGAAAGAAGATTTGAATATGGAAGAAGTATTCGGTTTAAGCACGCCAAAGAA GTTTCCTTTGGATGCTGTTGCCGAACCAAGACTGCCTCCACACCTGTATTCTATGTAG	883	
Pathway step 3 sequence N CYP92A127 Protein	MEAPSWVSYAAAWVATLALLLLSRRLRRRKLNLPPGPKPWPLIGNLNLIGSLPHQSIHQLSQKY GPIMHLRFGSFPVVVGSSVDMAKIFLKTQDLTFVSRPKTAAGKYTTYNYSNITWSQYGPYWRQA RKMCLMELFSARRLDSYEYIRKEEMNALLKEICKSSGKVIKLKDYLSTVSLNVISRMVLGKKYT	884	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	DESEDAIVSPDEFKKMLDELFLLSGVLNIGDSIPWIDFLDLQGYVKRMKALSKKFDRFLEHVLD EHNERRKGVKDYVAKDMVDVLLQLADDPDLEVKLERHGVKAFTQDLIAGGTESSAVTVEWAMSE LLKKPEMFEKASEELDRVIGRERWVEEKDIANLPYIDAIAKETMRLHPVAPMLVPRLCREDCQI AGYDIAKGTRVLVNVWTIGRDPTVWENPDEFNPERFLGKSIDVKGQDFELLPFGSGRRMCPGYS LGLKVIQSSLANLLHGFSWKLAGDTKKEDLNMEEVFGLSTPKKFPLDAVAEPRLPPHLYSM.		
Pathway step 3 sequence O CYP92A129 (codon optimized)	ATGGAGGCACCACCGTGGGTTTCATATGCAGCTGCGTGGGTAGCAACATTGGCTCTGTTACTTC TGTCTAGACATTTGCGTAGAAGAAAATTGAATTTACCACCTGGTCCAAAGCCTTGGCCTCTAAT TGGCAATCTGAACTTGATAGGATCGCTACCACATCAATCCATACATCAATTGAGTCAGAAATAT GGCCCAATTATGCAGTTAAGATTTGGTTCTTTTCCCGTTGTTGTTGGTTCAAGCGTAGATATGG CCAAAATTTTCCTGAAAACACACGATCTTACGTTTGTGAGCAGACCGAAAACTGCTGCAGGCAA ATACACCACGTATAACTGTTCCAATATAACTTGGTCGCAATATGGTCCGTATTGGAGACAAGCC AGGAAAATGTGTTTGATGGAGCTGTTTAGCGCTAGACGTCTGGATTCATACGAATACATCAGAA AAGAGGAAATGAATGCACTATTGAAGGAGATTTGCAAAAGTAGTGGGAAAGTAATCAAACTTAA AGACTATTTGTCTACTGTCTCGCTTAATGTCATCAGTAGAATGGTGCTAGGAAAGAAGTACACC GATGAGTCTGAAGATGCCATTGTTTCTCCCGATGAATTTAAGAAAATGTTGGATGAATTGTTTC TACTGGGCGGTGTTTTGAACATCGGTGATTCCATACCTTGGATCGACTTCTTAGATCTTCAAGG ATATGTCAAGAGAATGAAGGCTTTATCAAAGAAATTTGATCGTTTTCTAGAACACGTACTAGAT GAACACAACGAGCGTAGAAAAGGTGTGAAGGATTATGTTGCTAAGGACATGGTCGATGTGTTAT TGCAATTGGCTGACGATCCAGACTTGGAAGTCAGGTTAGAGAGGCATGGTGTTAAGGCGTTTAC CCAAGACTTGATTGCAGGAGGAACAGAATCATCCGCAGTAACAGTAGAATGGGCCATGTCTGAA TTGTTAAAGAAGCCCGAAATGTTCGAGAAAGCCTCAGAAGAGCTAGACAGAGTGATTGGTAGGG AAAGATGGGTTGAAGAGAAAGACATAGCCAATTTACCGTATATAGACGCCATCGCTAAAGAAAC CATGAGATTGCATCCAGTCGCACCTATGCTAGTTCCACGTTTATGCAGAGAAGATTGTCAGATT GCTGGATACGATATTGCTAAGGGTACTAGAGTCTTGGTGAACGTTTGGACAATTGGTAGGGATC CTACTGTATGGGAAAATCCTGATGAATTCAATCCCGAAAGATTCTTAGGGAAATCCATCGATGT CAAAGGTCAAGACTTCGAATTATTGCCATTCGGATCAGGCAGAAGAATGTGTCCAGGGTACTCC TTAGGCTTAAAGGTTATACAGAGTAGCTTAGCAAATCTTTTGCATGGTTTCTCTTGGAGACTTG CTGGGGACGTTAAGAAAGAAGATTTAAACATGGAAGAAGTGTTTGGTCTTTCTACTCCCAAGAA ATTTCCATTGGATGCGGTTGCTGAACCTAGGTTACCACCTCACTTGTACTCTATTTAG	885	
Pathway step 3 sequence P) CYP92A129 Protein	MEAPPWVSYAAAWVATLALLLLSRHLRRRKLNLPPGPKPWPLIGNLNLIGSLPHQSIHQLSQKY GPIMQLRFGSFPVVVGSSVDMAKIFLKTHDLTFVSRPKTAAGKYTTYNCSNITWSQYGPYWRQA RKMCLMELFSARRLDSYEYIRKEEMNALLKEICKSSGKVIKLKDYLSTVSLNVISRMVLGKKYT DESEDAIVSPDEFKKMLDELFLLGGVLNIGDSIPWIDFLDLQGYVKRMKALSKKFDRFLEHVLD EHNERRKGVKDYVAKDMVDVLLQLADDPDLEVRLERHGVKAFTQDLIAGGTESSAVTVEWAMSE LLKKPEMFEKASEELDRVIGRERWVEEKDIANLPYIDAIAKETMRLHPVAPMLVPRLCREDCQI AGYDIAKGTRVLVNVWTIGRDPTVWENPDEFNPERFLGKSIDVKGQDFELLPFGSGRRMCPGYS LGLKVIQSSLANLLHGFSWRLAGDVKKEDLNMEEVFGLSTPKKFPLDAVAEPRLPPHLYSI.	886	
Pathway step 3	ATGGAAATGTCATCATGTGTAGCCGCTACGATTAGCATCTGGATGGTGGTTGTTTGTATTGTGG	887	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
sequence Q CYP92A458 (codon optimized)	GTGTTGGATGGAGAGTGGTAAATTGGGTTTGGCTAAGACCCAAGAAATTGGAGAAAAGGTTAAG GGAACAAGGCTTGGCAGGGAACTCTTACAGATTGTTATTTGGTGACCTTAAAGAACGTGCAGCA ATGGCTGAACAAGCCAATTCAAAACCGATTAATTTTAGTCACGACATTGGTCCAAGAGTTTTCC CAAGTATGTACAAAACCATTCAGAATTATGGGAAGAATTCCTACATGTGGTTAGGTCCCTATCC AAGAGTGCATATAATGGATCCTCAACAGCTGAAAACCGTCTTTACATTGGTTTATGACATTCAA AAGCCGAATCTGAATCCACTGGTCAAATTCTTGTTAGATGGGATTGTCACTCATGAAGGAGAAA AGTGGGCAAAGCATAGAAAGATCATTAATCCAGCTTTTCACCTTGAAAAGTTGAAGGACATGAT TCCTGCCTTCTTTCACTCTTGCAATGAGATAGTTAATGAGTGGGAAAGACTAATTTCGAAGGAG GGTTCCTGTGAACTTGATGTTATGCCTTACTTGCAGAACTTAGCTGCTGATGCTATATCCAGAA CAGCGTTTGGTTCTAGCTATGAAGAGGGTAAAATGATATTCCAATTACTTAAGGAATTGACTGA TTTGGTCGTAAAAGTAGCGTTTGGTGTGTATATCCCTGGTTGGAGATTCTTACCAACCAAATCA AACAACAAAATGAAAGAAATCAACAGGAAAATCAAATCTCTGCTATTAGGAATCATTAACAAAC GTCAGAAAGCAATGGAAGAAGGCGAAGCTGGTCAATCTGATTTGTTAGGCATACTAATGGAATC GAATTCCAACGAAATTCAAGGAGAAGGAAACAATAAGGAGGACGGTATGTCTATAGAAGATGTA ATCGAGGAATGCAAGGTTTTCTATATAGGTGGACAAGAGACTACAGCCAGACTATTAATTTGGA CAATGATACTTTTAAGTTCACATACGGAATGGCAAGAGAGAGCAAGGACTGAAGTCTTGAAAGT CTTTGGCAATAAGAAGCCTGATTTTGATGGCTTGAACAGATTGAAAATCGTTAGTGAAATTCTA TAG		
Pathway step 3 sequence R CYP92A458 Protein	MEMSSCVAATISIWMVVVCIVGVGWRVVNWVWLRPKKLEKRLREQGLAGNSYRLLFGDLKERAA MAEQANSKPINFSHDIGPRVFPSMYKTIQNYGKNSYMWLGPYPRVHIMDPQQLKTVFTLVYDIQ KPNLNPLVKFLLDGIVTHEGEKWAKHRKIINPAFHLEKLKDMIPAFFHSCNEIVNEWERLISKE GSCELDVMPYLQNLAADAISRTAFGSSYEEGKMIFQLLKELTDLVVKVAFGVYIPGWRFLPTKS NNKMKEINRKIKSLLLGIINKRQKAMEEGEAGQSDLLGILMESNSNEIQGEGNNKEDGMSIEDV IEECKVFYIGGQETTARLLIWTMILLSSHTEWQERARTEVLKVFGNKKPDFDGLNRLKIVSEIL	888	
Pathway step 3 sequence S CYP88D6 Protein	MEVHWVCMCAATLLVCYIFGSKFVRNLNGWYYDVKLRRKEHPLPPGDMGWPLMGNLLSFIKDFS SGHPDSFINNLVLKYGRSGIYKTHLFGNPSIIVCEPQMCRRVLTDDVNFKLGYPKSIKELARCR PMIDVSNAEHRLFRRLITSPIVGHKALAMYLERLEEIVINSLEELSSMKHPVELLKEMKKVSFK AIVHVFMGSSNQDIIKKIGSSFTDLYNGMFSIPINVPGFTFHKALEARKKLAKIVQPVVDERRL MIENGQQEGDQRKDLIDILLEVKDENGRKLEDEDISDLLIGLLFAGHESTATSLMWSITYLTQH PHILKKAKEEQEEIMRTRLSSQKQLSFKEIKQMVYLSQVIDETLRCANIAFATFREATADVNIN GYIIPKGWRVLIWARAIHMDSEYYPNPEEFNPSRWDDYNAKAGTFLPFGAGSRLCPGADLAKLE ISIFLHYFLLNYRLERVNPECHVTSLPVSKPTDNCLAKVMKVSCA.	889	
Pathway step 3 sequence T CYP88D6 (codon optimized)	ATGGAAGTACATTGGGTTTGCATGTGCGCTGCCACTTTGTTGGTATGCTACATTTTTGGAAGCA AGTTTGTGAGGAATTTGAATGGGTGGTATTATGATGTAAAACTAAGAAGGAAAGAACACCCACT ACCCCCAGGTGACATGGGATGGCCTCTTATGGGCAATCTATTGTCCTTCATCAAAGATTTCTCA TCGGGTCACCCTGATTCATTCATCAACAACCTTGTTCTCAAATATGGACGAAGTGGTATCTACA AGACTCACTTGTTTGGGAATCCAAGCATCATTGTTTGCGAGCCTCAGATGTGTAGGCGAGTTCT CACTGATGATGTGAACTTTAAGCTTGGTTATCCAAAATCTATCAAAGAGTTGGCACGATGTAGA	890	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CCCATGATTGATGTCTCTAATGCGGAACATAGGCTTTTTCGACGCCTCATTACTTCCCCAATCG TGGGTCACAAGGCGCTAGCAATGTACCTAGAACGTCTTGAGGAAATTGTGATCAATTCGTTGGA AGAATTGTCCAGCATGAAGCACCCCGTTGAGCTCTTGAAAGAGATGAAGAAGGTTTCCTTTAAA GCCATTGTCCACGTTTTCATGGGCTCTTCCAATCAGGACATCATTAAAAAAATTGGAAGTTCGT TTACTGATTTGTACAATGGCATGTTCTCTATCCCCATTAACGTACCTGGTTTTACATTCCACAA AGCACTCGAGGCACGTAAGAAGCTAGCCAAAATAGTTCAACCCGTTGTGGATGAAAGGCGGTTG ATGATAGAAAATGGTCAACAAGAAGGGGACCAAAGAAAAGATCTTATTGATATTCTTTTGGAAG TCAAAGATGAGAATGGACGAAAATTGGAGGACGAGGATATTAGCGATTTATTAATAGGGCTTTT GTTTGCTGGCCATGAAAGTACAGCAACCAGTTTAATGTGGTCAATTACATATCTTACACAGCAT CCCCATATCTTGAAAAAGGCTAAGGAAGAGCAGGAAGAAATAATGAGGACAAGATTGTCCTCGC AGAAACAATTAAGTTTTAAGGAAATTAAACAAATGGTTTATCTTTCTCAGGTAATTGATGAAAC TTTACGATGTGCCAATATTGCCTTTGCAACTTTTCGAGAGGCAACTGCTGATGTGAACATCAAT GGTTATATCATACCAAAGGGATGGAGAGTGCTAATTTGGGCAAGAGCCATTCATATGGATTCTG AATATTACCCAAATCCAGAAGAATTTAATCCATCGAGATGGGATGATTACAATGCCAAAGCAGG AACCTTCCTTCCTTTTGGAGCAGGAAGTAGACTTTGTCCTGGAGCCGACTTGGCGAAACTTGAA ATTTCCATATTTCTTCATTATTTCCTCCTTAATTACAGGTTGGAGCGAGTAAATCCAGAATGTC ATGTTACCAGCTTACCAGTATCTAAGCCCACAGACAATTGCCTCGCTAAGGTGATGAAGGTCTC ATGTGCTTAG		
Pathway step 3 sequence U CYP1798 (codon optimized)	ATGGAAATGTCCTCTTCTGTTGCTGCCACCATTTCTATTTGGATGGTTGTTGTATGTATCGTTG GTGTTGGTTGGAGAGTTGTTAATTGGGTTTGGTTAAGACCAAAGAAGTTGGAAAAGAGATTGAG AGAACAAGGTTTGGCTGGTAACTCTTACAGATTGTTGTTCGGTGACTTGAAAGAAAGAGCTGCT ATGGAAGAACAAGCTAACTCTAAGCCAATCAACTTCTCCCATGATATTGGTCCAAGAGTTTTCC CATCTATGTACAAGACCATTCAAAACTACGGTAAGAACTCCTATATGTGGTTGGGTCCATACCC AAGAGTTCATATTATGGATCCACAACAATTGAAAACCGTCTTTACCTTGGTTTACGACATCCAA AAGCCAAACTTGAACCCATTGATCAAGTTCTTGTTGGATGGTATTGTCACCCATGAAGGTGAAA AATGGGCTAAACATAGAAAGATTATCAACCCAGCCTTCCACTTGGAAAAGTTGAAAGATATGAT TCCAGCCTTCTTCCACTCTTGCAACGAAATAGTTAATGAATGGGAAAGATTGATCTCCAAAGAA GGTTCTTGCGAATTGGATGTTATGCCATACTTGCAAAATTTGGCTGCTGATGCTATTTCTAGAA CTGCTTTTGGTTCCTCTTACGAAGAAGGTAAGATGATCTTCCAATTATTGAAAGAATTGACCGA CTTGGTTGTTAAGGTTGCTTTCGGTGTTTACATTCCAGGTTGGAGATTTTTGCCAACTAAGTCC AACAACAAGATGAAGGAAATCAACAGAAAGATCAAGTCTTTGTTGTTAGGTATCATCAACAAGA GACAAAAGGCCATGGAAGAAGGTGAAGCTGGTCAATCTGATTTGTTGGGTATTTTGATGGAATC CAACTCCAACGAAATTCAAGGTGAAGGTAACAACAAAGAAGATGGTATGTCCATCGAAGATGTT ATCGAAGAATGCAAGGTTTTCTACATCGGTGGTCAAGAAACTACCGCCAGATTATTGATTTGGA CCATGATCTTGTTGAGTTCCCATACTGAATGGCAAGAAAGAGCAAGAACTGAAGTCTTGAAGGT TTTCGGTAACAAAAAGCCAGATTTCGACGGTTTGTCTAGATTGAAGGTTGTCACCATGATTTTG AACGAAGTTTTGAGATTATACCCACCAGCTTCTATGTTGACCAGAATCATTCAAAAAGAAACCA GAGTCGGTAAGTTGACTTTGCCAGCTGGTGTTATTTTGATCATGCCAATCATCTTGATCCACAG	891	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AGATCATGATTTGTGGGGTGAAGATGCTAATGAATTCAAGCCAGAAAGATTCTCCAAGGGTGTT TCTAAAGCTGCTAAAGTTCAACCAGCTTTCTTTCCATTTGGTTGGGGTCCAAGAATATGTATGG GTCAAAATTTCGCTATGATCGAAGCTAAGATGGCCTTGTCTTTGATCTTGCAAAGATTTTCCTT CGAATTGTCCTCCTCATATGTTCATGCTCCAACTGTTGTTTTCACCACTCAACCACAACATGGT GCTCATATCGTTTTGAGAAAGTTGTAA		
Pathway step 3 sequence V CYP1798 Protein	MEMSSSVAATISIWMVVVCIVGVGWRVVNWVWLRPKKLEKRLREQGLAGNSYRLLFGDLKERAA MEEQANSKPINFSHDIGPRVFPSMYKTIQNYGKNSYMWLGPYPRVHIMDPQQLKTVFTLVYDIQ KPNLNPLIKFLLDGIVTHEGEKWAKHRKIINPAFHLEKLKDMIPAFFHSCNEIVNEWERLISKE GSCELDVMPYLQNLAADAISRTAFGSSYEEGKMIFQLLKELTDLVVKVAFGVYIPGWRFLPTKS NNKMKEINRKIKSLLLGIINKRQKAMEEGEAGQSDLLGILMESNSNEIQGEGNNKEDGMSIEDV IEECKVFYIGGQETTARLLIWTMILLSSHTEWQERARTEVLKVFGNKKPDFDGLSRLKVVTMIL NEVLRLYPPASMLTRIIQKETRVGKLTLPAGVILIMPIILIHRDHDLWGEDANEFKPERFSKGV SKAAKVQPAFFPFGWGPRICMGQNFAMIEAKMALSLILQRFSFELSSSYVHAPTVVFTTQPQHG AHIVLRKL.	892	
Pathway step 3 sequence W EPH2A (codon optimized)	ATGGATGAAATCGAACATATTACCATCAATACAAATGGAATCAAAATGCATATTGCGTCAGTCG GCACAGGACCAGTTGTTCTCTTGCTACACGGCTTTCCAGAATTATGGTACTCTTGGAGACACCA ACTACTTTACCTGTCCTCCGTTGGGTACAGAGCAATAGCTCCAGATTTGAGAGGCTATGGCGAT ACTGACAGTCCAGCTAGTCCTACCTCTTATACTGCTCTTCATATTGTAGGTGACCTGGTCGGCG CATTAGACGAATTGGGAATAGAAAAGGTCTTTTTAGTGGGTCATGACTGGGGTGCTATTATCGC ATGGTACTTTTGTTTGTTTAGACCAGATAGAATTAAAGCACTTGTGAATTTGTCTGTCCAGTTT ATCCCACGTAACCCAGCAATACCTTTTATAGAAGGTTTCAGAACAGCTTTTGGTGATGACTTCT ACATTTGTAGATTTCAAGTACCTGGGGAAGCTGAAGAGGATTTCGCGTCTATCGATACTGCTCA ATTGTTTAAAACTTCATTATGCAATAGAAGCTCAGCCCCTCCTTGTTTGCCTAAAGAGATTGGT TTTAGGGCTATCCCACCACCAGAAAATCTGCCATCTTGGCTCACAGAGGAAGATATCAACTTCT ACGCAGCCAAGTTTAAACAAACTGGTTTTACTGGTGCCCTTAACTATTATAGAGCATTCGACTT GACATGGGAATTAACAGCCCCATGGACAGGAGCCCAGATCCAAGTTCCTGTAAAGTTCATAGTT GGTGATTCAGATCTCACGTACCATTTCCCTGGTGCTAAGGAATACATCCACAACGGAGGGTTTA AAAGAGATGTGCCACTATTAGAGGAAGTTGTTGTGGTAAAAGATGCCTGCCACTTCATTAACCA AGAGCGACCACAAGAGATTAATGCTCATATTCATGACTTCATCAATAAGTTCTAA	893	
Pathway step 3 sequence X EPH2A Protein	MDEIEHITINTNGIKMHIASVGTGPVVLLLHGFPELWYSWRHQLLYLSSVGYRAIAPDLRGYGD TDSPASPTSYTALHIVGDLVGALDELGIEKVFLVGHDWGAIIAWYFCLFRPDRIKALVNLSVQF IPRNPAIPFIEGFRTAFGDDFYICRFQVPGEAEEDFASIDTAQLFKTSLCNRSSAPPCLPKEIG FRAIPPPENLPSWLTEEDINFYAAKFKQTGFTGALNYYRAFDLTWELTAPWTGAQIQVPVKFIV GDSDLTYHFPGAKEYIHNGGFKRDVPLLEEVVVVKDACHFINQERPQEINAHIHDFINKF	894	
Pathway step 3 sequence Y tDexT DNA (native DNA sequence)	ATGCCAGCAAATGCCCCAGATAAACAATCAGTGACTAATGCACCAGTAGTGCCGCCAAAGCATG ATACGGACCAGCAGGACGATTCACTAGAAAAACAGCAAGTATTAGAACCGAGCGTAAATAGTAA TATACCAAAAAAGCAGACAAATCAACAGTTAGCGGTTGTTACAGCACCAGCAAATTCAGCACCT CAAACCAAAACAACAGCAGAAATTTCTGCTGGTACAGAGTTAGACACGATGCCTAATGTTAAGC	895	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ATGTAGATGGCAAAGTTTATTTTTATGGAGATGATGGCCAACCAAAAAAGAATTTTACTACTAT TATAGATGGTAAACCTTACTACTTTGATAAAGATACAGGGGCACTATCTAATAACGATAAGCAA TATGTATCGGAATTATTCAGTATTGGCAATAAACATAACGCCGTCTATAACACATCATCAGATA ATTTTACGCAATTAGAAGGACATCTGACGGCAAGTAGTTGGTATCGTCCAAAAGATATTTTGAA AAATGGTAAACGTTGGGCACCTTCAACAGTGACTGATTTCAGACCATTATTGATGGCCTGGTGG CCGGATAAGAGTACGCAAGTCACTTATCTGAATTACATGAAAGATCAGGGCCTCTTGTCTGGTA CTCATCACTTTTCCGATAATGAAAATATGCGGACCTTAACGGCAGCTGCCATGCAGGCACAGGT AAACATTGAGAAAAAAATTGGGCAACTTGGCAATACGGATTGGTTGAAAACGGCGATGACGCAA TACATTGATGCCCAGCCCAATTGGAATATTGACAGTGAGGCGAAAGGAGATGATCATCTACAAG GTGGTGCACTACTTTATACAAATAGTGATATGTCGCCAAAGGCCAATTCTGATTATCGTAAGCT GAGCCGTACGCCTAAAAATCAAAAAGGTCAAATTGCTGATAAATATAAGCAAGGTGGGTTTGAA TTATTACTAGCAAACGATGTCGATAATTCTAATCCAGTTGTGCAAGCAGAACAACTTAATTGGT TACATTATATGATGAATATCGGTAGTATTTTACAAAATGATGACCAAGCTAATTTTGATGGTTA CCGTGTTGATGCTGTCGATAATGTGGACGCTGACTTACTACAGATTGCTGGTGAATATGCTAAG GCTGCCTATGGTGTTGACAAAAATGACGCGAGAGCGAATCAACATTTATCAATTTTGGAAGACT GGGGAGATGAAGATCCAGACTATGTCAAAGCACATGGCAACCAGCAAATTACAATGGATTTCCC CTTGCATTTAGCGATTAAATACGCGCTCAACATGCCTAATGATAAGCGGAGTGGCCTTGAGCCA ACCCGTGAACACAGTTTAGTCAAACGAATTACAGATGATAAAGAAAATGTTGCACAACCAAATT ATTCATTTATCCGAGCTCATGACAGTGAAGTACAAACGATTATTGCTGATATTATTAAAGATAA AATCAACCCGGCGTCAACAGGGCTAGATTCAACAGTGACTTTGGATCAAATTAAGCAGGCTTTT GACATCTATAATGCTGATGAATTGAAAGCAGATAAAGTTTACACACCTTACAATATTCCAGCAT CATACGCTTTGTTATTGACTAATAAAGACACAATTCCACGTGTTTATTATGGGGATATGTTCAC GGATGATGGCCAATACATGGCTAAACAATCACCTTACTATCAAGCGATTGATGCGTTGTTGAAA GCTCGTATCAAGTATGCTGCTGGTGGTCAAACCATGAAAATGAACTATTTTCCAGATGAACAAT CTGTTATGACATCAGTTCGTTATGGTAAGGGTGCAATGACGGCAAGTGACTCTGGTAACCAAGA GACACGCTATCAAGGTATTGGACTTGTTGTCAACAATCGCCCAGATTTGAAACTATCTGACAAA GATGAAGTCAAAATGGATATGGGTGCGGCACATAAAAACCAAGATTATCGCCCAGTTTTGTTGA CGACAAAATCAGGATTAAAAGTCTACAGCACTGATGCAAATGCACCTGTCGTTCGAACTGACGC CAATGGCCAATTAACTTTTAAGGCAGACATGGTATATGGTGTAAACGACCCACAAGTGTCAGGG TACATTGCGGCTTGGGTACCAGTAGGGGCTTCAGAAAATCAAGATGCTCGAACGAAAAGTGAAA CAACGCAGTCAACTGACGGGAGTGTTTATCATTCTAATGCAGCGTTAGATTCGCAAGTCATTTA TGAAGGCTTTTCAAATTTTCAAGACTTTCCAACAACACCCGATGAGTTTACGAACATTAAAATT GCTCAAAATGTTAACTTATTTAAGGATTGGGGTATTACTAGCTTTGAAATGGCGCCACAATATC GCGCCAGCTCAGATAAAAGTTTCTTAGATGCTATCGTACAAAATGGTTATGCATTTACAGATCG ATATGATATTGGTTACAACACACCAACAAAGTATGGGACAGCAGATAATTTGTTAGATGCTTTA CGTGCATTGCATGGTCAGGGTATTCAAGCGATTAACGACTGGGTACCAGATCAAATTTATAATC TACCCGATGAACAGTTAGTCACGGCTATTCGAACAGACGGTTCAGGTGATCATACTTATGGTTC AGTTATTGACCATACTTTGTATGCATCAAAGACAGTTGGCGGGGGCATTTATCAGCAACAATAT		

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GGTGGGGCCTTCTTGGAACAATTAAAAACACAGTACCCGCAACTTTTCCAGCAAAAACAGATTT CCACAGATCAGCCAATGAACCCAGATATTCAAATTAAGTCATGGGAAGCCAAGTATTTCAACGG TTCGAACATTCAGGGGCGTGGGGCTTGGTATGTTTTGAAGGACTGGGGCACACAACAGTATTTT AATGTGTCAGATGCGCAGACCTTCCTTCCAAAGCAATTATTGGGTGAAAAGGCCAAAACTGGTT TTGTTACGCGTGGTAAGGAGACTTCATTCTATTCCACTAGTGGCTATCAAGCAAAATCTGCCTT TATTTGTGATAACGGTAATTGGTACTACTTTGATGACAAAGGGAAAATGGTTGTTGGAAACCAA GTTATCAATGGCATCAATTATTACTTTTTACCGAATGGTATCGAATTACAAGATGCCTATCTAG TACATGATGGTATGTACTATTATTATAATAATATTGGCAAGCAACTGCACAACACATATTACCA AGATAAACAAAAAAATTTCCATTACTTCTTTGAAGATGGGCACATGGCACAGGGTATTGTCACC ATCATTCAAAGTGATGGCACCCCAGTCACACAGTACTTTGATGAGAATGGTAAGCAACAAAAAG GCGTGGCGGTCAAAGGATCAGATGGTCATTTGCATTACTTTGACGGTGCGTCAGGGAATATGCT CTTTAAATCATGGGGTAGACTAGCAGATGGCTCTTGGCTATATGTAGACGAGAAAGGTAATGCG GTTACAGGCAAACAAACCATTAATAATCAAACGGTTTACTTTAATGATGATGGTCGTCAAATCA AAAATAACTTTAAAGAATTAGCAGATGGTTCTTGGCTTTATCTTAACAATAAAGGTGTTGCAGT AACAGGAGAGCAAATAATTAATGGGCAGACACTTTATTTTGGTAACGATGGTCGTCAATTTAAA GGGACAACACATATAAATGCTACTGGTGAAAGCCGTTACTATGACCCAGACTCAGGTAATATGA TAACTGATCGTTTTGAACGTGTTGGTGATAATCAATGGGCTTATTTTGGTTATGATGGTGTTGC AGTAACAGGGGACCGAATCATTAAAGGGCAAAAACTCTATTTCAACCAAAATGGTATCCAAATG AAAGGCCACTTACGTCTTGAAAATGGTATCATGCGTTATTACGATGCTGATACTGGCGAATTAG TTCGTAATCGATTTGTATTGCTATCTGATGGTTCATGGGTTTACTTTGGCCAAGATGGCGTACC CGTAACTGGCGTGCAAGTGATTAATGGCCAAACATTATATTTTGACGCAGATGGTAGGCAAGTC AAAGGGCAGCAACGTGTAATCGGCAATCAACGCTATTGGATGGATAAAGACAATGGTGAAATGA AAAAAATAACATACTAG		
Pathway step 3 sequence Z tDexT Protein	MPANAPDKQSVTNAPVVPPKHDTDQQDDSLEKQQVLEPSVNSNIPKKQTNQQLAVVTAPANSAP QTKTTAEISAGTELDTMPNVKHVDGKVYFYGDDGQPKKNFTTIIDGKPYYFDKDTGALSNNDKQ YVSELFSIGNKHNAVYNTSSDNFTQLEGHLTASSWYRPKDILKNGKRWAPSTVTDFRPLLMAWW PDKSTQVTYLNYMKDQGLLSGTHHFSDNENMRTLTAAAMQAQVNIEKKIGQLGNTDWLKTAMTQ YIDAQPNWNIDSEAKGDDHLQGGALLYTNSDMSPKANSDYRKLSRTPKNQKGQIADKYKQGGFE LLLANDVDNSNPVVQAEQLNWLHYMMNIGSILQNDDQANFDGYRVDAVDNVDADLLQIAGEYAK AAYGVDKNDARANQHLSILEDWGDEDPDYVKAHGNQQITMDFPLHLAIKYALNMPNDKRSGLEP TREHSLVKRITDDKENVAQPNYSFIRAHDSEVQTIIADIIKDKINPASTGLDSTVTLDQIKQAF DIYNADELKADKVYTPYNIPASYALLLTNKDTIPRVYYGDMFTDDGQYMAKQSPYYQAIDALLK ARIKYAAGGQTMKMNYFPDEQSVMTSVRYGKGAMTASDSGNQETRYQGIGLVVNNRPDLKLSDK DEVKMDMGAAHKNQDYRPVLLTTKSGLKVYSTDANAPVVRTDANGQLTFKADMVYGVNDPQVSG YIAAWVPVGASENQDARTKSETTQSTDGSVYHSNAALDSQVIYEGFSNFQDFPTTPDEFTNIKI AQNVNLFKDWGITSFEMAPQYRASSDKSFLDAIVQNGYAFTDRYDIGYNTPTKYGTADNLLDAL RALHGQGIQAINDWVPDQIYNLPDEQLVTAIRTDGSGDHTYGSVIDHTLYASKTVGGGIYQQQY GGAFLEQLKTQYPQLFQQKQISTDQPMNPDIQIKSWEAKYFNGSNIQGRGAWYVLKDWGTQQYF	896	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	NVSDAQTFLPKQLLGEKAKTGFVTRGKETSFYSTSGYQAKSAFICDNGNWYYFDDKGKMVVGNQ VINGINYYFLPNGIELQDAYLVHDGMYYYYNNIGKQLHNTYYQDKQKNFHYFFEDGHMAQGIVT IIQSDGTPVTQYFDENGKQQKGVAVKGSDGHLHYFDGASGNMLFKSWGRLADGSWLYVDEKGNA VTGKQTINNQTVYFNDDGRQIKNNFKELADGSWLYLNNKGVAVTGEQIINGQTLYFGNDGRQFK GTTHINATGESRYYDPDSGNMITDRFERVGDNQWAYFGYDGVAVTGDRIIKGQKLYFNQNGIQM KGHLRLENGIMRYYDADTGELVRNRFVLLSDGSWVYFGQDGVPVTGVQVINGQTLYFDADGRQV KGQQRVIGNQRYWMDKDNGEMKKITY		
Pathway 1 sequence id A tHMG-CoA DNA	ATGGCAGCTGACCAATTGGTGAAAACTGAAGTCACCAAGAAGTCTTTTACTGCTCCTGTACAAA AGGCTTCTACACCAGTTTTAACCAATAAAACAGTCATTTCTGGATCGAAAGTCAAAAGTTTATC ATCTGCGCAATCGAGCTCATCAGGACCTTCATCATCTAGTGAGGAAGATGATTCCCGCGATATT GAAAGCTTGGATAAGAAAATACGTCCTTTAGAAGAATTAGAAGCATTATTAAGTAGTGGAAATA CAAAACAATTGAAGAACAAAGAGGTCGCTGCCTTGGTTATTCACGGTAAGTTACCTTTGTACGC TTTGGAGAAAAAATTAGGTGATACTACGAGAGCGGTTGCGGTACGTAGGAAGGCTCTTTCAATT TTGGCAGAAGCTCCTGTATTAGCATCTGATCGTTTACCATATAAAAATTATGACTACGACCGCG TATTTGGCGCTTGTTGTGAAAATGTTATAGGTTACATGCCTTTGCCCGTTGGTGTTATAGGCCC CTTGGTTATCGATGGTACATCTTATCATATACCAATGGCAACTACAGAGGGTTGTTTGGTAGCT TCTGCCATGCGTGGCTGTAAGGCAATCAATGCTGGCGGTGGTGCAACAACTGTTTTAACTAAGG ATGGTATGACAAGAGGCCCAGTAGTCCGTTTCCCAACTTTGAAAAGATCTGGTGCCTGTAAGAT ATGGTTAGACTCAGAAGAGGGACAAAACGCAATTAAAAAAGCTTTTAACTCTACATCAAGATTT GCACGTCTGCAACATATTCAAACTTGTCTAGCAGGAGATTTACTCTTCATGAGATTTAGAACAA CTACTGGTGACGCAATGGGTATGAATATGATTTCTAAAGGTGTCGAATACTCATTAAAGCAAAT GGTAGAAGAGTATGGCTGGGAAGATATGGAGGTTGTCTCCGTTTCTGGTAACTACTGTACCGAC AAAAAACCAGCTGCCATCAACTGGATCGAAGGTCGTGGTAAGAGTGTCGTCGCAGAAGCTACTA TTCCTGGTGATGTTGTCAGAAAAGTGTTAAAAAGTGATGTTTCCGCATTGGTTGAGTTGAACAT TGCTAAGAATTTGGTTGGATCTGCAATGGCTGGGTCTGTTGGTGGATTTAACGCACATGCAGCT AATTTAGTGACAGCTGTTTTCTTGGCATTAGGACAAGATCCTGCACAAAATGTTGAAAGTTCCA ACTGTATAACATTGATGAAAGAAGTGGACGGTGATTTGAGAATTTCCGTATCCATGCCATCCAT CGAAGTAGGTACCATCGGTGGTGGTACTGTTCTAGAACCACAAGGTGCCATGTTGGACTTATTA GGTGTAAGAGGCCCGCATGCTACCGCTCCTGGTACCAACGCACGTCAATTAGCAAGAATAGTTG CCTGTGCCGTCTTGGCAGGTGAATTATCCTTATGTGCTGCCCTAGCAGCCGGCCATTTGGTTCA AAGTCATATGACCCACAACAGGAAACCTGCTGAACCAACAAAACCTAACAATTTGGACGCCACT GATATAAATCGTTTGAAAGATGGGTCCGTCACCTGCATTAAATCCTAA	897	
Pathway 1 sequence id B tHMG-CoA Protein	MAADQLVKTEVTKKSFTAPVQKASTPVLTNKTVISGSKVKSLSSAQSSSSGPSSSSEEDDSRDI ESLDKKIRPLEELEALLSSGNTKQLKNKEVAALVIHGKLPLYALEKKLGDTTRAVAVRRKALSI LAEAPVLASDRLPYKNYDYDRVFGACCENVIGYMPLPVGVIGPLVIDGTSYHIPMATTEGCLVA SAMRGCKAINAGGGATTVLTKDGMTRGPVVRFPTLKRSGACKIWLDSEEGQNAIKKAFNSTSRF ARLQHIQTCLAGDLLFMRFRTTTGDAMGMNMISKGVEYSLKQMVEEYGWEDMEVVSVSGNYCTD KKPAAINWIEGRGKSVVAEATIPGDVVRKVLKSDVSALVELNIAKNLVGSAMAGSVGGFNAHAA	898	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	NLVTAVFLALGQDPAQNVESSNCITLMKEVDGDLRISVSMPSIEVGTIGGGTVLEPQGAMLDLL GVRGPHATAPGTNARQLARIVACAVLAGELSLCAALAAGHLVQSHMTHNRKPAEPTKPNNLDAT DINRLKDGSVTCIKS		
Pathway 1 sequence id C erg1 DNA	ATGTCTGCTGTTAACGTTGCACCTGAATTGATTAATGCCGACAACACAATTACCTACGATGCGA TTGTCATCGGTGCTGGTGTTATCGGTCCATGTGTTGCTACTGGTCTAGCAAGAAAGGGTAAGAA AGTTCTTATCGTAGAACGTGACTGGGCTATGCCTGATAGAATTGTTGGTGAATTGATGCAACCA GGTGGTGTTAGAGCATTGAGAAGTCTGGGTATGATTCAATCTATCAACAACATCGAAGCATATC CTGTTACCGGTTATACCGTCTTTTTCAACGGCGAACAAGTTGATATTCCATACCCTTACAAGGC CGATATCCCTAAAGTTGAAAAATTGAAGGACTTGGTCAAAGATGGTAATGACAAGGTCTTGGAA GACAGCACTATTCACATCAAGGATTACGAAGATGATGAAAGAGAAAGGGGTGTTGCTTTTGTTC ATGGTAGATTCTTGAACAACTTGAGAAACATTACTGCTCAAGAGCCAAATGTTACTAGAGTGCA AGGTAACTGTATTGAGATATTGAAGGATGAAAAGAATGAGGTTGTTGGTGCCAAGGTTGACATT GATGGCCGTGGCAAGGTGGAATTCAAAGCCCACTTGACATTTATCTGTGACGGTATCTTTTCAC GTTTCAGAAAGGAATTGCACCCAGACCATGTTCCAACTGTCGGTTCTTCGTTTGTCGGTATGTC TTTGTTCAATGCTAAGAATCCTGCTCCTATGCACGGTCACGTTATTCTTGGTAGTGATCATATG CCAATCTTGGTTTACCAAATCAGTCCAGAAGAAACAAGAATCCTTTGTGCTTACAACTCTCCAA AGGTCCCAGCTGATATCAAGAGTTGGATGATTAAGGATGTCCAACCTTTCATTCCAAAGAGTCT ACGTCCTTCATTTGATGAAGCCGTCAGCCAAGGTAAATTTAGAGCTATGCCAAACTCCTACTTG CCAGCTAGACAAAACGACGTCACTGGTATGTGTGTTATCGGTGACGCTCTAAATATGAGACATC CATTGACTGGTGGTGGTATGACTGTCGGTTTGCATGATGTTGTCTTGTTGATTAAGAAAATAGG TGACCTAGACTTCAGCGACCGTGAAAAGGTTTTGGATGAATTACTAGACTACCATTTCGAAAGA AAGAGTTACGATTCCGTTATTAACGTTTTGTCAGTGGCTTTGTATTCTTTGTTCGCTGCTGACA GCGATAACTTGAAGGCATTACAAAAAGGTTGTTTCAAATATTTCCAAAGAGGTGGCGATTGTGT CAACAAACCCGTTGAATTTCTGTCTGGTGTCTTGCCAAAGCCTTTGCAATTGACCAGGGTTTTC TTCGCTGTCGCTTTTTACACCATTTACTTGAACATGGAAGAACGTGGTTTCTTGGGATTACCAA TGGCTTTATTGGAAGGTATTATGATTTTGATCACAGCTATTAGAGTATTCACCCCATTTTTGTT TGGTGAGTTGATTGGTTAA	899	
Pathway 1 sequence id D erg1 protein	MSAVNVAPELINADNTITYDAIVIGAGVIGPCVATGLARKGKKVLIVERDWAMPDRIVGELMQP GGVRALRSLGMIQSINNIEAYPVTGYTVFFNGEQVDIPYPYKADIPKVEKLKDLVKDGNDKVLE DSTIHIKDYEDDERERGVAFVHGRFLNNLRNITAQEPNVTRVQGNCIEILKDEKNEVVGAKVDI DGRGKVEFKAHLTFICDGIFSRFRKELHPDHVPTVGSSFVGMSLFNAKNPAPMHGHVILGSDHM PILVYQISPEETRILCAYNSPKVPADIKSWMIKDVQPFIPKSLRPSFDEAVSQGKFRAMPNSYL PARQNDVTGMCVIGDALNMRHPLTGGGMTVGLHDVVLLIKKIGDLDFSDREKVLDELLDYHFER KSYDSVINVLSVALYSLFAADSDNLKALQKGCFKYFQRGGDCVNKPVEFLSGVLPKPLQLTRVF FAVAFYTIYLNMEERGFLGLPMALLEGIMILITAIRVFTPFLFGELIG	900	
Pathway 2 sequence id E Cmelo DNA	ATGTGGAGATTAAAAGTGGGAAAAGAGAGTGTTGGGGAAAAAGAAGAGAAATGGATTAAGAGTA TAAGCAATCACTTGGGACGTCAAGTTTGGGAATTTTGCAGTGGTGAAAATGAAAATGATGATGA TGAAGCCATTGCTGTTGCTAATAATTCTGCTTCAAAGTTCGAGAATGCCAGGAATCACTTTCGT	901	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AATAATCGTTTCCATCGCAAGCAATCTTCCGACCTCTTTCTTGCCATTCAGTGTGAAAAGGAAA TAATAAGAAACGGTGCAAAAAATGAAGGAACCACCAAAGTAAAAGAAGGGGAAGATGTGAAGAA AGAAGCAGTGAAGAATACATTAGAAAGAGCATTAAGTTTCTATTCGGCTGTTCAAACAAGCGAT GGGAATTGGGCTTCGGATCTTGGCGGGCCTATGTTTTTACTACCGGGTTTAGTGATTGCTCTAT ATGTCACTGGAGTCTTGAATTCTGTTCTGTCCAAGCACCATCGCCAAGAAATGTGTAGATATAT TTACAATCATCAGAATGAAGATGGGGGATGGGGTTTGCACATTGAAGGTTCGAGCACGATGTTT GGTTCGGCACTGAATTATGTTGCACTGAGACTGCTTGGAGAGGCTGCCGATGGCGGAGAGCACG GCGCAATGACAAAAGCTCGAAGTTGGATCTTGGAGCGTGGTGGAGCTACCGCAATCACTTCTTG GGGAAAATTGTGGCTGTCAGTACTTGGAGTCTATGAATGGAGTGGCAACAATCCTCTCCCACCT GAATTTTGGTTACTCCCATATAGCCTACCATTTCATCCTGGAAGAATGTGGTGCCATTGTCGAA TGGTTTATCTACCAATGTCGTACTTATATGGAAAGAGATTTGTTGGGCCAATCACACCCATAGT TTTATCTCTAAGAAAAGAGCTTTACACAATTCCATATCATGAAATTGATTGGAATAGATCTCGC AATACATGTGCAAAGGAGGATTTGTACTATCCACATCCGAAGATGCAAGATATTTTATGGGGAT CGATATACCACGTGTATGAGCCATTGTTTAGTGGTTGGCCAGGGAAAAGGTTGAGGGAAAAGGC AATGAAAATTGCAATGGAACATATACATTATGAAGATGAAAATAGTCGATATATATGTCTTGGT CCTGTCAATAAAGTACTTAATATGCTTTGTTGTTGGGTTGAAGATCCTTATTCAGATGCCTTCA AATTTCATCTACAAAGAATCCCTGACTATCTTTGGCTTGCTGAAGATGGCATGAGAATGCAGGG TTACAATGGGAGTCAATTGTGGGACACTGCTTTCTCTATTCAAGCAATTATATCCACCAAACTT ATAGACACCTTTGGCCCAACCTTAAGAAAAGCACATCATTTTGTTAAACACTCTCAGATCCAGG AGGACTGTCCTGGTGATCCTAACGTTTGGTTCCGTCACATTCATAAAGGTGCTTGGCCTTTTTC AACTCGAGATCATGGTTGGCTCATCTCTGACTGTACGGCCGAGGGACTAAAGGCTTCTTTGATG TTATCCAAACTTCCATCCAAAATAGTTGGGGAGCCATTAGAAAAGAATCGCCTTTGTGATGCTG TTAATGTTCTCCTTTCTTTACAAAACGAAAATGGTGGATTTGCATCATACGAGTTGACAAGATC ATACCCTTGGTTGGAGTTGATCAACCCTGCAGAAACATTTGGAGATATCGTCATCGATTATTCG TATGTGGAGTGCACCTCAGCGACAATGGAAGCATTGGCATTGTTTAAGAAGTTACATCCAGGGC ATAGGACCAAAGAGATTGATGCTGCTATTGCCAAGGCCGCCAACTTTCTTGAAAATATGCAAAA GACTGATGGCTCTTGGTATGGATGTTGGGGGGTATGCTTCACATATGCAGGGTGGTTTGGGATA AAGGGATTGGTTGCTGCAGGAAGAACATATAATAACTGTGTTGCAATTCGTAAGGCTTGTAATT TTCTTTTATCTAAAGAGTTACCTGGTGGTGGATGGGGGGAGAGTTACCTTTCATGTCAGAATAA GGTCTACACCAATCTTGAAGGAAACAAACCACACTTGGTTAATACTGCTTGGGTAATGATGGCT CTCATTGAAGCTGGCCAGGGTGAGAGAGACCCAGCCCCATTGCATCGTGCAGCAAGATTATTAA TCAATTCTCAATTGGAGAGTGGTGATTTTCCCCAACAGGAGATCATGGGAGTGTTTAATAAAAA CTGTATGATTACATATGCTGCATACCGAAACATTTTTCCCATTTGGGCTCTTGGAGAGTATTCC CATAGAGTTTTGGATATGTAA		
Pathway 2 sequence id F Cmelo Protein	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP	902	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDM		
Pathway 2 sequence id G PSXY118L DNA (codon optimized)	ATGTGGAAACTTAAAGTTGCTGAGGGTGGCACTCCATGGTTAAGAACCCTAAACAATCACGTGG GTAGACAGGTTTGGGAGTTTGACCCACATTCTGGTTCTCCTCAAGACTTGGACGATATTGAGAC AGCAAGAAGAAATTTCCATGACAATCGTTTCACTCATAAACACTCAGACGACTTACTTATGAGA TTGCAGTTTGCCAAAGAAAACCCCATGAATGAAGTACTGCCTAAGGTTAAGGTTAAAGACGTTG AAGATGTCACAGAAGAAGCAGTTGCTACCACTCTAAGAAGAGGCTTGAACTTCTACAGCACCAT ACAATCCCACGATGGTCATTGGCCCGGTGATTTGGGTGGTCCTATGTTCTTGATGCCTGGTTTA GTTATCACTTTGTCCGTTACTGGGGCTCTTAACGCTGTTTTAACCGATGAACATAGAAAAGAGA TGAGAAGATACTTATACAATCACCAAAACAAAGATGGAGGCTGGGGCTTGCATATTGAAGGTCC TAGTACGATGTTTGGTTCAGTGTTATGCTATGTTACCTTGAGACTATTGGGTGAAGGGCCAAAT GATGGTGAGGGTGACATGGAGAGAGGAAGAGATTGGATCCTAGAACATGGTGGAGCAACATATA TAACCTCTTGGGGCAAAATGTGGTTATCTGTATTGGGCGTGTTTGAATGGTCAGGGAACAATCC AATGCCACCAGAAATTTGGTTGTTGCCTTATGCTCTTCCAGTTCATCCAGGAAGAATGTGGTGT CATTGTAGGATGGTTTACTTACCGATGTCGTACTTATACGGAAAACGTTTTGTCGGTCCTATTA CACCGACCGTGCTTAGTCTTAGGAAAGAGCTATTTACAGTACCGTATCATGATATAGACTGGAA CCAAGCAAGAAATTTATGTGCCAAAGAAGATTTATATTACCCTCATCCACTAGTGCAGGATATA TTATGGGCTACACTTCACAAGTTTGTCGAACCCGTCTTTATGAATTGGCCTGGTAAGAAGCTAA GGGAAAAGGCGATCAAAACAGCAATTGAGCACATTCATTATGAGGATGAGAATACTAGGTATAT CTGCATTGGGCCCGTCAACAAAGTGTTGAATATGCTGTGTTGTTGGGTGGAAGATCCTAATTCC GAAGCTTTCAAACTGCATTTGCCGAGAATTTATGATTACCTATGGGTAGCTGAAGATGGCATGA AAATGCAAGGTTATAACGGATCGCAATTGTGGGATACAGCATTTGCTGCACAAGCCATTATTAG CACAAATCTAATTGACGAATTCGGACCCACGTTAAAGAAGGCGCACGCCTTCATTAAGAATAGT CAAGTATCCGAAGATTGTCCTGGTGATCTGAGCAAATGGTACAGACACATCTCAAAAGGTGCTT GGCCATTTTCTACTGCCGATCATGGCTGGCCAATTAGCGACTGTACTGCGGAAGGGCTTAAGGC AGTATTGTTATTATCGAAGATAGCACCTGAGATTGTTGGAGAACCATTGGATTCCAAGCGTTTG TATGATGCAGTTAATGTAATTCTGTCACTGCAGAACGAAAATGGAGGTTTGGCGACTTACGAAT TGACTAGATCATATACGTGGCTGGAAATAATCAACCCTGCCGAAACGTTTGGTGACATAGTCAT AGATTGTCCATATGTTGAATGCACAAGTGCTGCCATTCAGGCTCTAGCAACTTTTGGTAAATTG TATCCAGGTCATCGTCGTGAAGAAATACAATGTTGCATAGAGAAAGCCGTTGCCTTCATCGAGA AGATTCAAGCTTCTGATGGTTCTTGGTATGGATCATGGGGCGTCTGTTTTACCTACGGGACGTG	903	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GTTTGGTATCAAGGGTTTGATTGCTGCAGGGAAGAATTTCTCCAATTGCTTAAGTATAAGGAAA GCGTGTGAGTTCTTACTGTCTAAACAATTGCCAAGTGGTGGATGGGCCGAATCTTACTTGTCTT GTCAGAACAAAGTGTACTCTAACTTAGAAGGAAATAGGTCGCACGTCGTTAATACAGGATGGGC TATGCTTGCATTGATTGAAGCAGAGCAAGCTAAGAGAGATCCAACTCCACTACATAGAGCAGCC GTATGCTTAATCAACTCACAACTTGAAAATGGCGACTTTCCGCAAGAAGAAATCATGGGCGTAT TCAATAAGAACTGTATGATAACTTACGCTGCGTATAGGTGCATCTTTCCCATTTGGGCTTTGGG TGAATATAGAAGAGTCTTACAAGCTTGCTAG		
Pathway 2 sequence id H PSXY118L Protein	MWKLKVAEGGTPWLRTLNNHVGRQVWEFDPHSGSPQDLDDIETARRNFHDNRFTHKHSDDLLMR LQFAKENPMNEVLPKVKVKDVEDVTEEAVATTLRRGLNFYSTIQSHDGHWPGDLGGPMFLMPGL VITLSVTGALNAVLTDEHRKEMRRYLYNHQNKDGGWGLHIEGPSTMFGSVLCYVTLRLLGEGPN DGEGDMERGRDWILEHGGATYITSWGKMWLSVLGVFEWSGNNPMPPEIWLLPYALPVHPGRMWC HCRMVYLPMSYLYGKRFVGPITPTVLSLRKELFTVPYHDIDWNQARNLCAKEDLYYPHPLVQDI LWATLHKFVEPVFMNWPGKKLREKAIKTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNS EAFKLHLPRIYDYLWVAEDGMKMQGYNGSQLWDTAFAAQAIISTNLIDEFGPTLKKAHAFIKNS QVSEDCPGDLSKWYRHISKGAWPFSTADHGWPISDCTAEGLKAVLLLSKIAPEIVGEPLDSKRL YDAVNVILSLQNENGGLATYELTRSYTWLEIINPAETFGDIVIDCPYVECTSAAIQALATFGKL YPGHRREEIQCCIEKAVAFIEKIQASDGSWYGSWGVCFTYGTWFGIKGLIAAGKNFSNCLSIRK ACEFLLSKQLPSGGWAESYLSCQNKVYSNLEGNRSHVVNTGWAMLALIEAEQAKRDPTPLHRAA VCLINSQLENGDFPQEEIMGVFNKNCMITYAAYRCIFPIWALGEYRRVLQAC	904	
Pathway 2 sequence id I DdCASY80L DNA (codon optimized)	ATGACCACGACAAACTGGTCCCTAAAGGTAGACAGAGGGCGTCAAACTTGGGAATACTCTCAAG AAAAGAAGGAGGCCACTGATGTGGACATCCATTTGCTACGACTGAAGGAACCCGGCACACATTG CCCTGAAGGTTGTGATCTGAATCGCGCTAAAACTCCCCAACAAGCGATTAAGAAAGCATTTCAG TACTTCTCCAAAGTCCAAACAGAAGATGGTCATTGGGCTGGAGATTTGGGTGGGCCAATGTTCT TGTTACCCGGTTTGGTGATAACATGCTACGTTACTGGCTATCAATTGCCAGAATCCACTCAAAG GGAAATTATAAGGTATCTGTTCAATAGACAGAATCCGGTTGATGGTGGCTGGGGTTTGCATATA GAGGCCCACTCTGATATATTTGGAACTACGTTACAATATGTATCATTGAGATTACTTGGAGTTC CAGCCGACCATCCATCTGTTGTAAAGGCAAGAACCTTCTTATTACAGAATGGTGGAGCAACCGG TATTCCTTCATGGGGTAAATTCTGGTTGGCCACGTTGAATGCATACGACTGGAACGGGTTGAAT CCAATTCCTATTGAATTTTGGCTGTTACCCTACAACTTACCCATTGCTCCTGGTAGGTGGTGGT GTCACTGTCGGATGGTCTATCTCCCAATGTCTTATATCTACGCTAAGAAAACAACTGGTCCACT AACAGATTTGGTCAAGGATCTGAGGAGAGAAATCTATTGTCAAGAGTACGAAAAGATTAACTGG TCTGAACAAAGAAACAATATTTCGAAATTAGACATGTACTACGAGCATACATCTCTTTTAAATG TTATAAACGGATCATTGAATGCTTACGAGAAAGTTCATTCCAAATGGCTTAGGGATAAAGCCAT TGACTATACCTTTGACCATATACGCTATGAAGATGAGCAGACGAAATACATTGACATAGGTCCA GTCAATAAGACCGTCAATATGTTATGCGTTTGGGATAGAGAAGGCAAATCTCCTGCGTTTTACA AACATGCCGATCGACTTAAAGATTATCTATGGTTATCTTTCGATGGGATGAAAATGCAAGGCTA TAACGGTTCTCAATTGTGGGACACTGCTTTTACGATCCAAGCATTCATGGAATCTGGGATTGCC AATCAATTCCAGGATTGTATGAAATTAGCTGGTCACTATTTGGACATCTCCCAGGTACCAGAAG	905	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ATGCCAGAGATATGAAGCACTACCACAGACACTATTCGAAGGGTGCATGGCCTTTTAGTACCGT TGACCATGGATGGCCAATTTCAGATTGCACAGCAGAAGGTATCAAGTCAGCGCTTGCTCTCAGA TCTTTGCCTTTTATCGAACCAATATCCTTAGATAGAATTGCTGATGGCATTAATGTTCTATTAA CCTTGCAAAATGGGGATGGTGGATGGGCATCGTACGAGAACACAAGAGGACCGAAATGGCTGGA AAAGTTTAACCCTTCCGAAGTTTTCCAGAATATAATGATTGACTATAGCTATGTGGAATGTAGT GCTGCTTGTATTCAAGCTATGAGTGCGTTTCGTAAACATGCACCTAATCATCCAAGAATTAAGG AAATCAACAGATCTATTGCACGTGGAGTGAAATTTATCAAGAGCATTCAACGTCAGGATGGTTC ATGGCTGGGCAGTTGGGGAATTTGTTTTACCTACGGTACTTGGTTTGGCATAGAGGGCTTAGTA GCATCTGGTGAGCCTCTAACATCGCCATCGATCGTGAAGGCTTGCAAGTTTCTTGCGTCAAAAC AACGTGCAGATGGTGGTTGGGGAGAAAGCTTTAAAAGCAATGTGACTAAAGAATATGTTCAACA CGAAACTTCACAAGTAGTCAATACTGGTTGGGCTCTACTCAGTCTAATGAGTGCTAAATATCCG GACAGAGAGTGCATAGAGAGAGGTATCAAATTCTTAATACAGAGGCAATATCCGAACGGTGATT TTCCACAGGAATCCATTATTGGCGTTTTCAATTTTAACTGTATGATCTCATATTCAAACTATAA GAACATATTCCCTCTTTGGGCCTTGAGTAGGTATAATCAATTGTACCTTAAAAGCAAAATCTGA		
Pathway 2 sequence id J DdCASY80L protein	MTTTNWSLKVDRGRQTWEYSQEKKEATDVDIHLLRLKEPGTHCPEGCDLNRAKTPQQAIKKAFQ YFSKVQTEDGHWAGDLGGPMFLLPGLVITCYVTGYQLPESTQREIIRYLFNRQNPVDGGWGLHI EAHSDIFGTTLQYVSLRLLGVPADHPSVVKARTFLLQNGGATGIPSWGKFWLATLNAYDWNGLN PIPIEFWLLPYNLPIAPGRWWCHCRMVYLPMSYIYAKKTTGPLTDLVKDLRREIYCQEYEKINW SEQRNNISKLDMYYEHTSLLNVINGSLNAYEKVHSKWLRDKAIDYTFDHIRYEDEQTKYIDIGP VNKTVNMLCVWDREGKSPAFYKHADRLKDYLWLSFDGMKMQGYNGSQLWDTAFTIQAFMESGIA NQFQDCMKLAGHYLDISQVPEDARDMKHYHRHYSKGAWPFSTVDHGWPISDCTAEGIKSALALR SLPFIEPISLDRIADGINVLLTLQNGDGGWASYENTRGPKWLEKFNPSEVFQNIMIDYSYVECS AACIQAMSAFRKHAPNHPRIKEINRSIARGVKFIKSIQRQDGSWLGSWGICFTYGTWFGIEGLV ASGEPLTSPSIVKACKFLASKQRADGGWGESFKSNVTKEYVQHETSQVVNTGWALLSLMSAKYP DRECIERGIKFLIQRQYPNGDFPQESIIGVFNFNCMISYSNYKNIFPLWALSRYNQLYLKSKI	906	
Pathway 2 sequence id K Cmelo DNA (codon optimized)	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA	907	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGTGA		
Pathway 1 sequence id L SQE1 DNA (codon optimized)	ATGGTCGATCAATGTGCGTTAGGCTGGATATTAGCTAGTGCACTAGGATTGGTTATCGCTCTAT GCTTCTTCGTTGCACCAAGAAGAAACCACAGAGGTGTTGATAGCAAAGAAAGGGATGAGTGTGT GCAGTCTGCAGCCACAACTAAGGGCGAATGCAGATTTAACGATAGAGATGTGGATGTTATTGTG GTTGGTGCAGGAGTAGCTGGTTCGGCATTAGCCCATACACTTGGTAAAGACGGTAGAAGAGTTC ATGTCATTGAGAGGGATTTGACTGAACCAGACAGGATTGTTGGTGAACTTCTACAACCAGGAGG CTATTTGAAGTTGATTGAGTTAGGCTTACAAGACTGTGTGGAAGAAATAGATGCACAAAGAGTT TACGGGTATGCTTTGTTTAAAGATGGTAAGAACACCAGACTATCTTATCCACTTGAAAATTTTC ACTCAGATGTCTCCGGTAGAAGCTTTCACAACGGTAGATTCATTCAGAGGATGAGAGAAAAGGC TGCTTCGCTGCCAAATGTAAGATTGGAACAAGGGACGGTTACTAGTCTACTGGAAGAGAAAGGG ACGATCAAAGGAGTTCAGTATAAGTCCAAGAATGGAGAGGAGAAAACCGCGTATGCGCCTTTAA CGATAGTGTGTGATGGCTGTTTCTCTAACTTACGTAGATCATTATGTAATCCCATGGTTGACGT TCCGAGCTACTTTGTTGGTCTTGTGTTAGAAAATTGCGAACTGCCATTTGCCAATCATGGACAT GTTATCCTTGGTGATCCATCCCCAATCTTATTCTATCAGATCTCAAGAACCGAAATTAGGTGTT TGGTCGATGTACCCGGTCAAAAGGTCCCTTCAATTGCCAATGGCGAAATGGAGAAATATTTAAA	908	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GACTGTTGTAGCTCCACAAGTACCACCTCAGATTTACGACAGTTTTATAGCCGCCATTGACAAA GGGAATATCAGAACTATGCCTAATAGGTCTATGCCTGCAGCTCCCCATCCAACTCCAGGTGCGT TACTGATGGGCGATGCATTCAACATGAGACATCCTCTAACAGGAGGTGGCATGACAGTAGCACT GTCTGACATTGTGGTCTTGAGAAACTTGTTAAAACCGTTAAAAGACTTGTCTGACGCCTCTACT TTGTGCAAATACTTGGAATCCTTTTATACCCTTCGTAAACCAGTAGCTAGCACAATCAACACCT TAGCTGGAGCCTTGTACAAAGTCTTTTGCGCATCACCGGATCAAGCGAGAAAGGAAATGAGACA AGCTTGTTTTGATTACCTAAGTCTGGGAGGTATTTTCTCGAATGGTCCTGTCTCATTGTTGTCA GGGTTGAATCCCAGACCTTTATCCTTGGTATTGCACTTCTTCGCTGTCGCAATTTATGGTGTTG GTCGTTTGCTTCTACCTTTTCCAAGTGTTAAGGGTATATGGATTGGTGCAAGGTTGATCTACTC TGCCTCTGGTATAATATTTCCCATAATTAGAGCTGAAGGCGTTCGTCAAATGTTCTTTCCTGCT ACAGTGCCCGCTTACTATCGTTCCCCACCTGTATTTAAACCGATAGTGTAG		
Pathway 1 sequence id M SQE1 Protein	MVDQCALGWILASALGLVIALCFFVAPRRNHRGVDSKERDECVQSAATTKGECRFNDRDVDVIV VGAGVAGSALAHTLGKDGRRVHVIERDLTEPDRIVGELLQPGGYLKLIELGLQDCVEEIDAQRV YGYALFKDGKNTRLSYPLENFHSDVSGRSFHNGRFIQRMREKAASLPNVRLEQGTVTSLLEEKG TIKGVQYKSKNGEEKTAYAPLTIVCDGCFSNLRRSLCNPMVDVPSYFVGLVLENCELPFANHGH VILGDPSPILFYQISRTEIRCLVDVPGQKVPSIANGEMEKYLKTVVAPQVPPQIYDSFIAAIDK GNIRTMPNRSMPAAPHPTPGALLMGDAFNMRHPLTGGGMTVALSDIVVLRNLLKPLKDLSDAST LCKYLESFYTLRKPVASTINTLAGALYKVFCASPDQARKEMRQACFDYLSLGGIFSNGPVSLLS GLNPRPLSLVLHFFAVAIYGVGRLLLPFPSVKGIWIGARLIYSASGIIFPIIRAEGVRQMFFPA TVPAYYRSPPVFKPIV	909	
Pathway 1 sequence id N SQE2 DNA (codon optimized)	ATGGTCGATCAATGCGCGTTAGGCTGGATATTAGCTTCCGTCCTAGGAGCTGCAGCGTTGTATT TCTTGTTTGGTAGAAAGAATGGTGGTGTGTCTAATGAAAGAAGGCATGAAAGTATTAAGAACAT TGCAACTACCAATGGTGAGTATAAGTCAAGTAACTCCGATGGTGACATCATCATTGTTGGTGCT GGCGTTGCTGGATCTGCTTTGGCCTATACGCTAGGTAAAGATGGGAGAAGAGTGCATGTCATTG AAAGGGATTTGACAGAACCAGACCGTATAGTAGGTGAATTGTTACAACCAGGAGGGTATCTAAA ACTGACAGAGTTGGGTTTAGAAGATTGTGTGGATGATATAGATGCTCAACGTGTTTATGGGTAT GCATTATTCAAAGACGGTAAAGATACCAGATTGTCCTATCCCTTGGAAAAGTTTCACTCTGACG TCGCAGGCAGATCCTTTCATAATGGCAGATTCATTCAGCGTATGAGAGAGAAAGCTGCTTCATT GCCTAAAGTGAGCCTAGAGCAAGGGACTGTAACGTCACTGTTGGAGGAAAACGGAATAATCAAA GGGGTACAGTATAAAACTAAGACTGGTCAAGAGATGACTGCATATGCTCCTTTAACAATCGTCT GTGACGGCTGCTTTTCGAACCTTCGTAGAAGCTTGTGCAACCCAAAAGTCGATGTTCCCTCATG TTTTGTGGGATTAGTTCTAGAAAATTGCGATTTGCCTTACGCCAATCACGGACATGTGATCTTG GCTGATCCGTCACCTATTCTGTTCTACAGAATATCTAGTACCGAAATCAGGTGTTTGGTTGATG TTCCAGGTCAGAAAGTGCCTTCTATCAGTAATGGCGAAATGGCCAACTACTTGAAGAATGTTGT TGCACCTCAGATTCCAAGCCAACTTTACGACTCTTTTGTTGCAGCCATTGACAAGGGAAACATA AGAACAATGCCGAATAGATCTATGCCAGCAGATCCATATCCAACACCCGGTGCGCTGCTAATGG GTGATGCCTTTAACATGAGACATCCTCTAACAGGTGGTGGTATGACAGTCGCTTTATCGGATGT TGTCGTATTAAGAGACTTACTGAAACCACTTAGAGACTTGAATGATGCACCTACCTTGAGCAAG	910	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TATTTAGAAGCCTTTTACACTCTGCGTAAGCCTGTTGCTTCTACCATAAACACGTTAGCAGGAG CATTGTACAAGGTATTCTGTGCTTCTCCTGATCAAGCGAGAAAGGAAATGAGACAAGCCTGTTT TGACTACCTTTCACTTGGTGGCATATTCAGTAATGGACCAGTATCCTTATTGTCAGGTCTTAAT CCAAGGCCCATTTCCCTTGTTTTACACTTCTTTGCAGTGGCTATCTATGGTGTTGGAAGGCTAT TAATACCGTTTCCATCACCGAAAAGGGTATGGATTGGTGCTAGAATTATTTCGGGCGCGAGTGC AATTATTTTCCCCATTATCAAAGCTGAAGGCGTCAGACAAATGTTCTTTCCAGCCACTGTAGCT GCCTACTACAGAGCCCCAAGAGTTGTTAAAGGTAGGTAG		
Pathway 1 sequence id O SQE2 Protein	MVDQCALGWILASVLGAAALYFLFGRKNGGVSNERRHESIKNIATTNGEYKSSNSDGDIIIVGA GVAGSALAYTLGKDGRRVHVIERDLTEPDRIVGELLQPGGYLKLTELGLEDCVDDIDAQRVYGY ALFKDGKDTRLSYPLEKFHSDVAGRSFHNGRFIQRMREKAASLPKVSLEQGTVTSLLEENGIIK GVQYKTKTGQEMTAYAPLTIVCDGCFSNLRRSLCNPKVDVPSCFVGLVLENCDLPYANHGHVIL ADPSPILFYRISSTEIRCLVDVPGQKVPSISNGEMANYLKNVVAPQIPSQLYDSFVAAIDKGNI RTMPNRSMPADPYPTPGALLMGDAFNMRHPLTGGGMTVALSDVVVLRDLLKPLRDLNDAPTLSK YLEAFYTLRKPVASTINTLAGALYKVFCASPDQARKEMRQACFDYLSLGGIFSNGPVSLLSGLN PRPISLVLHFFAVAIYGVGRLLIPFPSPKRVWIGARIISGASAIIFPIIKAEGVRQMFFPATVA AYYRAPRVVKGR	911	
Pathway 1 sequence id P SQE3 DNA (codon optimized)	ATGGAATTCCAATCGGAACCCTTGTTTGGGGTTCTGTTGGCTAGTCTTTTAGCGCTGGTTTTCT TCTTTACTTTGAGAGATGGTACCAAGAACAAGAAAACCACAACTGGGTCATCTGTGGATCTGAA ACGTACTGACGCTGTCCTACAAATGTCTCCCGAAAACGATGCTAGAAGGCAGGAAATCATAGGG GATTCAGACGTGATTGTAGTAGGTGCAGGAGTTGCAGGAGCTGCATTAGCCTATACGTTGGGCA AAGATGGTAGAAAAGTTCACGTAATTGAAAGAGACTTGACAGAGCCAGATAGAATTGTAGGTGA ACTATTACAACCTGGTGGCTACTTGAAGCTAGTGGAGTTGGGTCTTGAAGATAGTGTTAAAGGT ATTGACGCTCAACAAGTCTTTGGATATGCGTTGTATAAGGACGGTAAACACACAAGACTTACGT ATCCTTTGGAAAAGTTCGACTCAACTGTATCAGGCAGATCCTTCCATAATGGCAGATTCATCCA AAGATTAAGGGAATCTGTGAGACTAGAACAAGGAACTGTTACCAGCATCTTAGAAGAGGATGGA ACAGTTAAAGGTGTTCAGTATAAGACGAAAATTGGAGAGGAGTTTACAGCTTATGCACCATTGA CAATCGTCTGTGATGGCGGGTTTAGTAACTTGAGAAGAAATTTATGCAAACCACAAATCGACAT TCCCTCGTGTTTTGTGGGATTAGTTTTGGAAAACTGCAAACTTCCCTTCGAGAATCATGGCCAT GTAGTACTGGCAGATCCGTCACCTATTCTGTTATACCCGATTAGTTCAACGGAAATTCGTTGTT TGGTTGACATTCCAGGTCAGAAAGTGCCCTCAGTAGCCAATGGCGAAATGGCCAGATACTTAAA GACTGTTGTCGCTCCGCAAGTTCCACCTGAACTACATGCTGCCTTTATAGCGGCTATAGAGAAA GGTAATATCAAGAGCACAACTAACAGATCTATGCCAGCAGCACCTCACCCAACACCTGGCGCCC TGTTGCTAGGTGATGCATTCAATATGAGACATCCCTTAACCGGTGGTGGTATGACTGTTGCCTT AGCGGACATTGTTGTGCTTAGAGATTTGTTGCGTCCTCTTGCTAATCTAAAGGATGCTGATGCC TTGTGTCACTATCTAGAGTCCTTTTACACCCTTCGTAAACCTGTCGCATCCACCATAAACACAT TAGCTGGCGCATTATACAAGGTCTTTTGTGCCTCTCCAGATTCTGCTAGAAAGGAAATGAGGGA AGCATGTTTTGATTACCTGAGTTTAGGTGGTGTCTTTTCGTCTGGACCTGTAGCTTTGTTATCC GGTTTGAATCCAAGACCTTTGTCCTTATTTTGCCATTTCTTTGCAGTGGCCATATATGGAGTTT	912	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTAGGTTGCTTATACCATTCCCAAGCCCAATGAGGATTTGGATTGGTGTTAGATTAATCACTGT TGCGGCCGGTATAATATTTCCGATTATCAAAGCTGAAGGGGTCAGACAGATGTTCTTTCCTGCT ACTGTCCCAGCTTATTACAGGGCACCACCAATGTAG		
Pathway 1 sequence id Q SQE3 Protein	MEFQSEPLFGVLLASLLALVFFFTLRDGTKNKKTTTGSSVDLKRTDAVLQMSPENDARRQEIIG DSDVIVVGAGVAGAALAYTLGKDGRKVHVIERDLTEPDRIVGELLQPGGYLKLVELGLEDSVKG IDAQQVFGYALYKDGKHTRLTYPLEKFDSTVSGRSFHNGRFIQRLRESVRLEQGTVTSILEEDG TVKGVQYKTKIGEEFTAYAPLTIVCDGGFSNLRRNLCKPQIDIPSCFVGLVLENCKLPFENHGH VVLADPSPILLYPISSTEIRCLVDIPGQKVPSVANGEMARYLKTVVAPQVPPELHAAFIAAIEK GNIKSTTNRSMPAAPHPTPGALLLGDAFNMRHPLTGGGMTVALADIVVLRDLLRPLANLKDADA LCHYLESFYTLRKPVASTINTLAGALYKVFCASPDSARKEMREACFDYLSLGGVFSSGPVALLS GLNPRPLSLFCHFFAVAIYGVSRLLIPFPSPMRIWIGVRLITVAAGIIFPIIKAEGVRQMFFPA TVPAYYRAPPM	913	
SS3e-E7 fusion protein, coding sequence, cucurbitadienol synthase	ATGTCCGAAAATCACGTTCCTGCCGTTGTCAAAACGCGTGGAAGTGCAGCTCCTGGAAGTGGAA GTGGTTCAGGAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTG GATTAAAAGCATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAA AACGACGACGATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAA ATCACTTCCGTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATG CGAGAAAGAGATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAA GATGTTAAGAAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTAC AGACCTCTGACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGT TATTGCGCTATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATG TGTCGTTACATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTT CTACTATGTTTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGG CGGTGAGCATGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCA ATAACTTCCTGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATC CATTGCCACCCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTG TCATTGTAGAATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATC ACTCCAATAGTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGA ATAGATCCAGAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATAT CCTATGGGGCAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTG AGGGAAAAGGCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACA TATGCCTTGGACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTC TGATGCTTTCAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATG AGAATGCAGGGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCT CAACGAAATTGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAG TCAGATTCAAGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCT TGGCCTTTTAGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAG CTTCACTGATGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTT	915	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ATGTGATGCAGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAA TTAACTAGGTCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAA TTGACTACAGTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTT GCACCCTGGGCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAG AATATGCAGAAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTT GGTTTGGCATCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAA GGCTTGTAACTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGT TGCCAAAACAAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGG TCATGATGGCCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGC CAGATTGCTAATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTG TTTAATAAGAACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAG GTGAATACTCCCATCGTGTCTTGGATATGTGA		
SS3e-E7 fusion protein, cucurbitadienol synthase	MSENHVPAVVKTRGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENE NDDDEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGE DVKKEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEM CRYIYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATA ITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPI TPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRL REKAMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGM RMQGYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGA WPFSTRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYE LTRSYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLE NMQKTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLS CQNKVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGV FNKNCMITYAAYRNIFPIWALGEYSHRVLDM-	916	
SS3e-E7 fusion protein, fusion domain	MSENHVPAVVKTR	917	
SS3d-G5 fusion protein, coding sequence, cucurbitadienol synthase	ATGACGACCCAGCAAGAGGAGCTCGATGTTGGAGACAGTGAGGGAAGTGCAGCTCCTGGAAGTG GAAGTGGTTCAGGAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAA GTGGATTAAAAGCATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAAT GAAAACGACGACGATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAA GAAATCACTTCCGTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACA ATGCGAGAAAGAGATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGT GAAGATGTTAAGAAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTG TACAGACCTCTGACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCT AGTTATTGCGCTATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAA ATGTGTCGTTACATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCT	919	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTTCTACTATGTTTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGA TGGCGGTGAGCATGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACA GCAATAACTTCCTGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACA ATCCATTGCCACCCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTG GTGTCATTGTAGAATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCA ATCACTCCAATAGTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATT GGAATAGATCCAGAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGA TATCCTATGGGGCAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGA TTGAGGGAAAAGGCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGT ACATATGCCTTGGACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTA CTCTGATGCTTTCAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGA ATGAGAATGCAGGGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAA TCTCAACGAAATTGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCA TAGTCAGATTCAAGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGA GCTTGGCCTTTTAGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGA AAGCTTCACTGATGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCG TTTATGTGATGCAGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTAT GAATTAACTAGGTCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCG TAATTGACTACAGTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAA GTTGCACCCTGGGCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTG GAGAATATGCAGAAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTG GTTGGTTTGGCATCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAG AAAGGCTTGTAACTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTA AGTTGCCAAAACAAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCAT GGGTCATGATGGCCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGC TGCCAGATTGCTAATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGT GTGTTTAATAAGAACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTC TAGGTGAATACTCCCATCGTGTCTTGGATATGTGA		
SS3d-G5 fusion protein, cucurbitadienol synthase	MTTQQEELDVGDSEGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGEN ENDDDEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEG EDVKKEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQE MCRYIYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGAT AITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGP ITPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKR LREKAMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDG MRMQGYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKG AWPFSTRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASY ELTRSYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFL	920	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ENMQKTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYL SCQNKVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMG VFNKNCMITYAAYRNIFPIWALGEYSHRVLDM-		
SS3d-G5 fusion protein, fusion domain	MTTQQEELDVGDSE	921	
SS3c-G8 fusion protein, coding sequence	ATGGAGGACGGTAAACAGGCCATCAGCGAGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAG GAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAG CATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGAC GATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCC GTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGA GATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAG AAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTG ACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCT ATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTAC ATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGT TTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCA TGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCC TGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCAC CCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAG AATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATA GTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCA GAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGG CAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAG GCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTG GACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTT CAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAG GGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAAT TGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCA AGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTT AGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGA TGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGC AGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGG TCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACA GTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGG GCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAG AAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCA TCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAA	923	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAAC AAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGG CCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCT AATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAG AACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACT CCCATCGTGTCTTGGATATGTGA		
SS3c-G8 fusion protein, cucurbitadienol synthase	MEDGKQAISEGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDD DEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVK KEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRY IYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITS WGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPI VLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREK AMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQ GYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPF STRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTR SYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQ KTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQN KVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNK NCMITYAAYRNIFPIWALGEYSHRVLDM-	924	
SS3c-G8 fusion protein, fusion domain	MEDGKQAISE	925	
SS3c-E5 fusion protein, coding sequence	ATGACGATCGGTGATAAGCTGAAAAAGAAGCTTGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTT CAGGAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAA AAGCATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGAC GACGATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACT TCCGTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAA AGAGATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTT AAGAAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCT CTGACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGC GCTATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGT TACATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTA TGTTTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGA GCATGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACT TCCTGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGC CACCCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTG TAGAATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCA ATAGTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGAT	927	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CCAGAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATG GGGCAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAA AAGGCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCC TTGGACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGC TTTCAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATG CAGGGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGA AATTGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGAT TCAAGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCT TTTAGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCAC TGATGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGA TGCAGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACT AGGTCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACT ACAGTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCC TGGGCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATG CAGAAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTG GCATCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTG TAACTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAA AACAAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGA TGGCCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATT GCTAATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAAT AAGAACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAAT ACTCCCATCGTGTCTTGGATATGTGA		
SS3c-E5 fusion protein, cucurbitadienol synthase	MTIGDKLKKKLGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENEND DDEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDV KKEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCR YIYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAIT SWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITP IVLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLRE KAMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRM QGYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWP FSTRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELT RSYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENM QKTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQ NKVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFN KNCMITYAAYRNIFPIWALGEYSHRVLDM-	928	
SS3c-E5 protein, fusion domain	MTIGDKLKKKL	929	
SS2c-E2 fusion	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA	931	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
protein, coding sequence	TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGATGTTGG AGCCGTCACCCTAA		
SS2c-E2 fusion	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR	932	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
protein, cucurbitadienol synthase	NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMMLEPSP-		
SS2c-E2 protein, fusion domain	MMLEPSP	933	
SS2c-A10b fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG	935	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGAATTCGA ATGAAGACATCATACCTGAACTATAA		
SS2c-A10b fusion protein, cucurbitadienol synthase	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMNSNEDIIPEL-	936	
SS2c-A10b protein, fusion domain	MNSNEDIIPE	937	
SS3b-C1 fusion protein, coding sequence	ATGTGGAACAAAACCAAAAAAACACAAGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAA TGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCAT TAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGAT GAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGTA ATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGAT CATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAAA GAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGACG GTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTATA CGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACATC	939	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTTG GGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATGG TGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTGG GGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCCG AATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAAT GGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGTT TTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGAA ACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCAG TATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGCC ATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGAC CCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCAA GTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGGT TATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTGA TTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAGA GGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAGC ACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATGT TATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAGT CAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTCC TATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGTT ATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGCA TAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAAA ACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATCA AAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACTT CCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAAA GTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCCT TGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAAT CAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAAC TGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCCC ATCGTGTCTTGGATATGTGA		
SS3b-C1 fusion protein, cucurbitadienol synthase	MWNKTKKTQGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDD EAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKK EAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYI YNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSW GKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIV LSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKA MKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQG YNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFS TRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRS	940	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	YPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQK TDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNK VYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKN CMITYAAYRNIFPIWALGEYSHRVLDM-		
SS3b-C1 protein, fusion domain	MWNKTKKTQ	941	
SS3b-B10 fusion protein, coding sequence	ATGGCCAAAGAAGATACTGTAAAACTAAAAAGGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTT CAGGAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAA AAGCATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGAC GACGATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACT TCCGTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAA AGAGATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTT AAGAAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCT CTGACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGC GCTATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGT TACATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTA TGTTTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGA GCATGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACT TCCTGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGC CACCCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTG TAGAATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCA ATAGTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGAT CCAGAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATG GGGCAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAA AAGGCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCC TTGGACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGC TTTCAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATG CAGGGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGA AATTGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGAT TCAAGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCT TTTAGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCAC TGATGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGA TGCAGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACT AGGTCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACT ACAGTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCC TGGGCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATG CAGAAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTG GCATCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTG	943	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TAACTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAA AACAAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGA TGGCCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATT GCTAATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAAT AAGAACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAAT ACTCCCATCGTGTCTTGGATATGTGA		
SS3b-B10 fusion protein, cucurbitadienol synthase	MAKEDTVKLKRGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENEND DDEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDV KKEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCR YIYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAIT SWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITP IVLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLRE KAMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRM QGYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWP FSTRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELT RSYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENM QKTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQ NKVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFN KNCMITYAAYRNIFPIWALGEYSHRVLDM-	944	
SS3b-B10 protein, fusion domain	MAKEDTVKLKR	945	
SS3a-D8 fusion protein, coding sequence	ATGTCATTTCAAATTGAAACGGTTCGTACTGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAG GAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAG CATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGAC GATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCC GTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGA GATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAG AAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTG ACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCT ATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTAC ATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGT TTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCA TGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCC TGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCAC CCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAG AATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATA GTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCA GAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGG	947	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAG GCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTG GACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTT CAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAG GGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAAT TGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCA AGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTT AGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGA TGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGC AGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGG TCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACA GTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGG GCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAG AAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCA TCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAA CTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAAC AAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGG CCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCT AATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAG AACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACT CCCATCGTGTCTTGGATATGTGA		
SS3a-D8 fusion protein, cucurbitadienol synthase	MSFQIETVRTGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDD DEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVK KEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRY IYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITS WGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPI VLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREK AMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQ GYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPF STRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTR SYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQ KTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQN KVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNK NCMITYAAYRNIFPIWALGEYSHRVLDM	948	
SS3a-D8 protein, fusion domain	MSFQIETVRT	949	
SS3a-A2 fusion protein (5'	ATGACCGGCTTGAATGGAGATGCTGACAGCGATCTACTAGGAAGTGCAGCTCCTGGAAGTGGAA GTGGTTCAGGAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTG	951	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
fusion), coding sequence	GATTAAAAGCATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAA AACGACGACGATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAA ATCACTTCCGTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATG CGAGAAAGAGATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAA GATGTTAAGAAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTAC AGACCTCTGACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGT TATTGCGCTATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATG TGTCGTTACATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTT CTACTATGTTTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGG CGGTGAGCATGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCA ATAACTTCCTGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATC CATTGCCACCCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTG TCATTGTAGAATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATC ACTCCAATAGTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGA ATAGATCCAGAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATAT CCTATGGGGCAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTG AGGGAAAAGGCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACA TATGCCTTGGACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTC TGATGCTTTCAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATG AGAATGCAGGGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCT CAACGAAATTGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAG TCAGATTCAAGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCT TGGCCTTTTAGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAG CTTCACTGATGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTT ATGTGATGCAGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAA TTAACTAGGTCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAA TTGACTACAGTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTT GCACCCTGGGCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAG AATATGCAGAAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTT GGTTTGGCATCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAA GGCTTGTAACTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGT TGCCAAAACAAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGG TCATGATGGCCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGC CAGATTGCTAATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTG TTTAATAAGAACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAG GTGAATACTCCCATCGTGTCTTGGATATGTGA		
SS3a-A2 fusion protein (5'	MTGLNGDADSDLLGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENE NDDDEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGE	952	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
fusion), cucurbitadienol synthase	DVKKEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEM CRYIYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATA ITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPI TPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRL REKAMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGM RMQGYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGA WPFSTRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYE LTRSYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLE NMQKTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLS CQNKVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGV FNKNCMITYAAYRNIFPIWALGEYSHRVLDM-		
SS3a-A2 protein (5' fusion), fusion domain	MTGLNGDADSDLL	953	
SS3f-A8 fusion protein (5' fusion), coding sequence	ATGGCAAGTAACCAGCTCGAGCCCCTGCAAACTGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTT CAGGAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAA AAGCATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGAC GACGATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACT TCCGTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAA AGAGATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTT AAGAAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCT CTGACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGC GCTATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGT TACATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTA TGTTTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGA GCATGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACT TCCTGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGC CACCCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTG TAGAATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCA ATAGTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGAT CCAGAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATG GGGCAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAA AAGGCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCC TTGGACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGC TTTCAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATG CAGGGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGA AATTGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGAT	955	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TCAAGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCT TTTAGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCAC TGATGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGA TGCAGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACT AGGTCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACT ACAGTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCC TGGGCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATG CAGAAAACAGACGGGTCTTGGTATGGTTGTTGGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTG GCATCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTG TAACTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAA AACAAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGA TGGCCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATT GCTAATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAAT AAGAACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAAT ACTCCCATCGTGTCTTGGATATGTGA		
SS3f-A8 fusion protein (5' fusion), cucurbitadienol synthase	MASNQLEPLQTGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENEND DDEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDV KKEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCR YIYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAIT SWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITP IVLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLRE KAMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRM QGYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWP FSTRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELT RSYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENM QKTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQ NKVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFN KNCMITYAAYRNIFPIWALGEYSHRVLDM-	956	
SS3f-A8 protein (5' fusion), fusion domain	MASNQLEPLQT	957	
SS4b-B8b fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC	959	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGAATTTAG ATTTAGATCAAGATTCAGACTAG		
SS4b-B8b fusion protein, cucurbitadienol synthase	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG	960	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMNLDLDQDSD-		
SS4b-B8b fusion protein, fusion domain	MNLDLDQDSD	961	
SS4b-B8a fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT	963	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGATAAAAC ATATAGTTTCGCCATTCAGGACGAATTTTGTTGGCATCAGCAAGTCCGTGCTGTCAAGGATGAT TCATCACAAGGTTACAATCATAGGTTCTGGCCCCGCTGCCCACACCGCTGCTATATACTTGGCA AGAGCAGAGATGAAGCCCACATTATATGAGGGAATGATGGCCAACGGAATTGCTGCTGGTGGCC AATTGACAACAACCACCGATATCGAAAATTTCCCAGGGTTTCCTGAATCGTTGAGTGGCAGTGA ACTGATGGAGAGGATGAGGAAACAATCTGCCAAGTTTGGCACTAACATAATTATCGAGACTGTC TCTAAAGTCGATTTATCTTCAAAACCATTCAGATTATGGACCGAATTTAATGAGGATGCAGAGC CTGTGACCACTGATGCTATAATCTTGGCCACGGGTGCTTCCGCTAAGAGAATGCATTTACCAGG GGAGGAAACCTACTGGCAGCAGGGAATATCTGCCTGTGCTGTATGTGATGGTGCAGTCCCTATC TTTAGAAACAAGCCATTGGCCGTTATTGGTGGTGGTGACTCTGCGTGTGAGGAAGCGGAATTTC TTACGAAGTATGCGTCGAAAGTATATATATTAGTAAGAAAGGATCATTTTCGTGCATCTGTAAT AATGCAGAGACGAATTGAGAAAAATCCAAACATCATTGTTTTGTTCAACACAGTTGCATTAGAA GCTAAGGGTGATGGTAAGTTATTGAATATGTTGAGAATTAAGAATACTAAAAGTAATGTGGAGA ACGATTTAGAAGTAAATGGACTATTTTACGCAATAGGTCACAGCCCTGCCACAGATATAGTTAA AGGACAAGTAGATGAAGAAGAGACGGGGTATATAAAAACTGTGCCTGGATCGTCTCTGACTTCT GTGCCAGGTTTTTTTGCTGCAGGTGACGTTCAGGACTCTAGGTATAGACAAGCAGTTACTTCTG CTGGTTCCGGATGCATTGCTGCTTTGGATGCAGAACGGTACCTAAGTGCCCAAGAGTAA		
SS4b-B8a fusion protein, cucurbitadienol synthase	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS	964	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	HRVLDMGSAAPGSGSGSGMIKHIVSPFRTNFVGISKSVLSRMIHHKVTIIGSGPAAHTAAIYLA RAEMKPTLYEGMMANGIAAGGQLTTTTDIENFPGFPESLSGSELMERMRKQSAKFGTNIIIETV SKVDLSSKPFRLWTEFNEDAEPVTTDAIILATGASAKRMHLPGEETYWQQGISACAVCDGAVPI FRNKPLAVIGGGDSACEEAEFLTKYASKVYILVRKDHFRASVIMQRRIEKNPNIIVLFNTVALE AKGDGKLLNMLRIKNTKSNVENDLEVNGLFYAIGHSPATDIVKGQVDEEETGYIKTVPGSSLTS VPGFFAAGDVQDSRYRQAVTSAGSGCIAALDAERYLSAQE-		
SS4b-D4 fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA	966	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGGCCGTAC AGAACCATATCTTGCCTCTAA		
SS4b-D4 fusion protein, cucurbitadienol synthase	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMAVQNHILPLTRVM-	967	
SS4b-D4 fusion protein, fusion domain	MAVQNHILPLTRVM	968	
SS4c-C4a fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA	970	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGTCTGCTT CAACTCATTCGCCTGAATAACCGTGTCTGA		
SS4c-C4a fusion protein, cucurbitadienol synthase	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMSASTHSPE-PCL	971	
SS4c-C4a fusion protein, fusion domain	MSASTHSPE	972	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
SS4c-C4b fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGGGTACTA GCATAGTAAATCTAAACCAAAAGATTGAACTGCCCCCAATCCAGGTCTTATTCGAGTCACTTAA	974	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CCGAGAAAATGAAACAAAACCCCACTTCGAGGAACGCAGGTTATATCAACCTAATCCTTCATTT GTTCCTAGAACAAATATAGCAGTTGGTAGCCCAGTTAACCCGGTTCCAGTATCATCCCCTGTTT TTTTCATTGGTCCTTCTCCACAGAGAAGCATTCAGAATCACAACGCTATTATGACTCAAAACAT ACGTCAGTATCCAGTTATATATAATAATAACCGAGAAGTTATATCTACGGGTGAGAGAAATTAC ATAATAACTGTAGGGGGGCCTCCGGTAACTTCTTCGCAGCCCGAGTATGAGCATATCTCAACTC CCAATTTCTATCAAGAGCAGAGACTGGCACAACCTCATCCGGTAAATGAGAGTATGATGATAGG TGGTTATACAAATCCTCAGCCTATTAGCATTTCCCGAGGTAAAATGCTATCCGGCAACATAAGT ACGAACTCGGTCCGCGGATCTAATAATGGATATTCCGCAAAAGAAAAAAAACATAAGGCACATG GTAAGAGGTCCAATTTACCAAAGGCCACCGTTTCAATTCTAAACAAATGGTTACATGAGCACGT AAACAACCCTTACCCAACCGTGCAGGAAAAAAGAGAACTGCTCGCGAAAACTGGTCTAACTAAA CTTCAAATTTCCAATTGGTTCATTAATGCTAGGAGAAGAAAAATATTTTCTGGCCAGAATGACG CAAATAATTTCAGAAGAAAATTCAGTTCTTCTACAAATTTAGCTAAGTTCTGA		
SS4c-C4b fusion protein, cucurbitadienol synthase	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMGTSIVNLNQKIELPPIQVLFESLNRENETKPHFEERRLYQPNPSF VPRTNIAVGSPVNPVPVSSPVFFIGPSPQRSIQNHNAIMTQNIRQYPVIYNNNREVISTGERNY IITVGGPPVTSSQPEYEHISTPNFYQEQRLAQPHPVNESMMIGGYTNPQPISISRGKMLSGNIS TNSVRGSNNGYSAKEKKHKAHGKRSNLPKATVSILNKWLHEHVNNPYPTVQEKRELLAKTGLTK LQISNWFINARRRKIFSGQNDANNFRRKFSSSTNLAKF-	975	
SS4e-B2 fusion protein, coding sequence	ATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCA TTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGA TGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGT AATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGA TCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAA AGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGAC GGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTAT ACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACAT CTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTT	978	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATG GTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTG GGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCC GAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAA TGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGT TTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGA AACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCA GTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGC CATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGA CCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCA AGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGG TTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTG ATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAG AGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAG CACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATG TTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAG TCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTC CTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGT TATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGC ATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAA AACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATC AAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACT TCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAA AGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCC TTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAA TCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAA CTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCC CATCGTGTCTTGGATATGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGGACCAGC CAAGGACCATTTAAGTAA		
SS4e-B2 fusion protein, cucurbitadienol synthase	MWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEAIAVANNSASKFENARNHFR NNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSD GNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMF GSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGKLWLSVLGVYEWSGNNPLPP EFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSR NTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMKIAMEHIHYEDENSRYICLG PVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKL IDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLM LSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYPWLELINPAETFGDIVIDYS	979	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	YVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGI KGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMA LIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYS HRVLDMGSAAPGSGSGSGMDQPRTI-V		
SS4e-B2 fusion protein, fusion domain	MDQPRTI	980	
SS5a-E8 fusion protein, coding sequence	ATGACTACGGATAACGCAGCAGCATATTCAGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAG GAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAG CATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGAC GATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCC GTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGA GATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAG AAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTG ACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCT ATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTAC ATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGT TTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCA TGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCC TGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCAC CCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAG AATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATA GTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCA GAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGG CAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAG GCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTG GACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTT CAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAG GGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAAT TGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCA AGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTT AGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGA TGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGC AGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGG TCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACA GTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGG GCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAG	982	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCA TCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAA CTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAAC AAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGG CCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCT AATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAG AACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACT CCCATCGTGTCTTGGATATGTGA		
SS5a-E8 fusion protein, cucurbitadienol synthase	MTTDNAAAYSGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDD DEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVK KEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRY IYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITS WGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPI VLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREK AMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQ GYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPF STRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTR SYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQ KTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQN KVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNK NCMITYAAYRNIFPIWALGEYSHRVLDM-	983	
SS5a-E8 fusion protein, fusion domain	MTTDNAAAYS	984	
SS5d-E7 fusion protein, coding sequence	ATGGCTCTCGGTAATGAGATAGCCAACTTGCAAGAGGGAAGTGCAGCTCCTGGAAGTGGAAGTG GTTCAGGAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGAT TAAAAGCATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAAC GACGACGATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATC ACTTCCGTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGA GAAAGAGATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGAT GTTAAGAAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGA CCTCTGACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTAT TGCGCTATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGT CGTTACATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTA CTATGTTTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGG TGAGCATGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATA ACTTCCTGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCAT TGCCACCCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCA	986	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	TTGTAGAATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACT CCAATAGTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATA GATCCAGAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCT ATGGGGCAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGG GAAAAGGCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATAT GCCTTGGACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGA TGCTTTCAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGA ATGCAGGGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAA CGAAATTGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCA GATTCAAGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGG CCTTTTAGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTT CACTGATGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATG TGATGCAGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTA ACTAGGTCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTG ACTACAGTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCA CCCTGGGCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAAT ATGCAGAAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGT TTGGCATCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGC TTGTAACTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGC CAAAACAAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCA TGATGGCCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAG ATTGCTAATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTT AATAAGAACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTG AATACTCCCATCGTGTCTTGGATATGTGA		
SS5d-E7 fusion protein, cucurbitadienol synthase	MALGNEIANLQEGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENEN DDDEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGED VKKEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMC RYIYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAI TSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPIT PIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLR EKAMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMR MQGYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAW PFSTRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYEL TRSYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLEN MQKTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSC QNKVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVF NKNCMITYAAYRNIFPIWALGEYSHRVLDM-	987	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
SS5d-E7 fusion protein, fusion domain	MALGNEIANLQE	988	
SS5d-G5 fusion protein, coding sequence	ATGCCAGTCTCTGTAATAACCACGTCAACACAGCCACATGTGAAGGAGCCTGTGGAAGAAGAGA GTGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGTGGAGACTGAAAGTAGGCAAAGA ATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCATTAGTAACCATTTGGGCAGACAAGTC TGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGATGAAGCAATTGCTGTAGCTAACAATT CAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGTAATAACAGATTCCATAGGAAGCAATC TTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGATCATTAGGAATGGTGCTAAGAATGAA GGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAAAGAAGCCGTAAAAAATACACTAGAAA GAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGACGGTAATTGGGCATCAGACTTGGGAGG ACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTATACGTTACAGGCGTTCTTAACAGTGTG TTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACATCTATAATCACCAAAACGAAGATGGAG GGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTTGGGTCTGCACTGAATTACGTTGCGTT AAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATGGTGCGATGACCAAAGCAAGGTCCTGG ATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTGGGGCAAACTTTGGCTTTCCGTATTAG GAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCCGAATTTTGGTTACTTCCATATAGCCT TCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAATGGTTTATCTGCCAATGTCGTATCTT TATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGTTTTATCGTTGAGAAAAGAGTTATATA CCATTCCGTACCACGAGATTGATTGGAATAGATCCAGAAACACGTGTGCTAAGGAGGACTTATA TTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCAGTATCTACCATGTCTACGAACCGCTA TTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGCCATGAAAATTGCAATGGAACATATCC ATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGACCCGTGAACAAAGTTCTAAATATGCT GTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCAAGTTTCACTTGCAGAGAATTCCTGAC TATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGGTTATAATGGTTCACAGCTTTGGGACA CTGCATTCAGCATACAAGCGATAATCTCAACGAAATTGATTGATACGTTTGGTCCGACATTACG TAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAGAGGATTGTCCAGGTGATCCAAACGTA TGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAGCACTAGAGATCATGGTTGGCTTATAT CGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATGTTATCTAAGTTGCCTTCTAAAATTGT GGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAGTCAATGTTTTGTTATCATTGCAAAAC GAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTCCTATCCCTGGTTAGAGTTGATTAACC CTGCCGAAACATTTGGTGATATCGTAATTGACTACAGTTATGTGGAATGCACTAGTGCGACGAT GGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGCATAGAACCAAAGAAATTGATGCCGCT ATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAAAACAGACGGGTCTTGGTATGGTTGTT GGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATCAAAGGCTTAGTAGCTGCTGGTAGAAC ATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACTTCCTGTTGTCAAAAGAATTGCCTGGC GGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAAAGTTTATACCAATCTGGAGGGAAACA AGCCACACTTAGTCAATACTGCATGGGTCATGATGGCCTTGATTGAAGCAGGACAAGGGGAAAG	990	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAATCAATAGTCAATTGGAGTCAGGTGAC TTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAACTGTATGATAACATATGCCGCATACA GAAACATATTCCCTATATGGGCTCTAGGTGAATACTCCCATCGTGTCTTGGATATGTGA		
SS5d-G5 fusion protein, cucurbitadienol synthase	MPVSVITTSTQPHVKEPVEEESGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQV WEFCSGENENDDDEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNE GTTKVKEGEDVKKEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSV LSKHHRQEMCRYIYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSW ILERGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYL YGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPL FSGWPGKRLREKAMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPD YLWLAEDGMRMQGYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNV WFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQN ENGGFASYELTRSYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAA IAKAANFLENMQKTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPG GGWGESYLSCQNKVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGD FPQQEIMGVFNKNCMITYAAYRNIFPIWALGEYSHRVLDM-	991	
SS5d-G5 fusion protein, fusion domain	MPVSVITTSTQPHVKEPVEEES	992	
SS5d-G7 fusion protein, coding sequence	ATGTCATCGAGCAAGAAAATCACCAGTGTCAAACAAGGAAGTGCAGCTCCTGGAAGTGGAAGTG GTTCAGGAATGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGAT TAAAAGCATTAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAAC GACGACGATGAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATC ACTTCCGTAATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGA GAAAGAGATCATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGAT GTTAAGAAAGAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGA CCTCTGACGGTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTAT TGCGCTATACGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGT CGTTACATCTATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTA CTATGTTTGGGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGG TGAGCATGGTGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATA ACTTCCTGGGGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCAT TGCCACCCGAATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCA TTGTAGAATGGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACT CCAATAGTTTTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATA GATCCAGAAACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCT ATGGGGCAGTATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGG GAAAAGGCCATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATAT	994	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	GCCTTGGACCCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGA TGCTTTCAAGTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGA ATGCAGGGTTATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAA CGAAATTGATTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCA GATTCAAGAGGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGG CCTTTTAGCACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTT CACTGATGTTATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATG TGATGCAGTCAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTA ACTAGGTCCTATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTG ACTACAGTTATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCA CCCTGGGCATAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAAT ATGCAGAAAACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGT TTGGCATCAAAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGC TTGTAACTTCCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGC CAAAACAAAGTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCA TGATGGCCTTGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAG ATTGCTAATCAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTT AATAAGAACTGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTG AATACTCCCATCGTGTCTTGGATATGTGA		
SS5d-G7 fusion protein, cucurbitadienol synthase	MSSSKKITSVKQGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENEN DDDEAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGED VKKEAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMC RYIYNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAI TSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPIT PIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLR EKAMKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMR MQGYNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAW PFSTRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYEL TRSYPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLEN MQKTDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSC QNKVYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVF NKNCMITYAAYRNIFPIWALGEYSHRVLDM-	995	
SS5d-G7 fusion protein, fusion domain	MSSSKKITSVKQ	996	
SS5e-C10 fusion protein, coding sequence	ATGCGAGGCTTGACACCTAAGGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGTGGA GACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCATTAGTAA CCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGATGAAGCA	998	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	ATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGTAATAACA GATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGATCATTAG GAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAAAGAAGCC GTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGACGGTAATT GGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTATACGTTAC AGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACATCTATAAT CACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTTGGGTCTG CACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATGGTGCGAT GACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTGGGGCAAA CTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCCGAATTTT GGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAATGGTTTA TCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGTTTTATCG TTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGAAACACGT GTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCAGTATCTA CCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGCCATGAAA ATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGACCCGTGA ACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCAAGTTTCA CTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGGTTATAAT GGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTGATTGATA CGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAGAGGATTG TCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAGCACTAGA GATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATGTTATCTA AGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAGTCAATGT TTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTCCTATCCC TGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGTTATGTGG AATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGCATAGAAC CAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAAAACAGAC GGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATCAAAGGCT TAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACTTCCTGTT GTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAAAGTTTAT ACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCCTTGATTG AAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAATCAATAG TCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAACTGTATG ATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCCCATCGTG TCTTGGATATGTGA		
SS5e-C10 fusion protein, cucurbitadienol	MRGLTPKGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDEA IAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKEA VKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIYN	999	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
synthase	HQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWGK LWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLS LRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAMK IAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGYN GSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFSTR DHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSYP WLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKTD GSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKVY TNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNCM ITYAAYRNIFPIWALGEYSHRVLDM-		
SS5e-C10 fusion protein, fusion domain	MRGLTPK	1000	
SS5e-G8 fusion protein, coding sequence	ATGGCAAACCAAATAGCAAATCAAGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGT GGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCATTAG TAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGATGAA GCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGTAATA ACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGATCAT TAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAAAGAA GCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGACGGTA ATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTATACGT TACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACATCTAT AATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTTGGGT CTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATGGTGC GATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTGGGGC AAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCCGAAT TTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAATGGT TTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGTTTTA TCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGAAACA CGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCAGTAT CTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGCCATG AAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGACCCG TGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCAAGTT TCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGGTTAT AATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTGATTG ATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAGAGGA TTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAGCACT AGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATGTTAT	1002	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	CTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAGTCAA TGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTCCTAT CCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGTTATG TGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGCATAG AACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAAAACA GACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATCAAAG GCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACTTCCT GTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAAAGTT TATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCCTTGA TTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAATCAA TAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAACTGT ATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCCCATC GTGTCTTGGATATGTGA		
SS5e-G8 fusion protein, cucurbitadienol synthase	MANQIANQGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDE AIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKE AVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIY NHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWG KLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVL SLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAM KIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGY NGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFST RDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSY PWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKT DGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKV YTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNC MITYAAYRNIFPIWALGEYSHRVLDM	1003	
SS5e-G8 fusion protein, fusion domain	MANQIANQ	1004	
SS5f-E11 fusion protein, coding sequence	ATGGTTGACGCTAGGGGTAGCAACGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAATGT GGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCATTAG TAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGATGAA GCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGTAATA ACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGATCAT TAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAAAGAA GCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGACGGTA ATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTATACGT TACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACATCTAT	1006	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTTGGGT CTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATGGTGC GATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTGGGGC AAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCCGAAT TTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAATGGT TTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGTTTTA TCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGAAACA CGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCAGTAT CTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGCCATG AAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGACCCG TGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCAAGTT TCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGGTTAT AATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTGATTG ATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAGAGGA TTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAGCACT AGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATGTTAT CTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAGTCAA TGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTCCTAT CCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGTTATG TGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGCATAG AACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAAAACA GACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATCAAAG GCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACTTCCT GTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAAAGTT TATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCCTTGA TTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAATCAA TAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAACTGT ATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCCCATC GTGTCTTGGATATGTGA		
SS5f-E11 fusion protein, cucurbitadienol synthase	MVDARGSNGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDDE AIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKKE AVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYIY NHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSWG KLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVL SLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKAM KIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQGY NGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFST RDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRSY	1007	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	PWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQKT DGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNKV YTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKNC MITYAAYRNIFPIWALGEYSHRVLDM-		
SS5f-E11 fusion protein, fusion domain	MVDARGSN	1008	
SS5f-F8 fusion protein, coding sequence	ATGCACGGCAAAGAGTTGGCTGGGCTAGGAAGTGCAGCTCCTGGAAGTGGAAGTGGTTCAGGAA TGTGGAGACTGAAAGTAGGCAAAGAATCTGTTGGCGAAAAGGAAGAAAAGTGGATTAAAAGCAT TAGTAACCATTTGGGCAGACAAGTCTGGGAGTTTTGCTCTGGTGAAAATGAAAACGACGACGAT GAAGCAATTGCTGTAGCTAACAATTCAGCCTCAAAATTTGAAAATGCAAGAAATCACTTCCGTA ATAACAGATTCCATAGGAAGCAATCTTCCGACTTATTCTTGGCTATACAATGCGAGAAAGAGAT CATTAGGAATGGTGCTAAGAATGAAGGTACTACCAAAGTGAAAGAAGGTGAAGATGTTAAGAAA GAAGCCGTAAAAAATACACTAGAAAGAGCATTGTCGTTCTATTCTGCTGTACAGACCTCTGACG GTAATTGGGCATCAGACTTGGGAGGACCTATGTTCCTTTTACCAGGGCTAGTTATTGCGCTATA CGTTACAGGCGTTCTTAACAGTGTGTTGTCAAAACATCACAGGCAAGAAATGTGTCGTTACATC TATAATCACCAAAACGAAGATGGAGGGTGGGGTTTACATATAGAAGGCTCTTCTACTATGTTTG GGTCTGCACTGAATTACGTTGCGTTAAGGTTACTAGGAGAAGCGGCAGATGGCGGTGAGCATGG TGCGATGACCAAAGCAAGGTCCTGGATATTGGAAAGAGGAGGTGCCACAGCAATAACTTCCTGG GGCAAACTTTGGCTTTCCGTATTAGGAGTGTATGAATGGTCGGGAAACAATCCATTGCCACCCG AATTTTGGTTACTTCCATATAGCCTTCCCTTTCATCCAGGGAGAATGTGGTGTCATTGTAGAAT GGTTTATCTGCCAATGTCGTATCTTTATGGTAAGAGATTTGTTGGTCCAATCACTCCAATAGTT TTATCGTTGAGAAAAGAGTTATATACCATTCCGTACCACGAGATTGATTGGAATAGATCCAGAA ACACGTGTGCTAAGGAGGACTTATATTACCCTCACCCTAAGATGCAAGATATCCTATGGGGCAG TATCTACCATGTCTACGAACCGCTATTTTCAGGTTGGCCAGGTAAGAGATTGAGGGAAAAGGCC ATGAAAATTGCAATGGAACATATCCATTACGAAGATGAGAATTCCAGGTACATATGCCTTGGAC CCGTGAACAAAGTTCTAAATATGCTGTGTTGCTGGGTTGAGGATCCTTACTCTGATGCTTTCAA GTTTCACTTGCAGAGAATTCCTGACTATCTATGGTTAGCTGAAGATGGAATGAGAATGCAGGGT TATAATGGTTCACAGCTTTGGGACACTGCATTCAGCATACAAGCGATAATCTCAACGAAATTGA TTGATACGTTTGGTCCGACATTACGTAAAGCACACCATTTCGTCAAGCATAGTCAGATTCAAGA GGATTGTCCAGGTGATCCAAACGTATGGTTCAGACACATTCATAAAGGAGCTTGGCCTTTTAGC ACTAGAGATCATGGTTGGCTTATATCGGATTGCACAGCTGAGGGATTGAAAGCTTCACTGATGT TATCTAAGTTGCCTTCTAAAATTGTGGGTGAACCCTTGGAGAAGAACCGTTTATGTGATGCAGT CAATGTTTTGTTATCATTGCAAAACGAAAATGGTGGGTTCGCTTCTTATGAATTAACTAGGTCC TATCCCTGGTTAGAGTTGATTAACCCTGCCGAAACATTTGGTGATATCGTAATTGACTACAGTT ATGTGGAATGCACTAGTGCGACGATGGAAGCCTTAGCCTTGTTTAAGAAGTTGCACCCTGGGCA TAGAACCAAAGAAATTGATGCCGCTATTGCTAAAGCCGCAAATTTTCTGGAGAATATGCAGAAA ACAGACGGGTCTTGGTATGGTTGTTGGGGTGTCTGTTTTACTTATGCTGGTTGGTTTGGCATCA	1010	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
	AAGGCTTAGTAGCTGCTGGTAGAACATACAATAATTGTGTTGCCATTAGAAAGGCTTGTAACTT CCTGTTGTCAAAAGAATTGCCTGGCGGTGGTTGGGGCGAAAGCTACCTAAGTTGCCAAAACAAA GTTTATACCAATCTGGAGGGAAACAAGCCACACTTAGTCAATACTGCATGGGTCATGATGGCCT TGATTGAAGCAGGACAAGGGGAAAGAGATCCAGCTCCGTTGCATAGAGCTGCCAGATTGCTAAT CAATAGTCAATTGGAGTCAGGTGACTTTCCACAACAAGAAATCATGGGTGTGTTTAATAAGAAC TGTATGATAACATATGCCGCATACAGAAACATATTCCCTATATGGGCTCTAGGTGAATACTCCC ATCGTGTCTTGGATATGTGA		
SS5f-F8 fusion protein, cucurbitadienol synthase	MHGKELAGLGSAAPGSGSGSGMWRLKVGKESVGEKEEKWIKSISNHLGRQVWEFCSGENENDDD EAIAVANNSASKFENARNHFRNNRFHRKQSSDLFLAIQCEKEIIRNGAKNEGTTKVKEGEDVKK EAVKNTLERALSFYSAVQTSDGNWASDLGGPMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYI YNHQNEDGGWGLHIEGSSTMFGSALNYVALRLLGEAADGGEHGAMTKARSWILERGGATAITSW GKLWLSVLGVYEWSGNNPLPPEFWLLPYSLPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIV LSLRKELYTIPYHEIDWNRSRNTCAKEDLYYPHPKMQDILWGSIYHVYEPLFSGWPGKRLREKA MKIAMEHIHYEDENSRYICLGPVNKVLNMLCCWVEDPYSDAFKFHLQRIPDYLWLAEDGMRMQG YNGSQLWDTAFSIQAIISTKLIDTFGPTLRKAHHFVKHSQIQEDCPGDPNVWFRHIHKGAWPFS TRDHGWLISDCTAEGLKASLMLSKLPSKIVGEPLEKNRLCDAVNVLLSLQNENGGFASYELTRS YPWLELINPAETFGDIVIDYSYVECTSATMEALALFKKLHPGHRTKEIDAAIAKAANFLENMQK TDGSWYGCWGVCFTYAGWFGIKGLVAAGRTYNNCVAIRKACNFLLSKELPGGGWGESYLSCQNK VYTNLEGNKPHLVNTAWVMMALIEAGQGERDPAPLHRAARLLINSQLESGDFPQQEIMGVFNKN CMITYAAYRNIFPIWALGEYSHRVLDM-	1011	
SS5f-F8 fusion protein, fusion domain	MHGKELAGL	1012	
EXG1 YLR300W *Saccharomyces cerevisiae*	MLSLKTLLCTLLTVSSVLATPVPARDPSSIQFVHEENKKRYYDYDHGSLGEPIRGVNIGGWLLL EPYITPSLFEAFRTNDDNDEGIPVDEYHFCQYLGKDLAKSRLQSHWSTFYQEQDFANIASQGFN LVRIPIGYWAFQTLDDDPYVSGLQESYLDQAIGWARNNSLKVWVDLHGAAGSQNGFDNSGLRDS YKFLEDSNLAVTTNVLNYILKKYSAEEYLDTVIGIELINEPLGPVLDMDKMKNDYLAPAYEYLR NNIKSDQVIIIHDAFQPYNYWDDFMTENDGYWGVTIDHHHYQVFASDQLERSIDEHIKVACEWG TGVLNESHWTVCGEFAAALTDCTKWLNSVGFGARYDGSWVNGDQTSSYIGSCANNDDIAYWSDE RKENTRRYVEAQLDAFEMRGGWIIWCYKTESSLEWDAQRLMFNGLFPQPLTDRKYPNQCGTISN	1013	
EXG1 *Yarrowia lipolytica*	MKLTKLVALAGAALASPIQLVPREGSFLGFNYGSEKVHGVNLGGWFVLEPFITPSLFEAFGNND ANVPVDEYHYTAWLGKEEAEKRLTDHWNTWITEYDIKAIAENYKLNLVRIPIGYWAFSLLPNDP YVQGQEAYLDRALGWCRKYGVKAWVDVHGVPGSQNGFDNSGLRDHWDWPNADNVQHSINVINYI AGKYGAPEYNDIVVGIELVNEPLGPAIGMEVIEKYFQEGFWTVRHAGSDTAVVIHDAFQEKNYF NNFMTTEQGFWNVVLDHHQYQVFSPGELARNIDQHIAEVCNVGRQASTEYHWRIFGEWSAALTD CTHWLNGVGKGPRLDGSFPGSYYQRSCQGRGDIQTWSEQDKQESRRYVEAQLDAWEHGGDGWIY WTYKTENALEWDFRRLVDNGIFPFPYWDRQFPNQCGF	1014	

Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
Glugan 1, 4, alpha glucosidase	MPRLSYALCALSLGHAAIAAPQLSARATGSLDSWLGTETTVALNGILANIGADGAYAKSAKPGI IIASPSTSEPDYYYTWTRDAALVTKVLVDLFRNGNLGLQKVITEYVNSQAYLQTVSNPSGGLAS GGLAEPKYNVDMTAFTGAWGRPQRDGPALRATALIDFGNWLIDNGYSSYAVNNIWPIVRNDLSY VSQYWSQSGFDLWEEVNSMSFFTVAVQHRALVEGSTFAKRVGASCSWCDSQAPQILCYMQSFWT GSYINANTGGGRSGKDANTVLASIHTFDPEAGCDDTTFQPCSPRALANHKVYTDSFRSVYAINS GIPQGAAVSAGRYPEDVYYNGNPWFLTTLAAAEQLYDAIYQWKKIGSISITSTSLAFFKDIYSS AAVGTYASSTSTFTDIINAVKTYADGYVSIVQAHAMNNGSLSEQFDKSSGLSLSARDLTWSYAA FLTANMRRNGVVPAPWGAASANSVPSSCSMGSATGTYSTATATSWPSTLTSGSPGSTTTVGTTT STTSGTAAETACATPTAVAVTFNEIATTTYGENVYIVGSISELGNWDTSKAVALSASKYTSSNN LWYVSVTLPAGTTFEYKYIRKESDGSIVWESDPNRSYTVPAACGVSTATENDTWQ	1015	
polygalacturonase 1 [Aspergillus aculeatus] CAE46193.1 GI:34366090	MHLNTTLLVSLALGAASVLASPAPPAITAPPTAEEIAKRATTCTFSGSNGASSASKSKTSCSTI VLSNVAVPSGTTLDLTKLNDGTHVIFSGETTFGYKEWSGPLISVSGSDLTITGASGHSINGDGS RWWDGEGGNGGKTKPKFFAAHSLTNSVISGLKIVNSPVQVFSVAGSDYLTLKDITIDNSDGDDN GGHNTDAFDIGTSTYVTISGATVYNQDDCVAVNSGENIYFSGGYCSGGHGLSIGSVGGRSDNTV KNVTFVDSTIINSDNGVRIKTNIDTTGSVSDVTYKDITLTSIAKYGIVVQQNYGDTSSTPTTGV PITDFVLDNVHGSVVSSGTNILISCGSGSCSDWTWTDVSVSGGKTSSKCTNVPSGASC	1016	
polygalacturonase 2 [Aspergillus aculeatus] CAE46194.1 GI:34366092	MHSFQLLGLAAVGSVVSAAPTASRVSDLVKKSSSTCTFTSASEASETSSSCSNVVLSNIEVPAG ETLDLSDAADGATITFEGTTSFGYEEWDGPLIRFGGKQLTITQSDGAVIDGDGSRWWDSEGTNG GKTKPKFMYVHDVEDSTIKGLQIKNTPVQAISVQATNVYLTDITIDNSDGDDNGGHNTDGFDIS ESTGVYISGATVKNQDDCIAINSGENILFTGGTCSGGHGLSIGSVGGRDDNTVKNVTISDSTVT DSANGVRIKTIYGDTGDVSEITYSNIQLSGITDYGIVIEQDYENGSPTGTPSTGVPITDVTVDG VTGSIEDDAVQVYILCGDGSCSDWTWSGVDITGGETSSDCENVPSGASC	1017	
polygalacturonase 3 [Aspergillus aculeatus] CAE46195.1 GI:34366094	MVRQLALACGLLAAVAVQAAPAEPAHPMVTEAPDASLLHKRATTCTFSGSEGASKVSKSKTACS TIYLSALAVPSGTTLDLKDLNDGTHVIFEGETTFGYEEWEGPLVSVSGTDITVEGASGAVLNGD GSRWWDGEGGNGGKTKPKFFAAHDLTSSTIKSIYIENSPVQVFSIDGATDLTLTDITIDNTDGD TDDLAANTDGFDIGESTDITITGAKVYNQDDCVAINSGENIYFSASVCSGGHGLSIGSVGGRDD NTVKNVTFYDVNVLKSQQAIRIKAIYGDTGSISDITYHEIAFSDATDYGIVIEQNYDDTSKTPT TGVPITDFTLENVIGTCADDDCTEVYIACGSGACSDWSWSSVSVTGGKVSSKCLNVPSGISCDL	1018	
Chain A, Crystal Structure Of Polygalacturonase From Aspergillus Aculeatus At Ph4.5 1IB4_A GI:15988280	ATTCTFSGSNGASSASKSKTSCSTIVLSNVAVPSGTTLDLTKLNDGTHVIFSGETTFGYKEWSG PLISVSGSDLTITGASGHSINGDGSRWWDGEGGNGGKTKPKFFAAHSLTNSVISGLKIVNSPVQ VFSVAGSDYLTLKDITIDNSDGDDNGGHNTDAFDIGTSTYVTISGATVYNQDDCVAVNSGENIY FSGGYCSGGHGLSIGSVGGRSDNTVKNVTFVDSTIINSDNGVRIKTNIDTTGSVSDVTYKDITL TSIAKYGIVVQQNYGDTSSTPTTGVPITDFVLDNVHGSVVSSGTNILISCGSGSCSDWTWTDVS VSGGKTSSKCTNVPSGASC	1019	
Chain B, Crystal Structure Of Polygalacturonase	ATTCTFSGSNGASSASKSKTSCSTIVLSNVAVPSGTTLDLTKLNDGTHVIFSGETTFGYKEWSG PLISVSGSDLTITGASGHSINGDGSRWWDGEGGNGGKTKPKFFAAHSLTNSVISGLKIVNSPVQ VFSVAGSDYLTLKDITIDNSDGDDNGGHNTDAFDIGTSTYVTISGATVYNQDDCVAVNSGENIY	1020	

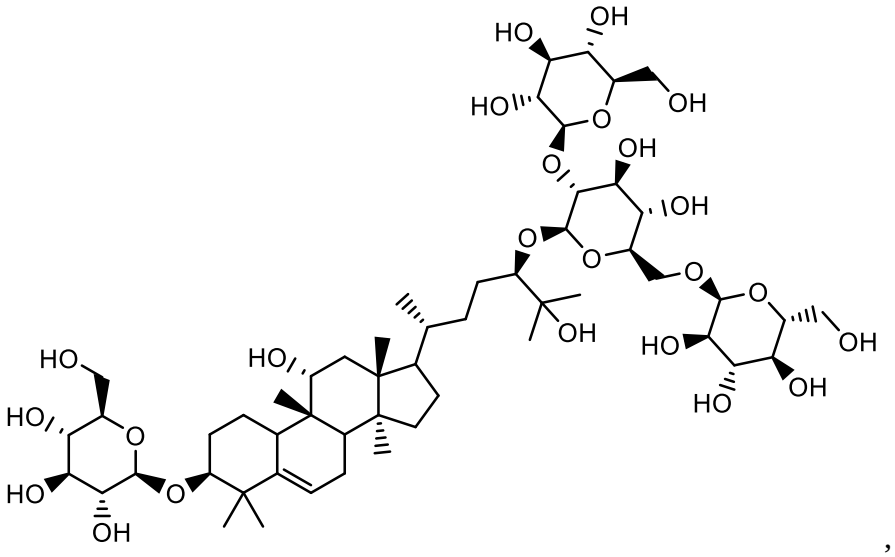
Protein/DNA Description	Protein/DNA Sequence	SEQ ID NO	Reference
From Aspergillus Aculeatus At Ph4.5 1IB4_B GI:15988281	FSGGYCSGGHGLSIGSVGGRSDNTVKNVTFVDSTIINSDNGVRIKTNIDTTGSVSDVTYKDITL TSIAKYGIVVQQNYGDTSSTPTTGVPITDFVLDNVHGSVVSSGTNILISCGSGSCSDWTWTDVS VSGGKTSSKCTNVPSGASC		
Chain A, Polygalacturonase From Aspergillus Aculeatus 1IA5_A GI:15988279	ATTCTFSGSNGASSASKSKTSCSTIVLSNVAVPSGTTLDLTKLNDGTHVIFSGETTFGYKEWSG PLISVSGSDLTITGASGHSINGDGSRWWDGEGGNGGKTKPKFFAAHSLTNSVISGLKIVNSPVQ VFSVAGSDYLTLKDITIDNSDGDDNGGHNTDAFDIGTSTYVTISGATVYNQDDCVAVNSGENIY FSGGYCSGGHGLSIGSVGGRSDNTVKNVTFVDSTIINSDNGVRIKTNIDTTGSVSDVTYKDITL TSIAKYGIVVQQNYGDTSSTPTTGVPITDFVLDNVHGSVVSSGTNILISCGSGSCSDWTWTDVS VSGGKTSSKCTNVPSGASC	1021	
polygalacturonase precursor [Aspergillus aculeatus] 378 aa protein AAC23565.1 GI:3220207	MHLNTTLLVSLALGAASVLASPAPPAITAPPTAEEIAKRATTCTFSGSNGASSASKSKTSCSTI VLSNVAVPSGTTLDLTKLNDGTHVIFSGETTFGYKEWSGPLISVSGSDLTITGASGHSINGDGS RWWDGEGGNGGKTKPKFFAAHSLTNSVISGLKIVNSPVQVFSVAGSDYLTLKDITIDNSDGDDN GGHNTDAFDIGTSTYVTISGATVYNQDDCVAVNSGENIYFSGGYCSGGHGLSIGSVGGRSDNTV KNVTFVDSTIINSDNGVRIKTNIDTTGSVSDVTYKDITLTSIAKYGIVVQQNYGDTSSTPTTGV PITDFVLDNVHGSVVSSGTNILISCGSGSCSDWTWTDVSVSGGKTSSKCTNVPSGASC	1022	
EXG2 YDR261C *Saccharomyces cerevisiae*	MPLKSFFFSAFLVLCLSKFTQGVGTTEKEESLSPLELNILQNKFASYYANDTITVKGITIGGWL VTEPYITPSLYRNATSLAKQQNSSSNISIVDEFTLCKTLGYNTSLTLLDNHFKTWITEDDFEQI KTNGFNLVRIPIGYWAWKQNTDKNLYIDNITFNDPYVSDGLQLKYLNNALEWAQKYELNVWLDL HGAPGSQNGFDNSGERILYGDLGWLRLNNTKELTLAIWRDMFQTFLNKGDKSPVVGIQIVNEPL GGKIDVSDITEMYYEAFDLLLKKNQNSSDNTTFVIHDGFQGIGHWNLELNPTYQNVSHHYFNLTG ANYSSQDILVDHHHYEVFTDAQLAETQFARIENIINYGDSIHKELSFHPAVVGEWSGAITDCAT WLNGVGVGARYDGSYYNTTLFTTNDKPVGTCISQNSLADWTQDYRDRVRQFIEAQLATYSSKTT GWIFWNWKTEDAVEWDYLKLKEANLFPSPFDNYTYFKADGSIEEKFSSSLSAQAFPRTTSSVLS STTTSRKSKNAAISNKLTTSQLLPIKNMSLTWKASVCALAITIAALCASL	1023	

[0418] While the invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. This includes embodiments which do not provide all of the benefits and features set forth herein. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto. Accordingly, the scope of the invention is defined only by reference to the appended claims.

CLAIMS

What is claimed is:

1. A method of producing Compound 1 having the structure of:



,

the method comprising:

contacting mogroside III_E with an enzyme capable of catalyzing a reaction for the production of Compound 1 from the mogroside III_E,

wherein the enzyme comprises an amino acid sequence having at least 90% sequence identity to any one of the sequences set forth in SEQ ID NOs: 2 and 103 having dextransucrase activity or wherein the enzyme having dextransucrase activity is encoded by a gene comprising a nucleic acid sequence having at least 90% sequence identity to any one of SEQ ID Nos: 104, and 105.

2. The method of Claim 1, wherein the mogroside III_E is contacted with a recombinant host cell recombinant host cell that comprises the gene comprising the nucleic acid sequence having at least 90% sequence identity to any one of SEQ ID Nos: 104, and 105 encoding the enzyme having dextransucrase activity.

3. The method of Claim 2, wherein the mogroside III_E is present in and/or produced by the recombinant host cell.