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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 December 2024 (19.12.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/256472 A2

(51) International Patent Classification:
C12P 17/04 (2006.01) *C12N 9/88* (2006.01)
C07D 307/92 (2006.01) *C12N 9/90* (2006.01)
C12N 9/04 (2006.01) *C12P 7/04* (2006.01)
C12N 9/02 (2006.01) *C12P 7/24* (2006.01)
C12N 9/18 (2006.01) *C12P 7/62* (2022.01)

(21) International Application Number:
PCT/EP2024/066253

(22) International Filing Date:
12 June 2024 (12.06.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
23178694.8 12 June 2023 (12.06.2023) EP

(71) Applicant: **FIRMENICH SA** [CH/CH]; 7, Rue de la Bergère, 1242 Satigny (CH).

(72) Inventors: **SCHALK, Michel**; Rue de la Bergère 7, 1242 Satigny (CH). **GÖRNER, Christian**; Rue de la Bergère 7, 1242 Satigny (CH). **SOLIS ESCALANTE, Daniel**; Rue de la Bergère 7, 1242 Satigny (CH). **ATHANASAKOGLOU, Anastasia**; Rue de la Bergère 7, 1242 Satigny (CH). **DE JONG, René Marcel**; c/o DSM Intellectual Property, P.O. Box 4, 6100 AA ECHT (NL). **MUCHLINSKI, Andrew**; DSM-Firmenich, 10636 Scripps Summit Court #201, San Diego, CA 92131 (US).

(74) Agent: **DSM-FIRMENICH IP**; Wurmisweg 576, 4303 Kaiseraugst (CH).

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

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

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

Published:
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
— *with sequence listing part of description (Rule 5.2(a))*
— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

(54) Title: PREPARATION OF TERPENOID COMPOUNDS

(57) Abstract: The present invention provides a process for the preparation of a compound of formula (I). Included in this invention is an *in vivo* process for the preparation of compound of formula (I) in recombinant cells. The present invention also provides recombinant cells which may be used in said process. The present invention also provides polypeptides which can be used in the present invention. The present invention also provides further processes for the preparation of further compounds.

WO 2024/256472 A2

PREPARATION OF TERPENOID COMPOUNDS

Technical field

The present invention relates to polypeptides and processes for the enzymatic preparation of ambroxide and other terpenoid compounds.

Background

In the perfumery industry, there is a constant need to provide methods for the preparation of compounds for industrial use in fragrances. Key amongst such compounds are ingredients relating to the amber class of olfactory compounds which are naturally found in ambergris and which can function as fixatives to allow scent to endure much longer.

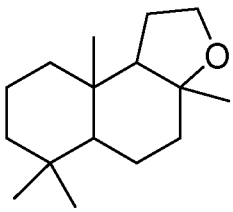
A key compound in ambergris is ambroxide, a terpenoid compound. Terpenoids are found in most organisms (microorganisms, animals and plants). These compounds are made up of five carbon units called isoprene units and are classified by the number of these units present in their structure. Thus monoterpenes, sesquiterpenes and diterpenes are terpenes containing 10, 15 and 20 carbon atoms respectively. Sesquiterpenes, for example, are widely found in the plant kingdom. Many sesquiterpene molecules are known for their flavor and fragrance properties and their cosmetic, medicinal and antimicrobial effects. Numerous sesquiterpene hydrocarbons and sesquiterpenoids have been identified. Commercially relevant compounds include Cetalox® ((3aRS,9aRS,9bRS)-3a,6,6,9a-tetramethyl-1,2,3a,4,6,7,8,9,9a,9b-decahydronaphtho[2,1-b]furan; origin: Firmenich SA, Geneva, Switzerland) or Ambrox® ((3aR,5aS,9aS,9bR)-3a,6,6,9a-tetramethyldodecahydronaphtho[2,1-b]furan; origin: Firmenich SA, Geneva, Switzerland), these compounds replicating ambroxide.

Chemical routes to the preparation of these compounds are known in the art. However, given the environmental and waste problems associated with chemical production of such compounds, there is a need to develop more sustainable processes for the production of ambroxide and other terpenoid compounds.

This problem is addressed by the present invention which provides polypeptides and processes for producing such compounds by *in vivo* and/or bioconversion methods. Said methods may use a multi-step enzymatic process.

Summary

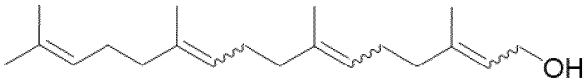
A first aspect of the invention provides a process for the preparation of a compound of formula (I)



(I)

in the form of any one of its stereoisomers or a mixture thereof, comprising:

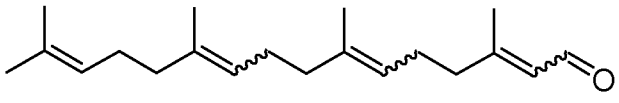
(i) contacting a compound of formula (II)



(II)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having alcohol dehydrogenase (ADH) enzyme activity to produce a compound of formula (III);

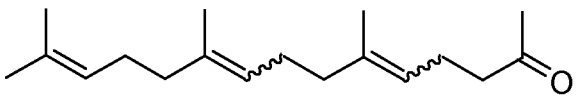
(ii) contacting the compound of formula (III)



(III)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV);

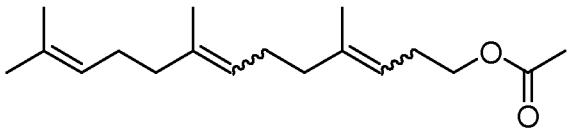
(iii) contacting the compound of formula (IV)



(IV)

In the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having Baeyer-Villiger monooxygenase (BVMO) enzyme activity to produce a compound of formula (V);

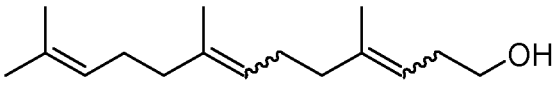
(iv) contacting the compound of formula (V)



(V)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having esterase enzyme activity to produce a compound of formula (VI); and

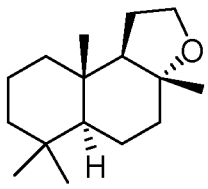
(v) contacting the compound of formula (VI)



(VI)

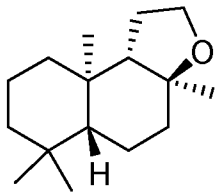
in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having terpene cyclase enzyme activity to produce a compound of formula (I).

An embodiment of the invention is wherein more than 97% of the compound of formula (I) is in the form of formula (Ia):



(formula Ia);

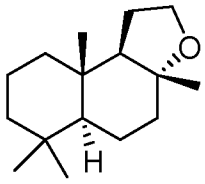
and/or formula (Ib):



(formula Ib).

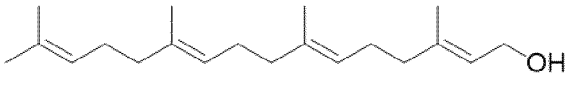
A further embodiment of the invention is wherein:

(i) the compound of formula (I) is in the form of formula (Ia):



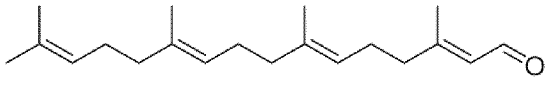
(formula Ia)

(ii) the compound of formula (II) is in the form of formula (IIa):



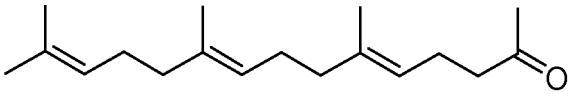
(formula IIa);

(iii) the compound of formula (III) is in the form of formula (IIIa):



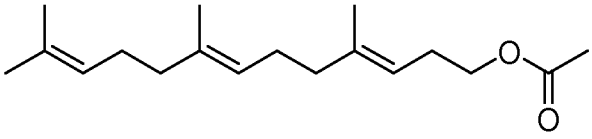
(formula IIIa);

(iv) the compound of formula (IV) is in the form of formula (IVa):



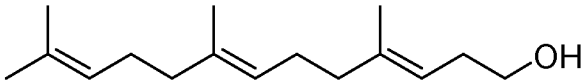
(formula IVa);

(v) the compound of formula (V) is in the form of formula (Va):



(formula Va);

(vi) the compound of formula (VI) is in the form of formula (VIa):



(formula VIa).

A further embodiment of the invention is wherein the process further comprises one or more steps prior to step (i), said step(s) comprising:

(a) preparing geranylgeranyl-diphosphate (GGPP) from isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) using one or more polypepides having prenyltransferase enzyme activity; and/or,

(b) preparing a compound of formula (II) from GGPP using one or more polypeptides having phosphatase enzyme activity.

A further embodiment of the invention is wherein the process is an *in vivo* or bioconversion process.

A further embodiment of the invention is wherein said process is performed in a recombinant cell capable of functionally expressing: (i) the polypeptide having ADH enzyme activity, (ii) the polypeptide having enal-cleaving enzyme activity, (iii) the polypeptide having BVMO enzyme activity, (iv) the polypeptide having esterase enzyme activity, and (v) the polypeptide having terpene cyclase enzyme activity.

A further aspect of the invention provides a recombinant cell comprising, capable of producing or producing a compound of formula (I). The recombinant cell may further comprise one or more compound(s) selected from the group of formula (II), formula (III), formula (IV), formula (V) and formula (VI).

A further aspect of the invention provides a cell culture fermentation medium comprising the recombinant cell of the invention. The cell culture fermentation medium may further comprise a compound of formula (I) and optionally, one or more compound(s) of formula (II), formula (III), formula (IV) formula (V) and/or formula (VI).

A further aspect of the invention provides a reaction mixture comprising compound of formula (I). The reaction mixture may further comprise one or more compound(s) of formula (II), formula (III), formula (IV) formula (V) and/or formula (VI).

Another aspect of the invention provides a process for the preparation of a compound of formula (VI); a recombinant cell comprising, capable of producing or producing a compound of formula (VI); a cell culture fermentation medium comprising said recombinant cell; a reaction mixture comprising compound of said formula (VI).

Another aspect of the invention provides a process for the preparation of a compound of formula (V); a recombinant cell comprising, capable of producing or producing a compound of formula (V); a cell culture fermentation medium comprising said recombinant cell; a reaction mixture comprising compound of said formula (V).

Another aspect of the invention provides a process for the preparation of a compound of formula (IV); a recombinant cell comprising, capable of producing or producing a compound of formula (IV); a cell culture fermentation medium comprising said recombinant cell; a reaction mixture comprising compound of said formula (IV).

A further aspect of the invention provides a compound obtained or obtainable by a process of the invention or from a recombinant cell of the invention or from a cell culture fermentation medium of the invention or from a reaction mixture of the invention as described herein above.

A further aspect of the invention provides the use of said compound as a perfumery, flavor or aroma ingredient, or as a precursor for making said ingredient.

A further aspect of the invention is the use of a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV), (V), (VI), (I) and/or a derivative thereof.

Description of the drawings

Figure 1. Biosynthetic pathway of (2E)-geranyl-diphosphate (GPP), (2E,6E)-farnesyl-diphosphate (FPP) and (2E,6E,10E)-geranylgeranyl-diphosphate (GGPP) from isopentenyl-diphosphate (IPP) and dimethylallyl-diphosphate (DMAPP).

Figure 2. Biosynthetic pathways of (2E,6E,10E)-geranylgeraniol from (2E,6E,10E)-geranylgeranyl-diphosphate (GGPP). Pi, inorganic phosphate; PPi, inorganic pyrophosphate.

Figure 3. New biochemical pathway to (3E,7E)-homofarnesol.

Figure 4. GC-MS analysis of terpenoids and derivatives produced using *E. coli* cells engineered to produce (3E,7E)-homofarnesol and expressing the proteins PsAerADH (SEQ ID NO: 11), SCH24-BVMO1 (SEQ ID NO: 23), SCH24-EST1 (SEQ ID NO: 27), CcrGGPPS2-del57 (SEQ ID NO: 1), PgpB (SEQ ID NO: 3) encoded by the plasmid pHFOL-5.

Figure 5. GC-MS analysis of terpenoids and derivatives produced using *E. coli* cells expressing the proteins PsAerADH (SEQ ID NO: 11), SCH24-EST1 (SEQ ID NO: 27), CcrGGPPS2-del57 (SEQ ID NO: 1), PgpB (SEQ ID NO: 3) encoded by the plasmid pF-Facetone-7 (A) and using the same cells expressing in addition a BVMO enzyme (AflaBVMO1, SEQ ID NO: 26) (B).

Figure 6. GC-MS chromatogram of YST403_HFOL strain engineered to produce (3E,7E)-homofarnesol. The final product (3E,7E)-homofarnesol, as well as the pathway intermediates (5E,9E)-farnesylacetone and (2E,6E,10E)-geranylgeraniol are shown.

Figure 7. New biochemical pathway of compound of formula Ia using squalene cyclases.

Figure 8: GC-MS analysis of terpenoids and derivatives produced using *E coli* DP1205 expressing the proteins PsAerADH (SEQ ID NO: 11), AflavBVMO1 (SEQ ID NO: 26), SCH24-EST1 (SEQ ID NO: 27), CcrGGPPS2-del57 (SEQ ID NO: 1), PgpB (SEQ ID NO: 3), the wild type A0A5P9HJ69 (SEQ ID NO: 42) (A) and the A0A5P9HJ69_V1 variant (SEQ ID NO: 43) (B). The compound of formula (Ia) and the pathway intermediate (3E,7E)-homofarnesol are shown.

Figure 9: Titers of compound of formula (Ia) being produced by *E coli* DP1205 expressing the proteins PsAerADH (SEQ ID NO: 11), AflavBVMO1 (SEQ ID NO: 26), SCH24-EST1 (SEQ ID NO: 27), CcrGGPPS2-del57 (SEQ ID NO: 1), PgpB (SEQ ID NO: 3) and different wild type or mutant squalene cyclases.

Figure 10. Titers of compound of formula (Ia) produced in *S. cerevisiae* cells expressing the geranylgeranyl diphosphate synthase CarG (SEQ ID NO: 2), the phosphatase PgpB (SEQ ID NO: 3),

the alcohol dehydrogenase SCH23-ADH1 (SEQ ID NO: 21), the Baeyer-Villiger monooxygenase AflavBVMO1 (SEQ ID NO: 26), the enal-cleaving enzymes SCH94-03944 (SEQ ID NO: 22) and different wild type or mutant variants of squalene cyclases.

Figure 11. GC-MS chromatogram of culture extracts of *S. cerevisiae* engineered to express the (3E,7E)-homofarnesol biosynthetic pathway genes and squalene cyclase mutant variant OYT72085_V1 (SEQ ID NO: 48) (A) or the wild type squalene cyclase OYT72085.1 (SEQ ID NO: 47) (B). Peaks that correspond to (3E,7E)-homofarnesol and compound of formula (Ia) are indicated.

Figure 12. Titers of compound of formula (Ia) being produced by *E coli* DP1205 expressing the proteins PsAerADH (SEQ ID NO: 11), AflavBVMO1 (SEQ ID NO: 26), SCH24-EST1 (SEQ ID NO: 27), CcrGGPPS2-del57 (SEQ ID NO: 1), PgpB (SEQ ID NO: 3) and bacterial membrane-integrated meroterpenoid cyclases.

Figure 13. New biochemical pathway of compound of formula (Ia) using meroterpenoid cyclases.

Figure 14. Titers of compound of formula (Ia) being produced by the yeast strain YST403 engineered to produce (3E,7E)-homofarnesol and expressing different bacterial membrane-integrated meroterpenoid cyclases.

Figure 15. GC-MS chromatogram of the yeast strain YST403 engineered to produce (3E,7E)-homofarnesol and to express WP_234754442.1 (SEQ ID NO: 51) (A) and GC-MS chromatograms of the control strain YST403 expressing only the (3E,7E)-homofarnesol pathway enzymes (B). The compound of formula (Ia) and the pathway intermediate (3E,7E)-homofarnesol are shown.

Figure 16. GC-MS chromatogram in single ion monitoring mode (221 Da) of the yeast strain YST403 engineered to express (3E,7E)-homofarnesol and to express A0A2P1DP74.1 (MacJ) (SEQ ID NO: 71) (A) and GC-MS chromatograms of the control strain YST403 expressing only the (3E,7E)-homofarnesol pathway enzymes (B). The compound of formula (Ia) and the pathway intermediate (3E,7E)-homofarnesol are shown.

Figure 17. The chiral GC-MS chromatogram in single ion monitoring mode (221 Da) of *E. coli* DP1205 engineered to express (3E,7E)-homofarnesol and expressing OKH29475.1 (SEQ ID NO: 74) (C) compared to authentic standards of compound of formula (Ia) (B) and compound of formula (Ib) (A).

Figure 18. The GC-MS chromatogram in single ion monitoring mode (221 Da) of the strain YST403 engineered to produce (3E,7E)-homofarnesol and expressing the soluble meroterpenoid cyclases OKH29475.1 (SEQ ID NO: 74) and NEQ07043.1 (SEQ ID NO: 75) as well as the (3E,7E)-homofarnesol producing control strain YST403 HFOL. Compound of formula (Ia) is shown.

Figure 19. The predicted structure of WP_234754442.1 (SEQ ID NO: 51) using ESMFold showing a pore like structure consisting of 7 helices. The N and C-Terminus are indicated as well as the entrance to the proposed active site, which is at the same side than the N-Terminus.

Figure 20. Titers of compound of formula (Ia) being produced by *E coli* DP1205 expressing the proteins PsAerADH (SEQ ID NO: 11), AflavBVMO1 (SEQ ID NO: 26), SCH24-EST1 (SEQ ID NO: 27), CcrGGPPS2-del57 (SEQ ID NO: 1), PgpB (SEQ ID NO: 3) and different mutant variants of the bacterial membrane-integrated meroterpenoid cyclases WP_234754442.1 (SEQ ID NO: 51). The titers are shown relative to the wild-type enzyme WP_234754442.1 (SEQ ID NO: 51).

Figure 21. Titers of compound of formula (Ia) being produced by the yeast strain YST403 engineered to produce (3E,7E)-homofarnesol and express the bacterial membrane-integrated meroterpenoid cyclases WP_234754442.1 (SEQ ID NO: 51) and the mutant variants WP_234754442.1_S9C (SEQ ID NO: 56) and WP_234754442.1_S9M (SEQ ID NO: 57). The titers are shown in percent relative to the wild type enzyme WP_234754442.1 (SEQ ID NO: 51).

Figure 22. Titers of the compound of formula (Ia) (Left), compound of formula (Ic) (middle) and compound of formula (Id) (right) produced by bioconversion of chemically synthesized homofarnesol with the bacterial membrane-integrated meroterpenoid cyclases WP_051467941.1 (SEQ ID NO: 50), WP_234754442.1 (SEQ ID NO: 51), WP_190963420.1 (SEQ ID NO: 52) and the squalene hopen cylase AacSHC_M132R_I432T_A224V (SEQ ID NO: 78) in the presence or absence of 0.06 (*w/v*) of the detergent sodium dodecyl sulfate (SDS).

Figure 23. GC-MS analysis of terpenoids and derivatives produced by the bioconversion of chemically synthesized (3E,7E)-homofarnesol containing (3Z,7E)-homofarnesol impurities with E coli Bl21(DE3)Star cells expressing the gene for the bacterial membrane-integrated meroterpenoid cyclases WP_234754442.1 (SEQ ID NO: 51) (A) as well as the mutant squalene cyclase AAcSHC_M132R_A224V_I432T (SEQ ID NO: 78) (B). The cyclisation products compound of formula (Ia) and compound of formula (Id) from (3E,7E)-homofarnesol as well as the cyclisation product compound of formula (Ic) from (3Z,7E)-homofarnesol are shown.

Figure 24. GC-MS analysis of the bioconversion of chemically synthesized (3E,7E)-homofarnesol with *E. coli* Bl21(DE3)Star cells expressing the genes encoding for the squalene cyclases ZmSHC_F437A_G600M (SEQ ID NO: 88), AacSHC_F437A_G600M (SEQ ID NO: 81), A0A0T6LPP7-V1 (SEQ ID NO: 265), A0A7V0I7Y5-V1 (SEQ ID NO: 266), UPI00248B5E40-V1 (SEQ ID NO: 267) and UPI002800B5BA-V1 (SEQ ID NO: 268). Control is in the absence of said squalene cyclases.

Figure 25. Titers of compound of formula (Ia) being produced by *E coli* DP1205 expressing PsAerADH (SEQ ID NO: 11), AflavBVMO1 (SEQ ID NO: 26), SCH24-EST1 (SEQ ID NO: 27), CcrGGPPS2-del57

(SEQ ID NO: 1), PgpB (SEQ ID NO: 3) and different bacterial membrane-integrated meroterpenoid cyclases. The titers are shown relative to the enzyme WP_234754442.1 (SEQ ID NO: 51).

Abbreviations used

ADH	alcohol dehydrogenase
BVMO	Baeyer-Villiger Monooxygenase
bp	base pair
kb	kilo base
DNA	deoxyribonucleic acid
cDNA	complementary DNA
DMAPP	dimethylallyl diphosphate
FMO	Flavin Monooxygenase
FPP	farnesyl diphosphate
GPP	geranyldiphosphate
GGPP	geranylgeranyl diphosphate
GGPS	geranylgeranyl diphosphate synthase
GC	gas chromatograph
IPP	isopentenyl diphosphate
iMS	mass spectrometer / mass spectrometry
MVA	mevalonic acid
PP	diphosphate, pyrophosphate
PCR	polymerase chain reaction
RNA	ribonucleic acid
SHC	squalene cyclase
mRNA	messenger ribonucleic acid
miRNA	micro RNA
siRNA	small interfering RNA
rRNA	ribosomal RNA
tRNA	transfer RNA
TPP	terpenyl diphosphate

Definitions

General terms:

For the descriptions herein and the appended claims, the use of “or” means “and/or” unless stated otherwise. Similarly, “comprise”, “comprises”, “comprising”, “include”, “includes”, and “including” are interchangeable and not intended to be limiting.

It is to be further understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of".

The terms "purified", "substantially purified", and "isolated" as used herein refer to the state of being free of other, dissimilar compounds with which a compound of the invention is normally associated in its natural state, so that the "purified", "substantially purified", and "isolated" subject comprises at least 0.5%, 1%, 5%, 10%, or 20%, or at least 50% or 75% of the mass, by weight, of a given sample. In one embodiment, these terms refer to the compound of the invention comprising at least 95, 96, 97, 98, 99 or 100%, of the mass, by weight, of a given sample. As used herein, the terms "purified," "substantially purified," and "isolated" when referring to a nucleic acid or protein, or nucleic acids or proteins, also refers to a state of purification or concentration different than that which occurs naturally, for example in a prokaryotic or eukaryotic environment, like, for example in a bacterial or fungal cell, or in the mammalian organism, especially human body. Any degree of purification or concentration greater than that which occurs naturally, including (1) the purification from other associated structures or compounds or (2) the association with structures or compounds to which it is not normally associated in said prokaryotic or eukaryotic environment, are within the meaning of "isolated". The nucleic acid or protein or classes of nucleic acids or proteins, described herein, may be isolated, or otherwise associated with structures or compounds to which they are not normally associated in nature, according to a variety of methods and processes known to those of skill in the art.

The term “about” indicates a potential variation of ± 25% of the stated value, in particular ± 15%, ± 10 %, more particularly ± 5%, ± 2% or ± 1%.

The term "substantially" describes a range of values from about 80 to 100%, such as, for example, 85-99.9%, in particular 90 to 99.9%, more particularly 95 to 99.9%, or 98 to 99.9% and especially 99 to 99.9%.

“Predominantly” refers to a proportion in the range of above 50%, as for example in the range of 51 to 100%, particularly in the range of 75 to 99,9%, more particularly 85 to 98,5%, like 95 to 99%.

A “main product” in the context of the present invention designates a single compound or a group of at least 2 compounds, like 2, 3, 4, 5 or more, particularly 2 or 3 compounds, which single compound or group of compounds is “predominantly” prepared by a reaction as described herein, and is contained in said reaction in a predominant proportion based on the total amount of the constituents of the product formed by said reaction. Said proportion may be a molar proportion, a weight proportion or, preferably based on chromatographic analytics, an area proportion calculated from the corresponding chromatogram of the reaction products.

A “side product” in the context of the present invention designates a single compound or a group of at least 2 compounds, like 2, 3, 4, 5 or more, particularly 2 or 3 compounds, which single compound or group of compounds is not “predominantly” prepared by a reaction as described herein.

Because of the reversibility of enzymatic reactions, the present invention relates, unless otherwise stated, to the enzymatic or biocatalytic reactions described herein in both directions of reaction.

The term "stereoisomers" includes conformational isomers and in particular configuration isomers.

Included in general are, according to the invention, all “stereoisomeric forms” of the compounds described herein, such as “constitutional isomers” and “stereoisomers”.

“Stereoisomeric forms” encompass in particular, “stereoisomers” and mixtures thereof, e.g. configuration isomers (optical isomers), such as enantiomers, or geometric isomers (diastereomers), such as E- and Z-isomers, and combinations thereof. If one or more asymmetric centers are present in one molecule, the invention encompasses all combinations of different conformations of these asymmetry centers, e.g. enantiomeric pairs.

“Stereoselectivity” describes the ability to produce a particular stereoisomer of a compound in a stereoisomerically pure form or to specifically convert a particular stereoisomer in an enzyme catalyzed method as described herein out of a plurality of stereoisomers. More specifically, this means that a product of the invention is enriched with respect to a specific stereoisomer, or an educt may be depleted with respect to a particular stereoisomer. This may be quantified via the purity %ee-parameter calculated according to the formula:

$$\%ee = [X_A - X_B]/[X_A + X_B] * 100,$$

wherein X_A and X_B represent the molar ratio (Molenbruch) of the stereoisomers A and B.

The terms “selectively converting” or “increasing the selectivity” in general means that a particular stereoisomeric form, as for example the E-form, of an unsaturated hydrocarbon, is converted in a higher proportion or amount (compared on a molar basis) than the corresponding other stereoisomeric form, as for example Z-form, either during the entire course of said reaction (i.e. between initiation and termination of the reaction), at a certain point of time of said reaction, or during an “interval” of said reaction. In particular, said selectivity may be observed during an “interval” corresponding 1 to 99%, 2 to 95%, 3 to 90%, 5 to 85%, 10 to 80%, 15 to 75%, 20 to 70%, 25 to 65%, 30 to 60%, or 40 to 50% conversion of the initial amount of the substrate. Said higher proportion or amount may, for example, be expressed in terms of:

- a higher maximum yield of an isomer observed during the entire course of the reaction or said interval thereof;
- a higher relative amount of an isomer at a defined % degree of conversion value of the substrate; and/or

- an identical relative amount of an isomer at a higher % degree of conversion value;

each of which preferably being observed relative to a reference method, said reference method being performed under otherwise identical conditions with known chemical or biochemical means.

"Yield" and / or the "conversion rate" of a reaction according to the invention is determined over a defined period of, for example, 4, 6, 8, 10, 12, 16, 20, 24, 36 or 48 hours, in which the reaction takes place. In particular, the reaction is carried out under precisely defined conditions, for example at "standard conditions" as herein defined.

If the present disclosure refers to features, parameters and ranges thereof of different degree of preference (including general, not explicitly preferred features, parameters and ranges thereof) then, unless otherwise stated, any combination of two or more of such features, parameters and ranges thereof, irrespective of their respective degree of preference, is encompassed by the disclosure of the present description.

Biochemical and biological terms

The term "domain" refers to a set of amino acids or a partial sequence of amino acids residues conserved at specific positions along an alignment of sequences of evolutionarily related proteins. While amino acids at other positions can vary between protein homologues, amino acids that are highly conserved at specific positions of such domain indicate amino acids that are likely essential in the structure, stability or function of a protein. Identified by their high degree of conservation in aligned sequences of a family of protein homologues, they can be used as identifiers to determine if any polypeptide in question belongs to a previously identified polypeptide family.

The term "motif " or consensus sequence" or "signature" refers to a short-conserved region in the sequence of evolutionarily related proteins. Motifs are frequently highly conserved parts of domains, but may also include only part of the domain. Signatures are predictive models which describe protein families, domains or sites.

The sequences of motifs can be described using the standard IUPAC one-letter codes for the amino acids. Ambiguities are indicated by listing the acceptable amino acids for a given position between brackets. For example, [LWI] stands for L (Leucine), W (Tryptophan) or I (Isoleucine). X represent positions where independently of each other any natural amino acid residue is present.

A "protein family" is defined as a group of proteins that share a common evolutionary origin reflected by their related functions, similarities in sequence, or similar primary, secondary or tertiary structure. Proteins within protein families are usually homologous and have similar structure of conserved functional domains and motifs.

Specialist databases exist for the identification of protein domains, for example, SMART (http://smart.embl-heidelberg.de/smart/set_mode.cgi?GENOMIC=1) (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95, 5857-5864; Letunic et al. (2020) Nucleic Acids Res 49, D458–D460), InterPro (Paysan-Lafosse et al, *Nucleic Acids Research, Nov 2022*; Mulder et al., (2003) Nucl. Acids. Res. 31, 315-318), or Pfam (Bateman et al., Nucleic Acids Research 30(1): 276-280 (2002)).

Useful tools to search or predict protein domains or protein family signatures in protein sequence are for example the NCBI conserved domain search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) or the InterProScan tool (http://www.ebi.ac.uk/interpro/search/sequence/). Domains or motifs may also be identified using routine techniques, such as by sequence alignment.

The term "Pfam" refers to a large collection of protein domains and protein families maintained by the Pfam Consortium and available at several sponsored world wide web sites, such as the InterPro consortium web site https://www.ebi.ac.uk/interpro/ (European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL_EBI). The latest release of Pfam is Pfam 35.0 (November 2021), based on the UniProt Reference Proteomes (El-Gebali S. et al, 2019, Nucleic Acids Res. 47, Database issue D427–D432). Pfam domains and families are identified using multiple sequence alignments and hidden Markov models (HMMs). Pfam-A family or domain assignments, are high quality assignments generated by a curated seed alignment using representative members of a protein family and profile hidden Markov models based on the seed alignment (Unless otherwise specified, matches of a queried protein to a Pfam domain or family are Pfam-A matches). All identified sequences belonging to the family are then used to automatically generate a full alignment for the family (Sonnhammer (1998) Nucleic Acids Research 26, 320-322; Bateman (2000) Nucleic Acids Research 26, 263-266; Bateman (2004) Nucleic Acids Research 32, Database Issue, D138-D141; Finn (2006) Nucleic Acids Research Database Issue 34, D247-251; Finn (2010) Nucleic Acids Research Database Issue 38, D211-222). By accessing the Pfam database, for example, using any of the above-reference websites, protein sequences can be queried against the HMMs using HMMER homology search software (e.g., HMMER2, HMMER3, or a higher version, hmmer.janelia.org/). Significant matches that identify a queried protein as being in a pfam family (or as having a particular Pfam domain) are those in which the bit score is greater than or equal to the gathering threshold for the Pfam domain. Expectation values (e-values) can also be used as a criterion for inclusion of a queried protein in a Pfam or for determining whether a queried protein has a particular Pfam domain, where low e-values, much less than 1.0, for example less than 0.1, or less.

InterPro is another database of protein families providing a classification of protein sequences into families and identifies functionally important domains and conserved sites (Blum et al, Nucleic Acids Res. 2021 49(D1):D344-D354). The protein signatures are provided by multiple databases such as Pfam or SMART (Simple Modular Architecture Research Tool). InterProScan is a software that allows protein and nucleic acid sequences to be searched against InterPro's signatures.

The “E-value” (expectation value) is the number of hits that would be expected to have a score equal to or better than this value, by chance alone. This means that a good E-value which gives a confident prediction is much less than 1. E-values around 1 is what is expected by chance. Thus, the lower the E-value, the more specific the search for domains will be. Only positive numbers are allowed.

A “precursor” compound or molecule of a target compound or molecule as described herein is converted to said target compound, preferably through the enzymatic action of a suitable polypeptide performing at least one structural or functional change on said precursor molecule. For example, a “diphosphate precursor” (as for example a “terpenyl diphosphate precursor”) is converted to said target compound (as for example a terpene alcohol) via enzymatic removal of the diphosphate moiety, for example by removal of mono- or diphosphate groups by a phosphatase enzyme. For example, a “non-cyclic precursor” (like a “non-cyclic terpenyl precursor”) may be converted to the cyclic target molecule (like a cyclic terpene compound) through the action of a cyclase or synthase enzyme, irrespective of the particular enzymatic mechanism of such enzyme, in one or more steps.

The enzyme nomenclature or enzyme classification (EC) established by the International Union of Biochemistry and Molecular Biology (IUBMB) is a system of naming and categorizing enzymes based on their catalytic activity and biochemical properties. The enzyme nomenclature is widely used in biochemistry to classify and categorize based on their function. The E.C. classification assigns each enzyme a number reflecting the reaction or the type of reaction catalyzed by this enzyme.

The enzyme classification can be explored using the ‘ExplorEnz’ database (https://www.enzyme-database.org/) or International Union of Biochemistry and Molecular Biology (IUBMB) web site (https://iubmb.qmul.ac.uk). Information can be found about the classification and nomenclature of enzymes, their functions and properties. The database can be searched to find information for a specific enzyme family or enzyme.

The terms “biological function,” “function”, “biological activity” or “activity” of a terpenyl synthase refer to the ability of a terpenyl diphosphate synthase as described herein to catalyze the formation of at least one terpenyl diphosphate from the corresponding precursor terpene.

The terms “biological function,” “function”, “biological activity” or “activity” of a terpenyl diphosphate phosphatase refer to the ability of the terpenyl diphosphate phosphatase as described herein to catalyze the removal of a diphosphate group from said terpenyl compound to form the corresponding terpene alcohol.

As used herein, the term “host cell”, “recombinant cell” or “transformed cell” refers to a cell (or organism) altered to harbor at least one nucleic acid molecule, for instance, a recombinant gene encoding a desired protein or nucleic acid sequence which upon transcription yields at least one functional

polypeptide of the present invention. The host cell is particularly a bacterial cell, a fungal cell or a plant cell or plants. The host cell may contain a recombinant gene or several genes, as for example organized as an operon, which has been integrated into the nuclear organelle genomes of the host cell. Alternatively, the host may contain the recombinant gene extra-chromosomally. Methods of introducing recombinant nucleic acid sequences into such host cells are well known in the art and constitute routine laboratory methodologies which do not need to be further described herein.

The term “organism” refers to any non-human multicellular or unicellular organism such as a plant, or a microorganism. Particularly, a micro-organism is a bacterium, a yeast, an algae or a fungus.

The term “plant” is used interchangeably to include plant cells including plant protoplasts, plant tissues, plant cell tissue cultures giving rise to regenerated plants, or parts of plants, or plant organs such as roots, stems, leaves, flowers, pollen, ovules, embryos, fruits and the like. Any plant can be used to carry out the methods of an embodiment herein.

Detailed Description

As described above, many sesquiterpene molecules are known for their flavor and fragrance properties and their cosmetic, medicinal and antimicrobial effects. Numerous sesquiterpene hydrocarbons and sesquiterpenoids have been identified. Commercially relevant compounds include Cetalox® ((3aRS,9aRS,9bRS)-3a,6,6,9a-tetramethyl-1,2,3a,4,6,7,8,9,9a,9b-decahydronaphtho[2,1-b]furan; origin: Firmenich SA, Geneva, Switzerland) or Ambrox® ((3aR,5aS,9aS,9bR)-3a,6,6,9a-tetramethyldodecahydronaphtho[2,1-b]furan; origin: Firmenich SA, Geneva, Switzerland), these compounds replicating ambroxide.

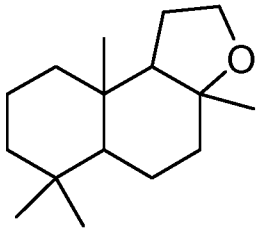
The present inventors sought to identify improved processes for the preparation of a compound of formula (I), also known as 3a,6,6,9a tetramethyldodecahydronaphtho[2,1-b]furan.

To prepare an improved process for the preparation of a compound of formula (I), they developed a deep understanding of the biochemical route to the production of these compounds by a multi-enzymatic reaction from precursor compounds. This multi-enzymatic reaction is the first time the preparation of this compound has been performed by such a step-wise reaction and constitutes a significant scientific and commercial advance in the preparation of sesquiterpene compound of formula (I). In particular, the combination of enzymes and their order in the process has not been described before in the prior art.

Included in this invention is an *in vivo* process for the preparation of compound of formula (I) in recombinant cells. This is the first time a wholly *in vivo* process for this production of this compound has been demonstrated by creating a biosynthetic pathway to the compound of formula (I) in recombinant cells.

Accordingly, this invention provides a solution to the the problem of the preparation of such compound.

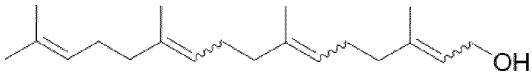
A first aspect of the invention provides a process for the preparation of a compound of formula (I)



(I)

in the form of any one of its stereoisomers or a mixture thereof, comprising:

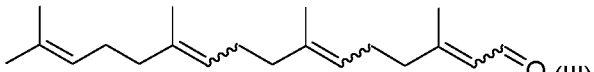
(i) contacting a compound of formula (II)



(II)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having ADH enzyme activity to produce a compound of formula (III);

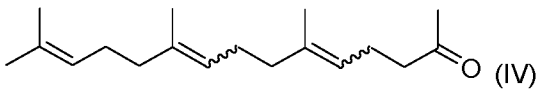
(ii) contacting a compound of formula (III)



(III)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV);

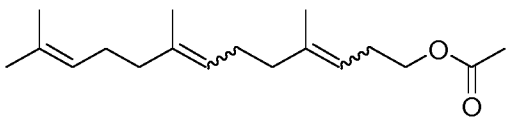
(iii) contacting the compound of formula (IV)



(IV)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having BVMO enzyme activity to produce a compound of formula (V);

(iv) contacting the compound of formula (V)



(V)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having esterase enzyme activity to produce a compound of formula (VI); and,

(v) contacting the compound of formula (VI)

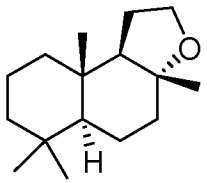
(VI)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having terpene cyclase enzyme activity to produce a compound of formula (I).

For the sake of clarity, by the expression "any one of its stereoisomers", or the similar, it is meant the normal meaning understood by a person skilled in the art, i.e. that the invention compound can be a pure stereoisomer such as an enantiomer or a diastereomer (e.g. in relation to the configuration E or Z of any of the double bonds or in relation to the configuration R or S of any of the chiral carbon centers).

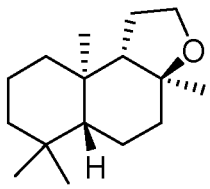
According to any of the aspects or embodiments of the invention, said compound can be in the form of any of its steroisomers or of a mixture thereof, e.g. the invention relates to compositions of matter comprising one or more forms of the compound of formula (I), having the same chemical structure but differing by the configuration of the chiral centers.

In particular, compound (I) can be in the form of a mixture comprising stereoisomer Ia (formula Ia) and wherein said stereoisomer Ia represents at least 50 %, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 99% or more of the total mixture.



(formula Ia)

Alternatively, compound (I) can be in the form of a mixture comprising stereoisomer Ib (formula Ib) and wherein said stereoisomer Ib represents at least 50 %, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 99% or more of the total mixture.



(formula Ib)

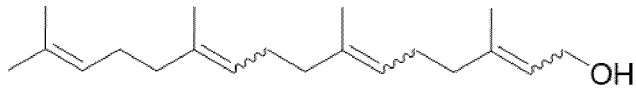
In one embodiment, more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

According to any of the aspects or embodiments of the invention, compounds of formula (II) to (VI) can be in the form of its E or Z isomer or of a mixture thereof. In particular, any one of compounds of formula

(II) to (VI) can be in the form of a mixture consisting of stereoisomer E and Z and wherein said stereoisomer IIa, IIIa, IVa, Va or VIa represent at least 50 % of the total mixture, or even at least 75% (i.e a mixture E/Z comprised between 75/25 and 100/0).

Step (i) of the process of the invention

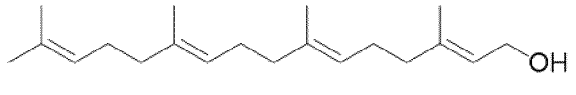
Step (i) of the process of the invention relates to contacting a compound of formula (II) with a polypeptide having ADH enzyme activity.



(formula II)

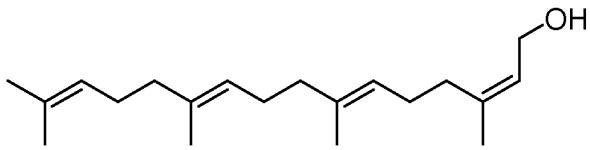
The compound of formula (II) is also known as geranylgeraniol, i.e. 3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-ol, CAS No 7614-21-3.

The compound of formula (II) may be present in any one of its stereoisomers or a mixture thereof. Specifically, the compound may have the following structures and isoforms.



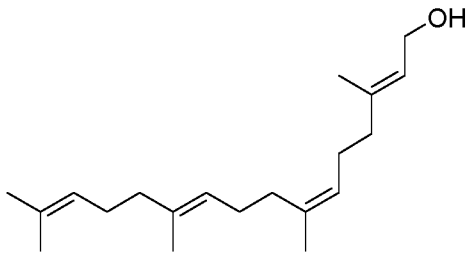
(formula IIa)

(2E,6E,10E)-Geranylgeraniol; (2E,6E,10E)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-ol; CAS No 24034-73-9.



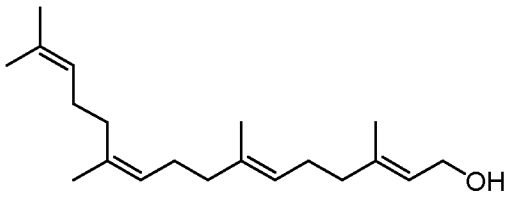
(formula IIb)

(2Z,6E,10E)-Geranylgeraniol; (2Z,6E,10E)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-ol; CAS No 57784-25-5.



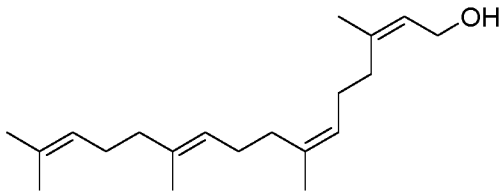
(formula IIc)

(2E,6Z,10E)-Geranylgeraniol; (2E,6Z,10E)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-ol; CAS No 83689-05-8.



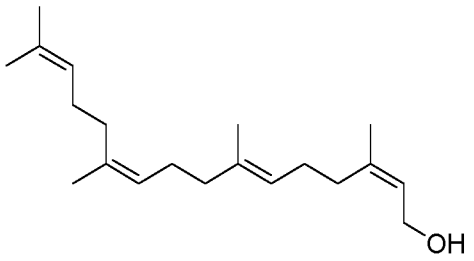
(formula IId)

2E,6E,10Z)-Geranylgeraniol; (2E,6E,10Z)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-ol; CAS No 68690-77-7.



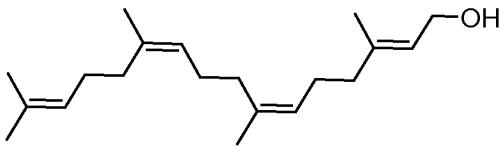
(formula IIe)

(2Z,6Z,10E)-Geranylgeraniol; (2Z,6Z,10E)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-ol; CAS No 83689-06-9.



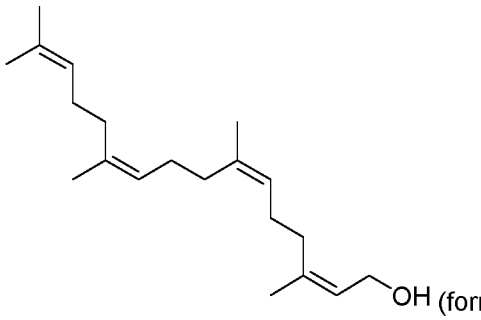
(formula IIf)

(2Z,6E,10Z)-Geranylgeraniol; (2Z,6E,10Z)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-ol; CAS No 83689-07-0.



(formula IIg)

(2E,6Z,10Z)-Geranylgeraniol; (2E,6Z,10Z)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-ol; CAS No 83689-08-1.



(formula IIh)

(2Z,6Z,10Z)-Geranylgeraniol; (2Z,6Z,10Z)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-ol; CAS No 1945-42-2.

Step (i) relates to the use of a polypeptide having ADH enzyme activity.

An “alcohol dehydrogenase” (ADH) in the context of the present invention refers to a polypeptide having the ability to oxidize an alcohol to the corresponding aldehyde in the presence of NAD^+ or $NADP^+$ as cofactor. Such enzymes are members of the E.C. families 1.1.1.1 (NAD^+ dependent) or 1.1.1.2 ($NADP^+$ dependent). More particularly, an ADH of the invention has the ability to oxidize linear terpenoid alcohols to the respective carbonyl compounds in particular to the corresponding aldehydes, like geranylgeraniol to geranylgeranial. ADHs, as used herein, may either be endogenously present in the respective biocatalytic process or may be exogenous.

“Alcohol dehydrogenase enzyme activity” is determined under “standard conditions” as described herein below: It can be determined using recombinant alcohol dehydrogenase (ADH) polypeptide expressing host cells, disrupted ADH polypeptide expressing cells, fractions of these or enriched or purified ADH polypeptide, in a culture medium or reaction medium, preferably buffered, having a pH in the range of 6 to 11, preferably 7 to 9, at a temperature in the range of about 20 to 45 °C, like about 25 to 40 °C, preferably 25 to 32 °C and in the presence of a reference substrate, here in particular geranylgeraniol, either added at an initial concentration in the range of 1 to 100 µM, preferably 5 to 50 µM, in particular 30 to 40 µM, or endogenously produced by the host cell. For *in-vitro* assays a cofactor selected from NADH and NADPH has to be added in a suitable easily to be determined concentration. The conversion reaction to form the respective aldehyde compounds, like geranylgeranial is conducted from 10 min to 5 h, preferably about 1 to 2 h. The oxidation product may then be determined in conventional matter, for example after extraction with an organic solvent, like ethyl acetate.

A further method to evaluate the oxidation of geranylgeraniol to geranylgeranial by ADHs is described in Example 3.

A preferred embodiment of the invention is wherein the polypeptide having said ADH enzyme activity comprises at least one or more sequence motifs selected from:

. CHTD (SEQ ID NO: 228) as for example in SEQ ID NO: 11, 12, 13, 14, 17, 18, 19 or 20;
. GHEGxG (SEQ ID NO: 229) as for example in SEQ ID NO: 11, 12, 13, 14, 17, 18, 19 or 20;
. LxCGxxTGxGA (SEQ ID NO: 230) as for example in SEQ ID NO: 11, 12, 13, 14, 17, 18, 19 or 20;
. Gx[VI]GL (SEQ ID NO: 231) as for example in SEQ ID NO: 11, 12, 13, 14, 15, 17, 18 , 19 or 20;
. LxxxG[LVI][PA] (SEQ ID NO: 232) as for example in SEQ ID NO: 11, 12, 15, 17, 18 , 19 or 20;
. GxVxAI (SEQ ID NO: 233) as for example in SEQ ID NO: 16 or 21; and
. YxATKxA (SEQ ID NO: 234) as for example in SEQ ID NO: 16 or 21;

wherein

in the above motifs, residues x represent independently of each other any natural amino acid residue in a polypeptide having ADH activity. Ambiguities are indicated by listing the acceptable amino acids for a given position between brakets. For example, [VI] stands for V (valine), or I (isoleucine).

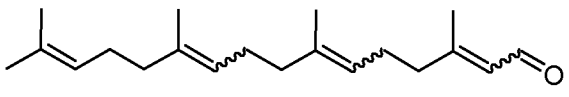
Preferably, the polypeptide having said ADH enzyme activity comprises: CHTD (SEQ ID NO: 228), GHEGxG (SEQ ID NO: 229), LxCGxxTGxGA (SEQ ID NO: 230) and Gx[VI]GL (SEQ ID NO: 231) motifs, as for example in SEQ ID NO: 11, 12, 13, 14, 17, 18, 19 or 20;

Preferably, the polypeptide having said ADH activity comprises: CHTD (SEQ ID NO 228), GHEGxG (SEQ ID NO: 229), LxCGxxTGxGA (SEQ ID NO: 230), Gx[VI]GL (SEQ ID NO: 231) and LxxxG[LVI][PA] (SEQ ID NO: 232) motifs, as for example in SEQ ID NO: 11, 12, 17, 18 ,19 or 20.

A preferred embodiment of the invention is wherein the polypeptide having ADH enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 11 to 21. Preferably, the polypepeptide has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to the amino acid sequence provided in SEQ ID NO: 11 or 21. Preferably, the polypeptide has the amino acid sequence provided in SEQ ID NO: 11 or 21.

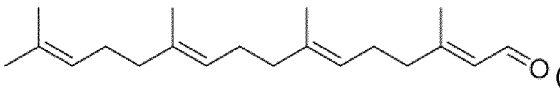
Step (ii) of the process of the invention

Step (ii) of the process of the invention relates to contacting a compound of formula (III) with a polypeptide having enal-cleaving enzyme activity.

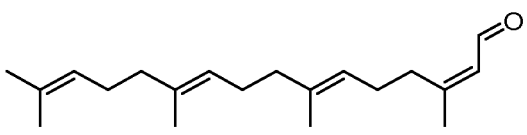

(formula III)

The compound of formula (III) is also known as geranylgeranial, i.e. 3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraenal; CAS No 32480-11-8.

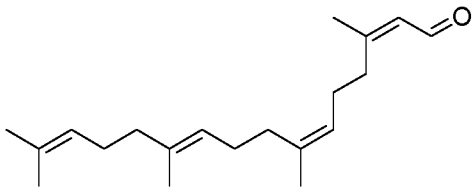
The compound of formula (III) may be present in any one of its stereoisomers or a mixture thereof. Specifically, the compound may have the following structures and isoforms:


(formula IIIa)

(2E,6E,10E)-geranylgeranial; (2E,6E,10E)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraenal; CAS No 13920-12-2.

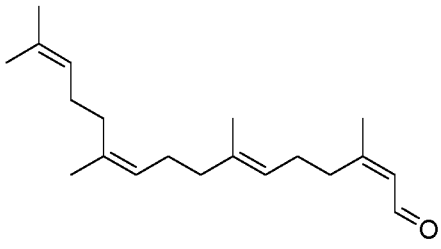

(formula IIIb)

(2Z,6E,10E)-geranylgeranial; (2Z,6E,10E)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraenal; CAS No 57784-38-0.



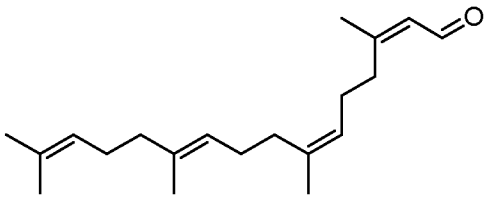
(formula IIIc)

(2Z,6Z,10E)-geranylgeranial; (2Z,6Z,10E)-3,7,11,15-Tetramethyl-2,6,10,14-hexadecatetraenal.



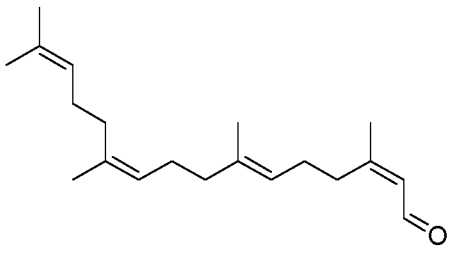
(formula IIId)

(2Z,6E,10Z)-geranylgeranial; (2Z,6E,10Z)-3,7,11,15-Tetramethyl-2,6,10,14-hexadecatetraenal.



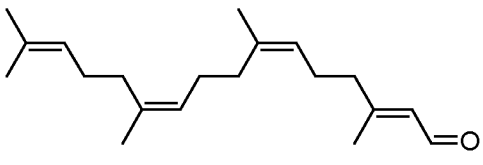
(formula IIIe)

(2Z,6Z,10E)-geranylgeranial; (2Z,6Z,10E)-3,7,11,15-Tetramethyl-2,6,10,14-hexadecatetraenal.



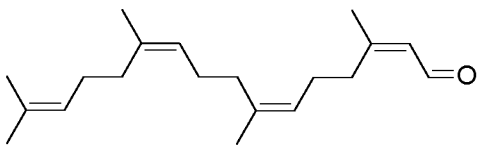
(formula IIIf)

(2Z,6E,10Z)-geranylgeranial; (2Z,6E,10Z)-3,7,11,15-Tetramethyl-2,6,10,14-hexadecatetraenal.



(formula IIIg)

(2E,6Z,10Z)-geranylgeranial; (2E,6Z,10Z)-3,7,11,15-Tetramethyl-2,6,10,14-hexadecatetraenal.



(formula IIIh)

(2Z,6Z,10Z)-geranylgeranial; (2Z,6Z,10Z)-3,7,11,15-Tetramethyl-2,6,10,14-hexadecatetraenal.

Step (ii) relates to the use of a polypeptide having enal-cleaving enzyme activity.

An “enal-cleaving enzyme” or “enal-cleaving protein” or “enal-cleaving polypeptide” in the context of the present invention designates an “α,β-unsaturated aldehyde carbon-carbon double bond-cleaving enzyme”, which also may be called a “α,β-unsaturated aldehyde C=C bond-cleaving enzyme” or “α,β-unsaturated aldehyde C=C-cleaving enzyme” or a “enal C=C-cleaving enzyme”. The enal-cleaving protein of the invention, based on protein domain organization, may also be described as a member of the “DUF4334 protein family” and/or as a member of the “GXWXG protein family” (SEQ ID NO: 263). Examples of such enzymes can be found in literature; for example, in WO2021005097.

More particularly, an enal cleaving enzyme of the invention has the ability to cleave terpenoid compounds containing an α,β-unsaturated aldehyde group, in particular geranylgeranial to farnesylacetone.

“Enal-cleaving enzyme activity” is determined under “standard conditions” as described herein below. It can be determined using recombinant enal-cleaving polypeptide expressing host cells, disrupted enal-cleaving polypeptide expressing cells, fractions of these or enriched or purified enal-cleaving polypeptide, in a culture medium or reaction medium, preferably buffered, having a pH in the range of 6 to 11, preferably 7 to 9, at a temperature in the range of about 20 to 45 °C, like about 25 to 40 °C, preferably 25 to 32 °C and in the presence of a reference substrate, here in particular geranylgeranial, either added at an initial concentration in the range of 1 to 100 µM, preferably 5 to 50 µM, in particular 30 to 40 µM, or endogenously produced by the host cell. The conversion reaction to form the respective cleavage product, like farnesylacetone is conducted from 10 min to 5 h, preferably about 1 to 2 h. The cleavage product may then be determined in conventional matter, for example after extraction with an organic solvent, like ethyl acetate.

The polypeptide having said enal-cleaving enzyme activity may be selected from the group of polypeptides containing:

a) at least one DUF4334 protein family domain having the Pfam ID number PF14232 (in particular within the C-terminal region of their amino acid sequence);
b) at least one GXWXG (SEQ ID NO: 263) protein family domain having the Pfam ID number PF14231 (in particular within the N-terminal region of their amino acid sequence); and/or
c) a domain retaining at least 90% sequence identity to PF14232 or PF14231.

In particular, a polypeptide of the invention having enal-cleaving enzyme activity is identified as a member of the DUF4334 protein family comprising said domain PF14232 if it matches with said domain with an e-value of less than $1x10^{-5}$, or less than $1x10^{-10}$, or less than $1x10^{-15}$, or less than $1x10^{-20}$, or less than $1x10^{-25}$, or less than $1x10^{-30}$, or less than or equal to $1x10^{-35}$, in particular in a range of $1x10^{-20}$ to $1x10^{-32}$ and more particular in a range of $1x10^{-25}$ to $1x10^{-31}$.

In particular, a polypeptide having enal-cleaving enzyme activity is identified as a member of GXWXG (SEW ID NO: 263) protein family comprising said domain PF14231 if it matches with an e-value of less than $1x10^{-5}$, or less than $1x10^{-10}$, or less than $1x10^{-15}$, or less than $1x10^{-20}$, or less than $1x10^{-25}$, or less than $1x10^{-30}$, or less than or equal to $1x10^{-35}$, in particular in a range of $1x10^{-20}$ to $1x10^{-30}$.

As the query sequence the sequence of a polypeptide having enal-cleaving enzyme activity is applied.

For example, the following website may be applied for the search and calculating such e-value: http://www.ebi.ac.uk/Tools/hmmer/search/hmmscan or http://www.ebi.ac.uk/Tools/pfa/pfamscan/.

Furthermore, the polypeptide having said enal-cleaving enzyme activity may be selected from the group of polypeptides that comprise at least one or more sequence motifs/domains selected from:

. G-[Y or "-"]-x-W-x-G-x-x-[F,L or I]-x-[T,S or R]-G-[H or D] (also expressed as GxxWxGxxxxxGx) set forth in SEQ ID NO: 235, or any partial motif thereof comprising up to 10 or up to 5 consecutive amino acid residues, as for example corresponding to residues in positions 1-8 or 9-13 of SEQ ID NO: 235 Here, X2 can be Y or can be deleted; X3 can be any naturally occurring amino acid; X5 can be any naturally occurring amino acid; X7 can be any naturally occurring amino acid; X8 can be any naturally occurring amino acid; X9 can be F, L, or I; X10 can be any naturally occurring amino acid; X11 can be R, S, or T; X13 can be H or D;

. W-[Y, A or V]-G-K-x-[F or Y]-x-[S or D] (also expressed as WxGKxxxx) set forth in SEQ ID NO: 236, or any partial motif thereof comprising up to 4 consecutive amino acid residues, as for example corresponding to residues in positions 1-4 or 5-8 of SEQ ID NO: 236. Here, X2 can be A, V, or Y; X5 can be any naturally occurring amino acid; X6 can be F or Y; X7 can be any naturally occurring amino acid; X8 can be D or S;

. [G or S]-x-[A or G]-x-[L or V]-x-x-x-x-[F, Y or L]-R-G-x-V (also expressed as xxxxxxxxxxRGxV) set forth in SEQ ID NO:237, or any partial motif thereof comprising up to 10 or up to 5 consecutive amino acid residues, as for example corresponding to residues in positions 1-8 or 9-14 of SEQ ID NO:237. Here X1 can be G or S; X2 can be any naturally occurring amino acid; X3 can be A or G; X4 can be any naturally occurring amino acid; X5 can be L or V; X6 can be any naturally occurring amino acid; X7 can be any naturally occurring amino acid; X8 can be any naturally occurring amino acid; X9 can be any naturally occurring amino acid; X10 can be F, L, or Y; X13 can be any naturally occurring amino acid; and

. [M or L]-[V or I]-Y-D-x-x-P-[I or V]-x-D-[H or S]-[F or L] (also expressed as xxYDxxPxxDxx) set forth in SEQ ID NO:238, or any partial motif thereof comprising up to 10 or up to 5 consecutive amino acid residues, as for example corresponding to residues in positions 1-6 or 7-12 of SEQ ID NO:238. Here X1 can be L or M; X2 can be I or V; X5 can be any naturally occurring amino acid; X6 can be any naturally occurring amino acid; X8 can be I or V; X9 can be any naturally occurring amino acid; X11 can be H or S; X12 can be F or L.

wherein

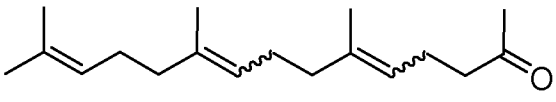
the numbering of X (e.g. X2) corresponds to its position in the relevant sequence. For example, X2 corresponds to X at position 2 in the relevant sequence; and

in the above motifs, residues x represent independently of each other any natural amino acid residue, and wherein optionally in each of the above motifs, 1, 2, 3, 4 or 5 amino acid residues different from the x residues may be modified, for example by amino acid substitution, in particular by conservative substitutions, provided that the enzyme retains, at least to analytically detectable extent, enal-cleaving enzyme activity. The function of the square brackets has been described above.

A preferred embodiment of the invention is wherein the polypeptide having enal-cleaving activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to the sequence provided in SEQ ID NO: 22.

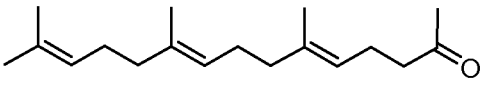
Step (iii) of the process of the invention

Step (iii) of the process of the invention relates to contacting a compound of formula (IV) with a polypeptide having BVMO enzyme activity.

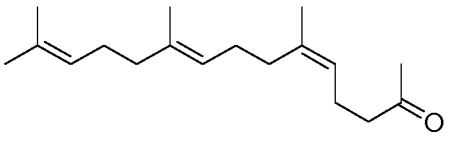

(formula IV)

The compound of formula (IV) is also known as farnesylacetone, i.e. 6,10,14-trimethylpentadeca-5,9,13-trien-2-one; CAS No 762-29-8.

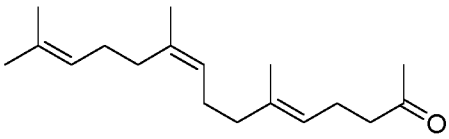
The compound of formula (IV) may be present in any one of its stereoisomers or a mixture thereof. Specifically, the compound may have the following structures and isoforms:


(formula IVa)

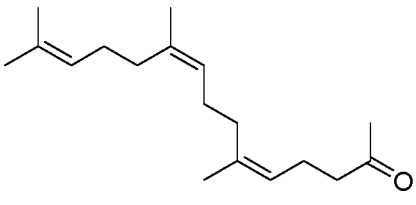
(5E,9E)-farnesylacetone; (5E,9E)-6,10,14-trimethylpentadeca-5,9,13-trien-2-one; CAS No 1117-52-8.


(formula IVb)

(5Z,9E)-farnesylacetone; (5Z,9E)-6,10,14-trimethylpentadeca-5,9,13-trien-2-one; CAS No 1117-51-7.


(formula IVc)

(5E,9Z)-farnesylacetone; (5E,9Z)-6,10,14-trimethylpentadeca-5,9,13-trien-2-one; CAS No 3053-35-3.



(formula IVd)

(5Z,9Z)-farnesylacetone; (5Z,9Z)-6,10,14-trimethylpentadeca-5,9,13-trien-2-one; Cas No 3796-69-8.

Step (iii) relates to the use of a polypeptide having BVMO enzyme activity.

"Baeyer-Villiger monooxygenases" (BVMOs) are flavoenzymes and belong to the class of refers to a polypeptide having oxidoreductase activity (EC 1.14.13.X). They catalyze the oxidation of linear, cyclic (aromatic or non-aromatic) aldehydes or ketones to the corresponding esters or lactones, highly similar to the chemical Baeyer-Villiger oxidation. During the enzymatic oxidation one atom of molecular oxygen is incorporated into a carbon-carbon bond of a non-activated carbonyl compound. The BVMOs require NADPH or NADH as cofactor or accept both. They also require molecular oxygen as co-substrate. More particularly, a BVMO of the invention has the ability to oxidize terpene- derived aldehydes or ketones, like for example linear terpenoid carbonyl compounds, in particular farnesylacetone to the respective carbonyl ester.

"BVMO enzyme activity" is determined under "standard conditions" as described herein below: It can be determined using recombinant BVMO expressing host cells, disrupted BVMO expressing cells, fractions of these or enriched or purified BVMO enzyme, in a culture medium or reaction medium, preferably buffered, having a pH in the range of 6 to 11, preferably 7 to 9, at a temperature in the range of about 20 to 45 °C, like about 25 to 40 °C, preferably 25 to 32 °C and in the presence of a reference substrate, here in particular farnesylacetone, either added at an initial concentration in the range of 1 to 100 µM, preferably 5 to 50 µM, in particular 30 to 40 µM, or endogenously produced by the host cell and in the presence of molecular oxygen. For *in-vitro* assays a cofactor selected from NADH and NADPH has to be added in a suitable easily to be determined concentration range of the conversion reaction to form the respective enzyme product, like homofarnesyl acetate in the case of farnesylacetone is conducted from 10 min to 5 h, preferably about 1 to 2 h. The BVMO product may then be determined in conventional matter, for example after extraction with an organic solvent, like ethyl acetate.

A further method to screen for BVMOs and evaluate the conversion of farnesylacetone to homofarnesyl acetate is described in Example 2.

The polypeptide having BVMO enzyme activity may be selected from:

(1) the group of polypeptides containing a flavin-containing monooxygenase (FMO) protein family domain having the Pfam ID number PF00743 within their amino acid sequence; or a domain retaining at least 90%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to PF00743;

In particular, a polypeptide having BVMO activity is identified as member of the FMO protein family comprising said domain PF00743 if it matches with said domain with an e-value of less than $1x10^{-5}$ or less than $1x10^{-10}$, or less than or equal to $1x10^{-15}$, or less than or equal to $1x10^{-18}$, in particular in a range of $1x10^{-10}$ to $1x10^{-18}$ and more particular, in a range of $1x10^{-14}$ to $1x10^{-17}$. As the query sequence, the sequence of a polypeptide having BVMO activity is applied.

For example, the following website may be applied for the search and calculating such e-value: http://www.ebi.ac.uk/Tools/hmmer/search/hmmscan or http://www.ebi.ac.uk/Tools/pfa/pfamscan/.

and/or

(2) the group of polypeptides that comprise at least one or more of the sequence motifs/domains selected from:

. GxGxxG (SEQ ID NO: 239), as for example in any of SEQ ID NOs: 23 to 26. Here, X4 can be any naturally occurring amino acid, particularly A or I. The numbering of X corresponds to its position in the sequence.

. [GS]GxWxxxxYPGxxxD (SEQ ID NO: 240), as for example in any of SEQ ID NOs: 23 to 26;

. Gxxx[FY]xGxxx[HS]xxxW (SEQ ID NO: 241), as for example in any of SEQ ID NOs: 23 to 26; and

. [KQ]x[VI]xx[IV]GxG (SEQ ID NO: 242), as for example in any of SEQ ID NOs: 23 to 26.

wherein

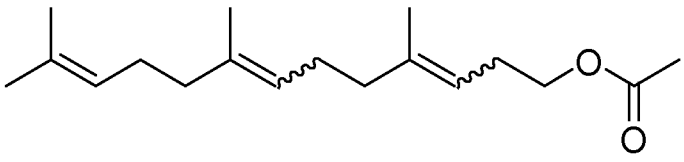
in the above motifs, residues x represent independently of each other any natural amino acid residue, and wherein optionally in each of the above motifs, 1, 2, 3, 4 or 5 of the conserved amino acid residues (i.e. different from the x residues) may be modified, for example by amino acid substitution, in particular by conservative substitutions, provided that the enzymes retains, at least to analytically detectable extent, BVMO enzyme activity. The function of the square brackets has been described above.

A preferred embodiment of the invention is wherein the polypeptide having BVMO enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 23 to 26. A preferred embodiment of the invention is wherein the polypeptide having BVMO enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 25 or 26. Preferably, the polypeptide has the amino acid sequence provided in SEQ ID NO: 25 or 26.

Alternatively, the polypeptide having BVMO enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 216 to 227.

Step (iv) of the process of the invention

Step (iv) of the process of the invention relates to contacting a compound of formula (V) with a polypeptide having esterase enzyme activity.

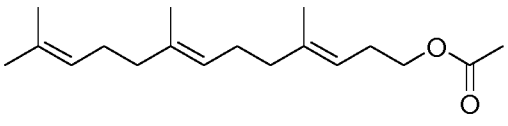


(formula V)

The compound of formula (V) is also known as homofarnesyl acetate, i.e. 4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate; CAS No 109813-25-4.

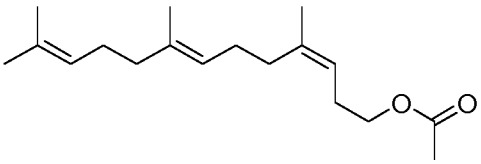
The compound of formula (V) may be present in any one of its stereoisomers or a mixture thereof.

Specifically, the compound may have the following structures and isoforms:



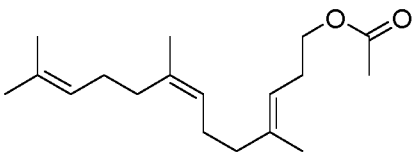
(formula Va)

(3E,7E)- homofarnesyl acetate; (3E,7E)-4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate; CAS No 944346-19-4.



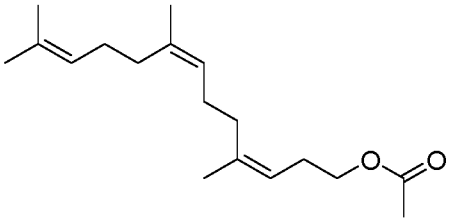
(formula Vb)

(3Z,7E)- homofarnesyl acetate; (3Z,7E)-4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate; CAS No 1467099-77-9.



(formula Vc)

(3E,7Z)- homofarnesyl acetate; (3E,7Z)-4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate.



(formula Vd)

(3Z,7Z)- homofarnesyl acetate; (3Z,7Z)-4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate.

Step (iv) relates to the use of a polypeptide having esterase enzyme activity.

An “esterase” refers to a polypeptide having hydrolase activity that splits esters into an acid and an alcohol in a chemical reaction with water (hydrolysis). Esterases in the context of the present invention are selected from the class of carboxylic ester hydrolases (EC 3.1.1.-), which splits off acyl groups, like acetyl or formyl groups, from the respective ester substrate. More particularly, an esterase of the invention has the ability to cleave terpenyl ester compounds, like homofarnesyl acetate, to form the corresponding alcohol, in particular homofarnesol.

“Esterase enzyme activity” is determined under “standard conditions” as described herein below: It can be determined using recombinant esterase polypeptide expressing host cells, disrupted esterase polypeptide expressing cells, fractions of these or enriched or purified esterase polypeptide, in a culture medium or reaction medium, preferably buffered, having a pH in the range of 6 to 11, preferably 7 to 9, at a temperature in the range of about 20 to 45 °C, like about 25 to 40°C, preferably 25 to 32 °C and in the presence of a reference substrate, here in particular homofarnesyl acetate, either added at an initial concentration in the range of 1 to 100 µM preferably 5 to 50 µM, in particular 30 to 40 µM, or endogenously produced by the host cell. The conversion reaction to form the respective alcohol, in particular homofarnesol is conducted from 10 min to 5 h, preferably about 1 to 2 h. The detection and quantification of esterase product may then be determined in conventional matter, for example after extraction with an organic solvent, like ethyl acetate.

A further method to evaluate the conversion of homofarnesyl acetate to homofarnesol by an esterase is described in Example 4.

A preferred embodiment of the invention is wherein the polypeptide having said esterase activity comprises at least one or more sequence motifs selected from:
. AxVVxVxxRLAPE (SEQ ID NO: 243), as for example in SEQ ID NO: 27 or 28;
. GASAGGGLxA (SEQ ID NO: 244), as for example in SEQ ID NO: 27 or 28;
. VxQLLxYPMLDDR (SEQ ID NO: 245), as for example in SEQ ID NO: 27 or 28; and,
. ARxxDLSGLPxT (SEQ ID NO: 246), as for example in SEQ ID NO: 27 or 28;
wherein
in the above motifs, residues x represent independently of each other any natural amino acid residue, and wherein optionally in each of the above motifs, 1, 2, 3 or 4 amino acid residues different from the x residues may be modified, for example by amino acid substitution, in particular by conservative substitutions, provided that the enzyme retains, at least to analytically detectable extent, esterase enzyme activity.

Preferably, the polypeptide having said esterase activity comprises: AxVVxVxxRLAPE (SEQ ID NO: 243), GASAGGGLxA (SEQ ID NO: 244), VxQLLxYPMLDDR (SEQ ID NO: 245), and ARxxDLSGLPxT (SEQ ID NO: 246), as for example in SEQ ID NO: 27 or 28.

A preferred embodiment of the invention is wherein the polypeptide having esterase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 27 or 28. A preferred embodiment of the invention is wherein the polypeptide having esterase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO 28. Preferably, the polypeptide has the amino acid sequence provided in SEQ ID NO: 28.

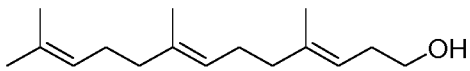
Step (v) of the process of the invention

Step (v) of the process of the invention relates to contacting a compound of formula (VI) with a polypeptide having terpene cyclase enzyme activity.

(formula VI)

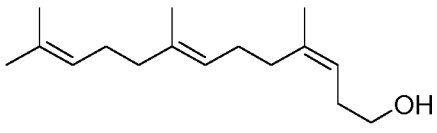
The compound of formula (VI) is also known as homofarnesol, 4 ,8,12-trimethyltrideca-3,7,11-trien-1-ol CAS No 35826-67-6.

The compound of formula (VI) may be present in any one of its stereoisomers or a mixture thereof. Specifically, the compound may have the following structures and isoforms:



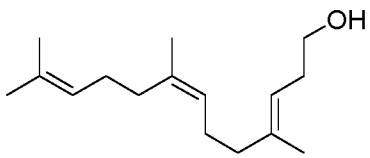
(formula VIa)

(3E,7E)-homofarnesol; (3E,7E)-4,8,12-trimethyltrideca-3,7,11-trien-1-ol; CAS No 459-89-2.



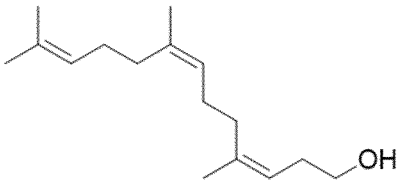
(formula VIb)

(3Z,7E)-homofarnesol; (3Z,7E)-4,8,12-trimethyltrideca-3,7,11-trien-1-ol; CAS No 138152-06-4



(formula VIc)

(3E,7Z)-homofarnesol; (3E,7Z)-4,8,12-trimethyltrideca-3,7,11-trien-1-ol; CAS No 2032064-12-1.



(formula VId)

(3Z,7Z)-homofarnesol; (3Z,7z)-4,8,12-trimethyltrideca-3,7,11-trien-1-ol; CAS No 138152-08-6.

Terpene cyclases are divided into two categories depending on the way the initial carbocation is generated. In class I (or type I) terpene cyclase, the diphosphate group of the linear terpenoid precursor is abstracted to form an allylic carbocation on the terpene moiety. In class II, the initial carbocation is formed by protonation of a double bond or epoxy group in the terpene carbon chain. Thus, class I cyclase necessarily use substrates with a diphosphate group, while class II cyclase (since they do not need a diphosphate group for the generation of the initial carbocation) can use terpenoids as substrates.

For all terpene cyclases, the generated reactive carbocation species triggers the subsequent cascade reaction including carbocation reactions with double bonds, alkyl-shifts, hydride shifts or carbon-carbon bound formation. The reaction can be terminated by deprotonation of a carbon atom adjacent to the carbocation or by quenching of the carbocation with a hydroxyl group or molecule of water.

The type II activity in terpene cyclases is associated with aspartate-rich conserved motifs.

Typical examples of class II terpene cyclases are the class II diterpene cyclases catalyzing the protonation-initiated cyclization of geranylgeranyl-diphosphate into for example, labdadienyl-diphosphate intermediates or other cyclic diphosphate intermediates (Peters, R. J. (2010). Nat. Prod. Rep. 27, 1521–1530; Zerbe, P. et al (2015). Plant J. 83, 783–793).

Squalene cyclases (SHCs) constitute a classical example of class II terpene cyclases where the substrate does not contain a diphosphate functional group. The squalene cyclase enzyme family comprise squalene cyclases and 2,3-oxidosqualene cyclases and enzymes catalyzing mechanistically related cyclization reactions. Squalene cyclases catalyze a protonation-initiated cyclization cascade of a linear terpene to a cyclic compound. Thus, squalene cyclases are class II terpene cyclases. The squalene family includes for example squalene-hopene cyclases catalyzing the cyclization of squalene to hopene (EC 5.4.99.17) and squalene-hopanol cyclases catalyzing the cyclization of squalene to hopan-22-ol (EC 4.2.1.129). Tetraprenyl-β-curcumene-sporulenol cyclases catalyze similar class II cyclization of linear terpene substrate (EC 4.2.1.137). It was shown that tetraprenyl-β-curcumene-sporulenol cyclase can also catalyze the cyclization of squalene (Sato, T., et al. (2011). Journal of the American Chemical Society 133(44): 17540-17543), thus tetraprenyl-β-curcumene-sporulenol cyclases are also members of the squalene cyclase family.

Squalene cyclase polypeptides have typically a length between 600 and 800 amino acids and are membrane-associated proteins. They bind to the surface of cellular membranes but do not contain a

transmembrane region. Squalene cyclases are classified in the IPR018333 family of the InterPro protein sequence classification database (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR018333/) (InterPro release 93.0, 2nd March 2023). The structure of squalene cyclases is organized in two domains comprising several alpha-helices, recognized as the β-domain and γ-domain or the βγ-domain architecture (Christianson DW, Chem. Rev, 2017, 117, 11570–11648). The two domains have characteristic sequence signatures as described in the Pfam database under the Pfam Squalene-hopene cyclase N-terminal domain (PF13249) and Squalene-hopene cyclase C-terminal domain and (PF13243) (Pfam 35.0 released, 19 November 2021). The presence of the IPR018333, PF13249 or PF13243 protein sequences signatures can be predicted using the NCBI conserved domain search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) or the InterProScan tool (http://www.ebi.ac.uk/interpro/search/sequence/).

The squalene cyclase polypeptide contains characteristic conserved amino acid motifs located along the sequence and associated with the protein architecture and enzymatic reaction. In particular, the squalene cyclase contains at least one or more amino acid motifs selected from:

. [SP][TP][VIL]WDTx[LWI] (SEQ ID NO: 247),
. PGG[WF][GYA]F (SEQ ID NO: 248),
. PDxDD[TAS][TIAS] (SEQ ID NO: 249),
. [MIL]QxxxG[GA][WF]x[AS][FY] (SEQ ID NO: 250),
. Qxxx[GH]xWxG[RK]WGxx[YF]xYG (SEQ ID NO: 251),
. Qxx[DN]G[GS][WF][GS]ExxxS (SEQ ID NO: 252), and
. [STA]xx[SFN][QC]T[AGT]W[AS][LIV]xx[LQ] (SEQ ID NO: 253)

The motif sequences are described using the standard IUPAC one-letter codes for the amino acids. Ambiguities are indicated by listing the acceptable amino acids for a given position between brackets. For example, [SP] or [S or P] stands for S (serine), or P (proline). The “x” represents positions where independently of each other any natural amino acid residue is present. The function of the square brackets has been described above.

Meroterpenoids are hybrid secondary metabolites derived from mixed biosynthetic pathways and are partially derived from a terpenoid co-substrate (Cornforth, J.W. Terpenoid biosynthesis. Chem. Br. 1968, 4, 102–106). The non-terpenoid part can originate for example from polyketides, alkaloids, phenols, or amino acids biosynthetic pathway. Large chemical diversity is found among meroterpenoids, in particular in bacteria and in fungi.

The meroterpenoids biosynthetic pathways follow several modular biosynthetic steps. In the first step, the building blocks are generated from the corresponding biosynthetic pathway (e.g. terpenoids, polyketides). The terpenoid and non-terpenoid parts are assembled by prenyltransferases. The precursors of the terpenoid parts are generally linear terpenoid-diphosphates such as geranyl-diphosphate, farnesyl-diphosphate or geranylgeranyl-diphosphate.

In the following step, the linear polyene terpenoid part of the hybrid precursor is cyclized to form a monocyclic or polycyclic structure. This cyclization is catalyzed by a specific class of non-canonical class II terpene cyclases named meroterpenoid cyclase first discovered in fungi (T. Itoh et al, 2010, 2, 858–864). The first discovered representative meroterpenoid cyclase is Pyr4 from *Aspergillus fumigatus* Af293 (Itoh, T., et al. (2010). Nature Chemistry 2(10): 858-864).

In many meroterpenoids, the linear terpenoid precursor is first activated by a stereoselective epoxidation by a monooxygenase of one of the double-bonds. The meroterpenoid cyclases catalyze then the protonation of the epoxide moiety generating a reactive carbocation species and triggering a subsequent cascade reaction similar to other terpene cyclases. Some meroterpenoid cyclases can convert the isoprenic precursors to cyclized products without the involvement of a prior epoxidation step. These meroterpenoid cyclases are able to directly protonate the terminal double bond generating a reactive carbocation and catalyzing a cyclization. For example, MacJ from the fungi *Penicillium terrestry* was the first identified fungi meroterpenoid cyclase using a type II double-bond protonation initiations reaction (Tang, M.-C., et al. (2017). Organic Letters 19(19): 5376-5379). Another example of meroterpenoid cyclase which initiates polyene cyclization by direct double bond protonation is DmtA1 from bacteria (*Streptomyces youssoufiensis* OUC68199) (Yao et al, Nat. Commun., 2018, 9, 4091).

Like other class II terpene cyclases, the carbocation generated by meroterpenoid cyclases triggers a cascade reaction generally starting by the attack of a double bond and generating monocyclic or polycyclic structure with a tertiary carbocation. The reaction is terminated either by deprotonation to form a double bond or by reacting with a water molecule to generate a tertiary alcohol. Typical cyclic structures found in meroterpenoids compounds contain drimane or labdane scafolds.

The largest group of meroterpenoid cyclases are compact membrane-integrated proteins containing several (generally seven) transmembrane helices. This protein architecture based on transmembrane helices can easily be predicted using for example the TMHMM 2.0 server available at https://dtu.biolib.com/DeepTMHMM (Krogh, A., et al. (2001) J Mol Biol 305(3): 567-580.). In addition to the protein architecture, meroterpenoid cyclases differ from other class II cyclases such as the squalene cyclases by their smaller polypeptide size. The bacterial and fungal meroterpenoid cyclase polypetides have a length ranging from 150 to 550 residues. The transmembrane helices are located over a portion of the polypetide covering 180 to 300 amino acid and carry the catalytic domains.

Recently meroterpenoid cyclases having a protein architecture different from the membrane-integrated meroterpenoid cyclases were described. For example, MstE from the bacteria *Scytonema sp.* PCC 1002 is a soluble cyclase having a structure similar to canonical cyclases such as diterpene synthases and squalene cyclases, but nevertheless different, since it is a monodomain protein with only an α-domain (Moosmann, P., et al. (2020). Nat Chem 12(10): 968-972). Soluble bacterial meroterpenoid cyclase polypetides have length ranging from 150 to 550.

Several meroterpenoid cyclases catalyze reactions of cyclisation of the terpenoid part of the meroterpenoid hybrid precursor to labdane cyclic structures. However, the cyclization of linear terpenoids by meroterpenoid cyclases to labdane compounds have so far not been shown.

Meroterpenoid cyclase polypeptides contain characteristic conserved amino acid motifs located along the sequence and associated with the protein architecture or enzymatic reaction as follows:

Membrane-integrated meroterpenoid cyclase of bacterial origin containing at least one or more amino acid motifs selected from:
. [W]xxx[D]xx[ILVMN] (SEQ ID NO: 254);
. PxxAxxxNxxWE (SEQ ID NO: 255);
. MxxxFxxMLxxR (SEQ ID NO: 256); and
. RxxxxGQS (SEQ ID NO: 257).

Membrane-integrated meroterpenoid cyclase of fungal origin containing a least one or more amino acid motifs selected from:
. [WY]Exx[YFW] (SEQ ID NO: 258); and
. [DNE]xSYxxP (SEQ ID NO: 259).

Soluble meroterpenoid cyclases of bacterial origin containing a least one or more amino acid motifs selected from:
. GxWxxxW[WG]xxxxY (SEQ ID NO: 260);
. WxxxHxxV[TSA] (SEQ ID NO: 261); and
. GxWxD[FY] (SEQ ID NO: 262).

The motif sequences are described using the standard IUPAC one-letter codes for the amino acids. Residues x represent independently of each other any natural amino acid residue, and wherein optionally in each of the above motifs, 1, 2, 3 or 4 amino acid residues different from the x residues may be modified, for example by amino acid substitution, in particular by conservative substitutions, provided that the enzyme retains, at least to analytically detectable extent, its enzyme activity. The function of the square brackets has been described above.

Meroterpenoid cyclase polypeptides can be searched in sequences databases using for example the BLAST search tools (Tatiana et al, *FEMS Microbiol Lett.*, 1999, 174:247-250, 1999) using as query sequences MacJ (SEQ ID NO: 71), DmTA1 (SEQ ID NO: 77) or MstE (SEQ ID NO: 76). The selection can further be refined by selecting sequences with appropriate length or containing the characteristic amino acid motifs as described above. The selection can also be refined bases using a prediction of the protein architecture, in particular by predicting the presence of the transmembrane helices as described above.

Particular examples of suitable standard conditions for each of the above-described enzyme activities may be taken from the Examples section below.

As discussed above, the terpene cyclase may be a squalene cyclase (SHC) or a meroterpenoid cyclase (MeroTPS).

For the avoidance of doubt, SHCs and meroterpenoid cyclases are distinct classes of enzymes which can be distinguished by physical characteristics. Furthermore, meroterpenoid cyclases can be classified as (i) bacterial membrane-integrated meroterpenoid cyclases; (ii) fungal membrane-integrated meroterpenoid cyclases; (iii) bacterial soluble meroterpenoid cyclases.

Table A below outlines the differences between SHCs and the different types of meroterpenoid cyclases.

Feature	**SHC**	**Bacterial membrane-integrated meroTPS**	**Fungal membrane-integrated meroTPS**	**Bacterial soluble MeroTPS**
Protein size (# amino acids)	600-800	150-550 (180-300 for region containing the helices)	150-550 (180-300 for region containing the helices)	150 to 550
Transmembrane region	No	Yes	Yes	No
	Membrane-associated	membrane-integrated	membrane-integrated	Soluble
Protein structure	βγ-domain architecture (2 domains)	transmembrane helices	transmembrane helices	α-domain architecture (monodomain protein)

Table A: enzyme characteristics

Hence the skilled person can, from the information provided herein, readily identify whether an enzyme is a SHC enzyme or a class of meroterpenoid cyclase enzyme.

A preferred embodiment of the invention is wherein the terpene cyclase is a SHC.

As provided in the accompanying examples, the inventors have demonstrated that SHC enzymes can be used in step (v) of the process of the invention. A preferred embodiment of the invention is wherein the polypeptide having SHC enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 29 to 49 and 265 to 279. Preferably, the polypeptide having SHC enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 29 to 49, 265 to 274 and 276 to 279. More

preferably, the polypeptide having SHC enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 48, 265, 266, 267, 268, 274, 276, and 279. More preferably, the polypeptide having SHC enzyme activity has the amino acid sequence of any of SEQ ID NOs: 48, 265, 266, 267, 268, 274, 276, and 279.

Alternatively, the polypeptide having SHC enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 79 to 89.

A preferred embodiment of the invention is wherein the terpene cyclase is a meroterpenoid cyclase.

When preparing the process of the present invention, the inventors sought to compare the isomeric profile of the compound of formula (I) synthesized via SHCs and meroterpenoid cyclases.

To their surprise, they found that the meroterpenoid cyclases when used in the process of the invention, produced compound of formula (I) having an isomeric bias towards isomers of compound of formula (I) having preferred olfactory profiles more than that produced by SHC enzymes.

In particular, they demonstrated that less than 1% of the compound of formula (I) produced by the process of the invention were in the isomeric form of formula I(c) and/or I(d). Hence, the use of meroterpenoid cyclases is associated with a surprising technical advantage over the use of SHC enzymes.

Preferably, the meroterpenoid cyclase is a membrane-integrated meroterpenoid cyclase. It can be seen from the accompanying examples that this class of meroterpenoid cyclases preferably produce compounds of formula (Ia) rather than (Ic) and/or (Id). Examples of membrane-integrated meroterpenoid cyclase include those provided in any of SEQ ID NOs: 50 to 73 and 280 to 289. SEQ ID NOs: 50 to 70 and 280 to 289 are membrane-integrated meroterpenoid cyclases of bacterial origin and SEQ ID NOs: 71 to 73 are membrane-integrated meroterpenoid cyclases of fungal origin.

Preferably, the meroterpenoid cyclase is a soluble meroterpenoid cyclase. It can be seen from the accompanying examples that this class of meroterpenoid cyclases preferably produce compounds of formula (Ib) rather than (Ic) and/or (Id). Examples of soluble meroterpenoid cyclase include those provided in SEQ ID NO: 74 or 75.

This is the first time meroterpenoid cyclases have been used to prepare a compound of formula (I), and the bias towards such isomeric forms is surprising and technically important commercially.

Furthermore, the present inventors also surprisingly found that meroterpenoid cyclases can be used in bioconversion processes without the need for the addition of any detergent to the reaction. The absence

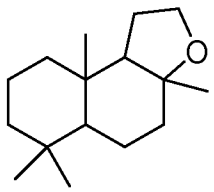
of detergent in the bioconversion reaction leads to a simplified and hence more cost-effective process for the preparation of a compound of formula (I).

As provided in the accompanying Examples, the inventors have demonstrated that meroterpenoid cyclase enzymes can be used in step (v) of the process of the invention. A preferred embodiment of the invention is wherein the polypeptide having meroterpenoid cyclase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 50 to 75 and 280 to 289. Preferably, the meroterpenoid cyclase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 57, 71, 74, 280, 281, 282, 283, 286, 287 and 288. Preferably, the meroterpenoid cyclase enzyme activity has the sequence provided in SEQ ID NO: 57, 71, 74, 280, 281, 282, 283, 286, 287 or 288.

Forms of compound of formula (I) produced by the process of the invention.

The process of the invention is for the preparation of a compound of formula (I).

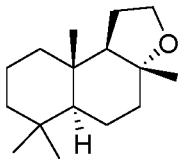
The compound of formula (I)



formula (I)

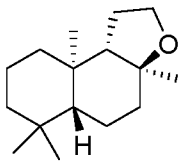
The compound of formula (I) is also known as 3a,6,6,9a tetramethyldodecahydronaphtho[2,1-b]furan; CAS No 3738-00-9.

The compound of formula (I) may be present in any one of its stereoisomers or a mixture thereof. Specifically, the compound may have the following structures and isoforms:



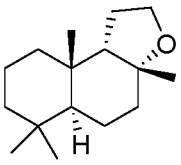
(formula Ia)

(3aR,5aS,9aS,9bR)-3a,6,6,9a-tetramethyldodecahydronaphtho[2,1-b]furan; CAS No 6790-58-5.



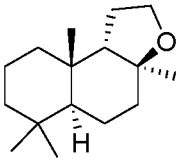
(formula Ib)

(3aS,5aR,9aR,9bS)-3a,6,6,9a-tetramethyldodecahydronaphtho[2,1-b]furan; CAS No 234431-64-2.



(formula Ic)

(3aR,5aS,9aS,9bS)-3a,6,6,9a-tetramethyldodecahydronaphtho[2,1-b]furan.



(formula Id)

(3aS,5aS,9aS,9bS)-3a,6,6,9a-tetramethyldodecahydronaphtho[2,1-b]furan.

A compound of formula (I) can have different enantiomeric forms. Some isomers have a preferred olfactory profile to alternative isomeric forms of the compound. In particular, the olfactory preferred form of compound (I) is the compound (Ia) and/or (Ib).

To improve the yield or ratio of compound (Ia) and/or (Ib) to other isomeric forms of compound (I), achieving high selectivity in the generation of compound (VI) plays an important role. Specifically, the ratio of the E/Z isomers at the 3,4-double bond of compound (VI) has significant importance. The preferred form of compound (VI) is that of formula (VIa).

Despite extensive efforts, using chemical methods to obtain a compound of formula (VI) with a E/Z ratio at the 3,4-double bond higher higher than 90:10 remains a challenge and is so far not accessible at large scale (Eichhorn, E. and F. Schroeder (2023). J Agric Food Chem).

The present invention achieves this goal by the use of highly selective enzymes. In the present invention, high selectivity of production of a compound of formula (VI) in the form of formula (VIa) has been achieved, as demonstrated in the accompanying Examples. Chemical methods (as described in Eichhorn, E. and F. Schroeder (2023). J Agric Food Chem) do not allow to achieve such high selectivity. Compounds of the pathway with high a E/Z ratio can be obtained using enzymatic pathways. In particular geranylgeranyl-diphopshate and geranylgeraniol with high E/Z ratio of the double bonds can be achieve using enzymes and in particular, geranylgeranyl-diphopshate synthase. The configuration of the double bonds is retained in all intermediates of the pathway.

This subsequently leads to a process in which olfactory preferred forms of a compound of formula (I) are prepared. Hence, the process of the invention to prepare a compound of formula (I) lacking substantial amount of undesirable side products is of significant commercial importance. As can be shown in the accompanying Examples, more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

Accordingly, an embodiment of the present invention is wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

Also included in the scope of the present invention is a compound of formula (I) obtained or obtainable by an *in vivo* process.

It can be appreciated that this is the first time a compound of formula (I) has been prepared by in *in vivo* process. For the reasons outlined herein, this is an important advance in the preparation of this compound, from technical and commercial aspects. The compound of formula (I) may be prepared via any *in vivo* process, preferably using recombinant cells expressing enzymes which can be used in the pathway to synthesise this molecule. An embodiment of the present invention is wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

Embodiments of the process of the invention.

The process of the invention comprises using (i) a polypeptide having ADH enzyme activity, (ii) a polypeptide having enal-cleaving enzyme activity; (iii) a polypeptide having BVMO enzyme activity; (iv) a polypeptide having esterase enzyme activity; and (v) a polypeptide having terpene cyclase enzyme activity.

In a further preferred embodiment of the process of the invention is wherein:

(i) the polypeptide having ADH enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 11 to 21;

(ii) the polypeptide having enal-cleaving enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 22;

(iii) the polypeptide having BVMO enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 23 to 26 and 216 to 227; preferably, to any of SEQ ID NOs: 23 to 26;

(iv) the polypeptide having esterase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 27 and 28; and/or,

(v) the polypeptide having terpene cyclase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 29 to 75, 79 to 89 and 265 to 289; preferably, to any of SEQ ID NOs: 29 to 75 and 265 to 289; more preferably, to any of SEQ ID NOs: 29 to 75, 265 to 274 and 276 to 289.

A further preferred embodiment of the invention is wherein:

(i) the polypeptide having ADH enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 11 or 21; preferably the ADH enzyme has the sequence of SEQ ID NO: 11 or 21;

(ii) the polypeptide having enal-cleaving enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 22; preferably the enal-cleaving enzyme has the sequence of SEQ ID NO: 22;

(iii) the polypeptide having BVMO enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 25 or 26; preferably the BVMO enzyme has the sequence of SEQ ID NO: 25 or 26;

(iv) the polypeptide having esterase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO 28; preferably the esterase enzyme has the sequence of SEQ ID NO: 28; and/or,

(v) the polypeptide having terpene cyclase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 48, 57, 71, 74, 265, 266, 267, 268, 274, 276, 279, 280, 281, 282, 283, 286, 287 or 288; preferably, the terpene cyclase enzyme has the sequence of SEQ ID NO: 48, 57, 71, 74, 265, 266, 267, 268, 274, 276, 279, 280, 281, 282, 283, 286, 287 or 288.

Alternatively, in this embodiment of the invention:

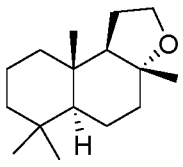
(v) the polypeptide having terpene cyclase enzyme activity is a SHC enzyme and has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 48, 265, 266, 267, 268, 274, 276 or 279; preferably, the SHC enzyme has the sequence of SEQ ID NO: 48, 265, 266, 267, 268, 274, 276 or 279.

Alternatively, in this embodiment of the invention:

(v) the polypeptide having terpene cyclase enzyme activity is a meroterpenoid cyclase enzyme and has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 57, 71, 74, 280, 281, 282, 283, 286, 287 or 288; preferably, the meroterpenoid cyclase enzyme has the sequence of SEQ ID NO: 57, 71, 74, 280, 281, 282, 283, 286, 287 or 288.

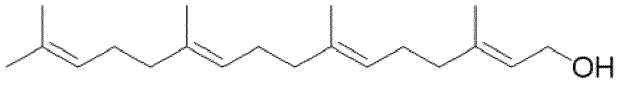
A further embodiment of the invention is wherein:

(i) the compound of formula (I) is in the form of formula (Ia):



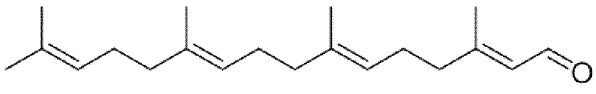
(formula Ia)

(ii) the compound of formula (II) is in the form of formula (IIa):



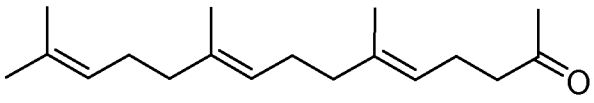
(formula IIa);

(iii) the compound of formula (III) is in the form of formula (IIIa):



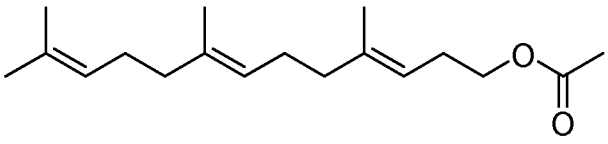
(formula IIIa);

(iv) the compound of formula (IV) is in the form of formula (IVa):



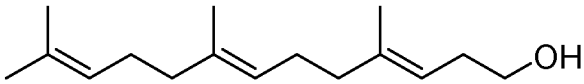
(formula IVa);

(v) the compound of formula (V) is in the form of formula (Va):



(formula Va);

(vi) the compound of formula (VI) is in the form of formula (Va):



(formula VIa).

Preparation of a compound of formula (II)

The present aspect of the invention provides a process of preparing a compound of formula (I) from the sequential biocatalytic conversion of a compound of formula (II).

The compound of formula (II) can be used as the starting substrate in the process of the invention, for example in the form of a purified compound preparation.

However, a preferred embodiment of the invention is wherein the process of the invention further comprises providing a compound of formula (II) by the sequential biocatalytic conversion of precursor compounds to the compound of formula (II).

Compounds of formulas (I) to (VI) are terpenoids.

Terpenoids is a large family of structurally diverse natural compounds. All terpenoids derive biosynthetically from two five-carbon units, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate

(DMAPP). IPP and DMAPP can be produced from different biosynthetic pathways such as the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway, the mevalonate (MVA) pathway or alternative MVA pathways (Dellas, N., et al. (2013) eLife 2: e00672). Alternatively, IPP and DMAPP can also be formed by successive enzymatic phosphorylation or by enzymatic pyrophosphorylation of their corresponding alcohols, isoprenol and prenol (Ma, X, et al. (2022). J Agric Food Chem 70(11): 3512-3520).

These terpene building blocks are condensed successively to form linear terpenoid precursors with various length and multiple of five carbon such as geranyl-diphosphate (GPP), farnesyl-diphosphate (FPP) or geranylgeranyl-diphosphate (GGPP) containing 10, 15 and 15 carbons, respectively. The condensation of the IPP and DMAPP is performed by a class of enzyme named prenyltransferases. Prenyltransferase enzymes catalyze the initial condensation reaction between IPP and DMAPP to give GPP, and the subsequent addition of IPP molecules to give FPP and then GGPP (Ogura, K., and Koyama, T. (1998). Chem. Rev. 98, 1263–1276). The successive condensation of DMAPP and IPP to GGPP can be performed by:

i. the successive action of 3 prenyltransferases, a GPP synthase, a FPP synthase catalyzing the addition of one IPP to GPP and a GGPP synthase catalyzing the addition of one IPP to FPP;

ii. the combination of 2 prenyltransferases, for example a FPP synthase catalyzing the condensation of one DMAPP and two IPP and a GGPP synthase catalyzing the addition of one IPP to FPP; and/or

iii. the action of 1 prenyltransferase, for example a GGPP synthase capable of catalyzing the the successive condensation of tree IPP and one DMAPP.

Some terpenoids have simple linear structures, for example geranylgeraniol is a linear diterpene (with 20 carbons) containing a terminal hydroxyl group. Geranylgeraniol can be made from GGPP using either:

i. an enzyme having pyrophosphatase activity such as a phosphatase; and/or

ii. two enzymes having phosphatase activity and successively cleaving the two phosphate groups of GGPP.

Alternatively, geranylgeraniol can be made from GGPP using an enzyme from the class I terpene cyclase family (as described below) able to cleave the pyrophosphate group of GGPP but lacking the ability to catalyze the successive cyclization.

These enzymatic pathways thus use enzymes having phosphatase activities and allow the cleavage of a diphosphate group of GGPP and release of geranylgeraniol.

Alternatively, geranylgeraniol can be made from another linear diterpene. For example, from geranyllinallol or from a corresponding polyene for example from beta-springene. Enzymes such as dehydratase-isomerases can be used for such reaction. Dehydratase-isomerases can catalyze the reversibly isomerization reactions between linear terpene compounds having a terminal alcohol group

or terminal double bound (Nestl, B. M., et al. (2017). Nature Chemical Biology 13(3): 275-281) (see figure 2).

Alternatively, geranylgeraniol can also be synthesized using chemical methods. For example, geranylgeraniol can be obtained from farnesene or farnesol by chain extension (Organic Syntheses, Vol. 84, p. 43-57 (2007).

The pathways leading to IPP and DMAPP and to geranyl-diphosphate (GPP), farnesyl-diphosphate (FPP) or geranylgeranyl-diphosphate (GGPP) is involved in the synthesis of terpenoids, a diverse class of molecules that play essential roles in primary metabolism and various cellular processes. Terpenoids are involved in numerous biological functions, including the synthesis of sterols, such as cholesterol in animals and phytosterols in plants, as well as the production of hormones, vitamins (such as vitamin E and K), and signaling molecules (such as ubiquinone and dolichol). Additionally, terpenoids are crucial for the formation of membrane lipids and post-translational modifications of proteins.

Therefore, the pathway leading to GPP, FPP and GGPP as described above is an important component of primary metabolism in all organisms as it provides the necessary precursors for the synthesis of essential terpenoid compounds involved in various physiological processes essential for the growth of the cells.

The majority of terpenoids compounds contain cyclic carbon scaffolds. The diversity of monocyclic and polycyclic carbon skeletons is due to the enzymatic conversion of the linear terpenoid precursors by Terpene Cyclases (TC), also referred to as terpene synthases. The cyclization reaction starts from a carbocation which reacts with electron rich double bonds leading to new carbon-bound formation. The outcome of the reaction is defined by the substrate folding and the nature of amino acid side chains in the enzyme active site.

Hence in a preferred embodiment of the invention, the process of the invention further comprises one or more steps prior to step (i), said step(s) comprising:
(a) preparing (producing) geranylgeranyl-diphosphate (GGPP) from IPP and DMAPP using one or more prenyltransferase enzyme(s); and/or,
(b) preparing (producing) a compound of formula (II) from GGPP using one or more enzymes having phosphatase enzyme activity.

Step (a) of this embodiment of the invention requires the use of one or more prenyltransferase enzyme(s).

The term ‘prenyltransferases’ represents a group of enzymes having the ability to condense successively five-carbon units such as isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) to form linear terpenyl-diphosphate compounds such as geranyl-diphosphate (GPP),

farnesyl-diphosphate (FPP) or geranylgeranyl-diphosphate (GGPP) containing 10, 15 and 15 carbons, respectively. Some prenyl transferases can add 5-carbon units to linear terpenyl-diphosphate compounds thereby extending the carbon chain length. An example of prenyl transferase are geranyl-diphosphate synthases (GGPP synthases) having the ability of producing GGPP from IPP and DMAPP or by adding 5 carbons to FPP. The term 'Prenyltransferases' also refers to a group of enzymes having the ability to transfer an isoprenoid subunit from a terpenyl-diphosphate compound, generally from a linear terpenyl-diphosphate compounds, to the non-terpenoid scaffold during the biosynthesis of meroterpenoids.

Step (a) of this embodiment of the invention may be performed by:

i. the successive action of 3 prenyltransferases: a GPP synthase, a FPP synthase catalyzing the addition of one IPP to GPP and a GGPP synthase catalyzing the addition of one IPP to FPP;

ii. the combination of 2 prenyltransferases: for example, a FPP synthase catalyzing the condensation of one DMAPP and two IPP and a GGPP synthase catalyzing the addition of one IPP to FPP; and/or

iii. the action of 1 prenyltransferase for example a GGPP synthase capable of catalyzing the the successive condensation of tree IPP and one DMAPP.

Examples of prenyltransferase enzymes that can be used in this step of the process of the invention are well known in the art. For example, a GGPP synthase from *Blakeslea trispora* can be used for the biosynthesis of GGPP (sun et al, Biotechnol. Lett. 34 (11), 2077-2082 (2012))

Preferably, the prenyltransferase is a GGPP synthase. Preferably, the GGPP synthase has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 1 or 2. Preferably, the GGPP synthase has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 2. Preferably, the GGPP synthase has the amino acid sequence of SEQ ID NO: 2.

Step (b) of this embodiment of the invention requires the use of one or more enzymes having phosphatase activity.

The term "phosphatase" represents a group of enzymes that are known to remove phosphate or diphosphate groups from a precursor containing a phosphate or diphosphate group. A particular subgroup of phosphatases has the ability of removing phosphate or diphosphate group from a terpenyl precursor releasing inorganic phosphate and the corresponding terpenyl alcohol. For example, some phosphatases are known to remove the diphosphate group of GGPP to form geranylgeraniol. Phosphatases acting on terpenyl diphosphate are found in several enzyme classes.

The term "protein tyrosine phosphatase" represents a group of enzymes that are generally known to remove phosphate groups from phosphorylated tyrosine residues on proteins. A particular subgroup of

said family as described in WO2020011883A1 are enzymes useful to remove diphosphate groups from phosphorylated terpene molecules (terpenyl diphosphate). In particular, phosphatases from the protein tyrosine phosphatase family having the Pfam ID number PF13350 can dephosphorylate GGPP to geranylgeraniol. Polypeptides can be scanned for matches against the Pfam protein family signature databases.

Phosphatases and in particular GGPP phosphatases can also be obtained from other protein families; for example, from the Phosphatidic Acid Phosphatases of type 2 (PAP2) protein family (IPR00326), Nudix-Hydrolase protein family (IPR015797) and Haloacid Dehalogenase-like (HAD-like) Hydrolases protein family (IPR041492). A method to screen for phosphatases and evaluate the conversion of geranylgeranyl diphosphate to geranylgeraniol is described in Example 5.

Examples of phosphatases enzymes that can be used in this step of the process of the invention are well known in the art.

Preferably, the phosphatase is a GGPP phosphatase. Preferably, the GGPP phosphatase has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any one of SEQ ID NOs: 3 to 10. Preferably, the GGPP phosphatase has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 3. Preferably, the GGPP phosphatase has the amino acid sequence of SEQ ID NO: 3.

As described above, this embodiment of the invention provides a compound of formula (II) by the sequential biocatalytic conversion of precursor compounds to the compound of formula (II).

In this embodiment of the invention, the precursor compounds to the compound of formula (II) are provided from IPP and DMAPP.

A further embodiment of the invention is wherein the process further comprises the preparation of IPP and DMAPP.

One means for the preparation of IPP and DMAPP is via the “mevalonate pathway”. The “mevalonate pathway” also known as the “isoprenoid pathway” or “HMG-CoA reductase pathway” is an essential metabolic pathway present in eukaryotes, archaea, and some bacteria. The mevalonate pathway begins with acetyl-CoA and produces two five-carbon building blocks called isopentenyl pyrophosphate (IPP) and dimethyl allyl pyrophosphate (DMAPP). Combining the mevalonate pathway with enzyme activity to generate the terpene precursors GPP, FPP or GGPP allows the recombinant cellular production of terpenes. The pathway is well known in the art. The list of enzymes required for the conversion of acetyl- CoA to IPP and DMAPP is provided below:

. Acetyl-CoA acetyltransferase (ACAT)

. 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA synthase)

. 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase)
. Mevalonate kinase
. Phosphomevalonate kinase
. Mevalonate diphosphate decarboxylase
. Isopentenyl diphosphate isomerase.

An alternative means for the preparation of IPP, and DMAPP is via the methylerythritol phosphate (MEP). The pathway is well known in the art. The list of enzymes required for the conversion of glyceraldehyde 3-phosphate (GAP) and pyruvate to IPP and DMAPP is provided below:
. 1-Deoxy-D-xylulose 5-phosphate synthase (DXS)
. 1-Deoxy-D-xylulose 5-phosphate reductoisomerase (DXR)
. 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT, IspD)
. 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK, IspE)
. 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS, IspF)
. 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS, IDS)
. 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (HDR)

Further alternative pathways to the preparation of IPP and DMAPP are known, see for example: Rinaldi, M. A., et al. (2022). Natural Product Reports 39(1): 90-118. https://doi.org/10.1039/D1NP00025J (see part 3 of this article).

The inventors have therefore provided a complete biocatalytic route for the preparation of a compound of formula (I) from acetyl-CoA or glyceraldehyde 3-phosphate (GAP) and pyruvate. This multistep biocatalytic process has for the first time been described herein and constitutes a significant advance in the preparation of such compounds.

Reaction conditions for the process of the invention

The process of the present invention may be an *in vivo* process or a bioconversion process.

The term *in vivo* process (or whole-cell production, or *in-vivo* production, or *in-vivo* biosynthesis) refers to a process of using a metabolically active cell where the primary metabolism is active to produce the precursors for the processes of the invention (preferably a microbial cell) to convert a carbon source to a new compound, such as the conversion of a carbon source to a terpene or terpene-derived compound.

Preferred sources of carbon are sugars, such as mono-, di- or polysaccharides. Very good sources of carbon are for example glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose. Sugars can also be added to the media via complex compounds, such as molasses, or other by-products from sugar refining. It may also be advantageous

to add mixtures of various sources of carbon. Other possible sources of carbon are oils and fats such as soybean oil, sunflower oil, peanut oil and coconut oil, fatty acids such as palmitic acid, stearic acid or linoleic acid, alcohols such as glycerol, methanol or ethanol and organic acids such as acetic acid or lactic acid.

The cells thus contain all enzymes of one or more biosynthetic pathways. At least some of the enzymes involved in the process are part of the cell`s primary metabolism. For example, the cells may contain the enzymes of a pathway to convert a carbon source (e.g., glucose, glycerol, isoprenol, prenol, CO_2) to terpenoid precursors (e.g., IPP, DMAPP, FPP, GGPP) and a pathway converting the terpene precursor to a terpene or terpene derived molecule such as a compound of formula (I). The enzymes may be present naturally in the cell or the cells can be transformed to produce the enzymes.

It is important to point out that, until the present invention, it had not been possible to prepare a compound of formula (I) using *in vivo* processes. The state of the art when devising the present invention may be found in, for example WO2016170099 and WO2010139719. Here, it can be seen that the existing processes for preparing the compound of formula (I) were not *in vivo* and hence, the present invention provides an important advance of the processes disclosed therein.

Alternatively, the processes of the present invention may be performed under bioconversion, also known as biotransformation conditions. Bioconversion processes refer to processes of conversion of compounds to different products using a biological process or agent such enzymes or whole cells (preferably a microbial cell). Bioconversion does not include the use of a cell`s primary metabolism (as defined above) to produce the precursors for the processes of the invention. A bioconversion process can comprise multistep reactions each performed by a different enzyme. The compounds used in bioconversion process can be extracted from a natural source or produced using a separate chemical or a biochemical process.

The at least one polypeptide/enzyme which is present during the bioconversion method of the invention or an individual step of the multistep method as defined herein above, can be present in living cells naturally or recombinantly producing the enzyme or enzymes, in harvested cells, dead cells, in permeabilized cells, in crude cell extracts, in purified extracts, or in essentially pure or completely pure form, i.e. under bioconversion conditions. Such extracts may comprise membrane fraction or a liquid fraction prepared from the recombinant host cell that expresses at least one polypeptide/enzyme. The cells may be immobilized on a suitable substrate as is known in the art. At least one polypeptide/enzyme may be present in solution or as an enzyme immobilized on a carrier. One or several enzymes may simultaneously be present in soluble and/or immobilized forms.

It can be understood by the skilled person that there may be advantages for the use of an *in vivo* process.

In particular, a bioconversion process involves multiple steps, typically:

- Preparation or isolation of the starting compound to be transformed. The compound can be prepared using a chemical or biochemical process or by extraction from a natural source.
- Production of the enzymes or (living) cells used for the bioconversion.
- Biotransformation reaction by contacting the compound with the enzymes or (living) cells.
- Product Recovery and Refinement.

In comparison, an *in vivo* process requires a limited number of steps, generally limited to:

- The cultivation of the microorganism under conditions suitable for the production of the desired compound.
- Harvesting the cells or growing medium and purification of the desired compound.

In a bioconversion process such as the bioconversion of a compound of formula (VI), the addition of a detergent is often required to facilitate the solubilization of the compound or to maximize the contact with the biocatalyst. In an *in vivo* process, the reactants and enzymes are produced in the cells and the addition of a detergent is not needed.

Therefore, the *in vivo* process is usually more efficient and cost-effective than a bioconversion process.

Laboratory methods that can be used in *in vivo* and bioconversion processes of the invention are well known in the art. There follows a discussion on some of the methods that can be used.

The bioconversion processes according to the invention can be performed in common reactors, which are known to those skilled in the art, and in different ranges of scale, e.g. from a laboratory scale (few milliliters to dozens of liters of reaction volume) to an industrial scale (several liters to thousands of cubic meters of reaction volume). If the polypeptide is used in a form encapsulated by non-living, optionally permeabilized cells, in the form of a more or less purified cell extract or in purified form, a chemical reactor can be used. The chemical reactor usually allows controlling the amount of at least one enzyme, the amount of at least one substrate, the pH, the temperature and the circulation of the reaction medium.

Where the process of the invention is *in vivo* then it is preferred that the reaction is performed in a fermenter, where parameters necessary for suitable living conditions for the living cells (e.g. culture medium with nutrients, temperature, aeration, presence or absence of oxygen or other gases, antibiotics, and the like) can be controlled.

The term "fermentative production" or "fermentation" refers to the ability of a microorganism (assisted by enzyme activity contained in or generated by said microorganism) to produce a chemical compound in cell culture utilizing at least one carbon source added to the incubation.

The term "fermentation broth" or "fermentation medium" is understood to mean a liquid, particularly aqueous or aqueous /organic solution which is based on a fermentative process and has not been worked up or has been worked up, for example, as described herein.

Those skilled in the art are familiar with chemical reactors or bioreactors, e.g. with procedures for up-scaling chemical or biotechnological methods from laboratory scale to industrial scale, or for optimizing process parameters, which are also extensively described in the literature (for biotechnological methods see e.g. Crueger und Crueger, Biotechnologie – Lehrbuch der angewandten Mikrobiologie, 2. Ed., R. Oldenbourg Verlag, München, Wien, 1984).

The culture medium that is to be used must satisfy the requirements of the particular strains in an appropriate manner. Descriptions of culture media for various microorganisms are given in the handbook "Manual of Methods for General Bacteriology" of the American Society for Bacteriology (Washington D. C., USA, 1981).

These media that can be used according to the invention may comprise one or more sources of carbon, sources of nitrogen, inorganic salts, vitamins and/or trace elements.

Sources of nitrogen are usually organic or inorganic nitrogen compounds or materials containing these compounds. Examples of sources of nitrogen include ammonia gas or ammonium salts, such as ammonium sulfate, ammonium chloride, ammonium phosphate, ammonium carbonate or ammonium nitrate, nitrates, urea, amino acids or complex sources of nitrogen, such as corn-steep liquor, soybean flour, soy-bean protein, yeast extract, meat extract and others. The sources of nitrogen can be used separately or as a mixture.

Inorganic salt compounds that may be present in the media comprise the chloride, phosphate or sulfate salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron.

Inorganic sulfur-containing compounds, for example sulfates, sulfites, di-thionites, tetrathionates, thiosulfates, sulfides, but also organic sulfur compounds, such as mercaptans and thiols, can be used as sources of sulfur.

Phosphoric acid, potassium dihydrogenphosphate or dipotassium hydrogenphosphate or the corresponding sodium-containing salts can be used as sources of phosphorus.

Chelating agents can be added to the medium, in order to keep the metal ions in solution. Especially suitable chelating agents comprise dihydroxyphenols, such as catechol or protocatechuate, or organic acids, such as citric acid.

The fermentation media used according to the invention may also contain other growth factors, such as vitamins or growth promoters, which include for example biotin, riboflavin, thiamine, folic acid, nicotinic acid, pantothenate and pyridoxine. Growth factors and salts often come from complex components of the media, such as yeast extract, molasses, corn-steep liquor and the like. In addition, suitable precursors can be added to the culture medium. The precise composition of the compounds in the medium is strongly dependent on the particular experiment and must be decided individually for each specific case. Information on media optimization can be found in the textbook "Applied Microbiol. Physiology, A Practical Approach" (1997) Growing media can also be obtained from commercial suppliers, such as Standard 1 (Merck) or BHI (Brain heart infusion, DIFCO) etc.

All components of the medium are sterilized, either by heating (20 min at 1.5 bar and 121 °C) or by sterile filtration. The components can be sterilized either together or if necessary, separately. All the components of the medium can be present at the start of growing, or optionally can be added continuously or by batch feed.

The temperature of the culture is normally between 15 °C and 45 °C, preferably 25 °C to 40 °C and can be kept constant or can be varied during the experiment. The pH value of the medium should be in the range from 5 to 8.5, preferably around 7.0. The pH value for growing can be controlled during growing by adding basic compounds such as sodium hydroxide, potassium hydroxide, ammonia or ammonia water or acid compounds such as phosphoric acid or sulfuric acid. Antifoaming agents, e.g. fatty acid polyglycol esters, can be used for controlling foaming. To maintain the stability of plasmids, suitable substances with selective action, e.g. antibiotics, can be added to the medium. Oxygen or oxygen-containing gas mixtures, e.g. the ambient air, are fed into the culture in order to maintain aerobic conditions. The temperature of the culture is normally from 20 °C to 45 °C. Culture is continued until a maximum of the desired product has formed. This is normally achieved within 1 hour to 160 hours.

Where the process of the invention is a bioconversion, cells containing the at least one enzyme can be permeabilized by physical or mechanical means, such as ultrasound or radiofrequency pulses, French presses, or chemical means, such as hypotonic media, lytic enzymes and detergents present in the medium, or combination of such methods. Examples for detergents are SDS, digitonin, n-dodecylmaltoside, octylglycoside, Triton® X-100, Tween ® 20, deoxycholate, CHAPS (3-[(3-Cholamidopropyl)dimethylammonio]-1-propansulfonate), Nonidet ® P40 (Ethylphenolpoly(ethyleneglycolether), and the like. As stated above, where the process of the invention is an *in vivo* process, then a detergent is not required for the reasons stated herein.

The conversion reaction can be carried out batch wise, semi-batch wise or continuously. Reactants (and optionally nutrients) can be supplied at the start of reaction or can be supplied subsequently, either semi-continuously or continuously.

The bioconversion reaction of the invention, depending on the particular reaction type, may be performed in an aqueous, aqueous-organic or non-aqueous reaction medium.

An aqueous or aqueous-organic medium may contain a suitable buffer in order to adjust the pH to a value in the range of 5 to 11, like 6 to 10.

In an aqueous-organic medium an organic solvent miscible, partly miscible or immiscible with water may be applied. Non-limiting examples of suitable organic solvents are listed below. Further examples are mono- or polyhydric, aromatic or aliphatic alcohols, in particular polyhydric aliphatic alcohols like glycerol.

The non-aqueous medium may contain is substantially free of water, i.e. will contain less than about 1 wt.-% or 0.5 wt.-% of water.

Bioconversion methods may also be performed in an organic non-aqueous medium. As suitable organic solvents there may be mentioned aliphatic hydrocarbons having for example 5 to 8 carbon atoms, like pentane, cyclopentane, hexane, cyclohexane, heptane, octane or cyclooctane; aromatic carbohydrates, like benzene, toluene, xylenes, chlorobenzene or dichlorobenzene, aliphatic acyclic and ethers, like diethylether, methyl-tert.-butylether, ethyl-tert.-butylether, dipropylether, diisopropylether, dibutylether; or mixtures thereof.

The concentration of the reactants/substrates may be adapted to the optimum bioconversion reaction conditions, which may depend on the specific enzyme applied. For example, the initial substrate concentration may be in the 0,1 to 0,5 M, as for example 10 to 100 mM.

The bioconversion reaction temperature may be adapted to the optimum reaction conditions, which may depend on the specific enzyme applied. For example, the reaction may be performed at a temperature in a range of from 0 to 70 °C, as for example 20 to 50 or 25 to 40 °C. Examples for reaction temperatures are about 30 °C, about 35 °C, about 37 °C, about 40 °C, about 45 °C, about 50 °C, about 55 °C and about 60 °C.

The bioconversion may proceed until equilibrium between the substrate and then product(s) is achieved, but may be stopped earlier. Usual process times are in the range from 1 minute to 25 hours, in particular 10 min to 6 hours, as for example in the range from 1 hour to 4 hours, in particular 1.5 hours to 3.5 hours. These parameters are non-limiting examples of suitable process conditions.

Advantageously, microorganisms such as bacteria, fungi or yeasts are used as host organisms. Advantageously, gram-positive or gram-negative bacteria are used, preferably bacteria of the families *Enterobacteriaceae*, *Pseudomonadaceae*, *Rhizobiaceae*, *Streptomycetaceae*, *Streptococcaceae* or *Nocardiaceae*, especially preferably bacteria of the genera *Escherichia*, *Pseudomonas*, *Streptomyces*,

Lactococcus, *Nocardia*, *Burkholderia*, *Salmonella*, *Agrobacterium*, *Clostridium* or *Rhodococcus*. The genus and species *Escherichia coli* is quite especially preferred. Furthermore, other advantageous bacteria are to be found in the group of alpha-Proteobacteria, beta-Proteobacteria or gamma-Proteobacteria. Advantageously also yeasts of families like *Saccharomyces* or *Pichia* are suitable hosts.

Preferably, the cell is a bacterium or a fungal cell, in particular a yeast cell. Preferably, the cell is a unicellular organism, a cultured cell derived from a multi-cellular organism, a cell present in a cultured tissue derived from a multicellular organism, or a cell present in a living multicellular organism. Preferably, the cell is a bacterial cell of the genus *Escherichia*, preferably *E. coli*, or a yeast cell of the genus *Saccharomyces*, preferably *S. cerevisiae*, of the genus *Yarrowia*, preferably *Y. lipolytica*, or of the genus *Pichia*, preferably *P. pastoris*.

Alternatively, entire plants or plant cells may serve as natural or recombinant host. As non-limiting examples, the following plants or cells derived therefrom may be mentioned: the genera *Nicotiana*, in particular *Nicotiana benthamiana* and *Nicotiana tabacum* (tobacco); as well as *Arabidopsis*, in particular *Arabidopsis thaliana*.

Product isolation

The methodology of the present invention can further include a step of recovering an end product or an intermediate product, optionally in stereoisomerically or enantiomerically substantially pure form. The term “recovering” includes extracting, harvesting, isolating or purifying the compound from culture or reaction media. Recovering the compound can be performed according to any conventional isolation or purification methodology known in the art including, but not limited to, treatment with a conventional resin (e.g., anion or cation exchange resin, non-ionic adsorption resin, etc.), treatment with a conventional adsorbent (e.g., activated charcoal, silicic acid, silica gel, cellulose, alumina, etc.), alteration of pH, solvent extraction (e.g., with a conventional solvent such as an alcohol, ethyl acetate, hexane and the like), distillation, dialysis, filtration, concentration, crystallization, recrystallization, pH adjustment, lyophilization and the like.

Identity and purity of the isolated product may be determined by known techniques, like High Performance Liquid Chromatography (HPLC), gas chromatography (GC), Spektroskopy (like IR, UV, NMR), Colouring methods, TLC, NIRS, enzymatic or microbial assays.(see for example: Patek et al. (1994) Appl. Environ. Microbiol. 60:133-140; Malakhova et al. (1996) Biotekhnologiya 11 27-32; und Schmidt et al. (1998) Bioprocess Engineer. 19:67-70. Ullmann's Encyclopedia of Industrial Chemistry (1996) Bd. A27, VCH: Weinheim, S. 89-90, S. 521-540, S. 540-547, S. 559-566, 575-581 und S. 581-587; Michal, G (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, Bd. 17.).

The compounds produced in any of the processes described herein can be converted to derivatives such as, but not limited to hydrocarbons, esters, amides, glycosides, ethers, epoxides, aldehydes, ketones, alcohols, diols, acetals or ketals. The terpene compound derivatives can be obtained by a chemical method such as, but not limited to oxidation, reduction, alkylation, acylation and/or rearrangement. Alternatively, the terpene compound derivatives can be obtained using a biochemical method by contacting the terpene compound with an enzyme such as, but not limited to an oxidoreductase, a monooxygenase, a dioxygenase, a transferase or a terpene cyclase. The biochemical conversion can be performed *in vitro* using isolated enzymes, enzymes from lysed cells or bioconversion using whole cells.

Recombinant cells of the invention

As discussed herein, the present inventors were able for the first time to innovate an *in vivo* process for the preparation of a compound of formula (I). To achieve this, the present inventors created a biosynthetic pathway to the compound of formula (I) in recombinant cells. Until the present invention, it had not previously been known to prepare this compound using an *in vivo* method involving the use of a recombinant cell. Hence, until the present invention, a recombinant cell producing a compound of formula (I) was not known in the art. The compound may be present within the recombinant cell or exported into the reaction medium.

Furthermore, as stated above in the present invention, high selectivity of production of a compound of formula (VI) in the form of formula (VIa) has been achieved, as demonstrated in the accompanying Examples.

This subsequently leads to a process in which olfactory preferred forms of a compound of formula (I) are prepared. Hence, the process of the invention to prepare a compound of formula (I) lacking substantial amount of undesirable side products is of significant commercial importance. As can be shown in the accompanying examples, more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

Accordingly therefore, an embodiment of the present invention is a recombinant cell wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

Accordingly therefore, a further aspect of the invention provides a recombinant cell comprising, capable of producing or producing a compound of formula (I) and one or more compound(s) of formula (II), formula (III), formula (IV), formula (V) and/or formula (VI).

Methods for preparing a recombinant cell comprising, capable of producing or producing a compound of formula (I) and one or more compound (s) of formula (II), formula (III), formula (IV), formula (V) and/or formula (VI) are provided herein. Preferably, said cell comprises (i) a polypeptide having ADH enzyme

activity, (ii) a polypeptide having enal-cleaving enzyme activity, (iii) a polypeptide having BVMO enzyme activity, (iv) a polypeptide having esterase enzyme activity, and (v) a polypeptide having terpene cyclase enzyme activity.

A preferred embodiment of the invention is wherein:

(i) the polypeptide having ADH enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 11 to 21;

(ii) the polypeptide having enal-cleaving enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 22;

(iii) the polypeptide having BVMO enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 23 to 26 and 216 to 227; preferably to SEQ ID NOs: 23 to 26;

(iv) the polypeptide having esterase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 27 and 28; and/or,

(v) the polypeptide having terpene cyclase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 29 to 75, 79 to 89 and 265 to 289; preferably, to SEQ ID NOs: 29 to 75 and 265 to 289; more preferably, to any of SEQ ID NOs: 29 to 75, 265 to 274 and 276 to 289.

A further preferred embodiment of the invention is wherein:

(i) the polypeptide having ADH enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 11 or 21; preferably, the ADH enzyme has the sequence of SEQ ID NO: 11 or 21;

(ii) the polypeptide having enal-cleaving enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 22; preferably, the enal-cleaving enzyme has the sequence of SEQ ID NO: 22;

(iii) the polypeptide having BVMO enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 25 or 26; preferably, the BVMO enzyme has the sequence of SEQ ID NO: 25 or 26;

(iv) the polypeptide having esterase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO 28; preferably, the esterase enzyme has the sequence of SEQ ID NO: 28; and/or,

(v) the polypeptide having terpene cyclase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 48, 57, 71, 74, 265, 266, 267, 268, 274, 276, 279, 280, 281, 282, 283, 286, 287 or 288; preferably, the terpene cyclase enzyme has the sequence of SEQ ID NO: 48, 57, 71, 74, 265, 266, 267, 268, 274, 276, 279, 280, 281, 282, 283, 286, 287 or 288.

Alternatively, in this embodiment of the invention:

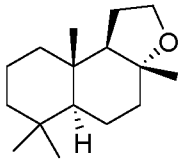
(v) the polypeptide having terpene cyclase enzyme activity is a SHC enzyme and has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 48, 265, 266, 267, 268, 274, 276 or 279; preferably, the SHC enzyme has the sequence of SEQ ID NO: 48, 265, 266, 267, 268, 274, 276 or 279.

Alternatively, in this embodiment of the invention:

(v) the polypeptide having terpene cyclase enzyme activity is a meroterpenoid cyclase enzyme and has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 57, 71, 74, 280, 281, 282, 283, 286, 287 or 288; preferably, the meroterpenoid cyclase enzyme has the sequence of SEQ ID NO: 57, 71, 74, 280, 281, 282, 283, 286, 287 or 288.

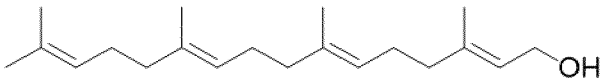
A further embodiment of the invention is wherein:

(i) the compound of formula (I) is in the form of formula (Ia):



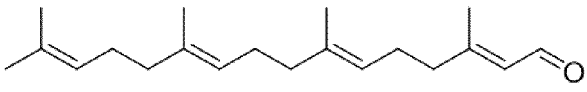
(formula Ia)

(ii) the compound of formula (II) is in the form of formula (IIa):



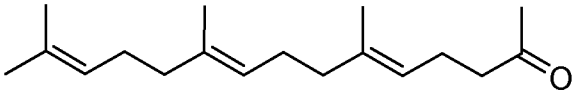
(formula IIa);

(iii) the compound of formula (III) is in the form of formula (IIIa):



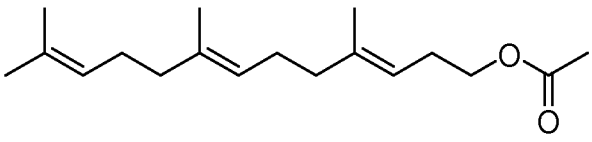
(formula IIIa);

(iv) the compound of formula (IV) is in the form of formula (IVa):



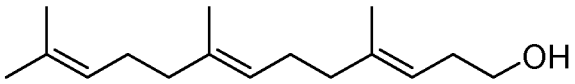
(formula IVa);

(v) the compound of formula (V) is in the form of formula (Va):



(formula Va);

(vi) the compound of formula (VI) is in the form of formula (VIa):



(formula VIa).

The recombinant cell may be any such cell suitable for the production of a compound of formula (I).

A list of suitable cells for the production of a compound of formula (I) is provided above in relation to the process of the invention and are also cells for this aspect of the invention.

Preferably, the cell is a bacterium or a fungal cell, in particular a yeast. Preferably, the cell is a unicellular organism, a cultured cell derived from a multi-cellular organism, a cell present in a cultured tissue derived from a multicellular organism, or a cell present in a living multicellular organism. Preferably, the cell is a bacterial cell of the genus *Escherichia*, preferably *E. coli*, or a yeast cell of the genus *Saccharomyces*, preferably *S. cerevisiae*, of the genus *Yarrowia*, preferably *Y. lipolytica*, or of the genus *Pichia*, preferably *P. pastoris*.

Methods of introducing recombinant nucleic acid sequences into such host cells are well known in the art and constitute routine laboratory methodologies which do not need to be further described herein.

An embodiment of this aspect of the invention is wherein the cell further comprises one or more prenyltransferase enzyme(s). Preferably, the cell further comprises one or more enzymes having phosphatase activity.

As mentioned above in relation to the process of the invention, the cells of the invention may also comprise further enzymes to provide a compound of formula (II). Such a process requires the presence of one or more prenyltransferase enzyme(s) and one or more enzymes having phosphatase activity.

Preferably, the prenyltransferase is a GGPP synthase. Preferably, the GGPP synthase has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 1 or 2.

Preferably, the phosphatase is a GGPP phosphatase. Preferably, the GGPP phosphatase has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 3 to 10.

A further embodiment of the invention is wherein the cell of the invention comprises enzymes for the IPP and DMAPP.

As mentioned above in relation to the process of the invention, the cells of the invention may also comprise still further enzymes to provide a compound of formula (II) via the "mevalonate pathway", methylerythritol phosphate (MEP) pathway or alternative pathways to the preparation of IPP and DMAPP.

In one embodiment of the invention, the cell comprises enzymes of the mevalonate pathway:

. Acetyl-CoA acetyltransferase (ACAT)

. 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA synthase)

. 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase)

. Mevalonate kinase

. Phosphomevalonate kinase

. Mevalonate diphosphate decarboxylase

. Isopentenyl diphosphate isomerase

. Dimethylallyl diphosphate synthase

In a further embodiment of the invention, the cell comprises enzymes of the MEP pathway:

. 1-Deoxy-D-xylulose 5-phosphate synthase (DXS)

. 1-Deoxy-D-xylulose 5-phosphate reductoisomerase (DXR)

. 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT, IspD)

. 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK, IspE)

. 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS, IspF)

. 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS, IDS)

. 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (HDR)

As stated above, meroterpenoid cyclases when used in the process of the invention, produce a compound of formula (I) having an isomeric bias towards isomers of compound of formula (I) having preferred olfactory profiles, i.e. formula (Ia) and/or (Ib) rather than (Ic) and/or (Id). Hence, the use of meroterpenoid cyclases is associated with a surprising technical advantage over the use of SHC enzymes. In particular, they demonstrated that less than 1% of compound of formula (I) produced by the process of the invention were in the isomeric form of formula I(c) and/or I(d). Hence, the use of meroterpenoid cyclases is associated with a surprising technical advantage over the use of SHC enzymes.

Accordingly therefore, a preferred embodiment of the invention is wherein the recombinant cell comprises a meroterpenoid cyclases as a terpene cyclase enzyme.

Preferably, the meroterpenoid cyclase is a membrane-integrated meroterpenoid cyclase. It can be seen from the accompanying Examples that this class of meroterpenoid cyclases preferably produces compounds of formula (Ia) rather than (Ic) and/or (Id). Examples of membrane-integrated meroterpenoid cyclase include those provided in any of SEQ ID NOs: 50 to 73 and 280 to 289.

Preferably, the meroterpenoid cyclase is a soluble meroterpenoid cyclase. It can be seen from the accompanying Examples that this class of meroterpenoid cyclases preferably produces compounds of formula (Ib) rather than (Ic) and/or (Id). Examples of soluble meroterpenoid cyclase include those provided in SEQ ID NO: 74 or 75.

Preferably, the recombinant cell of the invention comprises greater than 97% of the compound of formula (I) in the form of compound (Ia) and/or (b) rather than I(c) and/or I(d).

Cell culture fermentation media of the invention

As discussed herein, the present inventors were able for the first time to innovate an *in vivo* process for the preparation of a compound of formula (I). To achieve this, the present inventors created a biosynthetic pathway to the compound of formula (I) in recombinant cells which are subsequently cultured in appropriate cell culture fermentation media. Until the present invention, it had not previously been able to prepare this compound using an *in vivo* method involving the use of a recombinant cell in a cell culture fermentation medium. Hence, until the present invention, a cell culture fermentation medium comprising the recombinant cell of the invention and/or a compound of formula (I) was not known in the art.

Furthermore as stated above, in the present invention, high selectivity of production of a compound of formula (VI) in the form of formula (VIa) has been achieved, as demonstrated in the accompanying Examples.

This subsequently leads to a process in which olfactory preferred forms of compound of formula (I) are prepared. Hence, the process of the invention to prepare compound of formula (I) lacking substantial amount of undesirable side products is of significant commercial importance. As can be shown in the accompanying Examples, more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

Accordingly therefore, an embodiment of the present invention is a cell culture fermentation medium comprising compound of formula (I), wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

Accordingly, a further aspect of the invention provides a cell culture fermentation medium comprising the recombinant cell as described herein above. The cell culture fermentation medium may further comprise a compound of formula (I) and/or one or more compound(s) of formula (II), formula (III), formula (IV), formula (V) and/or formula (VI).

As discussed herein, the present inventors were able for the first time to innovate a biosynthetic pathway to the compound of formula (I) in recombinant cells. Such cells are then grown under conditions suitable for the production of said compound using cell culture fermentation media as appropriate for specific cell types.

The cell culture fermentation media can be a nutrient rich broth for the growth and maintenance of the cells during the production phase. Yeast culture conditions for maintaining and propagating various strains can require specific formulations of complex media for use in cloning and protein expression, and can be appreciated by those of skill in the art. Commercially available culture media can be used from ThermoFisher for example. The media can be YPD broth or can have a yeast nitrogen base. Yeast can be grown in YPD or synthetic media at 30 ºC.

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

Reaction mixture comprising compounds of the invention

The process of the invention may be a bioconversion process.

As discussed herein, the present inventors were able for the first time to innovate a bionconversion process for the preparation of a compound of formula (I). To achieve this, the present inventors created a biosynthetic pathway to the compound of formula (I). Until the present invention, it had not previously been able to prepare this compound using a bioconversion process.

Furthermore and as stated above, in the present invention, high selectivity of production of a compound of formula (VI) in the form of formula (VIa) has been achieved, as demonstrated in the accompanying Examples.

This subsequently leads to a bioconversion in which olfactory preferred forms of compound of formula (I) are prepared. Hence, the bioconversion of the invention to prepare a compound of formula (I) lacking substantial amount of undesirable side products is of significant commercial importance. As can be shown in the accompanying Examples, more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

Accordingly therefore, an embodiment of the present invention is a reaction mixture comprising compound of formula (I), wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib). The reaction mixture may further comprise one or more compounds of formula (II), formula (III), formula (IV), formula (V), and/or formula (VI).

Additional components of the reaction mixture may include detergents, co-factors, cells, cell-debris, cell culture media, and other such components well known to the person skilled in the art.

Preparation of a compound of formula (VI)

As set out above, to prepare an improved process for the preparation of a compound of formula (I), the inventors developed a deep understanding of the biochemical route to the production of this compound by a multi-enzymatic reaction from precursor compounds. This multi-enzymatic reaction is the first time the preparation of this compound has been performed by such a step-wise reaction and constitutes a significant scientific and commercial advance in the preparation of sesquiterpene compound of formula (I).

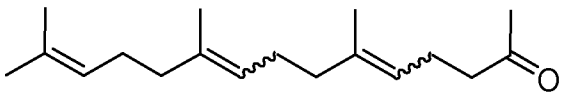
In addition to the preparation of a compound of formula (I), the inventors also devised a process for the preparation of a compound of formula (VI). This compound is a commercially important precursor for the preparation of a compound of formula (I), either by subsequent bioconversion or chemical methods.

Accordingly therefore, a further aspect of the invention provides a process for the preparation of a compound of formula (VI)

(VI)

in the form of any one of its stereoisomers or a mixture thereof, comprising:

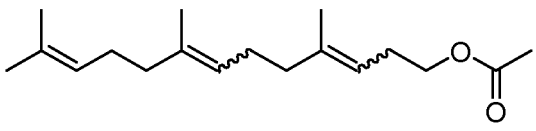
(i) contacting the compound of formula (IV)



(IV)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having BVMO enzyme activity to produce a compound of formula (V); and,

(ii) contacting the compound of formula (V)

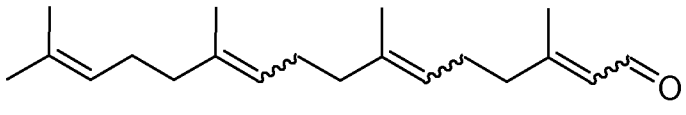


(V)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having esterase enzyme activity to produce a compound of formula (VI).

An embodiment of the invention is wherein the process comprises a prior step of:

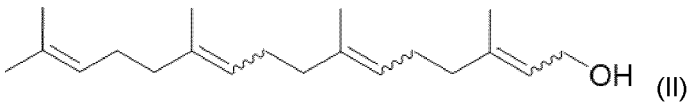
(a) contacting a compound of formula (III)


(III)

In the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce the compound of formula (IV).

A further embodiment of the invention is wherein the process comprises a prior step of:

(a) contacting a compound of formula (II)


(II)

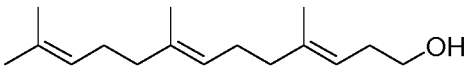
in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having ADH enzyme activity to produce a compound of formula (III).

As discussed herein, the present inventors were able for the first time to innovate the biosynthetic pathway to the compound of formula (I) in recombinant cells. In preparing such a pathway, the present inventors also prepared cells which could produce compound of formula (VI). Hence, this aspect of the invention is not disclosed in the state of the art.

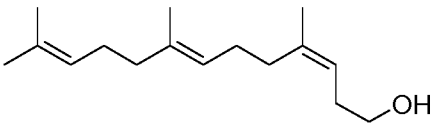
As can be appreciated, the process of preparing a compound of formula (VI) comprises steps (i) to (iv) of the process for preparing a compound of formula (I). Accordingly therefore, all the embodiments set out herein in relation to the process of preparing a compound of formula (I) can be used in this aspect of the invention, with the exception of the polypeptides relating to step (v) of the process of preparing a compound of formula (I).

For the avoidance of doubt, the compound of formula (VI) is also known as homofarnesol, 4 ,8,12-trimethyltrideca-3,7,11-trien-1-ol; CAS No 35826-67-6.

The compound of formula (VI) may be present in any one of its stereoisomers or a mixture thereof. Specifically, the compound may have the following structures and isoforms:

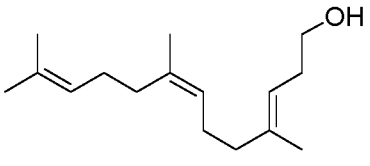

(formula VIa)

(3E,7E)-homofarnesol; (3E,7E)-4,8,12-trimethyltrideca-3,7,11-trien-1-ol; CAS No 459-89-2.



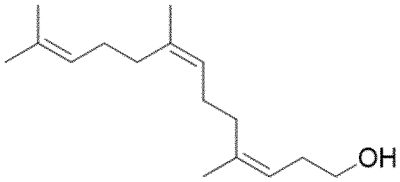
(formula VIb)

(3Z,7E)-homofarnesol; (3Z,7E)-4,8,12-trimethyltrideca-3,7,11-trien-1-ol; CAS No 138152-06-4.



(formula VIc)

(3E,7Z)-homofarnesol; (3E,7Z)-4,8,12-trimethyltrideca-3,7,11-trien-1-ol; CAS No 2032064-12-1.



(formula VId)

(3Z,7Z)-homofarnesol; (3Z,7Z)-4,8,12-trimethyltrideca-3,7,11-trien-1-ol; CAS No 138152-08-6.

Furthermore, and as stated above, in the present invention, high selectivity of production of a compound of formula (VI) in the form of formula (VIa) has been achieved, as demonstrated in the accompanying Examples. Hence, a preferred embodiment of the invention is wherein the process prepares a compound of formula (VI), wherein more than 99% of the compound of formula (VI) is in the form of formula (VIa).

A further preferred embodiment of the invention is wherein:

(i) the polypeptide having BVMO enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 25 or 26; preferably, the BVMO enzyme has the sequence of SEQ ID NO: 25 or 26; and/or

(ii) the polypeptide having esterase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO 28; preferably, the esterase enzyme has the sequence of SEQ ID NO: 28.

A further preferred embodiment of the invention is wherein:

(i) the polypeptide having ADH enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 11 or 21; preferably, the ADH enzyme has the sequence of SEQ ID NO: 11 or 21; and/or

(ii) the polypeptide having enal-cleaving enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 22; preferably, the enal-cleaving enzyme has the sequence of SEQ ID NO: 22.

Preferably, the process is an *in vivo* or bioconversion process.

Further included in this aspect of the invention is a recombinant cell comprising, capable of producing or producing a compound of formula (VI) and one or more compound(s) of formula (II), formula (III), formula (IV) and/or formula (V). The recombinant cell comprises (i) a polypeptide having ADH enzyme activity, (ii) a polypeptide having enal-cleaving enzyme activity, (iii) a polypeptide having BVMO enzyme activity, and (iv) a polypeptide having esterase enzyme activity.

Further included in this aspect of the invention is a process for making a compound of formula (VI) comprising growing the recombinant cell of this aspect of the invention under growth conditions suitable for the production of the compound of formula (VI) and as described herein above.

Further included in this aspect of the invention is a cell culture fermentation medium comprising the recombinant cell of this aspect of the invention. The cell culture fermentation medium may further comprise a compound of formula (VI) and/or one or more compound(s) of formula (II), formula (III), formula (IV) and/or formula (V).

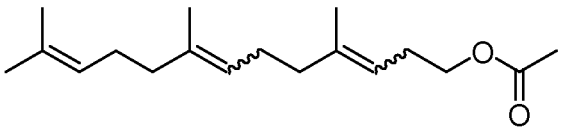
Further included in this aspect of the invention is a reaction mixture comprising compound of formula (VI); preferably, in the form of formula (VIa). More preferably, more than 99% of the compound of formula (VI) is in the form of formula (VIa). The reaction mixture may further comprise one or more compounds of formula (II), formula (III), formula (IV) and/or formula (V).

Further included in this aspect of the invention is a compound of formula (VI) obtained or obtainable by the process of this aspect of the invention or from the recombinant cell, the cell culture fermentation or the reaction mixture of this aspect of the invention.

Preparation of a compound of formula (V)

In addition to the preparation of a compound of formula (I), the inventors also devised a process for the preparation of a compound of formula (V). This compound is a commercially important precursor for the preparation of a compound of formula (I), either by subsequent bioconversion or chemical methods.

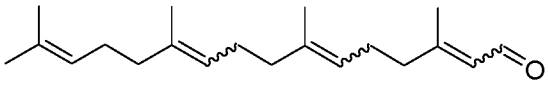
Accordingly therefore, a further aspect of the invention provides a process for the preparation of a compound of formula (V)



(V)

in the form of any one of its stereoisomers or a mixture thereof, comprising:

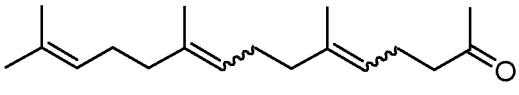
(i) contacting the compound of formula (III)



(III)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV); and,

(ii) contacting the compound of formula (IV)

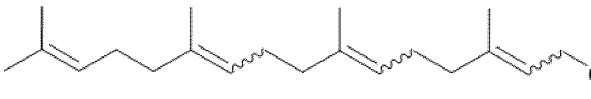


(IV)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having BVMO enzyme activity to produce a compound of formula (V).

An embodiment of the invention is wherein the process comprises a prior step of:

(a) contacting a compound of formula (II)



(II)

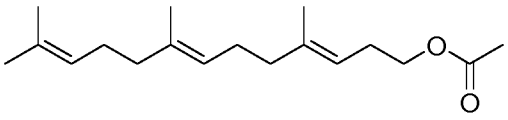
in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having ADH enzyme activity to produce a compound of formula (III).

As discussed herein, the present inventors were able for the first time to innovate the biosynthetic pathway to the compound of formula (I) in recombinant cells. In preparing such a pathway, the present inventors also prepared cells which could produce compound of formula (V). Hence, this aspect of the invention is not disclosed in the state of the art.

As can be appreciated, the process of preparing a compound of formula (V) comprises steps (i) to (iii) of the process for preparing a compound of formula (I). Accordingly therefore, all the embodiments set out herein in relation to the process of preparing a compound of formula (I) can be used in this aspect of the invention, with the exception of the polypeptides relating to step (iv) and (v) of the process of preparing a compound of formula (I).

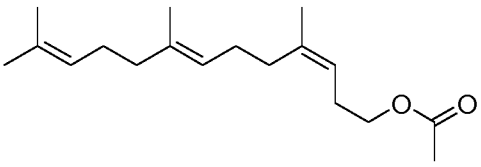
For the avoidance of doubt, the compound of formula (V) is also known as homofarnesyl acetate, i.e. 4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate; CAS No 109813-25-4.

The compound of formula (V) may be present in any one of its stereoisomers or a mixture thereof. Specifically, the compound may have the following structures and isoforms:



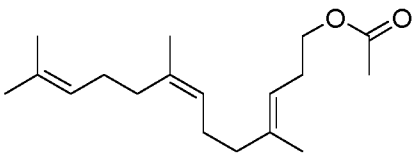
(formula Va)

(3E,7E)- homofarnesyl acetate; (3E,7E)-4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate; CAS No 944346-19-4.



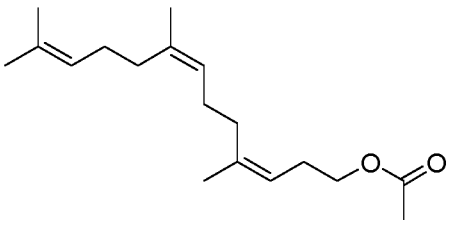
(formula Vb)

(3Z,7E)- homofarnesyl acetate; (3Z,7E)-4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate; CAS No 1467099-77-9.



(formula Vc)

(3E,7Z)- homofarnesyl acetate; (3E,7Z)-4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate.



(formula Vd)

(3Z,7Z)- homofarnesyl acetate; (3Z,7Z)-4,8,12-trimethyltrideca-3,7,11-trien-1-yl acetate.

Preferably the process is an *in vivo* or bioconversion process.

Further included in this aspect of the invention is a recombinant cell comprising, capable of producing or producing a compound of formula (V) and one or more compound(s) of formula (II), formula (III) and/or formula (IV). The recombinant cell comprises (i) a polypeptide having ADH enzyme activity, (ii) a polypeptide having enal-cleaving enzyme activity, and (iii) a polypeptide having BVMO enzyme activity.

Further included in this aspect of the invention is a process for making a compound of formula (V) comprising growing the recombinant cell of this aspect of the invention under growth conditions suitable for the production of the compound of formula (V) and as described herein above.

Further included in this aspect of the invention is a cell culture fermentation medium comprising the recombinant cell of this aspect of the invention. The cell culture fermentation medium may further comprise a compound of formula (V) and/or one or more compound(s) of formula (II), formula (III) and/or formula (IV).

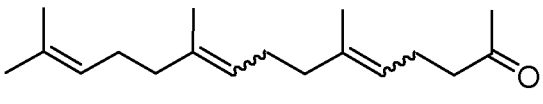
Further included in this aspect of the invention is a reaction mixture comprising compound of formula (V). The reaction mixture may further comprise one or more compounds of formula (II), formula (III) and/or formula (IV).

Further included in this aspect of the invention is a compound of formula (V) obtained or obtainable by the process of this aspect of the invention or from the recombinant cell, the cell culture fermentation or the reaction mixture of this aspect of the invention.

Preparation of a compound of formula (IV)

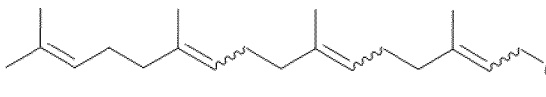
In addition to the preparation of a compound of formula (I), the inventors also devised a process for the preparation of a compound of formula (IV). This compound is a commercially important precursor for the preparation of a compound of formula (I), either by subsequent bioconversion or chemical methods.

Accordingly therefore, a further aspect of the invention provides a process for the preparation of a compound of formula (IV)


(IV)

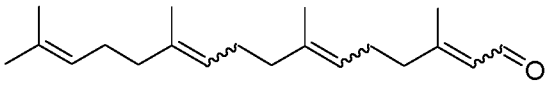
in the form of any one of its stereoisomers or a mixture thereof, comprising:

(i) contacting the compound of formula (II)


(II)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having ADH enzyme activity to produce a compound of formula (III); and,

(ii) contacting the compound of formula (III)


(III)

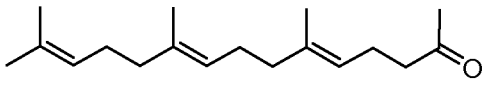
in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV).

As discussed herein, the present inventors were able for the first time to innovate the biosynthetic pathway to the compound of formula (I) in recombinant cells. In preparing such a pathway, the present inventors also prepared cells which could produce compound of formula (IV). Hence, this aspect of the invention is not disclosed in the state of the art.

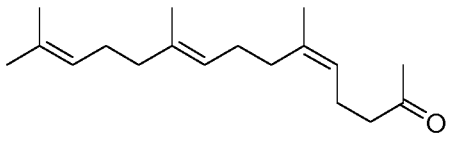
As can be appreciated, the process of preparing a compound of formula (IV) comprises steps (i) to (ii) of the process for preparing a compound of formula (I). Accordingly therefore, all the embodiments set out herein in relation to the process of preparing a compound of formula (I) can be used in this aspect of the invention, with the exception of the polypeptides relating to step (iii), (iv) and (v) of the process of preparing a compound of formula (I).

For the avoidance of doubt, the compound of formula (IV) is also known as farnesylacetone, i.e. 6,10,14-trimethylpentadeca-5,9,13-trien-2-one; CAS No 762-29-8.

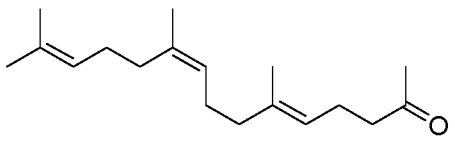
The compound of formula (IV) may be present in any one of its stereoisomers or a mixture thereof. Specifically, the compound may have the following structures and isoforms:


(formula IVa)

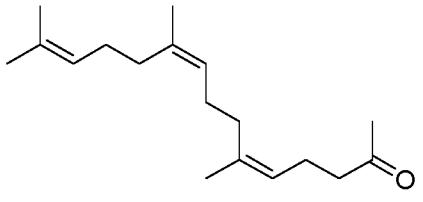
(5E,9E)-farnesylacetone; (5E,9E)-6,10,14-trimethylpentadeca-5,9,13-trien-2-one; CAS No 1117-52-8.


(formula IVb)

(5Z,9E)-farnesylacetone; (5Z,9E)-6,10,14-trimethylpentadeca-5,9,13-trien-2-one; CAS No 1117-51-7.


(formula IVc)

(5E,9Z)-farnesylacetone; (5E,9Z)-6,10,14-trimethylpentadeca-5,9,13-trien-2-one; CAS No 3053-35-3.


(formula IVd)

(5Z,9Z)-farnesylacetone; (5Z,9Z)-6,10,14-trimethylpentadeca-5,9,13-trien-2-one; Cas No 3796-69-8.

Preferably, the process is an *in vivo* or bioconversion process.

Further included in this aspect of the invention is a recombinant cell comprising, capable of producing or producing a compound of formula (IV) and one or more compound(s) of formula (II) and/or formula (III). The recombinant cell comprises (i) a polypeptide having ADH enzyme activity and (ii) a polypeptide having enal-cleaving enzyme activity.

Further included in this aspect of the invention is a process for making a compound of formula (IV) comprising growing the recombinant cell of this aspect of the invention under growth conditions suitable for the production of the compound of formula (IV) and as described herein above.

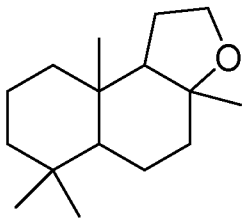
Further included in this aspect of the invention is a cell culture fermentation medium comprising the recombinant cell of this aspect of the invention. The cell culture fermentation medium may further comprise a compound of formula (IV) and/or one or more compound(s) of formula (II) and/or formula (III).

Further included in this aspect of the invention is a reaction mixture comprising compound of formula (IV). The reaction mixture may further comprise one or more compounds of formula (II) and/or formula (III).

Further included in this aspect of the invention is a compound of formula (IV) obtained or obtainable by the process of this aspect of the invention or from the recombinant cell, the cell culture fermentation or the reaction mixture of this aspect of the invention.

<u>Further aspect for the preparation of a compound of formula (I)</u>

A further aspect of the invention provides a process fo the preparation of a compound of formula (I)



(I)

in the form of any one of its stereoisomers or a mixture thereof,

comprising:

(i) contacting a compound of formula (VI)

(VI)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having terpene cyclase enzyme activity to produce a compound of formula (I).

An embodiment of this aspect of the invention is wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or formula (Ib).

Another embodiment of this aspect of the invention is wherein the compound of formula (VI) is in the form of formula (VIa).

A further aspect of the invention is wherein the polypeptide having terpene cyclase enzyme activity is a polypeptide that is not a squalene cyclase (SHC) enzyme and/or is a polypeptide that is a squalene cyclase enzyme. In the context of the invention, the polypeptide that is not a SHC enzyme is a meroterpenoid cyclase enzyme.

Hence, a further aspect of the invention is wherein the polypeptide having terpene cyclase enzyme activity is a meroterpenoid cyclase enzyme and/or a squalene cyclase enzyme.

As can be appreciated in this aspect of the invention, the process of preparing a compound of formula (I) from compound of formula (VI) comprises step (v) of the process for preparing a compound of formula (I) in the first aspect of the invention above. Accordingly therefore, all the embodiments set out herein in relation to step (v) of the process of preparing a compound of formula (I) in the first aspect of the invention can be used in this aspect of the invention.

In one embodiment of this aspect of the invention, the squalene cyclase enzyme has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 29 to 49 and 265 to 279; preferably, to any of SEQ ID NOs: 29 to 49, 265 to 274 and 276 to 279. More preferably, the squalene cyclase enzyme comprises the sequence provided in any of SEQ ID NOs: 29 to 49, 265 to 274 and 276 to 279.

In a further embodiment, the squalene cyclase enzyme has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 29, 31, 33, 34, 36 to 38, 40, 41, 43 to 46, 48, 49, 265 to 274 and 276 to 279. Preferably, the squalene cyclase enzyme comprises the sequence provided in any of SEQ ID NOs: 29, 31, 33, 34, 36 to 38, 40, 41, 43 to 46, 48, 49, 265 to 274 and 276 to 279.

When preparing the process of the present invention, the inventors sought to compare the isomeric profile of the compound of formula (I) synthesized via SHC and meroterpenoid cyclases.

To their surprise, they found that the meroterpenoid cyclases when used in the process of the invention, produced compound of formula (I) having an isomeric bias towards isomers of compound of formula (I) having preferred olfactory profiles more than that produced by SHC enzymes.

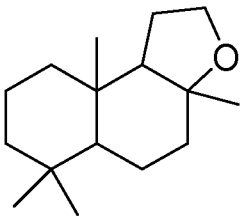
In particular, they demonstrated that less than 1% of compound of formula (I) produced by the process of the invention were in the isomeric form of formula I(c) and/or I(d). Hence, the use of meroterpenoid cyclases is associated with a surprising technical advantage over the use of SHC enzymes.

Preferably, the meroterpenoid cyclase is a membrane-integrated meroterpenoid cyclase. It can be seen from the accompanying Examples that this class of meroterpenoid cyclases preferably produce compounds of formula (Ia) rather than (Ic) and/or (Id). Examples of membrane integrated meroterpenoid cyclase include those provided in any of SEQ ID NOs: 50 to 73 and 280 to 289.

Preferably, the meroterpenoid cyclase is a soluble meroterpenoid cyclase. It can be seen from the accompanying Examples that this class of meroterpenoid cyclases preferably produce compounds of formula (Ib) rather than (Ic) and/or (Id). Examples of soluble meroterpenoid cyclase include those provided in SEQ ID NO: 74 or 75.

This is the first time meroterpenoid cyclases have been used to prepare compound of formula (I), and the bias towards such isomeric forms is surprising and technically important commercially.

Accordingly, a further aspect of this invention provides a process for the preparation of a compound of formula (I):

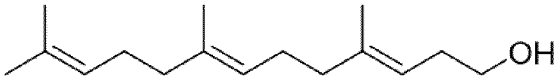


(I)

in the form of any one of its stereoisomers or a mixture thereof,

comprising:

(i) contacting the compound of formula (VI)



(VI)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having terpene cyclase enzyme activity to produce a compound of formula (I), wherein the polypeptide having terpene cyclase enzyme activity is not a SHC enzyme.

A preferred embodiment of this aspect of the invention is wherein the polypeptide that is not a SHC enzyme is a meroterpenoid cyclase enzyme.

As can be appreciated, meroterpenoid cyclase enzymes (or polypeptides having meroterpenoid cyclase enzyme activity) are described above in details in relation to the first aspect of the invention and this is incorporated herein to this aspect of the invention.

Preferably, the meroterpenoid cyclase enzyme has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 50 to 75 and 280 to 289. Preferably, the meroterpenoid cyclase enzyme comprises the sequence provided in any of SEQ ID NOs: 50 to 75 and 280 to 289.

Preferably, the meroterpenoid cyclase enzyme has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 57, 71, 74, 280, 281, 282, 283, 286, 287 or 288. Preferably, the meroterpenoid cyclase enzyme comprises the sequence provided in any of SEQ ID NOs: 57, 71, 74, 280, 281, 282, 283, 286, 287 or 288.

Preferably, the meroterpenoid cyclase enzyme is a membrane-integrated meroterpenoid cyclase. It can be seen from the accompanying Examples that this class of meroterpenoid cyclases preferably produce compounds of formula (Ia) rather than (Ic) and/or (Id). Hence, in a preferred embodiment, the meroterpenoid cyclase enzyme is a membrane-integrated meroterpenoid cyclase enzyme having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 50 to 73 and 280 to 289.

Preferably, the meroterpenoid cyclase is a soluble meroterpenoid cyclase. It can be seen from the accompanying Examples that this class of meroterpenoid cyclases preferably produce compounds of formula (Ib) rather than (Ic) and/or (Id). Hence, in another preferred embodiment, the meroterpenoid cyclase enzyme is a soluble meroterpenoid cyclase enzyme having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 74 and 75.

As can be appreciated in this aspect of the invention for the preparation of a compound of formula (I), the process may further comprise one or more steps prior to step (i), said one or more steps being described above in details in relation to the first aspect of the invention and this is incorporated herein to this aspect of the invention.

Preferably, the process is an *in vivo* or a bioconversion process.

Further included in this aspect of the invention is a recombinant cell comprising or capable of producing or producing a compound of formula (I); optionally, wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib). In one embodiment, the recombinant cell comprises or is capable of functionally expressing or is functionally expressing a polypeptide having SHC enzyme activity and/or a polypeptide having meroterpenoid cyclase enzyme activity as described herein above.

Further included in this aspect of the invention is a process for making a compound of formula (I) comprising growing the recombinant cell of this aspect of the invention under growth conditions suitable for the production of the compound of formula (I).

Further included in this aspect of the invention is a cell culture fermentation medium comprising the recombinant cell of this aspect of the invention. The cell culture fermentation medium may further comprise a compound of formula (I); optionally, wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib). The cell culture fermentation medium may further a compound of formula (VI).

Further included in this aspect of the invention is a reaction mixture comprising a compound of formula (I); optionally, wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib). The reaction mixture may further a compound of formula (VI).

Further included in this aspect of the invention is a compound of formula (I) obtained or obtainable by the process of this aspect of the invention or from the recombinant cell or from the cell culture fermentation medium or from the reaction mixture of this aspect of the invention.

Further included in this aspect of the invention is a compound (I), wherein more than 97% of said compound is in the form of formula (Ia) and/or (Ib).

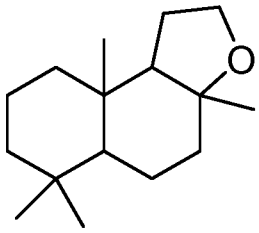
A further aspect of the invention is the use of a meroterpenoid cyclase enzyme for the production of a compound of formula (I) and/or a derivative thereof.

<u>Further aspect for the preparation of a compound of formula (I)</u>

As set out above, to prepare an improved process for the preparation of a compound of formula (I), the inventors developed a deep understanding of the biochemical route to the production of these compounds by a multi-enzymatic reaction from precursor compounds. This multi-enzymatic reaction is the first time the preparation of these compounds has been performed by such a step-wise reaction and constitutes a significant scientific and commercial advance in the preparation of sesquiterpene compound of formula (I).

In addition to the preparation of a compound of formula (I) from a compound of formula (II), the inventors also devised a process for the preparation of a compound of formula (I) from a compound of formula (V). This is also a commercially important process to the preparation of a compound of formula (I).

Accordingly therefore, a further aspect of the invention provides a process for the preparation of a compound of formula (I):

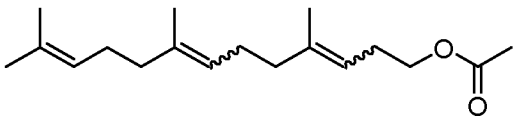


(I)

in the form of any one of its stereoisomers or a mixture thereof,

comprising:

(i) contacting a compound of formula (V)



(V)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having esterase enzyme activity to produce a compound of formula (VI); and

(ii) contacting the compound of formula (VI)

(VI)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having terpene cyclase enzyme activity to produce a compound of formula (I).

Esterase enzymes (or polypeptides having esterase enzyme activity) and terpene cyclase enzymes (or polypeptides having terpene cyclase enzyme activity) are described above in details in relation to the first aspect of the invention and this is incorporated herein to this aspect of the invention.

In this aspect of the invention, the process for preparing a compound of formula (I) uses a compound of formula (V) as a starting material. The process of this aspect of the invention may be an *in vivo* process or a bioconversion process. Preferably, the process is a bioconversion process.

A preferred embodiment of this aspect of the invention is wherein:

(i) the polypeptide having esterase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 27 or 28; preferably, to SEQ ID NO: 28; and/or

(ii) the polypeptide having terpene cyclase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 29 to 75, 79 to 89 and 265 to 289; preferably, to any of SEQ ID NOs: 29 to 75 and 265 to 289; more preferably, to any of SEQ ID NOs: 29 to 75, 265 to 274 and 276 to 289; even more preferably, to any of SEQ ID NO: 48, 57, 71, 74, 265, 266, 267, 268, 274, 276, 279, 280, 281, 282, 283, 286, 287 or 288.

Further included in this aspect of the invention is a recombinant cell comprising one or more compound(s) of formula (I), formula (V) and/or formula (VI). The recombinant cell may further comprise (i) a polypeptide having esterase enzyme activity and (ii) a polypeptide having terpene cyclase enzyme activity.

Further included in this aspect of the invention is a process for making a compound of formula (I) comprising growing a recombinant cell of this aspect of the invention under growth conditions suitable for the production of the compound of formula (I).

Further included in this aspect of the invention is a cell culture fermentation medium comprising the recombinant cell of this aspect of the invention. The cell culture fermentation medium may further comprise one or more compounds of formula (I), formula (V) and/or formula (VI).

Further included in this aspect of the invention is a reaction mixture comprising a compound of formula (I). The reaction mixture may further comprise one or more compounds of formula (V) and/or formula (VI)

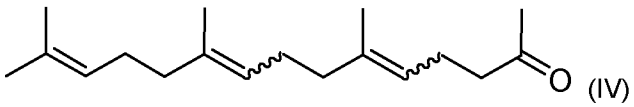
Further included in this aspect of the invention is a compound of formula (I) obtained or obtainable by the process of this aspect of the invention or from the recombinant cell, the cell culture fermentation medium or the reaction mixture of this aspect of the invention.

Furthermore as stated above, in the present invention high selectivity of production of a compound of formula (VI) in the form of compound (VIa) has been achieved, as demonstrated in the accompanying Examples. This subsequently leads to a bioconversion in which olfactory preferred forms of compounds of formula (I) are prepared. Hence the bioconversion of the invention to prepare compounds of formula (I) lacking substantial amount of undesirable side products is of significant commercial importance.

Accordingly therefore, and as can be shown in the accompanying Examples, an embodiment of this aspect of the invention is wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

A further embodiment of this aspect of the invention is wherein the process comprises a prior step:

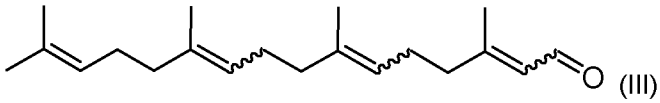
(a) contacting the compound of formula (IV)


(IV)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having BVMO enzyme activity to produce a compound of formula (V).

A further embodiment of this aspect of the invention is wherein the process comprises an additional prior step:

(a) contacting a compound of formula (III)


(III)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV).

Examples of BVMO and enal-cleaving enzymes are described above in the first aspect of the invention and may be used in this aspect of the invention.

Phosphatase enzymes for use in the process of the invention

As described herein above, one embodiment of the process of the first aspect of the invention further comprises one or more biocatalytic steps prior to step (i) to prepare a compound of formula (II), said biocatalytic step(s) comprising:

(a) preparing geranylgeranyl-diphosphate (GGPP) from IPP and DMAPP using one or more prenyltransferase enzyme(s); and,

(b) preparing a compound of formula (II) from GGPP using one or more enzymes having phosphatase activity.

When preparing the process of the invention, the present inventors identified phosphatase enzymes which can be used in step (b) of said process. This is the first time these enzymes were shown to be capable of catalysing the step of preparing a compound of formula (II) from GGPP.

Hence, included in this aspect of the invention is the use of a phosphatase for the preparation of a compound of formula (II).

A further aspect of the invention provides the use of a phosphatase enzyme comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in any of SEQ ID NOs: 3 to 10 for the preparation of a compound of formula (II) from GGPP.

Preferably, the phosphatase enzyme comprises 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in any of SEQ ID NOs: 3, 4, 5, 7 and 8. More preferably, the phosphatase enzyme comprises the sequences provided in any of SEQ ID NOs: 3, 4, 5, 7 and 8.

ADH enzymes for use in the process of the invention

As described herein, the process of the first aspect of the invention comprises as step (i) the preparation of a compound of formula (III) by contacting a compound of formula (II) with an ADH enzyme.

When preparing the process of the invention, the present inventors identified ADH enzymes which can be used in step (i) of said process. This is the first time these enzymes were shown to be capable of catalysing the step of preparing a compound of formula (III) from a compound of formula (II).

Hence, a further aspect of the invention provides the use of an ADH enzyme comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in any of SEQ ID NOs: 11 to 21 for the preparation of compound of formula (III) from a compound of formula (II). Preferably, the ADH enzyme comprises the sequence provided in any of SEQ ID NOs: 11 to 21.

Yet, another further aspect of the invention provides the use of an ADH enzyme comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 11 or 21 for the preparation of compound of formula (III) from a compound of formula (II). Preferably, the ADH enzyme comprises the sequence provided in SEQ ID NOs: 11 or 21.

Enal-cleaving enzymes for use in the process of the invention

As described herein, the process of the first aspect of the invention comprises as step (ii) the preparation of a compound of formula (IV) by contacting a compound of formula (III) with an enal-cleaving enzyme.

When preparing the process of the invention, the present inventors identified an enal-cleaving enzyme which can be used in step (ii) of said process. This is the first time this enzyme was shown to be capable of catalysing the step of preparing a compound of formula (IV) from a compound of formula (III).

Hence, a further aspect of the invention provides the use of a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV). A further aspect of the invention provides the use of a polypeptide having enal-cleaving enzyme activity for the preparation of a compound of formula (IV) from a compound of formula (III). A further aspect of the invention is the use of a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV), (V), (VI), (I) and/or a derivative thereof.

Hence, a further aspect of the invention provides the use of a polypeptide having enal-cleaving enzyme activity comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 22 to produce a compound of formula (IV). A further aspect of the invention provides the use of a polypeptide having enal-cleaving enzyme activity comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 22 for the preparation of a compound of formula (IV) from a compound of formula (III). A further aspect of the invention provides the use of a polypeptide having enal-cleaving enzyme activity comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 22 to produce a compound of formula (IV), (V), (VI), (I) and/or a derivative thereof.

Preferably, the polypeptide having enal-cleaving enzyme activity comprises the sequence provided in SEQ ID NO: 22.

<u>BVMO enzymes for use in the process of the invention</u>

As described herein, the process of the first aspect of the invention comprises as step (iii) the preparation of a compound of formula (V) by contacting a compound of formula (IV) with a BVMO enzyme.

When preparing the process of the invention, the present inventors identified BMVO enzymes which can be used in step (iii) of said process. This is the first time these enzymes were shown to be capable of catalysing the step of preparing a compound of formula (V) from a compound of formula (IV).

Hence, a further aspect of the invention provides the use of a BVMO enzyme comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 23 to 26 and 216 to 227 for the preparation of compound of formula (V) from a compound of formula (IV).

Preferably, the BVMO enzyme comprises 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 23 to 26. More preferably, the BVMO enzyme comprises the sequence provided in any of SEQ ID NOs: 23 to 26.

Preferably, the BVMO enzyme comprises 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 25 or 26. More preferably, the BVMO enzyme comprises the sequence provided in SEQ ID NO: 25 or 26.

Esterase enzymes for use in the process of the invention

As described herein, the process of the first aspect of the invention comprises as step (iv) the preparation of a compound of formula (VI) by contacting a compound of formula (V) with an esterase enzyme.

When preparing the process of the invention, the present inventors identified esterase enzymes which can be used in step (iv) of said process. This is the first time these enzymes were shown to be capable of catalysing the step of preparing a compound of formula (VI) from a compound of formula (V).

Hence a further aspect of the invention provides the use of an esterase enzyme comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 27 or 28 for the preparation of compound of formula (VI) from a compound of formula (V).

Preferably, the esterase enzyme comprises the sequence provided in SEQ ID NO: 27 or 28.

SHC enzymes according to the invention

When preparing the process of the invention, the present inventors identified new polypeptide sequences encoding SHC enzymes which can be used in said process. These polypeptides are therefore also part of the present invention.

When investigating the SHC enzymes, the present inventors identified enzymes having amino acid alanine at position 437 and amino acid methionine at position 600 relative to sequence provided in SEQ ID NO: 82 as being of particular utility.

Accordingly therefore, a further aspect of the present invention is a mutant SHC enzyme having an amino acid alanine at position 437 and amino acid methionine at position 600 relative to sequence provided in SEQ ID NO: 82. This is the first time this combination of mutations has been shown to be able to function in the reaction described in step (v) of the process of the first aspect of the invention.

SHC enzymes have been described above; for example, in relation to step (v) of the process of the first aspect of the invention. Using the information provided in that section of the present specification, the skilled person can identify any SHC enzyme which can be modified using standard laboratory techniques to arrive at the mutant SHC enzyme of this aspect of the invention.

A preferred embodiment of this aspect of the invention is wherein the mutant SHC enzyme is a polypeptide having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 29, 31, 33, 34, 36 to 38, 40, 41, 43 to 46, 48, 49, 265 to 274 and 276 to 279, wherein said polypeptide has amino acid alanine at position 437 and amino acid methionine at position 600 relative to the sequence provided in SEQ ID NO: 82. A preferred embodiment is wherein the mutant SHC enzyme is a polypeptide having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 29, 31, 33, 34, 36 to 38, 40, 41, 43 to 46, 48, 49, 265, 266, 267, 268, 274, 276 and 279, wherein said polypeptide has amino acid alanine at position 437 and amino acid methionine at position 600 relative to the sequence provided in SEQ ID NO: 82. Preferably, the mutant SHC enzyme is a polypeptide having the amino acid sequence provided in any of SEQ ID NOs: 29, 31, 33, 34, 36 to 38, 40, 41, 43 to 46, 48, 49, 265 to 274 and 276 to 279. Preferably, the mutant SHC enzyme is a polypeptide having the amino acid sequence provided in any of SEQ ID NOs: 29, 31, 33, 34, 36 to 38, 40, 41, 43 to 46, 48, 49, 265, 266, 267, 268, 274, 276 and 279. Also included in this aspect of the invention is polypeptide fragments, variants and functional equivalents thereof.

The polypeptides of this aspect of the invention are SHC enzymes which may be used in step (v) of the process of the first aspect of the invention or any futher aspects of the invention comprising prepraring a compound of formula (I) from a compound of formula (VI).

A further aspect of the invention provides a nucleic acid sequence encoding a mutant SHC enzyme having an amino acid alanine at position 437 and amino acid methionine at position 600 relative to sequence provided in SEQ ID NO: 82. Methods of preparing such nucleic acid sequence are known in the art.

A preferred embodiment of this aspect of the invention is wherein the nucleic acid sequence encoding the mutant SHC enzyme is a nucleic acid sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 123, 125, 126, 128 to 131, 134 to 137, 139 to 142, 144 to 148, 150 to 153, 290 to 299 and 301 to 304, wherein said nucleic acid sequence encodes a polypeptide having amino acid alanine at position 437 and amino acid methionine at position 600 relative to sequence provided in SEQ ID NO: 82. Preferably, the nucleic acid sequence is a nucleic sequence provided in any of SEQ ID NOs: 123, 125, 126, 128 to 131, 134 to 137, 139 to 142, 144 to 148, 150 to 153, 290 to 299 and 301

to 304. Also included in this aspect of the invention is expression vectors, cassettes and other such related technology comprising nucleic acid sequences of the invention.

Furthermore, as described herein, in the process of the invention comprising the preparation of a compound of formula (I) by contacting a compound of formula (VI) with a terpene cyclase enzyme, the terpene cyclase enzyme may be an SHC enzyme. When preparing the process of the invention, the present inventors identified SHC enzymes which can be used in said process. This is the first time these enzymes were shown to be capable of catalysing the step of preparing a compound of formula (I) from a compound of formula (VI).

Hence, a further aspect of the invention provides the use of a SHC enzyme comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 29 to 49, 79 to 89 and 265 to 279 for the preparation of compound of formula (I) from a compound of formula (VI).

Preferably, the SHC enzyme comprises 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 29 to 49 and 265 to 279. More preferably, the SHC enzyme comprises the sequence provided in any of SEQ ID NOs: 29 to 49 and 265 to 279.

Preferably, the SHC enzyme comprises 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 29 to 49, 265 to 274 and 276 to 279. More preferably, the SHC enzyme comprises the sequence provided in any of SEQ ID NOs: 29 to 49, 265 to 274 and 276 to 279.

Meroterpenoid cyclase enzymes according to the invention

When preparing the process of the invention, the present inventors identified new polypeptide sequences encoding meroterpenoid cyclase enzymes which can be used in said process. These polypeptides are therefore also part of the present invention.

One embodiment of this aspect of the invention is a mutant meroterpenoid cyclase enzyme having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in any of SEQ ID NOs: 56 to 70. Preferably, the mutant meroterpenoid cyclase enzyme has the amino acid sequence provided in any of SEQ ID NOs: 56 to 70. Also included in this aspect of the invention is polypeptide fragments, variants and functional equivalents thereof.

Accordingly, this aspect of the invention provides a nucleic acid sequence encoding the mutant meroterpenoid cyclase enzyme having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%,

95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 165 to 185. Preferably, the nucleic acid sequence is a nucleic acid sequence provided in any of SEQ ID NOs: 165 to 185. Also included in this aspect of the invention is expression vectors, cassettes and other such related technology comprising nucleic acid sequences of the invention.

When investigating the meroterpenoid cyclase enzymes, the present inventors identified enzymes having an amino acid substitution at amino acid position 9 relative to sequence provided in SEQ ID NO: 51.

Accordingly therefore, a further aspect of the invention is a mutant meroterpenoid cyclase enzyme having an amino acid substitution at amino acid position 9 relative to sequence provided in SEQ ID NO: 51. This is the first time a meroterpenoid cyclase enzyme having this mutation has been shown to be able to function in the step of preparing a compound of formula (I) from a compound of formula (VI).

Meroterpenoid cyclase enzymes have been described above in relation to step (v) of the process of the first aspect of the invention. Using the information provided in that section of the present specification, the skilled person can identify any meroterpenoid cyclase enzyme which can be modified using standard laboratory techniques to arrive at the mutant meroterpenoid cyclase enzyme of this aspect of the invention.

Preferably, the mutant meroterpenoid cyclase enzyme has a substitution at amino acid position 9 relative to the sequence provided in SEQ ID NO: 51 introducing a cysteine, methionine or threonine in this position.

Accordingly, a preferred embodiment of this aspect of the invention is wherein the mutant meroterpenoid cyclase enzyme is a polypeptide having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in any of SEQ ID NOs: 56 to 61, 69 and 70, wherein the mutant meroterpenoid cyclase enzyme has an amino acid substitution at amino acid position 9 relative to the sequence provided in SEQ ID NO: 51. Preferably, the mutant meroterpenoid cyclase enzyme has the amino acid sequence provided in any of SEQ ID NOs: 56 to 61, 69 and 70. Also included in this aspect of the invention is polypeptide fragments, variants and functional equivalents thereof.

The polypeptides of this aspect of the invention are meroterpenoid cyclase enzymes which may be used in step (v) of the process of the first aspect of the invention or any futher aspects of the invention comprising prepraring a compound of formula (I) from a compound of formula (VI).

A further aspect of the invention provides a nucleic acid sequence encoding a mutant meroterpenoid cyclase enzyme having an amino acid substitution at amino acid position 9 relative to sequence provided in SEQ ID NO: 51. Methods of preparing such nucleic acid sequence are known in the art.

A preferred embodiment of this aspect of the invention is wherein a nucleic acid sequence encoding the mutant meroterpenoid cyclase enzyme is a nucleic acid sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of the sequences provided in SEQ ID NOs: 165 to 176, 184 and 185, wherein said nucleic acid sequence encodes a polypeptide having an amino acid substitution at amino acid position 9 relative to sequence provided in SEQ ID NO: 51. Preferably, the nucleic acid sequence is a nucleic acid sequence provided in any of SEQ ID NOs: 165 to 176, 184 and 185. Also included in this aspect of the invention is expression vectors, cassettes and other such related technology comprising nucleic acid sequences of the invention.

Furthermore, as described herein, in the process of the invention comprising the preparation of a compound of formula (I) by contacting a compound of formula (VI) with a terpene cyclase enzyme, the terpene cyclase enzyme may be a meroterpenoid cyclase enzyme. When preparing the process of the invention, the present inventors identified meroterpenoid cyclase enzymes which can be used in said process. This is the first time these meroterpenoid cyclases were shown to be capable of catalysing the step of preparing a compound of formula (I) from a compound of formula (VI).

Hence, a further aspect of the invention provides the use of a meroterpenoid cyclase enzyme comprising 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ NOs: 50 to 75 and 280 to 289 for the preparation of a compound of formula (I) from a compound of formula (VI).

Preferably, the meroterpenoid cyclase enzyme comprises the sequence provided in any of SEQ NOs: 50 to 75 and 280 to 289.

Polypeptides and nucleic acids of the invention or used in the process of the invention

The generic terms “polypeptide” or “peptide”, which may be used interchangeably, refer to a natural or synthetic linear chain or sequence of consecutive, peptidically linked amino acid residues, comprising about 10 to up to more than 1.000 residues. Short chain polypeptides with up to 30 residues are also designated as “oligopeptides”.

The term “protein” refers to a macromolecular structure consisting of one or more polypeptides. The amino acid sequence of its polypeptide(s) represents the “primary structure” of the protein. The amino acid sequence also predetermines the “secondary structure” of the protein by the formation of special structural elements, such as alpha-helical and beta-sheet structures formed within a polypeptide chain. The arrangement of a plurality of such secondary structural elements defines the “tertiary structure” or spatial arrangement of the protein. If a protein comprises more than one polypeptide chains said chains are spatially arranged forming the “quaternary structure” of the protein. A correct spatial arrangement

or “folding” of the protein is prerequisite of protein function. Denaturation or unfolding destroys protein function. If such destruction is reversible, protein function may be restored by refolding.

A typical protein function referred to herein is an “enzyme function”, i.e. the protein acts as biocatalyst on a substrate, for example a chemical compound, and catalyzes the conversion of said substrate to a product. An enzyme may show a high or low degree of substrate and/or product specificity.

A “polypeptide” referred to herein as having a particular “activity” thus implicitly refers to a correctly folded protein showing the indicated activity, as for example a specific enzyme activity.

Thus, unless otherwise indicated the term “polypeptide” also encompasses the terms “protein” and “enzyme”.

Similarly, the term “polypeptide fragment” encompasses the terms “protein fragment“ and “enzyme fragment”.

The term “isolated polypeptide” refers to an amino acid sequence that is removed from its natural environment by any method or combination of methods known in the art and includes recombinant, biochemical and synthetic methods.

“Target peptide” refers to an amino acid sequence which targets a protein, or polypeptide to intracellular organelles, i.e., mitochondria, or plastids, or to the extracellular space (secretion signal peptide). A nucleic acid sequence encoding a target peptide may be fused to the nucleic acid sequence encoding the amino terminal end, e.g., N-terminal end, of the protein or polypeptide, or may be used to replace a native targeting polypeptide.

The present invention also relates to "functional equivalents" (also designated as “analogs” or “functional mutations”) of the polypeptides specifically described herein.

For example, "functional equivalents" refer to polypeptides which, in a test used for determining enzymatic activity display at least a 1 to 10 %, or at least 20 %, or at least 50 %, or at least 75 %, or at least 90 % higher or lower activity, as that of the polypeptides specifically described herein.

"Functional equivalents”, according to the invention, also cover particular mutants, which, in at least one sequence position of an amino acid sequences stated herein, have an amino acid that is different from that concretely stated one, but nevertheless possess one of the aforementioned biological activities, as for example enzyme activity. "Functional equivalents" thus comprise mutants obtainable by one or more, like 1 to 20, in particular 1 to 15 or 5 to 10 amino acid additions, substitutions, in particular conservative substitutions, deletions and/or inversions, where the stated changes can occur in any sequence position, provided they lead to a mutant with the profile of properties according to the

invention. Functional equivalence is in particular also provided if the activity patterns coincide qualitatively between the mutant and the unchanged polypeptide, i.e. if, for example, interaction with the same agonist or antagonist or substrate, however at a different rate, (i.e. expressed by a EC_{50} or IC_{50} value or any other parameter suitable in the present technical field) is observed. Examples of suitable (conservative) amino acid substitutions are shown in the following table:

Original residue	Examples of substitution
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn ; Gln
Ile	Leu; Val
Leu	Ile; Val
Lys	Arg ; Gln ; Glu
Met	Leu ; Ile
Phe	Met ; Leu ; Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp ; Phe
Val	Ile; Leu

"Functional equivalents" in the above sense are also "precursors" of the polypeptides described herein, as well as "functional derivatives" and "salts" of the polypeptides.

"Precursors" are in that case natural or synthetic precursors of the polypeptides with or without the desired biological activity.

The expression "salts" means salts of carboxyl groups as well as salts of acid addition of amino groups of the protein molecules. Salts of carboxyl groups can be produced in a known way and comprise inorganic salts, for example sodium, calcium, ammonium, iron and zinc salts, and salts with organic bases, for example amines, such as triethanolamine, arginine, lysine, piperidine and the like. Salts of acid addition, for example salts with inorganic acids, such as hydrochloric acid or sulfuric acid and salts with organic acids, such as acetic acid and oxalic acid, are also covered by the invention.

"Functional derivatives" of polypeptides according to the invention can also be produced on functional amino acid side groups or at their N-terminal or C-terminal end using known techniques. Such derivatives comprise for example aliphatic esters of carboxylic acid groups, amides of carboxylic acid groups, obtainable by reaction with ammonia or with a primary or secondary amine; N-acyl derivatives of free amino groups, produced by reaction with acyl groups; or O-acyl derivatives of free hydroxyl groups, produced by reaction with acyl groups.

”Functional equivalents” naturally also comprise polypeptides that can be obtained from other organisms, as well as naturally occurring variants. For example, areas of homologous sequence regions can be established by sequence comparison, and equivalent polypeptides can be determined on the basis of the concrete parameters of the invention.

"Functional equivalents" also comprise “fragments”, like individual domains or sequence motifs, of the polypeptides according to the invention, or N- and or C-terminally truncated forms, which may or may not display the desired biological function. Preferably such “fragments” retain the desired biological function at least qualitatively.

"Functional equivalents" are, moreover, fusion proteins, which have one of the polypeptide sequences stated herein or functional equivalents derived there from and at least one further, functionally different, heterologous sequence in functional N-terminal or C-terminal association (i.e. without substantial mutual functional impairment of the fusion protein parts). Non-limiting examples of these heterologous sequences are e.g. signal peptides, histidine anchors or enzymes.

“Functional equivalents” which are also comprised in accordance with the invention are homologs to the specifically disclosed polypeptides. These have at least 60%, preferably at least 75%, in particular at least 80 or 85%, such as, for example, 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99%, homology (or identity) to one of the specifically disclosed amino acid sequences, calculated by the algorithm of Pearson and Lipman, Proc. Natl. Acad, Sci. (USA) 85(8), 1988, 2444-2448. A homology or identity, expressed as a percentage, of a homologous polypeptide according to the invention means in particular an identity, expressed as a percentage, of the amino acid residues based on the total length of one of the amino acid sequences described specifically herein.

The identity data, expressed as a percentage, may also be determined with the aid of BLAST alignments, algorithm blastp (protein-protein BLAST), or by applying the Clustal settings specified herein below.

In the case of a possible protein glycosylation, "functional equivalents" according to the invention comprise polypeptides as described herein in deglycosylated or glycosylated form as well as modified forms that can be obtained by altering the glycosylation pattern.

Functional equivalents or homologues of the polypeptides according to the invention can be produced by mutagenesis, e.g. by point mutation, lengthening or shortening of the protein or as described in more detail below.

Functional equivalents or homologs of the polypeptides according to the invention can be identified by screening combinatorial databases of mutants, for example shortening mutants. For example, a variegated database of protein variants can be produced by combinatorial mutagenesis at the nucleic acid level, e.g. by enzymatic ligation of a mixture of synthetic oligonucleotides. There are a great many methods that can be used for the production of databases of potential homologues from a degenerated oligonucleotide sequence. Chemical synthesis of a degenerated gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic gene can then be ligated in a suitable expression vector. The use of a degenerated genome makes it possible to supply all sequences in a mixture, which code for the desired set of potential protein sequences. Methods of synthesis of degenerated oligonucleotides are known to a person skilled in the art.

In the prior art, several techniques are known for the screening of gene products of combinatorial databases, which were produced by point mutations or shortening, and for the screening of cDNA libraries for gene products with a selected property. These techniques can be adapted for the rapid screening of the gene banks that were produced by combinatorial mutagenesis of homologues according to the invention. The techniques most frequently used for the screening of large gene banks, which are based on a high-throughput analysis, comprise cloning of the gene bank in expression vectors that can be replicated, transformation of the suitable cells with the resultant vector database and expression of the combinatorial genes in conditions in which detection of the desired activity facilitates isolation of the vector that codes for the gene whose product was detected. Recursive Ensemble Mutagenesis (REM), a technique that increases the frequency of functional mutants in the databases, can be used in combination with the screening tests, in order to identify homologues.

An embodiment provided herein provides orthologs and paralogs of polypeptides disclosed herein as well as methods for identifying and isolating such orthologs and paralogs. A definition of the terms “ortholog” and “paralog” is given below and applies to amino acid and nucleic acid sequences.

The polypeptides of the invention include all active forms, including active subsequences, e.g., catalytic domains or active sites, of an enzyme of the invention. In one aspect, the invention provides catalytic domains or active sites as set forth below. In one aspect, the invention provides a peptide or polypeptide comprising or consisting of an active site domain as predicted through use of a database such as Pfam (http://pfam.wustl.edu/hmmsearch.shtml) (which is a large collection of multiple sequence alignments and hidden Markov models covering many common protein families, The Pfam protein families database, A. Bateman, E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. L. Sonnhammer, Nucleic Acids Research, 30(1):276-280, 2002) or

equivalent, as for example InterPro and SMART databases (http://www.ebi.ac.uk/interpro/scan.html, http://smart.embl-heidelberg.de/).

The invention also encompasses “polypeptide variant” having the desired activity, wherein the variant polypeptide is selected from an amino acid sequence having at least 40%, 45%, 50%. 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or more sequence identity to a specific, in particular natural, amino acid sequence as referred to by a specific SEQ ID NO and contains at least one substitution modification relative to said SEQ ID NO.

Coding nucleic acid sequences applicable according to the invention

In this context the following definitions apply:

The terms “nucleic acid sequence,” “nucleic acid,” “nucleic acid molecule” and “polynucleotide” are used interchangeably meaning a sequence of nucleotides. A nucleic acid sequence may be a single-stranded or double-stranded deoxyribonucleotide, or ribonucleotide of any length, and include coding and non-coding sequences of a gene, exons, introns, sense and anti-sense complimentary sequences, genomic DNA, cDNA, miRNA, siRNA, mRNA, rRNA, tRNA, recombinant nucleic acid sequences, isolated and purified naturally occurring DNA and/or RNA sequences, synthetic DNA and RNA sequences, fragments, primers and nucleic acid probes. The skilled artisan is aware that the nucleic acid sequences of RNA are identical to the DNA sequences with the difference of thymine (T) being replaced by uracil (U). The term “nucleotide sequence” should also be understood as comprising a polynucleotide molecule or an oligonucleotide molecule in the form of a separate fragment or as a component of a larger nucleic acid.

An “isolated nucleic acid” or “isolated nucleic acid sequence” relates to a nucleic acid or nucleic acid sequence that is in an environment different from that in which the nucleic acid or nucleic acid sequence naturally occurs and can include those that are substantially free from contaminating endogenous material.

The term “naturally-occurring” as used herein as applied to a nucleic acid refers to a nucleic acid that is found in a cell of an organism in nature and which has not been intentionally modified by a human in the laboratory.

A “fragment” of a polynucleotide or nucleic acid sequence refers to contiguous nucleotides that is particularly at least 15 bp, at least 30 bp, at least 40 bp, at least 50 bp and/or at least 60 bp in length of the polynucleotide of an embodiment herein. Particularly the fragment of a polynucleotide comprises at least 25, more particularly at least 50, more particularly at least 75, more particularly at least 100, more particularly at least 150, more particularly at least 200, more particularly at least 300, more particularly at least 400, more particularly at least 500, more particularly at least 600, more particularly at least 700,

more particularly at least 800, more particularly at least 900, more particularly at least 1000 contiguous nucleotides of the polynucleotide of an embodiment herein. Without being limited, the fragment of the polynucleotides herein may be used as a PCR primer, and/or as a probe, or for anti-sense gene silencing or RNAi.

"Recombinant nucleic acid sequences" are nucleic acid sequences that result from the use of laboratory methods (for example, molecular cloning) to bring together genetic material from more than on source, creating or modifying a nucleic acid sequence that does not occur naturally and would not be otherwise found in biological organisms.

"Recombinant DNA technology" refers to molecular biology procedures to prepare a recombinant nucleic acid sequence as described, for instance, in Laboratory Manuals edited by Weigel and Glazebrook, 2002, Cold Spring Harbor Lab Press; and Sambrook *et al.*, 1989, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

The term "gene" means a DNA sequence comprising a region, which is transcribed into a RNA molecule, e.g., an mRNA in a cell, operably linked to suitable regulatory regions, e.g., a promoter. A gene may thus comprise several operably linked sequences, such as a promoter, a 5' leader sequence comprising, e.g., sequences involved in translation initiation, a coding region of cDNA or genomic DNA, introns, exons, and/or a 3'non-translated sequence comprising, e.g., transcription termination sites.

"Polycistronic" refers to nucleic acid molecules, in particular mRNAs, that can encode more than one polypeptide separately within the same nucleic acid molecule.

A "chimeric gene" refers to any gene which is not normally found in nature in a species, in particular, a gene in which one or more parts of the nucleic acid sequence are present that are not associated with each other in nature. For example, the promoter is not associated in nature with part or all of the transcribed region or with another regulatory region. The term "chimeric gene" is understood to include expression constructs in which a promoter or transcription regulatory sequence is operably linked to one or more coding sequences or to an antisense, i.e., reverse complement of the sense strand, or inverted repeat sequence (sense and antisense, whereby the RNA transcript forms double stranded RNA upon transcription). The term "chimeric gene" also includes genes obtained through the combination of portions of one or more coding sequences to produce a new gene.

A "3' UTR" or "3' non-translated sequence" (also referred to as "3' untranslated region," or "3'end") refers to the nucleic acid sequence found downstream of the coding sequence of a gene, which comprises, for example, a transcription termination site and (in most, but not all eukaryotic mRNAs) a polyadenylation signal such as AAUAAA or variants thereof. After termination of transcription, the mRNA transcript may be cleaved downstream of the polyadenylation signal and a poly(A) tail may be added, which is involved in the transport of the mRNA to the site of translation, e.g., cytoplasm.

The term “primer” refers to a short nucleic acid sequence that is hybridized to a template nucleic acid sequence and is used for polymerization of a nucleic acid sequence complementary to the template.

The term “selectable marker” refers to any gene which upon expression may be used to select a cell or cells that include the selectable marker. Examples of selectable markers are described below. The skilled artisan will know that different antibiotic, fungicide, auxotrophic or herbicide selectable markers are applicable to different target species.

The invention also relates to nucleic acid sequences that code for polypeptides as defined herein.

In particular, the invention also relates to nucleic acid sequences (single-stranded and double-stranded DNA and RNA sequences, e.g. cDNA, genomic DNA and mRNA), coding for one of the above polypeptides and their functional equivalents, which can be obtained for example using artificial nucleotide analogs.

The invention relates both to isolated nucleic acid molecules, which code for polypeptides according to the invention or biologically active segments thereof, and to nucleic acid fragments, which can be used for example as hybridization probes or primers for identifying or amplifying coding nucleic acids according to the invention.

The present invention also relates to nucleic acids with a certain degree of “identity” to the sequences specifically disclosed herein. "Identity" between two nucleic acids means identity of the nucleotides, in each case over the entire length of the nucleic acid.

The “identity” between two nucleotide sequences (the same applies to peptide or amino acid sequences) is a function of the number of nucleotide residues (or amino acid residues) or that are identical in the two sequences when an alignment of these two sequences has been generated. Identical residues are defined as residues that are the same in the two sequences in a given position of the alignment. The percentage of sequence identity, as used herein, is calculated from the optimal alignment by taking the number of residues identical between two sequences dividing it by the total number of residues in the shortest sequence and multiplying by 100. The optimal alignment is the alignment in which the percentage of identity is the highest possible. Gaps may be introduced into one or both sequences in one or more positions of the alignment to obtain the optimal alignment. These gaps are then taken into account as non-identical residues for the calculation of the percentage of sequence identity. Alignment for the purpose of determining the percentage of amino acid or nucleic acid sequence identity can be achieved in various ways using computer programs and for instance publicly available computer programs available on the world wide web.

Particularly, the BLAST program (Tatiana et al, *FEMS Microbiol Lett.*, 1999, 174:247-250, 1999) set to the default parameters, available from the National Center for Biotechnology Information (NCBI) website at ncbi.nlm.nih.gov/BLAST/bl2seq/wblast2.cgi, can be used to obtain an optimal alignment of protein or nucleic acid sequences and to calculate the percentage of sequence identity.

Alternatively, the identity may be determined according to Chenna, et al. (2003), the web page: http://www.ebi.ac.uk/Tools/clustalw/index.html# and the following settings:

DNA Gap Open Penalty	15.0
DNA Gap Extension Penalty	6.66
DNA Matrix	Identity
Protein Gap Open Penalty	10.0
Protein Gap Extension Penalty	0.2
Protein matrix	Gonnet
Protein/DNA ENDGAP	-1
Protein/DNA GAPDIST	4

All the nucleic acid sequences mentioned herein (single-stranded and double-stranded DNA and RNA sequences, for example cDNA and mRNA) can be produced in a known way by chemical synthesis from the nucleotide building blocks, e.g. by fragment condensation of individual overlapping, complementary nucleic acid building blocks of the double helix. Chemical synthesis of oligonucleotides can, for example, be performed in a known way, by the phosphoamidite method (Voet, Voet, 2nd edition, Wiley Press, New York, pages 896-897). The accumulation of synthetic oligonucleotides and filling of gaps by means of the Klenow fragment of DNA polymerase and ligation reactions as well as general cloning techniques are described in Sambrook et al. (1989), see below.

The nucleic acid molecules according to the invention can in addition contain non-translated sequences from the 3' and/or 5' end of the coding genetic region.

The invention further relates to the nucleic acid molecules that are complementary to the concretely described nucleotide sequences or a segment thereof.

The nucleotide sequences according to the invention make possible the production of probes and primers that can be used for the identification and/or cloning of homologous sequences in other cellular types and organisms. Such probes or primers generally comprise a nucleotide sequence region which hybridizes under "stringent" conditions (as defined herein elsewhere) on at least about 12, preferably at least about 25, for example about 40, 50 or 75 successive nucleotides of a sense strand of a nucleic acid sequence according to the invention or of a corresponding antisense strand.

"Homologous" sequences include orthologous or paralogous sequences. Methods of identifying orthologs or paralogs including phylogenetic methods, sequence similarity and hybridization methods are known in the art and are described herein.

"Paralogs" result from gene duplication that gives rise to two or more genes with similar sequences and similar functions. Paralogs typically cluster together and are formed by duplications of genes within related plant species. Paralogs are found in groups of similar genes using pair-wise Blast analysis or during phylogenetic analysis of gene families using programs such as CLUSTAL. In paralogs, consensus sequences can be identified characteristic to sequences within related genes and having similar functions of the genes.

"Orthologs", or orthologous sequences, are sequences similar to each other because they are found in species that descended from a common ancestor. For instance, plant species that have common ancestors are known to contain many enzymes that have similar sequences and functions. The skilled artisan can identify orthologous sequences and predict the functions of the orthologs, for example, by constructing a polygenic tree for a gene family of one species using CLUSTAL or BLAST programs. A method for identifying or confirming similar functions among homologous sequences is by comparing of the transcript profiles in host cells or organisms, such as plants or microorganisms, overexpressing or lacking (in knockouts/knockdowns) related polypeptides. The skilled person will understand that genes having similar transcript profiles, with greater than 50% regulated transcripts in common, or with greater than 70% regulated transcripts in common, or greater than 90% regulated transcripts in common will have similar functions. Homologs, paralogs, orthologs and any other variants of the sequences herein are expected to function in a similar manner by making the host cells, organism such as plants or microorganisms producing terpene cyclase proteins.

A nucleic acid molecule according to the invention can be recovered by means of standard techniques of molecular biology and the sequence information supplied according to the invention. For example, cDNA can be isolated from a suitable cDNA library, using one of the concretely disclosed complete sequences or a segment thereof as hybridization probe and standard hybridization techniques (as described for example in Sambrook, (1989)).

In addition, a nucleic acid molecule comprising one of the disclosed sequences or a segment thereof, can be isolated by the polymerase chain reaction, using the oligonucleotide primers that were constructed on the basis of this sequence. The nucleic acid amplified in this way can be cloned in a suitable vector and can be characterized by DNA sequencing. The oligonucleotides according to the invention can also be produced by standard methods of synthesis, e.g. using an automatic DNA synthesizer.

To test a function of variant DNA sequences according to an embodiment herein, the sequence of interest is operably linked to a selectable or screenable marker gene and expression of said reporter

gene is tested in transient expression assays, for example, with microorganisms or with protoplasts or in stably transformed plants.

The invention also relates to derivatives of the concretely disclosed or derivable nucleic acid sequences.

Thus, further nucleic acid sequences according to the invention can be derived from the sequences specifically disclosed herein and can differ from it by one or more, like 1 to 20, in particular 1 to 15 or 5 to 10 additions, substitutions, insertions or deletions of one or several (like for example 1 to 10) nucleotides, and furthermore code for polypeptides with the desired profile of properties.

The invention also encompasses nucleic acid sequences that comprise so-called silent mutations or have been altered, in comparison with a concretely stated sequence, according to the codon usage of a special original or host organism.

According to a particular embodiment of the invention variant nucleic acids may be prepared in order to adapt its nucleotide sequence to a specific expression system. For example, bacterial expression systems are known to more efficiently express polypeptides if amino acids are encoded by particular codons. Due to the degeneracy of the genetic code, more than one codon may encode the same amino acid sequence, multiple nucleic acid sequences can code for the same protein or polypeptide, all these DNA sequences being encompassed by an embodiment herein. Where appropriate, the nucleic acid sequences encoding the polypeptides described herein may be optimized for increased expression in the host cell. For example, nucleic acids of an embodiment herein may be synthesized using codons particular to a host for improved expression.

The invention also encompasses naturally occurring variants, e.g. splicing variants or allelic variants, of the sequences described therein.

Allelic variants may have at least 60 % homology at the level of the derived amino acid, preferably at least 80 % homology, quite especially preferably at least 90 % homology over the entire sequence range (regarding homology at the amino acid level, reference should be made to the details given above for the polypeptides). Advantageously, the homologies can be higher over partial regions of the sequences.

The invention also relates to sequences that can be obtained by conservative nucleotide substitutions (i.e. as a result thereof the amino acid in question is replaced by an amino acid of the same charge, size, polarity and/or solubility).

The invention also relates to the molecules derived from the concretely disclosed nucleic acids by sequence polymorphisms. Such genetic polymorphisms may exist in cells from different populations or within a population due to natural allelic variation. Allelic variants may also include functional

equivalents. These natural variations usually produce a variance of 1 to 5 % in the nucleotide sequence of a gene. Said polymorphisms may lead to changes in the amino acid sequence of the polypeptides disclosed herein. Allelic variants may also include functional equivalents.

Furthermore, derivatives are also to be understood to be homologs of the nucleic acid sequences according to the invention, for example animal, plant, fungal or bacterial homologs, shortened sequences, single-stranded DNA or RNA of the coding and noncoding DNA sequence. For example, homologs have, at the DNA level, a homology of at least 40 %, preferably of at least 60 %, especially preferably of at least 70 %, quite especially preferably of at least 80 % over the entire DNA region given in a sequence specifically disclosed herein.

Moreover, derivatives are to be understood to be, for example, fusions with promoters. The promoters that are added to the stated nucleotide sequences can be modified by at least one nucleotide exchange, at least one insertion, inversion and/or deletion, though without impairing the functionality or efficacy of the promoters. Moreover, the efficacy of the promoters can be increased by altering their sequence or can be exchanged completely with more effective promoters even of organisms of a different genus.

Generation of functional polypeptide mutants

Moreover, a person skilled in the art is familiar with methods for generating functional mutants, that is to say nucleotide sequences which code for a polypeptide with at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or more sequence identity to anyone of amino acid related to SEQ ID NOs as disclosed herein and/or encoded by a nucleic acid molecule comprising a nucleotide sequence having at least 50% sequence identity to anyone of the nucleotide related to SEQ ID NOs as disclosed herein.

Depending on the technique used, a person skilled in the art can introduce entirely random or else more directed mutations into genes or else noncoding nucleic acid regions (which are for example important for regulating expression) and subsequently generate genetic libraries. The methods of molecular biology required for this purpose are known to the skilled worker and for example described in Sambrook and Russell, Molecular Cloning. 3rd Edition, Cold Spring Harbor Laboratory Press 2001.

Methods for modifying genes and thus for modifying the polypeptide encoded by them have been known to the skilled worker for a long time, such as, for example:

- site-specific mutagenesis, where individual or several nucleotides of a gene are replaced in a directed fashion (Trower MK (Ed.) 1996; In vitro mutagenesis protocols. Humana Press, New Jersey),
- saturation mutagenesis, in which a codon for any amino acid can be exchanged or added at any point of a gene (Kegler-Ebo DM, Docktor CM, DiMaio D (1994) Nucleic Acids Res 22:1593; Barettino

D, Feigenbutz M, Valcárel R, Stunnenberg HG (1994) Nucleic Acids Res 22:541; Barik S (1995) Mol Biotechnol 3:1),

- error-prone polymerase chain reaction, where nucleotide sequences are mutated by error-prone DNA polymerases (Eckert KA, Kunkel TA (1990) Nucleic Acids Res 18:3739);
- the SeSaM method (sequence saturation method), in which preferred exchanges are prevented by the polymerase. Schenk et al., Biospektrum, Vol. 3, 2006, 277-279
- the passaging of genes in mutator strains, in which, for example owing to defective DNA repair mechanisms, there is an increased mutation rate of nucleotide sequences (Greener A, Callahan M, Jerpseth B (1996) An efficient random mutagenesis technique using an *E.coli* mutator strain. In: Trower MK (Ed.) In vitro mutagenesis protocols. Humana Press, New Jersey), or
- DNA shuffling, in which a pool of closely related genes is formed and digested and the fragments are used as templates for a polymerase chain reaction in which, by repeated strand separation and reassociation, full-length mosaic genes are ultimately generated (Stemmer WPC (1994) Nature 370:389; Stemmer WPC (1994) Proc Natl Acad Sci USA 91:10747).

Using so-called directed evolution (described, inter alia, in Reetz MT and Jaeger K-E (1999), Topics Curr Chem 200:31; Zhao H, Moore JC, Volkov AA, Arnold FH (1999), Methods for optimizing industrial polypeptides by directed evolution, In: Demain AL, Davies JE (Ed.) Manual of industrial microbiology and biotechnology. American Society for Microbiology), a skilled worker can produce functional mutants in a directed manner and on a large scale. To this end, in a first step, gene libraries of the respective polypeptides are first produced, for example using the methods given above. The gene libraries are expressed in a suitable way, for example by bacteria or by phage display systems.

The relevant genes of host organisms which express functional mutants with properties that largely correspond to the desired properties can be submitted to another mutation cycle. The steps of the mutation and selection or screening can be repeated iteratively until the present functional mutants have the desired properties to a sufficient extent. Using this iterative procedure, a limited number of mutations, for example 1, 2, 3, 4 or 5 mutations, can be performed in stages and assessed and selected for their influence on the activity in question. The selected mutant can then be submitted to a further mutation step in the same way. In this way, the number of individual mutants to be investigated can be reduced significantly.

The results according to the invention also provide important information relating to structure and sequence of the relevant polypeptides, which is required for generating, in a targeted fashion, further polypeptides with desired modified properties. In particular, it is possible to define so-called “hot spots”, i.e. sequence segments that are potentially suitable for modifying a property by introducing targeted mutations.

Information can also be deduced regarding amino acid sequence positions, in the region of which mutations can be affected that should probably have little effect on the activity, and can be designated as potential “silent mutations”.

Constructs for expressing polypeptides of the invention and/or used in the process of the invention

In this context the following definitions apply:

“Expression of a gene” encompasses “heterologous expression” and “over-expression” and involves transcription of the gene and translation of the mRNA into a protein. Overexpression refers to the production of the gene product as measured by levels of mRNA, polypeptide and/or enzyme activity in transgenic cells or organisms that exceeds levels of production in non-transformed cells or organisms of a similar genetic background.

“Expression vector” as used herein means a nucleic acid molecule engineered using molecular biology methods and recombinant DNA technology for delivery of foreign or exogenous DNA into a host cell. The expression vector typically includes sequences required for proper transcription of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for an RNA, e.g., an antisense RNA, siRNA and the like.

An “expression vector” as used herein includes any linear or circular recombinant vector including but not limited to viral vectors, bacteriophages and plasmids. The skilled person is capable of selecting a suitable vector according to the expression system. In one embodiment, the expression vector includes the nucleic acid of an embodiment herein operably linked to at least one “regulatory sequence”, which controls transcription, translation, initiation and termination, such as a transcriptional promoter, operator or enhancer, or an mRNA ribosomal binding site and, optionally, including at least one selection marker. Nucleotide sequences are “operably linked” when the regulatory sequence functionally relates to the nucleic acid of an embodiment herein.

An “expression system” as used herein encompasses any combination of nucleic acid molecules required for the expression of one, or the co-expression of two or more polypeptides either *in vivo* of a given expression host, or *in vitro*. The respective coding sequences may either be located on a single nucleic acid molecule or vector, as for example a vector containing multiple cloning sites, or on a polycistronic nucleic acid, or may be distributed over two or more physically distinct vectors. As a particular example there may be mentioned an operon comprising a promotor sequence, one or more operator sequences and one or more structural genes each encoding an enzyme as described herein

As used herein, the terms "amplifying" and "amplification" refer to the use of any suitable amplification methodology for generating or detecting recombinant of naturally expressed nucleic acid, as described in detail, below. For example, the invention provides methods and reagents (e.g., specific degenerate

oligonucleotide primer pairs, oligo dT primer) for amplifying (e.g., by polymerase chain reaction, PCR) naturally expressed (e.g., genomic DNA or mRNA) or recombinant (e.g., cDNA) nucleic acids of the invention *in vivo*, *ex vivo* or *in vitro*.

"Regulatory sequence" refers to a nucleic acid sequence that determines expression level of the nucleic acid sequences of an embodiment herein and is capable of regulating the rate of transcription of the nucleic acid sequence operably linked to the regulatory sequence. Regulatory sequences comprise promoters, enhancers, transcription factors, promoter elements and the like.

A "promoter", a "nucleic acid with promoter activity" or a "promoter sequence" is understood as meaning, in accordance with the invention, a nucleic acid which, when functionally linked to a nucleic acid to be transcribed, regulates the transcription of said nucleic acid. "Promoter" in particular refers to a nucleic acid sequence that controls the expression of a coding sequence by providing a binding site for RNA polymerase and other factors required for proper transcription including without limitation transcription factor binding sites, repressor and activator protein binding sites. The meaning of the term promoter also includes the term "promoter regulatory sequence". Promoter regulatory sequences may include upstream and downstream elements that may influences transcription, RNA processing or stability of the associated coding nucleic acid sequence. Promoters include naturally-derived and synthetic sequences. The coding nucleic acid sequences is usually located downstream of the promoter with respect to the direction of the transcription starting at the transcription initiation site.

In this context, a "functional" or "operative" linkage is understood as meaning for example the sequential arrangement of one of the nucleic acids with a regulatory sequence. For example the sequence with promoter activity and of a nucleic acid sequence to be transcribed and optionally further regulatory elements, for example nucleic acid sequences which ensure the transcription of nucleic acids, and for example a terminator, are linked in such a way that each of the regulatory elements can perform its function upon transcription of the nucleic acid sequence. This does not necessarily require a direct linkage in the chemical sense. Genetic control sequences, for example enhancer sequences, can even exert their function on the target sequence from more remote positions or even from other DNA molecules. Preferred arrangements are those in which the nucleic acid sequence to be transcribed is positioned behind (i.e. at the 3'-end of) the promoter sequence so that the two sequences are joined together covalently. The distance between the promoter sequence and the nucleic acid sequence to be expressed recombinantly can be smaller than 200 base pairs, or smaller than 100 base pairs or smaller than 50 base pairs.

In addition to promoters and terminator, the following may be mentioned as examples of other regulatory elements: targeting sequences, enhancers, polyadenylation signals, selectable markers, amplification signals, replication origins and the like. Suitable regulatory sequences are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990).

The term “constitutive promoter” refers to an unregulated promoter that allows for continual transcription of the nucleic acid sequence it is operably linked to.

As used herein, the term “operably linked” refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is “operably linked” when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter, or rather a transcription regulatory sequence, is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous. The nucleotide sequence associated with the promoter sequence may be of homologous or heterologous origin with respect to the plant to be transformed. The sequence also may be entirely or partially synthetic. Regardless of the origin, the nucleic acid sequence associated with the promoter sequence will be expressed or silenced in accordance with promoter properties to which it is linked after binding to the polypeptide of an embodiment herein. The associated nucleic acid may code for a protein that is desired to be expressed or suppressed throughout the organism at all times or, alternatively, at a specific time or in specific tissues, cells, or cell compartment. Such nucleotide sequences particularly encode proteins conferring desirable phenotypic traits to the host cells or organism altered or transformed therewith. More particularly, the associated nucleotide sequence leads to the production of the product or products of interest as herein defined in the cell or organism. Particularly, the nucleotide sequence encodes a polypeptide having an enzyme activity as herein defined.

The nucleotide sequence as described herein above may be part of an “expression cassette”. The terms “expression cassette” and “expression construct” are used synonymously. The (preferably recombinant) expression construct contains a nucleotide sequence which encodes a polypeptide according to the invention and which is under genetic control of regulatory nucleic acid sequences.

In a process applied according to the invention, the expression cassette may be part of an “expression vector”, in particular of a recombinant expression vector.

An “expression unit” is understood as meaning, in accordance with the invention, a nucleic acid with expression activity which comprises a promoter as defined herein and, after functional linkage with a nucleic acid to be expressed or a gene, regulates the expression, i.e. the transcription and the translation of said nucleic acid or said gene. It is therefore in this connection also referred to as a “regulatory nucleic acid sequence”. In addition to the promoter, other regulatory elements, for example enhancers, can also be present.

An “expression cassette” or “expression construct” is understood as meaning, in accordance with the invention, an expression unit which is functionally linked to the nucleic acid to be expressed or the gene to be expressed. In contrast to an expression unit, an expression cassette therefore comprises not only

nucleic acid sequences which regulate transcription and translation, but also the nucleic acid sequences that are to be expressed as protein as a result of transcription and translation.

The terms “expression” or “overexpression” describe, in the context of the invention, the production or increase in intracellular activity of one or more polypeptides in a microorganism, which are encoded by the corresponding DNA. To this end, it is possible for example to introduce a gene into an organism, replace an existing gene with another gene, increase the copy number of the gene(s), use a strong promoter or use a gene which encodes for a corresponding polypeptide with a high activity; optionally, these measures can be combined.

Preferably such constructs according to the invention comprise a promoter 5’-upstream of the respective coding sequence and a terminator sequence 3’-downstream and optionally other usual regulatory elements, in each case in operative linkage with the coding sequence.

Nucleic acid constructs according to the invention comprise in particular a sequence coding for a polypeptide for example derived from the amino acid related SEQ ID NOs as described therein or the reverse complement thereof, or derivatives and homologs thereof and which have been linked operatively or functionally with one or more regulatory signals, advantageously for controlling, for example increasing, gene expression.

In addition to these regulatory sequences, the natural regulation of these sequences may still be present before the actual structural genes and optionally may have been genetically modified so that the natural regulation has been switched off and expression of the genes has been enhanced. The nucleic acid construct may, however, also be of simpler construction, i.e. no additional regulatory signals have been inserted before the coding sequence and the natural promoter, with its regulation, has not been removed. Instead, the natural regulatory sequence is mutated such that regulation no longer takes place and the gene expression is increased.

A preferred nucleic acid construct advantageously also comprises one or more of the already mentioned “enhancer” sequences in functional linkage with the promoter, which sequences make possible an enhanced expression of the nucleic acid sequence. Additional advantageous sequences may also be inserted at the 3’-end of the DNA sequences, such as further regulatory elements or terminators. One or more copies of the nucleic acids according to the invention may be present in a construct. In the construct, other markers, such as genes which complement auxotrophisms or antibiotic resistances, may also optionally be present so as to select for the construct.

Examples of suitable regulatory sequences are present in promoters such as cos, tac, trp, tet, trp-tet, lpp, lac, lpp-lac, $lacI^q$, T7, T5, T3, gal, trc, ara, rhaP ($rhaP_{BAD}$)SP6, lambda-P_R or in the lambda-P_L promoter, and these are advantageously employed in Gram-negative bacteria. Further advantageous regulatory sequences are present for example in the Gram-positive promoters amy and SPO2, in the

yeast or fungal promoters ADC1, MFalpha, AC, P-60, CYC1, GAPDH, TEF, rp28, ADH. Artificial promoters may also be used for regulation.

For expression in a host organism, the nucleic acid construct is inserted advantageously into a vector such as, for example, a plasmid or a phage, which makes possible optimal expression of the genes in the host. Vectors are also understood as meaning, in addition to plasmids and phages, all the other vectors which are known to the skilled worker, that is to say for example viruses such as SV40, CMV, baculovirus and adenovirus, transposons, IS elements, phasmids, cosmids and linear or circular DNA or artificial chromosomes. These vectors are capable of replicating autonomously in the host organism or else chromosomally. These vectors are a further development of the invention. Binary or cpo-integration vectors are also applicable.

Suitable plasmids are, for example, in E. coli pLG338, pACYC184, pBR322, pUC18, pUC19, pKC30, pRep4, pHS1, pKK223-3, pDHE19.2, pHS2, pPLc236, pMBL24, pLG200, pUR290, pIN-III^{113}-B1, λgt11 or pBdCI, in Streptomyces pIJ101, pIJ364, pIJ702 or pIJ361, in Bacillus pUB110, pC194 or pBD214, in Corynebacterium pSA77 or pAJ667, in fungi pALS1, pIL2 or pBB116, in yeasts 2alphaM, pAG-1, YEp6, YEp13 or pEMBLYe23 or in plants pLGV23, pGHlac^{+}, pBIN19, pAK2004 or pDH51. The abovementioned plasmids are a small selection of the plasmids which are possible. Further plasmids are well known to the skilled worker and can be found for example in the book Cloning Vectors (Eds. Pouwels P. H. et al. Elsevier, Amsterdam-New York-Oxford, 1985, ISBN 0 444 904018).

In a further development of the vector, the vector which comprises the nucleic acid construct according to the invention or the nucleic acid according to the invention can advantageously also be introduced into the microorganisms in the form of a linear DNA and integrated into the host organism's genome via heterologous or homologous recombination. This linear DNA can consist of a linearized vector such as a plasmid or only of the nucleic acid construct or the nucleic acid according to the invention.

For optimal expression of heterologous genes in organisms, it is advantageous to modify the nucleic acid sequences to match the specific "codon usage" used in the organism. The "codon usage" can be determined readily by computer evaluations of other, known genes of the organism in question.

An expression cassette according to the invention is generated by fusing a suitable promoter to a suitable coding nucleotide sequence and a terminator or polyadenylation signal. Customary recombination and cloning techniques are used for this purpose, as are described, for example, in T. Maniatis, E.F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) and in T.J. Silhavy, M.L. Berman and L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and in Ausubel, F.M. et al., Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience (1987).

For expression in a suitable host organism, the recombinant nucleic acid construct or gene construct is advantageously inserted into a host-specific vector which makes possible optimal expression of the genes in the host. Vectors are well known to the skilled worker and can be found for example in “cloning vectors” (Pouwels P. H. et al., Ed., Elsevier, Amsterdam-New York-Oxford, 1985).

An alternative embodiment of an embodiment herein provides a method to “alter gene expression” in a host cell. For instance, the polynucleotide of an embodiment herein may be enhanced or overexpressed or induced in certain contexts (e.g. upon exposure to certain temperatures or culture conditions) in a host cell or host organism.

Alteration of expression of a polynucleotide provided herein may also result in ectopic expression which is a different expression pattern in an altered and in a control or wild-type organism. Alteration of expression occurs from interactions of polypeptide of an embodiment herein with exogenous or endogenous modulators, or as a result of chemical modification of the polypeptide. The term also refers to an altered expression pattern of the polynucleotide of an embodiment herein which is altered below the detection level or completely suppressed activity.

In one embodiment, provided herein is also an isolated, recombinant or synthetic polynucleotide encoding a polypeptide or variant polypeptide provided herein.

In one embodiment, several polypeptide encoding nucleic acid sequences are co-expressed in a single host, particularly under control of different promoters. In another embodiment, several polypeptide encoding nucleic acid sequences can be present on a single transformation vector or be co-transformed at the same time using separate vectors and selecting transformants comprising both chimeric genes. Similarly, one or polypeptide encoding genes may be expressed in a single plant, cell, microorganism or organism together with other chimeric genes.

Recombinant production of polypeptides according to the invention

The invention further relates to methods for recombinant production of polypeptides according to the invention or functional, biologically active fragments thereof, wherein a polypeptide-producing microorganism is cultured, optionally the expression of the polypeptides is induced by applying at least one inducer inducing gene expression and the expressed polypeptides are isolated from the culture. The polypeptides can also be produced in this way on an industrial scale, if desired.

The microorganisms produced according to the invention can be cultured continuously or discontinuously in the batch method or in the fed-batch method or repeated fed-batch method. A summary of known cultivation methods can be found in the textbook by Chmiel (Bioprozesstechnik 1. Einführung in die Bioverfahrenstechnik [Bioprocess technology 1. Introduction to bioprocess technology] (Gustav Fischer Verlag, Stuttgart, 1991)) or in the textbook by Storhas (Bioreaktoren und

periphere Einrichtungen [Bioreactors and peripheral equipment] (Vieweg Verlag, Braunschweig/Wiesbaden, 1994)). Standard laboratory methods can be used for this purpose and are known in the art and also further described herein.

If the polypeptides are not secreted in the culture medium, the cells can also be lysed and the product can be obtained from the lysate by known methods for isolation of proteins. The cells can optionally be disrupted with high-frequency ultrasound, high pressure, for example in a French press, by osmolysis, by the action of detergents, lytic enzymes or organic solvents, by means of homogenizers or by a combination of several of the aforementioned methods.

The polypeptides can be purified by known chromatographic techniques, such as molecular sieve chromatography (gel filtration), such as Q-sepharose chromatography, ion exchange chromatography and hydrophobic chromatography, and with other usual techniques such as ultrafiltration, crystallization, salting-out, dialysis and native gel electrophoresis. Suitable methods are described for example in Cooper, T. G., Biochemische Arbeitsmethoden [Biochemical processes], Verlag Walter de Gruyter, Berlin, New York or in Scopes, R., Protein Purification, Springer Verlag, New York, Heidelberg, Berlin.

For isolating the recombinant protein, it can be advantageous to use vector systems or oligonucleotides, which lengthen the cDNA by defined nucleotide sequences and therefore code for altered polypeptides or fusion proteins, which for example serve for easier purification. Suitable modifications of this type are for example so-called "tags" functioning as anchors, for example the modification known as hexa-histidine anchor or epitopes that can be recognized as antigens of antibodies (described for example in Harlow, E. and Lane, D., 1988, Antibodies: A Laboratory Manual. Cold Spring Harbor (N.Y.) Press). These anchors can serve for attaching the proteins to a solid carrier, for example a polymer matrix, which can for example be used as packing in a chromatography column, or can be used on a microtiter plate or on some other carrier.

At the same time these anchors can also be used for recognition of the proteins. For recognition of the proteins, it is moreover also possible to use usual markers, such as fluorescent dyes, enzyme markers, which form a detectable reaction product after reaction with a substrate, or radioactive markers, alone or in combination with the anchors for derivatization of the proteins.

Uses of a compound of formula (I) and other compounds of the invention

A further aspect of the invention comprises the use of a compound of formula (I) or another (intermediate) compound obtained or obtainable from the embodiments of the invention as a perfumery, flavor or aroma ingredient or as a precursor for making said ingredient.

As mentioned above, the invention comprises the use of a compound of formula (I) as a perfuming ingredient. In other words, it concerns a method or a process to confer, enhance, improve or modify the

odor properties of a perfuming composition or of a perfumed article or of a surface, which method comprises adding to said composition or article an effective amount of at least a compound of formula (I), e.g. to impart its typical note. Understood that the final hedonic effect may depend on the precise dosage and on the organoleptic properties of the invention's compound, but anyway the addition of the invention's compound will impart to the final product its typical touch in the form of a note, touch or aspect depending on the dosage.

By "use of a compound of formula (I)" it has to be understood here also the use of any composition containing a compound (I) and which can be advantageously employed in the perfumery industry.

Said compositions, which in fact can be advantageously employed as perfuming ingredients, are also an object of the present invention.

Therefore, another object of the present invention is a perfuming composition comprising:

i) as a perfuming ingredient, at least one invention's compound as defined above;

ii) at least one ingredient selected from the group consisting of a perfumery carrier and a perfumery base; and

iii) optionally at least one perfumery adjuvant.

By "perfumery carrier" it is meant here a material which is practically neutral from a perfumery point of view, i.e. that does not significantly alter the organoleptic properties of perfuming ingredients. Said carrier may be a liquid or a solid.

As liquid carrier one may cite, as non-limiting examples, an emulsifying system, i.e. a solvent and a surfactant system, or a solvent commonly used in perfumery. A detailed description of the nature and type of solvents commonly used in perfumery cannot be exhaustive. However, one can cite as non-limiting examples, solvents such as butylene or propylene glycol, glycerol, dipropyleneglycol and its monoether, 1,2,3-propanetriyl triacetate, dimethyl glutarate, dimethyl adipate 1,3-diacetyloxypropan-2-yl acetate, diethyl phthalate, isopropyl myristate, benzyl benzoate, benzyl alcohol, 2-(2-ethoxyethoxy)-1-ethanol, tri-ethyl citrate or mixtures thereof, which are the most commonly used. For the compositions which comprise both a perfumery carrier and a perfumery base, other suitable perfumery carriers than those previously specified, can be also ethanol, water/ethanol mixtures, limonene or other terpenes, isoparaffins such as those known under the trademark Isopar™ (origin: Exxon Chemical) or glycol ethers and glycol ether esters such as those known under the trademark Dowanol™ (origin: Dow Chemical Company), or hydrogenated castors oils such as those known under the trademark Cremophor® RH 40 (origin: BASF).

Solid carrier is meant to designate a material to which the perfuming composition or some element of the perfuming composition can be chemically or physically bound. In general, such solid carriers are employed either to stabilize the composition, or to control the rate of evaporation of the compositions

or of some ingredients. Solid carriers are of current use in the art and a person skilled in the art knows how to reach the desired effect. However, by way of non-limiting examples of solid carriers, one may cite absorbing gums or polymers or inorganic materials, such as porous polymers, cyclodextrins, wood-based materials, organic or inorganic gels, clays, gypsum talc or zeolites.

As other non-limiting examples of solid carriers, one may cite encapsulating materials. Examples of such materials may comprise wall-forming and plasticizing materials, such as mono, di- or trisaccharides, natural or modified starches, hydrocolloids, cellulose derivatives, polyvinyl acetates, polyvinylalcohols, proteins or pectins, or yet the materials cited in reference texts such as H. Scherz, Hydrokolloide: Stabilisatoren, Dickungs- und Geliermittel in Lebensmitteln, Band 2 der Schriftenreihe Lebensmittelchemie, Lebensmittelqualität, Behr's Verlag GmbH & Co., Hamburg, 1996. The encapsulation is a well-known process to a person skilled in the art, and may be performed, for instance, by using techniques such as spray-drying, agglomeration or yet extrusion; or consists of a coating encapsulation, including coacervation and complex coacervation techniques.

As non-limiting examples of solid carriers, one may cite in particular the core-shell capsules with resins of aminoplast, polyamide, polyester, polyurea or polyurethane type or a mixture thereof (all of said resins are well known to a person skilled in the art) using techniques like phase separation process induced by polymerization, interfacial polymerization, coacervation or altogether (all of said techniques have been described in the prior art), optionally in the presence of a polymeric stabilizer or of a cationic copolymer.

Resins may be produced by the polycondensation of an aldehyde (e.g. formaldehyde, 2,2-dimethoxyethanal, glyoxal, glyoxylic acid or glycolaldehyde and mixtures thereof) with an amine such as urea, benzoguanamine, glycoluryl, melamine, methylol melamine, methylated methylol melamine, guanazole and the like, as well as mixtures thereof. Alternatively, one may use preformed resins alkylolated polyamines such as those commercially available under the trademark Urac® (origin: Cytec Technology Corp.), Cymel® (origin: Cytec Technology Corp.), Urecoll® or Luracoll® (origin: BASF).

Other resins are the ones produced by the polycondensation of an a polyol, like glycerol, and a polyisocyanate, like a trimer of hexamethylene diisocyanate, a trimer of isophorone diisocyanate or xylylene diisocyanate or a Biuret of hexamethylene diisocyanate or a trimer of xylylene diisocyanate with trimethylolpropane (known with the tradename of Takenate®, origin: Mitsui Chemicals), among which a trimer of xylylene diisocyanate with trimethylolpropane and a Biuret of hexamethylene diisocyanate are preferred.

Some of the seminal literature related to the encapsulation of perfumes by polycondensation of amino resins, namely melamine-based resins with aldehydes includes articles such as those published by K. Dietrich et al. Acta Polymerica, 1989, vol. 40, pages 243, 325 and 683, as well as 1990, vol. 41, page 91. Such articles already describe the various parameters affecting the preparation of such core-shell

microcapsules following prior art methods that are also further detailed and exemplified in the patent literature. US 4'396'670, to the Wiggins Teape Group Limited is a pertinent early example of the latter. Since then, many other authors have enriched the literature in this field and it would be impossible to cover all published developments here, but the general knowledge in encapsulation technology is very significant. More recent publications of pertinence, which disclose suitable uses of such microcapsules, are represented for example by the article of K. Bruyninckx and M. Dusselier, ACS Sustainable Chemistry & Engineering, 2019, vol. 7, pages 8041-8054.

By "perfumery base" what is meant here is a composition comprising at least one perfuming co-ingredient.

Said perfuming co-ingredient is not of formula (I). Moreover, by "perfuming co-ingredient" it is meant here a compound, which is used in a perfuming preparation or a composition to impart a hedonic effect. In other words such a co-ingredient, to be considered as being a perfuming one, must be recognized by a person skilled in the art as being able to impart or modify in a positive or pleasant way the odor of a composition, and not just as having an odor. The perfuming ingredient may impart an additional benefit beyond that of modifying or imparting an odor, such as long-lasting, blooming, malodour counteraction, antimicrobial effect, antiviral effect, microbial stability, or pest control.

The nature and type of the perfuming co-ingredients present in the base do not warrant a more detailed description here, which in any case would not be exhaustive, the skilled person being able to select them on the basis of his general knowledge and according to the intended use or application and the desired organoleptic effect. In general terms, these perfuming co-ingredients belong to chemical classes as varied as alcohols, lactones, aldehydes, ketones, esters, ethers, acetates, nitriles, terpenoids, nitrogenous or sulphurous heterocyclic compounds and essential oils, and said perfuming co-ingredients can be of natural or synthetic origin.

In particular one may cite perfuming co-ingredients which are commonly used in perfume formulations, such as:

- Aldehydic ingredients: decanal, dodecanal, 2-methyl-undecanal, 10-undecenal, octanal, nonanal and/or nonenal;
- Aromatic-herbal ingredients: eucalyptus oil, camphor, eucalyptol, 5-methyltricyclo[6.2.1.0~2,7~]undecan-4-one, 1-methoxy-3-hexanethiol, 2-ethyl-4,4-dimethyl-1,3-oxathiane, 2,2,7/8,9/10-Tetramethylspiro[5.5]undec-8-en-1-one, menthol and/or alpha-pinene;
- Balsamic ingredients: coumarin, ethylvanillin and/or vanillin;
- Citrus ingredients: dihydromyrcenol, citral, orange oil, linalyl acetate, citronellyl nitrile, orange terpenes, limonene, 1-p-menthen-8-yl acetate and/or 1,4(8)-p-menthadiene;
- Floral ingredients:methyl dihydrojasmonate, linalool, citronellol, phenylethanol, 3-(4-tert-butylphenyl)-2-methylpropanal, hexylcinnamic aldehyde, benzyl acetate, benzyl salicylate, tetrahydro-2-isobutyl-4-methyl-4(2H)-pyranol, beta ionone, methyl 2-(methylamino)benzoate, (E)-3-methyl-4-

(2,6,6-trimethyl-2-cyclohexen-1-yl)-3-buten-2-one, (1E)-1-(2,6,6-trimethyl-2-cyclohexen-1-yl)-1-penten-3-one, 1-(2,6,6-trimethyl-1,3-cyclohexadien-1-yl)-2-buten-1-one, (2E)-1-(2,6,6-trimethyl-2-cyclohexen-1-yl)-2-buten-1-one, (2E)-1-[2,6,6-trimethyl-3-cyclohexen-1-yl]-2-buten-1-one, (2E)-1-(2,6,6-trimethyl-1-cyclohexen-1-yl)-2-buten-1-one, 2,5-dimethyl-2-indanmethanol, 2,6,6-trimethyl-3-cyclohexene-1-carboxylate, 3-(4,4-dimethyl-1-cyclohexen-1-yl)propanal, 3-(3,3/1,1-dimethyl-5-indanyl)propanal, hexyl salicylate, 3,7-dimethyl-1,6-nonadien-3-ol, 3-(4-isopropylphenyl)-2-methylpropanal, verdyl acetate, geraniol, p-menth-1-en-8-ol, 4-(1,1-dimethylethyl)-1-cyclohexyle acetate, 1,1-dimethyl-2-phenylethyl acetate, 4-cyclohexyl-2-methyl-2-butanol, amyl salicylate , high cis methyl dihydrojasmonate, 3-methyl-5-phenyl-1-pentanol, verdyl proprionate, geranyl acetate, tetrahydro linalool, cis-7-p-menthanol, propyl (S)-2-(1,1-dimethylpropoxy)propanoate, 2-methoxynaphthalene, 2,2,2-trichloro-1-phenylethyl acetate, 4/3-(4-hydroxy-4-methylpentyl)-3-cyclohexene-1-carbaldehyde, amylcinnamic aldehyde, 8-decen-5-olide, 4-phenyl-2-butanone, isononyle acetate, 4-(1,1-dimethylethyl)-1-cyclohexyl acetate, verdyl isobutyrate and/or mixture of methylionones isomers;

- Fruity ingredients: gamma-undecalactone, 2,2,5-trimethyl-5-pentylcyclopentanone, 2-methyl-4-propyl-1,3-oxathiane, 4-decanolide, ethyl 2-methyl-pentanoate, hexyl acetate, ethyl 2-methylbutanoate, gamma-nonalactone, allyl heptanoate, 2-phenoxyethyl isobutyrate, ethyl 2-methyl-1,3-dioxolane-2-acetate, diethyl 1,4-cyclohexanedicarboxylate, 3-methyl-2-hexen-1-yl acetate, 1-[3,3-dimethylcyclohexyl]ethyl [3-ethyl-2-oxiranyl]acetate and/or diethyl 1,4-cyclohexane dicarboxylate;
- Green ingredients: 2-methyl-3-hexanone (E)-oxime, 2,4-dimethyl-3-cyclohexene-1-carbaldehyde, 2-tert-butyl-1-cyclohexyl acetate, styrallyl acetate, allyl (2-methylbutoxy)acetate, 4-methyl-3-decen-5-ol, diphenyl ether, (Z)-3-hexen-1-ol and/or 1-(5,5-dimethyl-1-cyclohexen-1-yl)-4-penten-1-one;
- Musk ingredients: 1,4-dioxa-5,17-cycloheptadecanedione, (Z)-4-cyclopentadecen-1-one, 3-methylcyclopentadecanone, 1-oxa-12-cyclohexadecen-2-one, 1-oxa-13-cyclohexadecen-2-one, (9Z)-9-cycloheptadecen-1-one, 2-{(1S)-1-[(1R)-3,3-dimethylcyclohexyl]ethoxy}-2-oxoethyl propionate, 3-methyl-5-cyclopentadecen-1-one, 4,6,6,7,8,8-hexamethyl-1,3,4,6,7,8-hexahydrocyclopenta[g]isochromene, (1S,1'R)-2-[1-(3',3'-dimethyl-1'-cyclohexyl)ethoxy]-2-methylpropyl propanoate, oxacyclohexadecan-2-one and/or (1S,1'R)-[1-(3',3'-dimethyl-1'-cyclohexyl)ethoxycarbonyl]methyl propanoate;
- Woody ingredients: 1-[(1RS,6SR)-2,2,6-trimethylcyclohexyl]-3-hexanol, 3,3-dimethyl-5-[(1R)-2,2,3-trimethyl-3-cyclopenten-1-yl]-4-penten-2-ol, 3,4'-dimethylspiro[oxirane-2,9'-tricyclo[6.2.1.02,7]undec[4]ene, (1-ethoxyethoxy)cyclododecane, 2,2,9,11-tetramethylspiro[5.5]undec-8-en-1-yl acetate, 1-(octahydro-2,3,8,8-tetramethyl-2-naphtalenyl)-1-ethanone, patchouli oil, terpenes fractions of patchouli oil, Clearwood®, (1'R,E)-2-ethyl-4-(2',2',3'-trimethyl-3'-cyclopenten-1'-yl)-2-buten-1-ol, 2-ethyl-4-(2,2,3-trimethyl-3-cyclopenten-1-yl)-2-buten-1-ol, methyl cedryl ketone, 5-(2,2,3-trimethyl-3-cyclopentenyl)-3-methylpentan-2-ol, 1-(2,3,8,8-tetramethyl-1,2,3,4,6,7,8,8a-octahydronaphthalen-2-yl)ethan-1-one and/or isobornyl acetate;
- Other ingredients (e.g. amber, powdery spicy or watery): dodecahydro-3a,6,6,9a-tetramethyl-naphtho[2,1-b]furan and any of its stereoisomers, heliotropin, anisic aldehyde, eugenol, cinnamic

aldehyde, clove oil, 3-(1,3-benzodioxol-5-yl)-2-methylpropanal, 7-methyl-2H-1,5-benzodioxepin-3(4H)-one, 2,5,5-trimethyl-1,2,3,4,4a,5,6,7-octahydro-2-naphthalenol, 1-phenylvinyl acetate, 6-methyl-7-oxa-1-thia-4-azaspiro[4.4]nonane and/or 3-(3-isopropyl-1-phenyl)butanal.

A perfumery base according to the invention may not be limited to the above mentioned perfuming co-ingredients, and many other of these co-ingredients are in any case listed in reference texts such as the book by S. Arctander, Perfume and Flavor Chemicals, 1969, Montclair, New Jersey, USA, or its more recent versions, or in other works of a similar nature, as well as in the abundant patent literature in the field of perfumery. It is also understood that said co-ingredients may also be compounds known to release in a controlled manner various types of perfuming compounds also known as properfume or profragrance. Non-limiting examples of suitable properfume may include 4-(dodecylthio)-4-(2,6,6-trimethyl-2-cyclohexen-1-yl)-2-butanone, 4-(dodecylthio)-4-(2,6,6-trimethyl-1-cyclohexen-1-yl)-2-butanone, trans-3-(dodecylthio)-1-(2,6,6-trimethyl-3-cyclohexen-1-yl)-1-butanone, 2-(dodecylthio)octan-4-one, 2-phenylethyl oxo(phenyl)acetate, 3,7-dimethylocta-2,6-dien-1-yl oxo(phenyl)acetate, (Z)-hex-3-en-1-yl oxo(phenyl)acetate, 3,7-dimethyl-2,6-octadien-1-yl hexadecanoate, bis(3,7-dimethylocta-2,6-dien-1-yl) succinate (2-((2-methylundec-1-en-1-yl)oxy)ethyl)benzene, 1-methoxy-4-(3-methyl-4-phenethoxybut-3-en-1-yl)benzene, (3-methyl-4-phenethoxybut-3-en-1-yl)benzene, 1-(((Z)-hex-3-en-1-yl)oxy)-2-methylundec-1-ene, (2-((2-methylundec-1-en-1-yl)oxy)ethoxy)benzene, 2-methyl-1-(octan-3-yloxy)undec-1-ene, 1-methoxy-4-(1-phenethoxyprop-1-en-2-yl)benzene, 1-methyl-4-(1-phenethoxyprop-1-en-2-yl)benzene, 2-(1-phenethoxyprop-1-en-2-yl)naphthalene, (2-phenethoxyvinyl)benzene, 2-(1-((3,7-dimethyloct-6-en-1-yl)oxy)prop-1-en-2-yl)naphthalene, (2-((2-pentylcyclopentylidene)methoxy)ethyl)benzene, 4-allyl-2-methoxy-1-((2-methoxy-2-phenylvinyl)oxy)benzene, (2-((2-pentylcyclopentylidene)methoxy)ethyl)benzene, (2-((2-heptylcyclopentylidene)methoxy)ethyl)benzene, 1-isopropyl-4-methyl-2-((2-pentylcyclopentylidene)methoxy)benzene, 2-methoxy-1-((2-pentylcyclopentylidene)methoxy)-4-propylbenzene, 3-methoxy-4-((2-methoxy-2-phenylvinyl)oxy)benzaldehyde, 4-((2-(hexyloxy)-2-phenylvinyl)oxy)-3-methoxybenzaldehyde or a mixture thereof.

By “perfumery adjuvant”, it is meant here an ingredient capable of imparting additional added benefit such as a color, a particular light resistance, chemical stability, etc. A detailed description of the nature and type of adjuvant commonly used in perfuming composition cannot be exhaustive, but it has to be mentioned that said ingredients are well known to a person skilled in the art. One may cite as specific non-limiting examples the following: viscosity agents (e.g. surfactants, thickeners, gelling and/or rheology modifiers), stabilizing agents (e.g. preservatives, antioxidant, heat/light and or buffers or chelating agents, such as BHT), coloring agents (e.g. dyes and/or pigments), preservatives (e.g. antibacterial or antimicrobial or antifungal or anti irritant agents), abrasives, skin cooling agents, fixatives, insect repellants, ointments, vitamins and mixtures thereof.

It is understood that a person skilled in the art is perfectly able to design optimal formulations for the desired effect by admixing the above-mentioned components of a perfuming composition, simply by applying the standard knowledge of the art as well as by trial and error methodologies.

An invention's composition consisting of at least one compound of formula (I) and at least one perfumery carrier consists of a particular embodiment of the invention as well as a perfuming composition comprising at least one compound of formula (I), at least one perfumery carrier, at least one perfumery base, and optionally at least one perfumery adjuvant.

According to a particular embodiment, the compositions mentioned above, comprise more than one compound of formula (I) and enable the perfumer to prepare accords or perfumes possessing the odor tonality of various compounds of the invention, creating thus new building block for creation purposes.

For the sake of clarity, it is also understood that any mixture resulting directly from a chemical synthesis, e.g. a reaction medium without an adequate purification, in which the compound of the invention would be involved as a starting, intermediate or end-product could not be considered as a perfuming composition according to the invention as far as said mixture does not provide the inventive compound in a suitable form for perfumery. Thus, unpurified reaction mixtures are generally excluded from the present invention unless otherwise specified.

The invention's compound can also be advantageously used in all the fields of modern perfumery, i.e. fine or functional perfumery, to positively impart or modify the odor of a consumer product into which said compound (I) is added. Consequently, another object of the present invention consists of a perfumed consumer product comprising, as a perfuming ingredient, at least one compound of formula (I), as defined above.

The invention's compound can be added as such or as part of an invention's perfuming composition.

For the sake of clarity, "perfumed consumer product" is meant to designate a consumer product which delivers at least a pleasant perfuming effect to the surface or space to which it is applied (e.g. skin, hair, textile, or home surface). In other words, a perfumed consumer product according to the invention is a perfumed consumer product which comprises a functional formulation, as well as optionally additional benefit agents, corresponding to the desired consumer product, and an olfactive effective amount of at least one invention's compound. For the sake of clarity, said perfumed consumer product is a non-edible product.

The nature and type of the constituents of the perfumed consumer product do not warrant a more detailed description here, which in any case would not be exhaustive, the skilled person being able to select them on the basis of his general knowledge and according to the nature and the desired effect of said product.

Non-limiting examples of suitable perfumed consumer products include a perfume, such as a fine perfume, a splash or eau de parfum, a cologne or a shave or after-shave lotion; a fabric care product, such as a liquid or solid detergent, a fabric softener, a liquid or solid scent booster, a fabric refresher, an ironing water, a paper, a bleach, a carpet cleaner, a curtain–care product; a body-care product, such as a hair care product (e.g. a shampoo, a coloring preparation or a hair spray, a color-care product, a hair shaping product, a dental care product), a disinfectant, an intimate care product; a cosmetic preparation (e.g. a skin cream or lotion, a vanishing cream or a deodorant or antiperspirant (e.g. a spray or roll on), a hair remover, a tanning or sun or after sun product, a nail product, a skin cleansing, a makeup); or a skin-care product (e.g. a soap, a shower or bath mousse, oil or gel, or a hygiene product or a foot/hand care products); an air care product, such as an air freshener or a “ready to use” powdered air freshener which can be used in the home space (rooms, refrigerators, cupboards, shoes or car) and/or in a public space (halls, hotels, malls, etc..); or a home care product, such as a mold remover, a furnisher care product, a wipe, a dish detergent or a hard-surface (e.g. a floor, bath, sanitary or a window-cleaning) detergent; a leather care product; a car care product, such as a polish, a wax or a plastic cleaner.

Some of the above-mentioned perfumed consumer products may represent an aggressive medium for the invention’s compounds, so that it may be necessary to protect the latter from premature decomposition, for example by encapsulation or by chemically binding it to another chemical which is suitable to release the invention’s ingredient upon a suitable external stimulus, such as an enzyme, light, heat or a change of pH.

The proportions in which the compounds according to the invention can be incorporated into the various aforementioned products or compositions vary within a wide range of values. These values are dependent on the nature of the article to be perfumed and on the desired organoleptic effect as well as on the nature of the co-ingredients in a given base when the compounds according to the invention are mixed with perfuming co-ingredients, solvents or additives commonly used in the art.

For example, in the case of perfuming compositions, typical concentrations are in the order of 0.001 % to 10 % by weight, or even more, of the compounds of the invention based on the weight of the composition into which they are incorporated. In the case of perfumed consumer product, typical concentrations are in the order of 0.01 % to 1 % by weight, or even more, of the compounds of the invention based on the weight of the consumer product into which they are incorporated.

Additionally, the intermediates compounds produced in any of the embodiments described herein can be converted to derivatives such as, but not limited to hydrocarbons, alcohols, diols, triols, acetals, ketals, aldehydes, acids, ethers, amides, ketones, lactones, epoxides, acetates, glycosides and/or an esters. These derivatives can be obtained by a chemical method such as, but not limited to oxidation, reduction, alkylation, acylation and/or rearrangement. Alternatively, the derivatives can be obtained

using a biochemical method by contacting the terpene compound with an enzyme such as, but not limited to an oxidoreductase, a monooxygenase, a dioxygenase, a transferase. The biochemical conversion can be performed in-vitro using isolated enzymes, enzymes from lysed cells or in vivo using whole cells. The conversion can be a cyclization reaction realized by chemical or biochemical method. The derivatives can be used as perfumery, flavor or aroma ingredients.

The numerous possible variations that will become immediately evident to a person skilled in the art after having considered the disclosure provided herein also fall within the scope of the invention.

The invention will now be described in further details by way of the following Examples. Said Examples are illustrative only and are not intended to limit the scope of the embodiments as described herein.

EXAMPLES

Materials and Methods

Unless otherwise stated, all chemical and biochemical materials and microorganisms or cells employed herein are commercially available products.

Unless otherwise specified, recombinant proteins are cloned and expressed by standard methods, such as, for example, as described by Sambrook, J., Fritsch, E.F. and Maniatis, T., Molecular cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Engineering of a recombinant *E. coli* strain for production of terpenoid precursors by chromosomal integration of the genes encoding mevalonate pathway enzymes.

An *E. coli* strain was engineered to produce farnesyl-pyrophosphate (FPP) by chromosomal integration of recombinant genes encoding mevalonate pathway enzymes.

An upper pathway operon (operon 1 from acetyl-CoA to mevalonate) was designed consisting of the *atoB* gene from *E. coli* encoding an acetoacetyl-CoA thiolase, and the *mvaA* and *mvaS* genes from *Staphylococcus aureus* encoding a HMG-CoA synthase and a HMG-CoA reductase, respectively.

As a lower mevalonate pathway operon (operon 2 from mevalonate to IPP/DMAPP), a natural operon from the gram-negative bacteria *Streptococcus pneumoniae* was selected, encoding a mevalonate kinase (*mvaK1*), a phosphomevalonate kinase (*mvaK2*), a phosphomevalonate decarboxylase (*mvaD*), and an isopentenyl diphosphate isomerase (*fni*).

A codon optimized *Saccharomyces cerevisiae* FPP synthase encoding gene (ERG20) was introduced at the 3'-end of the upper pathway operon to convert isopentenyl-diphosphate (IPP) and dimethylallyl-diphosphate (DMAPP) into FPP.

The above-described operons were synthesized by DNA 2.0 and integrated into the *araA* gene of the *Escherichia coli* strain BL21(DE3). The heterologous pathway was introduced in two separate recombination steps using the CRISPR/Cas9 genome engineering system. The first operon (lower pathway; operon 2) to be integrated carries a spectinomycin (Spec) marker which was used to screen for Spec resistant candidate integrants. The second operon was designed to displace the Spec marker of the previously integrated operon and was accordingly screened for Spec candidate integrants following the second recombination event. Guide RNA expression vectors targeting the *araA* gene were designed and synthetized by DNA 2.0. PCR was used to verify operon integration by designing PCR primers to amplify across the *araA* gene integration target and across recombination junctions of

integrants. One clone yielding correct PCR results was then fully sequenced and archived as strain DP1205.

Medium composition for *E. coli* cultivation.

The mineral AM medium used in the shake flask and lab-scale fermentation experiments consists of : KH2PO4 4.2 g/L; K2HPO4 • 3H2O 15.7 g/L; (NH4)2SO4 2.0 g/L; Citric acid 1.7 g/L; EDTA 8.4 mg/L; glycerol 30 g/L; yeast extract 5 g/L were dissolved in diH2O; dodecane at 10% (v/v) and sterilized at 121°C for 30 min. Concentrated stocks of MgSO4 • 7H2O 1M, 5 mL/L; vitamin (thiamine • HCl 4.5 g/L), 1 mL/L; and batch trace metal solution; 10 mL/L were added aseptically to the medium and the pH was adjusted at 7 with NaOH 5M. Batch trace metal solution in (per L 1M HCL): CoCl2 • 6H2O 0.25 g/L; MnCl2 • 4H2O 1.5 g/L; CuCl2 • 2H20 0.15 g/L; H3BO3 0.3 g/L; Na2MoO4 • 2H2O 0.25 g/L; Zn(CHCOO)2 •2H2O 1.3 g/L; Fe(III)citrate 10 g/L. For the fermentations carried out in feed-batch mode with the AM medium, 20 L glycerol feed solution containing 700 g/L glycerol, 12 g/L MgSO4 • 7H2O; 13 mg/L EDTA and 10 mL/L feed trace solution was prepared. The feed trace metal solution was prepared by dissolving 0.4 g CoCl2 • 6H2O, 2.35 g MnCl2 •4H2O, 0.25 g CuCl2 • 2H20, 0.5 g H3BO3; 0.4 g Na2MoO4 • 2H2O; 1.6 g Zn(CHCOO)2 •2H2O and 10 g Fe(III)citrate • H2O in 1L HCl 1M.

Cultivation of engineered bacteria cells under conditions enabling production of terpene compounds.

The DP1205 *E. coli* cells engineered to produce increased levels of the terpenoid precursor farnesyl diphosphate (FPP) (as described in WO2018/114839) were transformed with one or two expression plasmids carrying genes encoding for enzymes from a homofarnesol biosynthetic pathway enzymes and/or for terpene cyclases. The transformed cells were cultured with the appropriate antibiotics (kanamycin (50 µg/mL) and/or carbenicillin (50 µg/mL) and/or chloramphenicol (34 µg/mL)) and/or streptomycin (50 µg/mL)) on LB-agarose plates. Single colonies were used to inoculate 5 mL liquid LB medium supplemented with the same antibiotics, 4 g/L glucose and 10% (v/v) n-dodecane. The next day, 2 mL of AM medium supplemented with the same antibiotics and 10% (v/v) n-dodecane were inoculated with 0.2 mL of the overnight culture. The cultures were incubated at 37°C until an optical density of 3 was reached. The expression of the recombinant proteins was then induced by addition of 0.1 mM IPTG and the cultures were incubated for 72 h at 25°C.

The cultures were then extracted with one volume of methyl tert-butyl ether (MTBE) and the composition of the organic phase was analyzed by GC-MS as described below. For quantification, an internal standard (α-longipinene (Sigma-Aldrich, Missouri, USA)) was added to the extract prior to GC-MS analysis and concentrations of the components were estimated based on comparison of the peak areas.

GC-MS analysis methods.

Samples were analyzed using an Agilent 6890N GC system coupled with a 5975B series Mass Selective Detector (MSD) and equipped with a split/splitless injector (Agilent Technologies, CA) and a CombiPAL autosampler (PAL LSI 85 autosampler, Agilent Technologies, CA) injection system. The GC inlet temperature was set to 240 °C and 1.0 µL of sample was injected in split mode with a ratio of 25:1 (23.304 PSI) and analyzed on a DB-5ms capillary column (30 m x 0.25 mm inner diameter x 0.25 µm film thickness; Agilent J&W) using helium as a carrier gas at a constant flow of 1.2 mL/min. The initial temperature of the oven was set at 80 °C (hold 1 min) and was programmed to 300 °C (10 °C/min) and then to 300 °C (30°C/min; hold 1min).

General methods for *Saccharomyces cerevisiae*, genetic modification, cultivation and compound analysis.

A *Saccharomyces cerevisiae* strain producing increased levels of the terpenoid precursor farnesyl diphosphate (FPP) (as described in WO2018/114839) was used as base strain for the expression of homofarnesol biosynthetic pathway genes and terpene cyclases. In short, the strain contains all the endogenous mevalonate pathway genes integrated in its genome under the control of the native *GAL1* or *GAL10* promoters. Further increase of the Farnesyl diphosphate precursor pool in this strain was achieved through downregulation of the squalene synthase gene (*ERG9*) via replacement of its native promoter.

All genes (synthesized by ATUM, California, USA or Twist Bioscience, California, USA), together with relevant regulatory elements (eg. promoters and terminators) were introduced in the base strain through genomic integration or through plasmids constructed *in vivo* using the yeast homologous recombination machinery (Kuijpers et al., Microb Cell Fact., 2013, 12:47). All yeast transformations were performed with the lithium acetate method (Gietz and Woods, Methods Enzymol., 2002, 350:87–96).
Successfully transformed yeast colonies were grown for three days at 30 °C on media containing 6.7 g/L of Yeast Nitrogen Base without amino acids (BD Difco, New Jersey, USA), the appropriate antibiotic or nutrients according to the marker gene used, 20 g/L glucose and 20 g/L agar.

For metabolite production and analysis, single colonies of the modified yeast strains were inoculated in 2 mL of culture medium (Westfall et al., Proc Natl Acad Sci USA, 2012, 109:E111-118) with addition of 2% galactose and 10% (v/v) n-dodecane (Sigma-Aldrich, Missouri, USA). The cultures were incubated for 3 days at 30°C and shaking at 200 rpm. After the incubation period, the cultures were extracted with two volumes of MTBE (supplemented with α-longipinene standard for quantification as described above) and the composition of the organic phase was analyzed by GC-MS using an Agilent 7890A GC system coupled with a 5975C series Mass Selective Detector (MSD) and equipped with a split/splitless injector and a GC Injector 80 injection system (Agilent Technologies, CA). The GC inlet temperature was set to 260°C and 1.0 µl of sample was injected in splitless mode and analyzed on a HP-5 GC

column (30 m x 0.25 mm x 0.25 µm; Agilent J&W) using helium as a carrier gas at a constant flow of 1.2 mL/min. The initial temperature of the oven was set at 100 °C and was programmed to 300 °C (10 °C/min).

Example 1. *In vivo* production of (3E,7E)-homofarnesol and biosynthetic intermediates in engineered bacterial cells expressing a GGPP synthase, a phosphatase, an alcohol dehydrogenase, a BVMO, an enal-cleaving enzyme and an esterase.

The reaction scheme in Figure 3 depicts a biochemical pathway that can be used to produce (3E,7E)-homofarnesol *in vivo*. The universal isoprenoid precursors isopentenyl-diphosphate (IPP) and dimethylallyl-diphosphate (DMAPP) are condensed to form (2E,6E,10E)-geranylgeranyl-diphosphate (GGPP). This reaction can be catalyzed by a GGPP synthase or using a combination of a farnesyl-diphosphate synthase (FPP synthase) and a GGPP synthase. GGPP can be converted to (2E,6E,10E)-geranylgeraniol by a terpene synthase or by a phosphatase as for example described in WO2020011883A1. (2E,6E,10E)-geranylgeraniol undergoes then several enzymatic degradations steps. In the proposed pathway, (2E,6E,10E)-geranylgeraniol is first oxidized by an alcohol dehydrogenase (ADH) and cleaved to yield (5E,9E)-farnesylacetone. The following cleavage reaction can be catalyzed by an enal-cleaving enzyme (ENase) such as a protein containing a GXWXG (SEQ ID NO: 263) and DUF4334 domain as described in WO2021005097. (5E,9E)-farnesylacetone is further converted by a Baeyer-Villiger monooxygenase (BVMO) to (3E,7E)-homofarnesyl acetate. In the last step, the ester is hydrolyzed by an esterase to finally form (3E,7E)-homofarnesol.

To validate the (3E,7E)-homofarnesol pathway, *E. coli* cells were engineered to express the necessary enzymes. A plasmid was assembled to contain two operons.

The first operon was designed to contain three cDNAs encoding for:
- PsAerADH (SEQ ID NO: 11), an alcohol dehydrogenase from *Pseudomonas aeruginosa* (GeneBank accession number: WP_079868259.1) having the ability to oxidize (2E,6E,10E)-geranylgeraniol to (2E,6E,10E)-geranylgeranial,
- SCH24-BVMO1 (SEQ ID NO: 23), a Baeyer-Villiger Monooxygenase (BVMO) from *Filobasidium magnum* and described in WO2021005097, which allow the oxidation of (5E,9E)-farnesylacetone to (3E,7E)-homofarnesyl acetate, and
- SCH24-EST1 (SEQ ID NO: 27), an esterase from *Filobasidium magnum* and described in WO2021005097, which hydrolyses (3E,7E)-homofarnesylacetate to (3E,7E)-homofarnesol and acetic acid.

The second operon was constructed to contain 4 cDNAs encoding for:
- SCH94-03944 (SEQ ID NO: 22), a protein containing an enal-cleaving enzyme from *Rhodococcus erythropolis* and described in WO2021005097, which cleaves (2E,6E,10E)-geranylgeranial into (5E,9E)-farnesylacetone and acetaldehyde,

- CcrGGPPS2-del57, a truncated version of the (2E,6E,10E)-geranylgeranyl diphosphate synthase from *Cistus creticus* (GeneBank: AAM21639.1) (SEQ ID NO: 1), and
-Two copies of a cDNAs encoding for PgpB (SEQ ID NO: 3), a phosphatase from *Escherichia coli* (GeneBank: WP_089622241.1), which converts (2E,6E,10E)-geranylgeranyl diphosphate into (2E,6E,10E)-geranylgeraniol by cleaving of the diphosphate group.

The cDNAs encoding for PsAerADH, SCH24-BVMO1, SCH24-EST1, SCH94-03944, CcrGGPPS2-del57 and PgpB were codon optimized for *Escherichia coli (*SEQ ID NOs:102, 115, 120, 113, 90 and 93) and an RBS sequence (AAGGAGGTAAAAAA) (SEQ ID NO: 264) was placed upstream of each of these cDNAs. The first operon containing the cDNAs for PsAerADH, SCH24-BVMO1 and SCH24-EST1 were under the control of a T5 promotor and rrnB T1 terminator. The second operon containing the cDNA for SCH94-03944, CcrGGPPS2-del57 and the two cDNAs for PgpB was under the control of a T5 promotor and rrnB terminator. Both operons were synthesized and cloned into a vector backbone containing a pUC origin of replication, a kanamycin resistance gene and a LacI gene resulting in the vector pHFOL-5.

The farnesyl-diphosphate (FPP) producing *E. coli* strain DP1205, described in WO2021005097, was transformed with the vector pHFOL-5 described above. When cultivated in the conditions enabling production of terpene compounds, the resulting cells were capable of producing (3E,7E)-homofarnesol (Figure 4). Under the conditions described in the “materials and methods” section, 69 mg/L of (3E,7E)-homofarnesol were produced in the tube assay in the culture media.

The product profile of the cells (Figure 4) also shows the accumulation of several metabolic intermediates, such as (5E,9E)-farnesylacetone and (2E,6E,10E)-geranylgeraniol. The different enzymatic steps of the pathway can be optimized to limit the accumulation of intermediates and increase the concentration of the final product.

The following examples show methods of identifying suitable enzymes for each of the enzymatic step in the pathway to increase (3E,7E)-homofarnesol production and limit the accumulation of the metabolic intermediates.

Example 2. Screening of Baeyer-Villiger monooxygenases to improve the enzymatic conversion of (5E,9E)-farnesylacetone to (3E,7E)-homofarnesylacetate and the *in vivo* production of (3E,7E)-homofarnesol.

Example 1 shows that for the *in vivo* production of (3E,7E)-homofarnesol, (5E,9E)-farnesylacetone can also be detected due to the insufficient activity of the Baeyer-Villiger monooxygenase (SCH24-BVMO1 (SEQ ID NO: 23)) in this strain.

In this example, an *in vivo* screening of different BVMOs was conducted to identify enzyme candidates with higher efficiency compared to SCH24-BVMO1 (SEQ ID NO: 23). For the screening, a modified version of the vector pHFOL-5 (described in example 1) was created by removing SCH24-BVMO1. The new vector was called pF-Facetone-7. The *E. coli* strain DP1205 was transformed with the vector pF-Facetone-7, providing a strain that was capable of producing (5E,9E)-farnesylacetone when cultivated under conditions enabling the production of terpene compounds. Up to 410 mg/L of (5E,9E)-farnesylacetone were produced in the culture media in the tube assay (Figure 5A). When the cells are further transformed with a vector expressing an active BVMO, (5E,9E)-farnesylacetone is converted to (3E,7E)-homofarnesyl acetate, which itself is converted to (3E,7E)-homofanesol by the SCH24-EST1 esterase (Figure 5B).

In the next step, codon optimized cDNAs encoding for BVMOs were designed and cloned in the pJ423 expression plasmid (ATUM, Newark, California). The DP1205 *E. coli* cells were co-transformed with one of these plasmids and plasmid pF-Facetone-7. The activity of the BVMOs was determined by quantifying the production of (3E,7E)-homofarnesol for each BVMO tested and compared to SCH24-BVMO1 (SEQ ID NO: 23).

The table below (Table 1) shows the relative activity of (5E,9E)-farnesylacetone conversion of some BVMOs identified in this screening.

Enzyme Name	**Organisme Name**	**SEQ ID NO**	**Relative production of (3E,7E)-homofarnesol**
SCH24-BVMO1 (WO2021005097)	*Filobasidium magnum*	23	100%
SCH46-BVMO1 (WO2021005097)	Bensingtonia ciliata	24	92%
AraBVMO1	*Acinetobacter radioresistens*	25	128%
AflavBVMO1	*Aspergillus flavus*	26	159%

Table 1: Selected BVMOs and relative activity of conversion of (5E,9E)-farnesylacetone to (3E,7E)-homofarnesyl acetate, which was further converted to (3E,7E)-homofarnesol.

AraBVMO1 and AflavBVMO1 were found to produce significantly higher amounts of (3E,7E)-homofarnesol than SCH24-BVMO1 (SEQ ID NO: 23). AflavBVMO1 (SEQ ID NO: 26) increases the production of (3E,7E)-homofarnesol by 59 % compared to the reference BVMO.

Example 3. *In vivo* screening of alcohol dehydrogenases to improve the efficiency of the enzymatic conversion of (2E,6E,10E)-geranylgeraniol to (2E,6E,10E)-geranylgeranial and the *in vivo* production of (3E,7E)-homofarnesol.

In this Example, alcohol dehydrogenases (ADHs) were tested *in vivo* for their efficiency to oxidize (2E,6E,10E)-geranylgeraniol to (2E,6E,10E)-geranylgeranial. Alcohol dehydrogenases catalyze the reversible oxidation of alcohols.

To avoid the reverse alcohol-dehydrogenase reaction in this *in vivo* screening assay, the enzyme enal-cleaving enzyme SCH94-03944 described in Example 1 was co-expressed in the *E. coli* cells to enzymatically convert the (2E,6E,10E)-geranylgeranial form to (5E,9E)-farnesylacetone. The catalytic efficiency of the ADHs tested was then correlated to the amount of (2E,6E,10E)-geranylgeraniol being converted to (5E,9E)-farnesylacetone. The ADH candidates were codon optimized and cloned into the pJ423 expression plasmid (ATUM, Newark, California). The DP1205 *E. coli* cells were co-transformed with one of these plasmids and with the plasmid pJ401-SCH94-3944-PgpB-CcrGGPPS, which contained the necessary genes to produce (5E,9E)-farnesylacetone as described below, except the ADH encoding gene.

Thus, the plasmid pJ401-SCH94-3944-PgpB-CcrGGPPS2-del57, contained an operon harboring 3 cDNAs that encoded for:

- The enal-cleaving enzyme SCH94-03944 (SEQ ID NO: 22), a protein containing a GXWXG (SEQ ID NO: 263) and DUF4334 domain described in WO2021005097, which cleaves (2E,6E,10E)-geranylgeranial into (5E,9E)-farnesylacetone and acetaldehyde
- PgpB (SEQ ID NO: 3), a phosphatase from *Escherichia coli* (GeneBank: WP_089622241.1), which converts (2E,6E,10E)-geranylgeranyl diphosphate into (2E,6E,10E)-geranylgeraniol by cleaving off diphosphate, and
- CcrGGPPS2-del57 (SEQ ID NO: 1), a truncated version of the (2E,6E,10E)-geranylgeranyl diphosphate synthase from *Cistus criticus* (GeneBank: AAM21639.1).

The cDNAs encoding for SCH94-03944, PgpB and CcrGGPPS2-del57 were codon optimized for the expression in *Escherichia coli* (SEQ ID NOs: 113, 93 and 90). An operon was designed containing successively the three cDNAs and an RBS sequence (AAGGAGGTAAAAAA) (SEQ ID NO: 264) placed upstream of each cDNA. The operon was synthesized and cloned into the pJ401 expression plasmid (ATUM, Newark, California).

The resulting strains (DP1205 containing the plasmid pJ401-SCH94-3944-PgpB-CcrGGPPS2-del57 and a plasmid pJ423 harboring an alcohol dehydrogenase candidate) were cultivated under conditions enabling the production of terpenes. The amounts of (2E,6E,10E)-geranylgeraniol and (5E,9E)-farnesylacetone were measured and the conversion rate was calculated for each alcohol dehydrogenase tested.

Enzyme Name	Organism Name	Seq ID NO	Conversion from (2E,6E,10E)-geranylgeraniol to (5E,9E)-farnesylacetone
ThTerpADH1	*Thauera terpenica*	12	87.5%
VoADH1	*Valeriana officinalis*	13	12.1%
SCH80-05240	*Rhodococcus sp.*	14	49.4%
PsAerADH	*Pseudomonas aeruginosa*	11	78.6%
Ppseudo-alkJ	*Pseudomonas fluorescens*	15	82.1%
CymB	*Pseudomonas putida*	17	80.9%
CdGeoA	*Castellaniella defragrans*	18	42.8%
AroAroADH	*Aromatoleum aromaticum*	19	59.3%
AzTolADH1	*Azoarcus toluclasticus*	20	63.4%

Table 2: ADHs selected *in vivo* for their activity on (2E,6E,10E)-geranylgeraniol to produce (2E,6E,10E)-geranylgeranial, which was further converted to (5E,9E)-farnesylacetone.

The results are shown in Table 2. The highest conversion rates from (2E,6E,10E)-geranylgeraniol were detected in the culture medium in the tube assay for the alcohol dehydrogenases ThTerpADH1 (SEQ ID NO: 12), Ppseudo-alkJ (SEQ ID NO: 15) and CymB (SEQ ID NO: 17). The highest amount of (5E,9E)-farnesylacetone was 172 mg/L and was produced with the ADH Ppseudo-alkJ (SEQ ID NO: 15).

Example 4. *In vivo* testing of esterases to catalyze the hydrolysis of (3E,7E)-homofarnesyl acetate to (3E,7E)-homofarnesol and acetic acid.

In this example, different esterases were tested *in vivo* for their ability to catalyze the hydrolysis of (3E,7E)-homofarnesyl acetate to (3E,7E)-homofarnesol and acetic acid.

The esterases were tested in a two-plasmid system in *E. coli* DP1205. The first plasmid consisted of an operon containing 3 cDNAs encoding for:

- The enal-cleaving enzyme SCH94-03944 (SEQ ID NO: 22), a protein containing a GXWXG (SEQ ID NO: 263) and DUF4334 domain described in WO2021/005097, which cleaves (2E,6E,10E)-geranylgeranial into (5E,9E)-farnesylacetone and acetaldehyde,
- PgpB (SEQ ID NO: 3), a phosphatase from *Escherichia coli* (GeneBank: WP_089622241.1), which converts (2E,6E,10E)-geranylgeranyl diphosphate into (2E,6E,10E)-Geranylgeraniol by cleaving off diphosphate, and
- CcrGGPPS2-del57 (SEQ ID NO: 1), a truncated version of the (2E,6E,10E)-geranylgeranyl diphosphate synthase from *Cistus criticus* (GeneBank: AAM21639.1).

The cDNAs encoding for SCH94-03944, PgpB and CcrGGPPS2-del57 were codon optimized for the expression in *E. coli* (SEQ ID NOs: 113, 93 and 90) and contained an upstream placed RBS sequence (AAGGAGGTAAAAAA) (SEQ ID NO: 264) The operon was synthesized and cloned into the pJ401 expression plasmid (ATUM, Newark, California).

The second operon consisted of 3 cDNAs encoding for:

- PsAerADH, an alcohol dehydrogenase from *Pseudomonas aeruginosa* (GeneBank: WP_079868259.1) (SEQ ID NO: 11) having the ability to oxidize (2E,6E,10E)-Geranylgeraniol to (2E,6E,10E)-geranylgeranial,
- SCH24-BVMO1 (SEQ ID NO: 23), a Baeyer-Villiger Monooxygenase (BVMO) described in WO2021005097, which allows the oxidation of (5E,9E)-farnesylacetone to (3E,7E)-homofarnesyl acetate, and
- an esterase gene candidate.

The cDNAs encoding for PsAerADH, SCH24-BVMO1 and an esterase gene candidate were codon optimized for the expression in *Escherichia coli* (SEQ ID NOs: 102 and 115) and contained an upstream RBS sequence (AAGGAGGTAAAAAA) (SEQ ID NO: 264). The operons were synthesized and cloned into the pJ424 expression plasmid (ATUM, Newark, California).

The DP1205 *E. coli* cells were co-transformed with the plasmids pJ401-SCH94-03944-PgpB-CcrGGPPS2-del57 and pJ424-PsAerADH-SCH24-BVMO1-Esterase. In the resulting strains, the esterases SCH24-EST1 (SEQ ID NO: 27) from *Filobasidium magnum* and SCH23-EST1 (SEQ ID NO: 28) from *Hyphozyma roseonigra* were found to be able to convert (3E,7E)-homofarnesyl acetate to (3E,7E)-homofarnesol.

Example 5. *In vivo* screening of different phosphatases to catalyze the hydrolysis from (2E,6E,10E)-geranylgeranyl-disphosphate to (2E,6E,10E)-geranylgeraniol.

In this example, phosphatases from different protein families were tested *in vivo* for their ability to catalyze the hydrolysis of (2E,6E,10E)-geranylgeranyl diphosphate to (2E,6E,10E)-geranylgeraniol and diphosphate. The phosphatases were tested in DP1205 *E. coli* cells, which contained a geranylgeranyl-diphosphate synthase in an expression vector. Codon optimized versions of the DNA fragments coding for the phosphatase genes, each cloned into a second expression vector, were then introduced into this strain by transformation. Under terpene production enabling conditions, we identified 8 phosphatases showing activity on (2E,6E,10E)-geranylgeranyl-diphosphate and production of (2E,6E,10E)-geranylgeraniol. The results are shown in Table 3. The phosphatases PgpB (SEQ ID NO: 3), PeSubTPP1 (SEQ ID NO: 7) and TalVeTPP (SEQ ID NO: 8) showed the highest activity. PgpB (SEQ ID NO: 3) was found to be the most efficient enzyme for the production of (2E,6E,10E)-geranylgeraniol.

Enzyme Name	Organism Name	Protein family	SEQ ID NO	Production of (2E,6E,10E)-geranylgeraniol
PgpB	*Escherichia coli K-12*	PAP2	3	+++
NudB	*Escherichia coli*	Nudix-hydrolase	4	+
LPP1	*Saccharomyces cerevisiae*	PAP2	5	+
NUDX1	*Rosa hybrida*	Nudix-hydrolase	6	+
PeSubTPP1	*Penicillium subrubescens*	HAD-like hydrolase	7	++
TalVeTPP	*Talaromyces verruculosus*	HAD-like hydrolase	8	++
AstI	*Aspergillus oryzae RIB40*	HAD-like hydrolase	9	+
AstK	*Aspergillus oryzae RIB40*	HAD-like hydrolase	10	+

Table 3: Phosphatases selected *in vivo* for their activity to convert (2E,6E,10E)-geranylgeraniol diphosphate to (2E,6E,10E)-geranylgeraniol. The phosphatases were ranked and divided into 4 groups (- No activity; + Minor activity, ++ High activity; +++ very high activity) based on the productivity.

Example 6. Production of (3E,7E)-homofarnesol in engineered fungal cells.

The (3E,7E)-homofarnesol biosynthetic pathway (Figure 3) genes were introduced in a *Saccharomyces cerevisiae* strain producing high levels of the terpenoid precursor farnesyl diphosphate (FPP) in a two-step process. First, single copies of the genes coding for the alcohol dehydrogenase SCH23-ADH1 (SEQ ID: 21) (WO2021005097), the esterase SCH23-EST1 (SEQ ID: 28) (WO2021005097) and the Baeyer-Villiger monooxygenase AflavBVMO (SEQ ID: 26) were integrated in the genome of the yeast strain. The resulting strain, YST403, was subsequently used for the *in vivo* construction of a 2-micron plasmid containing the geranylgeranyl diphosphate synthase CarG (SEQ ID: 2) (from *Blakeslea trispora*, NCBI accession JQ289995.1), the phosphatase PgpB (SEQ ID: 3) (from *Escherichia coli*, NCBI accession WP_089622241.1) and the enal-cleaving enzyme SCH94-03944 (SEQ ID: 22). All genes coding for the different enzymes were codon optimized for their expression in *S. cerevisiae* (SEQ ID NO: 112, 122, 119, 92, 94 and 114) and controlled by galactose inducible promoters. After cultivation, it was confirmed that the final strain, termed YST403_HFOL, produced (3E,7E)-homofarnesol. The product profile (Figure 6) also shows the accumulation of several metabolic intermediate such as of (5E,9E)-farnesylacetone (44 mg/L) and (2E,6E,10E)-geranylgeraniol (52 mg/L). The pathway can be optimized by increasing the enzymatic activity of each enzymatic step in order to limit the accumulation of intermediates and thereby to increase the concentration of the final product.

Example 7. *In vivo* production of compound of formula (Ia) and biosynthetic intermediates in bacterial cells engineered to produce (3E,7E)-homofarnesol and expressing a wild type or mutant variant squalene cyclase.

In this example, compound of formula (Ia) is being produced *in vivo* in bacterial cells by using the artificial biochemical pathway shown in figure 7. This pathway encompasses the production of (3E,7E)-homofarnesol from a simple carbon source and its conversion to compound of formula (Ia) by the squalene cyclase A0A5P9HJ69 (SEQ ID NO: 42) and its mutant variant A0A5P9HJ69_V1 (SEQ ID NO: 43). A0A5P9HJ69 is annotated in the uniprot protein database as tetraprenyl-β-curcumene-sporulenol, a class of enzymes that has been shown to also cyclize squalene (Sato, T., et al. (2011). Journal of the American Chemical Society 133(44): 17540-17543).

The mutant variant A0A5P9HJ69_V1 (SEQ ID NO: 43) contained the two-point mutations E429A and G593M. These mutations are based on the previously described point mutations AAcSHC F437A and AAcSHC G600M in the squalene cyclase from *Alicyclobacillus acidocaldarius* (AAcSHC). The point mutation AAcSHC F437A has been reported to increase the catalytic efficiency of the cyclisation of squalene at lower temperatures (T. Sato, M. Kouda, T. Hoshino, Biosci. Biotechnol. Biochem. 2004, 68, 728–738), while the point mutation G600M in AAcSHC was shown to significantly increase the conversion of (3E,7E)-homofarnesol to compound of formula (Ia) (*Angewandte Chemie International Edition* (2023): e202301607).

The amino acid positions for the point mutations in A0A5P9HJ69 (SEQ ID NO: 42) were identified by protein sequence alignment of A0A5P9HJ69 with AAcSHC (SEQ ID NO: 82). The amino acid position relative to AAcSHC F437 and AAcSHC G600 were selected and changed to alanine and methionine respectively.

The DNAs encoding the wild type A0A5P9HJ69 (SEQ ID NO: 42) and its mutant variant A0A5P9HJ69_V1 (SEQ ID NO: 43) were codon optimized (SEQ ID NO: 143 and 144) and cloned into the expression plasmid pD424 (ATUM, Newark, California). DP1205 *E. coli* cells were then co-transformed with one of these plasmids and with the vector pHFOL-11. The vector pHFOL-11 was based on the vector pHFOL-5 as described in Example 1, but contained a DNA encoding the Baeyer–Villiger monooxygenase AflavBVMO1 (SEQ ID NO: 26 encoded by 118) instead of the DNA encoding SCH24-BVMO1 (SEQ ID NO: 23 encoded by 115). As a control, only the vector pHFOL-11 was transformed into DP1205 *E. coli* cells.

The transformed *E. coli* cells were cultivated under conditions enabling the production of terpene compounds. After the cultivation, the terpenes in the tube assay were analyzed by GC-MS/FID. As expected, the control strain (DP1205 pHFOL-11) produced (3E,7E)-homofarnesol and it did not produce the compound of formula (Ia). When pHFOL-11 was co-expressed either with A0A5P9HJ69 (SEQ ID NO: 42) or A0A5P9HJ69_V1 (SEQ ID NO: 43), the compound of formula (Ia) was detected (Figure 8).

As shown in Figure 8, the mutant variant A0A5P9HJ69_V1 (SEQ ID NO: 43) produced significantly more compound of formula (Ia) when compared to the wild type A0A5P9HJ69 (SEQ ID NO: 42). When A0A5P9HJ69_V1 (SEQ ID NO: 43) was co-expressed with pHFOL-11, 44 mg/L of compound of formula (Ia) and 122 mg/L of the metabolic precursor (3E,7E)-homofarnesol could be quantified using the FID detector. In contrast, when the wild type enzyme A0A5P9HJ69 was co-expressed with pHFOL-11, 5 mg/L of compound of formula (Ia) and 143 mg/L of (3E,7E)-homofarnesol were quantified by FID detector.

Example 8. *In vivo* production of compound of formula (Ia) and biosynthetic intermediates in bacterial cells engineered to produce (3E,7E)-homofarnesol and expressing different wild type or mutant squalene cyclases.

In this example, wild type squalene cyclases and their corresponding mutant variants were screened *in vivo* in bacterial cells that produced (3E,7E)-homofarnesol (as described in Example 7) for the production of compound of formula (Ia). All enzymes were selected from the NCBI and UniProt protein database and were part of the squalene cyclase or tetraprenyl-β-curcumene—sporulenol cyclase protein family.

The mutant cyclase variants were based on the mutations AAcSHC F437A and AAcSHC G600M in the squalene cyclase from *Alicyclobacillus acidocaldarius* (AAcSHC) (SEQ ID NO: 82) and were designed as described in Example 7.

The DNAs encoding the squalene cyclase candidates were ordered codon optimized and cloned into the expression plasmid pD424 (ATUM, Newark, California) or an expression vector containing the clodf13 origin of replication, the streptomycin resistance (SmR), T5 promoter, lambda T0 terminator and lac operator to control transcription as well as the Lactose operon repressor (lacI). The vectors containing the cyclase candidates were transformed into DP1205 *E. coli* cells containing the vector pF-HOL11 (as described in Example 7). The resulting strains were cultivated under conditions enabling the production of terpenoid compounds and were subsequently analyzed. Figure 9 shows the titers of compound of formula (Ia) obtained with different squalene cyclases (SEQ ID NOs: 31, 33-36, 40-43, 46, 48 and 49 encoded by SEQ ID NO: 125, 128, 130, 132, 134, 139, 141, 143, 144, 148, 150 and 152). The highest titers were observed for the mutant tetraprenyl-β-curcumene—sporulenol cyclase A0A5P9HJ69_V1 (SEQ ID NO: 43) and the mutant squalene cyclase A0A0J5GUC6_V1 (SEQ ID NO: 31).

Example 9. *In vivo* production of compound of formula (Ia) and biosynthetic intermediates in fungal cells engineered to produce (3E,7E)-homofarnesol and expressing different wild type or mutant squalene cyclases.

Wild type and mutant variants of the squalene cyclases were screened for (3E,7E)-homofarnesol cyclization and compound of formula (Ia) production in *Saccharomyces cerevisiae*. The sequences were codon optimized for expression in *S. cerevisiae* and were introduced in the strain YST403 in a plasmid containing CarG (SEQ ID NO: 2 encoded by SEQ ID NO: 92), PgpB (SEQ ID NO: 3 encoded by SEQ ID NO: 94) and SCH94-03944 (SEQ ID NO: 22 encoded by SEQ ID NO: 114).

Examples of 12 wild type and mutant variants (SEQ ID NOs: 29 to 41, 43 to 45 and 47 to 49 encoded by SEQ ID NOs: 123, 124, 126, 127, 129, 131, 133, 135, 136, 137, 138, 140, 142, 145, 146, 147, 149, 151 and 153) that showed to catalyze the conversion of (3E,7E)-homofarnesol to the compound of formula (Ia) are shown in Figure 10. Similar to Examples 7 and 8, it was evident that the mutations improved the titers of the compound of formula (Ia). A 10-fold improvement was observed when the mutant variant OYT72085_V1 (SEQ ID NO: 48) was expressed, in comparison to the wild-type OYT72085.1 (SEQ ID NO: 47) (Figure 11).

Example 10. *In vivo* production of compound of formula (Ia) and biosynthetic intermediates in bacterial cells engineered to produce (3E,7E)-homofarnesol and expressing different bacterial membrane-integrated meroterpenoid cyclases.

In this example, putative bacterial membrane-integrated meroterpenoid cyclases were tested *in vivo* in bacterial cells producing (3E,7E)-homofarnesol for the production of compound of formula (Ia). Enzyme candidates were selected based on a BLAST search using standard parameters with the protein sequence of the bacterial membrane-integrated meroterpenoid cyclase DmtA1 (*Nat Commun* 9, 4091 (2018). https://doi.org/10.1038/s41467-018-06411-x). Enzymes candidates having high to low protein sequence similarity with DmtA1 were randomly picked. The DNA fragments coding for enzyme candidates were codon optimized and cloned into the expression plasmid pD424 (ATUM, Newark, California). The vectors containing the enzyme candidates were transformed into DP1205 *E. coli* containing the (3E,7E)-homofarnesol producing vector pF-HOL11 (as described in Example 7). The resulting strains were cultivated under conditions enabling the production of terpene compounds and subsequently analyzed. We identified 6 bacterial meroterpenoid cyclases (SEQ ID NO: 50-55 encoded by SEQ ID NOs: 154, 156, 158, 160, 162 and 164) able to produce compound of formula (Ia). Their corresponding titers are shown in Figure 12. The highest titer observed was 92 mg/L when DP1205 *E. coli* co-expressed WP_190963420.1 (SEQ ID NO: 52) and pF-HOL11. This example shows that, surprisingly, meroterpenoid cyclases can convert (3E,7E)-homofarnesol to compound of formula (Ia) and can be used in the new biochemical pathway shown in Figure 13.

Example 11. *In vivo* production of compound of formula (Ia) and biosynthetic intermediates in fungal cells engineered to produce (3E,7E)-homofarnesol and expressing different bacterial membrane-integrated meroterpenoid cyclases.

In this example, the membrane-integrated meroterpenoid cyclases WP_051467941.1 (SEQ ID NO: 50), WP_234754442.1 (SEQ ID NO: 51), WP_190963420.1 (SEQ ID NO: 52), WP_093699331.1 (SEQ ID NO: 53) and WP_067007865.1 (SEQ ID NO: 54) from Example 10 were evaluated for their functional activity in *Saccharomyces cerevisiae* with the aim to produce the compound of formula (Ia) *in vivo*.

A codon optimized DNA coding sequence of each membrane-integrated meroterpenoid cyclases candidate (SEQ ID NOs: 155, 157, 159, 161, 163) was introduced in a plasmid that contained the genes coding for CarG (SEQ ID NO 2 encoded by SEQ ID NO: 92), PgpB (SEQ ID NO: 3 encoded by SEQ ID NO: 94) and SCH94-03944 (SEQ ID NO: 22 encoded by SEQ ID NO: 114) under galactose inducible promoters. The plasmids were *in vivo* constructed in the base strain YST403, which was previously engineered to produce high levels of FPP and which also harbored the genes for SCH23-ADH1 (SEQ ID NO: 21 encoded by SEQ ID NO: 112), SCH23-EST1 (SEQ ID NO: 28 encoded by SEQ ID NO: 122) and AflavBVMO1 (SEQ ID NO: 26 encoded by SEQ ID NO: 119). The resulting strains were called YST403 WP_051467941.1, YST403 WP_234754442.1, YST403 WP_190963420.1, YST403 WP_093699331.1 and YST403 WP_067007865.1. The strains were then cultivated under production conditions and the metabolites produced during cultivation were extracted and analyzed by GC-MS. The titers of compound of formula (Ia) for the strains are shown in Figure 14. All the bacterial membrane-integrated meroterpenoid cyclases that were tested in *Saccharomyces cerevisiae* were functionally active. However, the relative activity was different compared to the *in vivo* production of compound of formula (Ia) in *E. coli* (Example 10).

For example, the highest amount of compound of formula (Ia) that was produced *in vivo* was observed for WP_051467941.1 1 (SEQ ID NO: 50) when *Saccharomyces cerevisiae* was used as host strain. In contrast, when *E. coli* was the host strain the highest amount was detected for WP_190963420.1 (SEQ ID NO: 52). Figure 15 shows the chromatogram of YST403 WP_234754442.1 (SEQ ID NO: 51) compared to (3E,7E)-homofarnesol producing control YST403 HFOL.

Example 12. *In vivo* production of compound of formula (Ia) and biosynthetic intermediates in *Saccharomyces 123 erevisiae* cells engineered to produce (3E,7E)-homofarnesol and expressing a fungal membrane-integrated meroterpenoid cyclase.

In this example, fungal membrane-integrated meroterpenoid cyclases were evaluated for the *in vivo* production of compound of formula (Ia) in *Saccharomyces cerevisiae*.

Enzyme candidates were selected based on a blast search with the protein sequence of the fungal membrane-integrated meroterpenoid cyclase A0A2P1DP74.1 (MacJ) (SEQ ID NO: 71) (Org. Lett. 2017,

19, 5376−5379). Enzymes candidates having high to low protein sequence similarity with A0A2P1DP74.1 (MacJ) (SEQ ID NO: 71) were randomly picked and each candidate was introduced as codon optimized DNA coding sequence in a plasmid in the base strain YST403 as described in Example 11. The resulting strains were cultivated under production conditions and the metabolites produced during cultivation were extracted and analyzed by GC-MS. We found three fungal meroterpenoid cyclases (A0A2P1DP74.1 (MacJ) (SEQ ID NO: 71 encoded by SEQ ID NO: 186), XP_018029969.1 (SEQ ID NO: 72 encoded by SEQ ID NO: 187) and KAG0152682.1 (SEQ ID NO: 73 encoded by SEQ ID NO: 188)) which produced detectable amounts (0.035, 0.03 and 0.03 mg/l, respectively) of compound of formula (Ia).

The GC-MS chromatogram in the single ion monitoring mode of the strain YST403 expressing A0A2P1DP74.1 (MacJ) (SEQ ID NO: 71) compared to the (3E,7E)-homofarnesol producing control YST403 HFOL is shown in Figure 16.

Example 13. *In vivo* production of compound of formula (Ib) and biosynthetic intermediates in bacterial cells engineered to produce (3E,7E)-homofarnesol and expressing a bacterial MstE-like soluble meroterpenoid cyclase.

In this Example, 14 bacterial MstE-like, soluble meroterpenoid cyclases were evaluated in *E. coli* for the *in vivo* production of compound of formula (Ia). Enzyme candidates were selected based on a blast search with the protein sequence of the bacterial soluble meroterpenoid cyclase MstE (Angew. Chem. Int. Ed. Engl. 56 (18), 4987-4990 (2017)) (SEQ ID NO: 76).

Enzymes having high to low protein sequence similarity with MstE were randomly selected. DNA fragments coding for enzyme candidates were codon optimized and cloned into the expression plasmid pD424 (ATUM, Newark, California). The vectors containing the enzyme candidates were co-transformed into DP1205 *E. coli* strain together with the vector pF-HOL11 (as described in Example 7). The resulting strains were cultivated under conditions enabling the production of terpene compounds and subsequently analyzed. Surprisingly, it was found that bacterial soluble meroterpenoid cyclases were active and produced selectively compound of formula (Ib). In particular, *E. coli* cells that co-expressed pF-HOL11 and OKH29475.1 (SEQ ID NO: 74 encoded by SEQ ID NO: 189) or NEQ07043.1 (SEQ ID NO: 75 encoded by SEQ ID NO: 191) produced 2.91 mg/L and 1.83 mg/L of compound of formula (Ib), respectively. Figure 17 shows the chiral chromatogram measured on a Cyclosil-B column in selective ion mode for the *in vivo* production of compound of formula (Ib) by OKH29475.1 (SEQ ID NO: 74). The retention time of the *in vivo* produced compound of formula (Ib) is identical with an authentic standard and is clearly distinguishable from the retention time of authentic compound of formula (Ia) standard.

Example 14. *In vivo* production of compound of formula (Ib) and biosynthetic intermediates in fungal cells engineered to produce (3E,7E)-homofarnesol and expressing bacterial MstE-like soluble meroterpenoid cyclases.

The two bacterial MstE-like soluble meroterpenoid cyclases OKH29475.1 (SEQ ID NO: 74) and NEQ07043.1 (SEQ ID NO: 75), which showed the highest activities in *E. coli* were codon optimized (SEQ ID NOs: 190 and 192) and expressed in *Saccharomyces cerevisiae* cells engineered to produce (3E,7E)-homofarnesol (constructed as described in Example 6). GC-MS analysis of culture samples showed that both enzymes were capable to produce trace amounts of compound of formula (Ib) (Figure 18).

Example 15. *In vivo* production of compound of formula (Ia) and biosynthetic intermediates in bacterial cells engineered to produce (3E,7E)-homofarnesol and expressing a mutant bacterial membrane-integrated meroterpenoid cyclase.

In this example, compound of formula (Ia) is being produced *in vivo* in bacterial cells that produce (3E,7E)-homofarnesol (as described in Example 10) and contain different mutant variants of the bacterial membrane-integrated meroterpenoid cyclase WP_234754442.1 (SEQ ID NO: 51). It is shown that different mutant variants of WP_234754442.1 (SEQ ID NO: 51) are able to produce compound of formula (Ia) in significant higher quantities compared to the wild type sequence (SEQ ID NO: 51). The mutant variants of WP_234754442.1 (SEQ ID NO: 51) with increased activity were determined by screening of a mutant enzyme library. Mutant variants in this library were designed by a structure guided approach that was based on a protein structure model of WP_234754442.1 (SEQ ID NO: 51), which was build using ESMFold (Zeming et al. *Evolutionary-scale prediction of atomic level protein structure with a language* model. bioRxiv 2022.07.20.500902). End to end atomic level structure prediction from the protein sequence was based on the pretrained neural network model esm.pretrained.esmfold_v1 (https://github.com/facebookresearch/esm#esmfold). Extracted representations of the protein sequence are derived from the protein language model ESM-2. The protein model is shown in Figure 19. It adopts a pore like structure consisting of 7 helices. This finding was in agreement with the transmembrane helice prediction software TMHMM 2.0 (TMHMM 2.0 server available at https://dtu.biolib.com/DeepTMHMM (Krogh, A., et al. (2001) J Mol Biol 305(3)), which also predicted the presence of 7 transmembrane helices. According to the TMHMM 2.0 prediction, the C-terminus is outside the membrane and located in extracellular space.

It was assumed that the active site is oriented inside the pore like structure. Since the substrate is being produced in the cytosol, it was expected that the entrance to the active site is oriented towards the intracellular space and is in proximity to the N-terminus (shown in Figure 19).

For the creation of the mutant enzyme library, nine amino acids residues located inside and at the entrance of the active site of WP_234754442.1 (SEQ ID NO: 51) were then selected as target amino acids for the creation of a series of single point mutation variants.

Codon optimized mutant variants were ordered and cloned into an expression vector containing the clodf13 origin, the streptomycin resistance (SmR), T5 promoter, lambda T0 terminator and lac operator to control transcription as well as the lactose operon repressor (lacI). The vectors containing the genes of the mutant enzymes were then each transformed into DP1205 *E. coli* harboring the vector pF-HOL11 (as described in Example 7). The resulting strains were tested for their *in vivo* production of compound of formula (Ia) as previously described. Figure 20 shows the results for SEQ ID NOs: 56 to 58, 62 to 70 (encoded by SEQ ID NOs: 165, 167, 169, 177-185). It was found that mutations in the amino acid positions S7, S9, S51 and/or N63 were able to increase the *in vivo* production of compound of formula (Ia) by several fold. In particular, the mutants WP_234754442.1 S9M (SEQ ID NO: 57) and WP_234754442.1 S9C (SEQ ID NO: 56) were able to increase the titer by respectively 2508% and 2415% compared to the wild type control. This data clearly shows the engineering potential of transmembrane meroterpenoid cyclases for the conversion of (3E,7E)-homofarnesol to compound of formula (Ia).).

Example 16*: In vivo* production of compound of formula (Ia) and biosynthetic intermediates in fungal cells engineered to produce (3E,7E)-homofarnesol and expressing the bacterial membrane-integrated meroterpenoid cyclase WP_234754442.1 or the mutant variants WP_234754442.1 S9C or WP_234754442.1 S9M.

In this example, the mutant variants WP_234754442.1 S9C (SEQ ID NO: 56) and WP_234754442.1 S9M (SEQ ID NO: 57) that were shown to produce high amounts of compound of formula (Ia) compared to the wild type enzyme when expressed in *E.coli* cells producing (3E,7E)-homofarnesol (see Example 15) were tested for their functional activity *Saccharomyces cerevisiae.* Therefore, codon optimized versions of the nucleotide sequences (SEQ ID NOs: 166 and 168) were introduced in the base strain YST403 that harbours the (3E,7E)-homofarnesol biosynthetic genes and were tested for *in vivo* production of compound of formula (Ia) as described above.

It was found that both mutant enzyme variants are functionally active in *Saccharomyces cerevisiae* and show a significantly higher conversion of (3E,7E)-homofarnesol to compound of formula (Ia) compared to the wild type enzyme (Figure 21).

Example 17. Bioconversion of (3E,7E)-homofarnesol in bacterial cells expressing bacterial membrane-integrated meroterpenoid cyclases.

In this example, the three bacterial membrane-integrated meroterpenoid cyclases WP_051467941.1 (SEQ ID NO: 50), WP_234754442.1 (SEQ ID NO: 51) and WP_190963420.1 (SEQ ID NO: 52)

described in Example 10 were tested for the bioconversion of chemically synthesized (3E,7E)-homofarnesol to compound of formula (Ia). The bacterial membrane-integrated meroterpenoid cyclases was compared to the recently characterized mutant squalene cyclase AAcSHC_M132R_A224V_I432T (WO2016/170099) (SEQ ID NO: 78). The comparison was done either in the presence or absence of the detergent sodium dodecyl sulfate (SDS).

All enzymes were codon optimized (SEQ ID NO: 154, 156, 158, 193) and cloned into the expression plasmid pD424 (ATUM, Newark, California), which was transformed into *E.coli* Bl21(DE3)Star (Thermo Fisher Scientific Inc). Protein expression in the *E.coli* cells was done in 50mL scale with LB-medium at 25 °C for 24 hours and in the presence of 0.2 mM IPTG. The cells of each culture were centrifuged and resuspended in either 50 mM Natrium Phosphate (pH 6.9) or 50 mM Natrium Phosphate (pH 6.9) containing 0.06 % (w/v) SDS to reach the cell density of OD600 = 10. For the bioconversion 1 mL of the cell solution was added to a glass tube (Wheaton® 358646), which contained 10 µL of a (3E,7E)-homofarnesol dimethylsulfoxid (DMSO) solution (50 mg (87 % (3E,7E)-homofarnesol)/1 mL DMSO)). (3E,7E)-homofarnesol was prepared as described by (Eichhorn, Eric, and Fridtjof Schroeder. "From Ambergris to (−)-Ambrox: Chemistry Meets Biocatalysis for Sustainable (−)-Ambrox Production." *Journal of Agricultural and Food Chemistry* (2023)) and contained impurities of (3Z,7E)-homofarnesol. The class tubes were sealed and incubated for 48 hours at 25 °C at 250 rpm in a Minitron incubation shaker (Infors AG). After incubation the tubes were extracted with methyl-tert-butylether containing an internal standard. The organic overlays were analyzed by GC-MS/FID. The results are shown in Figure 22.

In Figure 22, it can be seen that the bacterial membrane-integrated meroterpenoid cyclase WP_234754442.1 (SEQ ID NO: 51) produced the highest amount of compound of formula (Ia) of all membrane-integrated meroterpenoid cyclases tested. Interestingly, all bacterial membrane-integrated meroterpenoid cyclases tested showed a significant increase in the bioconversion when SDS was absent. The opposite was observed for the squalene cyclase AAcSHC_M132R_A224V_I432T (SEQ ID NO: 78), which requires the presence of SDS to show significant functional activity at higher substrate concentration and to avoid substrate excess inhibition, which is in agreement with the literature (*Angewandte Chemie International Edition* (2023): e202301607). In contrast to the bioconversion of (3E,E7)-homofarnesol by a squalene cyclase, the bacterial meroterpenoid cyclase allowed for a detergent-free bioconversion. The elimination of detergents is cost saving and also facilitates the downstream process, as detergents tend to form emulsions that are difficult to break down and reduce the yield of the product extraction. Under detergent-free conditions, the bioconversion with the bacterial meroterpenoid cyclase WP_234754442.1 (SEQ ID NO: 51) resulted in 334 mg/L of compound of formula (Ia), while 8 mg/L of (3E,7E)-homofarnesol were still detected (see Figure 23).

In the bioconversion of synthetic (3E,7E)-homofarnesol, small amounts of the isomers compound of formula (Ic) and/or compound of formula (Id) were also observed (see Figure 23). Both isomers have been described as cyclisation products of AAcSHC_M123R_I432T_A224V (WO2016/170099) (SEQ ID

NO: 78). Further, it was described that AAcSHC_M132R_A224V_I432T produces compound of formula (Ic) from (3Z,7E)-homofarnesol. The formation of compound of formula (Id) was described as minor side product from the cyclisation of (3E,7E)-homofarnesol by the squalene cyclase AAcSHC_M132R_A224V_I432T (SEQ ID NO: 78). Interestingly, compound of formula (Id) was neither detected as a side product in the bioconversion of chemically synthesised homofarnesol *nor* when compound of formula (Ia) was produced *in vivo* (see Example 10) by any of the bacterial membrane-integrated meroterpenoid cyclases tested. This shows that bacterial membrane-integrated meroterpenoid cyclases are more selective than squalene cyclases and are able to selectively cyclize (3E,7E)-homofarnesol to a single product.

However, small amounts of compound of formula (Ic) were formed by bioconversion with the bacterial membrane-integrated meroterpenoid cyclases WP_234754442.1 (SEQ ID NO: 51) and WP_190963420.1 (SEQ ID NO: 52), respectively. This compound was not detectable *in vivo*. It is, therefore, very likely that compound of formula (Ic) derives from the cyclisation of (3Z,7E)-homofarnesol, which is an impurity present in the chemically synthesized (3E,7E)-homofarnesol. Due to the high enzymatic selectivity, the *in vivo* produced (3E,7E)-homofarnesol is free from impurities of homofarnesol isomers such as (3Z,7E)-homofarnesol, which represents a significant advance compared to the chemical synthesis.

Example 18. Bioconversion of (3E,7E)-homofarnesol and *in vivo* production of compound of formula (Ia) in bacterial cells expressing different squalene cyclases.

In this example, further squalene cyclases were tested *in vitro* in bacterial cells for the bioconversion of chemically synthesized (3E,7E)-homofarnesol (compound of formula (VIa)) to compound of formula (Ia). The squalene cyclase candidates were selected as described in Example 8 and mutations were introduced as described in Example 7. For some candidates, the N-terminal sequence was modified to add the 66 first N-terminal amino acids of SEQ ID NO: 89. The coding sequences encoding for squalene cyclase enzyme candidates were ordered as codon optimized DNA fragments and cloned into an expression vector containing the clodf13 origin, the streptomycin resistance (SmR), T5 promoter, lambda T0 terminator and lac operator to control transcription as well as the lactose operon repressor (lacI). The vectors containing the genes of the squalene cyclase enzymes were then each transformed into *E. coli* BI21(DE3)Star (Thermo Fisher Scientific Inc).

Protein expression in the *E. coli* cells was done in polypropylene DeepWell plates in 0.5 mL AM medium at 25 °C for 24 hours in the presence of 0.1 mM IPTG. After that, 5 µL of a (3E,7E)-homofarnesol in solution in dimethylsulfide (DMSO) (described in Example 17) was added to each well.

The DeepWell Plate was sealed and incubated for another 72 hours at 25 °C at 1000 rpm in a Minitron incubation shaker (Infors AG). After incubation the wells were extracted with ethyl acetate containing an internal standard. The organic overlays were analyzed by GC-MS/FID as described above and formation of compound of formula (Ia) was evaluated.

The squalene cyclases (SEQ ID NO: 265 to 279 encoded by SEQ ID NO: 290 to 304) showed cyclization activity. For the squalene cyclases A0A0T6LPP7-V1 (SEQ ID NO: 265), A0A7V0I7Y5-V1 (SEQ ID NO: 266), UPI00248B5E40-V1 (SEQ ID NO: 267) and UPI002800B5BA-V1 (SEQ ID NO: 268), cyclization of (3E,7E)-homofarnesol to compound of formula (Ia) is shown in Figure 24. It can be seen that for A0A0T6LPP7-V1 (SEQ ID NO: 265), the (3E,7E)-homofarnesol cyclization enzyme activity was similar or even superior to the ones of ZmSHC_F437A_G600M (SEQ ID NO: 88) and AacSHC_F437A_G600M (SEQ ID NO: 81). For the other squalene cyclases, (3E,7E)-homofarnesol cyclization to compound of formula (Ia), cyclization activity was also observed; in particular, for A0A0S8EHU4-V1 (SEQ ID NO: 269), A0A561EKN8-V1 (SEQ ID NO: 270), A0A6I1QM06-V1 (SEQ ID NO: 271), A0A7C2YSH8-V1 (SEQ ID NO: 272), A0A9C8DF34-V1 (SEQ ID NO: 273), A0A9W6UWM7-V1 (SEQ ID NO: 274), A0LVK9-V1 (SEQ ID NO: 275), UPI0004233625-V1 (SEQ ID NO: 276), WP_201153260.1-V1 (SEQ ID NO: 277), NPA95314.1-V1 (SEQ ID NO: 278), MDT8427665.1-V1 (SEQ ID NO: 279).

The above squalene synthases were then used in an *in vivo* experiment using bacterial cells engineered to produce (3E,7E)-homofarnesol as described in Example 7, except that the DNA fragments encoding the squalene cyclases were cloned in the streptomycin resistance vector described above. For all squalene synthases, *in vivo* production of compound of formula (Ia) was confirmed.

Example 19. *In vivo* production of compound of formula (Ia) and biosynthetic intermediates in bacterial cells engineered to produce (3E,7E)-homofarnesol and expressing different bacterial membrane-integrated meroterpenoid cyclases.

In this example, further bacterial membrane-integrated meroterpenoid cyclases were tested *in vivo* in bacterial cells producing (3E,7E)-homofarnesol for the production of compound of formula (Ia). Said bacterial membrane-integrated meroterpenoid cyclases were selected based on a BLAST search using standard parameters with the protein sequence of the bacterial membrane-integrated meroterpenoid cyclase WP_234754442.1 (Seq ID NO: 51). The enzymes having high to low protein sequence similarity with WP_234754442.1 (SEQ ID NO: 51) were randomly selected and were tested *in vivo* as described in Example 15. For this, the coding sequences encoding the bacterial membrane-integrated meroterpenoid cyclases were ordered as codon optimized DNA fragments and cloned into the expression vector containing the clodf13 origin, the streptomycin resistance (SmR), T5 promoter, lambda T0 terminator and lac operator to control transcription as well as the lactose operon repressor (lacI). The vectors containing the genes of bacterial membrane-integrated meroterpenoid cyclases were then each transformed into DP1205 *E. coli* harboring the vector pF-HOL11 (as described in Example 7). The resulting strains were tested for their *in vivo* production of compound of formula (Ia) as previously described.

The bacterial membrane-integrated meroterpenoid cyclases (SEQ ID NOs: 280 to 289 encoded by SEQ ID NO: 305 to 314) showed cyclization activity and were able to produce compound of formula (Ia). Their corresponding titers relative to WP_234754442.1 (SEQ ID NO: 51) are shown in Figure 25. In

particular, it was found that the bacterial membrane-integrated meroterpenoid cyclases HBB88633.1 (SEQ ID NO: 280), WP_033281172.1 1 (Seq ID NO: 281), WP_028277553.1 1 (Seq ID NO: 282), MBI4522454.1 (Seq ID NO: 283), WP_229232892.1 (SEQ ID NO: 286), WP_183976775.1 1 (Seq ID NO: 287) and WP_206725628.1 (Seq ID NO: 288) were able to produce similar or higher amounts of compound of formula (Ia) when compared to WP_234754442.1 (Seq ID NO: 51) under same conditions.

SEQUENCE LISTING

Sequences:

Sequence Number (ID): 1
Sequence Name: CcrGGPPS2-del57
Length: 313
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..313
> mol_type, protein
> organism, synthetic construct
Residues:
MASVLSGKDT MKGEEENSGF DFQSYMDQMA DSVNQALESA VSLREPLKIH EAMRYSLLAG
GKRVRPLLCI AACELVGGDV SVAMPAACAV EMIHTMSLIH DDLPCMDNDD LRRGKPTNHK
AFGEDIAVLA GDALLSFAFE HVAVSTVGAS PDKIVRAVGE LAKAVGKEGL VAGQVVDITS
EGLNDVGLDH LEYIHVHKTA VLLEAAVVLG AILGGGTDEE VERLRKFAIC IGLLFQVVDD
ILDVTKSSVE LGKTAGKDLV ADKVTYPKLM GLEKSREFAE KLRDDAVEQL RVFDQVKAAP
LIALAHYIAY RQN

Sequence Number (ID): 2
Sequence Name: CarG
Length: 320
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..320
> mol_type, protein
> organism, Blakeslea trispora
Residues:
MLTSSKSIES FPKNVQPYGK HYQNGLEPVG KSQEDILLEP FHYLCSNPGK DVRTKMIEAF
NAWLKVPKDD LIVITRVIEM LHSASLLIDD VEDDSVLRRG VPAAHHIYGT PQTINCANYV
YFLALKEIAK LNKPNMITIY TDELINLHRG QGMELFWRDT LTCPTEKEFL DMVNDKTGGL
LRLAVKLMQE ASQSGTDYTG LVSKIGIHFQ VRDDYMNLQS KNYADNKGFC EDLTEGKFSF
PIIHSIRSDP SNRQLLNILK QRSSSIELKQ FALQLLENTN TFQYCRDFLR VLEKEAREEI
KLLGGNIMLE KIMDVLSVNE

Sequence Number (ID): 3
Sequence Name: PgpB
Length: 254
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..254
> mol_type, protein
> organism, Escherichia coli
Residues:
MRSIARRTAV GAALLLVMPV AVWISGWRWQ PGEQSWLLKA AFWVTETVTQ PWGVITHLIL
FGWFLWCLRF RIKAAFVLFA ILAAAILVGQ GVKSWIKDKV QEPRPFVIWL EKTHHIPVDE
FYTLKRAERG NLVKEQLAEE KNIPQYLRSH WQKETGFAFP SGHTMFAASW ALLAVGLLWP
RRRTLTIAIL LVWATGVMGS RLLLGMHWPR DLVVATLISW ALVAVATWLA QRICGPLTPP
SEENREIAQR EQES

Sequence Number (ID): 4
Sequence Name: NudB
Length: 150
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..150
> mol_type, protein
> organism, Escherichia coli
Residues:
MKDKVYKRPV SILVVIYAQD TKRVLMLQRR DDPDFWQSVT GSVEEGETAP QAAMREIKEE
VTIDVVAEQL TLIDCQRTVE FEIFSHLRHR YAPGVTRNTE SWFCLALPHE RQIVFTEHLA
YKWLDAPAAA ALTKSWSNRQ AIEQFVINAA

Sequence Number (ID): 5
Sequence Name: LPP1
Length: 274
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..274
> mol_type, protein
> organism, Saccharomyces cerevisiae
Residues:
MISVMADEKH KEYFKLYYFQ YMIIGLCTIL FLYSEISLVP RGQNIEFSLD DPSISKRYVP
NELVGPLECL ILSVGLSNMV VFWTCMFDKD LLKKNRVKRL RERPDGISND FHFMHTSILC
LMLIISINAA LTGALKLIIG NLRPDFVDRC IPDLQKMSDS DSLVFGLDIC KQTNKWILYE
GLKSTPSGHS SFIVSTMGFT YLWQRVFTTR NTRSCIWCPL LALVVMVSRV IDHRHHWYDV
VSGAVLAFLV IYCCWKWTFT NLAKRDILPS PVSV

Sequence Number (ID): 6
Sequence Name: NUDX1
Length: 150
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..150
> mol_type, protein
> organism, Rosa hybrida
Residues:
MGNETVVVAE TAGSIKVAVV VCLLRGQNVL LGRRRSSLGD STFSLPSGHL EFGESFEECA
ARELKEETDL DIGKIELLTV TNNLFLDEAK PSQYVAVFMR AVLADPRQEP QNIEPEFCDG
WGWYEWDNLP KPLFWPLDNV VQDGFNPFPT

Sequence Number (ID): 7
Sequence Name: PeSubTPP1
Length: 290
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..290
> mol_type, protein
> organism, Penicillium subrubescens
Residues:
MQPFISVDGV VNFRDIGGYV CRNPAGLSSL PSNVDETPEK QWCIRPGFVF RAAQPSQITP
AGIEILKKTL AIQAIFDFRS ESEIQLVSKR YPDSLLDIPG TTRHAVPVFQ EGDYSPISLA
KRYGVTADES TNDQSFRPGF VKAYEAIARN AAQAGSFRAI IQHILQDSAG PVLFHCTVGK
DRTGVFSALI LKLCGVADED IVADYALTTQ GLGVWREHLI QRLLQRGEAT TKEQAEAIIS
SDPRDMKAFL SNVVEGEFGG ARNYFVNLCG LPEGEVDRVI TKLVVPKTTK

Sequence Number (ID): 8
Sequence Name: TalVeTPP
Length: 311
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..311
> mol_type, protein
> organism, Talaromyces verruculosus
Residues:
MSNDTTTTAS AGTATSSRFL SVGGVVNFRE LGGYPCDSVP PAPASNGSPD NASEATLWVG
HSSIRPGFLF RSAQPSQITP AGIETLIRQL GIQTIFDFRS RTEIELVATR YPDSLLEIPG
TTRYSVPVFS EGDYSPASLV KRYGVSSDTA TDSTSSKSAK PTGFVHAYEA IARSAAENGS
FRKITDHIIQ HPDRPILFHC TLGKDRTGVF AALLLSLCGV PDETIVEDYA MTTEGFGAWR
EHLIQRLLQR KDAATREDAE SIIASPPETM KAFLEDVVAA KFGGARNYFI QHCGFTEAEV
DKLSHTLAIT N

Sequence Number (ID): 9

Sequence Name: AstI
Length: 201
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..201
> mol_type, protein
> organism, Aspergillus oryzae RIB40
Residues:
MTRQSHYQAI ILDLGNVVFE WDTSQNPPTA APNQISLLRT SMKSPVYHSY ERGQLSTEEC
HRLLGESLHV DPGQIKEAFD LARQSLRSNP ALLDFIRQLK QTRGVAVYAM SNIPQAEIEY
LKESRAGDME VFDEVFASGY VGSRKPETEF YRRVMGEIGL KAERVVFVDD KEENVDVARG
LGLYGVCFGG VEELRGHLLG I

Sequence Number (ID): 10
Sequence Name: AstK
Length: 196
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..196
> mol_type, protein
> organism, Aspergillus oryzae RIB40
Residues:
MCTTFKAAIF DMGGVLFTWN PIVDTQVSLK DLGTIINSET WEQFERGKIE PDDCYHQLGS
QIGLPGSEIA ATFRQTTGCL RPDARMTSLL RELKGQGVAV YMMTNIPAPD FHQLREMHYE
WDLFDGIFAS ALEGMRKPDL EFYEHVLKQI DTSAAETIFV DDKLENVIAA QAVGMVGLHL
TDSLATCMEL RQLVGC

Sequence Number (ID): 11
Sequence Name: PsAerADH
Length: 372
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..372
> mol_type, protein
> organism, Pseudomonas aeruginosa
Residues:
MNSIQPTQAK AAVLRAVGGP FSIEPIRISP PKGDEVLVRI VGVGVCHTDV VCRDSFPVPL
PIILGHEGSG VIEAVGDQVT GLKPGDHVVL SFNSCGHCYN CGHDEPASCL QMLPLNFGGA
ERAADGTIED DQGAAVRGLF FGQSSFGSYA IARAVNTVKV DDDLPLALLG PLGCGIQTGA
GAAMNSLGLQ GGQSFIVFGG GAVGLSAVMA AKALGVSPLI VVEPNEARRA LALELGASHA
FDPFNTEDLV ASIREVVPAG ANHALDTTGL PKVIANAIDC IMSGGKLGLL GMANPEANVP
ATLLDLLSKN VTLKPITEGD ANPQEFIPRM LALYREGKFP FDKLITTFPF EHINEAMEAT
ESGKAIKPVL TL

Sequence Number (ID): 12
Sequence Name: ThTerpADH1
Length: 373
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..373
> mol_type, protein
> organism, Thauera terpenica
Residues:
MCSNHDFTAA RAAVLRKVGG PLEIEDVRIS APKGDEVLVR MVGVGVCHTD LVCRDAFPVP
LPIVLGHEGA GIVEAVGEGV RSLEPGDRVV LSFNSCGRCG NCGSGHPSNC LQMLPLNFGG
AQRVDGGRML DAAGNAVQGL FFGQSSFGTY AIAREINAVK VAEDLPLEIL GPLGCGIQTG
AGAAINSLGI GPGQSLAVFG GGGVGLSALL GARAVGAAQV VVVEPNAARR ALALELGASH
AFDPFAGDDL VAAIRAATGG GATHALDTTG LPSVIGNAID CTLPGGTVGM VGMPAPDAAV
PATLLDLLTK SVTLRPITEG DADPQAFIPQ MLRFYREGKF PFDRLITRYR FDQINEALHA
TEKGGAIKPV LVF

Sequence Number (ID): 13
Sequence Name: VoADH1
Length: 381
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..381
> mol_type, protein
> organism, Valeriana officinalis
Residues:
MTKSSGEVIS CKAAVIYKSG EPAKVEEIRV DPPKSSEVRI KMLYASLCHT DILCCNGLPV
PLFPRIPGHE GVGVVESAGE DVKDVKEGDI VMPLYLGECG ECLNCSSGKT NLCHKYPLDF
SGVLPSDGTS RMSVAKSGEK IFHHFSCSTW SEYVVIESSY VVKVDSRLPL PHASFLACGF
TTGYGAAWKE ADIPKGSTVA VLGLGAVGLG VVAGARSQGA SRIIGVDIND KKKAKAEIFG
VTEFLNPKQL GKSASESIKD VTGGLGVDYC FECTGVPALL NEAVDASKIG LGTIVMIGAG
METSGVINYI PLLCGRKLIG SIYGGVRIRS DLPLIIEKCI NKEIPLNELQ THEVSLEGIN
DAFGMLKQPD CVKIVIKFEQ K

Sequence Number (ID): 14
Sequence Name: SCH80-05240
Length: 373
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..373
> mol_type, protein
> organism, Rhodococcus sp. Ni2
Residues:
MIRAEQNSTS AMQMTAALSH GPHSPFTLDT VEIDEPRADE ILVRIVATGL CHTDLFTKSV
LPERLGPCVF GHEGAGVVEA VGSAIDKVVP GDHVLLSYRS CGVCRQCLSG HRAYCESSHG
LNSSGARTDG STPVRRSGTP IRSAFFGQSS FAEYVIATAD NTVVVDPAVD LTVAAPLGCG
FQTGAGAVLN LLRPEPDSTF VVFGAGSVGL AALLAARAAG VSTLVAVDPV AQRRALAEEF
GAVTVDPTTE DAVEAVRAAT DGGSTHSLDT TGIGSVINQA VTSLRARGTL AVVGLGASTV
EMNMADIMLS GKTIRGCIEG ESEVSTFIPE LVELFTGGRF PIDRLVTRYA FSDINKAVED
QASGRVIKPV LVW

Sequence Number (ID): 15
Sequence Name: Ppseudo-alkJ
Length: 558
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..558
> mol_type, protein
> organism, Pseudomonas fluorescens
Residues:
MYDYIIVGAG SAGCVLANRL SADPSKRVCL LEAGPRDTNP LIHMPLGIAL LSNSKKLNWA
FQTAPQQNLN GRSLFWPRGK TLGGSSSINA MVYIRGHEDD YHAWEQAAGR YWGWYRALEL
FKRLECNQRF DKSEHHGVDG ELAVSDLKYI NPLSKAFVQA GMEANINFNG DFNGEYQDGV
GFYQVTQKNG QRWSSARAFL HGVLSRPNLD IITDAHASKI LFEDRKAVGV SYIKKNMHHQ
VKTTSGGEVL LSLGAVGTPH LLMLSGVGAA AELKEHGVSL VHDLPEVGKN LQDHLDITLM
CAANSREPIG VALSFIPRGV SGLFSYVFKR EGFLTSNVAE SGGFVKSSPD RDRPNLQFHF
LPTYLKDHGR KIAGGYGYTL HICDLLPKSR GRIGLKSANP LQPPLIDPNY LSDHEDIKTM
IAGIKIGRAI LQAPSMAKHF KHEVVPGQAV KTDDEIIEDI RRRAETIYHP VGTCRMGKDP
ASVVDPCLKI RGLANIRVVD ASIMPHLVAG NTNAPTIMIA ENAAEIIMRN LDVEALEASA
EFAREGAELE LAMIAVCM

Sequence Number (ID): 16
Sequence Name: PfluoADHF1
Length: 296
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..296
> mol_type, protein

> organism, Pseudomonas fluorescens DSM50106
Residues:
MKSFNGRVAA ITGAASGMGR ALALALAREG CHLALADKNA QGLEQTLALI KTSTLSPVMV
TTQVLDVADR QAMEAWAARC VAEHGQVNLV FNNAGVALSS TVEGVDYADL EWIVGINFWG
VVHGTKAFLP HLKASGDGHV INTSSVFGLF AQPGMSGYNA TKFAVRGFTE ALRQELDLQR
CGVSATCVHP GGIRTDICRS SRIDANMTGF LIHSEQQARA DFEKLFITDA DQAAKVILQG
VRKNKRRVLI GRDAYFLDLL ARCLPAAYQA LVVLASKRMA PKQRRPVFET NDEPRL

Sequence Number (ID): 17
Sequence Name: CymB
Length: 374
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..374
> mol_type, protein
> organism, Pseudomonas putida
Residues:
MTINSIQPIQ AKAAVLRAVG SPFNIEPIRI SPPKGDEVLV RIVGVGVCHT DVVCRDSFPV
PLPIILGHEG SGVIEAIGDQ VTSLKPGDHV VLSFNSCGHC YNCGHAEPAS CLQMLPLNFG
GAERAADGTI QDDKGEAVRG MFFGQSSFGT YAIARAVNAV KVDDDLPLPL LGPLGCGIQT
GAGAAMNSLS LQSGQSFIVF GGGAVGLSAV MAAKALGVSP LIVVEPNESR RALALELGAS
HVFDPFNTED LVASIREVVP AGANHALDTT GLPKVIASAI DCIMSGGKLG LLGMASPEAN
VPATLLDLLS KNVTLKPITE GDANPQEFIP RMLALYREGK FPFEKLITTF PFEHINEAME
ATESGKAIKP VLTL

Sequence Number (ID): 18
Sequence Name: CdGeoA
Length: 373
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..373
> mol_type, protein
> organism, Castellaniella defragrans
Residues:
MNDTQDFISA QAAVLRQVGG PLAVEPVRIS MPKGDEVLIR IAGVGVCHTD LVCRDGFPVP
LPIVLGHEGS GTVEAVGEQV RTLKPGDRVV LSFNSCGHCG NCHDGHPSNC LQMLPLNFGG
AQRVDGGQVL DGAGHPVQSM FFGQSSFGTH AVAREINAVK VGDDLPLELL GPLGCGIQTG
AGAAINSLGI GPGQSLAIFG GGGVGLSALL GARAVGADRV VVIEPNAARR ALALELGASH
ALDPHAEGDL VAAIKAATGG GATHSLDTTG LPPVIGSAIA CTLPGGTVGM VGLPAPDAPV
PATLLDLLSK SVTLRPITEG DADPQRFIPR MLDFHRAGKF PFDRLITRYR FDQINEALHA
TEKGEAIKPV LVF

Sequence Number (ID): 19
Sequence Name: AroAroADH
Length: 375
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..375
> mol_type, protein
> organism, Aromatoleum aromaticum
Residues:
MGSIQDSLFI PARAAVLRAV GGPLEIEDVR ISPPKGDEVL VRMVGVGVCH TDVVCRDGFP
VPLPIVLGHE GAGIVEAVGE RVTKVKPGQR VVLSFNSCGH CSSCGEDHPA TCHQMLPLNF
GAAQRVDGGC VTDASGEAVH SLFFGQSSFC TFALAREVNT VPVGDGVPLE ILGPLGCGIQ
TGAGAAINSL AIKPGQSLAI FGGGSVGLSA LLGALAVGAG PVVVVEPNDR RRALALDLGA
SHVFDPFNTE DLVASIKAAT GGGVTHSLDS TGLPPVIAKA IDCTLPGGTV GLLGVPAPDA
AVPVTLLDLL VKSVTLRPIT EGDANPQEFI PRMVQLYRDG KFPFDKLITT YRFENINDAF
KATETGEAIK PVLVF

Sequence Number (ID): 20
Sequence Name: AzTolADH1

Length: 375
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..375
> mol_type, protein
> organism, Azoarcus toluclasticus ATCC 700605
Residues:
MGSIQDSLFI RARAAVLRTV GGPLEIENVR ISPPKGDEVL VRMVGVGVCH TDVVCRDGFP
VPLPIVLGHE GSGIVEAVGE RVTKVKPGQR VVLSFNSCGH CASCCEDHPA TCHQMLPLNF
GAAQRVDGGT VIDASGEAVQ SLFFGQSSFG TYALAREVNT VPVPDAVPLE ILGPLGCGIQ
TGAGAAINSL ALKPGQSLAI FGGGSVGLSA LLGALAVGAG PVVVIEPNER RRALALDLGA
SHAFDPFNTE DLVASIKAAT GGGVTHSLDS TGLPPVIANA INCTLPGGTV GLLGVPSPEA
AVPVTLLDLL VKSVTLRPIT EGDANPQEFI PRMVQLYRDG KFPFDKLITT YRFDDINQAF
KATETGEAIK PVLVF

Sequence Number (ID): 21
Sequence Name: SCH23-ADH1
Length: 351
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..351
> mol_type, protein
> organism, Hyphozyma roseonigra ATCC 20624
Residues:
MQFSIGDVLA IVDKTILNPL VVSAGLLSLH FLTNDKYAIT ANDGLFPYQI STPDSHRKAL
FALGFGLLLR ANRYMSRKAL NNNTAAQFDW NREIIVVTGG SGGIGAQAAQ KLAERGSKVI
VIDVLPLTFD KPKNLYHYKC DLTNYKELQE VAAKIEREVG TPTCVVANAG ICRGKNIFDA
TERDVQLTFG VNNLGLLWTA KTFLPSMAKA NHGHFLIIAS QTGHLATAGV VDYAATKAAA
IAIYEGLQTE MKHFYKAPAV RVSCISPSAV KTKMFAGIKT GGNFFMPMLT PDDLGDLIAK
TLWDGVAVNI LSPAAAYISP PTRALPDWMR VGMQDAGAEI MTELTPHKPL E

Sequence Number (ID): 22
Sequence Name: SCH94-03944
Length: 157
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..157
> mol_type, protein
> organism, Rhodococcus erythropolis DSM 6344
Residues:
MNLNEARTAF ARLRAAENGL SPAELDEVWA ALETVAAEEI LGEWKGDDFA TGHRLHEKLS
ASRWYGKTFN SVEDAKPLIC RDEDGNLYSD VKSGNGEASL WNIEFRGEVT ATMVYDGAPI
FDHFKKVDDS TLMGIMNGKS ALVLDGGQHY YFLLERA

Sequence Number (ID): 23
Sequence Name: SCH24-BVMO1
Length: 612
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..612
> mol_type, protein
> organism, Cryptococcus magnus ATCC 20918
Residues:
MTIDLQQPDA VPFTSSTFVV PDPSNLASQA QNSQLQSAQE GAEYPVNAHG VRGDGTIHER
PINDRRKMRV ICVGAGISGL YMAIKLPRST ENVELKIYEK NHDLGGTWLE NRYPGCACDV
PAHAYAYSFE NNPEFPRFFS SSEDIHKYLL RVADKYDCKK YIAFNTKVVE AIWDEEQGIY
NVKIERSDGT VFQDTCEVLL NASGILNAWR YPGIPGIKDY KGTLMHSATW DRSVSLKGKK
VALIGSGSSG IQILPNILDD CKEVVTYIID PAWIAPANLV TAGVSDDGEE PKEPTPEELA
SSSDFAYSQE QINGFKKDPK SLMDHRATLE RTMNQSFPIL LRGSPSNLYA ASLFEDLMRK
RLAKKPEVAD AIIPEWSIGC RRLTPGPHYL EALCNPKVKI LTQAIKSFSD KGMYTADGEH
EDFDVVICAT GFDVSFRPRF KFIGKDGYEV PENFGQTPKG YLALAYAGFP NSFIFMGPNG

PIANGSVVVS LEKQGDYFIK AINKIQRQNI KGMTVRFDAV DDFTNHVDKY MDRTVLTDDC
ISWYKNGKRD GRVSAVWPGS ALHYMEAIAD PRWEDYTYTY REPGHSFSFL GDGTSWVEHT
GGDTAWYLKE TL

Sequence Number (ID): 24
Sequence Name: SCH46-BVMO1
Length: 588
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..588
> mol_type, protein
> organism, Bensingtonia ciliata ATCC 20919
Residues:
MPSAITPPVD HRSLPGLFKP QRKLKVICVG AGASGLLLSY KIQRHFEDFE LQVFEKNPEV
SGTWYENRYP GCACDVPSHN YTWSFEPKTD WSANYASSKE IFKYFKDFTK KYGLSKYIKL
EHEVVGATWM EAEAQWKVDV KDLRSGNTQS SFAHILVNAG GILNAWRYPP IPGIKDFKGD
LVHSAAWPEH LDLNGKVVGL IGNGSSGIQI LPAIKKDVKQ LVTFIREATW VAPPLGQAYR
AFSTDEQAQF AQDPRHHLET RRATEATMNQ SFGIFHSGSE EQKGVRQYMQ DIMETKLNNK
QLESVLIPEW SVGCRRLTPG TNYLESLSDD NVKVVYGEIT QITESGVICD DGKGEYPVEV
LICATGFDTT FKPRFPLIGT TKEKLSDVWK DDPRGYFGIA TNNYPNYFFT LGPNCPIGNG
PVLCAIEAEV EYIINMLSKF QKENIRSFDI KADAVDAFND WKDDFMKDTI WAEQCRSWYK
AGSATGKILA LWPGSTLHYL EALKSPRWED WDFKYQPGRN RFHYFGNGHS CAEQDGDLSW
YIRNEDDSYI DPVLKPKPKA AVESEAHIAL PGIGPMLMED PRDVAVEA

Sequence Number (ID): 25
Sequence Name: AraBVMO1
Length: 496
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..496
> mol_type, protein
> organism, Acinetobacter radioresistens
Residues:
MDKHIDVLIV GAGISGLGLA AHLSKNCPQR SFEIVERREG IGGTWDLFRY PGIRSDSDMS
TFGYNFKPWR KAKILADGAS IRQYLHEVVD EFHLDRKIHF KHRVISANYD TALKLWIVEI
EDQQGQNQTW YANFLLGCTG YYNYDEGFMP EYPGQHQFKG TLVHPQHWPE KLDYTGKRVI
VIGSGATAIT LVPSMVKGGA AHVTMLQRSP TYIASIPSID FVYQKMRGFL SEEMAYKLTR
ARNIGMQRAV YALSQKQPKL VRKLLLKSIE MQLKGKVDMK HFTPSYNPWD QRLCVVPDGD
LFKALREGHA SVETDHIEKF TETGIQLKSG KHLEADIIIS ATGLQIQIMG GIQGTVDGQP
IDTSEHMLYN GILISDVPNM AMIIGYINAS WTLKVDVAAE YICRLLNYMD KHHYDEVIAP
TDHSEIEQDT VMGSLSAGYI RRAADVIPKQ GKHAPWQVTN NYLADRKALK QAGFEDGILQ
FTKRDKQLER KPKLVS

Sequence Number (ID): 26
Sequence Name: AflavBVMO1
Length: 543
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..543
> mol_type, protein
> organism, Aspergillus flavus
Residues:
MNGTQASNGV LHLDALIIGS GFSGIYLLHK LRDELKLKVK IFEAESDIGG TWNNNRYPGA
RVDCPVPFYA YSLPEVWQSW NWTELYPNQK EIKSYFDHVD RVLDVRKDCL FHSRVNEGTF
DEATGRWTVW TTDGKVATAK YLLVAVGFAS KSYLPDWKGL DSFKGTIYHS AHWPEAEEIS
VKGKKVAVIG TGSTGIQIFQ EWAREAEEAF LFQRTPNLCL PMRQQELHAG YQVKDKGEYA
DYLAECALTF GGLEYQQTPK NTFDASEEER EAFWEDLYQM GGFRFWQNNY QDLLTSLDAN
REAYNFWARK TRARIQDPKK RDLLAPLEPP YPFGTKRPSL EQDFYEQFNK SNVHIVDTKS
QPIVGVTPTG IVTADEKVHE VDIIAVATGF DAVTGGLLRL GLKDVNGVGL DERWKDGMST
YLGMAISGFP NMFLPYSLQA PTAFANGPTL IELQGDWITS LIRKMEMENV QSVTATPHAE
SAWNDEVNMI ANKTLLPLTD SWYMGSNIPG KPVQSLNYLG GLPTYRERCA KVLDEDFFGF

AKA

Sequence Number (ID): 27
Sequence Name: SCH24-EST1
Length: 310
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..310
> mol_type, protein
> organism, Cryptococcus magnus ATCC 20918
Residues:
MTHSPPLDAE LSLLRYAPAV PVGWQLGRKL LRMNTLMTRP MEGVMRDDVV IPNLDGTANI
RLFICRPQDP TETMPVILWL HGGGMVAGHY KQDSGFMDIW AKRLGAFVVS VDYRLAPEAK
APAALDDCIA AWQWITTQTA RGIDTTRMAV GGASAGGGLA ASTVQRLVDL GGVKPVFQLL
IYPMLDDRTV VRFDPDRRYY MWTPDCNRYG WTSYLGVPPG SAEVPPYASA ARRPDLSGLP
PTWIGVGSLD LFHDEDMDYA RRLRESGVPV EEYVAVGAPH AFDTIYGKAK VTLDFWDSHF
NALRRALCLD

Sequence Number (ID): 28
Sequence Name: SCH23-EST1
Length: 345
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..345
> mol_type, protein
> organism, Hyphozyma roseonigra ATCC 20624
Residues:
MPSDLPRPAY DPEIEPFLSM VPLPPTINAD IMKELRKAPL LSQAPDLDAL LSDKPITHRE
VSIPGLNSQD PQITLSIFSS TLEGGPKPCI YFVHGGGMII GCRFVGIEDY LQYVEQNDAV
VVAVEYRLAP EHPDPAPVND CYAGLLWTAA NAAELGIDLE RLLICGASAG GGLSAGVALM
ARDKKGPKLV GQLLCYPMLD DRNDSLSSQQ YVDEGVWSRG SNAFGWKQLL GDRAGKEGVS
IYAAPARATD LSGLPNTFID VGSAEVFRDE DIAYASRLWA VGVQAELHVW PGGYHAAENM
APGTDYSKKV KATRLAWMKR VFMKAPKSTT ESLPAPTVDE AVGTI

Sequence Number (ID): 29
Sequence Name: A0A0C2RLD5_V1
Length: 615
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..615
> mol_type, protein
> organism, synthetic construct
Residues:
MQNQINELIG QLINEIEAHQ APDGSFRFCF ENSLMTDTNM MLLLKTFTTD ESLMNELGAR
VASLQHPDGY WSIYPGDNGN LSATIEAYFA LLYTGYWKKS ERSAQKASRF ILAQGGLGAA
HSMTKIMLAI HGQYPWPNLL HLPVHFILFP SSSPVSFYDF SSYARVHMAP ILLLTDSRFQ
LGEKDIPDMS ALLIRTPSEQ LNEHSRSLLS GIYQTASSIA GLPHLVHQRA RKRLVTYMLQ
RIEGDGTLYS YVSSTFYMLF ALVSEGYSKQ HPLIQKAITG LKTHRCLTNN GWHIQNSTST
VWDTALLCHA LHDLPFSPYK RKKAEVYLLK HQHDKFGDWI LTADQTSPGG WGFSDSNTIH
PDVDDTTASL RALSPSITVD PSLKESYLRG VSWVLAIQNE DGGWPAFERA KTNQLLTFVP
MDGASHAAID PSTADLTGRT LEFLSSEARL PFQHSAIQHA IRWLKKNQQS DGSWYGKWGI
SFLYGTWSAV TGLSAAGLNG EDPAIKKAVS FLERVQNEDG GWGESCLSDQ VMHYIPLGSS
TPSQTAWALD ALLSVHKQKT PSIERGINCL LGQLKTKDWT YRYPTGAMLP GNFYVYYHSY
NYIWPLIVLK KYAAL

Sequence Number (ID): 30
Sequence Name: A0A0J5GUC6
Length: 631
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..631

> mol_type, protein
> organism, Bacillus sp. LL01

Residues:
MIDKVNGKMN SLTSELLRKQ AKDGSWKFCF EGSIMTDAYM IILIRVLEIK DEEDLVRRLV
NRIKTKQSPN GAWKLYEDEK DGNLSATVEG YFALLYSGYT ERKERNMRKA ENFINQHGGL
SSCDWLTRMM LALNGQIEWP GIVKSIPIEI MLVPKWAPIN IYHLVGYARA HWVPIIISSN
LNASLVTAQT PKLSHLQTRE SGSEDYRLLE EMKLLSHYVN SAMKKLAASP ELLRKRAFTK
AENYISERIE ENGTLYSYFS ASFFMIFAFL ALGYDKNHPK IKNAFRGMKS YICSTHSDDK
FFVQNSPSTV WDTALLTGAL VQAGVLVHHE AMMSAGEYLL SRQHHKYGDW ALKNPDAPPG
GWGFSDINTL VPDVDDTTAA LRVITLLAHS DERYKSAWNK GVEWLLSMQN DDGGWSAFEK
NTDNYLLSFV PFRYEDRVFF DPSTADLTGR TLYFLGEYTT IPQDSEKILE AVKWLNYHQE
KDGSWYGRWG NCYIYGTWAA VTGLKAVGVP STDSTITRAV EWLLSIQNED GGWGESCFSD
VRNKYVPLHH STPSQTAWAL DALISVSESP TPQIEKGIRT LLDLMETNDW RTDYPTGGAI
PGGYYIHYHS YKYIWPLQAL GHYRNKFDLK T

Sequence Number (ID): 31
Sequence Name: A0A0J5GUC6_V1
Length: 631
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..631
> mol_type, protein
> organism, synthetic construct

Residues:
MIDKVNGKMN SLTSELLRKQ AKDGSWKFCF EGSIMTDAYM IILIRVLEIK DEEDLVRRLV
NRIKTKQSPN GAWKLYEDEK DGNLSATVEG YFALLYSGYT ERKERNMRKA ENFINQHGGL
SSCDWLTRMM LALNGQIEWP GIVKSIPIEI MLVPKWAPIN IYHLVGYARA HWVPIIISSN
LNASLVTAQT PKLSHLQTRE SGSEDYRLLE EMKLLSHYVN SAMKKLAASP ELLRKRAFTK
AENYISERIE ENGTLYSYFS ASFFMIFAFL ALGYDKNHPK IKNAFRGMKS YICSTHSDDK
FFVQNSPSTV WDTALLTGAL VQAGVLVHHE AMMSAGEYLL SRQHHKYGDW ALKNPDAPPG
GWGFSDINTL VPDVDDTTAA LRVITLLAHS DERYKSAWNK GVEWLLSMQN DDGGWSAFEK
NTDNYLLSFV PFRYADRVFF DPSTADLTGR TLYFLGEYTT IPQDSEKILE AVKWLNYHQE
KDGSWYGRWG NCYIYGTWAA VTGLKAVGVP STDSTITRAV EWLLSIQNED GGWGESCFSD
VRNKYVPLHH STPSQTAWAL DALISVSESP TPQIEKGIRT LLDLMETNDW RTDYPTGGMI
PGGYYIHYHS YKYIWPLQAL GHYRNKFDLK T

Sequence Number (ID): 32
Sequence Name: A0A0M9GQE0
Length: 631
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..631
> mol_type, protein
> organism, Bacillus sp. CHD6a

Residues:
MKPVDKINEK LKDMINNLLD TQTEKGTWNF CFEGSIMTDA YMIILIRILE LTDEELLVKS
LVERIKSKQE SNGAWKVYPD EDGGNLSATI EGYFSLLHSG YVAKDAANMR KAERFIQENG
GLAESDWLTK MMLALTGQIK WPSIIKIIPI EIMLLPNWSP ISIYQFVGYA RAHWIPILIC
SNLNYSYLHS RTPNLIHLQG VGSDSEDQRF VEDRQHLQLY FKNALKKLTG SPEVLKRKAF
IKAEDYILER IEENGTLYSY FSASFFMVFA FLALGYDKKN PLIQNAFQGM KAYLCRNADH
AFIQNSPSTV WDTALLSAAL QQAGVPHQHA SIMKASNYLL SKQQQKYGDW AIKNPDVTPG
GWGFSDTNTF VPDIDDTTAA LRAITPLAGT ENHFKHAWNK GVEWVLTMQN DDGGWSAFEK
NTDNYLLSFI PFKYEDRVLF DPSTADLTGR TLYFLGEYTT IPQESDVFRK AMEWLEYNQE
DNGSWYGRWG NCYIYGTWAA ITGLKAIGVS SDDPLIRRAV KWLLSIQNED GGWGESCESD
SKKRYIPLHH STPSQTAWAL DALISASENP TPKIEKGIHS LLRLLKADDW RTTYPTGAGI
PGGYYIHYHS YQYIWPLLAL SHYKNKYDET E

Sequence Number (ID): 33
Sequence Name: A0A0M9GQE0_V1
Length: 631
Molecule Type: AA

Features Location/Qualifiers:
- source, 1..631
> mol_type, protein
> organism, synthetic construct

Residues:
MKPVDKINEK LKDMINNLLD TQTEKGTWNF CFEGSIMTDA YMIILIRILE LTDEELLVKS
LVERIKSKQE SNGAWKVYPD EDGGNLSATI EGYFSLLHSG YVAKDAANMR KAERFIQENG
GLAESDWLTK MMLALTGQIK WPSIIKIIPI EIMLLPNWSP ISIYQFVGYA RAHWIPILIC
SNLNYSYLHS RTPNLIHLQG VGSDSEDQRF VEDRQHLQLY FKNALKKLTG SPEVLKRKAF
IKAEDYILER IEENGTLYSY FSASFFMVFA FLALGYDKKN PLIQNAFQGM KAYLCRNADH
AFIQNSPSTV WDTALLSAAL QQAGVPHQHA SIMKASNYLL SKQQQKYGDW AIKNPDVTPG
GWGFSDTNTF VPDIDDTTAA LRAITPLAGT ENHFKHAWNK GVEWVLTMQN DDGGWSAFEK
NTDNYLLSFI PFKYADRVLF DPSTADLTGR TLYFLGEYTT IPQESDVFRK AMEWLEYNQE
DNGSWYGRWG NCYIYGTWAA ITGLKAIGVS SDDPLIRRAV KWLLSIQNED GGWGESCESD
SKKRYIPLHH STPSQTAWAL DALISASENP TPKIEKGIHS LLRLLKADDW RTTYPTGAMI
PGGYYIHYHS YQYIWPLLAL SHYKNKYDET E

Sequence Number (ID): 34
Sequence Name: A0A1Y0CKF4_V1
Length: 629
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..629
> mol_type, protein
> organism, synthetic construct

Residues:
MNTENKINEK LKEMISTLLS KQSDNGAWNF CFEGSIMTDA YMIILIRTLE ITDEEVLVKD
LVERIKSRQS PNGAWKVYPD ENKGNLSATI EGYFSLLYSG YVGEEASYMR KAERFIRDNG
GLAKSDWLTK MMLALTGQIK WPSIIKIIPI EIMLLPRWSP ITIYQLVGYA RAHWIPILIC
SNLNKSFVAP QTPNISHLQE RLMDSENDRI LEEMQNLQLY FKHALKKLSK SPEILKKEAF
IKAENYIIER IEENGTMYSY FSASFFMVFA FLALGYDANH PLIRNAFQGM KSYLCRNAEQ
PFIQNSPSTV WDTALLTAAL QQAGVSYRHS SIMKANNYLL SRQHQKYGDW AVNNPDVIPG
GWGFSDINTF VPDIDDTTAA LRAITPLTQT NILYKEAWNK GVEWILSMQN GDGGWSAFEK
NMDNYLLSLI PFKYADRVLF DPSTADLTGR TLYFLGEYTT IPLESEIFQT AKEWFERNQE
ANGSWYGRWG NCYIYGTWAA ITGLKAIGVS NDDPIISRAV RWLLSVQNED GGWGESCASD
IKKRYIPLPH STPSQTAWAL DALISASDNP TSRIEVGIHA LLNILEANDG RSNYPTGAMI
PGGYYIHYHS YKYIWPLQAF SHYKNKYDL

Sequence Number (ID): 35
Sequence Name: A0A2S5GD65
Length: 626
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..626
> mol_type, protein
> organism, Jeotgalibacillus proteolyticus

Residues:
MEKIIKEEIE RIVYQLEGIQ SSDGTWRFCF EGSVMTDAYL IILIKTLQLQ EDKLVKDLAE
RIASKQTENG AWKLFKDDDG NLSATVEAYY ALLMANYKKK SDPTMKKAED FIIKNGGLSK
VSSLTKVMLA TTGQYSWSKI IPIPIEVILL PQSCPINFFD MVGYARVHLM PILVLANNKF
SMKTAHLNLE YLNQSRDEEQ DECFISIQSD DTRSLLSFIK QNVQKLIGLP NELNRMALDQ
AKLFMLHRIE PDGTLYSYFS STFLMIFSLL SLGFTKDDPI IEKAINGLKG LACNTEDHIH
IQNSPSTVWD TALITHSLLS SGVDVRSNFI QLPTHYLLRK QQYLYGDWSI HNLNSLPGGW
GFSDSNTMNP DVDDTTAALR AIKPTISQHP NLSQSWFRGL NWVLSMQNND GGWPAFEKNT
DKEILKLIPF DGSESASIDP STADLTGRTL EFLGNDARLT VQHPQIKRAV DWLKDHQESD
GSWYGRWGIT YIYGTWAAIT GMRAVGEKSH HPTIVKAVQW LEEIQNADGG WGESCNSDIE
KKYIPLGAST PSQTAWALDS LISVYDHPTV EIKKGIGCLI NLLKEKDWTY SYPTGAGLPG
SFYIYYHSYN YIWPLLSLSR YLQKYT

Sequence Number (ID): 36
Sequence Name: A0A2S5GD65_V1

Length: 626
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..626
> mol_type, protein
> organism, synthetic construct
Residues:
MEKIIKEEIE RIVYQLEGIQ SSDGTWRFCF EGSVMTDAYL IILIKTLQLQ EDKLVKDLAE
RIASKQTENG AWKLFKDDDG NLSATVEAYY ALLMANYKKK SDPTMKKAED FIIKNGGLSK
VSSLTKVMLA TTGQYSWSKI IPIPIEVILL PQSCPINFFD MVGYARVHLM PILVLANNKF
SMKTAHLNLE YLNQSRDEEQ DECFISIQSD DTRSLLSFIK QNVQKLIGLP NELNRMALDQ
AKLFMLHRIE PDGTLYSYFS STFLMIFSLL SLGFTKDDPI IEKAINGLKG LACNTEDHIH
IQNSPSTVWD TALITHSLLS SGVDVRSNFI QLPTHYLLRK QQYLYGDWSI HNLNSLPGGW
GFSDSNTMNP DVDDTTAALR AIKPTISQHP NLSQSWFRGL NWVLSMQNND GGWPAFEKNT
DKEILKLIPF DGAESASIDP STADLTGRTL EFLGNDARLT VQHPQIKRAV DWLKDHQESD
GSWYGRWGIT YIYGTWAAIT GMRAVGEKSH HPTIVKAVQW LEEIQNADGG WGESCNSDIE
KKYIPLGAST PSQTAWALDS LISVYDHPTV EIKKGIGCLI NLLKEKDWTY SYPTGAMLPG
SFYIYYHSYN YIWPLLSLSR YLQKYT

Sequence Number (ID): 37
Sequence Name: A0A2W0HKM8_V1
Length: 619
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..619
> mol_type, protein
> organism, synthetic construct
Residues:
MHELLNDTIR RMTDSLRSLQ ANDGSWRFCF ETGPTTDAYT IILLRSLDTK GDEWLIQQLT
ERLLAIQDTS GSWKLYPDQA EGHLSATVEA YFALLYSGYV SETAPNMKKA RSFIEKNGGL
HKAGPFTRAM LALNGQISWP RLFRLPVESL LIPQNAPVNL YDIVSYARVH IVPVLAAANK
GYVCQLPSTP DLSALGKRNE ESAAEETERL FSTVAAEIHK LAETPGRLKA KAYKKAERLM
LERIEPDGLY FSYITSTVLM VYALLALGYS KNDAVIQRAL TAIRNQVCLT STRSHIEFAT
STVWDTALLS HALQRSGVPS EDPMIAGAGR YLLNRQHTKY GDWAFNCSGT LPGGWGFSDI
NTFLPDIDDT TASLRAVKDL IEAMPEYRIT WFRGTDWVLK MQNTDGGWAA FEKDTVKRRL
TLLPFPAADR VLIDPSTADL TGRALEFLSG EANLLLPHPA VDRAVNWLEK NQEQNGSWYG
RWGICYIYGT WAALTGLSAA GYEKENKTVK RGVEWFNSIQ NEDGGWGESC KSDTAGRYVP
LGSSTPSQTA WAVDALIAVH SRPTEAIDHG IRYLITNAGR SDRYPTGAML PGDFYIYYHS
YNHIWPLLAL ANYKSRYSS

Sequence Number (ID): 38
Sequence Name: A0A3D9KM71_V1
Length: 631
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..631
> mol_type, protein
> organism, synthetic construct
Residues:
MGNVISEIDE EIQRLASNVV RQQYPDGSWR YCFENGISID AYTIILIRVL EIESEEALIR
RLHDRILGAQ QPEGHWQWYR DEQNGNLSLT VEAYCALLFS RFSHSSDAPL QKAERYIRSQ
GGLGKTANIL TRCMLAATGQ TKWPLSLTSI PLEMLLLPDS FPINFFEFSG YSRVHLAPML
IMAHRHYSAR TGDTPELDAL RTDRSPSEAR PSRGVREWLD RLQVGVSKLI GAPKALHESA
LAKAEKFMRD RIEADGTLYS YASCTFLMVL ALLALGYDKR HPLIAKAVRG LIGMRFRTEE
GTTIQNSPST VWDTALLAYA LQEAGATEHH PAVRNASSYL LGLQHRKPGD WTRHNPNPVP
GGWGFSETNT INPDVDDTTA ALRAIQKLAR SDSTYRESWN RGLNWVLSMQ NRDGGWPAFE
RNVDQQLLTL VAIEGAKSAA IDPSEADLTG RTLEYLGNFT GLGRKHGFVK KAVDWLVEHQ
EEDGSWYGRW GICYIYGTWA ALTGLTSVID SPERHESIRK GAQWLLQIQN DDGGWGESCS
SDRQRHYVPL GKSTPSQTAW AIDALVSVYP EPTPALNQGI RRLIALLHDN DWPTSYPTGA
MLPGHFYVRY HSYNSIWPLL ALSHYRNKYG K

Sequence Number (ID): 39
Sequence Name: A0A559J1A9
Length: 668
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..668
> mol_type, protein
> organism, Paenibacillus sp. N4
Residues:
MNRRANIESG IDRLIQQLLM AQAPDGSWRF CFESGTMTDS YMIIIVRVLQ LSEDELVKQL
SQRIVSRQHP EGYWSVYPNE TNGNLSSTVE AYYALLYSGT MKKDDPILLK AKAYILSNGG
MQQANSVLTK TMLAATGQRP WPRSYTVPIE FLLLPEWSPI SFYDIVGYAR VHIAPILIMS
SLPDTTIPEG APDLSDLILP NRSWEYESES FNHTDDIQQQ QCLYEDYFAY DFSNFQSTYN
SVIQSGEAHR SLLQNVKREL FQLLPSPQSV KQEARNKAES FMLDRIEPNG TLYSYASATF
LMIFALLALG YDRNHPRITK AIQGLKSFVC PSDKHWHIQN SPPTIWDTAL ISHAFQQAGL
PVQHGAIQRA GAYLLSRQQH KFGDWQFHNP NTPPGGWGFS DINTIIPDID DTTAALRAIN
KLASSNTSYA AAYDKGLQWL LSMQNDDGGW PAFEKNTNKT ILTWLPYDGA NAALTDPSTA
DLTGRTLEYL GSTAQLKLEH AFVRRGADWL MNHQQQDGSW YGKWGISYIY GTWAAVTGLA
AVGVDASNPA LVRAARWLSR IQNQDGGWGE SCESDRKKTY IPLHLSTPSQ TAWALDALIA
VSPQPTEEIE RGIQNLLYML QHPNKQSNTY PTGAGLPGNF YIYYHSYNYI WPLLTLANYK
RKYSPPLG

Sequence Number (ID): 40
Sequence Name: A0A559J1A9_V1
Length: 668
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..668
> mol_type, protein
> organism, synthetic construct
Residues:
MNRRANIESG IDRLIQQLLM AQAPDGSWRF CFESGTMTDS YMIIIVRVLQ LSEDELVKQL
SQRIVSRQHP EGYWSVYPNE TNGNLSSTVE AYYALLYSGT MKKDDPILLK AKAYILSNGG
MQQANSVLTK TMLAATGQRP WPRSYTVPIE FLLLPEWSPI SFYDIVGYAR VHIAPILIMS
SLPDTTIPEG APDLSDLILP NRSWEYESES FNHTDDIQQQ QCLYEDYFAY DFSNFQSTYN
SVIQSGEAHR SLLQNVKREL FQLLPSPQSV KQEARNKAES FMLDRIEPNG TLYSYASATF
LMIFALLALG YDRNHPRITK AIQGLKSFVC PSDKHWHIQN SPPTIWDTAL ISHAFQQAGL
PVQHGAIQRA GAYLLSRQQH KFGDWQFHNP NTPPGGWGFS DINTIIPDID DTTAALRAIN
KLASSNTSYA AAYDKGLQWL LSMQNDDGGW PAFEKNTNKT ILTWLPYDGA NAALTDPSTA
DLTGRTLEYL GSTAQLKLEH AFVRRGADWL MNHQQQDGSW YGKWGISYIY GTWAAVTGLA
AVGVDASNPA LVRAARWLSR IQNQDGGWGE SCESDRKKTY IPLHLSTPSQ TAWALDALIA
VSPQPTEEIE RGIQNLLYML QHPNKQSNTY PTGAMLPGNF YIYYHSYNYI WPLLTLANYK
RKYSPPLG

Sequence Number (ID): 41
Sequence Name: A0A5D4SXS0_V1
Length: 631
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..631
> mol_type, protein
> organism, synthetic construct
Residues:
MNPVDKINEK LKDMISNLLS KQKDNGVWNF CFEGSIMTDA YMIILIRILE LTDEEVLVKS
LVERIKSQQE SNGAWKVYPD EDGGNLSVTI EGYFSLLYSG YVAKNAGYMR KAERFIQENG
GLSKSDWLTK MMLAITGQIK WPSIIKIIPI EIMLLPSRSP ISIYQFVGYA RAHWIPILIC
SNLNYSYVHS RTPNLTHLQG GGSDDEDQRL LEDMQNLQLY FKNALKKLAG STENLKRKAF
IKAEDYILER IEENGTLYSY FSASFFMVFA LLALGYDKKN PLIGNAFQGM KAYLCRSADQ
VFIQNSPSTV WDTALLSAAL QQAGVPHQHA SIMKASNYLL SRQQQKYGDW AIKNPDVTPG
GWGFSDTNTF VPDIDDTTAA LRAITPLAGT ENHFKHAWNK GVEWVLTMQN DDGGWSAFEK
NTDNYLLSFI PSKYADRVLF DPSTADLTGR TLYFLGEYTT MPQESDVFQK AREWFEHNQE

ENGSWYGRWG NCYIYGTWAA ITGLKAIGVS SDNPLISRGV KWLLSIQNED GGWGESCESD
SKKRYVPLHH STPSQTAWAL DALISASEIP TPKIEKGMHS LLSLLEASDW RSIYPTGAMI
PGGYYIHYHS YKYIWPLHAL SHYKNKYYEI E

Sequence Number (ID): 42
Sequence Name: A0A5P9HJ69
Length: 623
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..623
> mol_type, protein
> organism, Bacillus sp. THAF10
Residues:
MSINEKIHSL VTELLQKQEE NGSWIFCFEG TIMTDAYMII LIRVLQRTDE EELVKQLVKR
IKSKQQLDNG AWKVFKDEEE GNLNATIEGY FSLLYSGYVS QSDPLMQKAE QFIKRKGGLT
STDWLTRVML ALTGQIQWPT IIKSIPIEIM LLPKWSPVNL YQLVGYNRAH WVPIIISSSK
NISISTSSTP DISHLQVRAP KLEITKGLQL IQHYVKGFVN KLADTPEVLR DRAFSKAEKY
ITNRIEENGT LYSYFSSSFF MVFAFLALGY DRTHPLIQHA FQGMKSYVYK DENMIHVENS
PSTVWDTSLL TAALMQAGVS SNQEAIQKAA SYLLTLQQTK YGDWAVKNPN VAPGGWGFSE
SNTFVPDIDD TTAALRVLAA FVDKDSRYLD GWNKGISWLL SMQNDDGGWS AFEKNTDNYL
LFMIPFSYED RVLFDPSTAD LTGRALYFLG ENTTIPTDDK AVRRAKEWLV KNQEEDGSWY
GRWGVCYIYG TWAAVTGLTA VGETLRSQAL QRAVQWLYKI QNEDGGWGES CKSDFVKQYV
PLHASTASQT AWALDALISA SDVPSPEMKR GIKALLRLLD NEDWREEYPT GAGLPGGFYI
HYHSYNYIWP LQTLSHYRNK FGE

Sequence Number (ID): 43
Sequence Name: A0A5P9HJ69_V1
Length: 623
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..623
> mol_type, protein
> organism, synthetic construct
Residues:
MSINEKIHSL VTELLQKQEE NGSWIFCFEG TIMTDAYMII LIRVLQRTDE EELVKQLVKR
IKSKQQLDNG AWKVFKDEEE GNLNATIEGY FSLLYSGYVS QSDPLMQKAE QFIKRKGGLT
STDWLTRVML ALTGQIQWPT IIKSIPIEIM LLPKWSPVNL YQLVGYNRAH WVPIIISSSK
NISISTSSTP DISHLQVRAP KLEITKGLQL IQHYVKGFVN KLADTPEVLR DRAFSKAEKY
ITNRIEENGT LYSYFSSSFF MVFAFLALGY DRTHPLIQHA FQGMKSYVYK DENMIHVENS
PSTVWDTSLL TAALMQAGVS SNQEAIQKAA SYLLTLQQTK YGDWAVKNPN VAPGGWGFSE
SNTFVPDIDD TTAALRVLAA FVDKDSRYLD GWNKGISWLL SMQNDDGGWS AFEKNTDNYL
LFMIPFSYAD RVLFDPSTAD LTGRALYFLG ENTTIPTDDK AVRRAKEWLV KNQEEDGSWY
GRWGVCYIYG TWAAVTGLTA VGETLRSQAL QRAVQWLYKI QNEDGGWGES CKSDFVKQYV
PLHASTASQT AWALDALISA SDVPSPEMKR GIKALLRLLD NEDWREEYPT GAMLPGGFYI
HYHSYNYIWP LQTLSHYRNK FGE

Sequence Number (ID): 44
Sequence Name: F5LIR7_V1
Length: 649
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..649
> mol_type, protein
> organism, synthetic construct
Residues:
MRIMREVAHA IHNLSEWLLG KQEKDGTWKF CYDNGISTSA YLIILLRTLE TENADDEALI
RRLHDSIVRN QQPDGSWKLY ADEKDGNLAA SVEAYYALLF SGYSDDSSPS LVNARAFIRS
RGGITGVTNV LTRVILAATG QIPWPAAYSI PLEFLLLPAS APLSFFDLSA YARVHLAPVL
LMADRNFSVR TSRTPDLSAL GVRRDEDGKS GASSGSVPGS SAQPERGPGG LLAAIQAGIS
QLAALPGQLH GSAVKKAERF MLERIEPDGT LYSYASSTCL LVFALLSLGF ERRHPTIAQA
VKGLKAMLCV SEGRLLMQNA PPTVWDTALI AYALQEAGVR PEAPGIRKAA SYLLAKQQRK

IGDWGRKVSH PVPGGWGFSP SNTRNPDVDD TTAALRAVKF LRSEGTAGRE AWNRGLYWII
SMQNQDGGWA AFEKDTDEKL LTLLPMEGAK HAAIDPSTAD LTGRTLEFLG STAGLGVKHV
WIRRGADWLV ANQEKDGSWY GRWGICYLYG TWAALTGLAA VGLEPDHPAV AKGVRWLLSV
QNPDGGWGES CASDIVGRYE SLGASTPSQT AWALDALIAV HPRPSAAIDR GIQRLVAALD
ERDWTSLYPT GAMLPGSFYN TYESYRYIWP LLALSHYRNK YGEKSVKPE

Sequence Number (ID): 45
Sequence Name: MCH2311119_V1
Length: 670
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..670
> mol_type, protein
> organism, synthetic construct
Residues:
MESQVSHKLN KNLDSAIKKT QDYLFREQYD EGFWWGELES NPTMEAEFIL LHYFLGIRDE
KKFIKLSNHI KNQQREDGTW GQYYGAPGDL STSVECYLAL KIAGYSENDQ SMLKAKDFIV
SKGGIEETRV FTKIWLSLVD QWKWEGVPIM PAELILLPNW SPINIYEFSS WARSTIVPLL
ILMDKKPVRP LPKNLRVDEL FCDDVNNIDY SVKSPSIQIG WENFFYATDQ VLRLLDKLPI
KPTRDLALKK SEEWILQHQE ADGSWGGIQP PWVYSIMALY TAGYGLDHPV IDKALQGFKA
FEIEDQFSLR VQACVSPIWD TGLAIISLLD SGIKTDDDRI QKAGQWLINK QIKSEGDWQV
KANNVRSGGW AFEFENEHYP DIDDAAVVAT ALHKIDLTDE YGGLDKKSKS IKRCVQWIEG
MQSKNGGWAS FDKDNMRSFI ARIPFSDAGE TIDPPSVDVT AHVLELLGTL DAKKHATVIA
KALDYVLLEQ EQDGSWFGRW GVNYIYGIGS VLPALRAIGI DPSHEAMSKA TKWLEDHQNK
DGGWGETPAS YVDPSLHGKG PSTASQTAWS LISLIAADKG DSSHVLKGVN YLLSNQNEDG
SWDEPEFTGT MFPGYGIGMR PDISEQDDSK QHDIALPAGF MINYHMYRIY WPLCALGRFR
AWNVNRDSHH

Sequence Number (ID): 46
Sequence Name: NQW16228_V1
Length: 688
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..688
> mol_type, protein
> organism, synthetic construct
Residues:
MTSTLNRRSD IVPPVRLDHP PVPERAGGAA GEEFGDLLDS AIGRTQDYLF EHQVSDGYWW
GELESNPTME AEYVFLTHFL GLRDDVRWKK VQNYILSVQR PGGGWNQYHG APNDLSTSCE
CYLALKMTGL PASDPRMQQA REFILRKGGM EQTRVFTKIW FSLLGEWDWG GVPFLPPELM
LLPNRIPFNI YQFAMWSRTT IVPMSVLLSS KPIHPVAEEA TIDELYLNGR ENADYSMPTP
SGLGIERLMY AGDRILRLSN LLPWNPARGR ALRMAEKWIV NHQEKDGSWG GIQPPWVYSL
MALNELGYSN SHEVIKKGME GFELYGIERE DTWRLQASMS PLWDTCLSIN ALIDSGIEPD
HPAIVRAADY LLDRQASSPG DWQVKARNVE PGGWAFEFSN ETYPDTDDAA EVLLAIGAAG
VTDSARRDDS IKRGVNWVLA MQSKNGGWGA YDKDNTSTLV TKMPFFDAGE TIDPPSVDVT
AHIVEMLAKL RFPTDTPEIN AALDYIWKEQ EEDGCWFGRW GVNYVYGTAA VLPALESLGI
DQNDPRIQQA ADWLEMHQNS DGGWGESCAS YANPSLRGQG ASTASQTAWA LMGLISTGRA
AGESAKMGVE YLLGTQLADG SWTEDEYTGT MFPGYGIGER KFTGLETEDH DLISEELPAG
FMIKYHMYRI YWPLMALGRY RTCLSSGS

Sequence Number (ID): 47
Sequence Name: OYT72085.1
Length: 658
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..658
> mol_type, protein
> organism, Chloracidobacterium sp. CP2_5A
Residues:
MTGFAPRFVQ PVVESPIAPA PRAARPAPAS AASVGAAIAR AQSYLLSRQY PEGYWWAELE
ANVTLTAEYV FLHKILGTDG ARARQLGKIR TYLRRQQRDH GGWELYYGDG GELSTSVEAY

FALKLLGDAP DAPHMARARD FILARGGVAK ARVFTKIHLA LFGAFPWEGC PTLPPWIMLL
PDWFPFTIYE LASWARSSTV PLLLVGDKKP VVAVPGGAAD ELYAEGRASA DLALPNPDGM
LSLGGAFIAF DKALKLMERL NFSPRKAEAL ALAERWTLDH QDESGDWGGI IPAMLNSLLG
LHCRGYAPDH PAMRKGIEAV ERFCIETEDE FHTQPCVSPV WDTGLTILAL LDSGLPNDHP
ALVKAGEWLL SKQILRDGDW RFKNKTGPAG GWAFEFWNDF FPDVDDTAVV TMALHRLKLP
DEAEKQRRLK LATEWTLSMQ SKNGGWGAFD VDNDLAILNE IPYGDLKAMI DPPTADLTGH
ILEMLGVTGY PAPRERVERA IAFIKSQQEP EGCWWGRWGV NYIYGTHMVI CGLVALGLNP
REAFIMRGTQ WLNSCQNEDG GWGETCASYG DRRLMGIGAS APSQTAWALL GLIAGGEGKS
DCARRGIEYL VTRQNDDGGW TEAAFTGTGF PNHFYMNYHF YRHYFPLMAL GRYRPFAK

Sequence Number (ID): 48
Sequence Name: OYT72085_V1
Length: 658
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..658
> mol_type, protein
> organism, synthetic construct
Residues:
MTGFAPRFVQ PVVESPIAPA PRAARPAPAS AASVGAAIAR AQSYLLSRQY PEGYWWAELE
ANVTLTAEYV FLHKILGTDG ARARQLGKIR TYLRRQQRDH GGWELYYGDG GELSTSVEAY
FALKLLGDAP DAPHMARARD FILARGGVAK ARVFTKIHLA LFGAFPWEGC PTLPPWIMLL
PDWFPFTIYE LASWARSSTV PLLLVGDKKP VVAVPGGAAD ELYAEGRASA DLALPNPDGM
LSLGGAFIAF DKALKLMERL NFSPRKAEAL ALAERWTLDH QDESGDWGGI IPAMLNSLLG
LHCRGYAPDH PAMRKGIEAV ERFCIETEDE FHTQPCVSPV WDTGLTILAL LDSGLPNDHP
ALVKAGEWLL SKQILRDGDW RFKNKTGPAG GWAFEFWNDF FPDVDDTAVV TMALHRLKLP
DEAEKQRRLK LATEWTLSMQ SKNGGWGAFD VDNDLAILNE IPYGDAKAMI DPPTADLTGH
ILEMLGVTGY PAPRERVERA IAFIKSQQEP EGCWWGRWGV NYIYGTHMVI CGLVALGLNP
REAFIMRGTQ WLNSCQNEDG GWGETCASYG DRRLMGIGAS APSQTAWALL GLIAGGEGKS
DCARRGIEYL VTRQNDDGGW TEAAFTGTMF PNHFYMNYHF YRHYFPLMAL GRYRPFAK

Sequence Number (ID): 49
Sequence Name: WP_092048487_V1
Length: 710
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..710
> mol_type, protein
> organism, synthetic construct
Residues:
MNSGHMRNGS SGVASRLPRE STSRLDSSVK SAVSQARNWL LSEQNSEGYW LGELQGDTIL
ESEYILLLAW MGKSNTPIVQ ECANYIRQQQ LPEGGWAMFP GGPLEISSSV KAYWTLKIAG
DDPQAEHMQR ACAAIRAAGG AERVNSFTRY YMALLGIISY RQCPAVPPEL MLLPKWMPFN
IYEMSSWSRT IIVPLSLLWA FQPKTTLPRS QKIDELFLNS PEKLPVVMPP SGQLDKLKQQ
TWVPWDRIFR GIDVTWKFFE ALRMKPFRER AVRLATKWIV KRFEKSDGLG AIFPPIIWSV
IALRCLGHDE SSPMVQAALK ELEKLTIREG NTARLEPCRS PVWDTAIAVN ALRDAGVPAH
HPQLVRAVNW LLSKEVRSPG DWTVNHPDVE PGGWYFEFNN EFYPDVDDTI MVSMALARCL
PGDQHSNWSA SLLSKQGVQQ RSDFDLAIVI AGQTDAPERA VSDVERMQPM IAALRRAVKW
TVAMQSRNGG WGAFDADNDR EILTRVPFAD ANAMIDPPTA DITARVLEMF GRLGLTSREP
IFEKALKFVW DEQEPDHCWF GRWGVNYIYG TWQVLVGLTE FGVPRDDSRL QAAARWLKEK
QQSDGGWGET AQSYDEPALR GTGVTTPSQT AWAVLGLIAA GEGRSIAARR GIEFLLSRQT
EQGTWDETEF TGTMFPRVFY LRYHLYRHYF PLMALGRYAA QFESEDSEAY

Sequence Number (ID): 50
Sequence Name: WP_051467941.1
Length: 225
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..225
> mol_type, protein
> organism, Actinomadura oligospora

Residues:
MDVADVLSLG SGLAWTATYL LIIWTNWREK TYGMPIAALG ANLGWEFLFS FVRPGDGMQL
VVNYVWFGFD VAILALVVAY GPREFRFLPR WGFLAMLASV LVMGYLGVDL VSRQFDHGLA
TFAAFGQNLM MSGLFLSMLI ARGSTRGQSV WIALTKGVGT ALASGASWIW AQDEPWRHGS
LLPYLMITTA VLDLAYLVAV YAVARREAGG SASAPLRLNR VPEPV

Sequence Number (ID): 51
Sequence Name: WP_234754442.1
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, Arthrobacter ramosus
Residues:
MNLFLTILSG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VINIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 52
Sequence Name: WP_190963420.1
Length: 215
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..215
> mol_type, protein
> organism, Desmonostoc muscorum
Residues:
MGTYLMLGSG AFWILTYILL IERGFKDQTY GMPLVALCAN LSWEFIFSFI HPHQPPQLQI
NIVWLMLDLI ILYGFFKFGQ SELKDIPNKL FYPVFILTLF TSFCCVLLIT DEFQDWSGAY
TAFGQNLLMS ILFIDMLTKR NTVRGQSIFI AIFKMIGTLL ASIGFYINNP IQGRSLLFIF
LYTAIFVFDL IYVGMIAMKI KRFREKKLNH ALTRQ

Sequence Number (ID): 53
Sequence Name: WP_093699331.1
Length: 217
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..217
> mol_type, protein
> organism, Streptomyces sp.
Residues:
MHTAFLLGTG GFWTVAYVLL IRTGLRERTF GMPVVAFATN ISWEFMFAFV RPPTGVMHVV
NIVWFCFDVA IGYTLVRFGR AEFPYLPRSL FLPALLALLA LAYPGMNYAS ERFDEGAGAV
TAFGSNLAMS GMFLAMLAAR RGTRGQSAGI ALAKLLGTVC ASLSMLTDPG LEPRHDNALM
YYLYVGCFLL DAAYLGAVLA VRRAERAVAP VTVTVPV

Sequence Number (ID): 54
Sequence Name: WP_067007865.1
Length: 224
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..224
> mol_type, protein
> organism, Streptomyces cellostaticus
Residues:
MAAVHTAFLL GTGLFWTAAY VLLIRTGLRA RTFGMPVVAF ATNISWEFMF AFVRPPSGVM
HVINIVWFCF DLAIGYTVVR FGRAEFPYLP DRLFLPALAV LLALAYPGMN YVSEQFDEGV
GAITAFGSNL AMSGMFLAML AARRGTRGQS VGIAVTKLLG TACASLALLT DPDGDPRYDN
ALMYYFYIGC FLLDLAYAYA VFAVGRAERT TGSAQVPAQG ALQR

Sequence Number (ID): 55
Sequence Name: WP_220206969.1
Length: 220
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..220
> mol_type, protein
> organism, Reticulibacter mediterranei
Residues:
MFVLLMLGSG LFWTITYILI IRRSILDRTY GMPLAALCAN ISWEFIFSFI LPSSSIQRIV
NIIWFVLDAG ILVCFLRYGR NEFANLSKWI FFTTFGLTLA TSFGAVLLVT LEFHDSGAYS
AFGQNLMMSA LFILMLYRRG SLRGQSIAIA VTKLLGTALA SLAFFLYTTI SHNSVLLPFL
YVSILVYDMI YVAMVYKQQR AAKQTSIEAT ASASHVELSL

Sequence Number (ID): 56
Sequence Name: WP_234754442.1_S9C
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILCG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VINIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 57
Sequence Name: WP_234754442.1_S9M
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILMG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VINIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 58
Sequence Name: WP_234754442.1_S9T
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILTG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VINIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 59
Sequence Name: WP_190963420.1_S9C
Length: 215
Molecule Type: AA

Features Location/Qualifiers:
- source, 1..215
> mol_type, protein
> organism, synthetic construct
Residues:
MGTYLMLGCG AFWILTYILL IERGFKDQTY GMPLVALCAN LSWEFIFSFI HPHQPPQLQI
NIVWLMLDLI ILYGFFKFGQ SELKDIPNKL FYPVFILTLF TSFCCVLLIT DEFQDWSGAY
TAFGQNLLMS ILFIDMLTKR NTVRGQSIFI AIFKMIGTLL ASIGFYINNP IQGRSLLFIF
LYTAIFVFDL IYVGMIAMKI KRFREKKLNH ALTRQ

Sequence Number (ID): 60
Sequence Name: WP_190963420.1_S9M
Length: 215
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..215
> mol_type, protein
> organism, synthetic construct
Residues:
MGTYLMLGMG AFWILTYILL IERGFKDQTY GMPLVALCAN LSWEFIFSFI HPHQPPQLQI
NIVWLMLDLI ILYGFFKFGQ SELKDIPNKL FYPVFILTLF TSFCCVLLIT DEFQDWSGAY
TAFGQNLLMS ILFIDMLTKR NTVRGQSIFI AIFKMIGTLL ASIGFYINNP IQGRSLLFIF
LYTAIFVFDL IYVGMIAMKI KRFREKKLNH ALTRQ

Sequence Number (ID): 61
Sequence Name: WP_190963420.1_S9T
Length: 215
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..215
> mol_type, protein
> organism, synthetic construct
Residues:
MGTYLMLGTG AFWILTYILL IERGFKDQTY GMPLVALCAN LSWEFIFSFI HPHQPPQLQI
NIVWLMLDLI ILYGFFKFGQ SELKDIPNKL FYPVFILTLF TSFCCVLLIT DEFQDWSGAY
TAFGQNLLMS ILFIDMLTKR NTVRGQSIFI AIFKMIGTLL ASIGFYINNP IQGRSLLFIF
LYTAIFVFDL IYVGMIAMKI KRFREKKLNH ALTRQ

Sequence Number (ID): 62
Sequence Name: WP_234754442.1 T7M
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLMILSG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VINIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 63
Sequence Name: WP_234754442.1 S51T
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:

MNLFLTILSG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH TLSTGLSVQG
VINIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 64
Sequence Name: WP_234754442.1 N63Q
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILSG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VIQIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 65
Sequence Name: WP_234754442.1 N63F
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILSG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VIFIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 66
Sequence Name: WP_234754442.1 N63Y
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILSG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VIYIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 67
Sequence Name: WP_234754442.1 N63A
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILSG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VIAIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 68
Sequence Name: WP_234754442.1 N63M
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILSG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VIMIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 69
Sequence Name: WP_234754442.1 S9V
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILVG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VINIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 70
Sequence Name: WP_234754442.1 S9A
Length: 219
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..219
> mol_type, protein
> organism, synthetic construct
Residues:
MNLFLTILAG VAWTTVYICA IRIGFRDKTY AIPAAALGLN FAWEVIYSVH SLSTGLSVQG
VINIAWALAD VAIVYTFFAF GRRELPGFLT RPLFIGWAVL LGLASFAVQW LFIAEFDWDP
ASRYAAFLQN LLMSGLFIAM FAARRSLRGQ SLVIAVAKWI GTLAPTITFG VLESSLFILG
IGVLCSIFDL TYIGLLLWSK KNPGALSRDR DSGGLPAVP

Sequence Number (ID): 71
Sequence Name: A0A2P1DP74.1 (macJ)
Length: 258
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..258
> mol_type, protein
> organism, Penicillium terrestre
Residues:
MCFFALEEWA AANRDYENTP APYWHVKSVP DGFTAISGIL WSISYILMAK KAFKDRSYAM
PLHCLCLNIT WEAVYGFVYG PGLLNQVVFA QWMIVDVVLF YAILRSAPYA WKQSPLVAQH
LAGIIVVGCV ICLWLHLAIA ATFIPSIGRQ VVFMTAWPMQ VLINFSSIAQ LLSRGNTLGH
SWGIWWTRML GTIAAACCFF WRIHYWPERF GYAWTPYGKF LLLGSIGSDM VYAAVYVYVQ
RIEKQLDSLV NTKAQKAR

Sequence Number (ID): 72
Sequence Name: XP_018029969.1
Length: 255
Molecule Type: AA

Features Location/Qualifiers:
- source, 1..255
> mol_type, protein
> organism, Paraphaeosphaeria sporulosa

Residues:
MGLFALEEWA QANADYNNDT PPYWHAKIVP DLFTAISGIL WSVSYILMTL KGYKDRSYAM
PIYCLCLNIT WEFVFGFIYG PGLVNQIVFA QYMVVDVFLF HSILKFGPNE WRAHPLVARN
LSWIIGVGCA VCLGLHLVLA KTFVPVIGRQ VIFFTAWPMQ HMISLGCVAQ VLSRGHDAGQ
SMAIWWTRFL GTVTAGCCFY WRIYFWPERF GYAWTPYGAL LLVGSHVLDL AFPFALAYVR
KHGEGRQEKV NGKAA

Sequence Number (ID): 73
Sequence Name: KAG0152682.1
Length: 258
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..258
> mol_type, protein
> organism, Penicillium digitatum

Residues:
MGFFALEEWA AANRDYDNTP APYWHAKSVP DGFTAISGIL WSISYILMAK KAFKDRSYAM
PLHCLCLNIT WEAVYGFIYG PGLLNQVVFA QWMIVDVILF YAIVRSAPSA WKQSPLVAQH
LAGIIVVGCV VCLWLHLAIA ATFIPSIGRR VVFMTAWPMQ VLINLSSIAQ LLSRGNTLGH
SWGIWWTRML GTIAAACCFF WRVYYWPERF GYAWTPYGQF LLLGSIGSDV VYAVVYIYVH
RFANPLDTRV KIEAKKSG

Sequence Number (ID): 74
Sequence Name: OKH29475.1
Length: 343
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..343
> mol_type, protein
> organism, Nostoc calcicola

Residues:
MEDVTKQERG DADTIESAVT FLLTARNSEG WWIDFQLAAG LSDEWVTGYV GAMLANIRDI
RIPEAVSTAW NLLNSRRHRA NGKWGYNRLP PGDADSTGWV LQLAHAIGES NSERARQAMQ
SLAAHQRPDG GICTYESEES IRAFIHASPE IGFAGWCGSH TCVSAAIAAL PEYRFQLQDY
LRSTQQNDGS WLAYWWQDPE YVTALAAEAI AACYPNSDCI TSAVVWGMNR LNSQGFVATS
DRPSGSPFAT AWCLRLLILR RQDTPVQAAI AKATDWLLAQ QQPNGSWISS ARLQVPLPDD
LNPNKFNQWI YHGTIQGSLV FDKHCVFTTA TVLQALHRSL FGK

Sequence Number (ID): 75
Sequence Name: NEQ07043.1
Length: 358
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..358
> mol_type, protein
> organism, Moorea sp.

Residues:
MKVVPEQTAK SAVKSAIDRA IVFLLLSRDT QGWWKDFFLP AGASDAWVTG YVGTVLAHSQ
NSHAWKAAEK AWTLLAQQCH DREGWGYHAG VPADADSTLW GLQLAQALGR EGEESSHRGH
RFLRRHLKPD GGVTTYEQEA TIRNYIGLPP GLVPFTAWCH SHTCVTAAAA SLGEWREIVA
PYLLSQQQAD GSWHSYWWFE DEYCTALALT AVESQESIER AVKWGCHRLL YWLEASQPSE
FAIAWCLQIL SRDSTPSTQQ LVERGVKFLL QRQHSNGSWQ PSARLRVPRP DNFNPKSVKD
WQLWTGKFSG SVTLKNVLAN TFNIYSLDRQ SIFTTATVLY ALQSVTANAQ LRQQEVGM

Sequence Number (ID): 76
Sequence Name: MstE
Length: 366

Molecule Type: AA
Features Location/Qualifiers:
- source, 1..366
> mol_type, protein
> organism, Scytonema sp.
Residues:
MTLQPLENST RQEKLLYPKL NQLSNSINAA VAFLLEARNL EGWWQDFNFP QAASIGDEWV
TAYVGTMLAT LPYAHVHEAL MQAWELLKIR DHRPTGEWGY NYILCGDADT TGWALQLAAA
VGASDSERAQ QARAALATHL QPNGGIATFA EESIRAYIKV PDLANVSFQG WCGAHTCVSA
AVAALPEFRS RLHDYLRVTQ TSQGNWEGYW WSDHEYTTAL TAEALAAGGQ AADQPSIEQA
VAWGLKRLCP QGFVATSKHP NGSTFATAWC LRLLLLNTVD AEVKAARAAA IGWLLEQQRP
NGSWVSSAYL RIPYPFDRNP NQFPHWRYYD EIEGDKRFEG SIIFDHNSIF TTATVVNSLV
KAAPML

Sequence Number (ID): 77
Sequence Name: DmtA1
Length: 250
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..250
> mol_type, protein
> organism, Streptomyces youssoufiensis
Residues:
MRPHSSKAPR GDMVDGIRMA LAAVCGLGWT IAYVLAVRTG VRDKTYCIPL VALAMNICWE
FQFVFFRSAE HMSGSNSEVE GAEVFIGIIW LIVDCGLLYT VFRFGPNEFP YLPRRVFYAG
FIVVLGLAYA GIEVLSREFD DGDVVLTSFG MNVAMSGLFL AMLAARQSSR GQSMGIAVAK
FVGTTSTCIA WFFDTSVYPG PWLPYCTVAC VLLDVAYIAA LSAVLRKERG RSQLAGDRPE
IRDIERNASL

Sequence Number (ID): 78
Sequence Name: AAcSHC_M132R_A224V_I432T
Length: 631
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..631
> mol_type, protein
> organism, synthetic construct
Residues:
MAEQLVEAPA YARTLDRAVE YLLSCQKDEG YWWGPLLSNV TMEAEYVLLC HILDRVDRDR
MEKIRRYLLH EQREDGTWAL YPGGPPDLDT TIEAYVALKY IGMSRDEEPM QKALRFIQSQ
GGIESSRVFT RRWLALVGEY PWEKVPMVPP EIMFLGKRMP LNIYEFGSWA RATVVALSIV
MSRQPVFPLP ERARVPELYE TDVPPRRRGA KGGGGWIFDA LDRVLHGYQK LSVHPFRRAA
EIRALDWLLE RQAGDGSWGG IQPPWFYALI ALKILDMTQH PAFIKGWEGL ELYGVELDYG
GWMFQASISP VWDTGLAVLA LRAAGLPADH DRLVKAGEWL LDRQITVPGD WAVKRPNLKP
GGFAFQFDNV YYPDVDDTAV VVWALNTLRL PDERRRRDAM TKGFRWIVGM QSSNGGWGAY
DVDNTSDLPN HTPFCDFGEV TDPPSEDVTA HVLECFGSFG YDDAWKVIRR AVEYLKREQK
PDGSWFGRWG VNYLYGTGAV VSALKAVGID TREPYIQKAL DWVEQHQNPD GGWGEDCRSY
EDPAYAGKGA STPSQTAWAL MALIAGGRAE SEAARRGVQY LVETQRPDGG WDEPYYTGTG
FPGDFYLGYT MYRHVFPTLA LGRYKQAIER R

Sequence Number (ID): 79
Sequence Name: A0A1H2R2P0_9BACL
Length: 633
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..633
> mol_type, protein
> organism, Alicyclobacillus hesperidum
Residues:
MTKQLAEIPA YMQTLDNGVE YLLSRQHEEG YWWGPLLSNV TMEAEYVLLC HCLGKVDKGR
LEKIKTYLLH EQREDGTWAQ YPGGPQDLDT TIEAYVALKY IGLSPDDERM QKALAFIQSQ

GGIESARVFT RLWLAVVGEY PWRKLPVVPP EIMFLGKNMP LNIYDFGSWA RPTIVALTIV
MSRRAVFPLP AHAKVPELFE TNVPPRRRAA KGGNSSLFLS IDKLLQGYQN GSFHPFRKAA
EQRAIEWLIE HQAGDGSWGG IQPPWFYALL ALKVMNMTNH PAFIKGWEGL ELYGLELEYG
GWMFQASISP VWDTGLSILA LRAAGLAPDE PALVKAGKWL LDHRIATKGD WAVRRPNAKP
GGWAFQFDNP HYPDVDDTAV VVWALNGLKL PNEAERRDAM TAGFRWLTAM QSSNGGWGAY
DVDNNKELPN RIPFCDFGEV IDPPSEDVTA HVLECFGSFG YDEAWKVVAR AVNYLKREQK
PDGSWYGRWG VNYIYGIGAV VPALKSVGVD MKEPFVQKAL DWLVAHQNED GGWGEDCRSY
VDERFAGVGP STPSQTAWAL MALIAGGRVQ ADAVSRGVAY LVRTQRSDGG WDEPYYTGTG
FPGDFYLGYT LYRHIFPVMA LGRYKDALGR LTR

Sequence Number (ID): 80
Sequence Name: A0A1H2R2P0_9BACL_F437A G600M
Length: 633
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..633
> mol_type, protein
> organism, synthetic construct
Residues:
MTKQLAEIPA YMQTLDNGVE YLLSRQHEEG YWWGPLLSNV TMEAEYVLLC HCLGKVDKGR
LEKIKTYLLH EQREDGTWAQ YPGGPQDLDT TIEAYVALKY IGLSPDDERM QKALAFIQSQ
GGIESARVFT RLWLAVVGEY PWRKLPVVPP EIMFLGKNMP LNIYDFGSWA RPTIVALTIV
MSRRAVFPLP AHAKVPELFE TNVPPRRRAA KGGNSSLFLS IDKLLQGYQN GSFHPFRKAA
EQRAIEWLIE HQAGDGSWGG IQPPWFYALL ALKVMNMTNH PAFIKGWEGL ELYGLELEYG
GWMFQASISP VWDTGLSILA LRAAGLAPDE PALVKAGKWL LDHRIATKGD WAVRRPNAKP
GGWAFQFDNP HYPDVDDTAV VVWALNGLKL PNEAERRDAM TAGFRWLTAM QSSNGGWGAY
DVDNNKELPN RIPFCDAGEV IDPPSEDVTA HVLECFGSFG YDEAWKVVAR AVNYLKREQK
PDGSWYGRWG VNYIYGIGAV VPALKSVGVD MKEPFVQKAL DWLVAHQNED GGWGEDCRSY
VDERFAGVGP STPSQTAWAL MALIAGGRVQ ADAVSRGVAY LVRTQRSDGG WDEPYYTGTM
FPGDFYLGYT LYRHIFPVMA LGRYKDALGR LTR

Sequence Number (ID): 81
Sequence Name: AAcSHC_F437A_G600M
Length: 631
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..631
> mol_type, protein
> organism, synthetic construct
Residues:
MAEQLVEAPA YARTLDRAVE YLLSCQKDEG YWWGPLLSNV TMEAEYVLLC HILDRVDRDR
MEKIRRYLLH EQREDGTWAL YPGGPPDLDT TIEAYVALKY IGMSRDEEPM QKALRFIQSQ
GGIESSRVFT RMWLALVGEY PWEKVPMVPP EIMFLGKRMP LNIYEFGSWA RATVVALSIV
MSRQPVFPLP ERARVPELYE TDVPPRRRGA KGGGGWIFDA LDRALHGYQK LSVHPFRRAA
EIRALDWLLE RQAGDGSWGG IQPPWFYALI ALKILDMTQH PAFIKGWEGL ELYGVELDYG
GWMFQASISP VWDTGLAVLA LRAAGLPADH DRLVKAGEWL LDRQITVPGD WAVKRPNLKP
GGFAFQFDNV YYPDVDDTAV VVWALNTLRL PDERRRRDAM TKGFRWIVGM QSSNGGWGAY
DVDNTSDLPN HIPFCDAGEV TDPPSEDVTA HVLECFGSFG YDDAWKVIRR AVEYLKREQK
PDGSWFGRWG VNYLYGTGAV VSALKAVGID TREPYIQKAL DWVEQHQNPD GGWGEDCRSY
EDPAYAGKGA STPSQTAWAL MALIAGGRAE SEAARRGVQY LVETQRPDGG WDEPYYTGTM
FPGDFYLGYT MYRHVFPTLA LGRYKQAIER R

Sequence Number (ID): 82
Sequence Name: AAcSHC_wt
Length: 631
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..631
> mol_type, protein
> organism, Alicyclobacillus acidocaldarius
Residues:

MAEQLVEAPA YARTLDRAVE YLLSCQKDEG YWWGPLLSNV TMEAEYVLLC HILDRVDRDR
MEKIRRYLLH EQREDGTWAL YPGGPPDLDT TIEAYVALKY IGMSRDEEPM QKALRFIQSQ
GGIESSRVFT RMWLALVGEY PWEKVPMVPP EIMFLGKRMP LNIYEFGSWA RATVVALSIV
MSRQPVFPLP ERARVPELYE TDVPPRRRGA KGGGGWIFDA LDRALHGYQK LSVHPFRRAA
EIRALDWLLE RQAGDGSWGG IQPPWFYALI ALKILDMTQH PAFIKGWEGL ELYGVELDYG
GWMFQASISP VWDTGLAVLA LRAAGLPADH DRLVKAGEWL LDRQITVPGD WAVKRPNLKP
GGFAFQFDNV YYPDVDDTAV VVWALNTLRL PDERRRRDAM TKGFRWIVGM QSSNGGWGAY
DVDNTSDLPN HIPFCDFGEV TDPPSEDVTA HVLECFGSFG YDDAWKVIRR AVEYLKREQK
PDGSWFGRWG VNYLYGTGAV VSALKAVGID TREPYIQKAL DWVEQHQNPD GGWGEDCRSY
EDPAYAGKGA STPSQTAWAL MALIAGGRAE SEAARRGVQY LVETQRPDGG WDEPYYTGTG
FPGDFYLGYT MYRHVFPTLA LGRYKQAIER R

Sequence Number (ID): 83
Sequence Name: Gmo_SHC_F460A_G623M
Length: 685
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..685
> mol_type, protein
> organism, synthetic construct
Residues:
MSPADISTKS SSFQRLDNML PEAVSSACDW LIDQQKPDGH WVGPVESNAC MEAQWCLALW
FLGQEDHPLR PRLAQALLEM QREDGSWGIY VGADHGDINT TVEAYAALRS MGYAADMPIM
AKSAAWIQQK GGLRNVRVFT RYWLALIGEW PWDKTPNLPP EIIWLPDNFI FSIYNFAQWA
RATMMPLTIL SARRPSRPLL PENRLDGLFP EGRENFDYEL PVKGEEDLWG RFFRAADKGL
HSLQSFPVRR FVPREAAIRH VIEWIIRHQD ADGGWGGIQP PWIYGLMALS VEGYPLHHPV
LAKAMDALND PGWRRDKGDA SWIQATNSPV WDTMLAVLAL HDAGAEDRYS PQMDKAIGWL
LDRQVRVKGD WSIKLPDTEP GGWAFEYAND KYPDTDDTAV ALIALAGCRH RPEWRERDIE
GAISRGVNWL LAMQSSSGGW GAFDKDNNRS ILTKIPFCDA GEALDPPSVD VTAHVLEAFG
LLGISRNHPS VQKALAYIRS EQERNGAWFG RWGVNYVYGT GAVLPALAAI GEDMTQPYIV
RACDWLMSVQ QENGGWGESC ASYMDINAVG HGVATASQTA WALIGLLAAK RPKDREAIAR
GCQFLIERQE DGSWTEEEYT GTMFPGYGVG QAIKLDDPSL PDRLLQGAEL SRAFMLRYDL
YRQYFPVMAL SRARRMMKED ASAAA

Sequence Number (ID): 84
Sequence Name: Gmo_SHC_Q54E_M184I_V45L_T326S_F624Y
Length: 685
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..685
> mol_type, protein
> organism, synthetic construct
Residues:
MSPADISTKS SSFQRLDNML PEAVSSACDW LIDQQKPDGH WVGPLESNAC MEAEWCLALW
FLGQEDHPLR PRLAQALLEM QREDGSWGIY VGADHGDINT TVEAYAALRS MGYAADMPIM
AKSAAWIQQK GGLRNVRVFT RYWLALIGEW PWDKTPNLPP EIIWLPDNFI FSIYNFAQWA
RATIMPLTIL SARRPSRPLL PENRLDGLFP EGRENFDYEL PVKGEEDLWG RFFRAADKGL
HSLQSFPVRR FVPREAAIRH VIEWIIRHQD ADGGWGGIQP PWIYGLMALS VEGYPLHHPV
LAKAMDALND PGWRRDKGDA SWIQASNSPV WDTMLAVLAL HDAGAEDRYS PQMDKAIGWL
LDRQVRVKGD WSIKLPDTEP GGWAFEYAND KYPDTDDTAV ALIALAGCRH RPEWRERDIE
GAISRGVNWL LAMQSSSGGW GAFDKDNNRS ILTKIPFCDF GEALDPPSVD VTAHVLEAFG
LLGISRNHPS VQKALAYIRS EQERNGAWFG RWGVNYVYGT GAVLPALAAI GEDMTQPYIV
RACDWLMSVQ QENGGWGESC ASYMDINAVG HGVATASQTA WALIGLLAAK RPKDREAIAR
GCQFLIERQE DGSWTEEEYT GTGYPGYGVG QAIKLDDPSL PDRLLQGAEL SRAFMLRYDL
YRQYFPVMAL SRARRMMKED ASAAA

Sequence Number (ID): 85
Sequence Name: Gmo_SHC_WT
Length: 685
Molecule Type: AA
Features Location/Qualifiers:

- source, 1..685
> mol_type, protein
> organism, Gluconobacter morbifer

Residues:
MSPADISTKS SSFQRLDNML PEAVSSACDW LIDQQKPDGH WVGPVESNAC MEAQWCLALW
FLGQEDHPLR PRLAQALLEM QREDGSWGIY VGADHGDINT TVEAYAALRS MGYAADMPIM
AKSAAWIQQK GGLRNVRVFT RYWLALIGEW PWDKTPNLPP EIIWLPDNFI FSIYNFAQWA
RATMMPLTIL SARRPSRPLL PENRLDGLFP EGRENFDYEL PVKGEEDLWG RFFRAADKGL
HSLQSFPVRR FVPREAAIRH VIEWIIRHQD ADGGWGGIQP PWIYGLMALS VEGYPLHHPV
LAKAMDALND PGWRRDKGDA SWIQATNSPV WDTMLAVLAL HDAGAEDRYS PQMDKAIGWL
LDRQVRVKGD WSIKLPDTEP GGWAFEYAND KYPDTDDTAV ALIALAGCRH RPEWRERDIE
GAISRGVNWL LAMQSSSGGW GAFDKDNNRS ILTKIPFCDF GEALDPPSVD VTAHVLEAFG
LLGISRNHPS VQKALAYIRS EQERNGAWFG RWGVNYVYGT GAVLPALAAI GEDMTQPYIV
RACDWLMSVQ QENGGWGESC ASYMDINAVG HGVATASQTA WALIGLLAAK RPKDREAIAR
GCQFLIERQE DGSWTEEEYT GTGFPGYGVG QAIKLDDPSL PDRLLQGAEL SRAFMLRYDL
YRQYFPVMAL SRARRMMKED ASAAA

Sequence Number (ID): 86
Sequence Name: T0DEU9_ALIAG
Length: 633
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..633
> mol_type, protein
> organism, Alicyclobacillus acidoterrestris

Residues:
MTKQLLDTPM VQATLEAGVA HLLRRQAPDG YWWAPLLSNV CMEAEYVLLC HCLGKKNPER
EAQIRKYIIS QRREDGTWSI YPGGPSDLNA TVEAYVALKY LGEPASDPQM VQAKEFIQNE
GGIESTRVFT RLWLAMVGQY PWDKLPVIPP EIMHLPKSVP LNIYDFASWA RATIVTLTIV
MNRRPVTPLP DYAKVPELFE AKRPPKRRSA KGGDSGFFVA LDKFLKAYNK WPIQPGRKSG
EQKALEWILA HQEADGCWGG IQPPWFYALL ALKCLNMTDH PAFVKGFEGL EAYGVHTSDG
GWMFQASISP IWDTGLTVLA LRSAGLPPDH PALIKAGEWL VSKQILKDGD WKVRRRKAKP
GGWAFEFHCE NYPDVDDTAM VVLALNGIQL PDEGKRRDAL TRGFRWLREM QSSNGGWGAY
DVDNTRQLTN RIPFCDFGEV IDPPSEDVTA HVLECFGSFG YDEAWKVIRK AVEYLKAQQR
PDGSWFGRWG VNYVYGIGAV VPGLKAVGVD MREPWVQKSL DWLVEHQNED GGWGEDCRSY
DDPRLAGQGV STPSQTAWAL MALIAGGRVE SDAVLRGVTY LHDTQRADGG WDEEVYTGTG
FPGDFYLAYT MYRDIFPVWA LGRYQEAMQR IRG

Sequence Number (ID): 87
Sequence Name: T0DEU9_ALIAG_F437A G600M
Length: 633
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..633
> mol_type, protein
> organism, synthetic construct

Residues:
MTKQLLDTPM VQATLEAGVA HLLRRQAPDG YWWAPLLSNV CMEAEYVLLC HCLGKKNPER
EAQIRKYIIS QRREDGTWSI YPGGPSDLNA TVEAYVALKY LGEPASDPQM VQAKEFIQNE
GGIESTRVFT RLWLAMVGQY PWDKLPVIPP EIMHLPKSVP LNIYDFASWA RATIVTLTIV
MNRRPVTPLP DYAKVPELFE AKRPPKRRSA KGGDSGFFVA LDKFLKAYNK WPIQPGRKSG
EQKALEWILA HQEADGCWGG IQPPWFYALL ALKCLNMTDH PAFVKGFEGL EAYGVHTSDG
GWMFQASISP IWDTGLTVLA LRSAGLPPDH PALIKAGEWL VSKQILKDGD WKVRRRKAKP
GGWAFEFHCE NYPDVDDTAM VVLALNGIQL PDEGKRRDAL TRGFRWLREM QSSNGGWGAY
DVDNTRQLTN RIPFCDAGEV IDPPSEDVTA HVLECFGSFG YDEAWKVIRK AVEYLKAQQR
PDGSWFGRWG VNYVYGIGAV VPGLKAVGVD MREPWVQKSL DWLVEHQNED GGWGEDCRSY
DDPRLAGQGV STPSQTAWAL MALIAGGRVE SDAVLRGVTY LHDTQRADGG WDEEVYTGTM
FPGDFYLAYT MYRDIFPVWA LGRYQEAMQR IRG

Sequence Number (ID): 88
Sequence Name: ZmSHC_F503A_G667M

Length: 725
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..725
> mol_type, protein
> organism, synthetic construct
Residues:
MGIDRMNSLS RLLMKKIFGA EKTSYKPASD TIIGTDTLKR PNRRPEPTAK VDKTIFKTMG
NSLNNTLVSA CDWLIGQQKP DGHWVGAVES NASMEAEWCL ALWFLGLEDH PLRPRLGNAL
LEMQREDGSW GVYFGAGNGD INATVEAYAA LRSLGYSADN PVLKKAAAWI AEKGGLKNIR
VFTRYWLALI GEWPWEKTPN LPPEIIWFPD NFVFSIYNFA QWARATMVPI AILSARRPSR
PLRPQDRLDE LFPEGRARFD YELPKKEGID LWSQFFRTTD RGLHWVQSNL LKRNSLREAA
IRHVLEWIIR HQDADGGWGG IQPPWVYGLM ALHGEGYQLY HPVMAKALSA LDDPGWRHDR
GESSWIQATN SPVWDTMLAL MALKDAKAED RFTPEMDKAA DWLLARQVKV KGDWSIKLPD
VEPGGWAFEY ANDRYPDTDD TAVALIALSS YRDKEEWQKK GVEDAITRGV NWLIAMQSEC
GGWGAFDKDN NRSILSKIPF CDAGESIDPP SVDVTAHVLE AFGTLGLSRD MPVIQKAIDY
VRSEQEAEGA WFGRWGVNYI YGTGAVLPAL AAIGEDMTQP YITKACDWLV AHQQEDGGWG
ESCSSYMEID SIGKGPTTPS QTAWALMGLI AANRPEDYEA IAKGCHYLID RQEQDGSWKE
EEFTGTMFPG YGVGQTIKLD DPALSKRLLQ GAELSRAFML RYDFYRQFFP IMALSRAERL
IDLNN

Sequence Number (ID): 89
Sequence Name: ZmSHC_wt
Length: 725
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..725
> mol_type, protein
> organism, Zymomonas mobilis
Residues:
MGIDRMNSLS RLLMKKIFGA EKTSYKPASD TIIGTDTLKR PNRRPEPTAK VDKTIFKTMG
NSLNNTLVSA CDWLIGQQKP DGHWVGAVES NASMEAEWCL ALWFLGLEDH PLRPRLGNAL
LEMQREDGSW GVYFGAGNGD INATVEAYAA LRSLGYSADN PVLKKAAAWI AEKGGLKNIR
VFTRYWLALI GEWPWEKTPN LPPEIIWFPD NFVFSIYNFA QWARATMVPI AILSARRPSR
PLRPQDRLDE LFPEGRARFD YELPKKEGID LWSQFFRTTD RGLHWVQSNL LKRNSLREAA
IRHVLEWIIR HQDADGGWGG IQPPWVYGLM ALHGEGYQLY HPVMAKALSA LDDPGWRHDR
GESSWIQATN SPVWDTMLAL MALKDAKAED RFTPEMDKAA DWLLARQVKV KGDWSIKLPD
VEPGGWAFEY ANDRYPDTDD TAVALIALSS YRDKEEWQKK GVEDAITRGV NWLIAMQSEC
GGWGAFDKDN NRSILSKIPF CDFGESIDPP SVDVTAHVLE AFGTLGLSRD MPVIQKAIDY
VRSEQEAEGA WFGRWGVNYI YGTGAVLPAL AAIGEDMTQP YITKACDWLV AHQQEDGGWG
ESCSSYMEID SIGKGPTTPS QTAWALMGLI AANRPEDYEA IAKGCHYLID RQEQDGSWKE
EEFTGTGFPG YGVGQTIKLD DPALSKRLLQ GAELSRAFML RYDFYRQFFP IMALSRAERL
IDLNN

Sequence Number (ID): 90
Sequence Name: CcrGGPPS2-del57
Length: 942
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..942
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggcgagcg tgctgagcgg taaagatacg atgaaaggcg aagaagaaaa tagcggtttc
gacttccaga gctacatgga ccagatggct gactctgtga atcaagcact ggagagcgct
gtgagcctgc gtgagccgct caagattcat gaagcaatgc gctacagctt actggcgggt
ggcaaacgcg ttcgtccgct gctgtgcatc gccgcgtgcg aactggttgg cggcgatgtt
tcggttgcga tgccggctgc atgtgcagtt gagatgatcc acaccatgag cctgatccac
gacgatctgc cgtgtatgga caacgacgat ttgcgccgtg gtaagccaac gaatcataaa
gcattcggtg aagatatcgc ggttctggcg ggtgacgcgt tgctgagctt tgcctttgaa
cacgtcgctg tgagcactgt tggtgcgagc ccggacaaga tcgtgcgtgc ggtgggtgag

ctggcaaaag ctgtgggtaa ggaaggtctg gtggcaggcc aagtcgttga tattacgtcc
gagggcctga acgacgttgg cctggatcat ctggagtaca ttcacgtgca taagaccgca
gtcttgctgg aagcggccgt cgttctgggc gcgattttgg gtggcggcac cgatgaagag
gttgagcgtc tgcgtaagtt cgccatttgc atcggtttgc tgtttcaagt cgtcgatgac
atccttgacg ttaccaaatc ttccgtcgag ctgggtaaga ccgcgggtaa agacctggtg
gcggataaag tcacctatcc taagctgatg ggtctggaga aaagccgtga gttcgcggag
aaactgcgcg acgatgccgt ggaacagctg cgcgtgtttg accaagtaaa agccgcaccg
ctgattgcgc tggcccacta tattgcgtac cgtcagaact aa

Sequence Number (ID): 91
Sequence Name: CcrGGPPS2-del57
Length: 765
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..765
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgagatcta tcgctagaag aactgctgtt ggtgctgctt tgttgttggt tatgccagtt
gctgtttgga tctctggttg gagatggcaa ccaggtgaac aatcttggtt gttgaaggct
gctttctggg ttactgaaac tgttactcaa ccatggggtg ttatcactca cttgatcttg
ttcggttggt tcttgtggtg tttgagattc agaatcaagg ctgctttcgt tttgttcgct
atcttggctg ctgctatctt ggttggtcaa ggtgttaagt cttggatcaa ggacaaggtt
caagaaccaa gaccattcgt tatctggttg gaaaagactc accacatccc agttgacgaa
ttctacactt tgaagagagc tgaaagaggt aacttggtta aggaacaatt ggctgaagaa
aagaacatcc cacaatactt gagatctcac tggcaaaagg aaactggttt cgctttccca
tctggtcaca ctatgttcgc tgcttcttgg gctttgttgg ctgttggttt gttgtggcca
agaagaagaa ctttgactat cgctatcttg ttggtttggg ctactggtgt tatgggttct
agattgttgt tgggtatgca ctggccaaga gacttggttg ttgctacttt gatctcttgg
gctttggttg ctgttgctac ttggttggct caaagaatct gtggtccatt gactccacca
gctgaagaaa acagagaaat cgctcaaaga gaacaagaat cttaa

Sequence Number (ID): 92
Sequence Name: CarG
Length: 963
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..963
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgttgacat cttctaagtc catcgaatct ttcccaaaga acgttcaacc atacggtaaa
cactatcaaa acggtttaga accagtcggt aagtctcaag aagacatctt gttggaacct
ttccactact tatgttctaa tccaggtaag gatgttagaa ccaagatgat tgaagctttc
aacgcctggt tgaaagtccc aaaggacgat ttgattgtta tcaccagagt cattgaaatg
ttgcactccg cttctttgtt gattgatgac gtcgaggacg attctgtctt gagaagaggt
gtcccagccg cccaccatat ctacggtacc cctcaaacca tcaactgcgc taactacgtt
tatttcttgg ccttgaaaga aatcgccaag ttgaacaagc caaatatgat tactatttat
accgatgaat tgatcaactt gcacagaggt caaggtatgg aattgttctg gcgtgatacc
ttgacctgcc caactgagaa agagtttttg gatatggtta acgataagac tggtggtttg
ttgagattgg ccgtcaagtt gatgcaagag gcttctcaat ctggtaccga ctatactggt
ttggtttcta agatcggtat ccattttcaa gttagagatg actacatgaa cttgcaatcc
aaaaactacg ccgataataa gggtttctgt gaagatttga ccgaaggtaa gttctccttt
ccaattattc actctatcag atctgaccca tccaacagac aattattgaa tattttgaag
caaagatctt cttctattga attgaaacaa ttcgctttac aattgttaga aaacactaac
acttttcaat actgtagaga tttcttgaga gttttggaaa aggaagccag agaagagatc
aaattattgg gtggtaacat catgttggaa aagattatgg acgtcttgtc tgttaatgaa
taa

Sequence Number (ID): 93
Sequence Name: PgpB

Length: 765
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..765
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgcgttcga ttgccagacg taccgcagtg ggagctgcac tattgcttgt catgccagta
gccgtatgga tttctggctg gcgttggcaa cctggagaac aaagttggct actaaaagcg
gctttttggg ttactgaaac tgtcacccag ccctggggcg tcattacaca tttgatttta
ttcggctggt ttctctggtg tctgcgtttt cgcattaagg ctgcctttgt attatttgcc
attctggcgg ccgcaatcct tgtgggacaa ggcgttaaat cctggatcaa agacaaagtc
caggaaccac gacctttttgt tatctggctg gaaaaaacac atcatattcc ggttgatgag
ttctacactt taaagcgagc agaacgcgga aatctagtga aagaacagtt ggctgaagag
aaaaatatcc cacaatattt gcgttcacac tggcagaaag agacggggtt tgcctttcct
tccggtcaca cgatgtttgc tgccagttgg gcactgctgg ccgttggttt gctgtggccg
cgtcggcgaa cgttaaccat tgctatcttg ctggtctggg caacgggagt catgggaagc
cgcctgctgc tcgggatgca ttggccacgc gatctggtag tagctacgtt gatttcgtgg
gcgctggtgg cggtggcaac gtggcttgcg caacgaattt gtgggccatt aacaccacct
tcggaagaaa atcgcgaaat agcgcaacga gaacaagaaa gttaa

Sequence Number (ID): 94
Sequence Name: PgpB
Length: 765
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..765
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgagatcta tcgctagaag aactgctgtt ggtgctgctt tgttgttggt tatgccagtt
gctgtttgga tctctggttg gagatggcaa ccaggtgaac aatcttggtt gttgaaggct
gctttctggg ttactgaaac tgttactcaa ccatggggtg ttatcactca cttgatcttg
ttcggttggt tcttgtggtg tttgagattc agaatcaagg ctgctttcgt tttgttcgct
atcttggctg ctgctatctt ggttggtcaa ggtgttaagt cttggatcaa ggacaaggtt
caagaaccaa gaccattcgt tatctggttg gaaaagactc accacatccc agttgacgaa
ttctacactt tgaagagagc tgaaagaggt aacttggtta aggaacaatt ggctgaagaa
aagaacatcc cacaatactt gagatctcac tggcaaaagg aaactggttt cgctttccca
tctggtcaca ctatgttcgc tgcttcttgg gctttgttgg ctgttggttt gttgtggcca
agaagaagaa ctttgactat cgctatcttg ttggtttggg ctactggtgt tatgggttct
agattgttgt tgggtatgca ctggccaaga gacttggttg ttgctacttt gatctcttgg
gctttggttg ctgttgctac ttggttggct caaagaatct gtggtccatt gactccacca
gctgaagaaa acagagaaat cgctcaaaga gaacaagaat cttaa

Sequence Number (ID): 95
Sequence Name: NudB
Length: 453
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..453
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaaggata aagtgtataa gcgtcccgtt tcgatcttag tggtcatcta cgcacaagat
acgaaacggg tgctgatgtt gcagcggcgt gacgatcccg atttctggca gtcggtaacc
ggcagcgtgg aagagggtga aaccgcgccg caagctgcca tgcgcgaaat aaaggaagag
gtcaccattg atgttgtcgc tgaacaactg accttaattg actgtcagcg cacggtagag
tttgaaattt tttcacattt acgtcatcgc tatgcgccgg gcgtgacgcg taatacggaa
tcatggttct gtcttgcgct tccgcacgag cggcagatcg ttttcactga acatctggct
tacaagtggc ttgatgcgcc tgctgcggcg gcgctcacta agtcctggag caaccggcag
gcgattgaac agtttgtaat taacgctgcc taa

Sequence Number (ID): 96
Sequence Name: LPP1
Length: 825
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..825
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgattagcg tgatggcaga cgagaagcac aaagaatact ttaaactgta ctattttcag
tacatgatca tcggtctgtg caccattctg ttcctgtata gcgagatcag cctggtgccg
cgtggtcaaa acattgagtt tagcctggac gacccgtcta tcagcaagcg ctatgttccg
aatgaattgg tcggtccgct ggagtgcctg attctgtccg ttggtctgag caacatggtc
gtcttttgga cctgtatgtt cgacaaagat ctgctgaaga aaaatcgtgt caagcgcctg
cgtgagcgtc cggatggcat cagcaacgac tttcacttca tgcatacctc cattctgtgt
ctgatgctga tcatttccat caacgctgcg ctgaccggtg cgctgaagct catcattggc
aatctgcgtc cagacttcgt ggatcgctgt attcctgatc tgcagaaaat gtctgatagc
gacagcttgg tgttcggcct ggatatttgc aaacagacca ataagtggat cttatacgaa
ggcttgaaaa gcaccccgag cggccacagc tcttttatcg tgagcacgat gggtttcacg
tatctgtggc aacgcgtttt cacgacccgt aacacgcgtt cgtgtatttg gtgcccgttg
ctggccttgg ttgtgatggt cagccgcgtc attgaccatc gtcatcactg gtacgacgtt
gttagcggtg cggttctggc gttcctggta atctactgct gctggaaatg gacttttacc
aatcttgcaa agcgtgatat cctgccgtca ccggttagcg tgtaa

Sequence Number (ID): 97
Sequence Name: NUDX1
Length: 453
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..453
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgggtaatg aaactgttgt cgtcgcagaa accgctggtt ctattaaagt tgcagttgtt
gtgtgcctgc tgcgcggtca aaacgtcctg ctgggccgtc gccgtagcag cctgggtgat
tcgacgttca gcctgccgtc cggtcacttg gaatttggcg agagctttga agagtgcgcc
gcgcgcgagc tgaaagagga aaccgacctg gatatcggta aaatcgagtt gctgaccgtg
accaataacc tgtttctgga cgaagcgaag ccgagccaat acgtcgcggt gttcatgcgt
gcggtgttgg ccgacccgcg tcaggaaccg cagaatattg agcctgagtt ctgtgacggc
tggggttggt atgagtggga caatctgccg aagccgctgt tctggccgct ggataacgtt
gtgcaagatg gcttcaaccc atttccgacg taa

Sequence Number (ID): 98
Sequence Name: PeSubTPP1
Length: 873
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..873
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgcaacctt ttattagcgt cgatggtgtg gtgaattttc gtgatattgg tggttatgtt
tgccgtaatc cggccggttt gtcgagcctg ccgagcaacg ttgacgaaac cccggaaaag
caatggtgta tccgcccagg cttcgttttc cgtgcagcgc aaccgtccca aattacgccg
gctggtatcg agattcttaa gaaaacgctg gcgatccaag cgattttcga ttttcgtagc
gagtccgaga tccaactggt gagcaagcgt tacccggaca gcctgctgga catcccgggc
actacgcgtc atgctgttcc ggtgtttcag gagggtgatt acagcccgat ctcgttggcc
aaacgttacg gtgtgaccgc ggacgagagc accaacgatc agtccttccg tccgggtttt
gtcaaagcgt atgaagccat cgcacgcaac gcagcacagg ctggtagctt ccgcgccatt
atccagcata tcctgcagga ctccgctggc ccagttttgt ttcactgcac cgtaggcaaa

gatcgcacgg gtgttttctc tgcactgatt ctgaagctgt gcggtgtggc cgacgaagat
attgtggcag actatgcgct gaccactcag ggcctgggtg tctggcgtga gcacctgatc
cagcgcctgt tgcagcgtgg tgaagcgacc accaaagaac aagcggaagc gatcatctct
agcgacccgc gcgacatgaa agcgttcctg agcaacgtcg ttgagggcga gtttggtggc
gcacgcaact acttcgtgaa tctgtgtggc ctgcctgaag gcgaggttga ccgtgtcatt
accaaactgg tcgtcccgaa aaccaccaag taa

Sequence Number (ID): 99
Sequence Name: TalVeTPP
Length: 936
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..936
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgagcaatg acacgacgac caccgcgagc gccggtactg caacttctag ccgttttctg
agcgtcggcg gcgttgtgaa ttttcgcgag ctgggtggct atccatgcga cagcgtgccg
ccggctccgg caagcaacgg ttcgcctgat aatgcgtccg aggcaacgct gtgggttggt
cactccagca ttcgtccggg tttcctgttc cgcagcgcgc agccgagcca gattacgccg
gcgggtatcg aaacgctgat ccgccaactg ggcatccaga ccatttttga tttccgtagc
cgtaccgaga tcgaactggt ggcgacccgt tacccggact ctctgttgga aattccgggc
accacgcgct attccgtccc ggttttctcc gagggtgact attctccggc gagcctggtg
aagcgctatg gtgttagcag cgataccgcc acggacagca cctctagcaa gagcgcgaag
ccgaccggct tcgttcatgc atacgaagcc attgcgcgca gcgccgctga gaacggtagc
ttccgtaaaa ttaccgacca catcatccag catcctgatc gtccaatttt gttccactgt
accctgggta aagaccgtac gggtgtcttt gcggcgctgt tgctgagcct gtgtggtgtg
ccggacgaaa ccatcgtcga agattacgcg atgaccaccg aaggctttgg tgcatggcgt
gagcacctga tccaacgtct gctgcaacgt aaagacgctg caacccgtga agatgccgag
agcatcattg cgtcgccgcc ggagactatg aaagcatttc tggaagatgt tgtggcagcg
aaatttggtg gcgcgcgtaa ctacttcatt caacattgcg gcttcactga agctgaagtc
gataagctga gccacaccct ggcgatcacg aactaa

Sequence Number (ID): 100
Sequence Name: AstI
Length: 606
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..606
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgacgcgcc aaagccacta ccaagcaatc atcctggacc tgggcaatgt agtttttgag
tgggatacga gccagaaccc tccgaccgcg gcaccgaacc agatttctct gttgcgcacc
tcgatgaaaa gcccggttta tcacagctac gagcgtggtc agctgtctac cgaagagtgt
caccgcttgc tgggtgagag cctgcatgtc gatccgggtc agattaaaga ggccttcgac
ctggcgcgcc aaagcctgcg tagcaatccg gcactgctgg atttcatccg tcagttaaag
caaactcgtg gcgtcgcagt ctacgcgatg agcaatatcc cgcaagctga aattgaatat
cttaaagagt cccgtgcggg tgatatggaa gtgtttgacg aagttttcgc gagcggttat
gtgggcagcc gtaagccaga aaccgagttt taccgccgtg tgatgggcga gattggtctg
aaagccgagc gcgtggtttt tgttgacgat aaggaagaga acgttgacgt cgcccgcggt
ctgggtctgt acggcgtgtg cttcggtggc gtggaagaat tgcgtggcca tctgctgggc
atctaa

Sequence Number (ID): 101
Sequence Name: AstK
Length: 591
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..591
> mol_type, unassigned DNA

> organism, synthetic construct E.coli optimized

Residues:

atgtgtacga cgtttaaagc tgcaattttt gacatgggtg gtgttctgtt tacctggaat
ccaatcgtgg atacccaggt ttccctgaaa gacctgggta cgatcatcaa cagcgaaact
tgggagcaat ttgagcgtgg taagatcgag cctgacgact gctaccacca attgggcagc
cagattggcc tgccgggttc tgagatcgcg gcaacgttcc gtcagaccac gggttgcctg
cgtccggatg cgcgcatgac cagcctgctg cgcgagttga agggtcaggg tgttgcggtt
tacatgatga ccaatattcc ggccccggat ttccatcagc tgcgtgagat gcactatgag
tgggatttgt tcgatggcat ttttgcctcc gctctggaag gcatgcgtaa accggacctg
gaattttacg agcacgtgct gaagcaaatt gataccagcg cggcggaaac catcttcgtg
gacgataaac tggaaaacgt cattgcggcg caagccgtgg gtatggtcgg cctgcatctg
accgacagcc tggccacctg tatggaactg cgccaactgg tcggttgcta a

Sequence Number (ID): 102

Sequence Name: PsAerADH

Length: 1119

Molecule Type: DNA

Features Location/Qualifiers:

- source, 1..1119

> mol_type, unassigned DNA

> organism, synthetic construct E.coli optimized

Residues:

atgaactcga tccaacctac tcaagcaaaa gcagcagtct tgcgcgcagt cggcggcccg
ttctctattg agccgatccg catcagccca ccgaagggtg acgaagtgct ggttcgtatc
gttggtgtgg gtgtctgcca caccgacgtc gtctgtcgtg atagctttcc ggtgccgttg
ccgatcattc tgggtcacga gggctccggt gtgattgaag ctgtgggtga ccaagtgacc
ggtctgaaac cgggtgacca cgttgtgctg tccttcaata gctgcggcca ttgctacaac
tgtggtcatg acgagcctgc gtcttgtctg cagatgctgc cgttgaattt cggtggcgcg
gagcgtgcgg cggacggcac catcgaagat gaccagggcg cagctgttcg tggcctgttc
ttcggccaaa gctcctttgg tagctacgcg attgcacgtg cggttaacac tgtcaaagtt
gatgacgatc tgccgttggc gctgctgggt ccgctgggtt gcggtattca gaccggcgcg
ggtgcagcca tgaatagcct gggtttacag ggtggccaga gcttcattgt gtttggcggc
ggcgccgtcg gtctgagcgc ggtcatggcc gccaaggccc tgggtgttag cccgctgatt
gttgtggagc cgaacgaagc tcgccgtgcg ctggcactgg aattgggtgc gagccacgcg
tttgacccat ttaacaccga agatctggtc gcgagcattc gcgaagtcgt tccggctggc
gcaaaccacg cgctggacac gacgggtctg ccgaaagtta ttgccaacgc gatcgattgc
atcatgagcg gcggcaaact gggtctgctc ggtatggcga atccggaagc gaatgtgccg
gcgaccctgc tggatctgct gagcaaaaat gtgacgctga agccgatcac cgagggtgac
gcaaacccac aagaatttat tccgcgtatg ctggctctgt atcgtgaggg taagtttccg
ttcgataagc tgatcaccac gttcccgttc gagcatatca acgaagcaat ggaagctacc
gagagcggta aggccattaa accggttctg accctgtaa

Sequence Number (ID): 103

Sequence Name: ThTerpADH1

Length: 1122

Molecule Type: DNA

Features Location/Qualifiers:

- source, 1..1122

> mol_type, unassigned DNA

> organism, synthetic construct E.coli optimized

Residues:

atgtgtagca atcatgattt caccgcagcc cgtgcagcag tcttacgtaa agttggtggc
ccgttggaaa tcgaagatgt ccgtatttct gccccgaaag gcgacgaagt cctggtgcgt
atggttggcg tgggtgtgtg tcataccgac ctcgtctgcc gtgatgcgtt cccggtgccg
ctgcctattg ttctgggtca cgagggtgca ggcatcgttg aagccgtggg tgagggcgtg
cgctccctgg agccgggtga ccgtgttgtg ctgagcttca atagctgcgg ccgctgtggc
aactgcggta gcggtcaccc gagcaactgc ctgcaaatgc tgccgctgaa ttttggtggc
gcgcaacgcg ttgacggtgg ccgcatgttg gacgcggcgg gtaacgctgt ccagggtctg
ttttttggtc aatctagctt cggcacgtat gcgatcgcgc gtgagattaa cgccgtgaaa
gtcgccgaag atctgccgct ggaaatcctg ggtccgctgg gttgcggtat tcagaccggt
gcgggtgcgg cgattaacag cctgggtatt ggtccgggtc agtccttggc tgtgttcggt

ggcggcggcg tgggtcttag cgcgttgctg ggcgctcgtg ctgtgggtgc cgcccaagtt
gttgttgttg agccgaacgc cgcacgtcgc gcgctggcgc tggaactggg tgcgagccat
gcattcgacc cgtttgcggg tgacgacctg gtcgcggcga tccgcgcagc gacgggtggc
ggcgcaaccc acgcgctgga tacgaccggc ctgccgtcgg tgattggcaa tgcaatcgat
tgtactttgc cgggtggcac ggttggtatg gtcggcatgc cagcgcctga cgctgcggtc
ccggcgaccc tgctggattt gctgactaag agcgtcacgc tgcgtccgat caccgagggt
gacgcagatc cgcaggcctt catcccacag atgctgcgct tttaccgtga gggtaagttc
ccgtttgacc gtctgattac ccgttaccgt tttgatcaga tcaatgaagc tctgcacgca
accgaaaagg gtggcgcgat taaaccggtt ctggtgttct aa

Sequence Number (ID): 104
Sequence Name: VoADH1
Length: 1146
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1146
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgactaaat ccagcggtga agtgatttct tgtaaggcag cagtgatcta taagagcggt
gagcctgcta aagttgaaga aattcgtgtt gatccgccta agagcagcga agttcgtatt
aagatgctgt acgcctcctt gtgtcacacg gacattctgt gttgcaacgg cctgccggtg
ccgctgtttc cgcgcattcc gggtcacgag ggcgtgggtg ttgtggagag cgcgggtgaa
gatgtgaaag atgttaaaga gggcgacatc gttatgccac tgtacctggg cgagtgtggt
gagtgcctca attgcagcag cggtaagacg aatctgtgcc acaagtaccc actggacttc
tctggtgtgc tgccgagcga cggtacgagc cgcatgtcag tagcaaaatc cggtgagaaa
attttccatc acttcagctg tagcacctgg tccgaatatg ttgtcatcga gagctcgtat
gtcgtcaaag ttgatagccg tctgccgctg ccgcatgcgt cctttctggc atgcggcttc
accacgggtt acggcgcggc gtggaaagag gctgacattc cgaagggcag caccgtcgcg
gtgctgggcc tgggtgcggt cggtctgggt gtggttgctg gtgcgcgttc tcagggtgcg
agccgcatta ttggcgtgga catcaacgac aagaaaaaag caaaagccga gatctttggt
gttactgagt ttctgaatcc gaagcaactg ggtaaaagcg cgagcgaaag catcaaagac
gtcaccggcg gcctgggcgt tgactactgt ttcgagtgca ccggtgtccc ggccctgttg
aacgaagccg tggatgcgag caagatcggc ttgggtacga tcgtcatgat tggtgcgggt
atggaaacca gcggtgttat taactatatc ccgctgctgt gcggccgtaa actgatcggt
agcatttacg gtggcgttcg catccgtagc gacttaccgc tgatcattga gaaatgcatc
aacaaagaaa ttccgctgaa cgaactgcag acccacgaag tgagcttgga aggcattaat
gatgcattcg gcatgctgaa gcaaccggac tgcgttaaga tcgtcatcaa gttcgagcag
aaataa

Sequence Number (ID): 105
Sequence Name: SCH80-05240
Length: 1122
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1122
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgattagag cagaacagaa cagcaccagc gcgatgcaaa tgaccgcggc actgtcacat
ggtccgcaca gcccgtttac gctggatacg gttgagattg acgagccacg cgccgacgaa
attctggtac gcatcgttgc gactggtctg tgtcatacgg acttgtttac caagagcgtc
ctgccggagc gcctgggtcc gtgcgtgttc ggccacgagg gtgcgggcgt ggttgaggca
gttggctctg ccattgacaa agttgttccg ggtgatcacg tcctgttgtc ctaccgtagc
tgcggcgtct gccgtcagtg cctgagcggc caccgtgctt actgtgagag ctcccacggc
ctgaatagct ccggtgctcg taccgacggt agcaccccgg tgcgtcgtag cggtacgccg
attcgtagcg cgttcttcgg tcaatccagc ttcgcggaat atgttatcgc aaccgcagac
aacaccgttg tggtcgatcc ggccgtggac ttgaccgttg cagcaccgct gggttgtggc
tttcagaccg gcgccggcgc ggtgctgaat ctgctgcgcc ctgagccgga cagcactttc
gtcgtctttg gtgccggcag cgtcggtttg gcggcactgc tggcggcgcg tgcggcgggt
gtttcgaccc tggtcgcagt tgatccggtc gcgcagcgcc gtgcgttggc cgaagaattt

ggtgccgtta ccgtcgatcc gacgaccgaa gatgccgttg aagctgtgcg cgcggcgacc
gacggtggca gcacgcattc tctggatacc acgggcatcg gttctgtgat taaccaagcc
gtgacctctc tgcgtgcgcg tggtactctg gctgtggttg gcctgggtgc tagcacggtc
gagatgaata tggcagacat tatgctgagc ggtaaaacga tccgtggttg catcgagggc
gagagcgaag tttcgacgtt tatcccggaa ctggtcgagc tgttcaccgg tggccgtttc
ccgattgacc gcctggttac ccgttatgca ttcagcgata tcaacaaagc tgtggaagat
caagcgtccg gtcgcgtcat caagccagtg ctggtgtggt aa

Sequence Number (ID): 106
Sequence Name: Ppseudo-alkJ
Length: 1677
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1677
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgtatgatt atatcattgt gggcgctggt tctgcgggtt gtgttctggc taatcgtctg
agcgcggacc cgagcaagcg tgtctgcttg ctggaggctg gtccgcgcga cacgaacccg
ctgatccaca tgccgctggg tattgcgctg ctgtccaact ctaaaaagct gaactgggct
ttccagaccg caccgcaaca gaatctgaac ggtcgctctc tgttctggcc gcgtggtaaa
acgctgggcg gcagcagcag cattaatgcg atggtctaca ttcgtggtca cgaagatgac
tatcacgcgt gggaacaggc cgcgggccgc tactggggtt ggtatcgcgc gctggaactg
ttcaagcgtc tggaatgcaa tcagcgtttt gacaagagcg aacaccacgg cgttgatggt
gagctggccg ttagcgacct gaagtacatc aacccgctga gcaaagcgtt tgtccaagcg
ggcatggaag cgaacatcaa cttcaatggt gacttcaatg gcgagtatca ggatggcgtc
ggcttttatc aagtgacgca gaaaaatggt caacgctgga gcagcgcgcg tgcgttcctg
catggtgttc tgagccgtcc gaacttggac attattaccg acgcacatgc aagcaaaatc
ctcttcgagg accgcaaagc cgtgggtgtg agctacatta agaaaaatat gcaccaccaa
gtcaagacca cctctggcgg cgaagtgctg ctgagcttgg gtgcagtcgg cacgccgcac
ttgctgatgt tgagcggtgt cggtgcggct gcagagctga aagaacatgg tgttagcttg
gtgcatgacc ttccagaggt cggtaaaaat ctgcaggacc atctggatat cacgctgatg
tgtgcagcaa actcccgcga accgatcggt gtcgcactga gcttcattcc gcgcggtgtg
agcggtttgt tttcttatgt ttttaagcgt gagggcttcc tgaccagcaa cgttgctgag
agcggtggct tcgttaagtc gtccccagac cgtgatcgcc ctaatctgca atttcacttt
ctgcctacct acctgaaaga ccacggccgt aagatcgcgg gcggctacgg ctacaccctg
catatttgcg acctgttgcc gaaatcccgt ggtcgtattg gtttgaaatc cgccaacccg
ctgcaaccgc cactgatcga tccgaactac ctgtcagatc atgaagatat taagaccatg
attgccggca tcaagattgg tcgcgcgatc ctgcaagcgc cgagcatggc aaagcacttt
aaacacgaag ttgtgccggg ccaagccgtg aaaactgatg atgaaatcat cgaggacatc
cgtcgtcgcg cggaaactat ctaccaccct gtcggcacct gtcgcatggg caaagaccct
gcgtcggttg ttgatccgtg cctgaagatc cgtggtctgg cgaatatccg tgtggttgac
gccagcatca tgccgcactt agtggcaggc aataccaatg ccccgaccat tatgattgcc
gagaacgccg ctgagattat tatgcgcaat ttggatgtag aggctctgga agcgagcgcc
gagttcgcgc gcgagggtgc agagctggaa ctggcgatga tcgcagtctg catgtaa

Sequence Number (ID): 107
Sequence Name: PfluoADHF1
Length: 891
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..891
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaagtctt ttaacggtcg cgtcgcagca attacgggtg cagcatcagg tatgggtcgc
gcactggctc tggccttggc gcgtgagggt tgtcatctgg cgctggcgga taaaaacgcc
caaggcctgg agcaaaccct ggcgctgatc aagacgagca ccctgagccc ggttatggtt
accacccaag ttctggacgt tgccgaccgc caagcgatgg aagcgtgggc ggcgcgttgc
gtggccgagc atggtcaggt gaatttggtc tttaataatg ctggtgttgc cctgagcagc
actgtcgagg gtgtggatta tgcagacctg gagtggatcg tcggcattaa cttttggggt

gttgtgcatg gtacgaaagc gttcctgccg cacctgaaag cgagcggcga tggccacgtt
attaatacgt cctcggtttt tggcctgttc gcgcagccgg gtatgagcgg ttacaacgca
accaagttcg ccgtgcgtgg cttcaccgaa gcactgcgcc aagaactgga cttgcagcgc
tgcggcgtga gcgccacgtg cgtccacccg ggcggcatcc gtactgatat ttgtcgtagc
tcccgtattg atgcgaatat gaccggtttc ctgattcaca gcgagcaaca ggcgcgtgcg
gactttgaaa agctgtttat caccgatgct gatcaggccg caaaagtcat cctgcagggc
gtgcgtaaga acaaacgtcg cgtgctgatc ggtcgcgacg cttacttcct ggacttgttg
gcacgttgtc tgccggcggc gtatcaggct ctggtagtcc tggcgagcaa gcgtatggcg
ccgaaacagc gtcgcccagt gttcgaaacc aacgacgagc ctcgtctgta a

Sequence Number (ID): 108
Sequence Name: CymB
Length: 1125
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1125
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgactatca attctattca accgattcaa gcaaaagccg ctgtgctgcg tgccgtaggc
tccccgttta acattgagcc gattcgtatc agcccgccga agggtgatga agttctggtc
cgtattgtgg gtgtgggtgt ctgccatacc gacgtcgttt gccgtgacag cttcccggtt
ccgctgccaa tcatcctggg tcacgaaggc tcgggtgtga ttgaagcgat cggtgatcaa
gttacgagcc tgaagccagg tgaccacgtc gttctgagct tcaatagctg cggccactgt
tataactgcg gtcatgcgga gccggcaagc tgcctgcaga tgttaccgtt gaactttggt
ggcgcggagc gtgcggcgga cggcaccatc caagacgaca agggtgaagc cgtccgcggt
atgttctttg gccagtccag ctttggcacg tacgcaatcg cacgtgcggt gaatgctgtc
aaagttgacg acgatctgcc gctgcctctg ttgggcccgc tgggctgtgg tatccagacc
ggtgcgggtg cagcgatgaa cagcctgtct ctgcagagcg gtcagagctt catcgttttc
ggtggcggcg cggtcggtct gagcgctgtt atggcagcta aagcgctggg cgtgagcccg
ctgatcgttg tggagccgaa cgaaagccgc cgcgccctgg ccctggaact gggtgcatcc
cacgtgtttg atccgttcaa caccgaagat ctggttgcca gcattcgcga agtcgtgcct
gcgggtgcga accatgcact ggacacgacc ggtctgccga aagtgatcgc gagcgcgatt
gattgtatta tgagcggtgg caaactgggt ttgctgggta tggcgagccc ggaagcgaat
gtgccggcta ccctgttgga tttgctgagc aaaaatgtca cgctgaagcc gatcaccgag
ggcgatgcga acccacaaga gttcatcccg cgtatgctgg cactctaccg tgagggtaag
ttcccgtttg agaaactgat cacgaccttt ccgtttgagc acattaatga agcaatggaa
gccactgagt ccggtaaggc cattaaaccg gttctgacgc tgtaa

Sequence Number (ID): 109
Sequence Name: CdGeoA
Length: 1122
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1122
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaacgata cgcaggattt tattagcgcc caagccgcag tgttacgtca ggtcggtggc
ccgctggccg ttgagcctgt tcgtatcagc atgccgaagg gtgacgaagt cctgattcgt
atcgcgggtg ttggtgtgtg ccacaccgac ttggtgtgcc gtgatggctt cccggtgccg
ctgccaattg tgctgggtca cgagggtagc ggtactgtcg aagccgtcgg tgaacaagtc
cgtaccctga aaccgggcga tcgcgtcgtg ctgagcttta acagctgcgg tcattgcggt
aactgtcacg acggtcaccc gagcaattgc ctgcagatgc tgccgctgaa cttcggtggc
gcgcaacgcg tggacggtgg ccaagttttg gacggtgcgg gtcatccggt tcagtccatg
tttttcggcc agtccagctt tggcacccac gcagtagcgc gcgagatcaa cgcagtcaag
gtcggcgatg atctgccact ggaactgctg ggtccgttgg gttgtggcat tcaaaccggt
gcgggtgcag ctatcaattc tctgggcatt ggtccgggtc agtctctggc tatcttcggc
ggcggcggcg tgggtctgag cgcactgctg ggcgcccgtg cggtgggtgc cgaccgtgtt
gttgtcattg agccgaatgc agcgcgccgt gcgctggcat tggaactggg tgccagccac
gcactggacc cgcatgccga gggcgacctt gttgcggcga ttaaagctgc gacgggtggc

ggcgctacgc atagcttgga tacgaccggc ctgccgccag tcattggctc cgcgatcgcg
tgtactctgc cgggtggcac cgttggtatg gttggtctgc cggcgccgga cgcaccggtc
cctgcgacgc tgttggatct gctgagcaaa tcggttaccc tgcgtccgat taccgagggt
gacgctgacc cgcaacgctt catcccgcgt atgctggatt tccatcgtgc gggcaagttt
ccgttcgacc gcctgatcac ccgttaccgc tttgatcaga tcaatgaagc gctgcacgcg
accgagaaag gtgaagcaat caaaccggtt ctggtgtttt aa

Sequence Number (ID): 110
Sequence Name: AroAroADH
Length: 1130
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1130
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgggctcaa ttcaagattc tctgttcatc ccggctagag cggcagtgtt gcgtgcggtc
ggtggcccac tggaaatcga agatgttcgt atcagcccgc ctaagggcga cgaagttctg
gtccgtatgg ttggcgtggg cgtttgccac accgacgttg tgtgccgcga tggtttcccg
gtcccgctgc cgattgtctt gggtcacgag ggtgcgggta tcgtggaagc tgtgggtgag
cgtgtgacca aggtcaaacc tggccagcgt gtggtgctga gcttcaacag ctgcggtcac
tgcagctcct gtggtgagga tcacccggcg acgtgtcatc agatgctgcc gctgaatttt
ggtgcagcgc aacgtgttga cggtggctgt gtcaccgatg cgagcggtga agctgtacat
agcctgtttt tcggtcagag ctcttttttgc acctttgcac tggcgcgcga agtgaacacc
gttcctgtcg gtgacggcgt tccgctggaa attctgggtc cgctgggttg tggtattcaa
accggtgcag gcgcagcgat caacagcctg gccattaaac cgggtcagag cctggcgatt
ttcggtggcg gcagcgttgg tctgtccgcc ctgctgggcg cactggccgt gggcgcgggt
ccggttgttg tggtggagcc gaatgatcgt cgtcgtgcac tggccctgga cctgggtgcg
tcgcatgtgt ttgacccgtt caataccgaa gatctggttg cgagcattaa agccgcgacg
ggtggcggcg ttactcacag cctggacagc actggcttgc cgccggtgat cgcaaaggcc
attgattgta cgttgccggg tggcaccgtc ggtttactgg gtgttccggc tccggacgcc
gcagtgccgg tcacgctgct ggacttgctg gtgaagtccg ttaccctgcg cccgatcacc
gagggtgacg caaacccgca agaatttatt ccacgcatgg ttcagctcta ccgtgatggt
aagttcccat ttgataaact gatcaccacg tatcgttttg agaacatcaa tgacgcgttc
aaagcgacgg aaacgggtga agcgatcaaa ccggtcctgg ttttctataa

Sequence Number (ID): 111
Sequence Name: AzTolADH1
Length: 1128
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1128
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgggttcta ttcaagattc tctgttcatc cgtgcacgcg ccgctgttct gcgtactgtc
ggtggcccgc tggaaattga aaacgtccgc attagccctc cgaagggtga cgaagtgctc
gtgcgtatgg ttggtgttgg tgtgtgccat accgacgttg tgtgtcgcga tggcttcccg
gttccgctgc cgattgtgct gggtcacgag ggcagcggta ttgtcgaggc agtgggcgag
cgtgtgacca aggttaaacc gggtcagcgt gtcgttttat ccttcaatag ctgtggtcat
tgcgcgtcct gctgcgagga ccacccggcc acctgtcacc agatgctgcc actgaacttt
ggtgcggcgc agcgcgtgga tggtggcacc gttatcgacg cgagcggcga ggcagtgcag
agcctgtttt ttggtcaaag ctctttcggt acgtatgcat tggcgcgtga agtcaatacc
gtaccggtgc cggatgcagt tccgttggaa atcctgggcc cgttgggttg cggcatccag
acgggtgcgg gtgcggctat caacagcctg gcgctgaaac ctggtcaatc gctggcaatc
ttcggtggcg gcagcgtcgg tctgtccgcc ctgctgggcg cgctggccgt gggcgcgggc
ccggtcgttg tcattgagcc gaacgaacgt cgtcgtgcgt tggcgctgga cctgggtgcg
agccatgcat ttgatccgtt caacactgaa gatttggttg cgagcatcaa agccgctacg
ggtggcggcg ttacccacag cctggacagc acgggtctgc cgccggtcat cgcgaatgca
atcaactgta ccttgccggg cggcacggtc ggtctgctgg gcgtcccgag cccagaggct
gccgttccgg tgacgctgct ggatctgctg gttaaatcag ttaccctgcg tccgattacc

gagggtgacg ccaatccgca agaatttatt ccgcgtatgg tccagctgta ccgcgacggt
aaatttccgt ttgataagct gattacgacc taccgcttcg acgacatcaa tcaagcgttc
aaggcaaccg aaaccggtga agcgattaag ccagtgctgg tgttttaa

Sequence Number (ID): 112
Sequence Name: SCH23-ADH1
Length: 768
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..768
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgggtttgt tcgctttgga ggaatgggct caagctaatg ctgactacaa taacgatact
cctccatact ggcacgccaa gattgtccct gacttattca ccgccatctc tggtattttg
tggtctgtct cttacatttt gatgactcta aagggttaca aagacagatc ttacgccatg
ccaatctatt gtctatgttt gaacatcact tgggagttcg tcttcggttt catttacggt
cctggtttgg ttaaccaaat cgttttcgcc caatacatgg ttgtcgatgt cttcttgttc
cactctattt tgaaattcgg tccaaacgaa tggagagctc acccattagt tgctagaaac
ttgtcctgga tcattggtgt tggttgtgct gtttgtttgg gtttgcactt ggttttggct
aagaccttcg ttccagttat tggtagacaa gttattttct ttactgcttg gccaatgcaa
cacatgattt ccttaggttg tgttgctcaa gtcttgtcaa gaggtcatga tgctggtcaa
tctatggcta tctggtggac tagattcttg ggtactgtca ctgccggttg ctgtttttat
tggagaatct acttttggcc agaacgtttc ggttatgctt ggactccata cggtgctttg
ttattggttg gttctcatgt cttggactta gcttttccat tcgctttggc ttatgttaga
aagcatggtg aaggtagaca agaaaaagtt aacggtaagg ctgcttaa

Sequence Number (ID): 113
Sequence Name: SCH94-03944
Length: 474
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..474
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaacttaa atgaagcgcg taccgcattt gcacgtctga gagctgccga gaacggtctg
agcccggctg agctggatga agtgtgggca gcgctggaga ctgtggcggc tgaagaaatc
ctgggtgagt ggaagggtga cgattttgcg acgggtcacc gtctgcacga gaaactgtcg
gcgagccgct ggtatggtaa gaccttcaac tctgttgaag atgcaaagcc gctgatttgc
cgtgacgaag atggcaatct gtactccgat gtcaagagcg gtaacggtga ggccagcctg
tggaatatcg agtttcgcgg cgaagtcacc gcgacgatgg tttacgatgg tgccccgatc
ttcgaccatt tcaaaaaagt tgacgacagc accctgatgg gcattatgaa tggcaaaagc
gcgttggtgt tggacggtgg ccagcattac tatttcctgc tggagcgtgc gtaa

Sequence Number (ID): 114
Sequence Name: SCH94-03944
Length: 474
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..474
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaacttga acgaagctag aactgctttc gctagattga gagctgctga aaacggtttg
tctccagctg aattggacga agtttgggct gctttggaaa ctgttgctgc tgaagaaatc
ttgggtgaat ggaagggtga cgacttcgct actggtcaca gattgcacga aaagttgtct
gcttctagat ggtacggtaa gactttcaac tctgttgaag acgctaagcc attgatctgt
agagacgaag acggtaactt gtactctgac gttaagtctg gtaacggtga agcttctttg
tggaacatcg aattcagagg tgaagttact gctactatgg tttacgacgg tgctccaatc
ttcgaccact tcaagaaggt tgacgactct actttgatgg gtatcatgaa cggtaagtct

gctttggttt tggacggtgg tcaacactac tacttcttgt tggaaagagc ttaa

Sequence Number (ID): 115
Sequence Name: SCH24-BVMO1
Length: 1839
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1839
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaccatcg atttgcaaca gccagacgca gtcccgttta cgagcagcac tttcgtcgta
ccggacccgt ccaacctggc atcccaggct caaaacagcc aactgcagag cgcgcaagag
ggcgcagagt acccggtgaa tgcacacggt gtccgcggtg acggcaccat tcacgagcgt
ccgatcaatg accgtcgtaa aatgcgcgtc atctgcgttg gtgcgggtat tagcggcctg
tatatggcga tcaaactgcc gcgcagcacc gagaatgttg aactgaagat ctacgagaaa
aaccatgacc tcggcggcac gtggctggag aatcgctacc ctggctgcgc gtgcgatgtt
ccggcgcatg cgtatgcata ttcttttgag aataatccgg aatttccacg ctttttcagc
agcagcgagg atatccacaa gtacctgttg cgtgttgcgg acaagtacga ctgtaagaaa
tacatcgcct ttaacaccaa agtcgttgag gctatctggg acgaagaaca gggtatttac
aatgtgaaga ttgagcgtag cgacggcacc gtgttccagg acacctgtga ggtgctgctg
aacgcgagcg gtattctgaa tgcctggcgc tacccgggca tccctggcat taaggattac
aaaggtacgc tgatgcacag cgctacctgg gaccgtagcg tttctttgaa aggcaaaaaa
gtcgcactga ttggcagcgg tagcagcggt atccagattc tgccgaacat tctggacgac
tgcaaagaag tggtcacgta cattatcgac ccggcgtgga ttgctccggc taacctggtg
accgcgggtg tctccgatga tggtgaggaa ccgaaagagc caacccctga ggaactggcc
tcatcctccg acttcgctta tagccaggaa cagattaacg gcttcaagaa agatccgaag
tcgctgatgg atcaccgcgc cacgctggag cgtaccatga atcaatcgtt tccgattctg
ctgcgtggct ctccgagcaa cttgtatgcc gcaagcctgt tcgaggattt gatgcgtaag
cgtctggcga agaagccgga agttgcggac gcgattatcc cggagtggag catcggttgc
agacgcctga cgccgggtcc gcattacctg gaagcactgt gtaacccgaa agtgaagatc
ctgactcagg cgatcaagag ctttagcgat aagggcatgt atactgcgga cggtgagcat
gaagatttcg atgttgtcat ttgtgcgacc ggtttcgatg tgagctttcg tccgcgcttc
aagtttattg gtaaagatgg ctatgaagtc ccagagaatt tcggccaaac gccgaaaggt
tatctggcac tggcgtacgc cggcttcccg aacagcttca tctttatggg tccgaacggt
ccgattgcga acggtagcgt tgtggtgagc ctggagaagc aaggtgacta cttcattaaa
gcgatcaata agatccagcg tcaaaacatt aagggtatga ccgttcgttt cgacgccgtg
gatgatttta cgaatcacgt ggacaaatac atggaccgta cggtgctgac cgacgattgc
atcagctggt acaagaatgg taaacgtgac ggtcgtgtta gcgcagtttg gccgggttcc
gcgctgcact atatggaagc catcgcagac ccgcgttggg aagattacac ctacacctat
cgcgaaccgg gtcactcttt tagcttcctg ggtgatggca ccagctgggt tgagcatacg
ggtggcgata ccgcctggta tttgaaagaa accctgtaa

Sequence Number (ID): 116
Sequence Name: SCH46-BVMO1
Length: 1767
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1767
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgccatctg ccattactcc acctgttgat catcgtagcc tgccgggtct gttcaagccg
caacgtaagt tgaaagtgat ctgtgttggc gcgggcgcga gcggcctgtt gctgagctac
aagattcagc gtcactttga ggactttgag ttgcaagttt ttgagaaaaa ccctgaagtg
agcggcacct ggtacgagaa tcgctacccg ggttgcgcgt gcgatgttcc gagccataac
tatacctggt cttttgagcc gaaaacggat tggtccgcaa actatgccag cagcaaagaa
attttcaagt acttcaaaga tttcaccaag aaatatggtc tgtctaaata cattaaactg
gaacacgaag tcgtgggtgc gacgtggatg gaagcggaag ctcaatggaa agttgacgtc
aaagacttgc gtagcggcaa cacccagagc tccttcgcgc acattctggt caatgccggt
ggcattctga acgcttggcg ttacccgccg attccgggta tcaaagattt taagggtgac

ctggtgcact cggcagcgtg gccggagcat ctggatctga atggtaaagt cgttggcctg
attggtaacg gtagcagcgg catccaaatt ctgccggcca tcaaaaaaga cgtgaaacaa
ctggtcacgt ttatccgtga ggccacgtgg gtcgccccgc cgctgggcca agcgtaccgc
gcatttagca ccgacgaaca ggcgcagttt gcacaagacc cgcgtcacca tctggaaact
cgtcgcgcga ccgaagctac catgaatcag agcttcggta tcttccacag cggttcagag
gaacagaaag gtgtgcgtca gtacatgcag gatatcatgg aaacgaaatt gaataacaaa
cagctggaga gcgtgctgat tccggagtgg tccgtgggtt gtcgccgtct gaccccgggc
acgaactatc tggagagctt gagcgacgat aacgtgaaag ttgtttatgg cgagatcacc
cagatcaccg agtccggtgt gatttgcgat gatggcaagg gtgagtaccc ggtcgaagtg
ctgatttgcg ctaccggttt cgacaccacg ttcaaaccgc gcttcccgtt gattggcacc
accaaagaaa agctgagcga cgtctggaaa gatgaccctc gcggttattt cggtatcgcg
accaataact acccgaacta ctttttcacc ctgggtccga actgcccgat cggcaatggt
ccggtcctgt gtgcaatcga agctgaagtg gagtatatca tcaatatgct gagcaaattt
cagaaagaaa acattcgcag cttcgacatt aaagccgacg cagttgatgc gtttaacgac
tggaaagatg actttatgaa agacaccatc tgggcagagc agtgtcgttc ttggtacaag
gctggttctg cgacgggtaa gattttggca ctgtggccgg gcagcacgct gcattatctg
gaagccctga aaagcccacg ctgggaagat tgggacttca agtatcaacc gggtcgtaat
cgctttcact acttcggtaa cggccacagc tgcgcggagc aagatggtga tctgtcctgg
tatatccgta atgaagatga cagctacatt gacccggtac tgaagccgaa gccgaaggca
gcggtggaga gcgaagcaca catcgcgctg ccaggcattg gtccgatgct gatggaggac
ccgcgtgacg ttgcggttga ggcataa

Sequence Number (ID): 117
Sequence Name: AraBVMO1
Length: 1491
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1491
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggataagc atattgacgt tttgattgtt ggcgctggta tttcgggttt gggtttggca
gcccatctgt ccaagaactg cccacagcgt agcttcgaga ttgttgaacg tcgtgagggt
atcggtggca cctgggactt gttccgttat cctggtatcc gcagcgatag cgacatgagc
accttcggtt acaatttcaa gccgtggcgt aaggcaaaaa tcctggccga tggcgccagc
atccgtcagt atctgcacga agttgtggat gaatttcatc tggatcgtaa aatccacttt
aaacatcgtg tcatcagcgc gaattatgat accgcactga agctgtggat tgtcgagatt
gaagatcagc aaggtcagaa tcagacctgg tatgcgaact ttctgctggg ctgtacgggc
tattacaatt acgatgaagg ctttatgccg gaatacccgg gtcaacacca gtttaagggt
acgctggtgc acccgcaaca ctggccggag aaacttgact atacgggtaa acgcgtgatc
gttatcggca gcggtgcaac cgctattact ctggtgccga gcatggtgaa gggtggcgcg
gcacacgtta ccatgctgca acgtagcccg acgtacattg cgagcatccc gtcaattgat
tttgtttacc agaaaatgcg cggtttcctg tctgaagaaa tggcatacaa gctgacgcgt
gctagaaaca ttggtatgca acgcgctgtc tatgcgctgt cccagaagca accaaaactg
gtgcgtaaac tgctcctgaa atccattgag atgcagctga agggtaaagt cgatatgaag
catttcacgc cgtcctacaa cccttgggac caacgcctgt gcgtcgtacc ggatggtgac
ttgttcaagg cactgcgcga gggccacgcc agcgtcgaaa ccgaccacat cgagaaattc
accgagactg gtatccagct gaagtctggc aagcatttag aggccgatat catcattagc
gcgaccggtc tgcaaatcca gatcatgggc ggcattcagg gcacggttga cggtcagccg
atcgacacca gcgagcacat gctgtacaac ggtattctga tcagcgacgt gccgaacatg
gcgatgatta ttggctacat caacgcgagc tggaccctga aagttgacgt ggcggctgag
tacatttgtc gtctgctgaa ttacatggac aaacatcact atgacgaagt tattgcgccg
acggaccata gcgagattga gcaagatacc gtgatgggta gcctgagcgc gggttatatc
cgccgtgcgg cagacgtcat cccgaaacag ggtaaacacg cgccgtggca agtcaccaac
aattacctgg cggaccgtaa ggccctgaaa caggccggtt ttgaagatgg catcctgcaa
ttcaccaagc gcgataaaca attggagcgt aagccgaaac tggtgtctta a

Sequence Number (ID): 118
Sequence Name: AflavBVMO1
Length: 1632
Molecule Type: DNA

Features Location/Qualifiers:
- source, 1..1632
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized

Residues:
atgaatggta ctcaagcaag caacggtgtt ttgcatctgg acgcactgat tattggcagc
ggcttctccg gtatctacct gttgcacaag ctgcgcgatg aactgaagct gaaagtgaag
atctttgagg ccgagagcga tatcggtggc acctggaata acaaccgcta tccgggtgcc
cgtgtcgatt gcccggtccc gttctacgcg tattcactcc ctgaagtttg gcagagctgg
aattggaccg agctgtatcc aaatcagaaa gaaatcaagt cctacttcga ccatgtggat
cgtgttcttg acgtgcgtaa agattgcctg tttcactccc gtgtgaatga gggtacgttc
gacgaagcga ccggtcgttg gaccgtctgg accaccgatg gtaaagtcgc gacggcgaag
tacttgctgg tcgccgtggg tttcgcatcg aagtcttact tgccggactg gaagggtctg
gatagcttta agggcaccat ctaccatagc gcgcactggc ctgaggccga agaaatttcc
gttaaaggca aaaaagttgc agtaatcggt acgggtagca ccggtattca gatttttcaa
gagtgggcgc gcgaagcgga agaagccttt ctgttccagc gcacgccgaa cctgtgtctg
ccgatgcgtc agcaagaact gcacgcgggt tatcaagtca aagacaaggg cgagtacgca
gattacctgg ctgagtgcgc gctgacgttc ggtggcctgg agtatcagca gactccgaaa
aacacgtttg acgcgagcga agaggaacgt gaggctttct gggaagatct gtaccagatg
ggtggctttc gtttctggca aaacaactat caagacctgt tgactagcct ggatgcgaac
cgtgaggctt ataacttttg ggcgagaaag acccgtgcac gcatccaaga tccgaaaaag
cgtgacttac tggcgccgtt ggagccacca tacccgttcg gtacgaagcg cccgagcctg
gagcaagact tctacgagca attcaacaag tctaatgtcc acattgttga caccaagagc
cagccgatcg tgggcgttac cccgaccggt attgttaccg cagatgaaaa agtgcatgaa
gttgacatta ttgcggtcgc cacgggtttc gatgccgtta ccggtggctt gctgcgcctg
ggtctgaaag acgtgaatgg cgttggtctg gacgagcgtt ggaaagatgg catgagcacc
tatctgggta tggcaattag cggctttccg aatatgtttc tgccgtatag cctgcaagct
ccgaccgcct tcgcaaacgg cccgacgctg atcgagttgc agggtgactg gatcaccagc
ctgatccgta aaatggaaat ggaaaacgtc cagagcgtta ccgcaacccc gcacgcggag
agcgcgtgga acgacgaagt gaatatgatc gcgaataaga cgctgctgcc gctgacggat
tcttggtaca tgggcagcaa tattccgggt aaaccggtgc agagcctgaa ctacctgggt
ggcctgccga cctaccgtga gcgctgtgcg aaagtgctgg acgaagattt ctttggtttt
gcgaaagctt aa

Sequence Number (ID): 119
Sequence Name: AflavBVMO1
Length: 1632
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1632
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized

Residues:
atgaacggta cccaagcttc taatggtgtt ttgcacttgg atgctttaat tatcggttcc
ggtttctctg gtatctattt gttacataaa ttgcgtgatg aattgaagtt aaaagtcaag
attttcgaag ctgaatctga catcggcggt acttggaaca ataaccgtta tccaggtgcc
agagttgact gtccagttcc attttacgct tattccttgc cagaagtctg gcaatcttgg
aactggactg aattataccc aaaccaaaag gaaattaagt cctacttcga ccacgttgat
agagttctag acgttagaaa ggattgtttg tttcactcta gagtcaacga aggtaccttc
gacgaagcta ccggtagatg gacagtctgg accactgatg gtaaggtcgc tacagctaag
tacttattgg ttgctgttgg tttcgcctct aagtcttact tgccagactg gaaaggttta
gattctttca aaggtaccat ttaccactcc gcccactggc cagaagctga ggaaatttcc
gtcaagggta aaaaggttgc cgttatcggt actggttcta ctggtatcca aatttttcaa
gaatgggcca gagaagctga ggaagctttc ttatttcaaa gaacaccaaa cttatgtttg
ccaatgagac aacaagaatt acacgctggt taccaagtta aggataaggg tgagtacgcc
gactacttgg ctgaatgtgc cttgactttc ggcggtttgg aataccaaca aaccccaaag
aatacttttg atgcttccga agaggaaaga gaagcctttt gggaagactt gtaccaaatg
ggcggtttta gattctggca aaataactac caagatttat tgacatcatt ggacgctaat
agagaggcct acaacttttg ggccagaaag actagagcta gaattcaaga tccaaaaaag
agagatctat tggctccatt agagcctcca tacccattcg gcactaaaag accttctttg
gaacaagatt tctacgaaca attcaacaag tctaacgtcc acatcgttga cactaaatct

caaccaatcg ttggtgtcac tccaactggt atcgtcaccg ctgatgagaa ggttcacgag
gttgatatta tcgccgtcgc tactggtttt gatgctgtca ctggcggtct attgagattg
ggtttgaagg atgttaacgg tgttggtttg gacgaaagat ggaaagatgg tatgtccacc
tacttgggta tggctatctc cggttttcca aatatgttct tgccttattc tttgcaagct
ccaaccgctt tcgctaatgg tccaactcta atcgaactac aaggtgattg gattacctcc
ttaattcgta agatggaaat ggaaaacgtt caatccgtta ccgccactcc tcacgccgaa
tccgcttgga atgatgaagt caacatgatt gccaataaaa ctttgttacc attgactgat
tcctggtaca tgggttccaa tatcccaggt aaaccagttc aatccttaaa ctacttaggc
ggtttgccaa cctacagaga aagatgtgct aaggttttgg atgaagattt tttcggtttc
gctaaagctt aa

Sequence Number (ID): 120
Sequence Name: SCH24-EST1
Length: 933
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..933
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgacccact cgccgccact ggatgccgaa ctgagcttgc tgcgctacgc ccctgccgtt
ccggtgggtt ggcagctggg tcgcaaactg ctgcgtatga acaccttgat gacccgtccg
atggaaggtg tcatgcgcga cgatgtggtt attccgaatc tggacggcac ggctaacatc
cgtctgttta tctgtcgtcc gcaagacccg accgagacta tgccggttat cctgtggctg
cacggtggcg gcatggtcgc aggccactac aaacaagaca gcggtttcat ggacatttgg
gcgaagcgcc tgggtgcgtt tgttgttagc gttgattatc gcctggcgcc tgaggctaag
gcaccggcag cgctcgatga ctgcatcgcg gcgtggcagt ggattaccac ccagaccgcg
cgtggtattg acaccactcg tatggcagtg ggtggtgcga gcgcgggtgg cggtctggcg
gcaagcacgg ttcagcgtct tgtcgatctg ggcggtgtga aaccggtctt tcaactgctg
atctatccga tgctggacga tcgtaccgtg gtgcgcttcg acccggatcg tcgttattac
atgtggacgc cggactgcaa cagatacggc tggaccagct acctgggcgt gccaccgggt
agcgcagagg tcccgccgta tgcctccgcg gctcgtcgtc cggatctgtc cggcctgccg
ccgacgtgga tcggtgtcgg ctctctggat ctgttccatg acgaagatat ggattacgca
cgtcgtttgc gcgagagcgg tgtgccggtc gaagagtatg ttgctgtggg tgccccgcat
gcgttcgaca cgatttacgg caaggccaaa gttacgctgg acttttggga tagccacttc
aatgcgctgc gccgtgcgtt gtgtttagac taa

Sequence Number (ID): 121
Sequence Name: SCH23-EST1
Length: 1038
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1038
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgccgtctg acttgccacg cccagcttac gaccctgaaa ttgagccatt cttatctatg
gtcccactgc cgccgacgat caatgcagat atcatgaaag aactgcgtaa ggctccgctg
ctgagccaag cccctgacct ggacgcgctg ctgtccgaca agccgattac tcatcgtgaa
gtttcgattc cgggtctgaa ctcgcaggac ccgcagatca ccctgagcat cttttcgtct
actctggaag gcggcccgaa gccgtgtatt tactttgtgc atggtggtgg catgatcatc
ggttgtcgtt tcgttggtat cgaagattac ttgcaatatg tcgagcaaaa cgacgctgtc
gttgttgcgg ttgagtaccg tctggcacca gagcacccgg acccggctcc ggttaatgac
tgctatgcgg gcctgctgtg gacggctgca aacgctgcgg aactgggtat tgatttggag
cgtttgctga tttgcggtgc gagcgccggt ggcggcctga gcgcaggcgt ggcgttgatg
gcgcgcgata agaaaggtcc gaaactggtg ggccagctcc tgtgctatcc gatgctggac
gatcgcaatg actccctgtc tagccagcaa tacgtggatg agggtgtgtg gagccgtggt
agcaacgcgt tcggttggaa gcaacttctg ggtgaccgcg cgggcaagga aggtgtgagc
atctatgctg cgcctgcccg tgcgaccgat ttgtcaggtt tgcctaatac gttcatcgat
gttggtagcg ccgaagtgtt ccgcgacgaa gatattgctt acgcgagccg cctgtgggca
gttggcgtcc aggctgagct gcacgtctgg ccgggtggct accacgccgc tgagaacatg

gcaccaggca ccgactactc caaaaaagtt aaggcaaccc gtctggcgtg gatgaaacgt
gtctttatga aagcaccgaa aagcaccacc gagagcctgc cggcgccgac tgttgacgaa
gctgtgggca ccatctga

Sequence Number (ID): 122
Sequence Name: SCH23-EST1
Length: 1038
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1038
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgccatctg acttgccaag accagcttac gacccagaaa tcgaaccatt cttgtctatg
gttccattgc caccaactat caacgctgac atcatgaagg aattgagaaa ggctccattg
ttgtctcaag ctccagactt ggacgctttg ttgtctgaca agccaatcac tcacagagaa
gtttctatcc caggtttgaa ctctcaagac ccacaaatca ctttgtctat cttctcttct
actttggaag gtggtccaaa gccatgtatc tacttcgttc acggtggtgg tatgatcatc
ggttgtagat tcgttggtat cgaagactac ttgcaatacg ttgaacaaaa cgacgctgtt
gttgttgctg ttgaatacag attggctcca gaacacccag acccagctcc agttaacgac
tgttacgctg gtttgttgtg gactgctgct aacgctgctg aattgggtat cgacttggaa
agattgttga tctgtggtgc ttctgctggt ggtggtttgt ctgctggtgt tgctttgatg
gctagagaca agaagggtcc aaagttggtt ggtcaattgt tgtgttaccc aatgttggac
gacagaaacg actctttgtc ttctcaacaa tacgttgacg aaggtgtttg gtctagaggt
tctaacgctt tcggttggaa gcaattgttg ggtgacagag ctggtaagga aggtgtttct
atctacgctg ctccagctag agctactgac ttgtctggtt tgccaaacac tttcatcgac
gttggttctg ctgaagtttt cagagacgaa gacatcgctt acgcttctag attgtgggct
gttggtgttc aagctgaatt gcacgtttgg ccaggtggtt accacgctgc tgaaaacatg
gctccaggta ctgactactc taagaaggtt aaggctacta gattggcttg gatgaagaga
gttttcatga aggctccaaa gtctactact gaatctttgc cagctccaac tgttgacgaa
gctgttggta ctatctaa

Sequence Number (ID): 123
Sequence Name: A0A0C2RLD5_V1
Length: 1848
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1848
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgcaaaacc aaattaatga actaattggt caattaatta acgaaatcga agctcaccaa
gctccagacg gttctttcag attctgtttt gaaaattcct tgatgactga tactaacatg
atgttactat taaagacttt cactaccgac gaatctttga tgaacgaatt aggtgctcgt
gttgcttcct tgcaacaccc agatggttac tggtccattt acccaggtga taacggtaac
ctatccgcta ctattgaagc ttacttcgcc ttattgtaca ctggttattg gaaaaagtct
gaaagatccg cccaaaaggc ttctcgtttc atcttggccc aaggcggtct aggtgccgct
cactccatga ctaagatcat gttggctatc catggtcaat acccttggcc aaacttgtta
cacttgccag ttcacttcat cctatttcca tcttcctctc cagtttcctt ctacgatttt
tcttcctacg ctagagttca tatggctcca attctattgc taaccgattc aagattccaa
ttgggtgaaa aggacatccc agacatgtcc gctttgttaa ttagaacccc ttccgagcaa
ttgaacgagc attccagatc tttgttatct ggtatttacc aaaccgcctc ctctatcgct
ggtttaccac acttggtcca ccaaagagct cgtaaaagat tggtcactta catgttgcaa
cgtatcgaag gtgatggtac tttgtattct tacgtctcct ctactttcta catgttgttc
gctttggtct ctgaaggtta ctctaaacaa catccattaa ttcaaaaggc tattactggt
ttgaaaactc acagatgctt gaccaacaat ggttggcaca tccaaaactc tacctctaca
gtttgggaca ccgctctatt gtgtcacgct ttacacgact tgccattctc cccatacaag
agaaaaaagg ccgaagttta cctattaaaa catcaacatg acaagtttgg tgactggatc
ttgacagccg accaaacctc tcctggcggt tggggtttct ccgattccaa cactattcac
ccagatgttg atgacactac cgcttctttg agagctctat caccatctat tactgtcgac
ccttctttga aagaatctta cttaagaggt gtttcttggg tcttagctat ccaaaacgaa

gatggcggtt ggccagcttt tgaacgtgct aagactaacc aactattgac tttcgttcca
atggacggcg cttctcacgc tgccattgac ccatctactg ctgatttgac tggtcgtact
ttggaattct tgtcatccga agccagattg ccattccaac attccgctat tcaacacgct
atcagatggt taaaaaagaa tcaacaatcc gacggttctt ggtatggtaa gtggggtatt
tccttttttat atggtacctg gtctgctgtc accggtttgt ccgccgctgg tttgaatggt
gaagatccag ccattaaaaa ggccgtttct ttcttagaaa gagtccaaaa cgaagacggc
ggttggggtg aatcctgttt atccgatcaa gttatgcact acatcccttt aggttcttcc
acccccttctc aaactgcctg ggctttagac gccctattgt ccgttcacaa acaaaagact
ccttcaatcg aaagaggtat taactgttta ttgggtcaac taaagactaa ggactggact
tacagatacc caaccggtgc catgctacca ggtaacttct acgtttatta ccactcctac
aactacattt ggccattgat cgttttgaaa aagtacgccg ccttgtaa

Sequence Number (ID): 124
Sequence Name: A0A0J5GUC6
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgattgaca aggttaatgg taaaatgaac tcattaacct ccgaattatt gagaaagcaa
gctaaggatg gttcttggaa attctgcttc gaaggttcta tcatgaccga cgcttacatg
atcattttaa ttagagtctt ggaaattaag gatgaagagg acttggtccg tagattagtt
aacagaatta aaaccaagca atctccaaac ggcgcctgga aattgtatga agatgaaaag
gacggtaact tatcagctac cgttgaaggt tactttgctt tattgtattc aggttacact
gagagaaaag aaagaaacat gcgtaaggct gaaaacttca tcaaccaaca cggcggtttg
tcttcctgtg actggttgac tagaatgatg ttggctttga acggccaaat cgaatggcca
ggtatcgtca aatctattcc aattgaaatt atgttggttc caaagtgggc cccaatcaac
atttatcatt tagtcggcta cgctagagcc cattgggttc caatcattat ctcctctaat
ttgaacgctt ctttagtcac tgctcaaacc ccaaaattgt ctcacttgca aactagagaa
tctggttccg aagactacag attattggag gaaatgaagc tattgtccca ttacgtcaac
tccgctatga aaaagttagc cgcttctcca gaattattgc gtaagagagc ttttactaag
gctgaaaact atatttctga aagaatcgaa gagaacggta ctctatactc ttatttttct
gcttcttttt tcatgatctt tgccttctta gctttgggtt acgataagaa tcatccaaag
attaaaaacg ccttcagagg tatgaagtct tacatctgtt ctactcactc tgacgataag
tttttcgttc aaaattctcc ttcaaccgtc tgggacaccg ccttattgac tggtgctcta
gtccaagctg gtgtcttagt tcatcacgaa gctatgatgt ccgctggtga atacctattg
tctagacaac accataagta cggtgattgg gctttgaaga acccagacgc tccacctggc
ggttggggtt tttccgacat taacacccta gtcccagacg tcgacgatac taccgccgct
ttgagagtca tcactctatt ggcccattct gatgagagat acaaatctgc ttggaataaa
ggtgttgaat ggttgttatc tatgcaaaac gatgacggcg gttggtccgc tttcgaaaag
aacacagata actatttgct atcttttgtt ccattcagat acgaagacag agtcttcttt
gatccatcca ctgccgattt aacaggtaga accttgtact tcttgggtga atacactacc
atcccacaag actctgaaaa aatcttagaa gctgttaaat ggttgaatta ccaccaagaa
aaggatggtt cttggtacgg tcgttggggt aattgttaca tttatggtac ttgggccgct
gttaccggtt tgaaggctgt tggtgttcca tctaccgact ctactatcac tcgtgctgtt
gaatggttat tgtccattca aaacgaagat ggcggttggg gtgagtcttg tttttctgac
gtcagaaaca agtacgttcc attgcatcac tctactcctt ctcaaactgc ttgggctttg
gacgccttga tttctgtctc tgaatctcca actccacaaa ttgaaaaggg tattagaacc
ttattggact tgatggaaac taacgattgg agaaccgatt acccaactgg cggtgctatc
ccaggcggtt actatatcca ctatcactca tacaagtaca tctggccttt gcaagctttg
ggtcactacc gtaacaaatt cgatttgaag acttaa

Sequence Number (ID): 125
Sequence Name: A0A0J5GUC6_V1
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA

> organism, synthetic construct E.coli optimized

Residues:

atgattgata aagttaacgg caagatgaat tcgctgacgt ctgagctgct gagaaaacaa
gcgaaggatg gttcttggaa attctgtttt gagggtagca ttatgactga cgcttatatg
atcatcctga tccgcgtctt agagattaaa gatgaagaag atcttgtgcg tcgtctggtg
aaccgtatta agactaaaca gagcccaaat ggtgcgtgga agctgtatga agatgagaaa
gacggcaatt tgagcgcaac cgtcgagggc tacttcgcac tgctgtatag cggttacacc
gaacgtaaag agcgtaatat gcgtaaagcc gagaatttca ttaaccagca cggtggcctg
tccagctgcg actggctgac gcgtatgatg ctggcgctga atggtcaaat tgagtggcca
ggtattgtta agagcattcc gatcgagatt atgctggtcc cgaaatgggc accgattaac
atttaccatc tggttggtta tgctcgcgca cattgggtgc cgatcattat ctccagcaac
ctgaacgcaa gcctggtcac cgcccaaacg ccgaagttgt ctcatctgca aacccgtgag
agcggtagcg aagattatcg cctgctggaa gaaatgaaac tgctgtccca ctacgttaac
tccgcgatga agaaattggc ggcaagccca gagttgctgc gtaagcgtgc cttcaccaaa
gccgaaaact acattagcga acgcatcgaa gagaatggta cgctgtatag ctactttagc
gcgagctttt tcatgatctt cgcgtttctc gcattgggtt acgacaagaa tcatccgaag
atcaaaaacg cgtttcgtgg catgaaatca tatatctgca gcacccattc cgacgataag
tttttcgtcc agaatagccc gagcacggtg tgggataccg cgctgttgac cggtgcgctg
gttcaggctg gtgtcctggt tcaccacgaa gcgatgatga gcgcgggtga gtatctgctg
agccgtcaac accacaaata cggtgactgg gccctgaaga atccggacgc gccgccgggt
ggctggggct tcagcgatat caacacgctg gtgcctgatg ttgatgatac cacggcagcg
ctgcgtgtta ttacgctgct ggcacactcg gacgagcgct acaagagcgc gtggaataaa
ggcgtggagt ggctcttgag catgcagaat gacgacggtg gctggtcggc ttttgagaag
aacacggaca actacttgtt aagctttgtc ccgtttcgct acgccgaccg cgtgttcttc
gacccgtcta ccgcggatct gaccggtcgt accctgtatt tcctgggcga gtacaccacc
atccctcagg acagcgaaaa gatcttggaa gccgtaaaat ggctgaatta ccatcaggaa
aaggacggct cctggtacgg tcgttggggt aactgctaca tttatggcac gtgggcggca
gtgacgggtc tgaaagccgt tggtgttccg agcaccgaca gcaccattac ccgtgcggtc
gagtggctgc tgagcatcca gaacgaagat ggcggctggg gtgagagctg ttttagcgat
gttcgtaaca aatatgtgcc gctgcaccac agcaccccgt cccagaccgc gtgggcgctg
gacgcactga ttagcgtgag cgagagcccg accccgcaaa ttgagaaagg catccgcacc
ctgctggacc tgatggaaac gaacgactgg cgtactgatt acccgacggg tggcatgatc
ccgggtggct attacatcca ctatcattct tacaagtata tctggccgct gcaagctctg
ggtcactacc gcaataagtt cgatctgaaa acctaa

Sequence Number (ID): 126
Sequence Name: A0A0J5GUC6_V1
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized

Residues:

atgattgaca aggtcaacgg taagatgaat tccttaacct ccgaactatt gcgtaaacaa
gctaaggacg gttcctggaa gttttgtttc gaaggctcca ttatgactga cgcttacatg
atcattttga tcagagtttt ggaaatcaag gacgaggaag acttggtccg tagattggtt
aacagaatca aaactaagca atctccaaat ggtgcttgga agttgtatga ggacgaaaag
gacggtaact tgtccgccac agtcgaaggt tattttgctc tattgtactc tggttatact
gaaagaaagg aacgtaacat gcgtaaggct gaaaacttca ttaaccaaca cggcggtttg
tcctcttgcg attggttgac tcgtatgatg ttggctttga acggtcaaat cgaatggcca
ggtatcgtta agtctatccc aatcgaaatt atgttggtcc caaaatgggc tccaatcaac
atctaccatt tagttggtta cgctagagcc cactgggttc caatcattat ctcctctaac
ttgaacgcct ctttggttac agcccaaacc ccaaagttgt cacatttaca aactagagaa
tccggttcag aagattacag actattggag gaaatgaagt tgctatctca ttatgttaac
tctgctatga aaaagctagc cgcttcccca gaattattga gaaagagagc tttcactaag
gctgaaaact acatttctga acgtatcgag gaaaacggta ccttgtactc ttacttttct
gcctcttttt tcatgatttt tgctttcttg gccttgggtt atgataagaa ccatccaaag
atcaaaaacg ctttccgtgg tatgaagtcc tatatttgtt ccacccattc agacgataag
tttttcgttc aaaactctcc ttccaccgtt tgggacaccg ctctattgac cggtgcttta
gtccaagctg gtgttttggt tcaccatgaa gctatgatgt ctgccggtga atacttgtta

tctagacaac atcacaagta cggtgactgg gctttgaaga acccagacgc tccacctggc
ggttggggtt tctctgacat taacactttg gttccagacg tcgatgacac taccgctgcc
ttgagagtca tcactttatt ggctcactcc gacgaaagat acaagtccgc ctggaacaaa
ggtgtcgaat ggttattgtc catgcaaaac gatgacggcg gttggtctgc tttcgaaaag
aacaccgata actacttatt gtccttcgtc ccattccgtt acgctgaccg tgttttcttt
gacccatcta ccgccgacct aaccggtaga actttgtact tcttaggtga atacacaact
atcccacaag actccgaaaa gattttggaa gctgtcaagt ggttgaacta tcaccaagaa
aaagacggtt cttggtacgg tagatggggt aattgttaca tctacggtac ttgggccgct
gtcactggtc taaaggctgt cggtgttcct tctactgact ctactattac cagagctgtt
gaatggttat tgtcaatcca aaacgaagat ggcggttggg gtgaatcttg tttttctgac
gttcgtaaca agtatgtccc tttgcaccat tccactccat cccaaaccgc ctgggccttg
gatgccttga tctctgtttc tgaatcccca acccctcaaa ttgaaaaagg tattagaacc
ctattggact tgatggaaac caatgattgg agaaccgact acccaaccgg cggtatgatt
ccaggcggtt attacattca ctatcactca tacaaataca tttggccatt gcaagcctta
ggtcactaca gaaacaaatt tgacttgaag acctaa

Sequence Number (ID): 127
Sequence Name: A0A0M9GQE0
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaagccag ttgataagat caacgaaaag ctaaaggata tgattaataa cttattggac
acacaaactg aaaagggtac ttggaacttt tgtttcgaag gttctattat gactgatgct
tacatgatca ttttgattag aattttggag ttgacagatg aggaactatt ggtcaagtct
ttggttgaaa gaatcaagtc caagcaagag tccaacggtg cttggaaagt ttacccagat
gaggacggcg gtaacttgtc tgccactatt gaaggctact tttcattatt gcactctggt
tacgtcgcca aagatgccgc taacatgaga aaggctgaaa gattcattca agaaaacggc
ggtttggccg aatccgactg gctaaccaag atgatgttgg ctctaaccgg tcaaatcaag
tggccatcta tcattaagat cattccaatc gaaattatgc tattaccaaa ttggtcccct
atctccattt accaattcgt cggttacgct agagctcact ggattccaat tttgatctgt
tctaacttga actattctta cttgcactcc agaactccta acttgattca cttgcaaggt
gttggttctg actctgaaga tcaaagattt gtcgaagata gacaacactt gcaactatac
ttcaaaaacg ctttgaaaaa gttgactggt tcaccagaag ttttgaaaag aaaggctttc
atcaaggccg aagactacat cttggaaaga attgaggaaa acggtactct atactcctat
ttctccgcct ctttttttcat ggttttcgct ttcttagctt taggttacga caaaaagaac
ccattaatcc aaaacgcttt ccaaggcatg aaggcttacc tatgtagaaa cgctgaccat
gccttcattc aaaattctcc atccaccgtt tgggataccg ctttgttatc cgccgctttg
caacaagctg gtgtcccaca tcaacatgcc tccattatga aggcttccaa ctacttattg
tccaagcaac aacaaaagta cggtgattgg gctattaaga acccagatgt caccccaggt
ggctggggtt tctccgacac taacacattc gtcccagaca ttgatgacac taccgccgct
ttgagagcca tcactccatt ggctggtact gagaaccact tcaaacacgc ttggaacaag
ggtgttgaat gggttttaac tatgcaaaac gatgacggcg gttggtctgc ttttgaaaag
aacactgata actacttgtt atctttcatc ccattcaagt acgaagacag agttttattc
gacccatcta ctgctgactt gaccggtaga actttgtact tcttgggtga atacactacc
atcccacaag aatccgatgt tttccgtaaa gccatggaat ggttagaata caaccaagaa
gacaatggtt catggtacgg tagatggggt aactgttaca tttacggtac ttgggccgct
atcactggtt tgaaggctat tggtgtttct tccgatgacc ctctaatccg tagagctgtt
aagtggctat tgtctattca aaatgaagat ggcggttggg gtgaatcctg tgaatccgat
tctaaaaaga gatacatccc attgcaccat tcaaccccat ctcaaactgc ctgggccttg
gatgctttaa tttctgcctc cgagaatcca acaccaaaga ttgaaaaggg tatccactct
ttgctaagat tattgaaggc cgacgattgg cgtaccactt atccaactgg tgctggtatc
ccaggtggct actatatcca ttatcactca taccaatata tctggccact attggcttta
tcacattaca agaataaata tgacgaaacc gaataa

Sequence Number (ID): 128
Sequence Name: A0A0M9GQE0_V1
Length: 1892

Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1892
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaagcctg tggataagat taatgaaaag ctgaaagata tgattaacaa tttgctggat
acgcagaccg agaaaggtac gtggaatttc tgctttgagg gcagcatcat gaccgacgca
tacatgatca tcctgattcg catcctggaa ttgaccgacg aagaactgct ggttaagtcc
ctggttgagc gcattaagag caaacaagag agcaatggtg cctggaaagt gtatccggac
gaagatggcg gcaacctgag cgcgacgatc gaaggctatt tcagcctgtt gcatagcggt
tacgttgcta aagatgccgc taatatgcgt aaagccgagc gtttcatcca agagaacggt
ggccttgcgg aaagcgattg gctgacgaag atgatgctgg ccctgaccgg ccagatcaag
tggccgagca ttatcaagat tatcccgatt gagatcatgc tgctgccgaa ctggagcccg
atttcgattt accagttcgt gggctacgcg cgcgcgcact ggattccgat cttaatctgc
agcaacctga actattcgta tctgcacagc agaaccccga acctgatcca tctgcagggt
gtcggcagcg atagcgaaga tcaacgcttt gtcgaggacc gtcaacatct gcagctgtat
ttcaaaaatg cgctgaagaa actgaccggt agccctgaag ttctgaagcg caaggcattt
atcaaagctg aggactacat tctggagcgt atcgaagaga atggtacctt atacagctac
ttttccgcaa gcttcttcat ggtgtttgcg ttcctggccc tcggttacga caagaaaaat
ccgttgattc agaacgcgtt tcaaggcatg aaagcgtacc tgtgtcgtaa tgcagaccac
gcgtttatcc aaaacagccc gagcacggtt tgggataccg cgctgctgtc cgcggcactg
caacaggccg gtgtgccgca ccagcacgct agcattatga aagcgagcaa ctacctgttg
agcaagcaac agcagaaata tggtgattgg gcgatcaaaa acccagacgt cacgccgggc
ggctggggtt tctctgacac gaataccttt gtcccggaca ttgacgacac caccgccgca
ctgcgtgcaa tcacgccgct ggcgggtacc gagaaccatt tcaaacacgc gtggaataaa
ggtgtagaat gggtgctgac catgcaaaac gatgatggtg gttggagcgc tttcgagaag
aacaccgata actacttgct gtccttcatc ccattcaaat atgcagatcg tgttctgttt
gacccgtcta ccgcggattt gacgggtcgt accttgtatt ttctgggtga atatacgacc
attccgcaag agagcgacgt ctttcgtaag gcaatggaat ggttggagta caaccaagag
gataatggct cttggtatgg ccgttggggt aattgctaca tttacggtac gtgggcggca
attactggcc tgaaagctat cggtgtgagc tccgacgacc cgctcattcg tcgcgcagtt
aagtggctgc tgagcatcca gaatgaagat ggtggctggg gtgagagctg tgagagcgac
agcaaaaagc gttacattcc gctgcaccat agcaccccgt cacagaccgc gtgggcactg
gatgcgctga ttagcgcgtc tgagaacccg acgccgaaga tcgaaaaggg tattcactcc
ctgctgcgtc tgctgaaagc cgacgactgg cgcaccacct acccaactgg tgccatgatc
ccgggtggct attacatcca ctatcattct tatcagtaca tttggccgct gctggcgctg
agccactaca agaataaata tgacgaaact ga

Sequence Number (ID): 129
Sequence Name: A0A0M9GQE0_V1
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaagcctg ttgacaagat caacgaaaag ttgaaggata tgatcaataa cttattggat
acccaaaccg aaaagggtac ctggaacttt tgtttcgaag gttctattat gactgacgct
tacatgatta tcttgatcag aatcttagaa ttgaccgacg aggaattgct agttaagtct
ttggtcgaaa gaatcaagtc caagcaagaa tccaacggtg cttggaaggt ttacccagat
gaggacggcg gtaacttgtc tgctaccatt gaaggttact tctccttatt gcactccggc
tacgttgcta aggatgccgc taacatgcgt aaggctgaaa gattcattca agaaaacggc
ggtttggctg aatctgactg gttaactaaa atgatgttgg ccttgaccgg tcaaattaaa
tggccatcta tcattaaaat cattccaatt gaaatcatgt tgttacctaa ttggtcccca
atttccatct accaattcgt cggttatgcc agagctcact ggatcccaat tttgatttgt
tctaacttga actactctta tctacactcc agaaccccaa acttgattca cttacaaggt
gtcggttctg actctgaaga tcaaagattc gtcgaagaca gacaacattt gcaattgtac
ttcaagaacg ctttgaaaaa gttaactggt tctccagaag ttttaaagag aaaagctttc
atcaaggccg aagattacat tttggaaaga attgaggaaa acggtacctt gtactcttac

ttttctgctt ccttttttcat ggttttcgct tttctagctt tgggttacga caaaaagaac
ccattaatcc aaaacgcttt tcaaggtatg aaggcttact tgtgtagaaa cgctgatcac
gctttcatcc aaaactctcc atctactgtt tgggacactg ccttgttatc cgctgcctta
caacaagctg gtgttccaca ccaacatgct tctattatga aagcttctaa ctacttgtta
tcaaagcaac aacaaaagta cggtgactgg gctattaaaa acccagatgt cactccaggc
ggttggggtt tctctgatac taacactttc gttcctgata tcgatgacac taccgccgct
ttgagagcta tcactccatt ggctggtacc gagaaccatt ttaagcatgc ttggaacaag
ggtgttgaat gggttttgac tatgcaaaac gacgatggcg gttggtctgc tttcgaaaaa
aataccgata actacttatt gtccttcatt cctttcaagt acgctgatag agtcttgttc
gacccatcta ccgctgacct aactggtaga acattgtact tcttgggtga atacaccact
attccacaag aatctgacgt tttcagaaag gctatggaat ggttagaata caaccaagaa
gataacggtt cttggtatgg tagatggggt aactgttata tttacggtac ttgggccgct
attactggtt tgaaggctat tggtgtttca tctgacgatc cattgattcg tagagccgtc
aagtggctat tgtccatcca aaacgaagac ggcggttggg gtgaatcttg tgaatctgac
tccaaaaaga gatacatccc tttgcatcac tccaccccat ctcaaaccgc ctgggctttg
gatgccttga tctccgcttc tgaaaatcca actccaaaga ttgaaaaagg tatccactct
ttattgagac tattgaaggc tgacgattgg agaactacct acccaactgg cgctatgatt
ccaggcggtt attacattca ttaccactct tatcaataca tctggccatt attggctttg
tcccattata agaacaagta cgatgaaaca gaataa

Sequence Number (ID): 130
Sequence Name: A0A1Y0CKF4_V1
Length: 1889
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1889
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaatacgg aaaacaagat taatgaaaag ctgaaagaaa tgatttcgac gctgctgagc
aagcaaagcg acaacggcgc gtggaacttt tgcttcgagg gttctattat gactgatgcg
tacatgatta tcctgattcg taccctggaa atcaccgatg aagaggtcct ggtgaaagac
ctggttgagc gtattaagag ccgccagtcc ccgaacggtg cctggaaagt ttacccggat
gagaacaaag gcaatctgag cgcaaccatt gaaggctact ttagcttgct gtatagcggt
tacgtgggtg aagaagcgag ctatatgcgt aaagctgagc gcttcatccg cgataacggt
ggcctggcga agtccgattg gctgaccaag atgatgctgg cgctgaccgg ccagatcaag
tggccgtcca tcatcaagat tatcccgatc gagattatgc tgctgccgcg ttggagcccg
atcaccattt accaactggt tggttacgcg cgtgctcatt ggatcccaat tcttatttgc
agcaacttaa acaagtcttt tgttgctccg caaaccccga acatctctca tctgcaagag
cgtctgatgg acagcgagaa tgatcgtatc ctcgaagaaa tgcagaattt gcaactgtac
ttcaaacacg cgctgaaaaa gctgagcaag agccctgaga ttctgaaaaa agaagcgttc
attaaagctg agaactacat tattgagcgc atcgaagaga atggcaccat gtatagctat
ttctcagcat cctttttcat ggtgttcgca tttttggcac tgggttacga tgcaaaccac
ccgctgattc gtaacgcgtt tcagggtatg aaaagctatc tgtgtcgcaa tgcagagcag
ccgttcatcc agaatagccc gtctacggtt tgggacacgg cattgctgac cgcggcgctg
caacaggccg gtgtcagcta ccgtcacagc agcattatga aagccaataa ctatctgctg
agccgtcaac accagaaata tggtgactgg gccgtcaaca atccagacgt aattccgggt
ggctggggct ttagcgatat caacaccttc gttccggaca tcgatgacac caccgcagcc
ctccgtgcga tcaccccgtt gacccagacg aatattctgt acaaggaagc gtggaataaa
ggtgtcgagt ggatcctgtc tatgcagaat ggcgatggtg gctggagcgc cttcgagaag
aacatggaca actacctgtt gagcctgatt ccgtttaaat acgcggaccg cgtgctgttc
gacccgagca ctgcggattt gacgggtcgt accctgtatt tcctgggtga gtataccacc
attccgctgg agtccgagat ctttcagacg gcgaaggaat ggtttgagcg taaccaagaa
gccaatggtt cctggtacgg tcgctggggc aattgttaca tctacggtac gtgggctgcc
attacgggcc tgaaagcaat tggtgtgagc aatgatgacc caatcattag ccgtgcggtc
cgctggctgc tgagcgtgca aaatgaagat ggcggctggg gtgagagctg cgcaagcgac
atcaagaaac gttatatccc gctgccgcac agcactccgt cgcaaaccgc atgggcgctg
gacgcgttga ttagcgcgag cgacaatccg accagccgta tcgaagttgg catccacgcc
ttactgaaca ttctggaagc gaatgatggt agatccaact atcctacggg tgcaatgatc
ccgggtggct actatatcca ttaccatagc tataagtaca tttggccgct gcaggctttt
agccactata agaacaaata tgaccttaa

Sequence Number (ID): 131
Sequence Name: A0A1Y0CKF4_V1
Length: 1890
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1890
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaataccg aaaacaagat taacgaaaag ttgaaggaaa tgatctctac cttattgtcc
aagcaatctg ataatggtgc ttggaatttc tgtttcgaag gctcaattat gactgatgct
tatatgatca ttttaatcag aacattggaa attactgatg aggaagtttt ggttaaggac
ttggtcgaac gtatcaaatc tagacaatct cctaatggtg cttggaaggt ctacccagat
gaaaacaagg gtaatttatc tgctaccatt gaaggttact tctccttatt gtactctggt
tacgtcggtg aggaagcttc ttacatgaga aaagctgaaa gattcattag agacaacggc
ggtctagcta agtctgattg gctaaccaag atgatgttag ctttaactgg tcaaatcaag
tggccatcta ttatcaagat cattcctatt gagattatgc tattaccacg ttggtcccca
atcactattt accaattggt tggttatgct agagcccact ggatcccaat tttaatctgt
tctaatttaa acaaatcctt cgttgctcca caaactccaa acatttccca cttacaagaa
agattaatgg attccgaaaa cgaccgtatc ttggaggaaa tgcaaaactt acaattgtac
ttcaaacatg ccttgaaaaa gttgtctaaa tctcctgaaa ttttgaaaaa ggaagctttc
atcaaagctg aaaactatat cattgaaaga attgaggaaa acggtaccat gtactcttat
ttctctgctt ctttttttcat ggtttttgcc ttcttggctt tgggttacga tgctaaccac
cctttgatcc gtaacgcttt tcaaggtatg aagtcttact tgtgtagaaa cgctgaacaa
ccattcattc aaaactcccc atctacagtc tgggacactg ccttgttaac cgctgccttg
caacaagctg gtgtttctta cagacattcc tctatcatga aggctaataa ctatttgtta
tctagacaac accaaaaata cggtgattgg gctgttaaca atccagacgt tattccaggc
ggttggggtt tctccgatat taacaccttc gttccagata tcgatgacac aactgccgct
ttaagagcca ttactccatt gactcaaact aacattttat acaaggaagc ttggaataaa
ggtgttgaat ggattttgtc aatgcaaaac ggtgacggcg gttggtctgc tttcgaaaaa
aacatggaca actatttatt gtctttaatt ccattcaaat acgctgacag agttttattc
gatccatcaa ccgctgattt gactggtaga accttgtact tcttgggtga gtacaccact
atcccattgg aatctgagat cttccaaact gccaaggagt ggttcgagag aaaccaagag
gccaacggct cctggtatgg tagatggggt aattgttaca tttacggtac ttgggctgcc
attactggtt tgaaggctat tggtgtttcc aacgatgacc caattatctc aagagctgtt
cgttggttat tgtccgtcca aaatgaagat ggcggttggg gtgaatcttg tgcctctgac
attaaaaaga gatacattcc attaccacat tccactccat ctcaaacagc ttgggcccta
gatgccttga tctccgcttc cgacaaccca acttctagaa ttgaagttgg tattcacgcc
ttgttaaaca tcttagaagc caatgatggt agatctaact acccaaccgg tgctatgatt
cctggtggct attacattca ctaccattcc tacaagtaca tctggccatt gcaagctttt
tcccactata agaacaagta tgacttgtaa

Sequence Number (ID): 132
Sequence Name: A0A2S5GD65
Length: 1880
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1880
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggaaaaga ttatcaagga agaaattgag cgtattgttt atcagctgga aggcattcag
agctcggatg gtacctggcg tttctgcttc gagggtagcg ttatgaccga tgcgtacttg
atcattttga ttaagacgct gcagctccag gaagataaac tggtgaaaga cttggctgag
cgtatcgcca gcaagcaaac tgagaacggt gcgtggaaac tgtttaaaga cgacgatggt
aatctgagcg cgacggtcga ggcctactac gcgttgctga tggcaaacta taagaaaaag
tccgacccga ccatgaaaaa agctgaagat ttcatcatca aaaacggcgg cttgagcaaa
gttagcagcc tgaccaaagt gatgctggct accacgggtc agtattcttg gagcaagatc
attccgattc cgatcgaagt cattctgttg ccacagagct gtccgattaa cttcttcgac
atggtcggtt atgcccgtgt gcatttgatg ccgattctgg tgctggccaa caataagttc

tctatgaaaa cggcacacct gaatctggag tacctgaacc aaagccgtga cgaagaacaa
gatgagtgct ttatctccat tcaaagcgat gatacccgtt ccctgctgag ctttatcaag
caaaacgttc aaaagctgat tggtctgccg aatgaactga accgcatggc gctggaccaa
gcgaagctgt ttatgttgca ccgtatcgag ccggacggta ccctgtattc ctactttagc
tccacgttcc ttatgatctt ttctctgctg agcctgggtt ttactaaaga cgatccaatt
atcgagaaag caatcaacgg tctgaagggc ctggcgtgta acaccgaaga tcatattcat
atccagaata gcccgtctac ggtttgggac acggccctga ttacccacag cctgttaagc
tctggtgtcg atgtccgtag caactttatt caactgccta cccactatct gctgcgcaaa
caacagtacc tgtatggcga ttggagcatc cataatctca atagcctgcc gggtggctgg
ggtttcagcg acagcaatac catgaatcca gacgtagacg atacgacggc cgctctgcgt
gcgatcaagc cgaccatcag ccagcatccg aacctgagcc agtcctggtt tcgcggtctg
aactgggtgc tgagcatgca gaataatgac ggtggctggc cggcattcga gaagaacact
gacaaagaaa ttctgaagct gatcccgttc gacggtagcg agagcgcgag catcgacccg
agcaccgcgg atctgaccgg tcgcacgctg gaattcttgg gtaatgacgc ccgtttgacc
gtgcagcacc cgcaaattaa gcgtgcggtg gactggctga aagaccatca agagagcgat
ggctcgtggt atggccgctg gggtatcacc tacatctatg gtacgtgggc agcaatcacc
ggcatgcgcg cggttggcga gaagtcgcac caccctacca ttgtgaaagc tgttcagtgg
ctggaagaaa ttcagaacgc agatggcggc tggggcgaga gctgtaacag cgatatcgag
aagaaataca ttccgctggg tgcaagcacc ccgagccaga cggcgtgggc gctggatagc
ctgatcagcg tttatgacca cccgaccgtc gagatcaaga aaggcattgg ctgcctgatc
aatctgctga aagaaaaaga ctggacttac tcctatccga cgggtgcggg tctgccgggt
tccttctaca tttactacca cagctacaat tacatttggc cgctgctgtc actgtctcgt
tacttacaaa agtatactaa

Sequence Number (ID): 133
Sequence Name: A0A2S5GD65
Length: 1881
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1881
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atggaaaaga ttatcaagga ggaaattgaa cgtattgttt accaattgga aggtattcaa
tcctctgacg gtacttggcg tttttgtttt gaaggttctg tcatgaccga cgcctattta
atcattttaa ttaaaacttt acaattgcaa gaagacaaat tggtcaagga cttggctgaa
agaattgctt ccaagcaaac cgaaaacggt gcttggaaat tgtttaagga tgacgatggt
aacttgtctg ctactgttga agcttattac gctttgttaa tggccaacta caagaaaaag
tctgacccaa ccatgaaaaa ggctgaagat tttattatca agaacggcgg tttgtctaag
gtctcttcct tgactaaggt tatgctagcc accactggtc aatactcttg gtctaagatc
attccaatcc caatcgaggt tatcttattg ccacaatctt gtcctattaa ctttttcgac
atggttggtt acgccagagt tcacttgatg ccaattttgg tcttagccaa taacaagttt
tccatgaaga ccgcccactt aaatttagaa tacttgaacc aatctagaga cgaggaacaa
gacgaatgtt tcatctctat ccaatccgac gataccagat ctttattgtc ttttatcaag
caaaatgttc aaaagttaat tggtttgcca aacgagttga acagaatggc tttggatcaa
gctaaattat tcatgttaca cagaattgaa ccagatggta ccctatattc ctacttttca
tctactttct tgatgatttt ctctctattg tcactaggtt tcaccaaaga cgatccaatc
attgaaaagg ccattaacgg tttaaagggt ctagcttgta acactgaaga tcatatccac
atccaaaact ctccttccac cgtttgggac accgctttga ttactcactc tttattgtca
tctggtgttg atgttcgttc taacttcatc caattaccta cccactactt gctaagaaag
caacaatact tatatggtga ttggtccatt cacaatttga attctttacc aggcggttgg
ggtttctccg attctaacac tatgaaccca gacgttgacg ataccactgc cgctttaaga
gctatcaaac ctaccatttc tcaacatcca aatttgtctc aatcctggtt cagaggtcta
aattgggtcc tatccatgca aaataacgat ggcggttggc cagctttcga aaagaacact
gacaaagaaa tcttgaagtt gattccattc gacggttccg aatccgcttc catcgaccca
tctaccgccg atttaaccgg tagaacctta gaattcttgg gtaacgacgc tagattgacc
gtccaacacc ctcaaattaa gcgtgctgtc gactggctaa aggaccacca agaatccgat
ggttcctggt acggtcgttg gggtattact tatatctacg gtacttgggc cgctattacc
ggtatgcgtg ctgttggcga aaagtcccat cacccaacta ttgttaaagc cgtccaatgg
ttagaggaaa ttcaaaatgc tgacggcggt tggggtgaat cttgtaattc tgacattgaa
aagaaataca ttccattggg tgcttccact ccatctcaaa ctgcttgggc tttagattcc

ttgatctccg tttatgacca cccaactgtc gaaatcaaga aaggtatcgg ttgtttaatt
aacctattga aggaaaagga ctggacttac tcttatccaa ctggcgccgg cttgccaggt
tctttctaca tttattacca ttcttataac tacatctggc ctctattgtc tttgtctaga
tatttgcaaa agtacactta a

Sequence Number (ID): 134
Sequence Name: A0A2S5GD65_V1
Length: 1880
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1880
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggaaaaga ttatcaagga agaaattgag cgtattgttt atcagctgga aggcattcag
agctcggatg gtacctggcg tttctgcttc gagggtagcg ttatgaccga tgcgtacttg
atcattttga ttaagacgct gcagctccag gaagataaac tggtgaaaga cttggctgag
cgtatcgcca gcaagcaaac tgagaacggt gcgtggaaac tgtttaaaga cgacgatggt
aatctgagcg cgacggtcga ggcctactac gcgttgctga tggcaaacta taagaaaaag
tccgacccga ccatgaaaaa agctgaagat ttcatcatca aaaacggcgg cttgagcaaa
gttagcagcc tgaccaaagt gatgctggct accacgggtc agtattcttg gagcaagatc
attccgattc cgatcgaagt cattctgttg ccacagagct gtccgattaa cttcttcgac
atggtcggtt atgcccgtgt gcatttgatg ccgattctgg tgctggccaa caataagttc
tctatgaaaa cggcacacct gaatctggag tacctgaacc aaagccgtga cgaagaacaa
gatgagtgct ttatctccat tcaaagcgat gatacccgtt ccctgctgag ctttatcaag
caaaacgttc aaaagctgat tggtctgccg aatgaactga accgcatggc gctggaccaa
gcgaagctgt ttatgttgca ccgtatcgag ccggacggta ccctgtattc ctactttagc
tccacgttcc ttatgatctt ttctctgctg agcctgggtt ttactaaaga cgatccaatt
atcgagaaag caatcaacgg tctgaagggc ctggcgtgta acaccgaaga tcatattcat
atccagaata gcccgtctac ggtttgggac acggccctga ttacccacag cctgttaagc
tctggtgtcg atgtccgtag caactttatt caactgccta cccactatct gctgcgcaaa
caacagtacc tgtatggcga ttggagcatc cataatctca atagcctgcc gggtggctgg
ggtttcagcg acagcaatac catgaatcca gacgtagacg atacgacggc cgctctgcgt
gcgatcaagc cgaccatcag ccagcatccg aacctgagcc agtcctggtt tcgcggtctg
aactgggtgc tgagcatgca gaataatgac ggtggctggc cggcattcga gaagaacact
gacaaagaaa ttctgaagct gatcccgttc gacggtgcgg agagcgcgag catcgacccg
agcaccgcgg atctgaccgg tcgcacgctg gaattcttgg gtaatgacgc ccgtttgacc
gtgcagcacc cgcaaattaa gcgtgcggtg gactggctga aagaccatca agagagcgat
ggctcgtggt atggccgctg gggtatcacc tacatctatg gtacgtgggc agcaatcacc
ggcatgcgcg cggttggcga gaagtcgcac caccctacca ttgtgaaagc tgttcagtgg
ctggaagaaa ttcagaacgc agatggcggc tggggcgaga gctgtaacag cgatatcgag
aagaaataca ttccgctggg tgcaagcacc ccgagccaga cggcgtgggc gctggatagc
ctgatcagcg tttatgacca cccgaccgtc gagatcaaga aaggcattgg ctgcctgatc
aatctgctga aagaaaaaga ctggacttac tcctatccga cgggtgcgat gctgccgggt
tccttctaca tttactacca cagctacaat tacatttggc cgctgctgtc actgtctcgt
tacttacaaa agtatactaa

Sequence Number (ID): 135
Sequence Name: A0A2S5GD65_V1
Length: 1881
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1881
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atggagaaga ttatcaagga ggaaattgaa agaatcgtct accaattgga aggtattcaa
tcctctgacg gtacctggag attctgtttc gaaggctcag tcatgactga tgcttacttg
atcattttga tcaagacttt gcaactacaa gaagacaagt tggtcaaaga tttagctgag
agaatcgctt ctaagcaaac tgaaaacggt gcttggaagt tgttcaagga cgatgacggt
aacttgtctg ctactgttga agcttattac gctttattga tggccaacta caagaaaaag

tctgacccaa ctatgaagaa agctgaagat tttatcatta agaacggcgg tttgtccaaa
gtctcatcct tgaccaaagt catgttggcc accactggtc aatactcctg gtctaagatc
attccaattc caatcgaagt tatcttattg ccacaatcct gccctattaa tttctttgat
atggtcggtt acgccagagt ccacttgatg ccaattttag tcttggccaa caataagttt
tccatgaaaa ctgctcactt gaacctagaa tacttgaatc aatccagaga cgaggaacaa
gacgaatgtt tcatttctat tcaatctgac gatactagat ctttgctatc tttcatcaaa
caaaacgttc aaaagttgat tggcttacca aacgaattaa accgtatggc cttggatcaa
gccaaattgt tcatgttaca tagaattgag ccagacggta ccttatactc ctatttctcc
tctactttcc taatgatctt ctccttattg tccttgggtt tcactaaaga tgacccaatc
attgaaaagg ccattaacgg tttgaagggt ttggcctgta atactgaaga ccacattcac
attcaaaatt ctccatctac cgtttgggac accgccttga tcactcactc tttattgtct
tccggtgtcg acgtccgttc caacttcatt caattgccta ctcactactt gctaagaaag
caacaatact tgtacggtga ctggtccatt cacaatctaa actctttgcc aggcggttgg
ggtttctctg attcaaacac tatgaaccca gatgtcgacg ataccactgc tgccttgaga
gctatcaaac caaccatctc tcaacatcca aatctatctc aatcttggtt cagaggtttg
aactgggttc tatccatgca aaacaatgac ggcggttggc cagctttcga aaagaacact
gataaagaaa ttctaaagtt gattccattc gatggtgctg aatctgcttc catcgaccct
tccactgctg acttaaccgg tagaactttg gaatttttgg gtaacgacgc cagattgact
gttcaacacc ctcaaatcaa gcgtgctgtc gactggctaa aagatcatca agaatcagat
ggctcttggt atggtcgttg gggtattact tatatctacg gtacctgggc cgctattact
ggtatgagag ctgttggtga aaagtcccat cacccaacta tcgtcaaggc tgtccaatgg
ttggaggaaa ttcaaaacgc tgacggcggt tggggtgaat cctgtaactc cgatattgaa
aagaaataca tcccattggg tgcttctact ccatcacaaa ctgcttgggc cttagactct
ttgatttccg tctacgatca cccaactgtc gaaatcaaaa agggtattgg ttgcttgatc
aacttgttaa aagaaaaaga ttggacttac tcctacccaa ctggtgctat gttgcctggt
tccttctaca tctattacca ctcttacaat tacatttggc cattattgtc cttgtctaga
tacttacaaa agtacacttg a

Sequence Number (ID): 136
Sequence Name: A0A2W0HKM8_V1
Length: 1860
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1860
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgcacgaat tgttaaacga taccattcgt agaatgaccg attctttaag atctctacaa
gctaacgatg gttcctggag attttgtttc gaaaccggtc caactaccga cgcttatact
attatcttgc tacgttcttt agatactaag ggtgatgaat ggttgattca acaattgact
gaaagattat tggctattca agacacttct ggttcctgga aattatatcc agatcaagct
gagggccact tgtctgctac cgttgaggcc tactttgccc tattgtattc cggttatgtc
tctgagactg ctcctaacat gaaaaaggct agatcattca tcgaaaagaa cggcggttta
cataaagctg gtccattcac cagagctatg ttggctctaa acggtcaaat ctcatggcca
cgtttgttta gattgccagt cgaatcttta ttgatcccac aaaacgctcc tgtcaacttg
tacgacatcg tctcttacgc cagagtccat attgttccag tcttagctgc cgctaacaag
ggttacgttt gtcaattgcc atccactcca gacttgtccg ccttgggtaa gagaaatgag
gaatctgccg ctgaggaaac tgaacgtttg ttctctactg ttgctgccga aattcacaaa
ttggctgaaa ctccaggtcg tttgaaagct aaagcttata aaaaggccga aagattgatg
ctagaaagaa ttgaacctga cggtctatat ttctcttaca ttacttccac tgtcttgatg
gtttatgctt tattggcttt aggttactct aagaatgatg ctgttatcca aagagcttta
accgctatta gaaaccaagt ctgtttgacc tctaccagat ctcatatcga atttgctacc
tccaccgttt gggataccgc tttattgtcc cacgccttac aaagatctgg tgttccttcc
gaagatccaa tgattgctgg tgctggtaga tacttattga acagacaaca cacaaagtac
ggtgactggg ctttcaattg ttctggcact ttgccaggcg gttggggttt ctctgacatt
aacacttttc tacctgatat cgatgacact accgcttctt tgcgtgctgt caaagacttg
attgaggcta tgccagaata cagaattacc tggttcagag gtaccgattg ggtcctaaaa
atgcaaaaca ctgatggcgg ttgggctgcc tttgaaaagg acactgtcaa gcgtagattg
actttgctac catttccagc cgctgataga gttttgatcg acccatcaac agctgattta
actggtagag ccttggaatt cctatccggt gaagctaact tattgctacc acacccagct
gttgaccgtg ccgtcaactg gttggaaaag aatcaagaac aaaacggttc ttggtacggt

cgttggggta tctgttatat ctacggtact tgggctgcct tgactggttt gtctgccgct
ggttatgaaa aagaaaacaa aaccgttaag agaggtgtcg aatggttcaa ctctattcaa
aatgaagatg gcggttgggg tgaatcttgt aagtccgata ccgccggtcg ttacgtccca
ctaggttctt ccactccatc tcaaactgct tgggccgtcg atgctttgat cgctgtccat
tcaagaccaa cagaagctat cgaccacggt attagatacc taattaccaa cgctggcaga
tcagacagat acccaactgg cgctatgcta ccaggtgatt tctacatcta ttaccattct
tacaaccaca tctggccatt attggcttta gctaactata agtctagata ctcttcttaa

Sequence Number (ID): 137
Sequence Name: A0A3D9KM71_V1
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgggtaacg ttatttctga aattgatgag gaaattcaaa gattggcttc taacgtcgtt
cgtcaacaat acccagatgg ttcttggaga tattgtttcg agaacggtat ctctattgat
gcttacacta ttatcttgat cagagttttg gaaattgaat ctgaggaagc cttgatccgt
agattacacg accgtatctt aggtgctcaa caaccagaag gtcattggca atggtataga
gacgagcaaa acggtaactt atccttgacc gttgaagctt attgtgcttt gttattctct
cgtttctctc actcctctga tgctccatta caaaaggctg aaagatacat tagatctcaa
ggcggtttgg gtaaaactgc caacattttg accagatgca tgttggccgc taccggtcaa
accaaatggc cattgtcatt gacctctatc ccattggaaa tgttactatt gccagactca
ttcccaatta acttttcga attctctggt tactccagag tccatctagc tcctatgttg
attatggccc acagacacta ctctgctaga actggtgata ccccagaatt ggatgctttg
agaactgata gatctccatc tgaagccaga ccatctagag gtgttagaga atggttggac
agattgcaag ttggtgtttc caaattaatt ggtgctccaa aggccttaca tgaatccgct
ttagctaagg ctgaaaagtt tatgagagac agaatcgaag ccgacggtac cttgtactcc
tacgcctcat gcacttttt gatggttcta gctctattgg ccttgggtta cgacaagaga
cacccactaa ttgctaaggc cgtccgtggc ttgattggta tgagatttag aactgaggaa
ggtaccacta ttcaaaactc tccatcaact gtctgggata ccgctttgct agcttacgct
ttgcaagaag ctggtgccac tgaacatcac cctgctgtca gaaacgcttc ctcttattta
ttgggtttgc aacaccgtaa gccaggcgat tggacccgtc acaacccaaa cccagttcca
ggcggttggg gtttctcaga aactaacacc attaatccag acgtcgacga tactacagcc
gctttgagag ccatccaaaa gttggctaga tctgactcta cttacagaga atcttggaac
agaggtttga actgggtttt gtctatgcaa aacagagatg gcggttggcc agcctttgaa
agaaacgttg accaacaatt attgacctta gttgctatcg agggtgctaa gtcagccgct
atcgatccat ccgaagctga tttaaccggt agaaccttgg aatacctagg taacttcact
ggtttgggta gaaaacacgg tttcgtcaaa aaggccgttg actggttggt cgaacaccaa
gaggaagatg gttcctggta cggtagatgg ggtatctgtt atatttacgg tacctgggcc
gctttgactg gtttgacctc tgttattgac tctccagaaa gacatgaatc tattagaaag
ggtgctcaat ggctattgca aattcaaaat gatgacggcg gttggggtga atcttgttct
tccgacagac aacgtcacta cgttccacta ggcaagtcta ccccatcaca aaccgcctgg
gctatcgacg ctttggtttc tgtttatcca gaaccaactc cagctttgaa tcaaggtatt
cgtagattga tcgctttgtt acatgacaac gactggccaa cttcataccc aaccggtgcc
atgttgccag gtcacttcta cgtcagatat cactcttaca attcaatctg gcctctattg
gctttatccc attacagaaa caaatacggt aagtaa

Sequence Number (ID): 138
Sequence Name: A0A559J1A9
Length: 2007
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2007
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaaccgta gagccaacat cgaatccggt attgacagat taattcaaca attattgatg
gctcaagctc cagacggttc ctggagattc tgttttgaat ctggtaccat gactgactct

tatatgatta tcattgttag agtcttgcaa ttgtctgaag atgaattggt taagcaattg
tcccaaagaa ttgtttctcg tcaacatcca gaaggttact ggtctgttta cccaaacgaa
accaacggta atctatcctc aactgtcgaa gcctactatg ctttattgta ttctggtact
atgaagaaag atgacccaat cttattgaag gctaaggctt atatcttgtc aaatggtggc
atgcaacaag ctaattctgt tttaacaaag actatgttag ctgccactgg tcaaagacca
tggccaagat cttacactgt tccaattgaa ttcttgttat tgccagaatg gtccccaatt
tccttctacg atatcgttgg ttacgccaga gtccacattg ctccaatctt gatcatgtcc
tctttgccag ataccactat tccagaaggt gctcctgatt tgtctgactt gattttgcca
aacagatcct gggaatacga atctgaatcc ttcaaccaca ctgatgacat tcaacaacaa
caatgtttat acgaggatta ctttgcttac gacttctcta actttcaatc tacctacaat
tctgttatcc aatctggtga agctcataga tctttattgc aaaacgttaa aagagaatta
ttccaattgt taccatcccc acaatcagtt aagcaagaag ctagaaataa ggctgaatcc
ttcatgttgg atagaattga acctaatggt accttgtact catacgcttc tgctaccttc
ttaatgatct tcgctttatt ggctttgggt tacgacagaa accatccaag aatcaccaag
gctatccaag gtttgaagtc tttcgtttgt ccatctgaca agcattggca catccaaaat
tctcctccaa ctatttggga tactgctttg atctcccacg ctttccaaca agccggcttg
ccagttcaac acggtgctat ccaaagagcc ggtgcttact tattgtcaag acaacaacac
aagttcggtg attggcaatt tcacaaccca aacacccctc caggcggttg gggtttttca
gatatcaaca ctatcattcc tgatatcgac gataccactg ctgccttgag agctattaac
aagttggctt cctctaacac ttcctacgct gccgcttacg acaagggttt gcaatggcta
ttgtctatgc aaaatgatga cggcggttgg ccagctttcg aaaagaacac taacaagact
atcttaactt ggctaccttа cgacggtgcc aacgccgctt taacagaccc atctactgct
gatctaaccg gtcgtacttt ggaatatttg ggttctaccg cccaattaaa gttggaacac
gccttcgttc gtagaggtgc tgactggttg atgaaccacc aacaacaaga cggttcttgg
tatggtaagt ggggtatttc ttacatctac ggtacttggg ccgctgttac cggtttagct
gccgttggtg ttgacgcttc taacccagcc ttggttagag ccgctcgttg gttatctaga
atccaaaacc aagatggcgg ttggggtgaa tcctgtgaat ccgacagaaa aaagacttac
attccactac acttgtccac tccatcacaa actgcttggg ccttggacgc tttgattgct
gtttctccac aaccaactga ggaaattgaa agaggtattc aaaacttgtt atacatgttg
caacacccaa acaagcaatc taatacttat cctaccggtg ctggtttgcc tggtaacttc
tacatttatt accattctta caactacatc tggccattat tgactttggc taactacaag
agaaagtact ctccaccatt gggttaa

Sequence Number (ID): 139
Sequence Name: A0A559J1A9_V1
Length: 2006
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2006
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaatcgcc gtgcgaacat cgagtctggt atcgaccgtc tgatccaaca gctcctgatg
gcgcaggccc cggacggttc ttggcgcttc tgcttcgaaa gcggtactat gaccgactcc
tatatgatca ttatcgttcg tgttctgcag ctgtccgagg atgagctggt taaacagctg
tcccagcgta tcgtatcccg tcagcatcct gaaggttact ggagcgtgta cccgaacgaa
accaacggca atctgtcttc cactgtggaa gcatactatg cgctcctgta tagcggtacc
atgaagaaag atgacccaat cctgctcaaa gcgaaggctt atatcctgtc taacggtggc
atgcaacagg caaactccgt gctgaccaaa acgatgctgg cagctaccgg ccaacgtcct
tggccgcgtt cttacaccgt gccgattgag tttctcctgc tcccagaatg gtccccaatc
tccttctacg atatcgtggg ctacgctcgt gttcatatcg ctccaatcct gatcatgtcc
agcctgccgg acactaccat cccggaaggt gcgccggatc tgtccgacct gatcctgcct
aaccgtagct gggaatacga atctgaatct ttcaaccaca ccgatgacat ccaacagcaa
cagtgcctgt atgaagatta tttcgcatat gatttttcca actttcagtc cacttacaac
tctgtaatcc agagcggcga agcccaccgt agcctcctgc agaacgtgaa acgcgaactg
tttcagctcc tgccgtcccc acagtctgtt aaacaggaag ctcgtaacaa agcagagtct
tttatgctgg accgcatcga accaaacggc actctgtact cttacgcgtc cgcgactttc
ctgatgatct ttgcactcct ggcgctgggt tacgaccgta accacccgcg cattaccaaa
gcgatccagg gcctgaaatc tttcgtgtgc ccttctgaca agcactggca catccagaac
tccccgccta ctatttggga cactgccctc atctcccacg cgtttcaaca ggcgggtctg
cctgtccagc acggcgctat ccaacgcgca ggtgcttatc tcctgtctcg ccaacagcac

aaattcggtg actggcagtt ccacaacccg aacacccctc cgggcggttg gggtttctct
gacattaaca ccatcattcc ggatatcgat gacacgaccg ccgctctgcg cgctatcaat
aaactggcta gctctaacac ctcctatgct gcggcctacg acaaaggcct gcagtggctc
ctgtctatgc agaacgacga tggcggttgg ccggcattcg aaaaaaacac taacaaaacc
atcctgactt ggctgccgta tgatggtgcg aacgcggctc tgactgatcc gtctaccgcg
gacctgactg gtcgcaccct ggagtatctg ggctccaccg ctcagctgaa actggaacac
gcgtttgttc gccgtggtgc tgattggctc atgaaccatc aacagcaaga tggctcttgg
tacggcaagt ggggcatttc ctatatctac ggtacctggg cggctgttac gggcctggct
gcagttggtg tggatgcgag caacccggcg ctggtgcgtg cagcccgttg gctgagccgt
atccagaacc aggatggtgg ctggggcgaa tcttgcgagt ccgaccgcaa gaaaacttac
atcccgctgc acctctctac tccgagccaa accgcgtggg cactggacgc cctgatcgcc
gtatccccac agccgaccga ggaaatcgaa cgtggcatcc agaatctcct gtacatgctg
cagcacccga acaagcaatc taacacctac ccgaccggcg caatgctgcc tggtaacttt
tacatctact atcattccta caattatatc tggccgctcc tgaccctggc caattacaaa
cgtaaatatt ccctccgct gggtaa

Sequence Number (ID): 140
Sequence Name: A0A559J1A9_V1
Length: 2007
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2007
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaaccgta gagccaacat tgaatccggt attgacagat tgatccaaca actattgatg
gctcaagccc cagatggttc ctggcgtttc tgcttcgaat ctggtaccat gacagattcc
tatatgatta tcattgttag agttttgcaa ttatctgaag acgaattagt taagcaacta
tctcaaagaa ttgtctcacg tcaacaccca gaaggttact ggtctgttta cccaaacgaa
acaaacggca acctatcctc aaccgtcgaa gcttactatg ctttattgta ttccggcacc
atgaagaaag acgatccaat cctattgaag gccaaggcct acatcttatc caacggcggt
atgcaacaag ccaactctgt tttaactaag accatgttgg ccgctaccgg tcaaagacca
tggccaagat cttacactgt tccaatcgag ttcttgttat tgcctgaatg gtctccaatc
tcattctatg atatcgtcgg ttacgctaga gttcacatcg ctcctatttt gatcatgtcc
tctttgccag ataccactat tccagaaggt gctccagact tgtctgattt gatcttgcct
aacagatctt gggaatacga atctgagtcc tttaaccaca ctgacgatat ccaacaacaa
caatgtttgt acgaagacta cttcgcttac gatttctcta acttccaatc tacctacaac
tccgtcatcc aatctggtga agctcataga tccttattgc aaaatgttaa gagagagttg
ttccaattat tgccatcccc tcaatcagtt aagcaagaag ctagaaacaa ggccgaatcc
ttcatgttgg acagaattga acctaatggt actttgtact cctacgcttc tgctacattc
ttgatgattt tcgctttgtt agctttgggt tatgatagaa atcaccctag aattactaag
gccatccaag gtttaaagtc tttcgtttgt ccatctgata agcattggca cattcaaaac
tctcctccaa ccatttggga cactgctcta atctctcacg cttttcaaca agctggctta
cctgtccaac atggtgctat tcaacgtgct ggcgcttact tattgtccag acaacaacac
aaattcggtg attggcaatt ccataatcca aatacccctc caggcggttg gggtttctcc
gatattaata ctatcattcc agacattgac gataccactg ccgctttgag agctatcaac
aagttggcct cctctaacac ctcctacgcc gctgcctacg ataagggtct acaatggtta
ttgtctatgc aaaacgatga cggcggttgg cctgcctttg aaaagaacac caacaaaacc
atcttgactt ggttgcctta cgacggcgct aacgccgctt tgacagatcc atccaccgct
gatttgaccg gtagaacctt ggaatacttg ggctccaccg ctcaattgaa gttagaacac
gccttcgttc gtagaggtgc cgactggttg atgaaccatc aacaacaaga tggttcttgg
tacggtaagt ggggtatctc ctacatttac ggtacttggg ctgccgtcac cggtttagct
gccgtcggtg ttgatgcttc taacccagct ctagtcagag ccgctagatg gttgtctcgt
attcaaaacc aagacggcgg ttggggtgaa tcctgtgaat ctgatagaaa gaaaacctac
atcccattac atttgtctac tccatcccaa actgcctggg ccttggatgc tttgattgct
gtctccccac aaccaactga ggaaattgag cgtggtattc aaaacttgtt atatatgtta
caacacccta acaagcaatc caacacttac ccaactggtg ccatgttgcc aggtaacttt
tacatttact atcattccta caattacatt tggccattat tgactttagc caactataag
agaaagtact ctccaccttt gggttaa

Sequence Number (ID): 141

Sequence Name: A0A5D4SXS0_V1
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaacccgg ttgataaaat caacgaaaaa ctgaaggaca tgatctctaa cctcctgagc
aaacagaaag ataacggcgt ttggaacttt tgttttgagg gcagcattat gactgacgcg
tacatgatta tcctgatccg tatcctggaa ctgacggacg aagaggttct ggtaaaatct
ctggtggagc gcatcaaatc tcaacaggaa tccaacggtg cttggaaagt atacccggat
gaagacggtg gcaatctgag cgtaaccatt gaaggctact tctccctcct gtattccggt
tacgtggcta aaaacgcagg ctacatgcgc aaagcagaac gtttcattca agaaaatggc
ggtctgagca aatctgactg gctgaccaaa atgatgctgg ctatcaccgg tcaaatcaaa
tggccgtcta tcattaaaat tatcccgatc gaaatcatgc tcctgccgag ccgttccccg
atctctatct atcagttcgt tggttatgca cgtgcccact ggatcccaat cctgatctgt
agcaacctga actacagcta tgttcacagc cgcacgccga acctcaccca cctgcagggt
ggcggtagcg atgacgaaga ccagcgcctc ctggaggaca tgcagaatct gcaactgtac
ttcaaaaacg cgctgaagaa actggcgggt agcaccgaaa acctgaaacg taaagcattt
atcaaggctg aagactacat tctggaacgt atcgaagaga acggcactct gtactcctac
ttctctgcct ctttctttat ggtgttcgct ctcctggctc tgggttatga taagaaaaac
cctctgatcg gcaacgcttt ccagggtatg aaggcttacc tgtgccgctc tgccgatcaa
gttttcatcc agaactcccc gagcaccgtt tgggataccg cgctcctgtc tgcagccctc
caacaggccg gcgtcccgca ccaacacgca agcatcatga aggctagcaa ctatctcctg
tctcgtcaac agcaaaagta tggtgactgg gctattaaaa acccggatgt tactccgggt
ggctggggtt tctctgatac taacactttc gtaccggata ttgatgacac cactgcggct
ctgcgtgcga tcaccccgct ggcgggtacc gaaaaccatt tcaaacatgc gtggaacaag
ggtgttgagt gggttctgac gatgcagaac gacgatggtg gctggagcgc tttcgaaaag
aacaccgaca actatctcct gagcttcatc ccgtccaagt acgcagatcg tgttctgttc
gacccaagca cggcagacct gaccggccgt actctgtatt ttctgggtga atacactacc
atgcctcagg aatctgacgt tttccagaag gctcgcgaat ggttcgagca caaccaggag
gaaaacggct cttggtacgg tcgttggggc aactgctaca tttacggcac ttgggcggca
attaccggtc tgaaagcgat cggcgtttcc agcgacaatc cgctgatctc ccgtggtgta
aaatggctcc tgtctattca gaacgaggac ggcggttggg gtgagtcctg cgaatccgac
tctaagaaac gttatgtgcc gctgcatcac tccactcctt cccaaactgc gtgggccctg
gatgccctga tctctgcgtc tgaaatcccg acgccgaaaa tcgaaaaagg catgcactct
ctcctgagcc tgctcgaagc gtctgattgg cgtagcatct acccgaccgg cgcgatgatc
ccgggcggtt attacattca ctaccacagc tacaaataca tctggccgct gcacgcactg
tcccactaca aaaacaaata ttacgagatc gaataa

Sequence Number (ID): 142
Sequence Name: A0A5D4SXS0_V1
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaacccag ttgataagat caacgaaaag ttaaaggata tgatttctaa cttattgtct
aagcaaaaag acaacggcgt ctggaacttt tgtttcgaag gttctattat gaccgatgct
tatatgatca ttttgatcag aattttggaa ttgaccgatg aggaagtctt agtcaaatct
ttggttgaaa gaattaaatc ccaacaagaa tcaaacggtg cttggaaggt ttacccagac
gaagatggcg gtaatttgtc cgttaccatc gaaggttact tttctttgtt atactccggt
tacgttgcta agaacgctgg ttacatgaga aaggccgaaa gattcatcca agaaaacggc
ggtttgtcta agtctgactg gttaaccaag atgatgttgg ccattactgg tcaaatcaag
tggccatcca ttatcaaaat tatcccaatt gaaattatgt tgctaccatc tagatcccca
atttctatct accaattcgt cggttatgcc agagctcatt ggatcccaat tttgatttgt
tctaacttga actactccta cgttcactcc agaacaccaa atttgactca cttgcaaggt
ggcggttctg acgatgaaga ccaaagattg ttagaggaca tgcaaaattt gcaattgtac

ttcaagaacg ctttgaaaaa gctagccggt tctaccgaaa atttaaagag aaaggctttt
atcaaggctg aagattacat cttggaacgt attgaggaaa acggtacctt gtattcttac
ttttcagcct ctttttttcat ggtcttcgct ttattggccc taggttacga taaaaagaac
ccattaatcg gcaacgcctt ccaaggtatg aaggcctact tgtgcagatc tgccgatcaa
gttttcatcc aaaactcacc ttccaccgtc tgggatacag ccttattgtc tgccgctttg
caacaagctg gtgtcccaca tcaacatgct tccatcatga aggcctccaa ctacttgcta
tccagacaac aacaaaagta cggtgattgg gccatcaaga accctgacgt taccccaggc
ggttggggtt tctctgacac taacactttc gttccagata tcgatgacac cactgccgct
ctaagagcca ttaccccatt ggctggtacc gaaaaccatt ttaaacacgc ttggaacaaa
ggtgttgaat gggtcttgac tatgcaaaac gacgatggcg gttggtctgc ttttgaaaag
aacactgaca actatttgtt atccttcatt ccatctaagt atgccgacag agtcttgttc
gatccatcta ccgccgactt gaccggtaga accttgtact tcctaggtga atacaccact
atgccacaag agtctgacgt tttccaaaag gccagagaat ggttcgaaca caaccaagag
gaaaacggtt cttggtacgg tagatggggt aactgttata tttacggtac ctgggccgct
atcaccggtt taaaagctat cggtgtttcc tctgacaacc cattgatctc tagaggtgtc
aaatggttac tatctattca aaacgaagac ggcggttggg gtgaatcctg tgaatctgac
tctaaaaaga gatacgtccc attgcatcac tccactcctt ctcaaactgc ttgggctttg
gatgctttaa tttccgcttc tgaaattcca acaccaaaga tcgaaaaagg tatgcattct
ctattgtcct tgttagaagc ttctgactgg cgttctattt acccaactgg tgctatgatc
cctggtggct attacattca ctatcattct tacaagtaca tttggccatt acatgcttta
tcccattaca aaaacaagta ttacgaaatc gaataa

Sequence Number (ID): 143
Sequence Name: A0A5P9HJ69
Length: 1868
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1868
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgagcatta acgaaaagat ccacagcctg gtcacggaac tgctgcagaa acaagaagag
aacggtagct ggatttttctg ctttgagggc accatcatga ccgacgccta tatgatcatt
ctgatccgcg tgctgcaacg tactgacgaa gaagaactgg tgaaacaact ggttaagcgt
atcaaatcca aacaacagct ggataatggt gcgtggaaag tctttaaaga tgaagaagaa
ggcaatctga acgcgaccat tgaaggctac tttagcctgt tatacagcgg ttatgtgtct
cagagcgacc cgctgatgca gaaagccgag caattcatca agcgtaaggg tggcctgacg
agcaccgact ggctgacccg tgttatgctg gcactcacgg gtcagattca gtggccgacc
attatcaaaa gcattccaat tgagatcatg ctgctgccga agtggagccc ggttaacctg
taccaactgg ttggttacaa ccgtgcgcat tgggtgccga ttattatctc tagcagcaag
aatatcagca ttagcacctc tagcacgccg gacattagcc atctgcaagt gcgcgcaccg
aaactggaaa ttaccaaggg tctgcagctg atccagcact atgtcaaggg tttcgtcaac
aaactggcgg atacgccaga agtgctgcgt gaccgtgcgt tcagcaaagc tgagaagtat
atcaccaatc gcatcgaaga gaatggcact ctgtacagct atttcagctc cagcttttc
atggtgttcg cgttcttggc acttggttac gatcgcaccc acccgctgat tcagcacgcc
ttccaaggca tgaaatcgta tgtgtataaa gatgagaata tgattcatgt ggagaacagc
ccttccacgg tttgggacac ctctttactg acggcagcgt tgatgcaagc gggcgttagc
agcaaccaag aggctatcca gaaagccgcg agctacttgc tgaccctgca acagaccaaa
tacggtgact gggcagtgaa gaatccgaat gttgcaccgg gtggctgggg tttctccgag
tctaacacct ttgttccgga cattgacgat acgaccgcgg cactgcgtgt attggcggcg
ttcgtcgaca aagatagccg ttacctggac ggttggaaca aaggcatttc gtggctgctg
tccatgcaaa acgacgatgg tggctggagc gcctttgaga agaataccga taactatctg
ctgtttatga tcccgttcag ctatgaagat cgtgtcctgt ttgacccgag caccgcggac
ctgaccggtc gcgcgttgta cttcctgggt gagaatacca cgatcccgac cgacgataaa
gccgtccgtc gtgctaaaga gtggctggtt aagaatcaag aagaagatgg ctcgtggtac
ggccgctggg gtgtctgtta tatctacggt acctgggctg cggtcaccgg tttgaccgcg
gttggcgaga ctctgagaag ccaggccttg cagcgtgcag ttcagtggtt gtacaagatc
cagaacgaag atggcggctg gggtgagtcc tgcaagtccg actttgtgaa gcaatatgtc
ccgctgcatg ctagcacggc gtcacagacg gcgtgggcac tggacgcact gatcagcgcc
agcgacgttc cgagccctga gatgaagcgc ggtattaagg cactgctgcg cctgctggat
aatgaagatt ggcgtgagga atatccgacg ggtgcgggtc tgccgggtgg cttttacatt

cactatcaca gctacaatta catttggccg ctgcagacgc tgagccacta ccgtaacaag ttcggtga

Sequence Number (ID): 144
Sequence Name: A0A5P9HJ69_V1
Length: 1868
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1868
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgagcatta acgaaaagat ccacagcctg gtcacggaac tgctgcagaa acaagaagag
aacggtagct ggatttttctg ctttgagggc accatcatga ccgacgccta tatgatcatt
ctgatccgcg tgctgcaacg tactgacgaa gaagaactgg tgaaacaact ggttaagcgt
atcaaatcca aacaacagct ggataatggt gcgtggaaag tctttaaaga tgaagaagaa
ggcaatctga acgcgaccat tgaaggctac tttagcctgt tatacagcgg ttatgtgtct
cagagcgacc cgctgatgca gaaagccgag caattcatca agcgtaaggg tggcctgacg
agcaccgact ggctgacccg tgttatgctg gcactcacgg gtcagattca gtggccgacc
attatcaaaa gcattccaat tgagatcatg ctgctgccga agtggagccc ggttaacctg
taccaactgg ttggttacaa ccgtgcgcat tgggtgccga ttattatctc tagcagcaag
aatatcagca ttagcacctc tagcacgccg gacattagcc atctgcaagt gcgcgcaccg
aaactggaaa ttaccaaggg tctgcagctg atccagcact atgtcaaggg tttcgtcaac
aaactggcgg atacgccaga agtgctgcgt gaccgtgcgt tcagcaaagc tgagaagtat
atcaccaatc gcatcgaaga gaatggcact ctgtacagct atttcagctc cagcttttc
atggtgttcg cgttcttggc acttggttac gatcgcaccc acccgctgat tcagcacgcc
ttccaaggca tgaaatcgta tgtgtataaa gatgagaata tgattcatgt ggagaacagc
ccttccacgg tttgggacac ctctttactg acggcagcgt tgatgcaagc gggcgttagc
agcaaccaag aggctatcca gaaagccgcg agctacttgc tgaccctgca acagaccaaa
tacggtgact gggcagtgaa gaatccgaat gttgcaccgg gtggctgggg tttctccgag
tctaacacct ttgttccgga cattgacgat acgaccgcgg cactgcgtgt attggcggcg
ttcgtcgaca aagatagccg ttacctggac ggttggaaca aaggcatttc gtggctgctg
tccatgcaaa acgacgatgg tggctggagc gcctttgaga agaataccga taactatctg
ctgtttatga tcccgttcag ctatgcggat cgtgtcctgt ttgacccgag caccgcggac
ctgaccggtc gcgcgttgta cttcctgggt gagaatacca cgatcccgac cgacgataaa
gccgtccgtc gtgctaaaga gtggctggtt aagaatcaag aagaagatgg ctcgtggtac
ggccgctggg gtgtctgtta tatctacggt acctgggctg cggtcaccgg tttgaccgcg
gttggcgaga ctctgagaag ccaggccttg cagcgtgcag ttcagtggtt gtacaagatc
cagaacgaag atggcggctg gggtgagtcc tgcaagtccg actttgtgaa gcaatatgtc
ccgctgcatg ctagcacggc gtcacagacg gcgtgggcac tggacgcact gatcagcgcc
agcgacgttc cgagccctga gatgaagcgc ggtattaagg cactgctgcg cctgctggat
aatgaagatt ggcgtgagga atatccgacg ggtgcgatgc tgccgggtgg cttttacatt
cactatcaca gctacaatta catttggccg ctgcagacgc tgagccacta ccgtaacaag
ttcggtga

Sequence Number (ID): 145
Sequence Name: A0A5P9HJ69_V1
Length: 1872
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1872
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgtctatta atgaaaaaat tcattccttg gtcactgaat tgttacaaaa gcaagaggaa
aacggttctt ggatcttttg tttcgaaggt actatcatga ctgacgctta catgatcatt
ctaattagag ttttgcaaag aactgatgaa gaggaattgg ttaagcaatt ggttaagaga
atcaagtcta agcaacaatt ggacaacggc gcctggaagg ttttcaaaga tgaagaggaa
ggtaacttga acgctactat tgaaggttac ttctcactat tatactctgg ttacgtttcc
caatccgatc cattgatgca aaaggctgaa caattcatta agagaaaagg cggtttgact
tccactgact ggttgactag agttatgttg gctttgactg gtcaaattca atggccaacc

atcattaaat ctatccctat tgaaattatg ttgttaccta aatggtcccc tgtcaacttg
taccaattgg tcggttacaa cagagcccat tgggttccaa ttatcatttc ttcatccaaa
aatatttcca tttctacttc atctactcca gatatctctc acttgcaagt cagagctcca
aagttagaaa tcaccaaggg tttgcaacta atccaacatt acgttaaagg tttcgtcaac
aagttggctg ataccccaga ggtcttgaga gatagagcct tttccaaagc tgaaaagtac
atcactaaca gaattgaaga gaacggtacc ttgtactcat acttctcttc ctctttttc
atggttttcg ccttttagc cttgggttat gatagaaccc acccattgat tcaacacgct
tttcaaggta tgaagtctta cgtctataag gatgaaaaca tgattcacgt tgaaaattcc
ccatccactg tttgggacac ctctttattg accgccgctt tgatgcaagc tggtgtttct
tccaatcaag aagctatcca aaaggccgct tcctacttat tgactttgca acaaactaag
tatggtgatt gggctgttaa gaacccaaac gttgctccag gtggctgggg tttttccgaa
tccaacacct tcgtcccaga cattgatgac acaaccgccg ctttgagagt cttggccgct
ttcgttgata aggattccag atacttggac ggttggaaca agggtatttc ctggctattg
tctatgcaaa atgatgacgg cggttggtcc gccttcgaaa agaacactga caactacttg
ttattcatga ttccattttc ttacgctgat agagttttat ttgacccatc tactgctgac
ttgaccggta gagctttgta tttcttaggt gaaaacacaa ctattcctac tgatgacaaa
gccgttagac gtgctaaaga atggctagtc aagaaccaag aggaagatgg ttcttggtat
ggtagatggg gtgtttgtta tatttacggt acctgggccg ctgtcaccgg tttaactgcc
gtcggtgaaa ctttgcgttc tcaagctttg caaagagctg ttcaatggtt gtataagatt
caaaacgaag acggcggttg gggtgaatct tgtaagtctg attttgttaa gcaatacgtt
ccactacacg cttccaccgc ttcccaaacc gcctgggctt tggacgcttt gatttctgct
tccgacgttc catctccaga aatgaagcgt ggtatcaaag ctttattgcg tttgttagac
aatgaagatt ggcgtgagga atacccaact ggtgctatgt tgccaggcgg tttctatatc
cactaccatt cctacaacta tatttggcca ttgcaaacct tatcacacta cagaaacaag
tttggtgagt aa

Sequence Number (ID): 146
Sequence Name: F5LIR7_V1
Length: 1950
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1950
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgagaatta tgagagaagt tgctcacgct attcacaact tatctgaatg gttgttaggt
aaacaagaaa aggacggtac ctggaagttc tgctacgata acggtatttc cacttctgcc
tacttgatca ttttattgag aaccttggaa actgaaaatg ccgatgacga agctttgatc
cgtagattac atgactctat tgttagaaat caacaaccag atggttcatg gaagttgtat
gctgacgaaa aggatggtaa tttagccgct tccgttgaag cttattacgc tttattgttc
tctggttact ccgacgattc atccccatct ttagtcaacg ctagagcttt catccgttcc
cgtggcggta tcaccggcgt tactaacgtt ttgactagag ttatcttagc tgccactggt
caaatcccat ggccagccgc ttattctatc ccattagaat tcctattgtt accagcctcc
gctccattgt ccttctttga tttgtccgct tacgctagag ttcacttagc cccagttttg
ttaatggccg atagaaactt ctctgttaga acttccagaa ccccagattt gtctgctttg
ggtgtcagac gtgacgaaga tggcaagtct ggtgcttcct caggttccgt cccaggttcc
tcagcccaac cagaacgtgg tccaggcggt ttattggctg ccattcaagc tggtatctca
caattagctg ccttgcctgg tcaattgcac ggttctgctg tcaaaaaggc tgaaagattc
atgttagaga gaattgagcc agacggtact ttgtattctt acgcttcttc cacttgtcta
ttggtctttg ctttattgtc tttgggtttc gaacgtagac atccaactat cgctcaagcc
gtcaagggtt tgaaggccat gttatgtgtt tctgaaggta gattattgat gcaaaatgct
cctccaactg tttgggatac tgctttgatc gcctacgctt tgcaagaagc tggtgttaga
ccagaagctc ctggtatcag aaaggctgcc tcctacctat tggctaagca acaaagaaag
atcggtgact ggggtagaaa ggtctcccac cctgtccctg gcggttgggg tttctcccca
tcaaatacta gaaaccctga cgttgatgac actaccgccg ctttgagagc tgttaagttc
ttgagatcag aaggcaccgc cggtagagaa gcttggaaca gaggtttgta ctggattatc
tctatgcaaa accaagatgg cggttgggcc gcttttgaaa aggacaccga tgaaaagtta
ttgactttat tgccaatgga aggtgctaag cacgccgcta tcgatccatc caccgccgac
ttgactggta gaactttgga attcttgggt tctaccgctg gtttaggtgt caagcacgtt
tggattcgta gaggtgctga ctggttggtt gctaaccaag aaaaggacgg ttcctggtac
ggtagatggg gtatctgcta cttgtacggt acatgggccg ctttgactgg cttggccgct

gtcggtttgg aaccagatca tccagctgtt gccaagggtg tccgttggtt attgtctgtc
caaaacccag atggcggttg gggtgaatct tgtgcttctg acattgttgg tagatatgaa
tctttgggtg cctccacccc atcacaaact gcttgggccc tagacgcttt aattgctgtt
caccctagac catcagccgc tattgacaga ggcattcaaa gattggttgc cgctttagac
gagagagact ggacttctct ataccctact ggtgctatgt tgccaggttc cttttacaac
acctatgaat cctacagata catttggcca ttgctagctt tatctcacta ccgtaacaag
tatggtgaaa agtctgttaa acctgaataa

Sequence Number (ID): 147
Sequence Name: MCH2311119_V1
Length: 2013
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2013
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atggaatctc aagtttctca caagctaaat aaaaacttag actctgctat caaaaagact
caagattact tgttcagaga acaatacgat gaaggtttct ggtggggcga actagagtct
aatccaacca tggaagctga atttatcttg ttacactact tcctaggtat cagagacgag
aaaaagttta tcaagttgtc caaccacatc aagaaccaac aaagagagga tggtacttgg
ggtcaatact atggtgcccc aggtgacttg tctacttcag ttgaatgtta cttggcttta
aagattgccg gttattccga aaatgatcaa tctatgctaa aagctaagga cttcatcgtt
tctaagggcg gtatcgagga aaccagagtt tttactaaga tttggttgtc tttagttgac
caatggaaat gggaaggtgt tccaatcatg ccagctgaat tgattttatt gcctaactgg
tctccaatca acatctacga attctcctct tgggctagat ctactatcgt tccattattg
atcttgatgg ataaaaagcc agttagacca ttgccaaaga acttgagagt tgacgaatta
ttttgtgatg acgttaacaa tattgattac tctgttaagt ccccatctat ccaaattggt
tgggaaaact tcttttacgc caccgaccaa gttttgcgtt tattggacaa gttgccaatc
aagccaacca gagacttggc tttaaaaaag tccgaggaat ggattttgca acaccaagaa
gctgacggtt cttggggcgg tattcaacct ccatgggttt actctatcat ggccttgtac
accgccggtt acggtttgga ccacccagtt atcgacaaag ccttacaagg ttttaaggcc
ttcgaaattg aagatcaatt ttccttaaga gtccaagctt gtgtttctcc tatttgggac
accggtttgg ctatcatttc cttgttagat tccggtatta aaaccgatga cgatcgtatc
caaaaggctg gtcaatggct aatcaataaa caaatcaagt cagagggtga ttggcaagtt
aaggctaata acgtcagatc tggcggttgg gctttcgaat tcgagaacga acactaccca
gacattgacg atgccgctgt cgttgccact gccttacaca agatcgattt gactgatgaa
tacggcggtt tggacaaaaa gtctaagtct atcaagagat gtgtccaatg gattgaaggt
atgcaatcta aaaacggcgg ttgggcttcc ttcgacaagg ataacatgag atccttcatc
gccagaatcc cattctccga cgctggtgaa accatcgatc ctccatctgt tgacgtcacc
gcccatgttt tggaattatt gggtactcta gatgccaaaa agcacgctac cgttatcgcc
aaggctttgg attacgtctt attggaacaa gaacaagatg gttcatggtt cggtagatgg
ggtgtcaact acatctacgg tatcggttct gttctaccag ccttgcgtgc catcggtatt
gatccatccc acgaggctat gtctaaggct actaagtggt tggaagatca tcaaaacaag
gatggcggtt ggggtgaaac cccagcttcc tacgtcgatc catccttgca cggtaaaggt
ccatctaccg cttctcaaac cgcctggtct ttgatttctc taattgctgc cgacaagggt
gactcttcac acgtcttgaa gggtgtcaat tatttattgt ctaaccaaaa cgaagatggt
tcttgggacg aaccagaatt taccggtaca atgttcccag gttacggcat tggtatgaga
ccagatattt ctgaacaaga tgactctaaa caacacgata tcgccttgcc agctggtttt
atgattaact accacatgta cagaatttac tggccattgt gtgctttagg tagattcaga
gcttggaatg ttaacagaga ctcccatcac taa

Sequence Number (ID): 148
Sequence Name: NQW16228_V1
Length: 2067
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2067
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:

atgaccagca ctctgaaccg tcgctccgac attgttccgc cagtacgcct ggaccaccca
cctgttccgg aacgcgctgg cggtgccgct ggcgaagagt tcggcgacct cctggatagc
gctatcggtc gtacccagga ctacctgttc gaacaccagg taagcgatgg ttactggtgg
ggcgagctgg agtccaaccc gaccatggag gctgaatacg ttttcctgac tcatttcctg
ggtctgcgtg acgatgtacg ctggaagaaa gtgcaaaatt atattctgtc cgtgcaacgc
ccgggcggtg gctggaacca gtaccacggc gcaccgaatg acctgtctac ctcctgtgaa
tgttacctgg ctctgaaaat gaccggcctg ccggcttccg acccgcgtat gcagcaagcg
cgtgagttca ttctgcgtaa gggtggcatg gagcaaaccc gtgttttcac caaaatctgg
tttagcctcc tgggtgagtg ggattggggt ggcgttccgt tcctgcctcc ggaactgatg
ctcctgccga accgtattcc gttcaacatt tatcagtttg ccatgtggag ccgtacgact
attgtgccga tgtccgttct cctgtcttcc aagcctatcc atccagtagc ggaggaagca
accatcgatg aactgtatct gaacggtcgc gaaaacgctg actactccat gccgacccca
tctggcctcg gtatcgaacg tctgatgtac gcgggcgatc gtatcctgcg tctgagcaac
ctcctgccgt ggaacccagc gcgtggccgt gcactgcgta tggcggagaa atggattgtc
aaccatcagg aaaaagatgg tagctggggt ggcattcagc caccgtgggt ctactccctg
atggctctga acgaactggg ctatagcaac tcccacgaag tgatcaagaa aggtatggaa
ggtttcgaac tgtatggcat tgaacgtgaa gacacttggc gcctccaggc ttctatgagc
ccactgtggg acacctgtct gagcatcaac gctctgatcg attctggtat cgaacctgat
catccggcaa tcgtgcgtgc tgcagactac ctcctggacc gtcaagcgtc ctctccgggc
gactggcagg ttaaagctcg taacgtggag ccgggtggct gggcattcga gttctccaac
gaaacctacc ctgacactga tgacgctgcc gaagtactcc tggcaattgg tgccgcgggt
gtgaccgatt ccgcacgccg tgacgattct atcaaacgtg gtgtcaactg ggttctggct
atgcaatcca agaatggtgg ctggggtgcc tacgataaag ataacacttc taccctggtt
actaaaatgc cgttctttga tgctggcgaa accatcgatc caccgtccgt cgatgttact
gcccacattg ttgaaatgct ggcaaaactc cgtttcccaa ctgacactcc ggagatcaac
gcagccctgg actatatctg gaaagaacag gaagaggacg gttgttggtt cggccgttgg
ggtgtaaact acgtttatgg cactgcagcc gtcctgccgg cgctggaatc tctgggcatt
gaccagaacg acccgcgtat ccaacaggcg gctgattggc tggaaatgca tcagaactcc
gatggtggct ggggtgaatc ctgcgcctct tatgcaaacc cgagcctgcg tggtcagggt
gcctccaccg cctctcagac cgcttgggcg ctcatgggtc tgatttctac gggtcgtgcg
gctggtgaat ccgcgaagat gggcgtggaa tacctcctgg gcactcagct ggcggacggt
tcctggaccg aagacgaata taccggtacc atgtttccgg gttacggtat tggcgagcgt
aaatttaccg gcctcgaaac ggaagatcac gatctgatca gcgaagagct cccggctggt
ttcatgatca aataccacat gtatcgtatt tactggccgc tgatggctct gggtcgttat
cgcacctgtc tgtcttccgg tagctaa

Sequence Number (ID): 149
Sequence Name: OYT72085.1
Length: 1977
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1977
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgactggtt tcgctccaag attcgttcaa ccagtcgttg agtccccaat tgctccagct
ccacgtgccg ctagaccagc cccagcttca gctgcctccg ttggtgctgc cattgctaga
gctcaatctt acttgttatc tcgtcaatat cctgaaggtt actggtgggc tgaattggaa
gctaacgtta ccttaactgc tgaatacgtc ttcctacaca agattttggg tactgacggt
gctagagcca gacaattggg taagatcaga acttacttgc gtagacaaca aagagatcat
ggcggttggg aattatatta cggtgacggc ggtgaattgt caacttccgt cgaggcttac
ttcgctctaa agttattggg tgacgcccct gacgctcctc acatggctag agccagagac
ttcatcctag ctagaggcgg tgttgccaag gccagagtct tcaccaagat tcatttggct
ctattcggtg ctttcccatg ggaaggttgt ccaactttac ctccatggat tatgttgcta
cctgattggt tccctttcac tatttacgaa ttggcttctt gggctagatc ctctactgtt
ccattgctat tagtcggtga caaaaagcca gttgtcgctg ttccaggcgg tgccgctgat
gaactatatg ccgaaggtag agcttccgcc gatttggcct tgccaaaccc agacggcatg
ttgtccttag gcggtgcttt catcgctttc gacaaggctt taaaattgat ggaaagattg
aacttttctc caagaaaggc cgaagcttta gctttggccg aaagatggac tttagaccat
caagatgaat ccggtgactg gggcggtatc attccagcta tgttgaactc cttattgggt
ttgcactgta gaggttatgc tccagaccat ccagctatga gaaagggtat cgaagctgtc

gagagatttt gtattgaaac cgaagatgaa tttcacaccc aaccatgtgt ctcaccagtc
tgggataccg gtttgactat tttggcttta ttggattctg gtttaccaaa tgatcatcca
gctttagtca aagctggtga gtggctattg tctaaacaaa ttttaagaga cggtgattgg
agattcaaga ataaaacagg tccagccggc ggttgggcct ttgaattctg gaacgacttc
tttcctgatg ttgacgatac tgctgtcgtt accatggctt tacatagatt gaagttgcca
gatgaagctg aaaagcaacg tagattaaag ttggctactg aatggacttt gtctatgcaa
tccaaaaacg gcggttgggg tgcttttgac gttgataacg acctagctat cttgaacgaa
atcccatacg gtgatttgaa ggctatgatt gatcctccaa cagctgactt aaccggtcac
atcttagaaa tgttgggtgt cactggttac cctgctccta gagaaagagt tgaaagagct
attgctttta ttaagtccca acaagagcca gaaggttgtt ggtggggtag atggggtgtc
aactacatct atggtactca catggttatc tgtggcctag ttgccttggg tttgaaccca
agagaagcct tcattatgag aggtacccaa tggttgaact cttgtcaaaa cgaagatggc
ggttggggcg aaacctgtgc ttcttacggt gaccgtagat tgatgggtat cggtgcctca
gccccatcac aaactgcctg ggctttgtta ggtttgatcg ctggcggtga aggtaaatct
gactgtgcta gacgtggtat cgaatactta gttactagac aaaacgacga tggcggttgg
actgaagccg ctttcaccgg tactggtttc ccaaaccact tctacatgaa ttaccacttc
tatcgtcatt acttcccatt gatggctttg ggtcgttaca gaccattcgc taagtga

Sequence Number (ID): 150
Sequence Name: OYT72085_V1
Length: 1977
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1977
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaccggct tcgccccacg cttcgtccaa cctgttgttg aatccccgat cgcaccagcc
ccgcgtgcgg ctcgtccggc tcctgcgagc gccgcaagcg ttggtgcggc gatcgcacgc
gcgcaaagct atctgctgag cagacaatat ccggaaggct attggtgggc ggaattagag
gccaacgtca ccctgacggc tgaatacgtg ttcctgcaca aaattctggg tactgatggc
gctcgtgcgc gtcagctggg caagattcgc acctacttgc gtcgccagca gcgcgatcac
ggcggttggg aactgtacta cggcgacggt ggcgagctgt ccacctcagt cgaggcgtat
ttcgcgttga aattgctggg cgatgctccg gacgcaccgc acatggcgcg tgcgcgtgat
ttcattttgg cgcgtggcgg cgttgccaag gcccgtgtat ttaccaagat tcatctggcg
ctgttcggtg ctttcccgtg ggagggttgc ccaaccctcc cgccgtggat tatgctgctg
ccggactggt ttccgtttac catctatgaa ctggcatcgt gggcacgctc cagcaccgtt
ccgctgctgt tggttggtga taagaaaccg gtcgtcgccg tgccgggtgg cgcagcagac
gaactgtacg ctgagggtcg tgcgtctgcg gacctggctc tgccgaatcc ggatggcatg
ttgagcctgg gtggcgcatt catcgcgttc gacaaagcac tgaaactgat ggaacgcctg
aattttagcc cgcgtaaagc tgaggccctc gcgttagccg aacgttggac gctggaccat
caagatgaga gcggtgactg gggtggcatc atcccggcaa tgctgaacag cctgttgggt
ctgcattgcc gtggttatgc accggaccac ccggcgatgc gtaagggtat tgaagccgtt
gagcgttttt gcattgagac tgaagatgaa ttccacaccc agccgtgtgt ttctccggtg
tgggataccg gtctgacgat tctggcgctg ctggactccg gtctgccgaa cgaccatcct
gcactggtta aggccggtga gtggctgctg agcaaacaga tcctgcgtga tggtgattgg
cgctttaaga acaaaacggg cccggcgggt ggctgggcgt tcgagttctg gaacgatttc
tttccagacg tggacgacac cgcagtggtg acgatggcgc tgcaccgcct gaagctgccg
gacgaagcgg agaaacaacg tcgcctgaaa ttggcgaccg aatggaccct gagcatgcag
tctaaaaacg gtggctgggg cgcttttgac gtggataatg acctggcgat tctgaatgaa
atcccgtacg gcgacgcgaa ggcaatgatt gacccgccga ccgcggatct gaccggccac
atccttgaga tgctgggtgt cacgggttac ccggcaccgc gtgagcgtgt ggagcgtgcg
attgcgttta tcaagagcca acaagagcca gagggctgtt ggtggggtcg ctggggtgtc
aattacatct atggtaccca catggtgatt tgtggtctgg ttgcactggg tctgaacccg
cgtgaagcgt ttatcatgcg cggtacgcag tggttgaata gctgtcagaa tgaagatggc
ggttggggcg agacttgcgc gagctacggt gatcgtcgcc tgatgggtat cggtgccagc
gcaccgagcc aaacggcatg ggccctgctg ggtctgattg cgggtggcga gggtaagagc
gattgcgccc gtcgtggtat cgagtacctg gtcacccgtc agaatgacga cggcggctgg
accgaagccg cgttcacggg taccatgttt cctaaccatt tctatatgaa ctaccacttt
tatcgtcatt actttccgct gatggcgttg ggtcgctacc gcccgttcgc gaagtaa

Sequence Number (ID): 151
Sequence Name: OYT72085_V1
Length: 1977
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1977
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaccggtt ttgctccaag attcgtccaa ccagtcgttg aatctcctat tgctcctgct
cctagagccg ctagaccagc tccagcttct gccgcttccg ttggtgccgc tatcgctcgt
gctcaatcct acctattgtc tcgtcaatat cctgaaggtt actggtgggc cgaattagaa
gccaacgtta ctctaactgc tgaatacgtc tttttgcaca agattttagg tacagatggt
gccagagccc gtcaattggg taaaattaga acttacttgc gtagacaaca aagagaccac
ggcggttggg aattgtatta cggtgacggc ggtgagttgt ctacctctgt cgaagcttac
tttgctttga agctattggg tgacgctcct gatgctccac atatggctcg tgctagagac
ttcatcttgg ctcgtggcgg tgttgctaag gccagagttt tcaccaagat tcacttggct
ttgtttggtg ctttcccatg ggaaggttgt ccaaccttgc caccttggat tatgttattg
ccagattggt tcccatttac catctatgaa ttggcttcct gggccagatc ctctactgtt
cctttgctat tagttggtga taagaaacca gtcgttgccg ttccaggcgg tgctgccgat
gaattgtacg ctgaaggtag agcctctgct gacttggctt tgccaaaccc agacggtatg
ttgtccttag gcggtgcttt tattgctttc gataaggctt tgaagttgat ggaaagattg
aacttctctc ctagaaaggc tgaagcctta gctttggctg aaagatggac tctagatcac
caagacgaat ccggcgactg gggcggtatt atcccagcca tgttgaactc tttgttaggt
ttgcactgca gaggttacgc cccagaccac ccagctatga gaaagggtat cgaggccgtt
gaaagattct gtatcgaaac tgaagatgaa tttcacactc aaccatgtgt ttcaccagtc
tgggacactg gtttaaccat cttggcttta ttggactcag gtttgccaaa cgatcaccct
gccttagtta aggccggtga atggttgtta tctaagcaaa ttttgagaga cggtgattgg
agattcaaga acaagacagg tccagctggc ggttgggctt tcgaattctg gaacgatttt
ttcccagatg ttgatgacac cgccgtcgtt actatggcct tgcatcgttt gaaattacca
gacgaagctg aaaaacaacg tagactaaaa ctagctaccg aatggacttt gtctatgcaa
tccaagaacg gcggttgggg tgctttcgat gtcgataacg acttagctat tttgaacgaa
attccatacg gtgacgctaa ggctatgatc gatccaccta ctgctgactt gacaggtcac
attttggaaa tgttgggcgt cactggttac ccagccccac gtgaaagagt tgaaagagct
atcgctttta ttaagtctca acaagaacca gaaggctgct ggtggggtag atggggtgtc
aattacattt acggtaccca catggtcatt tgtggtttgg ttgccttggg tttaaaccca
agagaagctt ttatcatgag aggcactcaa tggttgaact cttgtcaaaa tgaggacggc
ggttggggtg aaacttgtgc ttcttacggt gacagacgtt taatgggtat tggtgcctcc
gctccatctc aaactgcctg ggctctattg ggtttgatcg ctggcggtga aggtaaatct
gattgtgcta gacgtggtat cgaatacttg gttaccagac aaaacgacga tggcggttgg
actgaagccg cttttaccgg tactatgttc ccaaaccatt tctacatgaa ttaccatttt
tacagacatt acttcccatt gatggccttg ggtagataca gacctttcgc taaataa

Sequence Number (ID): 152
Sequence Name: WP_092048487_V1
Length: 2133
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2133
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaacagcg gtcacatgcg caacggtagc tccggtgttg catctcgcct gccgcgtgaa
tccacctctc gcctcgactc ctctgtgaaa tccgccgtgt cccaggcgcg taactggctc
ctgtctgaac aaaattccga gggctattgg ctgggtgaac tgcagggcga caccattctg
gaaagcgaat acatcctgct cctggcgtgg atgggtaaat ccaatacgcc gattgttcag
gaatgtgcta actacatccg ccagcaacag ctgccggaag gcggttgggc tatgttcccg
ggtggcccac tggaaatttc ctcttccgtt aaagcatatt ggactctgaa aattgcaggt
gacgatccgc aggcggagca catgcagcgt gcgtgtgcag cgatccgtgc ggccggcggt
gcagaacgtg ttaactcctt tacgcgttac tatatggctc tcctgggcat tatctcctat
cgtcagtgcc ctgccgtgcc accggaactc atgctcctgc cgaagtggat gccgtttaac

atctatgaaa tgtcttcctg gagccgcact atcattgtgc cgctgagcct cctgtgggcg
ttccagccga aaactaccct cccgcgcagc cagaaaatcg acgaactgtt cctgaactct
ccggaaaaac tgcctgtcgt tatgccaccg agcggtcaac tggacaaact gaaacaacag
acttgggtac cgtgggatcg catctttcgt ggtatcgatg tgacctggaa gtttttcgaa
gctctgcgta tgaaaccgtt ccgtgaacgt gccgtgcgtc tggccaccaa atggatcgtg
aaacgctttg aaaaatctga cggcctgggc gcgatcttcc caccgatcat ttggagcgtt
atcgcgctcc gctgtctggg ccatgacgag agctctccga tggttcaggc ggctctgaag
gaactcgaga aactgaccat tcgcgagggc aacaccgcgc gcctggaacc gtgccgtagc
ccggtttggg acaccgctat cgcagttaac gctctgcgtg atgcgggtgt tccggctcat
caccctcagc tcgtgcgtgc ggtgaattgg ctcctgagca aagaagtgcg tagcccgggc
gactggaccg ttaaccatcc agacgtagag ccgggcggtt ggtattttga atttaacaat
gaattttacc cagacgttga cgataccatc atggtatcta tggccctggc gcgttgcctg
ccgggcgatc agcactctaa ctggtccgct agcctcctga gcaaacaggg cgtccagcaa
cgtagcgatt tcgacctggc gatcgtgatc gcgggtcaaa ccgacgcgcc tgagcgcgct
gttagcgacg ttgaacgtat gcagccgatg atcgctgcgc tgcgtcgcgc ggttaaatgg
accgtcgcaa tgcagtctcg taatggtggc tggggcgcat tcgatgctga caacgaccgt
gagatcctca cccgtgtccc gttcgccgac gcgaacgcga tgatcgaccc tccgaccgct
gacatcactg ctcgtgtcct ggaaatgttc ggtcgtctgg gtctgacttc tcgcgaaccg
atcttcgaaa aagcactgaa attcgtttgg gatgaacaag aaccggacca ctgctggttt
ggtcgctggg gtgtcaacta tatttacggt acctggcagg tactggtcgg tctgactgag
tttggtgtgc cgcgtgatga ctctcgtctg caggccgcgg cacgctggct gaaagaaaaa
caacagtctg atggcggttg gggcgaaacc gcgcaatcct atgacgaacc ggcactgcgt
ggcaccggcg taacgacccc atctcagacc gcttgggcgg tgctgggtct gatcgccgcg
ggtgagggcc gttccattgc cgctcgccgt ggcatcgagt ttctcctgag ccgccagact
gagcagggca cgtgggatga aaccgagttt acgggtacca tgtttccgcg tgttttctat
ctgcgttacc acctgtaccg ccattatttc ccgctgatgg ctctgggtcg ttacgctgcg
cagtttgaat ccgaagactc cgaagcgtac taa

Sequence Number (ID): 153
Sequence Name: WP_092048487_V1
Length: 2133
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2133
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaactctg gtcatatgcg taacggttct tccggtgttg cttcaagatt gccaagagaa
tctacttctc gtttggactc ctctgtcaaa tccgctgttt ctcaagctag aaattggtta
ttgtctgaac aaaattctga aggttactgg ttgggtgaat tacaaggtga taccatttta
gaatctgaat atattttatt gttagcttgg atgggtaagt ctaatactcc aatcgttcaa
gaatgtgcta actatattag acaacaacaa ttgccagaag gcggttgggc catgttccca
ggcggtccat tggaaatttc ctcttccgtc aaggcttact ggactttgaa gattgccggt
gacgatccac aagccgaaca catgcaaaga gcttgtgctg ccattagagc cgctggcggt
gctgaaagag ttaactcctt tactcgttat tacatggctt tgctaggtat tatctcttac
agacaatgtc cagctgtccc tccagaattg atgttgttac caaaatggat gccattcaac
atttatgaaa tgtcctcttg gtcaagaacc attatcgttc ctttgtcttt gttatgggct
ttccaaccaa aaaccacttt gccaagatct caaaaaatcg atgaattgtt cttgaactct
ccagaaaagt tgccagttgt catgcctcca tctggtcaat tagataagtt gaagcaacaa
acatgggtcc catgggatcg tatcttcaga ggtattgacg tcacttggaa gtttttcgaa
gctttgagaa tgaagccatt cagagaaaga gctgttagat tggctaccaa gtggattgtt
aagagattcg aaaagtccga cggtttaggt gctatcttcc ctccaatcat ttggtctgtt
atcgctttga gatgtttagg tcacgatgaa tcttccccta tggttcaagc cgctttgaag
gaattggaaa agttaaccat tcgtgaaggt aacactgcta gattggaacc atgtcgttcc
catcctcaat tggttcgtgc tgttaactgg ttattgtcca aagaagttag atccccaggt
gattggactg tcaaccaccc agacgttgag ccaggcggtt ggtactttga attcaataac
gaattctacc ctgacgtcga tgacactatt atggtttcta tggctttggc tcgttgtttg
ccaggtgacc aacactcaaa ctggtctgct tctttattgt ccaaacaagg tgttcaacaa
agatccgact tcgacttggc tattgtcatt gctggtcaaa ccgatgctcc agagagagcc
gtctctgacg tcgaaagaat gcaaccaatg attgctgcct tacgtagagc tgttaaatgg
acagttgcta tgcaatctag aaacggcggt tggggtgctt tcgacgccga caacgaccgt

gaaattttga ccagagttcc atttgctgat gccaacgcta tgattgaccc tccaactgcc
gatatcactg ctagagtttt ggaaatgttc ggcagattgg gtctaacttc tagagaacca
attttcgaaa aggctttgaa gttcgtttgg gatgaacaag aaccagatca ttgttggttc
ggcagatggg gtgtcaacta tatctatggt acatggcaag tcttggttgg tttgaccgaa
ttcggtgttc caagagatga ctctagattg caagctgccg ctagatggtt aaaggaaaag
caacaatccg atggcggttg gggcgaaact gcccaatctt acgatgaacc tgctttaaga
ggtaccggtg tcactacccc atctcaaacc gcctgggctg tcttgggttt aattgctgcc
ggtgaaggta gatctattgc cgctcgtaga ggtattgaat tcctattgtc tagacaaacc
gaacaaggta cttgggatga aactgaattc accggtacca tgtttccacg tgttttctac
ttaagatacc acttgtacag acattatttc ccattgatgg ccctaggtag atacgctgcc
caattcgaat ccgaagattc cgaagcttac taa

Sequence Number (ID): 154
Sequence Name: WP_051467941.1
Length: 678
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..678
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggatgttg cggatgtttt gagcttgggt agcggtttgg cgtggacggc cacttatttg
ctgattatct ggaccaattg gcgtgagaaa acctacggta tgccgatcgc ggcactgggt
gcgaacctgg gctgggaatt cctgttcagc tttgtgcgtc cgggcgatgg tatgcaactg
gttgttaact acgtctggtt cggttttgac gtggcgattc tggcgctggt tgtggcatac
ggtccgcgcg aatttcgttt cctgccgcgt tggggcttcc tggcaatgtt agcttccgtg
ctggtcatgg gttacctggg tgtagacctg gtcagccgcc agtttgacca tggcctggcg
acctttgcgg cgttcggtca gaatcttatg atgagcggcc tgttcctgag catgctgatc
gcgcgtggta gcacgcgtgg ccagagcgtc tggattgcac tgaccaaggg cgtcggtacg
gccctggcat ctggtgcctc ctggatctgg gcgcaagatg agccttggcg tcacggctcg
ctgctgccgt acctgatgat taccaccgct gttctggacc tcgcgtatct ggtggcagtg
tatgccgttg cgcgtcgcga ggccggtggc agcgctagcg caccactgcg cctgaaccgc
gtgccggaac cggtctaa

Sequence Number (ID): 155
Sequence Name: WP_051467941.1
Length: 678
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..678
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atggacgttg ctgatgtctt gtctttgggt tctggtttag cttggactgc tacttactta
ttgatcattt ggaccaattg gagagagaag acttacggta tgccaatcgc cgctttgggt
gctaacttgg gttgggaatt cttgttctcc tttgtcagac ctggtgatgg tatgcaattg
gttgtcaatt acgtctggtt cggtttcgat gttgctattt tggctttggt cgttgcttac
ggtccacgtg aattcagatt cttaccaaga tggggtttcc tagctatgtt agcttctgtt
ttggtcatgg gttacctagg tgttgatttg gtctctcgtc aattcgacca cggtttggcc
accttcgccg cttttggtca aaacttaatg atgtcaggct tgttcttgtc catgttaatc
gccagaggtt ccactagagg tcaatctgtt tggattgcct taactaaggg tgttggtacc
gctttggctt ccggtgcttc ttggatttgg gcccaagatg aaccatggag acacggttct
ttgttacctt acttgatgat cactaccgct gttctagatt tagcctactt agttgctgtc
tatgctgttg ctcgtagaga agctggcggt tcagcctccg ctcctttgag attgaacaga
gtcccagaac cagtctag

Sequence Number (ID): 156
Sequence Name: WP_234754442.1

Length: 660
Molecule Type: DNA

Features Location/Qualifiers:
- source, 1..660
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized

Residues:
atgaatctgt ttctgacgat tttgagcggt gtggcctgga cgacggttta tatttgtgcg
attcgtatcg gtttccgtga caaaacctat gcaatcccgg ctgcagcgct gggcctgaat
ttcgcctggg aagtgatcta cagcgttcac agcctgagca ccggtctgtc ggtccagggt
gttatcaaca tcgcgtgggc gctggcagat gtggcgatcg tatatacgtt cttcgcgttt
ggtcgtcgtg agctgccggg ttttctgacc cgtccgctgt tcatcggttg ggccgtctta
ttgggcttgg caagctttgc ggttcaatgg ctgttcattg cggaatttga ctgggacccg
gcgagccgct acgcggcatt tttgcagaac cttctgatgt ctggtctgtt catcgcaatg
tttgctgctc gccgctccct gcgtggccaa agcctggtca ttgccgttgc gaagtggatc
ggtaccctgg cgccgactat taccttcggc gtgctggaga gcagcctctt cattctgggc
attggtgtgc tgtgcagcat ctttgacctg acctacattg gcctgctgct gtggtccaag
aaaaaccctg gtgccctgtc tcgcgatcgt gatagcggtg gcctgccggc agtcccataa

Sequence Number (ID): 157
Sequence Name: WP_234754442.1
Length: 660
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..660
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized

Residues:
atgaacttgt tcttgactat tttgtccggt gtcgcttgga ctaccgtcta catttgtgct
attagaatcg gttttagaga caagacctac gctatcccag ctgccgcttt gggtttaaac
ttcgcttggg aagttattta ctctgttcac tccttatcta ctggtttgtc agttcaaggt
gttattaaca tcgcttgggc tttggctgac gtcgccatcg tctacacttt tttcgctttc
ggtcgtagag aattgccagg ttttttgacc agaccattgt ttatcggttg ggccgtccta
ttgggtttgg cttctttcgc tgtccaatgg ttattcatcg ccgaattcga ttgggaccct
gcttcaagat atgccgcttt cttgcaaaac ttgttaatgt ccggtttgtt tatcgctatg
ttcgccgctc gtagatcttt gcgtggtcaa tctttagtta ttgctgttgc taagtggatc
ggtaccttgg ctccaactat cactttcggt gtcttggaat cttccttgtt catcttgggt
attggcgttt tgtgttctat cttcgacttg acctacatcg gtttgttatt gtggtctaaa
aagaacccag gtgctttgtc tagagacaga gattctggcg gtctacctgc tgttccataa

Sequence Number (ID): 158
Sequence Name: WP_190963420.1
Length: 648
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..648
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized

Residues:
atgggtactt atttgatgtt gggcagcggc gcattctgga ttctgacgta tattctgttg
attgagcgtg gctttaaaga tcagacgtat ggtatgcctc tggtggcgct gtgtgctaac
ctgtcttggg aattcatctt tagcttcatt cacccgcacc aaccgccgca attgcagatt
aacattgtct ggctgatgct ggacctgatc atcctgtacg gctttttcaa attcggccag
agcgagctta aagatatccc gaacaagctg ttttacccgg ttttcatcct gacgctgttt
acctccttct gctgcgtgct gttaatcacc gacgaatttc aagactggag cggtgcgtac
accgccttcg gtcagaatct gctgatgagc attctgttca ttgacatgct caccaagcgc
aatacggtcc gtggtcagag catctttatt gcgatcttta agatgattgg caccctgctg
gcctcgattg gtttctacat taacaatcca atccagggtc gcagcctgct gttcatcttt
ctgtacaccg cgatttttgt tttcgatctg atctatgttg gtatgatcgc aatgaaaatc
aagcgcttcc gtgagaagaa actgaatcat gcgctgaccc gtcaataa

Sequence Number (ID): 159
Sequence Name: WP_190963420.1

Length: 648
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..648
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgggtacct atttgatgtt aggttctggt gctttctgga ttttaactta cattttattg
attgaaagag gtttcaagga ccaaacctat ggtatgcctt tggttgcctt atgtgctaac
ttgtcatggg aattcatctt ctctttcatc cacccacatc aaccacctca attgcaaatt
aatattgttt ggttgatgtt ggatttgatt atcttatacg gttttttcaa gtttggtcaa
tcagagctaa aggacatccc taataagttg ttctacccag ttttcatctt gactctattt
acctccttct gctgtgtttt attgatcact gacgaatttc aagactggtc tggtgcctac
actgctttcg gtcaaaactt attgatgtcc attctattca ttgacatgtt gactaagaga
aacaccgtta gaggtcaatc tatctttatc gctattttta agatgattgg taccttattg
gcctctattg gtttctatat taacaatcca atccaaggta gatctttatt gtttatcttc
ctatatactg ccatcttcgt tttcgattta atctatgttg gtatgatcgc tatgaagatt
aaaagattta gagaaaaaaa gctaaatcat gctttaacta gacaataa

Sequence Number (ID): 160
Sequence Name: WP_093699331.1
Length: 654
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..654
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgcatactg cttttctgtt gggtactgga ggtttttgga cggtcgcata cgttttgttg
atccgcacgg gcctgcgcga gcgtaccttc ggcatgccgg ttgttgcctt cgcgaccaat
atcagctggg aattcatgtt cgccttcgtc cgtccgccga ccggcgtgat gcacgtcgta
aacattgttt ggttctgctt tgacgtggca attggttata cgctggttcg ttttggtcgt
gcggaattcc cttatctgcc gcgttccctg ttcctgccag cgctgctggc cctgctggcg
ctggcgtacc cgggtatgaa ctatgccagc gagcgctttg acgagggtgc tggcgcggtc
accgcgtttg gtagcaactt ggcgatgagc ggtatgttcc ttgcaatgct ggcggctcgt
cgcggtaccc gtggtcagag cgccggtatc gcgctggcaa aactgctggg caccgtgtgc
gccagcctgt ctatgctgac cgatccgggc ctggaaccgc gtcacgataa tgcgctgatg
tactacctgt atgttggttg ttttctgtta gacgctgcgt acctcggcgc agtgctggca
gtgcgccgtg cggagcgtgc ggtcgcaccg gtgacggtga ccgttccggt ctaa

Sequence Number (ID): 161
Sequence Name: WP_093699331.1
Length: 654
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..654
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgcacactg ccttcttgct aggtacaggt ggcttttgga ccgtcgctta cgtcttgtta
attagaactg gtttgagaga gagaactttc ggtatgcctg ttgtcgcctt cgctacaaac
atttcctggg agttcatgtt cgctttcgtt agacctccaa ctggtgttat gcatgtcgtt
aacattgtct ggttctgttt tgatgtcgcc attggttaca ctttagtcag attcggtaga
gctgaattcc catacttgcc tcgttcttta ttcttaccag ctttattggc tttgttagct
ttggcttatc ctggtatgaa ctacgcttct gaaagattcg atgaaggtgc cggtgctgtt
accgcctttg gctccaactt ggctatgtct ggtatgttcc tagccatgtt ggccgctcgt
agaggtacca gaggtcaatc tgctggtatt gctctagcta agttattggg tactgtctgt
gcttctttgt ccatgttgac tgacccaggt ttggaaccaa gacatgacaa cgctttaatg
tattacttgt acgttggttg ttttttattg gacgccgctt acttgggtgc cgttttggct
gtccgtagag ctgaaagagc cgttgctcca gtcactgtta ctgtcccagt ttaa

Sequence Number (ID): 162
Sequence Name: WP_067007865.1
Length: 675
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..675
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggcagcag ttcacaccgc atttttgttg ggcaccggtt tattttggac ggcagcgtac
gtgctgctga ttcgtaccgg tctgcgtgcg cgtaccttcg gcatgccagt ggttgcattt
gcgaccaata tcagctggga attcatgttt gctttcgtcc gtccgccgtc cggcgtaatg
catgtgatta acatcgtctg gttttgcttc gacctggcaa ttggttatac tgttgtgcgc
tttggtcgtg ctgagttccc gtatctgcct gaccgcttat ttctgccggc actggccgtc
ctgctggccc tggcgtatcc gggtatgaac tacgtgagcg agcaattcga cgagggtgtc
ggtgcgatca cggcgttcgg cagcaatctg gccatgtctg gtatgttcct ggcgatgctg
gccgctcgtc gcggtacccg cggtcagagc gttggtattg cggttaccaa actgctcggc
acggcctgcg cgagcctggc gctgctgacg gatccggatg gtgacccgcg ttacgataac
gcgttgatgt attacttcta catcggctgt tttcttctgg atctggcgta cgcgtatgcc
gtgtttgcgg tcggccgtgc ggaacgcacc acgggtagcg ctcaggttcc ggcccagggc
gcactgcaac gttaa

Sequence Number (ID): 163
Sequence Name: WP_067007865.1
Length: 675
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..675
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atggccgctg tccataccgc ttttttgtta ggtactggtt tgttttggac cgccgcttac
gtcttgttaa tcagaactgg tttgagagcc agaacattcg gtatgccagt tgtcgctttc
gccactaaca tctcttggga atttatgttt gccttcgtca gacctccatc cggtgtcatg
cacgttatca acattgtttg gttctgtttt gacttggcta tcggttacac cgtcgttaga
ttcggtagag ccgaattccc atacttgcca gatagattat ttttaccagc tttggctgtt
ttgttagctt tggcctaccc aggtatgaac tacgtctccg aacaattcga cgaaggtgtc
ggtgctatta cagcttttgg ttctaacttg gctatgtcag gcatgttttt agctatgttg
gccgctcgta gaggtactag aggtcaatct gttggtattg ctgttactaa attattgggt
accgcctgtg cctctttggc tttgttaacc gatccagatg gtgacccaag atacgacaat
gctttgatgt actatttcta cattggttgt ttcctattgg acttggctta cgcctatgct
gttttcgctg ttggtagagc cgaaagaacc actggttcag cccaagttcc tgctcaaggt
gctttgcaaa gataa

Sequence Number (ID): 164
Sequence Name: WP_220206969.1
Length: 663
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..663
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgtttgtgc tgttgatgtt gggttccggt ttgttttgga cgatcactta cattctgatt
attcgtcgtt ccattctgga ccgtacctac ggcatgccgc tggcggcttt gtgtgcaaac
atcagctggg aattcatttt cagctttatc ctgccgagca gcagcattca gcgcattgtg
aatatcatct ggttcgtttt ggacgcgggc atcctggtgt gcttcctgcg ctacggccgc
aatgaattcg caaacctgag caaatggatt ttctttacga ccttcggtct gacgctggct
acctcctttg gcgcggttct gctggtcacc cttgagttcc acgattcggg tgcctactca
gcatttggcc agaacctgat gatgagcgcg ctgtttatcc tgatgctgta tcgccgtggt
agcctgcgtg gtcaaagcat cgcgattgct gttaccaagc tgttaggtac ggccctggcc

agcctggcgt tctttctgta taccaccatc tctcacaatt ctgtcctgct gccgttcctg
tatgtgagca tcctcgtcta cgatatgatc tatgttgcga tggtctacaa gcaacaacgt
gcggcgaaac agacgagcat tgaagccacc gcaagcgcaa gccatgttga gctgagcctg
taa

Sequence Number (ID): 165
Sequence Name: WP_234754442.1_S9C
Length: 660
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..660
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaatctgt ttctgacgat tttgtgcggt gtggcctgga cgacggttta tatttgtgcg
attcgtatcg gtttccgtga caaaacctat gcaatcccgg ctgcagcgct gggcctgaat
ttcgcctggg aagtgatcta cagcgttcac agcctgagca ccggtctgtc ggtccagggt
gttatcaaca tcgcgtgggc gctggcagat gtggcgatcg tatatacgtt cttcgcgttt
ggtcgtcgtg agctgccggg ttttctgacc cgtccgctgt tcatcggttg ggccgtctta
ttgggcttgg caagctttgc ggttcaatgg ctgttcattg cggaatttga ctgggacccg
gcgagccgct acgcggcatt tttgcagaac cttctgatgt ctggtctgtt catcgcaatg
tttgctgctc gccgctccct gcgtggccaa agcctggtca ttgccgttgc gaagtggatc
ggtaccctgg cgccgactat taccttcggc gtgctggaga gcagcctctt cattctgggc
attggtgtgc tgtgcagcat ctttgacctg acctacattg gcctgctgct gtggtccaag
aaaaaccctg gtgccctgtc tcgcgatcgt gatagcggtg gcctgccggc agtcccataa

Sequence Number (ID): 166
Sequence Name: WP_234754442.1_S9C
Length: 660
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..660
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaacttgt tcttgaccat cttgtgtggt gttgcctgga ccactgttta tatctgtgcc
attagaattg gtttcagaga caaaacttac gctatcccag ctgccgcttt gggtttgaac
ttcgcttggg aagtcatcta ctctgtccac tcattatcca ccggtttgtc tgttcaaggt
gttattaata ttgcttgggc cttggctgac gtcgctattg tctacacatt tttcgctttt
ggtagacgtg aattgccagg tttcttaact agaccactat ttattggttg ggctgtctta
ttgggcttag cttcttttgc tgtccaatgg ttgttcatcg ctgaattcga ttgggaccca
gcttccagat atgctgcctt cttgcaaaat ttattgatgt ccggtttgtt tattgctatg
ttcgccgcta gacgttcctt gagaggtcaa tctttggtta ttgccgttgc taaatggatt
ggtactttgg ctccaaccat cactttcggt gttttagaat cctctctatt tatcttgggt
attggtgttt tgtgttccat tttcgatttg acatacattg gtctattgtt atggtccaaa
aagaacccag gtgccttgtc tagagacaga gattccggtg gcttgccagc cgtcccatga

Sequence Number (ID): 167
Sequence Name: WP_234754442.1_S9M
Length: 660
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..660
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaatctgt ttctgacgat tttgatgggt gtggcctgga cgacggttta tatttgtgcg
attcgtatcg gtttccgtga caaaacctat gcaatcccgg ctgcagcgct gggcctgaat
ttcgcctggg aagtgatcta cagcgttcac agcctgagca ccggtctgtc ggtccagggt
gttatcaaca tcgcgtgggc gctggcagat gtggcgatcg tatatacgtt cttcgcgttt
ggtcgtcgtg agctgccggg ttttctgacc cgtccgctgt tcatcggttg ggccgtctta

ttgggcttgg caagctttgc ggttcaatgg ctgttcattg cggaatttga ctgggacccg
gcgagccgct acgcggcatt tttgcagaac cttctgatgt ctggtctgtt catcgcaatg
tttgctgctc gccgctccct gcgtggccaa agcctggtca ttgccgttgc gaagtggatc
ggtaccctgg cgccgactat taccttcggc gtgctggaga gcagcctctt cattctgggc
attggtgtgc tgtgcagcat ctttgacctg acctacattg gcctgctgct gtggtccaag
aaaaaccctg gtgccctgtc tcgcgatcgt gatagcggtg gcctgccggc agtcccataa

Sequence Number (ID): 168
Sequence Name: WP_234754442.1_S9M
Length: 660
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..660
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaacttgt tcttaaccat cttgatgggt gtcgcctgga ctaccgtcta tatctgtgct
attagaattg gttttcgtga caagacttac gccattccag ctgccgcttt gggtttgaac
tttgcctggg aggttatcta ctctgttcat tccttgtcta ccggtttgtc cgttcaaggt
gttattaaca tcgcttgggc tttggccgat gtcgctattg tttacacttt ctttgccttc
ggtagacgtg aattgccagg tttcttaacc agaccattgt tcattggttg ggctgtcttg
ttaggcttag cctcctttgc cgttcaatgg ttgttcattg ccgaatttga ttgggatcca
gcctctcgtt atgccgcttt cttgcaaaac ttgttaatgt ctggtctatt catcgctatg
ttcgccgctc gtagatcttt gagaggtcaa tctttggtta tcgctgtcgc taaatggatt
ggtactttgg ctccaacaat caccttcggt gttttggaat cctctttatt cattttgggt
attggtgtct tgtgctccat tttcgattta acttacatcg gcttgctatt gtggtctaaa
aagaacccag gtgccctatc tagagataga gactcaggcg gtttaccagc tgtcccataa

Sequence Number (ID): 169
Sequence Name: WP_234754442.1_S9T
Length: 660
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..660
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaatctgt ttctgacgat tttgaccggt gtggcctgga cgacggttta tatttgtgcg
attcgtatcg gtttccgtga caaaacctat gcaatcccgg ctgcagcgct gggcctgaat
ttcgcctggg aagtgatcta cagcgttcac agcctgagca ccggtctgtc ggtccagggt
gttatcaaca tcgcgtgggc gctggcagat gtggcgatcg tatatacgtt cttcgcgttt
ggtcgtcgtg agctgccggg ttttctgacc cgtccgctgt tcatcggttg ggccgtctta
ttgggcttgg caagctttgc ggttcaatgg ctgttcattg cggaatttga ctgggacccg
gcgagccgct acgcggcatt tttgcagaac cttctgatgt ctggtctgtt catcgcaatg
tttgctgctc gccgctccct gcgtggccaa agcctggtca ttgccgttgc gaagtggatc
ggtaccctgg cgccgactat taccttcggc gtgctggaga gcagcctctt cattctgggc
attggtgtgc tgtgcagcat ctttgacctg acctacattg gcctgctgct gtggtccaag
aaaaaccctg gtgccctgtc tcgcgatcgt gatagcggtg gcctgccggc agtcccataa

Sequence Number (ID): 170
Sequence Name: WP_234754442.1_S9T
Length: 660
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..660
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaacttgt tcttgactat cttgactggt gtcgcttgga ctaccgttta catctgtgcc
atcagaatcg gtttcagaga taagacttac gccatcccag ccgctgcctt gggtttgaac
tttgcttggg aagtcattta ctctgttcac tcattgtcta ccggtttgtc tgtccaaggt

gttattaaca ttgcttgggc tttggctgac gtcgccatcg tttacacttt tttcgccttt
ggtcgtagag aattgccagg ttttttgaca agaccactat tcatcggttg ggccgtttta
ttgggtttag cttcttttgc tgttcaatgg ttgttcatcg ccgaattcga ttgggaccca
gcctctcgtt acgccgcttt cttgcaaaac ttattgatgt ctggtttatt tatcgccatg
tttgccgctc gtagatcctt gagaggtcaa tctttggtca tcgctgttgc taagtggatt
ggtaccttgg ccctactat cactttcggt gtcttggaat cttccttgtt tatcttaggt
atcggtgtct tatgttccat tttcgatttg acttatattg gtttgctatt gtggtccaag
aaaaacccag gtgctttgtc cagagataga gattccggtg gcttgcctgc tgtcccataa

Sequence Number (ID): 171
Sequence Name: WP_190963420.1_S9C
Length: 648
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..648
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgggtactt atttgatgtt gggctgcggc gcattctgga ttctgacgta tattctgttg
attgagcgtg gctttaaaga tcagacgtat ggtatgcctc tggtggcgct gtgtgctaac
ctgtcttggg aattcatctt tagcttcatt cacccgcacc aaccgccgca attgcagatt
aacattgtct ggctgatgct ggacctgatc atcctgtacg gctttttcaa attcggccag
agcgagctta aagatatccc gaacaagctg ttttacccgg ttttcatcct gacgctgttt
acctccttct gctgcgtgct gttaatcacc gacgaatttc aagactggag cggtgcgtac
accgccttcg gtcagaatct gctgatgagc attctgttca ttgacatgct caccaagcgc
aatacggtcc gtggtcagag catctttatt gcgatcttta agatgattgg caccctgctg
gcctcgattg gtttctacat taacaatcca atccagggtc gcagcctgct gttcatcttt
ctgtacaccg cgatttttgt tttcgatctg atctatgttg gtatgatcgc aatgaaaatc
aagcgcttcc gtgagaagaa actgaatcat gcgctgaccc gtcaataa

Sequence Number (ID): 172
Sequence Name: WP_190963420.1_S9C
Length: 648
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..648
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgggcactt acttgatgtt aggttgtggt gctttttgga ttttgaccta catcttgcta
atcgaaagag gttttaaaga ccaaacttac ggtatgccat tggttgcctt gtgtgctaac
ttatcttggg aattcatttt ctccttcatt caccctcatc aacctccaca attgcaaatt
aatattgttt ggttgatgtt ggatctaatc attttgtacg gtttctttaa attcggtcaa
tctgaattga aggatattcc aaacaagttg ttctatccag ttttcatcct aactttgttc
acttctttct gctgtgtttt actaattact gatgaattcc aagactggtc tggtgcttac
accgcttttg gtcaaaactt attgatgtca attttgttca tcgacatgtt gactaagcgt
aacaccgtta gaggtcaatc tatcttcatc gctattttca aaatgatcgg taccttattg
gcctccatcg gtttctatat taataaccca atccaaggta gatccctatt gttcatcttc
ctatacactg ctatcttcgt cttcgatttg atttacgtcg gcatgattgc tatgaagatt
aaaagattca gagaaaaaaa gttgaaccat gctttgacta gacaatga

Sequence Number (ID): 173
Sequence Name: WP_190963420.1_S9M
Length: 648
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..648
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgggtactt atttgatgtt gggcatgggc gcattctgga ttctgacgta tattctgttg

attgagcgtg gctttaaaga tcagacgtat ggtatgcctc tggtggcgct gtgtgctaac
ctgtcttggg aattcatctt tagcttcatt cacccgcacc aaccgccgca attgcagatt
aacattgtct ggctgatgct ggacctgatc atcctgtacg gctttttcaa attcggccag
agcgagctta aagatatccc gaacaagctg ttttacccgg ttttcatcct gacgctgttt
acctccttct gctgcgtgct gttaatcacc gacgaatttc aagactggag cggtgcgtac
accgccttcg gtcagaatct gctgatgagc attctgttca ttgacatgct caccaagcgc
aatacggtcc gtggtcagag catctttatt gcgatcttta agatgattgg caccctgctg
gcctcgattg gtttctacat taacaatcca atccagggtc gcagcctgct gttcatcttt
ctgtacaccg cgatttttgt tttcgatctg atctatgttg gtatgatcgc aatgaaaatc
aagcgcttcc gtgagaagaa actgaatcat gcgctgaccc gtcaataa

Sequence Number (ID): 174
Sequence Name: WP_190963420.1_S9M
Length: 648
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..648
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgggtacct atttgatgtt gggtatgggt gcttttgga tcttaactta tatcttattg
attgagagag gttttaagga tcaaacctat ggtatgccat tggttgcttt atgtgctaat
ttgtcttggg aatttatttt ttcttttatc catccacacc aaccacctca attgcaaatc
aatattgtct ggttaatgtt ggatttgatc attctatacg gttttttcaa gttcggtcaa
tctgagttaa aggatatccc aaacaagttg ttctaccctg tctttatctt gaccttgttt
acttcttttt gctgtgtcct attgattacc gacgaatttc aagattggtc tggtgcttac
actgctttcg gtcaaaacct attgatgtcc atcttgttca ttgatatgtt aactaagaga
aataccgtta gaggtcaatc tattttcatc gctatcttta aaatgattgg tactttattg
gcctccattg gtttttatat caataaccca attcaaggta gatctttgtt atttatttt
ttatacactg ctattttcgt cttcgatttg atctatgtcg gtatgatcgc tatgaagatc
aagagattcc gtgaaaaaaa gttgaaccac gctttgactc gtcaataa

Sequence Number (ID): 175
Sequence Name: WP_190963420.1_S9T
Length: 648
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..648
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgggtactt atttgatgtt gggcaccggc gcattctgga ttctgacgta tattctgttg
attgagcgtg gctttaaaga tcagacgtat ggtatgcctc tggtggcgct gtgtgctaac
ctgtcttggg aattcatctt tagcttcatt cacccgcacc aaccgccgca attgcagatt
aacattgtct ggctgatgct ggacctgatc atcctgtacg gctttttcaa attcggccag
agcgagctta aagatatccc gaacaagctg ttttacccgg ttttcatcct gacgctgttt
acctccttct gctgcgtgct gttaatcacc gacgaatttc aagactggag cggtgcgtac
accgccttcg gtcagaatct gctgatgagc attctgttca ttgacatgct caccaagcgc
aatacggtcc gtggtcagag catctttatt gcgatcttta agatgattgg caccctgctg
gcctcgattg gtttctacat taacaatcca atccagggtc gcagcctgct gttcatcttt
ctgtacaccg cgatttttgt tttcgatctg atctatgttg gtatgatcgc aatgaaaatc
aagcgcttcc gtgagaagaa actgaatcat gcgctgaccc gtcaataa

Sequence Number (ID): 176
Sequence Name: WP_190963420.1_S9T
Length: 648
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..648
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized

Residues:
atgggtacct acttgatgtt aggtactggt gctttctgga ttttgactta cattctattg
atcgaacgtg gtttcaagga ccaaacatat ggtatgcctt tggtcgcttt gtgtgctaac
ttgtcatggg aattcatttt ctcattcatc cacccacatc aacctccaca attgcaaatc
aacattgttt ggttgatgtt ggatttgatt atcttatacg gttttttcaa gttcggtcaa
tctgaactaa aggatattcc taataagttg ttctacccag ttttcatttt gactctattc
acttctttct gctgtgtttt attgattacc gacgagtttc aagactggtc tggtgcctat
actgccttcg gtcaaaacct attgatgtcc atcttgttca tcgacatgtt gacaaagaga
aataccgtta gaggtcaatc tatcttcatc gctatcttca agatgatcgg taccttgtta
gcttccattg gtttctacat taataaccca atccaaggta gatccttatt gtttatttt
ttgtacactg ctatttttgt cttcgatttg atctacgtcg gtatgatcgc tatgaagatt
aagagattcc gtgaaaaaaa gttgaatcat gctttgacca gacaataa

Sequence Number (ID): 177
Sequence Name: WP_234754442.1 T7M
Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaacctgt tcctgatgat cctgagcggc gtagcgtgga cgaccgtgta catctgcgcg
atccgtattg gtttccgcga caaaacctac gcgatcccgg cggcagctct gggcctcaac
tttgcttggg aagtgatcta ctccgttcac tctctgtcca ccggtctgtc tgtgcagggc
gttattaaca tcgcgtgggc gctggcagat gtagcaattg tttacacctt tttcgctttt
ggtcgtcgcg aactgccggg tttcctgact cgtccactgt ttatcggttg ggcggtactc
ctgggcctgg cctctttcgc ggtccaatgg ctgttcatcg ctgagttcga ctgggatccg
gccagccgct acgctgcatt cctgcagaac ctcctgatga gcggcctgtt tattgcgatg
ttcgcggctc gtcgcagcct gcgtggtcag agcctggtta ttgcggttgc taagtggatc
ggtactctgg ccccgaccat caccttcggt gttctggaaa gctctctgtt catcctgggt
atcggcgtgc tgtgttctat cttcgatctg acttacatcg gcctgctcct gtggtccaag
aaaaacccgg gtgcactgtc tcgcgaccgt gactctggtg gcctgcctgc cgtaccg

Sequence Number (ID): 178
Sequence Name: WP_234754442.1 S51T
Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaacctgt tcctgaccat tctgagcggt gtggcgtgga ctaccgtata catttgtgct
atccgtatcg gtttccgtga taaaacctac gccattccag cagcggcact gggtctgaac
ttcgcgtggg aagtgatcta cagcgttcat accctgagca ctggcctgag cgttcagggt
gtaatcaaca ttgcgtgggc gctcgccgac gttgcgatcg tttatacttt ctttgcattc
ggtcgccgtg aactgccggg tttcctgact cgtccgctgt tcattggttg ggcagttctc
ctgggcctgg cgtctttcgc agtgcagtgg ctgttcattg cggagttcga ctgggaccca
gccagccgtt acgccgcatt cctgcaaaac ctcctgatga gcggcctgtt tatcgctatg
ttcgccgctc gccgttccct gcgcggtcag agcctggtaa tcgctgtcgc gaaatggatc
ggtaccctgg cgccgacgat caccttcggc gttctggaat cctctctgtt catcctgggc
atcggtgttc tctgttccat ttttgacctc acctatatcg gcctgctcct gtggagcaag
aaaaacccgg gtgccctgtc tcgcgaccgt gattccggcg gtctgcctgc ggttccg

Sequence Number (ID): 179
Sequence Name: WP_234754442.1 N63Q
Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657

> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized

Residues:
atgaacctgt tcctgaccat cctgtctggt gttgcgtgga ccactgtata catctgtgca
atccgtattg gctttcgtga caaaacctat gccatcccgg ctgcggctct gggtctgaac
tttgcgtggg aagttatcta ctctgtccac tccctgagca cgggtctgag cgtccagggt
gttatccaaa ttgcatgggc actggcggac gttgccatcg tgtacacctt tttcgccttc
ggtcgtcgcg aactgccggg ttttctgact cgcccgctgt tcattggctg ggctgtgctc
ctgggtctgg cttcttttgc agtacagtgg ctgttcatcg ctgagttcga ttgggatcca
gcatctcgtt acgctgcatt cctgcagaac ctcctgatgt ctggtctgtt cattgcgatg
ttcgcggccc gtcgcagcct gcgtggtcag tctctggtta tcgccgttgc taagtggatt
ggcactctcg caccgactat cacgtttggc gtactggaat ctagcctgtt cattctcggc
attggtgtac tgtgctctat ctttgatctg acctatatcg gcctgctcct gtggtctaag
aaaaaccctg gcgccctgtc ccgtgatcgt gacagcggtg gcctgccggc tgtccct

Sequence Number (ID): 180
Sequence Name: WP_234754442.1 N63F
Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaacctgt tcctgaccat cctgtctggt gttgcgtgga ccactgtata catctgtgca
atccgtattg gctttcgtga caaaacctat gccatcccgg ctgcggctct gggtctgaac
tttgcgtggg aagttatcta ctctgtccac tccctgagca cgggtctgag cgtccagggt
gttatcttta ttgcatgggc actggcggac gttgccatcg tgtacacctt tttcgccttc
ggtcgtcgcg aactgccggg ttttctgact cgcccgctgt tcattggctg ggctgtgctc
ctgggtctgg cttcttttgc agtacagtgg ctgttcatcg ctgagttcga ttgggatcca
gcatctcgtt acgctgcatt cctgcagaac ctcctgatgt ctggtctgtt cattgcgatg
ttcgcggccc gtcgcagcct gcgtggtcag tctctggtta tcgccgttgc taagtggatt
ggcactctcg caccgactat cacgtttggc gtactggaat ctagcctgtt cattctcggc
attggtgtac tgtgctctat ctttgatctg acctatatcg gcctgctcct gtggtctaag
aaaaaccctg gcgccctgtc ccgtgatcgt gacagcggtg gcctgccggc tgtccct

Sequence Number (ID): 181
Sequence Name: WP_234754442.1 N63Y
Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaacctgt tcctgaccat cctgtctggt gttgcgtgga ccactgtata catctgtgca
atccgtattg gctttcgtga caaaacctat gccatcccgg ctgcggctct gggtctgaac
tttgcgtggg aagttatcta ctctgtccac tccctgagca cgggtctgag cgtccagggt
gttatctata ttgcatgggc actggcggac gttgccatcg tgtacacctt tttcgccttc
ggtcgtcgcg aactgccggg ttttctgact cgcccgctgt tcattggctg ggctgtgctc
ctgggtctgg cttcttttgc agtacagtgg ctgttcatcg ctgagttcga ttgggatcca
gcatctcgtt acgctgcatt cctgcagaac ctcctgatgt ctggtctgtt cattgcgatg
ttcgcggccc gtcgcagcct gcgtggtcag tctctggtta tcgccgttgc taagtggatt
ggcactctcg caccgactat cacgtttggc gtactggaat ctagcctgtt cattctcggc
attggtgtac tgtgctctat ctttgatctg acctatatcg gcctgctcct gtggtctaag
aaaaaccctg gcgccctgtc ccgtgatcgt gacagcggtg gcctgccggc tgtccct

Sequence Number (ID): 182
Sequence Name: WP_234754442.1 N63A
Length: 657
Molecule Type: DNA

Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized

Residues:
atgaacctgt tcctgaccat cctgtctggt gttgcgtgga ccactgtata catctgtgca
atccgtattg gctttcgtga caaaacctat gccatcccgg ctgcggctct gggtctgaac
tttgcgtggg aagttatcta ctctgtccac tccctgagca cgggtctgag cgtccagggt
gttatcgcca ttgcatgggc actggcggac gttgccatcg tgtacacctt tttcgccttc
ggtcgtcgcg aactgccggg ttttctgact cgcccgctgt tcattggctg ggctgtgctc
ctgggtctgg cttcttttgc agtacagtgg ctgttcatcg ctgagttcga ttgggatcca
gcatctcgtt acgctgcatt cctgcagaac ctcctgatgt ctggtctgtt cattgcgatg
ttcgcggccc gtcgcagcct gcgtggtcag tctctggtta tcgccgttgc taagtggatt
ggcactctcg caccgactat cacgtttggc gtactggaat ctagcctgtt cattctcggc
attggtgtac tgtgctctat ctttgatctg acctatatcg gcctgctcct gtggtctaag
aaaaaccctg gcgccctgtc ccgtgatcgt gacagcggtg gcctgccggc tgtccct

Sequence Number (ID): 183
Sequence Name: WP_234754442.1 N63M
Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized

Residues:
atgaacctgt tcctgaccat cctgtctggt gttgcgtgga ccactgtata catctgtgca
atccgtattg gctttcgtga caaaacctat gccatcccgg ctgcggctct gggtctgaac
tttgcgtggg aagttatcta ctctgtccac tccctgagca cgggtctgag cgtccagggt
gttatcatga ttgcatgggc actggcggac gttgccatcg tgtacacctt tttcgccttc
ggtcgtcgcg aactgccggg ttttctgact cgcccgctgt tcattggctg ggctgtgctc
ctgggtctgg cttcttttgc agtacagtgg ctgttcatcg ctgagttcga ttgggatcca
gcatctcgtt acgctgcatt cctgcagaac ctcctgatgt ctggtctgtt cattgcgatg
ttcgcggccc gtcgcagcct gcgtggtcag tctctggtta tcgccgttgc taagtggatt
ggcactctcg caccgactat cacgtttggc gtactggaat ctagcctgtt cattctcggc
attggtgtac tgtgctctat ctttgatctg acctatatcg gcctgctcct gtggtctaag
aaaaaccctg gcgccctgtc ccgtgatcgt gacagcggtg gcctgccggc tgtccct

Sequence Number (ID): 184
Sequence Name: WP_234754442.1 S9V
Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized

Residues:
atgaatctgt tcctgaccat cctggtgggc gttgcttgga ccactgtcta tatctgcgcg
atccgtatcg gtttccgcga taaaacttat gctatcccgg cagccgcgct gggtctgaac
tttgcttggg aagtgattta cagcgtccac tctctgagca ccggcctgtc tgtacaaggc
gtaattaaca ttgcgtgggc gctggcggac gtcgctatcg tgtacacttt ctttgcgttc
ggtcgtcgcg aactgccggg tttcctgacc cgtcctctgt tcattggttg ggctgttctc
ctgggtctgg cgtctttcgc tgttcaatgg ctctttatcg ctgagttcga ttgggacccg
gcttcccgtt acgcggcctt tctgcagaac ctcctgatga gcggtctgtt tatcgcgatg
ttcgctgccc gccgttccct gcgtggccag agcctggtta tcgcggtggc aaaatggatt
ggtacgctgg ctccaaccat caccttggt gtcctggaat cttccctgtt catcctgggc
attggcgttc tgtgcagcat ctttgacctg acctatattg gtctgctcct gtggtctaag
aaaaacccgg gcgccctctc tcgtgaccgc gatagcggcg gtctgccggc tgtgccg

Sequence Number (ID): 185
Sequence Name: WP_234754442.1 S9A

Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaatctgt tcctgaccat cctggcgggc gttgcttgga ccactgtcta tatctgcgcg
atccgtatcg gtttccgcga taaaacttat gctatcccgg cagccgcgct gggtctgaac
tttgcttggg aagtgattta cagcgtccac tctctgagca ccggcctgtc tgtacaaggc
gtaattaaca ttgcgtgggc gctggcggac gtcgctatcg tgtacacttt ctttgcgttc
ggtcgtcgcg aactgccggg tttcctgacc cgtcctctgt tcattggttg ggctgttctc
ctgggtctgg cgtctttcgc tgttcaatgg ctctttatcg ctgagttcga ttgggacccg
gcttcccgtt acgcggcctt tctgcagaac ctcctgatga gcggtctgtt tatcgcgatg
ttcgctgccc gccgttccct gcgtggccag agcctggtta tcgcggtggc aaaatggatt
ggtacgctgg ctccaaccat caccttttggt gtcctggaat cttccctgtt catcctgggc
attggcgttc tgtgcagcat ctttgacctg acctatattg gtctgctcct gtggtctaag
aaaaacccgg gcgccctctc tcgtgaccgc gatagcggcg gtctgccggc tgtgccg

Sequence Number (ID): 186
Sequence Name: A0A2P1DP74.1 (macJ)
Length: 777
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..777
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgtgctttt tcgctttgga agagtgggct gccgctaata gagattacga aaacacccct
gccccatact ggcatgttaa gtcagttcca gacggtttca ctgctatttc aggtatttta
tggtccatct catacatttt gatggccaaa aaggctttca aagacagatc ttacgccatg
cctctacact gtttgtgttt gaacattacc tgggaagctg tttacggttt cgtttacggt
ccaggtttat tgaaccaagt tgtcttcgcc caatggatga ttgttgatgt cgttttgttc
tacgccatct tgagatctgc cccttatgct tggaagcaat cccctctagt cgcccaacac
ttggctggca tcattgttgt cggttgtgtt atttgtttgt ggctacatct agctattgcc
gctaccttca ttccatctat cggtcgtcaa gtcgttttca tgaccgcctg gccaatgcaa
gtcttgatca atttctcttc catcgcccaa ctattgtcca gaggtaatac tttgggtcac
tcctggggta tctggtggac tagaatgcta ggtactatcg ccgctgcctg ttgctttttc
tggcgtattc attattggcc agaaagattc ggctacgctt ggaccccata cggtaagttc
ttattgttag gctccattgg ttcagatatg gtttacgccg ctgtttacgt ttacgttcaa
cgtatcgaga agcaattaga ttcattagtt aataccaaag ctcaaaaagc tagataa

Sequence Number (ID): 187
Sequence Name: XP_018029969.1
Length: 768
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..768
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgggtttgt tcgctttgga ggaatgggct caagctaatg ctgactacaa taacgatact
cctccatact ggcacgccaa gattgtccct gacttattca ccgccatctc tggtattttg
tggtctgtct cttacatttt gatgactcta aagggttaca aagacagatc ttacgccatg
ccaatctatt gtctatgttt gaacatcact tgggagttcg tcttcggttt catttacggt
cctggtttgg ttaaccaaat cgttttcgcc caatacatgg ttgtcgatgt cttcttgttc
cactctattt tgaaattcgg tccaaacgaa tggagagctc acccattagt tgctagaaac
ttgtcctgga tcattggtgt tggttgtgct gtttgtttgg gtttgcactt ggttttggct
aagaccttcg ttccagttat tggtagacaa gttattttct ttactgcttg gccaatgcaa
cacatgattt ccttaggttg tgttgctcaa gtcttgtcaa gaggtcatga tgctggtcaa
tctatggcta tctggtggac tagattcttg ggtactgtca ctgccggttg ctgtttttat

tggagaatct actttggcc agaacgtttc ggttatgctt ggactccata cggtgctttg
ttattggttg gttctcatgt cttggactta gcttttccat tcgctttggc ttatgttaga
aagcatggtg aaggtagaca agaaaaagtt aacggtaagg ctgcttaa

Sequence Number (ID): 188
Sequence Name: KAG0152682.1
Length: 777
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..777
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgggttttt tcgctttaga ggaatgggct gccgctaaca gagactacga taatacccca
gctccatatt ggcacgccaa gtctgttcct gacggtttta ctgctatctc cggtatcttg
tggtccattt cttacatcct aatggccaaa aaggccttca aggatagatc ttacgctatg
ccattacact gtttgtgttt gaacattact tgggaagctg tttacggttt tatctatggc
ccaggtttat tgaaccaagt cgttttcgct caatggatga tcgttgatgt cattctattc
tacgctattg ttagatcagc tccatcagcc tggaagcaat ccccttggt cgcccaacat
ttagccggta tcattgttgt cggttgtgtc gtttgtctat ggttgcattt agctatcgcc
gctactttta ttccatctat tggtcgtaga gttgtcttca tgactgcctg gcctatgcaa
gttttgatta atttgtcttc catcgcccaa ctattgtctc gtggtaacac tttgggtcac
tcctggggca tctggtggac cagaatgttg ggcaccatcg ctgccgcttg ttgctttttc
tggagagtct attactggcc agaacgtttc ggttacgctt ggactccata cggtcaattt
ctattgttag gttccattgg ctctgacgtt gtctacgctg tcgtttacat ttacgttcac
agattcgcta acccattaga tactagagtt aaaattgagg ctaagaaatc tggctaa

Sequence Number (ID): 189
Sequence Name: OKH29475.1
Length: 1032
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1032
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggaagatg tcactaaaca agagcgtggt gatgcagaca cgattgagag cgcagttacg
ttcctgttga ccgctcgcaa ttctgagggt tggtggatcg acttccagct tgcggctggt
ctgagcgacg aatgggtgac gggctatgtt ggtgccatgc tggcgaacat tcgtgatatc
cgcatcccgg aagccgttag caccgcatgg aatttactga actctcgtcg ccatcgtgcc
aacggcaagt ggggttacaa tcgtttgccg ccgggcgacg cggacagcac cggttgggtg
ctgcagctgg cgcacgctat tggtgaaagc aacagcgagc gtgcgcgcca agcaatgcag
agcttggcag cgcaccaacg tccggacggt ggcatttgca cctatgaatc cgaggaatcc
attcgtgcgt tcattcacgc gtcccctgag atcggttttg ctggttggtg tggtagccac
acgtgcgtca gcgcagccat cgcggcactg ccggagtacc gcttccaact gcaagattat
ctgcgtagca cccagcaaaa cgatggtagc tggctggcgt attggtggca ggacccggaa
tacgttaccg ccctggcagc cgaggccatt gcggcgtgct acccgaactc agactgcatc
acttctgcag tggtctgggg catgaatcgt ctgaatagcc agggtttcgt tgcgacgagc
gatagaccat ccggttcgcc gtttgcgacc gcgtggtgtc tccgcctgct gatcctgcgt
cgccaggata ccccggtcca ggctgcaatc gccaaagcga ccgactggct gctggcgcag
cagcaaccga acggcagctg gatcagcagc gcgcgtctgc aagtaccgtt gccggatgat
ctgaacccaa ataaatttaa tcagtggatt taccacggca cgattcaggg cagcctggtt
tttgacaagc attgtgtgtt caccaccgcg acggtgctgc aagcactgca tcgtagcctg
tttggcaagt aa

Sequence Number (ID): 190
Sequence Name: OKH29475.1

Length: 1032
Molecule Type: DNA
Features Location/Qualifiers:

- source, 1..1032
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized

Residues:
atggaagatg tcaccaagca agaaagaggt gatgctgata ccattgaatc agctgttacc
tttctattga ctgccagaaa ttccgaaggt tggtggatcg acttccaatt ggccgctggt
ttgtctgatg aatgggtcac tggttatgtt ggtgccatgt tggctaacat tcgtgacatc
agaattccag aagccgtctc cactgcctgg aacttattga actctcgtag acatagagct
aacggtaagt ggggctacaa ccgtttgcct ccaggtgacg ctgattctac cggttgggtc
ttgcaattgg cccacgctat cggtgaatca aactctgaaa gagctagaca agccatgcaa
tctttagccg ctcaccaaag acctgacggc ggtatttgta cttacgaatc tgaggaatca
attagagctt ttatccatgc ttctccagaa attggtttcg ctggttggtg tggttctcat
acttgtgttt ccgccgctat tgctgccttg ccagaatata gatttcaatt acaagattac
ttgagatcca ctcaacaaaa cgacggttca tggctagctt actggtggca agacccagaa
tatgttaccg ccttggccgc tgaagctatc gccgcttgtt atccaaactc tgactgtatc
acttccgctg tcgtttgggg tatgaacaga ttgaactctc aaggtttcgt tgccacttcc
gatagacctt ccggttctcc ttttgctact gcttggtgtt tgagattact aattttacgt
agacaagata ctccagttca agccgctatc gctaaggcta ctgactggtt attggctcaa
caacaaccaa acggttcctg gatctcctct gctagattgc aagttccatt gccagacgat
ttgaacccaa acaagttcaa ccaatggatt taccatggta caattcaagg ttctttggtt
ttcgataagc actgtgtctt caccactgct actgtcttgc aagctctaca ccgttctttg
tttggcaagt aa

Sequence Number (ID): 191
Sequence Name: NEQ07043.1
Length: 1077
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1077
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized

Residues:
atgaaagttg ttccagaaca gaccgctaag tccgcagtta agtccgccat cgaccgtgcg
attgtttttc tgctgctgag cagagatacc cagggttggt ggaaagattt cttcctgccg
gcgggtgcat ctgatgcgtg ggtcacgggt tatgtgggca ccgttcttgc ccacagccaa
aatagccatg cgtggaaagc tgccgaaaag gcatggactc tgttagcgca gcagtgtcac
gaccgtgagg gttggggcta ccacgcgggt gtcccggctg acgcggacag caccctgtgg
ggcctgcagc tggcacaggc cttgggtcgt gagggtgagg aatctagcca tcgtggccac
cgtttcctgc gccgtcatct gaaaccggat ggcggcgtga ccacgtacga acaagaagca
acgattcgca attatattgg cctgccgcct ggtctggtgc cgttcaccgc gtggtgccat
agccacacct gtgttaccgc agcggcagcg tcgctgggtg agtggcgcga gatcgttgcc
ccgtacctgt tgagccagca gcaagcggac ggctcctggc acagctattg gtggttcgaa
gatgagtact gcaccgcgct ggcgttgact gccgtggaaa gccaagagag catcgagcgt
gcggtcaagt ggggttgtca tcgcctgctg tattggctgg aagcgagcca gccgtcagag
tttgcaattg cttggtgctt gcagattctg agccgcgact ccacgccgag cacgcaacaa
ctggtggagc gcggtgtgaa gtttttgctg caacgccaac acagcaatgg tagctggcag
ccgagcgccc gtctgcgtgt accgcgtccg gataacttca acccgaaaag cgttaaagac
tggcagctgt ggacgggcaa gtttagcggt tctgtgaccc tgaaaaacgt cctggcgaac
acgttcaata tctacagcct ggaccgtcaa agcatcttta cgaccgctac cgtgctgtac
gcgctgcaat cggtcaccgc aaacgcacag ctccgtcagc aagaggtcgg tatgtaa

Sequence Number (ID): 192
Sequence Name: NEQ07043.1
Length: 1077
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1077
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized

Residues:
atgaaggttg tcccagaaca aacagctaag tctgctgtca aatctgccat tgaccgtgcc

attgttttct tgttattgtc cagagacacc caaggttggt ggaaagattt ctttttgcca
gctggtgctt cagacgcttg ggttactggt tacgttggta ccgttttggc tcactctcaa
aactctcacg cctggaaggc tgccgaaaag gcttggactt tattggctca acaatgtcac
gatagagaag gttggggtta ccatgctggt gttccagctg acgctgactc taccttgtgg
ggtttgcaac tagctcaagc cttaggtaga gaaggtgagg aatcctctca tagaggtcac
agatttttgc gtagacactt gaaacctgac ggcggtgtta ccacttacga acaagaagcc
actattcgta actacattgg tttacctcca ggtttggttc cattcactgc ttggtgtcat
tctcacactt gtgttactgc tgccgctgcc tctttgggtg aatggagaga aatcgttgcc
ccatacttat tgtcccaaca acaagccgac ggttcatggc actcttattg gtggttcgaa
gatgaatatt gtacagcttt ggccttgact gctgttgagt ctcaagaatc catcgaaaga
gctgtcaagt ggggctgtca tagattgtta tactggttgg aagcttctca accatccgaa
tttgccattg cttggtgttt gcaaatccta tccagagact ctaccccatc tacccaacaa
ctagtcgaaa gaggtgttaa gttcttattg caaagacaac attctaatgg ttcttggcaa
ccttctgcta gattacgtgt tccaagacca gataacttca acccaaaatc tgttaaagat
tggcaattgt ggaccggcaa gttctctggt tctgtcacct tgaagaacgt tttagctaac
actttcaaca tttactcctt agaccgtcaa tctattttta ccacagccac tgtcttgtac
gctttgcaat cagtcactgc caacgcccaa ttacgtcaac aagaggtcgg tatgtaa

Sequence Number (ID): 193
Sequence Name: AAcSHC_M132R_A224V_I432T
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggcagaac aattggttga ggcaccagca tacgcccgca ctttagatcg cgcagtcgag
tatctgctgt cctgtcaaaa agatgaaggc tattggtggg gtccgctgct gagcaacgtt
acgatggaag ccgaatacgt cctgctgtgt catattctgg accgtgtaga tcgcgatcgt
atggaaaaga tccgtcgtta cttgctgcat gagcagcgtg aggacggtac ttgggcgttg
tacccgggtg gcccgccgga cctggacacg accatcgaag cctatgtggc tttgaagtac
attggcatga gccgcgacga agaaccgatg cagaaagcac tgcgtttcat tcagtctcag
ggtggcatcg agagctcccg tgtgttcacg cgtcgctggc tggcgcttgt tggtgagtac
ccgtgggaaa aagtcccgat ggtgccgccg gaaatcatgt ttctgggtaa acgtatgccg
ctgaacattt atgaatttgg ctcttgggcg cgcgcgaccg tggtcgcgct gagcatcgtt
atgagccgtc aaccggtttt ccctctgccg gagcgcgcac gcgttccaga gctgtacgag
actgacgtcc cgccgcgccg tcgtggtgcc aagggtggcg gcggctggat ttttgatgct
ctggaccgcg tgctccacgg ttaccaaaag ctgtcggtcc acccgttccg tcgtgcggca
gaaattcgtg ctctggactg gctcctggag cgtcaggctg gcgacggtag ctggggtggc
attcagccgc cgtggttcta cgcgctgatc gcactgaaga tcctggatat gacgcaacat
ccggcgttca ttaagggttg ggaaggcctg gaactgtacg gtgttgagct ggattatggt
ggctggatgt tccaggcgag catttccccg gtttgggaca ccggcctggc cgtgttggcg
ctgcgtgcgg ctggtctgcc ggccgatcac gatcgtctgg ttaaggccgg tgagtggtta
ctggatcgcc agattaccgt gccaggcgac tgggctgtta aaagaccgaa tctgaaaccg
ggtggcttcg cctttcaatt cgacaatgtg tactatcctg atgtggatga tacggccgtc
gtcgtttggg cgctgaacac cctgcgtttg ccggacgagc gtcgtcgtcg tgatgcgatg
accaaaggtt tccgctggat cgttggtatg cagtccagca atggcggctg gggtgcgtat
gatgtggaca ataccagcga tctgccgaac cacacccctt tttgcgactt tggtgaagtc
accgacccgc cgagcgaaga tgtgaccgcg cacgtgctgg agtgctttgg tagcttcggt
tatgacgacg cgtggaaagt catccgccgc gcggtggagt atctgaagcg tgagcaaaaa
ccggatggtt cttggtttgg ccgctggggt gtgaactatt tgtacggtac gggtgcggtt
gttagcgcgt tgaaagccgt cggtatcgac acgcgtgagc cgtacatcca aaaggcactg
gactgggttg agcagcatca gaacccggac ggcggttggg gtgaggattg ccgtagctac
gaagatcctg cgtacgcggg caagggtgcg agcacgccga gccaaacggc gtgggccttg
atggcactga ttgcgggtgg ccgtgcagaa agcgaagctg cgcgtcgtgg tgtccagtat
ctggtcgaaa cccagcgtcc ggatggtggt tgggacgagc cgtattacac cggcaccggc
ttcccgggtg acttttacct gggttacacc atgtaccgcc acgtgtttcc gaccctggca
ctgggtcgtt acaaacaagc gatcgagcgt cgttaa

Sequence Number (ID): 194

Sequence Name: A0A1H2R2P0_9BACL
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaccaaac agctggcgga aattccggca tatatgcaga ccctggacaa cggcgttgag
tatctcctgt ctcgtcaaca cgaggaaggc tattggtggg gcccactcct gagcaacgtg
accatggaag cagaatacgt tctcctgtgc cactgcctgg gcaaagttga caaaggtcgc
ctggaaaaaa tcaaaaccta cctcctgcac gaacaacgcg aggacggtac ctgggcgcaa
tacccgggtg gcccgcagga cctcgacacc acgatcgagg cttacgttgc actgaaatat
atcggtctgt ccccggatga cgagcgcatg cagaaagcgc tggcgtttat ccagagccag
ggcggtatcg aaagcgcacg cgtctttacc cgcctctggc tggcagtcgt aggcgaatac
ccgtggcgta aactgccggt ggttccaccg gaaattatgt tcctgggcaa aaacatgccg
ctgaacatct acgattttgg ctcttgggcc cgtccgacta ttgttgcgct gaccatcgtt
atgtctcgtc gcgcagtttt cccgctgccg gcacatgcga aagtaccgga actgttcgaa
accaatgtgc ctccgcgccg tcgcgccgcg aaaggcggta acagctccct gttcctgagc
atcgataaac tcctgcaggg ttaccagaac ggttcttttc atccgttccg caaagcagcc
gaacagcgtg cgattgagtg gctgatcgaa caccaggcgg gcgacggcag ctggggtggc
attcaaccgc cttggttcta cgcgctcctg gccctgaaag ttatgaacat gacgaatcac
ccggcgttta tcaaaggctg ggaaggcctg gaactgtacg gcctggagct cgaatatggt
ggctggatgt tccaggcatc tatctctcct gtttgggata ccggtctgag catcctggca
ctccgcgctg cgggcctcgc tccggacgag ccggcgctgg ttaaagctgg taaatggctc
ctggaccacc gtattgcaac caaaggcgat tgggctgtcc gtcgcccaaa cgcaaaaccg
ggcggttggg cgtttcagtt cgataacccg cactatccgg atgtggacga taccgcggtg
gttgtctggg cgctgaatgg cctgaaactg cctaacgaag ccgaacgccg tgatgcgatg
actgcaggct tccgctggct gaccgctatg cagagctcta acggcggttg gggcgcgtac
gacgtagata ataacaaaga actgccgaat cgcatcccat tctgcgattt cggcgaagtt
atcgaccctc catctgagga cgtgactgct cacgttctgg agtgctttgg ttctttcggt
tatgacgaag cctggaaagt tgtggcccgt gcggttaact acctgaaacg cgagcagaaa
ccggatggct cttggtatgg tcgctggggt gtaaactaca tttacggtat cggtgctgta
gtcccggcgc tgaaatccgt tggcgtggac atgaaagaac cgttcgttca gaaagcactg
gattggctgg tagcgcacca gaacgaagac ggcggttggg gcgaagactg ccgcagctac
gtagacgaac gtttcgcagg cgtcggccca tctaccccat ctcagactgc atgggcactg
atggcgctga tcgcaggcgg tcgtgttcaa gcggatgctg tgtctcgcgg cgtggcatac
ctggtgcgca cccagcgtag cgatggcggt tgggatgagc cgtattacac tggcaccggc
ttcccgggcg acttctatct gggctatact ctgtaccgtc acatcttccc agttatggct
ctgggtcgct acaaagacgc tctgggtcgc ctcacccgtt aa

Sequence Number (ID): 195
Sequence Name: A0A1H2R2P0_9BACL
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaccaagc aattggctga aattccagct tacatgcaaa ccttggataa tggtgttgaa
tacttgttat ctagacaaca cgaggaaggt tactggtggg gtccattatt gtctaacgtc
actatggaag ccgaatacgt cctattgtgt cattgtttag gtaaagttga taaaggtaga
ttggaaaaaa tcaagactta tttattgcac gaacaacgtg aagacggtac ttgggctcaa
tacccaggcg gtccacaaga tttggacacc actattgaag cttatgttgc tttgaagtac
attggtttat ctcctgatga cgaaagaatg caaaaggcct tggcctttat tcaatcccaa
ggcggtattg aatctgctag agttttcact agattgtggt tagctgtcgt tggtgaatac
ccatggagaa agttgccagt cgttcctcca gaaatcatgt tcttaggtaa aaacatgcca
ttgaatattt acgatttcgg ttcttgggct agaccaacta tcgtcgcctt gaccattgtt
atgtctcgta gagctgtctt cccattgcca gcccacgcta aagttccaga attgttcgaa
accaacgttc cacctagacg tagagccgct aagggcggta actcctcttt attcttgtct

attgacaaat tattgcaagg ttaccaaaac ggttcattcc acccattcag aaaggctgcc
gaacaaagag ctatcgaatg gttgatcgaa caccaagctg gtgatggctc ctggggcggt
attcaacctc catggtttta cgccttgtta gccttgaagg tcatgaacat gactaaccac
ccagctttca tcaaaggttg ggaaggtttg gaattgtatg gtttggaatt ggaatacggc
ggttggatgt ttcaagcttc tatctctcca gtttgggata ctggtttgtc catcttggcc
ttgagagccg ctggtttggc tccagacgaa ccagctttgg ttaaagctgg taagtggtta
ttggatcaca gaatcgctac caaaggtgac tgggctgtca gacgtccaaa cgctaagcca
ggcggttggg ctttccaatt cgacaaccca cactaccctg acgtcgatga cactgctgtc
gttgtctggg ctttgaatgg tctaaagttg ccaaacgagg ctgaacgtag agacgctatg
accgccggtt tcagatggtt gactgctatg caatcttcca acggcggttg gggtgcttac
gacgtcgata acaataaaga attgccaaac agaatcccat tctgtgattt tggtgaagtt
attgaccctc catctgaaga cgttactgct cacgtcttag aatgcttcgg ttctttcggt
tatgatgaag cttggaaggt tgtcgctaga gccgtcaact acttgaagag agagcaaaag
ccagatggtt cttggtacgg tagatggggt gtcaactata tttacggtat tggtgccgtt
gtccctgctt tgaagtctgt tggtgttgat atgaaggaac ctttcgttca aaaggctcta
gactggttgg ttgcccacca aaatgaagac ggcggttggg gtgaagattg tagatcttac
gtcgacgaaa gattcgctgg tgttggtcca tctacccctt cccaaactgc ttgggcttta
atggctttga tcgctggcgg tagagttcaa gctgatgccg tctccagagg tgttgcttac
ctagttagaa cacaaagatc cgacggcggt tgggacgagc catattacac tggtactggt
ttcccaggtg atttttactt gggttacacc ttgtacagac atattttccc tgtcatggct
ttgggtagat ataaggatgc tttgggtcgt ttgacccgtt aa

Sequence Number (ID): 196
Sequence Name: A0A1H2R2P0_9BACL_F437A G600M
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaccaaac agctggcaga aatcccggct tacatgcaga ctctggataa cggcgtggaa
tacctcctgt cccgtcagca tgaggaaggt tattggtggg gcccgctcct gtccaacgtc
accatggaag cggagtatgt tctcctgtgt cactgcctgg gtaaagtgga taaaggtcgt
ctggaaaaga tcaaaaccta cctcctgcac gaacagcgtg aagacggtac ctgggctcag
tatccgggtg gcccgcagga cctggatacg accattgagg cttacgtagc cctgaaatac
attggtctgt ctccggacga tgaacgcatg caaaaggccc tcgcgtttat ccagtcccaa
ggtggcattg aatctgcgcg cgttttcacc cgtctgtggc tggcggtagt tggcgaatac
ccgtggcgta agctgccggt tgtgcctccg gaaatcatgt tcctgggtaa aaacatgccg
ctcaatattt acgattttgg ttcttgggct cgtccaacca tcgtcgcact caccattgtg
atgtctcgtc gcgccgtttt cccgctgccg gcgcacgcta aagttccgga gctgttcgaa
accaacgtcc caccgcgtcg ccgtgcagcg aaaggcggta actcttccct gttcctgagc
atcgacaaac tcctgcaggg ttatcagaac ggctcctttc acccattccg taaagcagcg
gagcagcgtg caattgaatg gctgattgaa caccaggcgg gcgacggttc ctggggtggc
atccagcctc cgtggttcta cgcactcctg gctctgaagg ttatgaatat gactaaccac
ccggctttca tcaagggttg ggaaggcctg gaactgtacg gcctcgaact ggaatatggc
ggttggatgt tccaggcttc catctctccg gtttgggata ccggcctgtc catcctggct
ctgcgtgctg ccggtctggc accggatgaa ccggcactgg taaaagcggg taaatggctc
ctggatcatc gcatcgcaac caaaggcgac tgggctgttc gccgtccaaa cgcgaaaccg
ggcggttggg cattccagtt cgacaaccca cattacccag acgtggatga caccgcggtg
gttgtctggg cgctgaacgg tctgaagctg ccgaacgaag cagagcgtcg cgacgccatg
acggctggtt tccgttggct gacggcaatg cagtcttcca acggtggctg gggtgcgtac
gacgtagaca ataacaaaga actgccgaac cgcattccgt tctgtgacgc tggcgaagtt
atcgatccgc catctgaaga cgttactgct catgttctgg aatgttttgg ctccttcggt
tacgacgaag cgtggaaagt cgtggctcgc gcagttaact atctgaaacg tgaacagaaa
ccggacggct cttggtacgg tcgctggggc gttaattaca tttatggtat tggtgctgtt
gtaccggccc tgaaatctgt tggtgtggac atgaaagaac cgttcgttca gaaagcgctg
gactggctcg tcgcccacca gaacgaagat ggtggctggg gtgaggactg ccgcagctac
gttgacgaac gttttgcagg tgtgggcccg agcaccccga gccagacggc gtgggcgctg
atggcgctga tcgccggcgg tcgtgttcaa gctgacgcgg ttagccgcgg cgtagcatat
ctggtccgta cccagcgcag cgacggcggt tgggacgaac cgtactatac cggtactatg

tttccgggcg atttctatct cggttacact ctgtaccgtc atatcttccc ggttatggcg
ctgggccgtt acaaagatgc tctgggccgt ctgactcgct aa

Sequence Number (ID): 197
Sequence Name: A0A1H2R2P0_9BACL_F437A G600M
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgactaagc aattagctga aatcccagct tatatgcaaa ctctagataa cggtgttgaa
tacctattgt ctagacaaca cgaggaaggt tactggtggg gtcctttatt gtctaacgtc
actatggaag ctgaatacgt tctattgtgt cactgtctag gtaaggtcga taagggtaga
ttagaaaaga ttaaaaccta cttattgcac gaacaaagag aagatggtac ttgggctcaa
tacccaggcg gtcctcaaga cttagacacc actattgaag cctacgttgc tttaaaatac
atcggtttat ctccagacga tgaaagaatg caaaaagctc tagcttttat tcaatcccaa
ggcggtatcg aatcagccag agtctttact agattgtggt tggctgtcgt tggtgaatac
ccatggagaa agttgccagt cgttcctcca gaaatcatgt ttttgggtaa gaatatgcca
ttgaacattt acgacttcgg ttcctgggcc agaccaacta ttgtcgcctt gactatcgtt
atgtcccgta gagccgtctt cccattgcct gctcatgcta aggtcccaga attattcgaa
accaacgttc ctccaagacg tagagctgcc aagggcggta actcttcctt gtttctatct
attgataagt tgttacaagg ttaccaaaat ggttctttcc acccttttag aaaagccgct
gaacaacgtg ccatcgaatg gttgattgaa caccaagctg gtgacggttc ctggggcggt
atccaacctc catggtttta tgctctattg gccttgaagg ttatgaacat gactaatcac
ccagctttca ttaagggttg ggaaggtttg gaattgtacg gcttggagtt agaatacggc
ggttggatgt ttcaagcctc tatttctcca gtttgggata ccggtttatc tatcttggct
ttgagagccg ctggtttggc cccagatgaa ccagccttgg tcaaggctgg taagtggttg
ttagaccata gaatcgctac caaaggtgac tgggctgtca gacgtccaaa cgccaaacca
ggcggttggg cctttcaatt cgacaaccca cattatccag atgtcgatga cactgccgtc
gttgtctggg ctttgaacgg tttgaaacta ccaaacgaag ctgaaagacg tgatgctatg
accgctggtt ttagatggtt gactgctatg caatcctcta acggcggttg gggtgcttac
gatgttgata acaataagga attaccaaac cgtattccat tctgtgacgc cggtgaagtt
attgaccctc catccgaaga cgttacagcc cacgttttgg agtgtttcgg ttctttcggt
tacgacgaag cttggaaagt cgttgccaga gctgttaact acttaaagag agaacaaaag
ccagacggtt cttggtacgg tagatggggt gtcaactaca tttatggtat tggtgccgtc
gttccagctt tgaaatctgt cggtgtcgat atgaaggaac ctttcgtcca aaaggctttg
gattggttag ttgctcacca aaatgaagac ggcggttggg gtgaagattg tcgttcctac
gttgacgaaa gattcgctgg tgttggtcca tctactccat cacaaaccgc ttgggcccta
atggctttga tcgctggcgg tagagtccaa gccgatgccg tttctagagg tgttgcttat
ttggtcagaa ctcaaagatc tgatggcggt tgggacgaac catattacac cggtaccatg
ttcccaggtg acttctactt gggttacacc ttgtacagac atatcttccc tgtcatggct
ttgggcagat acaaagatgc tttgggtcgt ttgacccgtt aa

Sequence Number (ID): 198
Sequence Name: AAcSHC_F437A_G600M
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggcagaac aattggttga ggcaccagca tacgcccgca ctttagatcg cgcagtcgag
tatctgctgt cctgtcaaaa agatgaaggc tattggtggg gtccgctgct gagcaacgtt
acgatggaag ccgaatacgt cctgctgtgt catattctgg accgtgtaga tcgcgatcgt
atggaaaaga tccgtcgtta cttgctgcat gagcagcgtg aggacggtac ttgggcgttg
tacccgggtg gcccgccgga cctggacacg accatcgaag cctatgtggc tttgaagtac
attggcatga gccgcgacga agaaccgatg cagaaagcac tgcgtttcat tcagtctcag
ggtggcatcg agagctcccg tgtgttcacg cgtatgtggc tggcgcttgt tggtgagtac

ccgtgggaaa aagtcccgat ggtgccgccg gaaatcatgt ttctgggtaa acgtatgccg
ctgaacattt atgaatttgg ctcttgggcg cgcgcgaccg tggtcgcgct gagcatcgtt
atgagccgtc aaccggtttt ccctctgccg gagcgcgcac gcgttccaga gctgtacgag
actgacgtcc cgccgcgccg tcgtggtgcc aagggtggcg gcggctggat ttttgatgct
ctggaccgcg ctctccacgg ttaccaaaag ctgtcggtcc acccgttccg tcgtgcggca
gaaattcgtg ctctggactg gctcctggag cgtcaggctg gcgacggtag ctggggtggc
attcagccgc cgtggttcta cgcgctgatc gcactgaaga tcctggatat gacgcaacat
ccggcgttca ttaagggttg ggaaggcctg gaactgtacg gtgttgagct ggattatggt
ggctggatgt tccaggcgag catttccccg gtttgggaca ccggcctggc cgtgttggcg
ctgcgtgcgg ctggtctgcc ggccgatcac gatcgtctgg ttaaggccgg tgagtggtta
ctggatcgcc agattaccgt gccaggcgac tgggctgtta aaagaccgaa tctgaaaccg
ggtggcttcg cctttcaatt cgacaatgtg tactatcctg atgtggatga tacggccgtc
gtcgtttggg cgctgaacac cctgcgtttg ccggacgagc gtcgtcgtcg tgatgcgatg
accaaaggtt tccgctggat cgttggtatg cagtccagca atggcggctg gggtgcgtat
gatgtggaca ataccagcga tctgccgaac cacattcctt tttgcgacgc gggtgaagtc
accgacccgc cgagcgaaga tgtgaccgcg cacgtgctgg agtgctttgg tagcttcggt
tatgacgacg cgtggaaagt catccgccgc gcggtggagt atctgaagcg tgagcaaaaa
ccggatggtt cttggtttgg ccgctggggt gtgaactatt tgtacggtac gggtgcggtt
gttagcgcgt tgaaagccgt cggtatcgac acgcgtgagc cgtacatcca aaaggcactg
gactgggttg agcagcatca gaacccggac ggcggttggg gtgaggattg ccgtagctac
gaagatcctg cgtacgcggg caagggtgcg agcacgccga gccaaacggc gtgggccttg
atggcactga ttgcgggtgg ccgtgcagaa agcgaagctg cgcgtcgtgg tgtccagtat
ctggtcgaaa cccagcgtcc ggatggtggt tgggacgagc cgtattacac cggcaccatg
ttcccgggtg acttttacct gggttacacc atgtaccgcc acgtgtttcc gaccctggca
ctgggtcgtt acaaacaagc gatcgagcgt cgttaa

Sequence Number (ID): 199
Sequence Name: AAcSHC_F437A_G600M
Length: 2058
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2058
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgtctccag ccgatatctc tactaaatct tcctctttcc aaagattgga taatatgcta
cctgaagccg tttcatccgc ttgcgactgg ttgatcgatc aacaaaagcc agatggccat
tgggttggtc ctgttgaatc caacgcttgt atggaagctc aatggtgttt ggctttgtgg
tttttgggtc aagaagatca tccattacgt ccacgtttag cccaagcctt attggaaatg
caaagagaag atggttcttg gggtatctac gttggtgctg accacggtga tattaacact
accgttgagg cttatgccgc tttaagatca atgggttacg ccgctgatat gcctattatg
gctaagtctg ccgcttggat tcaacaaaag ggcggtttga gaaacgtccg tgttttcacc
agatattggt tagccttgat cggtgaatgg ccttgggata agactccaaa cctacctcca
gaaattatct ggttgcctga taatttcatt ttttctatct acaatttcgc tcaatgggcc
cgtgctacta tgatgccatt aaccatttta tccgctcgta gaccatctag accactattg
cctgaaaaca gattggacgg tttgttccca gaaggtagag aaaacttcga ctatgaacta
cctgtcaagg gtgaggaaga cctatggggt agattcttta gagccgctga taaaggtttg
cattctttgc aatcttttcc agttagacgt tttgtcccaa gagaggccgc tatcagacac
gtcattgaat ggatcattag acaccaagat gctgatggcg gttggggcgg tattcaacct
ccatggattt acggtttgat ggccctatcc gtcgaaggtt acccattgca tcacccagtt
ttggctaaag ccatggacgc tttgaacgac ccaggttggc gtagagacaa gggtgacgct
tcttggatcc aagctaccaa ctctcctgtt tgggacacca tgttggctgt cttggccttg
cacgatgccg gtgctgaaga tagatattct ccacaaatgg acaaggctat tggttggttg
ttagatagac aagttagagt taagggtgat tggtctatta agttgccaga tactgaacca
ggcggttggg ccttcgaata cgctaacgat aaatacccag acaccgatga cacagccgtt
gctttgatcg ccttggccgg ttgtagacac cgtccagagt ggagagaaag agacattgaa
ggtgctattt ccagaggcgt taactggttg ctagccatgc aatcctcttc aggcggttgg
ggtgccttcg ataaggacaa caatagatcc attttaacca aaattccatt ttgtgacgct
ggtgaagcct tggaccctcc atccgttgac gttactgccc atgtcttgga ggctttcggt
ttactaggca tctctagaaa ccacccatct gttcaaaaag ccttagccta tattagatct
gaacaagaaa gaaacggtgc ttggtttggt agatggggtg ttaactacgt ttatggtaca

ggtgccgttt tgccagcttt ggctgccatt ggtgaagaca tgacccaacc atacattgtt
agagcttgtg actggttaat gtctgttcaa caagaaaacg gcggttgggg tgaatcctgc
gcctcatata tggatatcaa cgctgttggt cacggtgttg ctaccgcctc tcaaactgct
tgggctttga tcggtctatt ggctgccaaa agaccaaagg atagagaagc tatcgccaga
ggttgtcaat tcttgattga aagacaagaa gacggttctt ggaccgaaga ggaatacact
ggtaccatgt tcccaggtta tggtgttggt caagccatta aattggacga tccatccttg
ccagatagat tgctacaagg cgctgaatta tcaagagctt tcatgttgcg ttatgacttg
tatagacaat acttccctgt tatggcttta tctagagccc gtagaatgat gaaagaagac
gcctctgccg ccgcttaa

Sequence Number (ID): 200
Sequence Name: AAcSHC_wt
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atggcagaac aattggttga ggcaccagca tacgcccgca ctttagatcg cgcagtcgag
tatctgctgt cctgtcaaaa agatgaaggc tattggtggg gtccgctgct gagcaacgtt
acgatggaag ccgaatacgt cctgctgtgt catattctgg accgtgtaga tcgcgatcgt
atggaaaaga tccgtcgtta cttgctgcat gagcagcgtg aggacggtac ttgggcgttg
tacccgggtg gcccgccgga cctggacacg accatcgaag cctatgtggc tttgaagtac
attggcatga gccgcgacga agaaccgatg cagaaagcac tgcgtttcat tcagtctcag
ggtggcatcg agagctcccg tgtgttcacg cgtatgtggc tggcgcttgt tggtgagtac
ccgtgggaaa aagtcccgat ggtgccgccg gaaatcatgt ttctgggtaa acgtatgccg
ctgaacattt atgaatttgg ctcttgggcg cgcgcgaccg tggtcgcgct gagcatcgtt
atgagccgtc aaccggtttt ccctctgccg gagcgcgcac gcgttccaga gctgtacgag
actgacgtcc cgccgcgccg tcgtggtgcc aagggtggcg gcggctggat ttttgatgct
ctggaccgcg ctctccacgg ttaccaaaag ctgtcggtcc acccgttccg tcgtgcggca
gaaattcgtg ctctggactg gctcctggag cgtcaggctg gcgacggtag ctggggtggc
attcagccgc cgtggttcta cgcgctgatc gcactgaaga tcctggatat gacgcaacat
ccggcgttca ttaagggttg ggaaggcctg gaactgtacg gtgttgagct ggattatggt
ggctggatgt tccaggcgag catttccccg gtttgggaca ccggcctggc cgtgttggcg
ctgcgtgcgg ctggtctgcc ggccgatcac gatcgtctgg ttaaggccgg tgagtggtta
ctggatcgcc agattaccgt gccaggcgac tgggctgtta aaagaccgaa tctgaaaccg
ggtggcttcg cctttcaatt cgacaatgtg tactatcctg atgtggatga tacggccgtc
gtcgtttggg cgctgaacac cctgcgtttg ccggacgagc gtcgtcgtcg tgatgcgatg
accaaaggtt tccgctggat cgttggtatg cagtccagca atggcggctg gggtgcgtat
gatgtggaca ataccagcga tctgccgaac cacattcctt tttgcgactt tggtgaagtc
accgacccgc cgagcgaaga tgtgaccgcg cacgtgctgg agtgctttgg tagcttcggt
tatgacgacg cgtggaaagt catccgccgc gcggtggagt atctgaagcg tgagcaaaaa
ccggatggtt cttggtttgg ccgctggggt gtgaactatt tgtacggtac gggtgcggtt
gttagcgcgt tgaaagccgt cggtatcgac acgcgtgagc cgtacatcca aaaggcactg
gactgggttg agcagcatca gaacccggac ggcggttggg gtgaggattg ccgtagctac
gaagatcctg cgtacgcggg caagggtgcg agcacgccga gccaaacggc gtgggccttg
atggcactga ttgcgggtgg ccgtgcagaa agcgaagctg cgcgtcgtgg tgtccagtat
ctggtcgaaa cccagcgtcc ggatggtggt tgggacgagc cgtattacac cggcaccggc
ttcccgggtg acttttacct gggttacacc atgtaccgcc acgtgtttcc gaccctggca
ctgggtcgtt acaaacaagc gatcgagcgt cgttaa

Sequence Number (ID): 201
Sequence Name: AAcSHC_wt
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:

atggctgaac aattggttga agctccagct tacgctcgta ccttggaccg tgctgtcgag
tacttattgt cctgtcaaaa ggatgaaggt tattggtggg gtccattatt gtccaatgtc
actatggagg ctgaatacgt cttgttatgt catattttgg acagagttga ccgtgacaga
atggaaaaga ttcgtagata cttgctacat gaacaacgtg aagatggtac ttgggctcta
tacccaggcg gtcctccaga cttggatact accatcgaag cctacgtcgc cttgaagtat
atcggtatgt ccagagatga agagccaatg caaaaagctt tgcgttttat tcaatctcaa
ggcggtatcg agtcctctag agttttcacc cgtatgtggt tggccttagt cggcgaatac
ccatgggaaa aggtcccaat ggtccctcca gagattatgt tcctaggtaa gagaatgcca
ttgaacatct acgaattcgg ttcttgggcc agagctactg tcgttgcctt gtctatcgtc
atgtctagac aaccagtctt cccattgcca gaacgtgcta gagtcccaga gttgtacgaa
accgacgttc ctccacgtag acgtggtgct aagggcggtg gcggttggat ctttgatgcc
ttggatagag ccttgcatgg ttatcaaaag ttgtcagttc acccattccg tagagccgct
gaaattcgtg ctttggattg gttattggaa cgtcaagctg gtgacggttc ttggggcggt
attcaacctc catggttcta cgccttgatt gctttgaaga ttttggacat gactcaacat
ccagctttca tcaagggttg ggagggtcta gaactatacg gtgttgaatt ggactacggc
ggttggatgt tccaagcttc catttcccca gtctgggaca ctggtttggc tgttttggcc
ttgcgtgctg ccggtttgcc agctgaccac gatcgtttgg ttaaagccgg tgaatggtta
ttggatagac aaattaccgt ccctggtgat tgggccgtta aaagaccaaa tttgaagcca
ggcggttttg ctttccaatt tgataacgtc tattacccag acgttgatga caccgctgtt
gtcgtttggg ctttaaatac tttaagattg ccagatgaac gtagacgtag agatgctatg
actaagggtt tcagatggat cgtcggtatg caatcctcta acggcggttg gggtgcctac
gatgttgaca acacatctga cttaccaaat cacatcccag cttgtgactt tggtgaagtt
acagatccac cttccgaaga cgttacagcc catgttttag aatgtttcgg ttcttttggt
tacgatgacg cctggaaagt cattcgtaga gctgtcgaat acttaaaaag agaacaaaag
ccagacggtt cttggttcgg tagatggggt gtcaactact tgtatggtac cggtgctgtt
gtctctgctt tgaaggctgt tggtatcgat accagagaac catacatcca aaaggcttta
gactgggtcg aacaacatca aaacccagac ggcggttggg gtgaagactg tagatcttac
gaagatccag cctacgccgg taaaggtgct tctaccccat ctcaaactgc ctgggctttg
atggccttga ttgctggcgg tagagctgaa tctgaagccg ctcgtagagg cgttcaatac
ttggttgaaa cccaaagacc tgacggcggt tgggatgaac catactatac tggtaccatg
ttccctggtg acttctacct aggttatact atgtaccgtc acgttttccc aaccttagct
ttaggtcgtt acaagcaagc tattgaaaga agataa

Sequence Number (ID): 202
Sequence Name: Gmo_SHC_F460A_G623M
Length: 2058
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2058
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgagcccag cggacattag caccaaatcg tcatcttttc agcgtctgga caatatgttg
ccggaagccg tcagctctgc gtgcgattgg ttgatcgacc agcagaaacc ggatggtcac
tgggtgggcc cagtggagtc caacgcatgt atggaagcgc agtggtgctt ggcactgtgg
tttctgggcc aagaggacca cccgctgcgt ccgcgcctgg cgcaagcatt gctggaaatg
cagcgcgaag atggcagctg gggtatttac gtgggcgcgg atcatggtga tatcaacacg
accgtcgagg cgtatgcagc gctgcgcagc atgggttacg cagcggatat gccgatcatg
gctaagtccg cagcgtggat tcaacaaaag ggtggcctgc gcaatgtccg tgtttttacg
cgctactggc ttgccctgat tggtgagtgg ccgtgggata agaccccgaa tctgccgcca
gagatcattt ggctgccgga caatttcatc ttcagcattt acaactttgc gcaatgggct
cgtgcgacga tgatgccgct caccatcctg tctgcacgcc gtccgagccg tccgctgctg
ccggagaacc gtctggacgg tctgttccct gagggtagag agaacttcga ttatgaattg
ccggtgaagg gcgaagaaga tctgtggggt cgtttcttcc gtgcagcaga caagggtctg
catagcctgc agagcttccc ggttcgccgt ttcgtcccgc gtgaggctgc gatccgccac
gtcatcgagt ggattattcg tcaccaagat gcggatggtg gttggggtgg cattcaaccg
ccgtggattt atggcctgat ggccctgtcc gttgagggtt acccgttgca ccacccggtt
ctggcgaaag cgatggacgc tctgaacgac ccgggttggc gtcgcgataa gggcgatgct
agctggattc aagcgaccaa tagcccggtg tgggacacca tgctggccgt gctggcactg
cacgacgcgg gtgctgagga tcgttatagc ccgcagatgg acaaggctat tggttggctg
ctggaccgtc aggtgcgtgt taagggtgac tggagcatta agctgccgga taccgagccg

ggtggctggg catttgaata tgcaaacgac aaataccctg acaccgacga cactgcggtc
gccttgatcg cactggcggg ctgccgtcat cgtccggagt ggcgcgaacg tgatatcgag
ggcgcgatca gccgtggtgt caactggctg ctggcgatgc agtccagcag cggtggctgg
ggcgcattcg ataaagataa caatcgttcg attctgacta aaatcccgtt ttgtgacgcg
ggtgaagccc tggacccgcc gagcgttgac gtgaccgccc atgtcctgga agcattcggt
ttactgggta tcagccgtaa ccatccgagc gtgcagaaag cactggcgta tatccgttct
gagcaagagc gcaatggtgc ctggttcggt cgctggggtg tcaattacgt atatggcacg
ggtgccgttc tcccggcgtt ggcagcgatt ggcgaagata tgacccagcc gtacattgtg
cgtgcctgcg actggctgat gagcgttcaa caagagaatg gcggctgggg tgagagctgt
gcgagctaca tggatatcaa cgccgtgggt cacggtgttg cgacggcgag ccagaccgcc
tgggcgctga ttggtctgct ggcggcgaag cgcccaaaag accgtgaagc gatcgctcgc
ggctgtcaat ttctgatcga gcgccaagaa gatggtagct ggaccgaaga agagtatacg
ggcacgatgt tcccgggcta cggcgttggt caggcaatca aactggatga cccgtccctg
cctgaccgcc tgctgcaggg tgccgagtta agccgtgctt tcatgctgcg ctatgacctg
taccgccagt actttcctgt tatggcgttg agccgtgcac gtcgtatgat gaaagaagat
gcgagcgctg cggcctaa

Sequence Number (ID): 203
Sequence Name: Gmo_SHC_F460A_G623M
Length: 2058
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2058
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgtctccag ctgacatttc taccaaatcc tcttcctttc aaagattgga caacatgttg
ccagaggctg tctcatctgc ttgtgattgg ttgatcgacc aacaaaagcc agacggtcat
tgggttggtc cattggaatc taacgcttgt atggaagctg agtggtgttt ggctttgtgg
tttttgggtc aagaagatca ccctttaaga cctagattgg ctcaagcctt gttagaaatg
caaagagaag acggttcctg gggtatttac gttggcgctg accacggtga cattaacact
acagtcgaag cctacgccgc tttgagatca atgggttacg ccgctgacat gccaattatg
gctaaatccg ccgcttggat ccaacaaaag ggcggtttga gaaacgtccg tgtttttact
agatactggt tggccttgat tggtgaatgg ccatgggaca agaccccaaa cttgcctcca
gaaatcattt ggttgcctga caacttcatt ttctctatct acaacttcgc tcaatgggct
agagctacta ttatgccatt gactatcttg tccgctagac gtccatccag accactattg
ccagaaaacc gtctagatgg tttattccca gaaggtcgtg aaaatttcga ctacgagtta
ccagttaagg gtgaggaaga cttgtggggt agatttttca gagccgctga taagggtttg
cactcattac aatctttccc agtccgtaga ttcgttccaa gagaagccgc tattagacac
gtcatcgaat ggatcattag acaccaagat gccgatggcg gttggggcgg tattcaacct
ccatggattt acggtttaat ggccttgtct gtcgaaggtt acccattgca tcacccagtt
ttggctaagg ctatggacgc tttgaacgac ccaggttggc gtagagataa gggtgacgcc
tcttggatcc aggcttccaa ctctccagtc tgggatacta tgttggccgt tttggctttg
cacgacgctg gtgctgaaga tagatactcc ccacaaatgg acaaggccat cggttggcta
ttggacagac aagttagagt caagggtgat tggtctatta agctaccaga tactgaacca
ggcggttggg ccttcgaata tgctaacgat aagtatccag atactgacga tactgctgtc
gctttgatcg ccttggctgg ttgtcgtcat agaccagaat ggagagaaag agatattgaa
ggtgccattt cccgtggtgt caactggtta ttggccatgc aatcttcctc tggcggttgg
ggtgccttcg ataaggataa caatcgttcc attttgacca aaatcccatt ttgcgacttt
ggtgaggctt tagatccacc ttctgttgat gttactgctc acgttttaga agccttcggt
ttattgggta tttctagaaa ccacccatct gtccaaaagg ctttagctta cattagatct
gaacaagagc gtaacggtgc ttggttcggt agatggggtg tcaactatgt ctacggcact
ggtgctgtct tgcctgcttt ggctgccatt ggcgaagata tgacccaacc atacattgtt
cgtgcctgcg attggttgat gtctgttcaa caagaaaacg gcggttgggg tgaatcctgt
gcttcctaca tggacatcaa cgctgttggt cacggtgttg ctactgcctc tcaaaccgct
tgggctttga ttggtttatt ggccgctaaa agaccaaagg atagagaagc tatcgctaga
ggttgtcaat ttctaattga aagacaagag gatggttctt ggacagaaga ggaatacact
ggtaccggtt acccaggtta cggtgttggt caagctatca agttggatga cccatccttg
ccagacagat tgttacaagg tgctgaattg tccagagcct tcatgttgag atatgactta
tacagacaat acttcccagt tatggccttg tctagagctc gtagaatgat gaaggaagac
gcttccgctg ctgcttaa

Sequence Number (ID): 204
Sequence Name: Gmo_SHC_Q54E_M184I_V45L_T326S_F624Y
Length: 2058
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2058
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgagcccag cggacattag caccaaatcg tcatcttttc agcgtctgga caatatgttg
ccggaagccg tcagctctgc gtgcgattgg ttgatcgacc agcagaaacc ggatggtcac
tgggtgggcc cactggagtc caacgcatgt atggaagcgg agtggtgctt ggcactgtgg
tttctgggcc aagaggacca cccgctgcgt ccgcgcctgg cgcaagcatt gctggaaatg
cagcgcgaag atggcagctg gggtatttac gtgggcgcgg atcatggtga tatcaacacg
accgtcgagg cgtatgcagc gctgcgcagc atgggttacg cagcggatat gccgatcatg
gctaagtccg cagcgtggat tcaacaaaag ggtggcctgc gcaatgtccg tgtttttacg
cgctactggc ttgccctgat tggtgagtgg ccgtgggata agaccccgaa tctgccgcca
gagatcattt ggctgccgga caatttcatc ttcagcattt acaactttgc gcaatgggct
cgtgcgacga tcatgccgct caccatcctg tctgcacgcc gtccgagccg tccgctgctg
ccggagaacc gtctggacgg tctgttccct gagggtagag agaacttcga ttatgaattg
ccggtgaagg gcgaagaaga tctgtggggt cgtttcttcc gtgcagcaga caagggtctg
catagcctgc agagcttccc ggttcgccgt ttcgtcccgc gtgaggctgc gatccgccac
780
gtcatcgagt ggattattcg tcaccaagat gcggatggtg gttggggtgg cattcaaccg
ccgtggattt atggcctgat ggccctgtcc gttgagggtt acccgttgca ccacccggtt
ctggcgaaag cgatggacgc tctgaacgac ccgggttggc gtcgcgataa gggcgatgct
agctggattc aagcgagcaa tagcccggtg tgggacacca tgctggccgt gctggcactg
cacgacgcgg gtgctgagga tcgttatagc ccgcagatgg acaaggctat tggttggctg
ctggaccgtc aggtgcgtgt taagggtgac tggagcatta agctgccgga taccgagccg
ggtggctggg catttgaata tgcaaacgac aaatacccctg acaccgacga cactgcggtc
gccttgatcg cactggcggg ctgccgtcat cgtccggagt ggcgcgaacg tgatatcgag
ggcgcgatca gccgtggtgt caactggctg ctggcgatgc agtccagcag cggtggctgg
ggcgcattcg ataaagataa caatcgttcg attctgacta aaatcccgtt ttgtgacttt
ggtgaagccc tggacccgcc gagcgttgac gtgaccgccc atgtcctgga agcattcggt
ttactgggta tcagccgtaa ccatccgagc gtgcagaaag cactggcgta tatccgttct
gagcaagagc gcaatggtgc ctggttcggt cgctggggtg tcaattacgt atatggcacg
ggtgccgttc tcccggcgtt ggcagcgatt ggcgaagata tgacccagcc gtacattgtg
cgtgcctgcg actggctgat gagcgttcaa caagagaatg gcggctgggg tgagagctgt
gcgagctaca tggatatcaa cgccgtgggt cacggtgttg cgacggcgag ccagaccgcc
tgggcgctga ttggtctgct ggcggcgaag cgcccaaaag accgtgaagc gatcgctcgc
ggctgtcaat ttctgatcga gcgccaagaa gatggtagct ggaccgaaga agagtatacg
ggcacgggtt acccgggcta cggcgttggt caggcaatca aactggatga cccgtccctg
cctgaccgcc tgctgcaggg tgccgagtta agccgtgctt tcatgctgcg ctatgacctg
taccgccagt actttcctgt tatggcgttg agccgtgcac gtcgtatgat gaaagaagat
gcgagcgctg cggcctaa

Sequence Number (ID): 205
Sequence Name: Gmo_SHC_Q54E_M184I_V45L_T326S_F624Y
Length: 2058
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2058
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgtcaccag ctgatatctc aacaaagtct tcctctttcc aaagattgga caacatgttg
ccagaagctg tttcttccgc ttgtgactgg ttgattgacc aacaaaagcc agacggtcac
tgggtcggtc cagtcgaatc taacgcttgt atggaagccc aatggtgttt ggccttgtgg
ttcttaggtc aagaagacca tccattgcgt ccaagattgg ctcaagctct attggaaatg
caaagagaag atggttcttg gggtatctac gttggtgctg accacggtga cattaacact

accgttgagg cttatgccgc tttgagatct atgggttacg ccgctgacat gccaattatg
gccaagtcag ctgcctggat tcaacaaaag ggcggtttgc gtaacgtcag agtctttact
agatactggt tggctttgat cggtgaatgg ccttgggata aaaccccaaa cttacctcca
gaaattatct ggttaccaga caatttcatc ttctctattt ataacttcgc tcaatgggcc
agagccacta tgatgccttt aaccatcttg tccgccagac gtccatctag acctttattg
ccagaaaata gattggacgg tttattccca gagggtagag aaaactttga ctacgagtta
ccagttaaag gtgaggaaga cttgtggggt cgttttttca gagctgccga taaaggtttg
cactccttgc aatcttttcc agttcgtaga ttcgtcccac gtgaagctgc catcagacac
gttattgaat ggatcattag acatcaagat gctgatggtg gctggggcgg tatccaacct
ccatggattt acggtttgat ggctttgtcc gtcgaaggtt atccattgca ccatccagtt
ttggccaagg ctatggacgc cttgaacgat cctggttggc gtagagataa gggtgacgct
tcttggatcc aagctactaa ctctccagtt tgggatacta tgttggctgt cttagctttg
cacgacgccg gtgccgaaga cagatactcc cctcaaatgg ataaggccat cggttggtta
ttggacagac aagttagagt caagggcgat tggtccatca agttgccaga cactgaacct
ggcggttggg ccttcgagta cgccaacgat aagtaccctg acaccgacga taccgccgtt
gccttgattg ccctagctgg ttgtagacac agaccagaat ggagagaaag agatatcgaa
ggtgccatct ccagaggtgt taactggtta ttggctatgc aatcttcctc tggcggttgg
ggtgctttcg acaaagacaa taacagatca atcttgacaa aaattccatt ctgtgacttc
ggtgaggctt tagatcctcc atctgttgac gttactgccc acgttttgga ggcttttggt
ctattgggta tttctagaaa ccacccatcc gtccaaaagg ccttagctta catcagatca
gaacaagaaa gaaacggtgc ctggtttggt agatggggtg ttaactacgt ttacggtacc
ggtgctgttt tgccagcttt ggccgctatt ggtgaagaca tgacccaacc atacattgtc
agagcttgtg actggttaat gtccgtccaa caagaaaacg gcggttgggg tgaatcttgc
gcttcttata tggacatcaa cgctgtcggt cacggtgtcg ctactgcctc tcaaactgct
tgggctttaa tcggtctatt ggctgccaaa agaccaaagg atagagaggc tattgctaga
ggttgtcaat ttttgatcga aagacaagaa gatggttctt ggacagaaga ggaatacacc
ggtactggtt tcccaggtta cggtgttggt caagctatca aattggacga tccatctttg
ccagatagac tattgcaagg tgctgaattg tccagagctt tcatgttgcg ttacgatttg
tacagacaat atttccctgt tatggctttg tccagagccc gtagaatgat gaaggaagat
gcttctgctg ccgcttaa

Sequence Number (ID): 206
Sequence Name: Gmo_SHC_WT
Length: 2058
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2058
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgagcccag cggacattag caccaaatcg tcatcttttc agcgtctgga caatatgttg
ccggaagccg tcagctctgc gtgcgattgg ttgatcgacc agcagaaacc ggatggtcac
tgggtgggcc cagtggagtc caacgcatgt atggaagcgc agtggtgctt ggcactgtgg
tttctgggcc aagaggacca cccgctgcgt ccgcgcctgg cgcaagcatt gctggaaatg
cagcgcgaag atggcagctg gggtatttac gtgggcgcgg atcatggtga tatcaacacg
accgtcgagg cgtatgcagc gctgcgcagc atgggttacg cagcggatat gccgatcatg
gctaagtccg cagcgtggat tcaacaaaag ggtggcctgc gcaatgtccg tgtttttacg
cgctactggc ttgccctgat tggtgagtgg ccgtgggata agaccccgaa tctgccgcca
gagatcattt ggctgccgga caatttcatc ttcagcattt acaactttgc gcaatgggct
cgtgcgacga tgatgccgct caccatcctg tctgcacgcc gtccgagccg tccgctgctg
ccggagaacc gtctggacgg tctgttccct gagggtagag agaacttcga ttatgaattg
ccggtgaagg gcgaagaaga tctgtggggt cgtttcttcc gtgcagcaga caagggtctg
catagcctgc agagcttccc ggttcgccgt ttcgtcccgc gtgaggctgc gatccgccac
gtcatcgagt ggattattcg tcaccaagat gcggatggtg gttggggtgg cattcaaccg
ccgtggattt atggcctgat ggccctgtcc gttgagggtt acccgttgca ccacccggtt
ctggcgaaag cgatggacgc tctgaacgac ccgggttggc gtcgcgataa gggcgatgct
agctggattc aagcgaccaa tagcccggtg tgggacacca tgctggccgt gctggcactg
cacgacgcgg gtgctgagga tcgttatagc ccgcagatgg acaaggctat tggttggctg
ctggaccgtc aggtgcgtgt taagggtgac tggagcatta agctgccgga taccgagccg
ggtggctggg catttgaata tgcaaacgac aaataccctg acaccgacga cactgcggtc
gccttgatcg cactggcggg ctgccgtcat cgtccggagt ggcgcgaacg tgatatcgag

ggcgcgatca gccgtggtgt caactggctg ctggcgatgc agtccagcag cggtggctgg
ggcgcattcg ataaagataa caatcgttcg attctgacta aaatcccgtt ttgtgacttt
ggtgaagccc tggacccgcc gagcgttgac gtgaccgccc atgtcctgga agcattcggt
ttactgggta tcagccgtaa ccatccgagc gtgcagaaag cactggcgta tatccgttct
gagcaagagc gcaatggtgc ctggttcggt cgctggggtg tcaattacgt atatggcacg
ggtgccgttc tcccggcgtt ggcagcgatt ggcgaagata tgacccagcc gtacattgtg
cgtgcctgcg actggctgat gagcgttcaa caagagaatg gcggctgggg tgagagctgt
gcgagctaca tggatatcaa cgccgtgggt cacggtgttg cgacggcgag ccagaccgcc
tgggcgctga ttggtctgct ggcggcgaag cgcccaaaag accgtgaagc gatcgctcgc
ggctgtcaat ttctgatcga gcgccaagaa gatggtagct ggaccgaaga agagtatacg
ggcacgggtt tcccgggcta cggcgttggt caggcaatca aactggatga cccgtccctg
cctgaccgcc tgctgcaggg tgccgagtta agccgtgctt tcatgctgcg ctatgacctg
taccgccagt actttcctgt tatggcgttg agccgtgcac gtcgtatgat gaaagaagat
gcgagcgctg cggcctaa

Sequence Number (ID): 207
Sequence Name: Gmo_SHC_WT
Length: 2178
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2178
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgggtatcg atagaatgaa ctccttgtct agattgttaa tgaaaaagat tttcggtgct
gaaaagactt cctacaagcc agcttctgac actatcattg gtaccgatac tttgaaaaga
ccaaacagac gtccagaacc aactgctaag gtcgacaaga ctatcttcaa gactatgggt
aactctttga ataacacctt ggtttctgct tgtgattggc taatcggtca acaaaaacca
gacggtcatt gggttggtgc tgttgaatcc aacgcttcta tggaagctga atggtgtttg
gctttatggt tcttgggttt ggaagatcac ccactaagac caagattggg caacgcttta
ttggaaatgc aaagagagga tggttcttgg ggtgtttatt tcggtgccgg taatggtgat
attaacgcca ctgttgaagc ctatgctgcc ttaagatcct tgggttactc cgctgataac
ccagttttga agaaagccgc tgcctggatt gctgaaaagg gtggcttgaa gaacatcaga
gttttcacca gatactggct agccttgatt ggtgaatggc catgggagaa gactccaaac
ttaccacctg aaatcatttg gtttccagac aactttgttt tctccattta caacttcgct
caatgggcta gagccaccat ggtcccaatc gctatcttat ccgcccgtag accatctaga
ccattaagac cacaagacag actagacgaa ttatttccag agggtagagc tagattcgac
tacgaactac caaaaaagga gggtatcgat ttgtggtcac aatttttcag aactaccgac
cgtggcttac attgggttca atctaactta ttgaaaagaa attccctaag agaagccgct
attagacatg ttctagaatg gattatcaga caccaagacg ctgacggcgg ttggggcggt
attcaacctc catgggtcta cggtttaatg gccttgcatg gtgaaggtta ccaattatac
catccagtta tggccaaggc tttgtctgct ttggacgatc caggttggag acacgacaga
ggtgaatcct cttggattca agctaccaat tccccagttt gggacaccat gttggccttg
atggctttga aggatgctaa ggccgaagat agattcactc cagaaatgga taaagccgct
gattggttgt tagccagaca agtcaaggtc aaaggtgatt ggtctatcaa gttgcctgac
gttgagccag gcggttgggc tttcgaatac gctaacgata gatacccaga taccgacgat
actgctgttg ctttgatcgc tttgtcatcc tacagagata aggaggaatg gcaaaaaaag
ggtgttgaag atgctatcac cagaggtgtt aactggttga ttgccatgca atctgaatgt
ggcggttggg gtgctttcga caaggataat aacagatcta tcctatctaa aatcccattc
tgtgatttcg gtgaatctat tgatccacct tctgttgacg tcactgctca cgttttggag
gcttttggta cattaggttt gtctagagat atgccagtta tccaaaaagc catcgattac
gtccgttccg aacaagaagc tgaaggtgct tggttcggta gatggggcgt caactacatt
tacggtactg gtgccgtctt gccagctttg gccgctatcg gtgaagacat gactcaacca
tacatcacca aggcttgtga ttggttagtc gctcaccaac aagaagacgg cggttggggt
gaatcctgtt catcttacat ggaaatcgac tctattggta agggtcctac taccccatca
caaactgctt gggctttgat gggtttaatt gccgctaaca gaccagaaga ttacgaagcc
atcgctaaag gttgtcatta cttgattgat cgtcaagaac aagatggttc ttggaaggaa
gaggaattta ccggtaccgg tttcccaggt tacggtgttg gtcaaaccat caagttggac
gatccagctt tgtccaagag attattgcaa ggtgctgaat tgtctagagc cttcatgttg
agatacgatt tctacagaca attctttcct atcatggctt tatctcgtgc cgaaagacta
attgatttga acaactaa

Sequence Number (ID): 208
Sequence Name: T0DEU9_ALIAG
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaccaagc agctcctgga caccccaatg gtgcaggcta cgctggaagc gggcgtagca
catctcctgc gccgtcaggc accagacggt tactggtggg cgccgctcct gtccaacgtt
tgcatggaag cggagtatgt actgctctgc cattgcctgg gtaagaaaaa cccggaacgc
gaggcacaga ttcgcaaata cattatctcc caacgccgtg aagacggtac ctggtccatt
tatccaggcg gtccgagcga cctcaacgcg accgttgaag cgtacgttgc tctgaaatac
ctgggtgaac cagcgtccga tccgcagatg gtccaggcga aagaatttat tcagaacgaa
ggtggcattg agtctactcg cgttttcact cgtctgtggc tggctatggt cggtcagtac
ccatgggata aactgccggt tatcccaccg gaaatcatgc atctgccgaa atccgtacca
ctgaacattt acgacttcgc ttcttgggct cgcgccacta ttgttaccct gacgattgtt
atgaaccgcc gtccggtgac cccgctgccg gattatgcaa aagttccgga actgtttgag
gcgaaacgtc ctccgaaacg ccgtagcgcg aaaggcggtg atagcggttt tttcgttgct
ctggataaat tcctcaaggc atacaacaaa tggccgatcc agccgggtcg taagtctggc
gaacagaaag ctctggaatg gattctggcc caccaggaag ctgacggttg ctggggcggt
atccagcctc cgtggttcta cgcgctcctg gctctcaagt gtctgaacat gaccgaccat
cctgcattcg ttaagggttt cgaaggtctg gaggcgtacg gtgttcacac ctccgacggc
ggttggatgt ttcaggcttc tattagcccg atttgggata ccggcctcac cgtactggca
ctgcgtagcg ctggtctgcc tccggatcac ccggcgctga ttaaagcagg cgaatggctg
gttagcaaac agattctgaa ggatggtgac tggaaagttc gtcgccgtaa agccaaaccg
ggtggctggg cgtttgagtt ccactgcgaa aactatccgg atgttgatga caccgcgatg
gtggtcctgg ctctgaacgg tatccagctg ccggacgaag gcaaacgccg tgacgctctg
acccgcggtt tccgctggct gcgcgaaatg cagtcctcta acggtggctg gggcgcatac
gacgtcgaca acacccgtca gctgactaac cgtatcccgt tttgcgactt cggcgaagtt
attgatccgc cttctgaaga cgtaaccgct catgttctgg aatgttttgg ttctttcggt
tacgacgaag catggaaagt tatccgcaaa gccgtagaat acctgaaagc tcaacagcgt
ccggatggca gctggttcgg ccgctggggt gtgaactacg tgtacggcat tggtgcagtt
gtcccaggcc tgaaggcagt tggcgttgac atgcgtgagc cgtgggtaca gaaatccctg
gactggctcg ttgagcacca gaacgaagac ggcggttggg gcgaggactg tcgttcttac
gatgacccac gcctggcggg tcagggcgtg agcacccctt cccaaactgc ttgggccctg
atggctctca tcgcgggtgg ccgtgtagaa agcgacgctg ttctgcgtgg tgtgacgtac
ctgcacgaca cgcagcgcgc cgacggtggc tgggatgagg aagtttatac cggtaccggc
ttcccgggtg acttctatct ggcgtacacc atgtaccgcg acatctttcc ggtttgggca
ctgggccgtt accaggaagc gatgcagcgt atccgcggtt aa

Sequence Number (ID): 209
Sequence Name: T0DEU9_ALIAG
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgaccaaac aattattgga taccccaatg gtccaagcta ctttggaggc cggtgttgct
cacttgttac gtagacaagc cccagatggt tattggtggg ctccattgct atctaacgtt
tgtatggaag ccgaatacgt tttattgtgt cactgtttgg gtaagaaaaa tccagaaaga
gaagctcaaa tcagaaagta tatcatttcc caacgtagag aagacggtac ttggtcaatt
tatccaggcg gtccatctga tttgaatgct accgttgagg cttatgttgc tttgaagtat
ttgggtgaac cagcttctga tccacaaatg gtccaagcta aggaatttat ccaaaacgaa
ggcggtatcg aatccaccag agttttcact agattgtggt tagctatggt cggtcaatac
ccatgggata agttaccagt tatccctcca gaaatcatgc acttgccaaa gtctgttcca
ttaaacatct atgacttcgc ctcttgggct cgtgctacca ttgtcactct aaccatcgtt

atgaaccgta gaccagtcac tccattgcca gactacgcta aagttccaga attattcgaa
gctaagagac cacctaaacg tagatccgct aagggcggtg actccggttt tttcgttgct
ttggacaagt tcttgaaagc ttacaacaag tggccaatcc aaccaggtag aaaatccggt
gaacaaaagg ctttggaatg gattctagct catcaagagg ctgacggttg ttggggcggt
attcaaccac cttggttcta cgctttattg gctctaaagt gtttgaatat gactgatcac
ccagcttttg tcaagggttt cgaaggtttg gaagcttatg gcgttcatac ttctgatggt
ggctggatgt tccaagcttc catttctcca atctgggata ctggtttgac tgttttggct
ttaagatccg ccggtttgcc tccagatcac ccagctctaa tcaaagctgg tgaatggttg
gtctctaaac aaatcttaaa ggacggtgat tggaaggtta gacgtagaaa ggccaagcca
ggcggttggg ctttcgaatt tcactgcgaa aactatcctg acgtcgacga tactgccatg
gttgtcttgg ctttgaacgg catccaattg ccagacgaag gtaaacgtag agacgctttg
acaagaggtt tcagatggct aagagaaatg caatcctcta acggcggttg gggtgcttac
gatgttgata acactagaca attgaccaac agaattccat tttgtgacgc tggtgaagtc
attgaccctc catctgaaga cgtcactgct catgttttgg aatgttttgg ctcttttggt
tatgacgaag cttggaaggt tattagaaag gctgtcgaat acttgaaggc tcaacaaaga
cctgatggtt cctggttcgg tagatggggt gtcaactacg tttacggcat cggtgctgtc
gttccaggtt taaaggccgt cggtgtcgat atgagagaac catgggttca aaagtccttg
gactggctag ttgaacacca aaacgaagac ggcggttggg gtgaagattg tagatcttac
gatgacccac gtctagctgg tcaaggtgtc tctaccccat cccaaaccgc ctgggcttta
atggctttaa ttgctggcgg tagagttgaa tctgacgctg tcctacgtgg tgtcacttac
ttacacgata ctcaaagagc tgatggcggt tgggacgagg aagtttacac aggcaccatg
tttccaggtg acttctactt agcttacacc atgtacagag atattttccc tgtttgggct
ctaggtagat accaagaagc catgcaaaga attcgtggtt aa

Sequence Number (ID): 210
Sequence Name: T0DEU9_ALIAG_F437A G600M
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgaccaaac agctcctgga taccccgatg gtacaagcta ccctggaggc gggtgtggcc
cacctcctgc gccgtcaggc accggacggc tactggtggg ccccgctcct gtctaacgtt
tgcatggagg cggaatacgt actcctgtgc cactgtctgg gcaagaaaaa cccggagcgt
gaagctcaga ttcgcaaata cattatctct cagcgccgtg aagacggcac ctggagcatc
tacccgggcg gtccttccga tctgaacgca accgtcgaag catacgtggc actgaagtat
ctgggtgaac cggctagcga cccgcaaatg gtgcaggcga aggagtttat ccagaacgaa
ggcggtatcg agtctacccg tgttttcacc cgtctgtggc tggcaatggt gggtcagtac
ccttgggaca aactgccggt tatcccaccg gaaatcatgc acctgccgaa atccgtcccg
ctgaacatct acgatttcgc ttcttgggca cgtgctacca tcgtgactct gaccattgta
atgaaccgtc gccctgttac gccgctgcct gactacgcga aagtaccgga gctgttcgaa
gcaaaacgcc caccgaagcg ccgtagcgct aagggcggtg acagcggctt tttcgttgca
ctggacaaat tcctgaaagc ctacaacaaa tggccgatcc agccgggtcg taagtccggc
gaacagaagg cgctggaatg gatcctggcg caccaggaag ctgatggctg ttggggcggt
atccagcctc cgtggttcta cgcactcctg gcgctgaaat gcctgaacat gaccgatcat
ccggcattcg taaaaggttt cgagggtctg gaggcgtatg gcgttcacac cagcgacggt
ggctggatgt tccaggcctc catcagccca atttgggaca ccggcctgac cgttctggcg
ctccgttctg cgggtctgcc gcctgatcat ccggctctga tcaaggcagg cgaatggctt
gtctctaagc aaattctgaa agacggtgac tggaaagttc gtcgccgtaa ggctaaacca
ggtggctggg ctttcgagtt ccactgcgaa aactacccgg atgtagatga caccgctatg
gtagttctgg ctctgaatgg tatccagctg ccggatgaag gtaaacgccg tgacgctctg
acccgcggct tccgttggct gcgtgaaatg cagtcttcca acggtggctg gggtgcctac
gatgtcgaca acacccgtca gctgacgaac cgtatcccgt tctgtgatgc gggcgaagtg
atcgatcctc cgtctgaaga tgtaaccgct cacgtgctgg aatgctttgg ctccttcggc
tatgacgaag cgtggaaagt aatccgcaaa gcagttgaat acctgaaagc tcaacagcgc
ccggacggta gctggttcgg tcgctggggt gttaactacg tctatggcat cggtgcagtt
gtaccgggcc tcaaagcagt gggcgtggat atgcgtgaac catgggtcca gaagtccctg
gactggctgg ttgaacacca gaacgaagat ggtggctggg gtgaagattg tcgtagctat
gacgatccgc gtctggccgg ccagggcgta tctacgccga gccagactgc ttgggcgctg

atggccctga ttgcgggcgg tcgcgtcgaa agcgacgcgg tgctgcgtgg cgttacctat
ctgcacgaca cccaacgcgc ggacggcggt tgggatgagg aagtctacac cggtaccatg
ttcccgggcg acttctatct ggcctacacg atgtaccgcg atatcttccc ggtatgggcg
ctgggtcgct atcaggaggc gatgcagcgt atccgtggct aa

Sequence Number (ID): 211
Sequence Name: T0DEU9_ALIAG_F437A G600M
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgactaagc aattgttaga caccccaatg gtccaagcta ctttagaagc cggtgtcgct
catctattgc gtagacaagc cccagatggt tactggtggg ctccactatt gtccaacgtt
tgtatggaag ctgaatatgt tttactatgt cactgtttgg gtaaaaagaa cccagaaaga
gaagcccaaa tcagaaagta catcatttct caaagacgtg aagacggtac ctggtctatt
tatcctggcg gtccatcaga cttgaacgct actgttgaag cctatgtcgc tttgaagtac
ttgggtgagc ctgcttccga cccacaaatg gttcaagcca aagaattcat tcaaaatgaa
ggcggtattg aatctaccag agtttttacc agattgtggt tggctatggt cggtcaatat
ccatgggata agttaccagt tatccctcca gaaatcatgc atttgccaaa atcagttcca
ttgaatattt atgactttgc ttcctgggct cgtgctacta ttgttacctt gaccatcgtc
atgaatcgta gaccagtcac accattacca gactatgcca aggttccaga attgttcgaa
gctaagagac ctccaaagcg tagatccgct aaaggcggtg attctggttt tttcgtcgct
ttggacaagt tcttgaaggc ttataacaag tggccaatcc aaccaggtcg taagtctggt
gaacaaaagg ctttggaatg gattttggct caccaagagg ctgacggttg ttggggcggt
attcaacctc catggttcta tgctttattg gctttgaagt gtctaaacat gactgaccac
ccagcttttg ttaagggttt cgaaggctta gaagcttacg gtgttcacac atctgatggc
ggttggatgt tccaagcttc tatctctcca atctgggata ccggtttgac tgttttggcc
ttgagatctg ctggtttgcc tccagaccac ccagctttga tcaaagctgg tgaatggtta
gtttctaaac aaatcttgaa ggacggtgac tggaaggtca gacgtagaaa agctaagcca
ggcggttggg ccttcgagtt ccactgtgaa aactacccag atgttgatga caccgctatg
gttgtcttgg ctttgaacgg tattcaattg ccagacgaag gtaagcgtag agatgctttg
accagaggtt tcagatggtt aagagaaatg caatcttcaa acggcggttg gggtgcctac
gacgtcgata acaccagaca attgactaac agaattccat tctgtgattt cggtgaagtt
atcgaccctc catccgaaga tgtcaccgct catgttttgg aatgtttcgg ttctttcggt
tacgatgaag cttggaaggt tatccgtaag gctgttgaat atttgaaagc tcaacaaaga
ccagacggtt catggttcgg tagatggggt gttaactatg tttacggtat cggtgctgtc
gttccaggct tgaaggccgt cggtgtcgac atgagagaac catgggtcca aaagtccttg
gattggctag ttgaacacca aaacgaagac ggcggttggg gtgaagattg tagatcctac
gacgatccaa gattggctgg tcaaggtgtt tccaccccat ctcaaactgc ttgggctttg
atggctttga ttgctggcgg tagagttgaa tctgatgctg tcttacgtgg tgttacctac
ttacatgaca ctcaaagagc tgatggcggt tgggacgagg aagtttacac aggtactggt
ttccctggtg acttttattt ggcctacaca atgtacagag atatcttccc agtttgggct
ttgggtagat accaagaagc tatgcaacgt attcgtggtt aa

Sequence Number (ID): 212
Sequence Name: ZmSHC_F503A_G667M
Length: 2178
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2178
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgggtattg atcgcatgaa ttcgctgagc cgtttgttga tgaaaaagat ctttggtgca
gagaaaacca gctataaacc ggcgtccgat accattattg gtacggacac cctgaagcgt
ccgaatcgtc gcccggagcc gactgcaaaa gtcgacaaaa ccattttcaa gacgatgggc
aacagcctga acaatacgct ggttagcgcg tgcgactggc tgatcggcca acagaaacct
gacggtcact gggttggtgc cgtcgaaagc aatgcgagca tggaagcgga gtggtgcttg

gctctgtggt tcttgggcct ggaagatcat cctctgcgtc cacgcctggg taacgcgctg
ctggaaatgc aacgcgaaga tggctcttgg ggcgtctatt ttggcgctgg taacggcgat
atcaatgcaa cggttgaggc ctacgcggct ctgcgtagcc tgggctatag cgcggacaat
ccggttctga agaaagcagc ggcgtggatt gccgagaagg gcggtctgaa gaatatccgt
gtgtttaccc gttactggct ggcgctgatt ggcgagtggc cgtgggaaaa gacgccgaat
ctgccaccgg agatcatctg gttcccggat aacttcgtgt ttagcatcta caatttcgcc
cagtgggcgc gtgccactat ggttccaatt gcaatcttga gcgcgcgtcg tccgagccgc
ccactgcgtc cgcaggatcg tctggacgaa ctgtttccgg aaggtcgtgc acgttttgat
tatgagctgc cgaagaaaga gggtatcgat ctgtggagcc aattcttccg caccaccgat
cgtggtttgc attgggtgca atctaacctg ttgaaacgta atagcctgcg tgaggcggcg
atccgtcacg tcctggagtg gattattcgt caccaggacg cagacggcgg ctggggtggc
atccaaccgc cgtgggttta cggtctgatg gccttgcatg gcgagggtta tcaactgtac
cacccggtta tggcgaaagc gctgagcgcc ctggatgacc cgggttggcg tcacgaccgt
ggtgagagca gctggattca agccaccaac agcccggttt gggataccat gctggcgttg
atggctctga aggacgcaaa agctgaggat cgcttcaccc cggaaatgga caaagcggcg
gactggttgc tggcacgtca ggttaaagtt aaaggtgact ggtccattaa gctgccggac
gtcgagccgg gtggttgggc ttttgagtat gctaacgacc gttacccgga tacggacgac
acggcagtgg cactgatcgc gctgagctct taccgcgata aagaagagtg gcagaagaag
ggtgtcgagg acgcgattac ccgcggtgtg aattggttga tcgctatgca aagcgagtgt
ggcggttggg gcgcttttga caaggacaac aaccgctcga tcctgagcaa gattccgttc
tgtgatgcgg gtgaatccat cgacccaccg agcgtggacg ttaccgcaca cgtgctggaa
gcgttcggta cgctgggtct gtcccgtgat atgccggtga ttcaaaaggc cattgactac
gtgcgttccg aacaagaggc agagggtgcc tggttcggcc gctggggcgt caactacatc
tacggtacgg gtgccgtcct gccggccttg gcggctattg gtgaggatat gacccagccg
tacatcacca aggcgtgcga ttggctggtc gctcatcaac aggaagatgg tggctggggc
gagtcttgta gctcgtatat ggagattgat tccattggta agggtcctac caccccgagc
cagaccgcgt gggcgctgat gggcctgatc gccgcaaacc gtcctgaaga ttatgaagcg
attgcgaagg gttgccacta cctgatcgat cgccaggagc aggacggtag ctggaaagaa
gaagagttta cgggtacgat gtttccgggt tatggtgtgg gtcagaccat taaactggac
gatccggcgt tgagcaaacg tctgctgcaa ggcgcggaac tgagccgtgc attcatgctg
cgttatgact tttaccgcca gttcttcccg atcatggcac tgagccgcgc agaacgcctg
atcgacctga ataactaa

Sequence Number (ID): 213
Sequence Name: ZmSHC_F503A_G667M
Length: 2178
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2178
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atgggtattg acagaatgaa ctcattgtcc cgtctattaa tgaaaaagat ctttggtgct
gaaaagacct cctacaagcc agcctctgac actatcattg gtactgacac cctaaagcgt
cctaacagac gtccagaacc aaccgctaag gttgacaaga ctatcttcaa gactatgggt
aactctttaa ataacacttt ggtttccgct tgcgactggc taatcggtca acaaaagcca
gatggtcact gggttggtgc tgttgaatct aacgcttcca tggaggctga atggtgtttg
gctttgtggt tcttgggttt agaagatcac ccattaagac caagattggg caacgcctta
ttggaaatgc aaagagaaga cggttcttgg ggtgtttact tcggtgccgg taacggtgat
attaatgcca cagtcgaagc ttatgctgcc ctaagatctt tgggttactc agccgataac
ccagttctaa aaaaggccgc tgcctggatc gctgaaaagg gcggtttgaa gaatatcaga
gtctttacaa gatactggtt ggctttgatt ggtgaatggc catgggaaaa gaccccaaac
ttacctccag aaatcatttg gttcccagac aatttcgttt tctctattta caactttgct
caatgggcca gagctaccat ggttccaatt gccatcttgt ccgctcgtag accttcaaga
ccattaagac cacaagatag actagatgaa ttgttcccag aaggtagagc cagattcgac
tacgaattac caaaaaagga aggtatcgat ttgtggtccc aatttttcag aaccactgat
agaggtttac actgggttca atctaactta ttgaagcgta actccttaag agaagccgct
atcagacatg ttttagaatg gatcattaga caccaagacg ctgacggcgg ttggggtggc
attcaacctc catgggtcta cggtttaatg gctttgcatg gtgaaggtta ccaattgtac
cacccagtca tggctaaggc tttgtctgcc ttagacgatc caggttggcg tcacgataga
ggcgaatcct cttggatcca agctactaac tccccagtct gggataccat gttggctttg

atggccttga aggatgctaa agctgaagat agattcactc cagaaatgga taaggctgcc
gactggttat tggctagaca agtcaaggtt aagggtgact ggtccattaa gttgccagat
gtcgaaccag gcggttgggc tttcgaatac gccaacgata gataccctga taccgatgac
accgctgttg ctttaattgc tttatcctct tacagagaca aggaggaatg gcaaaaaaag
ggtgttgaag atgctattac tcgtggtgtt aactggctaa tcgccatgca atctgaatgt
ggcggttggg gtgcctttga taaggacaat aacagatcta ttttgtccaa gatcccattc
tgcgacgctg gtgaatcaat cgacccacct tctgttgacg ttactgctca cgttttggaa
gcttttggca ccttgggttt gtctagagac atgccagtta ttcaaaaggc catcgattac
gttagatccg agcaagaagc cgaaggtgct tggttcggta gatggggtgt taactatatc
tacggtaccg gtgctgtttt gccagctttg gctgccatcg gtgaagatat gacccaacca
tacattacta aagcttgcga ctggttagtc gcccaccaac aagaagacgg cggttggggt
gaatcttgtt cttcctacat ggaaatcgac tcaatcggta agggtcctac taccccatct
caaaccgcct gggccttgat gggtttgatc gccgctaata gacctgaaga ctatgaagct
attgctaagg gttgtcacta cttaattgat agacaagaac aagacggttc ttggaaggaa
gaggaattta ccggtaccat gtttccaggt tatggtgtcg gtcaaactat taaattggat
gacccagcct tgtctaaaag attattgcaa ggtgctgaat tgtccagagc ttttatgtta
agatatgact tctacagaca atttttccca attatggctt tatccagagc tgaaagattg
atcgacttga acaactaa

Sequence Number (ID): 214
Sequence Name: ZmSHC_wt
Length: 2178
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2178
> mol_type, unassigned DNA
> organism, synthetic construct E.coli optimized
Residues:
atgggtattg atcgcatgaa ttcgctgagc cgtttgttga tgaaaaagat ctttggtgca
gagaaaacca gctataaacc ggcgtccgat accattattg gtacggacac cctgaagcgt
ccgaatcgtc gcccggagcc gactgcaaaa gtcgacaaaa ccattttcaa gacgatgggc
aacagcctga acaatacgct ggttagcgcg tgcgactggc tgatcggcca acagaaacct
gacggtcact gggttggtgc cgtcgaaagc aatgcgagca tggaagcgga gtggtgcttg
gctctgtggt tcttgggcct ggaagatcat cctctgcgtc cacgcctggg taacgcgctg
ctggaaatgc aacgcgaaga tggctcttgg ggcgtctatt ttggcgctgg taacggcgat
atcaatgcaa cggttgaggc ctacgcggct ctgcgtagcc tgggctatag cgcggacaat
ccggttctga agaaagcagc ggcgtggatt gccgagaagg gcggtctgaa gaatatccgt
gtgtttaccc gttactggct ggcgctgatt ggcgagtggc cgtgggaaaa gacgccgaat
ctgccaccgg agatcatctg gttcccggat aacttcgtgt ttagcatcta caatttcgcc
cagtgggcgc gtgccactat ggttccaatt gcaatcttga gcgcgcgtcg tccgagccgc
ccactgcgtc cgcaggatcg tctggacgaa ctgtttccgg aaggtcgtgc acgttttgat
tatgagctgc cgaagaaaga gggtatcgat ctgtggagcc aattcttccg caccaccgat
cgtggtttgc attgggtgca atctaacctg ttgaaacgta atagcctgcg tgaggcggcg
atccgtcacg tcctggagtg gattattcgt caccaggacg cagacggcgg ctggggtggc
atccaaccgc cgtgggttta cggtctgatg gccttgcatg gcgagggtta tcaactgtac
cacccggtta tggcgaaagc gctgagcgcc ctggatgacc cgggttggcg tcacgaccgt
ggtgagagca gctggattca agccaccaac agcccggttt gggataccat gctggcgttg
atggctctga aggacgcaaa agctgaggat cgcttcaccc cggaaatgga caaagcggcg
gactggttgc tggcacgtca ggttaaagtt aaaggtgact ggtccattaa gctgccggac
gtcgagccgg gtggttgggc ttttgagtat gctaacgacc gttacccgga tacggacgac
acggcagtgg cactgatcgc gctgagctct taccgcgata aagaagagtg gcagaagaag
ggtgtcgagg acgcgattac ccgcggtgtg aattggttga tcgctatgca aagcgagtgt
ggcggttggg gcgcttttga caaggacaac aaccgctcga tcctgagcaa gattccgttc
tgtgatttcg gtgaatccat cgacccaccg agcgtggacg ttaccgcaca cgtgctggaa
gcgttcggta cgctgggtct gtcccgtgat atgccggtga ttcaaaaggc cattgactac
gtgcgttccg aacaagaggc agagggtgcc tggttcggcc gctggggcgt caactacatc
tacggtacgg gtgccgtcct gccggccttg gcggctattg gtgaggatat gacccagccg
tacatcacca aggcgtgcga ttggctggtc gctcatcaac aggaagatgg tggctggggc
gagtcttgta gctcgtatat ggagattgat tccattggta agggtcctac caccccgagc
cagaccgcgt gggcgctgat gggcctgatc gccgcaaacc gtcctgaaga ttatgaagcg
attgcgaagg gttgccacta cctgatcgat cgccaggagc aggacggtag ctggaaagaa

gaagagttta cgggtacggg ttttccgggt tatggtgtgg gtcagaccat taaactggac
gatccggcgt tgagcaaacg tctgctgcaa ggcgcggaac tgagccgtgc attcatgctg
cgttatgact tttaccgcca gttcttcccg atcatggcac tgagccgcgc agaacgcctg
atcgacctga ataactaa

Sequence Number (ID): 215
Sequence Name: ZmSHC_wt
Length: 1896
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1896
> mol_type, unassigned DNA
> organism, synthetic construct S. cerevisiae optimized
Residues:
atggctgaac aattggtcga agctccagcc tacgctcgta ctttggatcg tgctgttgaa
tacttgttat cttgtcaaaa agatgaaggt tattggtggg gtccactatt gtctaacgtt
acaatggaag ctgaatatgt cctattgtgt catattttgg atagagttga cagagataga
atggaaaaga ttcgtagata cctattgcac gagcaaagag aagacggtac ctgggctttg
tacccaggcg gtcctccaga cttagacaca actatcgagg cttacgttgc cttgaagtac
atcggtatgt cacgtgacga ggaaccaatg caaaaggctt tgcgtttcat ccaatctcaa
ggcggtattg aatcttccag agtcttcacc agaatgtggt tggctttggt cggtgaatac
ccatgggaaa aagttccaat ggtccctcca gaaattatgt tcttgggtaa gagaatgcca
ttaaacatct acgagttcgg ctcttgggct agagctaccg tcgttgcttt gtctattgtt
atgtcaagac aaccagtttt ccctttgcca gaaagagcta gagttccaga attgtatgaa
accgacgttc ctccaagacg tagaggtgcc aagggcggtg gcggttggat cttcgatgcc
ttggacagag ctttgcacgg ttatcaaaaa ttgtctgttc acccatttag acgtgctgcc
gaaattcgtg ctttggactg gttattggaa cgtcaagctg gtgatggctc ttggggcggt
atccaacctc catggttcta cgctttgatt gccttgaaga tcttggatat gactcaacat
cctgccttca tcaagggctg ggagggtttg gaattgtatg gtgtcgagtt ggattacggc
ggttggatgt tccaggcttc tatctctcca gtttgggata caggtttggc tgttctagct
ttgagagctg ccggtttgcc agccgatcac gacagattag ttaaagccgg tgaatggcta
ttggacagac aaattaccgt cccaggtgat tgggccgtca aaagaccaaa cttgaagcct
ggcggtttcg cttttcaatt cgataacgtc tattacccag acgtcgatga caccgctgtt
gtcgtttggg ctctaaatac tttaagattg ccagatgaaa gacgtagacg tgacgctatg
accaaaggtt ttagatggat tgtcggtatg caatcttcca acggcggttg gggtgcttac
gatgtcgata acacttctga tttgcctaac cacattccat tctgtgactt tggtgaagtc
actgatcctc catccgaaga tgttactgct cacgtcttgg aatgtttcgg ttctttcggt
tacgatgacg cttggaaagt catccgtaga gctgtcgaat acttgaagcg tgagcaaaag
cctgacggtt cttggttcgg tcgttggggt gttaactatt tgtatggtac cggtgctgtt
gtctctgcct tgaaggctgt tggtattgac acccgtgaac catacattca aaaggcttta
gactgggttg aacaacatca aaacccagac ggcggttggg gtgaagactg tcgttcttat
gaagacccag cttacgccgg taaaggcgcc tctaccccat ctcaaaccgc ttgggccttg
atggctttaa ttgctggcgg tagagccgaa tctgaagccg ctcgtagagg tgttcaatac
ttagtcgaaa ctcaaagacc agatggcggt tgggacgaac catattacac aggtaccggt
tttccaggtg acttctactt gggttacacc atgtatagac atgtctttcc aactctagcc
ttgggtagat acaagcaagc tatcgaaaga agataa

Sequence Number (ID): 216
Sequence Name: EthA
Length: 489
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..489
> mol_type, protein
> organism, Mycobacterium tuberculosis CDC1551
Residues:
MTEHLDVVIV GAGISGVSAA WHLQDRCPTK SYAILEKRES MGGTWDLFRY PGIRSDSDMY
TLGFRFRPWT GRQAIADGKP ILEYVKSTAA MYGIDRHIRF HHKVISADWS TAENRWTVHI
QSHGTLSALT CEFLFLCSGY YNYDEGYSPR FAGSEDFVGP IIHPQHWPED LDYDAKNIVV
IGSGATAVTL VPALADSGAK HVTMLQRSPT YIVSQPDRDG IAEKLNRWLP ETMAYTAVRW
KNVLRQAAVY SACQKWPRRM RKMFLSLIQR QLPEGYDVRK HFGPHYNPWD QRLCLVPNGD

LFRAIRHGKV EVVTDTIERF TATGIRLNSG RELPADIIIT ATGLNLQLFG GATATIDGQQ
VDITTTMAYK GMMLSGIPNM AYTVGYTNAS WTLKADLVSE FVCRLLNYMD DNGFDTVVVE
RPGSDVEERP FMEFTPGYVL RSLDELPKQG SRTPWRLNQN YLRDIRLIRR GKIDDEGLRF
AKRPAPVGV

Sequence Number (ID): 217
Sequence Name: CPDMO
Length: 601
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..601
> mol_type, protein
> organism, Pseudomonas sp. HI-70
Residues:
MSQLIQEPAE AGVTSQKVSF DHVALREKYR QERDKRLRQD GQEQYLEVAV TCDEYLKDPY
ADPIVRDPVV RETDVFIIGG GFGGLLAAVR LQQAGVSDYV MVERAGDYGG TWYWNRYPGA
QCDIESYVYM PLLEEMGYIP TEKYAFGTEI LEYSRSIGRK FGLYERTYFQ TEVKDLSWDD
EAARWRITTD RGDKFSARFV CMSTGPLQRP KLPGIPGITS FKGHSFHTSR WDYSYTGGDQ
TGNLEGLKDK RVAIIGTGAT SIQAVPHLAA YAQELYVIQR TPISVGFRGN KPTDPEWAKS
LQPGWQQARM DNFNAITHGM PVDVDLVQDS WTKIFGEIGV FLGSDGSRAQ MVDFQLMEQI
RARVDQEVKD PATAESLKPY YNIMCKRPGF HDSYLPSFNK PNVTLVDTQG AGVERITEKG
LVVNGREYEV DCLIYATGFE YQTKLSRRNG YEIHGRNGQP LSDKWKDGLS TLWGYHIRDF
PNCFILGNGQ SAVTPNFTHM LNEAGKHVAY VVKHCLDERV DVFEPTAEAE QAWVDHVMSF
AGIKQQYDRE CTPSYYNNEG QVNDVALTRN NFYPGGAVAF INILREWREK GDFAQFQQRK
R

Sequence Number (ID): 218
Sequence Name: PAMO
Length: 542
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..542
> mol_type, protein
> organism, Thermobifida fusca YX
Residues:
MAGQTTVDSR RQPPEEVDVL VVGAGFSGLY ALYRLRELGR SVHVIETAGD VGGVWYWNRY
PGARCDIESI EYCYSFSEEV LQEWNWTERY ASQPEILRYI NFVADKFDLR SGITFHTTVT
AAAFDEATNT WTVDTNHGDR IRARYLIMAS GQLSVPQLPN FPGLKDFAGN LYHTGNWPHE
PVDFSGQRVG VIGTGSSGIQ VSPQIAKQAA ELFVFQRTPH FAVPARNAPL DPEFLADLKK
RYAEFREESR NTPGGTHRYQ GPKSALEVSD EELVETLERY WQEGGPDILA AYRDILRDRD
ANERVAEFIR NKIRNTVRDP EVAERLVPKG YPFGTKRLIL EIDYYEMFNR DNVHLVDTLS
APIETITPRG VRTSEREYEL DSLVLATGFD ALTGALFKID IRGVGNVALK EKWAAGPRTY
LGLSTAGFPN LFFIAGPGSP SALSNMLVSI EQHVEWVTDH IAYMFKNGLT RSEAVLEKED
EWVEHVNEIA DETLYPMTAS WYTGANVPGK PRVFMLYVGG FHRYRQICDE VAAKGYEGFV
LT

Sequence Number (ID): 219
Sequence Name: STMO
Length: 549
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..549
> mol_type, protein
> organism, Rhodococcus rhodochrous
Residues:
MNGQHPRSVV TAPDATTGTT SYDVVVVGAG IAGLYAIHRF RSQGLTVRAF EAASGVGGVW
YWNRYPGARC DVESIDYSYS FSPELEQEWN WSEKYATQPE ILAYLEHVAD RFDLRRDIRF
DTRVTSAVLD EEGLRWTVRT DRGDEVSARF LVVAAGPLSN ANTPAFDGLD RFTGDIVHTA
RWPHDGVDFT GKRVGVIGTG SSGIQSIPII AEQAEQLFVF QRSANYSIPA GNVPLDDATR
AEQKANYAER RRLSRESGGG SPHRPHPKSA LEVSEEERRA VYEERWKLGG VLFSKAFPDQ
LTDPAANDTA RAFWEEKIRA VVDDPAVAEL LTPKDHAIGA KRIVTDSGYY ETYNRDNVEL

VDLRSTPIVG MDETGIVTTG AHYDLDMIVL ATGFDAMTGS LDKLEIVGRG GRTLKETWAA
GPRTYLGLGI DGFPNFFNLT GPGSPSVLAN MVLHSELHVD WVADAIAYLD ARGAAGIEGT
PEAVADWVEE CRNRAEASLL NSANSWYLGA NIPGRPRVFM PFLGGFGVYR EIITEVAESG
YKGFAILEG

Sequence Number (ID): 220
Sequence Name: HPAMO
Length: 640
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..640
> mol_type, protein
> organism, Pseudomonas fluorescens
Residues:
MSAFNTTLPS LDYDDDTLRE HLQGADIPTL LLTVAHLTGD LQILKPNWKP SIAMGVARSG
MDLETEAQVR EFCLQRLIDF RDSGQPAPGR PTSDQLHILG TWLMGPVIEP YLPLIAEEAV
TAEEDLRAPR WHKDHVASGR DFKVVIIGAG ESGMIAALRF KQAGVPFVIY EKGNDVGGTW
RENTYPGCRV DINSFWYSFS FARGIWDDCF APAPQVFAYM QAVAREHGLY EHIRFNTEVS
DAHWDESTQR WQLLYRDSEG QTQVDSNVVV FAVGQLNRPM IPAIPGIETF KGPMFHSAQW
DHDVDWSGKR VGVIGTGASA TQFIPQLAQT AAELKVFART TNWLLPTPDL HEKISDSCKW
LLAHVPHYSL WYRVAMAMPQ SVGFLEDVMV DVGYPPTELA VSARNDRLRQ DISAWMEPQF
ADRPDLREVL IPDSPVGGKR IVRDNGTWIS TLKRDNVSMI RQPIEVITPK GICCVDGTEH
EFDLIVYGTG FHASKFLMPI NVTGRDGVAL HDVWKGDDAR AYLGMTVPQF PNMFCMYGPN
TGLVVYSTVI QFSEMTASYI VDAVRLLLEG GHQSMEVKTP VFESYNQRVD EGNALRAWGF
SKVNSWYKNS KGRVTQNFPF TAVEFWQRTH SVEPTDYQLG

Sequence Number (ID): 221
Sequence Name: ACMO
Length: 533
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..533
> mol_type, protein
> organism, Gordonia sp.
Residues:
MSTTTLDAAV IGTGVAGLYE LHMLREQGLE VRAYDKASGV GGTWYWNRYP GARFDSEAYI
YQYLFDEDLY KGWSWSQRFP GQEEIERWLN YVADSLDLRR DISLETEITS AVFDEDRNRW
TLTTADGDTI DAQFLITCCG MLSAPMKDLF PGQSDFGGQL VHTARWPKEG IDFAGKRVGV
IGNGATGIQV IQSIAADVDE LKVFIRTPQY ALPMKNPSYG PDEVAWYKSR FGELKDTLPH
TFTGFEYDFT DAWEDLTPEQ RRARLEDDYE NGSLKLWLAS FAEIFSDEQV SEEVSEFVRE
KMRARLVDPE LCDLLIPSDY GFGTHRVPLE TNYLEVYHRD NVTAVLVRDN PITRIRENGI
ELADGTVHEL DVIIMATGFD AGTGALTRID IRGRDGRTLA DDWSRDIRTT MGLMVHGYPN
MLTTAVPLAP SAALCNMTTC LQQQTEWISE AIRHLRATGK TVIEPTAEGE EAWVAHHDEL
ADANLISKTN SWYVGSNVPG KPRRVLSYVG GVGAYRDATL EAAAAGYKGF ALS

Sequence Number (ID): 222
Sequence Name: MekA
Length: 549
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..549
> mol_type, protein
> organism, Pseudomonas veronii
Residues:
MSAQSKLAAG SCAYGNVTSL DAMVIGAGVA GLYQLYRLRE MGLTVRAYDT ASGVGGTWYW
NRYPGARFDS QAEIYQYWFS EELYKSWQPT ERFPAQPETE EWLNFVANRL NLKKDIQFNT
RIASAHFCED SGRWVVTTAA GETINTQYLI SCCGMLSAPL SDRFPGQADF QGQIYHTGLW
PKDPVDFNGK RVAVVGTGAT GIQVIQTIAP TVGSMTVFVR TPQYVIPMRN PKYSKADWEK
WGTQFHQLKK RVRETFAGFD YDFDAGPWAE KTPDERQAVL EQLWKDGSLA MWLASFPEMF
FDEQVNEVVS QFVRIKMRER LRSRPDLCDL LIPTDYGFGT HRVPLENNYL EVYLQSNVKA
VDCKQSPIER IVPQGIQTAD GKIHEVDIIV LAVGFDAGSG ALSRIDIRGR DSRSLKEQWQ

QEIRTAMGLQ IHGYPNLFTT GAPLAPSAAL CNMTTCLQQQ VDWITGCIEF AAEHGKHVVE
ASKALEDNWV QHHDETAAKT LVVKTDSWYM GSNVDGKPRR LLSYIGGAGD YHRRCAEIAA
QGYPGFEMA

Sequence Number (ID): 223
Sequence Name: CDMO
Length: 603
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..603
> mol_type, protein
> organism, Rhodococcus ruber
Residues:
MTTSIDREAL RRKYAEERDK RIRPDGNDQY IRLDHVDGWS HDPYMPITPR EPKLDHVTFA
FIGGGFSGLV TAARLRESGV ESVRIIDKAG DFGGVWYWNR YPGAMCDTAA MVYMPLLEET
GYMPTEKYAH GPEILEHCQR IGKHYDLYDD ALFHTEVTDL VWQEHDQRWR ISTNRGDHFT
AQFVGMGTGP LHVAQLPGIP GIESFRGKSF HTSRWDYDYT GGDALGAPMD KLADKRVAVI
GTGATAVQCV PELAKYCREL YVVQRTPSAV DERGNHPIDE KWFAQIATPG WQKRWLDSFT
AIWDGVLTDP SELAIEHEDL VQDGWTALGQ RMRAAVGSVP IEQYSPENVQ RALEEADDEQ
MERIRARVDE IVTDPATAAQ LKAWFRQMCK RPCFHDDYLP AFNRPNTHLV DTGGKGVERI
TENGVVVAGV EYEVDCIVYA SGFEFLGTGY TDRAGFDPTG RDGVKLSEHW AQGTRTLHGM
HTYGFPNLFV LQLMQGAALG SNIPHNFVEA ARVVAAIVDH VLSTGTSSVE TTKEAEQAWV
QLLLDHGRPL GNPECTPGYY NNEGKPAELK DRLNVGYPAG SAAFFRMMDH WLAAGSFDGL
TFR

Sequence Number (ID): 224
Sequence Name: BpCHMO
Length: 537
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..537
> mol_type, protein
> organism, Brachymonas petroleovorans
Residues:
MSSSPSSAIH FDAIVVGAGF GGMYMLHKLR DQLGLKVKVF DTAGGIGGTW YWNRYPGALS
DTHSHVYQYS FDEAMLQEWT WKNKYLTQPE ILAYLEYVAD RLDLRPDIQL NTTVTSMHFN
EVHNIWEVRT DRGGYYTARF IVTALGLLSA INWPNIPGRE SFQGEMYHTA AWPKDVELRG
KRVGVIGTGS TGVQLITAIA PEVKHLTVFQ RTPQYSVPTG NRPVSAQEIA EVKRNFSKVW
QQVRESAVAF GFEESTVPAM SVSEAERQRV FQEAWNQGNG FYYMFGTFCD IATDPQANEA
AATFIRNKIA EIVKDPETAR KLTPTDVYAR RPLCDSGYYR TYNRSNVSLV DVKATPISAM
TPRGIRTADG VEHELDMLIL ATGYDAVDGN YRRIDLRGRG GQTINEHWND TPTSYVGVST
ANFPNMFMIL GPNGPFTNLP PSIEAQVEWI TDLVAHMRQH GLATAEPTRD AEDAWGRTCA
EIAEQTLFGQ VESWIFGANS PGKKHTLMFY LAGLGNYRKQ LADVANAQYQ GFAFQPL

Sequence Number (ID): 225
Sequence Name: Ocean
Length: 485
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..485
> mol_type, protein
> organism, Pseudooceanicola batsensis
Residues:
MNIQTENTKT VGADFDAVVI GAGFGGLYAV HKLRNEQGLN VRGYDSASDV GGTWWWNRYP
GALSDTESYV YRYSFDKELL RKGRWKTRYL TQPEILEYMN EVADHLDLRR SYKFDTKVDG
AHYNEKTGLW NVITDSGETV TAKYLVTGLG LLSATNVPKF KGIDDFKGRI LHTGAWPEGV
DLSNKRVGII GTGSTGVQVI TATAPIAKHL TVFQRSAQYV VPIGNTPQDD ATIAEQKANY
DNIWNQVKNS VVAFGFEESA EPAETASPEE RERVFEAAWQ RGGGFYFMFG TFCDIATSQV
ANDAAADFIK GKIKQIVKDP KVAEKLTPKD LYAKRPLCGN NYYEVYNRDN VTLADVKADP
GMMEVDFPNF FMILGPNGPF TNLPPSIETQ VEWIADTICA MEEEGVQSVE PTVEARDAWV
GTCREIADMT LFPKAESWIF GANIPGKKNA VMFYMAGIGN YRNAISAVKE EGYTSLIRDR

TAEKV

Sequence Number (ID): 226
Sequence Name: CPDMO
Length: 601
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..601
> mol_type, protein
> organism, Pseudomonas sp. HI-70
Residues:
MSQLIQEPAE AGVTSQKVSF DHVALREKYR QERDKRLRQD GQEQYLEVAV TCDEYLKDPY
ADPIVRDPVV RETDVFIIGG GFGGLLAAVR LQQAGVSDYV MVERAGDYGG TWYWNRYPGA
QCDIESYVYM PLLEEMGYIP TEKYAFGTEI LEYSRSIGRK FGLYERTYFQ TEVKDLSWDD
EAARWRITTD RGDKFSARFV CMSTGPLQRP KLPGIPGITS FKGHSFHTSR WDYSYTGGDQ
TGNLEGLKDK RVAIIGTGAT SIQAVPHLAA YAQELYVIQR TPISVGFRGN KPTDPEWAKS
LQPGWQQARM DNFNAITHGM PVDVDLVQDS WTKIFGEIGV FLGSDGSRAQ MVDFQLMEQI
RARVDQEVKD PATAESLKPY YNIMCKRPGF HDSYLPSFNK PNVTLVDTQG AGVERITEKG
LVVNGREYEV DCLIYATGFE YQTKLSRRNG YEIHGRNGQP LSDKWKDGLS TLWGYHIRDF
PNCFILGNGQ SAVTPNFTHM LNEAGKHVAY VVKHCLDERV DVFEPTAEAE QAWVDHVMSF
AGIKQQYDRE CTPSYYNNEG QVNDVALTRN NFYPGGAVAF INILREWREK GDFAQFQQRK
R

Sequence Number (ID): 227
Sequence Name: BVMO 24
Length: 564
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..564
> mol_type, protein
> organism, Rhodococcus jostii (strain RHA1)
Residues:
MTTSMKAANP MNFPSTSDTG IVDVLGVGAG FSGLYLSHRL TTAGWTFAGF EAGPSVGGTW
FWNTYPGARC DVESIYYSYS FDEALQQEWT WSQRFAPQAE ILSYINHVAD RFDLRKHFTF
NTRVVGATWN AAERLWEVQL DNGETRRGRY LISGAGGLST PKDFDVPGLG NFTGLQVSTS
RWNISLDDLA GKRVAVIGTG SSGVQAIPLI AEVAEHVTVF QRTPNYVMPA RNAELPLERV
DSIKDDYPAI REECRHSPGG IPDRPVTDKA FDVSAEERQR RYEAAYERSG FNGVGGEFAD
LLTDVEANRT ASEFIHDKIR EIVEDPATAE LLVPRYHPLG AKRSVFGTDY YETYNRPNVS
LVSLRDEPIE TMTANAIVTS KGTYEADAVV LAIGFDAFTG PLYGLGLTGA SGRKLQETWQ
DGIRTYLGMM TTDFPNFFMV AGPQSPALAS NVVMTIEQAV DWIADLIEHA RDSGATLVEA
TPEGQNDWVD ITEETVAQTL YATTDSWYRG SNVEGKPNTF MGYVGGVGKY RRMCTEIAKR
GYPGVRIDGE TESPHLGPIH REIS

Sequence Number (ID): 228
Sequence Name: ADH - CHTD motif
Length: 4
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..4
> mol_type, protein
> organism, synthetic construct
Residues:
CHTD

Sequence Number (ID): 229
Sequence Name: ADH - GHEGxG motif
Length: 6
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..6
> mol_type, protein

```
                > organism, synthetic construct
        - VARIANT, 5
                > note, this can be any amino acid
    Residues:
    GHEGXG

    Sequence Number (ID): 230
    Sequence Name: ADH - LxCGxxTGxG motif
    Length: 10
    Molecule Type: AA
    Features Location/Qualifiers:
        - source, 1..10
                > mol_type, protein
                > organism, synthetic construct
        - VARIANT, 2
    Residues:
    LXCGXXTGXG

    Sequence Number (ID): 231
    Sequence Name: ADH - Gx[VI]GL motif
    Length: 5
    Molecule Type: AA
    Features Location/Qualifiers:
        - source, 1..5
                > mol_type, protein
                > organism, synthetic construct
        - VARIANT, 2
                > note, any amino acid residue
        - VARIANT, 3
                > note, any amino acid residue
    Residues:
    GxxGL


    Sequence Number (ID): 232
    Sequence Name: ADH - LxxxG[LVI][PA] motif
    Length: 7
    Molecule Type: AA
    Features Location/Qualifiers:
        - source, 1..7
                > mol_type, protein
                > organism, synthetic construct
        - VARIANT, 2
                > note, any amino acid residue
        - VARIANT, 3
                > note, any amino acid residue
        - VARIANT, 4
                > note, any amino acid residue
        - VARIANT, 6
                > note, can be L, V or I
        - VARIANT, 7
                > note, can be P or A
    Residues:
    LxxxGxx

    Sequence Number (ID): 233
    Sequence Name: ADH - GxVxAI motif
    Length: 6
    Molecule Type: AA
    Features Location/Qualifiers:
        - source, 1..6
```

```
                > mol_type, protein
                > organism, synthetic construct
        Residues:
        GXVXAI

        Sequence Number (ID): 234
        Sequence Name: ADH - YxATKxA motif
        Length: 7
        Molecule Type: AA
        Features Location/Qualifiers:
              - source, 1..7
                    > mol_type, protein
                    > organism, synthetic construct
              - VARIANT, 2
                    > note, any amino acid residue
              - VARIANT, 6
                    > note, any amino acid residue
        Residues:
        YXATKXA

        Sequence Number (ID): 235
        Sequence Name: enal cleaving GxxWxGxxxxxGx
        Length: 13
        Molecule Type: AA
        Features Location/Qualifiers:
              - source, 1..13
                    > mol_type, protein
                    > organism, synthetic construct
              - VARIANT, 2
                    > note, can be Y or deleted
              - VARIANT, 3
                    > note, any amino acid residue
              - VARIANT, 5
                    > note, any amino acid residue
              - VARIANT, 7
                    > note, any amino acid residue
              - VARIANT, 8
                    > note, any amino acid residue
              - VARIANT, 9
                    > note, any amino acid residue
              - VARIANT, 10
                    > note, any amino acid residue
              - VARIANT, 11
                    > note, any amino acid residue
              - VARIANT, 13
                    > note, any amino acid residue
        Residues:
        GXXWXGXXXX  XGX

        Sequence Number (ID): 236
        Sequence Name: enal cleaving WxGKxxxx

        Length: 8
        Molecule Type: AA
        Features Location/Qualifiers:
              - source, 1..8
                    > mol_type, protein
                    > organism, synthetic construct
              - VARIANT, 2
                    > note, any amino acid residue
              - VARIANT, 3
```

> note, any amino acid residue
- VARIANT, 5
> note, any amino acid residue
- VARIANT, 6
> note, any amino acid residue
- VARIANT, 7
> note, any amino acid residue
- VARIANT, 8
> note, any amino acid residue

Residues:
WxGKxxxx

Sequence Number (ID): 237
Sequence Name: enal cleaving xxxxxxxxxxRGxV
Length: 14
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..14
> mol_type, protein
> organism, synthetic construct
- VARIANT, 1
> note, any amino acid residue
- VARIANT, 2
> note, any amino acid residue
- VARIANT, 3
> note, any amino acid residue
- VARIANT, 5
> note, any amino acid residue
- VARIANT, 6
> note, any amino acid residue
- VARIANT, 7
> note, any amino acid residue
- VARIANT, 8
> note, any amino acid residue
- VARIANT, 9
> note, any amino acid residue
- VARIANT, 10
> note, any amino acid residue
- VARIANT, 13
> note, any amino acid residue

Residues:
xxxxxxxxxxRGxV

Sequence Number (ID): 238
Sequence Name: enal cleaving xxYDxxPxxDxx
Length: 12
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..12
> mol_type, protein
> organism, synthetic construct
- VARIANT, 1
> note, any amino acid residue
- VARIANT, 2
> note, any amino acid residue
- VARIANT, 5
> note, any amino acid residue
- VARIANT, 6

```
                > note, any amino acid residue
          - VARIANT, 8
                > note, any amino acid residue
          - VARIANT, 9
                > note, any amino acid residue
          - VARIANT, 11
                > note, any amino acid residue
          - VARIANT, 12
                > note, any amino acid residue
    Residues:
    XXYDXXPXXD  XX

    Sequence Number (ID): 239
    Sequence Name: BVMO GxGxxG
    Length: 6
    Molecule Type: AA
    Features Location/Qualifiers:
          - source, 1..6
                > mol_type, protein
                > organism, synthetic construct
          - VARIANT, 2
                > note, any amino acid residue
          - VARIANT, 4
                > note, any amino acid residue
          - VARIANT, 5
                > note, any amino acid residue
    Residues:
    GxGxxG


    Sequence Number (ID): 240
    Sequence Name: BVMO [GS]GxWxxxxYPGxxxD
    Length: 15
    Molecule Type: AA
    Features Location/Qualifiers:
          - source, 1..15
                > mol_type, protein
                > organism, synthetic construct
          - VARIANT, 1
                > note, can be G or S
          - VARIANT, 3
                > note, any amino acid residue
          - VARIANT, 5
                > note, any amino acid residue
          - VARIANT, 6
                > note, any amino acid residue
          - VARIANT, 7
                > note, any amino acid residue
          - VARIANT, 8
                > note, any amino acid residue
          - VARIANT, 12
                > note, any amino acid residue
          - VARIANT, 13
                > note, any amino acid residue
          - VARIANT, 14
                > note, any amino acid residue
    Residues:
    XGXWXXXXYP  GXXXD

    Sequence Number (ID): 241
    Sequence Name: BVMO Gxxx[FY]xGxxx[HS]xxxW
```

Length: 15
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..15
> mol_type, protein
> organism, synthetic construct
- VARIANT, 2
> note, any amino acid residue
- VARIANT, 3
> note, any amino acid residue
- VARIANT, 4
> note, any amino acid residue
- VARIANT, 5
> note, can be F or Y
- VARIANT, 6
> note, any amino acid residue
- VARIANT, 7
> note, any amino acid residue
- VARIANT, 8
> note, any amino acid residue
- VARIANT, 9
> note, any amino acid residue
- VARIANT, 10
> note, any amino acid residue
- VARIANT, 11
> note, can be H or S
- VARIANT, 12
> note, any amino acid residue
- VARIANT, 13
> note, any amino acid residue
- VARIANT, 14
> note, any amino acid residue
Residues:
GxxxxxGxxxxxxxW

Sequence Number (ID): 242
Sequence Name: BVMO [KQ]x[VI]xx[IV]GxG
Length: 9
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..15
> mol_type, protein
> organism, synthetic construct
- VARIANT, 1
> note, can be K or Q
- VARIANT, 2
> note, any amino acid residue
- VARIANT, 3
> note, can be V or I
- VARIANT, 4
> note, any amino acid residue
- VARIANT, 5
> note, any amino acid residue
- VARIANT, 6
> note, can be V or I
- VARIANT, 8
> note, any amino acid residue
- VARIANT, 9
> note, any amino acid residue
Residues:

xxxxxxGxG

Sequence Number (ID): 243
Sequence Name: esterase AxVVxVXxRLAPE
Length: 13
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..13
> mol_type, protein
> organism, synthetic construct
- VARIANT, 2
> note, any amino acid residue
- VARIANT, 5
> note, any amino acid residue
- VARIANT, 8
> note, any amino acid residue
Residues:
AXVVXVXXRL APE

Sequence Number (ID): 244
Sequence Name: esterase GASAGGGLXA
Length: 10
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..10
> mol_type, protein
> organism, synthetic construct
- VARIANT, 9
> note, any amino acid residue
Residues:
GASAGGGLXA

Sequence Number (ID): 245
Sequence Name: esterase VxQLLXYPMLDDR
Length: 13
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..13
> mol_type, protein
> organism, synthetic construct
- VARIANT, 2
> note, any amino acid residue
Residues:
VXQLLXYPML DDR

Sequence Number (ID): 246
Sequence Name: esterase ARxxDLSGLPxT
Length: 12
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..12
> mol_type, protein
> organism, synthetic construct
- VARIANT, 3
> note, any amino acid residue
- VARIANT, 4
> note, any amino acid residue
- VARIANT, 11
> note, any amino acid residue
Residues:

ARXXDLSGLP XT

Sequence Number (ID): 247
Sequence Name: SHC [SP][TP][VIL]WDTx[LWI]
Length: 8
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..8
> mol_type, protein
> organism, synthetic construct
- VARIANT, 1
> note, S or P
- VARIANT, 2
> note, T or P
- VARIANT, 3
> note, V, I or L
- VARIANT, 7
> note, any amino acid residue
- VARIANT, 8
> note, L, W or I

Residues:
XXXWDTXX

Sequence Number (ID): 248
Sequence Name: SHC PGG[WF][GYA]F
Length: 6
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..6
> mol_type, protein
> organism, synthetic construct
- VARIANT, 4
> note, can be W or F
- VARIANT, 5
> note, can be G, Y or A
Residues:
PGGXXF

Sequence Number (ID): 249
Sequence Name: SHC PDxDD[TAS][TIAS]
Length: 7
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..7
> mol_type, protein
> organism, synthetic construct
- VARIANT, 3
> note, any amino acid residue
- VARIANT, 6
> note, can be T, A or S
- VARIANT, 7
> note, can be T, I , A or S
Residues:
PDXDDXX

Sequence Number (ID): 250
Sequence Name: SHC [MIL]QxxxG[GA][WF]x[AS][FY]
Length: 11
Molecule Type: AA
Features Location/Qualifiers:

```
            - source, 1..7
                  > mol_type, protein
                  > organism, synthetic construct
            - VARIANT, 1
                  > note, can be M I or L
            - VARIANT, 3
                  > note, any amino acid residue
            - VARIANT, 4
                  > note, any amino acid residue
            - VARIANT, 5
                  > note, any amino acid residue
            - VARIANT, 7
                  > note, can be G or A
            - VARIANT, 8
                  > note, can be W or F
            - VARIANT, 10
                  > note, can be A or S
            - VARIANT, 11
                  > note, can be F or Y
      Residues:11
      xQxxxGxxxxx

      Sequence Number (ID): 251
      Sequence Name: SHC Qxxx[GH]xWxG[RK]WGxx[YF]xYG
      Length: 18
      Molecule Type: AA
      Features Location/Qualifiers:
            - source, 1..18
                  > mol_type, protein
                  > organism, synthetic construct
            - VARIANT, 2
                  > note, any amino acid residue
            - VARIANT, 3
                  > note, any amino acid residue
            - VARIANT, 4
                  > note, any amino acid residue
            - VARIANT, 5
                  > note, can be G or H
            - VARIANT, 6
                  > note, any amino acid residue
            - VARIANT, 8
                  > note, any amino acid residue
            - VARIANT, 10
                  > note, can be R or K
            - VARIANT, 13
                  > note, any amino acid residue
            - VARIANT, 14
                  > note, any amino acid residue
            - VARIANT, 15
                  > note, can be Y or F
            - VARIANT, 16
                  > note, any amino acid residue
      Residues:
      QXXXXXWXGX  WGXXXXYG

      Sequence Number (ID): 252
      Sequence Name: SHC Qxx[DN]G[GS][WF][GS]ExxxS
      Length: 13
      Molecule Type: AA
      Features Location/Qualifiers:
            - source, 1..13
```

> mol_type, protein
> organism, synthetic construct
- VARIANT, 2
> note, any amino acid residue
- VARIANT, 3
> note, any amino acid residue
- VARIANT, 4
> note, can be D or N
- VARIANT, 6
> note, can be G or S
- VARIANT, 7
> note, can be W or F
- VARIANT, 8
> note, can be G or S
- VARIANT, 10
> note, any amino acid residue
- VARIANT, 11
> note, any amino acid residue
- VARIANT, 12
> note, any amino acid residue

Residues:
QXXXGXXXEX XXS

Sequence Number (ID): 253
Sequence Name: SHC [STA]xx[SFN][QC]T[AGT]W[AS][LIV]xx[LQ]
Length: 13
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..13
> mol_type, protein
> organism, synthetic construct
- VARIANT, 1
> note, can be S, T or A
- VARIANT, 2
> note, any amino acid residue
- VARIANT, 3
> note, any amino acid residue
- VARIANT, 4
> note, can be S, F, or N
- VARIANT, 5
> note, can be Q or C
- VARIANT, 7
> note, can be A. G or T
- VARIANT, 9
> note, can be A or S
- VARIANT, 10
> note, can be L, I or V
- VARIANT, 13
> note, can be L or Q

Residues:
xxxxxTxWxxxxx

Sequence Number (ID): 254
Sequence Name: meroM [W]xxx[D]xx[ILVMN]

Length: 8
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..8
> mol_type, protein

```
                > organism, synthetic construct
          - VARIANT, 1
                > note, can be W
          - VARIANT, 2
                > note, any amino acid residue
          - VARIANT, 3
                > note, any amino acid residue
          - VARIANT, 4
                > note, any amino acid residue
          - VARIANT, 5
                > note, can be D
          - VARIANT, 6
                > note, any amino acid residue
          - VARIANT, 7
                > note, any amino acid residue
          - VARIANT, 8
                > note, can be I, L, V, M or N
    Residues:
    xxxxxxxx

    Sequence Number (ID): 255
    Sequence Name: meroM PxxAxxxNxxWE
    Length: 12
    Molecule Type: AA
    Features Location/Qualifiers:
          - source, 1..12
                > mol_type, protein
                > organism, synthetic construct
          - VARIANT, 2
                > note, any amino acid residue
          - VARIANT, 3
                > note, 3
          - VARIANT, 5
                > note, any amino acid residue
          - VARIANT, 6
                > note, any amino acid residue
          - VARIANT, 7
                > note, any amino acid residue
          - VARIANT, 9
                > note, any amino acid residue
          - VARIANT, 10
                > note, any amino acid residue
    Residues:
    PXXAXXXNXX  WE

    Sequence Number (ID): 256
    Sequence Name: meroM MxxxFxxMLxxR
    Length: 12
    Molecule Type: AA
    Features Location/Qualifiers:
          - source, 1..12
                > mol_type, protein
                > organism, synthetic construct
          - VARIANT, 2
                > note, any amino acid residue
          - VARIANT, 3
                > note, any amino acid residue
          - VARIANT, 4
                > note, any amino acid residue
          - VARIANT, 6
```

```
                > note, any amino acid residue
          - VARIANT, 7
                > note, any amino acid residue
          - VARIANT, 10
                > note, any amino acid residue
          - VARIANT, 11
                > note, any amino acid residue
    Residues:
    MXXXFXXMLX  XR

    Sequence Number (ID): 257
    Sequence Name: meroM RxxxxGQS
    Length: 8
    Molecule Type: AA
    Features Location/Qualifiers:
          - source, 1..8
                > mol_type, protein
                > organism, synthetic construct
          - VARIANT, 2
                > note, any amino acid residue
          - VARIANT, 3
                > note, any amino acid residue
          - VARIANT, 4
                > note, any amino acid residue
          - VARIANT, 5
                > note, any amino acid residue
    Residues:
    RXXXXGQS

    Sequence Number (ID): 258
    Sequence Name: meroMF [WY]Exx[YFW]
    Length: 5
    Molecule Type: AA
    Features Location/Qualifiers:
          - source, 1..5
                > mol_type, protein
                > organism, synthetic construct
          - VARIANT, 1
                > note, can be W or Y
          - VARIANT, 3
                > note, any amino acid residue
          - VARIANT, 4
                > note, any amino acid residue
          - VARIANT, 5
                > note, can be Y,F or W
    Residues:
    xExxx


    Sequence Number (ID): 259
    Sequence Name: meroMF [DNE]xSYxxP

    Length: 7
    Molecule Type: AA
    Features Location/Qualifiers:
          - source, 1..7
                > mol_type, protein
                > organism, synthetic construct
          - VARIANT, 1
                > note, can be D, N or E
          - VARIANT, 2
```

```
                > note, any amino acid residue
          - VARIANT, 5
                > note, any amino acid residue
          - VARIANT, 6
                > note, any amino acid residue
    Residues:
    xxSYxxP


    Sequence Number (ID): 260
    Sequence Name: meroS GxWxxxW[WG]xxxxY
    Length: 13
    Molecule Type: AA
    Features Location/Qualifiers:
          - source, 1..13
                > mol_type, protein
                > organism, synthetic construct
          - VARIANT, 2
                > note, any amino acid residue
          - VARIANT, 4
                > note, any amino acid residue
          - VARIANT, 5
                > note, any amino acid residue
          - VARIANT, 6
                > note, any amino acid residue
          - VARIANT, 8
                > note, can be W or G
          - VARIANT, 9
                > note, any amino acid residue
          - VARIANT, 10
                > note, any amino acid residue
          - VARIANT, 11
                > note, any amino acid residue
          - VARIANT, 12
                > note, any amino acid residue
    Residues:
    GXWXXXWXXX  XXY

    Sequence Number (ID): 261
    Sequence Name: meroS WxxxHxxV[TSA]
    Length: 9
    Molecule Type: AA
    Features Location/Qualifiers:
          - source, 1..9
                > mol_type, protein
                > organism, synthetic construct
          - VARIANT, 2
                > note, any amino acid residue
          - VARIANT, 3
                > note, any amino acid residue
          - VARIANT, 4
                > note, any amino acid residue
          - VARIANT, 6
                > note, any amino acid residue
          - VARIANT, 7
                > note, any amino acid residue
          - VARIANT, 9
                > note, can be T, S or A
    Residues:
    WxxxHxxVx
```

Sequence Number (ID): 262
Sequence Name: meroS GxWxD[FY]
Length: 6
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..6
> mol_type, protein
> organism, synthetic construct
- VARIANT, 2
> note, any amino acid residue
- VARIANT, 4
> note, any amino acid residue
- VARIANT, 6
> note, can be F or Y
Residues:
GxWxDx

Sequence Number (ID): 263
Sequence Name: enal cleaving GXWXG
Length: 6
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..6
> mol_type, protein
> organism, synthetic construct
- VARIANT, 2
> note, any amino acid residue
- VARIANT, 5
> note, any amino acid residue
Residues:
GXWXG

Sequence Number (ID): 264
Sequence Name: RBS nucl acid
Length: 14
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..14
> mol_type, protein
> organism, synthetic construct
Residues:
AAGGAGGTAA AAAA

Sequence Number (ID): 265
Length: 688
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..688
> mol_type, protein
> organism, synthetic construct
Residues:
MTAKTDGKAQ DDDSRTDLNT DTTDRATEVA GRPADEEDPG GRTPARKESR VPQRRLSSEE 60
LASRATERAT AYLLSLQHQD GWWKGDLETN VTMDAEDLLL RQFLGIRTEE QTRATARWIR 120
SQQREDGTWA TFHGGPGELS TTVEAYVALR LAGDSPQDEH MAIAAKWVRE QGGIASARVF 180
TRIWLALFGW WRWEDLPEVP PEVIYLPKWM PLNIYSFGCW ARQTIVPLTI VSALRPVRPA 240
PFSLEELHTD PQQPNPRRRP APLASWDGVF QRLDKALHVY HRFAPRALRR AAMKSAAQWI 300
IKRQEADGCW GGIQPPAVYS LMALHLLGYP LDHPVMAKGL KAFDDFTVHT PDGMRWLEAC 360
QSPVWDTCLA TIALVDAGVP ADHPALVRAA DWMLGEQVLR RGDWSVRRPS LSPGGWAFEY 420

HNENYPDIDD TAEVVLALKR VEHPDPERVD TAVRRGVEWN LGMQSKNGAW GAFDVDNTSP 480
FPNRLPFCDA GEVIDPPSAD VTAHVVEMLA AVGRADDPRT RRGVSWLLAE QEPDGSWFGR 540
WGVNYVYGTG SVVPALVEAG LPTDHPAIRR AVRWLEQHQN EDGGWGEDLR SYRDPSWAGR 600
GASTASQTAW ALMALLAAGE RGPATERGVR WLAENQREDG GWDEPYFTGT MFPWDFSINY 660
HLYRIVFPVT ALGRYVHGGR PEGAAKGA 688

Sequence Number (ID): 266
Length: 642
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..642
> mol_type, protein
> organism, synthetic construct
Residues:
MVESSPGPPY HRDQLPSSFC HSRSWRAVDW LLKRQSPEGW WWAELESNAT ITAEHLFLTH 60
ILGIGSQELW DEIARQLLEW QNEDGSWSLW YGGPGELSTT VEAYVALKMA GVDPDSPEMR 120
RAREWILRRG GIERARNFTK IWLALLGEWP WEGLPVIPPE VVLLPRWFPI NIYKFASWAR 180
GTMVPLTIVY AYRPTFPIPK HARIDELFPR GRANADLSLP RKRSAWGRFF TFADKVLRVH 240
EHSRWKPLRK RAIKAAEEWI IARQEADGCW GGIQPAWVYS LIALYVLGHD PEGPILKKGI 300
EGLRRYSIEE EGKFRFQSCI SPVWDTALAM IGLQDAGLPR DHPALVKAGK WLLNEQIFVG 360
GDWQVKCKAR PGGWAFEFDN DVYPDTDDTA VVLMAILGTD LPKRAKDFAL SRGLEWLLGM 420
QSRNGGWGAF DRDNTAAFLR EIPFADAGEM IDPPSVDVTA HVVEFLGKMG YRPGFKPLDR 480
ALSYIFREQE PDGPWYGRWG VNYIYGTGYV LPALEAVGFP MDDPRVRKAV DWLLSRQNED 540
GGWGEDVMSY HRRELRGRGP STASQTAWAL LALIAAGGAR SEAVKRGIEY LIRTQNDEGT 600
WNEPYFTGTM FPTDFMIRYH LYRHHFPLMA LGRYRKAVMG DE 642

Sequence Number (ID): 267
Length: 636
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..636
> mol_type, protein
> organism, synthetic construct
Residues:
MTVTVPTGAT DAPDQAKDTL DRAVAHLLGL QDPAGWWKGE LETNVTMDAE DLLFREFLGI 60
RTAEQTEQSA RWIRANQHED GTWGTFRGGP AELSTTIEAY VALKLAGDSV DADHMRAAAA 120
YIRAQGGIAN SRVFTRFWLA FFGWWSWDEL PVMPPELIYL PKWFPLNVYD FACWARQTIV 180
PLTIVSALRP VKPAPFRLDE LRPTGTVRAK PVADDPWSRF FRTLDQALHV YSKRPVKPVR 240
KAAMRRAAQW IIERQEADGG WGGIQPPWVY SLIALHLSGY PLNHPVMAQG IKGLDGFTIW 300
EETDQGTVRR FEACQSPVWD TGLALVALRD AGFAADHPQV VKAAGWLLDE EIRETGDWAV 360
RRPNLEPGGW AFEFANDYYP DTDDTAEVVM ALRYAEHPDT DRMRDVLARA EVWTAGMRSK 420
EGAWGAFDAD NTSTLPLKLP FCDAGAVTDP PSADVTAHIV EMFAESGPEY RDQVVEGVRW 480
LLRNQEADGS WFGRWGANYI YGTGAVVPAL VLAGVDPRHS AVRRAVRWLH SIQNPDGGWG 540
EDLRSYADKG WSGRGDSTAS QTAWALLALL AAGERGAGVE RGVRWLAEHQ NDEGGWDEPQ 600
FTGTMFPGDF YIKYHMYRIV FPVMALGRYV EATGKR 636

Sequence Number (ID): 268
Length: 643
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..643
> mol_type, protein
> organism, synthetic construct
Residues:
MTQGSAATRH EERRPARAGT RAGDRTAEAL DRAVAHLRGL QDERGWWKGE LETNVTMDAE 60
DLLLREFLGI RGEQETAEAA RWIRSQQRDD GTWATFHGGP ADLSTTVEAW VALRLAGDPA 120
DAPHMVAARD FVLASGGLER TRVFTRIWLA LFGEWSWDRL PELPPEMIFL PKWFPLNIYN 180
WACWARQTVV PLTVVGSLRP VRPLPFSVAE LRTGRHARRD RPWSAAGAFQ RLDRVLHAYG 240
RRPVRPLREL ARRRAAEWIL ARQEADGSWG GIQPPWVYSL LALHLLGYPL DHPALRAGLA 300
GLDRFTVRER TPEGWVRRLE ACQSPVWDTG LAMTALLDAG APADDPALVR AAGWLLDEEI 360
RVPGDWAVRR PGLAPGGWAF EFDNDGYPDT DDTAEIVLAL RRTAYPDRDR LRAAIDRGVA 420

WTAGMRSRDG GWAAFDADNT RTLANKLPFC DAGEVIDPPS ADVTAHVVEM LAAEGRAGSA 480
ACRSGVAWLL KAQEPDGSWF GRWGANHVYG TGAAVPALVA AGTDRRAAPI RRAVRWLEAH 540
QNPDGGWGED LRSYDDPALA GRGTSTASQT AWALLALLAA GEESSSTVER GVRFLVDTQR 600
PDGTWDEPQF TGTMFPGDFY INYHLYRLVF PISAIGRYLA GRP 643

Sequence Number (ID): 269
Length: 647
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..647
> mol_type, protein
> organism, synthetic construct
Residues:
MRETQRETAS GRSVAGVSEA VEVAIQRAQD YLLSIQYPEG YWWGELETNV CMAAEYLLLT 60
HFLGAADRRR WDKIVEYLRR QQLPDGTWSI YHGGPSDLNA TVEAYFALKL AGVSPDEPSM 120
AKARQFVLSR GGVPKVRIFT KIWLALFGQW DWRGVPVLPP ELMLLPSWFP INIYEFASWA 180
RATVVPMLII LTRRPVCPIP GEAHIDELYP APREQVDYSL PKSDRLLSWK TLFLTTDKLL 240
RLYERWGWKP FRRRAARAAE EWIVEHQEAD GSWGGIQPPW VYSLIALKVL GYPLDHPVMA 300
KGLEGFEGFA IEDEETFNPQ ACLSPVWDTC LAMNALLDSG LPADHPALVK AGRWMLKEQI 360
LSGGDWQVKN RKGPPGGWAF EFANDLYPDT DDAAEVMIAL LRTRLPEEDE KAQALERGLR 420
WLLSMQSKNG GWGSFDVNNT RRIMTQIPFC DAGAVIDPPS EDVTAHIVEL LGQMGYDKTF 480
RPVRRALAYL RREQEPDGCW FGRWGVNYVY GTGAVLPALE AVGEDMGQES VRRAVRWLIE 540
HQNEDGGWGE TCASYADPGL RGKGASTASQ TAWALLGLLA GGQGASEAAL RGIGYLLETQ 600
AEDGSWDEPE FTGTMFPRDF FINYHLYRDY FPLMALGRYC ELAATTR 647

Sequence Number (ID): 270
Length: 705
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..705
> mol_type, protein
> organism, synthetic construct
Residues:
MTATADGRLD PEYEPEPVAV GDRPPVPERL NGRKAPVAGS APPLGSRRWE PDGSQAVRQS 60
DETGPDEPTP GEPTPAEALA KAAAHLLSLQ SPDGWWKGDL ETNVTMDAED LLLRQFLGIR 120
TEEQTDATAA WIRSQQREDG TWATFHGGPP ELSTTVEAYV ALKLAGDDPK APHMQAAARH 180
IRANGGIAAT RVFTRIWLAL FGWWPWERLP EVPPEIIFLP KWLPLNIYAF GCWARQTIVP 240
LAVVSAHRPV RPAPFALTEL HTDPADPYPL RPLAPPTGWD GLFERLDRLL HVYHRYAVRP 300
LRRLALAQAG RWIVERQEAD GCWGGIQPPA VYSLIALQLL GYDLDHPVMQ AGIASFDRFT 360
VHTEDGRRWM EACQSPVWDT CLATIALRDA GLPADHPAVV RAADWMLGEE ITKRGDWAVK 420
RPHLAPGGWA FEFENDNYPD IDDTAEVVLA LRRVTHPEQP RLDGAVRRAT EWTLGMQSRN 480
GAWGAFDVDN TSTLPNKLPF CDAGEVVDPP SADVTAHMVE MLGELGMAAD PRTRRGLAWL 540
LKNQEADGSW FGRWGTNYIY GTASVLPALV AVGLPTRHPA VRRAVRWLED RQNEDGGWGE 600
DMRSYQDPAS WSGRGESTAS QTAWALISLL AAGEGTDGSR SEAVERGVRW LVRTQLPSGS 660
WDEPQFTGTM FPWDFSINYH LYRLVFPVTA LGRYLHGNSL GGGRK 705

Sequence Number (ID): 271
Length: 637
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..637
> mol_type, protein
> organism, synthetic construct
Residues:
MTTVTPKLAA LAAGPVEDGI ERAVAWLRSR QQDGGYWWAP MDTNVCIEAE YLMLMRFLGH 60
EDPDQIAKMR RHILGTQRPD GSWATCFGGP PDLNCTVEAY FALKLTGSSP DNPPMRAARE 120
VVLSLGGVPA TRVFTRLWLA LFGQYPWSDL PAMPPEAVLL PGWSPINIYA FACWARQAVV 180
PILVVRTLEP VHTVPPDQAI PELYPAARRP QRSTGETGGI LSARNLLSVV DRFLRFYEPR 240
GPKPLRALAL RRCEQYIVTH QEADGSWGGI QPPWVYSLIA LTLLGHDLES PVVRKGIDGL 300
QGYLVEEDGR LWMQACISPI WDTCLAMIGM LDCGVPPDDP AVGKAAAYLV DRQIRKPGDW 360
QAQVSGVEPG GWAFEFANDW FPDTDDSAEV LLALDRARLP DDAGRLDAIE RGDRWLLAMQ 420

SANGGWGAFD KDNTRRLVTQ IPFADAGETI DPPSEDVTAH VIEYLGQRGY DRNFPPVARA 480
ITYLQSTQTI DGSWFGRWGV NHVYGTGAVL VGIAQVGEEP ILPYIQRAVG WLKSVQNDDG 540
GWGESCASYN DPSLKGVGPS TPSQTAWALL GLLAVGERES DAAARGVAYL VNNQRPDGTW 600
DEDQYTGTMF PGDFYLNYRL YRHYWPMMAL GRYRHGA 637

Sequence Number (ID): 272
Length: 647
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..647
> mol_type, protein
> organism, synthetic construct
Residues:
MSAGPLYRDR QRLAQAVREA VERSRDYFLR TQHPDGYWWG ELESNVTMAA EYLMLTHILG 60
VADPERWRKV ANYLRREARP DGTWSIYYGG PPDLNATVES YFALKMAGVP ADDPLLQKAR 120
EFVLSRGGVP RVRVFTKIWL AMLGQWDWRG VPVLPPEFMF LPTWFPINLY EFASWARATI 180
VPILIILDRR PVWPVPEHAR IDELFPVPPS QVDYRLPPPQ RLLSWKGLFW GLDHALRAYG 240
RVAFRPLREA AIRQAVKWII ARQEADGAWS GIQPPWVYSL IALKLLGYPL DHPVMKKGLE 300
AFEGSFRVED DVVYYPQACI SPVWDTALAM IGLLDAGLPP DHEALVRAGR WLLKEQIFTG 360
GDWQFKAKGV EPGGWAFEFD NDIYPDVDDT AEVMMALHRT RLPEERRKAM ALARGLEWVL 420
GMQCRDGGWG AFDKDNTQRL ITHIPFCDAG EVLDYPTEDV TAHVLEMLGL LGYGPDFPPA 480
RRALAFLRRK QDPDGPWWGR WGVNYIYGTG AVLPALRAIG EDMGRPYVRR AVRWLLSCQN 540
TDGGWGESCH SYDDPSWKGR GESTASQTAW ALMALLAALP EAEDDALREA VARGATFLVE 600
TQEEDGTWRE PQFTGTMFPR DFYINYHLYR NYWPLMALGR LQRLLGI 647

Sequence Number (ID): 273
Length: 649
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..649
> mol_type, protein
> organism, synthetic construct
Residues:
MSAKRAPQLQ TVTASPESLD SSIERGVSHL LDLQYDEGYW WEELESNVTI TSEHLFLTHI 60
LGVGNDVEWR KIANYLLNKQ REDGTWAIWY DGPADLSTTI EAYVALKMAG VSPDSPQMER 120
ARSFILSRGG VERARIFTKI WLAVLGEWDW RGTPMMPPEI ILLPKWSPIS IYDFGCWARG 180
TVVPMTIIRL LRPVFPLPRW AHIDELFTSG KSSADLSLPR KDTAWARLFA TLDRILRVYE 240
HSKVKPFRRA AIRKAVRWIE ERQEEDGSWG GIQPPWVYSL IALRSLGRDM DDPVVRKGFE 300
GFYGEKGFAI EEDDTFRMQS CLSPVWDTAL AAVALQDAGL PDDHPALVKA GKWLIDEQIF 360
VGGDWQVRCD AQPGGWTFEF ANDTYPDTDD SAIVMMAIQR IDLDKKRKSR ALDRGLEWLL 420
AMQSANGGWG AFDRNNTKAF LRQIPFADAG EMIDPPSVDV TAHIVEYLGR IGYRKGSNVI 480
DRALAYLKKE QDPDGAWFGR WGVNLTYGIG AVLPALAVIG EDMRLPYVRR AIDWLIAHQN 540
EDGGWGERIE GYVDEDWRGR GPSTSSQTAW ALLGLIAGGE IDHPSTRAGI DYLICSQRDD 600
GGWDEPYFTG TMFPVDFMIN YHLYRDIFPV MALGRYRRAL SKAFTPRTG 649

Sequence Number (ID): 274
Length: 632
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..632
> mol_type, protein
> organism, synthetic construct
Residues:
MTPTQTTTTH PSTDTAPLAD AAAAALRRAR DRLLELQSPE GWWKGELQTN VTMDAEDLLL 60
RQFLGIRTAD DTAEAARWIR SQQRDDGTWA TFHDGPPDLS TTIEAYAALR LAGDPVDAEH 120
MRRAAAFVRE AGGIEASRVF TRIWLALFGQ WPWDDLPVMP PEMVFLPSWF PLNVYDWACW 180
ARQTIVPLTV VGALRPVRPL PFDLAELRTG RRPRRSRDAW GHAFDALDRA LHVYQRRPVR 240
ALRTAALRRA AEWIIARQEA DGSWGGIQPP WVYSLIALNL LGYGLDHPVM KRGLDGLDRF 300
TIRDDKGRRL EACQSPVWDT VLAVTALADA GLPADHPALM NAAEWVLGQE IKGPGDWSVR 360
RPHLPPGGWA FEFDNDIYPD TDDTAEAILA LRRTAHPDAG PAIRRAVRWL TGMGSRDGGY 420
GAFDADNTRT LCTRLPFCDA GAVIDPPSAD VTAHVVEALC GEGLGGSRVV RRAVVWLLNA 480

QEPDGSWFGR WGANHVYGTG SVVPALVAAG VRPDKPAIRR AVAWLEAHQN PDGGWGEDMR 540
SYDDPAWIGR GVSTPSQTAW ALLALLAAGE ERSAAVRDGV AWLVQHQRED GGWDEDYFTG 600
TMFPGDFYIN YHLYRLVFPV SALGRYVRAA AT 632

Sequence Number (ID): 275
Length: 633
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..633
> mol_type, protein
> organism, synthetic construct
Residues:
MTQASVREDA KAALDRAVDY LLSLQDEKGF WKGELETNVT IEAEDLLLRE FLGIRTPDIT 60
AETARWIRAK QRSDGTWATF YDGPPDLSTS VEAYVALKLA GDDPAAPHME KAAAYIRGAG 120
GVERTRVFTR LWLALFGLWP WDDLPTLPPE MIFLPSWFPL NIYDWGCWAR QTVVPLTIVS 180
ALRPVRPIPL SIDEIRTGAP PPPRDPAWTI RGFFQRLDDL LRGYRRVADH GPARLFRRLA 240
MRRAAEWIIA RQEADGSWGG IQPPWVYSLI ALHLLGYPLD HPVLRRGLDG LNGFTIREET 300
ADGAVRRLEA CQSPVWDTAL AVTALRDAGL PADHPRVQAA ARWLVGEEVR VAGDWAVRRP 360
GLPPGGWAFE FANDNYPDTD DTAEVVLALR RVRLEDADQQ ALEAAVRRAT TWVIGMQSTD 420
GGWGAFDADN TRELVLRLPF CDAGAVIDPP SADVTAHIVE MLAALGMRDH PATVAGVRWL 480
LAHQEPDGSW FGRWGANHIY GTGAVVPALI AAGVSPDTPP IRRAIRWLEE HQNPDGGWGE 540
DLRSYTDPAL WVGRGVSTAS QTAWALLALL AAGEEASPAV DRGVRWLVTT QQPDGGWDEP 600
HYTGTMFPGD FYINYHLYRL VFPISALGRY VNR 633

Sequence Number (ID): 276
Length: 633
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..633
> mol_type, protein
> organism, synthetic construct
Residues:
MTQTADPAAR SAAGAAETLD RAVAHLKGLQ RDGGWWKGEL QTNVTMDAED LLMREFLGIR 60
TARETEEAAR WIRSQQRADG TWATFHGGPG DLSTTLEAWV ALRLAGDSPD EPHMRRAAEF 120
VRAGGGVEAS RVFTRIWLAL FGLWSWDDLP NMPPELVLLP SWVPLNVYDW GCWARQTVVP 180
LTVVSTLRPV RPLSFGIDEL RTGVKRGGGL VAPWTWSGAF HYLDKALHLY AKVAVKPVRE 240
FAMRQAAEWI LARQEADGGW GGIQPPWVYS LLALHLLGYS LDHPAMRAGL QGLEGFLIRE 300
ETSEGTVRRL EACQSPVWDT ALAITALLDA GAPADDPHVL KAVDWMLGEE ITVRGDWAVR 360
RPDLDPGGWA FEFANDLYPD TDDTAEVVLG LRRVAHPDRE RLSGALDRAV AWVTGMQSRD 420
GGWGAFDADN TQELTTKLPF CDAGAVIDPP SADVTAHVVE MLAKEGKTGS RECRRGVKWL 480
LDHQEPDGSW FGRWGANYVY GTGAVVPALV EAGVPASATA IRRAVRWLAE HQNADGGWGE 540
DLRSYRDPSW AGRGESTASQ TAWALLALLA AGERDSEVTA RGVRWLAETQ RPDGTWDEPQ 600
FTGTMFPGDF YINYHLYRLV FPVTALGRYL ERS 633

Sequence Number (ID): 277
Length: 720
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..720
> mol_type, protein
> organism, synthetic construct
Residues:
MGIDRMNSLS RLLMKKIFGA EKTSYKPASD TIIGTDTLKR PNRRPEPTAK VDKTIFKTMG 60
NSIDSAIDRA VDWSAARQTP EGYWMARVDT NACMEAQWVL ALWVLNQDDH PIMPGLVKGL 120
LDRQRDDGSW EIYHQAPAGD INTTVECYAA LRTAGLAADD ARLVRARHWI EARGGLRDIR 180
VFTRYWLALI GEWPWRNTPN LTPEVVFIPH EGIPFLSRFS IYNFASWARA TMMPLTVLSA 240
RRFARPLPAD RRLDELFPEG REKYDFGYKR NPPLVSWERF FLGTDRMLHK LQDMGLGLRR 300
ETAIRRVIDW IIDHQDADGV WGGIQPPWIY GLIALHAEGY GPDHPVMRKG LDALDDPRWA 360
FERDGGVIVQ ATVSPVWDTL LTLQAFQETG QDEAQIDRVE KAVDWLMSRE VRTAGDWSVK 420
IKGVEPGGWA FELENAHYPD TDDTAVAIMV LAPYRDHPRF KDKGIGAAVD RAVAWLRAMQ 480
CSNGGWGAFD KDNDDPFLTK IPFCDAGEVL DPPSVDVTAH ILEAFAVAGY GTDDPTVQRA 540

LKFLWDQQES DGSWWGRWGV NYVYGTGAVL PALARIGVDM RDERVLKAAD YLAATQDADG 600
GWGETCASYM DPSLSGKGEA TASQTAWGLM GLLAVGRGQD RKAVERGVGY LLDSQQDGSW 660
HEDQYTGTMF PGYGVGKLID LKNDKLEDDL NQSTELSRGF MINYHMYRHY FPMTALGRAK 720

Sequence Number (ID): 278
Length: 713
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..713
> mol_type, protein
> organism, synthetic construct
Residues:
MGIDRMNSLS RLLMKKIFGA EKTSYKPASD TIIGTDTLKR PNRRPEPTAK VDKTIFKTMG 60
NSLNKSIEKA LEWSRDNQAP EGYWIGGADS NCCMEAEWII AMYFLGLEDD PKMPRVIQAI 120
LNEQRSDGSW EIYYKAPTGD INTTVECYAA LRVAGFDKDH EALVKARKWI FKNGGLRNIR 180
VFTKYWLALI GEWPWEHTSN LPPEIIFLPK WFPLNIYDFA SWARATIVPL AILCSNRPCR 240
PLPPEKRLDE LFPEGRDAFD FSMPSKAKLF SLERLFILVD RLLNKYVNFP IKPLRKTAKK 300
YCLDWIIKHQ DADGVWGGIQ PPFIYSLMAL HTEGYYLDHP ILAAGLRAFD EHWSREKNGA 360
IYINATESIV WDTVLTMLAF LDCGEDPNKS EPLQKALRWL LDKFVDRPGD WQVKVKGVEP 420
GAWAFERANT WYPDVDDTAL VLIVLQRLLE SFPKTAEIDF KMTRATNWTV AMQSKNGGWA 480
AFDKDNTSLV VTKVPFCDAG EALDPPSADV TAHVLEALGL MGWPRSNPVV QRGLDYLLKE 540
QEEDGSWFGR WGVNYIYGTC AALCALKALG MDSSEEVIQR AAKWIVEHQN SDGGWGESCA 600
SYMDDSYRGK GPSTASQTSW AIMALLSVQD SRFDQAILKG LRFLVSTQKE NGTWDEPWYT 660
GTMFPGYGVG DRIDLSQWAD KLEQGAELSR GFMVNYNLYR HYFPLIAMGR ARR 713

Sequence Number (ID): 279
Length: 714
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..714
> mol_type, protein
> organism, synthetic construct
Residues:
MGIDRMNSLS RLLMKKIFGA EKTSYKPASD TIIGTDTLKR PNRRPEPTAK VDKTIFKTMG 60
NSLEQALDSG LHWLAEQQNP EGFWRGILES NCCMEAEWLM AFHILDIDFP HQQALARGIL 120
SKQRTDGAWE TFYDAPSGDI NTTVEAYVAL RISGQSPDAE PMRKARAWIL DHGGLSGIRV 180
FTRYWLALLG EWPWTRTPNL APEIIRMPLW MPFNIYRFSS WARATLMPLA VLSAKRYTRP 240
LPVDQRLDEL FPEGRDRLNY ELPRRGGLWS WDTFFRQLDR VLHGVQTLFG GFPGRQAALS 300
QCLEWIVRHQ DSDGAWGGIQ PPWIYSLMAL RASGYAADHP VMQKSLAALQ QHWSYWRDDK 360
LFIQASESPV WDTLLSLLAM QEAACSLDDS EPMNRALDWL LAHECRERGD WSHFTPQAQP 420
GGWAFERANH HYPDIDDTAV AILVLAGMKT SRRATEVAGP LQRAIDWVLA MQSDNGGWAA 480
FDRNNHTQII TKIPFCDAGE VLDPPSADVT AHVLEALIAA GMPADHPALT RALHYLWNEQ 540
ETNGSWFGRW GVNYIYGLGA VLPALKAAGE DMRQARIGRA VDWLLAHQNA DGGWGEQCSS 600
YMDLSLAGCG VSTASQTAWA MMALLASDDP RARDALHCGS EFLQQTQREG TWEEPYYTGT 660
MFPGYGFGER LESPEDNLAE RFMQGPELAR AFMINYNLYR HYFPLIALAR MSRL 714

Sequence Number (ID): 280
Length: 218
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..218
> mol_type, protein
> organism, Blastocatellia bacterium
Residues:
MKLILTLISG ICWTIVYIDG IRLGFKHRSY AIPFYALALN FAWELLYTYY GFQSTISVQA 60
LVNAVWLVFD AGILITYFKY GRKYFPARLP GSAPAAADNN ATPFIVWSAL TLIAACCVEY 120
AFRKEFGVRV GAGYSAFLQN LLMSVLFINM LVRRGSREGQ SLTIAVGKWL GTLAPTALFG 180
IIGDGGFPNG SFLIVVVGML CSIFDLIYIG LLLKTKPA 218

Sequence Number (ID): 281
Length: 231

Molecule Type: AA
Features Location/Qualifiers:
- source, 1..231
> mol_type, protein
> organism, Streptomyces sp. NRRL F-525
Residues:
MHIALILASG IAWTIVYIEA IRIGFRERTY AMPAVALGLN FAWEWTYAVH NLVFDPSVQG 60
GINLVWGIAD AVIVYTFFRY GRADFPSFVT PRMFAGLSVL LFGMSFAVQW LFLAKFGAED 120
GAGYSAFLQN LLMSALFIAM FVARRGLRGQ SVTIAVAKWL GTLAPTILFG ALQHDGFLLG 180
LGIMCSVLDL VYVWLCVGAR RDGGVGDMGD SGDVPEPGTV TTRADSASME A 231

Sequence Number (ID): 282
Length: 207
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..207
> mol_type, protein
> organism, Arthrobacter sp. I3
Residues:
MILFLTVVSG LAWTVVYVDA IRIGFKDRSY AIPAAALALN FAWEAIYATR SIATGITAQG 60
VFNIAWGLAD VVIVYTFLKF GRSELPDWVT RRLFIGWALL LGVTSFAVQL LFVVQFGWDD 120
AARYAAFLQN LLMSGLFIAM FAARAGTRGQ TLLIAVAKWV GTLAPTIALG WYGNSPLILG 180
VGVLCSVFDL VYIGLLWRAR QPGAARG 207

Sequence Number (ID): 283
Length: 215
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..215
> mol_type, protein
> organism, Deltaproteobacteria bacterium
Residues:
MTIPPEISWS LQIGSGVCWT LVYIFIIKLG FQEKTYGMPI AALCANISWE FIFSFIYPHE 60
PPQNIISVVW FIFDLAIVYQ ALRFGKSEFD REVAAGFFYP TFLLTLTLAF SAVLAITWEF 120
RDWDGKYAAF GQNLMMSILF IAMLLKRKNV RGQSIYIAFF KMVGTLLPSI LFFLSFPASV 180
LLNFLYISIF VFDLIYLVML GVKHRELGIN PWKRV 215

Sequence Number (ID): 284
Length: 218
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..218
> mol_type, protein
> organism, Subtercola sp.
Residues:
MQLFLTIVSG LAWTIVYIEA IRLGFRYKTY AMPVAALGLN IAWETIYGIN GLSGAIDPQT 60
IINLLWAAAD VVIVYTFFRF GRAELPAFVT RPLFIGWGIV IFATAFAVQA LFIAQFGWMD 120
GARYSAFLQN LLMSGLFIAM LVARRGARGQ SMVIAVAKWL GTLAPTILIG VLGDLPFIIG 180
IGLLCSVFDL IYIGMLWWAK KNPAAFAAAP VLSAAVQV 218

Sequence Number (ID): 285
Length: 217
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..217
> mol_type, protein
> organism, Arthrobacter bambusae
Residues:
MFLTILSGVA WTTVYICAIR IGFRDRTYAI PAAALGLNFA WEVIYSVHSL STRLSVQGVI 60
NIAWALADIA IVYTFFAFGR RELPGFVTRP LFIGWAVLLG LASFTVQWLF IAEFDWDPAS 120
RYAAFLQNLL MSGLFIAMFA ARRGVRGQSM VIAVAKWIGT LAPTITFGVL ENSLFILGVG 180

GLCSIFDLTY IGLLLWAKRN PGALTRTRDP GGLPAVP 217

Sequence Number (ID): 286
Length: 217
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..217
> mol_type, protein
> organism, Corynebacterium cyclohexanicum
Residues:
MNGSALGIAL TLFSGVAWTV AYIEAIRLGV RQRTYAMPVA ALGLNLAWEW LYAGVGFAEG 60
GSLQTVVNVA WGLADLAILA TFLRFGYREF SDRLGRTAFY VGAAVLILAC VLVQVLFLAE 120
FGPQLAPGYS AFLQNLLMSG LFIAMHLARG GNRGQSVLLA AAKWLGTLAP TLQFGLLSPS 180
SFILGIGLLC SVFDLAYLGL VVRARRTASV RPEKVSA 217

Sequence Number (ID): 287
Length: 212
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..212
> mol_type, protein
> organism, Runella defluvii
Residues:
MNTSLVLAGI SGICWTIVYI ECIRLGFKQK TYSMPFWALA LNIAWETLHT IIGYREEGLT 60
LQVGFNAVWC FFDIGILYTY FKYGQKYFPD FLSKNVFIAW SVLGLIVSYF IQYYFVEEFG 120
LVKGGSYSAF LQNLAMSILF IAMFVQRRGN EGQSLTLAIN KFIGTLTPTI LVGIVGLPAF 180
GKPNLFILVL GICIAVFDII YIGLLLGKQK ES 212

Sequence Number (ID): 288
Length: 247
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..247
> mol_type, protein
> organism, Pyxidicoccus parkwaysis
Residues:
MLNALFPENY VPGGPYDPLA WFNIVGEVGC VFWVLAYGFI IRQCFRDKSY GLPLVAICMN 60
LAWEFLASWV FPTPVPLWHL FDRVWFFVDL VIVYQLLRYG RGLQTIPEVK RHFFTVVAGT 120
TVLAGIGLYT FFVQYHDLLG LVGAFMINLV MSVSFVFFYF SRRQQGGVGL SWPAALCKLL 180
GTLGTSVECH HVIGMTQPWL GGLSFLHFLC VSIFLFDVLY LALVWKEARA HAPAAGQVRT 240
GAALAVA 247

Sequence Number (ID): 289
Length: 217
Molecule Type: AA
Features Location/Qualifiers:
- source, 1..217
> mol_type, protein
> organism, Methylomonas sp. LL1
Residues:
MSTQESWWFT LSGSASALLW IVAYGLIIRR GFKDRSYGMP FAPLCVNMSY ELIFGFVYPD 60
QPPMNYANQV WFAIDLIIFY QFIRFGKSEF ERLFPRAWFL PAVSLSVLLA FGGVLAVTLE 120
FHDFHGNYTG WGDQLLISIS FIWLLARRGS VAGQSVYIAL SRMLGSIVLI PGQMIQGPAD 180
SVLLGFIYVS FATLDAIYIA LLIRQCRLEG INPWRRL 217

Sequence Number (ID): 290
Length: 2067
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2067
> mol_type, other DNA

> organism, synthetic construct E. coli optimized

Residues:

```
atgacggcaa aaactgacgg caaggcgcag gatgatgatt ctcgtacaga tctgaacact 60
gatacaactg accgggcaac cgaagtcgct gggcgtcctg cagatgaaga agatccgggt 120
ggacgtacac ctgctcggaa ggaatcacgg gttccccagc gccgtttgtc ttccgaagaa 180
cttgcttcac gcgcgaccga acgtgccacc gcctatctct tatcgctgca gcatcaggat 240
ggatggtgga aaggcgacct tgagactaac gttaccatgg acgcggaaga tctgctgctt 300
cggcagtttt taggcattag aactgaggag caaacaagag ccacggcacg ttggattcgc 360
agccagcagc gggaagatgg tacctgggcg acgttccatg gcggtcccgg tgagttgtca 420
acgacagtag aggcttatgt tgccctgcgc ctggcagggg attcccctca agatgagcat 480
atggcaattg cagctaagtg ggttcgtgaa caaggtggca tagcaagcgc cagagttttt 540
acccggatct ggcttgcgct ctttggctgg tggcgttggg aagacctgcc ggaagttccc 600
ccggaagtga tctatctgcc aaagtggatg ccgttaaaca tctatagctt tgggtgttgg 660
gctcgccaaa ctattgtgcc gttaacaatt gtttcagcat tacgtccggt aagaccggcc 720
ccatttcct tggaagagtt gcacactgac cctcagcaac ctaacccccg tcggcgcccg 780
gcgccattag ccagttggga tggtgtcttt cagcgcctgg acaaagctct ccacgtgtat 840
cacagattcg cacctcgggc tcttcgccgc gcggccatga aaagtgccgc tcagtggatt 900
atcaaacgcc aggaagccga cgggtgttgg ggtggaatcc aaccaccggc ggtatactcc 960
cttatggcat tacacttgct ggggtatccc ttagatcatc cggtgatggc gaaaggctta 1020
aaagcattcg acgattttac cgtccacacg cccgatggta tgcgatggct cgaggcttgc 1080
cagagtcctg tctgggatac atgcctggct acgatagcgc tggtagatgc tggtgtgcct 1140
gccgatcatc cagcgttggt ccgagctgcc gattggatgc ttggggaaca ggtcctgcgt 1200
cgaggcgact ggtctgtaag acgcccaagc cttagcccag gcggctgggc gttcgaatac 1260
cataatgaga attacccgga tattgatgat accgcagagg tcgttttagc gttaaaacgc 1320
gttgagcatc cggacccgga acgagtagat acggcagtac gcagaggagt tgaatggaat 1380
ctgggtatgc aatcaaaaaa tggcgcctgg ggtgcctttg atgttgataa tacttcgccg 1440
tttccgaatc gtttgccctt ttgtgacgcc ggagaagtca tcgacccacc gagtgctgat 1500
gtgaccgcac atgtggtaga gatgctggcc gcggttggtc gcgctgatga cccgcggaca 1560
cgtcgcgggg tatcttggtt gctggccgaa caagaaccag acgggtcatg gttcggcaga 1620
tggggcgtga attacgttta tggaactggt tctgtggtgc ctgcactggt ggaagcggga 1680
ttgcctaccg accacccccgc tatacgtcgg gcggttcgtt ggctcgaaca gcatcagaac 1740
gaagacggcg gatggggtga agatctgcgt agttaccgag atcctagctg ggccggacgc 1800
ggtgcaagta ccgcatcgca aacggcctgg gccctgatgg cactgctcgc agctggtgag 1860
cgtggaccgg cgacggagcg aggtgtccgt tggcttgcgg aaaaccagcg cgaagatggc 1920
ggttgggatg agccgtattt tacaggcacc atgttccctt gggatttttc tattaattat 1980
catctgtacc gtatagtgtt cccagttaca gccctgggga gatatgtgca tggtggccgt 2040
ccagaggggg cagcgaaagg agcgtaa 2067
```

Sequence Number (ID): 291

Length: 1929

Molecule Type: DNA

Features Location/Qualifiers:

- source, 1..1929

> mol_type, other DNA

> organism, synthetic construct E. coli optimized

Residues:

```
atggtggaaa gctctccggg gcccccatat catcgagatc aacttccttc ttcgttctgc 60
cattcgcgtt catggagagc tgtcgactgg ctgttaaaac gccagtcccc cgaggggtgg 120
tggtgggctg agctggaatc taatgctacc ataaccgccg agcacctctt cctgacccac 180
atccttggta ttggaagcca agaactgtgg gatgaaattg cgagacaact tctggaatgg 240
cagaatgagg atggatcatg gtcattatgg tatgggggtc cgggggaact gtcaaccaca 300
gtcgaggcgt acgtagcatt gaaaatggcg ggcgttgatc ccgatagtcc ggaaatgaga 360
agagcccgcg agtggatctt acgccgtggt ggtatcgaac gggcaagaaa tttcacgaag 420
atatggctgg cgctgctcgg agagtggccg tgggaaggtt taccagtaat cccgccagag 480
gttgttctcc ttccgcgttg gtttcctatt aacatctaca aatttgcatc ttgggcccgt 540
ggtacgatgg tcccactgac tattgtttat gcctatcgac ctacgtttcc cattccaaaa 600
catgctcgca ttgatgaact gtttccgcgc gggcgtgcta acgccgatct gtcactgcct 660
cgcaaacgca gtgcctgggg acgcttcttt acgtttgccg ataaagtgct tcgcgtccat 720
gaacattcac gttggaagcc actgcgtaaa cgggcgatta aagcagcaga ggaatggatc 780
atagcgcgtc aggaagcaga tggttgttgg ggtgggattc agccagcctg ggtttatagt 840
cttattgcac tgtatgtgct cggccatgat cctgagggtc ctatcctgaa gaaaggtata 900
```

gaaggcctgc gccgatacag tatcgaagaa gaaggcaaat tccgctttca gtcgtgcata 960
agccctgtat gggatactgc cttggccatg attggcctgc aagacgcagg gctgccccgt 1020
gatcatccgg cattggttaa agcgggaaaa tggctgctta atgagcagat ttttgtgggc 1080
ggcgattggc aagttaagtg taaggcacgt cctggcggat gggcttttga atttgacaac 1140
gatgtttatc cggatacaga tgacaccgca gtagtcctca tggcaatttt aggcaccgac 1200
ctgccgaaac gcgcgaaaga tttcgcctta agtcggggat tggaatggct gttaggcatg 1260
cagtccagaa atgggggatg gggtgctttt gaccgggata acaccgctgc ttttttacgc 1320
gagatcccct ttgccgatgc cggggaaatg attgacccgc catctgtgga cgttacggcg 1380
catgtggtgg agttcttagg taagatgggt tatcgcccgg gattcaagcc gcttgacaga 1440
gctttatcct atatttttcg tgaacaagag ccagacggac cttggtatgg tcggtggggc 1500
gtgaactata tctacggcac aggttacgtt cttcctgcac tcgaagctgt gggttttccg 1560
atggatgatc cgcgtgttcg taaagctgta gattggttgt taagccgaca gaatgaagat 1620
gggggctggg gcgaagacgt aatgagctat catcgtcgtg aactgcgcgg ccgtggtccg 1680
agtacagcca gccagacagc atgggctttg ctggcgttaa ttgccgcggg cggtgcgcgg 1740
tccgaggcgg tcaaaagagg gattgaatac ttgatacgaa ctcagaatga tgaaggaaca 1800
tggaatgaac catattttac tggtactatg tttcccaccg acttcatgat acggtaccac 1860
ctgtatcggc accacttccc tttgatggca ttaggccgtt accggaaagc ggtgatgggt 1920
gacgagtaa 1929

Sequence Number (ID): 292
Length: 1911
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1911
> mol_type, other DNA
> organism, synthetic construct E. coli optimized
Residues:
atgaccgtca ccgtgcctac aggtgctaca gacgcaccag atcaagcaaa agatacctta 60
gatcgtgcag tggcccactt actcggactg caagatcctg caggttggtg gaaaggcgaa 120
ttagaaacga acgtaacaat ggacgcggaa gacctcttat ttcgagagtt tctggggata 180
cgcactgccg aacagaccga gcaatccgct cggtggatcc gtgctaatca gcatgaggat 240
ggtacctggg gtacattccg gggcggtccg gcagaactgt ctactaccat agaggcatat 300
gttgcccttа aacttgctgg agacagtgtt gatgcagacc atatgcgcgc cgcagccgca 360
tatattcgtg ctcagggcgg catagccaac tctcgagtct ttactcggtt ttggctggct 420
ttttttggat ggtggagttg ggatgaatta ccggtaatgc cgccggaatt gatctatctg 480
cccaagtggt tcccacttaa tgtttatgac tttgcatgtt gggcgcgtca gacaattgtt 540
ccgctgacaa ttgtgagtgc cttacggccc gtgaaaccag cgcctttccg cttagacgag 600
ttgcgtccca caggtactgt acgtgcgaag cctgtagcag acgatccctg gagccgcttt 660
tttcggactc tcgaccaggc cttgcatgtt tattccaaac gtccggtcaa accagtcaga 720
aaagcagcga tgcgccgcgc tgcgcagtgg attatcgaac gtcaagaagc tgatggcggt 780
tggggcggca tccagccccc atgggtgtac agtttaattg ccctgcacct gagcgggtat 840
ccgctgaacc accctgtgat ggctcagggg attaagggac tggacggatt cacgatttgg 900
gaagagactg atcaagggac tgtgcgcaga tttgaggcct gccagagccc ggtgtgggat 960
acgggtttgg cactcgttgc cctgagagat gcgggatttg cagcggatca tccgcaagtg 1020
gtgaaagcgg cgggctggct gcttgatgag gagattcgtg aaacgggtga ttgggccgtt 1080
cgccggccaa atctggaacc tggtggctgg gcttttgagt ttgccaacga ttattatcct 1140
gacacggatg atacggcgga agtggttatg gctctccggt acgcggaaca tccagatacg 1200
gatcgaatgc gagatgtctt ggcccgcgcc gaagtctgga ccgcaggtat gcgttcgaaa 1260
gagggtgcgt ggggcgcatt cgatgcggat aatacctcaa ccctgcctct gaaacttccc 1320
ttctgtgacg cgggtgccgt taccgatcca ccgtccgccg atgtgaccgc tcacattgtt 1380
gaaatgtttg cggagtcagg gccggagtat agagatcagg tggtagaagg tgtccgctgg 1440
ctgcttcgta atcaggaagc agatggctct tggttcggac gatggggagc taactatatc 1500
tacggcactg gggctgtcgt tcctgcgctc gttcttgcgg gtgtagaccc gcgtcactca 1560
gccgttcgcc gtgcagttcg gtggttgcat agcatccaga atccggacgg cggatggggc 1620
gaagacctgc gctcgtatgc tgataagggt tggtcgggtc gcggcgattc aaccgcctct 1680
cagacggcct gggccctgct tgcgttgctg gcggcagggg aaagaggcgc tggtgtagaa 1740
cgtggggtac gctggttagc agagcatcaa aatgatgaag ggggatggga tgaaccgcag 1800
ttcacaggaa caatgttccc gggtgacttt tacataaaat accatatgta cagaattgtc 1860
tttccagtaa tggcacttgg ccgttatgtt gaagctacag ggaagagata a 1911

Sequence Number (ID): 293

Length: 1932
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1932
> mol_type, other DNA
> organism, synthetic construct E. coli optimized
Residues:

```
atgacacaag  gatccgccgc  tacccgccat  gaagaacgcc  ggcctgctcg  cgctggtacg  60
cgcgcaggcg  atcgtaccgc  agaagcgtta  gatcgtgcag  tcgctcacct  gcgtggctta  120
caggatgaac  gtggttggtg  gaaaggagag  ctggaaacga  atgttactat  ggacgccgaa  180
gacctgctgc  ttcgtgagtt  tctgggtatc  cgtggagaac  aagagacagc  agaagccgct  240
cgctggattc  gcagccaaca  gcgcgatgac  ggtacatggg  cgactttcca  cggcggaccg  300
gctgacctct  caactaccgt  ggaagcctgg  gtggcgctgc  gtttggcagg  cgaccctgca  360
gatgccccac  atatggtggc  tgcgagagat  ttcgttctcg  ccagcggcgg  acttgaacgt  420
acacgcgttt  ttacccgtat  ttggttagct  ctcttcggcg  agtggtcgtg  ggatcgctta  480
cctgaattgc  caccggagat  gatttttctg  ccaaaatggt  ttcccttgaa  catctataac  540
tgggcttgct  gggcgcggca  gaccgtggtc  cctttaacag  tcgtaggatc  actgcggcca  600
gtccgtccat  taccgttttc  tgtagcagag  ctccgcacag  gtcgacatgc  tcgccgcgat  660
cggccgtggt  cggctgcagg  ggcctttcag  agattagacc  gcgtgctgca  tgcatatggt  720
cgtagaccgg  tgcgtccttt  gcgtgagtta  gcccgccgcc  gtgctgccga  gtggatcctt  780
gcgcgacagg  aagctgatgg  cagttggggc  gggatacagc  cgccctgggt  ttactctctg  840
ttggcccttc  acctgcttgg  ttatccatta  gatcatccgg  cattgagagc  cggtctggcg  900
ggtctggacc  ggtttacagt  acgcgagcgg  actccggaag  ggtgggtaag  acgtttggaa  960
gcgtgtcaga  gcccggtctg  ggatacaggc  ctcgcaatga  ccgccttgtt  agatgcgggg  1020
gcgccggccg  acgatcctgc  acttgttcgg  gcagccggtt  ggcttctgga  cgaagaaatt  1080
agagtacctg  gagattgggc  tgttcggcga  cccggactcg  cgccaggggg  ctgggcgttt  1140
gaatttgata  atgatgggta  tccggatact  gatgacacgg  ccgaaatagt  gcttgctctg  1200
cgtcgaaccg  catatccgga  tcgcgatcgg  ttgcgtgccg  cgattgaccg  gggcgtggcg  1260
tggacggcag  gcatgagatc  acgtgatggc  gggtgggccg  cctttgatgc  cgacaatacc  1320
cgtaccctgg  ctaataaact  gccttttttgc  gatgcgggag  aagtaattga  tccgccaagt  1380
gcagatgtca  cggcacatgt  agtggaaatg  cttgcagctg  aaggacgcgc  gggatcggcg  1440
gcatgtagat  caggggttgc  atggctgctg  aaggcacaag  aacccgacgg  gtcttggttt  1500
gggcgttggg  gtgccaatca  tgtttatggt  accggggccg  ctgttcccgc  gctggttgct  1560
gcgggtactg  atcggcgagc  agcaccgatt  cgacgggctg  tccggtggct  ggaagcgcac  1620
cagaaccctg  acggtggctg  gggtgaagat  ctgcgttctt  acgatgaccc  agccctggcc  1680
ggccgtggta  caagcacggc  tagccaaact  gcgtgggcct  tacttgcact  cctcgcggcg  1740
ggtgaggagt  ccagttccac  tgttgagcga  ggtgttcggt  tcctggtgga  tacgcagcga  1800
ccggatggta  cgtgggacga  gccccagttc  accggaacca  tgttccctgg  tgatttttac  1860
atcaactatc  atctctatag  attagtgttc  cccatcagtg  caataggccg  ctacctggcg  1920
ggcagaccat  aa                                                        1932
```

Sequence Number (ID): 294
Length: 1944
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1944
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:

```
atgcgtgaaa  cacagcgcga  gacagccagc  gggagatctg  ttgctggtgt  tagtgaagcc  60
gtggaagtgg  ccatacagcg  agcgcaggat  tatctgctca  gcattcaata  tcccgaaggc  120
tattggtggg  gtgaactgga  gactaatgtt  tgtatggccg  cggaatatct  ccttttaaca  180
cactttctgg  gcgcagcaga  ccggcgtcgg  tgggataaaa  ttgtagagta  cctcagacga  240
cagcaattgc  ctgatggaac  ctggtccatc  tatcacggtg  gaccctcgga  tctgaatgcc  300
acggtggaag  catactttgc  actgaagctg  gcaggcgtta  gcccggatga  accttccatg  360
gccaaagccc  gccaattcgt  actgagtcgc  ggtggcgttc  cgaaagtgcg  tatattcacg  420
aagatttggc  tggctttatt  tggacaatgg  gattggcggg  gagtgccagt  actgccccca  480
gaacttatgc  ttttgccctc  atggttcccg  attaacatct  atgagtttgc  aagctgggca  540
cgcgcaactg  tcgtccctat  gctgataatc  ttgactcgcc  gtccagtctg  cccgatccct  600
ggggaagcgc  atattgatga  gttatatccg  gcaccgcgcg  agcaggtgga  ctacagtctg  660
ccaaaaagtg  accgtctgct  gtcgtggaag  acgcttttc  tgacaaccga  taaacttctt  720
```

agactgtatg agcggtgggg ttggaaaccg tttcgtagac gtgcggctcg ggcggcagag 780
gaatggattg tcgaacatca ggaagccgac ggctcatggg gcggtatcca gccgccttgg 840
gtttattctc tgattgcgct gaaagtactc ggataccctt tggatcatcc agttatggcg 900
aaaggtttag aaggattcga gggtttcgcg attgaggatg aagaaacttt taatccgcag 960
gcatgcctga gtccagtctg ggatacatgt ctggcaatga atgcattatt agattccgga 1020
ttgcccgccg accatccagc gcttgtcaaa gcgggtagat ggatgcttaa ggaacagatc 1080
ctgtcgggtg gtgactggca agtaaaaaac cgcaaaggtc cgccaggcgg ttgggctttc 1140
gagtttgcga acgatcttta ccccgacacg gacgacgccg ctgaagttat gattgcgctg 1200
ctgcgtacaa gattacctga ggaagatgaa aaagcccagg ctttggaacg cggccttcgt 1260
tggctgttat caatgcagag caagaatggc ggctgggggt cctttgatgt taacaataca 1320
agacgcatta tgacccagat accattctgt gacgctggtg cggttattga tccgccctct 1380
gaagatgtga cggcacatat agttgaatta ctcgggcaga tgggatatga taagactttt 1440
cggcctgtac gacgtgcact cgcttatctg cggcgtgagc aggagccgga tggatgctgg 1500
ttcggtcgtt ggggcgtgaa ttatgtatac gggaccgggg cagtgcttcc tgcgttggaa 1560
gccgttggtg aagacatggg gcaggagtca gtgcgccgcg cggtccgttg gttgattgaa 1620
catcagaacg aagacggcgg ttggggcgaa acgtgcgcct cttacgctga tccggggtta 1680
cgtggaaaag gagcaagcac ggcttctcaa accgcttggg ccttactcgg cttgctggcc 1740
ggcggtcaag gggcctcaga agcagcttta cgaggcatcg gttatttgtt agaaacccaa 1800
gccgaggatg ggtcttggga tgagccggag tttaccggta ccatgtttcc gcgagatttt 1860
tttatcaatt atcacctgta ccgtgactat tttcctttaa tggctttagg ccgctattgt 1920
gaactggcgg ctactacccg ctaa 1944

Sequence Number (ID): 295
Length: 2118
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2118
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgacagcga cggctgatgg tcgtctggat cctgaatatg agcctgagcc agttgcggtg 60
ggtgatcggc caccggttcc ggaacgactg aatgggcgca aggcgccggt ggcgggtagt 120
gcacccccgt tgggttctag acggtgggaa ccggatggct cgcaagccgt aagacagtcc 180
gacgaaactg gcccggacga accgaccccc ggcgaaccaa ctcccgcaga ggcccttgca 240
aaagctgcag ctcatctgct ttctctgcaa tcaccggatg gctggtggaa aggcgatctg 300
gagacaaacg ttactatgga cgcggaagat ttgctgcttc gtcagttcct tggtataaga 360
accgaagaac agacagacgc gactgctgct tggatacgat ctcaacagcg cgaagatggc 420
acatgggcca cttttcatgg gggtccgcct gaactcagca cgaccgtcga agcctatgtc 480
gccttgaaac tggcagggga cgaccctaaa gcaccgcata tgcaggcagc ggcacgccat 540
attcgtgcga atgggggcat cgccgctact cgtgtcttta cacgcatttg gttagctctg 600
tttggctggt ggccctggga gcgtttgcca gaggtgccac cggaaattat cttcctgccg 660
aaatggttac cactgaatat ctatgccttt gggtgctggg cccggcagac cattgtacct 720
ctggcggttg tctcggccca ccgtcctgtg cgtcctgcac cttttgccct gacggagtta 780
catacggatc cagcagaccc gtacccccct cgtcctttag ccccaccgac aggttgggat 840
ggtttattcg agcgtttaga tcgcttgctc cacgtctacc acagatatgc ggttcgccct 900
ttgcggcgtt tagctctggc gcaggctggg cgttggatag tggagcgtca ggaagcggac 960
ggctgttggg gcggaatcca accaccggca gtatattcac tgattgccct gcaactgctg 1020
gggtatgacc tggaccaccc ggtgatgcag gcaggaattg cttcatttga tcgctttacc 1080
gttcacaccg aagatggccg gcgttggatg gaagcctgtc agtctccagt ctgggataca 1140
tgtctcgcca caattgcatt acgcgatgcc ggattgcctg ctgaccatcc agccgtagtt 1200
cgtgccgcag attggatgct cggcgaagag attaccaagc gcggggactg ggctgttaaa 1260
cgtccgcatc ttgcgcctgg gggttgggca ttcgagttcg aaaacgacaa ctatcccgat 1320
atagatgaca cggcagaagt agtgctggca ttaagacgag ttacgcatcc ggaacaaccc 1380
cggctggatg gagcggtgcg ccgggctacg gaatggactc tgggtatgca gagtcggaat 1440
ggcgcgtggg gtgcgtttga tgttgataac accagcacac tccctaataa actgcccttt 1500
tgcgatgcgg gagaggttgt agatccaccc tcagcggatg tcaccgcaca catggtggaa 1560
atgttagggg aattgggtat ggctgcggat ccgcgcacac gccgtggact cgcatggctt 1620
ttaaaaaacc aagaggctga cggatcctgg tttggtcgct ggggaaccaa ttacatctat 1680
ggtactgcca gcgttctgcc agcccttgta gcagtcggtt tgccaacgcg ccatccggca 1740
gtgcgcagag ctgttcgatg gttagaagac cgtcagaacg aggatggcgg ttggggcgag 1800
gatatgcgat cataccagga tcctgcatct tggtccggtc ggggtgaatc cactgccagt 1860

cagaccgcct gggctctgat cagtctcctt gctgcggggg aaggcacaga tggaagtcgt 1920
tcggaagcgg ttgaaagagg tgtacgctgg cttgtgcgga ctcagctccc ctcaggcagc 1980
tgggatgagc cgcaattcac cggaacgatg tttccgtggg atttcagcat taattatcat 2040
ctttaccgtt tagtgtttcc tgtgaccgcc ctgggtcgat atctgcatgg aaatagcctg 2100
ggtggcggaa gaaagtaa 2118

Sequence Number (ID): 296
Length: 1914
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1914
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgacaacgg ttactcctaa actggccgct ctcgcagctg gcccggtgga agacggtatc 60
gagagagcgg ttgcttggct gcgctcacgg caacaggatg ggggctattg gtgggcccct 120
atggacacaa atgtctgtat agaagcagag tacttgatgc tgatgcggtt cttgggtcat 180
gaggacccag atcaaattgc aaaaatgaga agacacatat taggcaccca gcgaccggat 240
ggcagttggg ccacctgctt tgggggtccc cctgacctga attgtacggt agaggcctac 300
ttcgcactga aactgactgg ttctagcccg gataatccac caatgcgggc tgcgcgggaa 360
gttgtgctga gccttggtgg ggttccggcg actcgcgtct tcactcgtct gtggttggcg 420
ttatttggcc agtacccctg gtctgatctg cctgcaatgc cgccagaggc cgttctgtta 480
cccggatggt cgccaatcaa tatctatgcc ttcgcgtgct gggcacggca ggcggtggtg 540
cccatacttg tagttcggac cctcgaacct gtacatacgg ttcccccgga tcaagcaatt 600
ccagaattat atccagcagc tcgtcgaccg caaagaagca ccggcgaaac cggtggaatt 660
ttatcagcgc gaaacttgct ttcagtcgtt gacagatttc tgcgtttcta tgaacctcgt 720
ggtcctaaac ctctgcgcgc cctggcttta cgccgctgcg aacagtatat cgtcacccac 780
caagaagcgg atggttcttg gggcgggatt cagccaccat gggtatattc attaattgcg 840
ttgaccctgc tgggacatga tcttgaaagt cctgtcgtgc gcaaaggtat tgatgggtta 900
caggggtacc tggtggaaga ggatggacgt ctttggatgc aagcatgtat ctcccccata 960
tgggacacgt gcttggctat gattggtatg ctggattgtg gcgtcccacc tgatgatccg 1020
gccgtaggca aagcggccgc gtatcttgtt gatcgtcaga ttcgtaagcc aggagattgg 1080
caggcacagg tgtccggtgt cgaaccgggc gggtgggctt ttgagttcgc caatgattgg 1140
tttcccgata ctgacgatag tgccgaagtg ctgctggcac tggaccgggc tcgtttaccg 1200
gacgacgcag gtcgccttga cgctatagaa cgcggcgatc gatggttgtt agctatgcaa 1260
agcgccaatg gcggttgggg cgcctttgat aaggacaaca cacgtcgtct ggtgacacag 1320
attccgtttg cagatgcagg ggaaacgatc gacccgccga gcgaagatgt gacagctcac 1380
gttatagaat acctgggtca acgtggttat gatcgtaatt ttccgccggt tgcgcgtgcc 1440
attacgtatt tgcagtccac ccagactatc gatggtagtt ggtttggccg atggggtgtg 1500
aatcatgtgt atggaacggg tgccgtactc gtcggtattg cccaggttgg agaggaaccg 1560
attttgccgt atatccaacg cgccgtaggc tggctgaagt cggttcagaa cgacgatggc 1620
ggatggggag agtcttgtgc tagttacaac gatccctcac ttaaaggcgt aggcccttcc 1680
acaccttctc agaccgcttg ggcactcctg ggactcctcg cagttggaga gcgtgaatcg 1740
gacgcagcgg cgagaggagt cgcatacctt gtgaacaacc agcgcccgga tggtacttgg 1800
gatgaggacc agtatacagg gacaatgttt ccgggtgatt tttatcttaa ttatcgctta 1860
taccgccatt attggccaat gatggcgctc gggcgttatc gacatggggc gtaa 1914

Sequence Number (ID): 297
Length: 1944
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1944
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgtcggcgg gccccttata ccgggatcgc caaagactgg cccaagccgt acgtgaggca 60
gttgagcgct cgcgcgacta tttcctgcgt acccagcacc cagatgggta ttggtgggga 120
gaattagaat ccaatgtgac catggctgct gagtatctca tgctgacgca tattctggga 180
gttgcagatc cggaacgatg gcggaaggtc gccaattatc ttcgccgtga ggcgcgtccg 240
gacggtactt ggagcatcta ttatggcggc ccgccggatc ttaatgccac agtagagtca 300
tactttgcac tcaaaatggc aggtgttccg gccgatgacc ccctcttaca gaaagctcgc 360

gagttcgtgt tatctcgcgg tggagtgcct cgtgtgcgag tttttactaa aatttggctt 420
gccatgcttg gccagtggga ttggcgtggc gtacctgttc ttccgcccga gtttatgttc 480
ctgcccacat ggtttcctat taacttgtat gaatttgctt cctgggctcg cgccacgatc 540
gtcccaatcc tgattatcct ggatcgccgt ccagtgtggc ccgttcctga acatgcacgg 600
atagacgagt tatttccagt tcccccatcc caggtggatt accggctgcc gccgccacaa 660
cgtctgttga gttggaaagg actgttttgg ggccttgatc atgcattaag agcgtacggc 720
agagtggcct tcagacctct tcgagaggcg gcgatacggc aggcagtaaa gtggataatt 780
gcacgccagg aagcggatgg ggcttggagt ggcattcagc cgccgtgggt ttatagctta 840
attgcattaa agctgctggg atatcctctt gaccacccgg tcatgaagaa aggtctggaa 900
gcgtttgagg gttcatttcg cgtcgaagat gatgtcgtgt attatccaca ggcatgcatc 960
tctccagttt gggatacagc gctggcaatg attgggttac tggatgcagg gttaccgcct 1020
gaccacgaag cgcttgtacg tgcgggtcgt tggttgctga aagaacagat tttcaccggt 1080
ggtgactggc aattcaaggc aaaaggagtg gaaccaggtg ggtgggcgtt tgaatttgat 1140
aatgacatat atcctgatgt ggacgatact gctgaagtta tgatggcact tcatagaacc 1200
cgtttgccag aagaacgacg taaagctatg gctctggccc gtgggcttga atgggttctg 1260
gggatgcaat gccgcgacgg tgggtggggt gcgtttgata aagataacac tcaacgttta 1320
attacccaca tacctttctg tgacgcaggt gaagtcctgg actatcccac cgaagatgtg 1380
acagctcatg ttttggagat gctcggcctg ctgggctatg gccctgactt tccgcccgcc 1440
cgtagagcat tagcattctt gcgccgtaaa caggacccag atggtccttg gtggggccgt 1500
tggggtgtga attacatcta tggcactggg gccgtcctgc cggccctcag agcaattggg 1560
gaagatatgg gacggccgta cgtacgacgc gcggtccggt ggttgctgag ctgtcagaat 1620
acggacggcg gatggggaga atcatgtcat agttacgatg atcctagctg gaaaggccgg 1680
ggcgagtcta cggcctcaca gacggcctgg gctttaatgg ccctgttagc cgctctcccg 1740
gaagcggaag atgatgcgtt gcgtgaagcg gtagcccgtg gtgctacctt tctggttgaa 1800
acacaagagg aagatggaac ctggcgcgag ccgcagttta ctggtacaat gtttccgcgc 1860
gatttctaca tcaactacca tctttatcgg aactattggc cattgatggc tctcggtaga 1920
ttgcagcgac tgctgggtat ctaa 1944

Sequence Number (ID): 298
Length: 1950
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1950
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgtctgcaa aacgggcccc acagcttcaa acggtcacgg ctagcccgga aagtttagat 60
tcctcaattg agcgcggcgt atcgcactta ttagacttac agtacgacga aggttattgg 120
tgggaagagc tggaaagcaa cgttacaatt acctctgagc acctgtttct tacccacatt 180
ctgggtgtgg gcaacgacgt tgaatggcgc aaaatcgcca attacctctt gaataagcaa 240
cgggaagatg gcacttgggc gatatggtac gacggccccg cagatctgtc aacaacaatc 300
gaagcctatg tagcgttaaa aatggccggt gtttctcccg atagtccaca aatggaacgc 360
gcccggtcct ttatcctcag cagaggtgga gtagaaagag ctcgtatctt taccaaaata 420
tggttggctg tgctgggaga atgggattgg cgcggaaccc cgatgatgcc ccctgaaatc 480
atcttattgc ccaaatggtc gcctatatct atctatgatt ttggttgctg ggccagaggc 540
acggtcgtac ctatgacaat catccgcctt ttaagaccag tttttcctct cccacgttgg 600
gcacatattg acgaactttt cactagtgga aaatcgtccg cggatctgtc acttccgaga 660
aaggatacgg cttgggcccg tctgttcgcc actctggatc gcattctgcg agtttacgag 720
cattctaaag ttaaaccgtt ccgccgtgca gccattcgga aagctgttcg ctggattgaa 780
gagcgtcagg aagaggatgg tagttggggc ggcatacaac cgccgtgggt atattccctc 840
attgcgttac gtagtttggg tcgagatatg gacgatccag tggtccgtaa aggcttcgaa 900
ggcttttacg gcgaaaaggg attcgccatt gaggaagatg acacctttcg tatgcagagt 960
tgtttaagcc cggtgtggga tacagcactg gctgctgtcg cgctgcagga tgcaggtctg 1020
ccggatgacc atccggctct ggttaaagcg ggaaaatggc tgattgatga gcagattttt 1080
gtcggtggcg actggcaagt tcggtgtgac gcacagccgg gtggttggac cttcgagttt 1140
gcgaacgaca catatcccga tacggacgat tcagctattg tgatgatggc catccagcgt 1200
attgatctgg acaaaaagcg gaagagccga gccctggatc gtggtcttga atggcttttg 1260
gctatgcaga gtgcgaatgg cggttggggt gcttttgatc gaaataacac taaggcgttc 1320
ctgcgtcaga tacctttcgc cgacgccggg gagatgattg atccgccaag cgtggacgtg 1380
actgcgcata ttgtggaata tctgggccgc atagggtatc gcaagggatc aaatgtcatc 1440
gatcgcgcgc tggcatatct taaaaaagaa caggatcctg atggggcgtg gtttgggcgt 1500

tggggagtaa accttaccta tggtattggg gcagtattgc cagcactggc agttattgga 1560
gaagacatgc gtttacctta tgtgcgccgt gcaatcgact ggttaatagc acatcaaaat 1620
gaagatgggg gatggggcga gcgtatagag gggtacgtcg acgaagattg gagagggcgg 1680
ggtccatcaa cctcgtcaca aactgcttgg gccctcttgg gcttgatcgc aggtggcgaa 1740
attgatcatc cttccacgcg cgcgggtatt gattatctga tatgcagcca gagagatgat 1800
ggtggatggg atgagcccta cttcaccggc actatgtttc ctgtggattt tatgattaat 1860
tatcacttat atcgggacat ctttccggtt atggctctgg ggcgctatcg tcgcgcactc 1920
tctaaagcgt ttacaccgcg aacggggtaa 1950

Sequence Number (ID): 299
Length: 1899
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1899
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgacgccta cccagaccac cactacccat ccgagcactg atacggcccc attagcagat 60
gcggcggctg ccgcattgag aagagctcgg gatcgtctgt tggaattgca gagcccagaa 120
ggctggtgga aaggggagct gcagaccaat gttactatgg atgcagagga tcttctgctc 180
cgtcagttcc tgggtattcg tacggctgat gacaccgcgg aagccgcacg gtggattcgt 240
agtcagcaac gtgatgacgg gacctgggcc acttttcacg acggtccgcc ggatctgagt 300
acaacaattg aagcgtatgc cgcactgcga ttagctggcg accctgtcga tgctgagcat 360
atgcgtagag cagccgcgtt cgtacgtgag gcaggcggaa ttgaagcatc acgcgttttt 420
actcggattt ggctcgcact gtttggccag tggccctggg atgatctccc agtgatgccg 480
ccggagatgg tatttctgcc ttcatggttt ccgttgaatg tttatgattg ggcctgttgg 540
gctcgccaga ccatagtacc acttacagtg gttggcgcac ttagaccggt tcgtcctttg 600
cccttcgatc ttgcggaact gcgcacagga cgtcgcccac gtcggtcccg ggatgcctgg 660
ggtcacgcat tcgacgcatt agaccgagca ttacacgtgt atcaacgccg gccggttcgt 720
gccctgcgta cagcggcgtt acgacgtgct gctgaatgga tcattgcgcg ccaggaagct 780
gatgggtctt ggggcggcat ccagccaccg tgggtctatt ctttgatcgc cctgaatctc 840
ctcggttacg gattagatca tcccgtgatg aaacgcggtc ttgatggcct ggatcgcttt 900
acgattcgcg atgacaaagg tcgtcggctt gaggcctgcc agtctccggt atgggatacc 960
gttctcgcag ttacggctct ggcggacgcc gggttgccag ctgaccatcc tgcgcttatg 1020
aatgctgccg agtgggtcct tggtcaagaa ataaagggtc ccggcgactg gtccgtgcga 1080
cgaccgcact tgccacctgg cggttgggcg tttgaatttg acaatgatat ctacccggat 1140
accgatgaca ccgcagaagc cattctggct ctgcgccgta cagcacatcc cgatgccgga 1200
ccggccatta gacgtgcagt acgctggctt actgggatgg ggtcaagaga tggcggatat 1260
ggggcattcg acgccgacaa tacaagaacc ctttgcacgc gactgccgtt ttgtgatgcg 1320
ggagccgtga tagatccgcc cagtgccgat gtgacggcac acgtggtgga agccttatgt 1380
ggagagggct taggcggttc tagagtcgtc cggcgtgcag tagtgtggct gttaaacgcg 1440
caagagccgg acggttcgtg gttcggccgc tggggtgcca accatgttta tggtacaggg 1500
agcgtcgtcc ctgctttggt ggcggccgga gttcgaccag acaaaccagc gataagacgc 1560
gcagttgcat ggttagaagc gcatcaaaac cctgacggtg gttggggtga agatatgcgg 1620
agctatgatg atccggcttg gatcggccgg ggtgtttcaa cacctagtca gactgcatgg 1680
gcactgttgg cgttactggc ggcgggagaa gagcgctcgg ctgctgtccg tgacggtgta 1740
gcgtggctgg ttcaacatca gcgcgaagac gggggctggg atgaagatta cttcactgga 1800
acaatgtttc ctggggattt ttacatcaac taccatctgt atcgcctggt atttcccgtt 1860
tccgccctgg gacgctatgt gcgtgctgct gcgacgtaa 1899

Sequence Number (ID): 300
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgactcaag ctagtgttag agaggacgcc aaagccgcct tggatcgggc cgtagattat 60
ctgctttcct tacaggacga gaaggggttt tggaaaggtg agctggaaac caatgttacc 120
atagaggcag aagatctttt actgcgagaa tttctgggga ttcgcacccc tgatattacg 180

gcggaaaccg cacgttggat tcgcgcgaaa cagcggtcag atggtacgtg ggccacattc 240
tatgatgggc cgcctgacct gagtacgagt gtggaagcgt atgtagcgtt gaaactggct 300
ggtgatgacc cggccgcccc gcacatggaa aaggcggcag catacatacg cggtgctggc 360
ggggtggagc ggactcgcgt attcactcgt ctgtggttag ccctgtttgg cctgtggccc 420
tgggatgatc tcccgacgtt accacccgaa atgatttttt tgccgagctg gtttcctctg 480
aacatctatg actggggctg ttgggctcgg cagaccgtgg ttccgcttac aattgtcagt 540
gcgttgcgcc ccgtccgccc tattccattg agcattgacg aaatacgcac cggcgcacca 600
ccgccgccgc gggatccggc ttggacaatc agaggtttct tccaacgcct ggatgatctg 660
ctgcgcggct atcgccgcgt tgcggaccac ggaccggcac gcctctttcg tcggctcgcg 720
atgcgccgag ccgcagaatg gattatagca cgacaggaag ctgatggatc atggggcggt 780
atacaaccac cgtgggtcta ttccttaatc gcactgcatc tgttaggata ccctttagac 840
catccggtac tgagaagagg ccttgatggg cttaatggtt ttactatacg agaagaaaca 900
gctgatggag cagtgcgccg tctggaagcc tgccaatctc ctgtgtggga taccgccctc 960
gcagtgacgg ccttgcgtga tgcgggcctg ccggcagatc atcctcgtgt ccaagcagct 1020
gcccgttggc tcgtcggtga agaggtccgt gttgcagggg attgggctgt tcgtcgtccg 1080
ggtctgccac ctgggggctg ggcttttgag tttgctaacg ataattatcc agatactgac 1140
gataccgccg aggtagtctt agcgctgcgc cgggtgcgtc ttgaggacgc ggaccagcag 1200
gcattggaag ccgcagttag acgtgccact acttgggtca tcggaatgca atctacggac 1260
ggcggatggg gcgcttttga tgcggataat acccgagagt tagttcttcg tttaccgttt 1320
tgtgacgcgg gcgcggtaat cgatccacca tctgctgatg taacagccca tattgttgaa 1380
atgctggcag ctctggggat gagagatcat cccgcgacgg ttgccggtgt gagatggctt 1440
cttgcgcatc aggaacctga cggttcgtgg tttgggcggt ggggtgcaaa tcatatctac 1500
ggtactggag ctgttgttcc agctttaatt gccgccgggg tgagccccga tacaccacct 1560
atcagacgag ccattcgttg gctggaagag caccagaatc cagacggtgg ctggggtgag 1620
gatctcagat catataccga cccagcactt tgggttggtc gtggagtttc gacagctagc 1680
cagacggcat gggcactttt ggcgctgctc gctgcgggag aagaagcgtc acccgcagtg 1740
gaccgtggtg tgcgctggtt ggtgaccacc cagcagcccg atggtggatg ggatgaaccg 1800
cactacacag gcacaatgtt cccgggcgat ttctacatca actatcatct ttatcgtctg 1860
gtgttcccta tttccgcgtt aggtcgttac gtaaaccggt aa 1902

Sequence Number (ID): 301
Length: 1902
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..1902
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgactcaaa cagccgatcc tgccgcgcgt agcgcggccg gagccgcaga aacgcttgat 60
agagccgtgg cccacctgaa aggattgcag cgcgacggcg ggtggtggaa aggagaactt 120
cagacgaatg ttacaatgga tgctgaagat cttttgatgc gagagttcct cggcattcgg 180
acggcgcggg agacagaaga ggccgcacgt tggatacgtt ctcaacagcg agccgacgga 240
acgtgggcga cgttccacgg tggacccggt gatctgagca cgacactgga agcgtgggtt 300
gctctgcgct tggcggggga ctctcctgac gaaccgcaca tgcggcgagc ggcagaattt 360
gttcgcgctg ggggcggagt cgaggcatcc agagttttca cccggatatg gctggcgctt 420
tttgggttat ggtcttggga tgaccttcct aatatgccac cggaattagt actgttacct 480
tcatgggtgc cgcttaacgt gtacgattgg ggttgctggg cccgtcagac tgtggttcca 540
ctgaccgtag tctctactct gcggcctgta cgtccactgt catttggtat cgatgagctg 600
cggacgggag tcaaaagagg gggtgggctc gtggcgccgt ggacgtggtc aggggcgttt 660
cattacttag ataaggctct tcatctttat gccaaagttg ccgtaaaacc ggtgcgcgag 720
tttgctatgc gtcaggcagc cgagtggatt cttgctagac aggaagcgga cggcggttgg 780
ggcggtattc agccgccgtg ggtgtactca ctcttagcac tgcatttgct gggctattcc 840
ctggatcatc cagcgatgcg tgccggcctg cagggcctgg aaggtttcct gattcgagaa 900
gagactagtg aaggtactgt acggcgactg gaagcctgtc aaagccccgt ttgggacacc 960
gcactcgcga ttaccgctct cttggatgca ggagccccgg cagacgatcc acatgtgctc 1020
aaagcggtcg actggatgtt aggagaagag atcaccgtcc gcggcgactg ggcagttcgt 1080
cgtccggact tagatccagg tgggtgggcg tttgaatttg caaatgacct gtatcctgat 1140
accgacgata cagccgaagt cgttctgggc ttgcgtagag tggcccaccc agatcgtgag 1200
cgtctgtcag gggcactgga ccgcgctgtg gcgtgggtca ccggtatgca gtcgcgggat 1260
gggggttggg gcgcttttga tgcagataat acacaggaac tgaccacaaa gttaccgttc 1320
tgcgacgccg gtgctgtcat cgatccgccg agcgctgacg ttacggctca cgttgtggag 1380

atgcttgcta aagaaggcaa gactggcagt cgtgaatgtc gccgcggagt aaaatggttg 1440
ttggatcatc aagaacccga tggcagttgg tttggtcgct ggggagcgaa ctatgtttat 1500
gggactggtg ccgtagtacc ggcattagtg gaagccggtg ttcccgcttc ggctacagca 1560
attcgccgcg cggtgagatg gcttgctgag catcagaacg cagatggcgg ctggggtgaa 1620
gatctgagat cttatcgaga tccaagttgg gcaggtcgcg gtgagtccac tgcatcgcaa 1680
accgcttggg ccctgttagc attgttagcg gcgggtgaac gggatagtga agtcacagct 1740
cgtggtgttc gctggttagc agaaacacaa cgccctgatg gcacctggga tgaaccccag 1800
tttactggga ccatgtttcc tggagatttc tatatcaatt accatctcta cagactggta 1860
tttcccgtta ccgcattagg ccgttatctg gagcgtagct aa 1902

Sequence Number (ID): 302
Length: 2163
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2163
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgggtattg atcgtatgaa ctcgctgagt cgtctgttga tgaaaaagat tttcggagcc 60
gaaaaaacat cctataagcc agcgagcgat acgattatag gtacagacac tttaaaacgt 120
cccaatcggc gacctgagcc caccgctaag gtagataaga ccatttttaa aacgatgggc 180
aatagtatag attccgctat tgatagagcc gtcgattgga gcgcggcccg tcaaacgccg 240
gaaggttact ggatggcccg tgtggatact aatgcttgta tggaagcaca gtgggttctg 300
gcactgtggg tgctgaacca ggatgatcat ccgattatgc cgggcctcgt taaaggcctg 360
ctggaccgac aacgcgatga cggttcatgg gagatctatc accaggcccc cgccggcgac 420
atcaatacga ccgttgaatg ctatgctgca ctgcgaactg caggactggc tgccgatgac 480
gccagactcg tccgggcacg tcattggatc gaggcgcggg gtggtttacg cgatattcgc 540
gtgttcactc gttactggtt ggcacttatt ggcgagtggc cttggagaaa cacaccaaat 600
ttgacaccgg aagtggtgtt tattcctcac gaggggattc cttttctgtc tcgctttagt 660
atctacaatt tcgcgtcgtg ggctcgtgcc accatgatgc cgctcacagt gctgtctgct 720
cgcagattcg cacgtccccc tcctgcagac cgccggctgg acgaattatt tccggaaggc 780
cgtgaaaagt atgattttgg ttataaacga aatccaccct tagtttcttg ggaacgtttc 840
tttctgggta cggatcgcat gttacataaa cttcaggata tgggactggg gctgcggaga 900
gagactgcga tacgtcgggt aattgattgg attatcgacc atcaagatgc ggacggggtt 960
tggggcggaa tccagccacc gtggatctac gggctgatcg cacttcacgc cgaaggctac 1020
gggccagatc accctgtcat gcgtaaagga ttggatgctt tagacgaccc acgctgggca 1080
tttgagcgcg acgggggcgt gattgtgcaa gctacggtgt caccggtttg ggataccctg 1140
ttgacactcc aggcatttca agaaacagga caggacgaag cccagataga tcgcgtagaa 1200
aaagcagtcg actggctcat gagccgtgaa gtgcgcacag ccggtgattg gtcagtcaaa 1260
ataaaagggg ttgagccggg cggatgggcg tttgaactgg aaaacgcaca ttatccggat 1320
acagatgata ccgcggttgc aattatggtg cttgcacctt atcgggacca tcctcgcttt 1380
aaagataaag gaataggcgc cgcggttgat cgggcggtag cctggcttcg cgcaatgcag 1440
tgttctaatg gcggctgggg tgcgtttgac aaagataatg acgatccttt cctcactaaa 1500
atcccatttt gtgacgctgg agaagtttta gatccaccct cagttgatgt caccgcgcac 1560
attttggaag cgtttgctgt ggcgggctat ggcaccgatg acccgaccgt tcagcgggcc 1620
ctgaagttcc tttgggatca gcaagaaagt gatggtagct ggtggggtcg ctggggtgtt 1680
aactatgtat atggtaccgg ggccgtactg ccggcattag ctagaatagg tgtagatatg 1740
cgagatgaac gtgtgctgaa agccgcagat tatttggcgg ccacgcaaga tgcggatggt 1800
ggctggggag aaacgtgcgc aagttatatg gatccgtctt taagcgggaa aggcgaagcg 1860
actgcttccc aaacggcttg gggcctgatg ggactcctgg cggtaggccg tggtcaggac 1920
agaaaagccg ttgagcgcgg ggtcggttat cttttggact cccagcagga tggatcgtgg 1980
catgaggacc agtataccgg tacaatgttt ccaggttacg gggtcggtaa acttattgat 2040
cttaaaaacg ataagttaga ggatgactta aaccagagca ctgagttatc acgtggtttc 2100
atgatcaatt accatatgta ccgacattat ttcccgatga ccgctctggg gagagctaag 2160
taa 2163

Sequence Number (ID): 303
Length: 2142
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2142

> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized

Residues:

```
atgggcatcg  accgaatgaa  tagtctgtcc  agactgctga  tgaaaaaaat  ttttggggca  60
gagaagacca  gctataaacc  ggcatccgat  accattattg  gaactgatac  tctgaagcga  120
cctaaccgca  gacccgaacc  tacagccaaa  gttgataaaa  ccattttcaa  gactatgggt  180
aactcgttga  ataaaagtat  tgaaaaagct  ctggaatggt  cacgcgataa  tcaggccccg  240
gaaggttatt  ggattggcgg  cgctgattct  aattgttgta  tggaagctga  atggattatc  300
gcgatgtact  tccttggttt  agaggacgat  ccgaaaatgc  ctcgcgttat  ccaggctatt  360
ctcaatgaac  agcggagcga  cggttcgtgg  gaaatatact  ataaagcgcc  gacaggagac  420
attaacacca  cagtggagtg  ctatgcagcc  ttgcgtgtag  caggatttga  taaagatcac  480
gaagctcttg  tgaaagcacg  aaaatggatt  tttaaaaacg  gtggtctgcg  taatatacgg  540
gtgtttacga  aatactggtt  agccttaatc  ggtgagtggc  cgtgggaaca  tacttcaaac  600
ctgccacccg  aaatcatttt  tctgccgaaa  tggttcccac  ttaatatcta  tgatttcgct  660
tcttgggcgc  gggcgaccat  agttccgttg  gcgattctgt  gtagtaatcg  cccttgtcgt  720
cccctgccgc  cggagaaacg  tttagacgag  cttttcccgg  aaggccggga  cgccttcgat  780
tttagtatgc  cctcgaaagc  aaaacttttt  agcctggagc  ggctgttcat  tttagttgat  840
cgcttactca  ataagtacgt  taattttcct  atcaagccgt  tgcgtaagac  agctaaaaag  900
tattgtctgg  actggataat  caaacaccaa  gacgcggatg  gtgtttgggg  tggcattcag  960
ccaccgttta  tctattcttt  gatggcactc  cataccgagg  gttattatct  ggatcatcca  1020
atacttgccg  caggactccg  cgctttcgat  gaacattggt  ctagagagaa  gaatggtgcc  1080
atctacatta  atgcgacgga  gagcattgtc  tgggatacgg  tattgacaat  gttagcattt  1140
ttagattgcg  gagaggatcc  aaacaaaagc  gaacctctgc  aaaaagcact  gcgttggctg  1200
ctcgacaagt  ttgtagatcg  tccgggggat  tggcaagtga  aagtcaaagg  tgtcgaacca  1260
ggtgcatggg  cgtttgagcg  cgcgaatact  tggtaccccg  atgttgacga  tactgcactc  1320
gtgcttatag  tgttacagag  attattggaa  tcctttccga  aaaccgctga  aattgacttt  1380
aaaatgaccc  gcgcgacaaa  ttggaccgta  gctatgcaga  gcaaaaatgg  cgggtgggca  1440
gctttcgata  aggataacac  gagtctggtg  gttactaaag  tgccattttg  cgatgcaggc  1500
gaagcactgg  atcccccgtc  cgctgacgtc  acggcgcatg  tcctggaagc  cctcgggttg  1560
atgggctggc  cacgttctaa  ccctgtagtt  caaagaggtc  tggattacct  tcttaaggaa  1620
caggaagaag  acggttcttg  gtttggccgg  tggggcgtta  attatatcta  tggtacatgc  1680
gcggccctgt  gtgcgttaaa  ggctcttgga  atggatagta  gcgaagaggt  gatccaacgc  1740
gccgccaaat  ggatcgtaga  acaccaaaac  tcggacgggg  gctggggaga  gtcgtgcgcc  1800
tcatatatgg  acgattcata  ccgtgggaaa  ggacctagta  cggcatccca  gacatcatgg  1860
gccattatgg  ccttactgag  cgtgcaagac  tctcgctttg  atcaggccat  attgaaaggg  1920
ctgcgttttc  ttgtttccac  gcagaaagaa  aatgggacct  gggatgaacc  ttggtatacc  1980
ggaacaatgt  tccctggcta  tggtgtcggg  gatcgcattg  acctgtcaca  gtgggcagat  2040
aagttagaac  agggcgccga  gctttcacgt  ggattcatgg  taaactacaa  cctttatcga  2100
cattattttc  cactgatagc  gatgggcaga  gcgcgccgtt  aa                      2142
```

Sequence Number (ID): 304
Length: 2145
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..2145
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized

Residues:

```
atggggattg  atcgtatgaa  tagtttaagc  agattactga  tgaaaaagat  ttttggcgcg  60
gaaaagacta  gttataaacc  ggcctctgat  acgataattg  gtacagatac  gctcaagcga  120
ccaaaccgtc  gtccggaacc  gactgcgaaa  gtggataaaa  ctatttttaa  aactatggga  180
aatagtctgg  agcaggctct  tgatagtggc  ctgcattggc  tggccgagca  gcaaaatccc  240
gaaggtttct  ggcgtggcat  cctggagtcg  aattgctgta  tggaagctga  atggctgatg  300
gccttccaca  tcctggatat  cgattttccc  catcagcagg  ccctcgcacg  tggcattctt  360
tctaagcaac  gtacggacgg  tgcttgggaa  acttttacg  atgctccatc  aggggacata  420
aacactaccg  tagaagccta  tgtggcgctc  cgtatatcgg  gacaatcccc  tgatgccgaa  480
ccaatgcgca  aagcccgtgc  gtggattctc  gaccatggcg  gtctgtctgg  cattcgggtc  540
ttcacacgct  attggcttgc  cctgctgggc  gagtggcctt  ggacccgcac  ccctaatctc  600
gcaccggaaa  ttattcgtat  gccattatgg  atgccgttta  atatctaccg  attttcatcc  660
tgggcccgcg  caaccttaat  gccgttagcc  gtcctgtccg  caaaacgtta  tacccggccg  720
ctgcccgtag  atcagcgcct  ggatgaattg  ttccctgaag  gtcgggacag  actgaactac  780
```

gaattgccac gccggggcgg attgtggtca tgggacacgt ttttccgaca gcttgatcgt 840
gtgctgcatg gagtgcaaac cctgttcggg ggctttccag gacgccaagc ggctctgagc 900
cagtgccttg agtggatcgt tagacaccaa gatagtgacg gcgcgtgggg cgggattcag 960
ccgccttgga tctattctct gatggccttg agagcgagtg gatatgcggc cgatcacccg 1020
gttatgcaaa aatcgctggc cgcccttcag cagcattggt cgtattggcg agatgataaa 1080
ctttttatcc aggccagcga gtcccccgtc tgggatacac ttctgagctt attggctatg 1140
caagaggcag cttgtagctt agatgactcc gaacctatga atcgtgcttt ggactggctt 1200
ttggcacatg aatgccgcga acggggtgat tggtcacact ttacccctca ggcacagccc 1260
ggcggatggg ctttcgagcg cgcaaatcac cactatcccg acattgacga cacagcagtt 1320
gcaatactcg tccttgcggg catgaagact tctcgccgcg cgaccgaagt agcgggtcct 1380
ctgcaacggg ctatagattg ggtgttagca atgcagtcag ataatggtgg ttgggcagca 1440
ttcgatagaa acaatcatac acagattatt acgaaaattc cgttttgtga cgcgggagag 1500
gttttggacc cgccaagcgc tgatgttaca gcccatgttt tggaagcact gatcgcagct 1560
gggatgcctg ctgatcatcc tgcattgaca cgcgcactgc attatctttg gaacgaacag 1620
gaaaccaatg gctcgtggtt tggtcgttgg ggcgtaaact atatctatgg gctgggtgct 1680
gttctgccgg cactcaaagc ggcgggtgag gacatgcgcc aagcgcgaat tggtagagcg 1740
gtggattggc tgctggctca ccagaacgcg gacgggggat ggggtgaaca atgttcttcc 1800
tacatggact taagcctcgc aggttgcgga gtgtcaacgg cctctcagac cgcatgggcg 1860
atgatggcgt tattagcctc agatgatccc cgcgcccgtg atgcattaca ttgtgggagc 1920
gagtttttgc agcaaacaca gcgagagggg acatgggaag aaccatacta cacgggtacc 1980
atgttcccgg gctatggttt tggtgaacgg ttagaaagtc cagaagataa cctggcggag 2040
cgttttatgc aggggccgga gctggctcgt gcctttatga taaactataa tttgtatcgg 2100
cattactttc cgttgatcgc tttagcccgc atgtcaagac tttaa 2145

Sequence Number (ID): 305
Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgaaactga tcctgacgct gatctctggt atctgctgga ccattgttta cattgacggc 60
atccgtctgg gcttcaaaca ccgtagctac gcaatcccgt tttatgcgct ggcgctgaac 120
ttcgcttggg agctcctgta cacctattac ggcttccagt ccaccatctc tgtgcaggcg 180
ctggttaacg ccgtatggct ggtttttgat gccggtatcc tcattaccta ctttaaatac 240
ggccgtaaat atttcccggc tcgtctgccg ggcagcgcgc cggcggctgc ggacaataac 300
gcaactccgt ttattgtatg gtctgcgctg accctgatcg cagcgtgctg tgtagagtac 360
gctttccgta aagagttcgg cgtgcgtgta ggcgctggtt actccgcatt tctccagaac 420
ctcctgatgt ccgttctgtt catcaacatg ctcgttcgcc gtggtagccg cgaaggccag 480
tccctgacca ttgcggttgg taaatggctg ggtactctgg ccccaaccgc actgttcggt 540
attatcggcg atggcggttt cccgaatggt tctttcctga tcgttgtcgt gggtatgctg 600
tgctctattt ttgacctcat ctacattggc ctcctgctca aaactaaacc ggcataa 657

Sequence Number (ID): 306
Length: 696
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..696
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgcatatcg cactgattct cgcgtctggt atcgcgtgga ccatcgtgta catcgaagcg 60
attcgtattg gcttccgtga acgcacctac gcgatgcctg ctgtagccct gggtctgaac 120
ttcgcctggg aatggaccta cgcagttcac aacctggtat tcgacccgtc tgtgcagggt 180
ggcatcaacc tggtttgggg cattgccgat gcggtcattg tctacacttt tttccgctac 240
ggtcgcgcgg acttcccgag ctttgttacc ccgcgtatgt tcgcaggtct gtctgtgctc 300
ctgtttggca tgtctttcgc tgtacaatgg ctgttcctcg caaaattcgg cgcggaggat 360
ggtgccggct acagcgcgtt cctgcagaac ctcctgatgt ccgcactgtt tatcgcgatg 420
ttcgtggcgc gccgtggtct gcgtggtcag tctgtaacta tcgccgtggc taaatggctg 480
ggtactctgg cgccaaccat cctgttcggt gcgctgcagc atgacggctt cctcctgggc 540

ctgggtatta tgtgttccgt gctggatctg gtttatgttt ggctgtgcgt tggcgctcgc 600
cgtgatggcg gtgtaggcga catgggtgat agcggcgacg ttccagaacc aggcaccgtg 660
acgacccgtg ccgacagcgc ttctatggaa gcttaa 696

Sequence Number (ID): 307
Length: 624
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..624
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgatcctgt ttctgaccgt tgtcagcggt ctggcatgga ccgttgtgta cgtggacgct 60
atccgcattg gttttaaaga tcgttcctac gccatcccgg ctgccgcgct ggcactgaac 120
ttcgcctggg aggctattta cgcaacccgt tccattgcga ccggtatcac ggcacagggt 180
gttttcaaca tcgcgtgggg tctggctgac gttgtaatcg tgtacacctt tctgaaattc 240
ggtcgttccg aactgccgga ctgggtgacc cgtcgcctgt tcattggctg ggctctgctc 300
ctgggcgtaa cctccttcgc agtccagctc ctgttcgttg tacagtttgg ttgggacgat 360
gcagctcgtt acgctgcatt cctgcagaac ctcctgatga gcggcctgtt cattgctatg 420
ttcgccgcgc gtgctggtac gcgtggtcaa acgctcctga tcgccgtggc taagtgggtt 480
ggcaccctcg ctccgacgat tgcactgggc tggtatggca actccccgct gatcctgggt 540
gtcggtgtgc tgtgttccgt gttcgatctg gtgtatattg gcctcctgtg gcgtgctcgt 600
caaccgggtg cggctcgtgg ctaa 624

Sequence Number (ID): 308
Length: 648
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..648
> mol_type, other DNA
> organism, synthetic construct
Residues:
atgaccattc ctccggagat ttcctggtcc ctgcagatcg gttctggcgt gtgttggacg 60
ctggtctata tcttcattat caaactgggt tttcaggaaa aaacttacgg catgccgatt 120
gctgcactgt gtgcgaacat ctcctgggag ttcattttta gctttatcta cccgcatgaa 180
cctccacaga acattatcag cgtcgtttgg ttcatcttcg acctggcaat cgtgtaccag 240
gctctgcgtt tcggcaaaag cgagttcgac cgcgaggtag ccgcgggctt tttctacccg 300
accttcctcc tgaccctgac cctggcgttt agcgcagttc tggccattac ttgggagttt 360
cgcgactggg acggtaaata cgcggctttt ggccaaaacc tgatgatgtc cattctgttc 420
atcgctatgc tgctcaagcg taaaaacgtg cgtggccagt ccatttatat cgcgttcttt 480
aaaatggtcg gcacgctcct gccgagcatc ctgtttttcc tgtctttccc ggctagcgta 540
ctcctgaatt tcctgtatat cagcattttt gtgtttgatc tgatttacct ggtgatgctg 600
ggcgttaaac accgtgaact gggcatcaac ccgtggaaac gtgtataa 648

Sequence Number (ID): 309
Length: 657
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..657
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgcagctct ttctcaccat cgtaagcggt ctggcatgga ccatcgtata tatcgaagcg 60
atccgtctgg gctttcgcta caaaacgtac gctatgccgg tagctgcgct cggtctgaac 120
atcgcatggg aaactatcta tggtatcaac ggtctgtccg gcgccattga tccgcagacc 180
attatcaatc tcctgtgggc tgcggctgac gtggttattg tttacacctt ctttcgcttt 240
ggtcgcgcgg agctgcctgc ttttgttacc cgtccgctgt tcattggttg gggtattgta 300
atctttgcta ccgccttcgc ggtgcaggcc ctgtttatcg cgcagttcgg ctggatggac 360
ggtgctcgtt actccgcttt cctgcagaat ctcctgatgt ccggcctgtt tatcgctatg 420
ctggtcgctc gtcgcggtgc acgtggtcag tctatggtaa ttgcagtggc gaaatggctg 480
ggcaccctcg ctccgactat cctgatcggt gtactgggtg acctcccgtt tattatcggt 540

atcggcctcc tgtgctccgt tttcgatctg atctacattg gcatgctgtg gtgggcgaaa 600
aagaacccgg cggccttcgc cgctgcgccg gtcctgagcg ccgcagtcca ggtgtaa 657

Sequence Number (ID): 310
Length: 654
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..654
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgttcctga ccattctgtc cggtgtggcg tggacgactg tctatatctg cgcaattcgt 60
atcggcttcc gcgaccgtac ctacgctatt ccggctgcgg ctctgggcct gaatttcgcg 120
tgggaagtca tctactctgt gcactctctg tccacccgtc tgtccgtcca gggcgtcatc 180
aacatcgctt gggcactggc ggatattgcg atcgtttata ctttttttcgc tttcggccgt 240
cgcgaactgc cgggttttgt tacccgtccg ctgttcatcg gctgggcggt actcctgggt 300
ctggcctcct tcactgtgca gtggctgttc attgcggagt tcgattggga cccggcttct 360
cgttacgcag cttttctgca gaacctcctg atgtctggtc tgtttattgc gatgttcgcg 420
gcacgtcgcg gcgttcgtgg ccagtccatg gtgattgcag tggctaaatg gatcggtacc 480
ctggcaccga ccattacttt cggtgtcctc gaaaactccc tgttcatcct gggtgttggt 540
ggcctgtgtt ctatttttga cctgacctat atcggtctgc tcctgtgggc aaaacgtaac 600
ccgggtgcac tgacccgtac tcgtgacccg ggtggcctgc cggcggtacc ataa 654

Sequence Number (ID): 311
Length: 654
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..654
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgaacggta gcgcactggg tatcgcactg accctgttct ctggtgtggc ttggaccgtg 60
gcttatatcg aagcgattcg cctgggcgtt cgtcagcgta cctatgccat gcctgttgcg 120
gccctgggcc tgaacctggc ctgggaatgg ctgtacgccg gcgttggttt tgcagaaggt 180
ggctccctgc agacggtggt aaacgtagca tggggtctgg ccgatctggc catcctggct 240
accttcctgc gtttcggcta ccgtgaattt tctgaccgcc tgggtcgtac cgctttctat 300
gtaggtgctg cagtactgat tctggcgtgt gttctggttc aggtactgtt tctcgctgag 360
ttcggccctc agctggcccc gggttattcc gctttcctgc agaacctcct gatgagcggt 420
ctgttcattg ctatgcatct ggcacgtggt ggcaaccgtg gtcagtctgt tctcctggca 480
gctgccaaat ggctgggtac cctggcgcct acgctgcagt tcggcctcct gagcccgtct 540
agcttcatcc tgggcattgg cctcctgtgc tctgtgttcg acctggcgta tctgggcctg 600
gttgtccgcg ctcgtcgcac cgcatctgtt cgcccggaaa aagtctctgc ataa 654

Sequence Number (ID): 312
Length: 639
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..639
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgaacacta gcctggttct ggcaggtatt tccggtatct gttggacgat tgtgtacatt 60
gaatgtatcc gtctgggttt taagcagaaa acctactcca tgccgttctg ggcgctggcg 120
ctgaacattg catgggaaac tctgcacacc atcattggct atcgtgagga aggtctgact 180
ctgcaagtgg gtttcaacgc ggtttggtgc ttctttgaca ttggtatcct gtatacctac 240
ttcaaatacg gtcagaagta tttcccggat tttctctcta aaaatgtttt catcgcgtgg 300
tctgttctgg gtctgatcgt ttcttacttc atccagtact attttgttga agagttcggc 360
ctcgtaaaag gcggtagcta ctccgcgttt ctgcagaacc tggcaatgag catcctgttc 420
atcgcaatgt tcgtccaacg ccgtggtaac gaaggtcagt ccctgactct ggcgatcaat 480
aaattcattg gtaccctgac ccctaccatc ctggtgggta tcgtgggcct gccggctttt 540
ggcaaaccga acctgttcat cctggtgctg ggcatctgta tcgctgtttt tgacattatc 600

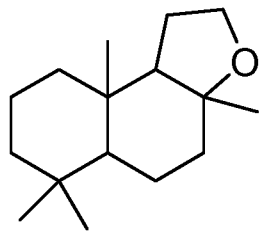
tacatcggcc tgctcctggg taaacagaag gagtcctaa 639

Sequence Number (ID): 313
Length: 744
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..744
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgctgaacg cgctcttccc agagaactac gtcccaggtg gcccgtacga tcctctggct 60
tggttcaaca tcgtgggtga agtgggttgc gtgttctggg tgctcgcgta cggcttcatt 120
atccgtcagt gttttcgtga caaatcctac ggcctgccgc tggtggcaat ttgtatgaac 180
ctggcatggg agtttctcgc aagctgggta tttccgaccc cggtgccgct gtggcacctg 240
ttcgaccgtg tatggttttt cgtggatctg gtcattgtat accagctcct gcgttatggc 300
cgcggtctgc agaccattcc tgaggtgaaa cgtcacttct ttaccgtggt cgcgggtacg 360
actgtactgg cgggcatcgg cctgtacacc tttttcgttc agtaccatga tctcctgggt 420
ctggttggcg cattcatgat caacctggtt atgtccgtta gctttgtctt tttctacttc 480
tctcgccgtc agcaaggcgg tgtgggcctg tcctggccgg cggccctgtg taaactcctg 540
ggtaccctgg gtacttccgt tgaatgtcat cacgttatcg gcatgactca accgtggctg 600
ggcggtctgt ccttcctgca cttcctgtgt gttagcattt tcctgtttga cgttctgtac 660
ctggcgctgg tttggaagga agcgcgtgca cacgctccgg cagctggcca ggttcgtact 720
ggtgctgcgc tcgcggttgc ctaa 744

Sequence Number (ID): 314
Length: 654
Molecule Type: DNA
Features Location/Qualifiers:
- source, 1..654
> mol_type, unassigned DNA
> organism, synthetic construct E. coli optimized
Residues:
atgagcaccc aggaatcttg gtggttcacc ctgagcggct ctgcatctgc actgctctgg 60
atcgtggcgt atggtctgat cattcgccgt ggcttcaaag accgttccta cggcatgccg 120
tttgcgccgc tctgtgtcaa catgagctat gagctgatct tcggcttcgt atacccggat 180
cagccaccga tgaactacgc aaaccaagtg tggtttgcta ttgacctgat cattttctac 240
cagttcatcc gctttggtaa atctgagttc gaacgcctct tcccgcgtgc gtggttcctg 300
ccggcagtgt ccctgtctgt gctcctggcc ttcggcggtg ttctggctgt taccctggaa 360
tttcatgact tccacggcaa ctacactggc tggggtgacc agctcctgat ctccatctct 420
tttatctggc tcctggcacg ccgtggttcc gtagccggtc agtccgttta catcgctctg 480
agccgtatgc tcggctctat tgtgctgatt cctggtcaga tgatccaggg cccggcggat 540
tctgttctcc tgggtttcat ttatgtctct ttcgcaaccc tggacgctat ttatatcgcc 600
ctcctgattc gccagtgtcg tctggagggc attaacccgt ggcgtcgtct gtaa 654

CLAIMS

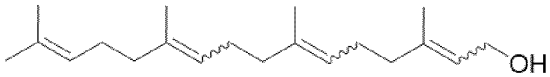
1. A process for the preparation of a compound of formula (I)



(I)

in the form of any one of its stereoisomers or a mixture thereof, comprising:

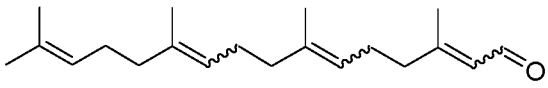
(i) contacting a compound of formula (II)



(II)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having alcohol dehydrogenase (ADH) enzyme activity to produce a compound of formula (III);

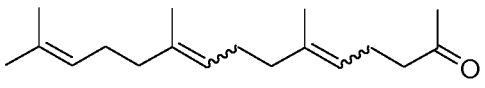
(ii) contacting the compound of formula (III)



(III)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV);

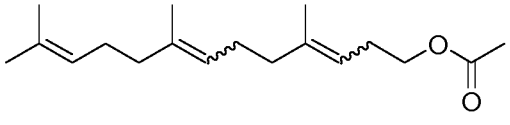
(iii) contacting the compound of formula (IV)



(IV)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having Baeyer-Villiger monooxygenase (BVMO) enzyme activity to produce a compound of formula (V);

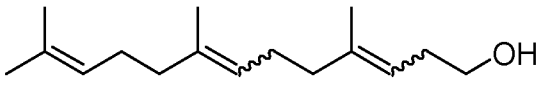
(iv) contacting the compound of formula (V)



(V)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having esterase enzyme activity to produce a compound of formula (VI); and

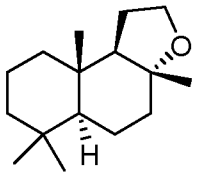
(v) contacting the compound of formula (VI)



(VI)

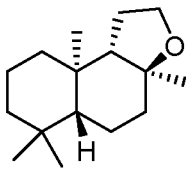
in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having terpene cyclase enzyme activity to produce a compound of formula (I).

2. The process of claim 1, wherein more than 97% of the compound of formula (I) is in the form of formula (Ia):



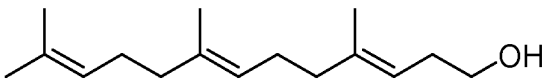
(formula Ia)

and/or formula (Ib):



(formula Ib).

3. The process of claim 1 or 2, wherein the compound of formula (VI) is in the form of formula (VIa)



(formula VIa).

4. The process of any of the previous claims, wherein:

(i) the polypeptide having ADH enzyme activity comprises at least one or more motifs selected from CHTD (SEQ ID NO: 228), GHEGxG (SEQ ID NO: 229), LxCGxxTGxGA (SEQ ID NO: 230), Gx[VI]GL (SEQ ID NO: 231), LxxxG[LVI][PA] (SEQ ID NO: 232), GxVxAI (SEQ ID NO: 233) and YxATKxA (SEQ ID NO: 234); wherein, residues x represent independently of each other any natural amino acid residue;

(ii) the polypeptide having enal-cleaving enzyme activity is selected from the group of polypeptides containing:

(a) at least one DUF4334 protein family domain having the Pfam ID number PF14232,

(b) at least one GXWXG (SEQ ID NO: 263) protein family domain having the Pfam ID number PF14231I, and/or

(c) a domain retaining at least 90% sequence identity to PF14232 or PF14231;

(iii) the polypeptide having BVMO enzyme activity is selected from the group of polypeptides comprising:

(a) a flavin-containing monooxygenase (FMO) protein family domain having the Pfam ID number PF00743 within their amino acid sequence or a domain retaining at least 90%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to PF00743, and/or

(b) at least one or more of motifs selected from:

. GxGxxG (SEQ ID NO: 239),

. [GS]GxWxxxxYPGxxxD (SEQ ID NO: 240),

. Gxxx[FY]xGxxx[HS]xxxW (SEQ ID NO: 241), and

. [KQ]x[VI]xx[IV]GxG (SEQ ID NO: 242);

wherein, residues x represent independently of each other any natural amino acid residue.

(iv) the polypeptide having esterase enzyme activity comprises at least one or more motifs selected from AxVVxVxxRLAPE (SEQ ID NO: 243), GASAGGGLxA (SEQ ID NO: 244), VxQLLxYPMLDDR (SEQ ID NO: 245) and ARxxDLSGLPxT (SEQ ID NO: 246); wherein, residues x represent independently of each other any natural amino acid residue; and/or

(v) the polypeptide having terpene cyclase enzyme activity is a squalene cyclase enzyme or a meroterpenoid cyclase enzyme.

5. The process of claim 4, wherein the squalene cyclase enzyme comprises at least one or more motifs selected from [SP][TP][VIL]WDTx[LWI] (SEQ ID NO: 247), PGG[WF][GYA]F (SEQ ID NO: 248), PDxDD[TAS][TIAS] (SEQ ID NO: 249), [MIL]QxxxG[GA][WF]x[AS][FY] (SEQ ID NO: 250), Qxxx[GH]xWxG[RK]WGxx[YF]xYG (SEQ ID NO: 251), Qxx[DN]G[GS][WF][GS]ExxxS (SEQ ID NO: 252) and [STA]xx[SFN][QC]T[AGT]W[AS][LIV]xx[LQ] (SEQ ID NO: 253); wherein, residues x represent independently of each other any natural amino acid residue.

6. The process of claim 4, wherein the meroterpenoid cyclase enzyme is selected from at least one or more of the following polypeptides:

(a) a membrane-integrated meroterpenoid cyclase of bacterial origin containing at least one or more amino acid motifs selected from:

. [W]xxx[D]xx[ILVMN] (SEQ ID NO: 254);

. PxxAxxxNxxWE (SEQ ID NO: 255);

. MxxxFxxMLxxR (SEQ ID NO: 256); and

. RxxxxGQS (SEQ ID NO: 257);

(b) a membrane-integrated meroterpenoid cyclase of fungal origin containing a least one or more amino acid motifs selected from:

. [WY]Exx[YFW] (SEQ ID NO: 258); and

. [DNE]xSYxxP (SEQ ID NO: 259);

(c) a soluble meroterpenoid cyclase of bacterial origin containing a least one or more amino acid motifs selected from:

. GxWxxxW[WG]xxxxY (SEQ ID NO: 260);

. WxxxHxxV[TSA] (SEQ ID NO: 261); and

. GxWxD[FY] (SEQ ID NO: 262);

wherein, residues x represent independently of each other any natural amino acid residue.

7. The process of any of the previous claims, wherein:

(i) the polypeptide having ADH enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 11 to 21;

(ii) the polypeptide having enal-cleaving enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 22;

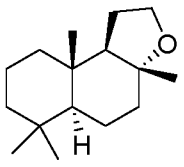
(iii) the polypeptide having BVMO enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 23 to 26 and 216 to 227;

(iv) the polypeptide having esterase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 27 and 28; and/or

(v) the polypeptide having terpene cyclase enzyme activity has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any of SEQ ID NOs: 29 to 75, 79 to 89 and 265 to 289.

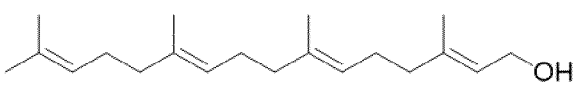
8. The process of any of the previous claims, wherein:

(i) the compound of formula (I) is in the form of formula (Ia):



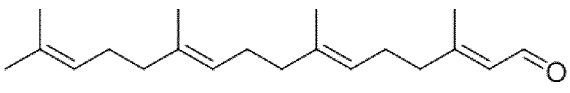
(formula Ia)

(ii) the compound of formula (II) is in the form of formula (IIa):



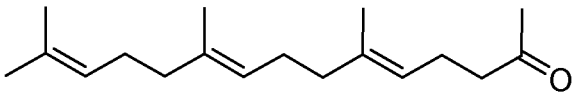
(formula IIa);

(iii) the compound of formula (III) is in the form of formula (IIIa):



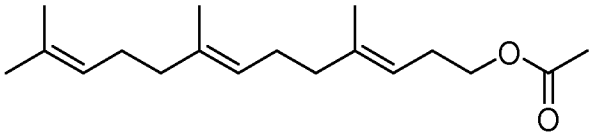
(formula IIIa);

(iv) the compound of formula (IV) is in the form of formula (IVa):



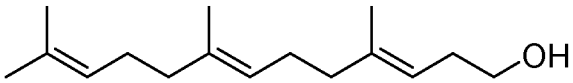
(formula IVa);

(v) the compound of formula (V) is in the form of formula (Va):



(formula Va);

(vi) the compound of formula (VI) is in the form of formula (VIa):



(formula VIa).

9. The process of any of the previous claims, wherein the process further comprises one or more steps prior to step (i), said step(s) comprising:
(a) preparing geranylgeranyl-diphosphate (GGPP) from isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) using one or more polypeptides having prenyltransferase enzyme activity; and/or,
(b) preparing a compound of formula (II) from GGPP using one or more polypeptides having phosphatase enzyme activity.

10. The process of claim 9, wherein the polypeptide having prenyltransferase enzyme activity is a GGPP synthase; preferably, said polypeptide has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to SEQ ID NO: 1 or 2.

11. The process of claim 9, wherein the polypeptide having phosphatase enzyme activity is a GGPP phosphatase; preferably, said polypeptide has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% or more sequence identity to any one of SEQ ID NOs: 3 to 10.

12. The process of any of the previous claims, wherein the process is an *in vivo* or a bioconversion process.

13. The process of any of the previous claims, wherein the process is performed in a recombinant cell capable of functionally expressing (i) the polypeptide having ADH enzyme activity, (ii) the polypeptide having enal-cleaving enzyme activity, (iii) the polypeptide having BVMO enzyme activity, (iv) the polypeptide having esterase enzyme activity, and (v) the polypeptide having terpene cyclase enzyme activity.

14. The process of claim 13, wherein the recombinant cell is a bacterial cell, a plant cell, a fungal cell such as a yeast cell; preferably, the recombinant cell is of the genus *Escherichia*, *Saccharomyces*, *Yarrowia* or *Pichia*.

15. A recombinant cell comprising a compound of formula (I) and one or more compound(s) selected from the group of formula (II), formula (III), formula (IV), formula (V) and formula (VI); preferably, wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

16. The recombinant cell of claim 15, wherein the cell comprises (i) a polypeptide having ADH enzyme activity, (ii) a polypeptide having enal-cleaving enzyme activity, (iii) a polypeptide having BVMO enzyme activity, (iv) a polypeptide having esterase enzyme activity, and (v) a polypeptide having terpene cyclase enzyme activity; preferably, said polypeptides are as defined in any of claims 4 to 7.

17. The recombinant cell of claim 16, further comprising a polypeptide having prenyltransferase enzyme activity and a polypeptide having phosphatase enzyme activity; preferably, said polypeptides are as defined in any of claims 9 to 11.

18. A cell culture fermentation medium comprising the recombinant cell of any of claims 15 to 17.

19. A reaction mixture comprising a compound of formula (I) and one or more compound(s) selected from the group of formula (II), formula (III), formula (IV), formula (V) and formula (VI); preferably, wherein more than 97% of the compound of formula (I) is in the form of formula (Ia) and/or (Ib).

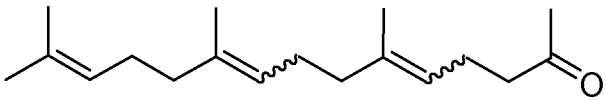
20. A compound of formula (I) obtained or obtainable by the process of claims 1 to 14, from the recombinant cell of any of claims 15 to 17, from the cell culture fermentation medium of claim 18, or from the reaction mixture of claim 19.

21. A process for the preparation of a compound of formula (VI)

(formula VI)

in the form of any one of its stereoisomers or a mixture thereof, comprising:

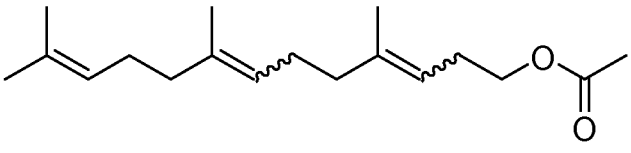
(i) contacting the compound of formula (IV)



(formula IV)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having BVMO enzyme activity to produce a compound of formula (V); and,

(ii) contacting the compound of formula (V)

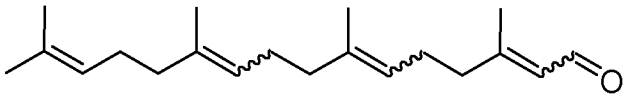


(V)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having esterase enzyme activity to produce a compound of formula (formula VI).

22. The process of claim 21, wherein the process further comprises one or more steps prior to step (i), said step(s) comprising:

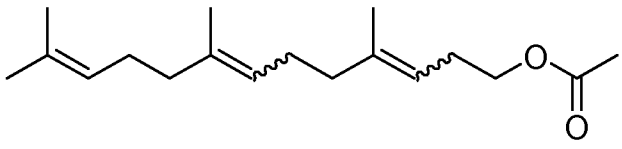
(a) contacting the compound of formula (III)



(formula III)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV).

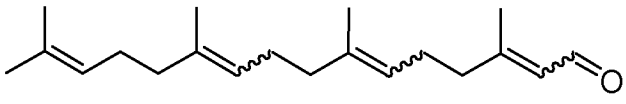
23. A process for the preparation of a compound of formula (V)



(formula V)

in the form of any one of its stereoisomers or a mixture thereof, comprising:

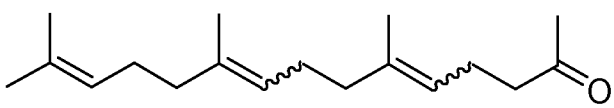
(i) contacting the compound of formula (III)



(formula III)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV); and,

(ii) contacting the compound of formula (IV)

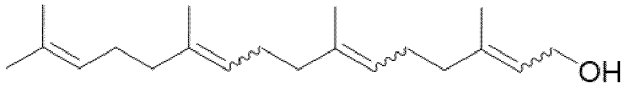


(formula IV)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having BVMO enzyme activity to produce a compound of formula (V).

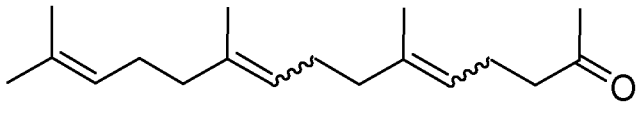
24. The process of claim 23, wherein the process further comprises one or more steps prior to step (i), said step(s) comprising:

(a) contacting the compound of formula (II)


(formula II)

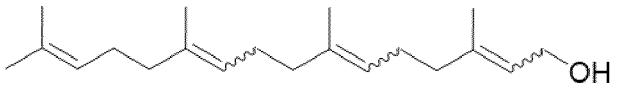
in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having ADH enzyme activity to produce a compound of formula (III).

25. A process for the preparation of a compound of formula (IV)


(formula IV)

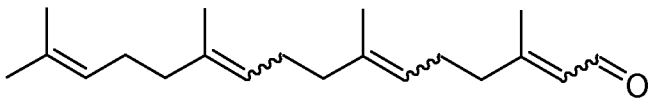
in the form of any one of its stereoisomers or a mixture thereof, comprising:

(i) contacting the compound of formula (II)


(formula II)

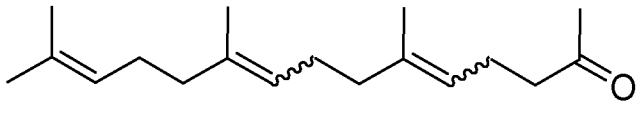
in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having ADH enzyme activity to produce a compound of formula (III); and,

(ii) contacting the compound of formula (III)


(formula III)

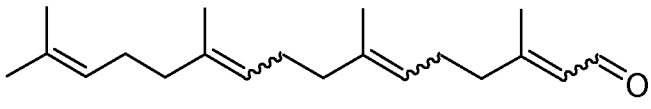
in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV).

26. A process for the preparation of a compound of formula (IV)


(formula IV)

in the form of any one of its stereoisomers or a mixture thereof, comprising:

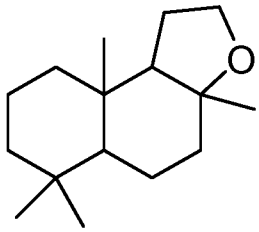
(i) contacting the compound of formula (III)



(formula III)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV).

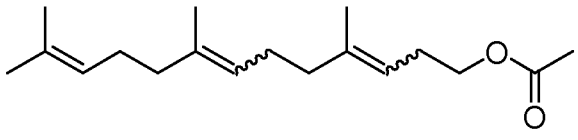
27. A process for the preparation of a compound of formula (I)



(I)

in the form of any one of its stereoisomers or a mixture thereof, comprising:

(i) contacting a compound of formula (V)



(formula V)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having esterase enzyme activity to produce a compound of formula (VI); and

(ii) contacting the compound of formula (VI)

(formula VI)

in the form of any one of its stereoisomers or a mixture thereof, with a polypeptide having terpene cyclase enzyme activity to produce a compound of formula (I).

28. The process of any of claims 21 to 27, wherein the polypeptides are as defined in any of the previous claims.

29. A compound obtained or obtainable by the process of any of claims 21 to 28.

30. Use of the compound of claim 20 or 29 as a perfumery, flavor or aroma ingredient, or as a precursor for making said ingredient.

31. Use of a polypeptide having enal-cleaving enzyme activity to produce a compound of formula (IV), (V), (VI), (I) and/or a derivative thereof.

Fig. 1

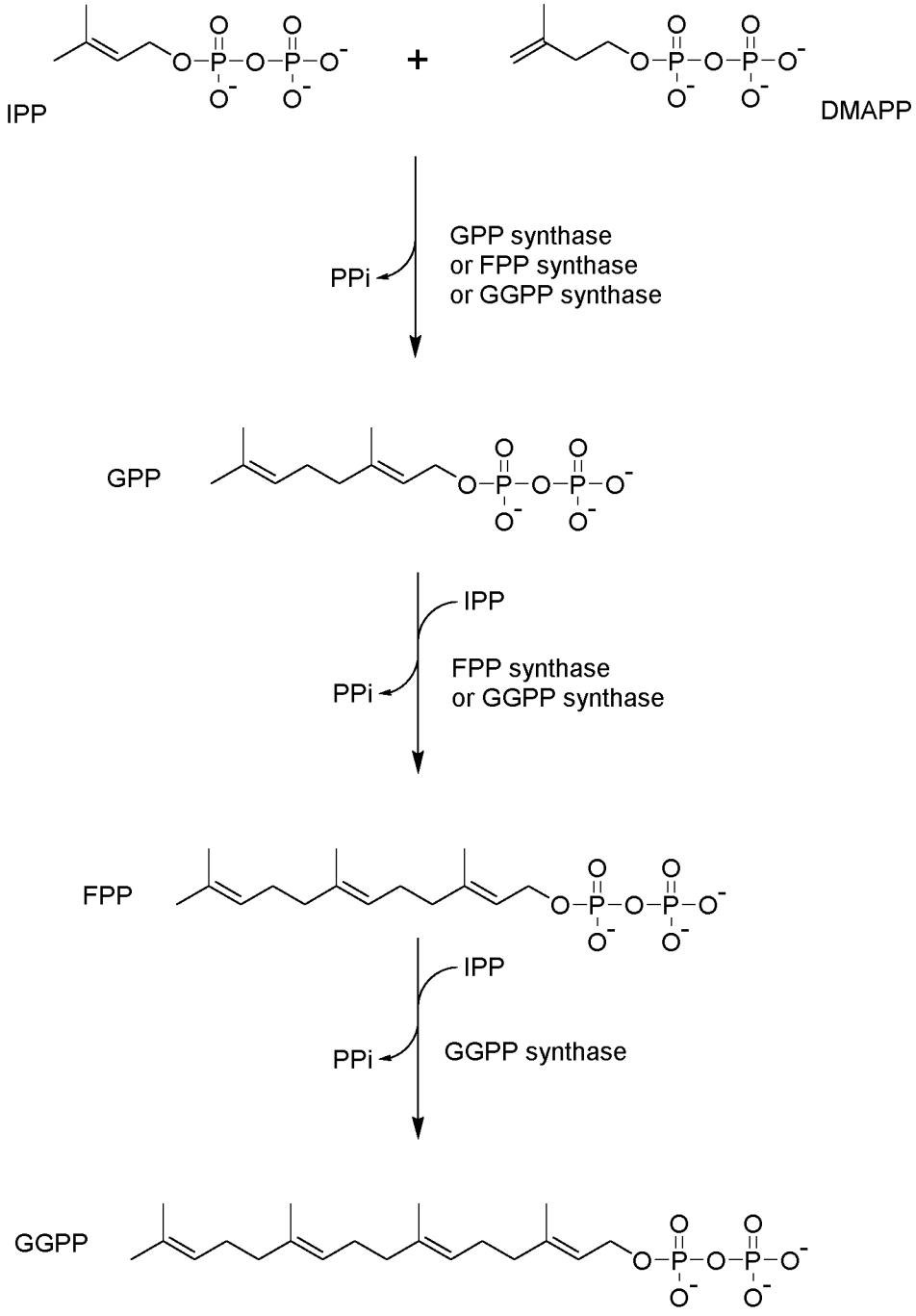
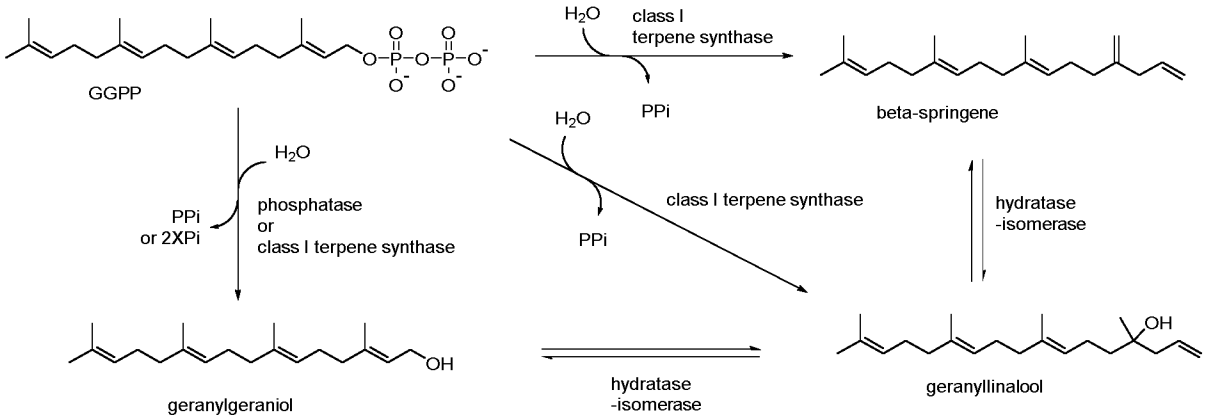
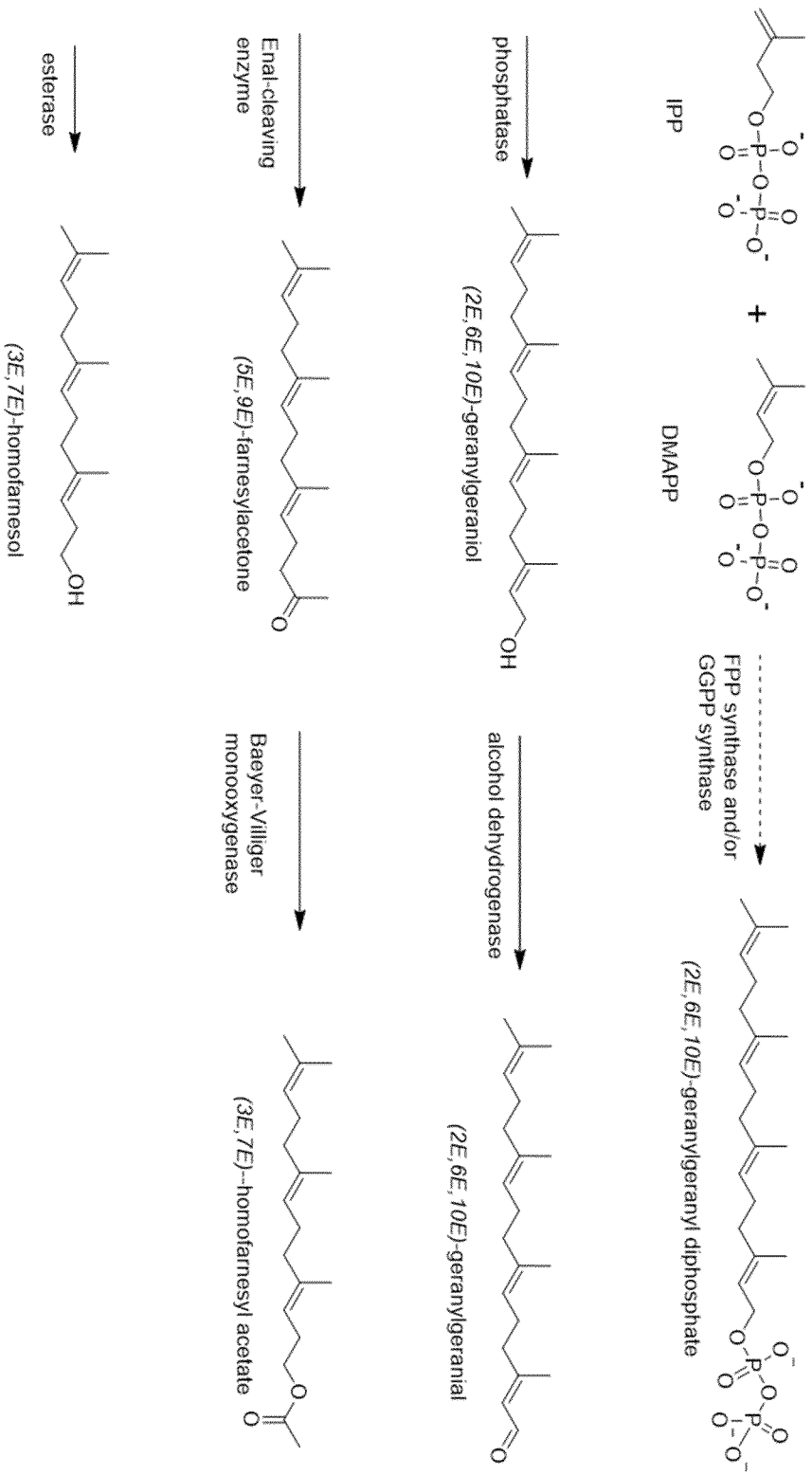

IPP
+
DMAPP
PPi
GPP synthase
or FPP synthase
or GGPP synthase
GPP
IPP
PPi
FPP synthase
or GGPP synthase
FPP
IPP
PPi
GGPP synthase
GGPP

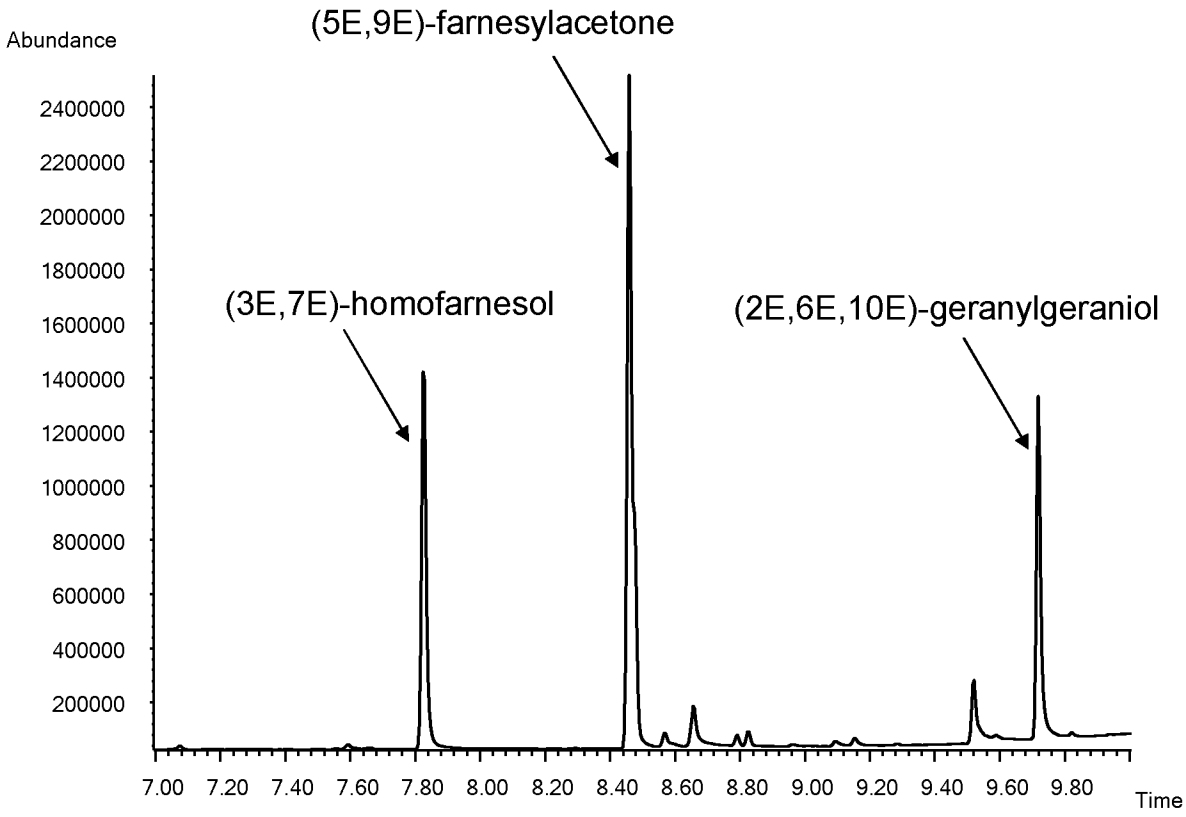
Fig. 2


GGPP
H2O
class I
terpene synthase
PPi
beta-springene
H2O
phosphatase
or
class I terpene synthase
PPi
or 2XPi
H2O
class I terpene synthase
PPi
hydratase
-isomerase
OH
geranylgeraniol
hydratase
-isomerase
OH
geranyllinalool

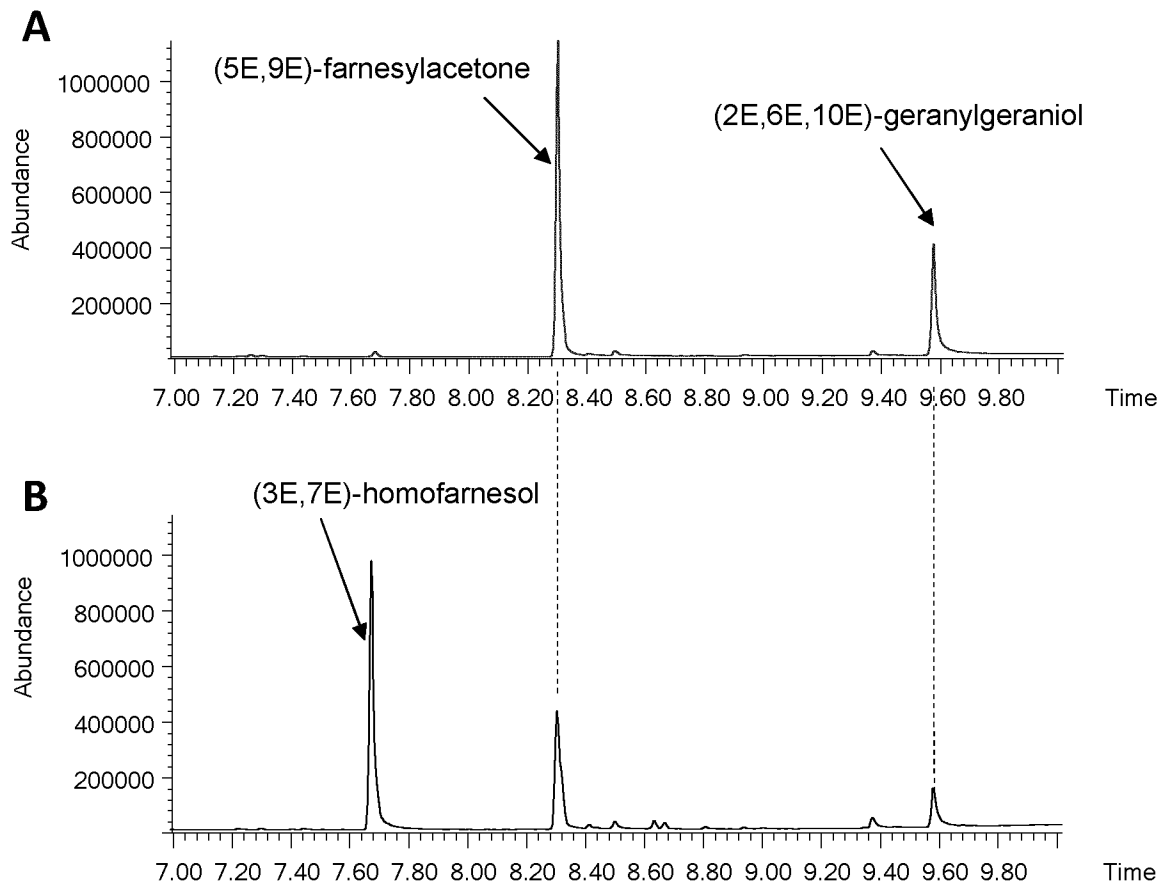
- Fig. 3 -


IPP
DMAPP
FPP synthase and/or
GGPP synthase
(2E,6E,10E)-geranylgeranyl diphosphate
phosphatase
(2E,6E,10E)-geranylgeraniol
alcohol dehydrogenase
(2E,6E,10E)-geranylgeranial
Enal-cleaving
enzyme
(5E,9E)-farnesylacetone
Baeyer-Villiger
monooxygenase
(3E,7E)-homofarnesyl acetate
esterase
(3E,7E)-homofarnesol

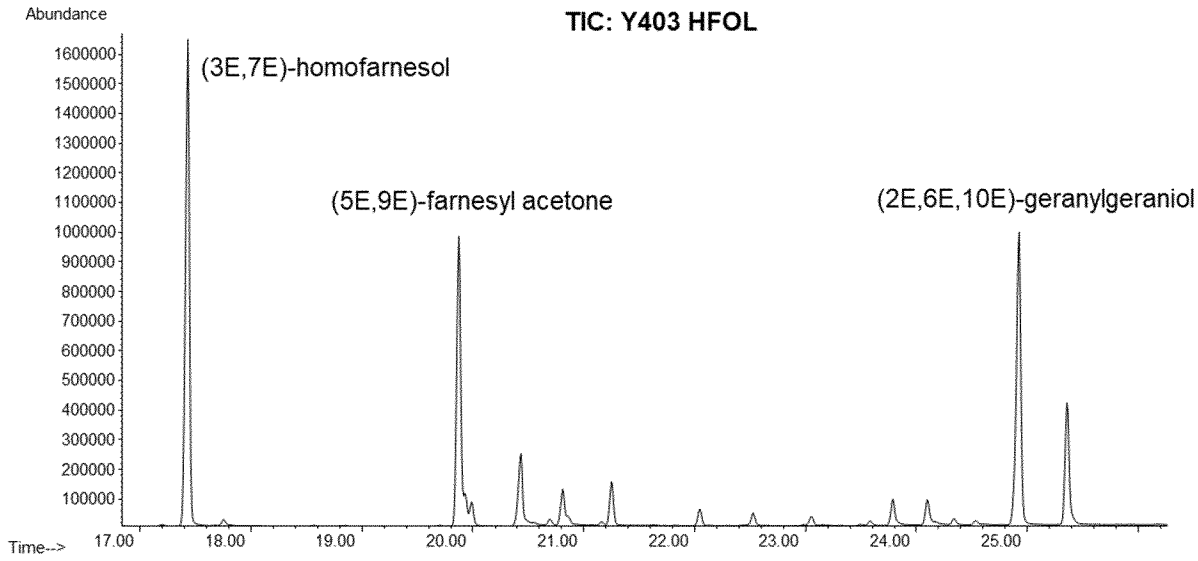
- Fig. 4 -


Abundance
(5E,9E)-farnesylacetone
(3E,7E)-homofarnesol
(2E,6E,10E)-geranylgeraniol
2400000
2200000
2000000
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
7.00 7.20 7.40 7.60 7.80 8.00 8.20 8.40 8.60 8.80 9.00 9.20 9.40 9.60 9.80
Time

- Fig. 5 -


A
(5E,9E)-farnesylacetone
(2E,6E,10E)-geranylgeraniol
Abundance
1000000
800000
600000
400000
200000
7.00 7.20 7.40 7.60 7.80 8.00 8.20 8.40 8.60 8.80 9.00 9.20 9.40 9.60 9.80
Time
B
(3E,7E)-homofarnesol
Abundance
1000000
800000
600000
400000
200000
7.00 7.20 7.40 7.60 7.80 8.00 8.20 8.40 8.60 8.80 9.00 9.20 9.40 9.60 9.80
Time

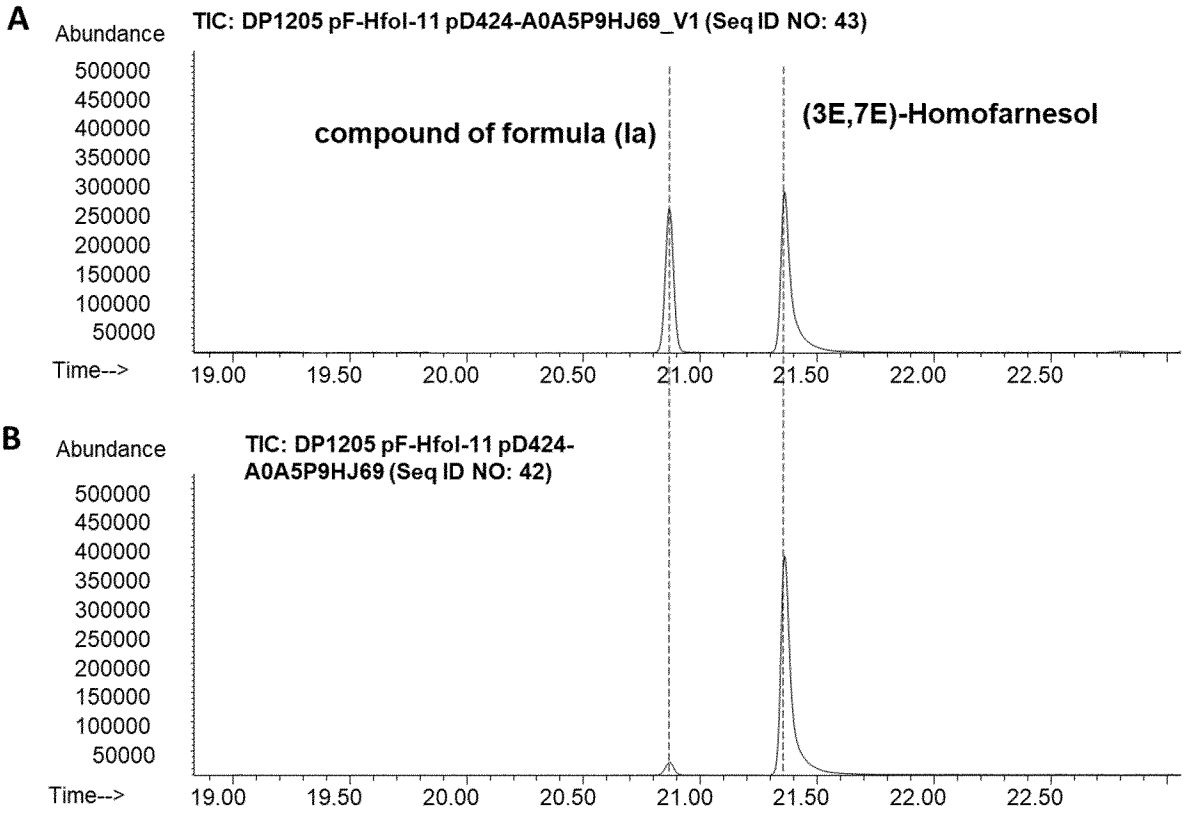
- Fig. 6 -


TIC: Y403 HFOL
Abundance
(3E,7E)-homofarnesol
(5E,9E)-farnesyl acetone
(2E,6E,10E)-geranylgeraniol
1600000
1500000
1400000
1300000
1200000
1100000
1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
Time-->
17.00
18.00
19.00
20.00
21.00
22.00
23.00
24.00
25.00

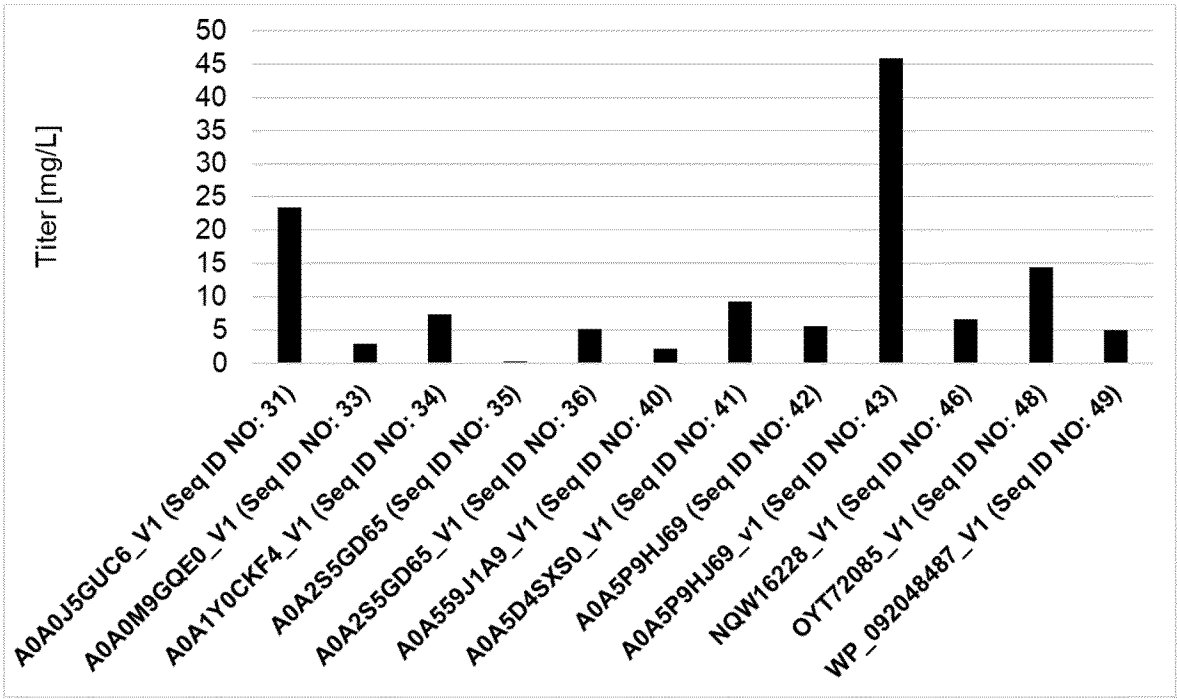
- Fig. 7 -

IPP
DMAPP
FPP synthase and/or
GGPP synthase
(2E,6E,10E)-geranylgeranyl diphosphate
phosphatase
(2E,6E,10E)-geranylgeraniol
alcohol dehydrogenase
(2E,6E,10E)-geranylgeranial
enal-cleaving
enzyme
(5E,9E)-farnesylacetone
baeyer-Villiger
monooxygenase
(3E,7E)-homofarnesyl acetate
esterase
(3E,7E)-homofarnesol
squalene cyclase
compound of formula (Ia)

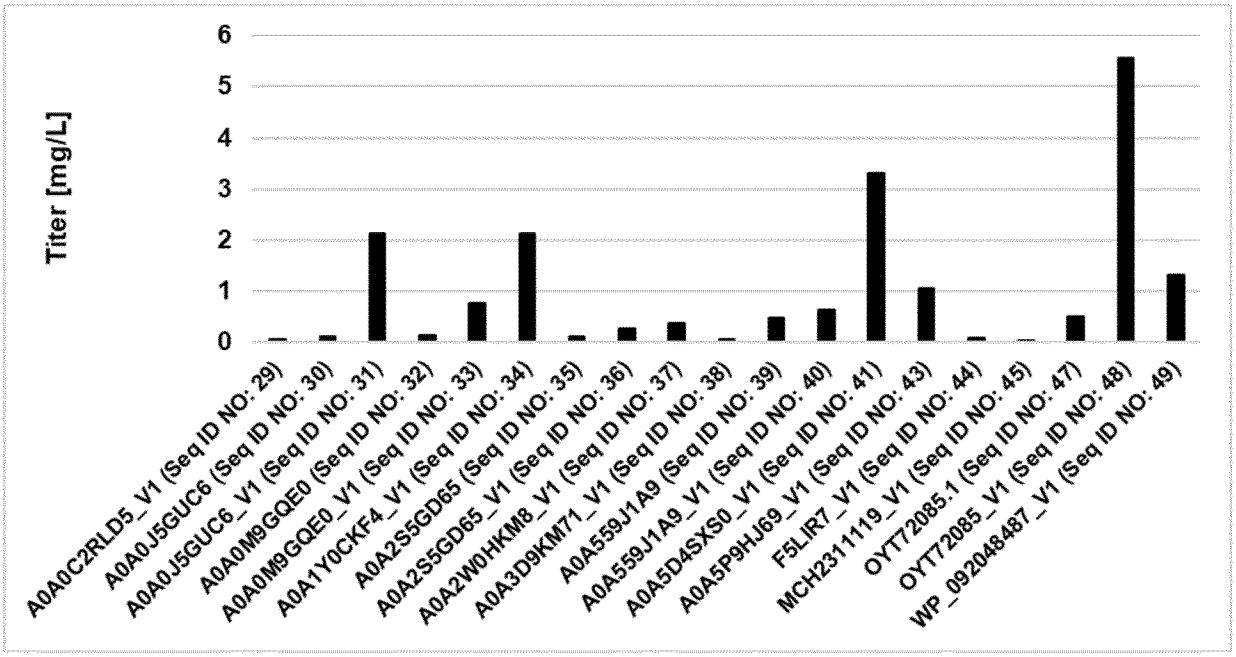
- Fig. 8 -


A
Abundance
TIC: DP1205 pF-Hfol-11 pD424-A0A5P9HJ69_V1 (Seq ID NO: 43)
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
Time-->
compound of formula (Ia)
(3E,7E)-Homofarnesol
19.00
19.50
20.00
20.50
21.00
21.50
22.00
22.50
B
Abundance
TIC: DP1205 pF-Hfol-11 pD424-A0A5P9HJ69 (Seq ID NO: 42)
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
Time-->
19.00
19.50
20.00
20.50
21.00
21.50
22.00
22.50

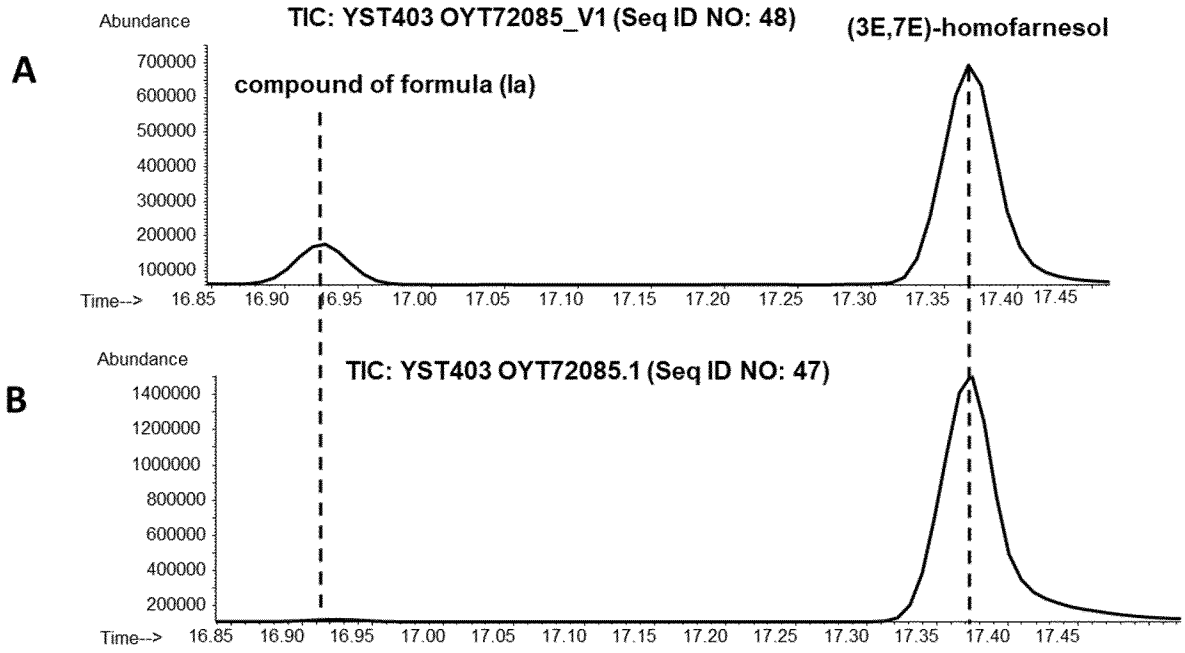
- Fig. 9 -


Titer [mg/L]
50
45
40
35
30
25
20
15
10
5
0
A0A0J5GUC6_V1 (Seq ID NO: 31)
A0A0M9GQE0_V1 (Seq ID NO: 33)
A0A1Y0CKF4_V1 (Seq ID NO: 34)
A0A2S5GD65 (Seq ID NO: 35)
A0A2S5GD65_V1 (Seq ID NO: 36)
A0A559J1A9_V1 (Seq ID NO: 40)
A0A5D4SXS0_V1 (Seq ID NO: 41)
A0A5P9HJ69 (Seq ID NO: 42)
A0A5P9HJ69_v1 (Seq ID NO: 43)
NQW16228_V1 (Seq ID NO: 46)
OYT72085_V1 (Seq ID NO: 48)
WP_092048487_V1 (Seq ID NO: 49)

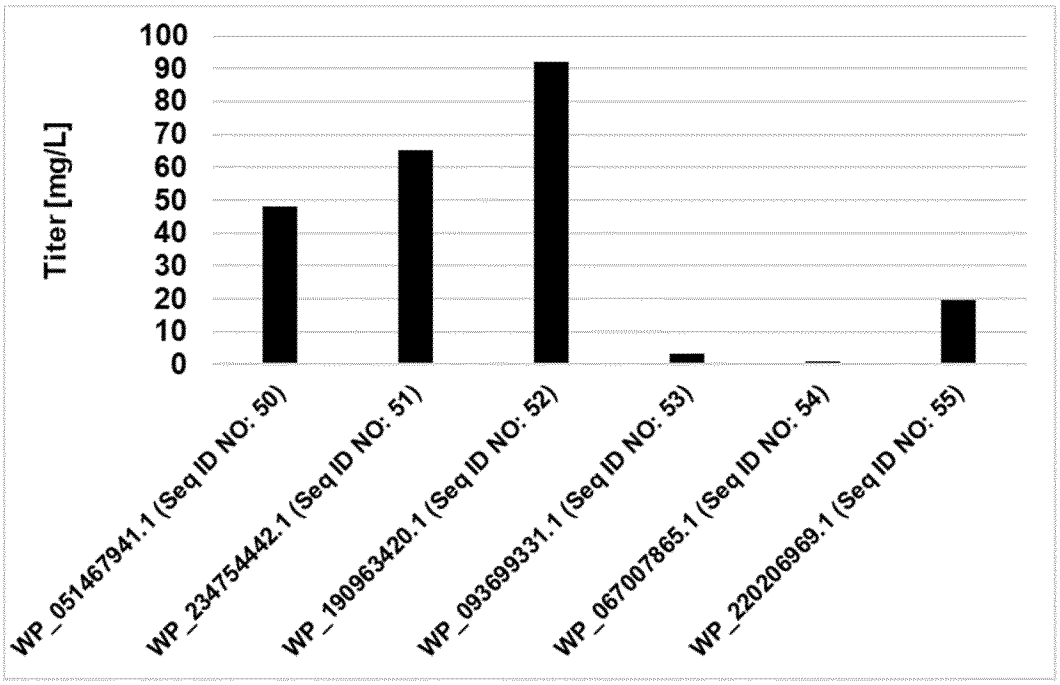
- Fig. 10 -


Titer [mg/L]
6
5
4
3
2
1
0
A0A0C2RLD5_V1 (Seq ID NO: 29)
A0A0J5GUC6 (Seq ID NO: 30)
A0A0J5GUC6_V1 (Seq ID NO: 31)
A0A0M9GQE0 (Seq ID NO: 32)
A0A0M9GQE0_V1 (Seq ID NO: 33)
A0A1Y0CKF4_V1 (Seq ID NO: 34)
A0A2S5GD65 (Seq ID NO: 35)
A0A2S5GD65_V1 (Seq ID NO: 36)
A0A2W0HKM8_V1 (Seq ID NO: 37)
A0A3D9KM71_V1 (Seq ID NO: 38)
A0A559J1A9 (Seq ID NO: 39)
A0A559J1A9_V1 (Seq ID NO: 40)
A0A5D4SXS0_V1 (Seq ID NO: 41)
A0A5P9HJ69_V1 (Seq ID NO: 43)
F5LIR7_V1 (Seq ID NO: 44)
MCH2311119_V1 (Seq ID NO: 45)
OYT72085.1 (Seq ID NO: 47)
OYT72085_V1 (Seq ID NO: 48)
WP_092048487_V1 (Seq ID NO: 49)

- Fig. 11 -


A
Abundance
TIC: YST403 OYT72085_V1 (Seq ID NO: 48)
(3E,7E)-homofarnesol
compound of formula (Ia)
700000
600000
500000
400000
300000
200000
100000
Time-->
16.85
16.90
16.95
17.00
17.05
17.10
17.15
17.20
17.25
17.30
17.35
17.40
17.45
B
Abundance
TIC: YST403 OYT72085.1 (Seq ID NO: 47)
1400000
1200000
1000000
800000
600000
400000
200000
Time-->
16.85
16.90
16.95
17.00
17.05
17.10
17.15
17.20
17.25
17.30
17.35
17.40
17.45

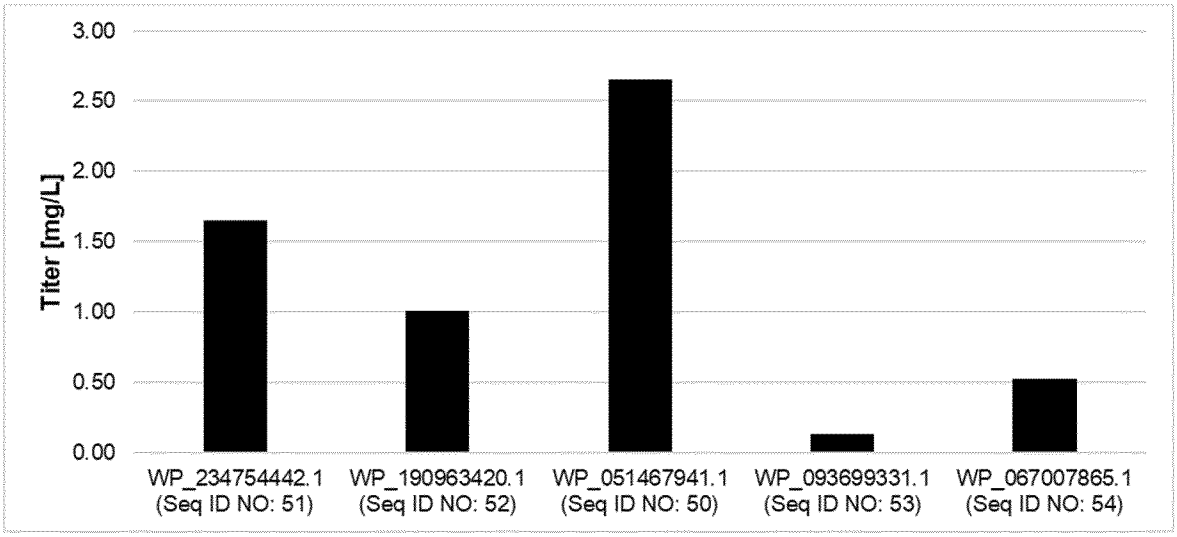
- Fig. 12 -


Titer [mg/L]
100
90
80
70
60
50
40
30
20
10
0
WP_051467941.1 (Seq ID NO: 50)
WP_234754442.1 (Seq ID NO: 51)
WP_190963420.1 (Seq ID NO: 52)
WP_093699331.1 (Seq ID NO: 53)
WP_067007865.1 (Seq ID NO: 54)
WP_220206969.1 (Seq ID NO: 55)

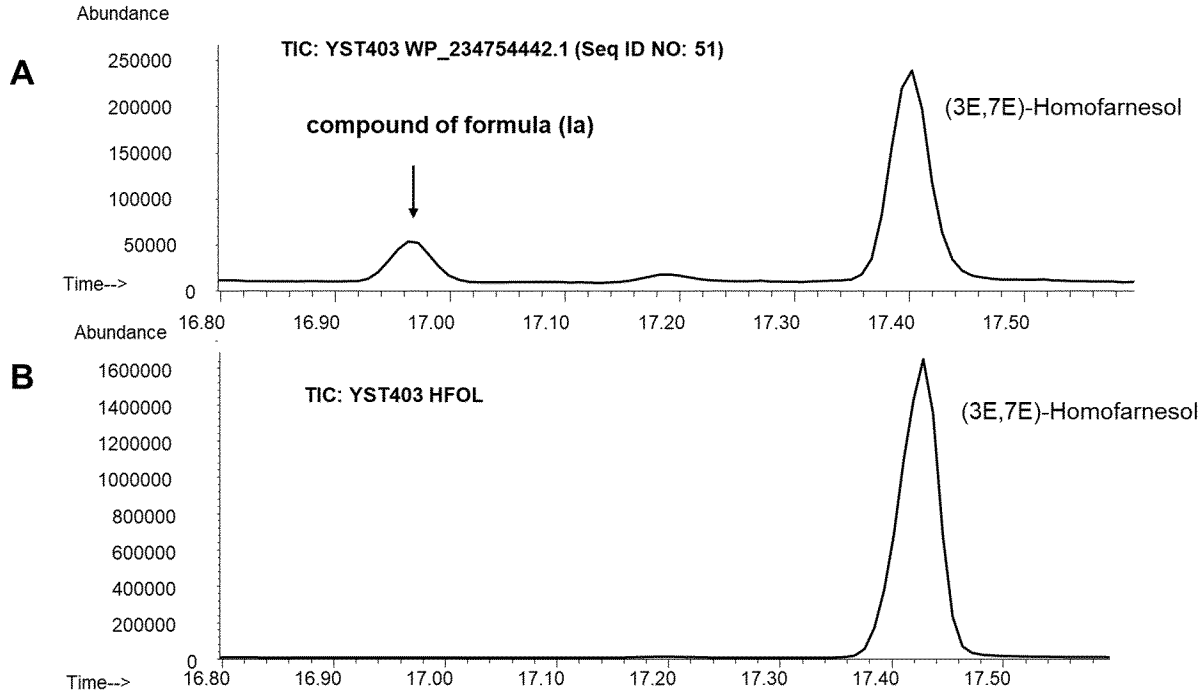
- Fig. 13 -

IPP
DMAPP
FPP synthase and/or
GGPP synthase
(2E,6E,10E)-geranylgeranyl diphosphate
phosphatase
(2E,6E,10E)-geranylgeraniol
alcohol dehydrogenase
(2E,6E,10E)-geranylgeranial
enal-cleaving
enzyme
(5E,9E)-farnesylacetone
baeyer-Villiger
monooxygenase
(3E,7E)-homofarnesyl acetate
esterase
(3E,7E)-homofarnesol
meroterpenoid
cyclase
compound of formula (Ia)

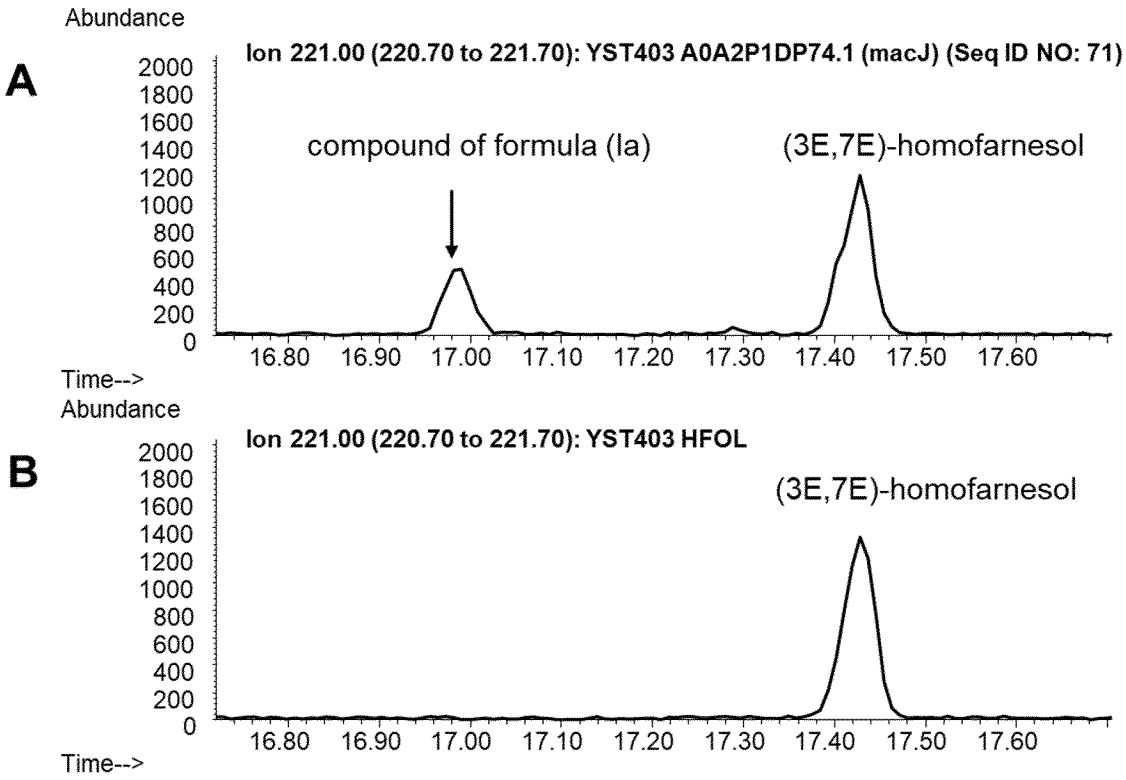
- Fig. 14 -


3.00
2.50
2.00
1.50
1.00
0.50
0.00
Titer [mg/L]
WP_234754442.1 (Seq ID NO: 51)
WP_190963420.1 (Seq ID NO: 52)
WP_051467941.1 (Seq ID NO: 50)
WP_093699331.1 (Seq ID NO: 53)
WP_067007865.1 (Seq ID NO: 54)

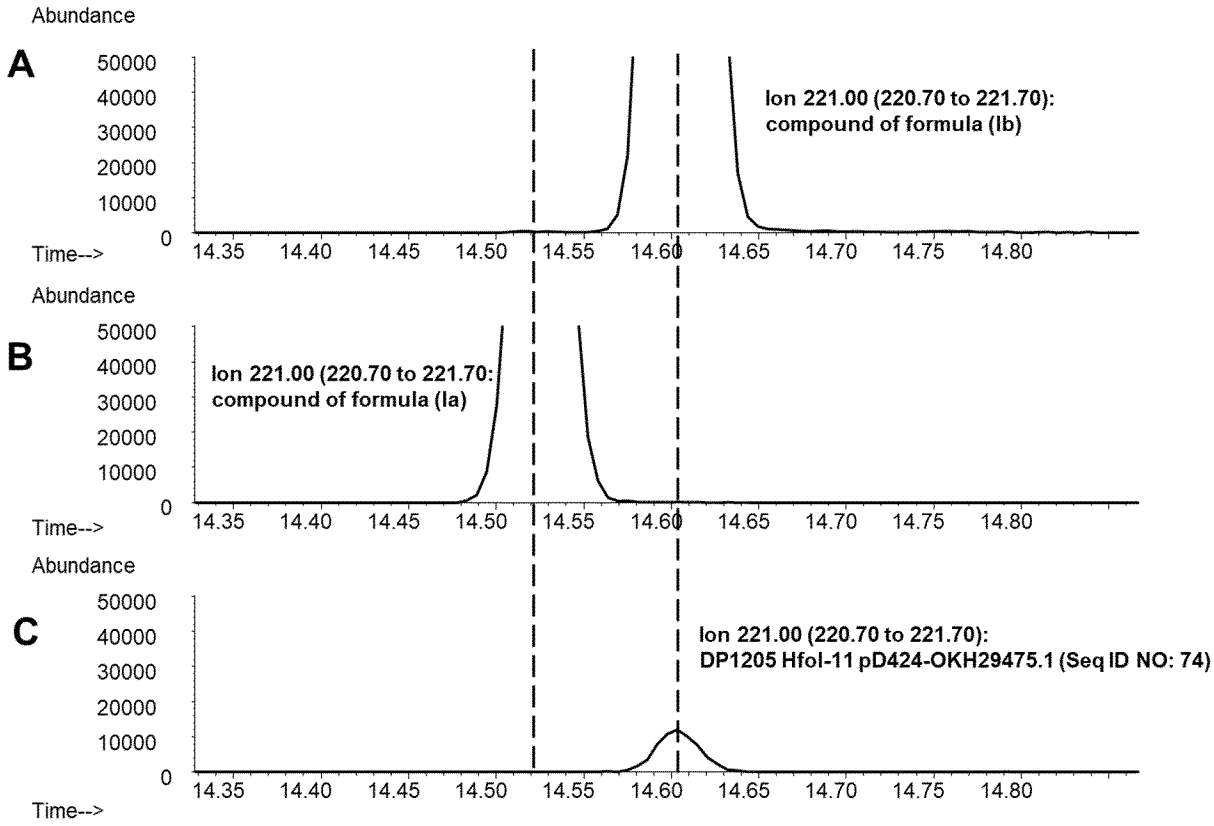
- Fig. 15 -


A
Abundance
250000
200000
150000
100000
50000
Time-->
0
TIC: YST403 WP_234754442.1 (Seq ID NO: 51)
compound of formula (Ia)
(3E,7E)-Homofarnesol
16.80
16.90
17.00
17.10
17.20
17.30
17.40
17.50
B
Abundance
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
Time-->
TIC: YST403 HFOL
(3E,7E)-Homofarnesol
16.80
16.90
17.00
17.10
17.20
17.30
17.40
17.50

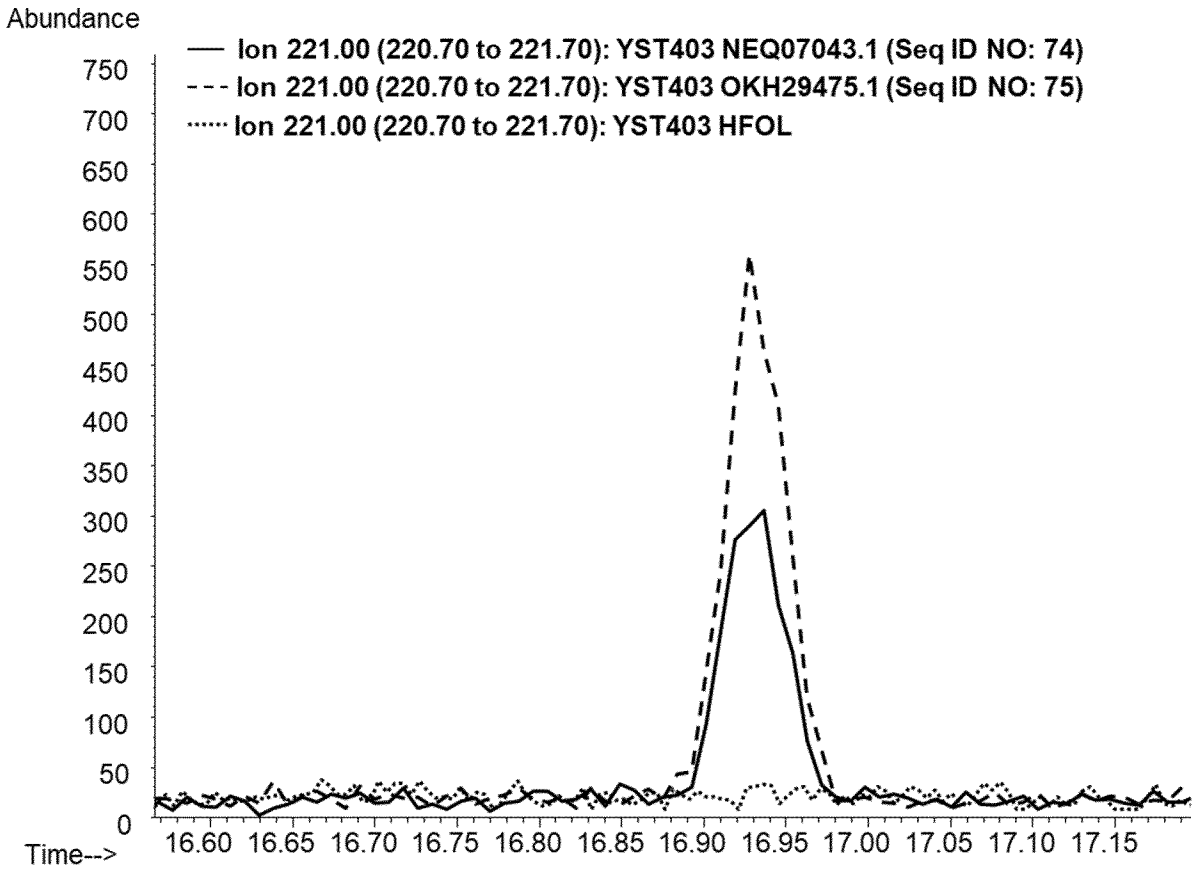
- Fig. 16 -


A
Abundance
Ion 221.00 (220.70 to 221.70): YST403 A0A2P1DP74.1 (macJ) (Seq ID NO: 71)
compound of formula (Ia)
(3E,7E)-homofarnesol
2000
1800
1600
1400
1200
1000
800
600
400
200
0
16.80 16.90 17.00 17.10 17.20 17.30 17.40 17.50 17.60
Time-->
B
Abundance
Ion 221.00 (220.70 to 221.70): YST403 HFOL
(3E,7E)-homofarnesol
2000
1800
1600
1400
1200
1000
800
600
400
200
0
16.80 16.90 17.00 17.10 17.20 17.30 17.40 17.50 17.60
Time-->

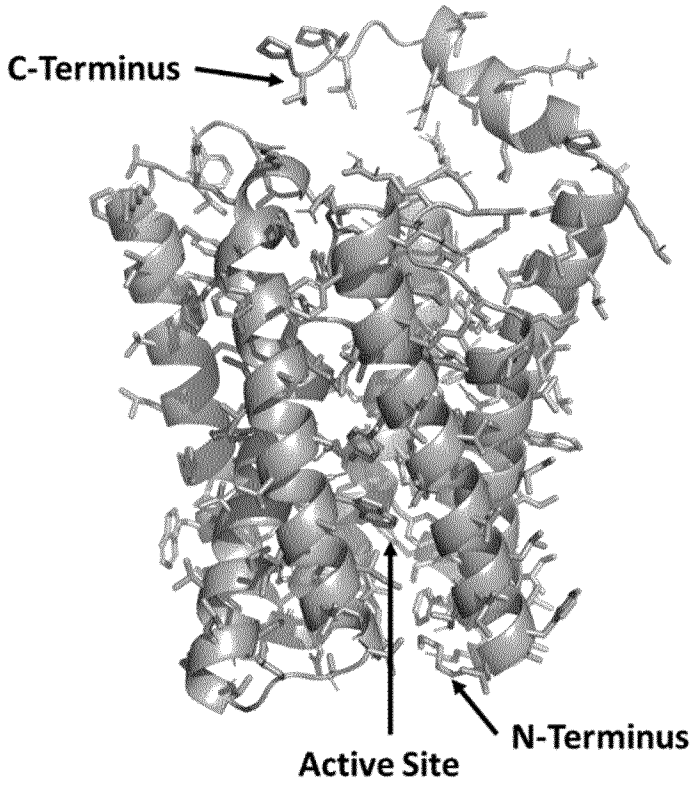
- Fig. 17 -


A
Abundance
50000
40000
30000
20000
10000
0
Time-->
14.35
14.40
14.45
14.50
14.55
14.60
14.65
14.70
14.75
14.80
Ion 221.00 (220.70 to 221.70):
compound of formula (Ib)
B
Abundance
Ion 221.00 (220.70 to 221.70:
compound of formula (Ia)
C
Abundance
Ion 221.00 (220.70 to 221.70):
DP1205 Hfol-11 pD424-OKH29475.1 (Seq ID NO: 74)

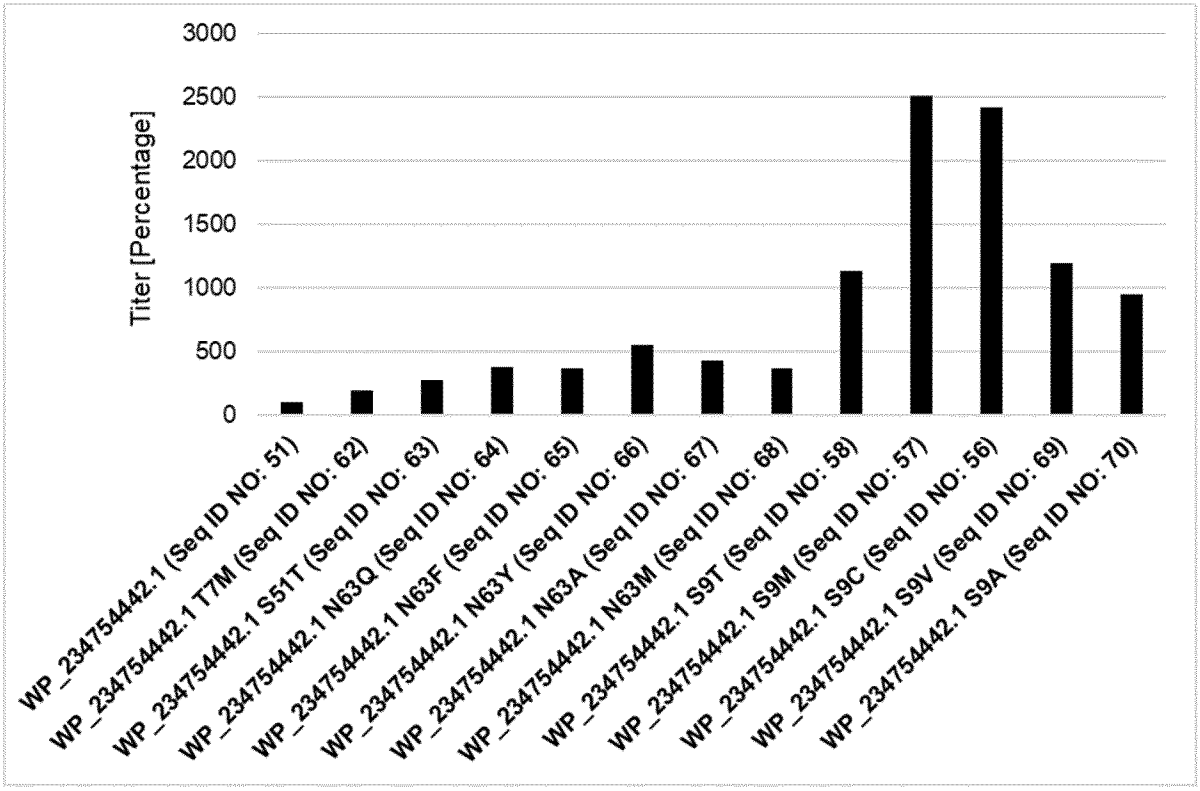
- Fig. 18 -


Abundance
Ion 221.00 (220.70 to 221.70): YST403 NEQ07043.1 (Seq ID NO: 74)
Ion 221.00 (220.70 to 221.70): YST403 OKH29475.1 (Seq ID NO: 75)
Ion 221.00 (220.70 to 221.70): YST403 HFOL
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
Time-->
16.60 16.65 16.70 16.75 16.80 16.85 16.90 16.95 17.00 17.05 17.10 17.15

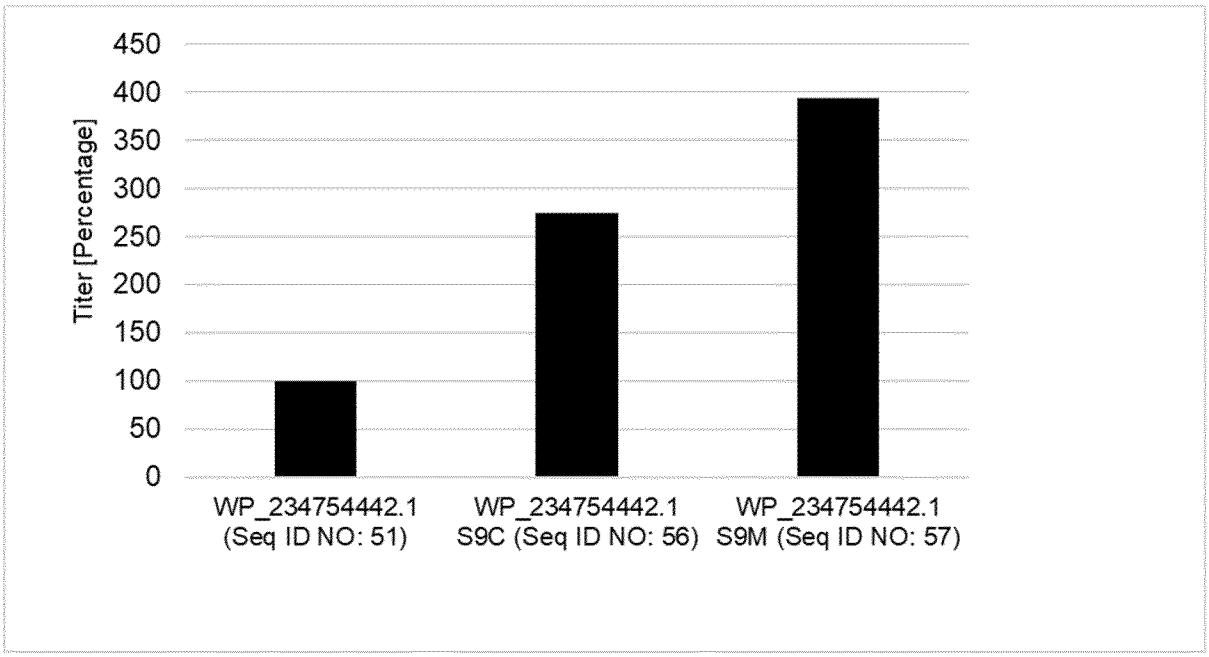
- Fig. 19 -



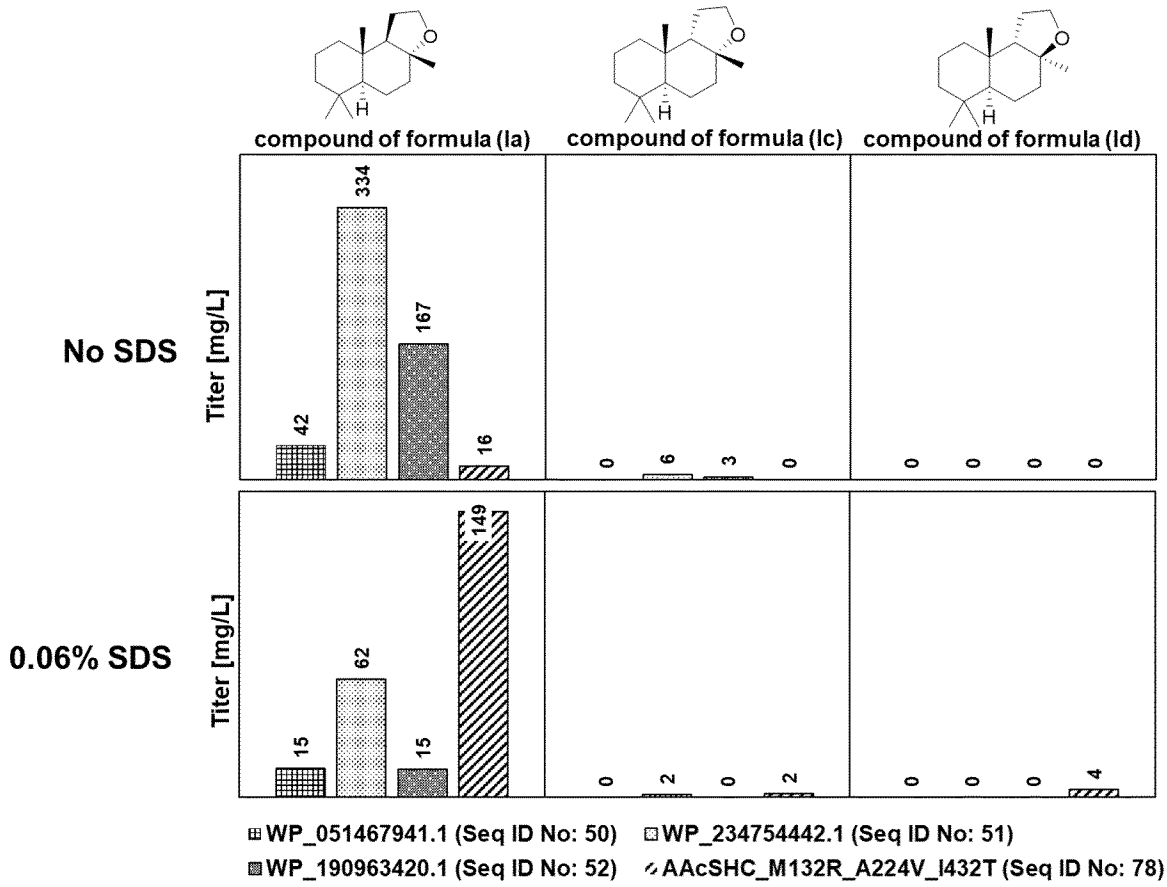
- Fig. 20 -


Titer [Percentage]
3000
2500
2000
1500
1000
500
0
WP_234754442.1 (Seq ID NO: 51)
WP_234754442.1 T7M (Seq ID NO: 62)
WP_234754442.1 S51T (Seq ID NO: 63)
WP_234754442.1 N63Q (Seq ID NO: 64)
WP_234754442.1 N63F (Seq ID NO: 65)
WP_234754442.1 N63Y (Seq ID NO: 66)
WP_234754442.1 N63A (Seq ID NO: 67)
WP_234754442.1 N63M (Seq ID NO: 68)
WP_234754442.1 S9T (Seq ID NO: 58)
WP_234754442.1 S9M (Seq ID NO: 57)
WP_234754442.1 S9C (Seq ID NO: 56)
WP_234754442.1 S9V (Seq ID NO: 69)
WP_234754442.1 S9A (Seq ID NO: 70)

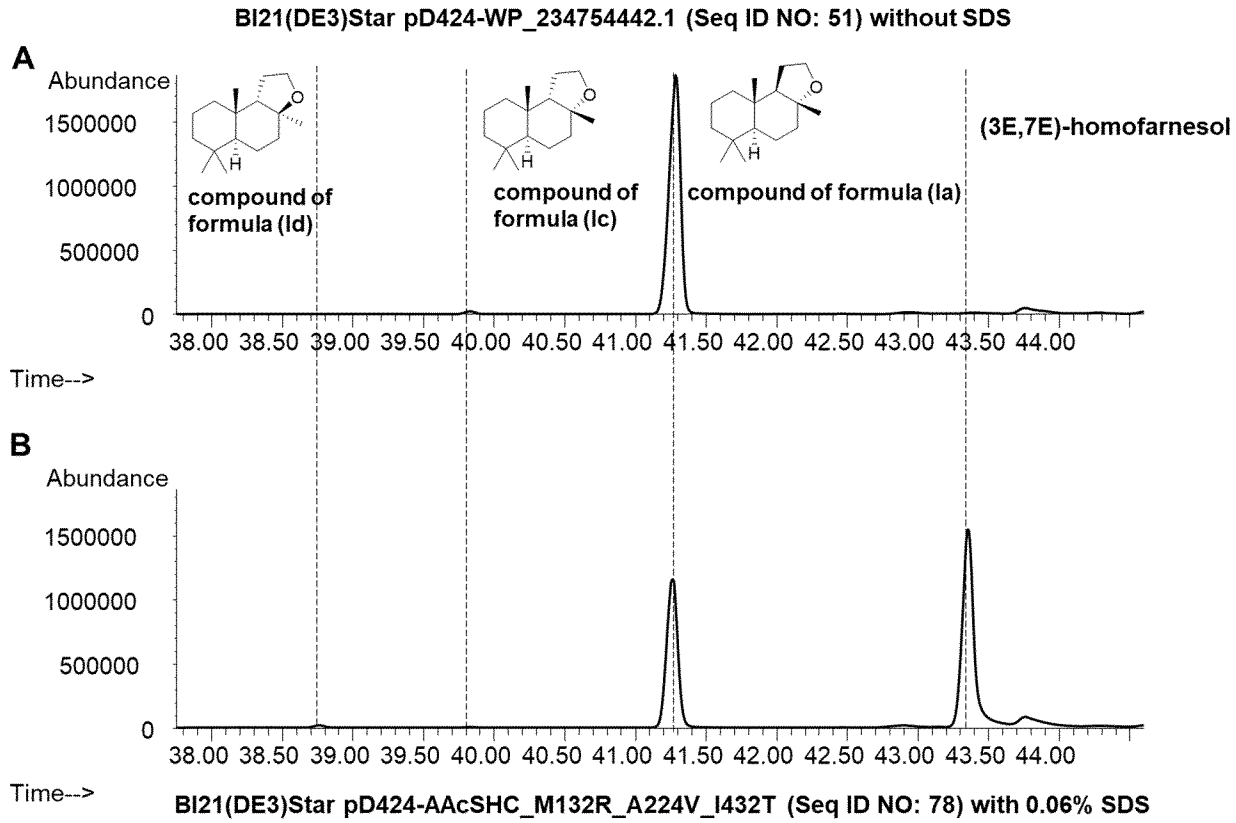
- Fig. 21 -


Titer [Percentage]
450
400
350
300
250
200
150
100
50
0
WP_234754442.1 (Seq ID NO: 51)
WP_234754442.1 S9C (Seq ID NO: 56)
WP_234754442.1 S9M (Seq ID NO: 57)

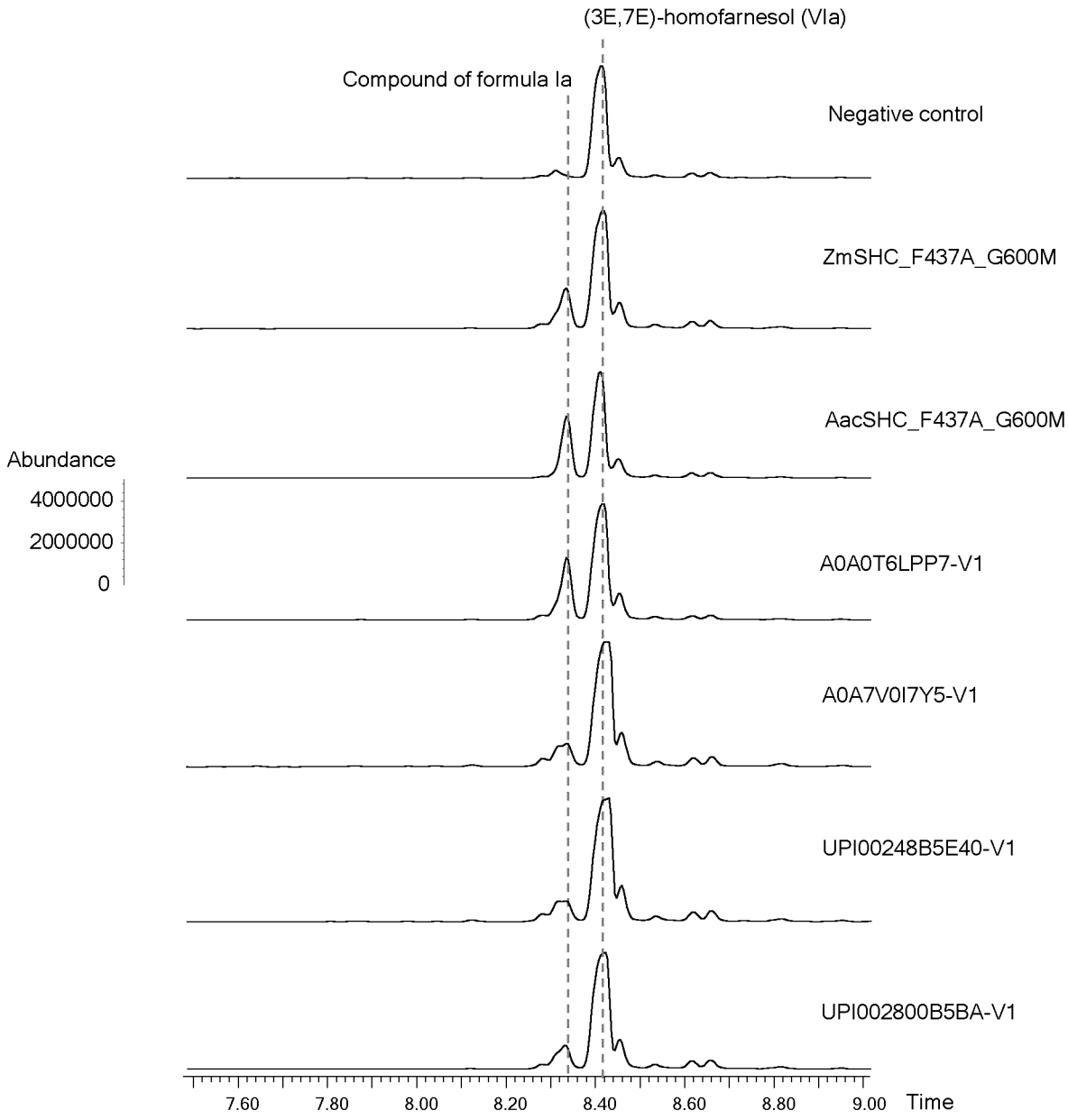
- Fig. 22 -


compound of formula (Ia)
compound of formula (Ic)
compound of formula (Id)
No SDS
0.06% SDS
Titer [mg/L]
42
334
167
16
0
6
3
0
0
0
0
0
15
62
15
149
0
2
0
2
0
0
0
4
WP_051467941.1 (Seq ID No: 50)
WP_234754442.1 (Seq ID No: 51)
WP_190963420.1 (Seq ID No: 52)
AAcSHC_M132R_A224V_I432T (Seq ID No: 78)

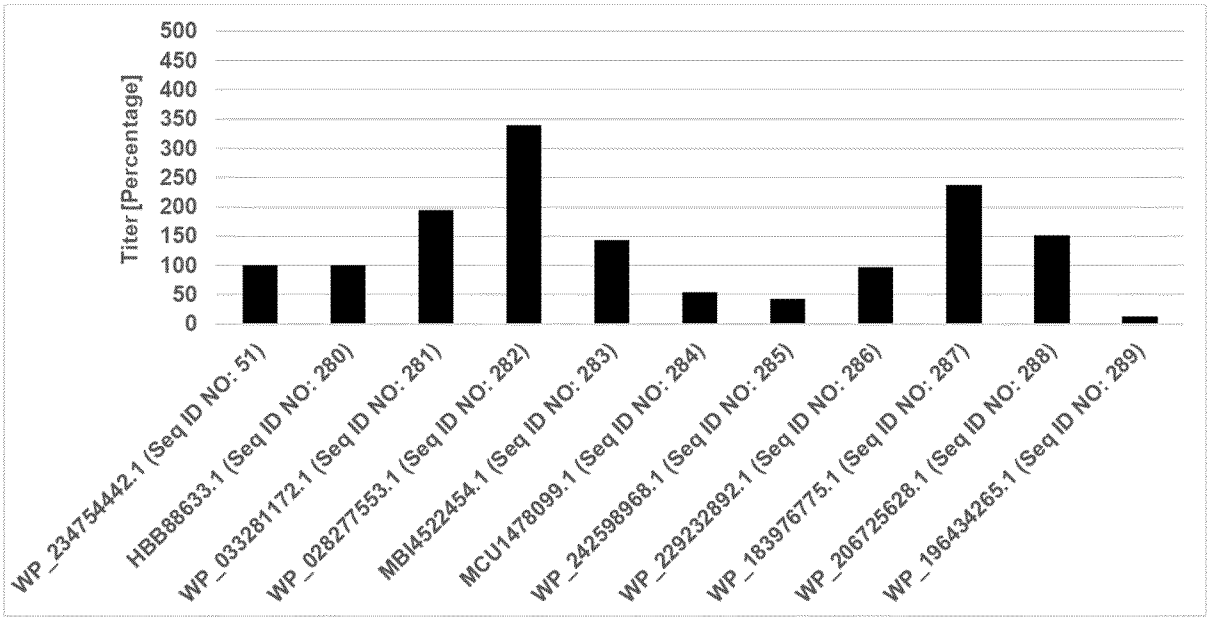
- Fig. 23 -


BI21(DE3)Star pD424-WP_234754442.1 (Seq ID NO: 51) without SDS
A
Abundance
1500000
1000000
500000
0
compound of formula (Id)
compound of formula (Ic)
compound of formula (Ia)
(3E,7E)-homofarnesol
38.00 38.50 39.00 39.50 40.00 40.50 41.00 41.50 42.00 42.50 43.00 43.50 44.00
Time-->
B
Abundance
1500000
1000000
500000
0
38.00 38.50 39.00 39.50 40.00 40.50 41.00 41.50 42.00 42.50 43.00 43.50 44.00
Time-->
BI21(DE3)Star pD424-AAcSHC_M132R_A224V_I432T (Seq ID NO: 78) with 0.06% SDS

- Fig. 24 -


(3E,7E)-homofarnesol (VIa)
Compound of formula Ia
Negative control
ZmSHC_F437A_G600M
AacSHC_F437A_G600M
A0A0T6LPP7-V1
A0A7V0I7Y5-V1
UPI00248B5E40-V1
UPI002800B5BA-V1
Abundance
4000000
2000000
0
7.60
7.80
8.00
8.20
8.40
8.60
8.80
9.00
Time

- Fig. 25 -


Titer [Percentage]
500
450
400
350
300
250
200
150
100
50
0
WP_234754442.1 (Seq ID NO: 51)
HBB88633.1 (Seq ID NO: 280)
WP_033281172.1 (Seq ID NO: 281)
WP_028277553.1 (Seq ID NO: 282)
MBI4522454.1 (Seq ID NO: 283)
MCU1478099.1 (Seq ID NO: 284)
WP_242598968.1 (Seq ID NO: 285)
WP_229232892.1 (Seq ID NO: 286)
WP_183976775.1 (Seq ID NO: 287)
WP_206725628.1 (Seq ID NO: 288)
WP_196434265.1 (Seq ID NO: 289)