(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2024 (13.06.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/124165 A2

(51) International Patent Classification:
C12P 17/06 (2006.01) *C07D 311/78* (2006.01)

(21) International Application Number:
PCT/US2023/083173

(22) International Filing Date:
08 December 2023 (08.12.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/431,560 09 December 2022 (09.12.2022) US

(71) Applicant: **AMYRIS, INC.** [US/US]; 5885 Hollis Street, Suite 100, Emeryville, CA 94608 (US).

(72) Inventors: **TEO ONG, Cara**; C/o Amyris, Inc., 5885 Hollis Street, Suite 100, Emeryville, CA 94608 (US). **GALEN, Brady**; C/o Amyris, Inc., 5885 Hollis Street, Suite 100, Emeryville, CA 94608 (US). **LLOYD-RANDOLFI, Jenna**; C/o Amyris, Inc., 5885 Hollis Street, Suite 100, Emeryville, CA 94608 (US). **DANGER NILL, Jennifer**; C/o Amyris, Inc., 5885 Hollis Street, Suite 100, Emeryville, CA 94608 (US). **DRAPER, William**; C/o Amyris, Inc., 5885 Hollis Street, Suite 100, Emeryville, CA 94608 (US). **HOLMES, Victor**; C/o Amyris, Inc., 5885 Hollis Street, Suite 100, Emeryville, CA 94608 (US). **WOGULIS, Mark**; C/o Amyris, Inc., 5885 Hollis Street, Suite 100, Emeryville, CA 94608 (US).

(74) Agent: **ELBING, Karen, L.** et al.; Clark & Elbing LLP, 101 Federal Street, 15th Floor, Boston, MA 02110 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *of inventorship (Rule 4.17(iv))*

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

(54) Title: METHODS AND COMPOSITIONS FOR PURIFYING CANNABINOIDS

(57) Abstract: Provided herein methods of producing and isolating a cannabinoid from a fermentation composition including are host cells capable of producing a cannabinoid such by combining a mixture including a cannabigerolic acid with a cannabichromenic acid (CBCa) synthase, optionally in an oil mixture. Also provided herein are CBCa synthases having improved activity in comparison to the wild-type enzyme.

WO 2024/124165 A2

METHODS AND COMPOSITIONS FOR PURIFYING CANNABINOIDS

Sequence Listing

The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on November 27, 2023, is named 51494-027WO2_Sequence_Listing_11_27_23, and is 217,381 bytes in size.

BACKGROUND OF THE INVENTION

Processes developed for the production of minor cannabinoids typically fall into two categories: 1) the extraction and purification from the cannabis plant, and 2) the conversion of a cannabinoid precursor into various products, either chemically or biochemically. Processes typically involved in the purification of plant-derived cannabinoids include extraction from the cannabis plant, decarboxylation to convert the acidic cannabinoids by heating the extract, evaporation to concentrate the product, crystallization, and a variety of chromatography steps to remove residual impurities. However, the state-of-the-art production of high purity minor cannabinoids from plants requires additional processing steps, with removal of trichomes from plant material and further fractionation of glandular from sessile trichomes prior to extraction, which can increase cost substantially and may not be feasibly scalable. Additionally, plant derived minor cannabinoids are typically <3% of the total plant mass, are coproduced with other cannabinoids that are difficult and expensive to separate, and with yields that decrease as purity increases.

Conventional synthetic methods for producing cannabichromene (CBC) and other minor cannabinoids are known but they lack selectivity, are difficult to scale, and, additionally, rely on toxic solvents and reagents or expensive catalysts. As a result, there remains a need for improved synthetic methods for producing high purity minor cannabinoids.

SUMMARY OF THE INVENTION

The present disclosure features methods of isolating and purifying one or more cannabinoids from a fermentation composition. The methods described herein include for the first time using an oil overlay in combination with a cannabichromenic acid (CBCa) synthase. It has been presently discovered that this combination results in near complete conversion of cannabigerolic acid (CBGa) to CBCa. The present disclosure provides variant CBCa synthase polypeptides having one or more amino acid substitutions. The variant CBCa synthases described herein show improved enzymatic activity in comparison to the wild-type enzyme.

In an aspect, the disclosure provides a method of making a cannabichromenic acid (CBCa). In some embodiments, the method includes (a) culturing a population of host cells capable of producing cannabigerolic acid (CBGa) in a culture medium comprising a fermentation broth and an overlay, under conditions suitable for the host cells to produce CBGa, and wherein the CBGa partitions into the overlay; (b) separating the overlay from the fermentation broth; (c) combining the separated

overlay of step (b), with a CBCa synthase, thereby producing a bioconversion mixture; and (d) purifying the CBCa from the bioconversion mixture.

In some embodiments, the overlay comprises a plant-based oil. In some embodiments, the plant-based oil is selected from soybean oil, sunflower oil, safflower oil, canola oil, grapeseed oil, or castor oil. In some embodiments, the plant-based oil is sunflower oil. In some embodiments, the overlay comprises a synthetic ester or a fatty alcohol. In some embodiments, the CBCa synthase is produced by culturing a population of host cells capable of producing a CBCa synthase in a culture medium and under conditions suitable for the host cells to produce the CBCa synthase, thereby producing the CBCa synthase.

. In some embodiments, the overlay and the fermentation broth are separated by centrifugation. In some embodiments, the overlay and the fermentation broth are separated by demulsification.

In some embodiments, the demulsification includes contacting the fermentation broth with an oil. In some embodiments, the oil includes a mineral oil, a vegetable oil, a synthetic ester, or a fatty alcohol. In some embodiments, the oil includes a vegetable oil. In some embodiments, the vegetable oil is soybean oil, sunflower oil, safflower oil, canola oil, grapeseed oil, or castor oil. In some embodiments, the oil includes a synthetic ester, optionally wherein the synthetic ester is ESTEREX™ A51. In some embodiments, the oil includes a fatty alcohol, optionally wherein the fatty alcohol is oleyl alcohol or JARCOL™ I-16. In some embodiments, the oil has a concentration of between about 1% (w/v) and about 10% (w/v) (e.g., about 1% (w/v), 2% (w/v), 3% (w/v), 4% (w/v), 5% (w/v), 6% (w/v), 7% (w/v), 8% (w/v), 9% (w/v), or 10% (w/v)). In some embodiments, the oil has a concentration of about 5% (w/v).

In some embodiments, the overlay includes CBGa. In some embodiments, the CBGa has a concentration of between about 0.1% (w/v) or and 10% (w/v) (e.g., between 0.1% (w/v) and 8% (w/v), 0.1% (w/v) and 6% (w/v), 0.1% (w/v) and 4% (w/v), 0.1% (w/v) and 2% (w/v), 2% (w/v) and 10% (w/v), 4% (w/v) and 10% (w/v), 6% (w/v) and 10% (w/v), or 8% (w/v) and 10% (w/v)). In some embodiments, the CBGa has a concentration of between about 0.5% (w/v) and 5% (w/v) (e.g., between 0.5% (w/v) and 4% (w/v), 0.5% (w/v) and 3% (w/v), 0.5% (w/v) and 2% (w/v), 0.5% (w/v) and 1% (w/v), 1% (w/v) and 5% (w/v), 2% (w/v) and 5% (w/v), 3% (w/v) and 5% (w/v), or 4% (w/v) and 5% (w/v)). In some embodiments, the method further includes stirring the bioconversion mixture for between 12 hours and 144 hours before performing step (d) (e.g., between 12 hours 120 hours, 12 hours and 96 hours, 12 hours and 72 hours, 12 hours and 48 hours, 12 hours and 24 hours, 24 hours and 144 hours, 48 hours and 144 hours, 72 hours and 144 hours, 96 hours and 144 hours, or 120 hours and 144 hours). In some embodiments, the method includes stirring the bioconversion mixture for between 24 hours and 96 hours (e.g., between 24 hours and 84 hours, 24 hours and 72 hours, 24 hours and 60 hours, 24 hours and 48 hours, 48 hours and 96 hours, 60 hours and 96 hours, 72 hours and 96 hours, or 84 hours and 96 hours). In some embodiments, the method includes stirring the bioconversion mixture for about 48 hours. In some embodiments, the bioconversion mixture is at a temperature of between 4 °C and 50 °C. In some embodiments, the bioconversion mixture is at a temperature of between 20 °C and 40 °C. In some embodiments, the bioconversion mixture is at a temperature of about 35 °C.

In some embodiments, the second mixture further includes one or more amphiphilic moieties. In some embodiments, the one or more amphiphilic moieties includes a cyclodextrin, plant derived silica, cellulose, or a combination thereof. In some embodiments, the cyclodextrin includes randomly methylated cyclodextrin, 2, 6-Di-O-methyl-β-cyclodextrin, or a combination thereof.

In some embodiments, the demulsification includes one or more of: (i) contacting the culture medium with an enzymatic composition comprising a serine protease, (ii)contacting the culture medium with a surfactant; and (iii) contacting the culture medium with NaOH to adjust the culture medium to a pH of between pH 7 and pH 9. In some embodiments, the demulsification includes contacting the culture medium with the enzymatic composition including the serine protease. In some embodiments, the purifying comprises contacting the culture medium with the surfactant. In some embodiments, the culture medium is contacted with the enzymatic composition or surfactant after the mixture is adjusted to a pH of between about pH 7 and pH 9. In some embodiments, the culture medium is contacted with the enzymatic composition or surfactant after the culture medium is adjusted to a pH of pH 8. In some embodiments, the final concentration of the enzymatic composition is between about 0.5% (w/v) to about 3% (w/v) (e.g., between about 0.5% (w/v) and 2% (w/v), 0.5% (w/v) and 1% (w/v), 1% (w/v) and 3% (w/v), or 2% (w/v) and 3% (w/v)) after contacting the culture medium including a cannabinoid with the enzymatic composition. In some embodiments, the culture medium is contacted with the enzymatic composition at a final concentration of about 1% (w/v). In some embodiments, the culture medium is mixed with the enzymatic composition for between 0.5 hours and 2 hours (e.g., between 0.5 hours and 1.5 hours, 0.5 hours and 1 hour, 1 hour and 2 hours, or 1.5 hours and 2 hours).

In some embodiments, the demulsification includes centrifugation of the culture medium. In some embodiments, the centrifugation includes liquid-liquid centrifugation. In some embodiments, the centrifugation results in a crude oil light phase and an aqueous heavy phase. In some embodiments, the demulsification further includes a decarboxylation step comprising evaporating the culture medium. In some embodiments, the decarboxylation includes evaporating the crude oil light phase. In some embodiments, evaporating includes one or more passes. In some embodiments, evaporating includes a first pass and a second pass. In some embodiments, the first pass is performed at a temperature of between about 100 °C and about 500 °C (e.g., between about 100 °C and 400 °C, 100 °C and 300 °C, 100 °C and 200 °C, 200 °C and 500 °C, 300 °C and 500 °C, or 400 °C and 500 °C). In some embodiments, the first pass is performed at a temperature of about 180 °C. In some embodiments, the first pass is performed at a pressure of between about 0.5 torr and 760 torr (e.g., between 0.5 torr and 700 torr, 0.5 torr and 500 torr, 0.5 torr and 200 torr, 0.5 torr and 50 torr, 50 torr and 760 torr, 200 torr and 760 torr, 400 torr and 760 torr, or 600 torr and 760 torr). In some embodiments, the first pass is performed at a pressure of about 1 torr.

In some embodiments, the second pass is performed at a temperature of between 150 °C and 300 °C (e.g., between 150 °C and 250 °C, 150 °C and 200 °C, 200 °C and 300 °C, or 250 °C and 300 °C). In some embodiments, the second pass is performed at a temperature of about 240 C. In some embodiments, the second pass is performed at a pressure of between about 0.5 torr and 760 torr (e.g., between 0.5 torr and 700 torr, 0.5 torr and 500 torr, 0.5 torr and 200 torr, 0.5 torr and 50

torr, 50 torr and 760 torr, 200 torr and 760 torr, 400 torr and 760 torr, or 600 torr and 760 torr). In some embodiments, the second pass is performed at a pressure of about 1 torr.

In some embodiments, purifying includes one or more of a liquid-liquid extraction, chromatography, or saponification.

In some embodiments, the host cells include a nucleic acid sequence encoding a cannabichromenic acid (CBCa) synthase. In some embodiments, CBCa synthase includes one or more amino acid substitutions relative to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the CBCa synthase includes an amino acid substitution at a residue selected from Q75, F82, T130, S140, V169, N240, V294, A299, K305, T335, R340, H354, L435, Y461, K535, S540, and T545 of SEQ ID NO: 1.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue Q75 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 substitutes Q75 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 is a Q75L substitution. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 substitutes Q75 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 is a Q75E substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue S140 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue S140 of SEQ ID NO: 1 substitutes S140 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue S140 of SEQ ID NO: 1 is a S140A substitution. In some embodiments, the amino acid substitution at residue S140 of SEQ ID NO: 1 is a S140T substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue V169 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue V169 of SEQ ID NO: 1 substitutes V169 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue V169 of SEQ ID NO: 1 is a V169E substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue N240 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue N240 of SEQ ID NO: 1 substitutes N240 with an amino acid including a polar, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue N240 of SEQ ID NO: 1 is a N240Q substitution. In some embodiments, the amino acid substitution at residue N240 of SEQ ID NO: 1 substitutes N240 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue N240 of SEQ ID NO: 1 is a N240M substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue V294 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 substitutes V294 with an amino acid including a polar, uncharged side

chain at physiological pH. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 is a V294S substitution. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 substitutes V294 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 is a V294E substitution. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 substitutes V294 with an amino acid including a cationic side chain at physiological pH. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 is a V294R substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue A299 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue A299 of SEQ ID NO: 1 substitutes A299 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue A299 of SEQ ID NO: 1 is an A299V substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue K305 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue K305 of SEQ ID NO: 1 substitutes K305 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue K305 of SEQ ID NO: 1 is a K305C substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue D328 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue D328 of SEQ ID NO: 1 is a D328P substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue T335 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue T335 of SEQ ID NO: 1 substitutes T335 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue T335 of SEQ ID NO: 1 is a T335L substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue R340 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue R340 of SEQ ID NO: 1 substitutes R340 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue R340 of SEQ ID NO: 1 is a R340M substitution. In some embodiments, the amino acid substitution at residue R340 of SEQ ID NO: 1 is a R340G substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue H354 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue H354 of SEQ ID NO: 1 substitutes H354 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue R340 of SEQ ID NO: 1 is a H354V substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue L435 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue L435 of SEQ ID NO: 1 substitutes L435 with an amino acid including a hydrophobic,

uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue L435 of SEQ ID NO: 1 is a L435A substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue Y461 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue Y461 of SEQ ID NO: 1 substitutes Y461 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue Y461 of SEQ ID NO: 1 is a Y461I substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue K535 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue K535 of SEQ ID NO: 1 substitutes K535 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue Y461 of SEQ ID NO: 1 is a K535M substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue S540 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue S540 of SEQ ID NO: 1 substitutes S540 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue S540 of SEQ ID NO: 1 is a S540D substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue T545 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue T545 of SEQ ID NO: 1 substitutes T545 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue T545 of SEQ ID NO: 1 is a T545E substitution.

In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A Y461I, K535, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, V294S, A299V, K305C, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, T335L, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, T130L, S140A, V169E, N240M, V294S, A299V, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, T130L, S140A, V169E, N240M, V294S, A299V, T335L, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, N240M, V294S, A299V, K305C, T335L, R340M, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V294S, A299V, K305C, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, S140A, V169E, N240M, K305C, T335L, R340M,

H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, V169E, V294S, A299V, R340M, H354V, L435A, K535M, or S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, V294S, A299V, T335L, H354V, L435A, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, T130L, S140A, V169E, V294S, A299V, K305C, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, A299V, K305C, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, N240M, V294S, T335L, R340M, H354V, L435A, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V294S, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, T335L, R340M, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including T130L, S140A, V169E, V294S, A299V, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, N240M, V294S, A299V, T335L, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, R340M, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, A299V, K305C, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, T130L, S140A, V169E, V294S, A299V, K305C, H354V,

L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, V169E, V294S, A299V, K305C, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, K305C, T335L, R340M, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including F82I, T130L, S140A, V169E, N240M, A299V, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, V294S, A299V, K305C, T335L, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, A299V, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, V169E, N240M, V294S, A299V, K305C, T335L, L435A, Y461I, and K535M of SEQ ID NO: 1.

In some embodiments, the CBCa synthase has an amino acid sequence that is from about 85% to about 99.7% identical to the amino acid sequence of SEQ ID NO: 1 (e.g., about 85%, 96%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.7%). In some embodiments, the CBCa synthase has an amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the CBCa synthase has an amino acid sequence that is from about 90% to about 99.7% (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.7%) identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the CBCa synthase has an amino acid sequence that is from about 95% to about 99.7% (e.g., about 95%, 95.5%, 96%, 96.5%, 97%, 97.5%, 98%, 98.5%, 99%, 99.5%, or 99.7%) identical to the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the CBCa synthase has an amino acid sequence that differs from the amino acid sequence of SEQ ID NO: 1 only by way of (i) the one or more amino acid substitutions or deletions and, optionally, (ii) one or more additional, conservative amino acid substitutions. In some embodiments, the CBCa synthase has an amino acid sequence that differs from the amino acid sequence of SEQ ID NO: 1 only by way of the one or more amino acid substitutions or deletions. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 85% identical to the amino acid sequence of any one of SEQ ID NO: 2-67. In some embodiments, the CBCa synthase

has an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 2-67. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 2-67. In some embodiments, the CBCa synthase has the amino acid sequence of any one of SEQ ID NO: 2-67.

In some embodiments, the CBCa synthase has an amino acid sequence that is at least 85% identical to the amino acid sequence of any one of SEQ ID NO: 29-67. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 29-67. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 29-67. In some embodiments, the CBCa synthase has the amino acid sequence of any one of SEQ ID NO: 29-67.

In some embodiments, the host cell includes one or more heterologous nucleic acids that each, independently, encode (a) an acyl activating enzyme (AAE), and/or (b) a tetraketide synthase (TKS), and/or (c) a cannabigerolic acid synthase (CBGaS), and/or (d) a geranyl pyrophosphate (GPP) synthase.

In some embodiments, the host cell includes heterologous nucleic acids that independently encode (a) an AAE, (b) a TKS, (c) a CBGaS, and (d) a GPP synthase.

In some embodiments, the host cell includes a heterologous nucleic acid that encodes an AAE having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 68-91. In some embodiments, the AAE has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 68-91. In some embodiments, the AAE has the amino acid sequence of any one of SEQ ID NO: 68-91. In some embodiments, the host cell includes a heterologous nucleic acid that encodes an AAE having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 68-80. In some embodiments, the AAE has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 68-80. In some embodiments, the AAE has the amino acid sequence of any one of SEQ ID NO: 68-80. In some embodiments, the host cell includes a heterologous nucleic acid that encodes an AAE having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 68-72. In some embodiments, the AAE has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 68-72. In some embodiments, the AAE has the amino acid sequence of any one of SEQ ID NO: 68-72.

In some embodiments, the host cell includes a heterologous nucleic acid that encodes a TKS having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 92-126. In some embodiments, the TKS has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 92-126. In some embodiments, the TKS has the amino acid sequence of any one of SEQ ID NO: 92-126. In some embodiments, the host cell includes a heterologous nucleic acid that encodes a TKS having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 92-95. In some embodiments, the TKS has an amino acid sequence that is at least 95% identical to the amino acid

sequence of any one of SEQ ID NO: 92-95. In some embodiments, the TKS has the amino acid sequence of any one of SEQ ID NO: 92-95. In some embodiments, the host cell includes a heterologous nucleic acid that encodes a TKS having an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 92. In some embodiments, the TKS has an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 92, optionally wherein the TKS has the amino acid sequence of SEQ ID NO: 92.

In some embodiments, the host cell includes a heterologous nucleic acid that encodes a CBGaS having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 127-131. In some embodiments, the CBGaS has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 127-131. In some embodiments, the CBGaS has the amino acid sequence of any one of SEQ ID NO: 127-131.

In some embodiments, the host cell includes a heterologous nucleic acid encoding a GPP synthase having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 132-137. In some embodiments, the GPP synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 132-137. In some embodiments, the GPP has the amino acid sequence of any one of SEQ ID NO: 132-137. In some embodiments, the host cell includes a heterologous nucleic acid encoding a GPP synthase having an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 132. In some embodiments, the GPP synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 132. In some embodiments, the GPP has the amino acid sequence of SEQ ID NO: 132.

In some embodiments, the host cell includes heterologous nucleic acids that independently encode an AAE having the amino acid sequence of any one of SEQ ID NO: 68-91, a TKS having the amino acid sequence of any one of SEQ ID NO: 92-126, a CBGaS having the amino acid sequences of any one of SEQ ID NO: 127-131, and a GPP synthase having the amino acid sequence of any one of SEQ ID NO: 132-137.

In some embodiments, the host cell further includes one or more heterologous nucleic acids that each, independently, encode an enzyme of the mevalonate biosynthetic pathway, wherein the enzyme is selected from an acetyl-CoA thiolase, an HMG-CoA synthase, an HMG-CoA reductase, a mevalonate kinase, a phosphomevalonate kinase, a mevalonate pyrophosphate decarboxylase, and an IPP:DMAPP isomerase. In some embodiments, the host cell includes heterologous nucleic acids that independently encode an acetyl-CoA thiolase, an HMG-CoA synthase, an HMG-CoA reductase, a mevalonate kinase, a phosphomevalonate kinase, a mevalonate pyrophosphate decarboxylase, and an IPP:DMAPP isomerase.

In some embodiments, the host cell further includes a heterologous nucleic acid that encodes an olivetolic acid cyclase (OAC). In some embodiments, the OAC has an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 138. In some embodiments, the OAC has an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 138. In some embodiments, the OAC has the amino acid sequence of SEQ ID NO: 138.

In some embodiments, the host cell further includes one or more heterologous nucleic acids that each, independently, encode an acetyl-CoA synthase, and/or an aldehyde dehydrogenase, and/or a pyruvate decarboxylase. In some embodiments, the acetyl-CoA synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 139. In some embodiments, the acetyl-CoA synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 139. In some embodiments, the acetyl-CoA synthase has the amino acid sequence of SEQ ID NO: 139. In some embodiments, the acetyl-CoA synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 140. In some embodiments, the acetyl-CoA synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 140. In some embodiments, the acetyl-CoA synthase has the amino acid sequence of SEQ ID NO: 140. In some embodiments, the aldehyde dehydrogenase has an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 141. In some embodiments, the aldehyde dehydrogenase has an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 141. In some embodiments, the aldehyde dehydrogenase synthase has the amino acid sequence of SEQ ID NO: 141. In some embodiments, the pyruvate decarboxylase has an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 142. In some embodiments, the pyruvate decarboxylase has an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 142. In some embodiments, the pyruvate decarboxylase has the amino acid sequence of SEQ ID NO: 142.

In some embodiments, expression of the one or more heterologous nucleic acids is regulated by an exogenous agent. In some embodiments, the exogenous agent includes a regulator of gene expression. In some embodiments, the exogenous agent decreases production of the cannabinoid. In some embodiments, the exogenous agent is maltose. In some embodiments, the exogenous agent increases production of the cannabinoid. In some embodiments, the exogenous agent is galactose. In some embodiments, the exogenous agent is galactose and expression of one or more heterologous nucleic acids encoding the AAE, TKS, and CBGaS enzymes is under the control of a GAL promoter. In some embodiments, expression of one or more heterologous nucleic acids encoding the AAE, TKS, and CBGaS enzymes is under the control of a galactose-responsive promoter, a maltose-responsive promoter, or a combination of both. In some embodiments, the method further includes culturing the host cell with a precursor required to make the cannabinoid. In some embodiments, the precursor required to make the cannabinoid is hexanoate.

In some embodiments, the cannabinoid is cannabichromene (CBC), cannabinol (CBN), cannabidivarin (CBDV), tetrahydrocannabivarin (THCV), cannabichromevarin (CBCV), cannabidiolic acid (CBDA), cannabidiol (CBD) or an acid form thereof, cannabigerolic acid (CBGA), cannabigerol (CBG) or an acid form thereof, tetrahydrocannabinol (THC) or an acid form thereof, or tetrahydrocannabinolic acid (THCa). In some embodiments, the cannabinoid is CBC.

In some embodiments, the CBC has a purity of at least about 50% (w/v) following purifying. In some embodiments, the CBC has a purity of between about 50% (w/w) and 100% (w/w) (e.g., between 50% (w/w) and 80 % (w/w), 50% (w/w) and 60% (w/w), 60% (w/w) and 100% (w/w), or 80%

(w/w) and 100% (w/w)) following purifying. In some embodiments, the CBC has a purity of about 70% (w/w).

In some embodiments, the host cell is a yeast cell or yeast strain. In some embodiments, the yeast cell is *S. cerevisiae*.

In another aspect, the disclosure provides a composition including CBC produced by any one of the methods described herein. In some embodiments, the CBC has a purity of at least about 50% (w/w). In some embodiments, the CBC has a purity of between about 50% (w/w) to about 100% (w/w) (e.g., between 50% (w/w) and 80 % (w/w), 50% (w/w) and 60% (w/w), 60% (w/w) and 100% (w/w), or 80% (w/w) and 100% (w/w)). In some embodiments, the CBC has a purity of between about 70% (w/w) to about 100% (w/w) (e.g., between 70% (w/w) and 90% (w/w), 70% (w/w) and 80% (w/w), 80% (w/w) and 100% (w/w), or 90% (w/w) and 100% (w/w)). In some embodiments, the CBC has a purity of from about 99.5% (w/w) to about 100% (w/w) (e.g., about 99.5% (w/w), 99.6% (w/w), 99.7% (w/w), 99.8% (w/w), or 99.9% (w/w)).

In another aspect, the disclosure provides a variant CBCa synthase polypeptide including one or more amino acid substitutions relative to the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the CBCa synthase includes an amino acid substitution at a residue selected from Q75, F82, T130, S140, V169, N240, V294, A299, K305, T335, R340, H354, L435, Y461, K535, S540, of SEQ ID NO: 1.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue Q75 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 substitutes Q75 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 is a Q75L substitution. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 substitutes Q75 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 is a Q75E substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue S140 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue S140 of SEQ ID NO: 1 substitutes S140 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue S140 of SEQ ID NO: 1 is a S140A substitution. In some embodiments, the amino acid substitution at residue S140 of SEQ ID NO: 1 is a S140T substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue V169 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue V169 of SEQ ID NO: 1 substitutes V169 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue V169 of SEQ ID NO: 1 is a V169E substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue N240 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue N240 of SEQ ID NO: 1 substitutes N240 with an amino acid including a polar, uncharged side

chain at physiological pH. In some embodiments, the amino acid substitution at residue N240 of SEQ ID NO: 1 is a N240Q substitution. In some embodiments, the amino acid substitution at residue N240 of SEQ ID NO: 1 substitutes N240 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue N240 of SEQ ID NO: 1 is a N240M substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue V294 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 substitutes V294 with an amino acid including a polar, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 is a V294S substitution. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 substitutes V294 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 is a V294E substitution. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 substitutes V294 with an amino acid including a cationic side chain at physiological pH. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 is a V294R substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue A299 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue A299 of SEQ ID NO: 1 substitutes A299 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue A299 of SEQ ID NO: 1 is an A299V substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue K305 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue K305 of SEQ ID NO: 1 substitutes K305 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue K305 of SEQ ID NO: 1 is a K305C substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue D328 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue D328 of SEQ ID NO: 1 is a D328P substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue T335 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue T335 of SEQ ID NO: 1 substitutes T335 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue T335 of SEQ ID NO: 1 is a T335L substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue R340 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue R340 of SEQ ID NO: 1 substitutes R340 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue R340 of SEQ ID NO: 1 is a R340M substitution. In some embodiments, the amino acid substitution at residue R340 of SEQ ID NO: 1 is a R340G substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue H354 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue H354 of SEQ ID NO: 1 substitutes H354 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue R340 of SEQ ID NO: 1 is a H354V substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue L435 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue L435 of SEQ ID NO: 1 substitutes L435 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue L435 of SEQ ID NO: 1 is a L435A substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue Y461 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue Y461 of SEQ ID NO: 1 substitutes Y461 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue Y461 of SEQ ID NO: 1 is a Y461I substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue K535 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue K535 of SEQ ID NO: 1 substitutes K535 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue Y461 of SEQ ID NO: 1 is a K535M substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue S540 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue S540 of SEQ ID NO: 1 substitutes S540 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue S540 of SEQ ID NO: 1 is a S540D substitution.

In some embodiments, the one or more amino acid substitutions include an amino acid substitution at residue T545 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue T545 of SEQ ID NO: 1 substitutes T545 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue T545 of SEQ ID NO: 1 is a T545E substitution.

In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A Y461I, K535, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, V294S, A299V, K305C, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, T335L, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, T130L, S140A, V169E, N240M, V294S, A299V, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, T130L, S140A, V169E, N240M, V294S, A299V, T335L,

R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, N240M, V294S, A299V, K305C, T335L, R340M, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V294S, A299V, K305C, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, S140A, V169E, N240M, K305C, T335L, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, V169E, V294S, A299V, R340M, H354V, L435A, K535M, or S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, V294S, A299V, T335L, H354V, L435A, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, T130L, S140A, V169E, V294S, A299V, K305C, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, A299V, K305C, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, V169E, N240M, V294S, T335L, R340M, H354V, L435A, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V294S, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, T335L, R340M, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including T130L, S140A, V169E, V294S, A299V, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, N240M, V294S, A299V, T335L, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M,

H354V, L435A, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, R340M, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, A299V, K305C, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, T130L, S140A, V169E, V294S, A299V, K305C, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, V169E, V294S, A299V, K305C, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, K305C, T335L, R340M, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including F82I, T130L, S140A, V169E, N240M, A299V, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, V294S, A299V, K305C, T335L, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, A299V, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, F82I, T130L, S140A, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. In some embodiments, the polypeptide has amino acid substitutions including Q75E, V169E, N240M, V294S, A299V, K305C, T335L, L435A, Y461I, and K535M of SEQ ID NO: 1. In some embodiments, the CBCa synthase comprises an amino acid substitution at a residue selected from Q75, F82, T130, S140, V169, N240, V294, A299, K305, T335, R340, H354, L435, Y461, K535, S540, and T545 of SEQ ID NO: 1

In some embodiments, the CBCa synthase has an amino acid sequence that is from about 85% to about 99.7% identical to the amino acid sequence of SEQ ID NO: 1 (e.g., about 85%, 96%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.7%). In some embodiments, the CBCa synthase has an amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the CBCa synthase has an amino acid sequence that is from about 90% to about 99.7% (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.7%)

identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the CBCa synthase has an amino acid sequence that is from about 95% to about 99.7% (e.g., about 95%, 95.5%, 96%, 96.5%, 97%, 97.5%, 98%, 98.5%, 99%, 99.5%, or 99.7%) identical to the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the CBCa synthase has an amino acid sequence that differs from the amino acid sequence of SEQ ID NO: 1 only by way of (i) the one or more amino acid substitutions or deletions and, optionally, (ii) one or more additional, conservative amino acid substitutions. In some embodiments, the CBCa synthase has an amino acid sequence that differs from the amino acid sequence of SEQ ID NO: 1 only by way of the one or more amino acid substitutions or deletions. In some embodiments, the CBCa synthase includes an amino acid sequence selected from SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, and SEQ ID NO: 28. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 85% identical to the amino acid sequence of any one of SEQ ID NO: 2-67. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 2-67. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 2-67. In some embodiments, the CBCa synthase has the amino acid sequence of any one of SEQ ID NO: 2-67.

In some embodiments, the CBCa synthase has an amino acid sequence that is at least 85% identical to the amino acid sequence of any one of SEQ ID NO: 29-67. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 29-67. In some embodiments, the CBCa synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 29-67. In some embodiments, the CBCa synthase has the amino acid sequence selected from SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, and SEQ ID NO: 67.

In another aspect, the disclosure provides a variant CBCa synthase polypeptide, wherein the polypeptide is capable producing a yield of CBCa concentration of at least 50 mg/L.

In another aspect, the disclosure provides a nucleic acid encoding any one of the variant polypeptides described herein.

In another aspect, the disclosure provides a host cell including any one of the variant polypeptides or nucleic acids described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a graph showing a schematic of the Production process for generating cannabichromene (CBC) by combining cannabigerolic acid (CBGa) in high oleic sunflower oil (HOSUN) overlay with aqueous CBCa synthase, chemically or mechanically demulsifying and decarboxylating CBCa, distilling the separated overlay to remove HOSUN, and purifying the resulting material.

FIG. 2 is a graph showing the bioconversion time course titers of cannabichromenic acid (CBCa), CBGa, and cannabigerol (CBG).

FIG. 3 is a graph showing the reaction rate as a function of CBGa concentration in the reaction and in the overlay for a given CBGa reaction, as the overlay % was decreased as the CBGa concentration in the overlay increased.

FIG. 4 is a graph showing the changeover base case bioconversion for Biosilica™ (silica derived from sugar cane bagasse) (purple square), microcrystalline cellulose, randomly methylated cyclodextrin, and 2,6-Di-O-methyl-beta-cyclodextrin.

FIG. 5 is a graph showing the reaction rates of the recycle experiments described in Example 1 without cyclodextrin and with cyclodextrin.

FIG. 6 is a graph showing the concentration of CBCa produced by CBGa synthase enzymes having a point mutation in comparison to the wild-type CBCa synthase.

FIG. 7 is a graph showing the concentration of CBCa produced by CBGa synthase enzymes having a combination of point mutations in comparison to the wild-type CBCa synthase.

DEFINITIONS

As used herein, the term “bioconversion mixture” refers to a composition including a substrate, for example, cannabigerolic acid (CBGa), and an enzyme capable of transforming the substrate, for example, cannabichromenic acid (CBCa) synthase, to generate a product, for example, cannabichromenic acid (CBCa).

As used herein, the term “cannabinoid” refers to a chemical substance that binds or interacts with a cannabinoid receptor (for example, a human cannabinoid receptor) and includes, without limitation, chemical compounds such endocannabinoids, phytocannabinoids, and synthetic cannabinoids. Synthetic compounds are chemicals made to mimic phytocannabinoids which are naturally found in the cannabis plant (e.g., *Cannabis sativa*). Cannabinoids includes but not limited to cannabigerols (CBG), cannabichromens (CBC), cannabidiol (CBD), tetrahydrocannabinol (THC), cannabinol (CBN), cannabinodiol (CBDL), cannabicyclol (CBL), cannabielsoin (CBE), cannabitriol (CBT), cannabinol (CBN), cannabichromene (CBC), cannabidioloic acid (CBDA), cannabigerolic acid (CBGA), tetrahydrocannabinolic acid (THCA), cannabinolic acid (CBNA), cannabidivarin (CBDV), tetrahydrocannabivarin (THCV), cannabigerovarin (CBGV), cannabichromevarin (CBCV), and others.

.

As used herein, the term “capable of producing” refers to a host cell that is genetically modified to express the enzyme(s) necessary for the production of a given compound in accordance

with a biochemical pathway that produces the compound. For example, a host cell (e.g., a yeast cell) that is “capable of producing” a cannabinoid is one that expresses the enzymes necessary for production of the cannabinoid according to the cannabinoid biosynthetic pathway.

As used herein, the term "endogenous" describes a molecule (e.g., a polypeptide, nucleic acid, or cofactor) that is found naturally in a particular organism (e.g., a human) or in a particular location within an organism (e.g., an organ, a tissue, or a cell, such as a human cell).

As used herein, the term "exogenous" describes a molecule (e.g., a polypeptide, nucleic acid, or cofactor) that is not found naturally in a particular organism (e.g., a human) or in a particular location within an organism (e.g., an organ, a tissue, or a cell, such as a human cell). Exogenous materials include those that are provided from an external source to an organism or to cultured matter extracted there from.

As used herein in the context of a gene, the term "express" refers to any one or more of the following events: (1) production of an RNA template from a DNA sequence (e.g., by transcription); (2) processing of an RNA transcript (e.g., by splicing, editing, 5' cap formation, and/or 3' end processing); (3) translation of an RNA into a polypeptide or protein; and (4) post-translational modification of a polypeptide or protein. Expression of a gene of interest in a cell, tissue sample, or subject can manifest, for example, as: an increase in the quantity or concentration of mRNA encoding a corresponding protein (as assessed, e.g., using RNA detection procedures described herein or known in the art, such as quantitative polymerase chain reaction (qPCR) and RNA seq techniques), an increase in the quantity or concentration of a corresponding protein (as assessed, e.g., using protein detection methods described herein or known in the art, such as enzyme-linked immunosorbent assays (ELISA), among others), and/or an increase in the activity of a corresponding protein (e.g., in the case of an enzyme, as assessed using an enzymatic activity assay described herein or known in the art).

The term "expression cassette" or “expression construct” refers to a nucleic acid construct that, when introduced into a host cell, results in transcription and/or translation of an RNA or polypeptide, respectively. In the case of expression of transgenes, one of skill will recognize that the inserted polynucleotide sequence need not be identical but may be only substantially identical to a sequence of the gene from which it was derived. As explained herein, these substantially identical variants are specifically covered by reference to a specific nucleic acid sequence. One example of an expression cassette is a polynucleotide construct that includes a polynucleotide sequence encoding a polypeptide for use in the invention operably linked to a promoter, e.g., its native promoter, where the expression cassette is introduced into a heterologous microorganism. In some embodiments, an expression cassette includes a polynucleotide sequence encoding a polypeptide of the invention where the polynucleotide that is targeted to a position in the genome of a microorganism such that expression of the polynucleotide sequence is driven by a promoter that is present in the microorganism.

As used herein, the term “gene” refers to the segment of DNA involved in producing or encoding a polypeptide chain. It may include regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments

(exons). Alternatively, the term “gene” can refer to the segment of DNA involved in producing or encoding a non-translated RNA, such as an rRNA, tRNA, gRNA, or micro RNA.

As used herein, the term “fermentation broth” refers to a composition which contains host cells and products, or metabolites produced by the host cells. An example of a fermentation broth is a whole cell broth, which may be the entire contents of a vessel, including cells, aqueous phase, and compounds produced from the host cells.

A “genetic pathway” or “biosynthetic pathway” as used herein refers to a set of at least two different coding sequences, where the coding sequences encode enzymes that catalyze different parts of a synthetic pathway to form a desired product (e.g., a cannabinoid). In a genetic pathway, a first encoded enzyme uses a substrate to make a first product which in turn is used as a substrate for a second encoded enzyme to make a second product. In some embodiments, the genetic pathway includes 3 or more members (e.g., 3, 4, 5, 6, 7, 8, 9, etc.), wherein the product of one encoded enzyme is the substrate for the next enzyme in the synthetic pathway.

The term "host cell" as used in the context of this disclosure refers to a microorganism, such as yeast, and includes an individual cell or cell culture including a heterologous vector or heterologous polynucleotide as described herein. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation and/or change. A host cell includes cells into which a recombinant vector or a heterologous polynucleotide of the invention has been introduced, including by transformation, transfection, and the like.

"Percent (%) sequence identity" with respect to a reference polynucleotide or polypeptide sequence is defined as the percentage of nucleic acids or amino acids in a candidate sequence that are identical to the nucleic acids or amino acids in the reference polynucleotide or polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid or amino acid sequence identity can be achieved in various ways that are within the capabilities of one of skill in the art, for example, using publicly available computer software such as BLAST, BLAST-2, or Megalign software. Those skilled in the art can determine appropriate parameters for aligning sequences, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. For example, percent sequence identity values may be generated using the sequence comparison computer program BLAST. As an illustration, the percent sequence identity of a given nucleic acid or amino acid sequence, A, to, with, or against a given nucleic acid or amino acid sequence, B, (which can alternatively be phrased as a given nucleic acid or amino acid sequence, A that has a certain percent sequence identity to, with, or against a given nucleic acid or amino acid sequence, B) is calculated as follows:

100 multiplied by (the fraction X/Y)

where X is the number of nucleotides or amino acids scored as identical matches by a sequence alignment program (e.g., BLAST) in that program's alignment of A and B, and where Y is the total number of nucleic acids in B. It will be appreciated that where the length of nucleic acid or amino acid sequence A is not equal to the length of nucleic acid or amino acid.

The terms "polynucleotide" and "nucleic acid" are used interchangeably and refer to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. A nucleic acid as used in the present disclosure will generally contain phosphodiester bonds, although in some cases, nucleic acid analogs may be used that may have alternate backbones, including, e.g., phosphoramidate, phosphorothioate, phosphorodithioate, or O-methylphosphoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press); positive backbones; non-ionic backbones, and non-ribose backbones. Nucleic acids or polynucleotides may also include modified nucleotides that permit correct read-through by a polymerase. "Polynucleotide sequence" or "nucleic acid sequence" includes both the sense and antisense strands of a nucleic acid as either individual single strands or in a duplex. As will be appreciated by those in the art, the depiction of a single strand also defines the sequence of the complementary strand; thus, the sequences described herein also provide the complement of the sequence. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The nucleic acid may be DNA, both genomic and cDNA, RNA, or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, isoguanine, etc. Nucleic acid sequences are presented in the 5' to 3' direction unless otherwise specified.

As used herein, the terms "polypeptide," "peptide," and "protein" are used interchangeably to refer to a polymer of amino acid residues. The terms encompass amino acid chains of any length, including full-length proteins, wherein the amino acid residues are linked by covalent peptide bonds.

As used herein the term "precursor cannabinoid" refers to a small molecule which by way of an enzymatic reaction is transformed to generate the cannabinoid of interest. The small molecule may be considered to be a cannabinoid itself.

Two sequences are "substantially identical" if two sequences have a specified percentage of amino acid residues or nucleotides that are the same (i.e. , 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity over a specified region, or, when not specified, over the entire sequence), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using a sequence comparison algorithm or by manual alignment and visual inspection as described above. Optionally, the identity exists over a region that is at least about 50 nucleotides (or 20 amino acids) in length, or more preferably over a region that is 100 to 500 or 1000 or more nucleotides (or 50, 100, or 200 or more amino acids) in length.

Nucleic acid or protein sequences that are substantially identical to a reference sequence include "conservatively modified variants." With respect to particular nucleic acid sequences, conservatively modified variants refer to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is

specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions in a nucleic acid, peptide, polypeptide, or protein sequence which alters a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Examples of amino acid groups defined in this manner can include: a "charged/polar group" including Glu (Glutamic acid or E), Asp (Aspartic acid or D), Asn (Asparagine or N), Gln (Glutamine or Q), Lys (Lysine or K), Arg (Arginine or R) and His (Histidine or H); an "aromatic or cyclic group" including Pro (Proline or P), Phe (Phenylalanine or F), Tyr (Tyrosine or Y) and Trp (Tryptophan or W); and an "aliphatic group" including Gly (Glycine or G), Ala (Alanine or A), Val (Valine or V), Leu (Leucine or L), Ile (Isoleucine or I), Met (Methionine or M), Ser (Serine or S), Thr (Threonine or T) and Cys (Cysteine or C). Within each group, subgroups can also be identified. For example, at pH 7, the group of charged/polar amino acids can be sub-divided into sub-groups including: the "positively-charged sub-group" including Lys, Arg and His; the "negatively-charged sub-group" comprising Glu and Asp; and the "polar sub-group" comprising Asn and Gln. In another example, the aromatic or cyclic group can be sub-divided into sub-groups including: the "nitrogen ring sub-group" comprising Pro, His and Trp; and the "phenyl sub-group" comprising Phe and Tyr. In another further example, the aliphatic group can be sub-divided into sub-groups including: the "large aliphatic non-polar sub-group" comprising Val, Leu, and Ile; the "aliphatic slightly-polar sub-group" comprising Met, Ser, Thr and Cys; and the "small-residue sub-group" comprising Gly and Ala. Examples of conservative mutations include amino acid substitutions of amino acids within the sub-groups above, such as, but not limited to: Lys for Arg or vice versa, such that a positive charge can be maintained; Glu for Asp or vice versa, such that a negative charge can be maintained; Ser for Thr or vice versa, such that a free -OH can be maintained; and Gln for Asn or vice versa, such that a free -NH_2 can be maintained. The following six groups each contain amino acids that further provide illustrative conservative substitutions for one another. 1) Ala, Ser, Thr; 2) Asp, Glu; 3) Asn, Gln; 4) Arg, Lys; 5) Ile, Leu, Met, Val; and 6) Phe, Try, and Trp (see, e.g., Creighton, Proteins: Structures and Molecular Principles. 1984, New York: W.H. Freeman).

Accordingly, the terms “conservative mutation,” “conservative substitution,” and “conservative amino acid substitution” refer to a substitution of one or more amino acids for one or more different amino acids that exhibit similar physicochemical properties, such as polarity, electrostatic charge, and steric volume. These properties are summarized for each of the twenty naturally occurring amino acids in Table 1, below.

Table 1. Representative physicochemical properties of naturally occurring amino acids

Amino Acid	3 Letter Code	1 Letter Code	Side-chain Polarity	Electrostatic character at physiological pH (7.4)	Steric Volume†
Alanine	Ala	A	nonpolar	neutral	small
Arginine	Arg	R	polar	cationic	large
Asparagine	Asn	N	polar	neutral	intermediate
Aspartic acid	Asp	D	polar	anionic	intermediate
Cysteine	Cys	C	nonpolar	neutral	intermediate
Glutamic acid	Glu	E	polar	anionic	intermediate
Glutamine	Gln	Q	polar	neutral	intermediate
Glycine	Gly	G	nonpolar	neutral	small
Histidine	His	H	polar	Both neutral and cationic forms in equilibrium at pH 7.4	large
Isoleucine	Ile	I	nonpolar	neutral	large
Leucine	Leu	L	nonpolar	neutral	large
Lysine	Lys	K	polar	cationic	large
Methionine	Met	M	nonpolar	neutral	large
Phenylalanine	Phe	F	nonpolar	neutral	large
Proline	Pro	P	non-polar	neutral	intermediate
Serine	Ser	S	polar	neutral	small
Threonine	Thr	T	polar	neutral	intermediate
Tryptophan	Trp	W	nonpolar	neutral	bulky
Tyrosine	Tyr	Y	polar	neutral	large
Valine	Val	V	nonpolar	neutral	intermediate

†based on volume in A^3: 50-100 is small, 100-150 is intermediate, 150-200 is large, and >200 is bulky

As used herein, the term "production" generally refers to an amount of compound produced by a genetically modified host cell provided herein. In some embodiments, production is expressed as a yield of the compound by the host cell. In other embodiments, production is expressed as a productivity of the host cell in producing the compound.

As used herein, the term "promoter" refers to a synthetic or naturally derived nucleic acid that is capable of activating, increasing, or enhancing expression of a DNA coding sequence, or

inactivating, decreasing, or inhibiting expression of a DNA coding sequence. A promoter may contain one or more specific transcriptional regulatory sequences to further enhance or repress expression and/or to alter the spatial expression and/or temporal expression of the coding sequence. A promoter may be positioned 5' (upstream) of the coding sequence under its control. A promoter may also initiate transcription in the downstream (3') direction, the upstream (5') direction, or be designed to initiate transcription in both the downstream (3') and upstream (5') directions. The distance between the promoter and a coding sequence to be expressed may be approximately the same as the distance between that promoter and the native nucleic acid sequence it controls. As is known in the art, variation in this distance may be accommodated without loss of promoter function. The term also includes a regulated promoter, which generally allows transcription of the nucleic acid sequence while in a permissive environment (e.g., microaerobic fermentation conditions, or the presence of maltose), but ceases transcription of the nucleic acid sequence while in a non-permissive environment (e.g., aerobic fermentation conditions, or in the absence of maltose). Promoters used herein can be constitutive, inducible, or repressible.

As used herein, the term “heterologous” refers to what is not normally found in nature. The term "heterologous nucleic acid" refers to a nucleic acid not normally found in a given cell in nature. A heterologous nucleic acid can be: (a) foreign to its host cell, i.e., exogenous to the host cell such that a host cell does not naturally contain the nucleic acid; (b) naturally found in the host cell, i.e., endogenous or native to the host cell, but present at an unnatural quantity in the cell (i.e., greater or lesser quantity than naturally found in the host cell); (c) be naturally found in the host cell but positioned outside of its natural locus. A “heterologous” polypeptide refers to a polypeptide that is encoded by a “heterologous nucleic acid.” Thus, for example, a “heterologous” polypeptide may be naturally produced by a host cell but is encoded by a heterologous nucleic acid that has been introduced into the host cell by genetic engineering. For example, a “heterologous” polypeptide can include embodiments in which an endogenous polypeptide is produced by an expression construct and is overexpressed in the host cell compared to native levels of the polypeptide produced by the host cell.

As used herein, the term “introducing” in the context of a nucleic acid or protein in a host cell refers to any process that results in the presence of a heterologous nucleic acid or polypeptide inside the host cell. For example, the term encompasses introducing a nucleic acid molecule (e.g., a plasmid or a linear nucleic acid) that encodes the nucleic acid of interest (e.g., an RNA molecule) or polypeptide of interest and results in the transcription of the RNA molecules and translation of the polypeptides. The term also encompasses integrating the nucleic acid encoding the RNA molecules or polypeptides into the genome of a progenitor cell. The nucleic acid is then passed through subsequent generations to the host cell, so that, for example, a nucleic acid encoding an RNA-guided endonuclease is “pre-integrated” into the host cell genome. In some cases, introducing refers to translocation of a nucleic acid or polypeptide from outside the host cell to inside the host cell. Various methods of introducing nucleic acids, polypeptides and other biomolecules into host cells are contemplated, including but not limited to, electroporation, contact with nanowires or nanotubes,

spheroplasting, PEG 1000-mediated transformation, biolistics, lithium acetate transformation, lithium chloride transformation, and the like.

As used herein, the term “transformation” refers to a genetic alteration of a host cell resulting from the introduction of exogenous genetic material, e.g., nucleic acids, into the host cell.

As used herein, the term “mutation” refers to a change in the nucleotide sequence of a gene. Mutations in a gene may occur naturally as a result of, for example, errors in DNA replication, DNA repair, irradiation, and exposure to carcinogens or mutations may be induced as a result of administration of a transgene expressing a mutant gene. Mutations may result from a single nucleotide substitution or deletion.

As used herein, the terms “oil,” “overlay oil,” or “overlay” refer to a biologically compatible hydrophobic, lipophilic, carbon-containing substance including but not limited to geologically-derived crude oil, distillate fractions of geologically-derived crude oil, vegetable oil, algal oil, microbial lipids, or synthetic oils. The oil is neither itself toxic to a biological molecule, a cell, a tissue, or a subject, nor does it degrade (if the oil degrades) at a rate that produces byproducts at toxic concentrations to a biological molecule, a cell, a tissue or a subject. Preferred examples of oils include but are not limited to avocado oil, canola oil, grapeseed oil, hemp oil, soybean oil, jojoba oil, and sunflower oil.

As used herein, the term "operably linked" refers to a functional linkage between nucleic acid sequences such that the sequences encode a desired function. For example, a coding sequence for a gene of interest is in operable linkage with its promoter and/or regulatory sequences when the linked promoter and/or regulatory region functionally controls expression of the coding sequence. It also refers to the linkage between coding sequences such that they may be controlled by the same linked promoter and/or regulatory region; such linkage between coding sequences may also be referred to as being linked in frame or in the same coding frame. "Operably linked" also refers to a linkage of functional but non-coding sequences, such as an autonomous propagation sequence or origin of replication. Such sequences are in operable linkage when they are able to perform their normal function, e.g., enabling the replication, propagation, and/or segregation of a vector bearing the sequence in a host cell.

As used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise.

As used herein, the term “about” is used herein to mean a value that is ±10% of the recited value.

DETAILED DESCRIPTION OF THE INVENTION

The present disclosure features methods of isolating and purifying one or more cannabinoids from a fermentation composition. The fermentation composition may include host cells capable of producing one or more cannabinoids, such as for example, cannabichromene (CBC). The methods described herein include using an oil overlay in combination with a cannabichromenic acid (CBCa) synthase to carry out near complete conversion of cannabigerolic acid (CBGa) to CBCa. The present disclosure additionally features variant CBCa synthase polypeptides having one or more amino acid

substitutions. The variant CBCa synthase may have increased enzymatic activity in comparison to the wild-type enzyme.

It has presently been discovered that the combination of the CBCa synthase with an oil overlay allowed for the efficient conversion of CBG to CBC. The following sections provide a detailed description of method of isolating the cannabinoid from a fermentation composition, the modified host cells that may be used to produce a cannabinoid, as well as variant CBCa synthase enzymes that show greater enzymatic efficiency.

Methods for Making a Cannabinoid

The disclosure provides a method of making a cannabichromenic acid (CBCa). The method may include (a) culturing a population of host cells capable of producing cannabigerolic acid (CBGa) in a culture medium comprising a fermentation broth and an overlay, under conditions suitable for the host cells to produce CBGa, and wherein the CBGa partitions into the overlay; (b) separating the overlay from the fermentation broth; (c) combining the separated overlay of step (b), with a CBCa synthase, thereby producing a bioconversion mixture; and (d) purifying the CBCa from the bioconversion mixture.

The overlay may include a plant-based oil. The plant-based oil may be, for example, soybean oil, sunflower oil, safflower oil, canola oil, grapeseed oil, or castor oil. The overlay may include a synthetic ester or a fatty alcohol. In some embodiments, the overlay and the fermentation broth are separated by centrifugation.

Once the fermentation broth and the oil are combined together, they may be demulsified, and optionally may undergo centrifugation. The demulsification may include contacting the second mixture containing the cannabinoid with an oil. The oil may be a mineral oil, a vegetable oil, a synthetic ester, or a fatty alcohol. For example, the oil may be a vegetable oil. In some embodiments, the vegetable oil is soybean oil, sunflower oil, safflower oil, canola oil, grapeseed oil, or castor oil. In some embodiments, the oil is a synthetic ester, optionally wherein the synthetic ester is ESTEREX™ A51. In some embodiments, the oil includes a fatty alcohol, optionally wherein the fatty alcohol is oleyl alcohol or JARCOL™ I-16. The oil may have a concentration of between about 1% (w/v) and about 10% (w/v) (e.g., about 1% (w/v), 2% (w/v), 3% (w/v), 4% (w/v), 5% (w/v), 6% (w/v), 7% (w/v), 8% (w/v), 9% (w/v), or 10% (w/v)). For example, the oil may have a concentration of is about 5% (w/v).

Th overlay further includes CBGa. The CBGa has a concentration of between about 0.1% (w/v) and 10% (w/v) (e.g., between about 0.1% (w/v) and 8% (w/v), 0.1% (w/v) and 6% (w/v), 0.1% (w/v) and 4% (w/v), 0.1% (w/v) and 2% (w/v), 0.1% (w/v) and 1% (w/v), 1% (w/v) and 10% (w/v), 2% (w/v) and 10% (w/v), 4% (w/v) and 10% (w/v), 6% (w/v) and 10% (w/v), or 8% (w/v) and 10% (w/v)). For example, the CBGa may have a concentration of between about 0.5% (w/v) or and 5% (w/v) (e.g., between about 0.5% (w/v) and 4% (w/v), 0.5% (w/v) and 3% (w/v), 0.5% (w/v) and 2% (w/v), 0.5% (w/v) and 1% (w/v), 1% (w/v) and 5% (w/v), 2% (w/v) and 5% (w/v), 3% (w/v) and 5% (w/v), or 4% (w/v) and 5% (w/v)).

The method may include stirring the bioconversion mixture. The bioconversion mixture may be stirred may include stirring the second mixture for between 12 hours and 144 hours (e.g., between 12 hours 120 hours, 12 hours and 96 hours, 12 hours and 72 hours, 12 hours and 48 hours, 12 hours and 24 hours, 24 hours and 144 hours, 48 hours and 144 hours, 72 hours and 144 hours, 96 hours and 144 hours, or 120 hours and 144 hours) prior to purifying the CBCa from the bioconversion mixture. For example, the bioconversion mixture may be stirred for between 24 hours and 96 hours (e.g., between 24 hours and 84 hours, 24 hours and 72 hours, 24 hours and 60 hours, 24 hours and 48 hours, 48 hours and 96 hours, 60 hours and 96 hours, 72 hours and 96 hours, or 84 hours and 96 hours). In some embodiments, the bioconversion mixture may be stirred for about 48 hours.

The bioconversion mixture may be held at a temperature of between 4 °C and 50 °C (e.g., between 4 °C and 40 °C, 4 °C and 30 °C, 4 °C and 20 °C, 4 °C and 10 °C, 10 °C and 50 °C, 20 °C and 50 °C, 30 °C and 50 °C, or 40 °C and 50 °C). For example, the bioconversion mixture may have a temperature of between 20 °C and 40 °C (e.g., about 35 °C).

The bioconversion mixture may be stirred at a rate of between about 50 rotations per minute (rpm) and 300 rpm (e.g., between about 50 rpm and 250 rpm, 50 rpm and 200 rpm, 50 rpm and 150 rpm, 50 rpm and 100 rpm, 100 rpm and 250 rpm, 150 rpm and 250 rpm, or 200 rpm and 250 rpm). For example, the bioconversion mixture may be stirred at a rate of about 150 rpm.

The second mixture may include one or more amphiphilic moieties in addition to the cannabinoid. The one or more amphiphilic moieties may include a cyclodextrin, plant derived silica, cellulose, or a combination thereof. For example, the cyclodextrin may include randomly methylated cyclodextrin, 2, 6-Di-O-methyl-β-cyclodextrin, or a combination thereof.

The demulsification includes one or more of: (i) contacting the culture medium with an enzymatic composition comprising a serine protease, (ii)contacting the culture medium with a surfactant; and (iii) contacting the culture medium with NaOH to adjust the culture medium to a pH of between pH 7 and pH 9.

The final concentration of the enzymatic composition may be from about 0.5% (w/v) to about 3% (w/v) (e.g., about 0.5% (w/v) and 2.5% (w/v), 0.5% (w/v) and 2% (w/v), 0.5% (w/v) and 1.5% (w/v), 0.5% (w/v) and 1% (w/v), 1% (w/v) and 3% (w/v), 1.5% (w/v) and 3% (w/v), 2% (w/v) and 3% (w/v), or 2.5% (w/v) and 3% (w/v)) after contacting the culture medium including a cannabinoid with the enzymatic composition. For example, the enzymatic composition may have a final concentration of about 1% (w/v).

The culture medium may be mixed with the enzymatic composition for between 0.5 hours and 2 hours (e.g., between 30 minutes and 90 minutes, 30 minutes and 70 minutes, 30 minutes and 50 minutes, 50 minutes and 90 minutes, or 70 minutes and 90 minutes.

The demulsification may include centrifugation of the culture medium including a cannabinoid. In some embodiments, the centrifugation may include liquid-liquid centrifugation, which may result in a crude oil light phase and an aqueous heavy phase.

The demulsification may further include a decarboxylation step comprising evaporating the culture medium. In some embodiments, the decarboxylation includes evaporating the crude oil light phase. The evaporating may include one or more passes; for example, the evaporating includes a

first pass and a second pass. The first pass may be performed at a temperature of between about 100 °C and about 500 °C (e.g., between about 100 °C and 400 °C, 100 °C and 300 °C, 100 °C and 200 °C, 200 °C and 500 °C, 300 °C and 500 °C, or 400 °C and 500 °C). For example, the first pass may be performed at a temperature of about 180 °C. The pressure of the first pass may be between about 0.5 torr and 760 torr (e.g., between 0.5 torr and 700 torr, 0.5 torr and 500 torr, 0.5 torr and 200 torr, 0.5 torr and 50 torr, 50 torr and 760 torr, 200 torr and 760 torr, 400 torr and 760 torr, or 600 torr and 760 torr). For example, the first pass may be performed at a pressure of about 1 torr.

A second pass of evaporating may be performed at a temperature of between 150 °C and 300 °C (e.g., between 200 °C and 300 °C, 250 °C and 300 °C, 150 °C and 250 °C, 150 °C, and 200 °C, or 150 °C and 175 °C). For example, the second pass may be performed at a temperature of about 240 °C. The second pass of evaporating may be performed at a pressure of between about 0.5 torr and 760 torr (e.g., between 0.5 torr and 700 torr, 0.5 torr and 500 torr, 0.5 torr and 200 torr, 0.5 torr and 50 torr, 50 torr and 760 torr, 200 torr and 760 torr, 400 torr and 760 torr, or 600 torr and 760 torr). For example, the first pass may be performed at a pressure of about 1 torr.

The purifying step may include one or more of a liquid-liquid extraction, chromatography, or saponification.

Cannabinoid Biosynthetic Pathway

In an aspect, a host cell described herein includes one or more nucleic acids encoding one or more enzymes of a heterologous genetic pathway that produces a cannabinoid or a precursor of a cannabinoid. The cannabinoid biosynthetic pathway may begin with hexanoic acid as the substrate for an acyl activating enzyme (AAE) to produce hexanoyl-CoA, which is used as the substrate of a tetraketide synthase (TKS) to produce tetraketide-CoA, which is used by an olivetolic acid cyclase (OAC) to produce olivetolic acid, which is then used to produce a cannabigerolic acid by a geranyl pyrophosphate (GPP) synthase and a cannabigerolic acid synthase (CBGaS). In some embodiments, the cannabinoid precursor that is produced is a substrate in the cannabinoid pathway (e.g., hexanoate or olivetolic acid). In some embodiments, the precursor is a substrate for an AAE, a TKS, an OAC, a CBGaS, or a GPP synthase. In some embodiments, the precursor, substrate, or intermediate in the cannabinoid pathway is hexanoate, olivetol, or olivetolic acid. In some embodiments, the precursor is hexanoate. In some embodiments, the host cell does not contain the precursor, substrate or intermediate in an amount sufficient to produce the cannabinoid or a precursor of the cannabinoid. In some embodiments, the host cell does not contain hexanoate at a level or in an amount sufficient to produce the cannabinoid in an amount over 10 mg/L. In some embodiments, the heterologous genetic pathway encodes at least one enzyme selected from the group consisting of an AAE, a TKS, an OAC, a CBGaS, or a GPP synthase. In some embodiments, the genetically modified host cell includes an AAE, TKS, OAC, CBGaS, and a GPP synthase. The cannabinoid pathway is described in Keasling et al. (U.S. Patent No. 10,563,211), the disclosure of which is incorporated herein by reference.

Cannabichromenic acid Synthase

The disclosure provided herein includes CBCa synthase enzymes which include one or more amino acid substitutions, additions, or deletions, which result in increased production of CBC. Some embodiments described herein concern a host cell that includes a heterologous CBCa synthase such that the host cell is capable of producing a cannabinoid. The CBCa synthase may have an amino acid sequence that differs from the amino acid sequence of SEQ ID NO: 1 only by way of the one or more amino acid substitutions or deletions. The variant polypeptide may have an amino acid sequence that differs from the amino acid sequence of SEQ ID NO: 1 only by way of (i) the one or more amino acid substitutions or deletions and, optionally, (ii) one or more additional, conservative amino acid substitutions. The CBCa synthase may include an amino acid substitution at a residue selected from Q75, F82, T130, S140, V169, N240, V294, A299, K305, T335, R340, H354, L435, Y461, K535, S540, of SEQ ID NO: 1.

The variant polypeptide may include an amino acid substitution at residue Q75 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 includes an amino acid having a hydrophobic, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue Q75 of SEQ ID NO: 1 substitutes Q75 with an amino acid including an anionic side chain at physiological pH. For example, the amino acid substitution may be a Q75L substitution or Q75E substitution.

The polypeptide may include an amino acid substitution at residue S140 of SEQ ID NO: 1. For example, the amino acid substitution at residue S140 of SEQ ID NO: 1 substitutes S140 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. The amino acid substitution at residue S140 of SEQ ID NO: 1 may be a S140A substitution or S140T substitution.

The variant polypeptide may include an amino acid substitution at residue V169 of SEQ ID NO: 1. The amino acid substitution at residue V169 of SEQ ID NO: 1 may substitute V169 with an amino acid including an anionic side chain at physiological pH. For example, the amino acid substitution at residue V169 of SEQ ID NO: 1 may be a V169E substitution.

The polypeptide may include an amino acid substitution at residue N240 of SEQ ID NO: 1. This amino acid substitution at residue N240 of SEQ ID NO: 1 may be an amino acid having a polar, uncharged side chain at physiological pH, or the amino acid substitution at residue N240 of SEQ ID NO: 1 may substitute N240 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. For example, the amino acid substitution may be a N240Q substitution or an N240M substitution.

The variant polypeptide may include an amino acid substitution at residue V294 of SEQ ID NO: 1. This substitution may include an amino acid including a polar, uncharged side chain at physiological pH. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 substitutes V294 with an amino acid including an anionic side chain at physiological pH. In some embodiments, the amino acid substitution at residue V294 of SEQ ID NO: 1 substitutes V294 with an amino acid including a cationic side chain at physiological pH. For example, the amino acid substitution at residue V294 of SEQ ID NO: 1 may be a V294S substitution, a V294E substitution, or V294R substitution.

The variant polypeptide may include an amino acid substitution at residue A299 of SEQ ID NO: 1. The amino acid substitution at residue A299 of SEQ ID NO: 1 may substitute A299 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. For example, the amino acid substitution at residue A299 of SEQ ID NO: 1 may be an A299V substitution.

The variant polypeptide may include an amino acid substitution at residue K305 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue K305 of SEQ ID NO: 1 substitutes K305 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. For example, the amino acid substitution at residue K305 of SEQ ID NO: 1 may be a K305C substitution.

The variant polypeptide may include an amino acid substitution at residue D328 of SEQ ID NO: 1. For example, the amino acid substitution at residue D328 of SEQ ID NO: 1 may be a D328P substitution.

The variant polypeptide may include an amino acid substitution at residue T335 of SEQ ID NO: 1. In some embodiments, wherein the amino acid substitution at residue T335 of SEQ ID NO: 1 substitutes T335 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. For example, the amino acid substitution at residue T335 of SEQ ID NO: 1 may be a T335L substitution.

The CBCa synthase may include one or more amino acid substitutions include an amino acid substitution at residue R340 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue R340 of SEQ ID NO: 1 substitutes R340 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. For example, the amino acid substitution at residue R340 of SEQ ID NO: 1 may be a R340M substitution or an R340G substitution.

The variant polypeptide may have an amino acid substitution at residue H354 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue H354 of SEQ ID NO: 1 includes an amino acid including a hydrophobic, uncharged side chain at physiological pH. For example, the amino acid substitution at residue R340 of SEQ ID NO: 1 may be a H354V substitution.

The variant CBCa synthase may include an amino acid substitution at residue L435 of SEQ ID NO: 1. The amino acid substitution at residue L435 of SEQ ID NO: 1 may substitute L435 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. For example, the amino acid substitution at residue L435 of SEQ ID NO: 1 may be a L435A substitution.

The variant CBCa synthase may include an amino acid substitution at residue Y461 of SEQ ID NO: 1. The amino acid substitution at residue Y461 of SEQ ID NO: 1 may substitute Y461 with an amino acid including a hydrophobic, uncharged side chain at physiological pH. For example, the amino acid substitution at residue Y461 of SEQ ID NO: 1 is a Y461I substitution.

The polypeptide may have one or more amino acid substitutions include an amino acid substitution at residue K535 of SEQ ID NO: 1. The amino acid substitution may be hydrophobic and/or include uncharged side chain at physiological pH. For example, the amino acid substitution at residue Y461 of SEQ ID NO: 1 may be a K535M substitution.

The variant polypeptide may include an amino acid substitution at residue S540 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue S540 of SEQ ID NO: 1

substitutes S540 with an amino acid including an anionic side chain at physiological pH. For example, the amino acid substitution at residue S540 of SEQ ID NO: 1 may have a S540D substitution.

The CBCa synthase may have an amino acid substitution at residue T545 of SEQ ID NO: 1. In some embodiments, the amino acid substitution at residue T545 of SEQ ID NO: 1 substitutes T545 with an amino acid including an anionic side chain at physiological pH. For example, the amino acid substitution at residue T545 of SEQ ID NO: 1 may be a T545E substitution.

The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A Y461I, K535, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, V294S, A299V, K305C, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, T335L, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have the polypeptide has amino acid substitutions including Q75E, T130L, S140A, V169E, N240M, V294S, A299V, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, T130L, S140A, V169E, N240M, V294S, A299V, T335L, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, V169E, N240M, V294S, A299V, K305C, T335L, R340M, L435A, Y461I, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V294S, A299V, K305C, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, S140A, V169E, N240M, K305C, T335L, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, V169E, V294S, A299V, R340M, H354V, L435A, K535M, or S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, V169E, V294S, A299V, T335L, H354V, L435A, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, T130L, S140A, V169E, V294S, A299V, K305C, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, V169E, A299V, K305C, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, V169E, N240M, V294S, T335L, R340M, H354V, L435A, and K535M of SEQ ID NO: 1. The

variant polypeptide may have amino acid substitutions including Q75E, F82I, S140A, V294S, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, T335L, R340M, L435A, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including T130L, S140A, V169E, V294S, A299V, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, N240M, V294S, A299V, T335L, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, R340M, L435A, Y461I, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, A299V, K305C, R340M, H354V, L435A, Y461I, and K535M of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, R340M, H354V, L435A, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, T130L, S140A, V169E, V294S, A299V, K305C, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, V169E, V294S, A299V, K305C, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, K305C, T335L, R340M, L435A, Y461I, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including F82I, T130L, S140A, V169E, N240M, A299V, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, S140A, V169E, N240M, V294S, A299V, K305C, T335L, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, V294S, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The variant polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, V169E, N240M, A299V, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The polypeptide may have amino acid substitutions including Q75E, F82I, T130L, S140A, N240M, V294S, A299V, K305C,

T335L, R340M, H354V, L435A, Y461I, K535M, and S540D of SEQ ID NO: 1. The polypeptide may have amino acid substitutions including Q75E, V169E, N240M, V294S, A299V, K305C, T335L, L435A, Y461I, and K535M of SEQ ID NO: 1.

The CBCa synthase may have an amino acid sequence that is from about 85% to about 99.7% (e.g., about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.7%) identical to the amino acid sequence of SEQ ID NO: 1. The CBCa synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 1. The CBCa synthase may have an amino acid sequence that is from about 90% to about 99.7% (e.g., about 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.7%) identical to the amino acid sequence of SEQ ID NO: 1. For example, the CBCa synthase may have an amino acid sequence that is from about 95% to about 99.7% (e.g., about 95%, 95.5%, 96%, 96.5%, 97%, 97.5%, 98%, 98.5%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, or 99.7%) identical to the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the CBCa synthase has an amino acid sequence that is at least 85% identical to the amino acid sequence of any one of SEQ ID NO: 2-67. The CBCa synthase may be an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 2-67. For example, the CBCa synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 2-67. In some embodiments, the CBCa synthase has the amino acid sequence of any one of SEQ ID NO: 2-67.

In some embodiments, the CBCa synthase has an amino acid sequence that is at least 85% identical to the amino acid sequence of any one of SEQ ID NO: 29-67. The CBCa synthase may have an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 29-67. The CBCa synthase may have an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 29-67. For example, the CBCa synthase may have the amino acid sequence of any one of SEQ ID NO: 29-67.

The variant CBCa synthase polypeptide may be capable of producing a yield of CBCa concentration of at least 50 mg/L.

Acyl Activating Enzymes

Some embodiments concern a host cell that includes a heterologous AAE such that the host cell is capable of producing a cannabinoid. The AAE may be from *Cannabis sativa* or may be an enzyme from another plant or fungal source which has been shown to have AAE activity in the cannabinoid biosynthetic pathway, resulting in the production of the cannabinoid precursor olivetolic acid. In some embodiments, the host cell contains a heterologous nucleic acid that encodes an AAE having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 68-91 (e.g., an amino acid sequence that is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 68-91). In some embodiments, the AAE has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 68-91 (e.g., an amino acid sequence that is 95%, 96%, 97%, 98%, 99%, or 100%

identical to any one of SEQ ID NO: 68-91). In some embodiments, the AAE has the amino acid sequence of any one of SEQ ID NO: 68-91.

In some embodiments, the host cell contains a heterologous nucleic acid that encodes an AAE having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 68-80 (e.g., an amino acid sequence that is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 68-80). In some embodiments, the AAE has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 68-80 (e.g., an amino acid sequence that is 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 68-80). In some embodiments, the AAE has the amino acid sequence of any one of SEQ ID NO: 68-80.

In some embodiments, the host cell contains a heterologous nucleic acid that encodes an AAE having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 68-72 (e.g., an amino acid sequence that is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 68-72). In some embodiments, the AAE has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 68-72 (e.g., an amino acid sequence that is 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 68-72). In some embodiments, the AAE has the amino acid sequence of any one of SEQ ID NO: 68-72.

Tetraketide Synthase Enzymes

Some embodiments concern a host cell that includes a heterologous TKS such that the host cell is capable of producing a cannabinoid. A TKS uses the hexanoyl-CoA precursor to generate tetraketide-CoA. The TKS may be from *Cannabis sativa* or may be an enzyme from another plant or fungal source which has been shown to have TKS activity in the cannabinoid biosynthetic pathway, resulting in the production of the cannabinoid precursor olivetolic acid. In some embodiments, the host cell contains a heterologous nucleic acid that encodes a TKS having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 92-126 (e.g., an amino acid sequence that is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 92-126). In some embodiments, the TKS has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 92-126 (e.g., an amino acid sequence that is 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 92-126). In some embodiments, the TKS has the amino acid sequence of any one of SEQ ID NO: 92-126.

In some embodiments, the host cell contains a heterologous nucleic acid that encodes a TKS having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 92-95 (e.g., an amino acid sequence that is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 92-95). In some embodiments, the TKS has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 92-95 (e.g., an amino acid sequence that is 95%, 96%, 97%, 98%, 99%, or 100%

identical to any one of SEQ ID NO: 92-95). In some embodiments, the TKS has the amino acid sequence of any one of SEQ ID NO: 92-95.

In some embodiments, the host cell contains a heterologous nucleic acid that encodes a TKS having an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 92 (e.g., an amino acid sequence that is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 92). In some embodiments, the TKS has an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 92 (e.g., an amino acid sequence that is 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 92). In some embodiments, the TKS has the amino acid sequence of SEQ ID NO: 92.

Cannabigerolic Acid Synthases

Some embodiments concern a host cell that includes a heterologous CBGaS such that the host cell is capable of producing a cannabinoid. A CBGaS uses the olivetolic acid precursor and GPP precursor to generate cannabigerolic acid. The CBGaS may be from *Cannabis sativa* or may be an enzyme from another plant or fungal source which has been shown to have CBGaS activity in the cannabinoid biosynthetic pathway, resulting in the production of the cannabinoid cannabigerolic acid. In some embodiments, the host cell contains a heterologous nucleic acid that encodes a CBGaS having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 127-131 (e.g., an amino acid sequence that is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 127-131). In some embodiments, the CBGaS has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 127-131 (e.g., an amino acid sequence that is 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 127-131). In some embodiments, the CBGaS has the amino acid sequence of any one of SEQ ID NO: 127-131.

Geranyl Pyrophosphate Synthase

Some embodiments concern a host cell that includes a heterologous GPP synthase such that the host cell is capable of producing a cannabinoid. A GPP synthase uses the product of the isoprenoid biosynthesis pathway precursor to generate cannabigerolic acid together with a prenyltransferase enzyme. The GPP synthase may be from *Cannabis sativa* or may be an enzyme from another plant or bacterial source which has been shown to have GPP synthase activity in the cannabinoid biosynthetic pathway, resulting in the production of the cannabinoid cannabigerolic acid.

In some embodiments, the host cell contains a heterologous nucleic acid that encodes a GPP synthase having an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 132-137 (e.g., an amino acid sequence that is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 132-137). In some embodiments, the GPP synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of any one of SEQ ID NO: 132-137 (e.g., an amino acid sequence that is 95%, 96%, 97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 132-137). In some embodiments, the GPP synthase has the amino acid sequence of any one of SEQ ID NO: 132-137.

In some embodiments, the host cell contains a heterologous nucleic acid that encodes a GPP synthase having an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO: 132 (e.g., an amino acid sequence that is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 132). In some embodiments, the GPP synthase has an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 132 (e.g., an amino acid sequence that is 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 132). In some embodiments, the GPP synthase has the amino acid sequence of SEQ ID NO: 132.

Additional Enzymes

The host cell may further express other heterologous enzymes in addition to the AAE, TKS, CBGaS, and/or GPP synthase. For example, the host cell may include an olivetolic acid cyclase (OAC) as part of the cannabinoid biosynthetic pathway. The OAC may have an amino acid sequence that is at least 90% (e.g., at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%) identical to SEQ ID NO: 138. In some embodiments, the OAC has an amino acid sequence of SEQ ID NO: 138. In some embodiments, the host cell may include a heterologous nucleic acid that encodes at least one enzyme from the mevalonate biosynthetic pathway. Enzymes which make up the mevalonate biosynthetic pathway may include but are not limited to an acetyl-CoA thiolase, an HMG-CoA synthase, an HMG-CoA reductase, a mevalonate kinase, a phosphomevalonate kinase, a mevalonate pyrophosphate decarboxylase, and an IPP:DMAPP isomerase. In some embodiments, the host cell includes a heterologous nucleic acid that encodes the acetyl-CoA thiolase, the HMG-CoA synthase, the HMG-CoA reductase, the mevalonate kinase, the phosphomevalonate kinase, the mevalonate pyrophosphate decarboxylase, and the IPP:DMAPP isomerase of the mevalonate biosynthesis pathway.

In some embodiments, the host cell may express heterologous enzymes of the central carbon metabolism. Enzymes of the central carbon metabolism may include an acetyl-CoA synthase, an aldehyde dehydrogenase, and a pyruvate decarboxylase. In some embodiments, the host cell includes heterologous nucleic acids that independently encode an acetyl-CoA synthase, and/or an aldehyde dehydrogenase, and/or a pyruvate decarboxylase. In some embodiments, the acetyl-CoA synthase and the aldehyde dehydrogenase from *Saccharomyces cerevisiae*, and the pyruvate decarboxylase from *Zymomonas mobilis*. In some embodiments, the acetyl-CoA synthase has an amino acid sequence that is at least 90% (e.g., at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%) identical to the amino acid sequence of SEQ ID NO: 139. In some embodiments, the acetyl-CoA synthase has the amino acid sequence of SEQ ID NO: 139. In some embodiments, the host cell expresses a heterologous acetyl-CoA synthase having an amino acid sequence that is at least 90% (e.g., at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%) identical to the amino acid sequence of SEQ ID NO: 140. In some embodiments, the acetyl-CoA synthase has the amino acid sequence of SEQ ID NO: 140. In some embodiments, the aldehyde dehydrogenase has an amino acid sequence that is at least 90% (e.g., at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%) identical to the amino acid sequence of SEQ ID NO: 141. In some embodiments, the aldehyde dehydrogenase has the amino acid sequence of SEQ ID NO: 141. In some embodiments, the

pyruvate dehydrogenase has an amino acid sequence that is at least 90% (e.g., at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%) identical to the amino acid sequence of SEQ ID NO: 142. In some embodiments, the pyruvate decarboxylase has an amino acid sequence of SEQ ID NO: 142.

Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or functionally equivalent polypeptides can also be used to clone and express the polynucleotides encoding the protein components of the heterologous genetic pathway described herein.

As will be understood by those of skill in the art, it can be advantageous to modify a coding sequence to enhance its expression in a particular host. The genetic code is redundant with 64 possible codons, but most organisms typically use a subset of these codons. The codons that are utilized most often in a species are called optimal codons, and those not utilized very often are classified as rare or low-usage codons. Codons can be substituted to reflect the preferred codon usage of the host, in a process sometimes called "codon optimization" or "controlling for species codon bias."

Optimized coding sequences containing codons preferred by a particular prokaryotic or eukaryotic host (Murray et al., 1989, Nucl Acids Res. 17: 477-508) can be prepared, for example, to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced from a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, typical stop codons for *S. cerevisiae* and mammals are UAA and UGA, respectively. The typical stop codon for monocotyledonous plants is UGA, whereas insects and *E. coli* commonly use UAA as the stop codon (Dalphin et al., 1996, Nucl Acids Res. 24: 216-8).

Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA molecules differing in their nucleotide sequences can be used to encode a given enzyme of the disclosure. Any one of the polypeptide sequences disclosed herein may be encoded by DNA molecules of any sequence that encode the amino acid sequences of the polypeptides and proteins of the enzymes utilized in the methods of the disclosure. In a similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The disclosure includes such polypeptides with different amino acid sequences than the specific proteins described herein so long as the modified or variant polypeptides have the enzymatic anabolic or catabolic activity of the reference polypeptide. Furthermore, the amino acid sequences encoded by the DNA sequences shown herein merely illustrate embodiments of the disclosure.

In addition, homologs of enzymes useful for the compositions and methods provided herein are encompassed by the disclosure. In some embodiments, two proteins (or a region of the proteins) are substantially homologous when the amino acid sequences have at least about 30%, 40%, 50% 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity. To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be

introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In one embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, typically at least 40%, more typically at least 50%, even more typically at least 60%, and even more typically at least 70%, 80%, 90%, 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

When "homologous" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art (e.g., Pearson W. R., 1994, Methods in Mol Biol 25: 365-89).

The following six groups each contain amino acids that are conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D), Glutamic Acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Alanine (A), Valine (V), and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Sequence homology for polypeptides, which is also referred to as percent sequence identity, is typically measured using sequence analysis software. A typical algorithm used for comparing a molecule sequence to a database containing a large number of sequences from different organisms is the computer program BLAST. When searching a database containing sequences from a large number of different organisms, it is typical to compare amino acid sequences.

Furthermore, any of the genes encoding the foregoing enzymes (or any others mentioned herein (or any of the regulatory elements that control or modulate expression thereof)) may be optimized by genetic/protein engineering techniques, such as directed evolution or rational mutagenesis, which are known to those of ordinary skill in the art. Such action allows those of ordinary skill in the art to optimize the enzymes for expression and activity in a host cell, for example, a yeast.

In addition, genes encoding these enzymes can be identified from other fungal and bacterial species and can be expressed in the host cell. A variety of organisms could serve as sources for these enzymes, including, but not limited to, Saccharomyces spp., including *S. cerevisiae* and *S.*

uvarum, Kluyveromyces spp., including *K. thermotolerans*, *K. lactis*, and *K. marxianus*, Pichia spp., Hansenula spp., including *H. polymorphs*, Candida spp., Trichosporon spp., Yamadazyma spp., including *Y. stipitis*, *Torulaspora pretoriensis*, *Issatchenkia orientalis*, Schizosaccharomyces spp., including *S. pombe*, Cryptococcus spp., Aspergillus spp., Neurospora spp., or Ustilago spp. Sources of genes from anaerobic fungi include, but are not limited to, Piromyces spp., Orpinomyces spp., or Neocallimastix spp. Sources of prokaryotic enzymes that are useful include, but are not limited to, *Escherichia coli*, *Zymomonas mobilis*, *Staphylococcus aureus*, Bacillus spp., Clostridium spp., Corynebacterium spp., Pseudomonas spp., Lactococcus spp., Enterobacter spp., and Salmonella spp.

Techniques known to those skilled in the art may be suitable to identify additional homologous genes and homologous enzymes. Generally, analogous genes and/or analogous enzymes can be identified by functional analysis and will have functional similarities. Techniques known to those skilled in the art may be suitable to identify analogous genes and analogous enzymes. For example, to identify homologous or analogous ADA genes, proteins, or enzymes, techniques may include, but are not limited to, cloning a gene by PCR using primers based on a published sequence of an ADA gene/enzyme or by degenerate PCR using degenerate primers designed to amplify a conserved region among ADA genes. Further, one skilled in the art can use techniques to identify homologous or analogous genes, proteins, or enzymes with functional homology or similarity. Techniques include examining a cell or cell culture for the catalytic activity of an enzyme through in vitro enzyme assays for said activity (e.g. as described herein or in Kiritani, K., Branched-Chain Amino Acids Methods Enzymology, 1970), then isolating the enzyme with said activity through purification, determining the protein sequence of the enzyme through techniques such as Edman degradation, design of PCR primers to the likely nucleic acid sequence, amplification of said DNA sequence through PCR, and cloning of said nucleic acid sequence. To identify homologous or similar genes and/or homologous or similar enzymes, analogous genes and/or analogous enzymes or proteins, techniques also include comparison of data concerning a candidate gene or enzyme with databases such as BRENDA, KEGG, JGI Phyzome v12.1, BLAST, NCBI RefSeq, UniProt KB, or MetaCYC Protein annotations in the UniProt Knowledgebase may also be used to identify enzymes which have a similar function in addition to the National Center for Biotechnology Information RefSeq database. The candidate gene or enzyme may be identified within the above-mentioned databases in accordance with the teachings herein.

Modified Host Cells

In one aspect, provided herein are host cells including at least one enzyme of the cannabinoid biosynthetic pathway. In some embodiments, the cannabinoid biosynthetic pathway contains a genetic regulatory element, such as a nucleic acid sequence, which is regulated by an exogenous agent. In some embodiments, the exogenous agent acts to regulate expression of the heterologous genetic pathway. Thus, in some embodiments, the exogenous agent can be a regulator of gene expression.

In some embodiments, the exogenous agent can be used as a carbon source by the host cell. For example, the same exogenous agent can both regulate production of a cannabinoid and provide a carbon source for growth of the host cell. In some embodiments, the exogenous agent is galactose. In some embodiments, the exogenous agent is maltose.

In some embodiments, the genetic regulatory element is a nucleic acid sequence, such as a promoter.

In some embodiments, the genetic regulatory element is a galactose-responsive promoter. Exemplary promoters are shown in Table 2 below. In some embodiments, galactose positively regulates expression of the cannabinoid biosynthetic pathway, thereby increasing production of the cannabinoid. In some embodiments, the galactose-responsive promoter is a GAL1 promoter. In some embodiments, the galactose-responsive promoter is a GAL10 promoter. In some embodiments, the galactose-responsive promoter is a GAL2, GAL3, or GAL7 promoter. In some embodiments, heterologous genetic pathway contains the galactose-responsive regulatory elements described in Westfall et al. (PNAS (2012) vol.109: E111-118). In some embodiments, the host cell lacks the gal1 gene and is unable to metabolize galactose, but galactose can still induce galactose-regulated genes.

Table 2: Exemplary GAL Promoter Sequences

Promoter	Sequence
pGAL1	SEQ ID NO: 143
pGAL10	SEQ ID NO: 144
pGAL2	SEQ ID NO: 145
pGAL3	SEQ ID NO: 146
pGAL7	SEQ ID NO: 147
pGAL4	SEQ ID NO: 148

In some embodiments, the galactose regulation system used to control expression of one or more enzymes of the cannabinoid biosynthetic pathway is re-configured such that it is no longer induced by the presence of galactose. Instead, the gene of interest will be expressed unless repressors, which may be maltose in some strains, are present in the medium.

In some embodiments, the genetic regulatory element is a maltose-responsive promoter. Exemplary promoters are shown in Table 3 below. In some embodiments, maltose negatively regulates expression of the cannabinoid biosynthetic pathway, thereby decreasing production of the cannabinoid. In some embodiments, the maltose-responsive promoter is selected from the group consisting of pMAL1, pMAL2, pMAL11, pMAL12, pMAL31 and pMAL32. The maltose genetic regulatory element can be designed to both activate expression of some genes and repress expression of others, depending on whether maltose is present or absent in the medium. Maltose regulation of gene expression and maltose-responsive promoters are described in U.S. Patent Publication 2016/0177341, which is hereby incorporated by reference. Genetic regulation of maltose

metabolism is described in Novak et al., "Maltose Transport and Metabolism in S. cerevisiae," Food Technol. Biotechnol. 42 (3) 213–218 (2004).

Table 3: Exemplary MAL Promoter Sequences

Promoter	Sequence
pMAL1	SEQ ID NO: 149
pMAL2	SEQ ID NO: 150
pMAL11	SEQ ID NO: 151
pMAL12	SEQ ID NO: 152
pMAL31	SEQ ID NO: 153
pMAL32	SEQ ID NO: 154

In some embodiments, the heterologous genetic pathway is regulated by a combination of the maltose and galactose regulons.

In some embodiments, the recombinant host cell does not contain, or expresses a very low level of (for example, an undetectable amount), a precursor (e.g., hexanoate) required to make the cannabinoid. In some embodiments, the precursor (e.g., hexanoate) is a substrate of an enzyme in the cannabinoid biosynthetic pathway.

Yeast Strains

In some embodiments, yeast strains useful in the present methods include yeasts that have been deposited with microorganism depositories (e.g. IFO, ATCC, etc.) and belong to the genera Aciculoconidium, Ambrosiozyma, Arthroascus, Arxiozyma, Ashbya, Babjevia, Bensingtonia, Botryoascus, Botryozyma, Brettanomyces, Bullera, Bulleromyces, Candida, Citeromyces, Clavispora, Cryptococcus, Cystofilobasidium, Debaryomyces, Dekkara, Dipodascopsis, Dipodascus, Eeniella, Endomycopsella, Eremascus, Eremothecium, Erythrobasidium, Fellomyces, Filobasidium, Galactomyces, Geotrichum, Guilliermondella, Hanseniaspora, Hansenula, Hasegawaea, Holtermannia, Hormoascus, Hyphopichia, Issatchenkia, Kloeckera, Kloeckeraspora, Kluyveromyces, Kondoa, Kuraishia, Kurtzmanomyces, Leucosporidium, Lipomyces, Lodderomyces, Malassezia, Metschnikowia, Mrakia, Myxozyma, Nadsonia, Nakazawaea, Nematospora, Ogataea, Oosporidium, Pachysolen, Phachytichospora, Phaffia, Pichia, Rhodosporidium, Rhodotorula, Saccharomyces, Saccharomycodes, Saccharomycopsis, Saitoella, Sakaguchia, Saturnospora, Schizoblastosporion, chizosaccharomyces, Schwanniomyces, Sporidiobolus, Sporobolomyces, Sporopachydermia, Stephanoascus, Sterigmatomyces, Sterigmatosporidium, Symbiotaphrina, Sympodiomyces, Sympodiomycopsis, Torulaspora, Trichosporiella, Trichosporon, Trigonopsis, Tsuchiyaea, Udeniomyces, Waltomyces, Wickerhamia, Wickerhamiella, Williopsis, Yamadazyma, Yarrowia, Zygoascus, Zygosaccharomyces, Zygowilliopsis, and Zygozyma, among others.

In some embodiments, the strain is *Saccharomyces cerevisiae*, *Pichia pastoris*, *Schizosaccharomyces pombe*, *Dekkera bruxellensis*, *Kluyveromyces lactis* (previously called *Saccharomyces lactis*), *Kluveromyces marxianus*, *Arxula adeninivorans*, or *Hansenula polymorphs*

(now known as *Pichia angusta*). In some embodiments, the host microbe is a strain of the genus Candida, such as *Candida lipolytica*, *Candida guilliermondii*, *Candida krusei*, *Candida pseudotropicalis*, or *Candida utilis*.

In a particular embodiment, the strain is *Saccharomyces cerevisiae*. In some embodiments, the host is a strain of *Saccharomyces cerevisiae* selected from the group consisting of Baker's yeast, CEN.PK, CEN.PK2, CBS 7959, CBS 7960, CBS 7961, CBS 7962, CBS 7963, CBS 7964, IZ-1904, TA, BG-1, CR-1, SA-1, M-26, Y-904, PE-2, PE-5, VR-1, BR-1, BR-2, ME-2, VR-2, MA-3, MA-4, CAT-1, CB-1, NR-1, BT-1, and AL-1. In some embodiments, the strain of *Saccharomyces cerevisiae* is CEN.PK.

In some embodiments, the strain is a microbe that is suitable for industrial fermentation. In particular embodiments, the microbe is conditioned to subsist under high solvent concentration, high temperature, expanded substrate utilization, nutrient limitation, osmotic stress due to sugar and salts, acidity, sulfite and bacterial contamination, or combinations thereof, which are recognized stress conditions of the industrial fermentation environment.

Culture and Fermentation Methods

Materials and methods for the maintenance and growth of microbial cultures are well known to those skilled in the art of microbiology or fermentation science (see, for example, Bailey et al., Biochemical Engineering Fundamentals, second edition, McGraw Hill, New York, 1986). Consideration must be given to appropriate culture medium, pH, temperature, and requirements for aerobic, microaerobic, or anaerobic conditions, depending on the specific requirements of the host cell, the fermentation, and the process.

The methods of producing cannabinoids provided herein may be performed in a suitable culture medium in a suitable container, including but not limited to a cell culture plate, a flask, or a fermentor. Further, the methods can be performed at any scale of fermentation known in the art to support industrial production of microbial products. Any suitable fermentor may be used including a stirred tank fermentor, an airlift fermentor, a bubble fermentor, or any combination thereof. In particular embodiments utilizing *Saccharomyces cerevisiae* as the host cell, strains can be grown in a fermentor as described in detail by Kosaric, et al, in Ullmann's Encyclopedia of Industrial Chemistry, Sixth Edition, Volume 12, pages 398-473, Wiley-VCH Verlag GmbH & Co. KDaA, Weinheim, Germany.

In some embodiments, the fermentation composition is fermented using a CBGa synthase expressing strain. In some embodiments, the fermentation composition is fermented using a CBCa synthase expressing strain.

In some embodiments, the culture medium is any culture medium in which a genetically modified microorganism capable of producing a heterologous product can subsist, i.e., maintain growth and viability. In some embodiments, the culture medium is an aqueous medium comprising assimilable carbon, nitrogen, and phosphate sources. Such a medium can also include appropriate salts, minerals, metals, and other nutrients. In some embodiments, the carbon source and each of the essential cell nutrients are added incrementally or continuously to the fermentation medium, and each required nutrient is maintained at essentially the minimum level needed for efficient assimilation

by growing cells, for example, in accordance with a predetermined cell growth curve based on the metabolic or respiratory function of the cells which convert the carbon source to a biomass.

Suitable conditions and suitable medium for culturing microorganisms are well known in the art. In some embodiments, the suitable medium is supplemented with one or more additional agents, such as, for example, an inducer (e.g., when one or more nucleotide sequences encoding a gene product are under the control of an inducible promoter), a repressor (e.g., when one or more nucleotide sequences encoding a gene product are under the control of a repressible promoter), or a selection agent (e.g., an antibiotic to select for microorganisms comprising the genetic modifications).

In some embodiments, the carbon source is a monosaccharide (simple sugar), a disaccharide, a polysaccharide, a non-fermentable carbon source, or one or more combinations thereof. Non-limiting examples of suitable monosaccharides include glucose, galactose, mannose, fructose, ribose, and combinations thereof. Non-limiting examples of suitable disaccharides include sucrose, lactose, maltose, trehalose, cellobiose, and combinations thereof. Non-limiting examples of suitable polysaccharides include starch, glycogen, cellulose, chitin, and combinations thereof. Non-limiting examples of suitable non-fermentable carbon sources include acetate and glycerol.

The concentration of a carbon source, such as glucose or sucrose, in the culture medium should promote cell growth, but not be so high as to repress growth of the microorganism used. Typically, cultures are run with a carbon source, such as glucose or sucrose, being added at levels to achieve the desired level of growth and biomass. Production of cannabinoids may also occur in these culture conditions, but at undetectable levels (with detection limits being about <0.1 g/l). In other embodiments, the concentration of a carbon source, such as glucose or sucrose, in the culture medium is greater than about 1 g/L, preferably greater than about 2 g/L, and more preferably greater than about 5 g/L. In addition, the concentration of a carbon source, such as glucose or sucrose, in the culture medium is typically less than about 100 g/L, preferably less than about 50 g/L, and more preferably less than about 20 g/L. It should be noted that references to culture component concentrations can refer to both initial and/or ongoing component concentrations. In some cases, it may be desirable to allow the culture medium to become depleted of a carbon source during culture.

Sources of assimilable nitrogen that can be used in a suitable culture medium include, but are not limited to, simple nitrogen sources, organic nitrogen sources and complex nitrogen sources. Such nitrogen sources include anhydrous ammonia, ammonium salts and substances of animal, vegetable and/or microbial origin. Suitable nitrogen sources include, but are not limited to, protein hydrolysates, microbial biomass hydrolysates, peptone, yeast extract, ammonium sulfate, urea, and amino acids. Typically, the concentration of the nitrogen sources in the culture medium is greater than about 0.1 g/L, preferably greater than about 0.25 g/L, and more preferably greater than about 1.0 g/L. Beyond certain concentrations, however, the addition of a nitrogen source to the culture medium is not advantageous for the growth of the microorganisms. As a result, the concentration of the nitrogen sources, in the culture medium is less than about 20 g/L, preferably less than about 10 g/L and more preferably less than about 5 g/L. Further, in some instances it may be desirable to allow the culture medium to become depleted of the nitrogen sources during culture.

The effective culture medium can contain other compounds such as inorganic salts, vitamins, trace metals, or growth promoters. Such other compounds can also be present in carbon, nitrogen, or mineral sources in the effective medium or can be added specifically to the medium.

The culture medium can also contain a suitable phosphate source. Such phosphate sources include both inorganic and organic phosphate sources. Preferred phosphate sources include, but are not limited to, phosphate salts such as mono or dibasic sodium and potassium phosphates, ammonium phosphate, and mixtures thereof. Typically, the concentration of phosphate in the culture medium is greater than about 1.0 g/L, preferably greater than about 2.0 g/L, and more preferably greater than about 5.0 g/L. Beyond certain concentrations, however, the addition of phosphate to the culture medium is not advantageous for the growth of the microorganisms. Accordingly, the concentration of phosphate in the culture medium is typically less than about 20 g/L, preferably less than about 15 g/L, and more preferably less than about 10 g/L.

A suitable culture medium can also include a source of magnesium, preferably in the form of a physiologically acceptable salt, such as magnesium sulfate heptahydrate, although other magnesium sources in concentrations that contribute similar amounts of magnesium can be used. Typically, the concentration of magnesium in the culture medium is greater than about 0.5 g/L, preferably greater than about 1.0 g/L, and more preferably greater than about 2.0 g/L. Beyond certain concentrations, however, the addition of magnesium to the culture medium is not advantageous for the growth of the microorganisms. Accordingly, the concentration of magnesium in the culture medium is typically less than about 10 g/L, preferably less than about 5 g/L, and more preferably less than about 3 g/L. Further, in some instances, it may be desirable to allow the culture medium to become depleted of a magnesium source during culture.

In some embodiments, the culture medium can also include a biologically acceptable chelating agent, such as the dihydrate of trisodium citrate. In such instances, the concentration of a chelating agent in the culture medium is greater than about 0.2 g/L, preferably greater than about 0.5 g/L, and more preferably greater than about 1 g/L. Beyond certain concentrations, however, the addition of a chelating agent to the culture medium is not advantageous for the growth of the microorganisms. Accordingly, the concentration of a chelating agent in the culture medium is typically less than about 10 g/L, preferably less than about 5 g/L, and more preferably less than about 2 g/L.

The culture medium can also initially include a biologically acceptable acid or base to maintain the desired pH of the culture medium. Biologically acceptable acids include, but are not limited to, hydrochloric acid, sulfuric acid, nitric acid, phosphoric acid, and mixtures thereof. Biologically acceptable bases include, but are not limited to, ammonium hydroxide, sodium hydroxide, potassium hydroxide, and mixtures thereof. In some embodiments, the base used is ammonium hydroxide.

The culture medium can also include a biologically acceptable calcium source, including, but not limited to, calcium chloride. Typically, the concentration of the calcium source, such as calcium chloride, dihydrate, in the culture medium is within the range of from about 5 mg/L to about 2000 mg/L, preferably within the range of from about 20 mg/L to about 1000 mg/L, and more preferably in the range of from about 50 mg/L to about 500 mg/L.

The culture medium can also include sodium chloride. Typically, the concentration of sodium chloride in the culture medium is within the range of from about 0.1 g/L to about 5 g/L, preferably within the range of from about 1 g/L to about 4 g/L, and more preferably in the range of from about 2 g/L to about 4 g/L.

In some embodiments, the culture medium can also include trace metals. Such trace metals can be added to the culture medium as a stock solution that, for convenience, can be prepared separately from the rest of the culture medium. Typically, the amount of such a trace metals solution added to the culture medium is greater than about 1 mL/L, preferably greater than about 5 mL/L, and more preferably greater than about 10 mL/L. Beyond certain concentrations, however, the addition of trace metals to the culture medium is not advantageous for the growth of the microorganisms. Accordingly, the amount of such a trace metals solution added to the culture medium is typically less than about 100 mL/L, preferably less than about 50 mL/L, and more preferably less than about 30 mL/L. It should be noted that, in addition to adding trace metals in a stock solution, the individual components can be added separately, each within ranges corresponding independently to the amounts of the components dictated by the above ranges of the trace metals solution.

The culture medium can include other vitamins, such as pantothenate, biotin, calcium, pantothenate, inositol, pyridoxine-HCl, and thiamine-HCl. Such vitamins can be added to the culture medium as a stock solution that, for convenience, can be prepared separately from the rest of the culture medium. Beyond certain concentrations, however, the addition of vitamins to the culture medium is not advantageous for the growth of the microorganisms.

The culture medium may be supplemented with hexanoic acid or hexanoate as a precursor for the cannabinoid biosynthetic pathway. The hexanoic acid may have a concentration of less than 3 mM hexanoic acid (e.g., from 1 nM to 2.9 mM hexanoic acid, from 10 nM to 2.9 mM hexanoic acid, from 100 nM to 2.9 mM hexanoic acid, or from 1 μM to 2.9 mM hexanoic acid).

The fermentation methods described herein can be performed in conventional culture modes, which include, but are not limited to, batch, fed-batch, cell recycle, continuous and semi-continuous. In some embodiments, the fermentation is carried out in fed-batch mode. In such a case, some of the components of the medium are depleted during culture, including pantothenate during the production stage of the fermentation. In some embodiments, the culture may be supplemented with relatively high concentrations of such components at the outset, for example, of the production stage, so that growth and/or production is supported for a period of time before additions are required. The preferred ranges of these components are maintained throughout the culture by making additions as levels are depleted by culture. Levels of components in the culture medium can be monitored by, for example, sampling the culture medium periodically and assaying for concentrations. Alternatively, once a standard culture procedure is developed, additions can be made at timed intervals corresponding to known levels at particular times throughout the culture. As will be recognized by those in the art, the rate of consumption of nutrient increases during culture as the cell density of the medium increases. Moreover, to avoid introduction of foreign microorganisms into the culture medium, addition is performed using aseptic addition methods, as are known in the art. In addition, anti-foaming agent may be added during the culture.

The temperature of the culture medium can be any temperature suitable for growth of the genetically modified cells and/or production of compounds of interest. For example, prior to inoculation of the culture medium with an inoculum, the culture medium can be brought to and maintained at a temperature in the range of from about 20 °C to about 45 °C, preferably to a temperature in the range of from about 25 °C to about 40 °C and more preferably in the range of from about 28 °C to about 32 °C.

The pH of the culture medium can be controlled by the addition of acid or base to the culture medium. In such cases when ammonia is used to control pH, it also conveniently serves as a nitrogen source in the culture medium. Preferably, the pH is maintained from about 3.0 to about 8.0, more preferably from about 3.5 to about 7.0, and most preferably from about 4.0 to about 6.5.

In some embodiments, the carbon source concentration, such as the glucose concentration, of the culture medium is monitored during culture. Glucose or sucrose concentration of the culture medium can be monitored using known techniques, such as, for example, use of the glucose oxidase enzyme test or high-pressure liquid chromatography, which can be used to monitor glucose concentration in the supernatant, e.g., a cell-free component of the culture medium. As stated previously, the carbon source concentration should be kept below the level at which cell growth inhibition occurs. Although such concentration may vary from organism to organism, for glucose as a carbon source, cell growth inhibition occurs at glucose concentrations greater than at about 60 g/L and can be determined readily by trial. Accordingly, when glucose is used as a carbon source the glucose is preferably fed to the fermenter and maintained below detection limits. Alternatively, the glucose concentration in the culture medium is maintained in the range of from about 1 g/L to about 100 g/L, more preferably in the range of from about 2 g/L to about 50 g/L, and yet more preferably in the range of from about 5 g/L to about 20 g/L. Although the carbon source concentration can be maintained within desired levels by addition of, for example, a substantially pure glucose solution, it is acceptable, and may be preferred, to maintain the carbon source concentration of the culture medium by addition of aliquots of the original culture medium. The use of aliquots of the original culture medium may be desirable because the concentrations of other nutrients in the medium (e.g., the nitrogen and phosphate sources) can be maintained simultaneously. Likewise, the trace metals concentrations can be maintained in the culture medium by addition of aliquots of the trace metals solution.

Examples

The following examples are put forth so as to provide those of ordinary skill in the art with a description of how the compositions and methods described herein may be used, made, and evaluated, and are intended to be purely exemplary of the invention and are not intended to limit the scope of what the inventors regard as their invention.

Example 1: Large scale bioconversion of cannabigerolic acid to cannabichromenic acid

This example shows the near complete conversion of cannabigerolic acid (CBGa) to cannabichromenic acid (CBCa) in a sunflower oil overlay by CBCa synthase at high titers, and the effects of reaction parameters and additives on the reaction rates and completion.

The production process for CBC involved producing CBGa with a HOSUN overlay, removing cells via liquid-solid centrifugation, followed by a separation of the aqueous phase from the overlay via liquid-liquid centrifugation. This process was preferably conducted without adjustment to alkaline pH or application of heat during liquid-liquid separation. An alternate preferred method was using a demulsification route that did not cause decarboxylation of CBGa into non-bioconvertable CBG or interfere with downstream enzymatic conversion of CBGa to CBCa (certain chemical additions such as solvents or surfactants may denature enzymes, for example). Fermentation of the Y81037 strain produces a CBCa synthase, which converted CBGa into CBCa. The synthase broth was separated from cells via liquid-solid centrifugation and combined with the CBGa containing HOSUN in a stirred reactor in a process referred to as bioconversion. These processes are summarized in FIG. 1.

At the 300 L scale, it was demonstrated that the strain Y81037 was capable of converting >97% of CBGa (61 g/kg) in overlay to CBCa over 24 hours, and there was >99.3% conversion at 48 hours (FIG. 2). This corresponded to ~60 g/kg of CBCa in the overlay (Table 4), or 3 g/L in the total reaction. The titer, rate, and purity were much higher than reported elsewhere.

Table 4: Bioconversion time course titers.

Time (hours)	CBCA (g/kg)	CBG (g/kg)	CBGA (g/kg)	CBCA (mol/kg)	CBGA (mol/kg)	Conversion by Gain (%)	Conversion by Loss (%)
0	0	0.74	61.47	0	0.17	0	0
0.5	6.5	0.74	59.70	0.02	0.17	9.87	2.88
24	58.97	0.74	1.8	0.16	0	97.05	97.07
48	59.83	0.74	0.4	0.17	0	99.34	99.35
72	59.5	0.74	0.36	0.17	0	99.40	99.41

Overlay concentration

With the previous strain produced at pilot scale, it was demonstrated that the reaction rate was higher both when the concentration of CBGa was increased in the overlay, as well as when the total concentration of CBGa was increased in the reaction (FIG. 3).

With the Y81037 strain, a >80% conversion of CBGa was achieved after 24 hours at bench scale when the overlay loading in the reaction was increased to 10 wt % while keeping $[CBGa]_{overlay}$ constant but doubling $[CBGa]_{reaction}$. The reaction conditions and results are summarized in Table 5. These results indicated that the origin of the decrease in reaction rate shown FIG. 4 with increasing overlay % has been overcome with the new strain.

Table 5: Effect of pH and overlay % on 24-hour conversion.

pH	Overlay (wt %)	$[CBGa]_{overlay}$ (g/kg)	$[CBGa]_{reaction}$ (g/L)	$[CBCa]_{Overlay}$ (g/kg)	$[CBCa]_{reaction}$ (g/L)	Conversion (%)
5	5	61	3.05	48.3	2.415	80
5	10	61	6.1	50.6	5.06	83
6.5	5	61	3.05	49.4	2.47	81
6.5	10	61	6.1	51.2	5.12	84

Temperature and stirring effects

The effects of temperature and agitation intensity on the conversion at 24 hours (Table 6.) was also assessed. Increasing the temperature to 35 °C and increasing agitation with increased stir rate and changing impeller geometry resulted in a 1.5X and 2.1X increase in conversion at 24 hours with Y78658. An intermediate timepoint (6-12 hours) would be needed to assess the impact of stirring and temperature on Y81037 because by 24 hours, near complete conversion was achieved.

Table 6: Effect of temperature and agitation on conversion comparted with the base case

Temperature (°C)	**Agitation**	**X Change Base Conversion at 24 hours**
30	Low	1
35	Low	1.5
30	Medium	1.05
30	High	2.1

Reaction additives

As shown in FIG. 4, addition of molecules with different hydrophobic moieties resulted in an increased rate of conversion that was especially prominent at early reaction timepoints. Soluble substrates such as cyclodextrins resulted in the highest rate increase, while solid particles with hydrophobic surfaces such as cellulose also increased the rate of bioconversion. This is the first time cyclodextrin has been used with CBCa synthase in an overlay system to improve bioconversion.

Recycle

To determine if the decline in rate was due to enzyme or substrate related phenomena, an enzyme recycling experiment was performed where fresh substrate was added to previously reacted enzyme, and the rates were measured (**Error! Reference source not found.**). The much higher rate of the system with cyclodextrin which acts as a carrier molecule of CBGa into the aqueous phase, along with the results in **Error! Reference source not found.**, where reaction rate decreases the larger the oil "reservoir" was for a given total CBGa loading, indicated that the system was initially substrate limited, and improved the accessibility of the substrate to the enzyme resulted in faster reactions. The fresh addition of substrate at later times should result in a reaction rate bump if the system continues to be substrate limited. However, the fact that both +cyclodextrin and -cyclodextrin

bioconversion showed no change in rate after new CBGa addition indicated that the longer-term rate decline was due to enzymatic inactivation.

Downstream purification

The downstream purification process for CBC was very similar to that for CBG. The overlay from the bioconversion was demulsified with Tergazyme® and adjusted to pH 8, which brok4 the emulsion and aided in decarboxylation of CBCa to CBG. The mixture was separated in a liquid-liquid centrifuge into a crude oil light phase and an aqueous heavy phase. The crude oil was distilled at the same conditions as for CBG (two WFE passes, the first at 180C, 1 torr to remove light boilers, and the second to distill cannabinoids from the oil at 240 °C, 1 torr). This resulted in ~70% pure CBC distillate, with <5% CBG and other cannabinoid impurities. Crystallization was attempted for CBC, but a preliminary screen of solvents and low temperatures indicated that CBC was not a crystallizable substance.

Example 2. Production of Cannabichromenic Acid Synthase Variants

To identify variant CBCa synthase enzymes capable of increasing production of cannabichromenic acid, a screen of the variants was performed.

Transformation of Heterologous Nucleic Acids into Yeast Cells

Each DNA construct was integrated into *Saccharomyces cerevisiae* (CEN.PK113-7D) using standard molecular biology techniques in an optimized lithium acetate transformation. Briefly, cells were grown overnight in yeast extract peptone dextrose (YPD) medium at 30 °C with shaking (200 rpm), diluted to an OD_{600} of 0.1 in 100 mL YPD, and grown to an OD_{600} of 0.6 – 0.8. For each transformation, 5 mL of culture were harvested by centrifugation, washed in 5 mL of sterile water, spun down again, resuspended in 1 mL of 100 mM lithium acetate, and transferred to a microcentrifuge tube. Cells were spun down (13,000x *g*) for 30 s, the supernatant was removed, and the cells were resuspended in a transformation mix consisting of 240 µL 50% PEG, 36 µL 1 M lithium acetate, 10 µL boiled salmon sperm DNA, and 74 µL of donor DNA. For transformations that require expression of the endonuclease F-Cph1, the donor DNA included a plasmid carrying the F-CphI gene expressed under the yeast TDH3 promoter. F-CphI endonuclease expressed in such a manner cuts a specific recognition site engineered in a host strain to facilitate integration of the target gene of interest. Following a heat shock at 42 °C for 40 min, cells were recovered overnight in YPD medium before plating on selective medium. When applicable, DNA integration was confirmed by colony PCR with primers specific to the integrations.

Culturing of Yeast

For routine strain characterization in a 96-well-plate format, yeast colonies were picked into a 1.1-mL-per-well capacity 96-well 'Pre-Culture plate' filled with 360 µL per well of pre-culture medium. Pre-culture medium consists of Bird Seed Media (BSM) at pH 5.05 with 14 g/L sucrose, 7 g/L maltose, 3.75g/L ammonium sulfate, and 1 g/L lysine. Cells were cultured at 28°C in a high-capacity

microtiter plate incubator shaking at 1000 rpm and 80% humidity for 3 days until the cultures reached carbon exhaustion.

The growth-saturated cultures were sub-cultured by taking 14.4 µL from the saturated cultures and diluting into a 2.2 mL per well capacity 96-well 'production plate' filled with 360 µL per well of production medium. Production medium consists of BSM at pH 5.05 with 40 g/L sucrose (for Examples 3-5 or 60 g/L raffinose for Example 6), 3.75g/L ammonium sulfate. Cannabigerolic acid (CBGa) in ethanol was added to the production medium to a final CBGa concentration of 200 mg/L (Examples 3 and 4) or 500 mg/L (Examples 5 and 6) and 1% EtOH (V/V) as a substrate for CBCaS made by the yeast strains.

Cells in the production medium were cultured at 30°C in a high-capacity microtiter plate shaker at 1000 rpm and 80% humidity for an additional 4 days prior to extraction and analysis.

Total activity of CBCaS in the culture was inferred from the amount of cannabichromenic acid (CBCa) measured at the end of the 4 days culture.

Analytical methods for Product Extraction and Titer Determination.

Production of cannabichromenic acid from cannabigerolic acid was initially analyzed in high-throughput by mass spectrometer (Agilent 6470-QQQ) with a mass spectrometry system autosampler with C4 cartridge. The conditions used are shown in Tables 7 and 8.

Table 7. Mass spectrometry system configuration

Pump 1: 0.1% acetic acid in water	0.8 mL/min
Pump 2: 0.1% formic acid in acetonitrile	1.5 mL/min
Pump 3: 0.1% formic acid in 40% acetone in water	0.8 mL/min
State 1: Aspirate	600 ms
State 2: Load/Wash	2000 ms
State 3: Extra wash	500 ms
State 4: Elute	6000 ms
State 5: Re-equilibrate	1000 ms

Table 8. 6470-QQQ MS method configurations

Ion Source	AJS ESI
Time Filtering peak width	0.02 min
Stop Time	No limit/as pump
Scan Type	MRM
Diverter Valve	To MS
Delta EMV	(+)0/ (-)0
Ion Mode (polarity)	Negative
Gas Temp	300 °C
Gas Flow	13 L/min
Nebulizer	30 psi
Sheath Gas Temp	30 °C
Sheath Gas Flow	12 L/min
Negative Capillary V	3500 V

The peak areas from a chromatogram from a mass spectrometer were used to generate the calibration curve using authentic standards. The amount in moles of each compound were generated through external calibration using an authentic standard.

Hit samples from the initial screen were then analyzed for cannabichromenic acid and cannabigerolic acid on a weight per volume basis, by the method below. All measurements were performed by reverse phase ultra-high pressure liquid chromatography and ultraviolet detection (UPLC-UV) using Thermo Vanquish Flex Binary UHPLC System with a Vanquish Diode Array Detector HL using the conditions in Tables 9-13.

Table 9. Mobile Phases and Column Information

Mobile Phase A:	99.9% water + 0.1% Formic Acid, 5mM ammonium formate
Mobile Phase B:	99.9% acetonitrile + 0.1% Formic acid
Column	Thermo Scientific Accucore Polar Premium C18 100mm x 2.1 mm X 2.6um, Thermo P/N 28026-103030
Guard Column	Thermo Scientific Guard Cartridge, 4 PK, P/N 28103014001
Guard Column holder	Uniguard Holder, ThermoFisher, 4.0-4.6 mm, P/N 850-00

Table 10. Gradient Method

Time	Flow [mL/min]	%A	%B	Curve
0.00	1.2	70	30	5
1.00	1.2	20	80	5
1.75	1.2	12.5	87.5	5
1.80	1.2	70	30	5
2.1	1.2	70	30	5

Table 11. Autosampler Parameters

Draw speed	2.00 µl/s
Dispense speed	5.00 µl/s
Injection wash mode	Both (before and after draw)
Injection wash time	5.0 s
Injection wash speed	10.0 µl/s
Sample puncture – puncture offset	0 um
Temperature control	8 °C

Table 12. Column compartment settings

Temperature control	On
Temperature	50.0 °C
Ready temp delta	0.50 °C
Equilibration time	1.0 min
Thermostatting mode	Still air
Fan Speed	5

Table 13. Detector Settings

UV-Vis Channel 1 Wavelength	270nm
Data collection rate	50.0 Hz
Response time	0.010 s
Peak width	0.100 min

Analytes were identified by retention time compared to an authentic standard. The peak areas were used to generate the linear calibration curve for each analyte. At the conclusion of the incubation of the production plate, methanol was added to each well such that the final concentration is 67% (v/v) methanol. An impermeable seal was added, and the plate is shaken at 1000 rpm for 30 seconds to lyse the cells and extract cannabinoids. The plate was centrifuged for 30 seconds at 200 × g to pellet cell debris. 300 µL of the clarified sample was moved to an empty 1.1-mL-capacity 96-well plate and sealed with a foil seal. The sample plate was stored at -20°C until analysis.

CBCaS expression construct

CBCaS enzyme was identified in Laverty, et al, Laverty et al. *Genome Res.* 2019 Jan;29(1):146-156. DNA sequence encoding the mature peptide (amino acids 29-545 of UNIPROT: Q33DQ2, SEQ ID:1) was optimized for yeast expression and synthesized with DNA coding for the 19

amino acid yeast mat-α signal sequence followed by a modified "L19" pro-region 5' to the mature peptide coding sequence.

The gene was inserted into the host genome by replacement of the GAS4 gene by homologous integration, using 540 nucleotide homology to genomic sequence upstream of the GAS4 gene and 555 nucleotide homology to genomic sequence downstream of the GAS4 gene.

Expression of the CBCaS or variants thereof was driven by either the promoter pGAL1 or the promoter pGAL1_7. The sequence of the first 167 nucleotides downstream of the SDH3 gene were used as the terminator.

Wild-type CBCaS or variants thereof were fused seamlessly to the promoter and terminator. The 3' end of the upstream GAS4 homology region was joined to the 5' end of the promoter by a cloning linker. The 3' end of tSDH3 was joined to the 5' end of the downstream GAS4 homology by a cloning linker.

CBCaS N and C terminal modifications

Various replacements for the mat-α/LA19Pro N-terminal sequence were tested, including an OST1 signal sequence/KEX2 motif, an OST1 signal sequence/KEX2 motif plus 5 amino acids, and a CWP2 signal sequence/KEX2 motif plus 28 amino acids. These were evaluated in combination with various C-terminal extensions: 6xHIS (3 histidine residues were added to the existing 3 C-terminal histidines in the original sequence), "SAG_short", "12aaLink4_SAG1", and "15aaLink2_FLO5". Combinations that improved over the parent are summarized in Table 14.

Table 14. Summary of improved constructs

N-terminus	C-Terminus	Promoter pGAL1	Promoter pGAL1_7
MA_LA19pro	6xHIS	improved	improved
MA_LA19pro	SAG_short	improved	improved
OST1/KEX2	12aaLink4_SAG1	improved	not improved
OST1/KEX2 extended	12aaLink4_SAG1	improved	improved
OST1/KEX2 extended	15aaLink2_FLO5	improved	improved
CWP2/KEX2 extended2	12aaLink4_SAG1	not improved	improved

CBCaS point mutants

Site saturation mutagenesis was used to improve CBCaS activity. 385 positions were chosen for mutagenesis based on previous experience with site saturation mutagenesis of the similar CBDaS enzyme. Each position was mutagenized using the degenerate codon NNK (where N can encode any of the 4 nucleotides and K encodes for either G or T) and transformed separately. The degenerate codon NNK can code for all 20 amino acids (and the TAG stop codon). Multiple isolates

from each transformation were screened to accumulate data on multiple substitutions at each position. Mutagenesis was performed on SEQ ID NO: 1

Top hits were re-transformed into the parent strain to ensure that performance improvements were due to the CBCaS mutation rather than any spontaneous mutations to the strain expressing the variant. Confirmed hits were sequenced by Sanger sequencing to determine which amino acid produced the improved activity. Improved variants are summarized in Table 15 and FIG. 6.

Table 15. Summary of improved CBCaS variants

Variant	Amino acid SEQ ID
Q75L	2
Q75E	3
F82I	4
T130L	5
T130L2	6
S140T	7
S140A	8
V169E	9
N240Q	10
N240M	11
V294S	12
V294R	13
V294E	14
A299V	15
K305C	16
D328P	17
T335L	18
T335L2	19
R340M	20
R340G	21
H354V	22
H354V2	23
L435A	24
Y461I	25
K535M	26
S540D	27
T545E	28

CBCaS combinatorial library mutants

Sixteen of the top individual CBDaS point mutants described above in Table 16 were consolidated together using a shuffling approach with 9 DNA segments. The 16 positions for recombination were Q75E, F82I, T130L, S140A, V169E, N240M, V294S, A299V, K305C, T335L, R340M, H354V, L435A, Y461I, K535M, and S540D. DNA fragments were combined so as to produce a library with a high percentage of variants. All combined variants tested were significantly improved

over the parent sequence. Forty-four variants with the greatest improvement in broth activity over the parent were sequenced by Sanger sequencing. Forty-two of these were unique variants, indicating a high diversity library. The number of mutations relative to the parent among these 42 variants ranged from 9 to 16, with this high number of mutations being expected based on the library construction. The improved variants are summarized in FIG. 7.

Other Embodiments

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the invention that come within known or customary practice within the art to which the invention pertains and may be applied to the essential features hereinbefore set forth, and follows in the scope of the claims. Other embodiments are within the claims

SEQUENCE LISTING

SEQ ID NO: 1 wild-type CBCa synthase

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFSFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKSFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH

SEQ ID NO: 2 -Q75L

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPLENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 3 - Q75E

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 4 – F82I

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 5 – T130L

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 6 – T130L

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 7 – S140T

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
TILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 8 – S140A

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
TILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 9 – V169E

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 10 - N240Q

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRQYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 11 – N290M

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 12 – V294S

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 13 – V294R

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVRVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG

IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 14 – V294E

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVEVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 15 – A299V

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 16 – K305C

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 17 – D328P

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYPKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 18 – T335L

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI

TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 19 – T335L

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 20 – R340G

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTGNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 21 – R340M

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTMNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 22 – H354V

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 23 – H354V

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN

EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 24 – L435A

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 25 – Y461I

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 26 – K535M

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 27 – S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 28 – T545E

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRAG
IMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYEQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 29 -
Q75E/F82I/T130L/S140A/V169E/N240M/V294S/A299V/K305C/T335L/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 30 -
Q75E/F82I/T130L/S140A/V169E/V294S/A299V/K305C/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 31 -
Q75E/F82I/T130L/S140A/V169E/N240M/V294S/A299V/T335L/R340M/H354V/L435A/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 32 -
Q75E/T130L/S140A/V169E/N240M/V294S/A299V/T335L/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK

EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 33 -
Q75E/T130L/S140A/V169E/N240M/V294S/A299V/T335L/R340M/H354V/L435A/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 34 -
Q75E/F82I/T130L/V169E/N240M/V294S/A299V/K305C/T335L/R340M/L435A/Y461I/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 35 - Q75E/F82I/T130L/S140A/V169E/T335L/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 36 - Q75E/F82I/T130L/S140A/V294S/A299V/K305C/H354V/L435A/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 37 - Q75E/S140A/V169E/N240M/K305C/T335L/R340M/H354V/L435A/Y461I/K535M/

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA

AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKATIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 38 - Q75E/V169E/V294S/A299V/R340M/H354V/L435A/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 39 -
Q75E/F82I/T130L/S140A/V169E/V294S/A299V/K305C/T335L/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 40 -
Q75E/F82I/T130L/V169E/N240M/V294S/A299V/K305C/T335L/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 41 - Q75E/F82I/T130L/V169E/V294S/A299V/T335L/H354V/L435A/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 42 -
Q75E/F82I/T130L/S140A/V169E/N240M/V294S/A299V/K305C/R340M/H354V/L435A/Y461I/K535M/

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 43 -
Q75E/T130L/S140A/V169E/V294S/A299V/K305C/R340M/H354V/L435A/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 44 - Q75E/F82I/T130L/V169E/A299V/K305C/R340M/H354V/L435A/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 45- Q75E/F82I/T130L/V169E/N240M/V294S/T335L/R340M/H354V/L435A/K535M/

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPESPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 46 -
Q75E/F82I/S140A/V294S/K305C/T335L/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW

AIRGGGGENFGIIAAWKIKLVSVPSKATIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 47 -
Q75E/F82I/T130L/S140A/V169E/N240M/V294S/A299V/T335L/R340M/L435A/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 48 -
F82I/T130L/S140A/V169E/N240M/V294S/A299V/K305C/H354V/L435A/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 49 -
Q75E/F82I/S140A/V169E/N240M/V294S/A299V/K305C/T335L/R340M/H354V/L435A/Y461I/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 50 -
Q75E/F82I/T130L/S140A/V169E/V294S/A299V/K305C/T335L/R340M/H354V/L435A/Y461I/K535M/

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 51 - T130L/S140A/V169E/V294S/A299V/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 52 - Q75E/F82I/T130L/N240M/V294S/A299V/T335L/H354V/L435A/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 53 -
Q75E/F82I/S140A/V169E/N240M/V294S/A299V/K305C/T335L/R340M/H354V/L435A/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLYPYGGIMDEISESAIPFPHRA
GIMYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPEDPNNYTQA
RIWGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 54 -
Q75E/F82I/T130L/S140A/V169E/N240M/V294S/A299V/K305C/R340M/L435A/Y461I/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTMNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 55 -
Q75E/F82I/S140A/V169E/N240M/A299V/K305C/R340M/H354V/L435A/Y461I/K535M/

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK

EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 56 -
Q75E/F82I/T130L/S140A/V169E/N240M/V294S/A299V/K305C/R340M/H354V/L435A/K535M/S540D

SEQ ID NO: 57 -
Q75E/T130L/S140A/V169E/V294S/A299V/K305C/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTTHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 58 - Q75E/F82I/V169E/V294S/A299V/K305C/T335L/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 59 -
Q75E/F82I/T130L/S140A/V169E/N240M/V294S/K305C/T335L/R340M/L435A/Y461I/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKATIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 60 -
F82I/T130L/S140A/V169E/N240M/V294S/A299V/K305C/T335L/R340M/H354V/L435A/Y461I/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI

MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 61 -
F82I/T130L/S140A/V169E/N240M/A299V/T335L/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 62-
Q75E/F82I/S140A/V169E/N240M/V294S/A299V/K305C/T335L/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 63 -
Q75E/F82I/T130L/S140A/V169E/N240M/V294S/T335L/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 64 -
F82I/T130L/S140A/V169E/N240M/V294S/A299V/K305C/T335L/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPQENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 65 -
Q75E/F82I/T130L/S140A/V169E/N240M/A299V/T335L/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVVVPSKVTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 66 - Q75E/F82I/T130L/S140A/N240M/V294S/A299V/K305C/T335L/R340M/H354V/L435A/Y461I/K535M/S540D

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCISEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVLPSNVSHIQA
AILCSKKVGLQIRTRSGGHDAEGLSYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTMNI
TDNHGKNKTTVVGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPEDPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 67 - Q75E/V169E/N240M/V294S/A299V/K305C/T335L/L435A/Y461I/K535M/

MRFPSIFTAVLFAASSALAQPIDDTESNTTSVNLMADDTESRFATNTTLALDVVNLISMAKREEAEAEA
EPKNPEENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQA
SILCSKKVGLQIRTRSGGHDAEGLSYISQEPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWIN
EMNENFTFPGGYCPTVGVGGHFSGGGYGALMRMYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVSVPSKVTIFSVCKNMEIHGLVKLFNKWQNIAYKYDKDLMLTLHFRTRNI
TDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKTFPELGIKKTDCKELSWIDTTIFYSGVVNYNTANFKK
EILLDRSAGKKTAFSIKLDYVKKAIPETAMVKILEKLYEEEVGVGMYVLIPYGGIMDEISESAIPFPHRAGI
MYELWYTATWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGMTNPESPNNYTQARI
WGEKYFGKNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*

SEQ ID NO: 68 AAE candidate isolated from Pseudonocardia sp. N23

Amino acid sequence

MTAAQAPDPAGVPLVERTVPRMLARSAALDPDRPFVVTRERTWSHTDAHRIVATLAAAFTDRGIGQG
SRVAVMMPTSPRHVWLLLALAHLRAVPVALNPDASGEVLRYFVADSECVLGVVDQERAAAFATAAG
PDGPPAIVLPPGADDLGELGSAGPGPLDPGAASFSDTFVVLYTSGSTGMPKATAVTHAQVITCGAVF
TDRLGLGPADRLYTCLPLFHINATAYSLSGALVSGASLALGPHFSATTFWDDVADLGATEVNAMGSM
VRILQSRPPRPAERAHRVRTMFVAPLPPDAVELSERFGLDFATCYAQTEWLPSSMTRPGEGYGRPG
ATGPVLPWTEVRIVGDDDRPLPAGQTGEIILRPRDPYTTFQGYLGKPQETVDAWRNLWFHTGDLGDI
GPDGWLHYRGRRKDVIRRRGENIPATVVEDLLAGHPDIAEVAAVSVPAHISEEEIFAFVVPGAGAALT
TADVEAHAHAVLPRYMVPSYLALVPDLPRTATNKIAKVELTERARAAVEGTGDPADAPTRTSAADRV
VVPAAE

SEQ ID NO: 69 – AAE candidate isolated from Pseudomonas putida

Amino acid sequence

MMVPTLEHELAPNEANHVPLSPLSFLKRAAQVYPQRDAVIYGARRYSYRQLHERSRALASALERVGV
QPGERVAILAPNIPEMLEAHYGVPGAGAVLVCINIRLEGRSIAFILRHCAAKVLICDREFGAVANQALAM
LDAPPLLVGIDDDQAERADLAHDLDYEAFLAQGDPARPLSAPQNEWQSIAINYTSGTTGDPKGVVLH
HRGAYLNACAGALIFQLGPRSVYLWTLPMFHCNGWSHTWAVTLSGGTHVCLRKVQPDAINAAIAEHA
VTHLSAAPVVMSMLIHAEHASAPPVPVSVITGGAAPPSAVIAAMEARGFNITHAYGMTESYGPSTLCL

WQPGVDELPLEARAQFMSRQGVAHPLLEEATVLDTDTGRPVPADGLTLGELVVRGNTVMKGYLHNP
EATRAALANGWLHTGDLAVLHLDGYVEIKDRAKDIIISGGENISSLEIEEVLYQHPEVVEAAVVARPDS
RWGETPHAFVTLRADALASGDDLVRWCRERLAHFKAPRHVSLVDLPKTATGKIQKFVLREWARQQE
AQIADAEH

SEQ ID NO: 70 – AAE candidate isolated from Streptomyces sp.ADI96-02

Amino acid sequence

MLSTMQDVPLTVTRILQHGMTIHGKSQVTTWTGEPEPHRRTFAEIGARATRLAHALRDELGIDGDQR
VATLMWNNAEHVEAYLAVPSMGAVLHTLNLRLPAEQLIWIVNHADDKVVIVNGSLLPLLVPLLPHLPTV
EHVVVSGPGDRSALAGVAPRVHEYEELIADRPTTYDWPELDERQAAAMCYTSGTTGDPKGVVYSHR
SVYLHSMQVNMTESMGLTDKDTTLVVVPQFHVNAWGLPHATFMAGVNMLMPDRFLQPAPLADMIE
RERPTHAAAVPTIWQGLLAEVTAHPRDLTSMASVTIGGAACPPSLMEAYDKLGVRLCHAWGMTETS
PLGTMANPPAGLSAEEEWPYRVTQGRFPAGVEARLVGPAGDHLPWDGRSAGELEVRGAWIAGAYY
GGADGEHLRPEDKFSADGWLKTGDVGVISADGFLTLTDRAKDVIKSGGEWISSVELENALMAHPDVA
EAAVVAVPDEKWGERPLATVVLKEGAEVGYEALKVFLADSGIAKWQLPERWTVIPAVPKTSVGKFDK
KVIRKQYADGELDITQL

SEQ ID NO: 71 – AAE candidate isolated from Erythrobacter citreus LAMA 915

Amino acid sequence

MSRAECRDRLTAPGERFEIETIDIRGVPTRVWKHAPTNMRQVAMAARTHGDRLFAIYEDERVTYEAW
FRAVARMAAELRERGVAKGDRVALAMRNLPEWPVAFFAATTIGAICVPLNAWWTGPELAFGLANSG
AKLLVCDAERWERIAPHRGELPDLEHALVSRSDAPLEGAEQLEDLLGTPKDYAALPSAALPQVDIDPE
DEATIFYTSGTTGQPKGALGTHRNLCTNIMSSAYNGAIAFLRRGEEPPAPVQKVGLTVIPLFHVTACSA
GLMGYVVAGHTMVFMHKWDPVKAFQLIEREKVNLTGGVPTIAWQLLEHPERANYDLSSLEAVAYGG
APAAPELVRKIHEEFGALPANGWGMTETMATVTGHSSEDYLNRPDSCGPPVAVADLKIVGDDGVTEL
PVGEVGELWARGPMVVKGYWNRPEATAETFVDGWVRTGDLARLDEEGWCYIVDRAKDMIIRGGENI
YSSEVENVLYDHPAVTDAALVAIAHPTLGEEPAAVVHLAPGMSATEDELREWVAARLAKFKVPVRIAF
VQDTLPRNANGKILKKDLGAFFA

SEQ ID NO: 72 – AAE candidate isolated from Saccharomyces cerevisiae

Amino acid sequence

MVAQYTVPVGKAANEHETAPRRNYQCREKPLVRPPNTKCSTVYEFVLECFQKNKNSNAMGWRDVK
EIHEESKSVMKKVDGKETSVEKKWMYYELSHYHYNSFDQLTDIMHEIGRGLVKIGLKPNDDDKLHLYA
ATSHKWMKMFLGAQSQGIPVVTAYDTLGEKGLIHSLVQTGSKAIFTDNSLLPSLIKPVQAAQDVKYIIH
FDSISSEDRRQSGKIYQSAHDAINRIKEVRPDIKTFSFDDILKLGKESCNEIDVHPPGKDDLCCIMYTSG
STGEPKGVVLKHSNVVAGVGGASLNVLKFVGNTDRVICFLPLAHIFELVFELLSFYWGACIGYATVKTL
TSSSVRNCQGDLQEFKPTIMVGVAAVWETVRKGILNQIDNLPFLTKKIFWTAYNTKLNMQRLHIPGGG
ALGNLVFKKIRTATGGQLRYLLNGGSPISRDAQEFITNLICPMLIGYGLTETCASTTILDPANFELGVAG
DLTGCVTVKLVDVEELGYFAKNNQGEVWITGANVTPEYYKNEEETSQALTSDGWFKTGDIGEWEAN
GHLKIIDRKKNLVKTMNGEYIALEKLESVYRSNEYVANICVYADQSKTKPVGIIVPNHAPLTKLAKKLGI
MEQKDSSINIENYLEDAKLIKAVYSDLLKTGKDQGLVGIELLAGIVFFDGEWTPQNGFVTSAQKLKRKD
ILNAVKDKVDAVYSSS

SEQ ID NO: 73 – AAE candidate isolated from Citreicella sp. SE45

Amino acid sequence

MSLSTEETARRRTLAEGAGYDALREGFRWPGAARVNMAEQVCDSWAAREPGRPAILDMRAGGAPE
VVSYGALQALSRRVEAWFRGQGVARGDRVGVLLSQSPLCAAAHIAAWRMGAISVPLFKLFKHDALE
SRLGDSGARVVVSDDEGAAMLAPFGLSVVTEAGLPQDGATEPAADTGPEDPAIIIYTSGTTGKPKGAL
HGHRVLTGHLPGVEMSHDLLGQPGDVLWTPADWAWIGGLFDVLMPGLYLGVPVVAARMPRFEISEC
LRICQQASVRNVFFPPTAFRMLKSEGAELPGLRSVASGGEPLGAEMLAWGRKAFGVEINEFYGQTE
CNMVASSCGALFRARPGCIGKPAPGFHIAVIDEDGNETDGEGDVAIRRGAGSMLLEYWQKPQETAD

KFRGDWLVTGDRGTWEDGYLRFVGREDDVITSAGYRIGPTEIEDCLMTHPAVATVGVVGKPCPLRTE
LVKAYVVLRPGVEVRASELQAWVKERLATYSYPREIAFLDALPMTVTGKVIRKELKAIAAAERTAEAAG
EVS

SEQ ID NO: 74 – AAE candidate isolated from Bacillus subtilis (strain 168)

Amino acid sequence

MNLVSKLEETASEKPDSIACRFKDHMMTYQELNEYIQRFADGLQEAGMEKGDHLALLLGNSPDFIIAF
FGALKAGIVVVPINPLYTPTEIGYMLTNGDVKAIVGVSQLLPLYESMHESLPKVELVILCQTGEAEPEAA
DPEVRMKMTTFAKILRPTSAAKQNQEPVPDDTAVILYTSGTTGKPKGAMLTHQNLYSNANDVAGYLG
MDERDNVVCALPMFHVFCLTVCMNAPLMSGATVLIEPQFSPASVFKLVKQQQATIFAGVPTMYNYLF
QHENGKKDDFSSIRLCISGGASMPVALLTAFEEKFGVTILEGYGLSEASPVTCFNPFDRGRKPGSIGT
SILHVENKVVDPLGRELPAHQVGELIVKGPNVMKGYYKMPMETEHALKDGWLYTGDLARRDEDGYF
YIVDRKKDMIIVGGYNVYPREVEEVLYSHPDVKEAVVIGVPDPQSGEAVKGYVVPKRSGVTEEDIMQH
CEKHLAKYKRPAAITFLDDIPKNATGKMLRRALRDILPQ

SEQ ID NO: 75 – AAE candidate isolated from Saccharomyces cerevisiae

Amino acid sequence

MTEQYSVAVGEAANEHETAPRRNIRVKDQPLIRPINSSASTLYEFALECFTKGGKRDGMAWRDIIDIH
ETKKTIVKRVDGKDKPIEKTWLYYELTPYITMTYEEMICVMHDIGRGLIKIGVKPNGENKFHIFASTSHK
WMKTFLGCMSQGIPVVTAYDTLGESGLIHSMVETDSVAIFTDNQLLSKLAVPLKTAKNVKFVIHNEPID
PSDKRQNGKLYKAAKDAVDKIKEVRPDIKIYSFDEIIEIGKKAKDEVELHFPKPEDPACIMYTSGSTGTP
KGVVLTHYNIVAGIGGVGHNVIGWIGPTDRIIAFLPLAHIFELTFEFEAFYWNGILGYANVKTLTPTSTRN
CQGDLMEFKPTVMVGVAAVWETVRKGILAKINELPGWSQTLFWTVYALKERNIPCSGLLSGLIFKRIR
EATGGNLRFILNGGSAISIDAQKFLSNLLCPMLIGYGLTEGVANACVLEPEHFDYGIAGDLVGTITAKLV
DVEDLGYFAKNNQGELLFKGAPICSEYYKNPEETAAAFTDDGWFRTGDIAEWTPKGQVKIIDRKKNLV
KTLNGEYIALEKLESIYRSNPYVQNICVYADENKVKPVGIVVPNLGHLSKLAIELGIMVPGEDVESYIHE
KKLQDAVCKDMLSTAKSQGLNGIELLCGIVFFEEEWTPENGLVTSAQKLKRRDILAAVKPDVERVYKE
NT

SEQ ID NO: 76 – AAE candidate isolated from Bhargavaea cecembensis DSE10

Amino acid sequence

MYTDHGWIMKRADITPDGTALIDVHTGQRWTYRELAGRTAAYMEQFRSAGLRKGERVAVLSHNRIDL
FAVLFACAGRGLIYVPMNWRLSESELRYIVSDSGPSLLLHDHEHAGRAAGLGIPAALLDSVPATSVNL
RTEQAAGRLDDPWMMIYTGGTTGRPKGVVLTFESVNWNAINTIISWNLSARDCTLNYMPLFHTGGLN
ALSLPILMAGGTVVIGRKFDPEEAIRALNDYRTTISLFVPTMHQAMLDTDLFWESDFPTVDVFLSGGAP
CPQTVYDAYRKKGVRFREGYGMTEAGPNNFIIDPDTAMRKRGAVGKSMQFNEVRILDAKGRPCRAG
EVGELHLRGRHLFSHYWNNEEATQEALKEGWFSTGDLASRDEDGDYFIVGRKKEMIISGGENIYPQE
VEQCLIGHDGVREIAVIGIADRKWGERVVAFIVAQPGNIPKTEELLKHCAQTLGSYKVPKDFFFVQELPI
TDIGKIDKKQLAIMAEELKKEEMQHPGQSG

SEQ ID NO: 77 – AAE candidate isolated from Deltaproteobacteria bacterium ADurb.Bin022

Amino acid sequence

MHKFTLDKPDNLVDWWGESVTRFADRPLFGTKNKEGVYKWATYKEIGNRIDNLRAGLTQLGIGKDD
VVGIIANNRPEWAVIGFATWGCLARYVPMYEAELVQVWKYIINDSGAKVLFVSNPAIYEKIKDFPKDIPT
LKHIFIIESDGDNSMASLEKKGAAKPVAPKSPKAEDVAELIYTSGTTGNPKGVLLMHMNFTSNSHAGL
KMYPELYENEVVSLTILPWAHVFGQTAELFAIIRLGGRMGLIESTKTIINDIVQIKPTFIIAVPTVFNRIYDG
LWNKMNKDGGLARALFVMGVEAAKKKRILAEKGQSDLMTNFKVAVADKIVFKKIRERMGGRMLGSM
TGSAAMNVEISKFFFDIGIPIYDCYGLTETSPGITMNGSQAYRIGSVGRPIDKVKVVIDSSVVEEGATDG
EIIAYGPNVMKGYHNRPEDTKAALTPDGGFRTGDRGRLDKDGYLFITGRIKEQYKLENGKFCFPVSLE
ENICLASFVQQAVVYGLNRPYNVCIVVPDFDVLLDYAKEKGLPTDIKTLVEREDIIHMISEAVTGQLKGK
FGGYEIPKKFIILPEAFSLDNGMLTQTMKLKRKVILDKLNDRIEALYKEDK

SEQ ID NO: 78 – AAE candidate isolated from Alcaligenes xylosoxydans (Achromobacter xylosoxidans)

Amino acid sequence

MYSRIHEPHACTLTDALREWAASRPAAPWLEDSQGIAFTVGQAFTSSQRFASFLHHQLGVQPEERV
GVFMSNSCAMVATTFGIGYLRATAVMLNTELRSSFLRHQLNDCQLATIVVDSALVEHVASLADELPHL
RTLVVVGDAPAAVPERWRQVAWMDSSACAPWEGPAPRPEDIFCIMYTSGTTGPSKGVLMPHCHCA
LLGLGAIRSLEITEADKYYICLPLFHANGLFMQLGATVLAGIPAFLKQRFSASTWLADIRRSGATLTNHL
GTTAMFVINQPPTEQDRDHRLRASLSAPNPAQHEAVFRERFGVKDVLSGFGMTEVGIPIWGRIGHAA
PNAAGWAHEDRFEICIADPETDVPVLAGQVGEILVRPKVPFGFMAGYLNVPAKTVEAWRNLWFHTG
DAGTRDEQGLITFVDRIKDCIRRRGENISATEVEVVVGQLPGVHEVAAYAVPAQGAGGEDEVMLALV
PSEGAALDMADIVRQASAQLPRFAKPRYLRQMDSLPKTATGKIQRAVLRQQGSAGAYDAEAAPAR

SEQ ID NO: 79 – AAE candidate isolated from Novosphingobium sp. MD-1

Amino acid sequence

MQFTQGLERAVQHHPDVTATICRARSQTFAELYERVTGLAGCLASRSLAKGARIAVLALNSDHYLEVY
LATAWAGGVIVPVNFRWSPAEIAYSLNDAGCVALMVDQHHAALVPTLREQCPGLQHIFLMGGTEESD
DLPGLDALIAAAEPLQNAGAGGDDLLGIFYTGGTTGRPKGVMLSHANLCSSGLSMLAEGVFNEGAVG
LHVAPMFHLADMLLTTCLVLRGCTHVMLPAFSPDAVLDHVARFGVTDTLVVPAMLQAIVDHPAIGNFD
TSSLCNILYGASPASETLLRRTMAAFPDVRLTQGYGMTESAAFICALPWHQHVVDNDGPNRLRAAGR
STFDVHLQIVDPDDRELPRGEIGEIIVKGPNVMQGYYNMPEATAETLRGGWLHTGDMAWMDEEGYV
FIVDRAKDMIISGGENIYSAEVENAVASHPAVAANAVIGIPHEQMGEAVHVALVLRPGSELSLEALQAH
CRALIAGYKVPRSMEVRPSLPLSGAGKILKTELREPFWKGRDRAVG

SEQ ID NO: 80 – AAE candidate isolated from Arabidopsis thaliana (Mouse-ear cress)

Amino acid sequence

MEDSGVNPMDSPSKGSDFGVYGIIGGGIVALLVPVLLSVVLNGTKKGKKRGVPIKVGGEEGYTMRHA
RAPELVDVPWEGAATMPALFEQSCKKYSKDRLLGTREFIDKEFITASDGRKFEKLHLGEYKWQSYGE
VFERVCNFASGLVNVGHNVDDRVAIFSDTRAEWFIAFQGCFRQSITVVTIYASLGEEALIYSLNETRVS
TLICDSKQLKKLSAIQSSLKTVKNIIYIEEDGVDVASSDVNSMGDITVSSISEVEKLGQKNAVQPILPSKN
GVAVIMFTSGSTGLPKGVMITHGNLVATAAGVMKVVPKLDKNDTYIAYLPLAHVFELEAEIVVFTSGSA
IGYGSAMTLTDTSNKVKKGTKGDVSALKPTIMTAVPAILDRVREGVLKKVEEKGGMAKTLFDFAYKRR
LAAVDGSWFGAWGLEKMLWDALVFKKIRAVLGGHIRFMLVGGAPLSPDSQRFINICMGSPIGQGYGL
TETCAGATFSEWDDPAVGRVGPPLPCGYVKLVSWEEGGYRISDKPMPRGEIVVGGNSVTAGYFNN
QEKTDEVYKVDEKGTRWFYTGDIGRFHPDGCLEVIDRKKDIVKLQHGEYVSLGKVEAALGSSNYVDN
IMVHADPINSYCVALVVPSRGALEKWAEEAGVKHSEFAELCEKGEAVKEVQQSLTKAGKAAKLEKFE
LPAKIKLLSEPWTPESGLVTAALKIKREQIKSKFKDELSKLYA

SEQ ID NO: 81 – AAE candidate isolated from Bradyrhizobium sp. CI-41S

Amino acid sequence

MDWSQHAIPPMRLEPRFGDRVVPAFVDRPASLWAMIADAVAQNGGGEALVCGDIRISWHEVARRAA
KVAAGFAKLGLNSGDRVAILLGNRIEFVLTMFAAAHAGLVTVLLSTRQQKPEIAYVLNDCGARALVHEA
TLAERIPDAADIPGLAHRIAVSDDAASQFAVLLDHPPAPAPAAVSEEDTAMILYTSGTTGRPKGAMLAH
CNIIHSSMVFASTLRLTQADRSIAAVPLAHVTGAVANITTMVRCAGTLIIMPEFKAAEYLKVAARERVSY
TVMVPAMYNLCLLQPDFDSYDLSSWRIGGFGGAPMPVATIERLDAKIPGLKLANCYGATETTSPSTLM
PGELTAAHIDSVGLPCPGAEIIVMGPDGRELPRGEIGELWIRSASVIKGYWNNPKATAESFTDGFWHS
GDLGSVDAENFVRVFDRQKDMINRGGLKIYSAEVESVLAGHPAVIESAIIAKPCPVLGERVHAVIVTRT
EVDAESLRAWCAERLSDYKVPETMTLTTTPLPRNANGKVVKRQLRETLAAGQAPA

SEQ ID NO: 82 – AAE candidate isolated from Bradyrhizobium sp. CI-41S

Amino acid sequence

MAGPAVLTVADTIARSFLLAVQTRGDRPAIREKKFGIWQPTSWREWLQISKDIAHGLHASGFRPGDVA
SIIANAVPEWVYADMGILCAGGVSSGIYPTDSTAQVEYLVNDSRTKIVFVEDEEQLDKVLACRARCPTL
EKIVVFDMEGLSGFSDPMVLSFAEFAALGRNHAHGNAALWDEMTGSRTASDLAILVYTSGTTGPPKG
AMHSNRSVTHQMRHANDLFPSTDSEERLVFLPLCHVAERVGGYYISIALGSVMNFAESPETVPDNLR
EVQPTAFLAVPRVWEKFYSGITIALKDATPFQNWMYGRALAIGNRMTECRLEGETPPLSLRLANRAAY
WLVFRNIRRMLGLDRCRIALTGAAPISPDLIRWYLALGLDMREVYGQTENCGVATIMPTERIKLGSVG
KAAPWGEVMICPKGEILIKGDFLFMGYLNQPERTAETIDAKGWLHTGDVGTIDNEGYVRITDRMKDIIIT
SGGKNVTPSEIENQLKFSPYVSDAVVIGDKRPYLTCLIMIDQENVEKFAQDHDIPFTNYASLCRAREIQ
DLIQREVEAVNTKFARVETIKKFYLIERQLTPEDEELTPTMKLKRSFVNKRYAAEIDAMYGARAVA

SEQ ID NO: 83 – AAE candidate isolated from Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579)

Amino acid sequence

MEGERMNAFPSTMMDEELNLWDFLERAAALFGRKEVVSRLHTGEVHRTTYAEVYQRARRLMGGLR
ALGVGVGDRVATLGFNHFRHLEAYFAVPGMGAVLHTANPRLSPKEIAYILNHAEDKVLLFDPNLLPLV
EAIRGELKTVQHFVVMDEKAPEGYLAYEEALGEEADPVRVPERAACGMAYTTGTTGLPKGVVYSHR
ALVLHSLAASLVDGTALSEKDVVLPVVPMFHVNAWCLPYAATLVGAKQVLPGPRLDPASLVELFDGE
GVTFTAGVPTVWLALADYLESTGHRLKTLRRLVVGGSAAPRSLIARFERMGVEVRQGYGLTETSPVV
VQNFVKSHLESLSEEEKLTLKAKTGLPIPLVRLRVADEEGRPVPKDGKALGEVQLKGPWITGGYYGN
EEATRSALTPDGFFRTGDIAVWDEEGYVEIKDRLKDLIKSGGEWISSVDLENALMGHPKVKEAAVVAI
PHPKWQERPLAVVVPRGEKPTPEELNEHLLKAGFAKWQLPDAYVFAEEIPRTSAGKFLKRALREQYK
NYYGGA

SEQ ID NO: 84 – AAE candidate isolated from Microbacterium oxydans

Amino acid sequence

MVRSTYPDVEIPEVSIHDFLFGDLSEAELDTVALVDGMSGATTTYRQLVGQIDLFAGALAARGVGVGT
TVGVLCPNVPAFATVFHGILRAGATATTINSLYTADEIANQLTDAGATWLVTVSPLLPGAQAAAEKLGF
DADHVIVLDGAEGHPSLPALLGEGRQAPDVSFDPSTHLAVLPYSSGTTGRPKGVMLTHRNLVANVSQ
CQPVLGVDASDRVLAVLPFFHIYGMTVLLNFALRQRAGLATMPRFDLPEFLRIIAEHRTSWVFVAPPIA
VALAKHPIVDQYDLSAVKVIFSGAAPLDGTLASAVANRLGCIVTQGYGMTETSPAVNLISEARTEIDRS
TIGPLVPNTEARLVDPDSGEDVVVPAEGASEPGELWVRGPQVMVGYLNRPDATAEMLDADGWLHT
GDVATVTHDGIYRIVDRLKELIKYKGYQVAPAVLEAVLLEHPAIADAAVIGAFDDDGQEVPKAFVVRQP
DADLDADAVMAHVTSHVAPHEKVRQVEFIDVIPKSSSGKILRKDLRAR

SEQ ID NO: 85 – AAE candidate isolated from Arabidopsis thaliana (Mouse-ear cress)

Amino acid sequence

MSLAADNVLLVEEGRPATAEHPSAGPVYRCKYAKDGLLDLPTDIDSPWQFFSEAVKKYPNEQMLGQ
RVTTDSKVGPYTWITYKEAHDAAIRIGSAIRSRGVDPGHCCGIYGANCPEWIIAMEACMSQGITYVPLY
DSLGVNAVEFIINHAEVSLVFVQEKTVSSILSCQKGCSSNLKTIVSFGEVSSTQKEEAKNQCVSLFSWN
EFSLMGNLDEANLPRKRKTDICTIMYTSGTTGEPKGVILNNAAISVQVLSIDKMLEVTDRSCDTSDVFF
SYLPLAHCYDQVMEIYFLSRGSSVGYWRGDIRYLMDDVQALKPTVFCGVPRVYDKLYAGIMQKISAS
GLIRKKLFDFAYNYKLGNMRKGFSQEEASPRLDRLMFDKIKEALGGRAHMLLSGAAPLPRHVEEFLRII
PASNLSQGYGLTESCGGSFTTLAGVFSMVGTVGVPMPTVEARLVSVPEMGYDAFSADVPRGEICLR
GNSMFSGYHKRQDLTDQVLIDGWFHTGDIGEWQEDGSMKIIDRKKNIFKLSQGEYVAVENLENTYSR
CPLIAQIWVYGNSFESFLVGVVVPDRKAIEDWAKLNYQSPNDFESLCQNLKAQKYFLDELNSTAKQY
QLKGFEMLKAIHLEPNPFDIERDLITPTFKLKRPQLLQHYKGIVDQLYSEAKRSMA

SEQ ID NO: 86 – AAE candidate isolated from Brevibacterium yomogidense

Amino acid sequence

MSWFDERPWLRTLGLTETEAVPLEPSTPLRDLADTVAAHPTTAAWTHYGQSATYAEFDRQTTAFAA
YLAESGIRPGDAVAVYAQNSPHFPIATYGIWKAGAVVVPLNPMYRDELTHAFADADVKAIVVQKALYL
MRVKEYAADLPLVVLAGDLDWAQDGPDAVFGAYADLPDVPLPDLRTVVDERLDTDFEPLTVRPEDP

ALIGYTSGTSGKAKGALHPHSSISSNSRMAARNAGLPQGAGVVSLAPLFHITGFICQMIASTANGSTLV
LNHRFDPASFLDLLRQEKPAFMAGPATVYTAMMASPSFGADAFDSFHSIMSGGAPLPEGLVKRFEEK
TGHYIGQGYGLTETAAQAVTVPHSLRAPVDPESGNLSTGLPQRDAMVRILDDDGNPVGPREVGEVAI
SGPMVATEYLGNPQATADSLPGGELRTGDVGFMDPDGWVFIVDRKKDMINASGFKVWPREVEDILY
MHPAVREGAVVGVPDEYRGETVVAFVSLQPDSQATAEDIIAHCKEHLASYKAPVEVTIVDELPKTSSG
KILRRTVRDEATQARQAQPDAH

SEQ ID NO: 87 – AAE candidate isolated from Nocardioides simplex (Arthrobacter simplex)

Amino acid sequence

MSFRYYRDLHPTFADRTEWALPTVLRHHAAERPDAVWLDCPEEGRTWTFAETLTAAERVGRSLLAA
GAEPGDRVVLVAQNSSAFVRTWLGTAVAGLVEVPVNTAYEHDFLAHQVSTVEATLAVVDDVYAARF
VAIAEAAKSIRKFWVIDTGSRDQALATLRDAGWEAAPFEELDEAATAPEVVDATLALPDVRPQDLASV
LFTSGTTGPSKGVAMPHAQMYFFADECVSLVRLTPDDAWMSVTPLFHGNAQFMAAYPTLVAGARFV
TRSRFSASRWVDQLRESRVTVTNFIGVMMDFIWKQDRRDDDADNPLRVVFAAPTAATLVGPMSERY
GIEAFVEVFGLTETSAPIISPYGVDRPAGAAGLAADEWFDVRLVDPETDEEVGVGEIGELVVRPKVPFI
CSMGYFNMPDKTVEAWRNLWFHTGDALRRDEDGWFYFVDRFKDALRRRGENISSYEIETSILAHPA
VVECAVIAVPASSEAGEDEVMAYVITGGDAPVPTPAELWAHCDGRIPSFAVPRYLRFVDEMPKTPSQ
RVQKAKLRALGVTPDTHDREA

SEQ ID NO: 88 – AAE candidate isolated from Brevibacterium linens

Amino acid sequence

MTVTEEFRAARDKLIELRSDYDAAREQFEWPRFDHFNFALDWFDKIAENNDKPALWIVEQDGSEGK
WSFAELSARSNQVANHFRRAGIKRGDHVMVMLNNQVELWETMLAGIKLGAVLMPATTQLGPIDLTD
RAERGHAEFVVAGAEDAAKFDDVDVEVVRIVVGGEPTRQQDYSYSDADDESTEFDPQGSSRADDL
MLLYFTSGTTSKAKMVAHTHVSYPVGHLSTMYWMGLTPGDVHLNVASPGWAKHAWSNIFTPWIAEA
CVFLYNYSRFDANALMETMDRVGVTSFCAPPTVWRMLIQADLKHLKTPPTKALGAGEPLNPEIIDRVH
SDWGVLIRDGFGQTESTLQIGNSPDQELKYGSMGKALPGFDVVLIDPATGEEGDEGEICLRLDPRPIG
LTTGYWSNPEKTAEAFEGGVYHTGDVASRDEDGFITYVGRADDVFKASDYRLSPFELESVLIEHEAV
AEAAVVPSPDPVRLAVPKAYVVVSSKFDADAETARSILAYCREHLAPYKRIRRLEFAELPKTISGKIRR
VELRAREDQLHPFSGEPVVEGNEYADTDFDLKS

SEQ ID NO: 89 – AAE candidate isolated from Pseudomonas putida (Arthrobacter siderocapsulatus)

Amino acid sequence

MNLGKIITRSARYWPDHTAVADSQTRLTYAQLERRSNRLASGLGALGVATGEHVAILAANRVELVEAE
VALYKAAMVKVPINARLSLDEVVRVLEDSCSVALITDATFAQALAERRAALPMLRQVIALEGEGGDLG
YAALLERGSEAPCSLDPADDALAVLHYTSGSSGVLKAAMLSFGNRKALVRKSIASPTRRSGPDDVMA
HVGPITHASGMQIMPLLAVGACNLLLDRYDDRLLLEAIERERVTRLFLVPAMINRLVNYPDVERFDLSS
LKLVMYGAAPMAPALVKKAIELFGPILVQGYGAGETCSLVTVLTEQDHLIEDGNYQRLASCGRCYFET
DLRVVNEAFEDVAPGEIGEIVVKGPDIMQGYWRAPALTAEVMRDGYYLTGDLATVDAQGYVFIVDRK
KEMIISGGFNVYPSEVEQVIYGFPEVFEAAVVGVPDEQWGEAVRAVVVLKPGAQLDAAELIERCGRAL
AGFKKPRGVDFVTELPKNPNGKVVRRLVREAYWQHSDRRI

SEQ ID NO: 90 – AAE candidate isolated from Drosophila melanogaster (Fruit fly)

Amino acid sequence

MNDLKPATSYRSTSLHDAVKLRLDEPSSFSQTVPPQTIPEFFKESCEKYSDLPALVWETPGSGNDGW
TTLTFGEYQERVEQAALMLLSVGVEERSSVGILAFNCPEWFFAEFGALRAGAVVAGVYPSNSAEAVH
HVLATGESSVCVVDDAQQMAKLRAIKERLPRLKAVIQLHGPFEAFVDHEPGYFSWQKLQEQTFSSEL
KEELLARESRIRANECAMLIFTSGTVGMPKAVMLSHDNLVFDTKSAAAHMQDIQVGKESFVSYLPLSH
VAAQIFDVFLGLSHAGCVTFADKDALKGTLIKTFRKARPTKMFGVPRVFEKLQERLVAAEAKARPYSR
LLLARARAAVAEHQTTLMAGKSPSIYGNAKYWLACRVVKPIREMIGVDNCRVFFTGGAPTSEELKQFF
LGLDIALGECYGMSETSGAITLNVDISNLYSAGQACEGVTLKIHEPDCNGQGEILMRGRLVFMGYLGL

PDKTEETVKEDGWLHSGDLGYIDPKGNLIISGRLKELIITAGGENIPPVHIEELIKKELPCVSNVLLIGDH
RKYLTVLLSLKTKCDAKTGIPLDALREETIEWLRDLDIHETRLSELLNIPADLQLPNDTAALAATLEITAK
PKLLEAIEEGIKRANKYAISNAQKVQKFALIAHEFSVATGELGPTLKIRRNIVHAKYAKVIERLYK

SEQ ID NO: 91 – AAE candidate isolated from Cannabis sativa

Amino acid sequence

MGKNYKSLDSVVASDFIALGITSEVAETLHGRLAEIVCNYGAATPQTWINIANHILSPDLPFSLHQMLFY
GCYKDFGPAPPAWIPDPEKVKSTNLGALLEKRGKEFLGVKYKDPISSFSHFQEFSVRNPEVYWRTVL
MDEMKISFSKDPECILRRDDINNPGGSEWLPGGYLNSAKNCLNVNSNKKLNDTMIVWRDEGNDDLPL
NKLTLDQLRKRVWLVGYALEEMGLEKGCAIAIDMPMHVDAVVIYLAIVLAGYVVVSIADSFSAPEISTRL
RLSKAKAIFTQDHIIRGKKRIPLYSRVVEAKSPMAIVIPCSGSNIGAELRDGDISWDYFLERAKEFKNCE
FTAREQPVDAYTNILFSSGTTGEPKAIPWTQATPLKAAADGWSHLDIRKGDVIVWPTNLGWMMGPW
LVYASLLNGASIALYNGSPLVSGFAKFVQDAKVTMLGVVPSIVRSWKSTNCVSGYDWSTIRCFSSSG
EASNVDEYLWLMGRANYKPVIEMCGGTEIGGAFSAGSFLQAQSLSSFSSQCMGCTLYILDKNGYPM
PKNKPGIGELALGPVMFGASKTLLNGNHHDVYFKGMPTLNGEVLRRHGDIFELTSNGYYHAHGRAD
DTMNIGGIKISSIEIERVCNEVDDRVFETTAIGVPPLGGGPEQLVIFFVLKDSNDTTIDLNQLRLSFNLGL
QKKLNPLFKVTRVVPLSSLPRTATNKIMRRVLRQQFSHFE

SEQ ID NO: 92 – TKS candidate isolated from Dendrobium catenatum

Amino acid sequence

MPSLESIRKAPRANGFASILAIGRANPENFIEQSTYPDFFFRITNSEHLVDLKKKFQRICDKTAIRKRHF
VWNEEFITTNPCLHTFMDKSLDVRQEVAIREIPKLGAKAAAKAIQEWGQPKSRITHLIFCTTSGMDLPG
ADYQLTQILGLNPNVERVMLYQQGCFAGGTTLRLAKCLAESRKGARVLVVCAETTTVLFRGPSEEHQ
EDLVTQALFADGASALIVGADPDEAAHERASFVIVSTSQVLLPDSAGAIGGHVSEGGLLATLHRDVPKI
VSKNVEKCLEEAFTPFGITDWNSIFWVPHPGGRAILDLVEERVGLKPEKLLVSRHVLAEYGNMSSVCV
HFALDEMRKRSAIEGKATTGEGLEWGVVFGFGPGLTVETVVLRSVPL

SEQ ID NO: 93 – TKS candidate isolated from Dictyostelium

Amino acid sequence

MNNSNVKSSPSIVKEEIVTLDKDQQPLLLKEHQHIIISPDIRINKPKRESLIRTPILNKFNQITESIITPSTPS
LSQSDVLKTPPIKSLNNTKNSSLINTPPIQSVQQHQKQQQKVQVIQQQQQPLSRLSYKSNNNSFVLGI
GISVPGEPISQQSLKDSISNDFSDKAETNEKVKRIFEQSQIKTRHLVRDYTKPENSIKFRHLETITDVNN
QFKKVVPDLAQQACLRALKDWGGDKGDITHIVSVTSTGIIIPDVNFKLIDLLGLNKDVERVSLNLMGCL
AGLSSLRTAASLAKASPRNRILVVCTEVCSLHFSNTDGGDQMVASSIFADGSAAYIIGCNPRIEETPLY
EVMCSINRSFPNTENAMVWDLEKEGWNLGLDASIPIVIGSGIEAFVDTLLDKAKLQTSTAISAKDCEFLI
HTGGKSILMNIENSLGIDPKQTKNTWDVYHAYGNMSSASVIFVMDHARKSKSLPTYSISLAFGPGLAF
EGCFLKNVV

SEQ ID NO: 94 – TKS candidate isolated from Arachis hypogaea

Amino acid sequence

MNNSNVKSSPSIVKEEIVTLDKDQQPLLLKEHQHIIISPDIRINKPKRESLIRTPILNKFNQITESIITPSTPS
LSQSDVLKTPPIKSLNNTKNSSLINTPPIQSVQQHQKQQQKVQVIQQQQQPLSRLSYKSNNNSFVLGI
GISVPGEPISQQSLKDSISNDFSDKAETNEKVKRIFEQSQIKTRHLVRDYTKPENSIKFRHLETITDVNN
QFKKVVPDLAQQACLRALKDWGGDKGDITHIVSVTSTGIIIPDVNFKLIDLLGLNKDVERVSLNLMGCL
AGLSSLRTAASLAKASPRNRILVVCTEVCSLHFSNTDGGDQMVASSIFADGSAAYIIGCNPRIEETPLY
EVMCSINRSFPNTENAMVWDLEKEGWNLGLDASIPIVIGSGIEAFVDTLLDKAKLQTSTAISAKDCEFLI
HTGGKSILMNIENSLGIDPKQTKNTWDVYHAYGNMSSASVIFVMDHARKSKSLPTYSISLAFGPGLAF
EGCFLKNVV

SEQ ID NO: 95 – TKS candidate isolated from Spinacia oleracea

Amino acid sequence

MASVDISEIHNVERAKGQANVLAIGTANPPNVMYQADYPDFYFRLTNSEHMTDLKAKFKRICEKTTIKK
RYMHISEDILKEKPDLCDYNASSLDIRQVILAKEVPKVGKDAAMKAIEEWGQAMSKITHLIFCTTSGVDI
PGADYQLTMLLGLNPSVKRYMLCQQGCHAGGTVLRLAKDLAENNYGSRVLVVCSENTTVCFRGPTE
THPDSMVAQALFADGAGAVIVGAYPDESLNERPIFQIVSTAQTILPNSQGAIEGHLRQIGLAIQLLPNVP
DLISNNIDKCLVEAFNPIGINDWNSIFWIAHPGGPAILGQVESKLGLQESKLTTTWHVLREFGNMSSAC
VFFIMDETRKRSLKEGKTTTGDGFDWGVLFGFGPGLTVETVVLRSFPLNQ

SEQ ID NO: 96 – TKS candidate isolated from Elaeis guineensis

Amino acid sequence

MSGLSRDMNPSLERSVGRAAVLGIGTANPPHVVEQSTFPDYYFKITNSEHMAHLKEKFTRICEKSKIA
KRYTVLTDEFLVANPTLTSFNAPSLDTRQQLLDVEVPRLGAEAATRAIKDWGRPMSDLTHLIFCNSYG
ASIPGADYELVKLLGLPLSTRRVMLYQQCCYAGGTVIRLAKDLAENNRDARVLVVCCELNTVGIRGPC
QSHLEDLVSQALFGDGAGALIIGADPRAGVERSIFEIVRTSQNIIAGSEGALVAKLREVGLVGRLKPEIP
MHLSCSIEKLASEALNPVGIADWNEAFWVMHPGGRAILDELEKKLGLGEEKLAATREVLRDYGNMSS
TSVLFVMEVMRRRSEERGLATAGEGLEWGVLLGFGPGLTMETVVLRCP

SEQ ID NO: 97 – TKS candidate isolated from Vitis pseudoreticulata

Amino acid sequence

MALVEEIRNAQRAKGPATVLAIGTATPDNCLYQSDFADYYFRVTKSEHMTELKKKFNRICDKSMIKKR
YIHLTEEMLEEHPNIGAYMAPSLNIRQEIITAEVPKLGKEAALKALKEWGQPKSKITHLVFCTTSGVEMP
GADYKLANLLGLEPSVRRVMLYHQGCYAGGTVLRTAKDLAENNAGARVLVVCSEITVVTFRGPSENA
LDSLVGQALFGDGSAAVIVGSDPDISIERPLFQLVSAAQTFIPNSAGAIAGNLREVGLTFQLWPNVPTLI
SENIEKCLTKAFDPIGISDWNSLFWIAHPGGPAILDAVEAKLNLDKQKLKATRHILSEYGNMSSACVLFI
LDEMRKKSLKEGKTTTGEGLDWGVLFGFGPGLTIETVVLHSVQMDSN

SEQ ID NO: 98 – TKS candidate isolated from Cannabis sativa

Amino acid sequence

MASISVDQIRKAQRANGPATVLAIGTANPPTSFYQADYPDFYFRVTKNQHMTELKDKFKRICEKTTIKK
RHLYLTEDRLNQHPNLLEYMAPSLNTRQDMLVVEIPKLGKEAAMKAIKEWGQPKSRITHLIFCSTNGV
DMPGADYECAKLLGLSSSVKRVMLYQQGCHAGGSVLRIAKDLAENNKGARILTINSEITIGIFHSPDET
YFDGMVGQALFGDGASATIVGADPDKEIGERPVFEMVSAAQEFIPNSDGAVDGHLTEAGLVYHIHKD
VPGLISKNIEKSLVEALNPIGISDWNSLFWIVHPGGPAILNAVEAKLHLKKEKMADTRHVLSEYGNMSS
VSIFFIMDKLRKRSLEEGKSTTGDGFEWGVLFGFGPGLTVETIVLHSLAN

SEQ ID NO: 99 – TKS candidate isolated from Chenopodium quinoa

Amino acid sequence

MASVQEIRNAQRADGPATILAIGTANPPNEMYQAEYPDFYFRVTESEHMTDLKKKFKRMCERSMIKK
RYMHVTEELLKENPHMCDYNASSLNTRQDILATEVPKLGKEAAIKAIKEWGQPRSKITHVIFCTTSGVD
MPGADYQLTKLLGLRPSVKRFMLYQQGCYAGGTVLRLAKDIAENNRGARVLVVCAEITVICFRGPTET
HLDSMIGQALFGDGAGAVIVGADVDESIERPIFQLVWAAQTILPDSEGAIDGHLREVGLAFHLLKDVPG
LISKNIEKALVEAFKPIGIDDWNSIFWVAHPGGPAILDQVESKLELKQDKLRDTRHVLSEFGNMSSACV
LFILDEMRNRSLKEGKTTTGEGLDWGVLFGFGPGLTVETVMLHSVPITN

SEQ ID NO: 100 – TKS candidate isolated from Ziziphus jujuba

Amino acid sequence

MVTVDEIREAQRAKGPATIMAIGTATPPNAIDQSTFTDYYFRITNSDHKTDLKKKFKTICDKSMIKKRYL
YLTEEHLKQNPNMSEYMAPSLDVRQEIVIAEVPKLGKEAANKAIKEWGQPKSKITHLVFSTISGVDAPG
ADYQLTKLLGLNPSVKRIMVYQQGCFAGGTSLRLAKDLAENNKGARVLVVCTEISAINFRGPSETYFD

SNVGQILFGDGASAVVVGSDPLVGVEKPLFELVSASQTIIPDSEGNIEGHICEVGLTIRLSKKVPSLISN
NIEKSLVEAFNPLGISDWNSIFWIAHPGGPAILDQIELKLGLKPEKLRASRHVLSEYGNMSSATVLFILD
EMRKKSIEDGLKTPGEGLEWGVLFGFGPGLTVETVVLHSVTA

SEQ ID NO: 101 – TKS candidate isolated from Marchantia polymorpha

Amino acid sequence

MSRSRLIAQAVGPATVLAMGKAVPANVFEQATYPDFFFNITNSNDKPALKAKFQRICDKSGIKKRHFY
LDQKILESNPAMCTYMETSLNCRQEIAVAQVPKLAKEASMNAIKEWGRPKSEITHIVMATTSGVNMPG
AELATAKLLGLRPNVRRVMMYQQGCFAGATVLRVAKDLAENNAGARVLAICSEVTAVTFRAPSETHI
DGLVGSALFGDGAAAVIVGSDPRPGIERPIYEMHWAGEMVLPESDGAIDGHLTEAGLVFHLLKDVPG
LITKNIGGFLKDTKNLVGASSWNELFWAVHPGGPAILDQVEAKLELEKG

SEQ ID NO: 102 – TKS candidate isolated from Caragana korshinskii

Amino acid sequence

MAYLEEIREVQRARGPATILAIGTANPSNCIYQADFTDYYFRVTNSDHMTKLKAKLKRICENSMIKKRH
VHLTEEILKENPNICTYKESSLDARQDMLIVEVPKLGEKAASKAIEEWGRPKSEITHLIFCSTSGVDMP
GADYQLINLLGLKPSTKRFMLYHQGCFAGGTVLRLAKDLAENNAGARVLVVCSEITVVTFRGPSETHL
DCLVGQALFGDGASSVIVGSDPDTSIERPLFHLVSASETILPNSEGAIEGHLREAGLMFQLKENVPQLI
GENIEKSLEEMFHPLGISDWNSLFWISHPGGPAILKRIEETAGLNPEKLKATKHVLSEYGNMSSACVLF
ILDEMRKRSMEEGKSTTGEGLNWGVLFGFGPGLTMETIALHSANIDTGY

SEQ ID NO: 103 – TKS candidate isolated from Glycine max

Amino acid sequence

MVSVAEIRQAQRAEGPATILAIGTANPPNCVAQSTYPDYYFRITNSEHMTELKEKFQRMCDKSMIKRR
YMYLNEEILKENPNMCAYMAPSLDARQDMVVVEVPKLGKEAAVKAIKEWGQPKSKITHLIFCTTSGVD
MPGADYQLTKQLGLRPYVKRYMMYQQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGPS
DTHLDSLVGQALFGDGAAAVIVGSDPIPQVEKPLYELVWTAQTIAPDSEGAIDGHLREVGLTFHLLKDV
PGIVSKNIDKALFEAFNPLNISDYNSIFWIAHPGGPAILDQVEQKLGLKPEKMKATRDVLSEYGNMSSA
CVLFILDEMRRKSAENGLKTTGEGLEWGVLFGFGPGLTIETVVLRSVAI

SEQ ID NO: 104 – TKS candidate isolated from Humulus lupulus

Amino acid sequence

MASVTVEQIRKAQRAEGPATILAIGTAVPANCFNQADFPDYYFRVTKSEHMTDLKKKFQRMCEKSTIK
KRYLHLTEEHLKQNPHLCEYNAPSLNTRQDMLVVEVPKLGKEAAINAIKEWGQPKSKITHLIFCTGSSI
DMPGADYQCAKLLGLRPSVKRVMLYQLGCYAGGKVLRIAKDIAENNKGARVLIVCSEITACIFRGPSE
KHLDCLVGQSLFGDGASSVIVGADPDESVGERPIFELVSAAQTILPNSDGAIAGHVTEAGLTFHLLRDV
PGLISQNIEKSLIEAFTPIGINDWNNIFWIAHPGGPAILDEIEAKLELKKEKMKASREMLSEYGNMSCAS
VFFIVDEMRKQSSKEGKSTTGDGLEWGALFGFGPGLTVETLVLHSVPTNV

SEQ ID NO: 105 – TKS candidate isolated from Humulus lupulus

Amino acid sequence

MVTVEEVRKAQRAEGPATILAIGTATPANCILQSEYPDYYFRITNSEHKTELKEKFKRMCDKSMIRKRY
MHLTEEILKENPNLCAYEAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQPKSKITHVVFCTTSGVD
MPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRVAKDLAENNKGARVLVVCSEITAVTFRGPN
DTHLDSLVGQALFGDGSAALIIGADPTPEIEKPIFELVSAAQTILPDSDGAIDGHLREVGLTFHLLKDVP
GLISKNIEKSLVEAFKPLGISDWNSLFWIAHPGGPAILDQVESKLALKPEKLRATRHVLGEYGNMSSAC
VLFILDEMRRKCAEDGLKTTGEGLEWGVLFGFGPGLTVETVVLHSVGI

SEQ ID NO: 106 – TKS candidate isolated from Trema orientale

Amino acid sequence

MASVTVDEIRKAQRAEGPATVLAIGTATPHNCVSQADYPDYYFRITNSEHMTELKEKFKRMCEKSMIK
KRYMHLTEEILKENPKMCEYMAPSLDARQDMVVVEVPKLGKEAATKAIKEWGLPKSKITHLVFCTTSG
VDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAENNRGARVLVVCSEITAVTFRG
PSDTHLDSMVGQALFGDGAAAVIVGADPDPSAGERPLFEMVSAAQTILPDSEGAIDGHLREAGLTFH
LLKDVPGLISKNIEKSLTEAFSPLGISDWNSLFWIAHPGGPAILDQVEAKLKLKEEKLRATRHVLSEYGN
MSSACVLFILDEMRKKSAEDGKPTTGEGLDWGVLFGFGPGLTVETVVLHSVAATATN

SEQ ID NO: 107 – TKS candidate isolated from Plumbago indica

Amino acid sequence

MAPAVQSQSHGGAYRSNGERSKGPATVLAIATAVPPNVYYQDEYADFFFRVTNSEHKTAIKEKFNRV
CGTSMIKKRHMYFTEKMLNQNKNMCTWDDKSLNARQDMVIPAVPELGKEAALKAIEEWGKPLSNITH
LIFCTTAGNDAPGADFRLTQLLGLNPSVNRYMIYQQGCFAGATALRIAKDLAENNKGARVLIVCCEIFA
FAFRGPHEDHMDSLICQLLFGDGAAAVIVGGDPDETENALFELEWANSTIIPQSEEAITLRMREEGLMI
GLSKEIPRLLGEQIEDILVEAFTPLGITDWSSLFWIAHPGGKAILEALEKKIGVEGKLWASWHVLKEYGN
LTSACVLFAMDEMRKRSIKEGKATTGDGHEYGVLFGVGPGLTVETVVLKSVPLN

SEQ ID NO: 108 – TKS candidate isolated from Artemisia annua

Amino acid sequence

MASLTDIAAIREAQRAQGPATILAIGTANPANCVYQADYPDYYFRITKSEHMVDIKEKFKRMCDKSMIR
KRYMHLTEEYLKENPSLCEYMAPSLDARQDVVVVEVPKLGKEAATKAIKEWGQPKSKITHLIFCTTSG
VDMPGADYQLTKLLGLRPSVKRFMMYQQGCFAGGTVLRLAKDLAENNKDARVLVVCSEITAVTFRG
PNDTHLDSLVGQALFGDGAAAVIVGSDPDLTKERPLFEMISAAQTILPDSEGAIDGHLREVGLTFHLLK
DVPGLISKNIEKALTQAFSPLGISDWNSIFWIAHPGGPAILDQVELKLGLKEEKMRATRHVLSEYGNMS
SACVLFIIDEMRKKSAEEGAATTGEGLDWGVLFGFGPGLTVETVVLHSLPTTISVVN

SEQ ID NO: 109 – TKS candidate isolated from Actinidia chinensis var. chinensis

Amino acid sequence

MAPSLEEILRAQRSQGPAEILGIGTATPPNCYDQADFPDFYFRVTNSEHMTHLKDKFKQICEKSTVKK
RYMYLTEEILKDNPSLCSYMGRSLDVRQNMVMTEVPKLGKEAAAKAIKEWGQPKSKITHLVFCTTSG
VDMPGADYHLTKLLGLQPSVKRIMMYQSSCYGGGTGLRLAKDLAENNAGARVLLVCSEISAINFRGP
PDTPARLDKLVAQALFGDGAAAVIVGADPDTSIERSLFQLISASQTIVPGSNGGIMGTFGEAGLMCHLI
KDVPRLISSNIEKCLMDAFTPIGINDWNSIFWIAHPGGPAILDMVEEKIGLEEEKLRATRHILSEYGNMS
SVCVLFILDEMRKKSAEEGKLTTGEGLEWGVLFGFGAGITVETVVLRSMSISNTTH

SEQ ID NO: 110 – TKS candidate isolated from Rhododendron dauricum

Amino acid sequence

MVTVEDVRKAQRAEGPATVMAIGTATPPNCVDQSTYPDFYFRITNSEHKAELKEKFQRMCDKSMIKK
RYMYLTEEILKENPSVCEYMAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQPKSKITHLVFCTTSGV
DMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGP
SDTHLDSLVGQALFGDGAAAIIVGADPVPEVEKPLFELVSAAQTILPDSDGAIDGHLREVGLTFHLLKD
VPGLISKNIEKALTEAFQPLGISDWNSIFWIAHPGGPAILDQVELKLSLKPEKLRATRHVLSEYGNMSSA
CVLFILDEMRRKSAEEGLKTTGEGLEWGVLFGFGPGLTVETVVLHSLCT

SEQ ID NO: 111 – TKS candidate isolated from Chenopodium quinoa

Amino acid sequence

MASASMNPATILAIGTANPPNVMCQSDYPDYHFRTTNSDHLTDLKAKFKRICDKSMIRKRHFYMNEEI
LKENPHLGDNNASSIGTRQALCANEIPKLGKEAAEKAIKEWGKPKSMITHLIFGTNSDFDLPGADFRLA
KLLGLQPTVKRFILPLGACHAGGTALRIAKDIAENNRGARVLVICSESTAISFHAPSETHLVSLAIFGDG

AGAMIVGTDPDEPSERPLFQLVSAGQITLPDSEDGIQARLSEIGMTIHLSPDVPKIIAKNIQTLLSESFDH
IGISNWNSIFWVAHPGGPAILDKVEAKLELETSKLSTSRHILSEYGNMWGASVIFVMDEMSKRSLKEG
KSTTGEGCEWGVLVAFGPGITVETIVLRSMPINY

SEQ ID NO: 112 – TKS candidate isolated from Cajanus cajan

Amino acid sequence

MVSVEDIRKAQRAEGPATVMAIGTATPPNCVDQSTYPDYYFRITNSEHKTELKEKFKRMCDKSMIKKR
YMYLNEEILKENPSVCEYMAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQPKSKITHLIFCTTSGVD
MPGADYQLTKLLGLRPSVKRYMMYQQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGPS
DTHLDSLVGQALFGDGAAAVIVGSDPLPVEKPFFELVWTAQTILPDSEGAIDGHLREVGLTFHLLKDV
PGLISKNIEKALVEAFQPLGISDYNSIFWIAHPGGPAILDQVEAKLGLKPEKMEATRHVLSEYGNMSSA
CVLFILDQMRKKSIENGLGTTGEGLEWGVLFGFGPGLTVETVVLRSVTV

SEQ ID NO: 113 – TKS candidate isolated from Lonicera japonica

Amino acid sequence

MGSVTVEEIRKAQRAQGPATVLAIGTATPANCVYQADYPDFYFRITKSEHKAELKEKFKRMCEKSMIR
KRYMHLNEEILKENPGICEYMAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQPKSKITHLVFCTTSG
VDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAENNAGARVLVVCSEITAVTFRG
PSDTHLDSLVGQALFGDGAAAVIIGADPDKSVERPLFELVSAAQTILPDSDGAIDGHLREVGLTFHLLK
DVPGLISKNIEKSLKEAFAPIGITDWNSLFWIAHPGGPAILDQVEIKLGLKEEKLRPTRHVLSEYGNMSS
ACVLFILDELRKKSIEEGKATTGDGLEWGVLFGFGPGLTVETVVLHSVPASI

SEQ ID NO: 114 – TKS candidate isolated from Ruta graveolens

Amino acid sequence

MESLKEMRKAQMSEGPAAILAIGTATPNNVYMQADYPDYYFRMTKSEHMTELKDKFRTLCEKSMIRK
RHMCFSEEFLKANPEVSKHMGKSLNARQDIAVVETPRLGNEAAVKAIKEWGQPKSSITHLIFCSSAGV
DMPGADYQLTRILGLNPSVKRMMVYQQGCYAGGTVLRLAKDLAENNKGSRVLVVCSELTAPTFRGP
SPDAVDSLVGQALFADGAAALVVGADPDSSIERALYYLVSASQMLLPDSDGAIEGHIREEGLTVHLKK
DVPALFSANIDTPLVEAFKPLGISDWNSIFWIAHPGGPAILDQIEEKLGLKEDKLRASKHVMSEYGNMS
SSCVLFVLDEMRSRSLQDGKSTTGEGLDWGVLFGFGPGLTVETVVLRSVPIEA

SEQ ID NO: 115 – TKS candidate isolated from Physcomitrella patens subsp. patens

Amino acid sequence

MASAGDVTRAALPRAQPRAEGPACVLGIGTAVPPAEFLQSEYPDFFFNITNCGEKEALKAKFKRICDK
SGIRKRHMFLTEEVLKANPGICTYMEPSLNVRHDIVVVQVPKLAAEAAQKAIKEWGGRKSDITHIVFAT
TSGVNMPGADHALAKLLGLKPTVKRVMMYQTGCFGGASVLRVAKDLAENNKGARVLAVASEVTAVT
YRAPSENHLDGLVGSALFGDGAGVYVVGSDPKPEVEKPLFEVHWAGETILPESDGAIDGHLTEAGLIF
HLMKDVPGLISKNIEKFLNEARKPVGSPAWNEMFWAVHPGGPAILDQVEAKLKLTKDKMQGSRDILS
EFGNMSSASVLFVLDQIRHRSVKMGASTLGEGSEFGFFIGFGPGLTLEVLVLRAAPNSA

SEQ ID NO: 116 – TKS candidate isolated from Rubus idaeus

Amino acid sequence

MGSVAKEAKYPATILAIATANPANCYHQKDYPDFLFRVTKSEDKTELKDKFKRICEKSMVKKRYLGITE
ESLNANPNICTYKAPSLDSRQDLLVHEVPKLGKEAALKAIEEWGQPISSITHLIFCTASCVDMPGADFQ
LVKLLGLDPTIKRFMIYQQGCFAGGTVLRIAKDVAENNAGARLLIVCCEITTMFFQQPSENHLDVLVGQ
ALFSDGAAALIVGTNPDPKSERQLFDIMSVRETIIPNSEHGVVAHLREMGFEYYLSSEVPKLVGGKIEE
YLNKGFEGIGVDGDWNSLFYSIHPGGPAILNKVEEELGLKEGKLRATRHVLSEFGNMGAPSVLFILDEI
RKRSMEEGKATTGEGFEWGVLIGIGPGLTVETVVLRSVSTAN

SEQ ID NO: 117 – TKS candidate isolated from Marchantia polymorpha subsp. ruderalis

Amino acid sequence

MATRVLSSQENFEKLMADLARPNGHVYSQSQSQSGSGQNGAGTSIVAKNTASILAIGKALPPNRICQ
STYTDFYFRVTHCSHKTELKNRMQRICDKSGINTRYLLLDEEALKEHSEFYTPGQASIEQRHDLLEEA
VPKLAAQAAASALEEWGRPACDVTHLIVVTLSGVAIPGADVRLVKLLGLREDVSRVMLYMLGCYAGV
TALRLAKDLAENNPGSRVLIACSEMTATTFRAPSEKSMYDIVGASLFGDGAVGVIVGAKPRPGIERSIF
EIHWAGVSLAPDTEHVVQGKLKPDGLYFFLDKSLPGLVGKHIAPFCRSLLDHAPENLNLGFNEVFWA
VHPGGPAILNTVEEQLLLNSEKLRASRDVLANYGNVSASSVLYVLDELRHRPGQEEWGAALAFGPGI
TFEGVLLRRNVNHR

SEQ ID NO: 118 – TKS candidate isolated from Oryza sativa

Amino acid sequence

MGKQGGQQLVAAILGIGTAVPPYVLPQSSFPDYYFDISNSNHLLDLKAKFADICEKTMIDKRHVHMSD
EFLRSNPSVAAYNSPSINVRQNLTDVTVPQLGAAAARLAIADWGRPACEITHLVMCTTVSGCMPGAD
FEVVKLLGLPLTTKRCMMYHIGCHGGGTALRLAKDLAENNPGGRVLVVCSEVVSMVFRGPCESHMG
NLVGQALFGDAAGAVVVGADPVEANGERTLFEMVSAWQDIIPETEEMVVAKLREEGLVYNLHRDVAA
RVAASMESLVKKAMVEKDWNEEVFWLVHPGGRDILDRVVLTLGLRDDKVAVCREVMRQHGNTLSS
CVIVAMEEMRRRSADRGLSTAGEGLEWGLLFGFGPGLTVETILLRAPPCNQAQAV

SEQ ID NO: 119 – TKS candidate isolated from Punica granatum

Amino acid sequence

MGYSQQAKGPATIMAIGTAIPSYVVYQADFPDYYFRLSGCDHMTELKEKFIRICEKSTIRKRHMHLTEE
ILKQNPAILTYDGPSLNVRQQLVASEVPKLAMEAASKAIEEWGQPVWKITHLVFSSVVGAATPGADYK
LIKLLGLEPSVKRVPLYQQGCYVGGTALRIAKDLAENNASARVLVVCVDNTISSFRGPSKHITNLVGQA
LFSDGASAAIVGADPIPSVERPIFQIAHTSMHLVPDSDSEVTLDFLDAGLIVHVSEKVPSLIADNLEKSLV
EALGPTGINDWNSLFWAAHPGGPKILDMIEAKLGLRKEKLRATRTVLREYGNMIGACLLFILDEIRQNS
LEAGMATTGEGFDWGILLGFGPGLTVEAVVLRSFPIAK

SEQ ID NO: 120 – TKS candidate isolated from Citrus x microcarpa

Amino acid sequence

MAKVKNFLNAKRSKGPASILAIGTANPPTCFNQSDYPDFYFRVTDCEHKTELKDKFKRICDRSAVKKR
YLHVTEEVLKENPSMRSYNAPSLDARQALLIEQVPKLGKEAAAKAIKEWGQPLSKITHLVFSAMAGVDI
PGADLRLMNLLGLEPSVKRLMIYSQGCFIGGAAIRCAKDFAENNAGARVLVVFSDIMNMYFHEPQEA
HLDILVGQAVFGDGAAAVIVGADPEVSIERPLFHVVSSTQMSVPDTNKFIRAHVKEMGMELYLSKDVP
ATVGKNIEKLLVDAVSPFGISDWNSLFYSVHPGGRAILDQVELNLGLGKEKLRASRHVLSEYGNMGG
SSVYFILDEIRKKSMQEAKPTTGDGLEWGVLFAIGPGLTVETVILLSVPIDSAC

SEQ ID NO: 121 – TKS candidate isolated from Rhododendron dauricum

Amino acid sequence

MALVNHRENVKGRAQILAIGTANPKNCFRQVDYPDYYFRVTKSDHLIDLKAKFKRMCEKSMIEKRYM
HVNEEILEQNPSMNHGGEKMVSSLDVRLDMEIMEIPKLAAEAATKAMDEWGQPKSRITHLVFHSTLG
TVMPGVDYELIKLLGLNPSVKRFMLYHLGCYGGGTVLRLAKDLAENNPGSRVLVLCCEMMPSGFHG
PPSLQHAHLDILTGHAIFGDGAGAVIVGCVDPSGGTNGVVERGVRRYEQPLFEIHSAYQTVLPDSKDA
VGGRLREAGLIYYLSKRLSNDVSGKIDECCLAEAFSAAIKDNFEDWNSLFWIVHPAGRPILDKLDAKLG
LNKEKLRASRNVLRDYGNMWSSSVLFVLDEMRKGSIAQRKTTTGEGFEWGVLLGFGPGVTVETVVL
RSVPTAKLK

SEQ ID NO: 122 – TKS candidate isolated from Curcuma zedoaria

Amino acid sequence

MEANGYRITHSADGPATILAIGTANPTNVVDQNAYPDFYFRVTNSEHLQELKAKFRRICEKAAIRKRHL
YLTEEILRENPSLLAPMAPSFDARQAIVVEAVPKLAKEAAEKAIKEWGRPKSDITHLVFCSASGIDMPG
SDLQLLKLLGLPPSVNRVMLYNVGCHAGGTALRVAKDLAENNRGARVLAVCSEVTVLSYRGPHPAHI
ESLFVQALFGDGAAALVVGSDPVDGVERPIFEIASASQVMLPESEEAVGGHLREIGLTFHLKSQLPSII
ASNIEQSLTTACSPLGLSDWNQLFWAVHPGGRAILDQVEARLGLEKDRLAATRHVLSEYGNMQSATV
LFILDEMRNRSAAEGHATTGEGLDWGVLLGFGPGLSIETVVLHSCRLN

SEQ ID NO: 123 – TKS candidate isolated from Garcinia mangostana

Amino acid sequence

MAPAMDSAQNGHQSRGSANVLAIGTANPPNVILQEDYPDFYFKVTNSEHLTDLKEKFKRICVKSKTR
KRHFYLTEQILKENPGIATYGAGSLDSRQKILETEIPKLGKEAAMVAIQEWGQPVSKITHVVFATTSGF
MMPGADYSITRLLGLNPNVRRVMIYNQGCFAGGTALRVAKDLAENNKGARVLVVCAENTAMTFHGP
NENHLDVLVGQAMFSDGAAALIIGANPNLPEERPVYEMVAAHQTIVPESDGAIVAHFYEMGMSYFLKE
NVIPLFGNNIEACMEAAFKEYGISDWNSLFYSVHPGGRAIVDGIAEKLGLDEENLKATRHVLSEYGNM
GSACVIFILDELRKKSKEEKKLTTGDGKEWGCLIGLGPGLTVETVVLRSVPIA

SEQ ID NO: 124 – TKS candidate isolated from Arachis hypogaea

Amino acid sequence

MGSLGATQEGNGAKGVATILAIGTANPPNIIRQDDYPDFYFRATKSNHMLHLKEKFQRLCKNSMIEKR
HFLYNEDLLMENPNIVTYGASSLNTRQNILIKEVPKLGKEAALKAINEWGQPLSEITHLIFYTTSCFGNM
PGPDYHLAKLLGLKPTVNRHMIFNNGCHGGGAVLRVAKDIVENNAGSRVLVVWVETMVASFHGPNP
NHMDVLVGQALFGDGAGALIIGTNPKPCIECPLFELVLASQTTIPNTESSINGNIQEMGLVYYLGKEIPIA
ISENIDKCLINAFRESSVDWNSLFYAIHPGGPSILNRIEEKLGLKKEKLRASRKVLSQYGNMWSPGVIFV
LDELRNWSKIEGKSTCGEGKEWGVLVGFGPGLSLELLVLRSFCFDG

SEQ ID NO: 125 – TKS candidate isolated from Aquilaria sinensis

Amino acid sequence

MAAQPVEWVRKADRAAGPAAVLAMATANPSNFYLQSDFPDFYFRVTRSDHMSDLKEKFKRICKKTT
VRKRHMILTEEILNKNPAIADYWSPSLAARHDLALANIPQLGKEAADKAIKEWGQPKSKITHLVFCTSA
GVLMPGADYQLTMLLGLNPSISRLMLHNLGCYAGGTALRVAKDLAENNGGARVLVVCSEANLLNFRG
PSETHIDALITQSLFADGAAALIVGSDPDLQTESPLYELISASQRILPESEDAIVGRLTEAGLVPYLPKDI
PKLVSTNIRSILEDALAPTGVQDWNSIFWIIHPGMPAILDQTEKLLQLDKEKLKATRHVLSEFGNMFSAT
VLFILDQLRKGAVAEGKSTTGEGCEWGVLFSFGPGFTVETVLLRSVATATLTDA

SEQ ID NO: 126 – TKS candidate isolated from Cs.

Amino acid sequence

MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEKFRKICDKSMIRKRNCFLNEE
HLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWGQPKSKITHLIFTSASTTDMPGADY
HCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACLFRGPSESDLELL
VGQAIFGDGAAAVIVGAEPDESVGERPIFELVSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNI
EKCLIEAFTPIGISDWNSIFWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDE
LRKRSLEEGKSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY

SEQ ID NO: 127 – CBGaS candidate isolated from Sb.PT (A0A193PS58)

Amino acid sequence

MPATRTPIHPEAAAYKNPRYQSGPLSVIPKSFVPYCELMRLELPHGNFLGYFPHLVGLLYGSSASPAR
LPANEVAFQAVLYIGWTFFMRGAGCAWNDVVDQDFDRKTTRCRVRPVARGAVSTTSANIFGFAMVA
LAFACISPLPAECQRLGLMTTVLSIIYPFCKRVTNFAQVILGMTLAINFILAAYGAGLPAIEAPYTVPTICV
TTAITLLVVFYDVVYARQDTADDLKSGVKGMAVLFRNYVEILLTSITLVIAGLIATTGVLVDNGPYFFVFS

VAGLLAALLAMIGGIRYRIFHTWNSYSGWFYALAIFNLLGGYLIEYLDQVPMLNKA

SEQ ID NO: 128 – CBGaS candidate isolated from Sc.PT (A0A084RYZ7)

Amino acid sequence

MSAKVSPMAYTNPRYETGPLSLIPKPIVPYFELMRFELPHGYYLGYFPHLVGIMYGASAGPERLPARD
LVFQALLYVGWTFAMRGAGCAWNDNIDQDFDRKTERCRTRPIARGAVSTTAGHVFAVAGVALAFLC
LSPLPTECHQLGVLVTVLSVIYPFCKRFTNFAQVILGMTLAANFILAAYGAGLPALEQPYTRPTMSATL
AITLLVVFYDVVYARQDTADDLKSGVKGMAVLFRNHIEVLLAVLTCTIGGLLAATGVSVGNGPYYFLFS
VAGLTVALLAMIGGIRYRIFHTWNGYSGWFYVLAIINLMSGYFIEYLDNAPILARGS

SEQ ID NO: 129 – CBGaS candidate A0A084B1B1

Amino acid sequence

MSAKVSPMAYTNPRYERGPLSLIPKPIVPYFELMRFELPHGYYLGYFPHLVGIMYGASAGPERLPARD
LVFQALLYVGWTFAMRGAGCAWNDNIDQDFDRKTERCRTRPIARGAVSTTAGHVFAVAGVALAFLC
LSPLPTECHQLGVLVTVLSVIYPFCKRFTNFAQVILGMTLAANFILAAYGAGLPALEQPYTRPTMSATL
AITLLVVFYDVVYARQDTADDLKSGVKGMAVLFRNHIEVLLAVLTCTIGGLLAATGVSVGNGPYYFLFS
VAGLTVALLAMIGGIRYRIFHTWNGYSGWFYVLAIINLMSGYFIEYLDNAPILARGS

SEQ ID NO: 130 – CBGaS candidate A0A084QZF6

Amino acid sequence

MSPKVSSMPYTNPRYESGPLSLIPKSIVPYFELMRFELPHGYYLGYFPHLVGIMYGASAGPERLPARD
LVFQALLYVGWTFAMRGAGCAWNDNIDQDFDRKTERCRTRPIARGAVSTTAGHIFAVAGVALAFLCL
SPLPTECHQLGVLVTVLSVIYPFCKRFTNFAQVILGMTLAANFILAAYGAGLPALEQPYTRPTMFATLAI
TLLVVFYDVVYARQDTADDLKSGVKGMAVLFRNHIEVLLAVLTCTIGGLLAATGVSVGNGPYYFLFSV
AGLTVALLAMIGGIRYRIFHTWNGYSGWFYVLAIINLMSGYFIEYLDNAPILARGS

SEQ ID NO: 131 – CBGaS candidate CBGaS 1 – Cs.PT4-T

Amino acid sequence

MAGSDQIEGSPHHESDNSIATKILNFGHTCWKLQRPYVVKGMISIACGLFGRELFNNRHLFSWGLMW
KAFFALVPILSFNFFAAIMNQIYDVDIDRINKPDLPLVSGEMSIETAWILSIIVALTGLIVTIKLKSAPLFVFI
YIFGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGLAFTSYSATTSALGLPFVWRPAFSFIIAFMTVMGM
TIAFAKDISDIEGDAKYGVSTVATKLGARNMTFVVSGVLLLNYLVSISIGIIWPQVFKSNIMILSHAILAFC
LIFQTRELALANYASAPSRQFFEFIWLLYYAEYFVYVFI

SEQ ID NO: 132 – GPPS candidate isolated from Streptomyces actuosus

Amino acid sequence

MTTEVTSFTGAGPHPAASVRRITDDLLQRVEDKLASFLTAERDRYAAMDERALAAVDALTDLVTSGG
KRVRPTFCITGYLAAGGDAGDPGIVAAAAGLEMLHVSALIHDDILDNSAQRRGKPTIHTLYGDLHDSH
GWRGESRRFGEGIGILIGNLALVYSQELVCQAPPAVLAEWHRLCSEVNIGQCLDVCAAAEFSADPEL
SRLVALIKSGRYTIHRPLVMGANAASRPDLAAAYVEYGEAVGEAFQLRDDLLDAFGDSTETGKPTGLD
FTQHKMTLLLGWAMQRDTHIRTLMTEPGHTPEEVRRRLEDTEVPKDVERHIADLVEQGRAAIADAPID
PQWRQELADMAVRAAYRTN

SEQ ID NO: 133 – GPPS candidate IpMSv3

Amino acid sequence

MAFKLAQRLPKSVSSLGSQLSKNAPNQLAAATTSQLINTPGIRHKSRSSAVPSSLSKSMYDHNEEMK
AAMKYMDEIYPEVMGQIEKVPQYEEIKPILVRLREAIDYTVPYGKRFKGVHIVSHFKLLADPKFITPENV
KLSGVLGWCAEIIQAYFCMLDDIMDDSDTRRGKPTWYKLPGIGLNAVTDVCLMEMFTFELLKRYFPKH

PSYADIHEILRNLLFLTHMGQGYDFTFIDPVTRKINFNDFTEENYTKLCRYKIIFSTFHNTLELTSAMANV
YDPKKIKQLDPVLMRIGMMHQSQNDFKDLYRDQGEVLKQAEKSVLGTDIKTGQLTWFAQKALSICND
RQRKIIMDNYGKEDNKNSEAVREVYEELDLKGKFMEFEEESFEWLKKEIPKINNGIPHKVFQDYTYGV
FKRRPE

SEQ ID NO: 134 – GPPS candidate SmGPPS_LSUv1

Amino acid sequence

MAFDFKRYMVEKADSVNKALEAVVQMKEPLKIHESMRYSLLAGGKRVRPMLCIAACELVGGEESTA
MPAACAVEMIHTMSLMHDDLPCMDNDDLRRGKPTNHKVFGEDVAVLAGDALLSLAFEHVAVATRGS
APERILRALGQLAKSIGAEGLVAGQVVDICSEGMAEVGLDHLEFIHLHKTAALLQGSVVMGAILGGAKE
EEVERLRKFAKCIGLMFQVVDDILDVTKSSHELGKTAGKDLVADKTTYPKLLGVQKSKEFADDLNREA
QEQLLHFDSHKAAPLIAIANYIAYRNN

SEQ ID NO: 135 – GPPS candidate SmGPPS_SSUv1

Amino acid sequence

MAQNHSYWAAIEADIDTYLKKSIAIRSPETVFEPMHHLTFAAPRTAASAICVAACELVGGERSQAIATA
SAIHIMHAAAYAHEHLPLTDRPRPNSKPAIQHKYGPNIELLTGDGMASFGFELLAGSIRSDHPNPERIL
RVIIEISRASGSEGIIDGFYREKEIVDQHSRFDFIEYLCRKKYGEMHACAAASGAILAGGAEEEIQKLRN
FGHYAGTLIGLLHKKIDTPQIQNVIGKLKDLALKELEGFHGKNVELLCSLVADASLCEAELEV

SEQ ID NO: 136 – GPPS candidate CrGPPA_LSUv1

Amino acid sequence

MAFDFKAYMIGKANSVNKALEDAVLVREPLKIHESMRYSLLAGGKRVRPMLCIAACELFGGTESVAM
PSACAVEMIHTMSLMHDDLPCMDNDDLRRGKPTNHKVFGEDVAVLAGDALLAFAFEHIATATKGVSS
ERIVRVVGELAKCIGSEGLVAGQVVDVCSEGIADVGLEHLEFIHIHKTAALLEGSVVLGAIVGGANDEQI
SKLRKFARCIGLLFQVVDDILDVTKSSQELGKTAGKDLVADKVTYPKLLGIDKSREFAEKLNREAQEQL
AEFDPEKAAPLIALANYIAYRDN

SEQ ID NO: 137 – GPPS candidate CrGPPS_SSUv1

Amino acid sequence

MAMKSNSWANIESDIQTHLKKSIPIRAPEDVFEPMHYLTFAAPRTTAPALCIAACEVVGGDGDQAMAA
AAAIHLVHAAAYAHENLPLTDRRRPKPPIQHKFNSNIELLTGDGIVPYGFELLAKSMDSNNSDRILRVIIE
ITQAAGSKGIIDGQFRELDVIDSEINMGLIEYVCKKKEGELNACGAACGAILGGGSEEEIGKLRKFGLYA
GMIQGLVHGVGKNREEIQELVRKLRYLAMEELKSLKNRKIDTISSLLETDLCSV

SEQ ID NO: 138 – Cs.OAC

Amino acid sequence

MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQ
DYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK

SEQ ID NO: 139 – Sc.ACS1

Amino acid sequence

MSPSAVQSSKLEEQSSEIDKLKAKMSQSASTAQQKKEHEYEHLTSVKIVPQRPISDRLQPAIATHYSP
HLDGLQDYQRLHKESIEDPAKFFGSKATQFLNWSKPFDKVFIPDSKTGRPSFQNNAWFLNGQLNACY
NCVDRHALKTPNKKAIIFEGDEPGQGYSITYKELLEEVCQVAQVLTYSMGVRKGDTVAVYMPMVPEAI
ITLLAISRIGAIHSVVFAGFSSNSLRDRINDGDSKVVITTDESNRGGKVIETKRIVDDALRETPGVRHVLV
YRKTNNPSVAFHAPRDLDWATEKKKYKTYYPCTPVDSEDPLFLLYTSGSTGAPKGVQHSTAGYLLGA
LLTMRYTFDTHQEDVFFTAGDIGWITGHTYVVYGPLLYGCATLVFEGTPAYPNYSRYWDIIDEHKVTQ

FYVAPTALRLLKRAGDSYIENHSLKSLRCLGSVGEPIAAEVWEWYSEKIGKNEIPIVDTYWQTESGSHL
VTPLAGGVTPMKPGSASFPFFGIDAVVLDPNTGEELNTSHAEGVLAVKAAWPSFARTIWKNHDRYLD
TYLNPYPGYYFTGDGAAKDKDGYIWILGRVDDVVNVSGHRLSTAEIEAAIIEDPIVAECAVVGFNDDLT
GQAVAAFVVLKNKSNWSTATDDELQDIKKHLVFTVRKDIGPFAAPKLIILVDDLPKTRSGKIMRRILRKIL
AGESDQLGDVSTLSNPGIVRHLIDSVKL

SEQ ID NO: 140 – Sc. ACS2

Amino acid sequence

MTIKEHKVVYEAHNVKALKAPQHFYNSQPGKGYVTDMQHYQEMYQQSINEPEKFFDKMAKEYLHW
DAPYTKVQSGSLNNGDVAWFLNGKLNASYNCVDRHAFANPDKPALIYEADDESDNKIITFGELLRKVS
QIAGVLKSWGVKKGDTVAIYLPMIPEAVIAMLAVARIGAIHSVVFAGFSAGSLKDRVVDANSKVVITCD
EGKRGGKTINTKKIVDEGLNGVDLVSRILVFQRTGTEGIPMKAGRDYWWHEEAAKQRTYLPPVSCDA
EDPLFLLYTSGSTGSPKGVVHTTGGYLLGAALTTRYVFDIHPEDVLFTAGDVGWITGHTYALYGPLTL
GTASIIFESTPAYPDYGRYWRIIQRHKATHFYVAPTALRLIKRVGEAEIAKYDTSSLRVLGSVGEPISPD
LWEWYHEKVGNKNCVICDTMWQTESGSHLIAPLAGAVPTKPGSATVPFFGINACIIDPVTGVELEGND
VEGVLAVKSPWPSMARSVWNHHDRYMDTYLKPYPGHYFTGDGAGRDHDGYYWIRGRVDDVVNVS
GHRLSTSEIEASISNHENVSEAAVVGIPDELTGQTVVAYVSLKDGYLQNNATEGDAEHITPDNLRRELI
LQVRGEIGPFASPKTIILVRDLPRTRSGKIMRRVLRKVASNEAEQLGDLTTLANPEVVPAIISAVENQFF
SQKKK

SEQ ID NO: 141 – Sc.ALD6

Amino acid sequence

MTKLHFDTAEPVKITLPNGLTYEQPTGLFINNKFMKAQDGKTYPVEDPSTENTVCEVSSATTEDVEYAI
ECADRAFHDTEWATQDPRERGRLLSKLADELESQIDLVSSIEALDNGKTLALARGDVTIAINCLRDAAA
YADKVNGRTINTGDGYMNFTTLEPIGVCGQIIPWNFPIMMLAWKIAPALAMGNVCILKPAAVTPLNALY
FASLCKKVGIPAGVVNIVPGPGRTVGAALTNDPRIRKLAFTGSTEVGKSVAVDSSESNLKKITLELGGK
SAHLVFDDANIKKTLPNLVNGIFKNAGQICSSGSRIYVQEGIYDELLAAFKAYLETEIKVGNPFDKANFQ
GAITNRQQFDTIMNYIDIGKKEGAKILTGGEKVGDKGYFIRPTVFYDVNEDMRIVKEEIFGPVVTVAKFK
TLEEGVEMANSSEFGLGSGIETESLSTGLKVAKMLKAGTVWINTYNDFDSRVPFGGVKQSGYGREM
GEEVYHAYTEVKAVRIKL

SEQ ID NO: 142 – Zm.PDC

Amino acid sequence

MSYTVGTYLAERLVQIGLKHHFAVAGDYNLVLLDNLLLNKNMEQVYCCNELNCGFSAEGYARAKGAA
AAVVTYSVGALSAFDAIGGAYAENLPVILISGAPNNNDHAAGHVLHHALGKTDYHYQLEMAKNITAAA
EAIYTPEEAPAKIDHVIKTALREKKPVYLEIACNIASMPCAAPGPASALFNDEASDEASLNAAVEETLKFI
ANRDKVAVLVGSKLRAAGAEEAAVKFADALGGAVATMAAAKSFFPEENPHYIGTSWGEVSYPGVEK
TMKEADAVIALAPVFNDYSTTGWTDIPDPKKLVLAEPRSVVVNGIRFPSVHLKDYLTRLAQKVSKKTG
ALDFFKSLNAGELKKAAPADPSAPLVNAEIARQVEALLTPNTTVIAETGDSWFNAQRMKLPNGARVEY
EMQWGHIGWSVPAAFGYAVGAPERRNILMVGDGSFQLTAQEVAQMVRLKLPVIIFLINNYGYTIEVMI
HDGPYNNIKNWDYAGLMEVFNGNGGYDSGAGKGLKAKTGGELAEAIKVALANTDGPTLIECFIGRED
CTEELVKWGKRVAAANSRKPVNKLL

SEQ ID NO: 143 pGAL1

TGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAAGCCGCCGA
GCGGGCGACAGCCCTCCGACGGAAGACTCTCCTCCGTGCGTCCTGGTCTTCACCGGTCGCGTT
CCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTT
TTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATCAACGAATCA
AATTAACAACCATAGGATAATAATGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGC
GAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCAAAAGCTGCATAACCACTTTAACTAAT
ACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTCATAAAAGTATCAACAAAAAATTGTT
AATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATA

SEQ ID NO: 144 pGAL10

CATCGCTTCGCTGATTAATTACCCCAGAAATAAGGCTAAAAAACTAATCGCATTATTATCCTATGG
TTGTTAATTTGATTCGTTGATTTGAAGGTTTGTGGGGCCAGGTTACTGCCAATTTTTCCTCTTCAT
AACCATAAAAGCTAGTATTGTAGAATCTTTATTGTTCGGAGCAGTGCGGCGCGAGGCACATCTGC
GTTTCAGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTCCGTCGGAGGGCTG
TCGCCCGCTCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGTATTACTGA
AAGTTCCAAAGAGAAGGTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAACATATAA
GTAAGATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGATTATTAAACTTCTTTGCG
TCCATCCAAAAAAAAAGTAAGAATTTTTGAAAATTCAATATAA

SEQ ID NO: 145 pGAL2

GGCTTAAGTAGGTTGCAATTTCTTTTTCTATTAGTAGCTAAAAATGGGTCACGTGATCTATATTCG
AAAGGGGCGGTTGCCTCAGGAAGGCACCGGCGGTCTTTCGTCCGTGCGGAGATATCTGCGCCG
TTCAGGGGTCCATGTGCCTTGGACGATATTAAGGCAGAAGGCAGTATCGGGGCGGATCACTCCG
AACCGAGATTAGTTAAGCCCTTCCCATCTCAAGATGGGGAGCAAATGGCATTATACTCCTGCTAG
AAAGTTAACTGTGCACATATTCTTAAATTATACAATGTTCTGGAGAGCTATTGTTTAAAAAACAAAC
ATTTCGCAGGCTAAAATGTGGAGATAGGATTAGTTTTGTAGACATATATAAACAATCAGTAATTGG
ATTGAAAATTTGGTGTTGTGAATTGCTCTTCATTATGCACCTTATTCAATTATCATCAAGAATAGCA
ATAGTTAAGTAAACACAAGATTAACATAATAAAAAAAATAATTCTTTCATA

SEQ ID NO: 146 pGAL3

TTTTACTATTATCTTCTACGCTGACAGTAATATCAAACAGTGACACATATTAAACACAGTGGTTTCT
TTGCATAAACACCATCAGCCTCAAGTCGTCAAGTAAAGATTTCGTGTTCATGCAGATAGATAACAA
TCTATATGTTGATAATTAGCGTTGCCTCATCAATGCGAGATCCGTTTAACCGGACCCTAGTGCAC
TTACCCCACGTTCGGTCCACTGTGTGCCGAACATGCTCCTTCACTATTTTAACATGTGGAATTCTT
GAAAGAATGAAATCGCCATGCCAAGCCATCACACGGTCTTTTATGCAATTGATTGACCGCCTGCA
ACACATAGGCAGTAAAATTTTTACTGAAACGTATATAATCATCATAAGCGACAAGTGAGGCAACAC
CTTTGTTACCACATTGACAACCCCAGGTATTCATACTTCCTATTAGCGGAATCAGGAGTGCAAAAA
GAGAAAATAAAAGTAAAAAGGTAGGGCAACACATAGT

SEQ ID NO: 147 pGAL7

GGACGGTAGCAACAAGAATATAGCACGAGCCGCGAAGTTCATTTCGTTACTTTTGATATCGCTCA
CAACTATTGCGAAGCGCTTCAGTGAAAAAATCATAAGGAAAAGTTGTAAATATTATTGGTAGTATT
CGTTTGGTAAAGTAGAGGGGGTAATTTTTCCCCTTTATTTTGTTCATACATTCTTAAATTGCTTTGC
CTCTCCTTTTGGAAAGCTATACTTCGGAGCACTGTTGAGCGAAGGCTCATTAGATATATTTTCTGT
CATTTTCCTTAACCCAAAAATAAGGGAAAGGGTCCAAAAAGCGCTCGGACAACTGTTGACCGTGA
TCCGAAGGACTGGCTATACAGTGTTCACAAAATAGCCAAGCTGAAAATAATGTGTAGCTATGTTC
AGTTAGTTTGGCTAGCAAAGATATAAAAGCAGGTCGGAAATATTTATGGGCATTATTATGCAGAG
CATCAACATGATAAAAAAAAAACAGTTGAATATTCCCTCAAAA

SEQ ID NO: 148 pGAL4

GCGACACAGAGATGACAGACGGTGGCGCAGGATCCGGTTTAAACGAGGATCCCTTAAGTTTAAA
CAACAACAGCAAGCAGGTGTGCAAGACACTAGAGACTCCTAACATGATGTATGCCAATAAAACAC
AAGAGATAAACAACATTGCATGGAGGCCCCAGAGGGGCGATTGGTTTGGGTGCGTGAGCGGCA
AGAAGTTTCAAAACGTCCGCGTCCTTTGAGACAGCATTCGCCCAGTATTTTTTTTATTCTACAAAC
CTTCTATAATTTCAAAGTATTTACATAATTCTGTATCAGTTTAATCACCATAATATCGTTTTCTTTGT
TTAGTGCAATTAATTTTTCCTATTGTTACTTCGGGCCTTTTTCTGTTTTATGAGCTATTTTTTCCGTC
ATCCTTCCCCAGATTTTCAGCTTCATCTCCAGATTGTGTCTACGTAATGCACGCCATCATTTTAAG
AGAGGACAGAGAAGCAAGCCTCCTGAAAG

SEQ ID NO: 149 pMAL1

GATGATGGACACTAGTGTGTCGAGAATGTATCAACTATATATAGTCCTAATGCCACACAAATATGA
AGTGGGGGAAGCCCATTCTTAATCCGGCTCAATTTTGGTGCGTGATCGCGGCCTATGTTTGCTTC

CAGAAAAAGCTTAGAATAATATTTCTCACCTTTGATGGAATGCTCGCGAGTGCTCGTTTTGATTAC
CCCATATGCATTGTTGCAGCATGCAAGCACTATTGCAAGCCACGCATGGAAGAAATTTGCAAACA
CCTATAGCCCCGCGTTGTTGAGGAGGTGGACTTGGTGTAGGACCATAAAGCTGTGCACTACTAT
GGTGAGCTCTGTCGTCTGGTGACCTTCTATCTCAGGCACATCCTCGTTTTTGTGCATGAGGTTCG
AGTCACGCCCACGGCCTATTAATCCGCGAAATAAATGCGAAATCTAAATTATGACGCAAGGCTGA
GAGATTCTGACACGCCGCATTTGCGGGGCAGTAATTATCGGGCAGTTTTCCGGGGTTCGGGATG
GGGTTTGGAGAGAAAGTTCAACACAGACCAAAACAGCTTGGGACCACTTGGATGGAGGTCCCCG
CAGAAGAGCTCTGGCGCGTTGGACAAACATTGACAATCCACGGCAAAATTGTCTACAGTTCCGT
GTATGCGGATAGGGATATCTTCGGGAGTATCGCAATAGGATACAGGCACTGTGCAGATTACGCG
ACATGATAGCTTTGTATGTTCTACAGACTCTGCCGTAGCAGTCTAGATATAATATCGGAGTTTTGT
AGCGTCGTAAGGAAAACTTGGGTTACACAGGTTTCTTGAGAGCCCTTTGACGTTGATTGCTCTGG
CTTCCATCCAGGCCCTCATGTGGTTCAGGTGCCTCCGCAGTGGCTGGCAAGCGTGGGGGTCAA
TTACGTCACTTCTATTCATGTACCCCAGACTCAATTGTTGACAGCAATTTCAGCGAGAATTAAATT
CCACAATCAATTCTCGCTGAAATAATTAGGCCGTGATTTAATTCTCGCTGAAACAGAATCCTGTCT
GGGGTACAGATAACAATCAAGTAACTATTATGGACGTGCATAGGAGGTGGAGTCCATGACGCAA
AGGGAAATATTCATTTTATCCTCGCGAAGTTGGGATGTGTCAAAGCGTCGCGCTCGCTATAGTGA
TGAGAATGTCTTTAGTAAGCTTAAGCCATATAAAGACCTTCCGCCTCCATATTTTTTTTTATCCCTC
TTGACAATATTAATTCCTT

SEQ ID NO: 150 pMAL2

AAGGAATTAATATTGTCAAGAGGGATAAAAAAAAATATGGAGGCGGAAGGTCTTTATATGGCTTA
AGCTTACTAAAGACATTCTCATCACTATAGCGAGCGCGACGCTTTGACACATCCCAACTTCGCGA
GGATAAAATGAATATTTCCCTTTGCGTCATGGACTCCACCTCCTATGCACGTCCATAATAGTTACT
TGATTGTTATCTGTACCCCAGACAGGATTCTGTTTCAGCGAGAATTAAATCACGGCCTAATTATTT
CAGCGAGAATTGATTGTGGAATTTAATTCTCGCTGAAATTGCTGTCAACAATTGAGTCTGGGGTA
CATGAATAGAAGTGACGTAATTGACCCCCACGCTTGCCAGCCACTGCGGAGGCACCTGAACCAC
ATGAGGGCCTGGATGGAAGCCAGAGCAATCAACGTCAAAGGGCTCTCAAGAAACCTGTGTAACC
CAAGTTTTCCTTACGACGCTACAAAACTCCGATATTATATCTAGACTGCTACGGCAGAGTCTGTA
GAACATACAAAGCTATCATGTCGCGTAATCTGCACAGTGCCTGTATCCTATTGCGATACTCCCGA
AGATATCCCTATCCGCATACACGGAACTGTAGACAATTTTGCCGTGGATTGTCAATGTTTGTCCA
ACGCGCCAGAGCTCTTCTGCGGGGACCTCCATCCAAGTGGTCCCAAGCTGTTTTGGTCTGTGTT
GAACTTTCTCTCCAAACCCCATCCCGAACCCCGGAAAACTGCCCGATAATTACTGCCCCGCAAAT
GCGGCGTGTCAGAATCTCTCAGCCTTGCGTCATAATTTAGATTTCGCATTTATTTCGCGGATTAAT
AGGCCGTGGGCGTGACTCGAACCTCATGCACAAAAACGAGGATGTGCCTGAGATAGAAGGTCA
CCAGACGACAGAGCTCACCATAGTAGTGCACAGCTTTATGGTCCTACACCAAGTCCACCTCCTCA
ACAACGCGGGGCTATAGGTGTTTGCAAATTTCTTCCATGCGTGGCTTGCAATAGTGCTTGCATGC
TGCAACAATGCATATGGGGTAATCAAAACGAGCACTCGCGAGCATTCCATCAAAGGTGAGAAATA
TTATTCTAAGCTTTTTCTGGAAGCAAACATAGGCCGCGATCACGCACCAAAATTGAGCCGGATTA
AGAATGGGCTTCCCCCACTTCATATTTGTGTGGCATTAGGACTATATATAGTTGATACATTCTCGA
CACACTAGTGTCCATCATC

SEQ ID NO: 151 pMAL11

GCGCCTCAAGAAAATGATGCTGCAAGAAGAATTGAGGAAGGAACTATTCATCTTACGTTGTTTGT
ATCATCCCACGATCCAAATCATGTTACCTACGTTAGGTACGCTAGGAACTAAAAAAAGAAAAGAA
AAGTATGCGTTATCACTCTTCGAGCCAATTCTTAATTGTGTGGGGTCCGCGAAAATTTCCGGATA
AATCCTGTAAACTTTAACTTAAACCCCGTGTTTAGCGAAATTTTCAACGAAGCGCGCAATAAGGA
GAAATATTATCTAAAAGCGAGAGTTTAAGCGAGTTGCAAGAATCTCTACGGTACAGATGCAACTT
ACTATAGCCAAGGTCTATTCGTATTACTATGGCAGCGAAAGGAGCTTTAAGGTTTTAATTACCCCA
TAGCCATAGATTCTACTCGGTCTATCTATCATGTAACACTCCGTTGATGCGTACTAGAAAATGACA
ACGTACCGGGCTTGAGGGACATACAGAGACAATTACAGTAATCAAGAGTGTACCCAACTTTAACG
AACTCAGTAAAAAATAAGGAATGTCGACATCTTAATTTTTTATATAAAGCGGTTTGGTATTGATTGT
TTGAAGAATTTTCGGGTTGGTGTTTCTTTCTGATGCTACATAGAAGAACATCAAACAACTAAAAAA
ATAGTATAAT

SEQ ID NO: 152 pMAL12

ATTATACTATTTTTTTAGTTGTTTGATGTTCTTCTATGTAGCATCAGAAAGAAACACCAACCCGAAA
ATTCTTCAAACAATCAATACCAAACCGCTTTATATAAAAAATTAAGATGTCGACATTCCTTATTTTTT
ACTGAGTTCGTTAAAGTTGGGTACACTCTTGATTACTGTAATTGTCTCTGTATGTCCCTCAAGCCC
GGTACGTTGTCATTTTCTAGTACGCATCAACGGAGTGTTACATGATAGATAGACCGAGTAGAATC
TATGGCTATGGGGTAATTAAAACCTTAAAGCTCCTTTCGCTGCCATAGTAATACGAATAGACCTTG
GCTATAGTAAGTTGCATCTGTACCGTAGAGATTCTTGCAACTCGCTTAAACTCTCGCTTTTAGATA
ATATTTCTCCTTATTGCGCGCTTCGTTGAAAATTTCGCTAAACACGGGGTTTAAGTTAAAGTTTAC
AGGATTTATCCGGAAATTTTCGCGGACCCCACACAATTAAGAATTGGCTCGAAGAGTGATAACGC
ATACTTTTCTTTTCTTTTTTTAGTTCCTAGCGTACCTAACGTAGGTAACATGATTTGGATCGTGGGA
TGATACAAACAACGTAAGATGAATAGTTCCTTCCTCAATTCTTCTTGCAGCATCATTTTCTTGAGG
CGCTCTGGGCAAGGTATAAAAAGTTCCATTAATACGTCTCTAAAAAATTAAATCATCCATCTCTTA
AGCAGTTTTTTTGATAATCTCAAATGTACATCAGTCAAGCGTAACTAAATTACATAA

SEQ ID NO: 153 pMAL31

TTATGTATTTTAGTTACGCTTGACTGATGTACATTTGAGATTATCAAAAAAACTGCTTAAGAGATAG
ATGGTTTAATTTTTTTAGAGACGTATTAATGGAACTTTTTATACCTTGCCCAGAGCGCCTCAAGAAA
ATGATGCTGAAAGAAGAATTGAGGAAGGAACTACTCATCTTACGTTGTTTGTATCATCCCACGAT
CCAAATCATGTTACCTACGTTAGGTACGCTAGGAACTGAAAAAAGAAAAGAAAAGTATGCGTTAT
CACTCTTCGAGCCAATTCTTAATTGTGTGGGGTCCGCGAAAACTTCCGGATAAATCCTGTAAACT
TAAACTTAAACCCCGTGTTTAGCGAAATTTTCAACGAAGCGCGCAATAAGGAGAAATATTATATAA
AAGCGAGAGTTTAAGCGAGGTTGCAAGAATCTCTACGGTACAGATGCAACTTACTATAGCCAAGG
TCTATTCGTATTGGTATCCAAGCAGTGAAGCTACTCAGGGGAAAACATATTTTCAGAGATCAAAGT
TATGTCAGTCTCTTTTTCATGTGTAACTTAACGTTTGTGCAGGTATCATACCGGCCTCCACATAAT
TTTTGTGGGGAAGACGTTGTTGTAGCAGTCTCCTTATACTCTCCAACAGGTGTTTAAAGACTTCTT
CAGGCCTCATAGTCTACATCTGGAGACAACATTAGATAGAAGTTTCCACAGAGGCAGCTTTCAAT
ATACTTTCGGCTGTGTACATTTCATCCTGAGTGAGCGCATATTGCATAAGTACTCAGTATATAAAG
AGACACAATATACTCCATACTTGTTGTGAGTGGTTTTAGCGTATTCAGTATAACAATAAGAATTAC
ATCCAAGACTATTAATTAACT

SEQ ID NO: 154 pMAL32

AGTTAATTAATAGTCTTGGATGTAATTCTTATTGTTATACTGAATACGCTAAAACCACTCACAACAA
GTATGGAGTATATTGTGTCTCTTTATATACTGAGTACTTATGCAATATGCGCTCACTCAGGATGAA
ATGTACACAGCCGAAAGTATATTGAAAGCTGCCTCTGTGGAAACTTCTATCTAATGTTGTCTCCAG
ATGTAGACTATGAGGCCTGAAGAAGTCTTTAAACACCTGTTGGAGAGTATAAGGAGACTGCTACA
ACAACGTCTTCCCCACAAAAATTATGTGGAGGCCGGTATGATACCTGCACAAACGTTAAGTTACA
CATGAAAAAGAGACTGACATAACTTTGATCTCTGAAAATATGTTTTCCCCTGAGTAGCTTCACTGC
TTGGATACCAATACGAATAGACCTTGGCTATAGTAAGTTGCATCTGTACCGTAGAGATTCTTGCAA
CCTCGCTTAAACTCTCGCTTTTATATAATATTTCTCCTTATTGCGCGCTTCGTTGAAAATTTCGCTA
AACACGGGGTTTAAGTTTAAGTTTACAGGATTTATCCGGAAGTTTTCGCGGACCCCACACAATTA
AGAATTGGCTCGAAGAGTGATAACGCATACTTTTCTTTTCTTTTTTCAGTTCCTAGCGTACCTAAC
GTAGGTAACATGATTTGGATCGTGGGATGATACAAACAACGTAAGATGAGTAGTTCCTTCCTCAA
TTCTTCTTTCAGCATCATTTTCTTGAGGCGCTCTGGGCAAGGTATAAAAAGTTCCATTAATACGTC
TCTAAAAAATTAAACCATCTATCTCTTAAGCAGTTTTTTTGATAATCTCAAATGTACATCAGTCAAG
CGTAACTAAAATACATAA

What is claimed:

1. A method of making a cannabichromenic acid (CBCa), the method comprising:
    (a) culturing a population of host cells capable of producing cannabigerolic acid (CBGa) in a culture medium comprising a fermentation broth and an overlay, under conditions suitable for the host cells to produce CBGa, and wherein the CBGa partitions into the overlay;
    (b) separating the overlay from the fermentation broth;
    (c) combining the separated overlay of step (b), with a CBCa synthase, thereby producing a bioconversion mixture; and
    (d) purifying the CBCa from the bioconversion mixture.

2. The method of claim 1, wherein the overlay comprises a plant-based oil.

3. The method of claim 2, wherein the plant-based oil is selected from soybean oil, sunflower oil, safflower oil, canola oil, grapeseed oil, or castor oil.

4. The method of claim 3, wherein the plant-based oil is sunflower oil.

5. The method of claim 1, wherein the overlay comprises a synthetic ester or a fatty alcohol.

6. The method of any one of claims 1-5, wherein the CBCa synthase is produced by culturing a population of host cells capable of producing a CBCa synthase in a culture medium and under conditions suitable for the host cells to produce the CBCa synthase, thereby producing the CBCa synthase.

7. The method of any one of claims 1-6, wherein the overlay and the fermentation broth are separated by centrifugation.

8. The method of any one of claims 1-7, wherein the overlay and the fermentation broth are separated by demulsification.

9. The method of any one of claims 1-8, wherein the overlay comprises CBGa.

10. The method of claim 9, wherein the CBGa has a concentration in the overlay of between about 0.1% (w/v) or and 10% (w/v).

11. The method of claim 10, wherein the CBGa has a concentration of between about 0.5% (w/v) or and 5% (w/v).

12. The method of any one of claims 1-11, further comprising stirring the bioconversion mixture for between 12 hours and 144 hours before performing step (d).

13. The method of claim 12, comprising stirring the bioconversion mixture for between 24 hours and 96 hours.

14. The method of claim 13, comprising stirring the bioconversion mixture for about 48 hours.

15. The method of any one of claims 1-14, wherein the bioconversion mixture is at a temperature of between 4 °C and 50 °C.

16. The method of claim 15, wherein the bioconversion mixture is at a temperature of between 20 °C and 40 °C.

17. The method of claim 16, wherein the bioconversion mixture is at a temperature of about 35 °C.

18. The method of any one of claims 1-17, wherein the bioconversion mixture comprises one or more amphiphilic moieties.

19. The method of claim 18, wherein the one or more amphiphilic moieties comprises a cyclodextrin, plant-derived silica, cellulose, or a combination thereof.

20. The method of claim 19, wherein the cyclodextrin comprises randomly methylated cyclodextrin, 2, 6-Di-O-methyl-β-cyclodextrin, or a combination thereof.

21. The method of any one of claims 8-20, wherein the demulsification comprises one or more of: (i) contacting the culture medium with an enzymatic composition comprising a serine protease, (ii)contacting the culture medium with a surfactant; and (iii) contacting the culture medium with NaOH to adjust the culture medium to a pH of between pH 7 and pH 9.

22. The method of claim 21, wherein the demulsification comprises contacting the culture medium with the enzymatic composition comprising the serine protease.

23. The method of claim 21 or 22, wherein the purifying comprises contacting the culture medium with the surfactant.

24. The method of any one of claims 21-23, wherein the culture medium is contacted with the enzymatic composition or surfactant after the culture medium is adjusted to a pH of between about pH 7 and pH 9.

25. The method of claim 24, wherein the culture medium is contacted with the enzymatic composition or surfactant after the culture medium is adjusted to a pH of pH 8.

26. The method of any one of claims 21-25, wherein the final concentration of the enzymatic composition is from about 0.5% (w/v) to about 3% (w/v) after contacting the culture medium comprising a cannabinoid with the enzymatic composition.

27. The method of claim 26, wherein the culture medium is contacted with the enzymatic composition at a final concentration of about 1% (w/v).

28. The method of any one of claims 21-27, wherein the culture medium is mixed with the enzymatic composition for between 0.5 hours and 2 hours.

29. The method of any one of claims 21-28, wherein the demulsification comprises centrifugation of the culture medium.

30. The method of claim 29, wherein the centrifugation comprises liquid-liquid centrifugation.

31. The method of claim 29 or 30, wherein the centrifugation results in a crude oil light phase and an aqueous heavy phase.

32. The method of any one of claims 1-31, further comprising a decarboxylation step comprising evaporating the culture medium.

33. The method of claim 32, wherein the decarboxylation comprises evaporating the crude oil light phase.

34. The method of claim 32 or 33, wherein evaporating comprises one or more passes.

35. The method of claim 34, wherein evaporating comprises a first pass and a second pass.

36. The method of claim 35, wherein the first pass is performed at a temperature of between about 100 °C and about 500 °C.

37. The method of claim 36, wherein the first pass is performed at a temperature of about 180 °C.

38. The method of any one of claims 35-37, wherein the first pass is performed at a pressure of between about 0.5 torr and 760 torr.

39. The method of claim 38, wherein the first pass is performed at a pressure of about 1 torr.

40. The method of any one of claims 35-39, wherein the second pass is performed at a temperature of between 150 °C and 300 °C.

41. The method of claim 40, wherein the second pass is performed at a temperature of about 240 C.

42. The method of any one of claims 35-41, wherein the second pass is performed at a pressure of between about 0.5 torr and 760 torr.

43. The method of claim 42, wherein the second pass is performed at a pressure of about 1 torr.

44. The method of any one of claims 1-43, wherein purifying comprises one or more of a liquid-liquid extraction, chromatography, or saponification.

45. The method of any one of claims 1-44, wherein the CBCa synthase comprises one or more amino acid substitutions relative to the amino acid sequence of SEQ ID NO: 1.

46. The method of claim 45, wherein the one or more amino acid substitutions are at a residue selected from Q75, F82, T130, S140, V169, N240, V294, A299, K305, T335, R340, H354, L435, Y461, K535, S540, and T545 of SEQ ID NO: 1.

47. The method of any one of claims 1-46, wherein the CBCa synthase comprises an amino acid selected from SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, and SEQ ID NO: 67.

48. The method of any one of claims 1-47, wherein the CBCa synthase has an amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 1.

49. The method of claim 48, wherein the CBCa synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 1.

50. The method of any one of claims 1-49, wherein the host cell comprises one or more heterologous nucleic acids that each, independently, encode (a) an acyl activating enzyme (AAE), and/or (b) a tetraketide synthase (TKS), and/or (c) a cannabigerolic acid synthase (CBGaS), and/or (d) a geranyl pyrophosphate (GPP) synthase.

51. The method of any one of claims 1-50, wherein the host cell comprises heterologous nucleic acids that independently encode:

a) an AAE having the amino acid sequence of any one of SEQ ID NO: 68-91;

b) a TKS having the amino acid sequence of any one of SEQ ID NO: 92-126;

c) a CBGaS having the amino acid sequences of any one of SEQ ID NO: 127-131; and

d) a GPP synthase having the amino acid sequence of any one of SEQ ID NO: 132-137.

52. The method of claim 1-51, wherein the CBCa has a purity of at least about 50% (w/v) following purifying.

53. The method of claim 52, wherein the CBCa has a purity of between about 50% (w/w) and 100% (w/w) following purifying.

54. The method of any one of claims 1-53, wherein the host cell is a yeast cell or yeast strain.

55. The method of claim 54, wherein the yeast cell is *S. cerevisiae*.

56. A composition comprising CBCa, wherein the composition is produced by the method of any one of claims 1-55.

57. A variant CBCa synthase polypeptide comprising one or more amino acid substitutions relative to the amino acid sequence of SEQ ID NO: 1.

58. The variant polypeptide of claim 57, wherein the CBCa synthase comprises an amino acid substitution at a residue selected from Q75, F82, T130, S140, V169, N240, V294, A299, K305, T335, R340, H354, L435, Y461, K535, S540, and T545 of SEQ ID NO: 1.

59. The variant polypeptide of claim 57, wherein the CBCa synthase comprises an amino acid sequence selected from SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, and SEQ ID NO: 28. .

60. The variant polypeptide of claim 57, wherein the polypeptide comprises an amino acid sequence selected from SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ

ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, and SEQ ID NO: 67.

61. The variant polypeptide of any one of claims 57-60, wherein the CBCa synthase has an amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 1.

62. The variant polypeptide of claim 61, wherein the CBCa synthase has an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NO: 1.

63. A nucleic acid encoding the variant polypeptide of any one of claims 57-62.

64. A host cell comprising the variant polypeptide of any one of claims 57-62 or the nucleic acid of claim 63.

FIG. 1

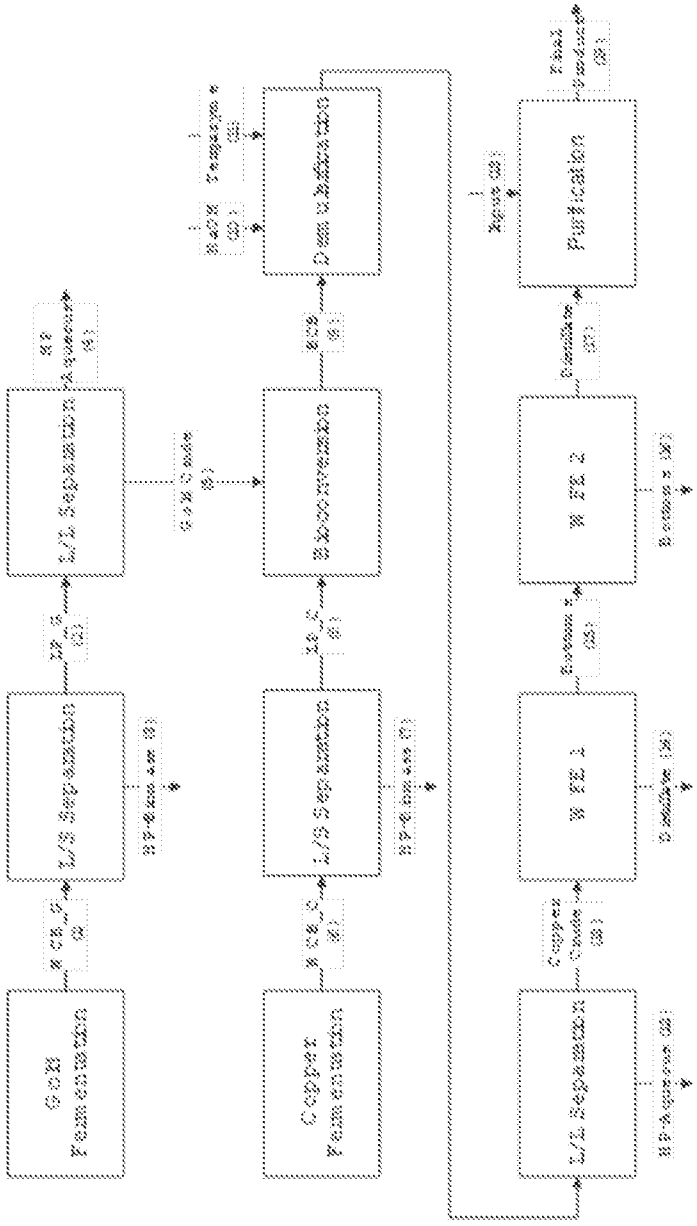


FIG. 2

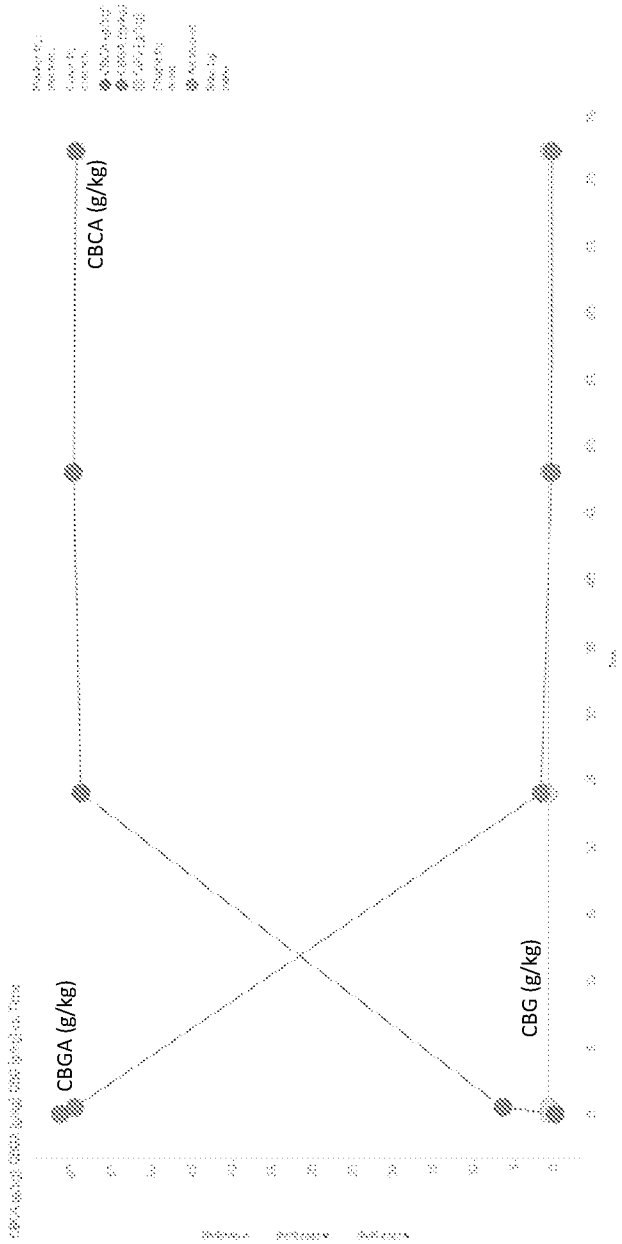


FIG. 3

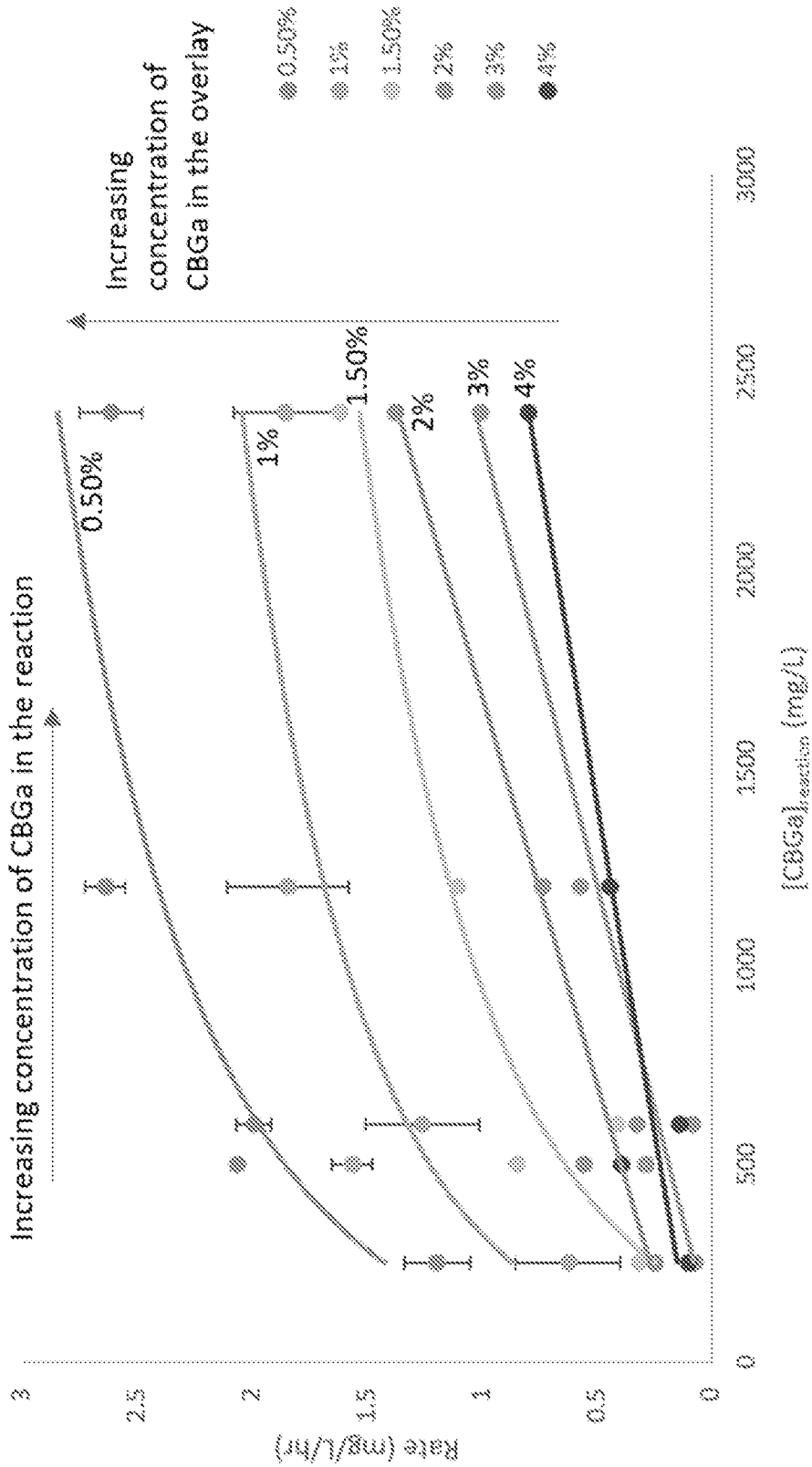

Increasing concentration of CBGa in the reaction
Increasing concentration of CBGa in the overlay
0.50%
1%
1.50%
2%
3%
4%
Rate (mg/L/hr)
$[CBGa]_{reaction}$ (mg/L)
0
0.5
1
1.5
2
2.5
3
500
1000
1500
2000
2500
3000

FIG. 4

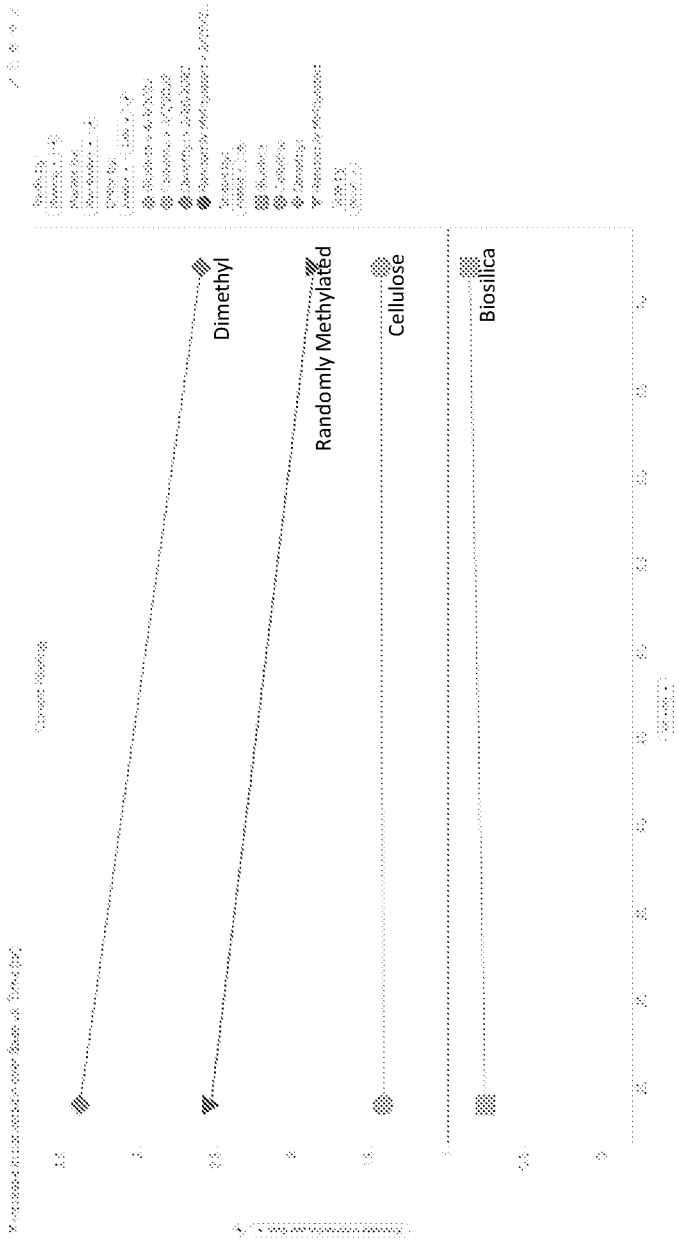


FIG. 5

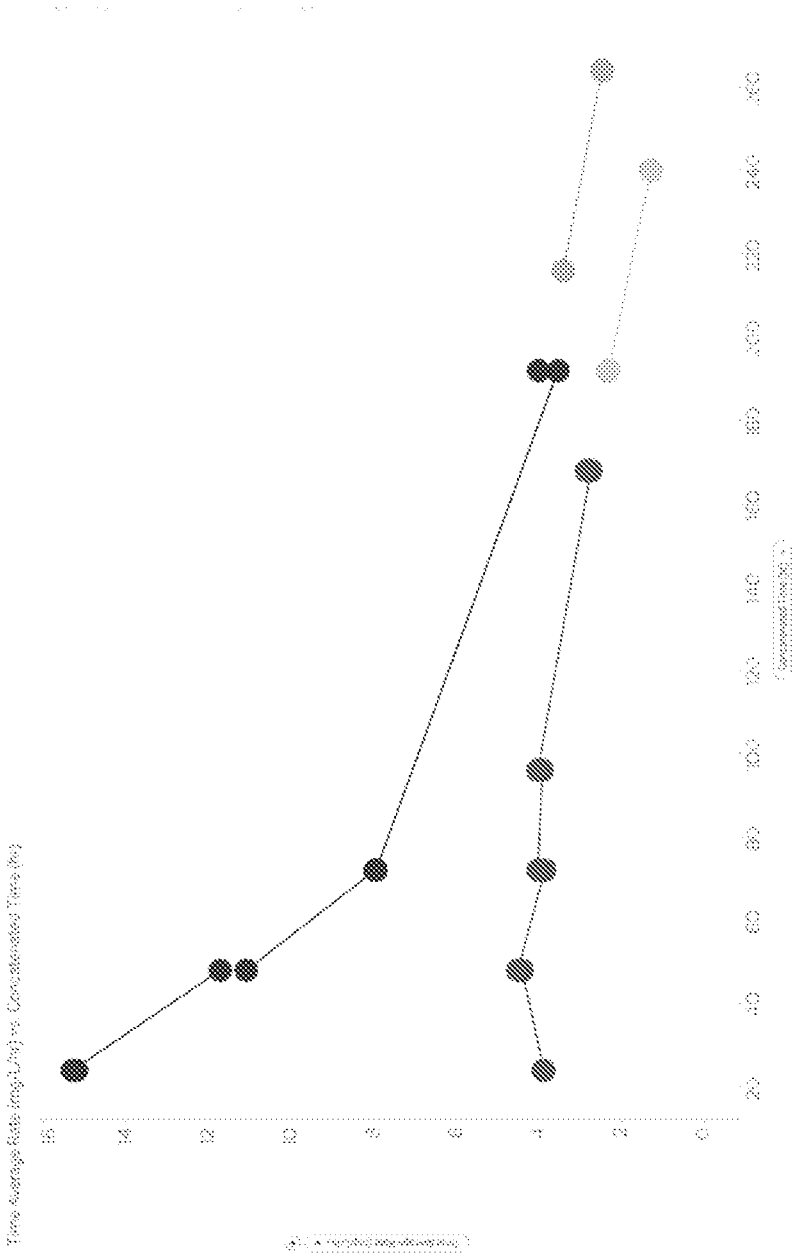


FIG. 6

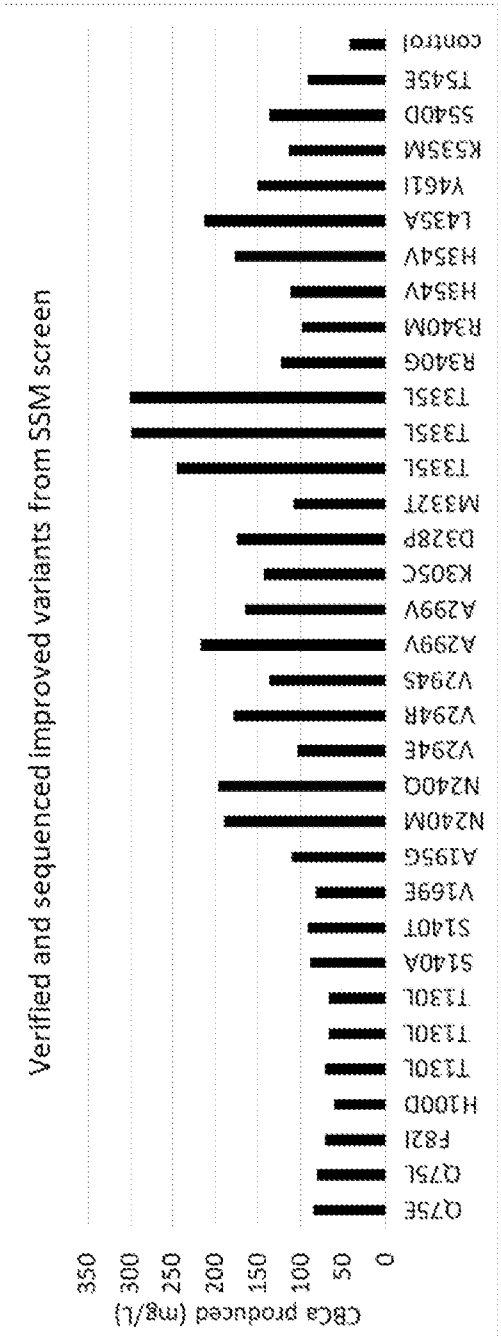
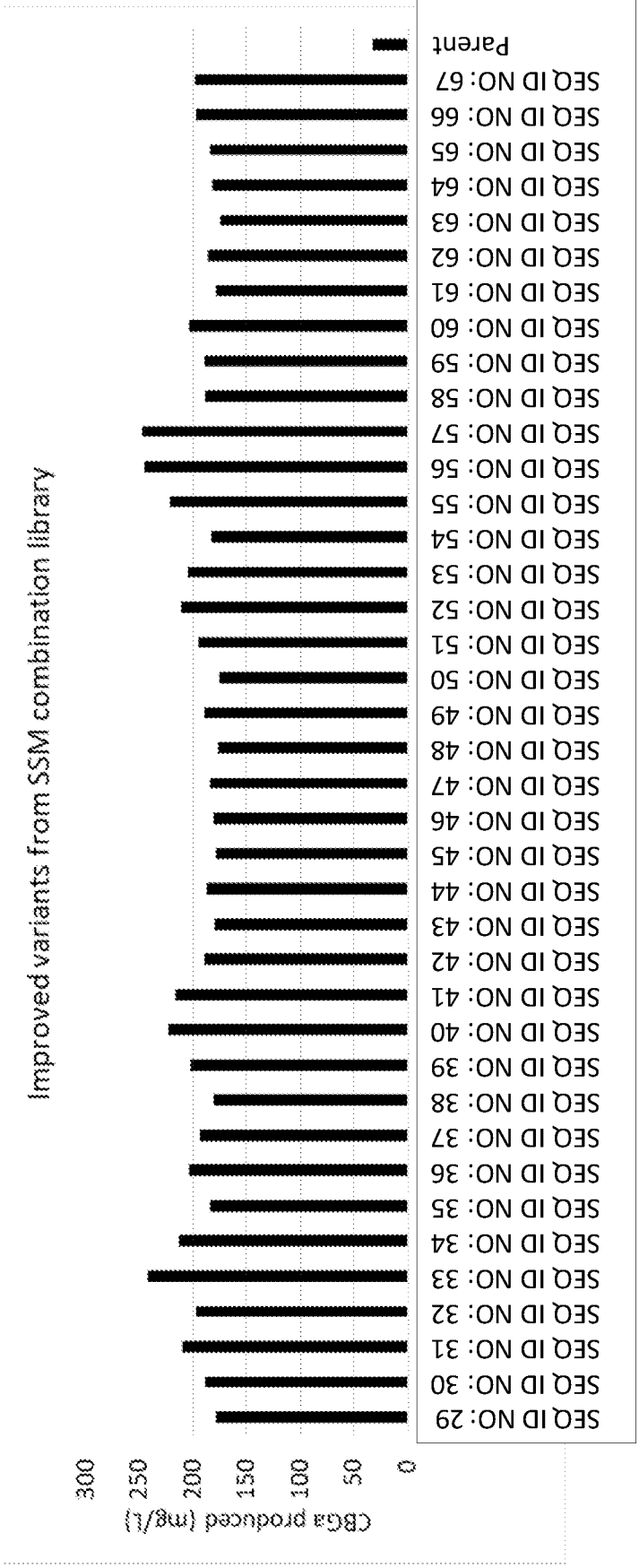

Verified and sequenced improved variants from SSM screen
CBCa produced (mg/L)
0
50
100
150
200
250
300
350
control
T545E
S540D
K535M
V461I
L435A
H354V
H354V
R340M
R340G
T335L
T335L
T335L
M332T
D328P
K305C
A299V
A299V
V294S
V294R
V294E
N240Q
N240M
A195G
V169E
S140T
S140A
T130L
T130L
T130L
H100D
F82I
Q75L
Q75E

FIG. 7


Improved variants from SSM combination library
CBGa produced (mg/L)
0
50
100
150
200
250
300
Parent
SEQ ID NO: 67
SEQ ID NO: 66
SEQ ID NO: 65
SEQ ID NO: 64
SEQ ID NO: 63
SEQ ID NO: 62
SEQ ID NO: 61
SEQ ID NO: 60
SEQ ID NO: 59
SEQ ID NO: 58
SEQ ID NO: 57
SEQ ID NO: 56
SEQ ID NO: 55
SEQ ID NO: 54
SEQ ID NO: 53
SEQ ID NO: 52
SEQ ID NO: 51
SEQ ID NO: 50
SEQ ID NO: 49
SEQ ID NO: 48
SEQ ID NO: 47
SEQ ID NO: 46
SEQ ID NO: 45
SEQ ID NO: 44
SEQ ID NO: 43
SEQ ID NO: 42
SEQ ID NO: 41
SEQ ID NO: 40
SEQ ID NO: 39
SEQ ID NO: 38
SEQ ID NO: 37
SEQ ID NO: 36
SEQ ID NO: 35
SEQ ID NO: 34
SEQ ID NO: 33
SEQ ID NO: 32
SEQ ID NO: 31
SEQ ID NO: 30
SEQ ID NO: 29