



(51) International Patent Classification:

C12N 9/10 (2006.01)

C12N 15/81 (2006.01)

(21) International Application Number:

PCT/US2023/066115

(22) International Filing Date:

24 April 2023 (24.04.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/334,307

25 April 2022 (25.04.2022)

US

(71) Applicant: GINKGO BIOWORKS, INC. [US/US]; 27 Drydock Avenue, 8th Floor, Boston, Massachusetts 02210 (US).

(72) Inventors: CARVALHO, Brian; c/o Ginkgo Bioworks, Inc., 27 Drydock Avenue, 8th Floor, Boston, Massachusetts 02210 (US). RITTER, Seth; c/o Ginkgo Bioworks, Inc., 27 Drydock Avenue, 8th Floor, Boston, Massachusetts 02210 (US). WRENBECK, Emily, E.; c/o Ginkgo Bioworks, Inc., 27 Drydock Avenue, 8th Floor, Boston, Massachusetts 02210 (US).

(74) Agent: LUO, Flora et al.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, Massachusetts 02210-2206 (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).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: BIOSYNTHESIS OF CANNABINOIDS AND CANNABINOID PRECURSORS

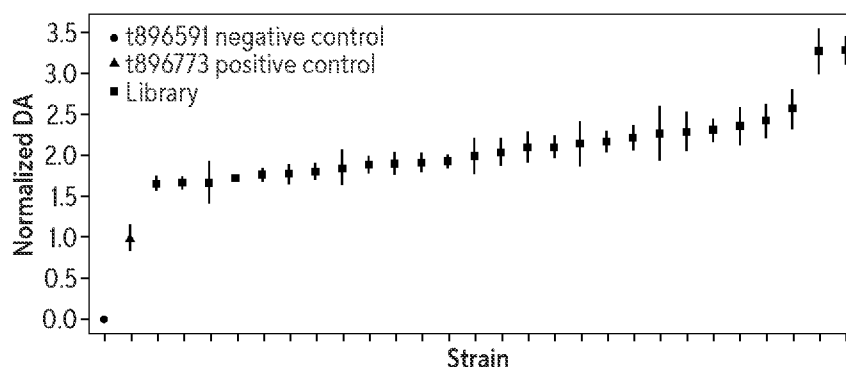


FIG. 6A

(57) Abstract: Aspects of the disclosure relate to biosynthesis of cannabinoids and cannabinoid precursors in recombinant cells and in vitro. Described are host cells comprising heterologous polynucleotides encoding polyketide synthase (PKS) variants that are capable of producing more divarinol in the presence of butyrate.

BIOSYNTHESIS OF CANNABINOIDS AND CANNABINOID PRECURSORS

CROSS-REFERENCE TO RELATED APPLICATIONS

[1] This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 63/334,307, filed April 25, 2022, entitled “BIOSYNTHESIS OF CANNABINOIDS AND CANNABINOID PRECURSORS,” the entire disclosure of which is hereby incorporated by reference in its entirety.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[2] The contents of the electronic sequence listing (G091970087WO00-SEQ-FL.xml; Size: 259,809 bytes; and Date of Creation: April 14, 2023) is herein incorporated by reference in its entirety.

FIELD OF INVENTION

[3] The present disclosure relates to the biosynthesis of cannabinoids and cannabinoid precursors, such as in recombinant cells.

BACKGROUND

[4] Cannabinoids are chemical compounds that may act as ligands for endocannabinoid receptors and have multiple medical applications. Traditionally, cannabinoids have been isolated from plants of the genus *Cannabis*. The use of plants for producing cannabinoids is inefficient, however, with isolated products often limited to the two most prevalent endogenous cannabinoids, THC and CBD, as other cannabinoids are typically produced in very low concentrations in *Cannabis* plants. Further, the cultivation of *Cannabis* plants is restricted in many jurisdictions. In addition, to obtain consistent results, *Cannabis* plants are often grown in a controlled environment, such as indoor grow rooms without windows, to provide flexibility in modulating growing conditions such as lighting, temperature, humidity, airflow, etc. Growing *Cannabis* plants in such controlled environments can result in high energy usage per gram of cannabinoid produced, especially for rare cannabinoids that the plants produce only in small amounts. For example, lighting in such grow rooms is provided by artificial sources, such as high-powered sodium lights. As many species of *Cannabis* have a vegetative cycle that requires 18 or more hours of light per day, powering such lights can result in significant energy expenditures. It has been estimated that between 0.88-1.34 kWh of energy is required to produce one gram of THC in dried *Cannabis* flower form (e.g., before

any extraction or purification). Additionally, concern has been raised over agricultural practices in certain jurisdictions, such as California, where the growing season coincides with the dry season such that the water usage may impact connected surface water in streams (Dillis, Christopher, Connor McIntee, Van Butsic, Lance Le, Kason Grady, and Theodore Grantham. "Water storage and irrigation practices for cannabis drive seasonal patterns of water extraction and use in Northern California." *Journal of Environmental Management* 272 (2020): 110955). See, also, Summers, H.M., Sproul, E. & Quinn, J.C. The greenhouse gas emissions of indoor cannabis production in the United States. *Nat Sustain* 4, 644–650 (2021).; and Zheng, Z., Fiddes, K. & Yang, L. A narrative review on environmental impacts of cannabis cultivation. *J Cannabis Res* 3, 35 (2021).

[5] Cannabinoids can be produced through chemical synthesis (see, *e.g.*, U.S. Patent No. 7,323,576 to Souza et al). However, such methods suffer from low yields and high cost. Production of cannabinoids, cannabinoid analogs, and cannabinoid precursors using engineered organisms may provide an advantageous approach to meet the increasing demand for these compounds.

SUMMARY

[6] Aspects of the present disclosure provide methods for production of cannabinoids and cannabinoid precursors from fatty acid substrates using genetically modified host cells. In some embodiments, the host cell comprises a heterologous polynucleotide encoding a polyketide synthase (PKS), wherein the host cell is capable of producing more divarinol in the presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.

[7] In some embodiments, the PKS comprises: an amino acid substitution at a residue corresponding to position 3, 4, 7, 11, 21, 24, 38, 44, 62, 66, 69, 72, 76, 77, 78, 79, 101, 103, 104, 124, 136, 137, 144, 150, 166, 225, 237, 262, 301, 319, and/or 385 in SEQ ID NO: 1; and a contiguous stretch of residues with amino acids unmutated relative to SEQ ID NO: 1 at positions corresponding to: positions 12-20 in SEQ ID NO: 1; positions 26-37 in SEQ ID NO: 1; positions 39-43 in SEQ ID NO: 1; positions 45-50 in SEQ ID NO: 1; positions 52-61 in SEQ ID NO: 1; positions 83-91 in SEQ ID NO: 1; positions 93-100 in SEQ ID NO: 1; positions 105-122 in SEQ ID NO: 1; positions 129-135 in SEQ ID NO: 1; positions 145-149 in SEQ ID NO: 1; positions 151-165 in SEQ ID NO: 1; positions 167-189 in SEQ ID NO: 1; positions

191-224 in SEQ ID NO: 1; positions 226-236 in SEQ ID NO: 1; positions 238-246 in SEQ ID NO: 1; positions 248-261 in SEQ ID NO: 1; positions 268-277 in SEQ ID NO: 1; positions 279-295 in SEQ ID NO: 1; positions 302-311 in SEQ ID NO: 1; positions 320-327 in SEQ ID NO: 1; positions 336-374 in SEQ ID NO: 1; and/or positions 376-383 in SEQ ID NO: 1.

[8] Aspects of the present disclosure provide host cells comprising a heterologous polynucleotide encoding a polyketide synthase (PKS), wherein the PKS comprises an amino acid substitution at a residue corresponding to position 3, 4, 7, 11, 21, 24, 38, 44, 62, 66, 69, 72, 76, 77, 78, 79, 101, 103, 104, 124, 136, 137, 144, 150, 166, 225, 237, 262, 301, 319, and/or 385 in SEQ ID NO: 1, and wherein the PKS does not comprise any of SEQ ID NOs: 91-94.

[9] In some embodiments, the PKS comprises one or more of the following substitutions relative to SEQ ID NO: 1: H3R or H3A; L4Q; E7P; S11T; E21P; L24I; K38N; Q44D; N62H; N66T; H69I; Q72E; L76M; V77C; E78T; H79Y; D101E; C103A; A104E; S124T; H136Q; C137L; S144P; V150T; I166L; V225A; G237A; H262S; K301P; V319R, V319K, or V319A; and/or Y385S, Y385T, or Y385N.

[10] In some embodiments, the PKS further comprises an amino acid substitution at a residue corresponding to position 25, 51, 63, 64, 75, 81, 82, 92, 123, 125, 128, 190, 247, 267, 278, 296, 312, 315, 316, 328, 334, 335, 375, and/or 384 in SEQ ID NO: 1.

[11] In some embodiments, the PKS comprises one or more of the following amino acid substitutions relative to SEQ ID NO: 1: L25C, L25E, L25F, L25G, L25H, L25K, L25M, L25N, L25P, L25Q, L25R, L25S, L25T, L25V, L25W, or L25Y; K51R; C63M, C63T, or C63F; F64Y; R75N, R75S, or R75G; M81D; Q82P; V92K or V92A; T123C or T123S; A125S; T128V; L190C; T247A; M267G; I278K or I278E; T296A; H312N or H312G; S315K or S315P; D316E; H328Y; S334P; T335C; R375S or R375T; and/or K384N.

[12] In some embodiments, the PKS does not comprise the amino acid I at the residue corresponding to position 25 in SEQ ID NO: 1.

[13] In some embodiments, the PKS comprises the amino acid glutamate (E), glycine (G), histidine (H), lysine (K), leucine (L), methionine (M), asparagine (N), proline (P),

glutamine (Q), arginine (R), serine (S), threonine (T), valine (V), tryptophan (W), or a tyrosine (Y) at the residue corresponding to position 25 in SEQ ID NO: 1.

[14] In some embodiments, the PKS comprises: the amino acid cysteine (C) at a residue corresponding to position 157 in SEQ ID NO: 1; the amino acid phenylalanine (F) at a residue corresponding to position 208 in SEQ ID NO: 1; the amino acid histidine (H) at a residue corresponding to position 297 in SEQ ID NO: 1; and/or the amino acid asparagine (N) at a residue corresponding to position 330 in SEQ ID NO: 1.

[15] In some embodiments, the PKS comprises the following amino acid substitutions relative to SEQ ID NO: 1: R75N, D101E, V319R, H328Y, and T335C; C63M, H69I, R75G, V77C, M267G, H312G, V319K, T335C, R375T, and Y385S; C63M, F64Y, S124T, H328Y, and T335C; K51R, H69I, R75N, V77C, D101E, C103A, T128V, K301P, H312G, and T335C; C63T, Q82P, D101E, C103A, and T123S; L25I, K51R, R75N, V77C, D101E, C103A, T128V, K301P, H312G, and T335C; K51R, R75N, V77C, D101E, C103A, T128V, K301P, H312G, D316E, and T335C; R75N, V77C, H79Y, D101E, C103A, T128V, K301P, H312G, D316E, and T335C; K51R, N66T, R75S, D101E, T123C, T296A, S315P, H328Y, T335C, and R375T; K51R, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, D316E, and T335C; H3R, S11T, K38N, Q44D, K51R, H69I, Q72E, R75N, L76M, V77C, E78T, H79Y, M81D, V92K, D101E, C103A, A104E, T128V, C137L, S144P, V150T, V225A, H262S, I278K, H312G, S315P, D316E, S334P, and K384N; K51R, R75N, V77C, H79Y, D101E, C103A, T128V, K301P, H312G, and D316E; L25I, K51R, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, and T335C; L4Q, E21P, N62H, H312N, and H328Y; L25I, H69I, R75N, V77C, H79Y, D101E, C103A, T128V, K301P, and H312G; H3A, L25Y, C63M, R75N, H136Q, M267G, H328Y, T335C, R375T, and Y385T; K51R, H69I, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, and T335C; D101E, T123C, H136Q, H312G, and T335C; H3R, E7P, S11T, K38N, Q44D, K51R, H69I, Q72E, R75N, L76M, V77C, E78T, H79Y, M81D, V92A, D101E, C103A, A104E, T128V, C137L, S144P, V150T, V225A, H262S, I278E, H312G, S315P, D316E, S334P, and K384N; K51R, R75N, V77C, H79Y, D101E, C103A, T128V, H312G, D316E, and T335C; H3A, L4Q, C63F, R75N, D101E, C103A, C137L, G237A, T247A, and Y385N; L4Q, L24I, Q44D, K51R, C63F, R75S, V77C, C103A, I166L, and V319A; L4Q, R75S, V319R, T335C, and R375S; R75S, V77C, S124T, T335C, and R375T; or H3A, R75S, D101E, S315K, and T335C.

[16] In some embodiments, the PKS comprises a sequence that is at least 90% identical to a sequence selected from SEQ ID NOs: 2-28. In some embodiments, the PKS comprises a sequence selected from SEQ ID NOs: 2-28, or a conservatively substituted version thereof.

[17] In some embodiments, the heterologous polynucleotide comprises a sequence that is at least 90% identical to a sequence selected from SEQ ID NOs: 30-56. In some embodiments, the heterologous polynucleotide comprises a sequence selected from SEQ ID NOs: 30-56.

[18] In some embodiments, the PKS comprises the following amino acid substitutions relative to SEQ ID NO: 1: K51R, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, and T335C.

[19] In some embodiments, the PKS comprises an amino acid substitution relative to SEQ ID NO: 1 selected from the group consisting of: L25E; L25G; L25K; L25M; L25N; L25P; L25Q; L25S; L25V; and L25Y.

[20] In some embodiments, the PKS comprises the following amino acid substitutions relative to SEQ ID NO: 1: K51R, R75N, V77C, D101E, C103A, T128V, K301P, H312G, and T335C.

[21] In some embodiments, the PKS comprises an amino acid substitution relative to SEQ ID NO: 1 selected from the group consisting of: L25E; L25H; L25K; L25M; L25N; L25P; L25R; L25T; L25W; and L25Y.

[22] In some embodiments, the PKS comprises a sequence that is at least 90% identical to a sequence selected from SEQ ID NOs: 104-130. In some embodiments, the PKS comprises a sequence selected from SEQ ID NOs: 104-130.

[23] In some embodiments, the heterologous polynucleotide encoding the PKS comprises a sequence that is at least 90% identical to a sequence selected from SEQ ID NOs: 131-157. In some embodiments, the heterologous polynucleotide encoding the PKS comprises a sequence selected from SEQ ID NOs: 131-157.

[24] In some embodiments, the heterologous polynucleotide is integrated into the genome of the host cell.

[25] In some embodiments, the host cell is a plant cell, an algal cell, a yeast cell, a bacterial cell, or an animal cell. In some embodiments, the host cell is a yeast cell. In some embodiments, the yeast cell is a *Saccharomyces* cell, a *Yarrowia* cell, a *Komagataella* cell, or a *Pichia* cell. In some embodiments, the *Saccharomyces* cell is a *Saccharomyces cerevisiae* cell. In some embodiments, the yeast cell is a *Yarrowia* cell. In some embodiments, the host cell is a bacterial cell. In some embodiments, the bacterial cell is an *E. coli* cell.

[26] In some embodiments, the PKS is an olivetol synthase (OLS) or a divarinol synthase.

[27] In some embodiments, the host cell further comprises one or more heterologous polynucleotides encoding one or more of: an acyl activating enzyme (AAE), a polyketide cyclase (PKC), a prenyltransferase (PT) and/or a terminal synthase (TS).

[28] In some embodiments, the PKC is an olivetolic acid cyclase (OAC) or divaric acid cyclase.

[29] In some embodiments, the PKC comprises a sequence that is at least 90% identical to SEQ ID NO: 61.

[30] In some embodiments, the PKC comprises SEQ ID NO: 61.

[31] In some embodiments, the host cell is capable of producing divaric acid in the presence of butyrate.

[32] In some embodiments, the host cell is capable of producing divarinol in the presence of butyrate.

[33] In some embodiments, the host cell is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17-fold more divaric acid in the presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1. In some embodiments, the host cell is capable of producing at least 500 µg/L divaric acid. In some embodiments, the host cell is capable of producing at least 700 µg/L divaric acid.

[34] In some embodiments, the host cell is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, or 9-fold more divarinol in the presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1. In some embodiments, the host cell is capable of producing at least 25,000 µg/L divarinol. In some embodiments, the host cell is capable of producing at least 30,000 µg/L divarinol. In some embodiments, the host cell is capable of producing at least 33,000 µg/L divarinol.

[35] In some embodiments, the host cell is capable of producing a varinolic cannabinoid. In some embodiments, the varinolic cannabinoid is CBGV, CBGVA, THCVA, CBDVA and/or CBCVA.

[36] In some embodiments, the host cell is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, or 9-fold more of the varinolic cannabinoid than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.

[37] In some embodiments, the AAE comprises a sequence that is at least 90% identical to SEQ ID NO: 86.

[38] In some embodiments, the AAE comprises SEQ ID NO: 86.

[39] In some embodiments, the heterologous polynucleotide encoding the AAE comprises a sequence that is at least 90% identical to SEQ ID NO: 87.

[40] In some embodiments, the heterologous polynucleotide encoding the AAE comprises SEQ ID NO: 87.

[41] In some embodiments, the PT comprises a sequence that is at least 90% identical to any one of SEQ ID NOs: 74, 76, 78, 80, 102, or 158.

[42] In some embodiments, the PT comprises the sequence of any one of SEQ ID NOs: 74, 76, 78, 80, 102, or 158.

[43] In some embodiments, the TS comprises a sequence that is at least 90% identical to SEQ ID NO: 84, 88, 95, 98, 100, or 161.

[44] In some embodiments, the TS comprises the sequence of SEQ ID NO: 84, 88, 95, 98, 100, or 161.

[45] Further aspects of the present disclosure provide host cells comprising a heterologous polynucleotide encoding a polyketide synthase (PKS), wherein the PKS comprises a sequence selected from SEQ ID NOs: 2-28 and 104-130 or a conservatively substituted version thereof.

[46] Further aspects of the present disclosure provide host cells comprising a heterologous polynucleotide encoding a polyketide synthase (PKS), wherein the heterologous polynucleotide comprises a sequence selected from SEQ ID NOs: 30-56 and 131-157.

[47] Further aspects of the present disclosure provide methods comprising culturing any of the host cells of the disclosure.

[48] Further aspects of the present disclosure provide methods of synthesizing divarinol or divaric acid comprising contacting butyryl-CoA with a polyketide synthase (PKS), wherein the PKS comprises an amino acid selected from any one of SEQ ID NOs: 2-28 and 104-130.

[49] Further aspects of the present disclosure provide methods of producing divaric acid (DA) comprising contacting butyrate and Coenzyme A (CoA) with: an acyl activating enzyme (AAE), a polyketide synthase (PKS), and a polyketide cyclase (PKC); or an acyl activating enzyme (AAE) and a PKS, wherein the PKS comprises the sequence of any one of SEQ ID NOs: 2-28 and 104-130.

[50] Further aspects of the present disclosure provide methods of producing divarinol (DL) comprising contacting butyrate with an acyl activating enzyme (AAE) and a polyketide synthase (PKS), wherein the PKS comprises the sequence of any one of SEQ ID NOs: 2-28 and 104-130.

[51] In some embodiments, the method occurs *in vitro*.

[52] In some embodiments, the method occurs within a host cell that expresses a heterologous polynucleotide encoding a polyketide synthase (PKS) that comprises an amino acid sequence selected from any one of SEQ ID NOs: 2-28 and 104-130.

[53] Further aspects of the disclosure provide methods of producing a cannabinoid compound or a cannabinoid precursor comprising culturing a host cell in the presence of butyrate, wherein the host cell comprises a heterologous polynucleotide encoding a polyketide synthase (PKS), and wherein the PKS comprises an amino acid sequence selected from any one of SEQ ID NOs: 2-28 and 104-130.

[54] In some embodiments, the host cell further comprises one or more heterologous polynucleotides encoding an acyl activating enzyme (AAE).

[55] In some embodiments, the AAE comprises a sequence that is at least 90% identical to the sequence of SEQ ID NO: 86.

[56] In some embodiments, the heterologous polynucleotide encoding the AAE comprises a sequence that is at least 90% identical to the sequence of SEQ ID NO: 87.

[57] In some embodiments, the heterologous polynucleotide comprises a sequence that is at least 90% identical to the sequence of any one of SEQ ID NOs: 30-56 and 131-157. In some embodiments, the heterologous polynucleotide comprises the sequence of any one of SEQ ID NOs: 30-56 and 131-157.

[58] In some embodiments, the heterologous polynucleotide is integrated into the genome of the host cell.

[59] In some embodiments, the host cell is a plant cell, an algal cell, a yeast cell, a bacterial cell, or an animal cell. In some embodiments, the host cell is a yeast cell. In some embodiments, the yeast cell is a *Saccharomyces* cell, a *Yarrowia* cell, a *Komagataella* cell, or a *Pichia* cell. In some embodiments, the *Saccharomyces* cell is a *Saccharomyces cerevisiae* cell. In some embodiments, the yeast cell is a *Yarrowia* cell. In some embodiments, the host cell is a bacterial cell. In some embodiments, the bacterial cell is an *E. coli* cell.

[60] In some embodiments, the AAE is capable of activating a short chain fatty acid in the presence of Coenzyme A (CoA). In some embodiments, the short chain fatty acid is a four-carbon fatty acid. In some embodiments, the four-carbon fatty acid is butyrate and the AAE is capable of catalyzing the production of butyryl-CoA from butyrate.

[61] In some embodiments, the host cell further comprises one or more heterologous polynucleotides encoding one or more of: a polyketide cyclase (PKC), a prenyltransferase (PT) and/or a terminal synthase (TS).

[62] In some embodiments, the PKS is an olivetol synthase (OLS) or a divarinol synthase.

[63] In some embodiments, the PKS is a divarinol synthase.

[64] In some embodiments, the PKS is a bifunctional PKS and is capable of producing divaric acid.

[65] In some embodiments, the PKC is an olivetolic acid cyclase (OAC) or a divaric acid cyclase.

[66] In some embodiments, the PKC comprises a sequence that is at least 90% identical to SEQ ID NO: 61.

[67] In some embodiments, the PKC comprises SEQ ID NO: 61.

[68] In some embodiments, the host cell is capable of producing divaric acid.

[69] In some embodiments, the host cell is capable of producing divarinol.

[70] In some embodiments, the host cell is capable of producing at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17 fold more divaric acid in the presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1. In some embodiments, the host cell is capable of producing at least 500 µg/L divaric acid. In some embodiments, the host cell is capable of producing at least 700 µg/L divaric acid.

[71] In some embodiments, the host cell is capable of producing at least 2, 3, 4, 5, 6, 7, 8, or 9 fold more divarinol in the presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1. In some embodiments, the host cell is capable of producing at least 25,000 µg/L divarinol. In some embodiments, the host cell is capable of producing at least 30,000 µg/L

divarinol. In some embodiments, the host cell is capable of producing at least 33,000 µg/L divarinol.

[72] In some embodiments, the host cell is capable of producing a varinolic cannabinoid. In some embodiments, the varinolic cannabinoid is CBGV, CBGVA, THCVA and/or CBCVA.

[73] In some embodiments, the host cell is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, or 9-fold more of the varinolic cannabinoid than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.

[74] In some embodiments, the PT comprises a sequence that is at least 90% identical to any one of SEQ ID NOs: 74, 76, 78, 80, 102, or 158. In some embodiments, the PT comprises the sequence of any one of SEQ ID NOs: 74, 76, 78, 80, 102, or 158.

[75] In some embodiments, the TS comprises a sequence that is at least 90% identical to SEQ ID NO: 84, 88, 95, 98, 100, or 161. In some embodiments, the TS comprises the sequence of SEQ ID NO: 84, 88, 95, 98, 100, or 161.

[76] Further aspects of the present disclosure provide non-naturally occurring polynucleotides encoding a polyketide synthase (PKS), wherein the polynucleotide comprises a sequence that is at least 90% identical to the sequence of any one of SEQ ID NOs: 30-56 and 131-157. In some embodiments, the polynucleotide comprises the sequence of any one of SEQ ID NOs: 30-56 and 131-157.

[77] Further aspects of the present disclosure provide vectors comprising any of the polynucleotide sequences of the disclosure.

[78] Further aspects of the present disclosure provide expression cassettes comprising any of the polynucleotide sequences of the disclosure.

[79] Further aspects of the present disclosure provide host cells transformed with the any of the polynucleotides of the disclosure, any of the vectors of the disclosure, or any of the expression cassettes of the disclosure.

[80] Further aspects of the present disclosure provide bioreactors for producing a cannabinoid compound or a cannabinoid precursor, wherein the bioreactor contains a polyketide synthase (PKS), and wherein the PKS comprises an amino acid sequence selected from any one of SEQ ID NOs: 2-28 and 104-130.

[81] In some embodiments, the bioreactor further comprises one or more of an acyl activating enzyme (AAE), a polyketide cyclase (PKC), a prenyltransferase (PT) and/or a terminal synthase (TS).

[82] In some embodiments, the bioreactor contains butyrate.

[83] In some embodiments, the bioreactor produces a varinolic cannabinoid. In some embodiments, the varinolic cannabinoid is CBGV, CBGVA, THCVA, CBDVA and/or CBCVA.

[84] In some embodiments, the bioreactor is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, or 9-fold more of the varinolic cannabinoid than a bioreactor comprising a control PKS that comprises the sequence of SEQ ID NO: 1.

[85] Each of the limitations of the invention can encompass various embodiments of the invention. It is, therefore, anticipated that each of the limitations of the invention involving any one element or combinations of elements can be included in each aspect of the invention. This disclosure is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced or of being carried out in various ways. Also, the phraseology and terminology used in this application is for the purpose of description and should not be regarded as limiting. The use of “including,” “comprising,” or “having,” “containing,” “involving,” and variations thereof, is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

BRIEF DESCRIPTION OF DRAWINGS

[86] The accompanying drawings are not intended to be drawn to scale. In the drawings, each identical or nearly identical component that is illustrated in various figures is represented by a like numeral. For purposes of clarity, not every component may be labeled in every drawing. In the drawings:

[87] **FIG. 1** is a schematic depicting the native *Cannabis* biosynthetic pathway for production of cannabinoid compounds, including five enzymatic steps mediated by: (R1a) acyl activating enzymes (AAE); (R2a) olivetol synthase enzymes (OLS); (R3a) olivetolic acid cyclase enzymes (OAC); (R4a) cannabigerolic acid synthase enzymes (CBGAS); and (R5a) terminal synthase enzymes (TS). Formulae 1a-11a correspond to hexanoic acid (1a), hexanoyl-CoA (2a), malonyl-CoA (3a), 3,5,7-trioxododecanoyl-CoA (4a), olivetol (5a), olivetolic acid (6a), geranyl pyrophosphate (7a), cannabigerolic acid (8a), cannabidiolic acid (9a), tetrahydrocannabinolic acid (10a), and cannabichromenic acid (11a). Hexanoic acid is an exemplary carboxylic acid substrate; other carboxylic acids may also be used (*e.g.*, butyric acid, isovaleric acid, octanoic acid, decanoic acid, *etc.*; see *e.g.*, **FIG. 3** below). The enzymes that catalyze the synthesis of 3,5,7-trioxododecanoyl-CoA and olivetolic acid are shown in R2a and R3a, respectively, and can include multi-functional enzymes that catalyze the synthesis of 3,5,7-trioxododecanoyl-CoA and olivetolic acid. The enzymes cannabidiolic acid synthase (CBDAS), tetrahydrocannabinolic acid synthase (THCAS), and cannabichromenic acid synthase (CBCAS) that catalyze the synthesis of cannabidiolic acid, tetrahydrocannabinolic acid, and cannabichromenic acid, respectively, and the varinolic cannabinoids cannabigerovarinic acid, tetrahydrocannabivarinic acid, and cannabichromevarinic acid, respectively, are shown in step R5a. The varinolic cannabinoids are not shown. **FIG. 1** is adapted from Carvalho et al. "Designing Microorganisms for Heterologous Biosynthesis of Cannabinoids" (2017) *FEMS Yeast Research* Jun 1;17(4), which is incorporated by reference in its entirety.

[88] **FIG. 2** is a schematic depicting a heterologous biosynthetic pathway for production of cannabinoid compounds, including five enzymatic steps mediated by: (R1) acyl activating enzymes (AAE); (R2) polyketide synthase enzymes (PKS) or bifunctional polyketide synthase-polyketide cyclase enzymes (PKS-PKC); (R3) polyketide cyclase enzymes (PKC) or bifunctional PKS-PKC enzymes; (R4) prenyltransferase enzymes (PT); and (R5) terminal synthase enzymes (TS). Any carboxylic acid of varying chain lengths, structures (*e.g.*, aliphatic, alicyclic, or aromatic) and functionalization (*e.g.*, hydroxylic-, keto-, amino-, thiol-, aryl-, or alogeno-) may also be used as precursor substrates (*e.g.*, thiopropionic acid, hydroxy phenyl acetic acid, norleucine, bromodecanoic acid, butyric acid, isovaleric acid, octanoic acid, decanoic acid, *etc.*).

[89] **FIG. 3** is a non-exclusive representation of select putative precursors for the cannabinoid pathway in **FIG. 2**.

[90] **FIG. 4** is a schematic depicting the biosynthetic pathway for production of varin cannabinoid compounds, including five enzymatic steps mediated by: (R1) acyl activating enzymes (AAE); (R2) polyketide synthase enzymes (PKS) or bifunctional polyketide synthase-polyketide cyclase enzymes (PKS-PKC); (R3) polyketide cyclase enzymes (PKC) or bifunctional PKS-PKC enzymes; (R4) prenyltransferase enzymes (PT); and (R5) terminal synthase enzymes (TS). The compounds of Formulae 1b-11b correspond to butyric acid (1b), butyryl-CoA (2b), malonyl-CoA (3b), 3,5,7-trioxodecanoyl-CoA (4b), divarinol (5b), divaric acid (6b), geranyl pyrophosphate (7b), cannabigerovarinic acid (8b), cannabidivarinic acid (9b), tetrahydrocannabivarinic acid (10b), and cannabichromevarinic acid (11b). Butyric acid is an exemplary carboxylic acid substrate; other carboxylic acids may also be used (*e.g.*, hexanoic acid, isovaleric acid, octanoic acid, decanoic acid, etc.; see *e.g.*, **FIG. 3** above). The enzymes that catalyze the synthesis of 3,5,7-trioxodecanoyl-CoA and divaric acid are shown in R2 and R3, respectively, and can include multi-functional enzymes that catalyze the synthesis of 3,5,7-trioxodecanoyl-CoA and divaric acid. The enzymes cannabidivarinic acid synthase (CBDVAS), tetrahydrocannabivarinic acid synthase (THCVAS), and cannabichromevarinic acid synthase (CBCVAS) that catalyze the synthesis of the varinolic cannabinoids cannabigerovarinic acid, tetrahydrocannabivarinic acid, and cannabichromevarinic acid, respectively, and their catalytic function thereof, are shown in step R5.

[91] **FIG. 5** is a schematic showing a plasmid used to express polyketide synthases (PKSs) in *S. cerevisiae*. The coding sequence for the PKS enzymes (labeled "Library gene") was driven by the GAL1 promoter. The plasmid contains markers for both yeast (URA3) and bacteria (ampR), as well as origins of replication for yeast (2 micron), and bacteria (pBR322).

[92] **FIGs. 6A-6B** depict results from a secondary screen of PKS mutants described in Example 1. **FIG. 6A** depicts divaric acid (DA) production. DA production for each test strain was normalized to DA production by positive control strain t896773, expressing a PKS with the amino acid sequence of SEQ ID NO: 1. Results for positive control strain t896773 and negative control strain t896591, expressing GFP are also shown. **FIG. 6B** depicts normalized divarinol (DL) production. DL production for each test strain was normalized to

DL production by positive control strain t896773, expressing a PKS with the amino acid sequence of SEQ ID NO: 1. Results for positive control strain t896773 and negative control strain t896591, expressing GFP are also shown.

[93] **FIGs. 7A-7B** depict results from a site saturation mutagenesis at residue 25 of synthetic polyketide synthase (PKS) mutants described in Example 2. **FIG. 7A** depicts cannabigerovarinic acid (CBGVA) production by strains harboring PKSs with SEQ ID NO: 7 as a template for mutation. Strain t1151829, expressing SEQ ID NO: 7, was used as a positive control. **FIG. 7B** depicts CBGVA production by strains harboring PKSs with SEQ ID NO: 14 as a template for mutation. Strain t1151830, expressing SEQ ID NO: 14, was used as a positive control.

DETAILED DESCRIPTION

[94] This disclosure provides methods for production of cannabinoids and cannabinoid precursors from fatty acid substrates using genetically modified host cells. The application describes PKSs that can be functionally expressed in host cells such as *S. cerevisiae*. As demonstrated in the Examples, multiple PKSs were identified that were active on butyryl-CoA in a host cell. The PKSs described in this disclosure may be useful in producing cannabinoids, and, in particular, for producing varinolic cannabinoids, such as, for example, CBGV, CBGVA, THCVA and CBCVA.

Definitions

[95] While the following terms are believed to be well understood by one of ordinary skill in the art, the following definitions are set forth to facilitate explanation of the disclosed subject matter.

[96] The term “a” or “an” refers to one or more of an entity, *i.e.*, can identify a referent as plural. Thus, the terms “a” or “an,” “one or more” and “at least one” are used interchangeably in this application. In addition, reference to “an element” by the indefinite article “a” or “an” does not exclude the possibility that more than one of the elements is present, unless the context clearly requires that there is one and only one of the elements.

[97] The terms “microorganism” or “microbe” should be taken broadly. These terms are used interchangeably and include, but are not limited to, the two prokaryotic domains,

Bacteria and Archaea, as well as certain eukaryotic fungi and protists. In some embodiments, the disclosure may refer to the “microorganisms” or “microbes” of lists/tables and figures present in the disclosure. This characterization can refer to not only the identified taxonomic genera of the tables and figures, but also the identified taxonomic species, as well as the various novel and newly identified or designed strains of any organism in the tables or figures. The same characterization holds true for the recitation of these terms in other parts of the specification, such as in the Examples.

[98] The term “prokaryotes” is recognized in the art and refers to cells that contain no nucleus or other cell organelles. The prokaryotes are generally classified in one of two domains, the Bacteria and the Archaea.

[99] “Bacteria” or “eubacteria” refers to a domain of prokaryotic organisms. Bacteria include at least 11 distinct groups as follows: (1) Gram-positive (gram+) bacteria, of which there are two major subdivisions: (a) high G+C group (*Actinomycetes*, *Mycobacteria*, *Micrococcus*, others) and (b) low G+C group (*Bacillus*, *Clostridia*, *Lactobacillus*, *Staphylococci*, *Streptococci*, *Mycoplasmas*); (2) Proteobacteria, e.g., Purple photosynthetic+non-photosynthetic Gram-negative bacteria (includes most “common” Gram-negative bacteria); (3) Cyanobacteria, e.g., oxygenic phototrophs; (4) Spirochetes and related species; (5) Planctomyces; (6) *Bacteroides*, *Flavobacteria*; (7) *Chlamydia*; (8) Green sulfur bacteria; (9) Green non-sulfur bacteria (also anaerobic phototrophs); (10) Radioresistant micrococci and relatives; and (11) *Thermotoga* and *Thermosiphon thermophiles*.

[100] The term “Archaea” refers to a taxonomic classification of prokaryotic organisms with certain properties that make them distinct from Bacteria in physiology and phylogeny.

[101] The term “Cannabis” refers to a genus in the family Cannabaceae. Cannabis is a dioecious plant. Glandular structures located on female flowers of Cannabis, called trichomes, accumulate relatively high amounts of a class of terpeno-phenolic compounds known as phytocannabinoids (described in further detail below). Cannabis has conventionally been cultivated for production of fibre and seed (commonly referred to as “hemp-type”), or for production of intoxicants (commonly referred to as “drug-type”). In drug-type Cannabis, the trichomes contain relatively high amounts of tetrahydrocannabinolic acid (THCA), which can convert to tetrahydrocannabinol (THC) via a decarboxylation reaction, for example upon

combustion of dried Cannabis flowers, to provide an intoxicating effect. Drug-type Cannabis often contains other cannabinoids in lesser amounts. In contrast, hemp-type Cannabis contains relatively low concentrations of THCA, often less than 0.3% THC by dry weight. Hemp-type Cannabis may contain non-THC and non-THCA cannabinoids, such as cannabidiolic acid (CBDA), cannabidiol (CBD), and other cannabinoids. Presently, there is a lack of consensus regarding the taxonomic organization of the species within the genus. Unless context dictates otherwise, the term “Cannabis” is intended to include all putative species within the genus, such as, without limitation, *Cannabis sativa*, *Cannabis indica*, and *Cannabis ruderalis* and without regard to whether the Cannabis is hemp-type or drug-type.

[102] The term “cyclase activity” in reference to a polyketide synthase (PKS) enzyme (e.g., an olivetol synthase (OLS) enzyme) or a polyketide cyclase (PKC) enzyme (e.g., an olivetolic acid cyclase (OAC) enzyme), refers to the activity of catalyzing the cyclization of an oxo fatty acyl-CoA (e.g., 3,5,7-trioxododecanoyl-CoA, 3,5,7-trioxodecanoyl-CoA) to the corresponding intramolecular cyclization product (e.g., olivetolic acid, divarinic acid). In some embodiments, the PKS or PKC catalyzes the C2-C7 aldol condensation of an acyl-CoA with three additional ketide moieties added thereto.

[103] A “cytosolic” or “soluble” enzyme refers to an enzyme that is predominantly localized (or predicted to be localized) in the cytosol of a host cell.

[104] A “eukaryote” is any organism whose cells contain a nucleus and other organelles enclosed within membranes. Eukaryotes belong to the taxon Eukarya or Eukaryota. The defining feature that sets eukaryotic cells apart from prokaryotic cells (i.e., bacteria and archaea) is that they have membrane-bound organelles, especially the nucleus, which contains the genetic material, and is enclosed by the nuclear envelope.

[105] The term “host cell” refers to a cell that can be used to express a polynucleotide, such as a polynucleotide that encodes an enzyme used in biosynthesis of cannabinoids or cannabinoid precursors. The terms “genetically modified host cell,” “recombinant host cell,” and “recombinant strain” are used interchangeably and refer to host cells that have been genetically modified by, e.g., cloning and transformation methods, or by other methods known in the art (e.g., selective editing methods, such as CRISPR). Thus, the terms include a host cell (e.g., bacterial cell, yeast cell, fungal cell, insect cell, plant cell, mammalian cell, human cell, etc.) that has been genetically altered, modified, or engineered, so that it exhibits an altered,

modified, or different genotype and/or phenotype, as compared to the naturally-occurring cell from which it was derived. It is understood that in some embodiments, the terms refer not only to the particular recombinant host cell in question, but also to the progeny or potential progeny of such a host cell.

[106] The term “control host cell,” or the term “control” when used in relation to a host cell, refers to an appropriate comparator host cell for determining the effect of a genetic modification or experimental treatment. In some embodiments, the control host cell is a wild type cell. In other embodiments, a control host cell is genetically identical to the genetically modified host cell, except for the genetic modification(s) differentiating the genetically modified or experimental treatment host cell. In some embodiments, the control host cell has been genetically modified to express a wild type or otherwise known variant of an enzyme being tested for activity in other test host cells.

[107] The term “heterologous” with respect to a polynucleotide, such as a polynucleotide comprising a gene, is used interchangeably with the term “exogenous” and the term “recombinant” and refers to: a polynucleotide that has been artificially supplied to a biological system; a polynucleotide that has been modified within a biological system, or a polynucleotide whose expression or regulation has been manipulated within a biological system. A heterologous polynucleotide that is introduced into or expressed in a host cell may be a polynucleotide that comes from a different organism or species from the host cell, or may be a synthetic polynucleotide, or may be a polynucleotide that is also endogenously expressed in the same organism or species as the host cell. For example, a polynucleotide that is endogenously expressed in a host cell may be considered heterologous when it is situated non-naturally in the host cell; expressed recombinantly in the host cell, either stably or transiently; modified within the host cell; selectively edited within the host cell; expressed in a copy number that differs from the naturally occurring copy number within the host cell; or expressed in a non-natural way within the host cell, such as by manipulating regulatory regions that control expression of the polynucleotide. In some embodiments, a heterologous polynucleotide is a polynucleotide that is endogenously expressed in a host cell but whose expression is driven by a promoter that does not naturally regulate expression of the polynucleotide. In other embodiments, a heterologous polynucleotide is a polynucleotide that is endogenously expressed in a host cell and whose expression is driven by a promoter that does naturally regulate expression of the polynucleotide, but the promoter or another regulatory region is

modified. In some embodiments, the promoter is recombinantly activated or repressed. For example, gene-editing based techniques may be used to regulate expression of a polynucleotide, including an endogenous polynucleotide, from a promoter, including an endogenous promoter. See, *e.g.*, Chavez *et al.*, Nat Methods. 2016 Jul; 13(7): 563–567. A heterologous polynucleotide may comprise a wild-type sequence or a mutant sequence as compared with a reference polynucleotide sequence.

[108] The term “at least a portion” or “at least a fragment” of a nucleic acid or polypeptide means a portion having the minimal size characteristics of such sequences, or any larger fragment of the full length molecule, up to and including the full length molecule. A fragment of a polynucleotide of the disclosure may encode a biologically active portion of an enzyme, such as a catalytic domain. A biologically active portion of a genetic regulatory element may comprise a portion or fragment of a full length genetic regulatory element and have the same type of activity as the full length genetic regulatory element, although the level of activity of the biologically active portion of the genetic regulatory element may vary compared to the level of activity of the full length genetic regulatory element.

[109] A coding sequence and a regulatory sequence are “operably joined” or “operably linked” when the coding sequence and the regulatory sequence are covalently linked and the expression or transcription of the coding sequence is under the influence or control of the regulatory sequence.

[110] The terms “link,” “linked,” or “linkage” means two entities (*e.g.*, two polynucleotides or two proteins) are bound to one another by any physicochemical means. Any linkage known to those of ordinary skill in the art, covalent or non-covalent, is embraced. In some embodiments, a nucleic acid sequence encoding an enzyme of the disclosure is linked to a nucleic acid encoding a signal peptide. In some embodiments, an enzyme of the disclosure is linked to a signal peptide. Linkage can be direct or indirect.

[111] The terms “transformed” or “transform” with respect to a host cell refer to a host cell in which one or more nucleic acids have been introduced, for example on a plasmid or vector or by integration into the genome. In some instances where one or more nucleic acids are introduced into a host cell on a plasmid or vector, one or more of the nucleic acids, or fragments thereof, may be retained in the cell, such as by integration into the genome of the cell, while the plasmid or vector itself may be removed from the cell. In such instances, the

host cell is considered to be transformed with the nucleic acids that were introduced into the cell regardless of whether the plasmid or vector is retained in the cell or not.

[112] The term “volumetric productivity” or “production rate” refers to the amount of product formed per volume of medium per unit of time. Volumetric productivity can be reported in gram per liter per hour (g/L/h).

[113] The term “specific productivity” of a product refers to the rate of formation of the product normalized by unit volume or mass or biomass and has the physical dimension of a quantity of substance per unit time per unit mass or volume $[M \cdot T^{-1} \cdot M^{-1} \text{ or } M \cdot T^{-1} \cdot L^{-3}]$, where M is mass or moles, T is time, L is length].

[114] The term “biomass specific productivity” refers to the specific productivity in gram product per gram of cell dry weight (CDW) per hour (g/g CDW/h) or in mmol of product per gram of cell dry weight (CDW) per hour (mmol/g CDW/h). Using the relation of CDW to OD600 for the given microorganism, specific productivity can also be expressed as gram product per liter culture medium per optical density of the culture broth at 600 nm (OD) per hour (g/L/h/OD). Also, if the elemental composition of the biomass is known, biomass specific productivity can be expressed in mmol of product per C-mole (carbon mole) of biomass per hour (mmol/C-mol/h).

[115] The term “yield” refers to the amount of product obtained per unit weight of a certain substrate and may be expressed as g product per g substrate (g/g) or moles of product per mole of substrate (mol/mol). Yield may also be expressed as a percentage of the theoretical yield. “Theoretical yield” is defined as the maximum amount of product that can be generated per a given amount of substrate as dictated by the stoichiometry of the metabolic pathway used to make the product and may be expressed as g product per g substrate (g/g) or moles of product per mole of substrate (mol/mol).

[116] The term “titer” refers to the strength of a solution or the concentration of a substance in solution. For example, the titer of a product of interest (*e.g.*, small molecule, peptide, synthetic compound, fuel, alcohol, *etc.*) in a fermentation broth is described as g of product of interest in solution per liter of fermentation broth or cell-free broth (g/L) or as g of product of interest in solution per kg of fermentation broth or cell-free broth (g/Kg).

[117] The term “total titer” refers to the sum of all products of interest produced in a process, including but not limited to the products of interest in solution, the products of interest in gas phase if applicable, and any products of interest removed from the process and recovered relative to the initial volume in the process or the operating volume in the process. For example, the total titer of products of interest (*e.g.*, small molecule, peptide, synthetic compound, fuel, alcohol, *etc.*) in a fermentation broth is described as g of products of interest in solution per liter of fermentation broth or cell-free broth (g/L) or as g of products of interest in solution per kg of fermentation broth or cell-free broth (g/Kg).

[118] The term “amino acid” refers to organic compounds that comprise an amino group, $-NH_2$, and a carboxyl group, $-COOH$. The term “amino acid” includes both naturally occurring and unnatural amino acids. Nomenclature for the twenty common amino acids is as follows: alanine (ala or A); arginine (arg or R); asparagine (asn or N); aspartic acid (asp or D); cysteine (cys or C); glutamine (gln or Q); glutamic acid (glu or E); glycine (gly or G); histidine (his or H); isoleucine (ile or I); leucine (leu or L); lysine (lys or K); methionine (met or M); phenylalanine (phe or F); proline (pro or P); serine (ser or S); threonine (thr or T); tryptophan (trp or W); tyrosine (tyr or Y); and valine (val or V). Non-limiting examples of unnatural amino acids include homo-amino acids, proline and pyruvic acid derivatives, 3-substituted alanine derivatives, glycine derivatives, ring-substituted phenylalanine derivatives, ring-substituted tyrosine derivatives, linear core amino acids, amino acids with protecting groups including Fmoc, Boc, and Cbz, β -amino acids (β_3 and β_2), and *N*-methyl amino acids.

[119] The term “aliphatic” refers to alkyl, alkenyl, alkynyl, and carbocyclic groups. Likewise, the term “heteroaliphatic” refers to heteroalkyl, heteroalkenyl, heteroalkynyl, and heterocyclic groups.

[120] The term “alkyl” refers to a radical of, or a substituent that is, a straight-chain or branched saturated hydrocarbon group having from 1 to 20 carbon atoms (“ C_{1-20} alkyl”). In certain embodiments, the term “alkyl” refers to a radical of, or a substituent that is, a straight-chain or branched saturated hydrocarbon group having from 1 to 10 carbon atoms (“ C_{1-10} alkyl”). In some embodiments, an alkyl group has 1 to 9 carbon atoms (“ C_{1-9} alkyl”). In some embodiments, an alkyl group has 1 to 8 carbon atoms (“ C_{1-8} alkyl”). In some embodiments, an alkyl group has 1 to 7 carbon atoms (“ C_{1-7} alkyl”). In some embodiments, an alkyl group has 2 to 7 carbon atoms (“ C_{2-7} alkyl”). In some embodiments, an alkyl group has 3 to 7 carbon atoms

("C₃₋₇ alkyl"). In some embodiments, an alkyl group has 1 to 6 carbon atoms ("C₁₋₆ alkyl"). In some embodiments, an alkyl group has 2 to 6 carbon atoms ("C₂₋₆ alkyl"). In some embodiments, an alkyl group has 3 to 5 carbon atoms ("C₃₋₅ alkyl"). In some embodiments, an alkyl group has 5 carbon atoms ("C₅ alkyl"). In some embodiments, the alkyl group has 3 carbon atoms ("C₃ alkyl"). In some embodiments, the alkyl group has 7 carbon atoms ("C₇ alkyl"). In some embodiments, an alkyl group has 1 to 5 carbon atoms ("C₁₋₅ alkyl"). In some embodiments, an alkyl group has 1 to 4 carbon atoms ("C₁₋₄ alkyl"). In some embodiments, an alkyl group has 1 to 3 carbon atoms ("C₁₋₃ alkyl"). In some embodiments, an alkyl group has 1 to 2 carbon atoms ("C₁₋₂ alkyl"). In some embodiments, an alkyl group has 1 carbon atom ("C₁ alkyl").

[121] Examples of C₁₋₆ alkyl groups include methyl (C₁), ethyl (C₂), propyl (C₃) (*e.g.*, n-propyl, isopropyl), butyl (C₄) (*e.g.*, n-butyl, tert-butyl, sec-butyl, iso-butyl), pentyl (C₅) (*e.g.*, n-pentyl, 3-pentanyl, amyl, neopentyl, 3-methyl-2-butanyl, tertiary amyl), and hexyl (C₆) (*e.g.*, n-hexyl). Additional examples of alkyl groups include n-heptyl (C₇), n-octyl (C₈), and the like. Unless otherwise specified, each instance of an alkyl group is independently unsubstituted (an "unsubstituted alkyl") or substituted (a "substituted alkyl") with one or more substituents (*e.g.*, halogen, such as F). In certain embodiments, the alkyl group is an unsubstituted C₁₋₁₀ alkyl (such as unsubstituted C₁₋₆ alkyl, *e.g.*, -CH₃ (Me), unsubstituted ethyl (Et), unsubstituted propyl (Pr, *e.g.*, unsubstituted n-propyl (n-Pr), unsubstituted isopropyl (i-Pr)), unsubstituted butyl (Bu, *e.g.*, unsubstituted n-butyl (n-Bu), unsubstituted tert-butyl (tert-Bu or t-Bu), unsubstituted sec-butyl (sec-Bu), unsubstituted isobutyl (i-Bu)). In certain embodiments, the alkyl group is a substituted C₁₋₁₀ alkyl (such as substituted C₁₋₆ alkyl, *e.g.*, -CF₃, benzyl).

[122] The term "acyl" refers to a group having the general formula -C(=O)R^{X1}, -C(=O)OR^{X1}, -C(=O)-O-C(=O)R^{X1}, -C(=O)SR^{X1}, -C(=O)N(R^{X1})₂, -C(=S)R^{X1}, -C(=S)N(R^{X1})₂, and -C(=S)S(R^{X1}), -C(=NR^{X1})R^{X1}, -C(=NR^{X1})OR^{X1}, -C(=NR^{X1})SR^{X1}, and -C(=NR^{X1})N(R^{X1})₂, wherein R^{X1} is hydrogen; halogen; substituted or unsubstituted hydroxyl; substituted or unsubstituted thiol; substituted or unsubstituted amino; substituted or unsubstituted acyl, cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic; cyclic or acyclic, substituted or unsubstituted, branched or unbranched heteroaliphatic; cyclic or acyclic, substituted or unsubstituted, branched or unbranched alkyl; cyclic or acyclic, substituted or unsubstituted, branched or unbranched alkenyl; substituted or unsubstituted alkynyl; substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl,

aliphaticoxy, heteroaliphaticoxy, alkyloxy, heteroalkyloxy, aryloxy, heteroaryloxy, aliphaticthioxy, heteroaliphaticthioxy, alkylthioxy, heteroalkylthioxy, arylthioxy, heteroarylthioxy, mono- or di- aliphaticamino, mono- or di- heteroaliphaticamino, mono- or di- alkylamino, mono- or di- heteroalkylamino, mono- or di-arylamino, or mono- or di- heteroarylamino; or two R^{X1} groups taken together form a 5- to 6-membered heterocyclic ring. Exemplary acyl groups include aldehydes ($-CHO$), carboxylic acids ($-CO_2H$), ketones, acyl halides, esters, amides, imines, carbonates, carbamates, and ureas. Acyl substituents include, but are not limited to, any of the substituents described in this application that result in the formation of a stable moiety (*e.g.*, aliphatic, alkyl, alkenyl, alkynyl, heteroaliphatic, heterocyclic, aryl, heteroaryl, acyl, oxo, imino, thiooxo, cyano, isocyano, amino, azido, nitro, hydroxyl, thiol, halo, aliphaticamino, heteroaliphaticamino, alkylamino, heteroalkylamino, arylamino, heteroarylamino, alkylaryl, arylalkyl, aliphaticoxy, heteroaliphaticoxy, alkyloxy, heteroalkyloxy, aryloxy, heteroaryloxy, aliphaticthioxy, heteroaliphaticthioxy, alkylthioxy, heteroalkylthioxy, arylthioxy, heteroarylthioxy, acyloxy, and the like, each of which may or may not be further substituted).

[123] “Alkenyl” refers to a radical of, or a substituent that is, a straight-chain or branched hydrocarbon group having from 2 to 20 carbon atoms, one or more carbon-carbon double bonds, and no triple bonds (“ C_{2-20} alkenyl”). In some embodiments, an alkenyl group has 2 to 10 carbon atoms (“ C_{2-10} alkenyl”). In some embodiments, an alkenyl group has 2 to 9 carbon atoms (“ C_{2-9} alkenyl”). In some embodiments, an alkenyl group has 2 to 8 carbon atoms (“ C_{2-8} alkenyl”). In some embodiments, an alkenyl group has 2 to 7 carbon atoms (“ C_{2-7} alkenyl”). In some embodiments, an alkenyl group has 2 to 6 carbon atoms (“ C_{2-6} alkenyl”). In some embodiments, an alkenyl group has 2 to 5 carbon atoms (“ C_{2-5} alkenyl”). In some embodiments, an alkenyl group has 2 to 4 carbon atoms (“ C_{2-4} alkenyl”). In some embodiments, an alkenyl group has 2 to 3 carbon atoms (“ C_{2-3} alkenyl”). In some embodiments, an alkenyl group has 2 carbon atoms (“ C_2 alkenyl”). The one or more carbon-carbon double bonds can be internal (such as in 2-butenyl) or terminal (such as in 1-butenyl). Examples of C_{2-4} alkenyl groups include ethenyl (C_2), 1-propenyl (C_3), 2-propenyl (C_3), 1-butenyl (C_4), 2-butenyl (C_4), butadienyl (C_4), and the like. Examples of C_{2-6} alkenyl groups include the aforementioned C_{2-4} alkenyl groups as well as pentenyl (C_5), pentadienyl (C_5), hexenyl (C_6), and the like. Additional examples of alkenyl include heptenyl (C_7), octenyl (C_8), octatrienyl (C_8), and the like. Unless otherwise specified, each instance of an alkenyl group is

independently optionally substituted, *i.e.*, unsubstituted (an “unsubstituted alkenyl”) or substituted (a “substituted alkenyl”) with one or more substituents. In certain embodiments, the alkenyl group is unsubstituted C₂₋₁₀ alkenyl. In certain embodiments, the alkenyl group is substituted C₂₋₁₀ alkenyl.

[124] “Alkynyl” refers to a radical of, or a substituent that is, a straight-chain or branched hydrocarbon group having from 2 to 20 carbon atoms, one or more carbon-carbon triple bonds, and optionally one or more double bonds (“C₂₋₂₀ alkynyl”). In some embodiments, an alkynyl group has 2 to 10 carbon atoms (“C₂₋₁₀ alkynyl”). In some embodiments, an alkynyl group has 2 to 9 carbon atoms (“C₂₋₉ alkynyl”). In some embodiments, an alkynyl group has 2 to 8 carbon atoms (“C₂₋₈ alkynyl”). In some embodiments, an alkynyl group has 2 to 7 carbon atoms (“C₂₋₇ alkynyl”). In some embodiments, an alkynyl group has 2 to 6 carbon atoms (“C₂₋₆ alkynyl”). In some embodiments, an alkynyl group has 2 to 5 carbon atoms (“C₂₋₅ alkynyl”). In some embodiments, an alkynyl group has 2 to 4 carbon atoms (“C₂₋₄ alkynyl”). In some embodiments, an alkynyl group has 2 to 3 carbon atoms (“C₂₋₃ alkynyl”). In some embodiments, an alkynyl group has 2 carbon atoms (“C₂ alkynyl”). The one or more carbon-carbon triple bonds can be internal (such as in 2-butyne) or terminal (such as in 1-butyne). Examples of C₂₋₄ alkynyl groups include, without limitation, ethynyl (C₂), 1-propynyl (C₃), 2-propynyl (C₃), 1-butyne (C₄), 2-butyne (C₄), and the like. Examples of C₂₋₆ alkenyl groups include the aforementioned C₂₋₄ alkynyl groups as well as pentynyl (C₅), hexynyl (C₆), and the like. Additional examples of alkynyl include heptynyl (C₇), octynyl (C₈), and the like. Unless otherwise specified, each instance of an alkynyl group is independently optionally substituted, *i.e.*, unsubstituted (an “unsubstituted alkynyl”) or substituted (a “substituted alkynyl”) with one or more substituents. In certain embodiments, the alkynyl group is unsubstituted C₂₋₁₀ alkynyl. In certain embodiments, the alkynyl group is substituted C₂₋₁₀ alkynyl.

[125] “Carbocyclyl” or “carbocyclic” refers to a radical of a non-aromatic cyclic hydrocarbon group having from 3 to 10 ring carbon atoms (“C₃₋₁₀ carbocyclyl”) and zero heteroatoms in the non-aromatic ring system. In some embodiments, a carbocyclyl group has 3 to 8 ring carbon atoms (“C₃₋₈ carbocyclyl”). In some embodiments, a carbocyclyl group has 3 to 6 ring carbon atoms (“C₃₋₆ carbocyclyl”). In some embodiments, a carbocyclyl group has 3 to 6 ring carbon atoms (“C₃₋₆ carbocyclyl”). In some embodiments, a carbocyclyl group has 5 to 10 ring carbon atoms (“C₅₋₁₀ carbocyclyl”). Exemplary C₃₋₆ carbocyclyl groups include, without limitation, cyclopropyl (C₃), cyclopropenyl (C₃), cyclobutyl (C₄), cyclobutenyl (C₄),

cyclopentyl (C₅), cyclopentenyl (C₅), cyclohexyl (C₆), cyclohexenyl (C₆), cyclohexadienyl (C₆), and the like. Exemplary C₃₋₈ carbocyclyl groups include, without limitation, the aforementioned C₃₋₆ carbocyclyl groups as well as cycloheptyl (C₇), cycloheptenyl (C₇), cycloheptadienyl (C₇), cycloheptatrienyl (C₇), cyclooctyl (C₈), cyclooctenyl (C₈), bicyclo[2.2.1]heptanyl (C₇), bicyclo[2.2.2]octanyl (C₈), and the like. Exemplary C₃₋₁₀ carbocyclyl groups include, without limitation, the aforementioned C₃₋₈ carbocyclyl groups as well as cyclononyl (C₉), cyclononenyl (C₉), cyclodecyl (C₁₀), cyclodecenyl (C₁₀), octahydro-1*H*-indenyl (C₉), decahydronaphthalenyl (C₁₀), spiro[4.5]decanyl (C₁₀), and the like. As the foregoing examples illustrate, in certain embodiments, the carbocyclyl group is either monocyclic (“monocyclic carbocyclyl”) or contain a fused, bridged or spiro ring system such as a bicyclic system (“bicyclic carbocyclyl”) and can be saturated or can be partially unsaturated. “Carbocyclyl” also includes ring systems wherein the carbocyclic ring, as defined above, is fused with one or more aryl or heteroaryl groups wherein the point of attachment is on the carbocyclic ring, and in such instances, the number of carbons continue to designate the number of carbons in the carbocyclic ring system. Unless otherwise specified, each instance of a carbocyclyl group is independently optionally substituted, *i.e.*, unsubstituted (an “unsubstituted carbocyclyl”) or substituted (a “substituted carbocyclyl”) with one or more substituents. In certain embodiments, the carbocyclyl group is unsubstituted C₃₋₁₀ carbocyclyl. In certain embodiments, the carbocyclyl group is a substituted C₃₋₁₀ carbocyclyl.

[126] In some embodiments, “carbocyclyl” is a monocyclic, saturated carbocyclyl group having from 3 to 10 ring carbon atoms (“C₃₋₁₀ cycloalkyl”). In some embodiments, a cycloalkyl group has 3 to 8 ring carbon atoms (“C₃₋₈ cycloalkyl”). In some embodiments, a cycloalkyl group has 3 to 6 ring carbon atoms (“C₃₋₆ cycloalkyl”). In some embodiments, a cycloalkyl group has 5 to 6 ring carbon atoms (“C₅₋₆ cycloalkyl”). In some embodiments, a cycloalkyl group has 5 to 10 ring carbon atoms (“C₅₋₁₀ cycloalkyl”). Examples of C₅₋₆ cycloalkyl groups include cyclopentyl (C₅) and cyclohexyl (C₆). Examples of C₃₋₆ cycloalkyl groups include the aforementioned C₅₋₆ cycloalkyl groups as well as cyclopropyl (C₃) and cyclobutyl (C₄). Examples of C₃₋₈ cycloalkyl groups include the aforementioned C₃₋₆ cycloalkyl groups as well as cycloheptyl (C₇) and cyclooctyl (C₈). Unless otherwise specified, each instance of a cycloalkyl group is independently unsubstituted (an “unsubstituted cycloalkyl”) or substituted (a “substituted cycloalkyl”) with one or more substituents. In certain

embodiments, the cycloalkyl group is unsubstituted C₃₋₁₀ cycloalkyl. In certain embodiments, the cycloalkyl group is substituted C₃₋₁₀ cycloalkyl.

[127] “Aryl” refers to a radical of a monocyclic or polycyclic (*e.g.*, bicyclic or tricyclic) $4n+2$ aromatic ring system (*e.g.*, having 6, 10, or 14 pi electrons shared in a cyclic array) having 6–14 ring carbon atoms and zero heteroatoms provided in the aromatic ring system (“C₆₋₁₄ aryl”). In some embodiments, an aryl group has six ring carbon atoms (“C₆ aryl”; *e.g.*, phenyl). In some embodiments, an aryl group has ten ring carbon atoms (“C₁₀ aryl”; *e.g.*, naphthyl such as 1-naphthyl and 2-naphthyl). In some embodiments, an aryl group has fourteen ring carbon atoms (“C₁₄ aryl”; *e.g.*, anthracyl). “Aryl” also includes ring systems wherein the aryl ring, as defined above, is fused with one or more carbocyclyl or heterocyclyl groups wherein the radical or point of attachment is on the aryl ring, and in such instances, the number of carbon atoms continue to designate the number of carbon atoms in the aryl ring system. Unless otherwise specified, each instance of an aryl group is independently optionally substituted, *i.e.*, unsubstituted (an “unsubstituted aryl”) or substituted (a “substituted aryl”) with one or more substituents. In certain embodiments, the aryl group is unsubstituted C₆₋₁₄ aryl. In certain embodiments, the aryl group is substituted C₆₋₁₄ aryl.

[128] “Aralkyl” is a subset of alkyl and aryl and refers to an optionally substituted alkyl group substituted by an optionally substituted aryl group. In certain embodiments, the aralkyl is optionally substituted benzyl. In certain embodiments, the aralkyl is benzyl. In certain embodiments, the aralkyl is optionally substituted phenethyl. In certain embodiments, the aralkyl is phenethyl. In certain embodiments, the aralkyl is 7-phenylheptanyl. In certain embodiments, the aralkyl is C₇ alkyl substituted by an optionally substituted aryl group (*e.g.*, phenyl). In certain embodiments, the aralkyl is a C₇-C₁₀ alkyl group substituted by an optionally substituted aryl group (*e.g.*, phenyl).

[129] “Partially unsaturated” refers to a group that includes at least one double or triple bond. A “partially unsaturated” ring system is further intended to encompass rings having multiple sites of unsaturation but is not intended to include aromatic groups (*e.g.*, aryl or heteroaryl groups) as defined in this application. Likewise, “saturated” refers to a group that does not contain a double or triple bond, *i.e.*, contains all single bonds.

[130] The term “optionally substituted” means substituted or unsubstituted.

[131] Alkyl, alkenyl, alkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl groups are optionally substituted (*e.g.*, “substituted” or “unsubstituted” alkyl, “substituted” or “unsubstituted” alkenyl, “substituted” or “unsubstituted” alkynyl, “substituted” or “unsubstituted” carbocyclyl, “substituted” or “unsubstituted” heterocyclyl, “substituted” or “unsubstituted” aryl or “substituted” or “unsubstituted” heteroaryl group). In general, the term “substituted,” whether preceded by the term “optionally” or not, means that at least one hydrogen present on a group (*e.g.*, a carbon or nitrogen atom) is replaced with a permissible substituent, *e.g.*, a substituent which upon substitution results in a stable compound, *e.g.*, a compound which does not spontaneously undergo transformation such as by rearrangement, cyclization, elimination, or other reaction. Unless otherwise indicated, a “substituted” group has a substituent at one or more substitutable positions of the group, and when more than one position in any given structure is substituted, the substituent is either the same or different at each position. The term “substituted” is contemplated to include substitution with all permissible substituents of organic compounds, any of the substituents described in this application that results in the formation of a stable compound. The present invention contemplates any and all such combinations in order to arrive at a stable compound. For purposes of this invention, heteroatoms such as nitrogen may have hydrogen substituents and/or any suitable substituent as described in this application which satisfy the valencies of the heteroatoms and results in the formation of a stable moiety.

[132] Exemplary carbon atom substituents include, but are not limited to, halogen, $-\text{CN}$, $-\text{NO}_2$, $-\text{N}_3$, $-\text{SO}_2\text{H}$, $-\text{SO}_3\text{H}$, $-\text{OH}$, $-\text{OR}^{\text{aa}}$, $-\text{ON}(\text{R}^{\text{bb}})_2$, $-\text{N}(\text{R}^{\text{bb}})_2$, $-\text{N}(\text{R}^{\text{bb}})_3^+\text{X}^-$, $-\text{N}(\text{OR}^{\text{cc}})\text{R}^{\text{bb}}$, $-\text{SH}$, $-\text{SR}^{\text{aa}}$, $-\text{SSR}^{\text{cc}}$, $-\text{C}(=\text{O})\text{R}^{\text{aa}}$, $-\text{CO}_2\text{H}$, $-\text{CHO}$, $-\text{C}(\text{OR}^{\text{cc}})_2$, $-\text{CO}_2\text{R}^{\text{aa}}$, $-\text{OC}(=\text{O})\text{R}^{\text{aa}}$, $-\text{OCO}_2\text{R}^{\text{aa}}$, $-\text{C}(=\text{O})\text{N}(\text{R}^{\text{bb}})_2$, $-\text{OC}(=\text{O})\text{N}(\text{R}^{\text{bb}})_2$, $-\text{NR}^{\text{bb}}\text{C}(=\text{O})\text{R}^{\text{aa}}$, $-\text{NR}^{\text{bb}}\text{CO}_2\text{R}^{\text{aa}}$, $-\text{NR}^{\text{bb}}\text{C}(=\text{O})\text{N}(\text{R}^{\text{bb}})_2$, $-\text{C}(=\text{NR}^{\text{bb}})\text{R}^{\text{aa}}$, $-\text{C}(=\text{NR}^{\text{bb}})\text{OR}^{\text{aa}}$, $-\text{OC}(=\text{NR}^{\text{bb}})\text{R}^{\text{aa}}$, $-\text{OC}(=\text{NR}^{\text{bb}})\text{OR}^{\text{aa}}$, $-\text{C}(=\text{NR}^{\text{bb}})\text{N}(\text{R}^{\text{bb}})_2$, $-\text{OC}(=\text{NR}^{\text{bb}})\text{N}(\text{R}^{\text{bb}})_2$, $-\text{NR}^{\text{bb}}\text{C}(=\text{NR}^{\text{bb}})\text{N}(\text{R}^{\text{bb}})_2$, $-\text{C}(=\text{O})\text{NR}^{\text{bb}}\text{SO}_2\text{R}^{\text{aa}}$, $-\text{NR}^{\text{bb}}\text{SO}_2\text{R}^{\text{aa}}$, $-\text{SO}_2\text{N}(\text{R}^{\text{bb}})_2$, $-\text{SO}_2\text{R}^{\text{aa}}$, $-\text{SO}_2\text{OR}^{\text{aa}}$, $-\text{OSO}_2\text{R}^{\text{aa}}$, $-\text{S}(=\text{O})\text{R}^{\text{aa}}$, $-\text{OS}(=\text{O})\text{R}^{\text{aa}}$, $-\text{Si}(\text{R}^{\text{aa}})_3$, $-\text{OSi}(\text{R}^{\text{aa}})_3$, $-\text{C}(=\text{S})\text{N}(\text{R}^{\text{bb}})_2$, $-\text{C}(=\text{O})\text{SR}^{\text{aa}}$, $-\text{C}(=\text{S})\text{SR}^{\text{aa}}$, $-\text{SC}(=\text{S})\text{SR}^{\text{aa}}$, $-\text{SC}(=\text{O})\text{SR}^{\text{aa}}$, $-\text{OC}(=\text{O})\text{SR}^{\text{aa}}$, $-\text{SC}(=\text{O})\text{OR}^{\text{aa}}$, $-\text{SC}(=\text{O})\text{R}^{\text{aa}}$, $-\text{P}(=\text{O})(\text{R}^{\text{aa}})_2$, $-\text{P}(=\text{O})(\text{OR}^{\text{cc}})_2$, $-\text{OP}(=\text{O})(\text{R}^{\text{aa}})_2$, $-\text{OP}(=\text{O})(\text{OR}^{\text{cc}})_2$, $-\text{P}(=\text{O})(\text{N}(\text{R}^{\text{bb}})_2)_2$, $-\text{OP}(=\text{O})(\text{N}(\text{R}^{\text{bb}})_2)_2$, $-\text{NR}^{\text{bb}}\text{P}(=\text{O})(\text{R}^{\text{aa}})_2$, $-\text{NR}^{\text{bb}}\text{P}(=\text{O})(\text{OR}^{\text{cc}})_2$, $-\text{NR}^{\text{bb}}\text{P}(=\text{O})(\text{N}(\text{R}^{\text{bb}})_2)_2$, $-\text{P}(\text{R}^{\text{cc}})_2$, $-\text{P}(\text{OR}^{\text{cc}})_2$, $-\text{P}(\text{R}^{\text{cc}})_3^+\text{X}^-$, $-\text{P}(\text{OR}^{\text{cc}})_3^+\text{X}^-$, $-\text{P}(\text{R}^{\text{cc}})_4$, $-\text{P}(\text{OR}^{\text{cc}})_4$, $-\text{OP}(\text{R}^{\text{cc}})_2$, $-\text{OP}(\text{R}^{\text{cc}})_3^+\text{X}^-$, $-\text{OP}(\text{OR}^{\text{cc}})_2$, $-\text{OP}(\text{OR}^{\text{cc}})_3^+\text{X}^-$, $-\text{OP}(\text{R}^{\text{cc}})_4$, $-\text{OP}(\text{OR}^{\text{cc}})_4$, $-\text{B}(\text{R}^{\text{aa}})_2$, $-\text{B}(\text{OR}^{\text{cc}})_2$, $-\text{BR}^{\text{aa}}(\text{OR}^{\text{cc}})$, C_{1-10} alkyl, C_{1-10} perhaloalkyl, C_2 .

10 alkenyl, C₂₋₁₀ alkynyl, heteroC₁₋₁₀ alkyl, heteroC₂₋₁₀ alkenyl, heteroC₂₋₁₀ alkynyl, C₃₋₁₀ carbocyclyl, 3-14 membered heterocyclyl, C₆₋₁₄ aryl, and 5-14 membered heteroaryl;

wherein:

each instance of R^{aa} is, independently, selected from C₁₋₁₀ alkyl, C₁₋₁₀ perhaloalkyl, C₂₋₁₀ alkenyl, C₂₋₁₀ alkynyl, heteroC₁₋₁₀ alkyl, heteroC₂₋₁₀ alkenyl, heteroC₂₋₁₀ alkynyl, C₃₋₁₀ carbocyclyl, 3-14 membered heterocyclyl, C₆₋₁₄ aryl, and 5-14 membered heteroaryl, or two R^{aa} groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups;

each instance of R^{bb} is, independently, selected from hydrogen, -OH, -OR^{aa}, -N(R^{cc})₂, -CN, -C(=O)R^{aa}, -C(=O)N(R^{cc})₂, -CO₂R^{aa}, -SO₂R^{aa}, -C(=NR^{cc})OR^{aa}, -C(=NR^{cc})N(R^{cc})₂, -SO₂N(R^{cc})₂, -SO₂R^{cc}, -SO₂OR^{cc}, -SOR^{aa}, -C(=S)N(R^{cc})₂, -C(=O)SR^{cc}, -C(=S)SR^{cc}, -P(=O)(R^{aa})₂, -P(=O)(OR^{cc})₂, -P(=O)(N(R^{cc})₂)₂, C₁₋₁₀ alkyl, C₁₋₁₀ perhaloalkyl, C₂₋₁₀ alkenyl, C₂₋₁₀ alkynyl, heteroC₁₋₁₀ alkyl, heteroC₂₋₁₀ alkenyl, heteroC₂₋₁₀ alkynyl, C₃₋₁₀ carbocyclyl, 3-14 membered heterocyclyl, C₆₋₁₄ aryl, and 5-14 membered heteroaryl, or two R^{bb} groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups; wherein X⁻ is a counterion;

each instance of R^{cc} is, independently, selected from hydrogen, C₁₋₁₀ alkyl, C₁₋₁₀ perhaloalkyl, C₂₋₁₀ alkenyl, C₂₋₁₀ alkynyl, heteroC₁₋₁₀ alkyl, heteroC₂₋₁₀ alkenyl, heteroC₂₋₁₀ alkynyl, C₃₋₁₀ carbocyclyl, 3-14 membered heterocyclyl, C₆₋₁₄ aryl, and 5-14 membered heteroaryl, or two R^{cc} groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups;

each instance of R^{dd} is, independently, selected from halogen, -CN, -NO₂, -N₃, -SO₂H, -SO₃H, -OH, -OR^{cc}, -ON(R^{ff})₂, -N(R^{ff})₂, -N(R^{ff})₃X⁻, -N(OR^{cc})R^{ff}, -SH, -SR^{cc}, -SSR^{cc}, -C(=O)R^{cc}, -CO₂H, -CO₂R^{cc}, -OC(=O)R^{cc}, -OCO₂R^{cc}, -C(=O)N(R^{ff})₂,

$-\text{OC}(=\text{O})\text{N}(\text{R}^{\text{ff}})_2$, $-\text{NR}^{\text{ff}}\text{C}(=\text{O})\text{R}^{\text{ee}}$, $-\text{NR}^{\text{ff}}\text{CO}_2\text{R}^{\text{ee}}$, $-\text{NR}^{\text{ff}}\text{C}(=\text{O})\text{N}(\text{R}^{\text{ff}})_2$, $-\text{C}(=\text{NR}^{\text{ff}})\text{OR}^{\text{ee}}$,
 $-\text{OC}(=\text{NR}^{\text{ff}})\text{R}^{\text{ee}}$, $-\text{OC}(=\text{NR}^{\text{ff}})\text{OR}^{\text{ee}}$, $-\text{C}(=\text{NR}^{\text{ff}})\text{N}(\text{R}^{\text{ff}})_2$, $-\text{OC}(=\text{NR}^{\text{ff}})\text{N}(\text{R}^{\text{ff}})_2$,
 $-\text{NR}^{\text{ff}}\text{C}(=\text{NR}^{\text{ff}})\text{N}(\text{R}^{\text{ff}})_2$, $-\text{NR}^{\text{ff}}\text{SO}_2\text{R}^{\text{ee}}$, $-\text{SO}_2\text{N}(\text{R}^{\text{ff}})_2$, $-\text{SO}_2\text{R}^{\text{ee}}$, $-\text{SO}_2\text{OR}^{\text{ee}}$, $-\text{OSO}_2\text{R}^{\text{ee}}$,
 $-\text{S}(=\text{O})\text{R}^{\text{ee}}$, $-\text{Si}(\text{R}^{\text{ee}})_3$, $-\text{OSi}(\text{R}^{\text{ee}})_3$, $-\text{C}(=\text{S})\text{N}(\text{R}^{\text{ff}})_2$, $-\text{C}(=\text{O})\text{SR}^{\text{ee}}$, $-\text{C}(=\text{S})\text{SR}^{\text{ee}}$, $-\text{SC}(=\text{S})\text{SR}^{\text{ee}}$,
 $-\text{P}(=\text{O})(\text{OR}^{\text{ee}})_2$, $-\text{P}(=\text{O})(\text{R}^{\text{ee}})_2$, $-\text{OP}(=\text{O})(\text{R}^{\text{ee}})_2$, $-\text{OP}(=\text{O})(\text{OR}^{\text{ee}})_2$, C_{1-6} alkyl, C_{1-6} perhaloalkyl,
 C_{2-6} alkenyl, C_{2-6} alkynyl, hetero C_{1-6} alkyl, hetero C_{2-6} alkenyl, hetero C_{2-6} alkynyl, C_{3-10}
carbocyclyl, 3-10 membered heterocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, wherein each
alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl,
aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{eg} groups, or two
geminal R^{dd} substituents can be joined to form $=\text{O}$ or $=\text{S}$; wherein X^- is a counterion;

each instance of R^{ee} is, independently, selected from C_{1-6} alkyl, C_{1-6}
perhaloalkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, hetero C_{1-6} alkyl, hetero C_{2-6} alkenyl, hetero C_{2-6} alkynyl,
 C_{3-10} carbocyclyl, C_{6-10} aryl, 3-10 membered heterocyclyl, and 3-10 membered heteroaryl,
wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl,
heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{eg} groups;

each instance of R^{ff} is, independently, selected from hydrogen, C_{1-6} alkyl, C_{1-6}
perhaloalkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, hetero C_{1-6} alkyl, hetero C_{2-6} alkenyl, hetero C_{2-6} alkynyl,
 C_{3-10} carbocyclyl, 3-10 membered heterocyclyl, C_{6-10} aryl and 5-10 membered heteroaryl, or
two R^{ff} groups are joined to form a 3-10 membered heterocyclyl or 5-10 membered heteroaryl
ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl,
carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or
5 R^{eg} groups; and

each instance of R^{eg} is, independently, halogen, $-\text{CN}$, $-\text{NO}_2$, $-\text{N}_3$, $-\text{SO}_2\text{H}$,
 $-\text{SO}_3\text{H}$, $-\text{OH}$, $-\text{OC}_{1-6}$ alkyl, $-\text{ON}(\text{C}_{1-6}$ alkyl) $_2$, $-\text{N}(\text{C}_{1-6}$ alkyl) $_2$, $-\text{N}(\text{C}_{1-6}$ alkyl) $_3^+\text{X}^-$, $-\text{NH}(\text{C}_{1-6}$
alkyl) $_2^+\text{X}^-$, $-\text{NH}_2(\text{C}_{1-6}$ alkyl) $^+\text{X}^-$, $-\text{NH}_3^+\text{X}^-$, $-\text{N}(\text{OC}_{1-6}$ alkyl)(C_{1-6} alkyl), $-\text{N}(\text{OH})(\text{C}_{1-6}$ alkyl),
 $-\text{NH}(\text{OH})$, $-\text{SH}$, $-\text{SC}_{1-6}$ alkyl, $-\text{SS}(\text{C}_{1-6}$ alkyl), $-\text{C}(=\text{O})(\text{C}_{1-6}$ alkyl), $-\text{CO}_2\text{H}$, $-\text{CO}_2(\text{C}_{1-6}$ alkyl),
 $-\text{OC}(=\text{O})(\text{C}_{1-6}$ alkyl), $-\text{OCO}_2(\text{C}_{1-6}$ alkyl), $-\text{C}(=\text{O})\text{NH}_2$, $-\text{C}(=\text{O})\text{N}(\text{C}_{1-6}$ alkyl) $_2$,
 $-\text{OC}(=\text{O})\text{NH}(\text{C}_{1-6}$ alkyl), $-\text{NHC}(=\text{O})(\text{C}_{1-6}$ alkyl), $-\text{N}(\text{C}_{1-6}$ alkyl) $\text{C}(=\text{O})(\text{C}_{1-6}$ alkyl),
 $-\text{NHCO}_2(\text{C}_{1-6}$ alkyl), $-\text{NHC}(=\text{O})\text{N}(\text{C}_{1-6}$ alkyl) $_2$, $-\text{NHC}(=\text{O})\text{NH}(\text{C}_{1-6}$ alkyl), $-\text{NHC}(=\text{O})\text{NH}_2$,
 $-\text{C}(=\text{NH})\text{O}(\text{C}_{1-6}$ alkyl), $-\text{OC}(=\text{NH})(\text{C}_{1-6}$ alkyl), $-\text{OC}(=\text{NH})\text{OC}_{1-6}$ alkyl, $-\text{C}(=\text{NH})\text{N}(\text{C}_{1-6}$
alkyl) $_2$, $-\text{C}(=\text{NH})\text{NH}(\text{C}_{1-6}$ alkyl), $-\text{C}(=\text{NH})\text{NH}_2$, $-\text{OC}(=\text{NH})\text{N}(\text{C}_{1-6}$ alkyl) $_2$, $-\text{OC}(\text{NH})\text{NH}(\text{C}_{1-6}$

₆ alkyl), $-\text{OC}(\text{NH})\text{NH}_2$, $-\text{NHC}(\text{NH})\text{N}(\text{C}_{1-6} \text{ alkyl})_2$, $-\text{NHC}(\text{=NH})\text{NH}_2$, $-\text{NH}\text{SO}_2(\text{C}_{1-6} \text{ alkyl})$, $-\text{SO}_2\text{N}(\text{C}_{1-6} \text{ alkyl})_2$, $-\text{SO}_2\text{NH}(\text{C}_{1-6} \text{ alkyl})$, $-\text{SO}_2\text{NH}_2$, $-\text{SO}_2\text{C}_{1-6} \text{ alkyl}$, $-\text{SO}_2\text{OC}_{1-6} \text{ alkyl}$, $-\text{OSO}_2\text{C}_{1-6} \text{ alkyl}$, $-\text{SOC}_{1-6} \text{ alkyl}$, $-\text{Si}(\text{C}_{1-6} \text{ alkyl})_3$, $-\text{OSi}(\text{C}_{1-6} \text{ alkyl})_3$, $-\text{C}(\text{=S})\text{N}(\text{C}_{1-6} \text{ alkyl})_2$, $\text{C}(\text{=S})\text{NH}(\text{C}_{1-6} \text{ alkyl})$, $\text{C}(\text{=S})\text{NH}_2$, $-\text{C}(\text{=O})\text{S}(\text{C}_{1-6} \text{ alkyl})$, $-\text{C}(\text{=S})\text{SC}_{1-6} \text{ alkyl}$, $-\text{SC}(\text{=S})\text{SC}_{1-6} \text{ alkyl}$, $-\text{P}(\text{=O})(\text{OC}_{1-6} \text{ alkyl})_2$, $-\text{P}(\text{=O})(\text{C}_{1-6} \text{ alkyl})_2$, $-\text{OP}(\text{=O})(\text{C}_{1-6} \text{ alkyl})_2$, $-\text{OP}(\text{=O})(\text{OC}_{1-6} \text{ alkyl})_2$, $\text{C}_{1-6} \text{ alkyl}$, $\text{C}_{1-6} \text{ perhaloalkyl}$, $\text{C}_{2-6} \text{ alkenyl}$, $\text{C}_{2-6} \text{ alkynyl}$, $\text{heteroC}_{1-6}\text{alkyl}$, $\text{heteroC}_{2-6}\text{alkenyl}$, $\text{heteroC}_{2-6}\text{alkynyl}$, $\text{C}_{3-10} \text{ carbocyclyl}$, $\text{C}_{6-10} \text{ aryl}$, 3-10 membered heterocyclyl, 5-10 membered heteroaryl; or two geminal R^{gg} substituents can be joined to form =O or =S ; wherein X^- is a counterion. Alternatively, two geminal hydrogens on a carbon atom are replaced with the group =O , =S , $\text{=NN}(\text{R}^{\text{bb}})_2$, $\text{=NNR}^{\text{bb}}\text{C}(\text{=O})\text{R}^{\text{aa}}$, $\text{=NNR}^{\text{bb}}\text{C}(\text{=O})\text{OR}^{\text{aa}}$, $\text{=NNR}^{\text{bb}}\text{S}(\text{=O})_2\text{R}^{\text{aa}}$, =NR^{bb} , or =NOR^{cc} ; wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups; wherein X^- is a counterion;

wherein:

each instance of R^{aa} is, independently, selected from $\text{C}_{1-10} \text{ alkyl}$, $\text{C}_{1-10} \text{ perhaloalkyl}$, $\text{C}_{2-10} \text{ alkenyl}$, $\text{C}_{2-10} \text{ alkynyl}$, $\text{heteroC}_{1-10} \text{ alkyl}$, $\text{heteroC}_{2-10}\text{alkenyl}$, $\text{heteroC}_{2-10}\text{alkynyl}$, $\text{C}_{3-10} \text{ carbocyclyl}$, 3-14 membered heterocyclyl, $\text{C}_{6-14} \text{ aryl}$, and 5-14 membered heteroaryl, or two R^{aa} groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups;

each instance of R^{bb} is, independently, selected from hydrogen, $-\text{OH}$, $-\text{OR}^{\text{aa}}$, $-\text{N}(\text{R}^{\text{cc}})_2$, $-\text{CN}$, $-\text{C}(\text{=O})\text{R}^{\text{aa}}$, $-\text{C}(\text{=O})\text{N}(\text{R}^{\text{cc}})_2$, $-\text{CO}_2\text{R}^{\text{aa}}$, $-\text{SO}_2\text{R}^{\text{aa}}$, $-\text{C}(\text{=NR}^{\text{cc}})\text{OR}^{\text{aa}}$, $-\text{C}(\text{=NR}^{\text{cc}})\text{N}(\text{R}^{\text{cc}})_2$, $-\text{SO}_2\text{N}(\text{R}^{\text{cc}})_2$, $-\text{SO}_2\text{R}^{\text{cc}}$, $-\text{SO}_2\text{OR}^{\text{cc}}$, $-\text{SOR}^{\text{aa}}$, $-\text{C}(\text{=S})\text{N}(\text{R}^{\text{cc}})_2$, $-\text{C}(\text{=O})\text{SR}^{\text{cc}}$, $-\text{C}(\text{=S})\text{SR}^{\text{cc}}$, $-\text{P}(\text{=O})(\text{R}^{\text{aa}})_2$, $-\text{P}(\text{=O})(\text{OR}^{\text{cc}})_2$, $-\text{P}(\text{=O})(\text{N}(\text{R}^{\text{cc}})_2)_2$, $\text{C}_{1-10} \text{ alkyl}$, $\text{C}_{1-10} \text{ perhaloalkyl}$, $\text{C}_{2-10} \text{ alkenyl}$, $\text{C}_{2-10} \text{ alkynyl}$, $\text{heteroC}_{1-10}\text{alkyl}$, $\text{heteroC}_{2-10}\text{alkenyl}$, $\text{heteroC}_{2-10}\text{alkynyl}$, $\text{C}_{3-10} \text{ carbocyclyl}$, 3-14 membered heterocyclyl, $\text{C}_{6-14} \text{ aryl}$, and 5-14 membered heteroaryl, or two R^{bb} groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups; wherein X^- is a counterion;

each instance of R^{cc} is, independently, selected from hydrogen, C_{1-10} alkyl, C_{1-10} perhaloalkyl, C_{2-10} alkenyl, C_{2-10} alkynyl, hetero C_{1-10} alkyl, hetero C_{2-10} alkenyl, hetero C_{2-10} alkynyl, C_{3-10} carbocyclyl, 3-14 membered heterocyclyl, C_{6-14} aryl, and 5-14 membered heteroaryl, or two R^{cc} groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups;

each instance of R^{dd} is, independently, selected from halogen, $-CN$, $-NO_2$, $-N_3$, $-SO_2H$, $-SO_3H$, $-OH$, $-OR^{ee}$, $-ON(R^{ff})_2$, $-N(R^{ff})_2$, $-N(R^{ff})_3^-X^-$, $-N(OR^{ee})R^{ff}$, $-SH$, $-SR^{ee}$, $-SSR^{ee}$, $-C(=O)R^{ee}$, $-CO_2H$, $-CO_2R^{ee}$, $-OC(=O)R^{ee}$, $-OCO_2R^{ee}$, $-C(=O)N(R^{ff})_2$, $-OC(=O)N(R^{ff})_2$, $-NR^{ff}C(=O)R^{ee}$, $-NR^{ff}CO_2R^{ee}$, $-NR^{ff}C(=O)N(R^{ff})_2$, $-C(=NR^{ff})OR^{ee}$, $-OC(=NR^{ff})R^{ee}$, $-OC(=NR^{ff})OR^{ee}$, $-C(=NR^{ff})N(R^{ff})_2$, $-OC(=NR^{ff})N(R^{ff})_2$, $-NR^{ff}C(=NR^{ff})N(R^{ff})_2$, $-NR^{ff}SO_2R^{ee}$, $-SO_2N(R^{ff})_2$, $-SO_2R^{ee}$, $-SO_2OR^{ee}$, $-OSO_2R^{ee}$, $-S(=O)R^{ee}$, $-Si(R^{ee})_3$, $-OSi(R^{ee})_3$, $-C(=S)N(R^{ff})_2$, $-C(=O)SR^{ee}$, $-C(=S)SR^{ee}$, $-SC(=S)SR^{ee}$, $-P(=O)(OR^{ee})_2$, $-P(=O)(R^{ee})_2$, $-OP(=O)(R^{ee})_2$, $-OP(=O)(OR^{ee})_2$, C_{1-6} alkyl, C_{1-6} perhaloalkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, hetero C_{1-6} alkyl, hetero C_{2-6} alkenyl, hetero C_{2-6} alkynyl, C_{3-10} carbocyclyl, 3-10 membered heterocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{gg} groups, or two geminal R^{dd} substituents can be joined to form $=O$ or $=S$; wherein X^- is a counterion;

each instance of R^{ee} is, independently, selected from C_{1-6} alkyl, C_{1-6} perhaloalkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, hetero C_{1-6} alkyl, hetero C_{2-6} alkenyl, hetero C_{2-6} alkynyl, C_{3-10} carbocyclyl, C_{6-10} aryl, 3-10 membered heterocyclyl, and 3-10 membered heteroaryl, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{gg} groups;

each instance of R^{ff} is, independently, selected from hydrogen, C_{1-6} alkyl, C_{1-6} perhaloalkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, hetero C_{1-6} alkyl, hetero C_{2-6} alkenyl, hetero C_{2-6} alkynyl, C_{3-10} carbocyclyl, 3-10 membered heterocyclyl, C_{6-10} aryl and 5-10 membered heteroaryl, or two R^{ff} groups are joined to form a 3-10 membered heterocyclyl or 5-10 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl,

carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{ss} groups; and

each instance of R^{ss} is, independently, halogen, -CN, -NO₂, -N₃, -SO₂H, -SO₃H, -OH, -OC₁₋₆ alkyl, -ON(C₁₋₆ alkyl)₂, -N(C₁₋₆ alkyl)₂, -N(C₁₋₆ alkyl)₃⁺X⁻, -NH(C₁₋₆ alkyl)₂⁺X⁻, -NH₂(C₁₋₆ alkyl)⁺X⁻, -NH₃⁺X⁻, -N(OC₁₋₆ alkyl)(C₁₋₆ alkyl), -N(OH)(C₁₋₆ alkyl), -NH(OH), -SH, -SC₁₋₆ alkyl, -SS(C₁₋₆ alkyl), -C(=O)(C₁₋₆ alkyl), -CO₂H, -CO₂(C₁₋₆ alkyl), -OC(=O)(C₁₋₆ alkyl), -OCO₂(C₁₋₆ alkyl), -C(=O)NH₂, -C(=O)N(C₁₋₆ alkyl)₂, -OC(=O)NH(C₁₋₆ alkyl), -NHC(=O)(C₁₋₆ alkyl), -N(C₁₋₆ alkyl)C(=O)(C₁₋₆ alkyl), -NHCO₂(C₁₋₆ alkyl), -NHC(=O)N(C₁₋₆ alkyl)₂, -NHC(=O)NH(C₁₋₆ alkyl), -NHC(=O)NH₂, -C(=NH)O(C₁₋₆ alkyl), -OC(=NH)(C₁₋₆ alkyl), -OC(=NH)OC₁₋₆ alkyl, -C(=NH)N(C₁₋₆ alkyl)₂, -C(=NH)NH(C₁₋₆ alkyl), -C(=NH)NH₂, -OC(=NH)N(C₁₋₆ alkyl)₂, -OC(NH)NH(C₁₋₆ alkyl), -OC(NH)NH₂, -NHC(NH)N(C₁₋₆ alkyl)₂, -NHC(=NH)NH₂, -NH₂SO₂(C₁₋₆ alkyl), -SO₂N(C₁₋₆ alkyl)₂, -SO₂NH(C₁₋₆ alkyl), -SO₂NH₂, -SO₂C₁₋₆ alkyl, -SO₂OC₁₋₆ alkyl, -OSO₂C₁₋₆ alkyl, -SOC₁₋₆ alkyl, -Si(C₁₋₆ alkyl)₃, -OSi(C₁₋₆ alkyl)₃, -C(=S)N(C₁₋₆ alkyl)₂, C(=S)NH(C₁₋₆ alkyl), C(=S)NH₂, -C(=O)S(C₁₋₆ alkyl), -C(=S)SC₁₋₆ alkyl, -SC(=S)SC₁₋₆ alkyl, -P(=O)(OC₁₋₆ alkyl)₂, -P(=O)(C₁₋₆ alkyl)₂, -OP(=O)(C₁₋₆ alkyl)₂, -OP(=O)(OC₁₋₆ alkyl)₂, C₁₋₆ alkyl, C₁₋₆ perhaloalkyl, C₂₋₆ alkenyl, C₂₋₆ alkynyl, heteroC₁₋₆alkyl, heteroC₂₋₆alkenyl, heteroC₂₋₆alkynyl, C₃₋₁₀ carbocyclyl, C₆₋₁₀ aryl, 3-10 membered heterocyclyl, 5-10 membered heteroaryl; or two geminal R^{ss} substituents can be joined to form =O or =S; wherein X⁻ is a counterion.

[133] A “counterion” or “anionic counterion” is a negatively charged group associated with a positively charged group in order to maintain electronic neutrality. An anionic counterion may be monovalent (*i.e.*, including one formal negative charge). An anionic counterion may also be multivalent (*i.e.*, including more than one formal negative charge), such as divalent or trivalent. Exemplary counterions include halide ions (*e.g.*, F⁻, Cl⁻, Br⁻, I⁻), NO₃⁻, ClO₄⁻, OH⁻, H₂PO₄⁻, HCO₃⁻, HSO₄⁻, sulfonate ions (*e.g.*, methanesulfonate, trifluoromethanesulfonate, *p*-toluenesulfonate, benzenesulfonate, 10-camphor sulfonate, naphthalene-2-sulfonate, naphthalene-1-sulfonic acid-5-sulfonate, ethan-1-sulfonic acid-2-sulfonate, and the like), carboxylate ions (*e.g.*, acetate, propanoate, benzoate, glycerate, lactate, tartrate, glycolate, gluconate, and the like), BF₄⁻, PF₄⁻, PF₆⁻, AsF₆⁻, SbF₆⁻, B[3,5-(CF₃)₂C₆H₃]₄⁻, B(C₆F₅)₄⁻, BPh₄⁻, Al(OC(CF₃)₃)₄⁻, and carborane anions (*e.g.*, CB₁₁H₁₂⁻ or (HCB₁₁Me₅Br₆)⁻). Exemplary counterions which may be multivalent include CO₃²⁻, HPO₄²⁻,

PO_4^{3-} , $\text{B}_4\text{O}_7^{2-}$, SO_4^{2-} , $\text{S}_2\text{O}_3^{2-}$, carboxylate anions (e.g., tartrate, citrate, fumarate, maleate, malate, malonate, gluconate, succinate, glutarate, adipate, pimelate, suberate, azelate, sebacate, salicylate, phthalates, aspartate, glutamate, and the like), and carboranes.

[134] The term “pharmaceutically acceptable salt” refers to those salts which are, within the scope of sound medical judgment, suitable for use in contact with the tissues of humans and lower animals without undue toxicity, irritation, allergic response and the like, and are commensurate with a reasonable benefit/risk ratio. Pharmaceutically acceptable salts are well known in the art. For example, Berge *et al.*, describe pharmaceutically acceptable salts in detail in *J. Pharmaceutical Sciences*, 1977, 66, 1–19, incorporated by reference. Pharmaceutically acceptable salts of the compounds disclosed in this application include those derived from suitable inorganic and organic acids and bases. Examples of pharmaceutically acceptable, nontoxic acid addition salts are salts of an amino group formed with inorganic acids such as hydrochloric acid, hydrobromic acid, phosphoric acid, sulfuric acid, and perchloric acid or with organic acids such as acetic acid, oxalic acid, maleic acid, tartaric acid, citric acid, succinic acid, or malonic acid or by using other methods known in the art such as ion exchange. Other pharmaceutically acceptable salts include adipate, alginate, ascorbate, aspartate, benzenesulfonate, benzoate, bisulfate, borate, butyrate, camphorate, camphorsulfonate, citrate, cyclopentanepropionate, digluconate, dodecylsulfate, ethanesulfonate, formate, fumarate, glucoheptonate, glycerophosphate, gluconate, hemisulfate, heptanoate, hexanoate, hydroiodide, 2-hydroxy-ethanesulfonate, lactobionate, lactate, laurate, lauryl sulfate, malate, maleate, malonate, methanesulfonate, 2-naphthalenesulfonate, nicotinate, nitrate, oleate, oxalate, palmitate, pamoate, pectinate, persulfate, 3-phenylpropionate, phosphate, picrate, pivalate, propionate, stearate, succinate, sulfate, tartrate, thiocyanate, p-toluenesulfonate, undecanoate, valerate salts, and the like. Salts derived from appropriate bases include alkali metal, alkaline earth metal, ammonium and $\text{N}^+(\text{C}_{1-4} \text{ alkyl})_4^+$ salts. Representative alkali or alkaline earth metal salts include sodium, lithium, potassium, calcium, magnesium, and the like. Further pharmaceutically acceptable salts include, when appropriate, nontoxic ammonium, quaternary ammonium, and amine cations formed using counterions such as halide, hydroxide, carboxylate, sulfate, phosphate, nitrate, lower alkyl sulfonate, and aryl sulfonate.

[135] The term “solvate” refers to forms of a compound that are associated with a solvent, usually by a solvolysis reaction. This physical association may include hydrogen

bonding. Conventional solvents include water, methanol, ethanol, acetic acid, DMSO, THF, diethyl ether, and the like. The compounds of Formula (1), (9), (10), and (11) may be prepared, *e.g.*, in crystalline form, and may be solvated. Suitable solvates include pharmaceutically acceptable solvates and further include both stoichiometric solvates and non-stoichiometric solvates. In certain instances, the solvate will be capable of isolation, for example, when one or more solvent molecules are incorporated in the crystal lattice of a crystalline solid. "Solvate" encompasses both solution-phase and isolable solvates. Representative solvates include hydrates, ethanolates, and methanolates.

[136] The term "hydrate" refers to a compound that is associated with water. Typically, the number of the water molecules contained in a hydrate of a compound is in a definite ratio to the number of the compound molecules in the hydrate. Therefore, a hydrate of a compound may be represented, for example, by the general formula $R \cdot x H_2O$, wherein R is the compound and wherein x is a number greater than 0. A given compound may form more than one type of hydrates, including, *e.g.*, monohydrates (x is 1), lower hydrates (x is a number greater than 0 and smaller than 1, *e.g.*, hemihydrates ($R \cdot 0.5 H_2O$)), and polyhydrates (x is a number greater than 1, *e.g.*, dihydrates ($R \cdot 2 H_2O$) and hexahydrates ($R \cdot 6 H_2O$)).

[137] The term "tautomers" refer to compounds that are interchangeable forms of a particular compound structure, and that vary in the displacement of hydrogen atoms and electrons. Thus, two structures may be in equilibrium through the movement of π electrons and an atom (usually H). For example, enols and ketones are tautomers because they are rapidly interconverted by treatment with either acid or base. Another example of tautomerism is the aci- and nitro- forms of phenylnitromethane, which are likewise formed by treatment with acid or base. Tautomeric forms may be relevant to the attainment of the optimal chemical reactivity and biological activity of a compound of interest.

[138] It is also to be understood that compounds that have the same molecular formula but differ in the nature or sequence of bonding of their atoms or the arrangement of their atoms in space are termed "isomers." Isomers that differ in the arrangement of their atoms in space are termed "stereoisomers."

[139] Stereoisomers that are not mirror images of one another are termed "diastereomers" and those that are non-superimposable mirror images of each other are termed

“enantiomers.” When a compound has an asymmetric center, for example, it is bonded to four different groups, a pair of enantiomers is possible. An enantiomer can be characterized by the absolute configuration of its asymmetric center and described by the R- and S-sequencing rules of Cahn and Prelog. An enantiomer can also be characterized by the manner in which the molecule rotates the plane of polarized light, and designated as dextrorotatory or levorotatory (*i.e.*, as (+) or (-)-isomers respectively). A chiral compound can exist as either an individual enantiomer or as a mixture of enantiomers. A mixture containing equal proportions of the enantiomers is called a “racemic mixture.”

[140] The term “co-crystal” refers to a crystalline structure comprising at least two different components (*e.g.*, a compound described in this application and an acid), wherein each of the components is independently an atom, ion, or molecule. In certain embodiments, none of the components is a solvent. In certain embodiments, at least one of the components is a solvent. A co-crystal of a compound and an acid is different from a salt formed from a compound and the acid. In the salt, a compound described in this application is complexed with the acid in a way that proton transfer (*e.g.*, a complete proton transfer) from the acid to a compound described in this application easily occurs at room temperature. In the co-crystal, however, a compound described in this application is complexed with the acid in a way that proton transfer from the acid to a compound described in this application does not easily occur at room temperature. In certain embodiments, in the co-crystal, there is no proton transfer from the acid to a compound described in this application. In certain embodiments, in the co-crystal, there is partial proton transfer from the acid to a compound described in this application. Co-crystals may be useful to improve the properties (*e.g.*, solubility, stability, and ease of formulation) of a compound described in this application.

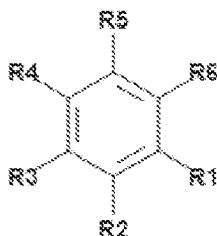
[141] The term “polymorphs” refers to a crystalline form of a compound (or a salt, hydrate, or solvate thereof) in a particular crystal packing arrangement. All polymorphs of the same compound have the same elemental composition. Different crystalline forms usually have different X-ray diffraction patterns, infrared spectra, melting points, density, hardness, crystal shape, optical and electrical properties, stability, and solubility. Recrystallization solvent, rate of crystallization, storage temperature, and other factors may cause one crystal form to dominate. Various polymorphs of a compound can be prepared by crystallization under different conditions.

[142] The term “prodrug” refers to compounds, including derivatives of the compounds of Formula (X), (8), (9), (10), or (11), that have cleavable groups and become by solvolysis or under physiological conditions the compounds of Formula (X), (8), (9), (10), or (11) and that are pharmaceutically active *in vivo*. The prodrugs may have attributes such as, without limitation, solubility, bioavailability, tissue compatibility, or delayed release in a mammalian organism. Examples include, but are not limited to, derivatives of compounds described in this application, including derivatives formed from glycosylation of the compounds described in this application (*e.g.*, glycoside derivatives), carrier-linked prodrugs (*e.g.*, ester derivatives), bioprecursor prodrugs (a prodrug metabolized by molecular modification into the active compound), and the like. Non-limiting examples of glycoside derivatives are disclosed in and incorporated by reference from PCT Publication No. WO2018/208875 and U.S. Patent Publication No. 2019/0078168. Non-limiting examples of ester derivatives are disclosed in and incorporated by reference from U.S. Patent Publication No. US2017/0362195.

[143] Other derivatives of the compounds of this invention have activity in both their acid and acid derivative forms, but the acid sensitive form often offers advantages of solubility, bioavailability, tissue compatibility, or delayed release in a mammalian organism (see, Bundgard, H., *Design of Prodrugs*, pp. 7-9, 21-24, Elsevier, Amsterdam 1985). Prodrugs include acid derivatives well known to practitioners of the art, such as, for example, esters prepared by reaction of the parent acid with a suitable alcohol, or amides prepared by reaction of the parent acid compound with a substituted or unsubstituted amine, or acid anhydrides, or mixed anhydrides. Simple aliphatic or aromatic esters, amides, and anhydrides derived from acidic groups pendant on the compounds of this invention are particular prodrugs. In some cases it is desirable to prepare double ester type prodrugs such as (acyloxy)alkyl esters or ((alkoxycarbonyl)oxy)alkylesters. C₁-C₈ alkyl, C₂-C₈ alkenyl, C₂-C₈ alkynyl, aryl, C₇-C₁₂ substituted aryl, and C₇-C₁₂ arylalkyl esters of the compounds of Formula (X), (8), (9), (10), or (11) may be preferred.

Cannabinoids

[144] As used in this application, the term “cannabinoid” includes compounds of Formula (X):

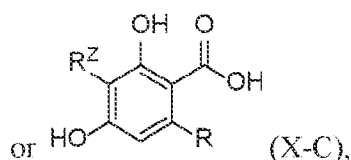
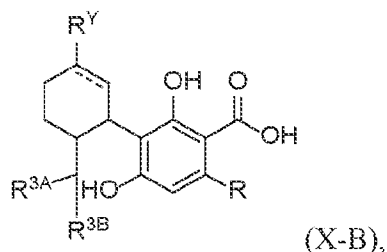
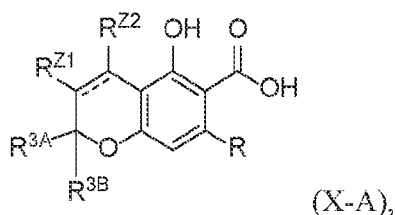


Formula (X)

or a pharmaceutically acceptable salt, co-crystal, tautomer, stereoisomer, solvate, hydrate, polymorph, isotopically enriched derivative, or prodrug thereof, wherein R1 is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl; R2 and R6 are, independently, hydrogen or carboxyl; R3 and R5 are, independently, hydroxyl, halogen, or alkoxy; and R4 is a hydrogen or an optionally substituted prenyl moiety; or optionally R4 and R3 are taken together with their intervening atoms to form a cyclic moiety, or optionally R4 and R5 are taken together with their intervening atoms to form a cyclic moiety, or optionally both 1) R4 and R3 are taken together with their intervening atoms to form a cyclic moiety and 2) R4 and R5 are taken together with their intervening atoms to form a cyclic moiety. In certain embodiments, R4 and R3 are taken together with their intervening atoms to form a cyclic moiety. In certain embodiments, R4 and R5 are taken together with their intervening atoms to form a cyclic moiety. In certain embodiments, “cannabinoid” refers to a compound of Formula (X), or a pharmaceutically acceptable salt thereof. In certain embodiments, both 1) R4 and R3 are taken together with their intervening atoms to form a cyclic moiety and 2) R4 and R5 are taken together with their intervening atoms to form a cyclic moiety.

[145] In some embodiments, cannabinoids may be synthesized via the following steps: a) one or more reactions to incorporate three additional ketone moieties onto an acyl-CoA scaffold, where the acyl moiety in the acyl-CoA scaffold comprises between four and fourteen carbons; b) a reaction cyclizing the product of step (a); and c) a reaction to incorporate a prenyl moiety to the product of step (b) or a derivative of the product of step (b). In some embodiments, non-limiting examples of the acyl-CoA scaffold described in step (a) include hexanoyl-CoA and butyryl-CoA. In some embodiments, non-limiting examples of the product of step (b) or a derivative of the product of step (b) include olivetolic acid, divarinic acid, and sphaerophorolic acid.

[146] In some embodiments, a cannabinoid compound of Formula (X) is of Formula (X-A), (X-B), or (X-C):



or a pharmaceutically acceptable salt, solvate, hydrate, polymorph, co-crystal, tautomer, stereoisomer, isotopically labeled derivative, or prodrug thereof;

wherein --- is a double bond or a single bond, as valency permits;

R is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl;

R^{Z1} is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl;

R^{Z2} is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl;

or optionally, R^{Z1} and R^{Z2} are taken together with their intervening atoms to form an optionally substituted carbocyclic ring;

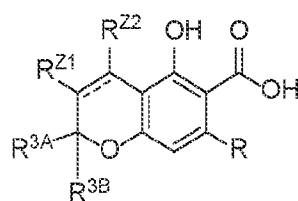
R^{3A} is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, or optionally substituted alkynyl;

R^{3B} is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, or optionally substituted alkynyl;

R^Y is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, or optionally substituted alkynyl;

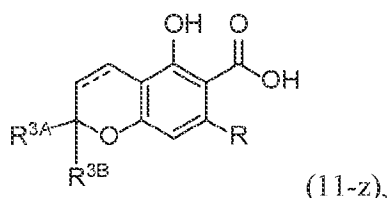
R^Z is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, or optionally substituted alkynyl.

[147] In certain embodiments, a cannabinoid compound is of Formula (X-A):

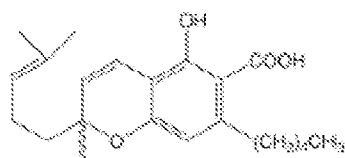


(X-A), wherein --- is a double bond, and each of R^{Z1} and R^{Z2} is hydrogen, one of R^{3A} and R^{3B} is optionally substituted C₂₋₆ alkenyl, and the other one of R^{3A} and R^{3B} is optionally substituted C₂₋₆ alkyl. In some embodiments, a cannabinoid compound of Formula (X) is of Formula (X-A), wherein each of R^{Z1} and R^{Z2} is hydrogen, one of R^{3A} and R^{3B} is a prenyl group, and the other one of R^{3A} and R^{3B} is optionally substituted methyl.

[148] In certain embodiments, a cannabinoid compound of Formula (X) of Formula (X-A) is of Formula (11-z):

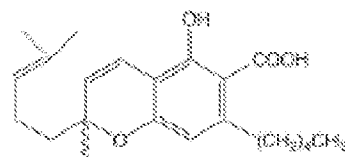


wherein --- is a double bond or single bond, as valency permits; one of R^{3A} and R^{3B} is C₁₋₆ alkyl optionally substituted with alkenyl, and the other of R^{3A} and R^{3B} is optionally substituted C₁₋₆ alkyl. In certain embodiments, in a compound of Formula (11-z), --- is a single bond; one of R^{3A} and R^{3B} is C₁₋₆ alkyl optionally substituted with prenyl; and the other of one of R^{3A} and R^{3B} is unsubstituted methyl; and R is as described in this application. In certain embodiments, in a compound of Formula (11-z), --- is a single bond; one of R^{3A} and R^{3B} is ; and the other of one of R^{3A} and R^{3B} is unsubstituted methyl; and R is as described in this application. In certain embodiments, a cannabinoid compound of Formula (11-z) is of Formula (11a):



(11a).

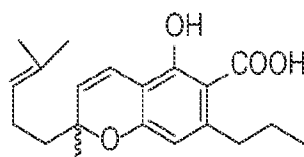
[149] In certain embodiments, a cannabinoid compound of Formula (X) of Formula



(X-A) is of Formula (11a):

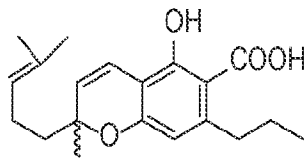
(11a).

[150] In certain embodiments, a cannabinoid compound of Formula (11-z) is of Formula (11b):



(11b).

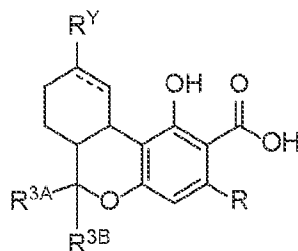
[151] In certain embodiments, a cannabinoid compound of Formula (X) of Formula



(X-A) is of Formula (11b):

(11b).

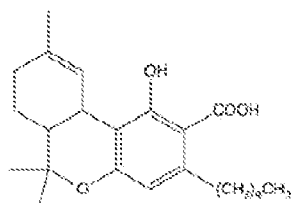
[152] In certain embodiments, a cannabinoid compound of Formula (X-A) is of



Formula (10-z):

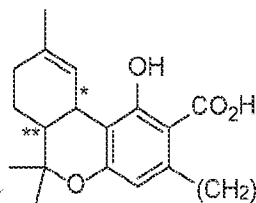
(10-z), wherein --- is a double bond or single bond,

as valency permits; R^Y is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, or optionally substituted alkynyl; and each of R^{3A} and R^{3B} is independently optionally substituted C_{1-6} alkyl. In certain embodiments, in a compound of Formula (10-z), --- is a single bond; each of R^{3A} and R^{3B} is unsubstituted methyl, and R is as described in this application. In certain embodiments, a cannabinoid compound of Formula

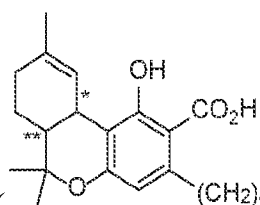


(10-z) is of Formula (10a):

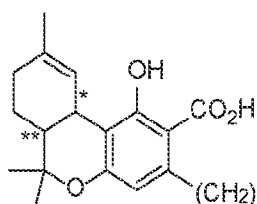
(10a). In certain embodiments, a



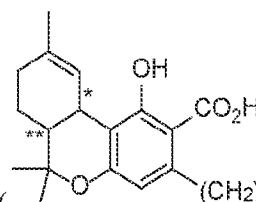
compound of Formula (10a) () has a chiral atom labeled with * at carbon 10 and a chiral atom labeled with ** at carbon 6. In certain embodiments, in a compound



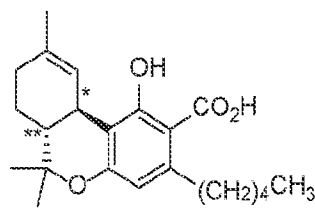
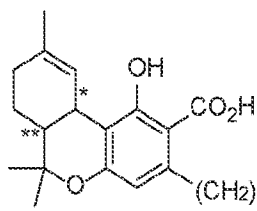
of Formula (10a) (), the chiral atom labeled with * at carbon 10 is of the *R*-configuration or *S*-configuration; and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration. In certain embodiments, in a compound of Formula (10a) (



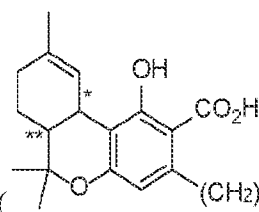
), the chiral atom labeled with * at carbon 10 is of the *S*-configuration; and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration or *S*-configuration. In

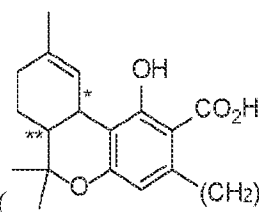


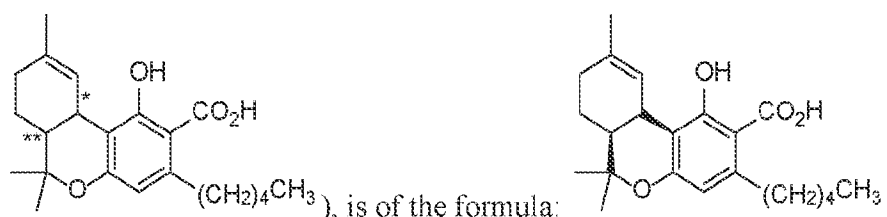
certain embodiments, in a compound of Formula (10a) (), the chiral atom labeled with * at carbon 10 is of the *R*-configuration and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration. In certain embodiments, a compound of Formula (10a) (



is of the formula: In certain

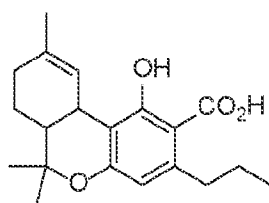


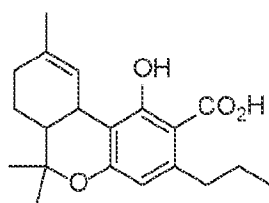
embodiments, in a compound of Formula (10a) (, the chiral atom labeled with * at carbon 10 is of the *S*-configuration and a chiral atom labeled with ** at carbon 6 is of the *S*-configuration. In certain embodiments, a compound of Formula (10a) (

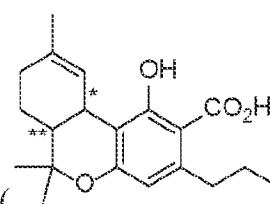


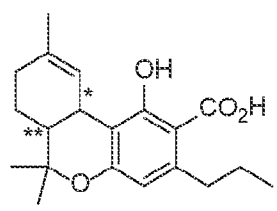
), is of the formula:

[153] In certain embodiments, a cannabinoid compound of Formula (10-z) is of

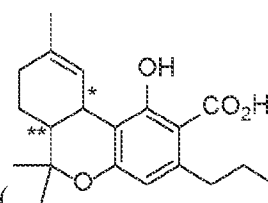


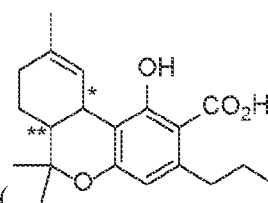
Formula (10b): (, (10b). In certain embodiments, a compound of Formula

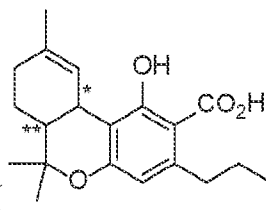
(10b) () has a chiral atom labeled with * at carbon 10 and a chiral atom labeled with ** at carbon 6. In certain embodiments, in a compound of Formula (10b) (



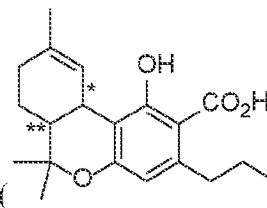
), the chiral atom labeled with * at carbon 10 is of the *R*-configuration or *S*-configuration; and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration. In



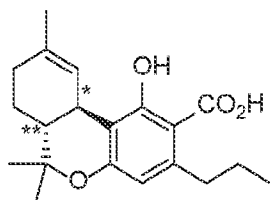
certain embodiments, in a compound of Formula (10b) (, the chiral atom labeled with * at carbon 10 is of the *S*-configuration; and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration or *S*-configuration. In certain embodiments, in a compound of



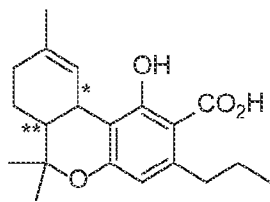
Formula (10b) (), the chiral atom labeled with * at carbon 10 is of the *R*-configuration and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration. In certain



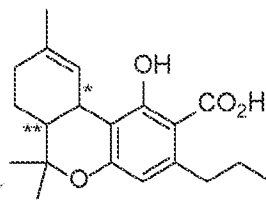
embodiments, a compound of Formula (10b) (), is of the formula:



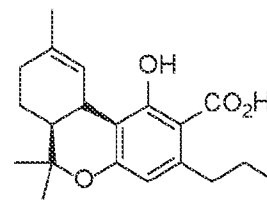
. In certain embodiments, in a compound of Formula (10b) (



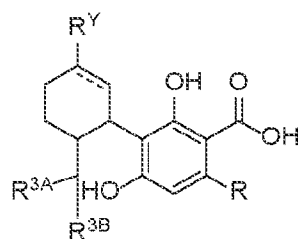
), the chiral atom labeled with * at carbon 10 is of the *S*-configuration and a chiral atom labeled with ** at carbon 6 is of the *S*-configuration. In certain embodiments,



a compound of Formula (10b) (), is of the formula:

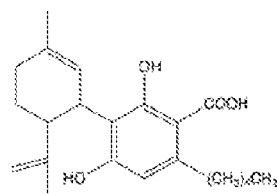


[154] In certain embodiments, a cannabinoid compound is of Formula (X-B):



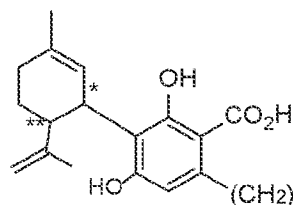
(X-B), wherein is a double bond; R^Y is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, or optionally substituted alkynyl; and each of R^{3A} and R^{3B} is independently optionally substituted C_{1-6} alkyl.

In certain embodiments, in a compound of Formula (X-B), R^Y is optionally substituted C_{1-6} alkyl; one of R^{3A} and R^{3B} is ; and the other one of R^{3A} and R^{3B} is unsubstituted methyl, and R is as described in this application. In certain embodiments, a compound of Formula (X-

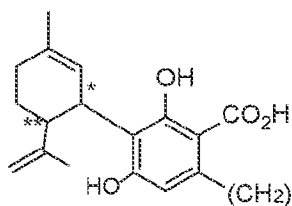


B) is of Formula (9a):

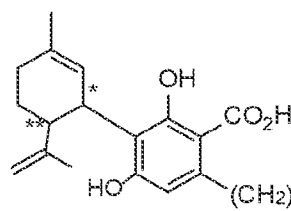
(9a). In certain embodiments, a compound



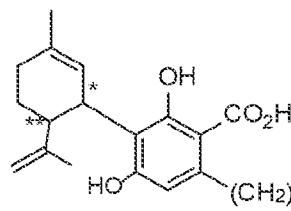
of Formula (9a) () has a chiral atom labeled with * at carbon 3 and a chiral atom labeled with ** at carbon 4. In certain embodiments, in a compound of



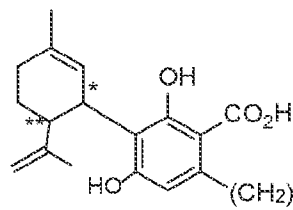
Formula (9a) (), the chiral atom labeled with * at carbon 3 is of the *R*-configuration or *S*-configuration; and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration. In certain embodiments, in a compound of Formula (9a) (



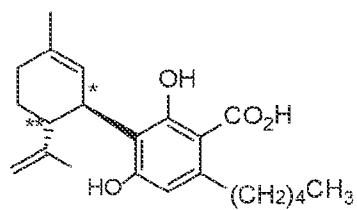
the chiral atom labeled with * at carbon 3 is of the *S*-configuration; and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration or *S*-configuration. In certain embodiments, in a compound of Formula (9a) (



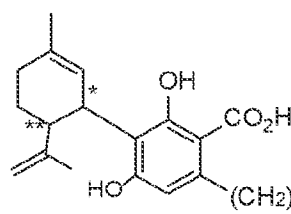
the chiral atom labeled with * at carbon 3 is of the *R*-configuration and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration. In certain



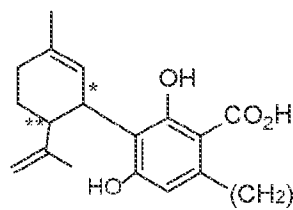
embodiments, a compound of Formula (9a) (



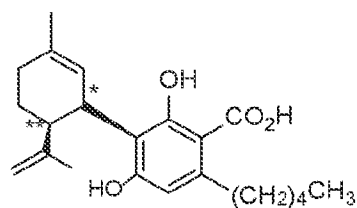
. In certain embodiments, in a compound of Formula (9a) (



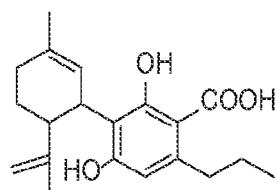
), the chiral atom labeled with * at carbon 3 is of the *S*-configuration and a chiral atom labeled with ** at carbon 4 is of the *S*-configuration. In certain



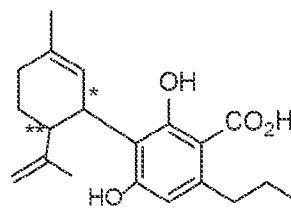
embodiments, a compound of Formula (9a) (



[155] In certain embodiments, a compound of Formula (X-B) is of Formula (9b):

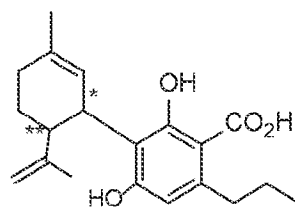


(9b). In certain embodiments, a compound of Formula (9b) (

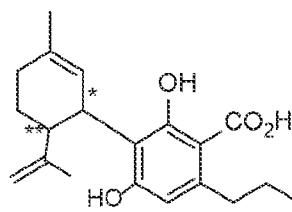


) has a chiral atom labeled with * at carbon 3 and a chiral atom labeled

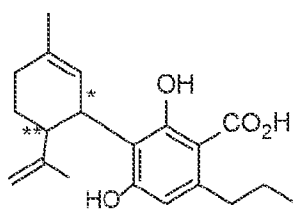
with ** at carbon 4. In certain embodiments, in a compound of Formula (9b) (



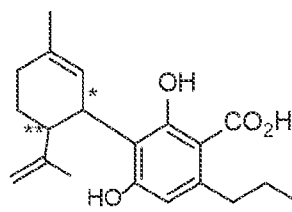
), the chiral atom labeled with * at carbon 3 is of the *R*-configuration or *S*-configuration; and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration. In



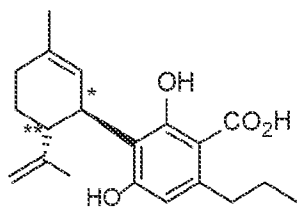
certain embodiments, in a compound of Formula (9b) (), the chiral atom labeled with * at carbon 3 is of the *S*-configuration; and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration or *S*-configuration. In certain embodiments, in a compound



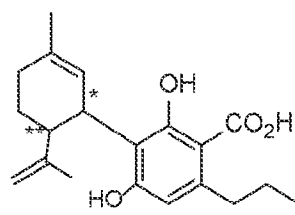
of Formula (9b) (), the chiral atom labeled with * at carbon 3 is of the *R*-configuration and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration. In



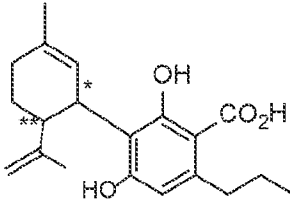
certain embodiments, a compound of Formula (9b) (), is of the

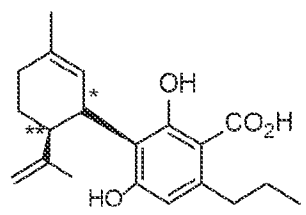


formula: . In certain embodiments, in a compound of Formula (9a) (

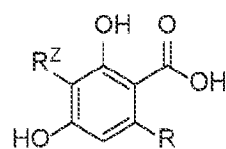


), the chiral atom labeled with * at carbon 3 is of the *S*-configuration and a chiral atom labeled with ** at carbon 4 is of the *S*-configuration. In certain embodiments,

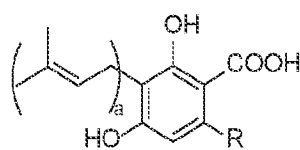
a compound of Formula (9b) () is of the formula:



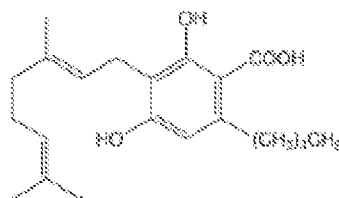
[156] In certain embodiments, a cannabinoid compound is of Formula (X-C):



(X-C), wherein R^Z is optionally substituted alkyl or optionally substituted alkenyl. In certain embodiments, a compound of Formula (X-C) is of formula:



(8'), wherein a is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. In certain embodiments, a is 1. In certain embodiments, a is 2. In certain embodiments, a is 3. In certain embodiments, a is 1, 2, or 3 for a compound of Formula (X-C). In certain embodiments, a cannabinoid compound is of Formula (X-C), and a is 1, 2, 3, 4, or 5. In certain embodiments, a compound



of Formula (X-C) is of Formula (8a): (8a).

[157] In some embodiments, cannabinoids of the present disclosure comprise cannabinoid receptor ligands. Cannabinoid receptors are a class of cell membrane receptors in the G protein-coupled receptor superfamily. Cannabinoid receptors include the CB_1 receptor and the CB_2 receptor. In some embodiments, cannabinoid receptors comprise GPR18, GPR55, and PPAR. (See Bram et al. "Activation of GPR18 by cannabinoid compounds: a tale of biased agonism" *Br J Pharmacol* v171 (16) (2014); Shi et al. "The novel cannabinoid receptor GPR55 mediates anxiolytic-like effects in the medial orbital cortex of mice with acute stress"

Molecular Brain 10, No. 38 (2017); and O'Sullivan, Elizabeth. "An update on PPAR activation by cannabinoids" *Br J Pharmacol* v. 173(12) (2016)).

[158] In some embodiments, cannabinoids comprise endocannabinoids, which are substances produced within the body, and phytocannabinoids, which are cannabinoids that are naturally produced by plants of genus *Cannabis*. In some embodiments, phytocannabinoids comprise the acidic and decarboxylated acid forms of the naturally-occurring plant-derived cannabinoids, and their synthetic and biosynthetic equivalents.

[159] Over 94 phytocannabinoids have been identified to date (Berman, Paula, et al. "A new ESI-LC/MS approach for comprehensive metabolic profiling of phytocannabinoids in Cannabis." *Scientific reports* 8.1 (2018): 14280; El-Alfy et al., 2010, "Antidepressant-like effect of delta-9-tetrahydrocannabinol and other cannabinoids isolated from *Cannabis sativa* L", *Pharmacology Biochemistry and Behavior* 95 (4): 434-42; Rudolf Brenneisen, 2007, *Chemistry and Analysis of Phytocannabinoids*, Citti, Cinzia, et al. "A novel phytocannabinoid isolated from Cannabis sativa L. with an in vivo cannabimimetic activity higher than Δ^9 -tetrahydrocannabinol: Δ^9 -Tetrahydrocannabiphorol." *Sci Rep* 9 (2019): 20335, each of which is incorporated by reference in this application in its entirety). In some embodiments, cannabinoids comprise Δ^9 - tetrahydrocannabinol (THC) type (e.g., (-)-trans-delta-9-tetrahydrocannabinol or dronabinol, (+)-trans-delta-9-tetrahydrocannabinol, (-)-cis-delta-9-tetrahydrocannabinol, or (+)-cis-delta-9-tetrahydrocannabinol), cannabidiol (CBD) type, cannabigerol (CBG) type, cannabichromene (CBC) type, cannabicyclol (CBL) type, cannabinodiol (CBND) type, or cannabitrilol (CBT) type cannabinoids, or any combination thereof (see, e.g., R Pertwee, ed, *Handbook of Cannabis* (Oxford, UK: Oxford University Press, 2014)), which is incorporated by reference in this application in its entirety). A non-limiting list of cannabinoids comprises: cannabiorcol-C1 (CBNO), CBND-C1 (CBNDO), Δ^9 -trans-Tetrahydrocannabiorcolic acid-C1 (Δ^9 -THCO), Cannabidiorcol-C1 (CBDO), Cannabiorchromene-C1 (CBCO), (-)- Δ^8 -trans-(6aR,10aR)-Tetrahydrocannabiorcol-C1 (Δ^8 -THCO), Cannabiorcyclol C1 (CBLO), CBG-C1 (CBGO), Cannabinol-C2 (CBN-C2), CBND-C2, Δ^9 -THC-C2, CBD-C2, CBC-C2, Δ^8 -THC-C2, CBL-C2, Bisnor-cannabielsoin-C1 (CBEO), CBG-C2, Cannabivarin-C3 (CBNV), Cannabinodivarin-C3 (CBNDV), (-)- Δ^9 -trans-Tetrahydrocannabivarin-C3 (Δ^9 -THCV), (-)-Cannabidivarin-C3 (CBDV), ($\pm$)-Cannabichromevarin-C3 (CBCV), (-)- Δ^8 -trans-THC-C3 (Δ^8 -THCV), ($\pm$)-(1aS,3aR,8bR,8cR)-Cannabicyclovarin-C3 (CBLV), 2-Methyl-2-(4-methyl-2-pentenyl)-7-propyl-2H-1-

benzopyran-5-ol, Δ^7 -tetrahydrocannabivarin-C3 (Δ^7 -THCV), CBE-C2, Cannabigerovarin-C3 (CBGV), Cannabitrinol-C1 (CBTO), Cannabinol-C4 (CBN-C4), CBND-C4, (-)- Δ^9 -*trans*-Tetrahydrocannabinol-C4 (Δ^9 -THC-C4), Cannabidiol-C4 (CBD-C4), CBC-C4, (-)-*trans*- Δ^8 -THC-C4, CBL-C4, Cannabielsoin-C3 (CBEV), CBG-C4, CBT-C2, Cannabichromanone-C3, Cannabiglendol-C3 (OH-iso-HHCV-C3), Cannabioxepane-C5 (CBX), Dehydrocannabifuran-C5 (DCBF), Cannabinol-C5 (CBN), Cannabinodiol-C5 (CBND), (-)- Δ^9 -*trans*-Tetrahydrocannabinol-C5 (Δ^9 -THC), (-)- Δ^8 -*trans*-(6aR,10aR)-Tetrahydrocannabinol-C5 (Δ^8 -THC), ($\pm$)-Cannabichromene-C5 (CBC), (-)-Cannabidiol-C5 (CBD), ($\pm$)-(1aS,3aR,8bR,8cR)-Cannabicyclol-C5 (CBL), Cannabicitran-C5 (CBR), (-)- Δ^9 -(6aS,10aR-*cis*)-Tetrahydrocannabinol-C5 ((-)-*cis*- Δ^9 -THC), (-)- Δ^7 -*trans*-(1R,3R,6R)-Isotetrahydrocannabinol-C5 (*trans*-iso Δ^7 -THC), CBE-C4, Cannabigerol-C5 (CBG), Cannabitrinol-C3 (CBTV), Cannabinol methyl ether-C5 (CBNM), CBNDM-C5, 8-OH-CBN-C5 (OH-CBN), OH-CBND-C5 (OH-CBND), 10-Oxo- $\Delta^{6a(10a)}$ -Tetrahydrocannabinol-C5 (OTHC), Cannabichromanone D-C5, Cannabicumaronone-C5 (CBCON-C5), Cannabidiol monomethyl ether-C5 (CBDM), Δ^9 -THCM-C5, ($\pm$)-3"-hydroxy- Δ^{4m} -cannabichromene-C5, (5aS,6S,9R,9aR)-Cannabielsoin-C5 (CBE), 2-geranyl-5-hydroxy-3-n-pentyl-1,4-benzoquinone-C5, 5-geranyl olivetolic acid, 5-geranyl olivetolate, 8 α -Hydroxy- Δ^9 -Tetrahydrocannabinol-C5 (8 α -OH- Δ^9 -THC), 8 β -Hydroxy- Δ^9 -Tetrahydrocannabinol-C5 (8 β -OH- Δ^9 -THC), 10 α -Hydroxy- Δ^8 -Tetrahydrocannabinol-C5 (10 α -OH- Δ^8 -THC), 10 β -Hydroxy- Δ^8 -Tetrahydrocannabinol-C5 (10 β -OH- Δ^8 -THC), 10 α -hydroxy- $\Delta^{9,11}$ -hexahydrocannabinol-C5, 9 β ,10 β -Epoxyhexahydrocannabinol-C5, OH-CBD-C5 (OH-CBD), Cannabigerol monomethyl ether-C5 (CBGM), Cannabichromanone-C5, CBT-C4, ($\pm$)-6,7-*cis*-epoxycannabigerol-C5, ($\pm$)-6,7-*trans*-epoxycannabigerol-C5, (-)-7-hydroxycannabichromane-C5, Cannabimovone-C5, (-)-*trans*-Cannabitrinol-C5 ((-)-*trans*-CBT), (+)-*trans*-Cannabitrinol-C5 ((+)-*trans*-CBT), ($\pm$)-*cis*-Cannabitrinol-C5 (($\pm$)-*cis*-CBT), (-)-*trans*-10-Ethoxy-9-hydroxy- $\Delta^{6a(10a)}$ -tetrahydrocannabivarin-C3 [(-)-*trans*-CBT-OEt], (-)-(6aR,9S,10S,10aR)-9,10-Dihydroxyhexahydrocannabinol-C5 [(-)-Cannabiripsol] (CBR), Cannabichromanone C-C5, (-)-6a,7,10a-Trihydroxy- Δ^9 -tetrahydrocannabinol-C5 [(-)-Cannabitetrol] (CBTT), Cannabichromanone B-C5, 8,9-Dihydroxy- $\Delta^{6a(10a)}$ -tetrahydrocannabinol-C5 (8,9-Di-OHCBT), ($\pm$)-4-acetoxycannabichromene-C5, 2-acetoxy-6-geranyl-3-n-pentyl-1,4-benzoquinone-C5, 11-Acetoxy- Δ^9 -Tetrahydrocannabinol-C5 (11-OAc- Δ^9 -THC), 5-acetyl-4-hydroxycannabigerol-C5, 4-acetoxy-2-geranyl-5-hydroxy-3-n-pentylphenol-C5, (-)-*trans*-10-Ethoxy-9-hydroxy- $\Delta^{6a(10a)}$ -tetrahydrocannabinol-C5 ((-)-*trans*-CBTOEt),

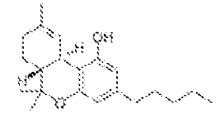
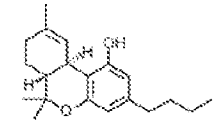
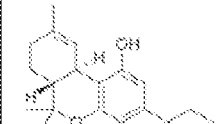
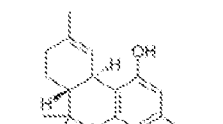
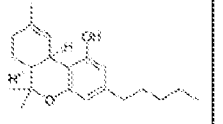
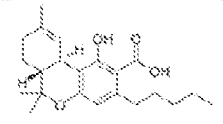
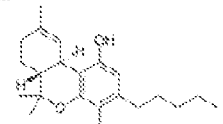
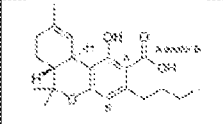
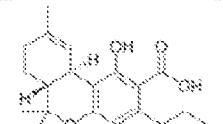
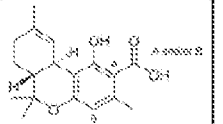
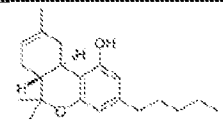
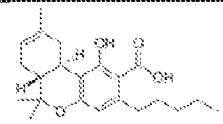
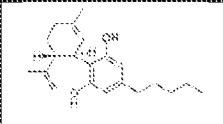
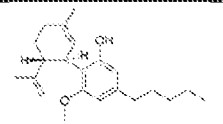
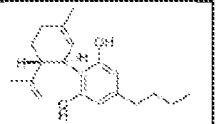
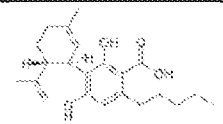
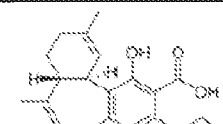
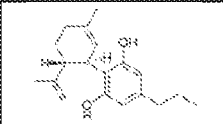
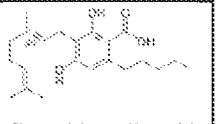
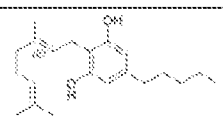
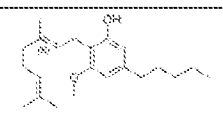
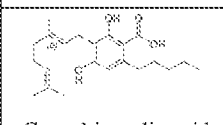
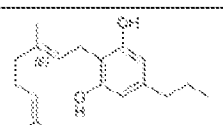
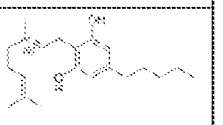
sesquicannabigerol-C5 (SesquiCBG), carmagerol-C5, 4-terpenyl cannabinolate-C5, β -fenchyl- Δ^9 -tetrahydrocannabinolate-C5, α -fenchyl- Δ^9 -tetrahydrocannabinolate-C5, epi-bornyl- Δ^9 -tetrahydrocannabinolate-C5, bornyl- Δ^9 -tetrahydrocannabinolate-C5, α -terpenyl- Δ^9 -tetrahydrocannabinolate-C5, 4-terpenyl- Δ^9 -tetrahydrocannabinolate-C5, 6,6,9-trimethyl-3-pentyl-6H-dibenzo[b,d]pyran-1-ol, 3-(1,1-dimethylheptyl)-6,6a,7,8,10,10a-hexahydro-1-hydroxy-6,6-dimethyl-9H-dibenzo[b,d]pyran-9-one, (–)-(3S,4S)-7-hydroxy- Δ^6 -tetrahydrocannabinol-1,1-dimethylheptyl, (+)-(3S,4S)-7-hydroxy- Δ^6 -tetrahydrocannabinol-1,1-dimethylheptyl, 11-hydroxy- Δ^9 -tetrahydrocannabinol, and Δ^8 -tetrahydrocannabinol-11-oic acid)); certain piperidine analogs (e.g., (–)-(6S,6aR,9R,10aR)-5,6,6a,7,8,9,10,10a-octahydro-6-methyl-3-[(R)-1-methyl-4-phenylbutoxy]-1,9-phenanthridinediol 1-acetate)), certain aminoalkylindole analogs (e.g., (R)-(+)-[2,3-dihydro-5-methyl-3-(4-morpholinylmethyl)-pyrrolo[1,2,3-de]-1,4-benzoxazin-6-yl]-1-naphthalenyl-methanone), certain open pyran ring analogs (e.g., 2-[3-methyl-6-(1-methylethenyl)-2-cyclohexen-1-yl]-5-pentyl-1,3-benzenediol and 4-(1,1-dimethylheptyl)-2,3'-dihydroxy-6' α -(3-hydroxypropyl)-1',2',3',4',5',6'-hexahydrobiphenyl, tetrahydrocannabiphorol (THCP), cannabidiphorol (CBDP), CBGP, CBCP, their acidic forms, salts of the acidic forms, dimers of any combination of the above, trimers of any combination of the above, polymers of any combination of the above, or any combination thereof.

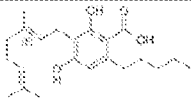
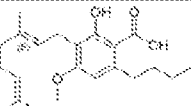
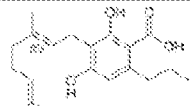
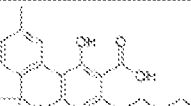
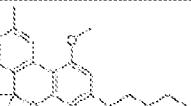
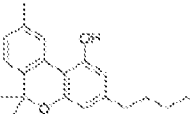
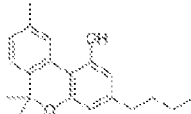
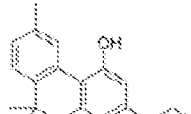
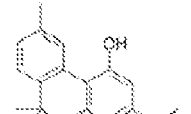
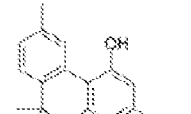
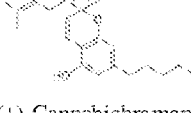
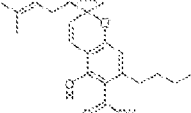
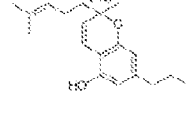
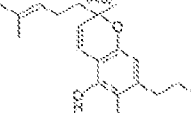
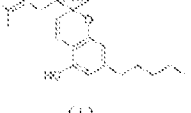
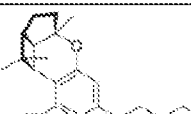
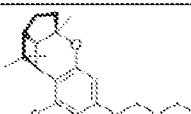
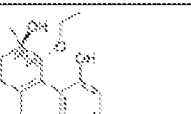
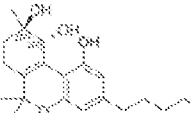
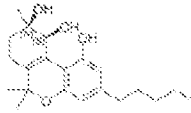
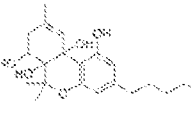
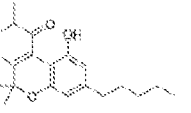
[160] A cannabinoid described in this application can be a rare cannabinoid. For example, in some embodiments, a cannabinoid described in this application corresponds to a cannabinoid that is naturally produced in conventional *Cannabis* varieties at concentrations of less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.25%, or 0.1% by dry weight of the female flower. In some embodiments, rare cannabinoids include CBGA, CBGVA, THCVA, CBDVA, CBCVA, and CBCA. In some embodiments, rare cannabinoids are cannabinoids that are not THCA, THC, CBDA or CBD.

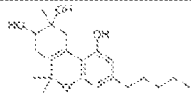
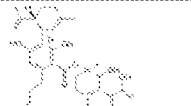
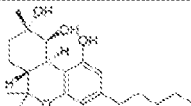
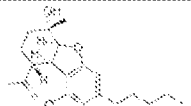
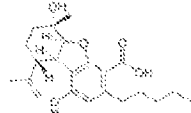
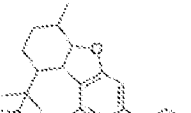
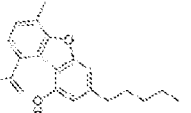
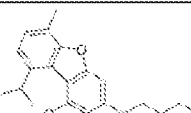
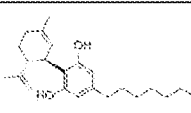
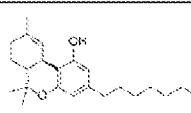
[161] A cannabinoid described in this application can also be a non-rare cannabinoid.

In some embodiments, the cannabinoid is selected from the cannabinoids listed in **Table 1**.

Table 1. Non-limiting examples of cannabinoids according to the present disclosure.

 Δ ⁹ -Tetrahydro- cannabinol Δ ⁹ -THC-C ₅	 Δ ⁹ -Tetrahydro- cannabinol-C ₄ Δ ⁹ -THC-C ₄	 Δ ⁹ -Tetrahydro- cannabivarin Δ ⁹ -THCV-C ₃	 Δ ⁹ -Tetrahydro- cannabiorcol Δ ⁹ -THCO-C ₁	 (-)-(6a <i>S</i> ,10a <i>R</i>)-Δ ⁹ - Tetrahydro- cannabinol (-)- <i>cis</i> -Δ ⁹ -THC-C ₅
 Δ ⁹ -Tetrahydro- cannabinolic acid A Δ ⁹ -THCA-C ₅ A	 Δ ⁹ -Tetrahydro- cannabinolic acid B Δ ⁹ -THCA-C ₅ B	 Δ ⁹ -Tetrahydro- cannabinolic acid-C ₄ A and/or B Δ ⁹ -THCA-C ₄ A and/or B	 Δ ⁹ -Tetrahydro- cannabivarinic acid A Δ ⁹ -THCVA-C ₃ A	 Δ ⁹ -Tetrahydro- cannabiorcolic acid A and/or B Δ ⁹ -THCOA-C ₁ A and/or B
 (-)-Δ ⁸ - <i>trans</i> - (6a <i>R</i> ,10a <i>R</i>)- Δ ⁸ -Tetrahydro- cannabinol Δ ⁸ -THC-C ₅	 (-)-Δ ⁸ - <i>trans</i> - (6a <i>R</i> ,10a <i>R</i>)- Tetrahydro- cannabinolic acid A Δ ⁸ -THCA-C ₅ A	 (-)-Cannabidiol CBD-C ₅	 Cannabidiol monomethyl ether CBDM-C ₅	 Cannabidiol-C ₄ CBD-C ₄
 Cannabidiolic acid CBDA-C ₅	 Cannabidivarinic acid CBDVA-C ₃	 (-)-Cannabidivarin CBDV-C ₃	 Cannabidiorcol CBD-C ₁	 Cannabigerolic acid A (<i>E</i>)-CBGA-C ₅ A
 Cannabigerol (<i>E</i>)-CBG-C ₅	 Cannabigerol monomethyl ether (<i>E</i>)-CBGM-C ₅ A	 Cannabinerolic acid A (<i>Z</i>)-CBGA-C ₅ A	 Cannabigerovarin (<i>E</i>)-CBGV-C ₃	 Cannabigerol (<i>E</i>)-CBG-C ₅

				
Cannabigerolic acid A (<i>E</i>)-CBGA-C ₅ A	Cannabigerolic acid A monomethyl ether (<i>E</i>)-CBGAM-C ₅ A	Cannabigerovarinic acid A (<i>E</i>)-CBGVA-C ₃ A	Cannabinolic acid A CBNA-C ₅ A	Cannabinol methyl ether CBNM-C ₅
				
Cannabinol CBN-C ₅	Cannabinol-C4 CBN-C ₄	Cannabivarin CBN-C ₃	Cannabinol-C2 CBN-C ₂	Cannabiorcol CBN-C ₁
				
(±)-Cannabichromene CBC-C ₅	(±)-Cannabichromenic acid A CBCA-C ₅ A	(±)-Cannabivarichrome, (±)- Cannabichromevarin CBCV-C ₃	(±)-Cannabichromevarinic acid A CBCVA-C ₃ A	(±)- Cannabichromene CBC-C ₅
				
(±)- (1 <i>aS</i> ,3 <i>aR</i> ,8 <i>bR</i> ,8 <i>cR</i>)- Cannabicyclol CBL-C ₅	(±)- (1 <i>aS</i> ,3 <i>aR</i> ,8 <i>bR</i> ,8 <i>cR</i>)- Cannabicyclic acid A CBLA-C ₅ A	(±)- (1 <i>aS</i> ,3 <i>aR</i> ,8 <i>bR</i> ,8 <i>cR</i>)- Cannabicyclovarin CBLV-C ₃	(-)-(9 <i>R</i> ,10 <i>R</i>)-trans- 10-O-Ethyl- cannabitrinol (-)-trans-CBT-OEt- C ₅	(±)- (9 <i>R</i> ,10 <i>R</i> /9 <i>S</i> ,10 <i>S</i>)- Cannabitrinol-C ₃ (±)-trans-CBT-C ₃
				
(-)-(9 <i>R</i> ,10 <i>R</i>)-trans- Cannabitrinol (-)-trans-CBT-C ₅	(+)-(9 <i>S</i> ,10 <i>S</i>)- Cannabitrinol (+)-trans-CBT-C ₅	(±)- (9 <i>R</i> ,10 <i>S</i> /9 <i>S</i> ,10 <i>R</i>)- Cannabitrinol (±)-cis-CBT-C ₅	(-)-6 <i>a</i> ,7,10 <i>a</i> - Trihydroxy- Δ ⁹ -tetrahydro- cannabinol (-)-Cannabitetrol	10-Oxo-Δ ^{6a} (10 <i>a</i>)- tetrahydro- cannabinol OTHC

 8,9-Dihydroxy- Δ6a(10a)- tetrahydro-cannabinol 8,9-Di-OH-CBT-C5	 Cannabidiolic acid A cannabitrinol ester CBDA-C5 9-OH- CBT-C5 ester	 (-)- (6aR,9S,10S,10aR)- 9,10-Dihydroxy- hexahydro- cannabinol, Cannabiripsol Cannabiripsol-C5	 (5aS,6S,9R,9aR)- Cannabielsoic acid B CBEA-C5 B	 (5aS,6S,9R,9aR)- C3-Cannabielsoic acid B CBEA-C3 B
 (5aS,6S,9R,9aR)- Cannabielsoin CBE-C5	 (5aS,6S,9R,9aR)- C3-Cannabielsoin CBE-C3	 (5aS,6S,9R,9aR)- Cannabielsoic acid A CBEA-C5 A	 Cannabiglendol-C3 OH-iso-HHCV-C3	 Dehydro- cannabifuran DCBF-C5
 Cannabifuran CBF-C5	 Cannabidiphorol (CBDP)	 Tetrahydro- cannabiphorol (THCP)		

[162] Cannabinoids are often classified by “type,” i.e., by the topological arrangement of their prenyl moieties (See, for example, M. A. Elsohly and D. Slade, *Life Sci.*, 2005, 78, 539–548; and L.O. Hanus et al. *Nat. Prod. Rep.*, 2016, 33, 1357). Generally, each “type” of cannabinoid includes the variations possible for ring substitutions of the resorcinol moiety at the position meta to the two hydroxyl moieties. As used in this disclosure, a “CBG-type” cannabinoid is a 3-[(2E)-3,7-dimethylocta-2,6-dienyl]-2,4-dihydroxybenzoic acid optionally substituted at the 6 position of the benzoic acid moiety. As used in this disclosure, “CBC-type” cannabinoids refer to 5-hydroxy-2-methyl-2-(4-methylpent-3-enyl)-chromene-6-carboxylic acid optionally substituted at the 7 position of the chromene moiety. As used in this disclosure, a “THC-type” cannabinoid is a (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-6a,7,8,10a-tetrahydrobenzo[c]chromene-2-carboxylic acid optionally substituted at the 3 position of the benzo[c]chromene moiety. As used in this disclosure, a “CBD-type” cannabinoid is a 2,4-dihydroxy-3-[(1R,6R)-3-methyl-6-prop-1-en-2-ylcyclohex-2-en-1-yl]-benzoic acid optionally substituted at the 6 position of the benzoic acid moiety. In some embodiments, the optional

ring substitution for each “type” is an optionally substituted C1-C11 alkyl, an optionally substituted C1-C11 alkenyl, an optionally substituted C1-C11 alkynyl, or an optionally substituted C1-C11 aralkyl.

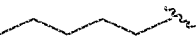
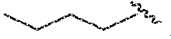
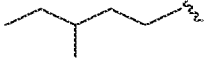
[163] The terms “varinolic cannabinoid” and “varin cannabinoid” are interchangeable, and mean a cannabinoid that is a derivative of divaric acid or divarinol, a cannabinoid of Formula (X) where R1 is propyl (*e.g.*, *n*-propyl), a cannabinoid of Formula (X-A), (X-B), (X-C), (11-z), (10-z), where R is propyl (*e.g.*, *n*-propyl), or any combination of thereof. Exemplary, varinolic cannabinoids and varin cannabinoids include, but are not limited to, CBGV, CBCV (cannabichromevarin), CBDV, CBGVA, THCV, THCVA and/or CBCVA.

Biosynthesis of Cannabinoids and Cannabinoid Precursors

[164] Aspects of the present disclosure provide tools, sequences, and methods for the biosynthetic production of cannabinoids in host cells. In some embodiments, the present disclosure teaches expression of enzymes that are capable of producing cannabinoids by biosynthesis.

[165] As a non-limiting example, one or more of the enzymes depicted in **FIG. 2** may be used to produce a cannabinoid or cannabinoid precursor of interest. **FIG. 1** shows a cannabinoid biosynthesis pathway for the most abundant phytocannabinoids found in *Cannabis*. See also, de Meijer et al. I, II, III, and IV (I: 2003, *Genetics*, 163:335-346; II: 2005, *Euphytica*, 145:189-198; III: 2009, *Euphytica*, 165:293-311; and IV: 2009, *Euphytica*, 168:95-112), and Carvalho et al. “Designing Microorganisms for Heterologous Biosynthesis of Cannabinoids” (2017) *FEMS Yeast Research* Jun 1;17(4), each of which is incorporated by reference in this application in its entirety. **FIG. 4** shows a biosynthetic pathway for production of varin cannabinoid compounds.

[166] It should be appreciated that a precursor substrate for use in cannabinoid biosynthesis is generally selected based on the cannabinoid of interest. Non-limiting examples of cannabinoid precursors include compounds of Formulae (1)-(8) in **FIG. 2**. In some embodiments, polyketides, including compounds of Formula (5), could be prenylated. In certain embodiments, the precursor is a precursor compound shown in **FIGs. 1- 4**. Substrates in which R contains 1-40 carbon atoms are preferred. In some embodiments, substrates in which R contains 3-8 carbon atoms are most preferred.

[167] As used in this application, a cannabinoid or a cannabinoid precursor may comprise an R group. See, *e.g.*, **FIG. 2**. In some embodiments, R may be a hydrogen. In certain embodiments, R is optionally substituted alkyl. In certain embodiments, R is optionally substituted C1-40 alkyl. In certain embodiments, R is optionally substituted C2-40 alkyl. In certain embodiments, R is optionally substituted C2-40 alkyl, which is straight chain or branched alkyl. In certain embodiments, R is optionally substituted C3-8 alkyl. In certain embodiments, R is optionally substituted C1-C40 alkyl, C1-C20 alkyl, C1-C10 alkyl, C1-C8 alkyl, C1-C5 alkyl, C3-C5 alkyl, C3 alkyl, or C5 alkyl. In certain embodiments, R is optionally substituted C1-C20 alkyl. In certain embodiments, R is optionally substituted C1-C10 alkyl. In certain embodiments, R is optionally substituted C1-C8 alkyl. In certain embodiments, R is optionally substituted C1-C5 alkyl. In certain embodiments, R is optionally substituted C1-C7 alkyl. In certain embodiments, R is optionally substituted C3-C5 alkyl. In certain embodiments, R is optionally substituted C3 alkyl. In certain embodiments, R is unsubstituted C3 alkyl. In certain embodiments, R is n-C3 alkyl. In certain embodiments, R is n-propyl. In certain embodiments, R is n-butyl. In certain embodiments, R is n-pentyl. In certain embodiments, R is n-hexyl. In certain embodiments, R is n-heptyl. In certain embodiments, R is of formula: . In certain embodiments, R is optionally substituted C4 alkyl. In certain embodiments, R is unsubstituted C4 alkyl. In certain embodiments, R is optionally substituted C5 alkyl. In certain embodiments, R is unsubstituted C5 alkyl. In certain embodiments, R is optionally substituted C6 alkyl. In certain embodiments, R is unsubstituted C6 alkyl. In certain embodiments, R is optionally substituted C7 alkyl. In certain embodiments, R is unsubstituted C7 alkyl. In certain embodiments, R is of formula: . In certain embodiments, R is of formula: . In certain embodiments, R is of formula: . In certain embodiments, R is of formula: . In certain embodiments, R is of formula: . In certain embodiments, R is optionally substituted n-propyl. In certain embodiments, R is n-propyl optionally substituted with optionally substituted aryl. In certain embodiments, R is n-propyl optionally substituted with optionally substituted phenyl. In certain embodiments, R is n-propyl substituted with unsubstituted phenyl. In certain embodiments, R is optionally substituted butyl. In certain embodiments, R is optionally substituted n-butyl. In certain embodiments, R is n-butyl optionally substituted with optionally substituted aryl. In certain embodiments, R is n-butyl optionally substituted with optionally

substituted phenyl. In certain embodiments, R is n-butyl substituted with unsubstituted phenyl. In certain embodiments, R is optionally substituted pentyl. In certain embodiments, R is optionally substituted n-pentyl. In certain embodiments, R is n-pentyl optionally substituted with optionally substituted aryl. In certain embodiments, R is n-pentyl optionally substituted with optionally substituted phenyl. In certain embodiments, R is n-pentyl substituted with unsubstituted phenyl. In certain embodiments, R is optionally substituted hexyl. In certain embodiments, R is optionally substituted n-hexyl. In certain embodiments, R is optionally substituted n-heptyl. In certain embodiments, R is optionally substituted n-octyl. In certain embodiments, R is alkyl optionally substituted with aryl (*e.g.*, phenyl). In certain embodiments, R is optionally substituted acyl (*e.g.*, -C(=O)Me).

[168] In certain embodiments, R is optionally substituted alkenyl (*e.g.*, substituted or unsubstituted C₂₋₆ alkenyl). In certain embodiments, R is substituted or unsubstituted C₂₋₆ alkenyl. In certain embodiments, R is substituted or unsubstituted C₂₋₅ alkenyl. In certain

embodiments, R is of formula: . In certain embodiments, R is optionally substituted alkynyl (*e.g.*, substituted or unsubstituted C₂₋₆ alkynyl). In certain embodiments, R is substituted or unsubstituted C₂₋₆ alkynyl. In certain embodiments, R is of formula:



. In certain embodiments, R is optionally substituted carbocyclyl. In certain embodiments, R is optionally substituted aryl (*e.g.*, phenyl or naphthyl).

[169] The chain length of a precursor substrate can be from C1-C40. Those substrates can have any degree and any kind of branching or saturation or chain structure, including, without limitation, aliphatic, alicyclic, and aromatic. In addition, they may include any functional groups including hydroxy, halogens, carbohydrates, phosphates, methyl-containing or nitrogen-containing functional groups.

[170] For example, **FIG. 3** shows a non-exclusive set of putative precursors for the cannabinoid pathway. Aliphatic carboxylic acids including four to eight total carbons (“C4”-“C8” in **FIG. 3**) and up to 10-12 total carbons with either linear or branched chains may be used as precursors for the heterologous pathway. Non-limiting examples include methanoic acid, butyric acid, pentanoic acid, hexanoic acid, heptanoic acid, isovaleric acid, octanoic acid, and decanoic acid. Additional precursors may include ethanoic acid and propanoic acid. In some embodiments, in addition to acids, the ester, salt, and acid forms may all be used as

substrates. Substrates may have any degree and any kind of branching, saturation, and chain structure, including, without limitation, aliphatic, alicyclic, and aromatic. In addition, they may include any functional modifications or combination of modifications including, without limitation, halogenation, hydroxylation, amination, acylation, alkylation, phenylation, and/or installation of pendant carbohydrates, phosphates, sulfates, heterocycles, or lipids, or any other functional groups.

[171] Substrates for any of the enzymes disclosed in this application may be provided exogenously or may be produced endogenously by a host cell. In some embodiments, the cannabinoids are produced from a glucose substrate, so that compounds of Formula 1 shown in **FIG. 2** and CoA precursors are synthesized by the cell. In other embodiments, a precursor is fed into the reaction. In some embodiments, a precursor is a compound selected from Formulae 1-8 in **FIG. 2**.

[172] Cannabinoids produced by methods disclosed in this application include rare cannabinoids. Due to the low concentrations at which cannabinoids, including rare cannabinoids occur in nature, producing industrially significant amounts of isolated or purified cannabinoids from the Cannabis plant may become prohibitive due to, *e.g.*, the large volumes of Cannabis plants, and the large amounts of space, labor, time, and capital requirements to grow, harvest, and/or process the plant materials (see, for example, Crandall, K., 2016. A Chronic Problem: Taming Energy Costs and Impacts from Marijuana Cultivation. EQ Research; Mills, E., 2012. The carbon footprint of indoor Cannabis production. Energy Policy, 46, pp.58-67; Jourabchi, M. and M. Lahet. 2014. Electrical Load Impacts of Indoor Commercial Cannabis Production. Presented to the Northwest Power and Conservation Council; O'Hare, M., D. Sanchez, and P. Alstone. 2013. Environmental Risks and Opportunities in Cannabis Cultivation. Washington State Liquor and Cannabis Board; 2018. Comparing Cannabis Cultivation Energy Consumption. New Frontier Data; and Madhusoodanan, J., 2019. Can cannabis go green? Nature Outlook: Cannabis; all of which are incorporated by reference in this disclosure). The disclosure provided in this application represents a potentially efficient method for producing high yields of cannabinoids, including rare cannabinoids. The disclosure provided in this application also represents a potential method for addressing concerns related to agricultural practices and water usage associated with traditional methods of cannabinoid production (Dillis et al. "Water storage and irrigation practices for cannabis drive seasonal patterns of water extraction and use in Northern

California." *Journal of Environmental Management* 272 (2020): 110955, incorporated by reference in this disclosure).

[173] Cannabinoids produced by the disclosed methods also include non-rare cannabinoids. Without being bound by a particular theory, the methods described in this application may be advantageous compared with traditional plant-based methods for producing non-rare cannabinoids. For example, methods provided in this application represent potentially efficient means for producing consistent and high yields of non-rare cannabinoids. With traditional methods of cannabinoid production, in which cannabinoids are harvested from plants, maintaining consistent and uniform conditions, including airflow, nutrients, lighting, temperature, and humidity, can be difficult. For example, with plant-based methods, there can be microclimates created by branching, which can lead to inconsistent yields and by-product formation. In some embodiments, the methods described in this application are more efficient at producing a cannabinoid of interest as compared to harvesting cannabinoids from plants. For example, with plant-based methods, seed-to-harvest can take up to half a year, while cutting-to-harvest usually takes about 4 months. Additional steps including drying, curing, and extraction are also usually needed with plant-based methods. In contrast, in some embodiments, the fermentation-based methods described in this application only take about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 days. In some embodiments, the fermentation-based methods described in this application only take about 3-5 days. In some embodiments, the fermentation-based methods described in this application only take about 5 days. In some embodiments, the methods provided in this application reduce the amount of security needed to comply with regulatory standards. For example, a smaller secured area may be needed to be monitored and secured to practice the methods described in this application as compared to the cultivation of plants. In some embodiments, the methods described in this application are advantageous over plant-sourced cannabinoids.

[174] Any of the enzymes, host cells, and methods described in this application may be used for the production of cannabinoids and cannabinoid precursors, such as those provided in **Table 1**. In general, the term "production" is used to refer to the generation of one or more products (*e.g.*, products of interest and/or by-products/off-products), for example, from a particular substrate or reactant. The amount of production may be evaluated at any one or more steps of a pathway, such as a final product or an intermediate product, using metrics familiar to one of ordinary skill in the art. For example, the amount of production may be assessed for

a single enzymatic reaction. Alternatively or in addition, the amount of production may be assessed for a series of enzymatic reactions (*e.g.*, the biosynthetic pathway shown in FIG. 1, FIG. 2, and/or FIG. 4). Production may be assessed by any metrics known in the art, for example, by assessing volumetric productivity, enzyme kinetics/reaction rate, specific productivity biomass-specific productivity, titer, yield, and total titer of one or more products (*e.g.*, products of interest and/or by-products/off-products).

[175] In some embodiments, the metric used to measure production may depend on whether a continuous process is being monitored (*e.g.*, several cannabinoid biosynthesis steps are used in combination) or whether a particular end product is being measured. For example, in some embodiments, metrics used to monitor production by a continuous process may include volumetric productivity, enzyme kinetics and reaction rate. In some embodiments, metrics used to monitor production of a particular product may include specific productivity biomass-specific productivity, titer, yield, and total titer of one or more products (*e.g.*, products of interest and/or by-products/off-products).

[176] Production of one or more products (*e.g.*, products of interest and/or by-products/off-products) may be assessed indirectly, for example by determining the amount of a substrate remaining following termination of the reaction/fermentation. In some embodiments, the production of a product (*e.g.*, products of interest and/or by-products/off-products) may be assessed as relative production, for example relative to a control.

Cannabinoid Pathway Enzymes

[177] Methods for production of cannabinoids and cannabinoid precursors can include expression of one or more of: an acyl activating enzyme (AAE); a polyketide synthase (PKS) (*e.g.*, OLS); a polyketide cyclase (PKC); a prenyltransferase (PT) and a terminal synthase (TS).

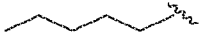
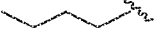
Acyl Activating Enzyme (AAE)

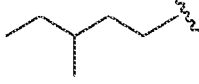
[178] A host cell described in this disclosure may comprise an AAE. As used in this disclosure, an AAE refers to an enzyme that is capable of catalyzing (“activating”) the esterification between a thiol and a substrate (*e.g.*, optionally substituted aliphatic or aryl group) that has a carboxylic acid moiety. In some embodiments, an AAE is capable of using Formula (1):

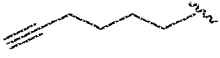


or a salt, solvate, hydrate, polymorph, co-crystal, tautomer, stereoisomer, isotopically labeled derivative thereof to produce a product of Formula (2):



[179] R is as defined in this application. In certain embodiments, R is hydrogen. In certain embodiments, R is optionally substituted alkyl. In certain embodiments, R is optionally substituted C1-40 alkyl. In certain embodiments, R is optionally substituted C2-40 alkyl. In certain embodiments, R is optionally substituted C2-40 alkyl, which is straight chain or branched alkyl. In certain embodiments, R is optionally substituted C2-10 alkyl, optionally substituted C10-C20 alkyl, optionally substituted C20-C30 alkyl, optionally substituted C30-C40 alkyl, or optionally substituted C40-C50 alkyl, which is straight chain or branched alkyl. In certain embodiments, R is optionally substituted C3-8 alkyl. In certain embodiments, R is optionally substituted C1-C40 alkyl, C1-C20 alkyl, C1-C10 alkyl, C1-C8 alkyl, C1-C5 alkyl, C3-C5 alkyl, C3 alkyl, or C5 alkyl. In certain embodiments, R is optionally substituted C1-C20 alkyl. In certain embodiments, R is optionally substituted C1-C20 branched alkyl. In certain embodiments, R is optionally substituted C1-C20 alkyl, optionally substituted C1-C10 alkyl, optionally substituted C10-C20 alkyl, optionally substituted C20-C30 alkyl, optionally substituted C30-C40 alkyl, or optionally substituted C40-C50 alkyl. In certain embodiments, R is optionally substituted C1-C10 alkyl. In certain embodiments, R is optionally substituted C3 alkyl. In certain embodiments, R is optionally substituted n-propyl. In certain embodiments, R is unsubstituted n-propyl. In certain embodiments, R is optionally substituted C1-C8 alkyl. In some embodiments, R is a C2-C6 alkyl. In certain embodiments, R is optionally substituted C1-C5 alkyl. In certain embodiments, R is optionally substituted C3-C5 alkyl. In certain embodiments, R is optionally substituted C3 alkyl. In certain embodiments, R is optionally substituted C5 alkyl. In certain embodiments, R is of formula: . In certain embodiments, R is of formula: . In certain embodiments, R is of formula:

. In certain embodiments, R is of formula: . In certain embodiments, R is optionally substituted propyl. In certain embodiments, R is optionally substituted n-propyl. In certain embodiments, R is n-propyl optionally substituted with optionally substituted aryl. In certain embodiments, R is n-propyl optionally substituted with optionally substituted phenyl. In certain embodiments, R is n-propyl substituted with unsubstituted phenyl. In certain embodiments, R is optionally substituted butyl. In certain embodiments, R is optionally substituted n-butyl. In certain embodiments, R is n-butyl optionally substituted with optionally substituted aryl. In certain embodiments, R is n-butyl optionally substituted with optionally substituted phenyl. In certain embodiments, R is n-butyl substituted with unsubstituted phenyl. In certain embodiments, R is optionally substituted pentyl. In certain embodiments, R is optionally substituted n-pentyl. In certain embodiments, R is n-pentyl optionally substituted with optionally substituted aryl. In certain embodiments, R is n-pentyl optionally substituted with optionally substituted phenyl. In certain embodiments, R is n-pentyl substituted with unsubstituted phenyl. In certain embodiments, R is optionally substituted hexyl. In certain embodiments, R is optionally substituted n-hexyl. In certain embodiments, R is optionally substituted n-heptyl. In certain embodiments, R is optionally substituted n-octyl. In certain embodiments, R is alkyl optionally substituted with aryl (*e.g.*, phenyl). In certain embodiments, R is optionally substituted acyl (*e.g.*, -C(=O)Me).

[180] In certain embodiments, R is optionally substituted alkenyl (*e.g.*, substituted or unsubstituted C₂₋₆ alkenyl). In certain embodiments, R is substituted or unsubstituted C₂₋₆ alkenyl. In certain embodiments, R is substituted or unsubstituted C₂₋₅ alkenyl. In certain embodiments, R is of formula: . In certain embodiments, R is optionally substituted alkynyl (*e.g.*, substituted or unsubstituted C₂₋₆ alkynyl). In certain embodiments, R is substituted or unsubstituted C₂₋₆ alkynyl. In certain embodiments, R is of formula: . In certain embodiments, R is optionally substituted carbocyclyl. In certain embodiments, R is optionally substituted aryl (*e.g.*, phenyl or naphthyl).

[181] In some embodiments, a substrate for an AAE is produced by fatty acid metabolism within a host cell. In some embodiments, a substrate for an AAE is provided exogenously.

[182] In some embodiments, an AAE is capable of catalyzing the formation of hexanoyl-coenzyme A (hexanoyl-CoA) from hexanoic acid and coenzyme A (CoA). In some embodiments, an AAE is capable of catalyzing the formation of butanoyl-coenzyme A (butanoyl-CoA) from butanoic acid and coenzyme A (CoA). In some embodiments, an AAE is capable of catalyzing the formation of butyryl-coenzyme A (butyryl-CoA) from butyric acid and coenzyme A (CoA).

[183] As one of ordinary skill in the art would appreciate, an AAE could be obtained from any source, including naturally occurring sources and synthetic sources (e.g., a non-naturally occurring AAE). In some embodiments, an AAE is a Cannabis enzyme. Non-limiting examples of AAEs include *C. sativa* hexanoyl-CoA synthetase 1 (CsHCS1) and *C. sativa* hexanoyl-CoA synthetase 2 (CsHCS2) as disclosed in US Patent No. 9,546,362, which is incorporated by reference in this application in its entirety.

[184] CsHCS1 has the sequence:

MGKNYKSLDSVVASDFIALGITSEVAETLHGRLAEIVCNYGAATPQTTWINIANHILSP
 DLPFSLHQMLFYGCYKDFGPAPPAWIPDPEKVKSTNLGALLEKRGKEFLGVKYKDPI
 SSFSHFQEFVSRNPEVYWRVTVLMDKISFSKDPECILRRDDINNPGGSEWLPGGYL
 NSAKNCLNVNSNKKLNDTMIVWRDEGNDDLPLNKLTLQDLRKRVLVGYALEEM
 GLEKGCAlAIDMPMHVDAVVIYLAIVLAGYVVVSIADSFSAPEISTRRLRSKAKAIFTQ
 DHIIRGKKRIPLYSRVVEAKSPMAIVIPCSGSNIGAELRDGDISWDYFLERAKEFKNCE
 FTAREQPVDAYTNILFSSGTTGEPKAIPWTQATPLKAAADGWSHLDIRKGDVIVWPT
 NLGWMMGPWL VYASLLNGASIALYNGSPLVSGFAKFVQDAKVTMLGVVPSIVRSW
 KSTNCVSGYDWSTIRCFSSSGEASNVEYLWLMGRANYKPVLEMCGGTEIGGAFFSA
 GSFLQAQSLSSFSQCMGCTLYILDKNGYPMPKNKPGIGELALGPVMFGASKTLLNG
 NHHDVYFKGMPTLNGEVLRRHGDIFELTSNGYYHAHGRADDTMNIGGIKISSIEIERV
 CNEVDDR VFETTAIGVPPLGGGPEQLVIFVFLKDSNDTTIDLNLRLSFNLGLQKKLN
 PLFKVTRVVPLSSLPRATNKMRRVLRQFSHFE (SEQ ID NO: 58).

[185] CsHCS2 has the sequence:

MEKSGYGRDGIYRSLRPPLHLPNNNNLSMVSFLFRNSSSYQKPALIDSETNQILSFSH
 FKSTVIK VSHGFLNLGIKKNDVVLIIYAPNSIHFPVCFLGIIASGAIAATTSNPLYTVSELS
 KQVKDSNPKLIIITVPQLLEKVKGFNLPTILIGPDSEQESSDKVMTFNDLVNLGGSSGS

EFPIVDDFKQSDTAALLYSSGTTGMSKGVVLTHKNFIASSLMVTMEQDLVGEMDNV
 FLCFLPMFHVFGGLAIITYAQLQRGNTVISMARFDLEKMLKDVEKYKVTHLWVVPPI
 LALSKNSMVKKFNLSSIKYIGSGAAPLGKDLMEESKVVYPYGIVAQGYGMTETCGIV
 SMEDIRGGKRNSGSAGMLASGVEAQIVSVDTLKPLPPNQLGEIWVKGPNNMMQGYFN
 NPQATKL TIDKKGWVHTGDLGYFDEDGHLVYVVDRIKELIKYKGFQVAPAELEGLLV
 SHPEILDAVVIPFPDAEAGEVPVAYVVRSPNSSLTENDVKKFIAGQVASFKRLRKVTFI
 NSVPKSASGKILRRELIQKVRSNM (SEQ ID NO: 59).

[186] Additional AAEs are described in and incorporated by reference from PCT Publication NO. WO/2020/176547 and U.S. Patent No. 11,274,320, which are incorporated by reference in this application in their entireties.

[187] In some embodiments, an AAE is from *Cicer arietinum* (Chickpea) (Garbanzo), corresponding to UniProt Accession No. A0A1S2XHV8, the protein sequence for which is provided as SEQ ID NO: 86.

MAYKSLSSISVSDIESLGIEQEAATLHQQLTEIIGIHQTDSPATWQSISRSILNPELPFS
 FHQMLYYGCFVDYGPDPAPWIPDPESVTSTNVGRLLMRGKEFLGSAYKDPITSFAD
 FQKFSVSNPEVYWKTVLGMENISFSKPPECILCESISDDGSSSYPSGQWLPGASINPAH
 NCLNLNGERSLNDTVILWRNELQDDLPLQRMTEELRQEVWLVAAYALESGLGKGS
 AIAIDMPMHCKSVVIYLAIVLAGYVVVSIADSFAPREISSRLKISNAKVIFTQDLILRGD
 KTLPLYSRIVDAESPMIAVIPTRGSEFSMKLRDGD LAWCFMDGVNLIKKEFIAVEE
 PVETFTNILFSSGTTGDPKAIPWTNISPLKAAADAWCHLDVRKGDVVS WPTNLGWM
 MGPWL VYASLLNGASMALYNGSPLGSGFAKFVQDSKVTMLGVIPSLVRSWRNANS
 TSGFDWSAIRCFASTGEASNIDEYLWLMGRAHYKPIIEYCGGTEIGGGFVTGSL LQAQ
 SLAAFSTPAMCCSLFILDQGHPIPNVPGMGELALGPLMLGASNTLLNADHYGVYF
 KGMPIWNGKVLRRHGDVFERTARGYYHAHGRADDTMNLGGIKVSSVEIERICNGA
 DSNILETAAIGIPPSGGGPEQLALAVVLKNSNVTSQDLLTLRMSFNSALQKTLNPLFR
 VSQVVPVPSLPRITASNKVMRRVLRQQLVENTQSSRI (SEQ ID NO: 86)

[188] A non-limiting example of nucleic acid sequence encoding SEQ ID NO: 86 is provide by SEQ ID NO: 87.

atggcttataaatctttatcatccatctctgtctccgatagagagctctgggcattgaacaggaacacgcagccacccttcatcaacaatt
 gactgaaattatcggtattcaccaaacggacagcccagctacatggcaatctatctcgagatctatcttgaatcccgaaactaccattctcc

tttcatcaaatgtgtactacgggtgcttcgttgattatggacctgacccaccagcttggattcctgatccagagtcggtaacctctactaac
 gtgggtcggttggtaaatgaggggtaaggaatttttaggttctgcttacaaggacccaattacttcattcgccgatttccagaaatttca
 gttccaacccccgaagttactggaagaccgtcttgggtgaaatgaacatttcttctgaagccacctgaatgtatccttgtgagtctatt
 tcagatgacgggttctccagttacccaagcggccaatgggtaccaggtgctagtataaatccagcacacaactgtttgaacttaaatgggtg
 aaagatccttgaatgacactgtcatcttgggagaaacgaactacaagacgatctaccttgcagaagaatgaccttagaagaattgaga
 caagaagtttgggtagttgctatgccctggaatcttgggttggaaaagggtagcgcctatcgctatcgatagccaatgcactgtaaatc
 tgtgtcatttacttggccatcggttagctggttatgctgtagtatcaattgctgactcttttggccctagagagatctcttcggtttaaagat
 ttccaatgcaaaagttattttcactcaagacttgatattgagaggagataagacactaccattgtactctaggattgtcgacgctgaaagtc
 ccatggctatcggttattccaactagaggttctgaattttccatgaagttgagagatgggtgatctagcctgggtgaacttcattgatgggtgta
 acaaaaataaagggtgaaggagttcatcgctgtggaagaaccagtcgagactttcaccaacattcttttctccggtaccacaggcgatc
 caaaagctataccatggacgaatatctcccccttgaaggctgctgctgacgcgatgggtccatctggatgttagaaaaggcgacgtagct
 catggcctactaacttgggttggatgatgggtccatgggttggctacgctagttattgaatgggtgcctctatggcactatataacggtagc
 ccacttggatctgggttggccaagtgcgttcaggatagtaagggtactatgttgggtgttatccaagtttagttcgttcttgagaaacgct
 aattcaacciccggtttagttgggtctgcaataagatgttgcgtagtagcagggtgaagcttcaacattgatgaatacttgggttaattggga
 agagctcactacaagccaatcattgaatattgtgggtgtacagagatcgggcggtggtttgttactgggtcgttgttacaagcccaatcatt
 ggccgctttcagcacccccgaatgtgctgttctgttcatccttgacgaccaaggacaccaatccacagaatgtgcctggcatgg
 gtgaacttgccttgggtcccttaattgttgggagcctctaacactcttttgaatgcagaccattacgggtgtctatttcaaagggtatgccaatctg
 gaacggtaaggctcctctgtagacacgggtgacgttttgaagaactgctcgtggttactaccatgctcacggtagagctgatgataccat
 gaacttgggtgggtataaaagtctcttccgttgaatgaagaatttgaacgggtgcagactcgaacatcctagaaacggctgccattggt
 ataccaccttgggttgggtggtccagaacaattggctttagctgttgaagaactccaatgttacctctcaagatttgcataacattaaga
 atgtcattcaattccgccttgcagaagacactgaaccattgtttagagtctctcaagtagttcctgtaccatcttggccaagaactgccagt
 aacaaagtcagaggagagttttagacaacaattagttgaaaatactcagtcctcaaggatc (SEQ ID NO: 87)

[189] In some embodiments, an AAE comprises a protein or nucleic acid sequence that is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or is 100% identical, including all values in between, to SEQ ID NO: 86 or 87. In some embodiments, an AAE comprises a sequence that is a conservatively substituted version of SEQ ID NO: 86.

[190] In some embodiments, an AAE acts on multiple substrates, while in other embodiments, it exhibits substrate specificity. For example, in some embodiments, an AAE exhibits substrate specificity for one or more of hexanoic acid, butyric acid, isovaleric acid, octanoic acid, or decanoic acid. In other embodiments, an AAE exhibits activity on at least two of hexanoic acid, butyric acid, isovaleric acid, octanoic acid, and decanoic acid.

[191] A host cell that expresses a heterologous polynucleotide encoding an AAE described in this disclosure may be capable of activating more of a short chain fatty acid (e.g., a four-carbon fatty acid) in the presence of Coenzyme A than a control. In some embodiments, the host cell is capable of catalyzing the production of more, for example, producing at least 1% (e.g., at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 100%, at least 125%, at least 150%, at least 175%, at least 200%, at least 300%, at least 400%, at least 500%, at least 600%, at least 700%, at least 800%, at least 900%, or at least 1,000%) more butyryl-CoA relative to a control. In some embodiments, a control is a host cell that does not express a heterologous gene encoding an AAE. In some embodiments, a control is a host cell that does not express a heterologous polynucleotide encoding an AAE of interest.

[192] A host cell that expresses a heterologous polynucleotide encoding an AAE described in this disclosure and that also expresses one or more other enzymes involved in cannabinoid biosynthesis may be capable of producing a varinolic cannabinoid.

Polyketide Synthases (PKS)

[193] A host cell described in this application may comprise a PKS. As used in this application, a “PKS” refers to an enzyme that is capable of producing a polyketide. In certain embodiments, a PKS converts a compound of Formula (2) to a compound of Formula (4), (5), and/or (6). In certain embodiments, a PKS converts a compound of Formula (2) to a compound of Formula (4). In certain embodiments, a PKS converts a compound of Formula (2) to a compound of Formula (5). In certain embodiments, a PKS converts a compound of Formula (2) to a compound of Formula (4) and/or (5). In certain embodiments, a PKS converts a compound of Formula (2) to a compound of Formula (5) and/or (6).

[194] In some embodiments, a PKS is a tetraketide synthase (TKS). In certain embodiments, a PKS is an olivetol synthase (OLS). As used in this application, an “OLS” refers to an enzyme that is capable of using a substrate of Formula (2a) to form a compound of Formula (4a), (5a) or (6a) as shown in **FIG. 1**.

[195] In certain embodiments, a PKS is a divarinol synthase.

[196] In certain embodiments, polyketide synthases can use hexanoyl-CoA or any acyl-CoA or a product of Formula (2):

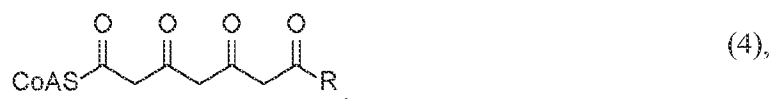


and three malonyl-CoAs as substrates to form 3,5,7-trioxododecanoyl-CoA or other 3,5,7-trioxo-acyl-CoA derivatives; or to form a compound of Formula (4):



wherein R is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl; depending on substrate. R is as defined in this application. In some embodiments, R is a C2-C6 optionally substituted alkyl. In some embodiments, R is a propyl or pentyl. In some embodiments, R is pentyl. In some embodiments, R is propyl. A PKS may also bind isovaleryl-CoA, octanoyl-CoA, hexanoyl-CoA, and butyryl-CoA. In some embodiments, a PKS is capable of catalyzing the formation of a 3,5,7- trioxoalkanoyl-CoA (e.g. 3,5,7-trioxododecanoyl-CoA). In some embodiments, an OLS is capable of catalyzing the formation of a 3,5,7- trioxoalkanoyl-CoA (e.g. 3,5,7-trioxododecanoyl-CoA).

[197] In some embodiments, a PKS uses a substrate of Formula (2) to form a compound of Formula (4):



wherein R is unsubstituted pentyl.

[198] As one of ordinary skill in the art would appreciate a PKS, such as an OLS, could be obtained from any source, including naturally occurring sources and synthetic sources (e.g., a non-naturally occurring PKS). In some embodiments a PKS is from *Cannabis*. In some embodiments a PKS is from *Dictyostelium*. Non-limiting examples of PKS enzymes may be found in U.S. Patent No. 6,265,633; PCT Publication No. WO2018/148848 A1; PCT Publication No. WO2018/148849 A1; U.S. Patent No. 10,435,727, WO 2020/176547, and U.S. Patent No. 11,274,320, which are incorporated by reference in this application in their entireties.

[199] A non-limiting example of an OLS is provided by UniProtKB - B1Q2B6 from *C. sativa*. In *C. sativa*, this OLS uses hexanoyl-CoA and malonyl-CoA as substrates to form 3,5,7-trioxododecanoyl-CoA. OLS (e.g., UniProtKB - B1Q2B6) in combination with olivetolic acid cyclase (OAC) produces olivetolic acid (OA) in *C. sativa*.

[200] The amino acid sequence of UniProtKB - B1Q2B6 is:
 MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEKFRKICDKSM
 IRKRNCFLNEEHLKQNPRLVEHEMQTLDDARQDMLVVEVPKLGKDACAIAKEWGQ
 PKSKITHLIFTSASTDMPGADYHCAKLLGLSPSVKRVMYQLGCYGGGTVLRIAKD
 IAENNKGARVLAVCCDIMACLFRGPSESDLELLVGQAIFGDGAAAVIVGAEPDES
 VGERPIFELVSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIG
 ISDWNSIFWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDLRK
 RSLEEGKSTTGDGFEGVLFVFGFPGPLTVERVVVRSVPIKY (SEQ ID NO: 1).

[201] In some embodiments, a PKS comprises the sequence of SEQ ID NO: 57:

MPSLESVKKSNRADGFASILAIGRANPENFIEQSTYPDFFFRVTNSEHLVNLKKKFQRI
 CDKTAIRKRHFVWNEELNANPCLGTMDNSLNVRQEFAIREIPKLGAEAAATKAIQE
 WGQPKSRITHLIFCTTSGMDLPGADYQLTQILGLNPNIERVMLYQQGCFAGGTTLRL
 AKCLAESRKGARVLVCAETTAVLFRAPSEEHQDDLVTQALFADGASALIVGADPD
 ETAHERASFVIVSTSQVLLPDSAGAIGHVSEGGLIATLHRDVPQIVSKNVGKCLEEA

FTPLGISDWNISIFWVPHPGGRAILDQVEERVGLKPEKLIVSRHVLAEYGNMSSVCVH
FALDEMRRKRSKKEGKATTGEGLDWGVLFVGFPGPLTVETVVLHSVPI (SEQ ID NO: 57)

[202] In some embodiments, a PKS comprises the sequence of SEQ ID NO: 82:

MNHLRAEGPASVLAIGTANPENILLQDEFDPYYFRVTKSEHMTQLKEKFRKICDKSM
IRKRNCFLNEEHLKQNPRLVEHEMQTLDDARQDMLVVEVPKLGKDACAIAKEWGQ
PKSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVM MYQLGCYGGGTVLRIAKD
IAENNKGARVLAVCCDIMACLFRGPSESDLELLVGQAIFGDGAAAVIVGAEPDES VG
ERPIFELVSTGQTILPNSEG TIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGID
WNSIFWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSCVLFVMDLRRK
RSLEEGKSTTGDFEWDGVLFVGFPGPLTVERVVVRSVPIKY (SEQ ID NO: 82).

[203] In some embodiments, the PKS is encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 83:

atgaatcatcttagagcagaaggtcctgcctccgtattagctatcggaaccgctaaccagagaaacatactactgcaagacgaatttcca
gattactatttccgtgttacgaaatcggaacacatgacacagttgaaggaaaagttcaggaagattttgtgacaaatcaatgattagaaag
agaaattgcttttgaacgaggaacacttaaaagcaaaacccagattggctgaacatgaaatgcaaaactttggatgctagacaagatatg
ctgtgttggaagtcccaaaattgggcaaggacgctgtgtgccaaggctatcaaggagtgagggtcaacaaaatctaaaattactcacttg
atcttcacctctgcaagcactactgatatgcctgggtccgactaccattgtgctaagttattgggtttgtccccatctgttaagagagttatg
atgtatcaactgggtgttacgggtggcaccgtcttgcgtattgctaaggatatcgctgaaaataacaagggtgctagagctcttagccg
tttgcgtgatataatggctgtctattcagaggaccaagtgaaatcgacttggaattgctagtgaggctaggttatgttgatgggtgccg
ctgctgttatgtcggtgccgaaccagacgaatccgttggtgagaggccaatctttgaattggtttcaaccgggtcaaacctcttacctaa
ctccgaaggcactataggtgggtcacaltagagaagccgggtctgatcttcgacttgcataaagatgtccaatgcttattagtaataacatt
gagaaatgtttgatcgaagcattcacaccaatcggtatttctgactggaactcgatcttctggattactcaccctggcggttaaggccatcc
tggaacaaggtagaagaaaagttgcacttaagagcgataaattcgtggactcccgatcatgtctgtctgaacacggaaatatgagctctt
cttgcgttctttttgtatggatgaactaagaaagagaagtttgaagaaggtaaatctaccacaggtgacgggtttgagtggggtgtcttg
ttcggttccggtcctggactaactgttgaagagttgtcgaagatccgttccaattaagtac (SEQ ID NO: 83).

[204] Example 1 describes the surprising identification of multiple PKS mutants that can produce more divarinol and more divaric acid than a PKS comprising SEQ ID NO: 1. These PKSs can be functionally expressed in host cells such as *S. cerevisiae*.

[205] Nucleic acid and protein sequences for PKSs identified in this application are provided in **Table 7**.

[206] In some embodiments, a PKS comprises a protein or nucleic acid sequence that is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or is 100% identical, including all values in between, to any one of SEQ ID NOs: 57, 82, 83, 1-56, 104-130, or 131-157. In some embodiments, a PKS comprises a sequence that is a conservatively substituted version of a sequence selected from SEQ ID NOs: 2-28 and 104-130.

[207] In some embodiments, a PKS comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, at least 30, at least 31, at least 32, at least 33, at least 34, at least 35, at least 36, at least 37, at least 38, at least 39, at least 40, at least 41, at least 42, at least 43, at least 44, at least 45, at least 46, at least 47, at least 48, at least 49, at least 50, at least 51, at least 52, at least 53, at least 54, or at least 55 amino acid substitutions relative to SEQ ID NO: 1. In some embodiments, a PKS comprises an amino acid substitution at a residue corresponding to position 3, 4, 7, 11, 21, 24, 25, 38, 44, 51, 62, 63, 64, 66, 69, 72, 75, 76, 77, 78, 79, 81, 82, 92, 101, 103, 104, 123, 124, 125, 128, 136, 137, 144, 150, 166, 190, 225, 237, 247, 262, 267, 278, 296, 301, 312, 315, 316, 319, 328, 334, 335, 375, 384, and/or 385 in SEQ ID NO: 1. In some embodiments, a PKS comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, at least 30, at least 31, at least 32, at least 33, at least 34, at least 35, at least 36, at least 37, at least 38, at least 39, at least 40, at least 41, at least 42, at least 43, at least 44, at least 45, at least 46, at least 47, at least 48, at least 49, at least 50, at least 51, at least 52, at least 53, at least 54, or at least 55 amino acid substitutions from Table

6. In some embodiments, a PKS comprises a combination of amino acid substitutions from Table 3, 4, or 5.

[208] In some embodiments, a PKS comprises the amino acid C at a residue corresponding to position 335 in SEQ ID NO: 1. In some embodiments, the PKS comprises the amino acid substitution T335C relative to SEQ ID NO: 1.

[209] In some embodiments, a PKS comprises an amino acid substitution at a residue corresponding to position 3, 4, 7, 11, 21, 24, 38, 44, 62, 66, 69, 72, 76, 77, 78, 79, 101, 103, 104, 124, 136, 137, 144, 150, 166, 225, 237, 262, 301, 319, and/or 385 in SEQ ID NO: 1.

[210] In some embodiments, a PKS comprises one or more of the following substitutions relative to SEQ ID NO: 1: H3R or H3A; L4Q; E7P; S11T; E21P; L24I; K38N; Q44D; N62H; N66T; H69I; Q72E; L76M; V77C; E78T; H79Y; D101E; C103A; A104E; S124T; H136Q; C137L; S144P; V150T; I166L; V225A; G237A; H262S; K301P; V319R, V319K, or V319A; and/or Y385S, Y385T, or Y385N.

[211] In some embodiments, a PKS comprises an amino acid substitution at a residue corresponding to position 25, 51, 63, 64, 75, 81, 82, 92, 123, 125, 128, 190, 247, 267, 278, 296, 312, 315, 316, 328, 334, 335, 375, and/or 384 in SEQ ID NO: 1.

[212] In some embodiments, a PKS comprises one or more of the following amino acid substitutions relative to SEQ ID NO: 1: L25I or L25Y; K51R; C63M, C63T, or C63F; F64Y; R75N, R75S, or R75G; M81D; Q82P; V92K or V92A; T123C or T123S; A125S; T128V; L190C; T247A; M267G; I278K or I278E; T296A; H312N or H312G; S315K or S315P; D316E; H328Y; S334P; T335C; R375S or R375T; and/or K384N.

[213] In some embodiments, a PKS comprises: the amino acid C at a residue corresponding to position 157 in SEQ ID NO: 1; the amino acid F at a residue corresponding to position 208 in SEQ ID NO: 1; the amino acid H at a residue corresponding to position 297 in SEQ ID NO: 1; and/or the amino acid N at a residue corresponding to position 330 in SEQ ID NO: 1.

[214] In some embodiments, a PKS comprises the following amino acid substitutions relative to SEQ ID NO: 1: R75N, D101E, V319R, H328Y, and T335C; C63M, H69I, R75G, V77C, M267G, H312G, V319K, T335C, R375T, and Y385S; C63M, F64Y, S124T, H328Y,

and T335C; K51R, H69I, R75N, V77C, D101E, C103A, T128V, K301P, H312G, and T335C; C63T, Q82P, D101E, C103A, and T123S; L25I, K51R, R75N, V77C, D101E, C103A, T128V, K301P, H312G, and T335C; K51R, R75N, V77C, D101E, C103A, T128V, K301P, H312G, D316E, and T335C; R75N, V77C, H79Y, D101E, C103A, T128V, K301P, H312G, D316E, and T335C; K51R, N66T, R75S, D101E, T123C, T296A, S315P, H328Y, T335C, and R375T; K51R, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, D316E, and T335C; H3R, S11T, K38N, Q44D, K51R, H69I, Q72E, R75N, L76M, V77C, E78T, H79Y, M81D, V92K, D101E, C103A, A104E, T128V, C137L, S144P, V150T, V225A, H262S, I278K, H312G, S315P, D316E, S334P, and K384N; K51R, R75N, V77C, H79Y, D101E, C103A, T128V, K301P, H312G, and D316E; L25I, K51R, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, and T335C; L4Q, E21P, N62H, H312N, and H328Y; L25I, H69I, R75N, V77C, H79Y, D101E, C103A, T128V, K301P, and H312G; H3A, L25Y, C63M, R75N, H136Q, M267G, H328Y, T335C, R375T, and Y385T; K51R, H69I, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, and T335C; D101E, T123C, H136Q, H312G, and T335C; H3R, E7P, S11T, K38N, Q44D, K51R, H69I, Q72E, R75N, L76M, V77C, E78T, H79Y, M81D, V92A, D101E, C103A, A104E, T128V, C137L, S144P, V150T, V225A, H262S, I278E, H312G, S315P, D316E, S334P, and K384N; K51R, R75N, V77C, H79Y, D101E, C103A, T128V, H312G, D316E, and T335C; H3A, L4Q, C63F, R75N, D101E, C103A, C137L, G237A, T247A, and Y385N; L4Q, L24I, Q44D, K51R, C63F, R75S, V77C, C103A, I166L, and V319A; L4Q, R75S, V319R, T335C, and R375S; R75S, V77C, S124T, T335C, and R375T; or H3A, R75S, D101E, S315K, and T335C.

[215] In some embodiments, a PKS comprises a contiguous stretch of amino acids from SEQ ID NO: 1. In some embodiments, a PKS comprises a contiguous stretch of at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, at least 30, at least 31, at least 32, at least 33, at least 34, at least 35, at least 36, at least 37, at least 38, at least 39, at least 40, at least 41, at least 42, at least 43, at least 44, at least 45, at least 46, at least 47, at least 48, at least 49, at least 50, at least 51, at least 52, at least 53, at least 54, at least 55, at least 56, at least 57, at least 58, at least 59, at least 60, at least 61, at least 62, at least 63, at least 64, at least 65, at least 66, at least 67, at least 68, at least 69, at least 70, at least 71, at least 72, at least 73, at least 74, at least 75, at least 76, at least 77, at least 78, at least 79, at least 80, at

[illegible]

330, at least 331, at least 332, at least 333, at least 334, at least 335, at least 336, at least 337, at least 338, at least 339, at least 340, at least 341, at least 342, at least 343, at least 344, at least 345, at least 346, at least 347, at least 348, at least 349, at least 350, at least 351, at least 352, at least 353, at least 354, at least 355, at least 356, at least 357, at least 358, at least 359, at least 360, at least 361, at least 362, at least 363, at least 364, at least 365, at least 366, at least 367, at least 368, at least 369, at least 370, at least 371, at least 372, at least 373, at least 374, at least 375, at least 376, at least 377, at least 378, at least 379, at least 380, at least 381, at least 382, at least 383, or at least 384 contiguous amino acids from SEQ ID NO: 1.

[216] It was observed that all PKS mutants identified in Example 1 comprise the following contiguous stretches of amino acids from SEQ ID NO: 1: contiguous stretches of residues with amino acids corresponding to positions 12-20 in SEQ ID NO: 1; positions 26-37 in SEQ ID NO: 1; positions 39-43 in SEQ ID NO: 1; positions 45-50 in SEQ ID NO: 1; positions 52-61 in SEQ ID NO: 1; positions 83-91 in SEQ ID NO: 1; positions 93-100 in SEQ ID NO: 1; positions 105-122 in SEQ ID NO: 1; positions 129-135 in SEQ ID NO: 1; positions 145-149 in SEQ ID NO: 1; positions 151-165 in SEQ ID NO: 1; positions 167-189 in SEQ ID NO: 1; positions 191-224 in SEQ ID NO: 1; positions 226-236 in SEQ ID NO: 1; positions 238-246 in SEQ ID NO: 1; positions 248-261 in SEQ ID NO: 1; positions 268-277 in SEQ ID NO: 1; positions 279-295 in SEQ ID NO: 1; positions 302-311 in SEQ ID NO: 1; positions 320-327 in SEQ ID NO: 1; positions 336-374 in SEQ ID NO: 1; and positions 376-383 in SEQ ID NO: 1.

[217] In some embodiments, a PKS comprises a contiguous stretch of residues with amino acids corresponding to positions 12-20 in SEQ ID NO: 1; positions 26-37 in SEQ ID NO: 1; positions 39-43 in SEQ ID NO: 1; positions 45-50 in SEQ ID NO: 1; positions 52-61 in SEQ ID NO: 1; positions 83-91 in SEQ ID NO: 1; positions 93-100 in SEQ ID NO: 1; positions 105-122 in SEQ ID NO: 1; positions 129-135 in SEQ ID NO: 1; positions 145-149 in SEQ ID NO: 1; positions 151-165 in SEQ ID NO: 1; positions 167-189 in SEQ ID NO: 1; positions 191-224 in SEQ ID NO: 1; positions 226-236 in SEQ ID NO: 1; positions 238-246 in SEQ ID NO: 1; positions 248-261 in SEQ ID NO: 1; positions 268-277 in SEQ ID NO: 1; positions 279-295 in SEQ ID NO: 1; positions 302-311 in SEQ ID NO: 1; positions 320-327 in SEQ ID NO: 1; positions 336-374 in SEQ ID NO: 1; and/or positions 376-383 in SEQ ID NO: 1. As a non-limiting example, a PKS that comprises a contiguous stretch of residues with

amino acids corresponding to positions 12-20 in SEQ ID NO: 1 comprises the amino acid sequence VLAIGTANP (SEQ ID NO: 97).

[218] In some embodiments, a PKS does not comprise one or more sequences selected from SEQ ID NOs: 91-94.

MATKSVAVEEMCKAQKAGGPATILAIPTAVPSNCCYYQSEYPDFYFRVTKSDHLTDL
KSKFKRMCERSSIKKRYMHLTEEILEENPNMCTFAAPSIDGRQDIVVKEIPKLAKEAA
SKAIKEWGQPKSNITHLVFCTTSGVDMPGCDYQLTRLLGLRPSIKRLMMYQQGCHA
GGTGLRLAKDLAENNKGARVLVVCSEMTVINFRGPSEAHMDSL VGQSLFGDGASA
VIVGSDPDLSTEHPYQIMSASQIIVADSEGAIDGHLRQEGLTFHLRKDVPSLVSDNIE
NTLVEAFTPILMDSIDSIIDWNSIFWIAHPGGPAILNQVQAKVGLKEEKLRVSRHILSE
YGNMSSACVFFIMDEMRRKRSMEEGKGTGEGLEWGVLF GFGPGFTVETIVLHSVPI
(SEQ ID NO: 91);

MATKSVAVEEMCKAQKAGGPATILAIPTAVPSNCCYYQSEYPDFYFRVTKSDHLTDL
KSKFKRMCERSSIKKRYMHLTEEILKENPNMCSFAAPSIDGRQDIVVKEIPKLAKEAA
SKAIKEWGQPKSNITHLVFCTTSGVDMPGCDYQLTRLLGLRPSIKRLMMYQQGCHA
GGTGLRLAKDLAENNKGARVLVVCSEMTVINFRGPSEAHMDSL VGQSLFGDGASA
VIVGSDPDLSTEHPYQIMSASQIIVADSEGVIDGHLRQEGLTFHLRKDVPSLVSDNIE
NTLVEAFTPILMDSIDSIIDWNSIFWIAHPGGPAILNQVQAKVGLKEEKLRVSRHILSE
YGNMSSACVFFIMDEMRRKRSVEEGKGTGEGLEWGVLF GFGPGFTVETIVLHSVPI
(SEQ ID NO: 92);

MATKSVAVEEMCKAQKAGGPATILAIPTAVPSNCCYYQSEYPDFYFRVTKSDHLTDL
KSKFKRMCERSSITKRYMHLTEEILEENPNMCTFAAPSIDGRQDIVVKEIPKLAKEAA
SKAIKEWGQPKSNITHLVFCTTSGVDMPGCDYQLTRLLGLRPSIKRLMMYQQGCHA
GGTGLRLAKDLAENNKGARVLVVCSEMTVINFRGPSEAHMDSL VGQSLFGDGASA
VIVGSDPDLSTEHPYQIMSASQIIVADSEGAIDGHLRQEGLTFHLRKDVPSLVSDNIE
NTLVEAFTPILMDSIDSIIDWNSIFWIAHPGGPAILNQVQAKVGLKEEKLRVSRHILSE
YGNMSSACVFFIMDEMRRKRSVEEGKGTGEGLEWGVLF GFGPGFTVETIVLHSVPI
(SEQ ID NO: 93); and

MSRSRLIAQAVGPATVLAMGKAVPANVFEQATYPDFFFNITNSNDKPALKAKFQRIC
DKSGIKKRHFYLDQKILESNPAMCTYMETSLNCRQEIAVAQVPKLAKEASMNAIKE

WGRPKSEITHIVMATTSGVNMPGAELATAKLLGLRPNVRRVMMYQQGCFAGATVL
 RVAKDLAENNAGARVLAICSEVTAVTFRAPSETHIDGLVGSALFGDGAAAVIVGSDP
 RPGIERPIYEMHWAGEMVLPESDGAIDGHLTEAGLVFHLLKDVPGLITKNIGGFLKDT
 KNLVGASSWNELFWAVHPGGPAILDQVEAKLELEKKGKFQASRDILSDYGNMSSASV
 LFLVDRVRERSLESNKSTFGEGSEWGFLLIGFGPGLTVETLLLRALPLQQAERV (SEQ
 ID NO: 94).

[219] PKS enzymes described in this application may or may not have cyclase activity. In some embodiments where the PKS enzyme does not have cyclase activity, one or more exogenous polynucleotides that encode a polyketide cyclase (PKC) enzyme may also be co-expressed in the same host cells to enable conversion of hexanoic acid or butyric acid or other fatty acid conversion into olivetolic acid or divarinolic acid or other precursors of cannabinoids. In some embodiments, the PKS enzyme and a PKC enzyme are expressed as separate distinct enzymes. In some embodiments, a PKS enzyme that lacks cyclase activity and a PKC are linked as part of a fusion polypeptide that is a bifunctional PKS. In some embodiments, a bifunctional PKS is referred to as a bifunctional PKS-PKC. In some embodiments, a bifunctional PKC is referred to as a bifunctional PKS-PKC. In some embodiments, a bifunctional PKC is a bifunctional tetraketide synthase (TKS-TKC). As used in this application, a bifunctional PKS is an enzyme that is capable of producing a compound of Formula (6):



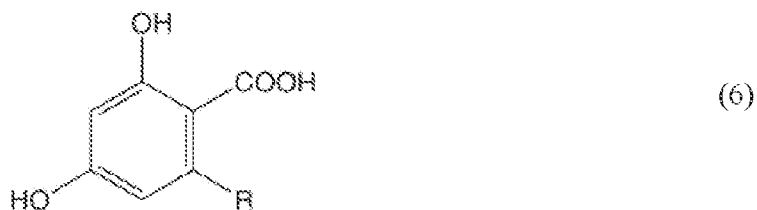
from a compound of Formula (2):



and a compound of Formula (3):



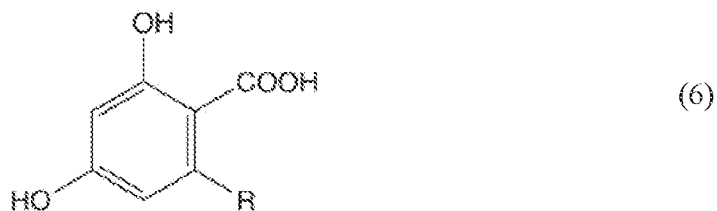
In some embodiments, a PKS produces more of a compound of Formula (6):



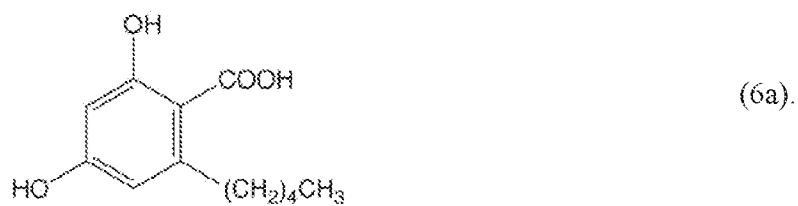
as compared to a compound of Formula (5):



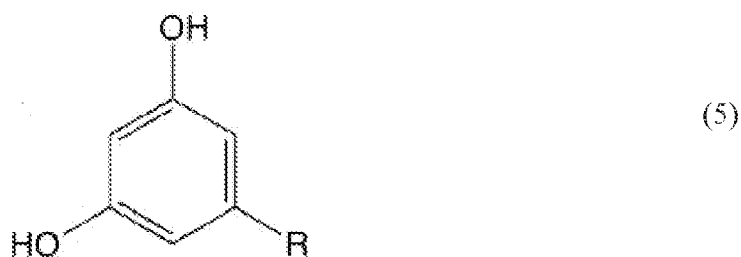
As a non-limiting example, a compound of Formula (6):



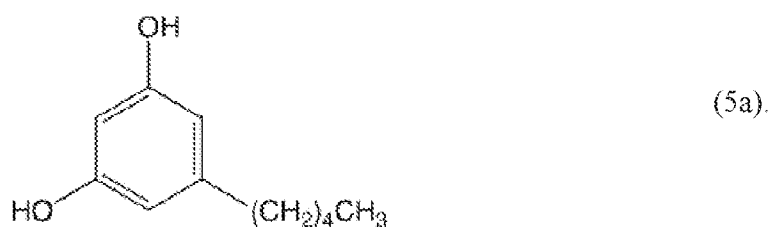
is olivetolic acid (Formula (6a)):



As a non-limiting example, a compound of Formula (5):



is olivetol (Formula (5a)):



[220] In some embodiments, a polyketide synthase of the present disclosure is capable of catalyzing a compound of Formula (2):



and a compound of Formula (3):



to produce a compound of Formula (4):



and also further catalyzes a compound of Formula (4):



to produce a compound of Formula (6):



In some embodiments, the PKS is not a fusion protein. In some embodiments, a PKS that is capable of catalyzing a compound of Formula (2):



and a compound of Formula (3):



to produce a compound of Formula (4):



and is also capable of further catalyzing the production of a compound of Formula (6):



from the compound of Formula (4):



is preferred because it avoids the need for an additional polyketide cyclase to produce a compound of Formula (6):



In some embodiments, such an enzyme that is a bifunctional PKS eliminates the transport considerations needed with addition of a polyketide cyclase, whereby the compound of Formula (4), being the product of the PKS, must be transported to the PKS for use as a substrate to be converted into the compound of Formula (6).

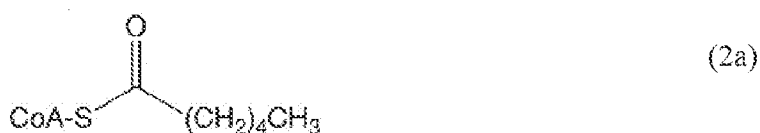
[221] In some embodiments, a PKS is capable of producing olivetolic acid in the presence of a compound of Formula (2a):



and Formula (3a):



[222] In some embodiments, an OLS is capable of producing olivetolic acid in the presence of a compound of Formula (2a):



and Formula (3a):



[223] In some embodiments, a PKS is a divarinol synthase. For example, a PKS may use butyryl-CoA and three malonyl-CoAs as substrates to form a compound of Formula (4):

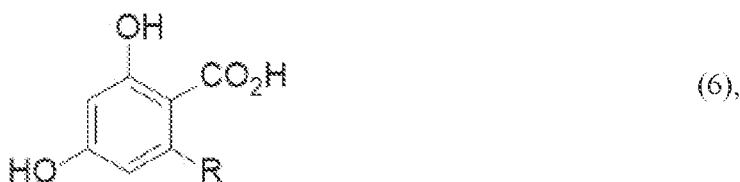


wherein R is n-propyl; and/or to form divarinol (a compound of Formula (5):



wherein R is n-propyl).

[224] In some embodiments, a PKS is a bifunctional PKS and can use butyryl-CoA and three malonyl-CoAs as substrates to form divaric acid (a compound of Formula (6):



wherein R is n-propyl).

[225] A host cell that expresses a heterologous polynucleotide encoding a PKS described in this disclosure, may be capable of producing more (*e.g.*, at least 1.5, 2, 3, 4, 5, 6, 7, 8 or 9-fold more) divarinol in the presence of butyrate relative to a control. A host cell that expresses a heterologous polynucleotide encoding a PKS described in this disclosure may be capable of producing more (*e.g.*, at least 1%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 100%, at least 125%, at least 150%, at least 175%, at least 200%, at

least 300%, at least 400%, at least 500%, at least 600%, at least 700%, at least 800%, at least 900%, or at least 1,000% more) divarinol in the presence of butyrate relative to a control. In some embodiments, a control is a host cell that does not express a heterologous gene encoding a PKS. In some embodiments, a control is a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1. In some embodiments, the host cell further comprises a heterologous polynucleotide encoding an AAE. In some embodiments, the AAE comprises a sequence that is at least 90% identical to the sequence of SEQ ID NO: 86. In some embodiments, the AAE comprises the sequence of SEQ ID NO: 86.

[226] In some embodiments, a host cell that expresses a heterologous polynucleotide encoding a PKS described in this disclosure may be capable of producing at least 5000, 6000, 7000, 8000, 9000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21,000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000, 30,000, 31,000, 32,000 or 33,000 µg/L divarinol. In some embodiments, the host cell further comprises a heterologous polynucleotide encoding an AAE. In some embodiments, the AAE comprises a sequence that is at least 90% identical to the sequence of SEQ ID NO: 86. In some embodiments, the AAE comprises the sequence of SEQ ID NO: 86.

[227] A host cell that expresses a heterologous polynucleotide encoding a PKS described in this disclosure may be capable of producing more (e.g., at least 1%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 100%, at least 125%, at least 150%, at least 175%, at least 200%, at least 300%, at least 400%, at least 500%, at least 600%, at least 700%, at least 800%, at least 900%, or at least 1,000% more) divaric acid in the presence of butyrate relative to a control. In some embodiments, the PKS is a bifunctional PKS. In some embodiments, the host cell further comprises a PKC. In some embodiments, a control is a host cell that does not express a heterologous gene encoding a PKS. In some embodiments, a control is a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1. In some embodiments, the host cell further comprises a heterologous polynucleotide encoding an AAE. In some embodiments, the AAE comprises a sequence that is at least 90% identical to the sequence of SEQ ID NO: 86. In some embodiments, the AAE comprises the sequence of SEQ ID NO: 86.

[228] A host cell that expresses a heterologous polynucleotide encoding a PKS described in this disclosure may be capable of producing more (*e.g.*, at least 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17-fold more) divaric acid in the presence of butyrate relative to a control. In some embodiments, the PKS is a bifunctional PKS. In some embodiments, the host cell further comprises a PKC. In some embodiments, a control is a host cell that does not express a heterologous gene encoding a PKS. In some embodiments, a control is a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1. In some embodiments, a host cell that expresses a heterologous polynucleotide encoding a PKS described in this disclosure and that also expresses an AAE, may be capable of producing at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 or 700 µg/L divaric acid. In some embodiments, the host cell further comprises a heterologous polynucleotide encoding an AAE. In some embodiments, the AAE comprises a sequence that is at least 90% identical to the sequence of SEQ ID NO: 86. In some embodiments, the AAE comprises the sequence of SEQ ID NO: 86.

[229] A host cell that expresses a heterologous polynucleotide encoding a PKS described in this disclosure and that also expresses one or more other enzymes involved in cannabinoid biosynthesis may be capable of producing a varinolic cannabinoid. In some embodiments, the varinolic cannabinoid is CBGV, CBGVA, THCVA and/or CBCVA.

[230] Methods for production of cannabinoids and cannabinoid precursors can further include expression of one or more of: a polykeide cyclase (PKC); a prenyltransferase (PT) and a terminal synthase (TS).

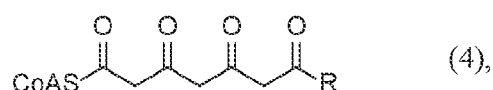
[231] In some embodiments, a method described in this disclosure is capable of catalyzing the production of more, for example, producing at least 1% (*e.g.*, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 100%, at least 125%, at least 150%, at least 175%, at least 200%, at least 300%, at least 400%, at least 500%, at least 600%, at least 700%, at least 800%, at least 900%, or at least 1,000%) more of a varinolic cannabinoid relative to a control. In some embodiments, a host cell and/or method described in this disclosure is capable of producing more (*e.g.*, at least 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17-fold more) of a varinolic cannabinoid relative to a control. In some embodiments, a control

is a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1. In some embodiments, the varinolic cannabinoid is CBGV, CBGVA, THCVA and/or CBCVA.

Polyketide Cyclase (PKC)

[232] A host cell described in this disclosure may comprise a PKC. As used in this application, a “PKC” refers to an enzyme that is capable of cyclizing a polyketide.

[233] In certain embodiments, a polyketide cyclase (PKC) catalyzes the cyclization of an oxo fatty acyl-CoA (e.g., a compound of Formula (4):



[234] or 3,5,7-trioxododecanoyl-CoA, 3,5,7-trioxodecanoyl-CoA) to the corresponding intramolecular cyclization product (e.g., compound of Formula (6), including olivetolic acid and divarinic acid). In some embodiments, a PKC catalyzes the formation of a compound which occurs in the presence of a PKS. PKC substrates include trioxoalkanoyl-CoA, such as 3,5,7-Trioxododecanoyl-CoA, or a compound of Formula (4):



wherein R is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl. In certain embodiments, a PKC catalyzes a compound of Formula (4):



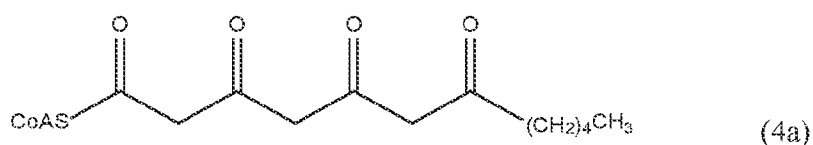
wherein R is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl; to form a compound of Formula (6):



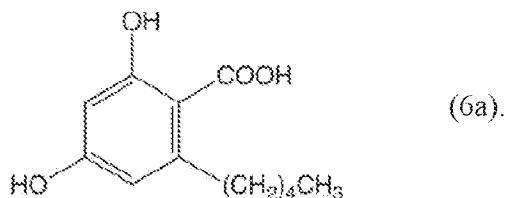
wherein R is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl; as substrates. R is as defined in this application. In some embodiments, R is a C2-C6 optionally substituted alkyl. In some embodiments, R is a propyl or pentyl. In some embodiments, R is pentyl. In some embodiments, R is propyl. In certain embodiments, a PKC is an olivetolic acid cyclase (OAC). In certain embodiments, a PKC is a divarinic acid cyclase (DAC).

[235] As one of ordinary skill in the art would appreciate a PKC could be obtained from any source, including naturally occurring sources and synthetic sources (*e.g.*, a non-naturally occurring PKC). In some embodiments, a PKC is from Cannabis. Non-limiting examples of PKCs include those disclosed in U.S. Patent No. 9,611,460; U.S. Patent No. 10,059,971; U.S. Patent Publication No. 2019/0169661, and PCT Publication No. WO2021/257915, which are incorporated by reference in this application in their entireties.

[236] In some embodiments, a PKC is an OAC. As used in this application, an “OAC” refers to an enzyme that is capable of catalyzing the formation of olivetolic acid (OA). In some embodiments, an OAC is an enzyme that is capable of using a substrate of Formula (4a) (3,5,7-trioxododecanoyl-CoA):



to form a compound of Formula (6a) (olivetolic acid):



[237] Olivetolic acid cyclase from *C. sativa* (CsOAC) is a 101 amino acid enzyme that performs non-decarboxylative cyclization of the tetraketide product of olivetol synthase (**FIG. 1** Structure 4a) via aldol condensation to form olivetolic acid (**FIG. 1** Structure 6a). CsOAC was identified and characterized by Gagne et al. (PNAS 2012) via transcriptome mining, and its cyclization function was recapitulated *in vitro* to demonstrate that CsOAC is required for formation of olivetolic acid in *C. sativa*. A crystal structure of the enzyme was published by Yang et al. (FEBS J. 2016 Mar;283(6):1088-106), which revealed that the enzyme is a homodimer and belongs to the α + β barrel (DABB) superfamily of protein folds. CsOAC is the only known plant polyketide cyclase. Multiple fungal Type III polyketide synthases have been identified that perform both polyketide synthase and cyclization functions (Funa *et al.*, J Biol Chem. 2007 May 11;282(19):14476-81); however, in plants such a dual function enzyme has not yet been discovered.

[238] A non-limiting example of an amino acid sequence of an OAC in *C. sativa* is provided by UniProtKB - I6WU39 (SEQ ID NO: 60), which catalyzes the formation of olivetolic acid (OA) from 3,5,7-Trioxododecanoyl-CoA.

[239] The sequence of UniProtKB - I6WU39 (SEQ ID NO: 60) is:

MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYT
HIVEVTFESVETIQDYIIHPAHVGFQDVYRSFWEKLLIFDYTPRK (SEQ ID NO: 60)

[240] A non-limiting example of a nucleic acid sequence encoding *C. sativa* OAC is:

atggcagtggaagcatttgattgtattgaagttcaagatgaaatcacagaagcccaaaaggaagaattttcaagacgtatgtgaatcttg
tgaatatcatcccagccatgaaagatgtatactggggtaaagatgtgactcaaaagaataaggaagaagggtacactcacatagttgag
gtaacatttgagagtggtgagactattcaggactacattatcatcctgcccatgttgattggagatgtctatcggtcttctctgggaaaaa
cttctcattttgactacacaccacgaaag (SEQ ID NO: 63).

[241] In some embodiments, a PKC comprises:

MAVKHLIVLKFKDEITNDQKEEFFKTYVNLLNIIPAMKDVYWGKDVTQKNKEEGYT
HIVEVTFESVETIQSYIIHPAHVGFAGFYRSFWEKLLIFDYTPRK (SEQ ID NO: 61).

[242] A non-limiting example of a nucleic acid sequence encoding SEQ ID NO: 61 is:

atggccgtcaagcatctgattgtattgaaatttaaggacgagatcacgaacgatcaaaaggaagaattttcaaaacttatgtaatttattg
aacataatccctgctatgaaggatgtctactggggtaaggatgtgacacagaaaaacaaggaagaaggctacactcacattgtggaag
ttacctcgagagcgttgaaactatcaatcctacatcatccaccccgctcatgtcggattcgggtccttctatcgttcctttgggaaaaatt
gttgattttcgactataccccaagaaag (SEQ ID NO: 90).

[243] In some embodiments, a PKC comprises a protein or nucleic acid sequence that is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or is 100% identical, including all values in between, to SEQ ID NO: 61 or 90. In some embodiments, a PKC comprises a sequence that is a conservatively substituted version of SEQ ID NO: 61.

Prenyltransferase (PT)

[244] A host cell described in this application may comprise a prenyltransferase (PT). As used in this application, a “PT” refers to an enzyme that is capable of transferring prenyl groups to acceptor molecule substrates. Non-limiting examples of prenyltransferases are described in U.S. Patent No. 7,544,498 and Kumano et al., *Bioorg Med Chem.* 2008 Sep 1; 16(17): 8117–8126 (e.g., NphB), PCT Publication No. WO2018/200888 (e.g., CsPT4), U.S. Patent No. 8,884,100 (e.g., CsPT1); Canadian Patent No. CA2718469; Valliere et al., *Nat Commun.* 2019 Feb 4;10(1):565 (e.g., NphB variants); PCT Publication Nos: WO2019/173770, WO2019/183152, and WO2020/210810 (e.g., NphB variants); Luo et al., *Nature* 2019 Mar;567(7746):123-126 (e.g., CsPT4); WO 2021/034848; PCT Publication No. WO2022/081615 (e.g., CsPT variants and chimeras), and PCT Publication No. WO2023/056350, which are incorporated by reference in their entireties. In some embodiments, a PT is capable of producing cannabigerolic acid (CBGA), cannabigerovarinic

acid (CBGVA), or other cannabinoids or cannabinoid-like substances. In some embodiments, a PT is cannabigerolic acid synthase (CBGAS). In some embodiments, a PT is cannabigerovarinic acid synthase (CBGVAS).

[245] In some embodiments, the PT is an NphB prenyltransferase. See, e.g., U.S. Patent No. 7,544,498; and Kumano et al., *Bioorg Med Chem.* 2008 Sep 1; 16(17): 8117–8126, which are incorporated by reference in this application in their entireties. In some embodiments, a PT corresponds to NphB from *Streptomyces sp.* (see, e.g., UniprotKB Accession No. Q4R2T2; see also SEQ ID NO: 2 of U.S. Patent No. 7,361,483). The protein sequence corresponding to UniprotKB Accession No. Q4R2T2 is provided by SEQ ID NO: 62:

MSEAADVERVYAAMEEAAGLLGVACARDKIYPLLSTFQDTLVEGGSVVVFMSMASG
RHSTELDFSISVPTSHGDPYATVVEKGLFPATGHPVDDLLADTQKHLPVSMFAIDGE
VTGGFKKTYAFFPTDNMPGVAELSAIPSPMPAVAENAELFARYGLDKVQMTSMDYK
KRQVNLYFSELSAQTLAESVLALVRELGLHVPNELGLKFKRSFSVYPTLNWETGK
IDRLCFAVISNDPTLVPSSDEGDIEKFHNYATKAPYAYVGEKRTL VYGLTLSPKEEYY
KLGAYYHITDVQRGLLKAFDSLED (SEQ ID NO:62).

[246] A non-limiting example of a nucleic acid sequence encoding NphB is:

atgtcagaagccgcagatgtcgaaagagttacgccgctatggaagaagccgccggttgtaggtgtgcctgtgccagagataagat
ctacccattgtgtctactttcaagatacattagttgaagggtggtcagttgtgtttctctatggcttcaggtagacattctacagaattgga
tttctctatctcagttccaacatcacatgggtgatccatacgtactgtgtgaaaaaggtttatttcagcaacaggtcatccagttgatgatt
tgttggctgatactcaaaagcatttgccagtttctatgtttgcaattgatgggtgaagttactgggtggttcaagaaaacttacgctttcttcca
actgataacatgccagggtgttcagaattatctgctattccatcaatgccaccagctgttcagaaaaatgcagaattattgctagatacgg
tttgataaggttcaaatgacatctatggattacaagaaaagacaagttaattgtactttctgaattatcagcacaaaactttggaagctga
atcagttttggcattagttagagaattgggtttacatgttccaaacgaattgggttgaagttttgtaaagatctttctcagtttatccaacttt
aaactgggaaacaggcaagatcgatagattatgtttcgagttatctctaacgatccaacattgggtccatcttcagatgaagggtgatac
gaaaagtttcataactacgctactaaagcaccataatgcttacgttggtgaaaagagaacattagtttaagggttgactttatcaccaaagga
agaatactacaagttgggtgcttactaccacattaccgacgtacaaagaggtttattgaaagcattcgatagttagaagac (SEQ ID
NO: 64).

[247] In other embodiments, a PT corresponds to CsPT1, which is disclosed as SEQ ID NO:2 in U.S. Patent No. 8,884,100 (*C. sativa*, corresponding to SEQ ID NO: 65 in this application):

MGLSSVCTFSFQTNHYHTLLNPHNNNPKTSLLCYRHPKTPIKYSYNNFPSKHCSTKSFH
LQNKCSSELSIAKNSIRAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTHAFTS
CACGLFGKELLHNTNLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRINKPDLPL
ASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGIVYSVPPFRWKQNPS
TAFLLNFLAHIITNFTFYASRAALGLPFELRPSFTFLLAFMKSMGSALALIKDASDVE
GDTKFGISTLASKYGSRNLTLCGIVLLSYVAAILAGIWPQAFNSNVMLLSHAILAF
WLILQTRDFALTNYDPEAGRFRFYEFMWKLYYAEYLVYVFI (SEQ ID NO: 65).

[248] In some embodiments, a PT corresponds to CsPT4, which is disclosed as SEQ ID NO:1 in PCT Publication No. WO2019/071000, corresponding to SEQ ID NO: 66 in this application:

MGLSLVCTFSFQTNHYHTLLNPHNKNPKNSLLSYQHPKTPHIKSSYDNFPSKYCLTKNF
HLLGLNSHNRISQSRISIRAGSDQIEGSPHHESDNSIATKILNFGHTCWKLQRPYVVK
GMISIAACGLFGRELFNNRHLFSWGLMWKAFFALVPILSFNFFAAIMNQIYDVIDRIN
KPDLPVSGEMSIETAWILSIIVALTGLIVTIKLSAPLFVFIYIFGIFAGFAYSVPPIRW
KQYPFTNFLITISSHVGLAFTSYSATTSALGLPFVWRPAFSFIHAFMTVMGMTIAFAKD
ISDIEGDAKYGVSTVATKLGARNMTFVVSGVLLLNLYLSISIGIWPQVFKNMILSH
AILAFCLIFQTRELALANYASAPSRQFFEFIWLLYYAEYFVYVFI (SEQ ID NO: 66).

[249] In some embodiments, a PT corresponds to a truncated CsPT4, which is provided as SEQ ID NO: 67:

MSAGSDQIEGSPHHESDNSIATKILNFGHTCWKLQRPYVVKGMISIAACGLFGRELFNN
RHLFSWGLMWKAFFALVPILSFNFFAAIMNQIYDVIDRINKPDLPLVSGEMSIETAW
ILSIIVALTGLIVTIKLSAPLFVFIYIFGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGLA
FTSYSATTSALGLPFVWRPAFSFIHAFMTVMGMTIAFAKD
ISDIEGDAKYGVSTVATKLGARNMTFVVSGVLLLNLYLSISIGIWPQVFKNMILSHAILAFCLIFQTRELALANY
ASAPSRQFFEFIWLLYYAEYFVYVFI (SEQ ID NO: 67).

[250] In some embodiments, a PT comprises the sequence of SEQ ID NO: 74:

MSAGSDQIEGSPHHESDNSIATKILNFGHTFWKLQRPYVVKGLISACGLFGKELLHN
 TNLISWSLMFKAFFALVPILSFNGFAAIMNQIYDVEIDRINKPDLPLVSGEMSIRTAWI
 LSHI VALTGLIVTIKMKGGPLFVFIYITGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGL
 AFTSYYASRAALGLPFELRPSFSFIIAFMTVMGMTIAFAKDISDIEGDAKYGVSTVAT
 KLGARNMTRVVSGVLLLNLYLSILAGIIWPF AFNSNVMILSHAILAFCLIFRTRELALA
 NYASAPSRQFFEFIWLLYYAEYLVYVFI (SEQ ID NO: 74).

[251] In some embodiments, the PT is encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 75:

atgtcgcaggatcagatcaaatagaaggagcccccatcagagtcggacaacagtatcgcgacaaaattcttaattttggccatac
 tttctggaagtacagcgtccttacgtagttaaaggtttgatttctatcgcttgcggtttgttcggttaaggaaactactccacaatacgaactg
 atatcatggctttaaattttaagccttcttgcctcgggtcccaattctaagttcaatggattcgcagccattatgaatcaaatctatgatgtt
 gaaatagataggattaacaagccggattaccactcgtgtccggagaaatgtctatacgcacagcttggatcctgtcgattattgtcgtc
 ttactggcttgatagtaccattaaaatgaaaggtggcctttatttggtttatctatataacgggtatattcgcaggttttgcctactctgtac
 ctccaattcgatggaagcagtatccgttcactaacttttgcacacatttcaagccatgttggactagccctcacatcctactatgcttctag
 agccgcgcttggcttaccatttgaactgcgccccctcatttagtttataaattgcattcatgacagttatggggatgactattgcctttgcaaag
 gacatttcagatattgaaggagatgcaaaatatgggtgtctctacggtagctactaaattaggtgctagaaatatgaccaggggtgtgtcgg
 ggggtgtgtctattaaattacttagtcagiattttggcaggaataatctggcccttcgcttttaattcaaatgcatgatcttatcgcacgcaattt
 tggccttctgtctcatctttagaacgagagaattagcccttgcctaactacgcaagtgtccttcccgtaatttttcgagtttatttgggtgtg
 tattatgcggaataacctgggtatcgttttcatt (SEQ ID NO: 75).

[252] In some embodiments, a PT comprises the sequence of SEQ ID NO: 76:

MSAGSDQIEGSPHHESDNSIATKILNFGHTFWKLQRPYVVKGLISACGLFGKELLHN
 TNLISWSLMFKAFFALVPILSFNFFAAAMNQIYDVEIDRINKPDLPLVSGEMSIRTAWI
 LSHI VALTGLIVTIKMKGGPLFVFIYIFGLFAGFAYSVPPIRWKQYPFTNFLITISSHVGL
 AFTSYYASRAALGLPFELRPSFSFIIAFMTVMGMTIAFAKDISDIEGDAKYGVSTVAT
 KLGARNMTRVVSGVLLLNLYLSILAGIIWPF AFNSNVMILSHAILAFCLIFRTRELALA
 NYASAPSRQFFEFIWLLYYAEYLVYVFI (SEQ ID NO: 76).

[253] In some embodiments, the PT is encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 77:

atgtccgcaggatcagatcaaatagaaggagccccatcacgagtcggacaacagtatcgcgacaaaattcttaattttggccatac
 ttcttgaagttacagcgctcttacgtagttaagggtctgatttctatcgcttgcggtttgttcggttaaggaaactactccacaatacgaacttg
 atatcatggctcttaattgtttaaagccttcttggctctggccccaaattctaagtttcaatttttcgcagccgcgatgaatcaaatctatgatgttg
 aaatagataggattaacaagccggattaccactcgtgtccggagaaatgtctataagaacagcttggatcctgtcgattattgtcgtctt
 actggcttgatagttaccattaaaaatgaaagggtggcccttatttggcttattctatataatttggcttattcgcaggttttgcctactctgtacctc
 caattcgatggaagcagtatccgttcactaacttttgcacacatttcaagccatgttggactagccttcacatcctactatgcttctagagc
 cgcgcttggctaccatttgaactgcgccccctatttagtttataattgcattcatgacagttatggggatgactattgcctttgcaaaggac
 atttcagatattgaaggagatgcaaaatatgggtgtctctacggtagctactaaattagggtgctagaaatatgacccgggtgtgtcgggggg
 tgttgcattataaattacttagtcagattttggcaggaataatctggcccttgcctttaaattcaaatgtcatgatcttatcgacgcaattttggc
 ctctgtctcatctttagaacgagagaattagcccttgcctaactcgcaagtgtccttcccgtcaattttcgagttatttgggtgctgtatta
 tgcggaatacctgggtatacgttttcatt (SEQ ID NO: 77).

[254] In some embodiments, a PT comprises the sequence of SEQ ID NO: 78:

MSAGSDQIEGSPHHESDNSIATKILNFGHTFWKLQRPYVVKGVISIAACGLFGKELLHN
 TNLISWSLMFKAFFALVPILSFNFFAAIVNQIYDVEIDRINKPDLPLVSGEMSIRTAWIL
 SIIVALTGLIVTIKMKGGPLFVFIYILGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGLA
 FTSYYASRAALGLPFELRPSFSFIIFMTVMGMTIAFAKDISDIEGDAKYGVSTVATKL
 GARNMTWVVSGLVLLNYLVSILAGIWPFAFNSNVMILSHAILAFCLIFRTRELALAN
 YASAPSRQFFEFIWLLYYAEYLVYVFI (SEQ ID NO: 78).

[255] In some embodiments, the PT is encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 79:

atgtccgcaggatcagatcaaatagaaggagccccatcacgagtcggacaacagtatcgcgacaaaattcttaattttggccatac
 ttcttgaagttacagcgctcttacgtagttaagggtctgatttctatcgcttgcggtttgttcggttaaggaaactactccacaatacgaacttg
 atatcatggctcttaattgtttaaagccttcttggctctggccccaaattctaagtttcaatttttcgcagccatttgaatcaaatctatgatgttg
 aatagataggattaacaagccggattaccactcgtgtccggagaaatgtctataagaacagcttggatcctgtcgattattgtcgtcttta
 ctggcttgatagttaccattaaaaatgaaagggtggcccttatttggcttattctatataatttgggtatattcgcaggttttgcctactctgtacctcc
 aattcgatggaagcagtatccgttcactaacttttgcacacatttcaagccatgttggactagccttcacatcctactatgcttctagagcc
 gcgcttggctaccatttgaactgcgccccctatttagtttataattgcattcatgacagttatggggatgactattgcctttgcaaaggaca
 tttcagatattgaaggagatgcaaaatatgggtgtctctacggtagctactaaattagggtgctagaaatatgacctgggttgtgtcggggggt
 gttgctattataaattacttagtcagattttggcaggaataatctggcccttgcctttaaattcaaatgtcatgatcttatcgacgcaattttggc
 ctctgtctcatctttaggacgagagaattagcccttgcctaactcgcaagtgtccttcccgtcaattttcgagttatttgggtgctgtatta
 tgcggaatacctgggtatacgttttcatt (SEQ ID NO: 79).

[256] In some embodiments, a PT comprises the sequence of SEQ ID NO: 80:

MSAGSDQIEGSPHHESDNSIATKILNFGHTCWKLQRPYVVKGLISACGLFGKELLHN
TNLISWSLMFKAFFALVPILSFNGFAAIMNQIYDVEIDRINKPDLPLVSGEMSIRTAWI
LSIIVALTGLIVTIKMKGGPLFVFIYIFGLFAGFAYSVPPIRWKQYPFTNFLITISSHVGL
AFTSYYASRAALGLPFELRPSFSFHAFMTVMGMTIAFAKDISDIEGDAKYGVSTVAT
KLGARNMTRVVSGVLLLNLYLSILAGIIWPFAFNSNVMILSHAILAFCLIFRTRELALA
NYASAPSRQFFFEFIWKLYYAEYLVYVFI (SEQ ID NO: 80).

[257] In some embodiments, the PT is encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 81:

atgtccgcaggatcagatcaatagaaggagccccatcacgagtcggacaacagtatcgcgacaaaattcttaattttggccatac
ttgttgaagttacagcgctcttacgtagftaaaggcttaatttctatcgcttgcggtttgttcggaaggaactactccacaatacgaactg
atatcatggtctttaatgtttaagccttcttgcctcgtcccaattctaagtttcaatgggttcgcagccattatgaatcaaatctatgatgtt
gagatagataggattaacaagccggattaccactcgtgtccggagaaatgtctataagaacagcttggatcctgtcgattattgtcgtc
ttactggcttgatagtaccattaaaatgaaagggtggccctttattgtgtttatctatatatttggttgttcgcaggttttgcttactctgtacct
ccaattcgaaggagcagatccgttcactaacttttgcaccatttcaagccatgttggactagccttcacatcctactatgcttctagag
ccgcgcttggctaccatttgaactgcgcccctcatttagttttataattgcattcatgacagttatggggatgactattgcctttgcaaagga
catttcagatattgaaggagatgcaaaatatggtgtctctacggtagctactaaattaggtgctagaaatatgaccagggtgtgtcgggg
gtgttgctatiaaattacttagtcagtatittggcaggaataatctggcccttcgcttttaattcaaatgtcatgatcttatcgcacgcaattttg
gccttctgtctcatctttagaacgagagaattagcccttgctaactacgcaagtgtccttcccgtaatttttcgagtttatttggaaactgt
attatgcggaataacctggtatagcttttcatt (SEQ ID NO: 81).

[258] In some embodiments, a PT comprises SEQ ID NO: 102:

MSAGSDQIEGSPHHESDNSIATKILNFGHTCWKLQRPYVVKGMISACGLFGRELFHN
TNLLSWSLMLKAFSSLMVILSFNGFAAIMNQIYDVIDRINKPDLPLVSGEMSIETAW
ILSIIVALTGLILTIKLNCGPLFISLYCVSIFAGFAYSVPPIRWKQYPFTNFLITISSHVGV
NFTCYASRAAFGLPFEMSPPTFILAFMTVMGMTIAFAKDISDIEGDAKYGVSTVAT
KLGARNMTRVVSGVLLLTYSAILAAIIWPQAFKSNVMLLSHAILAFCLIFQTRELAL
ANYASAPSRQFFFEFIWKLYHAEYLVYVFI.

[259] In some embodiments, a nucleic acid sequence encoding SEQ ID NO: 102 comprises:

[260] atgtcagccgggagcgaccagatagaaggaagtcctcatcacgagctcgataactccattgcgacaaaaatcctgaatttcgggtacacgtgctggaagttacaaaggccctatgtcgtgaaaggtatgatttcgatatgcttggcttggttgctgcgaactattcataataccaaccttttacttggtaattgatgtcaaggcattctcacttaattggtaatcctttctttaatggatttgctgccataatgaatcaaatttacgatgttgatattgaccgaattaacaaaccagatttggcactcgtagtgagagatgagtatcgaaactgcattggatactatcgatcattgtcgtctgacaggcttaataattgactattaagcttaatttggttcgctatttatatcattgtattgcgtatctatttcgctggattgcatacagcgttctccaattcgttggaacagtatcctttcaccaacttcttaataactatcagctcacacgtgggtgtaactttacatgtttatagcctccagagcagcccttggttaccctcgaaatgtctccgccatttactttatctcgttttatgaccgtaatgggaatgacaatagctttcgcgaaagacatttcagatatcgaaaggtgatgcgaaatatgggggtgtctactgttgctacgaagttagggtgcaagaaatatgacgagagttgtcagtggtgtctgttgctgacatacgttaagcgtatacttgacgtatcatttggccgcaagccttaagtcaaatgtgatgttggtatctcatgccatcctggcgttctgcctaattttccaaactagagagctagcgttgcaaatatgcttctgcacctagtaggcagtttttgaaattatctggaattgcattatgcagagtacttagtttatgttttcac (SEQ ID NO: 103).

[261] In some embodiments, a PT comprises SEQ ID NO: 158:

MSAGSDQIEGSPHHESDNSIATKILNFGHACWKFRPYVVKGLILACGLFGKELLHN
TNLISWSLMFKAFFALVPILSFNFFAAAINQIYDVEIDRINKPDLPLVSGEMSIRTAWIL
SIIVALTGLIVTIKMKGGPLFIFLYIFGLFLGFAYSVPPIRWKQYPFTNFLITVSSHVGLA
FTVYYASRAALGLPFELRPSFSFIHAFMTVMSITIAFAKDISDIEGDAKYGVSTLATKLG
AKNMTRVVSGLLLLNYVVAILAGINWPFAFNSNVMILSHAILAFCLIFRTRELALANY
ASAPSRQFFFEFIWNLYYAEYLVYVFI

[262] In some embodiments, a nucleic acid sequence encoding SEQ ID NO: 158 comprises:

atgtctgctggctctgaccaaattgaaggtccccgcacgaatcagataatagattgccacaaagatctaaactttgggcattgcatgttggaattccaaaggccctacgtcgtcaaggcattatactgatcgcttgcgtctgttcggaagggaattattgcataataactaacctaaataagctggagtttaattgttaaagctttcttcgcttagtgccaatcctaagcttaacttttcgccgcgcaattaaccagatttatgatgttgagatcgacaggataaataagccagatcttcattgggtatccgggtgaaatgtcaatacgtactgcattggatattatctattatcgttgcgtgaccggactgattgtaacaatcaaaatgaaaggtggccccctgtttattttctttatatatttggttggcttcctcggattcgttactcagtcacacatatcaggtggaagcagtagccattcacgaattttctgatcacgggttctagccacgtcgggttagcgttcacagtgactatgcacgagagctgccttggggcttcttcgagttgcgtccaagtttagtttatcattgcccttatgaccgtaatgagcattacgatcgcatcgcaaaaggacatttctgacatagagggggatgcaaaatacgggtgtctccactcttgcgacaaaattaggagctaaaaatatgactcgagtggtgtccggfllgttattactaaattatgttagctatattggcaggcatcaattggccggttgcgtttaaactccaacgttatgatactgagtcattgcatttttggtttgtctgatttttcgcacgcgtgagttggcgttgcaactatgcctctgcgccagcaggcagtttttgagttcatatggaatttgtaactatgccagtagtttagtctacgtatttatt (SEQ ID NO: 159).

[263] In some embodiments, a PT comprises a protein or nucleic acid sequence that is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or is 100% identical, including all values in between, to any one of SEQ ID NOs: 74-81, 102-103, and 158-159. In some embodiments, a PT comprises a sequence that is a conservatively substituted version of any one of SEQ ID NOs: 74, 76, 78, 80, 102, or 158.

[264] Functional expression of paralog *C. sativa* CBGAS enzymes in *S. cerevisiae* and production of the major cannabinoid CBGA has been reported (U.S. Patent Publication No. 2012/0144523, and Luo et al., *Nature* 2019 Mar;567(7746):123-126). Luo et al. reported the production of CBGA in *S. cerevisiae* by expressing a truncated version of a *C. sativa* CBGAS, CsPT4, with its native signal peptide removed. Without being bound by a particular theory, the integral-membrane nature of *C. sativa* CBGAS enzymes may render functional expression of *C. sativa* CBGAS enzymes in heterologous hosts challenging. Removal of transmembrane domain(s) or signal sequences or use of prenyltransferases that are not associated with the membrane and are not integral membrane proteins may facilitate increased interaction between the enzyme and available substrate, for example in the cellular cytosol and/or in organelles that may be targeted using peptides that confer localization.

[265] In some embodiments, the PT is a soluble PT. In some embodiments, the PT is a cytosolic PT. In some embodiments, the PT is a secreted protein. In some embodiments, the PT is not a membrane-associated protein. In some embodiments, the PT is not an integral membrane protein. In some embodiments, the PT does not comprise a transmembrane domain or a predicted transmembrane. In some embodiments, the PT may be primarily detected in the cytosol (e.g., detected in the cytosol to a greater extent than detected associated with the cell membrane). In some embodiments, the PT is a protein from which one or more transmembrane domains have been removed and/or mutated (e.g., by truncation, deletions, substitutions, insertions, and/or additions) so that the PT localizes or is predicted to localize in the cytosol of the host cell, or to cytosolic organelles within the host cell, or, in the case of bacterial hosts, in the periplasm. In some embodiments, the PT is a protein from which one or more

transmembrane domains have been removed or mutated (*e.g.*, by truncation, deletions, substitutions, insertions, and/or additions) so that the PT has increased localization to the cytosol, organelles, or periplasm of the host cell, as compared to membrane localization.

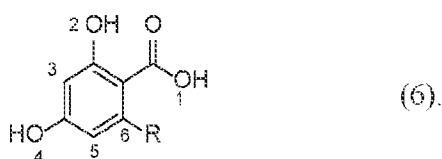
[266] Within the scope of the term “transmembrane domains” are predicted or putative transmembrane domains in addition to transmembrane domains that have been empirically determined. In general, transmembrane domains are characterized by a region of hydrophobicity that facilitates integration into the cell membrane. Methods of predicting whether a protein is a membrane protein or a membrane-associated protein are known in the art and may include, for example amino acid sequence analysis, hydropathy plots, and/or protein localization assays.

[267] In some embodiments, the PT is a protein from which a signal sequence has been removed and/or mutated so that the PT is not directed to the cellular secretory pathway. In some embodiments, the PT is a protein from which a signal sequence has been removed and/or mutated so that the PT is localized to the cytosol or has increased localization to the cytosol (*e.g.*, as compared to the secretory pathway).

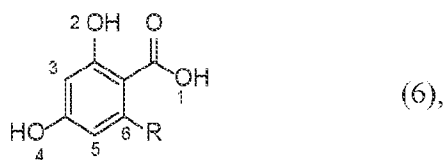
[268] In some embodiments, the PT is a secreted protein. In some embodiments, the PT contains a signal sequence.

[269] In some embodiments, a PT is a fusion protein. For example, a PT may be fused to one or more genes in the metabolic pathway of a host cell. In certain embodiments, a PT may be fused to mutant forms of one or more genes in the metabolic pathway of a host cell.

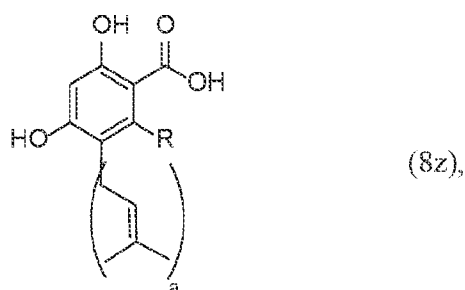
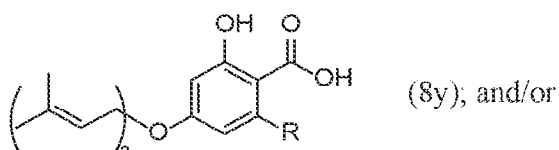
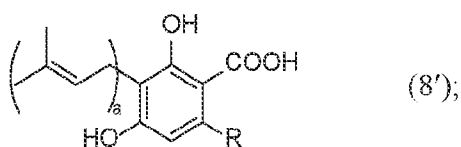
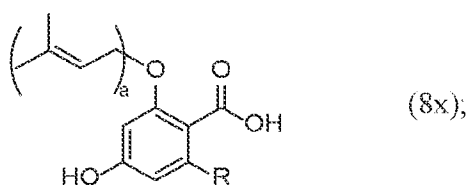
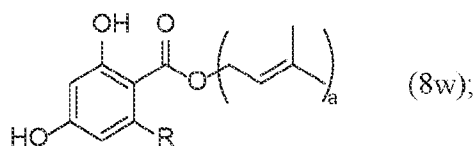
[270] In some embodiments, a PT described in this application transfers one or more prenyl groups to any of positions 1, 2, 3, 4, or 5 in a compound of Formula (6), shown below:



[271] In some embodiments, the PT transfers a prenyl group to any of positions 1, 2, 3, 4, or 5 in a compound of Formula (6), shown below:



to form a compound of one or more of Formula (8w), Formula (8x), Formula (8'), Formula (8y), Formula (8z):



or a pharmaceutically acceptable salt, solvate, hydrate, polymorph, co-crystal, tautomer, stereoisomer, isotopically labeled derivative, or prodrug thereof, wherein a is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

Terminal Synthases (TS)

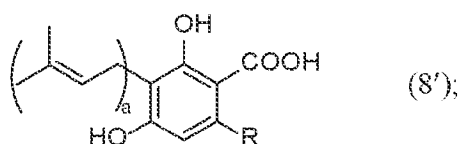
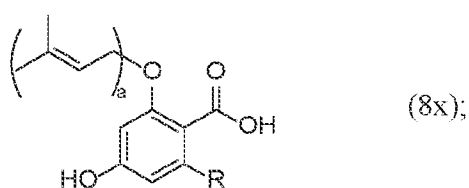
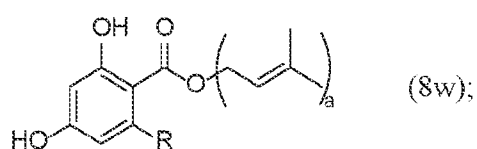
[272] A host cell described in this application may comprise a terminal synthase (TS). As used in this application, a “TS” refers to an enzyme that is capable of catalyzing oxidative

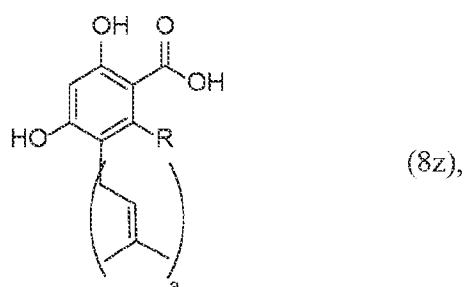
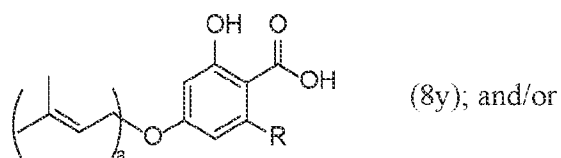
cyclization of a prenyl moiety (*e.g.*, terpene) to produce a ring-containing product (*e.g.*, heterocyclic ring-containing product). In certain embodiments, a TS is capable of catalyzing oxidative cyclization of a prenyl moiety (*e.g.*, terpene) to produce a carbocyclic-ring containing product (*e.g.*, cannabinoid). In certain embodiments, a TS is capable of catalyzing oxidative cyclization of a prenyl moiety (*e.g.*, terpene) to produce a heterocyclic-ring containing product (*e.g.*, cannabinoid). In certain embodiments, a TS is capable of catalyzing oxidative cyclization of a prenyl moiety (*e.g.*, terpene) to produce a cannabinoid. In some embodiments, a terminal synthase is a terpene cyclase that uses a terpenophenolic compound as a substrate.

[273] In some embodiments, a TS is a tetrahydrocannabinolic acid synthase (THCAS), a tetrahydrocannabivarinic acid synthase (THCVAS), a cannabidiolic acid synthase (CBDAS), a cannabidivarinic acid synthase (CBDVAS), a cannabichromenic acid synthase (CBCAS) and/or a cannabichromevarinic acid synthase (CBCVAS). As one of ordinary skill in the art would appreciate a TS could be obtained from any source, including naturally occurring sources and synthetic sources (*e.g.*, a non-naturally occurring TS).

a. Substrates

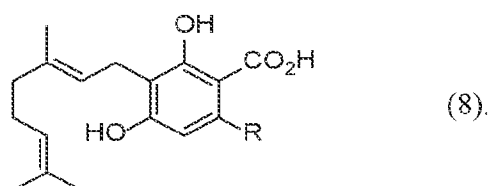
[274] A TS may be capable of using one or more substrates. In some instances, the location of the prenyl group and/or the R group differs between TS substrates. For example, a TS may be capable of using as a substrate one or more compounds of Formula (8w), Formula (8x), Formula (8'), Formula (8y), and/or Formula (8z):





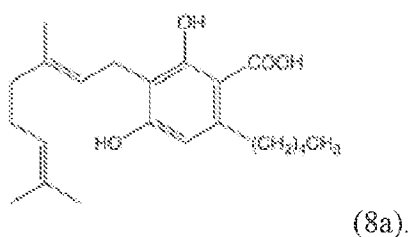
or a pharmaceutically acceptable salt, solvate, hydrate, polymorph, co-crystal, tautomer, stereoisomer, isotopically labeled derivative, or prodrug thereof, wherein *a* is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

[275] In certain embodiments, a compound of Formula (8') is a compound of Formula (8):



[276] In some embodiments, a TS catalyzes oxidative cyclization of the prenyl moiety (*e.g.*, terpene) of a compound of Formula (8) described in this application and shown in **FIG.**

2. In certain embodiments, a compound of Formula (8) is a compound of Formula (8a):

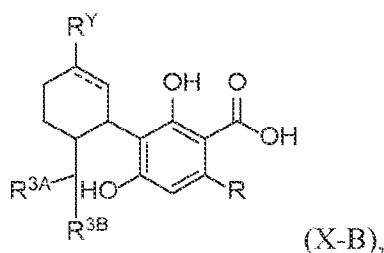
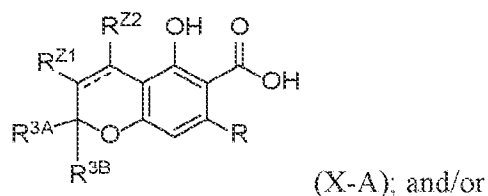


b. Products

[277] [185] In some embodiments, TS enzymes catalyze the formation of CBD-type cannabinoids, THC-type cannabinoids and/or CBC-type cannabinoids from CBG-type cannabinoids. In embodiments wherein CBGA is the substrate, the TS enzymes CBDAS, THCAS and CBCAS would generally catalyze the formation of cannabidiolic acid (CBDA), Δ^9 -tetrahydrocannabinolic acid (THCA) and cannabichromenic acid (CBCA), respectively. However, in some embodiments, a TS can produce more than one different product depending on reaction conditions. For example, the pH of the reaction environment may cause a THCAS or a CBDAS to produce CBCA in greater proportions than THCA or CBDAS, respectively (see, for example, U.S. Patent No. 9,359,625 to Winnicki and Donsky, incorporated by reference in its entirety). In some embodiments, a TS has a predetermined product specificity in intracellular conditions, such as cytosolic conditions or organelle conditions. By expressing a TS with a predetermined product specificity based on intracellular conditions, *in vivo* products produced by a cell expressing the TS may be more predictably produced. In some embodiments, a TS produces a desired product at a pH of 5.5. In some embodiments, a TS produces a desired product at a pH of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14. In some embodiments, a TS produces a desired product at a pH that is between 4.5 and 8.0. In some embodiments, a TS produces a desired product at a pH that is between 5 and 6. In some embodiments, a TS produces a desired product at a pH that is around 4.5, 4.6, 4.7, 4.8, 4.9, 5.0, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 7.0, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9, or 8.0, including all values in between. In some embodiments, the product profile of a TS is dependent on the TS's signal peptide because the signal peptide targets the TS to a particular intracellular location having particular intracellular conditions (*e.g.* a particular organelle) that regulate the type of product produced by the TS.

[278] A TS may be capable of using one or more substrates described in this application to produce one or more products. Non-limiting example of TS products are shown in **Table 1**. In some instances, a TS is capable of using one substrate to produce 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 different products. In some embodiments, a TS is capable of using more than one substrate to produce 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 different products.

[279] In some embodiments, a TS is capable of producing a compound of Formula (X-A) and/or a compound of Formula (X-B):



or a pharmaceutically acceptable salt, solvate, hydrate, polymorph, co-crystal, tautomer, stereoisomer, isotopically labeled derivative, or prodrug thereof;

wherein $\equiv$ is a double bond or a single bond, as valency permits;

R is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl;

R^{Z1} is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl;

R^{Z2} is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted aryl;

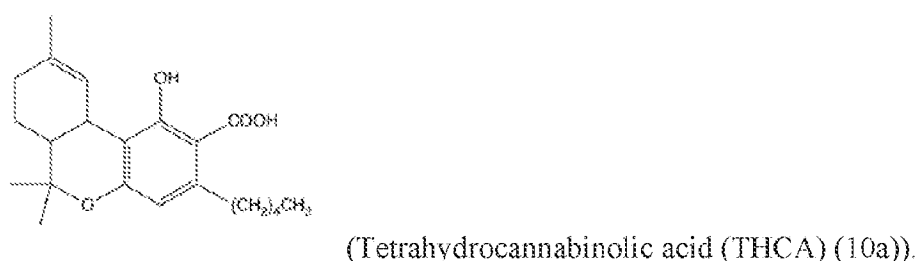
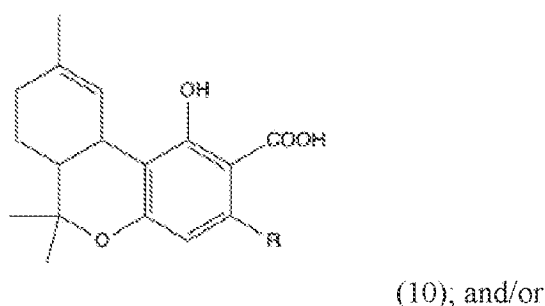
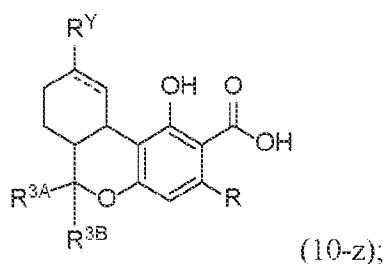
or optionally, R^{Z1} and R^{Z2} are taken together with their intervening atoms to form an optionally substituted carbocyclic ring;

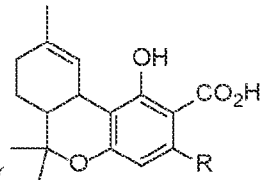
R^{3A} is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, or optionally substituted alkynyl;

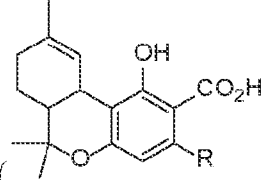
R^{3B} is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, or optionally substituted alkynyl; and/or

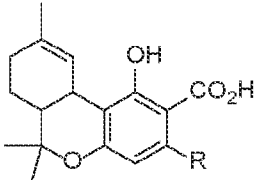
R^Y is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, or optionally substituted alkynyl.

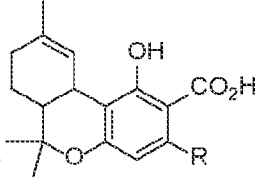
[280] In some embodiments, a compound of Formula (X-A) is:

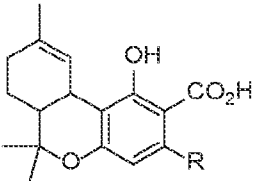
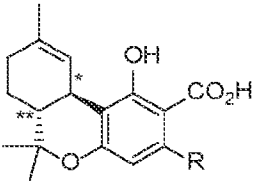


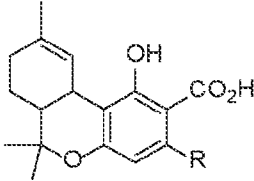
[281] In certain embodiments, a compound of Formula (10) () has a chiral atom labeled with * at carbon 10 and a chiral atom labeled with ** at carbon 6. In

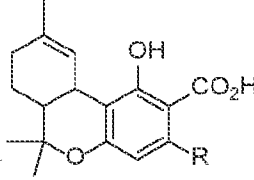
certain embodiments, in a compound of Formula (10) (), the chiral atom labeled with * at carbon 10 is of the *R*-configuration or *S*-configuration; and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration. In certain embodiments, in a compound

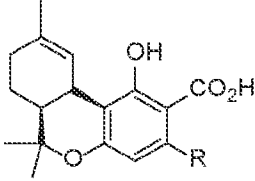
of Formula (10) () , the chiral atom labeled with * at carbon 10 is of the *S*-configuration; and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration or *S*-

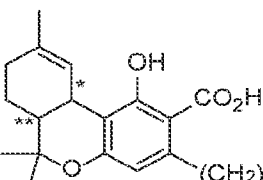
configuration. In certain embodiments, in a compound of Formula (10) () , the chiral atom labeled with * at carbon 10 is of the *R*-configuration and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration. In certain embodiments, a compound of Formula

(10) () , is of the formula:  . In certain embodiments,

in a compound of Formula (10) () , the chiral atom labeled with * at carbon 10 is of the *S*-configuration and a chiral atom labeled with ** at carbon 6 is of the *S*-

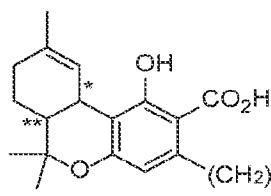
configuration. In certain embodiments, a compound of Formula (10) () , is

of the formula:  .

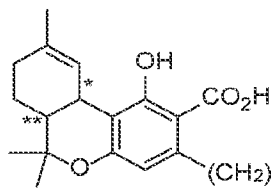
[282] In certain embodiments, a compound of Formula (10a) ()

has a chiral atom labeled with * at carbon 10 and a chiral atom

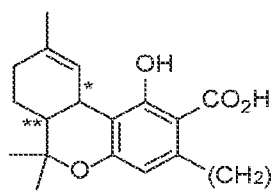
labeled with ** at carbon 6. In certain embodiments, in a compound of Formula (10a) (



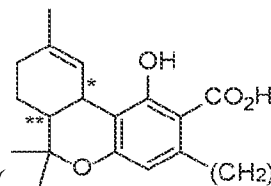
), the chiral atom labeled with * at carbon 10 is of the *R*-configuration or *S*-configuration; and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration. In certain embodiments, in a compound of Formula (10a) (



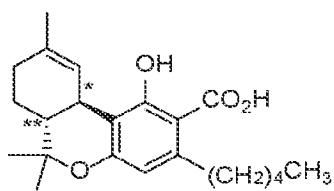
), the chiral atom labeled with * at carbon 10 is of the *S*-configuration; and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration or *S*-configuration. In certain embodiments, in a compound of Formula (10a) (



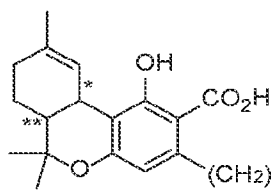
), the chiral atom labeled with * at carbon 10 is of the *R*-configuration and a chiral atom labeled with ** at carbon 6 is of the *R*-configuration. In certain



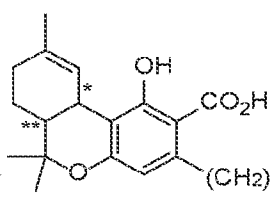
embodiments, a compound of Formula (10a) (

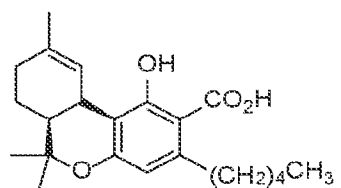


. In certain embodiments, in a compound of Formula (10a) (

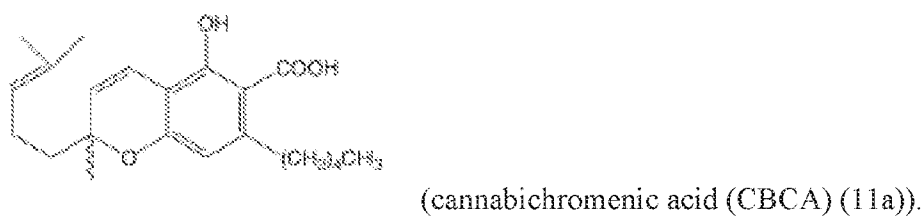
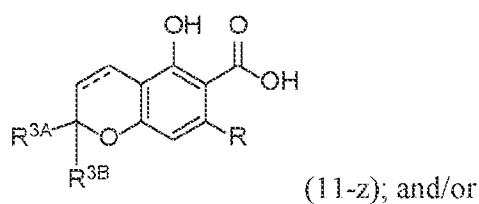
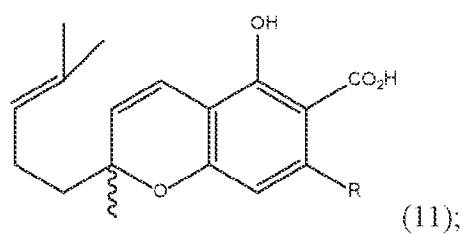


), the chiral atom labeled with * at carbon 10 is of the *S*-configuration and a chiral atom labeled with ** at carbon 6 is of the *S*-configuration. In certain

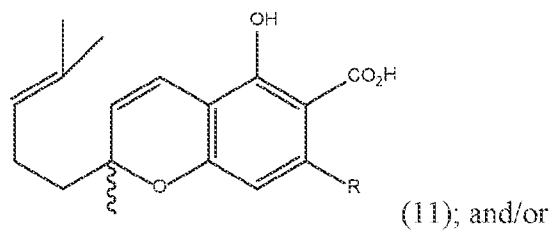
embodiments, a compound of Formula (10a) () is of the formula:

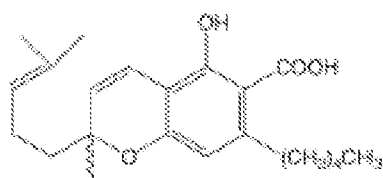


[283] In some embodiments, a compound of Formula (X-A) is:



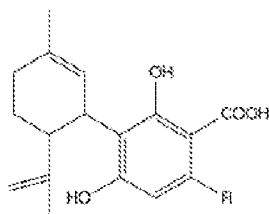
[284] In some embodiments, a compound of Formula (X-A) is:



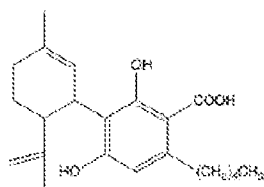


(cannabichromenic acid (CBCA) (11a)).

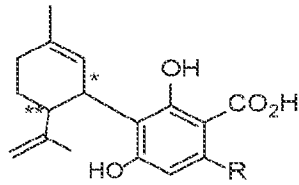
[285] In some embodiments, a compound of Formula (X-B) is:

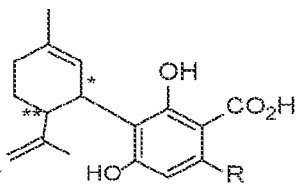


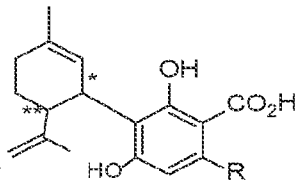
(9); and/or

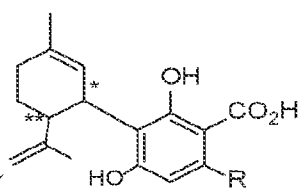


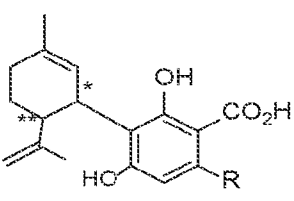
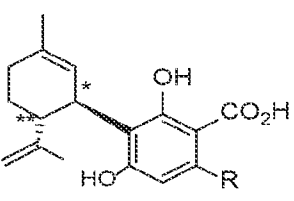
(cannabidiolic acid (CBDA) (9a)).

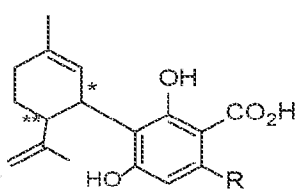
[286] In certain embodiments, a compound of Formula (9) () has a chiral atom labeled with * at carbon 3 and a chiral atom labeled with ** at carbon 4. In

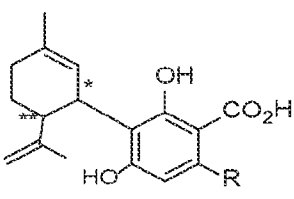
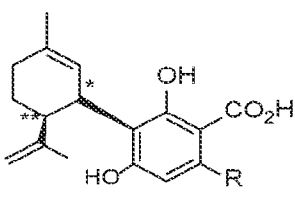
certain embodiments, in a compound of Formula (9) (), the chiral atom labeled with * at carbon 3 is of the *R*-configuration or *S*-configuration; and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration. In certain embodiments, in a compound of

Formula (9) (), the chiral atom labeled with * at carbon 3 is of the *S*-configuration; and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration or *S*-

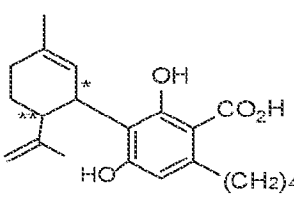
configuration. In certain embodiments, in a compound of Formula (9) () , the chiral atom labeled with * at carbon 3 is of the *R*-configuration and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration. In certain embodiments, a compound of Formula

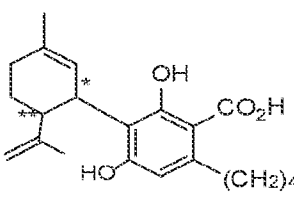
(9) () , is of the formula:  . In certain

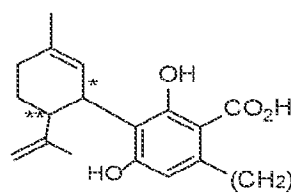
embodiments, in a compound of Formula (9) () , the chiral atom labeled with * at carbon 3 is of the *S*-configuration and a chiral atom labeled with ** at carbon 4 is of the *S*-configuration. In certain embodiments, a compound of Formula (9) (

) , is of the formula:  .

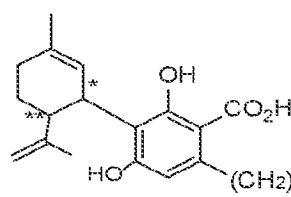
[287] In certain embodiments, a compound of Formula (9a) (CBDA) (

) has a chiral atom labeled with * at carbon 3 and a chiral atom labeled with ** at carbon 4. In certain embodiments, in a compound of Formula (9a) (

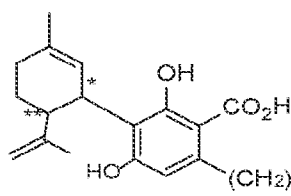
) , the chiral atom labeled with * at carbon 3 is of the *R*-configuration or *S*-configuration; and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration. In certain embodiments, in a compound of Formula (9a) (



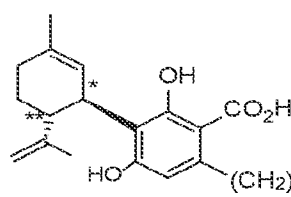
, the chiral atom labeled with * at carbon 3 is of the *S*-configuration; and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration or *S*-configuration. In certain embodiments, in a compound of Formula (9a) (



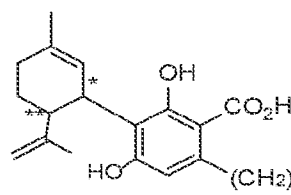
, the chiral atom labeled with * at carbon 3 is of the *R*-configuration and a chiral atom labeled with ** at carbon 4 is of the *R*-configuration. In certain



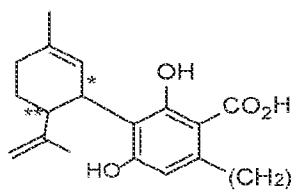
embodiments, a compound of Formula (9a) (



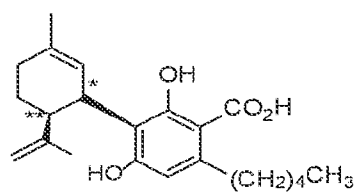
. In certain embodiments, in a compound of Formula (9a) (



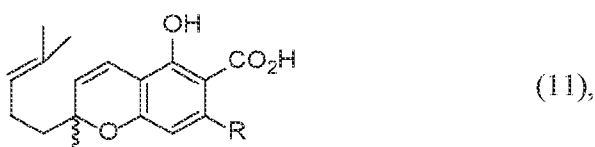
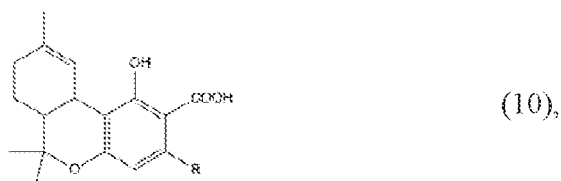
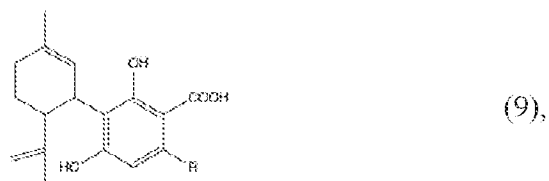
, the chiral atom labeled with * at carbon 3 is of the *S*-configuration and a chiral atom labeled with ** at carbon 4 is of the *S*-configuration. In certain



embodiments, a compound of Formula (9a) (



[288] In some embodiments, as shown in FIG. 2, a TS is capable of producing a cannabinoid from the product of a PT, including, without limitation, an enzyme capable of producing a compound of Formula (9), (10), or (11):



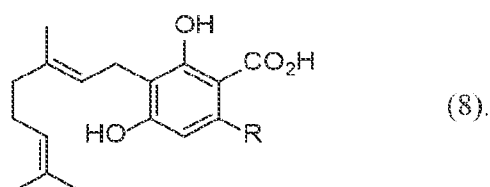
or a pharmaceutically acceptable salt, solvate, hydrate, polymorph, co-crystal, tautomer, stereoisomer, isotopically labeled derivative, or prodrug thereof, wherein R is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, or optionally substituted

aryl; produced from a compound of Formula (8').

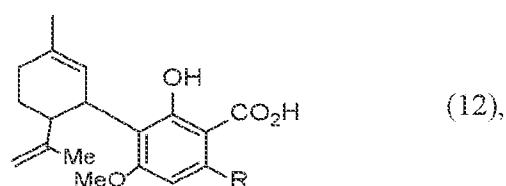
(8'),

wherein a is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10; and R is hydrogen, optionally substituted acyl, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl,

optionally substituted carbocyclyl, or optionally substituted aryl; or using any other substrate. In certain embodiments, a compound of Formula (8') is a compound of Formula (8):



[289] In certain embodiments, a compound of Formula (9), (10), or (11) is produced using a TS from a substrate compound of Formula (8') (e.g., compound of Formula (8)), for example. Non-limiting examples of substrate compounds of Formula (8') include but are not limited to cannabigerolic acid (CBGA), cannabigerovarinic acid (CBGVA), or cannabinerolic acid. In certain embodiments, at least one of the hydroxyl groups of the product compounds of Formula (9), (10), or (11) is further methylated. In certain embodiments, a compound of Formula (9) is methylated to form a compound of Formula (12):



or a pharmaceutically acceptable salt, solvate, hydrate, polymorph, co-crystal, tautomer, stereoisomer, isotopically labeled derivative, or prodrug thereof.

Tetrahydrocannabinolic acid synthase (THCAS)

[290] A host cell described in this application may comprise a TS that is a tetrahydrocannabinolic acid synthase (THCAS). As used in this application “tetrahydrocannabinolic acid synthase (THCAS)” or “ Δ^1 -tetrahydrocannabinolic acid (THCA) synthase” refers to an enzyme that is capable of catalyzing oxidative cyclization of a prenyl moiety (e.g., terpene) of a compound of Formula (8) to produce a ring-containing product (e.g., heterocyclic ring-containing product, carbocyclic-ring containing product) of Formula (10). In certain embodiments, a THCAS refers to an enzyme that is capable of producing Δ^9 -tetrahydrocannabinolic acid (Δ^9 -THCA, THCA, Δ^9 -Tetrahydro-cannabivarinic acid A (Δ^9 -THCVA-C3 A), THCVA, THCP, or a compound of Formula 10(a), from a compound of Formula (8). In certain embodiments, a THCAS is capable of producing Δ^9 -

tetrahydrocannabinolic acid (Δ^9 -THCA, THCA, or a compound of Formula 10(a)). In certain embodiments, a THCAS is capable of producing Δ^9 -tetrahydrocannabivarinic acid (Δ^9 -THCVA, THCVA, or a compound of Formula 10 where R is n-propyl).

[291] In some embodiments, a THCAS may catalyze the oxidative cyclization of substrates, such as 3-prenyl-2,4-dihydroxy-6-alkylbenzoic acids. In some embodiments, a THCAS may use cannabigerolic acid (CBGA) as a substrate. In some embodiments, the THCAS produces Δ^9 -THCA from CBGA. In some embodiments, a THCAS may catalyze the oxidative cyclization of cannabigerovarinic acid (CBGVA). In some embodiments, a THCAS exhibits specificity for CBGA substrates as compared to other substrates. In some embodiments, a THCAS may use a compound of Formula (8) of FIG. 2 where R is C4 alkyl (e.g., n-butyl) or R is C7 alkyl (e.g., n-heptyl) as a substrate. In some embodiments, a THCAS may use a compound of Formula (8) where R is C4 alkyl (e.g., n-butyl) as a substrate. In some embodiments, a THCAS may use a compound of Formula (8) of FIG. 2 where R is C7 alkyl (e.g., n-heptyl) as a substrate. In some embodiments, the THCAS exhibits specificity for substrates that can result in THCP as a product.

[292] In some embodiments, a THCAS is from *C. sativa*. *C. sativa* THCAS performs the oxidative cyclization of the geranyl moiety of Cannabigerolic Acid (CBGA) (FIG. 1 Structure 8a) to form Tetrahydrocannabinolic Acid (FIG. 1 Structure 10a) using covalently bound flavin adenine dinucleotide (FAD) as a cofactor and molecular oxygen as the final electron acceptor. THCAS was first discovered and characterized by Taura et al. (JACS. 1995) following extraction of the enzyme from the leaf buds of *C. sativa* and confirmation of its THCA synthase activity *in vitro* upon the addition of CBGA as a substrate. Additional analysis indicated that the enzyme is a monomer and possesses FAD binding and Berberine Bridge Enzyme (BBE) sequence motifs. A crystal structure of the enzyme published by Shoyama et al. (*J Mol Biol.* 2012 Oct 12;423(1):96-105) revealed that the enzyme covalently binds to a molecule of the cofactor FAD. See also, e.g., Sirikantarams et al., *J. Biol. Chem.* 2004 Sept 17; 279(38):39767-39774. There are several THCAS isozymes in *Cannabis sativa*.

[293] In some embodiments, a *C. sativa* THCAS (Uniprot KB Accession No.: I1V0C5) comprises the amino acid sequence shown below, in which the signal peptide is underlined and bolded:

MNCSAFSEWFYCKIIFFLSFNIOISIANNPQENFLKCFSEYIPNNPANPKFIYTQHDQL
YMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQASILCSKKVGLQIRTRSGGHDAEG
MSYISQVPFVVVDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENFSFPGG
YCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFW
AIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDK
DLVLMTHFITKNITDNHGKNKTTVHGYFSSIFHGGVDSLVDLMNKSFPPELGIIKTDC
KEFSWIDTTIFYSGVVNFENTANFKKEILLDRSAGKKTAFSIKLDYVKKPIPETAMVKIL
EKL YEEDVGVGMYVLYPYGGIMEEISESAIPFPHRAGIMYELWYTASWEKQEDNEK
HINWVRSVYNFTTPYVSQNPRLAYLNRYDLDLGKTNPESPNNYTQARIWGEKYFGK
NFNRLVKVKTADPNNFRNEQSIPPLPHHH (SEQ ID NO: 68).

[294] In some embodiments, a THCAS comprises the sequence shown below:
NPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTP
SNVSHIQASILCSKKVGLQIRTRSGGHDAEGMSYISQVPFVVVDLRNMHSIKIDVHSQ
TAWVEAGATLGEVYYWINEKNENFSFPGGYCPTVGVGGHFSGGGYGALMRNYGLA
ADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTI
FSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYF
SSIFHGGVDSLVDLMNKSFPPELGIIKTDCKEFSWIDTTIFYSGVVNFENTANFKKEILL
RSAGKKTAFSIKLDYVKKPIPETAMVKILEKL YEEDVGVGMYVLYPYGGIMEEISES
AIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNRY
DLDLGKTNPESPNNYTQARIWGEKYFGKNFNRLVKVKTADPNNFRNEQSIPPLPH
HHH (SEQ ID NO: 69).

[295] A non-limiting example of a nucleotide sequence encoding SEQ ID NO: 69 is:
aaccgcgaagaaactttctaaaatgcttttctgaatacattcctaacaacctgccaacccgaagttatctacacacaacacgatcaatt
gtatatgagcgtgtgaatagtacaatacagaacctgaggtttacatccgacacaacgccgaacccgtagtgatcgacacacctcca
acgtaagccacattcaggcaagcattttatgcagcaagaaagtcggactgcagataaggacgaggtccggaggacacgacgccgaa
gggatgagctatatctcccaggfacctttgtggtggttagacttgagaaatgcactctatcaagatagacgttcactccaaacccgctt
gggttgaggcgggagccaccttggtgaggtctactactggatcaacgaaaagaatgaaaatttagctttcctgggggatattgccca
actgtaggtgttgccggccacttctcaggaggcgggttatggggccttgatgcgtaactacggacttgcggccgacaacattatagacg
cacacttagtgaatgtagacggcaagtttagacaggaagagcatgggtgaggatctttttgggcaatttagaggcggaggggggaga
aaatttggaaattatcgctgcttgaaaattaagctagtlcgcggtaccgagcaaaagcactatattctctgtaaaaaagaacatggagala
catgggttggtgaagctttttaataagtgcaaaacatcgcggtacaagtacgacaaaagatctgggtctgatgacgcattttataacgaaaa
atatcaccgacaaccacggaaaaaacaaccacagtacatggctacttctctagtatattcatgggggagtcgattctctggttgattt

aatgaacaaatcattcccagagttgggtataaagaagacagactgtaaggagttctcttgattgacacaaactatattctattcaggcgta
 gtcaactttaacacggcgcaatttcaaaaaagagatccttctggacagatccgcaggttaagaaaactgcgttctctatcaaattggactatg
 tgaagaagcctattcccgaaccgcgatgggtcaagatacttgagaaattatacaggaagatgtgggagttggaatgtacgtactttatc
 cctatgggtgggataatggaagaaatcagcgagagcgccattccattccccatcgtgccggcatcatgtacgagctgtggtatactgcg
 agttgggagaagcaagaagacaacgaaaagcacattaactgggtcagatcagtttacaatttcaccaccccatacgtgtcccagaatc
 cgcgtctggcttacttgaactaccgtgatcttgacctgggtaaaacgaacccggagtcacccaacaattacactcaagctagaatctgg
 ggagagaaatacttgggaagaactcaacaggttagtaaagggttaaaaccaaggcagatccaaacaacttttttagaaatgaacaatc
 cattcccccgctacccccgcaccatcac (SEQ ID NO: 70).

[296] In some embodiments, a *C. sativa* THCAS comprises the amino acid sequence set forth in UniProtKB - Q8GTB6 (SEQ ID NO: 71):

[297] MNCSAFSFWFVCKIIFFLSFHIQISIANPRENFLKCFSKHIPNNVANPKL
 VYTQHDQLYMSILNSTIQNLRFISDTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRS
 GGHDAEGMSYISQVPFVVVDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKN
 ENLSFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKS
 MGEDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQN
 IAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFSSIFHGGVDSLVDLMNKSFP
 ELGIKKTDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYVKKPI
 PETAMVKILEKLYEEDVGAGMYVLYPYGGIMEEISESAIPFPHRAGIMYELWYTASWEK
 QEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNRYRDLDLGKTNHASPNNYTQARIW
 GEKYFGKNFNRLVKVTKVDPNFFRNEQSIPPLPPHHH (SEQ ID NO: 71)

[298] In some embodiments, a THCAS comprises the amino acid sequence:

MNPQENFLKCFSKHIPNNVANPKLVYTQNDSLYMSILNSTIQNLRFTSDTPKPLVIV
 TPSNVSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQVPFVVVDLRNMHSIKIDVE
 SQTAWVEAGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRNYG
 LAADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAAWKIKLVVPVSKV
 TIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGY
 FSSIFHGGVDSLVDLMNKSFP
 ELGIKKTDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYVKKPI
 PETAMVKILEKLYEEDVGAGMYVLYPYGGIMEEISESAIPFPHRAGIMYELWYTASWEK
 QEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNRYRDLDLGKNNEESPNNYTQARIW
 GEKYFGKNFNRLVKVTKVDPNFFRNEQSIPPLPPHHH (SEQ ID NO: 88).

[299] A non-limiting example of a nucleotide sequence encoding SEQ ID NO: 88 is:

atgaatccccaggagaactttcttaagtgtttcagtaaacaatccccaaataacgtggcgaatccgaaattagtatacacgcaaaacgatt
ccctatatatgagcattctgaatagtaccattcaaaacttgcgcttacttcggacacaaactccaaagccttagtcatagttactcctcaaa
tgtttcccatattcaggctacaatcctctgctctaaaaagttgggttgcaaatacgtactagatcaggaggccacgatccgaaggtatg
tcttatatctcccaagtgcattcggtgtcgtcgacttgaggaacatgcattctattaagattgatgtagaatccagaccgcatgggttgaa
gctgggtgccacattaggtgaagtatactattggataaacgagaaaaatgaaaattatcgtttcctggaggctattgtccaactgttgggtg
cgggtggctatttctcaggcgggggctacggagcattgatgcgaaactacggactagcagctgataacattatagacgccatctcgtg
aatgtggatggtaaagttcttgatagaaagagcatgggtgaagatttgtttggcgatcagaggcgggtggaggggaaaaactttggatt
attgcggcttgaaaaataaagctgggtcccggtccctcgaaagtaacaatctttctgtaaagaaaaatatggaatacatgggttggtca
agttattcaataaatggcaaaacattgcttataagtatgataaagacctcgttctgatgactcattttattacgaaaaatattaccgacaatca
cggtaagaataaaactacagtgcacgggtacttttctcaattttccatgggtggcgttgactctctagtagatttaataacaaaaagtttccct
gagttagggaalcaagaaaacggattgtaagaactgtcttgatcgacacacacatattctattcgggtgtgttaactttaataaccgcaa
actttaaaaaggaaattttgttagatcgctctgctggaagaaaacagcatttagtatcaaaactgactatgtaagaaaccgataccgga
gactgccatgggttaagatacttgagaagctatacgaagaggatgtgggagctggcatgtacgttctctacccctacggcgggtataatgg
aagaaattcagaatccgcaatccattcccacacagagcaggtattatgtatgaactgtggtatactgcctcatgggagaaacaagaa
gataacgaaaaacacattaattgggtccgtagcgtctacaatttactacacatatgtatccaagaacccgagattggcttacttgaatta
tagagacttggatctaggaaaaataatgaggaaagccgaataattatacacaggcaaggatttggggcgagaagtatttcggcaag
aactcaacagattagtaaaagttaagaccaaagtgacccaataacttttttagaaatgaacaaagtatccctccactcccaccacatc
atcat (SEQ ID NO: 89).

[300] In some embodiments, a THCAS comprises the amino acid sequence:

MNPQENFLKCF SKHIPNNVANPKLVYTQNDQLYMSILNSTIQNLRFTSDTTPKPLVIV
TPSNVSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQVPFVVVDLRNMHSIKIDVH
SQTAWVEAGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRNYG
LAADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAAWKIKLVPVPSKV
TIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHG
YFSSIFHGGVDSLVDLMNKSFPGLIKKTDCKEFSWIDTTIFYSGVVNFNTANFKKEIL
LDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGAGMYVLYPYGGIMEEIS
ESAIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSKNPRLAYLN
YRDLDLGKNNEESPNNYTQARIWGEKYFGKNFNRLVKVTKVDPNNFFRNEQSIPP
LPPHHH (SEQ ID NO: 95).

[301] A non-limiting example of a nucleotide sequence encoding SEQ ID NO: 95 is:

atgaatccccaggagaactttcttaagtgtttcagtaaacatatccaaataacgtggcgaatccgaaattagtatacacgcaaaacgatac
agctatatatgagcattctgaatagtaccattcaaaacttgcgtttacatcggacacaactccaaagcctttagtcatagttactccttcaa
atgtttcccatattcaggctacaatcctctgctctaaaaaagtgtgggttgcaatacgtactagatcaggaggccacgatgccgaaggat
gtcttatatctcccaagtgccattcggtgtcgtcgacttgaggaaacatgcattctattaagattgatgtacactccagaccgcatgggttga
agctgggtgccacattaggigaagtatactattggataaacgagaaaaatgaaaatttactgtttcctggaggctattgtccaactgttgggtg
tcgggtggtcattttcaggcggggggctacggagcattgatgcgaaactacggactagcagctgataacattatagacgccccatctcgtg
aatgtggatggtaaagtcttgatagaaagagcatgggtgaagattgttttggcgatcagaggcggtggagggggaaaactttgggtatt
attgctggcttggaaaaataagctgggtcccggttccctcgaaagtaacaatcttttctgtaaagaaaaatatggaaatacatgggttgggtca
agttattcaataaatggcaaaacattgcttataagtatgataaacctcgttctgatgactcattttattacgaaaaatattaccgacaatca
cggtagaataaaaactacagtgcacgggtacttttctcaattttccatgggtggcgttgactctctagtagatttaataacaaaaagtttccct
gagttaggatcaagaaaacgggattgtaagaattttctggatcgacacaaccatattctattcggtgttgttaactttaataccgcaaa
ctttaaaaaggaaattttgttagatcgctctgctggaaagaaaacagcatttagtatcaaaacttgactatgtaagaaaccgatacccgag
actgccatgggttaagatacttgagaagctatacgaagaggatgtgggagctggcatgtactgttctctacccttacggcggtataatggaa
gaaatttcagaatccgcaatccattccacacagagcaggtattatgtatgaactgtggtatactgectcatgggagaaacaagaagat
aacgaaaaacacattaattgggtccgtagcgtctacaatttcactacaccatatgtatccaagaaccgagattggcttacttgaattatag
agacttggatctaggaaaaataatgaagaaagccgaataattatacacaggcaaggatttggggcgagaagtatttcggcaagaac
ttcaacagattagtaaaagttaagaccaaagttgaccccaataacttttttagaaatgaacaaagfatccctccactcccaccacatcatca
t (SEQ ID NO: 96).

[302] Additional non-limiting examples of THCAS enzymes may also be found in U.S. Patent No. 9,512,391, U.S. Patent Application Publication No. 2018/0179564, and PCT Publication No. WO 2022/011175, which are incorporated by reference in this application in their entireties.

[303] In some embodiments, a THCAS comprises a protein or nucleic acid sequence that is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or is 100% identical, including all values in between, to SEQ ID NO: 88-89 or 95-96. In some embodiments, a THCAS comprises a sequence that is a conservatively substituted version of SEQ ID NO: 88 or 95.

Cannabidiolic acid synthase (CBDAS)

[304] A host cell described in this application may comprise a TS that is a cannabidiolic acid synthase (CBDAS). As used in this application, a “CBDAS” refers to an enzyme that is capable of catalyzing oxidative cyclization of a prenyl moiety (*e.g.*, terpene) of a compound of Formula (8) to produce a compound of Formula (9). In some embodiments, a compound of Formula 9 is a compound of Formula (9a) (cannabidiolic acid (CBDA)), CBDVA, or CBDP. A CBDAS may use cannabigerolic acid (CBGA) or cannabimerolic acid as a substrate. In some embodiments, a cannabidiolic acid synthase is capable of oxidative cyclization of cannabigerolic acid (CBGA) to produce cannabidiolic acid (CBDA). In some embodiments, the CBDAS may catalyze the oxidative cyclization of other substrates, such as 3-geranyl-2,4-dihydro-6-alkylbenzoic acids like cannabigerovarinic acid (CBGVA) or a substrate of Formula (8) with R as a C7 alkyl (heptyl) group (cannabigerophoric acid (CBGPA)). In some embodiments, the CBDAS exhibits specificity for CBGA substrates.

[305] In some embodiments, a CBDAS is from *Cannabis*. In *C. sativa*, CBDAS is encoded by the CBDAS gene and is a flavoenzyme. A non-limiting example of a CBDAS is provided by UniProtKB - A6P6V9 (SEQ ID NO: 72) from *C. sativa*:

MKCSTFSFWFVCKIIFFFSFENIQTSIANPRENFLKCFSQYIPNNATNLKLVYTQNNPLY
MSVLNSTIHNLRFTSDTTPKPLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGHDSEGM
SYISQVPFVIVDLRNMRSIKIDVHSQTAWVEAGATLGEVYYWVNEKNENLSLAAGY
CPTVCAGGHFGGGGYGPLMRNYGLAADNIIDAHLVNVHGKVLDRKSMGEDLFWAL
RGGGAESFGIIVAWKIRLVAVPKSTMFSVKKIMEIHELKLVNWKQNIAYKYDKDLL
LMTHFITRNITDNQGKNKTAIHTYFSSVFLGGVDSLVDLMNKSFPGLGIKKTDQRQLS
WIDTHIFYSGVVNYDTDNFNKEILLDRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLY
EEDIGAGMYALYPYGGIMDEISESAIPFPHRAGILYELWYICSWEKQEDNEKHLNWIR
NIYNFMTPYVSKNPRLAYLNYRDLDIGINDPKNPNNYTQARIWGEKYFGKNFDRLV
KVKTLVDPNNFFRNEQSIPPLPRHRH (SEQ ID NO: 72).

[306] In some embodiments, a CBDAS is a CBDVAS and comprises SEQ ID NO: 98:

MNPRENFLKCFSQYIPNNATNLKLVYTQNNPLYMSVLNSTIHNLRFTSDTTPKPLVIV
TPSHVSHIQATILCSKKVGLQIRTRSGGHDSEGM SYISQVPFVIVDLRNMRSIKIDVHS
QTAWVQAGATLGEVYYWINEKNENLSLAAGYCPTVGAGGHFGGGGYGPLMRKYG

LAADNIIDAHLVDVHGKVLDRKSMGEDLFWALRGGGAESFGHIVAWKIRLVAVPKS
 TMFSVKKIMEIHEL VKLVHKWQNIAYKYDKDLLMTHFITRNITDNQGKNKTAIHTY
 FSSVFLGGVDSLVDLMNKSFPGLGIKKEDCRELSWIDTIIFYSGVVNYDITDNFNKEILL
 DRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLYEEDIGAGMYVLVPYGGIMDEISES
 AIPFPHRAGILYELWYICSWEKQEDNEKHLNWIRNIYNFMTPYVSKNPRLAYLNYRD
 LDIGINDPKNPNNYTQARIWGEKYFGKNFDRLVKVKTLVDPNNFFRNEQSIPPLPRHR
 H.

[307] In some embodiments, SEQ ID NO: 98 is encoded by a nucleic acid comprising:
 atgaatccaagagaaaatttctgaagtgtttctcagtatatcccgaaataatgcgacgaaccttaagttagatatactcagaacaacct
 ctatatatgagcgttctaaattctacaatccacaacctgaattacgtccgacacgactccgaaaccttagttatagtgacaccgtcac
 atgttagccatatacagggcaccatactatgttccaaaaagtgggttacaaatacgtacccgtagcgggggacacgacagtgggg
 gatgagttatattagtcaggtgccttcgtcatagtggttaagaaatatgaggtcaatlaaatcgacgttcactcacaactgcctggg
 ttcaagcgggggccaattgggtgaagtattactggattaatgagaagaacgagaatcttcactagcagccgttatgtccacag
 tcggagccggcggtcacttggcgcgcggtacgggtccctaatgagaaaatacgggcttgcgcagacaatatcatagatgtca
 cttagttgatgttcatggaaaagtgttagaccgtaaaagcatgggggaggatctgtttggcgcttagagggggaggggcagaatcat
 ttggaataatagtggtcatggaaaatcaggttgggtgttccaaagagtaccatgttctcagtaaaagaaataatggagatccatgagc
 tagttaaacttgtgcataaatggcaaacatagccatataatgataaggacttgcgttatgactattcataaccagaaacaftacg
 gataaccaagggaagaacaaaacagccatccatactacttagctccgtttcttgggtggtgtagacagcttagttgacctgatgaac
 aagagtttccggaactaggtatcaagaaggaagattgtagagaactttcctggattgataccataatctttacagcggagtcgtcaatta
 tgacactgacaactcaacaaggaaattttattagataggagtcggggtcaaaatggggccttcaagatcaaactagactacgttaaaaa
 acccattcctgaaagtgttttgtcagattctggagaagctgtatgaagaagatattggcgcggtatgtacgtactttatccgtacggcg
 gcataatggatgagattagtgaaagcgccatcccttccccacagagctggtatcctgtacgagttgtggtatatctgctcctgggaga
 aacaggaggataacgaaaagcacttaaatggattaggaatatctacaatticatgacgccctacgtttccaagaacccaggttggcct
 atttgaactacagggatcttgatattggaatcaacgaccccaaaacccaaacaactacaccaggaaggattggggagagaagta
 ctccgggaagaacttcgacaggctagttaagggtgaaaacgctagttgatccaaataatttttcagaaacgaacagagtatccctccctta
 ccgctcataggcac (SEQ ID NO: 99).

[308] In some embodiments, a CBDAS is a CBDVAS and comprises SEQ ID NO:
 100:

MNPRENFLKCFSSQYIPNNAANPKLVYTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIV
 TPSHVSHIQATILCSKKVGLQIRTRSGGHDSEGMSYISQVPFVIVDLRNMRSIKIDVHS
 QTAWVQAGATLGEVYYWINEKNENLSLAAGYCPTVCAGGHFGGGGYGPLMRKYG
 LAADNIIDAHLVDVHGKVLDRKSMGEDLFWALRGGGAESFGHIVAWKIRLVAVPSK
 STMFSVKKIMEIHEL VKLVNWKQNIAYKYDKDLLMTHFITRNITDNQGKNKTAIHT

YFSSVFLGGVDSLVDLMNKSFPELGIKKTDCRELSWIDTIIFYSGVVNYDTDNFNKEI
LLDRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLYEEDIGAGMYALYPYGGIMDEIS
ESAIPFPHRAGILYELWYICSWEKQEDNEKHLNWIRNIYNFMTPYVSKNPRLAYLNY
RDLDIGINDPKNPNNYTQARIWGEKYFGKNFDRLVKVKTLVDPNNFFRNEQSIPPLPR
HRH.

[309] In some embodiments, SEQ ID NO: 100 is encoded by a nucleic acid comprising:

atgaacccgagagaaaatttccttaagtgttttcgcaatatattcctaataacgctgcaaatcccaaattggtttactcagcatgaccaa
ctgtatatgagcgtattaaactcaacaatacaaaatctacgctttacctcgatacgactccaaaacctttagtgatcgtcacccaagtca
cgtttctcatattcaggcgacaatcctctgctctaagaaagtaggcttgcaataaggacaagaagtgaggggcacgattcagagggtta
tgctttacatttcccaagttccattcgctcatagtggtatcgtaatatgcatgattaaaattgacgttcatagcctaaactgcttgggttca
ggccgggtgcaacgttgggagaagtgtattactggatcaacgaaaagaatgagaacctctcattggctgccgggttatgtccgacctgat
gcgcaggcggtcatttggaggaggcggttacgggtccttaagtgaagagtaggggttagcagctgataataatcgacgctcaccttg
tcgatgtccatggtaaagtctggatcgtaagicaatgggcgaagacttggtctgggcccttaggggtggcgagctgaaagtgttggtta
ttatcgtggcctggaaaattagattagtgggcggttccatccaagtctactatgttctctgtaaaaaaattatggaaatacacgagctggta
agcttgaataaaatggcaaacattgcgtataataacgataaggatctactgcttatgacacatttataacacggaatattactgacaatc
agggaaaaaataaaacgcgaatccatacgtattttcatcagtggtcttggagggtgtagattccctagttgacttgatgaacaagtcgttc
cccgaattaggtatcaaaaagacggattgtcgggagctgagttggatagatactattttttactctggcgttgcactatgatacaga
caatttaacaaagaataactcttagatagaagtgcgtgggcaaacggggcatttaagattaaacttgactacgtcaaaaagccaatccc
tgaatccgtgtttgtacaaattttagagaaattgtatgaagaggacatcggtgcaggtatgtatgcgctatacccatatggtggcatcatgg
atgaaattagcgaatccgctattccgttcccccatcgtccggcactactctacgaactatggatatatgttcttgggaaaaacaagaaga
caatgagaagcatttgaattggattaggaacatttacaacttcatgacaccttatgtgtctaaaaatccaagacttgcattttaaactaccg
ggatttagatatcgggattaatgatccaaagaacccaaataactatacacaagccaggatatggggagaaaaaacttttgtaagaattt
tgatagattggttaaggtaaagactcttggatgcccaataatttttcagaaatgagcagagcatacctccattacctagacaccgacac
(SEQ ID NO: 101).

[310] In some embodiments, a CBDAS comprises a protein or nucleic acid sequence that is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or is 100% identical, including all values in between, to SEQ ID NO: 98-

101. In some embodiments, a CBDAS comprises a sequence that is a conservatively substituted version of SEQ ID NO: 98 or 100.

[311] Additional non-limiting examples of CBDAS enzymes may also be found in U.S. Patent No. 9,512,391, U.S. Patent Application Publication No. 2018/0179564 and PCT Publication No. WO 2022/011175, which are incorporated by reference in this application in their entireties.

Cannabichromenic acid synthase (CBCAS)

[312] A host cell described in this application may comprise a TS that is a cannabichromenic acid synthase (CBCAS). As used in this application, a “CBCAS” refers to an enzyme that is capable of catalyzing oxidative cyclization of a prenyl moiety (*e.g.*, terpene) of a compound of Formula (8) to produce a compound of Formula (11). In some embodiments, a compound of Formula (11) is a compound of Formula (11a) (cannabichromenic acid (CBCA)), CBCVA, or a compound of Formula (8) with R as a C7 alkyl (heptyl) group. A CBCAS may use cannabigerolic acid (CBGA) as a substrate. In some embodiments, a CBCAS produces cannabichromenic acid (CBCA) from cannabigerolic acid (CBGA). In some embodiments, the CBCAS may catalyze the oxidative cyclization of other substrates, such as 3-geranyl-2,4-dihydro-6-alkylbenzoic acids like cannabigerovarinic acid (CBGVA), or a substrate of Formula (8) with R as a C7 alkyl (heptyl) group. In some embodiments, the CBCAS exhibits specificity for CBGA substrates.

[313] In some embodiments, a CBCAS is from *Cannabis*. In *C. sativa*, an amino acid sequence encoding CBCAS is provided by, and incorporated by reference from, SEQ ID NO:2 disclosed in U.S. Patent Publication No. 2017/0211049. In other embodiments, a CBCAS may be a THCAS described in and incorporated by reference from U.S. Patent No. 9,359,625. SEQ ID NO:2 disclosed in U.S. Patent Application Publication No. 2017/0211049 (corresponding to SEQ ID NO: 73 in this application) has the amino acid sequence:

MNCSTFSFWFVCKIIFFLSFNIQISIANPQENFLKCFSEYIPNNPANPKFIYTQHDQLY
MSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQASILCSKKVGLQIRTRSGGHDAEGL
SYISQVPFAIVDLRNMHTVKVDIHSQTAWVEAGATLGEVYYWINEMNENFSFPGGY
CPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMGEDLFWAI
RGGGGENFGIIAACKIKLVVVPSKATIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDL

MLTTHFRTRNITDNHGKNKTTVHGYFSSIFLGGVDSLVDLMNKSFPGLGIKKTDCKE
LSWIDTTIFYSGVVNYNTANFKKEILLDRSAGKKTAFSIKLDYVKKLIPETAMVKILE
KLYEEVGVGMVLYPYGGIMDEISESAIPFPHRAGIMYELWYTATWEKQEDNEKHI
NWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARIWGEKYFGKNF
NRLVKVKTADPNNFFRNEQSIPPLPPRHH (SEQ ID NO: 73).

[314] Additional CBCASs are disclosed in and incorporated by reference from PCT Publication No. WO2021/195520, PCT Publication No. WO 2022/011175 and PCT Publication No. WO2023/056350.

[315] In some embodiments a CBCAS is from *Aspergillus vadensis*, corresponding to UniProt Accession No. A0A319B6X5.

[316] In some embodiments, a CBCAS comprises the sequence of SEQ ID NO: 84:

MGNTTSIAGRDCLVSALGGNSALAAFPNQLLWTADVHEYNLNLPVTPAAITYPETAE
QIAGIVKCASDYDYKVQARSGGHSFGNYGLGGTDGAVVVDMMKHFNQFSMNDQTYE
AVIGPGTTLNDVDIELYNNGKRAMAHGVCPTIKTGGHFTIGGLGPTARQWGLALDH
VEEVEVVLANSIVRASNTQNQDVFFAVKGAAADFGIVTEFKVRTEPAPGLAVQYSY
TFNLGSTAEKAQFVKDWQSFISAKNLTRQFYNNMVIFDGDIILEGFFGSKEQYDAL
GLEDHFAPKNPGNILLVLTDWLGMVGHAELEDITLKLVGNTPTWIFYAKSLGFRQDTLIP
SAGIDEFFEYIANHTAGTPAWFVTLSEGGAINDAEDATAYAHRDVLFWVQLFMV
NPLGPISSETTYEFTDGLYDVLARAVPESVGHAYLGCPDPRMENAPQKYWRTNLPRL
QELKEELDPKNTFHHPQGVIPA (SEQ ID NO: 84).

[317] A non-limiting example of a nucleotide sequence encoding SEQ ID NO: 84 is:

atggglaacacaactccatcgagcagagattgcttagtctcagcccttgaggtaattctgcttggctgcttcccaaaccaattgct
gtggaccgccgacgttcacgagtataattgaacctacctgtaacgccagctgccataacctacccgaaactgctgaacagattgctg
gtatcgftaagtgtgctagtgtacgactataaagtgaagctaggtctggtggtcattccttggtaattacgggttgggaggtactgatg
gtgccgttgctcgatgaagcacttcaaccaattctcgatgaacgatcaaacctacgaagcagttatggtccaggtactaccttaaa
cgacgttgacattgaattgtacaacaatggcaagagagctatggctcatggtgttgcctcaactatcaaaacaggtggtcactttacaatt
ggcggtctgggtcctactgccagacaatggggttggctttagaiaacgtcgaagaagtgaagtagcttggccaactcttctatcgctc
gtgctagcaataacccaaaaccaggatgtcttcttgcgtcaagggcgagctgccgactcggtatcggtacggaggtcaaggttagaa
ctgagccagcacctggttagctgttcaatattcgatataccttaacttggtagtactgctgaaaaagcccaattgtcaaggattggcaaa
gcttcattccgctaaaaactgactcgtcaattctacaacaatatgggtatattgacgggtgacattattttagaaggttgttttcggatcaa

aggaacaatacgaatgccttgggttgaagatcatcttgcctcaagaatccaggaatacctagtgctgacggactgggtgggaatgg
taggtcatgcttgggaagacaccatttgaagctagttggaacacacccacttgggtctacgctaatacttgggttcagacaagatacc
ctaatacccatctgctgggtattgacgaattttcgaatatatagcaaacacaccgctgggtactccagcttgggtcgttaccttctctggaa
ggcggcgctataaacgatgtggctgaagatgccacagcatacgcacacagagatgtcctatttgggttcagttggtcatggtcaatcca
ctaggtccaatctcagaaactacctaagagttcactgacgggttataatgacgtcttagcaagagctgtccctgaatctgttggtcatgccta
tttgggtgtccagaccaagaatggaaaacgctccacaaaagtactggcgtactaatttgcctagattacaagaattgaaagaggaatt
ggatccaaagaacaccttccaccatccacaaggtgtgattccagct (SEQ ID NO: 85).

[318] In some embodiments, a CBCAS comprises the sequence of SEQ ID NO: 161:

MNPQENFLKCFSSQYIPNNVANPKLVYTQHDQLYMSILNSTIQNLRFTSDTTPKPLVIV
TPSNVSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQVPFVIVDLRNMHSIKIDVES
QTAWVEAGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRNYG
LAADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAAWKIKLVAVPSK
VTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLLLFTHFITRNITDNHGKKNKTTVHG
YFSSIFHGGVDSLVDLMNKSFPGLGIKKTDCKELSWIDTTFYSGVVNYNTANFKKEI
LLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGVGMYVLYPYGGIMDEI
SESAIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYL
NYRDLDLGKNNPESPNNYTQARIWGEKYFGKNFNRLVKVKTKVDPNNFFRNEQSIP
PLPPHHH.

[319] A non-limiting example of a nucleotide sequence encoding SEQ ID NO: 161 is:

atgaatcccaagagaacttcttaagtgttcagtcagtatatcccaataacgtggcgaatccgaaattagtalacacgaacacgatcagct
atataatgagcattctgaatagtaccattcaaaacttgcgcttacttcggacacaactccaaagcctttagtcatagttactccttcaaatgttccc
atattcaggctacaatcctctgctctaaaaagttgggttgcaatacgtactagatcaggaggccacgatgccgaaggtatgtcttatactcc
caagtgccattcgttattgtcacttgaggaacatgcattctattaaagttagttagaatccagaccgatgggtgaagctggtgccacatt
aggtgaagtatactattggataaacgagaaaaatgaaattatcgtttctggaggctattgtccaactgttgggtgcgggtgctatttctcagg
cgggggctacggagcattgatgcgaaactacggactagcagctgataacattatagacgcccatctcgtgaatgtggatggtaaagtcttga
tagaaagagcatgggtgaagattgttttggcgatcagaggcgggtggaggggaaaacttgggtatttgcgggttgaaaaataaagctggt
cgcagttccctcgaaagtaacaatctttctgtaagaaaaatatggaatacatgggttggtcaagttattcaataatggcaaacattgcttat
aagtatgataaagacctctactgttactcattttattacgaggaataattaccgacaatcacggtaagaataaaactacagtgcacgggtactttc
ttcaattttcatggtggcgttgactctctagtagatttaataacaaaagtgtccctgagttaggatcaagaaaacggattgtaagaacttctt
ggatcgacacaaccataattctattcgggtgttgaactacaataccgcaacttfaaaaggaaatttggtagatcgctctgctggaaagaaaa
cagcatttagtatcaacttgactatgtaagaaaccgatacccgagactgcatggttaagatacttgagaagctatacgaagaggatgtgg

gagtaggcatgtacgttctacccttacggcggtataatggatgaaatttcagaatccgcaatccattcccacacagagcaggtattatgtat
 gaactgtgggtatactgcctcatgggagaaacaagaagataacgaaaaacacattaattgggtccgtagcgtctacaatttcactacaccatag
 tatcccaaaacccgagattggcttactgaattatagagacttggatctaggaaaaataatccgaaagcccgaataattatacacaggcaag
 gatttggggcgagaagtatttcggcaagaactcaacagattagtaaaagttaagaccaaagttgacccaataacttttttagaaatgaacaa
 agtatccctccactcccaccacatcatcat (SEQ ID NO: 160).

[320] In some embodiments, a CBCAS comprises a protein or nucleic acid sequence that is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or is 100% identical, including all values in between, to SEQ ID NO: 84 or 85, or 160-161. In some embodiments, a CBCAS comprises a sequence that is a conservatively substituted version of SEQ ID NO: 84 or 161.

Variants

[321] Aspects of the disclosure relate to nucleic acids encoding any of the polypeptides (*e.g.*, AAE, PKS, PKC, PT, or TS) described in this application. In some embodiments, a nucleic acid encompassed by the disclosure is a nucleic acid that hybridizes under high or medium stringency conditions to a nucleic acid encoding an AAE, PKS, PKC, PT, or TS and is biologically active. For example, high stringency conditions of 0.2 to 1 x SSC at 65 °C followed by a wash at 0.2 x SSC at 65 °C can be used. In some embodiments, a nucleic acid encompassed by the disclosure is a nucleic acid that hybridizes under low stringency conditions to a nucleic acid encoding an AAE, PKS, PKC, PT, or TS and is biologically active. For example, low stringency conditions of 6 x SSC at room temperature followed by a wash at 2 x SSC at room temperature can be used. Other hybridization conditions include 3 x SSC at 40 or 50 °C, followed by a wash in 1 or 2 x SSC at 20, 30, 40, 50, 60, or 65 °C.

[322] Hybridizations can be conducted in the presence of formaldehyde, *e.g.*, 10%, 20%, 30% 40% or 50%, which further increases the stringency of hybridization. Theory and practice of nucleic acid hybridization is described, *e.g.*, in S. Agrawal (ed.) Methods in

Molecular Biology, volume 20; and Tijssen (1993) Laboratory Techniques in biochemistry and molecular biology-hybridization with nucleic acid probes, *e.g.*, part I chapter 2 “Overview of principles of hybridization and the strategy of nucleic acid probe assays,” Elsevier, New York provide a basic guide to nucleic acid hybridization.

[323] Variants of enzyme sequences described in this application (*e.g.*, AAE, PKS, PKC, PT, or TS, including nucleic acid or amino acid sequences) are also encompassed by the present disclosure. A variant may share at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with a reference sequence, including all values in between.

[324] Unless otherwise noted, the term “sequence identity,” which is used interchangeably in this disclosure with the term “percent identity,” as known in the art, refers to a relationship between the sequences of two polypeptides or polynucleotides, as determined by sequence comparison (alignment). In some embodiments, sequence identity is determined across the entire length of a sequence (*e.g.*, AAE, PKS, PKC, PT, or TS sequence). In some embodiments, sequence identity is determined over a region (*e.g.*, a stretch of amino acids or nucleic acids, *e.g.*, the sequence spanning an active site) of a sequence (*e.g.*, AAE, PKS, PKC, PT, or TS sequence). For example, in some embodiments, sequence identity is determined over a region corresponding to at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or over 100% of the length of the reference sequence.

[325] Identity measures the percent of identical matches between two or more sequences with gap alignments (if any) addressed by a particular mathematical model, algorithm, or computer program.

[326] Identity of related polypeptides or nucleic acid sequences can be readily calculated by any of the methods known to one of ordinary skill in the art. The percent identity of two sequences (*e.g.*, nucleic acid or amino acid sequences) may, for example, be determined

using the algorithm of Karlin and Altschul *Proc. Natl. Acad. Sci. USA* 87:2264-68, 1990, modified as in Karlin and Altschul *Proc. Natl. Acad. Sci. USA* 90:5873-77, 1993. Such an algorithm is incorporated into the NBLAST[®] and XBLAST[®] programs (version 2.0) of Altschul et al., *J. Mol. Biol.* 215:403-10, 1990. BLAST[®] protein searches can be performed, for example, with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to the proteins described in this application. Where gaps exist between two sequences, Gapped BLAST[®] can be utilized, for example, as described in Altschul *et al.*, *Nucleic Acids Res.* 25(17):3389-3402, 1997. When utilizing BLAST[®] and Gapped BLAST[®] programs, the default parameters of the respective programs (*e.g.*, XBLAST[®] and NBLAST[®]) can be used, or the parameters can be adjusted appropriately as would be understood by one of ordinary skill in the art.

[327] Another local alignment technique which may be used, for example, is based on the Smith-Waterman algorithm (Smith, T.F. & Waterman, M.S. (1981) "Identification of common molecular subsequences." *J. Mol. Biol.* 147:195-197). A general global alignment technique which may be used, for example, is the Needleman–Wunsch algorithm (Needleman, S.B. & Wunsch, C.D. (1970) "A general method applicable to the search for similarities in the amino acid sequences of two proteins." *J. Mol. Biol.* 48:443-453), which is based on dynamic programming.

[328] More recently, a Fast Optimal Global Sequence Alignment Algorithm (FOGSAA) was developed that purportedly produces global alignment of nucleic acid and amino acid sequences faster than other optimal global alignment methods, including the Needleman–Wunsch algorithm. In some embodiments, the identity of two polypeptides is determined by aligning the two amino acid sequences, calculating the number of identical amino acids, and dividing by the length of one of the amino acid sequences. In some embodiments, the identity of two nucleic acids is determined by aligning the two nucleotide sequences and calculating the number of identical nucleotide and dividing by the length of one of the nucleic acids.

[329] For multiple sequence alignments, computer programs including Clustal Omega (Sievers *et al.*, *Mol Syst Biol.* 2011 Oct 11;7:539) may be used.

[330] In preferred embodiments, a sequence, including a nucleic acid or amino acid sequence, is found to have a specified percent identity to a reference sequence, such as a

sequence disclosed in this application and/or recited in the claims when sequence identity is determined using the algorithm of Karlin and Altschul *Proc. Natl. Acad. Sci. USA* 87:2264-68, 1990, modified as in Karlin and Altschul *Proc. Natl. Acad. Sci. USA* 90:5873-77, 1993 (e.g., BLAST®, NBLAST®, XBLAST® or Gapped BLAST® programs, using default parameters of the respective programs).

[331] In some embodiments, a sequence, including a nucleic acid or amino acid sequence, is found to have a specified percent identity to a reference sequence, such as a sequence disclosed in this application and/or recited in the claims when sequence identity is determined using the Smith-Waterman algorithm (Smith, T.F. & Waterman, M.S. (1981) "Identification of common molecular subsequences." *J. Mol. Biol.* 147:195-197) or the Needleman-Wunsch algorithm (Needleman, S.B. & Wunsch, C.D. (1970) "A general method applicable to the search for similarities in the amino acid sequences of two proteins." *J. Mol. Biol.* 48:443-453) using default parameters.

[332] In some embodiments, a sequence, including a nucleic acid or amino acid sequence, is found to have a specified percent identity to a reference sequence, such as a sequence disclosed in this application and/or recited in the claims when sequence identity is determined using a Fast Optimal Global Sequence Alignment Algorithm (FOGSAA) using default parameters.

[333] In some embodiments, a sequence, including a nucleic acid or amino acid sequence, is found to have a specified percent identity to a reference sequence, such as a sequence disclosed in this application and/or recited in the claims when sequence identity is determined using Clustal Omega (Sievers et al., *Mol Syst Biol.* 2011 Oct 11;7:539) using default parameters.

[334] As used in this application, a residue (such as a nucleic acid residue or an amino acid residue) in sequence "X" is referred to as corresponding to a position or residue (such as a nucleic acid residue or an amino acid residue) "Z" in a different sequence "Y" when the residue in sequence "X" is at the counterpart position of "Z" in sequence "Y" when sequences X and Y are aligned using amino acid sequence alignment tools known in the art.

[335] As used in this application, variant sequences may be homologous sequences. As used in this application, homologous sequences are sequences (e.g., nucleic acid or amino

acid sequences) that share a certain percent identity (*e.g.*, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% percent identity, including all values in between). Homologous sequences include but are not limited to paralogous or orthologous sequences. Paralogous sequences arise from duplication of a gene within a genome of a species, while orthologous sequences diverge after a speciation event.

[336] In some embodiments, a polypeptide variant (*e.g.*, AAE, PKS, PKC, PT, or TS enzyme variant) comprises a domain that shares a secondary structure (*e.g.*, alpha helix, beta sheet) with a reference polypeptide (*e.g.*, a reference AAE, PKS, PKC, PT, or TS enzyme). In some embodiments, a polypeptide variant (*e.g.*, AAE, PKS, PKC, PT, or TS enzyme variant) shares a tertiary structure with a reference polypeptide (*e.g.*, a reference AAE, PKS, PKC, PT, or TS enzyme). As a non-limiting example, a polypeptide variant (*e.g.*, AAE, PKS, PKC, PT, or TS enzyme) may have low primary sequence identity (*e.g.*, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, or less than 5% sequence identity) compared to a reference polypeptide, but share one or more secondary structures (*e.g.*, including but not limited to loops, alpha helices, or beta sheets), or have the same tertiary structure as a reference polypeptide. For example, a loop may be located between a beta sheet and an alpha helix, between two alpha helices, or between two beta sheets. Homology modeling may be used to compare two or more tertiary structures.

[337] Functional variants of the recombinant AAE, PKS, PKC, PT, or TS enzyme disclosed in this application are encompassed by the present disclosure. For example, functional variants may bind one or more of the same substrates or produce one or more of the same products. Functional variants may be identified using any method known in the art. For example, the algorithm of Karlin and Altschul *Proc. Natl. Acad. Sci. USA* 87:2264-68, 1990 described above may be used to identify homologous proteins with known functions.

[338] Putative functional variants may also be identified by searching for polypeptides with functionally annotated domains. Databases including Pfam (Sonnhammer *et al.*, *Proteins*. 1997 Jul;28(3):405-20) may be used to identify polypeptides with a particular domain.

[339] Homology modeling may also be used to identify amino acid residues that are amenable to mutation (e.g., substitution, deletion, and/or insertion) without affecting function. A non-limiting example of such a method may include use of position-specific scoring matrix (PSSM) and an energy minimization protocol.

[340] Position-specific scoring matrix (PSSM) uses a position weight matrix to identify consensus sequences (e.g., motifs). PSSM can be conducted on nucleic acid or amino acid sequences. Sequences are aligned and the method takes into account the observed frequency of a particular residue (e.g., an amino acid or a nucleotide) at a particular position and the number of sequences analyzed. See, e.g., Stormo *et al.*, *Nucleic Acids Res.* 1982 May 11;10(9):2997-3011. The likelihood of observing a particular residue at a given position can be calculated. Without being bound by a particular theory, positions in sequences with high variability may be amenable to mutation (e.g., substitution, deletion, and/or insertion; e.g., PSSM score ≥ 0) to produce functional homologs.

[341] PSSM may be paired with calculation of a Rosetta energy function, which determines the difference between the wild-type and the single-point mutant. The Rosetta energy function calculates this difference as ($\Delta\Delta G_{calc}$). With the Rosetta function, the bonding interactions between a mutated residue and the surrounding atoms are used to determine whether a mutation increases or decreases protein stability. For example, a mutation that is designated as favorable by the PSSM score (e.g. PSSM score ≥ 0), can then be analyzed using the Rosetta energy function to determine the potential impact of the mutation on protein stability. Without being bound by a particular theory, potentially stabilizing amino acid mutations are desirable for protein engineering (e.g., production of functional homologs). In some embodiments, a potentially stabilizing amino acid mutation has a $\Delta\Delta G_{calc}$ value of less than -0.1 (e.g., less than -0.2, less than -0.3, less than -0.35, less than -0.4, less than -0.45, less than -0.5, less than -0.55, less than -0.6, less than -0.65, less than -0.7, less than -0.75, less than -0.8, less than -0.85, less than -0.9, less than -0.95, or less than -1.0) Rosetta energy units (R.e.u.). See, e.g., Goldenzweig *et al.*, *Mol Cell*. 2016 Jul 21;63(2):337-346. Doi: 10.1016/j.molcel.2016.06.012.

[342] In some embodiments, a coding sequence comprises an amino acid mutation at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more than 100 positions relative to a reference coding sequence. In some embodiments, the coding sequence comprises an amino acid mutation in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more codons of the coding sequence relative to a reference coding sequence. As will be understood by one of ordinary skill in the art, a mutation within a codon may or may not change the amino acid that is encoded by the codon due to degeneracy of the genetic code. In some embodiments, the one or more substitutions, insertions, or deletions in the coding sequence do not alter the amino acid sequence of the coding sequence relative to the amino acid sequence of a reference polypeptide.

[343] In some embodiments, the one or more mutations in a coding sequence do alter the amino acid sequence of the corresponding polypeptide relative to the amino acid sequence of a reference polypeptide. In some embodiments, the one or more mutations alters the amino acid sequence of the polypeptide relative to the amino acid sequence of a reference polypeptide and alter (enhance or reduce) an activity of the polypeptide relative to the reference polypeptide.

[344] The activity (*e.g.*, specific activity) of any of the recombinant polypeptides described in this application (*e.g.*, AAE, PKS, PKC, PT, or TS) may be measured using routine methods. As a non-limiting example, a recombinant polypeptide's activity may be determined by measuring its substrate specificity, product(s) produced, the concentration of product(s) produced, or any combination thereof. As used in this application, "specific activity" of a recombinant polypeptide refers to the amount (*e.g.*, concentration) of a particular product produced for a given amount (*e.g.*, concentration) of the recombinant polypeptide per unit time.

[345] Mutations in a recombinant polypeptide coding sequence may result in conservative amino acid substitutions to provide functionally equivalent variants of the recombinant polypeptides, *e.g.*, variants that retain the activities of the polypeptides.

[346] In some instances, an amino acid is characterized by its R group (see, *e.g.*, Table 2). For example, an amino acid may comprise a nonpolar aliphatic R group, a positively charged R group, a negatively charged R group, a nonpolar aromatic R group, or a polar uncharged R group. Non-limiting examples of an amino acid comprising a nonpolar aliphatic R group include alanine, glycine, valine, leucine, methionine, and isoleucine. Non-limiting examples of an amino acid comprising a positively charged R group includes lysine, arginine, and histidine. Non-limiting examples of an amino acid comprising a negatively charged R group include aspartate and glutamate. Non-limiting examples of an amino acid comprising a nonpolar, aromatic R group include phenylalanine, tyrosine, and tryptophan. Non-limiting examples of an amino acid comprising a polar uncharged R group include serine, threonine, cysteine, proline, asparagine, and glutamine.

[347] Non-limiting examples of functionally equivalent variants of polypeptides may include conservative amino acid substitutions in the amino acid sequences of proteins disclosed in this application. As used in this application, a “conservative amino acid substitution” or “conservative substitution,” which are used interchangeably, refer to an amino acid substitution that does not alter the relative charge or size characteristics or functional activity of the protein in which the amino acid substitution is made. Non-limiting examples of conservative amino acid substitutions are provided in **Table 2**.

[348] In some embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more than 20 residues can be changed when preparing variant polypeptides. In some embodiments, amino acids are replaced by conservative amino acid substitutions.

Table 2. Conservative Amino Acid Substitutions

Original Residue*	R Group Type	Conservative Amino Acid Substitutions
Ala (A)	nonpolar aliphatic R group	Cys, Gly, Ser
Arg (R)	positively charged R group	His, Lys
Asn (N)	polar uncharged R group	Asp, Gln, Glu
Asp (D)	negatively charged R group	Asn, Gln, Glu
Cys (C)	polar uncharged R group	Ala, Ser

Gln (Q)	polar uncharged R group	Asn, Asp, Glu
Glu (E)	negatively charged R group	Asn, Asp, Gln
Gly (G)	nonpolar aliphatic R group	Ala, Ser
His (H)	positively charged R group	Arg, Tyr, Trp
Ile (I)	nonpolar aliphatic R group	Leu, Met, Val
Leu (L)	nonpolar aliphatic R group	Ile, Met, Val
Lys (K)	positively charged R group	Arg, His
Met (M)	nonpolar aliphatic R group	Ile, Leu, Phe, Val
Pro (P)	polar uncharged R group	
Phe (F)	nonpolar aromatic R group	Met, Trp, Tyr
Ser (S)	polar uncharged R group	Ala, Gly, Thr
Thr (T)	polar uncharged R group	Ala, Asn, Ser
Trp (W)	nonpolar aromatic R group	His, Phe, Tyr, Met
Tyr (Y)	nonpolar aromatic R group	His, Phe, Trp
Val (V)	nonpolar aliphatic R group	Ile, Leu, Met, Thr

*in parentheses are the corresponding single letter designations for amino acids; “R” when used in reference to an amino acid residue means an arginine. When used in the term “R group” in Table 2 and throughout the application, “R” refers to an amino acid side chain or an atom or group of atoms bonded to a central atom of a small molecule.

[349] Amino acid substitutions in the amino acid sequence of a polypeptide to produce a recombinant polypeptide (*e.g.*, AAE, PKS, PKC, PT, or TS) variant having a desired property and/or activity can be made by alteration of the coding sequence of the polypeptide (*e.g.*, AAE, PKS, PKC, PT, or TS). Similarly, conservative amino acid substitutions in the amino acid sequence of a polypeptide to produce functionally equivalent variants of the polypeptide typically are made by alteration of the coding sequence of the recombinant polypeptide (*e.g.*, AAE, PKS, PKC, PT, or TS).

[350] Mutations (*e.g.*, substitutions, insertions, additions, or deletions) can be made in a nucleic acid sequence by a variety of methods known to one of ordinary skill in the art. For example, mutations (*e.g.*, substitutions, insertions, additions, or deletions) can be made by PCR-directed mutation, site-directed mutagenesis, such as according to the method of Kunkel (Kunkel, *Proc. Nat. Acad. Sci. U.S.A.* 82: 488-492, 1985), by chemical synthesis of a gene or polypeptide, by gene editing techniques such as CRISPR, or by insertions, such as insertion of a tag (*e.g.*, a HIS tag or a GFP tag). Mutations can include, for example, substitutions, insertions, additions, deletions, and translocations, generated by any method known in the art. Methods for producing mutations may be found in references such as *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Fourth Edition, Cold Spring Harbor Laboratory

Press, Cold Spring Harbor, New York, 2012, or Current Protocols in Molecular Biology, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York, 2010.

[351] In some embodiments, methods for producing variants include circular permutation (Yu and Lutz, *Trends Biotechnol.* 2011 Jan;29(1):18-25). In circular permutation, the linear primary sequence of a polypeptide can be circularized (*e.g.*, by joining the N-terminal and C-terminal ends of the sequence) and the polypeptide can be severed (“broken”) at a different location. Thus, the linear primary sequence of the new polypeptide may have low sequence identity (*e.g.*, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, less or less than 5%, including all values in between) as determined by linear sequence alignment methods (*e.g.*, Clustal Omega or BLAST). Topological analysis of the two proteins, however, may reveal that the tertiary structure of the two polypeptides is similar or dissimilar. Without being bound by a particular theory, a variant polypeptide created through circular permutation of a reference polypeptide and with a similar tertiary structure as the reference polypeptide can share similar functional characteristics (*e.g.*, enzymatic activity, enzyme kinetics, substrate specificity or product specificity). In some instances, circular permutation may alter the secondary structure, tertiary structure or quaternary structure and produce an enzyme with different functional characteristics (*e.g.*, increased or decreased enzymatic activity, different substrate specificity, or different product specificity). See, *e.g.*, Yu and Lutz, *Trends Biotechnol.* 2011 Jan;29(1):18-25.

[352] It should be appreciated that in a protein that has undergone circular permutation, the linear amino acid sequence of the protein would differ from a reference protein that has not undergone circular permutation. However, one of ordinary skill in the art would be able to determine which residues in the protein that has undergone circular permutation correspond to residues in the reference protein that has not undergone circular permutation by, for example, aligning the sequences and detecting conserved motifs, and/or by comparing the structures or predicted structures of the proteins, *e.g.*, by homology modeling.

[353] In some embodiments, an algorithm that determines the percent identity between a sequence of interest and a reference sequence described in this application accounts for the presence of circular permutation between the sequences. The presence of circular

permutation may be detected using any method known in the art, including, for example, RASPODOM (Weiner et al., Bioinformatics. 2005 Apr 1;21(7):932-7). In some embodiments, the presence of circulation permutation is corrected for (e.g., the domains in at least one sequence are rearranged) prior to calculation of the percent identity between a sequence of interest and a sequence described in this application. The claims of this application should be understood to encompass sequences for which percent identity to a reference sequence is calculated after taking into account potential circular permutation of the sequence.

Expression of Nucleic Acids in Host Cells

[354] Aspects of the present disclosure relate to recombinant enzymes, functional modifications and variants thereof, as well as their uses. For example, the methods described in this application may be used to produce cannabinoids and/or cannabinoid precursors. The methods may comprise using a host cell comprising an enzyme disclosed in this application, cell lysate, isolated enzymes, or any combination thereof. Methods comprising recombinant expression of genes encoding an enzyme disclosed in this application in a host cell are encompassed by the present disclosure. *In vitro* methods comprising reacting one or more cannabinoid precursors or cannabinoids in a reaction mixture with an enzyme disclosed in this application are also encompassed by the present disclosure. In some embodiments, the enzyme is a TS.

[355] A nucleic acid encoding any of the recombinant polypeptides (e.g., AAE, PKS, PKC, PT, or TS enzyme) described in this application may be incorporated into any appropriate vector through any method known in the art. For example, the vector may be an expression vector, including but not limited to a viral vector (e.g., a lentiviral, retroviral, adenoviral, or adeno-associated viral vector), any vector suitable for transient expression, any vector suitable for constitutive expression, or any vector suitable for inducible expression (e.g., a galactose-inducible or doxycycline-inducible vector).

[356] A vector encoding any of the recombinant polypeptides (e.g., AAE, PKS, PKC, PT, or TS enzyme) described in this application may be introduced into a suitable host cell using any method known in the art. Non-limiting examples of yeast transformation protocols are described in Gietz *et al.*, Yeast transformation can be conducted by the LiAc/SS Carrier DNA/PEG method. *Methods Mol Biol.* 2006;313:107-20, which is hereby incorporated by reference in its entirety. Host cells may be cultured under any conditions suitable as would be

understood by one of ordinary skill in the art. For example, any media, temperature, and incubation conditions known in the art may be used. For host cells carrying an inducible vector, cells may be cultured with an appropriate inducible agent to promote expression.

[357] In some embodiments, a vector replicates autonomously in the cell. In some embodiments, a vector integrates into a chromosome within a cell. A vector can contain one or more endonuclease restriction sites that are cut by a restriction endonuclease to insert and ligate a nucleic acid containing a gene described in this application to produce a recombinant vector that is able to replicate in a cell. Vectors can be composed of DNA or RNA. Cloning vectors include, but are not limited to: plasmids, fosmids, phagemids, virus genomes and artificial chromosomes. As used in this application, the terms “expression vector” or “expression construct” refer to a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell (*e.g.*, microbe), such as a yeast cell. In some embodiments, the nucleic acid sequence of a gene described in this application is inserted into a cloning vector so that it is operably joined to regulatory sequences and, in some embodiments, expressed as an RNA transcript. In some embodiments, the vector contains one or more markers, such as a selectable marker as described in this application, to identify cells transformed or transfected with the recombinant vector. In some embodiments, a host cell has already been transformed with one or more vectors. In some embodiments, a host cell that has been transformed with one or more vectors is subsequently transformed with one or more vectors. In some embodiments, a host cell is transformed simultaneously with more than one vector. In some embodiments, a cell that has been transformed with a vector or an expression cassette incorporates all or part of the vector or expression cassette into its genome. In some embodiments, the nucleic acid sequence of a gene described in this application is recoded. Recoding may increase production of the gene product by at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100%, including all values in between) relative to a reference sequence that is not recoded.

[358] In some embodiments, introduction of a polynucleotide, such as a polynucleotide encoding a recombinant polypeptide, into a host cell results in genomic integration of the polynucleotide. In some embodiments, a host cell comprises at least 1 copy,

at least 2 copies, at least 3 copies, at least 4 copies, at least 5 copies, at least 6 copies, at least 7 copies, at least 8 copies, at least 9 copies, at least 10 copies, at least 11 copies, at least 12 copies, at least 13 copies, at least 14 copies, at least 15 copies, at least 16 copies, at least 17 copies, at least 18 copies, at least 19 copies, at least 20 copies, at least 21 copies, at least 22 copies, at least 23 copies, at least 24 copies, at least 25 copies, at least 26 copies, at least 27 copies, at least 28 copies, at least 29 copies, at least 30 copies, at least 31 copies, at least 32 copies, at least 33 copies, at least 34 copies, at least 35 copies, at least 36 copies, at least 37 copies, at least 38 copies, at least 39 copies, at least 40 copies, at least 41 copies, at least 42 copies, at least 43 copies, at least 44 copies, at least 45 copies, at least 46 copies, at least 47 copies, at least 48 copies, at least 49 copies, at least 50 copies, at least 60 copies, at least 70 copies, at least 80 copies, at least 90 copies, at least 100 copies, or more, including any values in between, of a polynucleotide sequence, such as a polynucleotide sequence encoding any of the recombinant polypeptides described in this application, in its genome.

[359] In some embodiments, the nucleic acid encoding any of the proteins described in this application is under the control of regulatory sequences (*e.g.*, enhancer sequences). In some embodiments, a nucleic acid is expressed under the control of a promoter. In some embodiments, the promoter can be a native promoter, *e.g.*, the promoter of the gene in its endogenous context. Alternatively, a promoter can be a promoter that is different from the native promoter of the gene, *e.g.*, the promoter is different from the promoter of the gene in its endogenous context.

[360] In some embodiments, the promoter is a eukaryotic promoter. Non-limiting examples of eukaryotic promoters include TDH3, PGK1, PKC1, PDC1, TEF1, TEF2, RPL18B, SSA1, TDH2, PYK1, TPI1, GAL1, GAL10, GAL7, GAL3, GAL2, MET3, MET25, HXT3, HXT7, ACT1, ADH1, ADH2, CUP1-1, ENO2, and SOD1, as would be known to one of ordinary skill in the art (see, *e.g.*, Addgene website: blog.addgene.org/plasmids-101-the-promoter-region). In some embodiments, the promoter is a prokaryotic promoter (*e.g.*, bacteriophage or bacterial promoter). Non-limiting examples of bacteriophage promoters include PlsI_{con}, T3, T7, SP6, and PL. Non-limiting examples of bacterial promoters include P_{bad}, P_{mgrB}, P_{trc2}, P_{lac/ara}, P_{tac}, and P_m.

[361] In some embodiments, the promoter is an inducible promoter. As used in this application, an “inducible promoter” is a promoter controlled by the presence or absence of a

molecule. This may be used, for example, to controllably induce the expression of an enzyme. In some embodiments, an inducible promoter linked to an enzyme may be used to regulate expression of the enzyme(s), for example to reduce cannabinoid production in certain scenarios (*e.g.*, during transport of the genetically modified organism to satisfy regulatory restrictions in certain jurisdictions, or between jurisdictions, where cannabinoids may not be shipped). In some embodiments, an inducible promoter linked to an enzyme may be used to regulate expression of the enzyme(s), for example to reduce cannabinoid production in certain scenarios (*e.g.*, during transport of the genetically modified organism to satisfy regulatory restrictions in certain jurisdictions, or between jurisdictions, where cannabinoids may not be shipped). Non-limiting examples of inducible promoters include chemically regulated promoters and physically regulated promoters. For chemically regulated promoters, the transcriptional activity can be regulated by one or more compounds, such as alcohol, tetracycline, galactose, a steroid, a metal, an amino acid, or other compounds. For physically regulated promoters, transcriptional activity can be regulated by a phenomenon such as light or temperature. Non-limiting examples of tetracycline-regulated promoters include anhydrotetracycline (aTc)-responsive promoters and other tetracycline-responsive promoter systems (*e.g.*, a tetracycline repressor protein (tetR), a tetracycline operator sequence (tetO) and a tetracycline transactivator fusion protein (tTA)). Non-limiting examples of steroid-regulated promoters include promoters based on the rat glucocorticoid receptor, human estrogen receptor, moth ecdysone receptors, and promoters from the steroid/retinoid/thyroid receptor superfamily. Non-limiting examples of metal-regulated promoters include promoters derived from metallothionein (proteins that bind and sequester metal ions) genes. Non-limiting examples of pathogenesis-regulated promoters include promoters induced by salicylic acid, ethylene or benzothiadiazole (BTH). Non-limiting examples of temperature/heat-inducible promoters include heat shock promoters. Non-limiting examples of light-regulated promoters include light responsive promoters from plant cells. In certain embodiments, the inducible promoter is a galactose-inducible promoter. In some embodiments, the inducible promoter is induced by one or more physiological conditions (*e.g.*, pH, temperature, radiation, osmotic pressure, saline gradients, cell surface binding, or concentration of one or more extrinsic or intrinsic inducing agents). Non-limiting examples of an extrinsic inducer or inducing agent include amino acids and amino acid analogs, saccharides and polysaccharides, nucleic acids, protein transcriptional activators and repressors, cytokines, toxins, petroleum-based compounds, metal containing compounds, salts, ions, enzyme substrate analogs, hormones or any combination.

[362] In some embodiments, the promoter is a constitutive promoter. As used in this application, a “constitutive promoter” refers to an unregulated promoter that allows continuous transcription of a gene. Non-limiting examples of a constitutive promoter include TDH3, PGK1, PKC1, PDC1, TEF1, TEF2, RPL18B, SSA1, TDH2, PYK1, TPI1, HXT3, HXT7, ACT1, ADH1, ADH2, ENO2, and SOD1.

[363] Other inducible promoters or constitutive promoters, including synthetic promoters, that may be known to one of ordinary skill in the art are also contemplated.

[364] Regulatory sequences for gene expression may also include a terminator sequence. In some embodiments, a terminator sequence marks the end of a gene in DNA during transcription. The choice and design of one or more appropriate vectors suitable for inducing expression of one or more genes described in this application in a host cell is within the ability and discretion of one of ordinary skill in the art.

[365] Expression vectors containing the necessary elements for expression are commercially available and known to one of ordinary skill in the art (see, *e.g.*, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Fourth Edition, Cold Spring Harbor Laboratory Press, 2012).

Host cells

[366] The disclosed cannabinoid biosynthetic methods and host cells are exemplified with *S. cerevisiae*, but are also applicable to other host cells, as would be understood by one of ordinary skill in the art.

[367] Suitable host cells include, but are not limited to: yeast cells, bacterial cells, algal cells, plant cells, fungal cells, insect cells, and animal cells, including mammalian cells. In one illustrative embodiment, suitable host cells include *E. coli* (*e.g.*, Shuffle™ competent *E. coli* available from New England BioLabs in Ipswich, Mass.).

[368] Other suitable host cells of the present disclosure include microorganisms of the genus *Corynebacterium*. In some embodiments, preferred *Corynebacterium* strains/species include: *C. efficiens*, with the deposited type strain being DSM44549, *C. glutamicum*, with the deposited type strain being ATCC13032, and *C. ammoniagenes*, with the deposited type strain

being ATCC6871. In some embodiments the preferred host cell of the present disclosure is *C. glutamicum*.

[369] Suitable host cells of the genus *Corynebacterium*, in particular of the species *Corynebacterium glutamicum*, are in particular the known wild-type strains: *Corynebacterium glutamicum* ATCC13032, *Corynebacterium acetoglutamicum* ATCC15806, *Corynebacterium acetoacidophilum* ATCC13870, *Corynebacterium melassecola* ATCC17965, *Corynebacterium thermoaminogenes* FERM BP-1539, *Brevibacterium flavum* ATCC14067, *Brevibacterium lactofermentum* ATCC13869, and *Brevibacterium divaricatum* ATCC14020; and L-amino acid-producing mutants, or strains, prepared therefrom, such as, for example, the L-lysine-producing strains: *Corynebacterium glutamicum* FERM-P 1709, *Brevibacterium flavum* FERM-P 1708, *Brevibacterium lactofermentum* FERM-P 1712, *Corynebacterium glutamicum* FERM-P 6463, *Corynebacterium glutamicum* FERM-P 6464, *Corynebacterium glutamicum* DM58-1, *Corynebacterium glutamicum* DG52-5, *Corynebacterium glutamicum* DSM5714, and *Corynebacterium glutamicum* DSM12866.

[370] Suitable yeast host cells include, but are not limited to: *Candida*, *Hansenula*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, *Kluyveromyces*, and *Yarrowia*. In some embodiments, the yeast cell is *Hansenula polymorpha*, *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, *Saccharomyces diastaticus*, *Saccharomyces norbensis*, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe*, *Komagataella phaffii*, formerly known as *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia kodamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia quercuum*, *Pichia pijperi*, *Pichia stipitis*, *Pichia methanolica*, *Pichia angusta*, *Kluyveromyces lactis*, *Candida albicans*, or *Yarrowia lipolytica*.

[371] In some embodiments, the yeast strain is an industrial polyploid yeast strain. Other non-limiting examples of fungal cells include cells obtained from *Aspergillus* spp., *Penicillium* spp., *Fusarium* spp., *Rhizopus* spp., *Acremonium* spp., *Neurospora* spp., *Sordaria* spp., *Magnaporthe* spp., *Allomyces* spp., *Ustilago* spp., *Botrytis* spp., and *Trichoderma* spp.

[372] In certain embodiments, the host cell is an algal cell such as, *Chlamydomonas* (e.g., *C. Reinhardtii*) and *Phormidium* (*P. sp.* ATCC29409).

[373] In other embodiments, the host cell is a prokaryotic cell. Suitable prokaryotic cells include gram positive, gram negative, and gram-variable bacterial cells. The host cell may be a species of, but not limited to: *Agrobacterium*, *Alicyclobacillus*, *Anabaena*, *Anacystis*, *Acinetobacter*, *Acidothermus*, *Arthrobacter*, *Azobacter*, *Bacillus*, *Bifidobacterium*, *Brevibacterium*, *Butyrivibrio*, *Buchnera*, *Campestris*, *Campylobacter*, *Clostridium*, *Corynebacterium*, *Chromatium*, *Coprococcus*, *Escherichia*, *Enterococcus*, *Enterobacter*, *Erwinia*, *Fusobacterium*, *Faecalibacterium*, *Francisella*, *Flavobacterium*, *Geobacillus*, *Haemophilus*, *Helicobacter*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, *Ilyobacter*, *Micrococcus*, *Microbacterium*, *Mesorhizobium*, *Methylobacterium*, *Methylobacterium*, *Mycobacterium*, *Neisseria*, *Pantoea*, *Pseudomonas*, *Prochlorococcus*, *Rhodobacter*, *Rhodopseudomonas*, *Rhodopseudomonas*, *Roseburia*, *Rhodospirillum*, *Rhodococcus*, *Scenedesmus*, *Streptomyces*, *Streptococcus*, *Syneccoccus*, *Saccharomonospora*, *Saccharopolyspora*, *Staphylococcus*, *Serratia*, *Salmonella*, *Shigella*, *Thermoanaerobacterium*, *Tropheryma*, *Tularensis*, *Temecula*, *Thermosynechococcus*, *Thermococcus*, *Ureaplasma*, *Xanthomonas*, *Xylella*, *Yersinia*, and *Zymomonas*.

[374] In some embodiments, the bacterial host strain is an industrial strain. Numerous bacterial industrial strains are known and suitable for the methods and compositions described in this application.

[375] In some embodiments, the bacterial host cell is of the *Agrobacterium* species (e.g., *A. radiobacter*, *A. rhizogenes*, *A. rubi*), the *Arthrobacter* species (e.g., *A. aureus*, *A. citreus*, *A. globiformis*, *A. hydrocarboglutamicus*, *A. mysorens*, *A. nicotianae*, *A. paraffineus*, *A. protophominae*, *A. roseoparaffinus*, *A. sulfureus*, *A. ureafaciens*), the *Bacillus* species (e.g., *B. thuringiensis*, *B. anthracis*, *B. megaterium*, *B. subtilis*, *B. lentus*, *B. circularis*, *B. pumilus*, *B. lautus*, *B. coagulans*, *B. brevis*, *B. firmus*, *B. alkaophilus*, *B. licheniformis*, *B. clausii*, *B. stearothermophilus*, *B. halodurans* and *B. amyloliquefaciens*). In particular embodiments, the host cell will be an industrial *Bacillus* strain including but not limited to *B. subtilis*, *B. pumilus*, *B. licheniformis*, *B. megaterium*, *B. clausii*, *B. stearothermophilus* and *B. amyloliquefaciens*. In some embodiments, the host cell will be an industrial *Clostridium* species (e.g., *C. acetobutylicum*, *C. tetani* E88, *C. lituseburense*, *C. saccharobutylicum*, *C. perfringens*, *C. beijerinckii*). In some embodiments, the host cell will be an industrial *Corynebacterium* species (e.g., *C. glutamicum*, *C. acetoacidophilum*). In some embodiments, the host cell will be an industrial *Escherichia* species (e.g., *E. coli*). In some

embodiments, the host cell will be an industrial *Erwinia* species (e.g., *E. uredovora*, *E. carotovora*, *E. ananas*, *E. herbicola*, *E. punctata*, *E. terreus*). In some embodiments, the host cell will be an industrial *Pantoea* species (e.g., *P. citrea*, *P. agglomerans*). In some embodiments, the host cell will be an industrial *Pseudomonas* species, (e.g., *P. putida*, *P. aeruginosa*, *P. mevalonii*). In some embodiments, the host cell will be an industrial *Streptococcus* species (e.g., *S. equisimilis*, *S. pyogenes*, *S. uberis*). In some embodiments, the host cell will be an industrial *Streptomyces* species (e.g., *S. ambofaciens*, *S. achromogenes*, *S. avermitilis*, *S. coelicolor*, *S. aureofaciens*, *S. aureus*, *S. fungicidicus*, *S. griseus*, *S. lividans*). In some embodiments, the host cell will be an industrial *Zymomonas* species (e.g., *Z. mobilis*, *Z. lipolytica*), and the like.

[376] The present disclosure is also suitable for use with a variety of animal cell types, including mammalian cells, for example, human (including 293, HeLa, WI38, PER.C6 and Bowes melanoma cells), mouse (including 3T3, NS0, NS1, Sp2/0), hamster (CHO, BHK), monkey (COS, FRhL, Vero), insect cells, for example fall armyworm (including Sf9 and Sf21), silkworm (including BmN), cabbage looper (including BTI-Tn-5B1-4) and common fruit fly (including Schneider 2), and hybridoma cell lines.

[377] In various embodiments, strains that may be used in the practice of the disclosure including both prokaryotic and eukaryotic strains, and are readily accessible to the public from a number of culture collections such as American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL). The present disclosure is also suitable for use with a variety of plant cell types. In some embodiments, the plant is of the *Cannabis* genus in the family Cannabaceae. In certain embodiments, the plant is of the species *Cannabis sativa*, *Cannabis indica*, or *Cannabis ruderalis*. In other embodiments, the plant is of the genus *Nicotiana* in the family Solanaceae. In certain embodiments, the plant is of the species *Nicotiana rustica*.

[378] The term “cell,” as used in this application, may refer to a single cell or a population of cells, such as a population of cells belonging to the same cell line or strain. Use of the singular term “cell” should not be construed to refer explicitly to a single cell rather than a population of cells. The host cell may comprise genetic modifications relative to a wild-type

counterpart. Reduction of gene expression and/or gene inactivation in a host cell may be achieved through any suitable method, including but not limited to, deletion of the gene, introduction of a point mutation into the gene, selective editing of the gene and/or truncation of the gene. For example, polymerase chain reaction (PCR)-based methods may be used (see, *e.g.*, Gardner *et al.*, *Methods Mol Biol.* 2014;1205:45-78). As a non-limiting example, genes may be deleted through gene replacement (*e.g.*, with a marker, including a selection marker). A gene may also be truncated through the use of a transposon system (see, *e.g.*, Poussu *et al.*, *Nucleic Acids Res.* 2005; 33(12): e104). A gene may also be edited through the use of gene editing technologies known in the art, such as CRISPR-based technologies.

Culturing of Host Cells

[379] Any of the cells disclosed in this application can be cultured in media of any type (rich or minimal) and any composition prior to, during, and/or after contact and/or integration of a nucleic acid. The conditions of the culture or culturing process can be optimized through routine experimentation as would be understood by one of ordinary skill in the art. In some embodiments, the selected media is supplemented with various components. In some embodiments, the concentration and amount of a supplemental component is optimized. In some embodiments, other aspects of the media and growth conditions (*e.g.*, pH, temperature, etc.) are optimized through routine experimentation. In some embodiments, the frequency that the media is supplemented with one or more supplemental components, and the amount of time that the cell is cultured, is optimized.

[380] Culturing of the cells described in this application can be performed in culture vessels known and used in the art. In some embodiments, an aerated reaction vessel (*e.g.*, a stirred tank reactor) is used to culture the cells. In some embodiments, a bioreactor or fermenter is used to culture the cell. Thus, in some embodiments, the cells are used in fermentation. As used in this application, the terms “bioreactor” and “fermenter” are interchangeably used and refer to an enclosure, or partial enclosure, in which a biological, biochemical and/or chemical reaction takes place that involves a living organism or part of a living organism. A “large-scale bioreactor” or “industrial-scale bioreactor” is a bioreactor that is used to generate a product on a commercial or quasi-commercial scale. Large scale bioreactors typically have volumes in the range of liters, hundreds of liters, thousands of liters, or more.

[381] Non-limiting examples of bioreactors include: stirred tank fermenters, bioreactors agitated by rotating mixing devices, chemostats, bioreactors agitated by shaking devices, airlift fermenters, packed-bed reactors, fixed-bed reactors, fluidized bed bioreactors, bioreactors employing wave induced agitation, centrifugal bioreactors, roller bottles, and hollow fiber bioreactors, roller apparatuses (for example benchtop, cart-mounted, and/or automated varieties), vertically-stacked plates, spinner flasks, stirring or rocking flasks, shaken multi-well plates, MD bottles, T-flasks, Roux bottles, multiple-surface tissue culture propagators, modified fermenters, and coated beads (*e.g.*, beads coated with serum proteins, nitrocellulose, or carboxymethyl cellulose to prevent cell attachment).

[382] In some embodiments, the bioreactor includes a cell culture system where the cell (*e.g.*, yeast cell) is in contact with moving liquids and/or gas bubbles. In some embodiments, the cell or cell culture is grown in suspension. In other embodiments, the cell or cell culture is attached to a solid phase carrier. Non-limiting examples of a carrier system includes microcarriers (*e.g.*, polymer spheres, microbeads, and microdisks that can be porous or non-porous), cross-linked beads (*e.g.*, dextran) charged with specific chemical groups (*e.g.*, tertiary amine groups), 2D microcarriers including cells trapped in nonporous polymer fibers, 3D carriers (*e.g.*, carrier fibers, hollow fibers, multicartridge reactors, and semi-permeable membranes that can comprising porous fibers), microcarriers having reduced ion exchange capacity, encapsulation cells, capillaries, and aggregates. In some embodiments, carriers are fabricated from materials such as dextran, gelatin, glass, or cellulose.

[383] In some embodiments, industrial-scale processes are operated in continuous, semi-continuous or non-continuous modes. Non-limiting examples of operation modes are batch, fed batch, extended batch, repetitive batch, draw/fill, rotating-wall, spinning flask, and/or perfusion mode of operation. In some embodiments, a bioreactor allows continuous or semi-continuous replenishment of the substrate stock, for example a carbohydrate source and/or continuous or semi-continuous separation of the product, from the bioreactor.

[384] In some embodiments, the bioreactor or fermenter includes a sensor and/or a control system to measure and/or adjust reaction parameters. Non-limiting examples of reaction parameters include biological parameters (*e.g.*, growth rate, cell size, cell number, cell density, cell type, or cell state, etc.), chemical parameters (*e.g.*, pH, redox-potential, concentration of reaction substrate and/or product, concentration of dissolved gases, such as

oxygen concentration and CO₂ concentration, nutrient concentrations, metabolite concentrations, concentration of an oligopeptide, concentration of an amino acid, concentration of a vitamin, concentration of a hormone, concentration of an additive, serum concentration, ionic strength, concentration of an ion, relative humidity, molarity, osmolarity, concentration of other chemicals, for example buffering agents, adjuvants, or reaction by-products), physical/mechanical parameters (*e.g.*, density, conductivity, degree of agitation, pressure, and flow rate, shear stress, shear rate, viscosity, color, turbidity, light absorption, mixing rate, conversion rate, as well as thermodynamic parameters, such as temperature, light intensity/quality, etc.). Sensors to measure the parameters described in this application are well known to one of ordinary skill in the relevant mechanical and electronic arts. Control systems to adjust the parameters in a bioreactor based on the inputs from a sensor described in this application are well known to one of ordinary skill in the art in bioreactor engineering.

[385] In some embodiments, the method involves batch fermentation (*e.g.*, shake flask fermentation). General considerations for batch fermentation (*e.g.*, shake flask fermentation) include the level of oxygen and glucose. For example, batch fermentation (*e.g.*, shake flask fermentation) may be oxygen and glucose limited, so in some embodiments, the capability of a strain to perform in a well-designed fed-batch fermentation is underestimated. Also, the final product (*e.g.*, cannabinoid or cannabinoid precursor) may display some differences from the substrate in terms of solubility, toxicity, cellular accumulation and secretion and in some embodiments can have different fermentation kinetics.

[386] In some embodiments, the cells of the present disclosure are adapted to produce cannabinoids or cannabinoid precursors *in vivo*. In some embodiments, the cells are adapted to secrete one or more enzymes for cannabinoid synthesis (*e.g.*, AAE, PKS, PKC, PT, or TS). In some embodiments, the cells of the present disclosure are lysed, and the remaining lysates are recovered for subsequent use. In such embodiments, the secreted or lysed enzyme can catalyze reactions for the production of a cannabinoid or precursor by bioconversion in an *in vitro* or *ex vivo* process. In some embodiments, any and all conversions described in this application can be conducted chemically or enzymatically, *in vitro* or *in vivo*.

[387] In some embodiments, the host cells of the present disclosure are adapted to produce cannabinoids or cannabinoid precursors *in vivo*. In some embodiments, the host cells are adapted to secrete one or more cannabinoid pathway substrates, intermediates, and/or

terminal products (e.g., olivetol, THCA, THC, CBDA, CBD, CBGA, CBGVA, THCVA, CBDVA, CBCVA, or CBCA). In some embodiments, the host cells of the present disclosure are lysed, and the lysate is recovered for subsequent use. In such embodiments, the secreted substrates, intermediates, and/or terminal products may be recovered from the culture media.

Purification and further processing

[388] In some embodiments, any of the methods described in this application may include isolation and/or purification of the cannabinoids and/or cannabinoid precursors produced (e.g., produced in a bioreactor). For example, the isolation and/or purification can involve one or more of cell lysis, centrifugation, extraction, column chromatography, distillation, crystallization, and lyophilization.

[389] The methods described in this application encompass production of any cannabinoid or cannabinoid precursor known in the art. Cannabinoids or cannabinoid precursors produced by any of the recombinant cells disclosed in this application or any of the *in vitro* methods described in this application may be identified and extracted using any method known in the art. Mass spectrometry (e.g., LC-MS, GC-MS) is a non-limiting example of a method for identification and may be used to extract a compound of interest.

[390] In some embodiments, any of the methods described in this application further comprise decarboxylation of a cannabinoid or cannabinoid precursor. As a non-limiting example, the acid form of a cannabinoid or cannabinoid precursor may be heated (e.g., at least 90°C) to decarboxylate the cannabinoid or cannabinoid precursor. See, e.g., U.S. Patent No. 10,159,908, U.S. Patent No. 10,143,706, U.S. Patent No. 9,908,832 and U.S. Patent No. 7,344,736. See also, e.g., Wang *et al.*, Cannabis Cannabinoid Res. 2016; 1(1): 262–271.

Compositions, kits, and administration

[391] The present disclosure provides compositions, including pharmaceutical compositions, comprising a cannabinoid or a cannabinoid precursor, or pharmaceutically acceptable salt thereof, produced by any of the methods described in this application, and optionally a pharmaceutically acceptable excipient.

[392] In certain embodiments, a cannabinoid or cannabinoid precursor described in this application is provided in an effective amount in a composition, such as a pharmaceutical

composition. In certain embodiments, the effective amount is a therapeutically effective amount. In certain embodiments, the effective amount is a prophylactically effective amount.

[393] Compositions, such as pharmaceutical compositions, described in this application can be prepared by any method known in the art. In general, such preparatory methods include bringing a compound described in this application (*i.e.*, the “active ingredient”) into association with a carrier or excipient, and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping, and/or packaging the product into a desired single- or multi-dose unit.

[394] Pharmaceutical compositions can be prepared, packaged, and/or sold in bulk, as a single unit dose, and/or as a plurality of single unit doses. A “unit dose” is a discrete amount of the pharmaceutical composition comprising a predetermined amount of the active ingredient. The amount of the active ingredient is generally equal to the dosage of the active ingredient which would be administered to a subject and/or a convenient fraction of such a dosage, such as one-half or one-third of such a dosage.

[395] Relative amounts of the active ingredient, the pharmaceutically acceptable excipient, and/or any additional ingredients in a pharmaceutical composition described in this application will vary, depending upon the identity, size, and/or condition of the subject treated and further depending upon the route by which the composition is to be administered. The composition may comprise between 0.1% and 100% (w/w) active ingredient.

[396] Pharmaceutically acceptable excipients used in the manufacture of pharmaceutical compositions include inert diluents, dispersing and/or granulating agents, surface active agents and/or emulsifiers, disintegrating agents, binding agents, preservatives, buffering agents, lubricating agents, and/or oils. Excipients such as cocoa butter and suppository waxes, coloring agents, coating agents, sweetening, flavoring, and perfuming agents may also be present in the composition. Exemplary excipients include diluents, dispersing and/or granulating agents, surface active agents and/or emulsifiers, disintegrating agents, binding agents, preservatives, buffering agents, lubricating agents, and/or oils (*e.g.*, synthetic oils, semi-synthetic oils) as disclosed in this application.

[397] Exemplary diluents include calcium carbonate, sodium carbonate, calcium phosphate, dicalcium phosphate, calcium sulfate, calcium hydrogen phosphate, sodium

phosphate lactose, sucrose, cellulose, microcrystalline cellulose, kaolin, mannitol, sorbitol, inositol, sodium chloride, dry starch, cornstarch, powdered sugar, and mixtures thereof.

[398] Exemplary granulating and/or dispersing agents include potato starch, corn starch, tapioca starch, sodium starch glycolate, clays, alginic acid, guar gum, citrus pulp, agar, bentonite, cellulose, and wood products, natural sponge, cation-exchange resins, calcium carbonate, silicates, sodium carbonate, cross-linked poly(vinyl-pyrrolidone) (crospovidone), sodium carboxymethyl starch (sodium starch glycolate), carboxymethyl cellulose, cross-linked sodium carboxymethyl cellulose (croscarmellose), methylcellulose, pregelatinized starch (starch 1500), microcrystalline starch, water insoluble starch, calcium carboxymethyl cellulose, magnesium aluminum silicate (Veegum), sodium lauryl sulfate, quaternary ammonium compounds, and mixtures thereof.

[399] Exemplary surface active agents and/or emulsifiers include natural emulsifiers (*e.g.*, acacia, agar, alginic acid, sodium alginate, tragacanth, chondrux, cholesterol, xanthan, pectin, gelatin, egg yolk, casein, wool fat, cholesterol, wax, and lecithin), colloidal clays (*e.g.*, bentonite (aluminum silicate) and Veegum (magnesium aluminum silicate)), long chain amino acid derivatives, high molecular weight alcohols (*e.g.*, stearyl alcohol, cetyl alcohol, oleyl alcohol, triacetin monostearate, ethylene glycol distearate, glyceryl monostearate, and propylene glycol monostearate, polyvinyl alcohol), carbomers (*e.g.*, carboxy polymethylene, polyacrylic acid, acrylic acid polymer, and carboxyvinyl polymer), carrageenan, cellulosic derivatives (*e.g.*, carboxymethylcellulose sodium, powdered cellulose, hydroxymethyl cellulose, hydroxypropyl cellulose, hydroxypropyl methylcellulose, methylcellulose), sorbitan fatty acid esters (*e.g.*, polyoxyethylene sorbitan monolaurate (Tween[®] 20), polyoxyethylene sorbitan (Tween[®] 60), polyoxyethylene sorbitan monooleate (Tween[®] 80), sorbitan monopalmitate (Span[®] 40), sorbitan monostearate (Span[®] 60), sorbitan tristearate (Span[®] 65), glyceryl monooleate, sorbitan monooleate (Span[®] 80), polyoxyethylene esters (*e.g.*, polyoxyethylene monostearate (Myrj[®] 45), polyoxyethylene hydrogenated castor oil, polyethoxylated castor oil, polyoxymethylene stearate, and Solutol[®]), sucrose fatty acid esters, polyethylene glycol fatty acid esters (*e.g.*, Cremophor[®]), polyoxyethylene ethers, (*e.g.*, polyoxyethylene lauryl ether (Brij[®] 30)), poly(vinyl-pyrrolidone), diethylene glycol monolaurate, triethanolamine oleate, sodium oleate, potassium oleate, ethyl oleate, oleic acid, ethyl laurate, sodium lauryl sulfate, Pluronic[®] F-68, poloxamer P-188, cetrimonium bromide, cetylpyridinium chloride, benzalkonium chloride, docusate sodium, and/or mixtures thereof.

[400] Exemplary binding agents include starch (*e.g.*, cornstarch and starch paste), gelatin, sugars (*e.g.*, sucrose, glucose, dextrose, dextrin, molasses, lactose, lactitol, mannitol, *etc.*), natural and synthetic gums (*e.g.*, acacia, sodium alginate, extract of Irish moss, panwar gum, ghatti gum, mucilage of isapol husks, carboxymethylcellulose, methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropyl cellulose, hydroxypropyl methylcellulose, microcrystalline cellulose, cellulose acetate, poly(vinyl-pyrrolidone), magnesium aluminum silicate (Veegum[®]), and larch arabogalactan), alginates, polyethylene oxide, polyethylene glycol, inorganic calcium salts, silicic acid, polymethacrylates, waxes, water, alcohol, and/or mixtures thereof.

[401] Exemplary preservatives include antioxidants, chelating agents, antimicrobial preservatives, antifungal preservatives, antiprotozoan preservatives, alcohol preservatives, acidic preservatives, and other preservatives. In certain embodiments, the preservative is an antioxidant. In other embodiments, the preservative is a chelating agent.

[402] Exemplary antioxidants include alpha tocopherol, ascorbic acid, ascorbyl palmitate, butylated hydroxyanisole, butylated hydroxytoluene, monothioglycerol, potassium metabisulfite, propionic acid, propyl gallate, sodium ascorbate, sodium bisulfite, sodium metabisulfite, and sodium sulfite.

[403] Exemplary chelating agents include ethylenediaminetetraacetic acid (EDTA) and salts and hydrates thereof (*e.g.*, sodium edetate, disodium edetate, trisodium edetate, calcium disodium edetate, dipotassium edetate, and the like), citric acid and salts and hydrates thereof (*e.g.*, citric acid monohydrate), fumaric acid and salts and hydrates thereof, malic acid and salts and hydrates thereof, phosphoric acid and salts and hydrates thereof, and tartaric acid and salts and hydrates thereof. Exemplary antimicrobial preservatives include benzalkonium chloride, benzethonium chloride, benzyl alcohol, bronopol, cetrimide, cetylpyridinium chloride, chlorhexidine, chlorobutanol, chlorocresol, chloroxylonol, cresol, ethyl alcohol, glycerin, hexetidine, imidurea, phenol, phenoxyethanol, phenylethyl alcohol, phenylmercuric nitrate, propylene glycol, and thimerosal.

[404] Exemplary antifungal preservatives include butyl paraben, methyl paraben, ethyl paraben, propyl paraben, benzoic acid, hydroxybenzoic acid, potassium benzoate, potassium sorbate, sodium benzoate, sodium propionate, and sorbic acid.

[405] Exemplary alcohol preservatives include ethanol, polyethylene glycol, phenol, phenolic compounds, bisphenol, chlorobutanol, hydroxybenzoate, and phenylethyl alcohol.

[406] Exemplary acidic preservatives include vitamin A, vitamin C, vitamin E, beta-carotene, citric acid, acetic acid, dehydroacetic acid, ascorbic acid, sorbic acid, and phytic acid.

[407] Other preservatives include tocopherol, tocopherol acetate, deteroxime mesylate, cetrimide, butylated hydroxyanisol (BHA), butylated hydroxytoluene (BHT), ethylenediamine, sodium lauryl sulfate (SLS), sodium lauryl ether sulfate (SLES), sodium bisulfite, sodium metabisulfite, potassium sulfite, potassium metabisulfite, Glydant[®] Plus, Phenonip[®], methylparaben, Germall[®] 115, Germaben[®] II, Neolone[®], Kathon[®], and Euxyl[®].

[408] Exemplary buffering agents include citrate buffer solutions, acetate buffer solutions, phosphate buffer solutions, ammonium chloride, calcium carbonate, calcium chloride, calcium citrate, calcium gluconate, calcium gluceptate, calcium gluconate, D-gluconic acid, calcium glycerophosphate, calcium lactate, propanoic acid, calcium levulinate, pentanoic acid, dibasic calcium phosphate, phosphoric acid, tribasic calcium phosphate, calcium hydroxide phosphate, potassium acetate, potassium chloride, potassium gluconate, potassium mixtures, dibasic potassium phosphate, monobasic potassium phosphate, potassium phosphate mixtures, sodium acetate, sodium bicarbonate, sodium chloride, sodium citrate, sodium lactate, dibasic sodium phosphate, monobasic sodium phosphate, sodium phosphate mixtures, tromethamine, magnesium hydroxide, aluminum hydroxide, alginic acid, pyrogen-free water, isotonic saline, Ringer's solution, ethyl alcohol, and mixtures thereof.

[409] Exemplary lubricating agents include magnesium stearate, calcium stearate, stearic acid, silica, talc, malt, glyceryl behenate, hydrogenated vegetable oils, polyethylene glycol, sodium benzoate, sodium acetate, sodium chloride, leucine, magnesium lauryl sulfate, sodium lauryl sulfate, and mixtures thereof.

[410] Exemplary natural oils include almond, apricot kernel, avocado, babassu, bergamot, black current seed, borage, cade, camomile, canola, caraway, carnauba, castor, cinnamon, cocoa butter, coconut, cod liver, coffee, corn, cotton seed, emu, eucalyptus, evening primrose, fish, flaxseed, geraniol, gourd, grape seed, hazel nut, hyssop, isopropyl myristate, jojoba, kukui nut, lavandin, lavender, lemon, litsea cubeba, macademia nut, mallow, mango seed, meadowfoam seed, mint, nutmeg, olive, orange, orange roughy, palm, palm kernel,

peach kernel, peanut, poppy seed, pumpkin seed, rapeseed, rice bran, rosemary, safflower, sandalwood, sasquana, savoury, sea buckthorn, sesame, shea butter, silicone, soybean, sunflower, tea tree, thistle, tsubaki, vetiver, walnut, and wheat germ oils. Exemplary synthetic or semi-synthetic oils include, but are not limited to, butyl stearate, medium chain triglycerides (such as caprylic triglyceride and capric triglyceride), cyclomethicone, diethyl sebacate, dimethicone 360, isopropyl myristate, mineral oil, octyldodecanol, oleyl alcohol, silicone oil, and mixtures thereof. In certain embodiments, exemplary synthetic oils comprise medium chain triglycerides (such as caprylic triglyceride and capric triglyceride).

[411] Liquid dosage forms for oral and parenteral administration include pharmaceutically acceptable emulsions, microemulsions, solutions, suspensions, syrups and elixirs. In addition to the active ingredients, the liquid dosage forms may comprise inert diluents commonly used in the art such as, for example, water or other solvents, solubilizing agents and emulsifiers such as ethyl alcohol, isopropyl alcohol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate, propylene glycol, 1,3-butylene glycol, dimethylformamide, oils (*e.g.*, cottonseed, groundnut, corn, germ, olive, castor, and sesame oils), glycerol, tetrahydrofurfuryl alcohol, polyethylene glycols and fatty acid esters of sorbitan, and mixtures thereof. Besides inert diluents, the oral compositions can include adjuvants such as wetting agents, emulsifying and suspending agents, sweetening, flavoring, and perfuming agents. In certain embodiments for parenteral administration, the conjugates described in this application are mixed with solubilizing agents such as Cremophor[®], alcohols, oils, modified oils, glycols, polysorbates, cyclodextrins, polymers, and mixtures thereof.

[412] Injectable preparations, for example, sterile injectable aqueous or oleaginous suspensions can be formulated according to the known art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation can be a sterile injectable solution, suspension, or emulsion in a nontoxic parenterally acceptable diluent or solvent, for example, as a solution in 1,3-butanediol. Among the acceptable vehicles and solvents that can be employed are water, Ringer's solution, U.S.P., and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose, any bland fixed oil can be employed including synthetic mono- or diglycerides. In addition, fatty acids such as oleic acid are used in the preparation of injectables.

[413] The injectable formulations can be sterilized, for example, by filtration through a bacterial-retaining filter, or by incorporating sterilizing agents in the form of sterile solid compositions which can be dissolved or dispersed in sterile water or other sterile injectable medium prior to use.

[414] In order to prolong the effect of a drug, it is often desirable to slow the absorption of the drug from subcutaneous or intramuscular injection. This can be accomplished by the use of a liquid suspension of crystalline or amorphous material with poor water solubility. The rate of absorption of the drug then depends upon its rate of dissolution, which, in turn, may depend upon crystal size and crystalline form. Alternatively, delayed absorption of a parenterally administered drug form may be accomplished by dissolving or suspending the drug in an oil vehicle.

[415] Compositions for rectal or vaginal administration are typically suppositories which can be prepared by mixing the conjugates described in this application with suitable non-irritating excipients or carriers such as cocoa butter, polyethylene glycol, or a suppository wax which are solid at ambient temperature but liquid at body temperature and therefore melt in the rectum or vaginal cavity and release the active ingredient.

[416] Solid dosage forms for oral administration include capsules, tablets, pills, powders, and granules. In such solid dosage forms, the active ingredient is mixed with at least one inert, pharmaceutically acceptable excipient or carrier such as sodium citrate or dicalcium phosphate and/or (a) fillers or extenders such as starches, lactose, sucrose, glucose, mannitol, and silicic acid, (b) binders such as, for example, carboxymethylcellulose, alginates, gelatin, polyvinylpyrrolidinone, sucrose, and acacia, (c) humectants such as glycerol, (d) disintegrating agents such as agar, calcium carbonate, potato or tapioca starch, alginic acid, certain silicates, and sodium carbonate, (e) solution retarding agents such as paraffin, (f) absorption accelerators such as quaternary ammonium compounds, (g) wetting agents such as, for example, cetyl alcohol and glycerol monostearate, (h) absorbents such as kaolin and bentonite clay, and (i) lubricants such as talc, calcium stearate, magnesium stearate, solid polyethylene glycols, sodium lauryl sulfate, and mixtures thereof. In the case of capsules, tablets, and pills, the dosage form may include a buffering agent.

[417] Solid compositions of a similar type can be employed as fillers in soft and hard-filled gelatin capsules using such excipients as lactose or milk sugar as well as high molecular

weight polyethylene glycols and the like. The solid dosage forms of tablets, dragées, capsules, pills, and granules can be prepared with coatings and shells such as enteric coatings and other coatings well known in the art of pharmacology. They may optionally comprise opacifying agents and can be of a composition that they release the active ingredient(s) only, or preferentially, in a certain part of the intestinal tract, optionally, in a delayed manner. Examples of encapsulating compositions which can be used include polymeric substances and waxes. Solid compositions of a similar type can be employed as fillers in soft and hard-filled gelatin capsules using such excipients as lactose or milk sugar as well as high molecular weight polyethylene glycols and the like.

[418] The active ingredient can be in a micro-encapsulated form with one or more excipients as noted above. The solid dosage forms of tablets, dragées, capsules, pills, and granules can be prepared with coatings and shells such as enteric coatings, release controlling coatings, and other coatings well known in the pharmaceutical formulating art. In such solid dosage forms the active ingredient can be admixed with at least one inert diluent such as sucrose, lactose, or starch. Such dosage forms may comprise, as is normal practice, additional substances other than inert diluents, *e.g.*, tableting lubricants and other tableting aids such as magnesium stearate and microcrystalline cellulose. In the case of capsules, tablets and pills, the dosage forms may comprise buffering agents. They may optionally comprise opacifying agents and can be of a composition that they release the active ingredient(s) only, or preferentially, in a certain part of the intestinal tract, optionally, in a delayed manner. Examples of encapsulating agents which can be used include polymeric substances and waxes.

[419] Dosage forms for topical and/or transdermal administration of a compound described in this application may include ointments, pastes, creams, lotions, gels, powders, solutions, sprays, inhalants, and/or patches. Generally, the active ingredient is admixed under sterile conditions with a pharmaceutically acceptable carrier or excipient and/or any needed preservatives and/or buffers as can be required. Additionally, the present disclosure contemplates the use of transdermal patches, which often have the added advantage of providing controlled delivery of an active ingredient to the body. Such dosage forms can be prepared, for example, by dissolving and/or dispensing the active ingredient in the proper medium. Alternatively or additionally, the rate can be controlled by either providing a rate controlling membrane and/or by dispersing the active ingredient in a polymer matrix and/or gel.

[420] Suitable devices for use in delivering intradermal pharmaceutical compositions described in this application include short needle devices. Intradermal compositions can be administered by devices which limit the effective penetration length of a needle into the skin. Alternatively or additionally, conventional syringes can be used in the classical mantoux method of intradermal administration. Jet injection devices which deliver liquid formulations to the dermis *via* a liquid jet injector and/or *via* a needle which pierces the stratum corneum and produces a jet which reaches the dermis are suitable. Ballistic powder/particle delivery devices which use compressed gas to accelerate the compound in powder form through the outer layers of the skin to the dermis are suitable.

[421] Formulations suitable for topical administration include, but are not limited to, liquid and/or semi-liquid preparations such as liniments, lotions, oil-in-water and/or water-in-oil emulsions such as creams, ointments, and/or pastes, and/or solutions and/or suspensions. Topically administrable formulations may, for example, comprise from about 1% to about 10% (w/w) active ingredient, although the concentration of the active ingredient can be as high as the solubility limit of the active ingredient in the solvent. Formulations for topical administration may further comprise one or more of the additional ingredients described in this application.

[422] A pharmaceutical composition described in this application can be prepared, packaged, and/or sold in a formulation suitable for pulmonary administration *via* the buccal cavity. Such a formulation may comprise dry particles which comprise the active ingredient and which have a diameter in the range from about 0.5 to about 7 nanometers, or from about 1 to about 6 nanometers. Such compositions are conveniently in the form of dry powders for administration using a device comprising a dry powder reservoir to which a stream of propellant can be directed to disperse the powder and/or using a self-propelling solvent/powder dispensing container such as a device comprising the active ingredient dissolved and/or suspended in a low-boiling propellant in a sealed container. Such powders comprise particles wherein at least 98% of the particles by weight have a diameter greater than 0.5 nanometers and at least 95% of the particles by number have a diameter less than 7 nanometers. Alternatively, at least 95% of the particles by weight have a diameter greater than 1 nanometer and at least 90% of the particles by number have a diameter less than 6 nanometers. Dry powder compositions may include a solid fine powder diluent such as sugar and are conveniently provided in a unit dose form.

[423] Low boiling propellants generally include liquid propellants having a boiling point of below 65° F at atmospheric pressure. Generally, the propellant may constitute 50 to 99.9% (w/w) of the composition, and the active ingredient may constitute 0.1 to 20% (w/w) of the composition. The propellant may further comprise additional ingredients such as a liquid non-ionic and/or solid anionic surfactant and/or a solid diluent (which may have a particle size of the same order as particles comprising the active ingredient).

[424] Although the descriptions of pharmaceutical compositions provided in this application are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals of all sorts. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with ordinary experimentation.

[425] Compounds provided in this application are typically formulated in dosage unit form for ease of administration and uniformity of dosage. It will be understood, however, that the total daily usage of the compositions described in this application will be decided by a physician within the scope of sound medical judgment. The specific therapeutically effective dose level for any particular subject or organism will depend upon a variety of factors including the disease being treated and the severity of the disorder; the activity of the specific active ingredient employed; the specific composition employed; the age, body weight, general health, sex, and diet of the subject; the time of administration, route of administration, and rate of excretion of the specific active ingredient employed; the duration of the treatment; drugs used in combination or coincidental with the specific active ingredient employed; and like factors well known in the medical arts.

[426] The compounds and compositions provided in this application can be administered by any route, including enteral (*e.g.*, oral), parenteral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, subcutaneous, intraventricular, transdermal, interdermal, rectal, intravaginal, intraperitoneal, topical (as by powders, ointments, creams, and/or drops), mucosal, nasal, bucal, sublingual; by intratracheal instillation, bronchial instillation, and/or inhalation; and/or as an oral spray, nasal spray, and/or aerosol. Specifically contemplated routes are oral administration, intravenous administration (*e.g.*, systemic

intravenous injection), regional administration *via* blood and/or lymph supply, and/or direct administration to an affected site. In general, the most appropriate route of administration will depend upon a variety of factors including the nature of the agent (*e.g.*, its stability in the environment of the gastrointestinal tract), and/or the condition of the subject (*e.g.*, whether the subject is able to tolerate oral administration).

[427] In some embodiments, compounds or compositions disclosed in this application are formulated and/or administered in nanoparticles. Nanoparticles are particles in the nanoscale. In some embodiments, nanoparticles are less than 1 μm in diameter. In some embodiments, nanoparticles are between about 1 and 100 nm in diameter. Nanoparticles include organic nanoparticles, such as dendrimers, liposomes, or polymeric nanoparticles. Nanoparticles also include inorganic nanoparticles, such as fullerenes, quantum dots, and gold nanoparticles. Compositions may comprise an aggregate of nanoparticles. In some embodiments, the aggregate of nanoparticles is homogeneous, while in other embodiments the aggregate of nanoparticles is heterogeneous.

[428] The exact amount of a compound required to achieve an effective amount will vary from subject to subject, depending, for example, on species, age, and general condition of a subject, severity of the side effects or disorder, identity of the particular compound, mode of administration, and the like. An effective amount may be included in a single dose (*e.g.*, single oral dose) or multiple doses (*e.g.*, multiple oral doses). In certain embodiments, when multiple doses are administered to a subject or applied to a tissue or cell, any two doses of the multiple doses include different or substantially the same amounts of a compound described in this application. In certain embodiments, when multiple doses are administered to a subject or applied to a tissue or cell, the frequency of administering the multiple doses to the subject or applying the multiple doses to the tissue or cell is three doses a day, two doses a day, one dose a day, one dose every other day, one dose every third day, one dose every week, one dose every two weeks, one dose every three weeks, or one dose every four weeks. In certain embodiments, the frequency of administering the multiple doses to the subject or applying the multiple doses to the tissue or cell is one dose per day. In certain embodiments, the frequency of administering the multiple doses to the subject or applying the multiple doses to the tissue or cell is two doses per day. In certain embodiments, the frequency of administering the multiple doses to the subject or applying the multiple doses to the tissue or cell is three doses per day. In certain embodiments, when multiple doses are administered to a subject or applied to a tissue or cell,

the duration between the first dose and last dose of the multiple doses is one day, two days, four days, one week, two weeks, three weeks, one month, two months, three months, four months, six months, nine months, one year, two years, three years, four years, five years, seven years, ten years, fifteen years, twenty years, or the lifetime of the subject, tissue, or cell. In certain embodiments, the duration between the first dose and last dose of the multiple doses is three months, six months, or one year. In certain embodiments, the duration between the first dose and last dose of the multiple doses is the lifetime of the subject, tissue, or cell. In certain embodiments, a dose (e.g., a single dose, or any dose of multiple doses) described in this application includes independently between 0.1 μ g and 1 μ g, between 0.001 mg and 0.01 mg, between 0.01 mg and 0.1 mg, between 0.1 mg and 1 mg, between 1 mg and 3 mg, between 3 mg and 10 mg, between 10 mg and 30 mg, between 30 mg and 100 mg, between 100 mg and 300 mg, between 300 mg and 1,000 mg, or between 1 g and 10 g, inclusive, of a compound described in this application. In certain embodiments, a dose described in this application includes independently between 1 mg and 3 mg, inclusive, of a compound described in this application. In certain embodiments, a dose described in this application includes independently between 3 mg and 10 mg, inclusive, of a compound described in this application. In certain embodiments, a dose described in this application includes independently between 10 mg and 30 mg, inclusive, of a compound described in this application. In certain embodiments, a dose described in this application includes independently between 30 mg and 100 mg, inclusive, of a compound described in this application.

[429] Dose ranges as described in this application provide guidance for the administration of provided pharmaceutical compositions to an adult. The amount to be administered to, for example, a child or an adolescent can be determined by a medical practitioner or person skilled in the art and can be lower or the same as that administered to an adult.

[430] A compound or composition, as described in this application, can be administered in combination with one or more additional pharmaceutical agents (e.g., therapeutically and/or prophylactically active agents). The compounds or compositions can be administered in combination with additional pharmaceutical agents that improve their activity, improve bioavailability, improve safety, reduce drug resistance, reduce and/or modify metabolism, inhibit excretion, and/or modify distribution in a subject or cell. It will also be appreciated that the therapy employed may achieve a desired effect for the same disorder,

and/or it may achieve different effects. In certain embodiments, a pharmaceutical composition described in this application including a compound described in this application and an additional pharmaceutical agent shows a synergistic effect that is absent in a pharmaceutical composition including one of the compound and the additional pharmaceutical agent, but not both.

[431] The compound or composition can be administered concurrently with, prior to, or subsequent to one or more additional pharmaceutical agents, which may be useful as, *e.g.*, combination therapies. Pharmaceutical agents include therapeutically active agents. Pharmaceutical agents also include prophylactically active agents. Pharmaceutical agents include small organic molecules such as drug compounds (*e.g.*, compounds approved for human or veterinary use by the U.S. Food and Drug Administration as provided in the Code of Federal Regulations (CFR)), peptides, proteins, carbohydrates, monosaccharides, oligosaccharides, polysaccharides, nucleoproteins, mucoproteins, lipoproteins, synthetic polypeptides or proteins, small molecules linked to proteins, glycoproteins, steroids, nucleic acids, DNAs, RNAs, nucleotides, nucleosides, oligonucleotides, antisense oligonucleotides, lipids, hormones, vitamins, and cells. In certain embodiments, the additional pharmaceutical agent is a pharmaceutical agent useful for treating and/or preventing a disease (*e.g.*, proliferative disease, neurological disease, painful condition, psychiatric disorder, or metabolic disorder). Each additional pharmaceutical agent may be administered at a dose and/or on a time schedule determined for that pharmaceutical agent. The additional pharmaceutical agents may also be administered together with each other and/or with the compound or composition described in this application in a single dose or administered separately in different doses. The particular combination to employ in a regimen will take into account compatibility of the compound described in this application with the additional pharmaceutical agent(s) and/or the desired therapeutic and/or prophylactic effect to be achieved. In general, it is expected that the additional pharmaceutical agent(s) in combination be utilized at levels that do not exceed the levels at which they are utilized individually. In some embodiments, the levels utilized in combination will be lower than those utilized individually.

[432] In some embodiments, one or more of the compositions described in this application are administered to a subject. In certain embodiments, the subject is an animal. The animal may be of either sex and may be at any stage of development. In certain embodiments, the subject is a human. In other embodiments, the subject is a non-human

animal. In certain embodiments, the subject is a mammal. In certain embodiments, the subject is a non-human mammal. In certain embodiments, the subject is a domesticated animal, such as a dog, cat, cow, pig, horse, sheep, or goat. In certain embodiments, the subject is a companion animal, such as a dog or cat. In certain embodiments, the subject is a livestock animal, such as a cow, pig, horse, sheep, or goat. In certain embodiments, the subject is a zoo animal. In another embodiment, the subject is a research animal, such as a rodent (*e.g.*, mouse, rat), dog, pig, or non-human primate.

[433] Also encompassed by the disclosure are kits (*e.g.*, pharmaceutical packs). The kits provided may comprise a composition, such as a pharmaceutical composition, or a compound described in this application and a container (*e.g.*, a vial, ampule, bottle, syringe, and/or dispenser package, or other suitable container). In some embodiments, provided kits may optionally further include a second container comprising a pharmaceutical excipient for dilution or suspension of a pharmaceutical composition or compound described in this application. In some embodiments, the pharmaceutical composition or compound described in this application provided in the first container and the second container a combined to form one unit dosage form.

[434] Thus, in one aspect, provided are kits including a first container comprising a compound or composition described in this application. In certain embodiments, the kits are useful for treating a disease in a subject in need thereof. In certain embodiments, the kits are useful for preventing a disease in a subject in need thereof. In certain embodiments, the kits are useful for reducing the risk of developing a disease in a subject in need thereof.

[435] In certain embodiments, a kit described in this application further includes instructions for using the kit. A kit described in this application may also include information as required by a regulatory agency such as the U.S. Food and Drug Administration (FDA). In certain embodiments, the information included in the kits is prescribing information. In certain embodiments, the kits and instructions provide for treating a disease in a subject in need thereof. In certain embodiments, the kits and instructions provide for preventing a disease in a subject in need thereof. In certain embodiments, the kits and instructions provide for reducing the risk of developing a disease in a subject in need thereof. A kit described in this application may include one or more additional pharmaceutical agents described in this application as a separate composition.

[436] In some embodiments, the compositions include consumer product, such as comestible, cosmetic, toiletry, potable, inhalable, and wellness products. Exemplary consumer products include salves, waxes, powdered concentrates, pastes, extracts, tinctures, powders, oils, capsules, skin patches, sublingual oral dose drops, mucous membrane oral spray doses, makeup, perfume, shampoos, cosmetic soaps, cosmetic creams, skin lotions, aromatic essential oils, massage oils, shaving preparations, oils for toiletry purposes, lip balm, cosmetic oils, facial washes, moisturizing creams, moisturizing body lotions, moisturizing face lotions, bath salts, bath gels, bath soaps in liquid form, shower gels, bath bombs, hair care preparations, shampoos, conditioner, chocolate bars, brownies, chocolates, cookies, crackers, cakes, cupcakes, puddings, honey, chocolate confections, frozen confections, fruit-based confectionery, sugar confectionery, gummy candies, dragées, pastries, cereal bars, chocolate, cereal based energy bars, candy, ice cream, tea-based beverages, coffee-based beverages, and herbal infusions.

[437] The present invention is further illustrated by the following Examples, which in no way should be construed as limiting. The entire contents of all of the references (including literature references, issued patents, published patent applications, and co-pending patent applications) cited throughout this application are hereby expressly incorporated by reference. If a reference incorporated in this application contains a term whose definition is incongruous or incompatible with the definition of same term as defined in the present disclosure, the meaning ascribed to the term in this disclosure shall govern. However, mention of any reference, article, publication, patent, patent publication, and patent application cited in this application is not, and should not be taken as an acknowledgment or any form of suggestion that they constitute valid prior art or form part of the common general knowledge in any country in the world.

EXAMPLES

*Example 1: Functional characterization of a library of synthetic polyketide synthase (PKS) mutants in *S. cerevisiae**

[438] To identify polyketide synthase (PKS) mutants with enhanced activity against butyryl-CoA, a library of approximately 1004 PKSs containing one or more amino acid substitutions relative to a PKS with the sequence of SEQ ID NO: 1 was designed. The genes within the library were codon-optimized for expression in *S. cerevisiae* and synthesized in the

integrative yeast expression vector shown in FIG. 5. Each PKS mutant was transformed into an *S. cerevisiae* CEN.PK strain also expressing an AAE from *C. arietinum* (SEQ ID NO: 86) and a synthetic polyketide cyclase (SEQ ID NO: 61). Strain t896773, expressing a *C. sativa* PKS was used as a positive control to establish hit ranking. Strain t896591, expressing a GFP fluorescent reporter, was used as a negative control.

[439] An assay for measuring production of divaric acid and divarinol was conducted as follows: each thawed glycerol stock of PKS transformants was stamped into a well of YPD + 4% dextrose media. Samples were incubated at 30°C in a shaking incubator for 2 days. A portion of each of the resulting cultures was stamped into a well of YEP + 2% raffinose + 2% galactose + 1 mM sodium butyrate (C4). Samples were incubated at 30°C in a shaking incubator for 4 days. A portion of each of the resulting production cultures was stamped into a well of PBS. Optical measurements were taken on a plate reader, with absorbance measured at 600 nm and fluorescence at 528 nm with 485 nm excitation. A portion of each of the production cultures was stamped into a well of 100% methanol in half-height deepwell plates. Plates were heat sealed and frozen at - 80°C for two hours. Samples were then thawed and spun down. A portion of the supernatant was stamped into half-area 96 well plates. Divaric acid (DA) and divarinol (DL) production in the samples were quantified via LC-MS.

[440] 265 PKS mutants were selected from this high-throughput screen based on levels of production of DA by strains expressing the PKS mutants. To confirm the activity of these PKS mutants, a secondary screen was performed to verify and further quantify DA and DL production. The experimental protocols for the secondary screen were the same as for the primary screen except that additional biological replicates were included per strain. All strains were screened in quadruplicate. Strain t896773, also described above, expressing a *C. sativa* PKS, was used as a positive control to establish hit ranking. Strain t896591, also described above, expressing a GFP fluorescent reporter, was used as a negative control.

[441] 166 strains expressing PKS mutants assayed in the secondary screen produced on average more DA than the positive control strains t896773. The top 10% of strains shown in FIGs. 6A-6B demonstrate an average DA and DL titer $\geq 50\%$ greater than that of the positive control strain t896773. Results for strain t896591, expressing GFP as a negative control, are also shown.

[442] PKS mutants assayed in the secondary screen that produced on average more DA than the positive control strains t896773 comprised a contiguous stretch of residues with amino acids corresponding to positions 12-20 in SEQ ID NO: 1; positions 26-37 in SEQ ID NO: 1; positions 39-43 in SEQ ID NO: 1; positions 45-50 in SEQ ID NO: 1; positions 52-61 in SEQ ID NO: 1; positions 83-91 in SEQ ID NO: 1; positions 93-100 in SEQ ID NO: 1; positions 105-122 in SEQ ID NO: 1; positions 129-135 in SEQ ID NO: 1; positions 145-149 in SEQ ID NO: 1; positions 151-165 in SEQ ID NO: 1; positions 167-189 in SEQ ID NO: 1; positions 191-224 in SEQ ID NO: 1; positions 226-236 in SEQ ID NO: 1; positions 238-246 in SEQ ID NO: 1; positions 248-261 in SEQ ID NO: 1; positions 268-277 in SEQ ID NO: 1; positions 279-295 in SEQ ID NO: 1; positions 302-311 in SEQ ID NO: 1; positions 320-327 in SEQ ID NO: 1; positions 336-374 in SEQ ID NO: 1; and/or positions 376-383 in SEQ ID NO: 1.

Table 3. Activity data for synthetic PKSs in *S. cerevisiae* described in Example 1 and FIGs. 6A-6B

Strain	Amino Acid Substitutions relative to SEQ ID NO: 1	t896773 normalized Divaric Acid (μg/L) Average	t896773 normalized Divaric Acid (μg/L) Standard Deviation	t896773 normalized Divarinol (μg/L) Average	t896773 normalized Divarinol (μg/L) Standard Deviation
t896591	N/A (GFP control)	0	0	0	0
t896773	N/A (wild-type)	1	0.159	1	0.141
t896290	R75N; D101E; V319R; H328Y; T335C	3.292	0.182	2.992	0.122
t896396	C63M; H69I; R75G; V77C; M267G; H312G; V319K; T335C; R375T; Y385S	3.288	0.312	2.680	0.180
t896586	C63M; F64Y; S124T; H328Y; T335C	1.723	0.047	2.569	0.002
t896709	K51R; H69I; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	2.359	0.249	2.492	0.157
t896258	C63T; Q82P; D101E; C103A; T123S	2.433	0.225	2.359	0.110
t896561	L25I; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	2.274	0.359	2.311	0.184
t896517	K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; D316E; T335C	2.051	0.171	2.256	0.196
t896676	R75N; V77C; H79Y; D101E; C103A; T128V; K301P; H312G; D316E; T335C	1.927	0.081	2.143	0.079
t896360	K51R; N66T; R75S; D101E; T123C; T296A; S315P; H328Y; T335C; R375T	2.577	0.259	2.110	0.345

t896493	K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; D316E; T335C	1.890	0.105	2.058	0.126
t897093	H3R; S11T; K38N; Q44D; K51R; H69I; Q72E; R75N; L76M; V77C; E78T; H79Y; M81D; V92K; D101E; C103A; A104E; T128V; C137L; S144P; V150T; V225A; H262S; I278K; H312G; S315P; D316E; S334P; K384N	2.293	0.254	2.018	0.204
t896294	K51R; R75N; V77C; H79Y; D101E; C103A; T128V; K301P; H312G; D316E	2.108	0.146	1.994	0.022
t896188	L25I; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	2.220	0.162	1.980	0.154
t896324	L4Q; E21P; N62H; H312N; H328Y	2.106	0.202	1.955	0.168
t896386	L25I; H69I; R75N; V77C; H79Y; D101E; C103A; T128V; K301P; H312G	2.144	0.293	1.952	0.297
t896701	H3A; L25Y; C63M; R75N; H136Q; M267G; H328Y; T335C; R375T; Y385T	1.998	0.251	1.903	0.222
t896451	K51R; H69I; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	1.896	0.148	1.897	0.082
t896801	D101E; T123C; H136Q; H312G; T335C	2.318	0.145	1.883	0.116
t897022	H3R; E7P; S11T; K38N; Q44D; K51R; H69I; Q72E; R75N; L76M; V77C; E78T; H79Y; M81D; V92A; D101E; C103A; A104E; T128V; C137L; S144P; V150T; V225A; H262S; I278E; H312G; S315P; D316E; S334P; K384N	1.916	0.110	1.827	0.094
t897014	K51R; R75N; V77C; H79Y; D101E; C103A; T128V; H312G; D316E; T335C	1.777	0.078	1.794	0.035
t896793	H3A; L4Q; C63F; R75N; D101E; C103A; C137L; G237A; T247A; Y385N	1.844	0.240	1.747	0.133
t896232	L4Q; L24I; Q44D; K51R; C63F; R75S; V77C; C103A; I166L; V319A	1.783	0.123	1.716	0.077
t896392	L190C	2.176	0.133	1.705	0.103
t896962	L4Q; R75S; V319R; T335C; R375S	1.806	0.113	1.701	0.060
t896580	A125S	1.675	0.286	1.654	0.204
t896509	R75S; V77C; S124T; T335C; R375T	1.662	0.093	1.633	0.095
t896223	H3A; R75S; D101E; S315K; T335C	1.667	0.062	1.614	0.025

Example 2: Functional characterization by mutagenesis at residue 25 of synthetic polyketide synthase (PKS) mutants in *S. cerevisiae*

[443] Several of the top strains identified in Example 1 expressed a PKS that comprised the amino acid substitution L25I. To examine the impact of amino acid substitutions at this position and the plasticity of their effects on cannabinoid metabolite production, two PKS mutants were selected for mutagenesis of position 25. These two PKS mutants were SEQ ID NO: 7 (expressed in strain t896561), which comprises the following mutations relative to SEQ ID NO: 1: L25I, K51R, R75N, V77C, D101E, C103A, T128V, K301P, H312G, and T335C, and SEQ ID NO: 14 (expressed in strain t896188), which comprises the following mutations relative to SEQ ID NO: 1: and L25I, K51R, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, and T335C. The genes in the site saturation mutagenesis library were codon-optimized for expression in *S. cerevisiae* and integrated into the host genome. Each PKS mutant was transformed into an *S. cerevisiae* CEN.PK strain also expressing an AAE from *C. arbutum* (SEQ ID NO: 86), a synthetic polyketide cyclase (SEQ ID NO: 61), and a synthetic prenyltransferase (SEQ ID NO: 102). Strain t1151829, expressing the synthetic PKS (SEQ ID NO: 7) expressed in strain t896561, was used as a positive control. Strain t1151830, expressing the synthetic PKS (SEQ ID NO: 14) expressed in strain t896188, was also used as a positive control.

[444] An assay for measuring production of divaric acid, divarinol, and cannabigerovarinic acid (CBGVA) was conducted as described above. Results are shown in Tables 4-5 and FIGs. 7A-7B. Among strains comprising variants of SEQ ID NO: 7 expressed in strain t1151829, eleven demonstrated greater average CBGVA titers relative to the t1151829 parent strain expressing SEQ ID NO: 7. The amino acid substitutions at position 25 of SEQ ID NO: 7 that increased CBGVA production were: E, H, K, L, M, N, P, R, T, W, and Y. Similarly, among strains comprising variants of SEQ ID NO: 14 expressed in strain t1151830, ten demonstrated greater average CBGVA titers relative to strain t1151830 expressing SEQ ID NO: 14. The amino acid substitutions at position 25 of SEQ ID NO: 14 that increased CBGVA production were: E, G, K, M, N, P, Q, S, V, and Y. This suggests that there is plasticity in the identity of residue 25 and that substitution of position 25 in SEQ ID NO: 1 with an amino acid other than isoleucine may be used to increase cannabinoid production.

Table 4. Activity data for synthetic PKSs in *S. cerevisiae* described in Example 2 and FIG. 7A

Strain	Amino Acid Substitutions Relative To SEQ ID NO: 1	Divarinol [μmol]		Divaric Acid [μmol]		CBGVA [μmol]		Total Activity* [μmol]
		AVG	STDEV	AVG	STDEV	AVG	STDEV	
t1212770	L25E; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	25592.2	2697.3	33836.7	3570.6	40124.8	3932.2	99553.7
t1212780	L25H; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	27749.6	4345.1	37596.9	5351.2	39521.8	4915.0	104868.3
t1212773	L25K; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	26117.6	4000.7	37783.9	8218.1	39459.6	5068.5	103361.1
t1212784	K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	27086.7	2266.8	37673.9	4713.6	38517.7	882.0	103278.3
t1212785	L25R; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	26574.3	2854.8	36675.6	3772.2	37385.6	1315.5	100635.5
t1212786	L25W; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	25809.8	2498.9	35536.4	4069.1	36958.9	790.1	98305.1
t1212783	L25N; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	24382.4	1810.3	34222.7	3965.3	36446.7	1224.2	95051.8
t1212767	L25M; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	23807.5	2243.1	33146.8	4022.5	36358.1	1643.0	93312.4
t1212790	L25T; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	23764.8	1460.0	32047.9	2266.7	36105.5	1507.7	91918.2
t1212761	L25Y; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	23042.2	2886.4	33407.5	5384.7	35385.7	2402.1	91835.4

t1212789	L25P; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	24558.9	2049.5	34563.5	2439.0	35291.9	1739.7	94414.3
t1151829 (positive control)	L25I; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	22840.3	631.5	34159.8	3314.5	35039.9	1336.6	92040.0
t1212792	L25Q; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	23847.8	3720.2	34687.1	4929.3	33889.5	4122.0	92424.4
t1212787	L25G; K51R; R75N; V77C; D101E; C103A; T128V; K301P; H312G; T335C	15770.8	9329.8	20757.6	13030.0	25227.9	11788.8	61756.3

*Combined production of divarinol, divaric acid and CBGVA

Table 5. Activity data for synthetic PKSs in *S. cerevisiae* described in Example 2 and FIG.

7B

Strain	Amino Acid Substitutions Relative To SEQ ID NO: 1	Divarinol [μmol]		Divaric Acid [μmol]		CBGVA [μmol]		Total Activity* [μmol]
		AVG	STDEV	AVG	STDEV	AVG	STDEV	
t1212765	L25P; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	18928.8	2229.8	24874.6	2681.5	34128.0	3376.2	77931.4
t1212795	L25S; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	18531.7	1690.3	26217.7	2880.7	29192.3	205.1	73941.7
t1212796	L25N; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	17133.5	806.3	25168.0	1523.1	29055.8	506.1	71357.3
t1212793	L25M; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	16389.1	334.5	24114.1	677.2	29033.4	654.3	69536.6
t1212779	L25Q; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	17532.4	1294.5	23136.4	1851.5	28920.6	840.7	69589.4

t1212774	L25K; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	17478.1	2078.5	23368.9	2381.0	28466.1	897.3	69313.1
t1212781	L25E; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	16708.7	2210.4	22261.0	2125.2	28354.0	1017.8	67323.7
t1212768	L25G; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	17282.6	1449.6	23460.5	1289.8	28234.7	638.9	68977.8
t1212794	L25Y; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	16817.0	2527.5	23757.0	3820.8	28111.8	2255.6	68685.8
t1212771	L25V; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	16808.6	1441.7	22542.0	1692.2	27955.6	922.9	67306.2
t1151830	L25I; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	15589.1	945.6	24091.5	1776.7	27780.3	396.1	67460.9
t1212777	L25F; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	16319.5	963.7	21837.6	1635.3	27586.4	859.7	65743.5
t1212778	L25C; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	16282.9	1980.7	21347.2	3194.2	27448.2	476.5	65078.3
t1212776	L25W; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	15657.6	3116.1	20335.2	3534.3	25105.4	3867.4	61098.2
t1212788	L25R; K51R; R75N; V77C; H79Y; D101E; C103A; K301P; H312G; T335C	14951.0	3740.5	19900.9	5787.5	24731.9	5342.1	59583.8

*Combined production of divarinol, divaric acid and CBGVA

Table 6. Non-limiting examples of PKS amino acid substitutions relative to SEQ ID NO: 1

Position in SEQ ID NO: 1	Amino acid substitutions
--------------------------------	--------------------------

3	H3R	H3A					
4	L4Q						
7	E7P						
11	S11T						
21	E21P						
24	L24I						
25	L25C	L25E	L25F	L25G	L25H	L25K	L25M
	L25N	L25P	L25Q	L25R	L25S	L25T	L25V
	L25W	L25Y					
38	K38N						
44	Q44D						
51	K51R						
62	N62H						
63	C63M	C63T	C63F				
64	F64Y						
66	N66T						
69	H69I						
72	Q72E						
75	R75N	R75S	R75G				
76	L76M						
77	V77C						
78	E78T						
79	H79Y						
81	M81D						
82	Q82P						
92	V92K	V92A					
101	D101E						
103	C103A						
104	A104E						
123	T123C	T123S					
124	S124T						
125	A125S						
128	T128V						
136	H136Q						
137	C137L						
144	S144P						
150	V150T						
166	I166L						
190	L190C						
225	V225A						
237	G237A						
247	T247A						

262	H262S						
267	M267G						
278	I278K	I278E					
296	T296A						
301	K301P						
312	H312N	H312G					
315	S315K	S315P					
316	D316E						
319	V319R	V319K	V319A				
328	H328Y						
334	S334P						
335	T335C						
375	R375S	R375T					
384	K384N						
385	Y385S	Y385T	Y385N				

Table 7. Amino acid and nucleic acid sequences of PKSs

Strain	PKS Protein SEQ ID NO:	PKS Nucleic Acid SEQ ID NO:
t896773	1	29
t896290	2	30
t896396	3	31
t896586	4	32
t896709	5	33
t896258	6	34
t896561	7	35
t896517	8	36
t896676	9	37
t896360	10	38
t896493	11	39
t897093	12	40
t896294	13	41
t896188	14	42
t896324	15	43
t896386	16	44
t896701	17	45
t896451	18	46
t896801	19	47
t897022	20	48
t897014	21	49
t896793	22	50
t896232	23	51
t896392	24	52
t896962	25	53
t896580	26	54
t896509	27	55
t896223	28	56
t1212770	104	131

t1212780	105	132
t1212773	106	133
t1212784	107	134
t1212785	108	135
t1212786	109	136
t1212783	110	137
t1212767	111	138
t1212790	112	139
t1212761	113	140
t1212789	114	141
t1151829	7	35
t1212765	115	142
t1212792	116	143
t1212795	117	144
t1212796	118	145
t1212793	119	146
t1212779	120	147
t1212774	121	148
t1212781	122	149
t1212768	123	150
t1212794	124	151
t1212771	125	152
t1151830	14	42
t1212777	126	153
t1212778	127	154
t1212787	128	155
t1212776	129	156
t1212788	130	157

[445] It should be appreciated that sequences disclosed in this application may or may not contain signal sequences. The sequences disclosed in this application encompass versions with or without signal sequences. It should also be understood that protein sequences disclosed in this application may be depicted with or without a start codon (M). The sequences disclosed in this application encompass versions with or without start codons. Accordingly, in some instances amino acid numbering may correspond to protein sequences containing a start codon, while in other instances, amino acid numbering may correspond to protein sequences that do not contain a start codon. It should also be understood that sequences disclosed in this application may be depicted with or without a stop codon. The sequences disclosed in this application encompass versions with or without stop codons. Aspects of the disclosure encompass host cells comprising any of the sequences described in this application and fragments thereof.

EQUIVALENTS

[446] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described here. Such equivalents are intended to be encompassed by the following claims.

[447] All references, including patent documents, are incorporated by reference in their entirety.

CLAIMS

1. A host cell that comprises a heterologous polynucleotide encoding a polyketide synthase (PKS), wherein the host cell is capable of producing more divarinol in the presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.
2. The host cell of claim 1, wherein the PKS comprises:
 - (a) an amino acid substitution at a residue corresponding to position 3, 4, 7, 11, 21, 24, 38, 44, 62, 66, 69, 72, 76, 77, 78, 79, 101, 103, 104, 124, 136, 137, 144, 150, 166, 225, 237, 262, 301, 319, and/or 385 in SEQ ID NO: 1; and
 - (b) a contiguous stretch of residues with amino acids unmutated relative to SEQ ID NO: 1 at positions corresponding to:
 - i. positions 12-20 in SEQ ID NO: 1;
 - ii. positions 26-37 in SEQ ID NO: 1;
 - iii. positions 39-43 in SEQ ID NO: 1;
 - iv. positions 45-50 in SEQ ID NO: 1;
 - v. positions 52-61 in SEQ ID NO: 1;
 - vi. positions 83-91 in SEQ ID NO: 1;
 - vii. positions 93-100 in SEQ ID NO: 1;
 - viii. positions 105-122 in SEQ ID NO: 1;
 - ix. positions 129-135 in SEQ ID NO: 1;
 - x. positions 145-149 in SEQ ID NO: 1;
 - xi. positions 151-165 in SEQ ID NO: 1;
 - xii. positions 167-189 in SEQ ID NO: 1;
 - xiii. positions 191-224 in SEQ ID NO: 1;
 - xiv. positions 226-236 in SEQ ID NO: 1;
 - xv. positions 238-246 in SEQ ID NO: 1;
 - xvi. positions 248-261 in SEQ ID NO: 1;
 - xvii. positions 268-277 in SEQ ID NO: 1;
 - xviii. positions 279-295 in SEQ ID NO: 1;
 - xix. positions 302-311 in SEQ ID NO: 1;
 - xx. positions 320-327 in SEQ ID NO: 1;
 - xxi. positions 336-374 in SEQ ID NO: 1; and/or

xxii. positions 376-383 in SEQ ID NO: 1.

3. A host cell comprising a heterologous polynucleotide encoding a polyketide synthase (PKS), wherein the PKS comprises an amino acid substitution at a residue corresponding to position 3, 4, 7, 11, 21, 24, 38, 44, 62, 66, 69, 72, 76, 77, 78, 79, 101, 103, 104, 124, 136, 137, 144, 150, 166, 225, 237, 262, 301, 319, and/or 385 in SEQ ID NO: 1, and wherein the PKS does not comprise any of SEQ ID NOs: 91-94.

4. The host cell of any one of claims 1-3, wherein the PKS comprises one of more of the following substitutions relative to SEQ ID NO: 1:

- (a) H3R or H3A;
- (b) L4Q;
- (c) E7P;
- (d) S11T;
- (e) E21P;
- (f) L24I;
- (g) K38N;
- (h) Q44D;
- (i) N62H;
- (j) N66T;
- (k) H69I;
- (l) Q72E;
- (m) L76M;
- (n) V77C;
- (o) E78T;
- (p) H79Y;
- (q) D101E;
- (r) C103A;
- (s) A104E;
- (t) S124T;
- (u) H136Q;
- (v) C137L;
- (w) S144P;

- (x) V150T;
- (y) I166L;
- (z) V225A;
- (aa) G237A;
- (bb) H262S;
- (cc) K301P;
- (dd) V319R, V319K, or V319A; and/or
- (ee) Y385S, Y385T, or Y385N.

5. The host cell of any one of claims 1-4, wherein the PKS further comprises an amino acid substitution at a residue corresponding to position 25, 51, 63, 64, 75, 81, 82, 92, 123, 125, 128, 190, 247, 267, 278, 296, 312, 315, 316, 328, 334, 335, 375, and/or 384 in SEQ ID NO: 1.

6. The host cell of claim 5, wherein the PKS comprises one or more of the following amino acid substitutions relative to SEQ ID NO: 1:

- (a) L25C, L25E, L25F, L25G, L25H, L25K, L25M, L25N, L25P, L25Q, L25R, L25S, L25T, L25V, L25W, or L25Y;
- (b) K51R;
- (c) C63M, C63T, or C63F;
- (d) F64Y;
- (e) R75N, R75S, or R75G;
- (f) M81D;
- (g) Q82P;
- (h) V92K or V92A;
- (i) T123C or T123S;
- (j) A125S;
- (k) T128V;
- (l) L190C;
- (m) T247A;
- (n) M267G;
- (o) I278K or I278E;
- (p) T296A;

- (q) H312N or H312G;
- (r) S315K or S315P;
- (s) D316E;
- (t) H328Y;
- (u) S334P;
- (v) T335C;
- (w) R375S or R375T; and/or
- (x) K384N.

7. The host cell of claim 5, wherein the PKS does not comprise the amino acid I at the residue corresponding to position 25 in SEQ ID NO: 1.

8. The host cell of claim 7, wherein the PKS comprises the amino acid glutamate (E), glycine (G), histidine (H), lysine (K), leucine (L), methionine (M), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S), threonine (T), valine (V), tryptophan (W), or a tyrosine (Y) at the residue corresponding to position 25 in SEQ ID NO: 1.

9. The host cell of any one of claims 1-8, wherein the PKS comprises:

- (a) the amino acid cysteine (C) at a residue corresponding to position 157 in SEQ ID NO: 1;
- (b) the amino acid phenylalanine (F) at a residue corresponding to position 208 in SEQ ID NO: 1;
- (c) the amino acid histidine (H) at a residue corresponding to position 297 in SEQ ID NO: 1; and/or
- (d) the amino acid asparagine (N) at a residue corresponding to position 330 in SEQ ID NO: 1.

10. The host cell of any one of claims 1-9, wherein the PKS comprises the following amino acid substitutions relative to SEQ ID NO: 1:

- (a) R75N, D101E, V319R, H328Y, and T335C;
- (b) C63M, H69I, R75G, V77C, M267G, H312G, V319K, T335C, R375T, and Y385S;
- (c) C63M, F64Y, S124T, H328Y, and T335C;
- (d) K51R, H69I, R75N, V77C, D101E, C103A, T128V, K301P, H312G, and T335C;

- (e) C63T, Q82P, D101E, C103A, and T123S;
- (f) L25I, K51R, R75N, V77C, D101E, C103A, T128V, K301P, H312G, and T335C;
- (g) K51R, R75N, V77C, D101E, C103A, T128V, K301P, H312G, D316E, and T335C;
- (h) R75N, V77C, H79Y, D101E, C103A, T128V, K301P, H312G, D316E, and T335C;
- (i) K51R, N66T, R75S, D101E, T123C, T296A, S315P, H328Y, T335C, and R375T;
- (j) K51R, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, D316E, and T335C;
- (k) H3R, S11T, K38N, Q44D, K51R, H69I, Q72E, R75N, L76M, V77C, E78T, H79Y, M81D, V92K, D101E, C103A, A104E, T128V, C137L, S144P, V150T, V225A, H262S, I278K, H312G, S315P, D316E, S334P, and K384N;
- (l) K51R, R75N, V77C, H79Y, D101E, C103A, T128V, K301P, H312G, and D316E;
- (m) L25I, K51R, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, and T335C;
- (n) L4Q, E21P, N62H, H312N, and H328Y;
- (o) L25I, H69I, R75N, V77C, H79Y, D101E, C103A, T128V, K301P, and H312G;
- (p) H3A, L25Y, C63M, R75N, H136Q, M267G, H328Y, T335C, R375T, and Y385T;
- (q) K51R, H69I, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, and T335C;
- (r) D101E, T123C, H136Q, H312G, and T335C;
- (s) H3R, E7P, S11T, K38N, Q44D, K51R, H69I, Q72E, R75N, L76M, V77C, E78T, H79Y, M81D, V92A, D101E, C103A, A104E, T128V, C137L, S144P, V150T, V225A, H262S, I278E, H312G, S315P, D316E, S334P, and K384N;
- (t) K51R, R75N, V77C, H79Y, D101E, C103A, T128V, H312G, D316E, and T335C;
- (u) H3A, L4Q, C63F, R75N, D101E, C103A, C137L, G237A, T247A, and Y385N;
- (v) L4Q, L24I, Q44D, K51R, C63F, R75S, V77C, C103A, I166L, and V319A;
- (w) L4Q, R75S, V319R, T335C, and R375S;
- (x) R75S, V77C, S124T, T335C, and R375T; or
- (y) H3A, R75S, D101E, S315K, and T335C.

11. The host cell of any one of claims 1-10, wherein the PKS comprises a sequence that is at least 90% identical to a sequence selected from SEQ ID NOs: 2-28.

12. The host cell of any one of claims 1-11, wherein the PKS comprises a sequence selected from SEQ ID NOs: 2-28, or a conservatively substituted version thereof.

13. The host cell of any one of claims 1-12, wherein the heterologous polynucleotide comprises a sequence that is at least 90% identical to a sequence selected from SEQ ID NOs: 30-56.

14. The host cell of any one of claims 1-13, wherein the heterologous polynucleotide comprises a sequence selected from SEQ ID NOs: 30-56.

15. The host cell of any one of claims 1-9, wherein the PKS comprises the following amino acid substitutions relative to SEQ ID NO: 1: K51R, R75N, V77C, H79Y, D101E, C103A, K301P, H312G, and T335C.

16. The host cell of claim 15, wherein the PKS comprises an amino acid substitution relative to SEQ ID NO: 1 selected from the group consisting of:

- (a) L25E;
- (b) L25G;
- (c) L25K;
- (d) L25M;
- (e) L25N;
- (f) L25P;
- (g) L25Q;
- (h) L25S;
- (i) L25V; and
- (j) L25Y.

17. The host cell of any one of claims 1-9, wherein the PKS comprises the following amino acid substitutions relative to SEQ ID NO: 1: K51R, R75N, V77C, D101E, C103A, T128V, K301P, H312G, and T335C.

18. The host cell of claim 17, wherein the PKS comprises an amino acid substitution relative to SEQ ID NO: 1 selected from the group consisting of:

- (a) L25E;
- (b) L25H;
- (c) L25K;

- (d) L25M;
- (e) L25N;
- (f) L25P;
- (g) L25R;
- (h) L25T;
- (i) L25W; and
- (j) L25Y.

19. The host cell of any one of claims 15-18, wherein the PKS comprises a sequence that is at least 90% identical to a sequence selected from SEQ ID NOs: 104-130.

20. The host cell of any one of claims 15-19, wherein the PKS comprises a sequence selected from SEQ ID NOs: 104-130.

21. The host cell of any one of claims 15-20, wherein the heterologous polynucleotide encoding the PKS comprises a sequence that is at least 90% identical to a sequence selected from SEQ ID NOs: 131-157.

22. The host cell of any one of claims 15-21, wherein the heterologous polynucleotide encoding the PKS comprises a sequence selected from SEQ ID NOs: 131-157.

23. The host cell of any one of claims 1-22, wherein the heterologous polynucleotide is integrated into the genome of the host cell.

24. The host cell of any one of claims 1-23, wherein the host cell is a plant cell, an algal cell, a yeast cell, a bacterial cell, or an animal cell.

25. The host cell of claim 24, wherein the host cell is a yeast cell.

26. The host cell of claim 25, wherein the yeast cell is a *Saccharomyces* cell, a *Yarrowia* cell, a *Komagataella* cell, or a *Pichia* cell.

27. The host cell of claim 26, wherein the *Saccharomyces* cell is a *Saccharomyces cerevisiae* cell.
28. The host cell of claim 26, wherein the yeast cell is a *Yarrowia* cell.
29. The host cell of claim 24, wherein the host cell is a bacterial cell.
30. The host cell of claim 29, wherein the bacterial cell is an *E. coli* cell.
31. The host cell of any one of claims 1-30, wherein the PKS is an olivetol synthase (OLS) or a divarinol synthase.
32. The host cell of any one of claims 1-31, wherein the host cell further comprises one or more heterologous polynucleotides encoding one or more of: an acyl activating enzyme (AAE), a polyketide cyclase (PKC), a prenyltransferase (PT) and/or a terminal synthase (TS).
33. The host cell of claim 32, wherein the PKC is an olivetolic acid cyclase (OAC) or divaric acid cyclase.
34. The host cell of claim 33, wherein the PKC comprises a sequence that is at least 90% identical to SEQ ID NO: 61.
35. The host cell of claim 34, wherein the PKC comprises SEQ ID NO: 61.
36. The host cell of any one of claims 1-35, wherein the host cell is capable of producing divaric acid in the presence of butyrate.
37. The host cell of any one of claims 1-36, wherein the host cell is capable of producing divarinol in the presence of butyrate.
38. The host cell of any one of claims 1-37, wherein the host cell is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17-fold more divaric acid in the

presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.

39. The host cell of any one of claims 1-38, wherein the host cell is capable of producing at least 500 µg/L divaric acid.

40. The host cell of claim 39, wherein the host cell is capable of producing at least 700 µg/L divaric acid.

41. The host cell of any one of claims 1-40, wherein the host cell is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, or 9-fold more divarinol in the presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.

42. The host cell of any one of claims 1-41, wherein the host cell is capable of producing at least 25,000 µg/L divarinol.

43. The host cell of any one of claims 1-42, wherein the host cell is capable of producing at least 30,000 µg/L divarinol.

44. The host cell of any one of claims 1-43, wherein the host cell is capable of producing at least 33,000 µg/L divarinol.

45. The host cell of any one of claims 1-44, wherein the host cell is capable of producing a varinolic cannabinoid.

46. The host cell of 45, wherein the varinolic cannabinoid is CBGV, CBGVA, THCVA, CBDVA and/or CBCVA.

47. The host cell of claim 45 or claim 46, wherein the host cell is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, or 9-fold more of the varinolic cannabinoid than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.

48. The host cell of any one of claims 32-47, wherein the AAE comprises a sequence that is at least 90% identical to SEQ ID NO: 86.
49. The host cell of claim 48, wherein the AAE comprises SEQ ID NO: 86.
50. The host cell of any one of claims 32-49, wherein the heterologous polynucleotide encoding the AAE comprises a sequence that is at least 90% identical to SEQ ID NO: 87.
51. The host cell of claim 50, wherein the heterologous polynucleotide encoding the AAE comprises SEQ ID NO: 87.
52. The host cell of any one of claims 32-51, wherein the PT comprises a sequence that is at least 90% identical to any one of SEQ ID NOs: 74, 76, 78, 80, 102, or 158.
53. The host cell of claim 52, wherein the PT comprises the sequence of any one of SEQ ID NOs: 74, 76, 78, 80, 102, or 158.
54. The host cell of any one of claims 32-53, wherein the TS comprises a sequence that is at least 90% identical to SEQ ID NO: 84, 88, 95, 98, 100, or 161.
55. The host cell of claim 54, wherein the TS comprises the sequence of SEQ ID NO: 84, 88, 95, 98, 100, or 161.
56. A host cell comprising a heterologous polynucleotide encoding a polyketide synthase (PKS), wherein the PKS comprises a sequence selected from SEQ ID NOs: 2-28 and 104-130 or a conservatively substituted version thereof.
57. A host cell comprising a heterologous polynucleotide encoding a polyketide synthase (PKS), wherein the heterologous polynucleotide comprises a sequence selected from SEQ ID NOs: 30-56 and 131-157.
58. A method comprising culturing the host cell of any one of claims 1-57.

59. A method of synthesizing divarinol or divaric acid comprising contacting butyryl-CoA with a polyketide synthase (PKS), wherein the PKS comprises an amino acid selected from any one of SEQ ID NOs: 2-28 and 104-130.
60. A method of producing divaric acid (DA) comprising contacting butyrate and Coenzyme A (CoA) with:
an acyl activating enzyme (AAE), a polyketide synthase (PKS), and a polyketide cyclase (PKC); or
an acyl activating enzyme (AAE) and a PKS,
wherein the PKS comprises the sequence of any one of SEQ ID NOs: 2-28 and 104-130.
61. A method of producing divarinol (DL) comprising contacting butyrate with an acyl activating enzyme (AAE) and a polyketide synthase (PKS), wherein the PKS comprises the sequence of any one of SEQ ID NOs: 2-28 and 104-130.
62. The method of any one of claims 59-61, wherein the method occurs *in vitro*.
63. The method of any one of claims 59-61, wherein the method occurs within a host cell that expresses a heterologous polynucleotide encoding a polyketide synthase (PKS) that comprises an amino acid sequence selected from any one of SEQ ID NOs: 2-28 and 104-130.
64. A method of producing a cannabinoid compound or a cannabinoid precursor comprising culturing a host cell in the presence of butyrate, wherein the host cell comprises a heterologous polynucleotide encoding a polyketide synthase (PKS), and wherein the PKS comprises an amino acid sequence selected from any one of SEQ ID NOs: 2-28 and 104-130.
65. The method of any one of claims 59-64, wherein the host cell further comprises one or more heterologous polynucleotides encoding an acyl activating enzyme (AAE).
66. The method of claim 65, wherein the AAE comprises a sequence that is at least 90% identical to the sequence of SEQ ID NO: 86.

67. The method of claim 65 or 66, wherein the heterologous polynucleotide encoding the AAE comprises a sequence that is at least 90% identical to the sequence of SEQ ID NO: 87.
68. The method of any one of claims 63-67, wherein the heterologous polynucleotide comprises a sequence that is at least 90% identical to the sequence of any one of SEQ ID NOs: 30-56 and 131-157.
69. The method of any one of claims 63-68, wherein the heterologous polynucleotide comprises the sequence of any one of SEQ ID NOs: 30-56 and 131-157.
70. The method of any one of claims 63-69, wherein the heterologous polynucleotide is integrated into the genome of the host cell.
71. The method of any one of claims 63-70, wherein the host cell is a plant cell, an algal cell, a yeast cell, a bacterial cell, or an animal cell.
72. The method of claim 71, wherein the host cell is a yeast cell.
73. The method of claim 72, wherein the yeast cell is a *Saccharomyces* cell, a *Yarrowia* cell, a *Komagataella* cell, or a *Pichia* cell.
74. The method of claim 73, wherein the *Saccharomyces* cell is a *Saccharomyces cerevisiae* cell.
75. The method of claim 73, wherein the yeast cell is a *Yarrowia* cell.
76. The method of claim 71, wherein the host cell is a bacterial cell.
77. The method of claim 76, wherein the bacterial cell is an *E. coli* cell.
78. The method of any one of claims 65-77, wherein the AAE is capable of activating a short chain fatty acid in the presence of Coenzyme A (CoA).

79. The method of claim 78, wherein the short chain fatty acid is a four-carbon fatty acid.
80. The method of claim 79, wherein the four-carbon fatty acid is butyrate and wherein the AAE is capable of catalyzing the production of butyryl-CoA from butyrate.
81. The method of any one of claims 63-80, wherein the host cell further comprises one or more heterologous polynucleotides encoding one or more of: a polyketide cyclase (PKC), a prenyltransferase (PT) and/or a terminal synthase (TS).
82. The method of any one of claims 59-81, wherein the PKS is an olivetol synthase (OLS) or a divarinol synthase.
83. The method of claim 82, wherein the PKS is a divarinol synthase.
84. The method of claim 83, wherein the PKS is a bifunctional PKS and is capable of producing divaric acid.
85. The method of claim 81, wherein the PKC is an olivetolic acid cyclase (OAC) or a divaric acid cyclase.
86. The method of claim 85, wherein the PKC comprises a sequence that is at least 90% identical to SEQ ID NO: 61.
87. The method of claim 86, wherein the PKC comprises SEQ ID NO: 61.
88. The method of any one of claims 63-87, wherein the host cell is capable of producing divaric acid.
89. The method of any one of claims 63-88, wherein the host cell is capable of producing divarinol.
90. The method of any one of claims 63-89, wherein the host cell is capable of producing at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17 fold more divaric acid in the

presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.

91. The method of any one of claims 63-90, wherein the host cell is capable of producing at least 500 µg/L divaric acid.

92. The method of any one of claims 63-91, wherein the host cell is capable of producing at least 700 µg/L divaric acid.

93. The method of any one of claims 63-92, wherein the host cell is capable of producing at least 2, 3, 4, 5, 6, 7, 8, or 9 fold more divarinol in the presence of butyrate than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.

94. The method of any one of claims 63-93, wherein the host cell is capable of producing at least 25,000 µg/L divarinol.

95. The method of any one of claims 63-94, wherein the host cell is capable of producing at least 30,000 µg/L divarinol.

96. The method of any one of claims 63-95, wherein the host cell is capable of producing at least 33,000 µg/L divarinol.

97. The method of any one of claims 63-96, wherein the host cell is capable of producing a varinolic cannabinoid.

98. The method of claim 97, wherein the varinolic cannabinoid is CBGV, CBGVA, THCVA and/or CBCVA.

99. The method of claim 97 or claim 98, wherein the host cell is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, or 9-fold more of the varinolic cannabinoid than a host cell that expresses a heterologous polynucleotide encoding a control PKS that comprises the sequence of SEQ ID NO: 1.

100. The method of any one of claims 81-99, wherein the PT comprises a sequence that is at least 90% identical to any one of SEQ ID NOs: 74, 76, 78, 80, 102, or 158.

101. The method of claim 100, wherein the PT comprises the sequence of any one of SEQ ID NOs: 74, 76, 78, 80, 102, or 158.

102. The method of any one of claims 81-101, wherein the TS comprises a sequence that is at least 90% identical to SEQ ID NO: 84, 88, 95, 98, 100, or 161.

103. The method of claim 102, wherein the TS comprises the sequence of SEQ ID NO: 84, 88, 95, 98, 100, or 161.

104. A non-naturally occurring polynucleotide encoding a polyketide synthase (PKS), wherein the polynucleotide comprises a sequence that is at least 90% identical to the sequence of any one of SEQ ID NOs: 30-56 and 131-157.

105. The non-naturally occurring polynucleotide of claim 104, wherein the polynucleotide comprises the sequence of any one of SEQ ID NOs: 30-56 and 131-157.

106. A vector comprising the polynucleotide sequence of claim 104 or claim 105.

107. An expression cassette comprising the polynucleotide sequence of claim 104 or claim 105.

108. A host cell transformed with the polynucleotide of claim 104 or claim 105, the vector of claim 106, or the expression cassette of claim 107.

109. A bioreactor for producing a cannabinoid compound or a cannabinoid precursor, wherein the bioreactor contains a polyketide synthase (PKS), and wherein the PKS comprises an amino acid sequence selected from any one of SEQ ID NOs: 2-28 and 104-130.

110. The bioreactor of claim 109, wherein the bioreactor further comprises one or more of an acyl activating enzyme (AAE), a polyketide cyclase (PKC), a prenyltransferase (PT) and/or a terminal synthase (TS).

111. The bioreactor of claim 109 or claim 110, wherein the bioreactor contains butyrate.

112. The bioreactor of any one of claims 109-111, wherein the bioreactor produces a varinolic cannabinoid.

113. The bioreactor of claim 112, wherein the varinolic cannabinoid is CBGV, CBGVA, THCVA, CBDVA and/or CBCVA.

114. The bioreactor of claim 112 or claim 113, wherein the bioreactor is capable of producing at least 1.5, 2, 3, 4, 5, 6, 7, 8, or 9-fold more of the varinolic cannabinoid than a bioreactor comprising a control PKS that comprises the sequence of SEQ ID NO: 1.

1/7

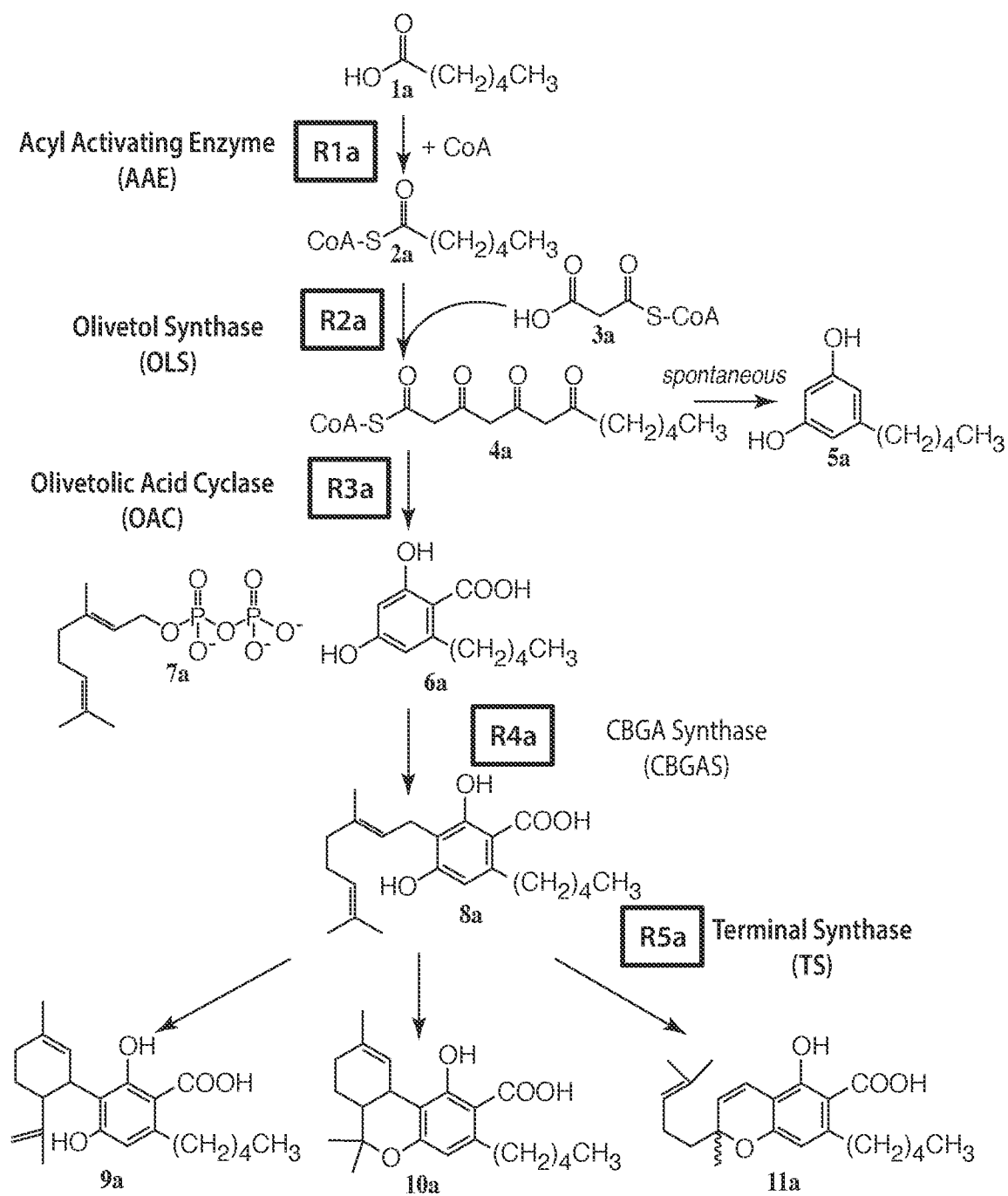


FIG. 1

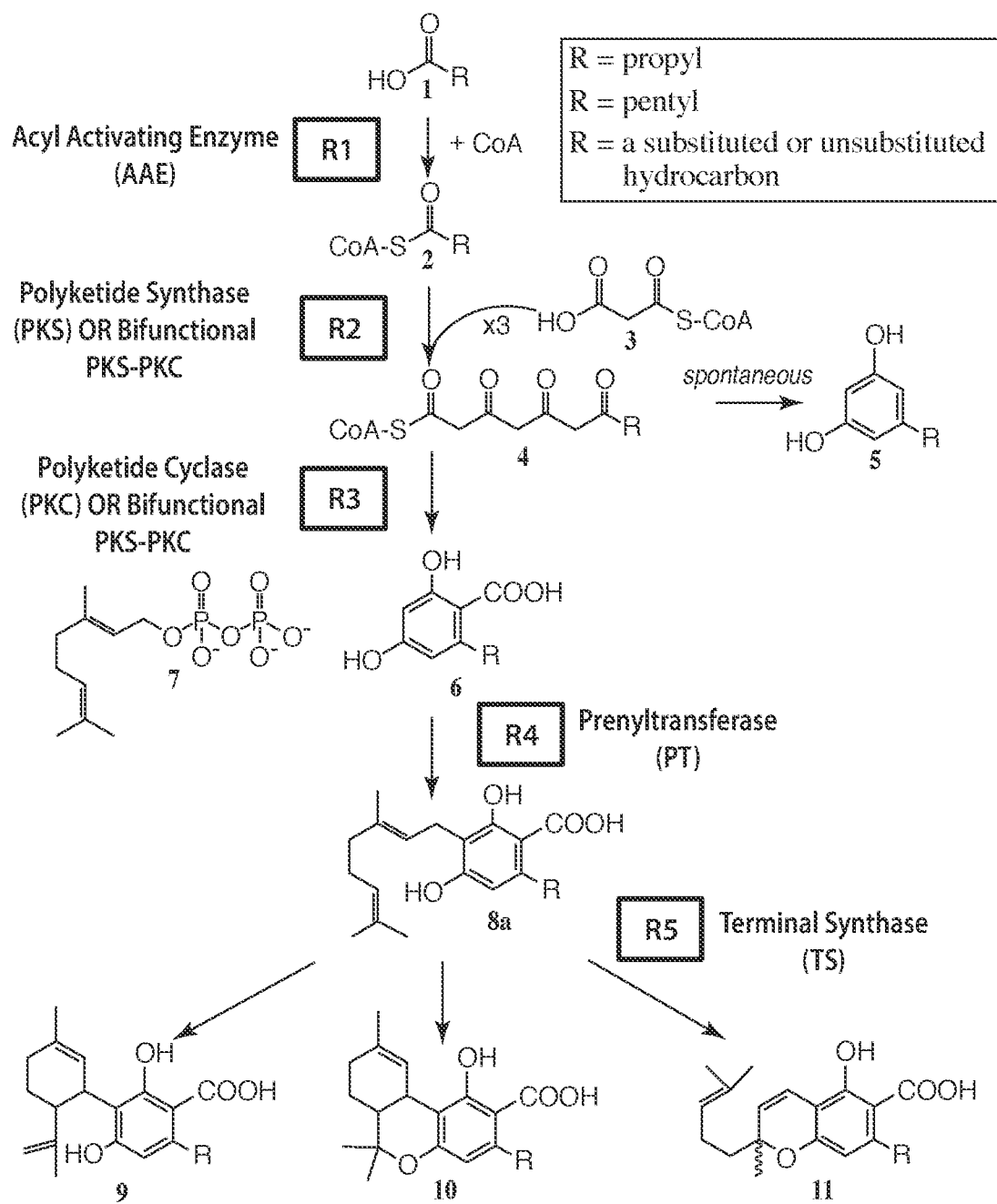


FIG. 2

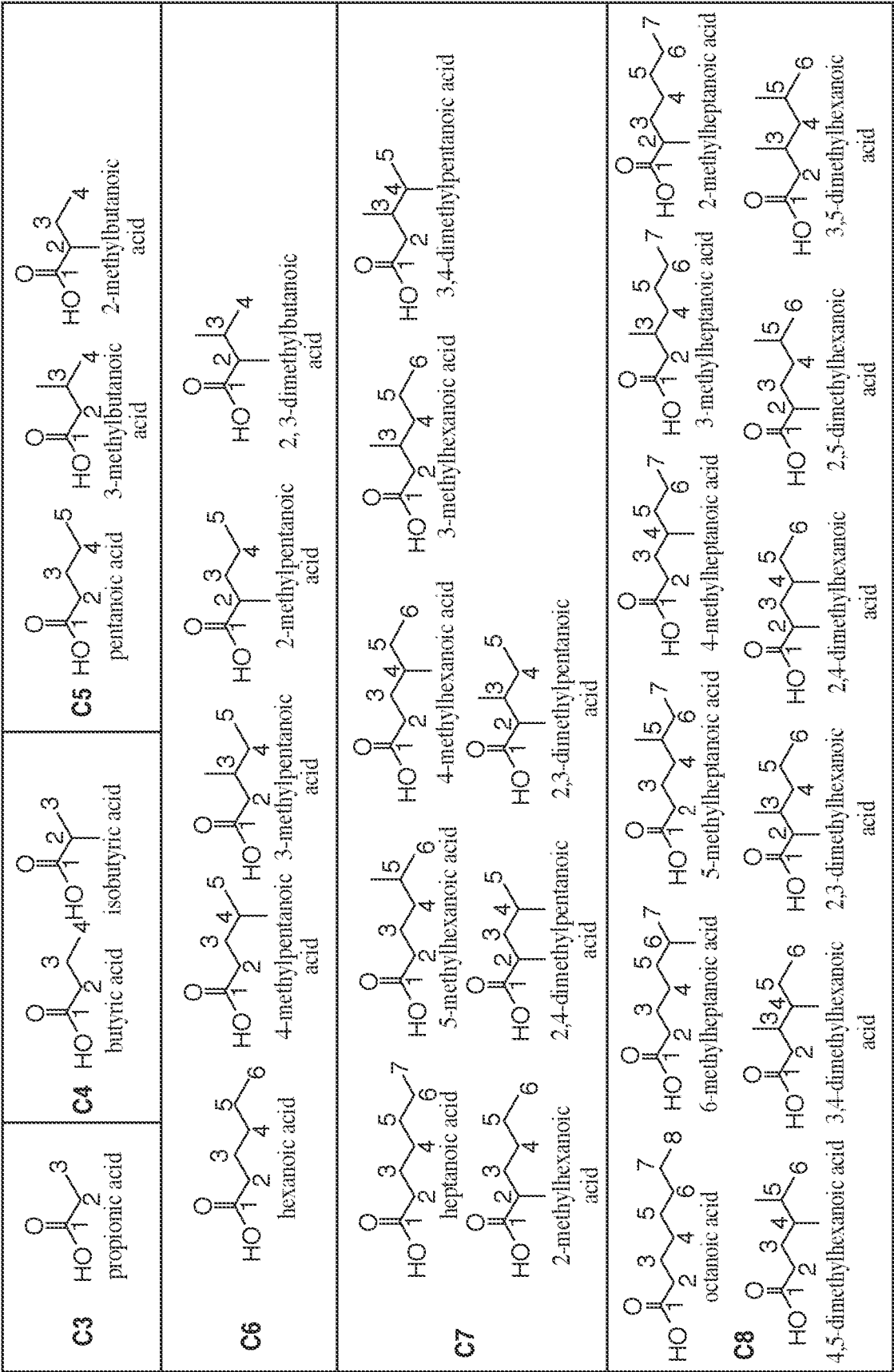


FIG. 3

4/7

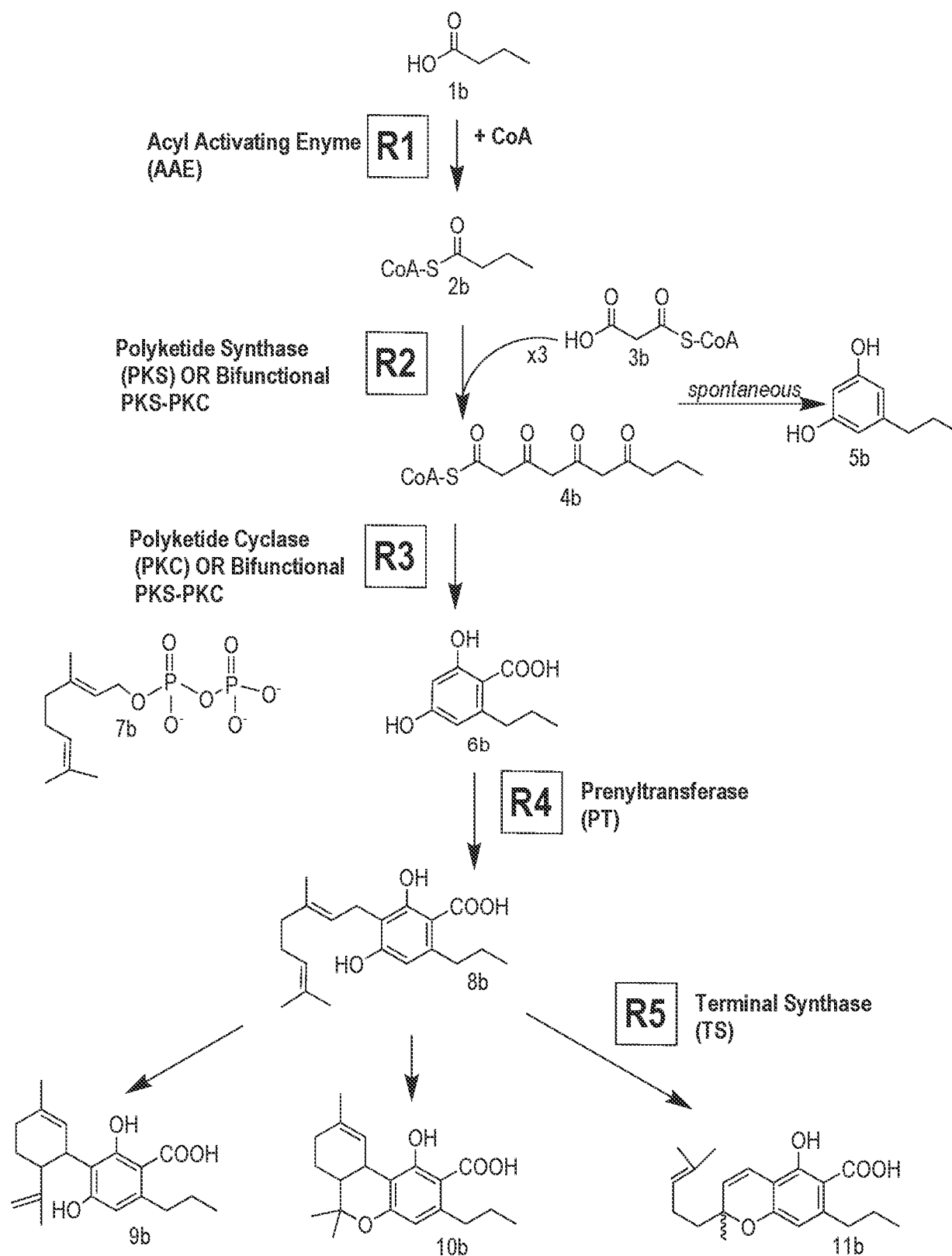


FIG. 4

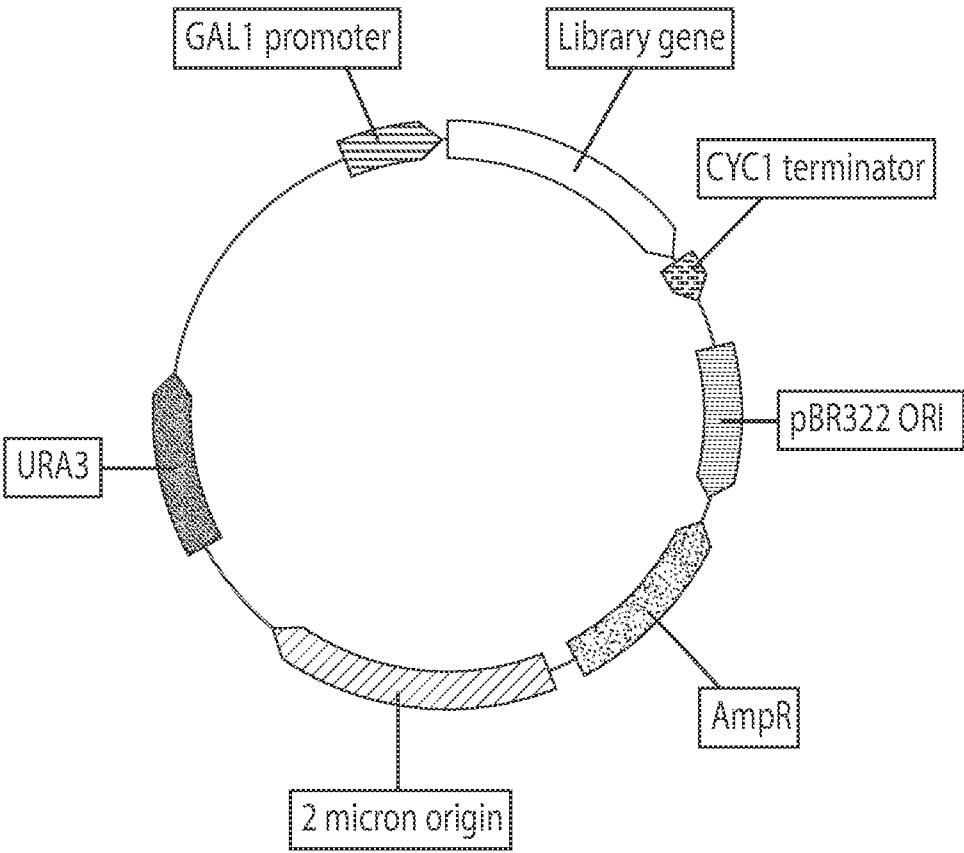


FIG. 5

6/7

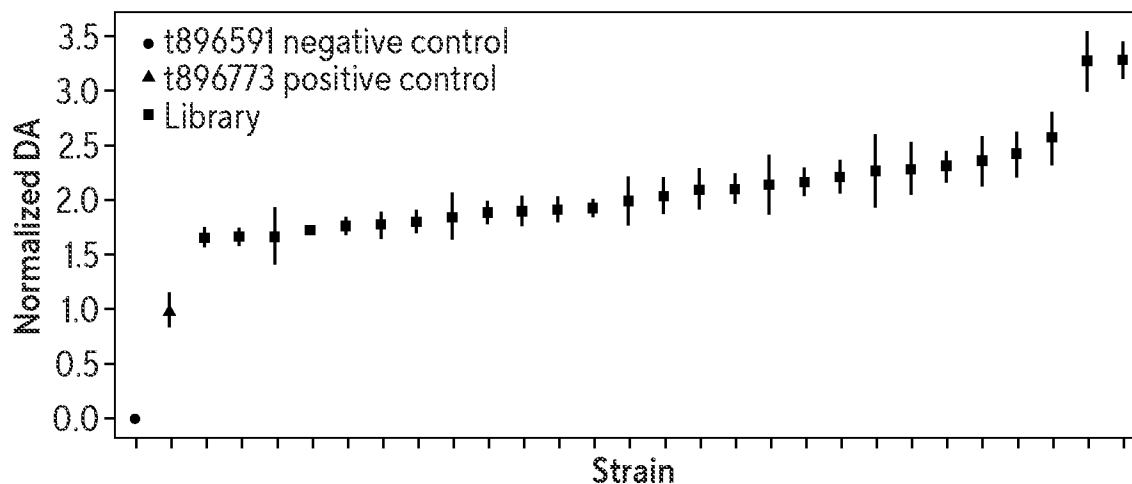


FIG. 6A

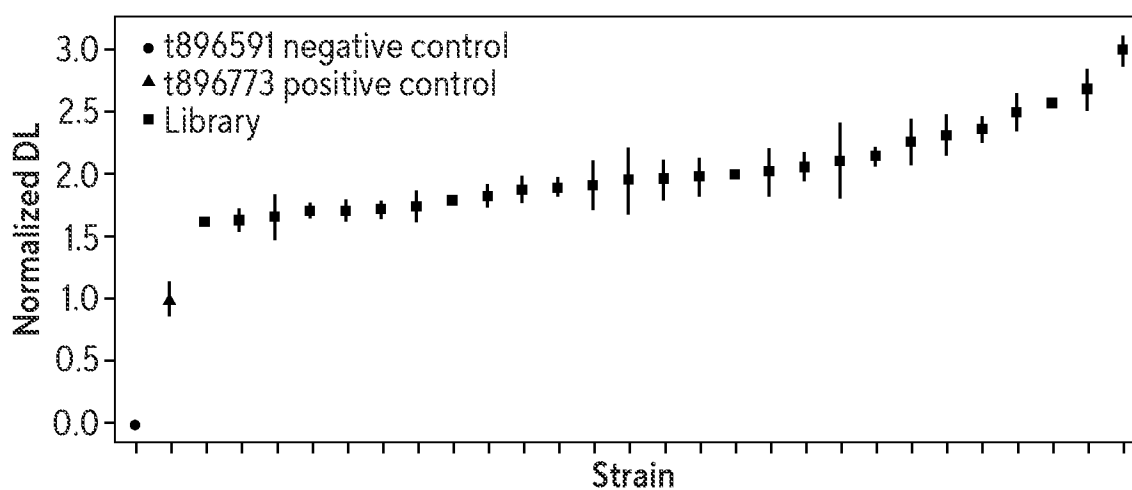


FIG. 6B

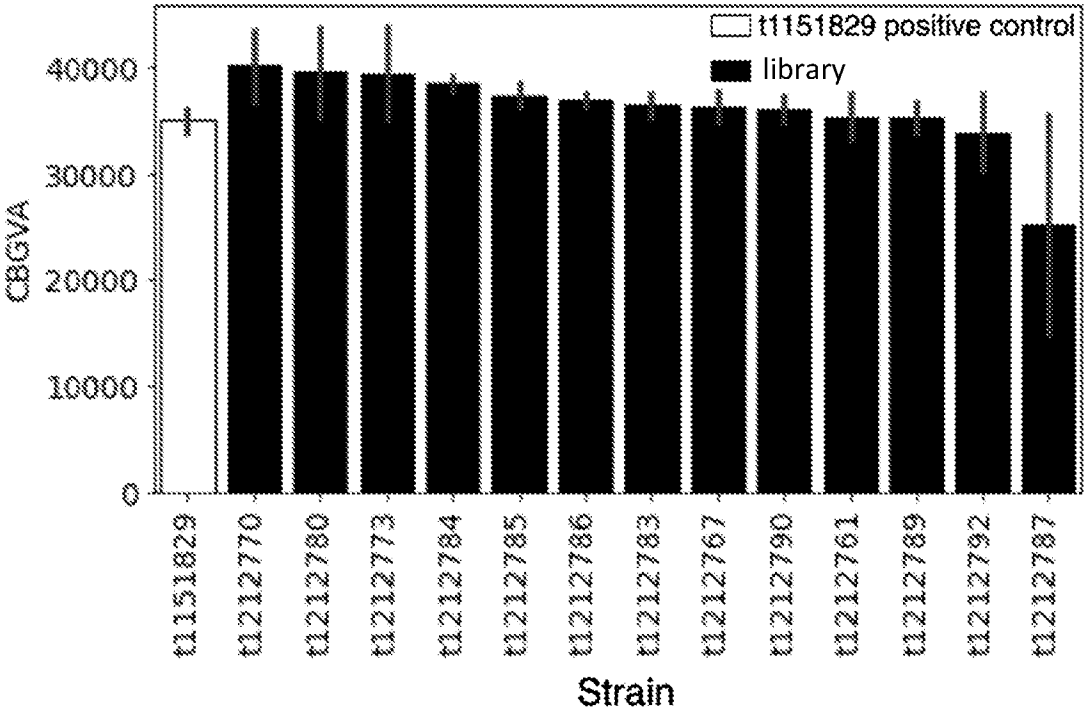


FIG. 7A

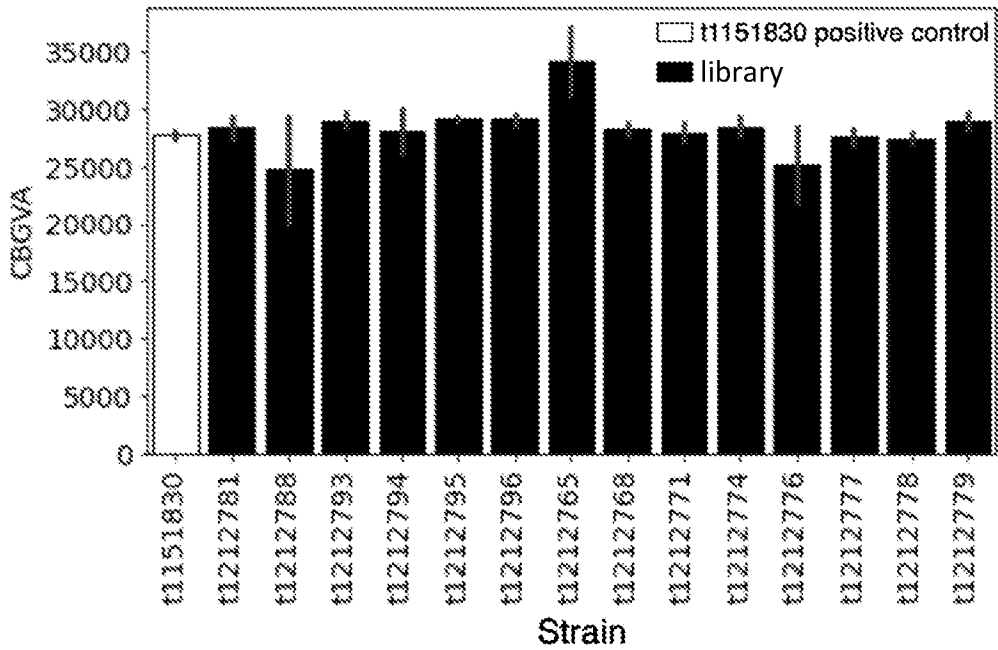


FIG. 7B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2023/066115

A. CLASSIFICATION OF SUBJECT MATTER

C12N 9/10 (2006.01) C12N 15/81 (2006.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, CAPLUS, BIOSIS, EMBASE, AGRICOLA, PATENW: PKS, divarinol, divaric acid, variant and like terms. GENOMEQUEST: SEQ ID NO: 1-28, 30-56, 104-157. PUBMED and internal databases: applicant/inventor search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

☒ Further documents are listed in the continuation of Box C

☒ See patent family annex

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"D" document cited by the applicant in the international application	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
29 June 2023

Date of mailing of the international search report
29 June 2023

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
Email address: pct@ipaustralia.gov.au

Authorised officer

Ella Jungerth
AUSTRALIAN PATENT OFFICE
(ISO 9001 Quality Certified Service)
Telephone No. +61 2 6283 2250

INTERNATIONAL SEARCH REPORT		International application No.
C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		PCT/US2023/066115
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2020/1765471 A1 (GINKO BIOWORKS, INC) 03 September 2020 abstract; [0112]; [0361]; [0371]; SEQ ID NO: 5, 87, 270, 227	1-19, 21, 23-31, 36-47, 56-58, 104-110, 112-114
X	WO 2020/102541 A1 (MANUS BIO, INC) 22 May 2020 abstract; page 20, line 9; claim 18, 19; SEQ ID NO: 50-52, 96-99	1-10, 16, 23-34, 36-47, 54 and 58
X	WO 2020/232553 A1 (HYASYNTH BIOLOGICALS INC.) 26 November 2020 abstract; ([00359]-[00360], [00414], claim 67; SEQ ID NOs: 240, 114, 215, 146, 249 and 247	1-11, 15-18, 23-30, 32-34, 36-47, 54 and 58
A	YANG, X. et al. 'Structural basis for olivetolic acid formation by a polyketide cyclase from <i>Cannabis sativa</i> ' The FEBS Journal, 2016, vol. 283, pages 1088-1106	
P,X	WO 2022/241299 A2 (CELLIBRE, INC.) 17 November 2022 abstract; [0092]; claim 34	1, 5-9, 23-29, 31, 36, 37 and 58

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2023/066115

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).
☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT		International application No.	
Information on patent family members		PCT/US2023/066115	
This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.			
Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	Publication Date	Publication Number	Publication Date
WO 2020/1765471 A1	03 September 2020		
WO 2020/102541 A1	22 May 2020	WO 2020102541 A1	22 May 2020
		CN 113227353 A	06 Aug 2021
		EP 3880799 A1	22 Sep 2021
		US 2022002764 A1	06 Jan 2022
WO 2020/232553 A1	26 November 2020	WO 2020232553 A1	26 Nov 2020
		AU 2020280176 A1	16 Dec 2021
		CA 3137451 A1	26 Nov 2020
		CN 114502734 A	13 May 2022
		EP 3973061 A1	30 Mar 2022
		IL 288157 A	01 Jan 2022
		JP 2022533449 A	22 Jul 2022
		KR 20220012327 A	03 Feb 2022
		SG 11202112690Y A	30 Dec 2021
		US 2022259603 A1	18 Aug 2022
WO 2022/241299 A2	17 November 2022	WO 2022241299 A2	17 Nov 2022
End of Annex			